

UNIVERSIDADE FEDERAL DO PARANÁ

ERNANI GOTTARDO

INFERÊNCIA DE ESTADOS AFETIVOS EM AMBIENTES
EDUCACIONAIS: PROPOSTA DE UM MODELO HÍBRIDO
BASEADO EM INFORMAÇÕES COGNITIVAS E FÍSICAS

CURITIBA PR

2018

ERNANI GOTTARDO

INFERÊNCIA DE ESTADOS AFETIVOS EM AMBIENTES
EDUCACIONAIS: PROPOSTA DE UM MODELO HÍBRIDO
BASEADO EM INFORMAÇÕES COGNITIVAS E FÍSICAS

Tese apresentada como requisito parcial à obtenção do grau de Doutor em Informática, no Programa de Pós-Graduação em Informática, setor de Ciências Exatas, da Universidade Federal do Paraná.

Área de concentração: *Ciência da Computação*.

Orientador: Prof. Dr. Andrey Ricardo Pimentel.

CURITIBA PR

2018

Catálogo na Fonte: Sistema de Bibliotecas, UFPR
Biblioteca de Ciência e Tecnologia

G685i

Gottardo, Ernani

Inferência de estados afetivos em ambientes educacionais: proposta de um modelo híbrido baseado em informações cognitivas e físicas / Ernani Gottardo. – Curitiba, 2018.

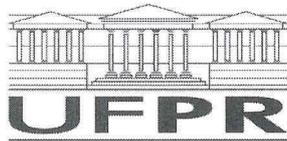
Tese - Universidade Federal do Paraná, Setor de Ciências Exatas, Programa de Pós-Graduação em Informática, 2018.

Orientador: Andrey Ricardo Pimentel.

1. Inferência (Lógica). 2. Emoções. 3. Emoções e cognição. 4. Aprendizagem cognitiva. 5. Computação. I. Universidade Federal do Paraná. II. Pimentel, Andrey Ricardo. III. Título.

CDD: 371.39

Bibliotecária: Vanusa Maciel CRB- 9/1928



TERMO DE APROVAÇÃO

Os membros da Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em INFORMÁTICA da Universidade Federal do Paraná foram convocados para realizar a arguição da tese de Doutorado de **ERNANI GOTTARDO** intitulada: **Inferência de estados afetivos em ambientes educacionais: proposta de um modelo híbrido baseado em informações cognitivas e físicas.**, após terem inquirido o aluno e realizado a avaliação do trabalho, são de parecer pela sua aprovação no rito de defesa.

A outorga do título de doutor está sujeita à homologação pelo colegiado, ao atendimento de todas as indicações e correções solicitadas pela banca e ao pleno atendimento das demandas regimentais do Programa de Pós-Graduação.

Curitiba, 11 de Dezembro de 2018.


ANDREY RICARDO PIMENTEL
Presidente da Banca Examinadora


ROBERTO PEREIRA
Avaliador Interno (UFPR)


EDUARDO TODT
Avaliador Interno (UFPR)


PATRICIA AUGUSTIN JAQUES MAILLARD
Avaliador Externo (UNISINOS)


ROBINSON VIDA NORONHA
Avaliador Externo (UTFPR)



A Deus por me conceder saúde e persistência e a minha família, esposa e filhas pelo apoio, incentivo e carinho.

AGRADECIMENTOS

Agradeço a Deus por ter me dado forças, saúde e persistência para conclusão do doutorado e me protegido nas inúmeras viagens de Erechim-RS a Curitiba-PR durante o Doutorado.

Agradeço a minha esposa Elenice Gottardo e minhas filhas Bruna Gottardo e Luiza Gottardo que me incentivaram e apoiaram durante as longas horas de estudo e viagens feitas durante o período do curso. Vocês me deram um suporte imprescindível para conseguir superar todas as dificuldades.

Sou muito grato ao meu orientador, Prof. Dr. Andrey Ricardo Pimentel pelo apoio, incentivo e direcionamento prestados sempre com muita empatia, prestesa e sabedoria.

Agradeço também aos professores Patrícia Jaques, Robinson Vida Noronha, Roberto Pereira e Eduardo Todt pela participação na banca de defesa da tese e pelos apontamentos realizados que contribuíram para a melhoria deste trabalho.

Sou grato ao Instituto Federal do Rio Grande do Sul - IFRS pela concessão de afastamento remunerado concedido nos dois últimos anos do curso. Esta concessão permitiu a realização de um trabalho de excelência, fato comprovado pelas publicações relevantes realizadas no âmbito desta pesquisa.

RESUMO

Na comunidade científica é comum o entendimento de que os softwares educacionais precisam evoluir para garantir um suporte mais efetivo ao processo de aprendizagem. Uma limitação recorrente destes softwares refere-se à falta de funcionalidades de adaptação às reações afetivas dos estudantes. Esta limitação torna-se relevante pois as emoções têm influência direta no processo de aprendizagem.

Reconhecer as emoções dos estudantes é o primeiro passo em direção a construção de software educativos sensíveis ao afeto. Trabalhos correlatos reportam relativo sucesso na tarefa de reconhecimento automático das emoções dos estudantes. No entanto, grande parte dos trabalhos correlatos utiliza sensores pouco práticos, intrusivos e caros que normalmente monitoram apenas reações físicas.

No contexto educacional o conjunto de emoções a ser considerado no processo de reconhecimento deve observar as singularidades deste domínio. Sendo assim, neste trabalho a inferência é realizada utilizando uma abordagem de quadrantes formadas pelas dimensões valência e ativação. Estes quadrantes representam situações relevantes para a aprendizagem e podem ser utilizados para embasar adaptações no ambiente computacional.

Diante disto, esta pesquisa apresenta a proposta de um modelo híbrido de inferência de emoções de estudantes durante o uso softwares educacionais. Este modelo tem como principal característica a utilização simultânea de informações oriundas de reações físicas (expressões faciais) e cognitivas (eventos no software educacional). Esta abordagem fundamenta-se na perspectiva teórica de que as emoções humanas são fortemente relacionadas com reações físicas, mas também são influenciadas por processos racionais ou cognitivos.

A combinação de expressões faciais e informações sobre os eventos do software educacional permite a construção de uma solução de baixo custo e intrusividade. Além disso, esta solução apresenta viabilidade de utilização em larga escala e em ambientes reais de ensino.

Experimentos realizados com estudantes em um ambiente real de ensino demonstraram a viabilidade desta proposta. Este fato é importante, considerando-se que a abordagem proposta neste trabalho é pouco explorada na comunidade científica e requer a fusão de informações bastante distintas. Nestes experimentos, foram obtidas acurácia e índice Cohen Kappa próximas de 66% e 0,55, respectivamente, na tarefa de inferência de cinco classes de emoções. Embora esses resultados sejam promissores quando comparados a trabalhos correlatos, entende-se que eles podem ser aprimorados no futuro, incorporando-se novos dados ao modelo proposto.

Palavras-chave: Inferência de Emoção, Emoção Relacionada à Aprendizagem, Tutoria Afetiva, Computação Afetiva.

ABSTRACT

In the scientific community there is a common understanding that educational software must evolve to ensure more effective support to the learning process. A common limitation of these software refers to the lack of adaptive features to students' affective reactions. This limitation becomes relevant because the emotions have a direct influence on the learning process.

Recognizing students' emotions is the first step toward building affect-sensitive educational software. Related work reports relatively successful in the task of automatically recognize students' emotions. However, most studies use impractical, intrusive and expensive sensors that typically monitor only physical reactions.

In the educational context the set of emotions to be considered in the recognition process must observe the singularities of this domain. Thus, in this work the inference is performed using a quadrant approach formed by the valence and activation dimensions. These quadrants represent situations relevant to learning and can be used to support adaptations in the computational environment.

So, this research presents a proposal of a hybrid model to infer emotions of students while using educational software. This model has as its main feature the simultaneous use of information coming from physical reactions (facial expressions) and cognitive (events in the educational software). This approach is based on the theoretical perspective that human emotions are strongly related with physical reactions, but are also influenced by rational or cognitive processes.

Combining facial expressions and information about the events of educational software allows the construction of a low-cost and intrusiveness solution. In addition, this solution presents feasibility for use in large scale in real learning environments.

Experiments with students in a real classroom demonstrated the feasibility of this proposal. This is important, considering that the approach proposed in this work is little explored in the scientific community and requires the fusion of quite different information. In these experiments, accuracy and Cohen Kappa index close to 66% and 0,55, respectively, were obtained in the inference of five emotion classes. Although these results are promising when compared to related works, it is understood that they can be improved in the future by incorporating new data into the proposed model.

Keywords: Emotion inference, Learning related emotion, Affective tutoring, Affective Computing.

Sumário

1	Introdução	16
1.1	Objetivos	19
1.2	Justificativa e relevância da pesquisa	20
1.3	Organização do documento	22
2	Revisão Bibliográfica	23
2.1	Computação Afetiva	23
2.2	Sistemas Tutores Afetivos	26
2.3	Aspectos conceituais sobre emoções	26
2.4	Inferências de emoções baseada em reações físicas	30
2.5	Inferência de emoções utilizando informações cognitivas	33
2.5.1	Modelo OCC	34
2.6	Possibilidades de utilização das inferências afetivas	36
2.7	Estado da Arte	38
2.8	Conclusão	46
3	Conjunto de Emoções Relacionadas à Aprendizagem	48
3.1	Dimensões das emoções	48
3.2	Quadrantes de emoções	50
3.3	Conjunto de emoções baseadas em reações físicas - expressões faciais	54
3.4	Conjunto de emoções baseadas em informações cognitivas	56
3.5	Conclusão	57
4	Modelo Híbrido de Inferência de Emoções	59
4.1	Arquitetura do Modelo Híbrido de Inferência de Emoções	59
4.1.1	Componente Cognitivo	60
4.1.2	Componente Físico	61
4.1.3	Componente Fusão	61
4.2	Inferência de emoções cognitivas	62
4.2.1	Modelo ALMA	63
4.3	Inferência de emoções físicas	65
4.3.1	Detecção da Face	65
4.3.2	Geração de atributos para inferência de emoções	66
4.3.3	Treinamento de classificadores para detecção de emoções	67
4.4	Fusão dos componentes cognitivo e físico	69
4.5	Exemplo ilustrativo do processo de inferência do ModHemo	71
4.6	Conclusão	73

5	Experimentos e Resultados	74
5.1	Metodologia	74
5.2	Materiais e procedimentos utilizados no experimento	77
5.2.1	Software Educativo	77
5.2.2	Rotulagem da base de dados dos experimentos	79
5.2.3	O Experimento	82
5.3	Resultados dos experimentos	85
5.3.1	Informações gerais dos experimentos	85
5.3.2	Resultados detalhados de um estudante específico	88
5.3.3	Dinâmica de transição e persistência das emoções	90
5.3.4	Descrição do indicadores utilizados para avaliar os resultados dos exper- imentos	92
5.3.5	Inferências do ModHEmo com a fusão dos componentes físico e cognitivo	96
5.3.6	Inferências do componente físico do ModHEmo	99
5.3.7	Inferências do componente cognitivo do ModHEmo	102
5.3.8	Inferências usando a dimensão valência	104
5.3.9	Inferências usando a dimensão ativação	107
5.4	Ameaças à validade dos resultados	109
5.5	Conclusão	110
6	Considerações Finais	114
6.1	Considerações gerais sobre a pesquisa	114
6.2	Perspectivas de trabalhos futuros	116
6.3	Contribuições científicas referentes a esta pesquisa	117
	Referências Bibliográficas	119

Lista de Figuras

2.1	Tela de um navegador web exibindo erro de carregamento da página	24
2.2	Diferentes emoções exibidas por um agente pedagógico animado. Fonte: Adaptado de Woolf et al. (2009)	25
2.3	Arquitetura de um Sistema Tutor Afetivo. Fonte: Adaptado de Sarrafzadeh et al. (2008)	27
2.4	Estudante sendo monitorado por múltiplos sensores enquanto interage com um ambiente educacional. Fonte: D’Mello et al. (2007)	33
2.5	Estrutura do modelo OCC. Fonte: Adaptado de Ortony et al. (1990).	35
3.1	Curva clássica da lei de Yerkes-Dodson	49
3.2	Modelo de aprendizagem em espiral. Fonte: Adaptado de Kort et al. (2001)	49
3.3	Mapeamento de um conjunto de emoções relacionadas à aprendizagem para o modelo dimensional valência x ativação: Fonte: Adaptado de Shen et al. (2007)	52
3.4	Mapeamento de um conjunto de emoções relacionadas à aprendizagem para o modelo dimensional valência x ativação. Fonte: Adaptado de Baker et al. (2010)	52
3.5	Abordagem de representação das emoções em quadrantes	53
3.6	Representação gráfica do modelo circumplexo de afeto. Fonte: Adaptado de Posner et al. (2005)	55
3.7	Mapeamento das sete emoções básicas para o espaço dimensional valência x ativação	55
3.8	Componentes do modelo OCC considerados relevantes para este trabalho	57
3.9	Relação entre emoções no modelo OCC e quadrantes das dimensões valência e ativação	58
4.1	Esquema resumido da organização do ModHEmo	59
4.2	Esquema detalhado do Modelo Híbrido de Inferência de Emoções - ModHEmo	60
4.3	Visão geral das etapas do processo de inferência de emoções via expressões faciais. Fonte: adaptado de Bartlett et al. (2006)	65
4.4	Resultado da aplicação da técnica de detecção de face em vídeo. Fonte: (Gottardo e Pimentel, 2016)	66
4.5	Exemplo de um banco de filtros Gabor com 4 escalas e oito orientações	67
4.6	Imagens do banco de dados Cohn-Kanade em que um participante simula três emoções básicas: a) medo b) alegria c) neutro Fonte: Kanade et al. (2000)	68
4.7	Quadros do vídeo demonstrando experimento com o autor em que as imagens foram classificadas corretamente como: a) neutro, b) surpresa e c) felicidade Fonte: Gottardo e Pimentel (2016)	68
4.8	Quadros do vídeo demonstrando experimento com o autor em que as imagens foram classificadas incorretamente como: a) desprezo, b) desgosto e c) raiva Fonte: Gottardo e Pimentel (2016)	69

4.9	Representação esquemática das diferentes abordagens de fusão de classificadores: fusão antecipada (a), fusão intermediária(b) e fusão tardia(c). Fonte: Adaptado de Schels et al. (2013)	70
4.10	Simulação de execução do ModHEmo	72
5.1	Formulário com questões respondidas pelos estudantes antes de iniciar o jogo .	76
5.2	Estudantes durante os experimentos	77
5.3	Exemplo da Interface do TuxMath.	78
5.4	Interface da ferramenta de rotulagem da base de dados no instante da ocorrência de um evento monitorado.	80
5.5	Representação Clássica da Ferramenta SAM: Fonte Bradley e Lang (1994) . . .	81
5.6	Proposta de Alterações ao SAM. Fonte: Hayashi et al. (2016)	81
5.7	Ferramenta de Avaliação Afetiva Utilizada nesta Pesquisa	82
5.8	Distribuição das classes obtidas no processo de rotulagem	88
5.9	Aderência entre a rotulagem e a inferência do ModHEmo para um estudante específico	89
5.10	Estudante com a mão ocultando parte da face	90
5.11	Grafo de sessão completa do estudante com as classes reais	91
5.12	Grafo de sessão completa do estudante com as classes inferidas pelo ModHEmo	92
5.13	Grafo dinâmico com as classes reais no instante “30/11/2017 09:20:08”	93
5.14	Grafo dinâmico das classes inferidas pelo ModHEmo no instante “30/11/2017 09:20:08”	93
5.15	Grafo dinâmico com as classes reais no instante “30/11/2017 09:22:37”	94
5.16	Grafo dinâmico das classes inferidas pelo ModHEmo no instante “30/11/2017 09:22:37”	94
5.17	Exemplos de Curvas ROC	96
5.18	Curvas ROC por classe do algoritmo RandomForest utilizando o conjunto de dados completo	97
5.19	Curvas ROC por classe do algoritmo IBK utilizando o conjunto de dados completo	98
5.20	Curvas ROC por classe do algoritmo RandomForest utilizando atributos do componente físico	101
5.21	Curvas ROC por classe do algoritmo IBK utilizando atributos do componente físico	101
5.22	Curvas ROC por classe do algoritmo RandomForest utilizando atributos do componente cognitivo	103
5.23	Curvas ROC por classe do algoritmo IBK utilizando atributos do componente cognitivo	104
5.24	Curvas ROC por classe do algoritmo RandomForest com 3 classes da dimensão valência	105
5.25	Curvas ROC por classe do algoritmo IBK com 3 classes da dimensão valência .	106
5.26	Curvas ROC por classe do algoritmo RandomForest com 3 classes da dimensão ativacao	108
5.27	Curvas ROC por classe do algoritmo IBK com 3 classes da dimensão ativacao .	109

Lista de Tabelas

2.1	Delimitação das características dos construtos psicológicos relativos a estados afetivos. Fonte: Adaptado de Scherer (2000)	28
2.2	Características que diferenciam as emoções medo e esperança/expectativa . Fonte: adaptado de Picard (1997)	29
2.3	Exemplos de mapeamento de emoções primárias (categóricas) para o modelo dimensional Fonte:Becker-Asano e Wachsmuth (2010).	30
2.4	Exemplos de reações físicas e o grau de aparência para outras pessoas Fonte:Picard (1997)	31
2.5	Exemplos de emoções e possíveis intervenções Fonte: D’Mello et al. (2007) . .	37
2.6	Estratégias adaptativas que combinam elementos afetivos e cognitivos. Fonte:Woolf et al. (2009).	37
2.7	Quantidade de artigos recuperados para avaliação	40
2.8	Trabalhos correlatos selecionados no processo de revisão sistemática da literatura	45
3.1	Conjunto de emoções tratadas em trabalhos correlatos	51
3.2	Mapeamento da Emoções do modelo OCC para o espaço dimensional. Fonte: Adaptado de (Gebhard, 2005)	57
4.1	Visão geral das condições de elicitação das emoções no modelo OCC. Fonte: Ortony et al. (1990)	62
4.2	Etiquetas de avaliação utilização na geração de emoções no modelo ALMA. Fonte : Gebhard et al. (2003)	64
4.3	Mapeamento das etiquetas de avaliação para emoções no modelo OCC Fonte:Gebhard (2005)	64
5.1	Eventos monitorados no TuxMath durante os experimentos	84
5.2	Seções de resultados e seus objetivos	85
5.3	Detalhamento da ocorrência de eventos monitorados nos experimentos	86
5.4	Detalhamento do conjunto de dados obtidos no processo de fusão dos componentes físico e cognitivo do ModHEmo	87
5.5	Matriz de confusão do algoritmo RandomForest usando conjunto de dados completo	96
5.6	Índices de desempenho por classe do algoritmo RandomForest utilizando o conjunto de dados completo	97
5.7	Matriz de confusão do algoritmo IBK usando conjunto de dados completo . . .	97
5.8	Índices de desempenho por classe do algoritmo IBK utilizando o conjunto de dados completo	98
5.9	Matriz de confusão do algoritmo RandomForest usando atributos do componente físico	100

5.10	Índices de desempenho por classe do algoritmo RandomForest utilizando os atributos do componente físico	100
5.11	Matriz de confusão do algoritmo IBK usando atributos do componente físico	100
5.12	Índices de desempenho por classe do algoritmo IBK utilizando os atributos do componente físico	100
5.13	Matriz de confusão do algoritmo RandomForest usando atributos do componente cognitivo	102
5.14	Índices de desempenho por classe do algoritmo RandomForest utilizando os atributos do componente cognitivo	102
5.15	Matriz de confusão do algoritmo IBK usando atributos do componente cognitivo	103
5.16	Índices de desempenho por classe do algoritmo IBK utilizando os atributos do componente cognitivo	103
5.17	Matriz de confusão do algoritmo RandomForest com 3 classes da dimensão valência	105
5.18	Índices de desempenho por classe do algoritmo RandomForest com 3 classes da dimensão valência	105
5.19	Matriz de confusão do algoritmo IBK com 3 classes da dimensão valência	106
5.20	Índices de desempenho por classe do algoritmo IBK com 3 classes da dimensão valência	106
5.21	Matriz de confusão do algoritmo RandomForest com 3 classes da dimensão ativação	107
5.22	Índices de desempenho por classe do algoritmo RandomForest com 3 classes da dimensão ativação	108
5.23	Matriz de confusão do algoritmo IBK com 3 classes da dimensão ativação	108
5.24	Índices de desempenho por classe do algoritmo IBK com 3 classes da dimensão ativação	109
5.25	Comparação dos índices de acurácia dos algoritmos e conjunto de dados	110
5.26	Comparação do índice Cohen Kappa dos algoritmos e conjunto de dados	111
5.27	Comparação da Área Sobre a Curva (AUC) média dos algoritmos e conjunto de dados	111
5.28	Comparação de resultados do ModHEmo com trabalhos correlatos.	113

Lista de Acrônimos

AdaBoost	Adaptive Boosting
ALMA	A Layered Model of Affect
STA	Sistema Tutor Afetivo
AU	Action Unit
AUC	Area Under the Curve
EaD	Educação à Distância
ECG	Eletrocardiograma
EDA	Atividade Eletrodermal - condutividade elétrica da pele
EEC	Condição de Elicitação de Emoção (Emotion Eliciting Condition)
EKG	Eletrocardiograma
EMG	Eletromiograma - atividade muscular e
FACS	Facial Action Coding System
FPS	Frames por segundo
MCS	Multiple Classifier System
OCC	Modelo Cognitivo de Ortony, Clore e Collons
OPENCV	Open Source Computer Vision Library
ROC	Receiver Operating Characteristic

Capítulo 1

Introdução

Ambientes educacionais têm usado historicamente computadores como ferramenta para apoio ou mesmo viabilização do processo de ensino-aprendizagem. No Brasil, observam-se incentivos à políticas públicas de apoio à Educação a Distância (EaD), tais como a Universidade Aberta do Brasil (UAB) e o Programa Nacional de Informática na Educação (PROINFO) (Jaques et al., 2012). Estes programas governamentais são exemplos de projetos educacionais que requerem o uso intensivo de tecnologia para sua viabilização.

Entretanto, questionamentos são recorrentes quanto à efetividade e possíveis contribuições dos ambientes computacionais no processo de ensino e aprendizagem. Uma das principais limitações dos softwares educacionais refere-se à falta de tratamento personalizado e adaptação às necessidades individuais de cada aprendiz (Khan et al., 2010; Hagemann et al., 2018).

Evidentemente, o ambiente computacional pode prover uma forma para o estudante fazer solicitações sob demanda para adaptação às suas necessidades. Todavia, são reportados melhores resultados em relação à usabilidade do ambiente (Becker-Asano e Wachsmuth, 2010) ou melhor experiência de aprendizagem (Baker et al., 2006; Hagemann et al., 2018) com a utilização de sistemas que pró ativamente adaptam-se aos seus usuários. Desta maneira, por exemplo, um ambiente de aprendizagem não deveria interromper aprendizes que estão avançando bem, mas poderia oferecer ajuda para outros que demonstram estar entediados (Baker et al., 2014).

Uma alternativa que busca superar algumas destas limitações são os Sistemas Tutores Inteligentes – STI. Um objetivo dos STI é prover o engajamento dos aprendizes em atividades de raciocínio sustentáveis e interagir com eles baseando-se no conhecimento sobre o comportamento dos estudantes (Corbett et al., 1997). Bouchet et al. (2018) completam que a adaptação é fundamental para o aprendizado bem-sucedido com STI. Contudo, mesmo com todos os avanços apresentados pelos STI, estudantes utilizando estes ambientes podem apresentar o comportamento *gaming the system*¹ (Baker et al., 2010) ou perceber o sistema como maçante, levando a situações de baixo aprendizado ou abandono de atividades (Baker et al., 2006).

Uma das limitações apresentadas atualmente pelos STI é a falta de funcionalidades para adaptação às reações emocionais dos estudantes (Alexander, 2008; D’Mello et al., 2010; Khan et al., 2010; Baker et al., 2012; Bursell, 2011). Atualmente os STI dão pouca ou nenhuma atenção às experiências emocionais dos estudantes, enfatizando atividades como o diagnóstico e correção de erros considerando apenas aspectos lógicos ou procedurais (Bosch et al., 2014).

Estas limitações são consideradas importantes dado que as emoções humanas são essenciais no processo de aprendizagem (Alexander, 2008; Harley et al., 2015; Baker et al., 2014). Conforme observam Frasson e Chalfoun (2010), Baker et al. (2010) e Damásio (2009) as funções cognitivas ligadas à aprendizagem e funções afetivas são intrinsicamente integradas no cérebro

¹Tentar avançar rapidamente nas atividades explorando propriedades ou limitações do sistema

humano. Como exemplo desta integração, DeFalco et al. (2018) destacam que intervenções baseadas em informações afetivas melhoram os resultados da aprendizagem. Os trabalhos de Baker et al. (2010) e Pardos et al. (2013) concluem que o tédio experimentado durante o processo de aprendizagem está negativamente correlacionado com ganhos de aprendizagem. Woolf et al. (2010) defendem que as intervenções afetivas estimulam o aprendizado e oferecem apoio e motivação que auxiliam os estudantes a superar os aspectos desagradáveis dos erros ou fracassos.

De acordo com Kort et al. (2001), bons professores são peritos em observar e reconhecer os estados emocionais dos estudantes e, baseando-se neste reconhecimento, tomar ações que impactam positivamente no aprendizado. Lajoie e Derry (2013) observam que aproximadamente metade das interações entre tutores humanos e aprendizes tem foco em aspectos relacionados a questões afetivas e de engajamento.

Desta forma, um professor, percebendo que os estudantes estão visivelmente confusos ou frustrados, poderia rever suas estratégias de ensino com o objetivo de atender às necessidades imediatas dos aprendizes, podendo resultar em melhorias na aprendizagem (Alexander, 2008). Em contrapartida, oferecer ajuda que interrompa a concentração ou engajamento de um estudante pode ser prejudicial (Baker et al., 2010), indo de encontro à estratégia que é chamada de 'intervenção mínima' por Vanlehn (2006).

Isto posto, Picard (1997) salienta que ambientes computacionais incapazes de reconhecer emoções ficam severamente restritos, especialmente em tarefas como aprendizagem ou tutoria. Assim como bons professores em um ambiente tradicional, sistemas computacionais eficazes de apoio às atividades de ensino precisam ser capazes de identificar as reações emocionais dos aprendizes e adaptar as atividades considerando estas informações. Se isto for alcançável, estes ambientes poderiam evoluir de artefatos lógico-analíticos, incorporando presença motivacional e até mesmo psicológica (Picard et al., 2004).

Nos últimos anos, conforme observam Grafsgaard et al. (2013b) e Bouchet et al. (2018), a investigação do impacto no processo de aprendizado de fatores como motivação e emoções dos estudantes têm emergido como uma linha de pesquisa promissora. Mesmo assim, pesquisas como (Kort et al., 2001; D'Mello, 2012; Wixon et al., 2014) apontam que estamos ainda em fase inicial de entendimento da complexa relação entre afeto e aprendizagem sendo este tema ainda um desafio em pesquisas que envolvem STI ou softwares educacionais adaptativos.

Mesmo com a existência pesquisas correlatas enfocando o processo automático de inferência de emoções em ambientes computacionais de ensino, este processo ainda apresenta-se como um desafio (D'Mello et al., 2010; Baker et al., 2012; Grafsgaard et al., 2013b; Reis et al., 2018b). Nesta linha, Picard et al. (2004) observam que facilmente pode-se medir a habilidade de alguém lembrar-se de conteúdos ensinados, contudo, é mais difícil medir as emoções de alguém durante este processo.

As principais barreiras para o avanço de ambientes computacionais adaptáveis às emoções dos usuários são: i) definições conceituais relacionadas às emoções, ii) operacionalização baseada nas definições conceituais e iii) mapeamento dos sinais em emoções (Van Den Broek, 2012).

DeFalco et al. (2018) notam que a criação de repostas afetivas de sucesso requer um considerável esforço para detectar o afeto do estudante e implementar respostas afetivas apropriadas. Neste sentido, é importante destacar que reconhecer as emoções dos estudantes representa a etapa inicial no processo de criação de ambientes computacionais de ensino adaptáveis. Todavia, para que um ambiente de ensino informatizado possa prover qualquer ação automatizada de adaptação ou resposta às reações emocionais dos aprendizes é necessário primeiramente que estas emoções sejam reconhecidas corretamente. Conforme observam Baker et al. (2010), a inferência das reações emocionais dos estudantes é uma tarefa fundamental, pois

o correto entendimento dos estados cognitivo-afetivos serve como base para o desenvolvimento de ambientes de ensino sensíveis aos estados afetivos *affect-sensitive*.

O conhecimento sobre as emoções dos aprendizes poderia ser utilizado, por exemplo, pelo módulo tutor de um STI para adaptar sua estratégia de intervenção. As intervenções poderiam ser com o objetivo de evitar o que D’Mello et al. (2007) denominam ‘ciclo vicioso’. O ‘ciclo vicioso’ ocorre quando um ou mais estados cognitivo-afetivos negativos sucedem-se repetidamente. Por outro lado, intervenções não devem ser feitas caso o estudante apresente a chamada "Concentração Engajada" (Baker et al., 2010), que é um estado afetivo-cognitivo que envolve atenção e foco na tarefa. Adicionalmente, o tédio pode ser tratado sugerindo-se que o estudante tente um novo problema ou ajustando-se a dificuldade (Bosch et al., 2015).

É importante destacar que grande parte os sensores fisiológicos utilizados em pesquisas correlatas para obtenção de informações que subsidiem o processo de inferência tem como inconvenientes sua intrusividade, custo e limitada disponibilidade destes sensores em condições reais de aprendizagem. Neste sentido, Wixon et al. (2014) notam que, no melhor dos casos, os computadores ou laptops de estudantes são equipados com microfone e câmeras de vídeo, mas não com sensores fisiológicos. Diante disto, a presente proposta baseia-se em estratégias de baixo custo, não invasiva (logs de eventos do sistema) ou minimamente invasivas (expressões faciais).

Considerando o contexto exposto até aqui, o presente trabalho considera como base para seu desenvolvimento o seguinte Problema de Pesquisa:

Como conceber e implementar um método multimodal de inferência de emoções de estudantes que agregue melhorias na acurácia dos métodos existentes com baixa intrusividade e custo, sendo viável de utilização em ambientes reais de ensino?

Um aspecto importante relativo ao processo de inferência de emoções diz respeito à inter relação e complementaridade dos aspectos cognitivos e físicos envolvidos no processo de geração de emoções nos seres humanos. Neste sentido, (Picard, 1997) lembra que é comum utilizarmos os termos “cabeça” e “coração” como referência à “pensamento/lógica” e “emoção”, respectivamente. Estes componentes são usualmente polarizados, como se fossem fenômenos opostos ou não relacionados. Porém, a referida autora observa que pesquisas da área de neurociência concordam largamente que no cérebro humano o componente cognitivo (e.g. pensamento) e físico (e.g. emoções) são altamente conectados e influenciam diretamente um ao outro.

Mesmo assim, pesquisas correlatas comumente utilizam separadamente informações físicas (Shen et al., 2007; Woolf et al., 2009; Baker et al., 2010; Bosch et al., 2014; Jaques et al., 2014) ou cognitivas (Jaques, 2004; Paquette et al., 2014; Muñoz et al., 2010; Jaques et al., 2011). Diante disto, o presente trabalho tem como principal diferencial a integração entre informações físicas e cognitivas, buscando o aprimoramento do processo de inferência de emoções em ambientes educacionais.

A relevância de considerar os aspectos cognitivos/contextuais juntamente com as reações físicas é ilustrada pelos três exemplos seguintes:

1. As lágrimas podem ser reconhecidas a partir de um vídeo ou imagem da face, mas não necessariamente correspondem à tristeza e podem também representar alegria (Picard, 1997).
2. As emoções com negativas tendem a aumentar a frequência cardíaca, mas apenas a frequência cardíaca fornece pouca informação sobre emoções específicas (Conati, 2002).

3. Pesquisas mostram que emoções como frustração ou aborrecimento não são claramente distinguíveis do estado neutro usando apenas expressões faciais (McDaniel et al., 2007; Gottardo e Pimentel, 2016).

Tomando-se como base a premissa de que as emoções nos seres humanos são influenciadas por aspectos físicos e cognitivos, definiu-se a seguinte Hipótese de Pesquisa:

A combinação de informações físicas e cognitivas é uma abordagem que poderia trazer melhorias na acurácia do processo de inferência de emoções de estudantes.

Considerando-se a hipótese descrita acima, o processo de inferência proposto neste trabalho foi fortemente embasado na concepção e implementação de um Modelo Híbrido de Inferência de Emoções, chamado de ModHEmo. Neste modelo, as informações físicas são obtidas usando as expressões faciais dos estudantes e as informações cognitivas tem como base os eventos ocorridos no ambiente computacional. Desta forma, a abordagem híbrida a ser utilizada na implementação do ModHEmo permite que sejam examinadas emoções de diferentes modalidades. Abordagens multimodais (usando mais de uma modalidade para medir emoções) estão alinhadas com teorias que definem as emoções como multi-componenciais (Harley et al., 2015).

Para a realização das inferências emocionais do ModHEmo, foram utilizadas estratégias de coleta de dados que observam questões de custo, intrusividade e viabilidade de uso fora de laboratórios específicos. Com base nestas inferências foi possível construir representações da dinâmica ou trajetória das emoções dos estudantes durante um processo de aprendizagem. Considerando-se esta dinâmica das emoções, estratégias de adaptações ou intervenções tutoriais podem ser implementadas em um software educacional, buscando melhorar a experiência de aprendizagem dos estudantes.

Dado o contexto exposto acima, os objetivos do presente trabalho serão definidos no próximo item. Na sequência serão apresentadas as justificativas e relevância da presente pesquisa. Finalizando este capítulo, a organização do restante deste documento será descrita.

1.1 Objetivos

Esta pesquisa tem como **Objetivo Geral** a concepção, desenvolvimento e avaliação de um proposta para inferência de emoções de estudantes representadas por quadrantes formados pelas dimensões valência e ativação utilizando uma abordagem multimodal (cognitiva-física).

Como estratégia para atingir o Objetivo Geral exposto acima, os seguintes **Objetivos Específicos** são considerados relevantes:

1. Avaliar as principais definições teóricas relacionadas com abordagens multimodais de inferência de emoções para escolha da abordagem conceitual mais adequada.
2. Escolher um conjunto de emoções relacionadas à aprendizagem a ser tratado no processo de inferência que seja consistente e adequado ao domínio educacional.
3. Investigar as opções de técnicas computacionais para a inferência das emoções cognitivas e físicas e a fusão destas.
4. Apresentar e avaliar criticamente os resultados obtidos nos experimentos com estudantes e discutir possíveis aplicações dos resultados alcançados.

1.2 Justificativa e relevância da pesquisa

Mesmo considerando a importância das emoções na aprendizagem, a análise dos trabalhos correlatos demonstra que não existe ainda uma abordagem consolidada de como inferir as emoções relevantes em um ambiente educacional. Tampouco existe um método estabelecido de como consolidar as informações físicas e cognitivas como pretendidas nesta proposta.

É fundamental reforçar que o processo de reconhecimento de emoções em aplicações computacionais, mesmo com os avanços e trabalhos sendo desenvolvidos nesta área, ainda carece de evolução e aprimoramentos (Baker et al., 2012). Nesta linha, Baker et al. (2010) e Wixon et al. (2014) observam que, apesar dos avanços, os métodos existentes para reconhecimento de emoções não são ainda suficientemente desenvolvidos para uso generalizado em ambientes reais.

Schels et al. (2013) destacam a dificuldade em estabelecer controle no registro de dados reais ("*in-the-wild*") e a grande quantidade de modalidades de dados que tornam o processo de reconhecimento de emoções uma tarefa desafiadora. Dentre os principais desafios, os autores destacam:

- Não restrição de comportamento dos sujeitos envolvidos.
- Emoções que ocorrem raramente e de difícil distinção.
- Dificuldade de criação de base de dados rotuladas, pois, mesmo observadores humanos, tendem a apresentar discordância no processo de classificação de emoções.

Em pesquisas correlatas, como por exemplo (Grafsgaard et al., 2013b; D’Mello et al., 2007; Pham e Wang, 2018; Xiao et al., 2017), os dados para inferência de emoções de aprendizes são obtidos fundamentalmente via dispositivos externos intrusivos, de alto custo ou inviáveis de utilização em larga escala. Dentre estes sensores, pode-se citar, sensores fisiológicos de batimentos cardíacos, impulsos elétricos do cérebro, condutividade de pele, cadeiras com sensor de pressão, entre outros. Conforme observam Baker et al. (2012) e Wixon et al. (2014) existem desafios práticos e econômicos para a implantação destes tipos de sensores afetivos, reduzindo a possibilidade de implementação em ambientes educacionais abertos e de larga escala.

Técnicas de inferência de emoções utilizando apenas dados de expressões faciais (reações físicas) apresentam muitas limitações para diferenciar emoções menos destacadas (e.g. desgosto) do estado neutro (Gottardo e Pimentel, 2016), assim como frustração e aborrecimento estado neutro (McDaniel et al., 2007; Woolf et al., 2009). Por conseguinte, em muitas situações, mudanças na expressão faciais podem ser sutis para serem detectadas, especialmente se a detecção é feita por um computador (Conati, 2002). Da mesma forma, alta variabilidade (Kappa entre 0,056 e 0,469) no desempenho preditivo entre as diferentes emoções foi obtido ao utilizar-se apenas dados cognitivos ou contextuais (logs de uso e interação) em um ambiente educacional (Paquette et al., 2016).

Trabalhos como Gottardo e Pimentel (2016), Woolf et al. (2009), Bosch et al. (2014) e McDaniel et al. (2007) investigam métodos para automatizar o processo de inferência de emoções utilizando dados provenientes de expressões faciais. Entretanto, é importante observar que reconhecer uma expressão facial não é o mesmo que reconhecer a emoção que a gerou (Picard, 1997). Como exemplo, a autora cita duas situações:

1. Não se pode garantir que uma expressão facial reconhecida como “triste” corresponde a qualquer emoção genuína de tristeza.
2. Métodos de baseados em expressões faciais falharão no reconhecimento de um estado de alegria se a pessoa alegre suprimir todas as expressões faciais relacionadas.

Estas limitações, conforme observa Picard (1997), têm forte relação com o fato de que as técnicas de inferência via expressões faciais não levam em consideração a semântica da situação que ocasionou ou contribuiu para a geração de uma resposta emocional. Conforme observam Nicolaou et al. (2011) nas suas interações diárias as pessoas reiteradamente exibem um vasto, complexo e sutil conjunto de estados afetivos ou mentais, como depressão, constrangimento, vergonha, entre outros. Estas reações afetivas são de difícil inferência ou observação utilizando apenas sensores de sinais físicos e sem a integração de informações contextuais ou cognitivas (Nicolaou et al., 2011).

Além disto, dispositivos de coleta de dados podem interferir nas emoções que o usuário está sentindo (Jaques et al., 2012). Estes dispositivos podem ser viáveis em situações que o usuário está habituado com a presença de sensores, como em um laboratório específico para este fim. Entretanto, não são satisfatórios em cenários aonde sensores externos podem distrair os usuários e até mesmo interferir nas tarefas principais (D’Mello et al., 2008).

Nicolaou et al. (2011) e Paquette et al. (2016) observam que existe um número limitado de pesquisas que investigam como diferentes modalidades de dados podem ser combinados e complementados entre si, envolvendo diferentes contextos, domínios e populações. Neste sentido, Reis et al. (2018b) defendem que a integração das informações de múltiplos sensores representa atualmente desafios e oportunidades para do desenvolvimento de futuras investigações.

Pesquisas como (Woolf et al., 2009), Glodek et al. (2013), Grafsgaard et al. (2013b), Schels et al. (2013) e Paquette et al. (2016) têm investigado os resultados obtidos com a coleta e integração de dados de um conjunto de sensores. De uma maneira geral, estas pesquisas apontam que os resultados obtidos com conjunto de sensores normalmente têm resultados superiores quando comparado aos resultados de sensores individuais.

No entanto, uma das principais limitações observadas nos experimentos que fazem uso de um conjunto de sensores é que eles usualmente incluem apenas dispositivos designados para capturar sinais fisiológicos. Sendo assim, mesmo utilizando-se um conjunto sensores, pode-se estar capturando apenas aspectos de reações físicas, ignorando-se completamente o contexto cognitivo ou racional em que estas reações aconteceram.

Picard (1997) apresenta evidências obtidas através de pesquisas da área de neurociência e psicologia demonstrando que os seres humanos respondem a certos eventos ou estímulos primeiramente através das emoções primárias, que são fortemente relacionadas à reações físicas. Porém, o processo de geração e controle das emoções envolve também um desenvolvimento individual, racional, cognitivo, consciente, dando origem às emoções chamadas de secundárias.

Desta forma, o aspecto racional ou cognitivo torna-se um fator importante a ser tratado no processo de inferência de emoções, considerando-se que existe uma intrínseca relação entre aspectos físicos e cognitivos observados no processo natural de geração das emoções nos seres humanos (Ortony et al., 1990). Mesmo com muitas discussões ainda em andamento na comunidade científica, parece haver concordância da intrínseca relação dos aspectos cognitivos/racionais e físicos, considerando-se o fato de que no cérebro humano não há uma separação explícita entre estes componentes (Picard, 1997). Adicionalmente, Damásio (2009) observa que componentes responsáveis em tratar reações físicas e racionais a eventos compartilham mecanismos cerebrais (não podem ser vistos como completamente separados).

Desta maneira, poder-se-ia esperar que um processo eficiente e natural de inferência de emoções via dispositivos computacionais deva considerar em conjunto estes dois aspectos. Entendeu-se então que uma abordagem híbrida teria o potencial de apresentar resultados promissores no processo de reconhecimento de emoções. Ao mesmo tempo, esta abordagem evitaria a utilização de métodos de coleta de dados de difícil aplicabilidade fora de laboratórios ou excessivamente intrusivos.

1.3 Organização do documento

Após a apresentação do Capítulo 1 com uma introdução ao trabalho, a sequência deste documento está estruturada em mais cinco capítulos.

No Capítulo 2 é apresentada uma revisão bibliográfica envolvendo trabalhos correlatos e definições conceituais diretamente ligados ao assunto principal desta pesquisa.

O Capítulo 3 é dedicado à discussão e detalhamento do processo utilizado para escolha do conjunto de emoções que farão parte do modelo de híbrido de inferência.

A proposta de um modelo híbrido de inferência é apresentada e detalhada no Capítulo 4, juntamente com uma descrição das técnicas computacionais utilizadas na implementação computacional do modelo.

No Capítulo 5 é apresentada a metodologia utilizada nos experimentos realizados. Neste capítulo também são apresentados os resultados obtidos com experimentos, bem como uma discussão dos principais resultados.

Finalizando o documento, o Capítulo 6 é dedicado às considerações finais, limitações e possibilidades de trabalhos futuros.

Capítulo 2

Revisão Bibliográfica

Neste capítulo serão detalhados os principais conceitos e técnicas computacionais que apresentam relação direta com o trabalho a ser desenvolvido. Será apresentada também uma discussão relativa a possibilidade de utilização dos resultados desta pesquisa. Além disto, trabalhos correlatos que representam o estado da arte no contexto da presente proposta serão apresentados, juntamente com uma discussão sobre os principais resultados e lacunas existentes nestes trabalhos.

2.1 Computação Afetiva

Nos últimos anos observa-se o desenvolvimento de pesquisas que aplicam conceitos conhecidos como Computação Afetiva (em inglês *Affective Computing*) em ambientes de software educacionais. O termo Computação Afetiva foi introduzido por Picard (1997) que o define como “computação que está relacionada com, que surge de ou deliberadamente influencia emoções”(p. 3).

Para Van Den Broek (2012), Computação Afetiva é a compreensão científica dos mecanismos subjacentes ao afeto e a incorporação em máquinas. A Computação Afetiva é uma área multidisciplinar que busca desenvolver métodos para dotar ambientes computacionais de várias características sociais, tais como reconhecer emoções e responder apropriadamente a elas, expressar emoções, mostrar empatia, entre outros (Jaques et al., 2012).

Calvo e D’Mello (2010) observam que automaticamente reconhecer e responder a estados afetivos de usuários durante a interação com um computador pode melhorar a qualidade da interação, tornando a interface computacional mais usável, agradável e afetiva. Além da aplicação em ambientes educacionais, que é o foco deste trabalho, técnicas de computação afetiva têm sido utilizadas em área como entretenimento, marketing, medicina, jogos, interação humano-computador, entre outros.

É importante destacar que, considerando as características da computação afetiva apresentadas acima, diferentes categorias de ambientes computacionais podem ser obtidas. Determinados ambientes podem ser capazes apenas de reconhecer emoções e realizar algum tipo de processamento ou adaptação. Por outro lado, outros ambientes podem ser capazes de expressar emoções, utilizando, por exemplo, agentes animados. Considerando as diferentes características de ambientes computacionais afetivos, Picard (1997) classifica estes ambientes nas quatro categorias seguintes:

- I Nenhuma funcionalidade que considere estados afetivos (a maior parte dos ambientes enquadram-se nesta categoria).

- II Ambientes que procuram expressar afetividade, sem ter capacidade de reconhecer estados afetivos, em resposta a uma situação pré-determinada (e.g. browser que não consegue carregar uma página web).
- III Ambientes que integram funções de reconhecimento de estados afetivos e que buscam adaptar-se aos estados afetivos inferidos (nesta categoria poderiam se enquadrar sistema tutores que, sem expressar emoções, realizam adaptações às emoções dos aprendizes).
- IV Ambientes capazes de reconhecer emoções e também expressar emoções (e.g. agente pedagógico emocional animado).

Cabe ressaltar, que a proposta deste trabalho enquadra-se na categoria III da lista apresentada acima, tendo em vista que o foco é no reconhecimento das emoções e que não está no escopo do trabalho a expressão de emoções. Contudo, serão discutidas neste trabalho alternativas de adaptação do ambiente computacional de aprendizagem em resposta às reações afetivas dos aprendizes.



Figura 2.1: Tela de um navegador web exibindo erro de carregamento da página

Na Figura 2.1 é mostrado um exemplo de um navegador web que procura expressar sinais faciais de tristeza ou aborrecimento, utilizando uma imagem e um texto, ao não conseguir carregar uma página. Este tipo de ambiente apenas procura expressar emoção, mesmo sem ser capaz de reconhecer as reações afetivas do usuário, enquadrando-se, portanto, na categoria II descrita acima.

Por outro lado, na Figura 2.2 é apresentado um exemplo extraído do trabalho de Woolf et al. (2009) mostrando diferentes emoções expressadas por um agente pedagógico animado. Estas diferentes expressões são geradas como resposta às emoções do aprendiz que o ambiente é capaz de reconhecer.

Considerando o domínio educacional, Morais et al. (2017) observam que para que um ambiente computacional de aprendizagem possa adaptar suas estratégias de ensino aos estados afetivos do estudante é necessário que este possua, no mínimo, as seguintes funcionalidades:

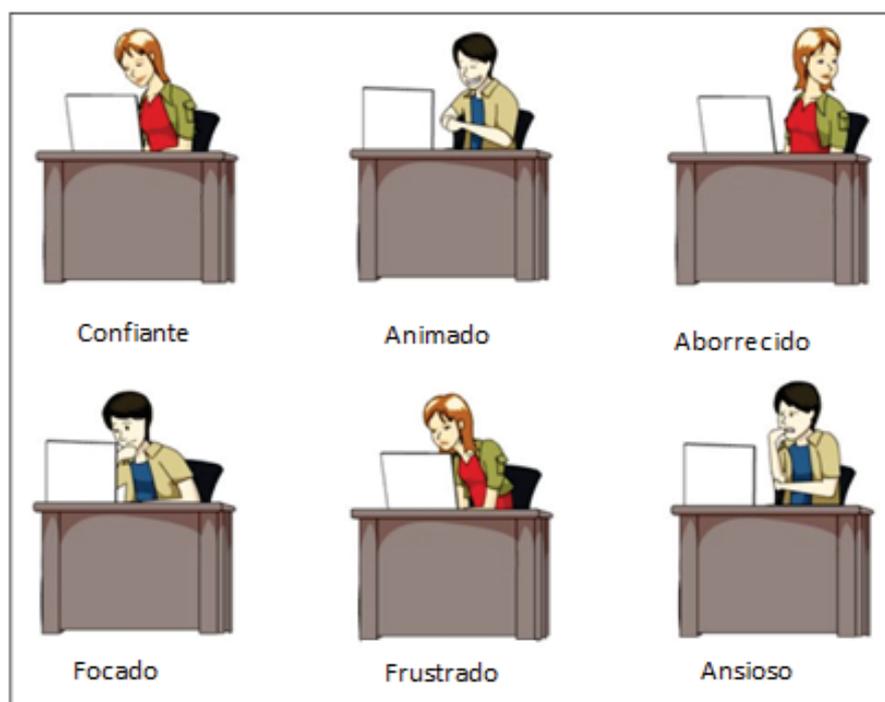


Figura 2.2: Diferentes emoções exibidas por um agente pedagógico animado. Fonte: Adaptado de Woolf et al. (2009)

1. Detecção dos estados afetivos dos estudantes.
2. Adaptação das suas estratégias cognitivas e afetivas de acordo com os estados afetivos do estudante.

A construção de computadores que reconhecem a emoção de um usuário tem se mostrado um desafio (Scherer, 2005; Paquette et al., 2016; D’Mello et al., 2008). A dificuldade principal é devido ao alto nível de ambiguidade existente no processo de mapeamento entre as emoções e os dispositivos que podem ser usados para detectá-las. Neste sentido, cabe destacar que mesmo seres humanos podem apresentar dificuldades e ambiguidades no reconhecimento de emoções (Conati, 2002).

Como exemplo de um desafio prático encontrado no processo de reconhecimento de emoções, D’Mello et al. (2008) observam que por pressões sociais, emoções negativas como frustração tendem a ser disfarçadas, dificultando a correta identificação destas emoções. Adicionalmente, certas emoções apresentam pouca diferenciação entre si (e.g. aborrecimento e frustração), tornando difícil sua correta distinção (Gottardo e Pimentel, 2016; McDaniel et al., 2007).

Picard (1997) propõe uma forma de representar uma arquitetura de computação afetiva, utilizando para isto três níveis de representação: baixo nível para sinais, nível médio para padrões e alto nível para conceitos. Esta proposta toma como base as três etapas que naturalmente são utilizadas quando uma pessoa tenta reconhecer as emoções de outra:

1. Sinais de baixo nível que carregam informações são detectados (e.g. expressões faciais, respiração, voz).
2. Padrões dos sinais (nível médio) podem ser combinados para prover um reconhecimento mais confiável (e.g. movimentos corporais, gestos).

3. Finalmente, sinais não expressivos do ambiente (alto nível) são levados em consideração para o raciocínio de alto nível ou cognitivo, relacionando que tipo de comportamento é comum nestas situações (e.g. pessoa em um escritório, um estudante em um período de exames finais).

Nesta arquitetura proposta por Picard (1997), o fluxo de informações ocorre nos dois sentidos entre baixo e alto nível, conforme ilustrado pela referida autora no dois exemplos a seguir. Um exemplo desta bidirecionalidade pode ocorrer quando pelo contexto de alto nível espera-se que a pessoa esteja mau humorada (e.g. estudante reprovado em uma disciplina). Neste caso, percepções de baixo nível podem ser distorcidas, percebendo expressões ambíguas como negativas. Outro exemplo é quando um ator cognitivamente representa um personagem triste.

2.2 Sistemas Tutores Afetivos

Os conceitos da Computação Afetiva serviram de base para o desenvolvimento de pesquisas em uma área que foi chamada de Sistemas Tutores Afetivos (STA) (Sarrafzadeh et al., 2008). A área de pesquisa em tutores afetivos investiga técnicas para permitir que os computadores reconheçam, modelem, compreendam e respondam efetivamente à emoção dos estudantes (Woolf et al., 2010).

Além de incluir todas as características dos ambientes STI, Sistemas Tutores Afetivos agregam funcionalidades de adaptação aos estados afetivos dos estudantes. Isto representa um passo adiante no desenvolvimento de sistemas tutores totalmente conscientes dos estados cognitivos e afetivos dos estudantes e perfeitamente capazes de se adaptar a estes estados.

Sistemas Tutores Afetivos integram aspectos cognitivos e afetivos dos estudantes em suas estratégias pedagógicas e motivacionais. Com isto, busca-se manter os aprendizes engajados, aumentar sua auto confiança e interesse e desta forma, presumidamente, facilitar a aprendizagem (D’Mello et al., 2010).

A arquitetura de um STA é composta por quatro componentes principais: Modelo de Estudante, Estratégias de Tutoria, Domínio de Conhecimento e Módulo Tutor. A Figura 2.3 apresenta uma proposta simplificada de arquitetura de um STA com seus quatro componentes principais. A principal diferença em relação a um STI convencional está no Modelo de Estudante, que inclui informações sobre o estado afetivo do estudante.

O processo para implementação de Sistemas Tutores Afetivos requer a modelagem de um ciclo conhecido como “affective loop” (D’Mello et al., 2008). Este ciclo envolve dois importantes desafios: i) inferir de forma confiável o estado afetivo dos estudantes e ii) conhecendo o estado afetivo do estudante, como adaptar da melhor maneira o sistema em resposta aos estados afetivos detectados.

Destaca-se que o objetivo principal deste trabalho relaciona-se com o primeiro desafio apresentado anteriormente. No entanto, será também apresentada uma discussão sobre a utilização das inferências obtidas neste trabalho para implementar estratégias de adaptação no ambiente educacional.

2.3 Aspectos conceituais sobre emoções

Mesmo com os resultados promissores alcançados nos últimos anos, ainda existem barreiras a serem vencidas para a disseminação de aplicações baseadas em técnicas da computação afetiva (Van Den Broek, 2012). Neste contexto, Scherer (2005) defende que uma definição única

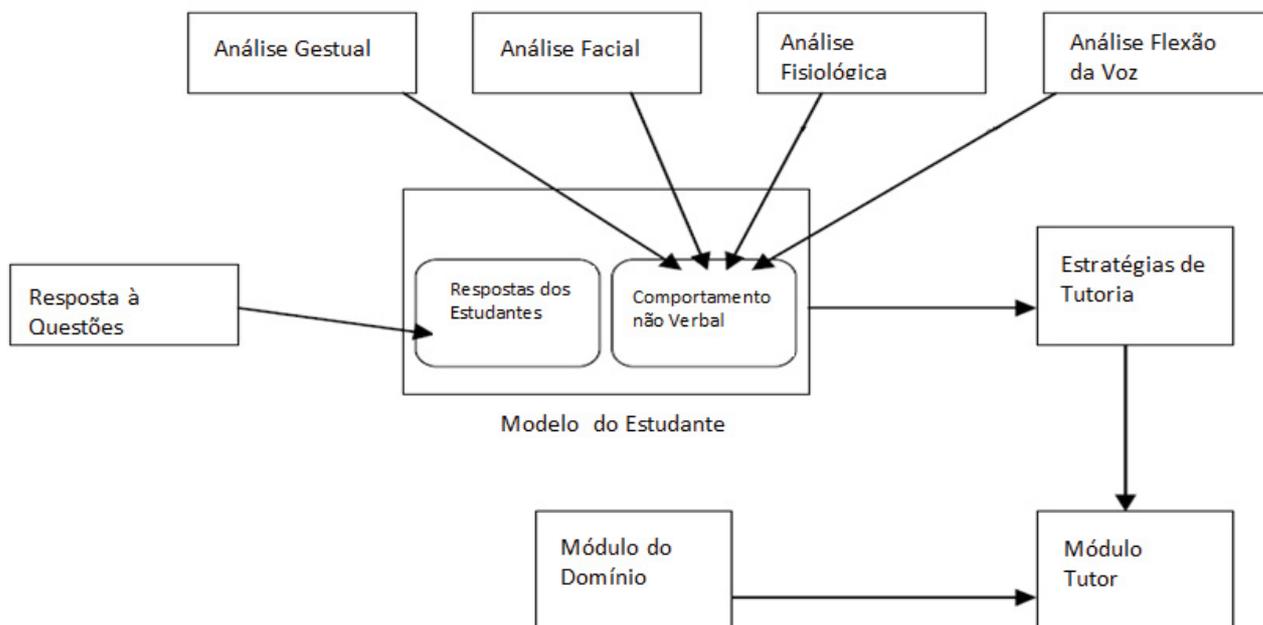


Figura 2.3: Arquitetura de um Sistema Tutor Afetivo. Fonte: Adaptado de Sarrafzadeh et al. (2008)

e concisa de emoção é um problema notório, sem consenso na comunidade científica, dificultando o progresso de pesquisas.

No último século, psicólogos e cientistas da área cognitiva têm desenvolvido teorias e definições relativas à emoções, motivação e outros fenômenos afetivos. Desde 1981, aproximadamente uma centena de definições de emoções foram propostas (Van Den Broek, 2012).

Emoções nos seres humanos são geradas a partir de mecanismos complexos que envolvem um complicado número de fatores como características individuais, ambientais e sociais (Picard, 1997). Estes fatores ainda não são totalmente conhecidos pela comunidade científica, demonstrando que o conhecimento já disponível nesta área precisa evoluir e se consolidar (Van Den Broek, 2012).

Mesmo estando fora do escopo e dos objetivos deste trabalho uma discussão exaustiva em relação às variadas teorias e definições relativas a emoções, considera-se importante detalhar terminologias e conceitos relevantes. Desta maneira, serão apresentados conceitos básicos de emoções relevantes para a área de pesquisa de Computação Afetiva.

Inicialmente, considera-se importante a definição do termo “Estado Afetivo”. Para Scherer (2005) estado afetivo é um termo abrangente que engloba um conjunto de outros construtos relacionados como emoções, humor, postura interpessoal, atitudes e disposições de afeto. A Tabela 2.1 apresenta um resumo de características e delimitações de emoção, humor e personalidade, incluindo uma breve descrição, exemplos e dimensões nas quais estes construtos psicológicos podem ser caracterizados (Scherer, 2000). Nesta tabela os símbolos indicam o grau com que as características estão presentes, com 0 indicando o valor mais baixo (ausência) e +++ indicando o valor mais alto. As setas indicam intervalos hipotéticos. Por exemplo, a duração de emoção é menor que humor que, por sua vez, tem duração menor que personalidade.

Tomando-se como base a abordagem híbrida de inferência de emoções proposta neste trabalho, torna-se relevante o conceito de emoções primárias e secundárias. Neste sentido, é

Tabela 2.1: Delimitação das características dos construtos psicológicos relativos a estados afetivos. Fonte: Adaptado de Scherer (2000)

Construtos	Breve definição e exemplos	Intensidade	Duração	Foco no evento	Desencadeamento da Avaliação	Velocidade de mudança	Impacto comportamental
Emoção	Episódio relativamente curto de resposta sincronizada de todos ou da maior parte dos subsistemas orgânicos em resposta a avaliação de um evento externo ou interno considerado significativo (e.g. raiva, tristeza, alegria, orgulho, vergonha, deses- pero)	++ -> +++	+	+++	+++	+++	+++
Humor	Estado Afetivo difuso, que se destaca por mudanças em sentimentos subjetivos, de baixa intensidade, mas de relativa longa duração, frequentemente sem causa aparente. (e.g., animado, tristonho, irritável, apático, depressivo).	+ -> ++	++	+	+	++	+
Personalidade	Disposições estáveis da personalidade e tendências de comportamentos típicos de uma pessoa (e.g. nervosa, ansiosa, hostil, imprudente, invejosa, ciumenta)	0 -> +	+++	0	0	0	+

fundamental considerar que não existe consenso na comunidade científica da complexa relação entre emoções geradas cognitivamente (secundárias) e não cognitivamente (primárias) (Scherer, 2005).

Todavia, a teoria de Damásio (Damásio, 2009) traz importante contribuição no entendimento e diferenciação entre emoções primárias e secundárias. De acordo com esta teoria, emoções primárias são natas, disparadas automaticamente em respostas reativas e emoções secundárias relacionam-se a processos cognitivos (e.g. reflexão, avaliação) e experiências anteriores. Como exemplo de emoção primária, pode-se citar medo ou raiva, enquanto frustração e esperança são exemplos de emoções secundárias. Para ilustrar determinadas diferenças entre emoções primárias e secundárias a Tabela 2.2 apresenta características que diferenciam “medo” (emoção primária) e “esperança” ou “expectativa” (emoção secundária).

Tabela 2.2: Características que diferenciam as emoções medo e esperança/expectativa . Fonte: adaptado de Picard (1997)

Características	Emoções	
	Medo (fear)	Esperança/Expectativa (hope)
Nível	Baixo nível	Alto Nível
Tipo de resposta	Respostas Rápidas/Automáticas	Processo Cognitivo/+ Lento
Coexistência	Deve sobrepor outras emoções	Deve coexistir com outras emoções
Sistema Orgânico Gerador	Sistema Límbico	Sistema Cortical
Tipo de Representação mais adequada	Sinal	Regras

Levando-se em conta o conhecimento existente e as diferentes teorias e definições, a modelagem e classificação de emoções tem sido feita seguindo duas abordagens teóricas principais: categóricas e dimensionais.

Teorias que usam a abordagem categórica propõem a definição de um conjunto discreto de categorias ou rótulos de emoções como, por exemplo, desgosto, alegria, entre outras (Picard, 1997). Como exemplo de teorias que usam a abordagem categórica pode-se citar Ekman (1992), Izard (1990) e Plutchik (1980). De maneira geral, estas teorias propõem um conjunto de emoções básicas ou primárias, variando a forma de organização e o número de emoções. O modelo de Izard (1990) propõe um conjunto de dez emoções, Plutchik (1980) oito e Ekman (1992) sete emoções.

Considerando a revisão da literatura a ser apresentada na última seção deste capítulo (ver Tabela 2.8, pode-se verificar que o modelo de Ekman (1992) é o mais utilizado em trabalhos correlatos. Neste modelo inicialmente foram representadas um conjunto de seis emoções que são também conhecidas como emoções “básicas”. Este conjunto inclui as seguintes emoções: raiva, desgosto, medo, alegria, surpresa e tristeza. Uma sétima emoção chamada ‘desprezo’ foi adicionada posteriormente ao conjunto de seis emoções iniciais (Matsumoto e Ekman, 2004). Além disso, um estado chamado “neutro” é incluído usualmente em aplicações computacionais, como a utilizada nesta pesquisa (ver Seção 4.3), representando a ausência de expressões faciais que indiquem emoções específicas.

Por outro lado, o modelo dimensional propõe uma forma diferente para representação das emoções. Segundo a abordagem dimensional, as emoções não são independentes entre si e se relacionam umas com as outras de uma maneira sistemática (Nicolaou et al., 2011). Calvo e D’Mello (2012) destacam que uma representação comum das emoções em um espaço

dimensional envolve duas dimensões: valência e ativação. A dimensão valência relaciona-se ao quão positiva (alegria) ou negativa (tristeza) uma emoção pode ser. Por outro lado, a dimensão ativação mede a intensidade de uma emoção, variando de agitação até calma. Por exemplo, a emoção medo tem valor negativo para a dimensão valência e valor positivo para a dimensão ativação.

Conforme observam Gunes et al. (2011), propostas que usam a abordagem dimensional podem incluir uma terceira dimensão chamada de Dominância (*dominance*) cujos valores variam de maior controle até menor controle.

É importante destacar que tanto o modelo categórico quanto dimensional são simplificações de emoções reais (Picard, 1997) e que existem propostas (e.g. (Shen et al., 2009; Baker et al., 2010; Becker-Asano e Wachsmuth, 2010)) de mapeamento entre as duas formas de representação. Um exemplo de mapeamento consiste em considerar categorias discretas de emoções como regiões em um espaço contínuo (Picard, 1997). Outro exemplo de mapeamento de modelo dimensional para categórico pode ser encontrado em Becker-Asano e Wachsmuth (2010). Neste trabalho é apresentada uma proposta de mapeamento do modelo dimensional com três dimensões: valência (V), ativação (A) e dominância (D). O intervalo de valores para cada uma das dimensões varia de -100 até 100. Desta forma, para a dimensão valência -100 representa emoções mais negativas e 100 emoções mais positivas. Para ativação -100 representa baixa ativação e 100 alta ativação. Na dimensão dominância, -100 são emoções com baixa dominância e 100 com alta dominância. Na Tabela 2.3 são demonstrados exemplos de emoções primárias e seus respectivos valores das três dimensões VAD (Valence, Arousal e Dominance) (Becker-Asano e Wachsmuth, 2010).

Tabela 2.3: Exemplos de mapeamento de emoções primárias (categóricas) para o modelo dimensional Fonte:Becker-Asano e Wachsmuth (2010).

Emoções Primárias	Valores VAD
Raiva	(80,80,100)
Tristeza	(-50,0,100)
Medo	(-80,80,100)

Modelos categóricos têm sido adotados por trabalhos correlatos que envolvem a inferência de emoções em contexto educacional (ver Tabela 2.8). Processos de inferência de emoções representadas em modelos categóricos, como o proposto neste trabalho, podem ser modelados como um problema de classificação. Sendo assim, uma variedade de técnicas de reconhecimento de padrões e de aprendizagem de máquina podem ser utilizadas (e.g. redes neurais, modelos probabilísticos, entre outros).

Levando-se em consideração o aspecto da simplicidade e a maneira pela qual pretende-se apresentar os resultados deste trabalho optou-se por adotar como base um modelo categórico de representação de emoções. Entretanto, estratégias de mapeamento entre o modelo categórico e dimensional foram necessárias para viabilizar a utilização de modelos e teorias considerados na construção desta pesquisa. No Capítulo 3 serão apresentados detalhes das estratégias de mapeamento utilizadas.

2.4 Inferências de emoções baseada em reações físicas

Pesquisas baseadas em informações físicas usam como referência as repostas fisiológicas que ocorrem simultaneamente ou que sucedem rapidamente uma emoção. As principais

expressões corporais utilizadas para inferências de emoções são: expressões faciais, voz, gestos, movimentos (Picard, 1997).

A expressão das emoções através de reações físicas é um processo natural e usualmente inconsciente, fazendo com que pesquisas e teorias tenham investigado como as emoções correlacionam-se com as reações físicas (Calvo e D’Mello, 2010; D’Mello et al., 2007; Ekman, 1992; Bosch et al., 2015; Paquette et al., 2016). Neste sentido, o trabalho de revisão apresentado por Poria et al. (2017) demonstra que aproximadamente 90% da pesquisa relevante em computação afetiva foca primariamente em áudio, vídeo/imagens (expressões faciais, gestos, etc) e texto. Porém, pode-se observar na revisão da literatura (ver tabela 2.8) que existe a utilização de uma diversidade de sinais de baixo nível como batimentos cardíacos, pressão sanguínea, condutividade da pele, entre outros.

É importante notar que certas reações físicas podem ser percebidas naturalmente por outras pessoas ou captados por dispositivos computacionais de imagem, enquanto outras requerem contato físico ou dispositivos específicos para sua medição, como, por exemplo, um sensor de batimentos cardíacos (Picard, 1997). Na Tabela 2.4 são descritos exemplos de reações físicas considerando-se o aspecto da aparência das reações para outras pessoas.

Tabela 2.4: Exemplos de reações físicas e o grau de aparência para outras pessoas Fonte:Picard (1997)

Reações físicas facilmente aparentes	Reações físicas menos aparentes
Expressões Faciais	Respiração
Entonação de Voz	Taxa de Batimentos Cardíacos
Gestos	Temperatura
Movimentos corporais	Transpiração
Postura	Pressão Sanguínea
Dilatação da pupila	Condutividade da Pele

Cada modalidade descrita na Tabela 2.4 tem suas vantagens e desvantagens para utilização como fonte viável para inferência de emoções. Dentre os fatores que interferem na escolha destaca-se: i) a validade do sinal como uma forma natural de identificação de emoções, ii) a confiabilidade do sinal em ambientes reais, iii) o tempo de obtenção e iv) custo e intrusividade para captura do sinal (Calvo e D’Mello, 2010).

A utilização das expressões faciais para a inferência de emoções é uma das técnicas mais amplamente utilizadas atualmente (Shen et al., 2009; Bosch et al., 2014; D’Mello, 2012; Hoque et al., 2012). Isto se deve ao fato de que a face é um dos principais mecanismos de comunicação de emoções (Picard, 1997) e desempenha um papel fundamental no processo de comunicação e interação dos seres humanos (Ekman, 1992). Desta forma as expressões faciais mantêm-se como uma das formas mais promissoras de detecção de emoções utilizando computadores (Shen et al., 2009). Expressões faciais apresentam a vantagem de não serem dependentes de ambiente computacional. Além disso apresentam fortes semelhanças entre as diferentes culturas e povos (Ekman, 1992). Por outro lado, Padrón-Rivera et al. (2016) aspectos culturais interferem na forma como as pessoas expressam ou reprimem as suas.

Embora existam outros métodos (e.g. MAX - *Maximally Discriminative Affect Coding System* (Izard, 1990)), método FACS (*Facial Action Coding System* (Ekman, 1992)) é a técnica mais utilizada na área de computação afetiva em pesquisas que utilizam informações de expressões faciais (Picard et al., 2004). O FACS é uma técnica baseada em observador humano projetada para detectar mudanças sutis em características faciais. No FACS cada movimento independente da face é denominado *Action Unit* - AU (ex. movimento de sobrancelhas, pálpebras, lábios). As

AU's representam os movimentos dos músculos usados para gerar expressões. São definidas 44 AU's cujas combinações representam protótipos (rótulos) de emoções específicas, como, por exemplo, tristeza.

A tarefa de rotulagem de banco de dados de imagens por observadores humanos é um exemplo de utilização do FACS. Como exemplo, a rotulagem do banco de dados Cohn-Kanade (Kanade et al., 2000) utilizado para ilustrar o processo de inferência do componente físico do ModHEmo (ver Seção 4.3) foi realizada usando o método FACS.

Considerando sua abrangência, o método FACS tem sido útil para descoberta de movimentos faciais que são indicativos de emoções (Bosch et al., 2014). Entretanto, determinadas limitações no modelo FACS devem ser consideradas. Conforme observa (Picard, 1997), esta técnica falhará no reconhecimento do estado de alegria caso a pessoa que estiver experimentando alegria suprimir todas as expressões faciais. Outra limitação importante do modelo FACS refere-se ao tempo necessário para identificação das expressões e também ao fato de que esta técnica foi desenvolvida originalmente para especialistas humanos codificar emoções baseadas em figuras estáticas.

A acurácia na inferência utilizando o modelo FACS é variável entre as diferentes emoções. Conforme concluem (McDaniel et al., 2007), encantamento é facilmente detectado por apresentar um conjunto de AU's pronunciados. Por outro lado, emoções como desprezo e desgosto mostraram-se difíceis de serem diferenciadas do estado 'neutro' por apresentarem expressões menos destacadas (Gottardo e Pimentel, 2016).

A voz também tem sido utilizada como uma importante fonte para identificar aspectos afetivos uma vez que as emoções modulam uma variedade de características da voz (Nicolaou et al., 2011). Certos parâmetros de efeitos vocais são habitualmente utilizados, como: velocidade da fala, intensidade e características da onda sonora (eg. frequência, amplitude, variações).

A inferência de informações afetivas em fala tende a ser mais difícil que a identificação através de expressões faciais. Por outro lado, expressões vocais são mais difíceis de controlar que expressões faciais (Picard, 1997).

Além de expressões faciais e voz, pesquisas têm utilizado uma diversidade de sensores e técnicas para coleta de dados visando a inferência de emoções. Fundamentando-se em teorias que relacionam emoções com reações físicas ou corporais, pesquisas (Haag et al., 2004; Woolf et al., 2009; Picard et al., 2016) focam na medida de sinais fisiológicos e a descoberta de padrões que correspondem à expressão de diferentes emoções.

Dentre os sinais fisiológicos utilizados, pode-se destacar, entre outros: i) atividade Eletrodermal (EDA) que mede a condutividade elétrica da pele, ii) eletromiograma (EMG) que mede atividade muscular e iii) eletrocardiograma (EKG ou ECG) que mede atividade cardíaca.

Sensores fisiológicos caracterizam-se por gerar um volume significativo de dados brutos que podem ser coletados continuamente enquanto as emoções são experimentadas. Neste sentido, um dos principais desafios consiste em identificar, em um universo potencialmente grande, qual é conjunto de atributos mais adequado para o reconhecimento das emoções (Picard, 1997).

Como principal problema na utilização de sensores fisiológicos, destaca-se que grande parte destes dispositivos são intrusivos, podendo, inclusive, interferir nas emoções. Buscando superar esta limitação, pesquisas têm sido desenvolvidas com o objetivo de inferir emoções utilizando apenas registros de uso e interação no ambiente (Paquette et al., 2016). Porém, os modelos de classificação gerados apenas com registro de interação tendem a capturar padrões específicos do ambiente de aprendizagem, sendo, portanto, pouco generalizáveis.

Dados obtidos utilizando padrões de comportamento também têm sido utilizados em pesquisas relacionadas à inferência de emoções. Dentre estes estudos, pode-se destacar a utilização de: i) sensores de postura e movimentos (D'Mello et al., 2007; Woolf et al., 2009;

Nicolaou et al., 2011; Grafsgaard et al., 2013b); ii) mouse com sensor de pressão (Woolf et al., 2009); iii) sensor de movimentos oculares (Jaques et al., 2014). Nesta linha, Adams et al. (2015) demonstram que a expressividade emocional nos movimentos da cabeça não se limitam a acenos ou balanços, e que gestos adicionais (como inclinações da cabeça, e quantidade geral de movimento) podem ser úteis para sistemas de reconhecimento automatizados.

Buscando melhorar os resultados no processo de reconhecimento de emoções, técnicas podem ser utilizadas em conjunto. Como exemplo, (Picard, 1997) destaca a combinação de expressões faciais e voz como uma alternativa promissora, dado que ativação (*arousal*) correlaciona-se mais fortemente com a voz e valência (*valence*) com expressões faciais. Entretanto, esta combinação é um desafio, devido ao fato de que a detecção de expressões faciais é dificultada quando simultaneamente a pessoa está falando.

Na Figura 2.4 é mostrado um exemplo da utilização conjunta de sensores para inferência de emoções de um estudante enquanto interage com o ambiente STI AutoTutor (D’Mello et al., 2007).

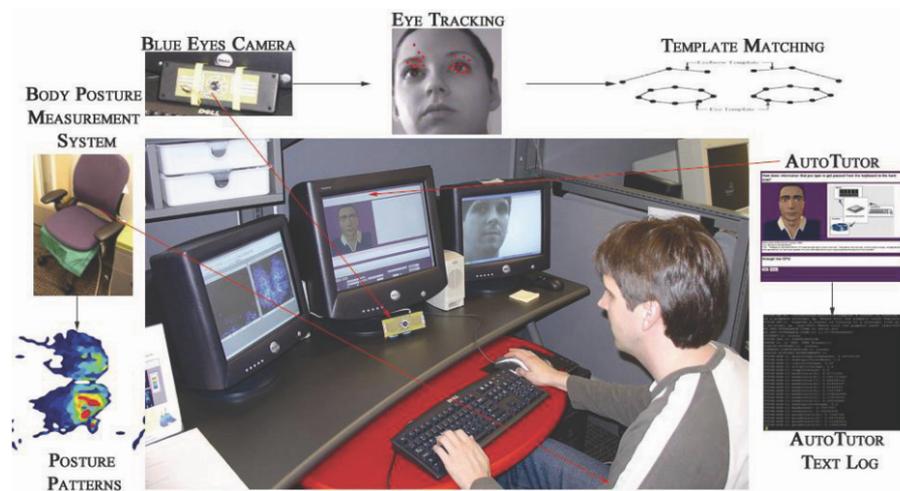


Figura 2.4: Estudante sendo monitorado por múltiplos sensores enquanto interage com um ambiente educacional. Fonte: D’Mello et al. (2007)

A fusão de um conjunto de sensores mostra-se uma alternativa com potencial interessante no processo de inferência de emoções. Todavia, no contexto educacional, a escolha dos sensores a serem usados torna-se um desafio tendo em vista aspectos como custo, intrusividade e viabilidade de utilização em ambientes reais. Como exemplo, o ambiente de laboratório apresentado na figura 2.4 claramente apresenta pouca viabilidade de implantação em cenários reais de ensino no Brasil nos dias atuais.

2.5 Inferência de emoções utilizando informações cognitivas

Teóricos da área cognitivista de psicologia das emoções como Ortony et al. (1990) defendem a importância de se avaliar os aspectos cognitivos envolvido no processo de geração e controle das emoções nos seres humanos. A Teoria da Avaliação (*Appraisal¹ Theory*) constitui uma importante base da abordagem cognitivista (Picard, 1997). A linha central desta teoria é que as emoções dependem do raciocínio sobre situações e da avaliação (*appraisal*) do valor ou do

¹No contexto deste trabalho o termo *appraisal* será traduzido como avaliação.

significado emocional de cada situação (Ortony et al., 1990). As situações podem envolver objetos ou eventos que afetam uma pessoa, considerando seus objetivos, preferências ou expectativas.

Scherer et al. (2001) destacam que as emoções são geradas e diferenciadas com base em uma avaliação subjetiva da significância pessoal de uma situação, evento ou objeto em um número variado de dimensões ou critérios. Picard (1997) completa que as emoções geradas cognitivamente envolvem elementos cognitivos como comparação, categorização, inferência, atribuição ou julgamento.

O aspecto da percepção ou avaliação pode ser observado na definição de emoção seguindo a abordagem cognitivista encontrada em (Ortony et al., 1990). Estes autores definem emoções como estados mentais internos que são suscitados quando as pessoas reagem a situações em curso que são percebidas como boas ou ruins para seus interesses pessoais. Adicionalmente, conforme Gunes et al. (2011) as emoções são geradas através de um mecanismo contínuo e com avaliação subjetiva e recursiva de estados internos e do estado do mundo exterior.

Um exemplo da dinâmica de geração de emoções cognitivamente pode ser observado ao sentirmos raiva ou irritação quando alguém (agente externo) está impedindo ou interferindo negativamente no atingimento de objetivos pessoais.

Seguindo a abordagem cognitivista, sabendo-se os objetivos e a percepção de eventos relevantes, poder-se-ia inferir as emoções de uma pessoa. Sendo assim, presume-se que as emoções são ativadas baseando-se na percepção individual em relação a aspectos positivos ou negativos de um evento ou objeto (Jaques, 2004). Todavia, conforme observa Conati (2011), estas informações não são sempre facilmente disponíveis ou obtidas.

O modelo OCC proposto por Ortony, Clore e Collins (Ortony et al., 1990) é uma das propostas mais amplamente adotada em trabalhos que seguem a abordagem cognitivista descrita acima (Frasson e Chalfoun, 2010). Na literatura consultada para construção desta pesquisa, envolvendo ambiente computacionais afetivos, trabalhos correlatos (Conati, 2002, 2011; Gebhard, 2005; Becker-Asano e Wachsmuth, 2010; Steunebrink et al., 2009; Jaques et al., 2011; Jaques e Vicari, 2007) têm usado este modelo com relativo sucesso.

O modelo de Roseman (Roseman et al., 1990) é outro exemplo de modelo cognitivo em que um conjunto de seis condições de ativação interagem para dar origem às dezessete emoções descritas neste modelo. Porém, este modelo não foi utilizado em nenhum dos trabalhos correlatos (ver Tabela 2.8).

Desta forma, o modelo OCC (Ortony et al., 1990) foi escolhido pela sua popularidade e destacada utilização em modelos computacionais baseados na abordagem cognitivista (Conati, 2011, 2002; Jaques, 2004; Jaques et al., 2014; Gebhard, 2005). Dada sua importância no contexto desta pesquisa, o modelo OCC será apresentado com mais detalhes na próxima seção.

2.5.1 Modelo OCC

O modelo OCC (Ortony et al., 1990) foi construído baseando-se na teoria cognitivista das emoções, explicando as origens das emoções e as agrupando de acordo com o processo cognitivo que as gera. Um aspecto importante a considerar refere-se ao fato de que no modelo OCC as emoções experimentadas dependem da interpretação pessoal de cada situação, alinhando-se, portanto, à teoria cognitivista das emoções.

Conforme observa Jaques (2004), o modelo OCC provê informações sobre como construir uma interpretação de uma situação do ponto de vista da pessoa (e.g. usuário de um sistema) e para que emoção esta interpretação conduz. No modelo OCC, assume-se que as emoções originam-se de reações de valência (positiva ou negativa) frente eventos, agentes e objetos. Desta maneira, conforme mostrado na Figura 2.5, formam-se três ramos de emoções,

caracterizados como: i) reações à Consequências de Eventos; ii) reações à Ações de Agentes e iii) reações à Aspectos de Objetos.

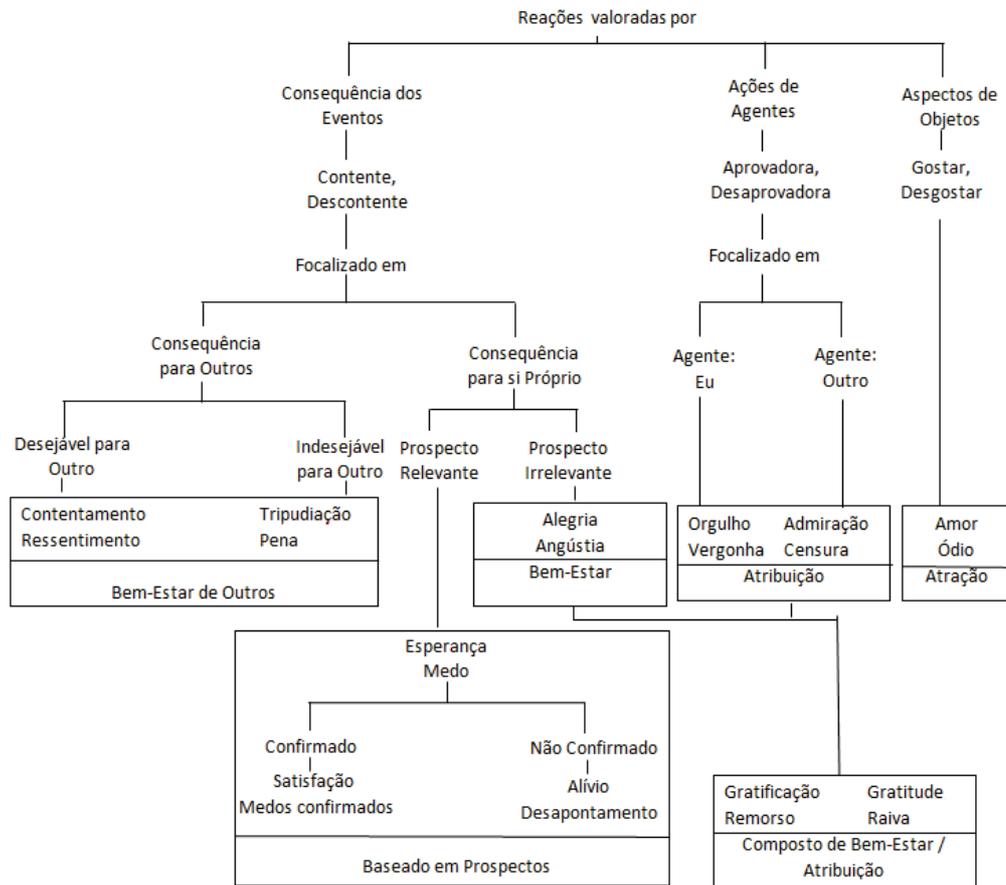


Figura 2.5: Estrutura do modelo OCC. Fonte: Adaptado de Ortony et al. (1990).

Além disto, no modelo OCC as emoções são diferenciadas considerando os seguintes fatores:

- Reações relacionadas a como um evento ocorrido impacta nos objetivos próprios (contente/descontente em relação ao evento, aprovação/reprovação se o evento foi gerado por uma terceira parte e orgulho/vergonha se o evento foi gerado pelo próprio autor).
- Reações relativas à como um evento impacta outras pessoas que o autor pode desejar ou não. (contentamento / ressentimento se o impacto é positivo ou pena / tripudiação se o impacto é negativo).
- Emoções relacionadas a efeitos esperados de um evento (esperança / medo ou alívio / desapontamento).

Observando-se a estrutura da Figura 2.5, pode-se verificar, por exemplo, que a emoção esperança (*hope*) é gerada quando uma pessoa desenvolve uma expectativa que um evento bom para si próprio (agente) acontecerá no futuro (prospecto). Caso este evento seja confirmado, a emoção satisfação será gerada, ou, em caso negativo, desapontamento.

Apesar de não ter sido desenvolvido com este propósito, o modelo OCC (Ortony et al., 1990) alcançou destaque na comunidade científica de inteligência artificial por ser o primeiro a apresentar um sistema baseado em regras para a geração das emoções (Picard, 1997).

Considerando os motivos elencados acima, entendeu-se que a abordagem OCC seria a melhor alternativa para a construção do componente cognitivo do modelo híbrido de inferência de emoções que é a proposta central deste trabalho de pesquisa.

No modelo OCC as inferências das emoções podem ser realizadas considerando-se o comportamento observável do usuário, podendo-se usar para isto as ações do usuário na interface de um ambiente computacional. Esta será a abordagem utilizada nesta pesquisa.

É importante destacar também que o modelo OCC apresenta limitações e nem sempre provê indicações inequívocas da real reação emocional (Conati, 2011, 2002). Dentre estas limitações pode-se destacar a necessidade de conhecimento dos objetivos e percepção de eventos relevantes para o usuário (Conati, 2011) ou dificuldades para definição da intensidade das emoções, como é o caso da pesquisa de (Jaques, 2004). Estas informações podem não ser facilmente disponíveis em determinados ambientes.

Entretanto, este problema particular pode ser aliviado se o ambiente for capaz de detectar ações do estudante que indicam movimentos em direção a um estado de valência negativa ou positiva (Conati, 2011). No contexto desta pesquisa, estes movimentos serão detectados por eventos ocorridos no ambiente (e.g. resposta errada, solicitações repetidas de ajuda, etc). Espera-se também que as inferências obtidas através das expressões faciais sejam úteis para a superação das limitações do modelo OCC destacadas acima.

2.6 Possibilidades de utilização das inferências afetivas

Cabe reforçar que escopo e o objetivo principal desta pesquisa estão relacionados com o processo de inferência de emoções. Mesmo assim, entendeu-se que seria importante uma discussão sobre as possibilidades de uso na prática das informações afetivas inferidas. Desta forma, baseando-se em concepções teóricas obtidas em trabalhos correlatos, nesta Seção serão apresentadas potenciais alternativas de uso das informações afetivas inferidas.

As intervenções pedagógicas representam uma aplicação prática em que as informações afetivas poderiam ter um uso relevante. Neste sentido, diferentes tipos de intervenções pedagógicas podem ser realizadas com o objetivo de superar estados emocionais que afetam negativamente a aprendizagem ou reforçar estados que facilitam o aprendizado. Um exemplo de intervenção pedagógica em um ambiente de aprendizagem baseado em jogos pode ser encontrado em Conati (2002). A autora cita o caso em que, se o ambiente educacional detectar que o estudante está envergonhado porque cometeu erros durante o jogo, poderá ser tentado tomar ações que façam o estudante sentir-se mais confiante sobre seu desempenho. Por outro lado, se o ambiente detectar um estado de ansiedade em um ponto específico do jogo, poderá ser iniciada uma interação com o estudante como único propósito de diverti-lo.

Outro exemplo pode ser encontrado no trabalho de Jaques (2004) em que um agente pedagógico animado promove emoções positivas nos estudantes. Para isto, são apresentados comportamentos emocionais e mensagens, chamadas de táticas afetivas, que são aplicadas de acordo com as emoções dos aprendizes.

Emoções como confusão e frustração destacam-se pela sua importância em situações de aprendizagem. Neste sentido, Reis et al. (2018a) observa que quando a confusão não é adequadamente monitorada e gerenciado pode avançar para o estado de tédio, dificultando ou mesmo impedindo o aprendizado. Bosch et al. (2015) defendem que confusão e frustração oferecem oportunidades de intervenção na forma de dicas ou revisitando material introdutório relacionado aos conceitos do problema.

D'Mello et al. (2007) completam que é fundamental que os ambientes de aprendizagem pró-ativamente antecipem e tentem evitar um conjunto de emoções negativas ao processo de

aprendizagem. Desta maneira, exemplos de emoções e intervenções possíveis discutidas neste trabalho são apresentadas na Tabela 2.5.

Tabela 2.5: Exemplos de emoções e possíveis intervenções Fonte: D’Mello et al. (2007)

Emoção do Aprendiz	Tipo de intervenção possível
Tédio	Tentar engajar o estudante em uma atividade que aumente o interesse (ex. simulações, desafios, jogos)
Frustração	Feedbacks diretos, declarações e correção de concepções errôneas
Engajamento	Evitar intervenções (<i>lay low</i>)
Confusão	É um momento chave para encorajar o aprendiz, possivelmente com feedbacks positivos. Não é necessariamente um estado a ser evitado, pois tem sido correlacionado positivamente com aprendizagem.

De maneira complementar à abordagem de intervenção proposta acima por D’Mello et al. (2007), Woolf et al. (2009) apresentam uma estratégia de intervenção tutorial que combina elementos cognitivos e afetivos. A tabela 2.6 apresenta exemplos desta estratégia, contendo o elemento cognitivo e afetivo que norteia a escolha da estratégia de intervenção mais adequada.

Tabela 2.6: Estratégias adaptativas que combinam elementos afetivos e cognitivos. Fonte: Woolf et al. (2009).

Elemento Cognitivo	Elemento Afetivo	Intervenção Tutorial Adequada
Estudante comete um erro	Estudante demonstra estar curioso e focado	Nenhuma intervenção é requerida. Estudante está engajado no aprendizado e exploração
	Estudante franze a testa, fica inquieto e olha ao redor	Ação de intervenção é plausível, pois, possivelmente o estudante está confuso.
Estudante está resolvendo problemas corretamente	Estudante não mostra sinais de frustração ou aborrecimento	Intervenção não é necessária. Estudante está no controle, concentrado e focado.
	Estudante mostra sinais de tédio ou decepção	A intervenção é necessária, possivelmente aumentando a dificuldade dos problemas

No trabalho de Vanlehn (2006) são discutidos os conceitos de *outer loop* (laço externo) *inner loop* (laço interno) e que são relacionados a estratégias de tutoria, intervenções e feedback em um STI. Conforme Vanlehn (2006), o *outer loop* externo é executado uma vez para cada tarefa, onde uma tarefa geralmente consiste em resolver um problema complexo com várias etapas. O *inner loop* é executado uma vez para cada passo dado pelo estudante na solução de uma tarefa específica.

No *inner loop* são compreendidas ações como feedback e dicas com o objetivo de auxiliar o aprendiz na realização das tarefas (Vanlehn, 2006). Neste ponto, entende-se que as informações afetivas inferidas poderiam ser úteis para a decisão de quando apresentar e também qual o tipo de feedback ou dica a ser exibido.

Como exemplo, Vanlehn (2006) cita o caso do STI Transom (Roberts et al., 2000) usado para ensinar a dirigir um veículo subaquático operado remotamente. Devido à entrada sensorial limitada do veículo (vídeo e sonar), geralmente os estudantes ficam confusos sobre a orientação e posição do veículo (Vanlehn, 2006). Nestes casos, a inferência da emoção confusão poderia disparar dicas sobre a posição e orientação do veículo. No STI original, o estudante precisa solicitar estas dicas.

Informações relativas ao número de ocorrência, persistência e a dinâmica das transições entre as emoções é a base para a inferência de padrões de mais alto nível como “Humor do Aprendiz” (*Learning Mood*) (positivo/negativo/neutro) (Baker et al., 2010), “Ciclo Vicioso” (*Vicious Cycle*) (D’Mello et al., 2007) ou “Concentração Engajada” (*Engaged Concentration*) (Baker et al., 2010). Um exemplo da importância de considerar a dinâmica das emoções é apresentado em D’Mello (2012). Estes autores observam que quando a confusão não é propriamente acompanhada e administrada o estudante pode ficar entediado, situação que dificulta ou até impede a aprendizagem.

Emoções com incidência/persistência preponderantes podem ser usados para inferir o “humor do aprendiz”, enquanto que emoções com valência negativa que dificultam o aprendizado como tédio e frustração indicam um ciclo vicioso (Baker et al., 2010). Por outro lado, “Concentração Engajada” é um estado afetivo-cognitivo com valência positiva que envolve atenção e foco na tarefa (Baker et al., 2010).

O trabalho de Reis et al. (2018a) apresenta um estudo que investiga a dinâmica do processo de transição entre emoções de estudantes relacionando com seu perfil de personalidade e conhecimento prévio do conteúdo. Como exemplo, no trabalho citado foi reportado que um estudante com perfil de personalidade “neuroticismo” e baixo conhecimento do conteúdo tende a transitar do estado confusão para tédio em um tempo de 19 segundos. Nesta mesma situação um estudante com alto conhecimento no conteúdo e perfil de personalidade “extrovertido” demora em torno de 313 segundos para a mesma transição.

Contudo, é importante ressaltar que não foi encontrada na literatura pesquisada uma indicação clara da quantidade ou tempo de reincidência de emoções com valência negativa para ser considerado um ciclo vicioso ou concentração engajada. Desta forma, estudos adicionais necessitam ser realizados para determinar estes parâmetros.

2.7 Estado da Arte

Nesta seção serão analisados os principais trabalhos que representam o estado da arte relativo à tarefa de inferência automática de emoções em ambientes computacionais de aprendizagem. Esta tarefa será realizada utilizando uma revisão sistemática da literatura com o objetivo de responder às seguintes questões:

1. Quais as informações utilizadas para a tarefa de inferência das emoções?
2. Quais os modelos de representação das emoções utilizado?
3. Que técnicas e ferramentas computacionais são mais utilizadas?
4. Quais os principais resultados obtidos?

Foi realizada uma revisão sistemática da literatura buscando verificar a existência de trabalhos recentes (de 2013 a 2018). Este processo de revisão sistemática foi desenvolvido seguindo as seguintes etapas: identificação das bases de dados para pesquisa, definição dos

critérios de busca e exclusão, realização das pesquisas, análise crítica dos trabalhos e organização de um resumo dos trabalhos correlatos.

Para a escolha das bases de dados primárias foram consideradas as principais bases digitais da área de Computação que incluem tanto periódicos como anais de congressos relevantes. Desta forma, optou-se por pesquisar nos seguintes repositórios digitais:

- ACM Digital Library
- IEEE Explore
- Science Direct (Elsevier)
- Springer Link (Springer)

A definição dos critérios de busca nas bases escolhidas envolveu um processo iterativo (escrita, testes e refinamento) para a obtenção de uma *string* de busca eficiente. Ao final deste processo, obteve-se a seguinte *string* de busca:

((emotion OR affect) AND (detection OR recognition OR infer OR predict) AND (learning OR education) AND NOT (speech OR "text mining" OR music))

Com exceção do repositório ACM Digital Library, todos os demais incluem pesquisas de todas as áreas do conhecimento e não foi encontrada funcionalidade que possibilitasse aplicar um filtro para selecionar somente computação. Assim, a *string* de busca descrita acima retorna um número elevado de trabalhos de áreas como psicologia e neurociência, por exemplo. Desta forma, o principal **critério de inclusão** foi a escolha somente dos periódicos ou anais de congressos da área de computação. Para isso, optou-se por incluir apenas os periódicos da área de computação que continham artigos na lista dos 30 mais relevantes em cada repositório. Adicionalmente, foram incluídos os trabalhos relacionadas nas recentes revisões sistemática da literatura de Reis et al. (2018b) e Moraes et al. (2017).

No contexto internacional foi encontrada revisão sistemática recente de Reis et al. (2018b) que apresenta um panorama global dos trabalhos que relacionados a inferência de emoções em Sistemas Tutores Afetivos. Neste trabalho a busca foi realizada utilizando as palavras chave “*emotion*”, “*affective*” e “*sentiment*”, além das palavras “*Intelligent Tutoring Systems*”, “*Affective Tutoring Systems*”, “ITS” e “ATS”. As bases de dados digitais usadas foram: ACM Digital Library, Scopus, Web of Knowledge, Elsevier (via Science Direct) e IEEEExplore. Esta busca foi realizada entre 05/08/2016 e 10/08/2016.

Para avaliação das pesquisas publicadas no Brasil, utilizou-se como referência o trabalho de Moraes et al. (2017). Este trabalho apresenta um levantamento sistemático dos trabalhos científicos brasileiros na área de Computação Afetiva no contexto da Informática na Educação. No referido trabalho, foram considerados os principais veículos brasileiros de divulgação científica na área, sendo eles: Simpósio Brasileiro de Informática na Educação (SBIE), Workshop de Informática na Escola (WIE) e Revista Brasileira de Informática na Educação (RBIE). A pesquisa nestes veículos de divulgação foi feita utilizando a busca por artigos que apresentavam em seus títulos as raízes das palavras-chave *afe*, *aff*, *emo*, *mot*, *humo*, *sent*, *perso*, *cog* e *adapt*. Esta revisão foi feita em junho de 2017 e englobou publicações a partir de 2001.

A Tabela 2.7 mostra o resultado do processo de busca com o critério de inclusão descrito acima (última atualização em 08/08/2018) e os dois trabalhos de revisão sistemática destacados. Nesta tabela são apresentados os repositórios, os periódicos ou congressos e o número total de artigos. Nesta tabela também são apresentados os artigos encontrados nos trabalhos de Reis et al. (2018b) e Moraes et al. (2017).

Como **critérios de exclusão** considerou-se as seguintes condições:

Tabela 2.7: Quantidade de artigos recuperados para avaliação

Repositório	Número de Artigos	Título dos Periódicos e Congressos
ACM Digital Library	26	ACM International Conference on Multimodal Interaction Workshop on Emotion Representation and Modelling in HCI ACM on International Conference on Multimedia Retrieval International Conference on Intelligent User Interfaces International Conference on Machine Learning and Soft Computing International Workshop on Audio/Visual Emotion Challenge Annual Workshop on Audio/Visual Emotion Challenge International Symposium on Information and Communication Tech. International Conference on Learning Analytics and Knowledge Brazilian Symposium on Human Factors in Computing Systems
IEEE Explore	52	IEEE Transactions on Affective Computing IEEE International Conference on Automatic Face & Gesture Recog.
Science Direct (Elsevier)	91	Computers & Education Computers in Human Behavior
SpringerLink	217	ITS - Intelligent Tutoring Systems AIED - Artificial Intelligence In Education IJAIED - International Journal of Artificial Intelligence in Education
Reis et al. (2018b)	40	International Conference on Knowledge Science, Engineering and Management Entertainment Computing International Conference on Affective Computing and Intelligent Interaction AIED - Artificial Intelligence in Education International journal of human-computer studies Mexican International Conference on Artificial Intelligence (MICAI) International Conference on Learning Analytics & Knowledge IEEE Transactions on Affective Computing International Conference on User Modeling, Adaptation, and Personalization
Morais et al. (2017)	24	SBIE - Simpósio Brasileiro de Informática na Educação WIE - Workshop de Informática na Escola RBIE - Revista Brasileira de Informática na Educação
Total	450	

1. Trabalhos que não apresentem método de inferência automática de emoções em ambientes educacionais.
2. Trabalhos que não apresentem estudos empíricos.
3. Trabalhos que não reportem informações objetivas sobre o desempenho do processo de inferência.
4. Trabalhos de revisão da literatura ou que apenas cite resultados de outros trabalhos.

A aplicação dos critérios de exclusão acima foi feita baseando-se no título e resumo dos trabalhos. Desta forma, foram descartados 419 dos 450 trabalhos listados na Tabela 2.7, restando 31 artigos para análise completa. Durante a leitura destes 31 artigos, foram encontradas referências reiteradas aos trabalhos de Woolf et al. (2009) e D’Mello et al. (2007). Após a leitura destes dois trabalhos, identificou-se que apresentam propostas relevantes de inferências automática de emoções e, por isso, foram incluídos nesta revisão. Desta forma, esta revisão sistemática resultou na identificação de um total de 33 trabalhos.

Cabe ressaltar que dois trabalhos dos autores desta tese publicados nos congressos *International Conference on Intelligent Tutoring Systems, ITS - 2018* e *Brazilian Symposium on Human Factors in Computing Systems - IHC 2017* (ver Seção 6.3) não foram incluídos por tratarem do mesmo tema e resultados descritos neste documento.

Visando permitir a análise crítica e responder às questões propostas nesta revisão sistemática, apresenta-se na Tabela 2.8 um lista de todos os trabalhos selecionados. Esta tabela contém um total de 33 trabalhos, com as seguintes origens (coluna origem): 19 revisão própria, 11 revisão de Reis et al. (2018b) e 3 revisão de Moraes et al. (2017). A tabela apresenta a referência dos trabalhos, fonte das informações e modelo de representação de emoções. Além disso, são apresentadas também as técnicas e ferramentas computacionais e os principais resultados reportados.

Tomando-se como base as informações dos trabalhos resumidos na Tabela 2.8, serão apresentadas respostas para as quatro questões norteadoras desta revisão sistemática destacadas no início desta seção.

Relativamente à **primeira questão** que trata das fontes de informações utilizados para inferência das emoções, pode-se observar que: i) expressões faciais são o método mais utilizado (52%), seguidos por fusão de fontes variadas de informações (46%), logs de uso e interação (33%), movimentos corporais (30%) e dados fisiológicos individuais (EMG, EEG, PPG, GSR, etc) (27%). A soma dos percentuais apresentados não corresponde a 100%, pois existem trabalhos que utilizam mais de uma fonte de informação. É importante destacar que apenas cinco dos 33 trabalhos (15%) utilizam uma abordagem cognitivista para inferência de emoções e nenhum investiga uma abordagem híbrida (física e cognitiva).

A **segunda questão** refere-se aos modelos de representação de emoções utilizado. Neste sentido, observa-se que o modelo categórico é utilizado por 32 dos 33 (97%) trabalhos. Apenas o trabalho de Woolf et al. (2009) utiliza o modelo dimensional. Dos 17 trabalhos que utilizam expressões faciais, 11 (65%) baseiam-se na teoria de Ekman (1992). Todos os cinco trabalhos que usam abordagem cognitivista usam como referência o modelo OCC (Ortony et al., 1990) com as interações dos estudantes servindo como base para o processo de inferência.

Em relação a **questão de número três** que busca investigar as técnicas e ferramentas computacionais mais utilizadas, pode-se observar que algoritmos de classificação (e.g. SVM, Redes Neurais, C4.5, etc) são utilizados por 22 dos 33 (67%) trabalhos. Este fato justifica-se levando-se em conta a opção pelo modelo categórico de representação das emoções que favorecem a utilização de técnicas de classificação.

Origem	Referência	Informações para inferência de Emoções	Modelo de Emoções	Técnicas e Ferramentas Computacionais	Principais Resultados
Revisão Própria	Woolf et al. (2009)	Expressões faciais, sensores de movimentos e condutividade da pele	Modelo dimensional: valência e ativação	Redes Bayesianas e Correlação Estatística	Correlação entre posições e movimentos de cabeça com emoções. Conjuntos dos sensores foram melhores para determinadas emoções
Revisão Própria	Bosch et al. (2014)	Expressões Faciais	Catégorico Ekman (1992)	15 Algoritmos de Classificação no WEKA	Taxas Kappa 0.038 e 0.232
Revisão Própria	Paquette et al. (2014)	Registros de interação dos estudantes	Modelo catégorico	Algoritmos de classificação J48, JRip, Step Regression	Índices Kappa=0,354 e A'=0.720
Revisão Própria	Grafsgaard et al. (2013b)	Gestos (uma-mão-na-face e duas-mãos-na-face) e mudanças posturais	Modelo catégorico	Microsoft Kinect	Mudanças posturais apresentaram alta frequência e ocorrência com emoções negativas
Revisão Própria	D'Mello et al. (2007)	Diálogos tutoriais, sensor de postura e expressões faciais	Modelo Catégorico de Ekman (1992)	Algoritmos Naive Bayes, MLP, C4.5	Sensor de postura para as emoções tédio e engajamento, expressões faciais para encantamento/eureka e texto para frustração
Revisão Própria	Jaques et al. (2014)	Sensor de movimentos oculares	Modelo Catégorico	Algoritmos Random Forests, Naive Bayes, Support Vector Machines (SVM)	Acurácia média de 69% para tédio e 73% para curiosidade
Revisão Própria	Pham e Wang (2018); Xiao et al. (2017)	Sinal PPG (<i>photoplethysmography</i>) e expressões faciais	Modelo Catégorico	Affdex SDK (baseado em AU's), <i>photoplethysmography</i> para sinais PPG	Acurácia global de 84,4% e Kappa global de 0.35 na inferência de 6 emoções
Revisão Própria	Salmeron-Majadas et al. (2015)	Dados fisiológicos, interação com mouse e teclado	Modelo catégorico	Algoritmos SVM, C4.5, RandomForest, Redes Neurais e Rede Bayesianas	A detecção de emoções usando a abordagem multimodal alcançou melhores resultados

Revisão Própria	Padrón-Rivera e Rebolledo-Mendez (2015)	Expressões Faciais	Modelo Categórico de Ekman	Algoritmo ID3	Acurácia de 62,86% na identificação de trajetórias afetivas relacionadas com aprendizagem
Revisão Própria	Okur et al. (2017)	Expressões faciais e movimentos obtidos com câmara 3D e Log de interação	Modelo Categórico (on/off task)	Algoritmo Random Forest	F1-score de 82% na detecção de dois estados de engajamento: on/off task
Revisão Própria	Vail et al. (2016)	Expressões faciais, atividade eletrodermal, postura corporal e gestos	Modelo Categórico (ganho de aprendizado)	FACET (software comercial)	Subconjunto de tipos de expressão facial e condutividade da pele são altamente preditivos do desempenho futuro dos estudantes.
Revisão Própria	Tiam-Lee e Sumi (2018)	Expressões faciais, padrão de digitação, resultado de compilação de funções	Modelo Categórico (Ekman)	HMM (<i>Hidden Markov Chain</i>)	Acurácia de 77,78% na inferência de duas emoções (confusão e engajamento)
Revisão Própria	Doubouya et al. (2018)	Expressões faciais, Eletroencefalograma (EEG) e movimento do olhos	Modelo categórico	Algoritmo k-NN e Random Forest	Acurácia bastante variável, entre 30% e 96% na identificação da emoções durante um jogo computacional
Revisão Própria	DeFalco et al. (2018)	Postura corporal, condutividade e temperatura da pele	Modelo categórico	J48, JRip, Naive Bayes, Logistic Regression e KStar	Resultados bastante variáveis entre as diferentes emoções e algoritmos com Kappa entre 0,056 e 0,469
Revisão Própria	Perusquia-Hernandez et al. (2017)	Eletromiografia (EMG) dos músculos da face	Modelo Categórico (micro expressões de sorriso)	Algoritmo de Redes Neurais	Potencial da EMG para detectar expressões espontâneas rápidas e sutis como micro sorrisos (não considera emoções) alcançando acurácia entre 100% e 90%
Revisão Própria	Happy e Routray (2015)	Expressões Faciais	Modelo Categórico - Emoções Básicas de Ekman	Algoritmo SVM	Acurácia entre 91,8% e 94,09% em banco de dados de emoções simuladas

Revisão Própria	Bosch et al. (2015)	Expressões Faciais e Movimentos corporais	Modelo Categórico (tédio, confusão, encanto, engajamento e frustração)	Algoritmos SVM, C4.5, Classificador bayesiano	AUC e Acurácia global média de 0,655 e 54% respectivamente
Revisão Própria	Zhang et al. (2018)	Expressões Faciais	Modelo Categórico Ekman	Algoritmo SVM	Acurácia média de 80,22% em experimentos com banco de dados de imagens
Revisão Própria	Akputu et al. (2018)	Expressões Faciais	Modelo Categórico Ekman	Algoritmo SVM	Acurácia entre 77,5% e 83% em quatro banco de dados de imagens
Reis et al. (2018b)	Muñoz et al. (2010)	Dados de interação com o ITS PlayPhysics - Modelo OCC	Modelo Categórico	Bayesian Networks (BNs)	Acurácia de 60.71% na inferência de 6 emoções
Reis et al. (2018b)	Muñoz et al. (2011)	Dados de Interação no ITS PlayPhysics - Modelo OCC	Modelo Categórico	Bayesian Networks (BNs)	Acurácia entre 66,67% e 75% na inferência de 2 emoções
Reis et al. (2018b)	Jaques et al. (2011)	Interação do estudante com o ambiente - Modelo OCC	Modelo Categórico	X-BDI	Taxa de acurácia entre 77% e 0% no reconhecimento de 4 grupos de emoções
Reis et al. (2018b)	Doddannara et al. (2013)	Interação do estudante com o ambiente	Modelo Categórico	Jrip, Naive-Bayes, KNN, RepTree	Kappa entre 0,23 e 0,40 na detecção de quatro emoções
Reis et al. (2018b)	Grafsgaard et al. (2013a)	Expressões Faciais	Modelo categórico Ekman (1992)	Regressão linear	Correlação significativa entre expressão facial, frustração e aprendizado: AU 2 (elevar a sobrancelha externa) foi negativamente correlacionada com ganho de aprendizagem
Reis et al. (2018b)	Harley et al. (2013)	Expressões Faciais	Modelo categórico Ekman (1992)	ANOVA	Concordância entre expressões faciais e questionários foi de 75,6%

Reis et al. (2018b)	Kapoor et al. (2007)	Expressões Faciais, movimentos corporais, condutividade da pele, pressão mouse, pressão na cadeira	Modelo Categórico	SVM, Gaussian Process, KNN	Acurácia de 79% na inferência de frustração
Reis et al. (2018b)	Zatarain-Cabada et al. (2013)	Expressões Faciais (Ekman, 1992)	Modelo Categórico	Rede Neural	Acurácia média de 70% usando banco de dados de emoções
Reis et al. (2018b)	Allen et al. (2016)	Perfil do estudante, texto e padrões de digitação	Modelo categórico	NLP, Coh-Metrix e Ferramenta SEANCE	Inferência de engajamento e tédio em sessões escrita com acurácia entre 76,5% e 77,3%
Reis et al. (2018b)	AlZoubi et al. (2012)	ECC, EMG, e GSR	Modelo categórico	kNN e Linear Bayes Normal Classifier	Kappa média de 0,25
Reis et al. (2018b)	Baker et al. (2014)	Logs de interação	Modelo categórico	Algoritmos JRip, J48, REPTree, Linear Regression)	Média Kappa de 0,25 e A' = 0.61
Morais et al. (2017)	Jaques et al. (2008)	Interação do estudante com o sistema- Modelo OCC	Modelo Categórico	Modelo BDI (<i>Belief, Desire and Intention</i>)	resultados indicam a efetividade do suporte afetivo do agente no desempenho dos estudantes
Morais et al. (2017)	de Melo et al. (2014)	Expressões faciais de uma base de imagens	Modelo Categórico (Ekman, 1992)	Algoritmo MLP, RBF, Redes Bayesianas	As três técnicas foram comparadas e identidou-se alta taxa de precisão para as redes Bayesianas
Morais et al. (2017)	Dantas et al. (2015)	Expressões faciais de uma base de imagens	Categórico (Ekman, 1992)	Árvore de decisão	resultados foram satisfatórios para todas as emoções, atingido acurácia geral de 86,4%

Tabela 2.8: Trabalhos correlatos selecionados no processo de revisão sistemática da literatura

Por fim, a **última questão** trata da avaliação dos principais resultados das pesquisas. Neste sentido, Observa-se que os resultados são amplamente variados com a acurácia variando de entre 0% e 96% e o índice Kappa entre 0,056 e 0,469. Estas variações refletem as diferenças de configuração dos experimentos, tipos de sensores, quantidade e tipo de emoções consideradas e técnicas de avaliação. Por, exemplo, o trabalho de Happy e Routray (2015) reporta taxas de acurácia de até 94,09% no reconhecimento de emoções básicas. Porém, estas taxas são obtidas usando banco de dados simulados com condições restritas de movimentos e iluminação adequada. Estas condições, conforme mostrado no trabalho de Gottardo e Pimentel (2016), diferem amplamente de um cenário real, impactando de forma significativa os resultados. Entretanto, um resultado recorrente indica que a fusão de sensores apresenta resultados melhores que sensores individuais.

2.8 Conclusão

Neste capítulo foram apresentadas as bases conceituais e teóricas que embasam a construção da presente pesquisa. Foram abordadas questões relativas ao conceito de computação afetiva e de emoções que definem técnicas e abordagens para o processo de inferência de emoções, que é o objetivo geral desta pesquisa.

O processo de inferência de emoções é a etapa inicial para a implementação de Sistemas Tutores Afetivos que sejam capazes de utilizar métodos de adaptação às reações afetivas individuais dos estudantes. Utilizando as inferências afetivas, estratégias específicas de intervenção pedagógica e feedbacks podem ser adotadas pelo ambiente computacional, buscando auxiliar o estudante a superar emoções que dificultam o aprendizado.

É importante reforçar as evidências apresentadas neste capítulo de que o processo de inferência de emoções baseado apenas na abordagem cognitivista acaba negligenciando as reações físicas individuais envolvidas no processo de geração das emoções. Da mesma forma, ao utilizar apenas elementos físicos, acaba-se desconsiderando o papel fundamental desempenhado pelo processo cognitivo de geração e controle das emoções no seres humanos.

Porém, considerando-se os trabalhos que compõem o estado da arte e que foram apresentados nesta seção, pode-se observar que nenhum trabalho usa uma abordagem híbrida, como proposto na presente pesquisa. Grande parte dos trabalhos atuais utiliza sensores para capturar informações relacionadas à reações físicas ou posturais. Por outro lado, um menor grupo de pesquisas, utiliza a abordagem cognitivista para o processo de inferência. Diante disso, esta é a principal lacuna observada nos trabalhos atuais e que o presente trabalho procura explorar.

Considerando todo o processo de busca de trabalhos correlatos, os únicos trabalhos que utilizam uma abordagem híbrida são os de Conati (2011, 2002). Porém, estes trabalhos não estão listados na Tabela 2.8 por apresentar apenas propostas de utilização de sensores fisiológicos (EMG - eletromiografia na testa) e informações cognitivas (não são apresentados resultados objetivos de experimentos com esta abordagem).

Outro diferencial relevante em relação aos trabalhos de Conati (2011, 2002) refere-se à abordagem utilizada para escolha do conjunto de emoções a ser tratado, bem como as técnicas de inferência a serem utilizadas. No trabalho de Conati (2011) é utilizada a técnica de inferência DDN (*Dynamic Decision Networks*) para inferência de emoções do modelo OCC como orgulho/vergonha e admiração/reprovação. Estas emoções são incomuns no contexto educacional (conforme será destacado no Capítulo 3) e o sensor EMG na face apresenta-se como altamente intrusivo e difícil de operacionalizar em ambientes reais de aprendizagem.

Uma limitação importante dos trabalhos correlatos que utilizam sensores fisiológicos relaciona-se ao fato da intrusividade e da difícil aplicação em cenários reais de aprendizagem.

Nota-se também que o modelo OCC (Ortony et al., 1990) foi o método utilizado nos trabalhos que usam a abordagem cognitivista para inferência das emoções. Por outro lado, o modelo de Ekman (Ekman, 1992) foi utilizado nos trabalhos que usam as expressões faciais como fonte de informação para a inferência das emoções.

Desta maneira, buscando-se explorar a possível complementariedade entre estes dois elementos, o modelo híbrido de inferência de emoções a ser apresentado no Capítulo 4 deste documento, apresenta componentes dedicados a aspectos físicos e cognitivos.

Capítulo 3

Conjunto de Emoções Relacionadas à Aprendizagem

A aplicação de técnicas de computação afetiva em ambientes computacionais de aprendizagem requer o tratamento de características e requisitos específicos do domínio educacional (Bosch et al., 2015; Baker et al., 2012). Trabalhos correlatos (e.g. (Baker et al., 2010; Paquette et al., 2016; Bosch e D’Mello, 2017)) têm dado ênfase às emoções relacionadas à aprendizagem. Todavia, Picard et al. (2004); Pour et al. (2010); Bosch et al. (2015) observam que existe um entendimento ainda limitado relativo a que emoções são as mais importantes no contexto educacional e como elas influenciam a aprendizagem.

Sendo assim, o conjunto de emoções a ser levado em consideração deve ser cuidadosamente avaliado e escolhido considerando as especificidades do domínio educacional (Baker et al., 2010). Neste capítulo será descrita a fundamentação e o método utilizado para escolha do conjunto de emoções relacionadas à aprendizagem a ser considerado nesta pesquisa.

3.1 Dimensões das emoções

Autores como Nicolaou et al. (2011) e Calvo e D’Mello (2012) defendem que a variabilidade entre as emoções pode ser representada de forma significativa por duas dimensões: Valência (*valence*) e Ativação (*arousal*). A dimensão Valência refere-se a quanto uma emoção é positiva ou negativa e varia de alegria até tristeza. Por outro lado, a dimensão Ativação refere-se ao nível de ativação ou apatia a que a emoção está relacionada, variando de sonolência até excitação frenética.

Em relação à importância da dimensão ativação, no trabalho de Picard (1997) encontra-se uma analogia para estudantes, considerando a frase comum entre os atletas que diz: “*no pain, no gain*”. A frase poderia ser traduzida como “sem dor ou esforço não há resultados ou vitórias”. Esta regra poderia ser aplicada para o aprendizado, embora, neste caso, dor ou esforço não se referem a aspectos físicos, mas intelectuais ou mentais (Picard, 1997). Desta forma, Picard (1997) defende que é possível aplicar ao contexto educacional a curva “U” invertida proposta na lei de Yerkes e Dodson (1908) (ver Figura 3.1) que relaciona desempenho com ativação (*arousal*).

Conforme pode ser visualizado na Figura 3.1, para atividades complexas (como é o caso da aprendizagem), o desempenho diminui nos níveis extremos (baixo e alto) de ativação. Em contrapartida, os melhores níveis de desempenho são obtidos com níveis intermediários de ativação. Nesta linha, Shen et al. (2007) validaram o relacionamento entre desempenho e ativação proposto na curva de Yerkes-Dodson, revelando que a ativação mantém-se relativamente estável em níveis intermediários durante um processo efetivo de aprendizagem.

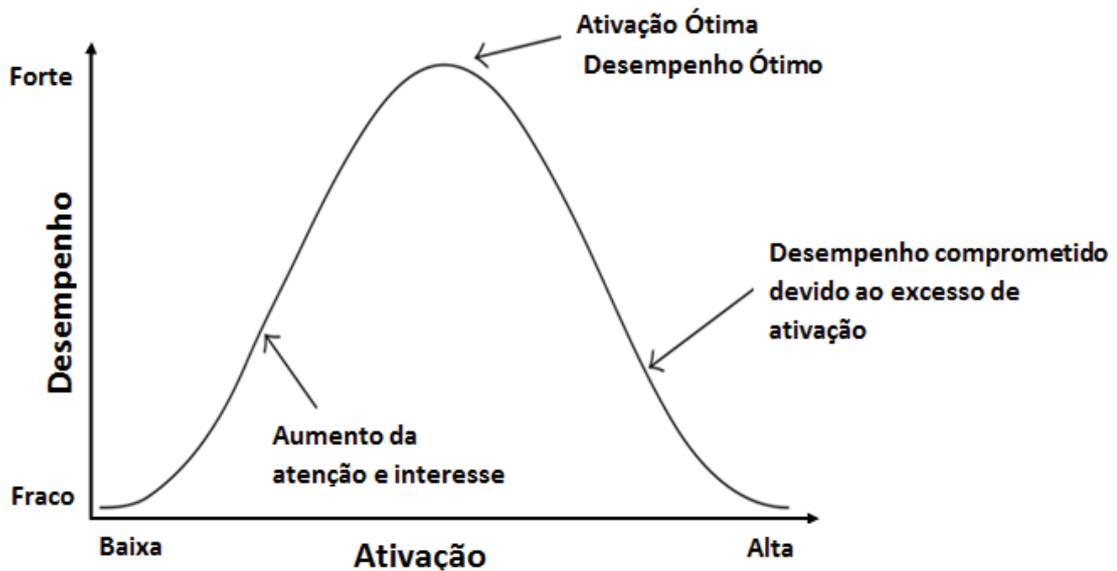


Figura 3.1: Curva clássica da lei de Yerkes-Dodson

Relativamente à dimensão valência, pesquisas como de McDaniel et al. (2007) e Baker et al. (2010) têm demonstrado que o estado cognitivo-afetivo confusão (valência negativa) apresenta correlação positiva com aprendizagem. Tomando como base a valência das emoções, Kort et al. (2001) definem um modelo de aprendizagem em espiral composto de quadrantes. Estes quadrantes representam os espaços no qual as emoções de um aprendiz movem-se dinamicamente durante a aprendizagem. Na Figura 3.2, é apresentado o modelo de Kort et al. (2001) em que o eixo horizontal representa a dimensão “valência”. Emoções com valência positiva estão à direita e negativas à esquerda. O eixo vertical é chamado de “Eixo da Aprendizagem” e simboliza construção do conhecimento na parte superior e descarte de conceitos errados na parte inferior.

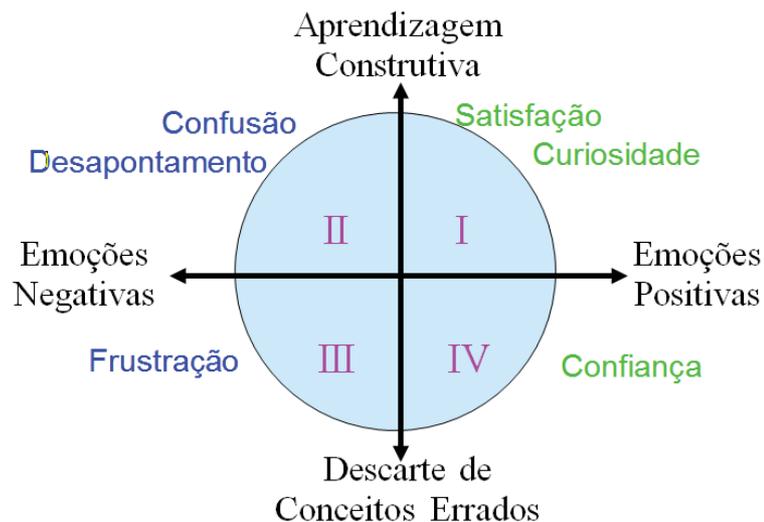


Figura 3.2: Modelo de aprendizagem em espiral. Fonte: Adaptado de Kort et al. (2001)

Conforme pode ser observado na Figura 3.2, a emoção confusão encontra-se em um quadrante (II) de aprendizado construtivo, mesmo sendo uma emoção com valência negativa. De acordo com Lisetti e Gmytrasiewicz (2002), emoções como surpresa, confusão, frustração e

satisfação podem auxiliar no acompanhamento do processo de aprendizado, tendo em vista que professores geralmente baseiam-se nestes aspectos para adaptar sua metodologia de ensino.

Considerando-se que o aprendizado naturalmente envolve emoções com valência negativa, como falhas ou erros, Kort et al. (2001) e Picard (1997) observam que uma estratégia eficiente de aprendizagem deve reconhecer que encontrar e superar momentos de confusão e frustração é uma parte natural do aprendizado. Entretanto, Picard (1997) observa que não existem definições relativas aos limites em que a frustração passa a ser prejudicial. Pour et al. (2010) observam que emoções como confusão, aborrecimento, frustração, curiosidade, interesse, surpresa e alegria têm emergido como altamente relevantes e influentes nas experiências de aprendizagem.

No entanto, tomando-se como referência trabalhos correlatos consultados, pode-se verificar que atualmente não existe o que poderia ser chamado “conjunto de emoções relevantes para a aprendizagem”. Este cenário pode ser observado analisando-se a Tabela 3.1 em que são apresentados trabalhos correlatos, juntamente com o conjunto de emoções considerado e o tipo de método de inferência destas emoções. Conforme pode-se observar nesta tabela, o número e o tipo de emoções variam entre os diferentes trabalhos.

Considerando-se os argumentos apresentados até aqui neste capítulo, entendeu-se que uma abordagem racional, eficiente e inovadora poderia ser a realização de inferência considerando não um conjunto específico de emoções, mas sim o agrupamento das emoções em quadrantes formados pelas dimensões valência e ativação.

Na sequência deste capítulo será apresentada a abordagem de representação das emoções em quadrantes e também o processo de mapeamento das emoções cognitivas e físicas para os quadrantes.

3.2 Quadrantes de emoções

Para viabilização da proposta de representação das emoções em quadrantes, foram considerados relevantes os trabalhos de Shen et al. (2009, 2007) e Baker et al. (2010) que apresentam modelos de mapeamento de emoções relacionadas à aprendizagem. Os modelos de mapeamento utilizados nestes dois trabalhos usam como referência o “modelo circunplexo” de Russel (1980) (ver Figura 3.6) e o “modelo de aprendizagem em espiral” de Kort et al. (2001) (ver Figura 3.2). Estes dois modelos apresentam o mapeamento de emoções relacionadas a aprendizagem em um espaço dimensional formado pelas dimensões valência e ativação. Esta escolha considerou o fato que o “modelo circunplexo” e o “modelo de aprendizagem em espiral” demonstram-se consolidados e repetidamente referenciadas em trabalhos correlatos como (Posner et al., 2005; D’Mello et al., 2007; Baker et al., 2010; Pour et al., 2010; Conati, 2011).

Na Figura 3.3 é apresentado o modelo de Shen et al. (2009) em que as principais emoções relacionadas à aprendizagem são mapeadas para um espaço dimensional valência (eixo horizontal) e ativação (eixo vertical). Emoções relacionadas a aprendizagem são posicionadas nesta figura, considerando-se o valor das dimensões valência e ativação de cada emoção. De maneira análoga, a Figura 3.4 apresenta o modelo de mapeamento de emoções relevantes para a aprendizagem proposto por Baker et al. (2010), que também se baseia nas dimensões valência e ativação.

Desta forma, na Figura 3.5 é apresentada a abordagem a ser utilizada neste trabalho para representação das emoções relacionadas à aprendizagem. Nesta abordagem as dimensões Valência e Ativação são utilizadas para a representação das emoções nos Quadrantes “Q1”, “Q2”, “Q3” e “Q4”. Foi atribuído um nome representativo para cada um dos quadrantes e também as

Tabela 3.1: Conjunto de emoções tratadas em trabalhos correlatos

Trabalho	Tipo de método de inferência	Conjunto de emoções tratadas
Baker et al. (2010)	Físico	Tédio, frustração, confusão, concentração, alegria e surpresa
Bosch et al. (2014)	Físico	Engajamento, aborrecimento, confusão, frustração e neutro
Conati (2011)	Cognitivo	alegria/tristeza, admiração/reprovação, orgulho/vergonha
Dantas et al. (2015)	Físico	Alegria, surpresa, raiva, medo, desgosto, tristeza e neutra
D’Mello et al. (2007)	Cognitivo e Físico	Tédio, confusão, frustração, encantamento, engajamento e neutro
Jaques et al. (2014)	Físico	Tédio e curiosidade
Jaques (2004)	Cognitivo	Alegria, angústia, satisfação, desapontamento, gratidão, raiva e vergonha
Jaques et al. (2003)	Cognitivo	Satisfação e Frustração
Jaques et al. (2008)	Cognitivo	Satisfação, frustração, alegria, tristeza, gratidão, raiva, vergonha
McDaniel et al. (2007)	Físico	aborrecimento, confusão, encantamento, engajamento, frustração, surpresa e neutro
Paquette et al. (2014)	Cognitivo	Tédio, confusão, frustração, concentração e “?” (indefinido)
de Melo et al. (2014)	Físico	Alegria, tristeza, raiva, desgosto, medo e surpresa
Shen et al. (2007)	Físico	Interesse, engajamento, confusão, frustração, aborrecimento, confiança
Shen et al. (2009)	Cognitivo e Físico	Engajamento, confusão, aborrecimento, confiança
Woolf et al. (2009)	Físico	Alegria, raiva, surpresa e medo
DeFalco et al. (2018)	Físico	Tédio, confusão, frustração, concentração engajada e surpresa
Jaques et al. (2011)	Cognitivo	Alegria/satisfação, angústia/desapontamento, gratidão e raiva
Bosch e D’Mello (2017)	Avaliação do estudante	Raiva, ansiedade, tédio, confusão, curiosidade, desgosto, medo, frustração, engajamento, alegria, tristeza, surpresa e neutro
Bosch et al. (2015)	Físico	Tédio, confusão, encanto, engajamento e frustração
Muñoz et al. (2010)	Cognitivo	Alívio antecipado, esperança, neutro, ansiedade, alegria antecipada, desespero



Figura 3.3: Mapeamento de um conjunto de emoções relacionadas à aprendizagem para o modelo dimensional valência x ativação: Fonte: Adaptado de Shen et al. (2007)

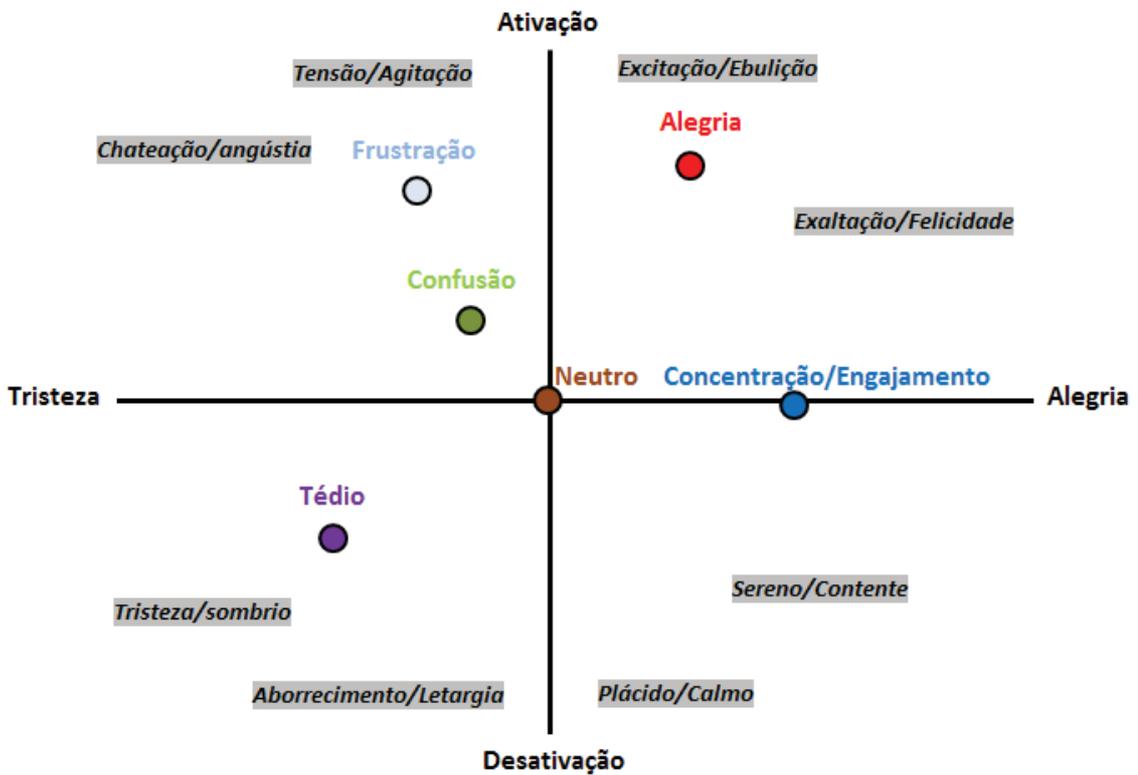


Figura 3.4: Mapeamento de um conjunto de emoções relacionadas à aprendizagem para o modelo dimensional valência x ativação. Fonte: Adaptado de Baker et al. (2010)

principais emoções contidas no mesmo considerando trabalhos correlatos (Russel, 1980; Kort et al., 2001; Shen et al., 2009; Baker et al., 2010).

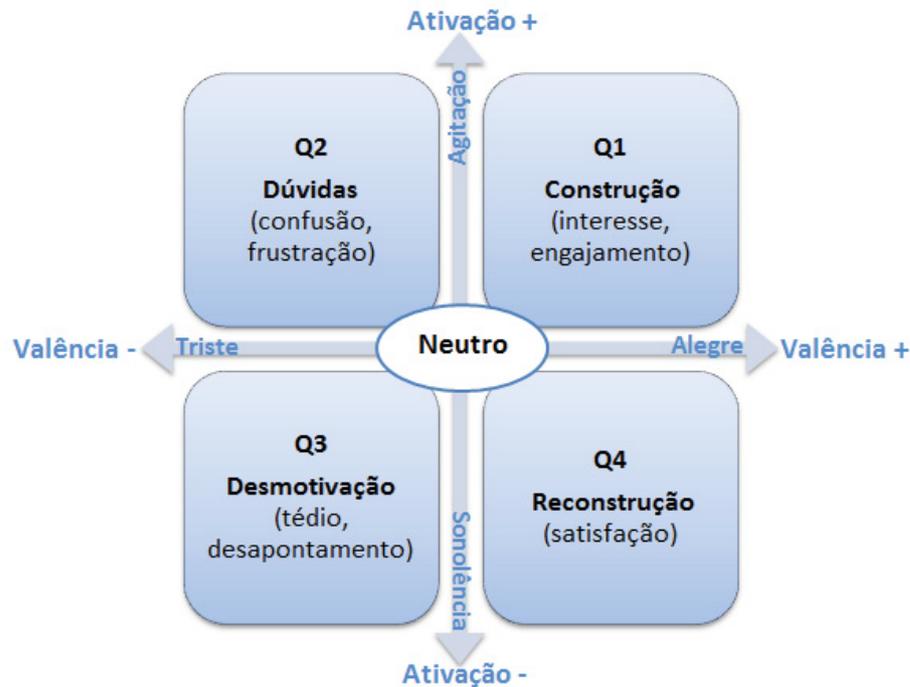


Figura 3.5: Abordagem de representação das emoções em quadrantes

Entende-se que com a abordagem em quadrantes, além do embasamento em propostas consolidadas na área de pesquisa (Russel, 1980; Shen et al., 2009; Baker et al., 2010; Kort et al., 2001), considera-se e diferencia-se através dos quadrantes os principais aspectos das emoções relacionadas à aprendizagem. Corroborando com esta proposta, Kort et al. (2001) observam que professores adaptam-se para assistir aos estudantes baseados em um pequeno conjunto de reações afetivas em oposição a um grande número de fatores complexos.

Levando-se em conta o exposto acima, entendeu-se que a abordagem utilizada para escolha do conjunto de emoções a serem tratadas esteja perfeitamente alinhada com os objetivos e justificativa desta pesquisa. Mesmo não sendo tratadas emoções individualmente, considera-se que a inferência dos quadrantes, pode ser uma informação importante para detectar estados como ciclo vicioso ou engajamento/desengajamento e subsidiar ações de adaptação e estratégias de intervenção.

Conforme já destacado nos capítulos anteriores, neste trabalho busca-se inferir o conjunto de emoções relacionadas à aprendizagem utilizando uma abordagem híbrida através da fusão de reações físicas (expressões faciais) e cognitivas (modelo OCC). No entanto, nenhum destas duas abordagens oferece uma forma de inferir diretamente o conjunto de emoções relacionadas a aprendizagem apresentados na Figura 3.5. Desta forma, na sequência deste capítulo, serão apresentadas propostas de mapeamento do conjunto de emoções disponíveis no modelo OCC e expressões faciais para a abordagem de quadrantes descrita acima.

3.3 Conjunto de emoções baseadas em reações físicas - expressões faciais

O processo de inferência de emoções usando expressões faciais pode ser realizado utilizando-se métodos criados para categorizar sistematicamente a expressão física das emoções, como o FACS (ver seção 2.4). A utilização do método FACS implica que as emoções inferidas normalmente representam o conjunto das sete emoções chamadas de “básicas” (raiva, desgosto, medo, alegria, tristeza, surpresa e desprezo) (Ekman, 1992). Pode-se observar que o conjunto de emoções relacionadas à aprendizagem destacadas na Figura 3.5 difere significativamente do conjunto das sete emoções básicas. Esta diferença deve-se ao fato de que recentes trabalhos (e.g. (Reis et al., 2018a; Baker et al., 2010; D’Mello, 2012)) têm identificado um conjunto específico de emoções que são relevantes para o domínio educacional e que diferem do conjunto de emoções básicas de Ekman (1992).

Desta forma, para realização desta pesquisa, inicialmente buscou-se verificar se existem alternativas viáveis para inferência de emoções relacionadas à aprendizagem diretamente a partir das expressões faciais. No entanto, não foram encontradas propostas consolidadas e disponíveis para esta tarefa. Corroborando com esta conclusão, Bosch et al. (2015) observam que ainda não está claro se emoções relacionadas a aprendizagem podem ser detectadas com fidelidade semelhante às emoções básicas, para as quais os vínculos entre emoção e expressão facial foram cuidadosamente mapeados por décadas. Bosch et al. (2015) completam que inexistem métodos semelhantes ao FACS para emoções centradas na aprendizagem, sendo esta uma questão totalmente aberta.

Neste contexto, mesmo considerando-se uma possível limitação, optou-se por realizar a inferência das emoções básicas e posteriormente mapeá-las para o conjunto de emoções relacionadas à aprendizagem distribuídas em quadrantes (ver Figura 3.5). Para realizar o referido mapeamento cada uma das emoções básicas foi representada utilizando o modelo dimensional valência e ativação. É importante observar que a abordagem de quadrantes proposta permite facilmente integrar possíveis futuras pesquisas que façam a inferência direta das emoções relacionadas à aprendizagem usando expressões faciais.

A proposta chamada de “Modelo Circumplexo do Afeto” de Russel (1980) é um exemplo de modelo que representa as emoções básicas usando a abordagem dimensional. Baseando-se no modelo de Russel (1980), trabalhos como de Posner et al. (2005) e Zagalo et al. (2005) foram desenvolvidos, identificando áreas no espaço dimensional como rótulos que representam as emoções básicas. Na Figura 3.6 é apresentado o modelo descrito em Posner et al. (2005) que se baseia no “Modelo Circumplexo do Afeto” de Russel (1980). No círculo externo desta figura foram inseridos rótulos (com fundo cinza) contendo as sete emoções básicas de Ekman (1992) mapeadas para o espaço dimensional valência (eixo horizontal) x ativação (eixo vertical).

Nos trabalhos de Russel (1980) e Posner et al. (2005) somente são consideradas as seis emoções básicas inicialmente descritas por Ekman (1992), não incluindo, portanto, a emoção desprezo. Para o mapeamento desta última emoção foi tomado como referência o trabalho de Zagalo et al. (2005) que também usa como referência Russel (1980) e mapeia desprezo como uma emoção de valência negativa e ativação positiva. Além disso, para atender aos requisitos desta pesquisa, foi inserido na Figura 3.6 o estado central neutro.

Desta forma, foi possível mapear as emoções básicas para a abordagem de quadrantes proposta neste trabalho (ver Figura 3.5). A Figura 3.7 apresenta uma representação deste mapeamento em que cada uma das sete emoções básicas são mapeadas para o respectivo quadrante do conjunto de emoções relacionadas à aprendizagem.

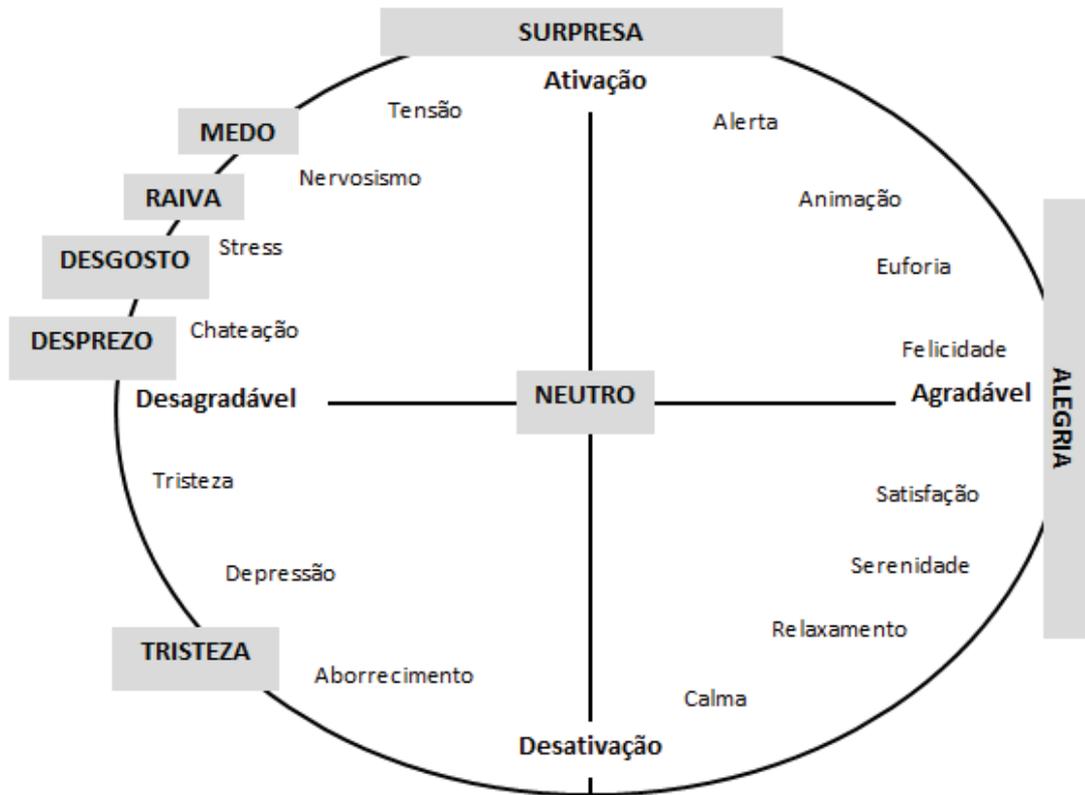


Figura 3.6: Representação gráfica do modelo circunplexo de afeto. Fonte: Adaptado de Posner et al. (2005)

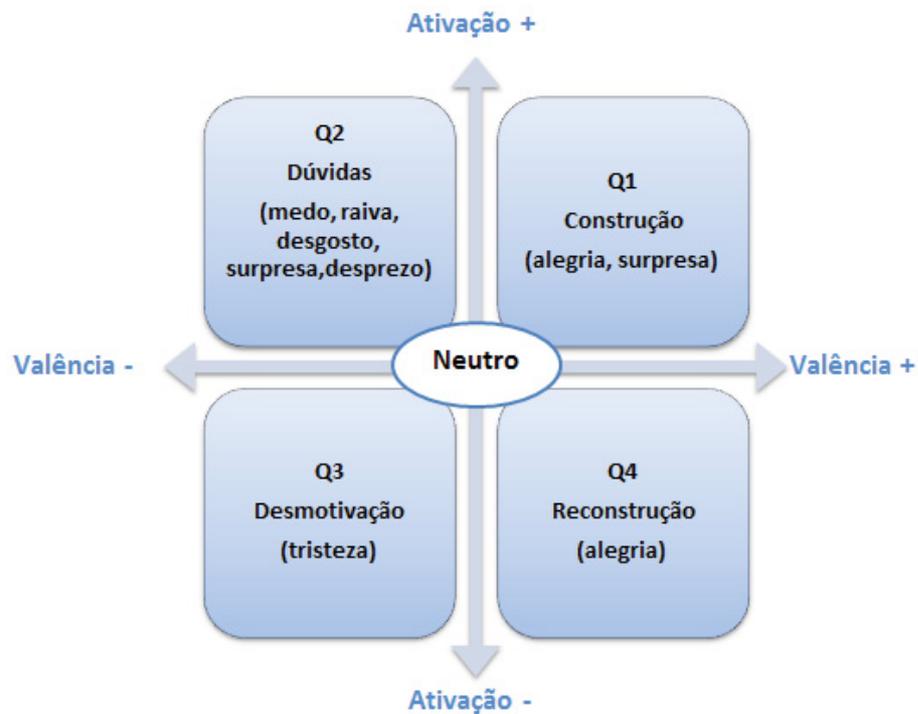


Figura 3.7: Mapeamento das sete emoções básicas para o espaço dimensional valência x ativação

É fundamental observar na Figura 3.7 e 3.6 que as emoções “alegria” e “surpresa” aparecem repetidas em dois quadrantes, pois podem ter alta variabilidade na dimensão ativação ou valência, respectivamente. Entendeu-se que é possível resolver esta ambiguidade considerando adaptações na implementação da inferência via expressões faciais.

Para a emoção “surpresa” foi verificado se o evento ocorrido no ambiente tem valência positiva ou negativa (e.g. acerto ou erro de uma questão). Se o evento tiver valência positiva, classificou-se “surpresa” no Quadrante 1 e, caso contrário, no Quadrante 2.

Para a emoção alegria, foi criada uma medida de intensidade relativa desta emoção. Esta medida considera a relação da emoção “alegria” em comparação com a soma das demais emoções. Caso a alegria tenha intensidade maior que a soma das demais emoções então foi classificada no Quadrante 1 e, caso contrário, no Quadrante 4.

3.4 Conjunto de emoções baseadas em informações cognitivas

O modelo OCC original (ver Figura 2.5) define um conjunto de vinte e duas emoções. Contudo, conforme observa Conati (2011), nem todas as emoções são sempre relevantes para todos os ambientes de aprendizagem. Mesmo quando uma emoção específica é potencialmente relevante, a inclusão no modelo deve ser avaliada considerando sua importância e também o custo de modelar a dinâmica que gera esta emoção. Como exemplo, (Jaques, 2004) utiliza um conjunto de sete emoções enquanto que Conati (2011) utiliza apenas seis das vinte e duas emoções do modelo OCC.

Avaliando-se o contexto e os objetivos deste trabalho, verificou-se que as emoções mais relevantes do modelo OCC são aquelas relacionadas a consequência de eventos que têm impacto no próprio agente (e.g. estudante). Desta maneira, analisando-se a estrutura do modelo OCC, foram selecionadas as seguintes oito emoções: Alegria, Angústia, Desapontamento, Alívio, Esperança, Medo, Satisfação e Medo Confirmado.

As emoções Gratificação, Remorso, Raiva e Gratidão do modelo OCC não foram incluídas por serem emoções mistas que envolvem, além da consequência de eventos, as ações de outros agentes. Além disso, em trabalhos correlatos (ver Tabela 3.1) não é comum a inclusão deste conjunto de emoções. Para melhorar ilustrar e destacar o conjunto de emoções do modelo OCC a ser considerado neste trabalho, a Figura 3.8 representa o modelo OCC simplificado a ser utilizado nesta pesquisa.

Para viabilizar a abordagem de utilização dos quadrantes de emoções relacionadas a aprendizagem (ver Figura 3.5), o conjunto de emoções escolhidos no modelo OCC precisa ser mapeado para um destes quadrantes. Neste sentido, o trabalho de Gebhard (2005) apresenta o mapeamento de todo o conjunto de emoções do modelo OCC para a abordagem dimensional valência e ativação. A Tabela 3.2 apresenta a lista das oito emoções do modelo OCC consideradas relevantes no âmbito deste trabalho, com os respectivos valores das dimensões valência e ativação.

Tomando-se como referência a Tabela 3.2, realizou-se o mapeamento das oito emoções selecionadas do modelo OCC com a abordagem baseada em quadrantes, conforme pode ser observado na Figura 3.9. Nesta figura, além do estado Neutro, foram colocadas cada uma das oito emoções do modelo OCC no quadrante correspondente aos valores das dimensões Valência e Ativação.

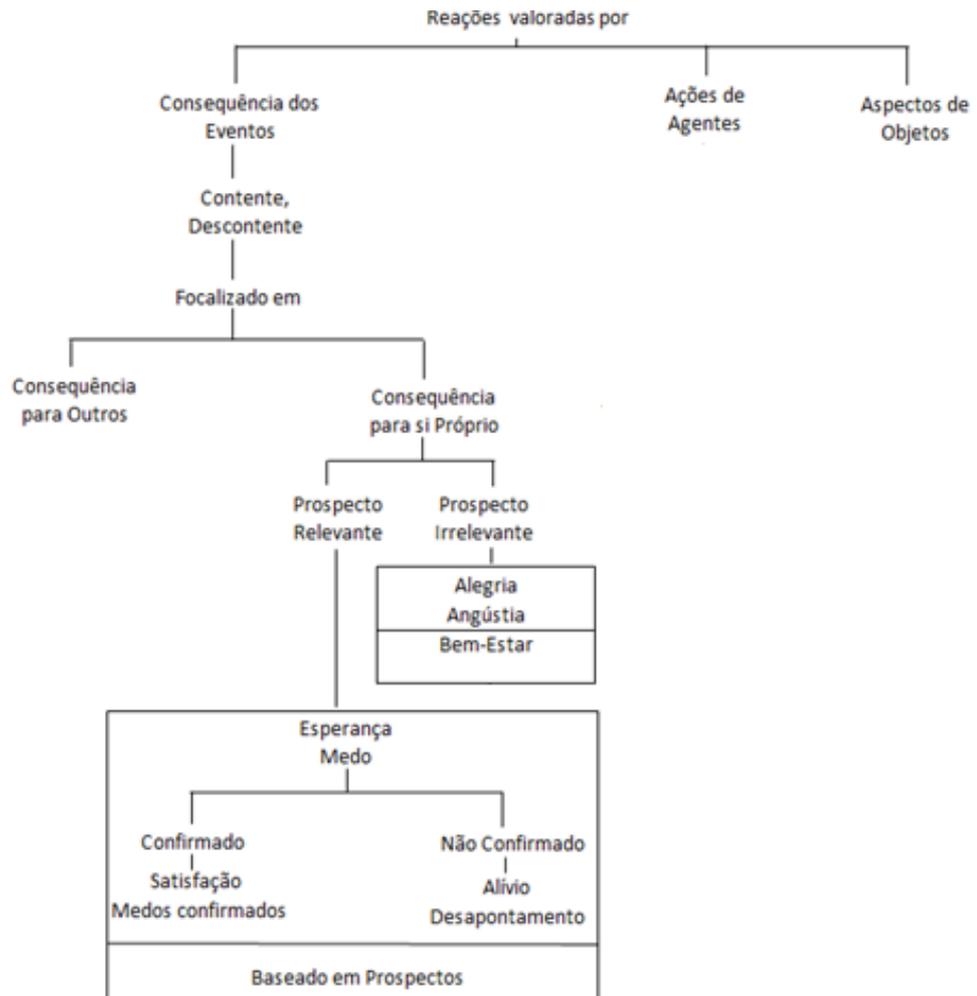


Figura 3.8: Componentes do modelo OCC considerados relevantes para este trabalho

Tabela 3.2: Mapeamento da Emoções do modelo OCC para o espaço dimensional. Fonte: Adaptado de (Gebhard, 2005)

Emoção	Valência	Ativação
Alegria	0,4	0,2
Alívio	0,2	-0,3
Angústia	-0,4	-0,2
Desapontamento	-0,3	0,1
Esperança	0,2	0,2
Medo	-0,64	0,6
Medo Confirmado	-0,5	-0,3
Satisfação	0,3	-0,2

3.5 Conclusão

Neste capítulo foram abordadas questões importantes relacionadas ao método de escolha do conjunto de emoções a serem consideradas no modelo de inferência. Este processo foi realizado norteando-se em pilares fundamentais, tais como: i) considerar as peculiaridades do domínio educacional, ii) construir uma proposta coesa, objetiva, implementável computacionalmente e

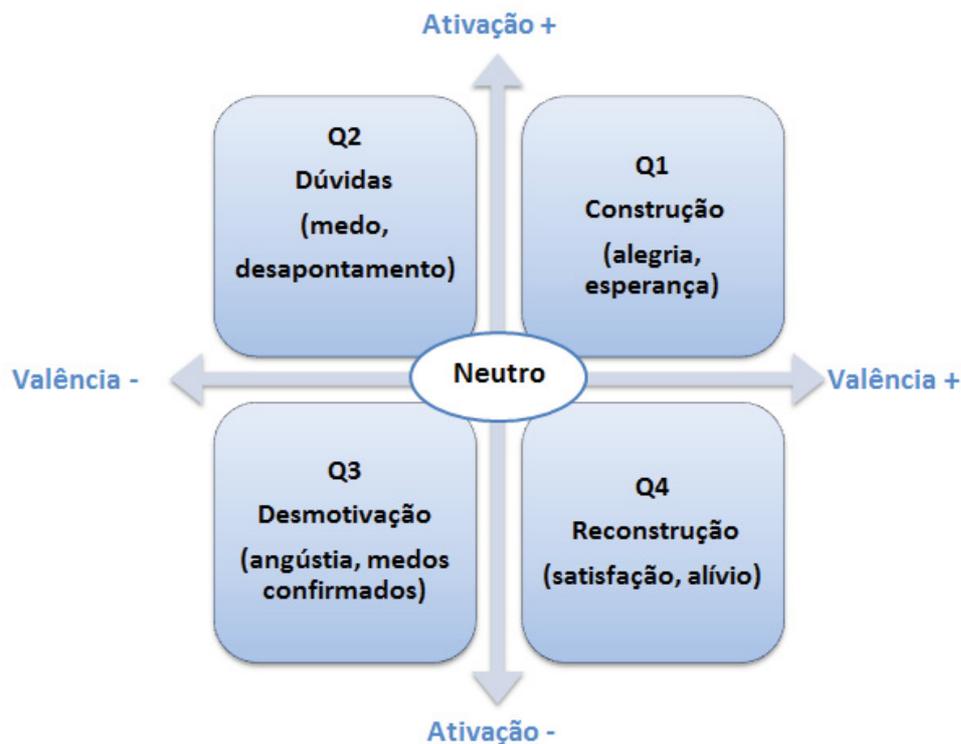


Figura 3.9: Relação entre emoções no modelo OCC e quadrantes das dimensões valência e ativação

embasada em aspectos teóricos consolidados, iii) possibilitar a fusão de modelos de emoções baseados em reações físicas e também aspectos cognitivos e iv) considerar a aplicabilidade das informações inferidas para o desenvolvimento de funcionalidades de adaptação em ambientes computacionais de aprendizagem.

Neste contexto, verificou-se que em trabalhos correlatos atuais não há um conjunto totalmente consolidado de emoções a serem incluídas em modelos de inferência aplicados ao domínio educacional. Desta forma, neste trabalho foi proposta uma abordagem baseada em quadrantes formados pelas dimensões valência e ativação. Cada quadrante agrega um conjunto de emoções que tem potencial de impactar o processo de aprendizagem.

A abordagem em quadrantes tem como fundamentação teórica dois modelos conceituais aplicáveis em processos de aprendizagem: modelo circumplexo de Russel (1980) e modelo de aprendizagem em espiral de Kort et al. (2001). Desta forma, buscou-se a construção de uma proposta coesa, objetiva e fundamentada em teorias consolidadas e testadas.

Para tornar viável a utilização de uma estratégia híbrida de inferência e fusão, foi apresentado um esquema de mapeamento de emoções físicas e cognitivas para a abordagem baseada em quadrantes. Este mapeamento permite a implementação computacional da proposta desta pesquisa.

Por fim, é importante ressaltar a possível aplicabilidade das inferências obtidas nesta pesquisa. Como os quadrantes representam situações com potencial impacto (positivo ou negativo) direto na aprendizagem, um sistema tutor afetivo poderia utilizar estas informações para detectar, por exemplo, ciclo vicioso, engajamento/desengajamento e implementar intervenções tutoriais adequadas para esta situação.

Capítulo 4

Modelo Híbrido de Inferência de Emoções

Neste capítulo será apresentado e discutido em detalhes o processo de concepção, organização e implementação de um Modelo Híbrido de Inferência de Emoções - ModHEmo. Este modelo servirá como base para realização dos experimentos e avaliação dos resultados desta pesquisa em um ambiente real de aprendizado.

4.1 Arquitetura do Modelo Híbrido de Inferência de Emoções

O processo de inferência de emoções proposto neste trabalho tem sua implementação organizada na forma de um modelo híbrido. A arquitetura do ModHEmo apresenta como característica principal a divisão do processo de inferência em três componentes: cognitivo, físico e fusão. Os componentes cognitivo e físico realizam paralelamente a inferência das emoções usando como entrada os eventos do ambiente computacional e imagens da face, respectivamente. Na sequência as inferências dos dois componentes são combinadas usando técnicas de classificação para, finalmente, obter-se a inferência final representada pelos quadrantes (ver Seção 3.2).

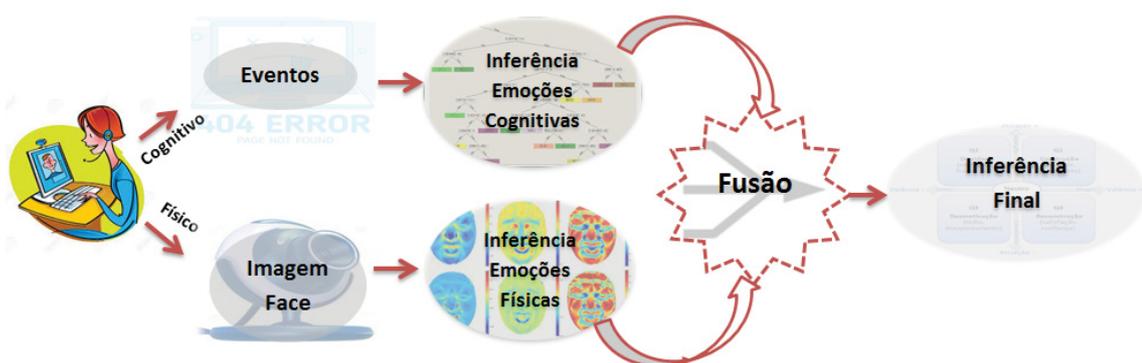


Figura 4.1: Esquema resumido da organização do ModHEmo

A Figura 4.1 apresenta um esquema resumido da organização do processo de inferência do ModHEmo. O processo de inferência inicia-se (esquerda da figura) com a obtenção dos eventos do software educacional e imagens da face do estudante. Com estas informações são realizadas as inferências das emoções cognitivas e físicas que, na sequência, são combinadas para obtenção do resultado final (direita da figura).

Na Figura 4.2 é mostrada a arquitetura detalhada da estrutura do ModHEmo, incluindo seus componentes e subcomponentes além das principais teorias e técnicas computacionais

utilizadas. Os componentes principais são: I)cognitivo, II)físico e III)fusão. Para facilitar a descrição dos subcomponentes do modelo foram inseridas na Figura 4.2 as letras de ‘A’ até ‘K’. Estas letras estão posicionadas próximas de cada subcomponente e serão referenciadas a seguir no detalhamento do ModHEmo.

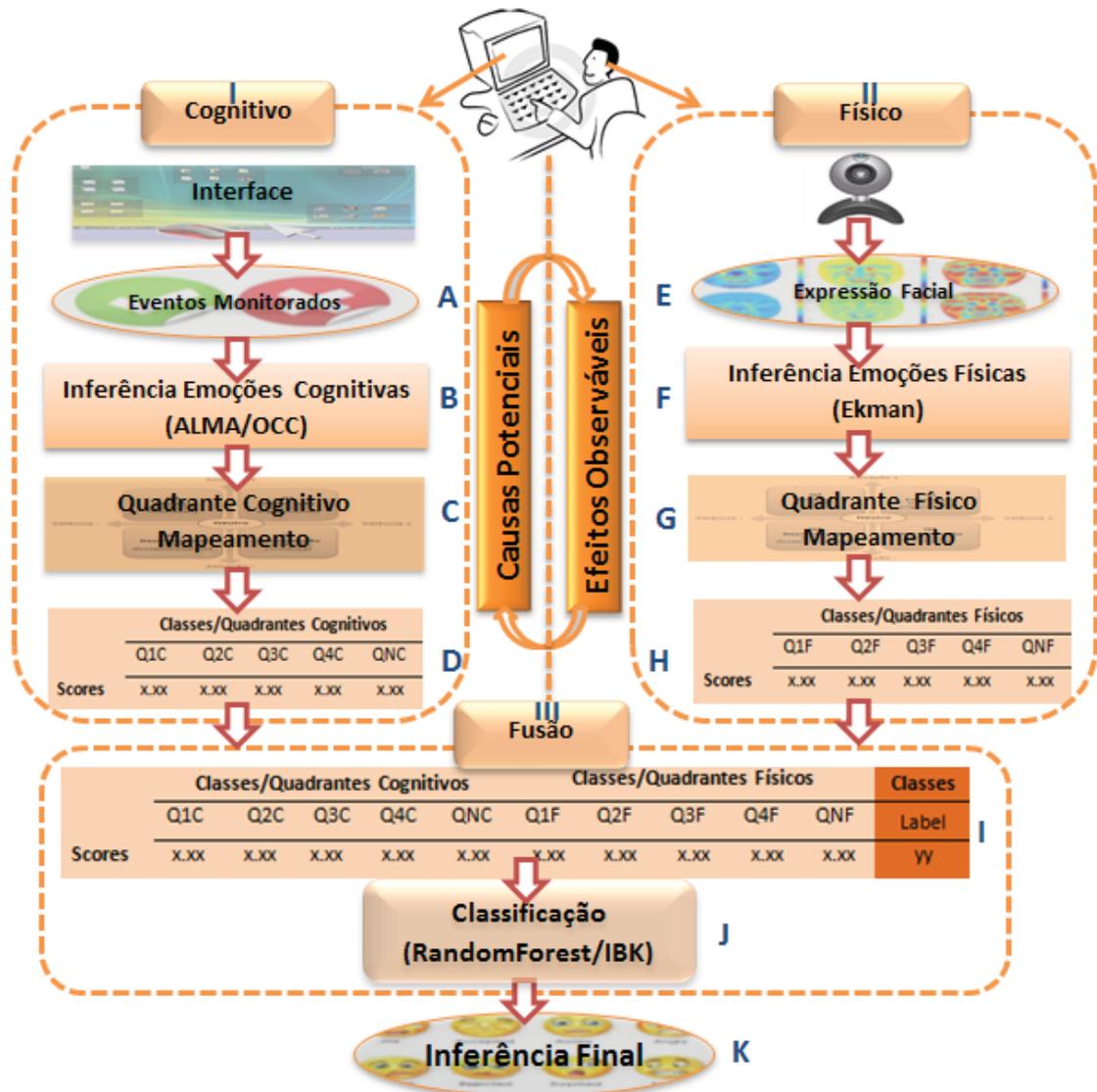


Figura 4.2: Esquema detalhado do Modelo Híbrido de Inferência de Emoções - ModHEmo

4.1.1 Componente Cognitivo

O componente cognitivo do ModHEmo é responsável por tratar os eventos relevantes no ambiente computacional ('A') que são indicativos das possíveis causas das emoções. O processo de inferência de emoções cognitivas ('B'), baseado na teoria OCC (Ortony et al., 1990), retorna valores no intervalo [0..1] para cada uma das oito emoções cognitivas consideradas neste trabalho (Alegria, Angústia, Desapontamento, Alívio, Esperança, Medo, Satisfação e Medo Confirmado). Este processo de inferência ('B') será apresentado com mais detalhes na Seção 4.2.

Após a inferência das emoções cognitivas, é realizado um mapeamento dos *scores* das emoções ('C') para os respectivos quadrantes, observando a distribuição das emoções cognitivas

nos quadrantes mostrados na Figura 3.9. Após este mapeamento cada quadrante cognitivo terá um valor formado pela soma dos valores individuais das emoções que compõem cada quadrante.

Considerando que cada emoção cognitiva pode ter valores no intervalo [0..1], foi realizado um processo de normalização ('D') para que a soma dos quadrantes seja igual a 1. Com esta normalização, os valores dos quadrantes cognitivos tornam-se compatíveis com os quadrantes físicos, viabilizando o processo de fusão. A função *Softmax* (Kuncheva, 2004) mostrada na Equação 4.1 foi o método usado para normalizar os valores do componente cognitivo do ModHEmo. Nesta equação, $g_1(x), \dots, g_c(x)$ são valores de cada quadrante do componente cognitivo do ModHEmo. Então, na Equação 4.1 os valores são calculados por $g'_1(x), \dots, g'_c(x), g'_j(x) \in [0, 1], \sum_{j=1}^C g'_j(x) = 1$. Desta forma o valor de cada quadrante fica no intervalo [0,1] e a soma de todos os quadrantes é igual a 1.

$$g'_j(x) = \frac{\exp\{g_j(x)\}}{\sum_{K=1}^C \exp\{g_K(x)\}} \quad (4.1)$$

4.1.2 Componente Físico

O componente físico do ModHEmo trata as reações observáveis, usando imagens da face dos estudantes ('E'), capturadas após (próximo *frame*) a ocorrência de um evento monitorado ('A') na interface do ambiente computacional. O componente físico realiza a inferência das emoções físicas ('F') e retorna *scores* para cada uma das oito emoções do componente físico consideradas nesta pesquisa (raiva, desgosto, medo, alegria, tristeza, surpresa, desprezo e neutro). Este processo de inferência ('F') será apresentado com mais detalhes na Seção 4.3.

Após a inferência das emoções físicas, é realizado um mapeamento dos *scores* das emoções ('G') para os respectivos quadrantes, observando a distribuição das emoções físicas nos quadrantes mostrados na Figura (3.7). Após este mapeamento cada quadrante físico terá um valor formado pela soma dos valores individuais das emoções que compõem cada quadrante. O componente físico já retorna *scores* normalizados no intervalo [0,1] de forma que a etapa 'H' é obtida diretamente da etapa 'G' sem a necessidade de aplicação de função de normalização (e.g. Softmax).

4.1.3 Componente Fusão

O processo de fusão dos componentes físico e cognitivo inicia-se com a criação de um conjunto de dados único ('I') contendo os *scores* dos quadrantes cognitivo e físico. Sendo assim, este conjunto de dados contém 10 atributos (5 físicos + 5 cognitivos) com *scores* dos quadrantes (mais Neutro). A classe (último atributo letra 'I') obtida através do processo de rotulagem (a ser descrito na seção 5.2.2) também faz parte do conjunto de dados. Com base neste conjunto de dados, foram treinados e testados dois algoritmos de classificação ('J') que realizam a inferência do resultado final do modelo ('K'). A inferência final é dada na forma de um dos quadrantes (Q1,Q2,Q3,Q4) ou neutro (QN) representando o grupo de emoções inferidas.

Cabe ressaltar que a implementação do processo de fusão utilizando classificadores, como utilizada neste trabalho, considera a possibilidade de falha ou indisponibilidade de dados dos componentes a serem fundidos. Isto se dá devido ao fato que na base de treinamento destes classificadores esta situação é observada e então o modelo é criado considerando estas condições. No contexto desta pesquisa, as informações das expressões faciais podem mais facilmente ficar indisponíveis. Isto poderia acontecer, por exemplo, em situações de posicionamento inadequado do estudante frente à câmera ou deficiência na iluminação.

Na sequência deste capítulo, será descrito o conjunto de técnicas identificadas como adequadas para a construção do processo de inferência de emoções cognitivas ('B') e físicas ('F'). Será feita também uma descrição detalhando as técnicas utilizadas no componente fusão (III)

4.2 Inferência de emoções cognitivas

Modelos cognitivos de emoções como o OCC (Ortony et al., 1990) baseiam-se em variáveis para controlar o processo de ativação das emoções. O trabalho de Gratch et al. (2009) apresenta um estudo do conjunto de variáveis mais utilizadas nos principais modelos cognitivos. O referido trabalho aponta que as variáveis desejabilidade (*desirability*) e prospecto (*prospect*) são as duas mais relevantes e amplamente utilizadas nos modelos cognitivos pesquisados.

Especificamente considerando o modelo OCC, estas duas variáveis constituem-se como componentes fundamentais para a geração e diferenciação das emoções. Desta maneira, as regras de ocorrência das emoções no modelo OCC utilizam estas variáveis como condição de disparo. Na Tabela 4.1 são descritas as condições ou regras de elicitación – EECs (*Emotion Eliciting Condition*) para cada uma das oito emoções utilizadas nesta pesquisa. Nesta tabela os termos desejável/indesejável referem-se à variável desejabilidade, enquanto confirmação/desconfirmação referem-se à variável expectativa.

Tabela 4.1: Visão geral das condições de elicitación das emoções no modelo OCC. Fonte: Ortony et al. (1990)

Condição ou Regras de Elicitación das emoções (EEC)	Emoção
(contente com) um evento desejável	Alegria
(descontente com) um evento indesejável	Angústia
(contente com) a expectativa de um evento desejável	Esperança
(descontente com) a expectativa de um evento indesejável	Medo
(contente com) a confirmação da expectativa de um evento desejável	Satisfação
(descontente com) a confirmação da expectativa de um evento indesejável	Medo Confirmado
(contente com) a desconfirmação da expectativa de um evento indesejável	Alívio
(descontente com) a desconfirmação da expectativa de um evento desejável	Desapontamento

Considerando as definições apresentadas na Tabela 4.1, as chamadas Condições de Elicitación de Emoções – EECs podem ser abstraídas. Desta forma, o modelo OCC pode ser implementado computacionalmente na forma de regras, que consideram as condições de elicitación para inferir determinada emoção (Picard, 1997). Existem propostas como Affective Reasoner (Elliott, 1997), extensões do modelo Belief-Desire-Intention - BDI (Bratman, 1990) e ALMA (Gebhard, 2005) que apresentam a implementação computacional do modelo OCC. No contexto desta pesquisa, para implementação computacional do componente cognitivo do ModHEmo será tomado como base o modelo ALMA (Gebhard, 2005) que será descrito a seguir. Esta escolha deveu-se principalmente ao fato de que o ALMA ser um modelo mais recente e disponibilizar o código fonte e licença que permite seu uso e customização.

4.2.1 Modelo ALMA

A abordagem cognitiva do modelo OCC é utilizada como base para implementação do modelo ALMA - *A Layered Model of Affect* (Gebhard, 2005). Este modelo tem como objetivo a implementação de atores virtuais afetivos em ambientes multiagentes. No entanto, os módulos que atuam no processo de raciocínio e inferência de emoções foram considerados particularmente relevantes no contexto deste trabalho.

Além da inferência de uma emoção específica, o modelo ALMA permite calcular a intensidade das emoções. Este cálculo pode ser feito setando-se, de forma global ou por evento, os parâmetros *intensity* (intensidade) e *likelihood* (probabilidade). Os valores possíveis para estas variáveis são no intervalo [0..1]. Em trabalhos correlatos que usam o modelo OCC (e.g. Conati (2011); Jaques (2004)) sugere-se que os valores para intensidade sejam obtidos usando sensores fisiológicos como, por exemplo, EEG, GSR, EMG, etc. Como neste trabalho não são utilizados este tipo de sensores, optou-se por setar a variável intensidade ¹ de forma estática por evento. Para isto foi considerado o possível impacto de cada evento no contexto do software educacional e, por consequência, no atingimento do objetivo do estudante. A definição da intensidade dos eventos é dependente do software educacional necessitando assim de ajustes para cada situação, a não ser que seja utilizada estratégia automática para obtenção deste parâmetro. Os valores da intensidade de cada evento para o software educacional utilizado neste trabalho serão apresentados na Tabela 5.1 do Capítulo 5.

Entende-se que a forma estática usada para setar a intensidade dos eventos pode representar uma possível limitação. Contudo, no âmbito desta pesquisa o impacto desta limitação tende a ser reduzido pois o componente físico do ModHEmo inclui a intensidade das emoções. Adicionalmente, a inferência final consiste em um rótulo (*label* dos quadrantes) que já considera os valores da dimensão intensidade.

O modelo ALMA possibilita ajustar a intensidade das emoções baseando-se no perfil de personalidade, conforme o modelo '*Big Five Personality Traits*' (McCrae e John, 1992). Como não estavam disponíveis informações sobre a personalidade dos estudantes, optou-se por desabilitar este ajuste, setando-se o valor zero para os cinco traços de personalidade do modelo '*Big Five*' (*openness, conscientiousness, extraversion, neuroticism, agreeableness*). Desta forma, neste trabalho a intensidade das emoções não é influenciada pela personalidade.

Em relação à variável probabilidade (*likelihood*) verificou-se que na implementação do modelo ALMA esta variável não tem interferência no processo de inferência do conjunto das emoções cognitivas consideradas neste trabalho. Esta variável é considerada em outras emoções do modelo OCC. Desta forma, setou-se com (*likelihood*) como o valor 0.5, apenas por ser um parâmetro obrigatório.

O processo de inferência de emoções no modelo ALMA também pode ser ajustado usando o parâmetro "EmotionDecay" que define o processo de declínio da intensidade das emoções. O processo de declínio é um componente importante do modelo ALMA, pois permite considerar a consequência de eventos anteriores no processo de inferência. No contexto deste trabalho, isto significa que é possível obter *scores* com as probabilidades dos quadrantes que variam conforme a sequência de eventos e não apenas do último evento. O processo de declínio ocorre até que todas as emoções estejam zeradas, situação que configura o estado neutro.

É possível escolher o tipo de função a ser utilizado para o cálculo do declínio das emoções, cujas opções incluem: Linear, Exponencial e Tangente Hiperbólica. Neste trabalho, foi utilizado a função Linear que é a padrão do modelo ALMA. Além disso, pode-se especificar a cada quanto tempo o processo de declínio será executado (*decayperiod*) e em quanto tempo as

¹Foram utilizados valores próximos ao centro (0,5) procurando evitar grandes disparidades entre os eventos

emoções ficarão zeradas (decaytime). Neste trabalho utilizou-se os valores padrão de 500ms para decayperiod e 20000ms para decaytime.

Com o objetivo de simplificar o processo de inferência de emoções, o modelo ALMA introduz uma linguagem básica chamada AffectML (Gebhard, 2005) que é baseada nas chamadas “etiquetas de avaliação” (*appraisal tags*). Estas etiquetas são, na realidade, atalhos para as variáveis relacionadas com as condições de elicitación de emoções (EEC) do modelo OCC (ver Tabela 4.1). Na Tabela 4.2 são apresentadas as etiquetas de avaliação relacionadas com cada um das condições de geração de emoções do modelo OCC (EEC).

Tabela 4.2: Etiquetas de avaliação utilização na geração de emoções no modelo ALMA. Fonte : Gebhard et al. (2003)

Condições de Geração de Emoções	Etiquetas de Avaliação
Evento	GoodEvent, GoodEventForGoodOther, GoodEventForBadOther, BadEvent, BadEventForGoodOther, BadEventForBadOther, GoodLikelyFutureEvent, GoodUnlikelyFutureEvent, BadLikelyFutureEvent, BadUnlikelyFutureEvent, EventConfirmed, EventDisconfirmed
Ação	GoodActSelf, GoodActOther, BadActSelf, BadActOther
Objeto	NiceThing, NastyThing

Considerando as etiquetas de avaliação descritas na Tabela 4.1, o modelo ALMA estabelece o mapeamento entre as emoções definidas no Modelo OCC e as etiquetas de avaliação. Na Tabela 4.3 é apresentado este mapeamento com as emoções do modelo OCC consideradas neste trabalho. As emoções do modelo OCC não consideradas neste trabalho não foram incluídas na Tabela 4.3 e também foram desabilitadas na versão customizada do modelo ALMA.

Tabela 4.3: Mapeamento das etiquetas de avaliação para emoções no modelo OCC Fonte:Gebhard (2005)

Etiqueta(s) de Avaliação	Emoção no Modelo OCC
GoodEvent	Alegria
BadEvent	Angústia
GoodLikelyFutureEvent	Esperança
GoodLikelyFutureEvent + EventConfirmed	Satisfação
GoodLikelyFutureEvent + EventDisconfirmed	Desapontamento
BadLikelyFutureEvent	Medo
BadLikelyFutureEvent + EventConfirmed	Medo Confirmado
BadLikelyFutureEvent + EventDisconfirmed	Alívio

Por fim, para que o processo de inferência das emoções cognitivas possa ser realizado, foi necessário definir a associação entre os eventos a serem monitorados no software educacional com suas respectivas etiquetas de avaliação descritas na Tabela 4.2. No contexto deste trabalho, esta associação foi realizada como uma customização do software educacional a ser descrito no Capítulo 5.

4.3 Inferência de emoções físicas

A abordagem apresentada em Littlewort et al. (2006) e Bartlett et al. (2006) pode ser utilizada para realizar a tarefa de inferência de emoções usando expressões faciais. Nestes trabalhos é demonstrada a aplicação de um conjunto de técnicas e algoritmos, organizados e estruturados para realizar a tarefa de inferência de emoções. Uma visão geral do processo é representado na Figura 4.3 e pode ser dividida em etapas, conforme descrito abaixo:

- Detecção e extração automática da região da face a partir de imagem ou vídeo.
- Extração de atributos da região da face utilizando e seleção de atributos mais relevantes.
- Treinamento de classificadores usando técnicas como SVM (*Support Vector Machine*) para a inferência das emoções, utilizando dados rotulados.

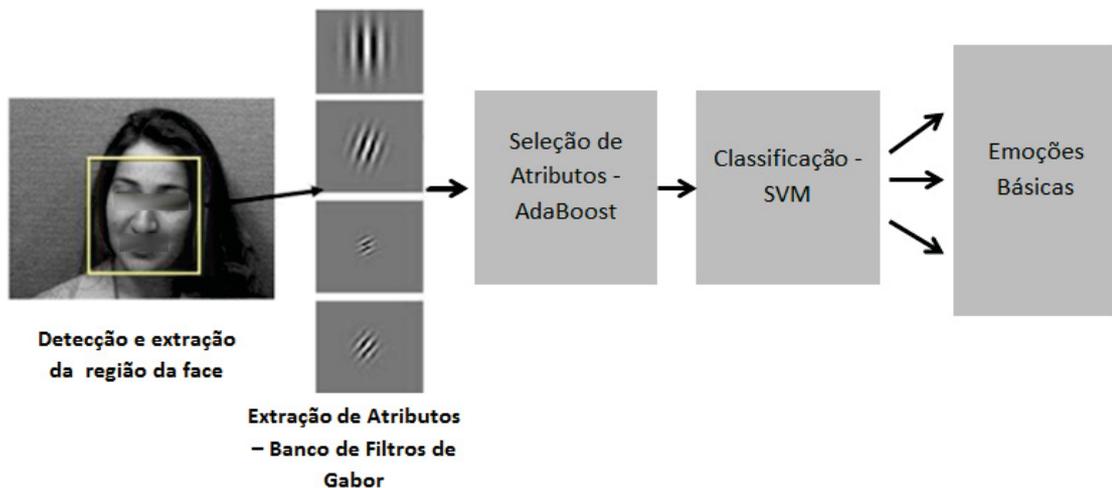


Figura 4.3: Visão geral das etapas do processo de inferência de emoções via expressões faciais. Fonte: adaptado de Bartlett et al. (2006)

Na sequência serão detalhadas as etapas envolvidas no processo de inferência de emoções apresentado na Figura 4.3.

4.3.1 Detecção da Face

A primeira tarefa no processo de inferência consiste em localizar em uma imagem a região de interesse que, neste caso, é a face. Uma técnica para detecção de objetos em imagens chama-se “*Haar Feature-Based Cascade Classifiers*” (Viola e Jones, 2001). Esta técnica é baseada em uma abordagem de aprendizagem de máquina na qual classificadores são treinados em um conjunto de imagens positiva e negativa (i.e. imagens que contêm e outras que não contêm faces).

A técnica proposta por Viola e Jones (2001) destaca-se pela velocidade de execução e também pela acurácia dos resultados. O algoritmo de Viola e Jones (2001) faz parte da ferramenta OPENCV ² (*Open Source Computer Vision Library*), que é uma plataforma aberta voltada para o desenvolvimento de aplicativos na área de visão computacional. Além disso, o OPENCV disponibiliza modelos de classificação já treinados para a tarefa de reconhecimento facial. Estas

²<https://opencv.org/>

características tornam a técnica descrita acima perfeitamente apropriada para aplicações em tempo real, adequando-se, portanto, aos requisitos necessários no contexto da presente pesquisa.

No contexto desta pesquisa, foram realizados experimentos de inferência de emoções básicas utilizando a abordagem de detecção da face apresentada acima, cujos resultados e discussões foram publicados em Gottardo e Pimentel (2016). Um exemplo do resultado destes experimentos pode ser visualizado na Figura 4.4 que demonstra um quadro de um vídeo do autor deste trabalho capturado via webcam. O retângulo sobreposto à imagem demonstra a área da face detectada pela técnica descrita acima.

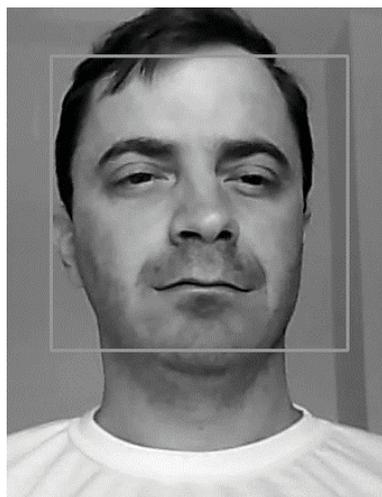


Figura 4.4: Resultado da aplicação da técnica de detecção de face em vídeo. Fonte: (Gottardo e Pimentel, 2016)

4.3.2 Geração de atributos para inferência de emoções

Após a detecção da face em uma imagem, a próxima etapa consiste em gerar atributos desta face que possam ser úteis para distinguir diferentes expressões faciais e correlacioná-las com emoções. Neste processo a técnica chamada de filtros de Gabor apresenta-se como uma abordagem utilizada em pesquisas que tratam do processo de reconhecimento de expressões faciais (Whitehill e Omlin, 2006; Akputu et al., 2018).

Filtros de Gabor são utilizados na área de processamento de imagens para tarefas como representação de imagens, detecção de bordas, representação e classificação de texturas (Mella e Bellavista, 2016). Utilizando um conjunto de classes de funções, os filtros de Gabor permitem representar qualquer tipo de imagem, considerando simultaneamente tanto a frequência quanto à orientação. Estas funções são compostas por senóides complexas e bidimensionais que são moduladas por uma função Gaussiana. Com isto torna-se possível a extração de atributos para caracterizar diferentes tipos de estruturas presentes nas imagens (Ji et al., 2004).

Bancos de filtros de Gabor podem ser gerados através da manipulação de parâmetros dos filtros de Gabor individuais como frequência, orientação, simetria e excentricidade (Ji et al., 2004). Na Figura 4.5 é apresentado um exemplo de banco de filtros de Gabor com 4 escalas (linhas) e 8 orientações (colunas) diferentes em cada uma das quatro escalas.

Assim como na primeira etapa de detecção da face, o processo de geração de atributos da face detectada utilizando filtros Gabor tem como resultado um grande número de atributos. Para melhorar os resultados do processo de classificação pode-se utilizar técnicas de redução de

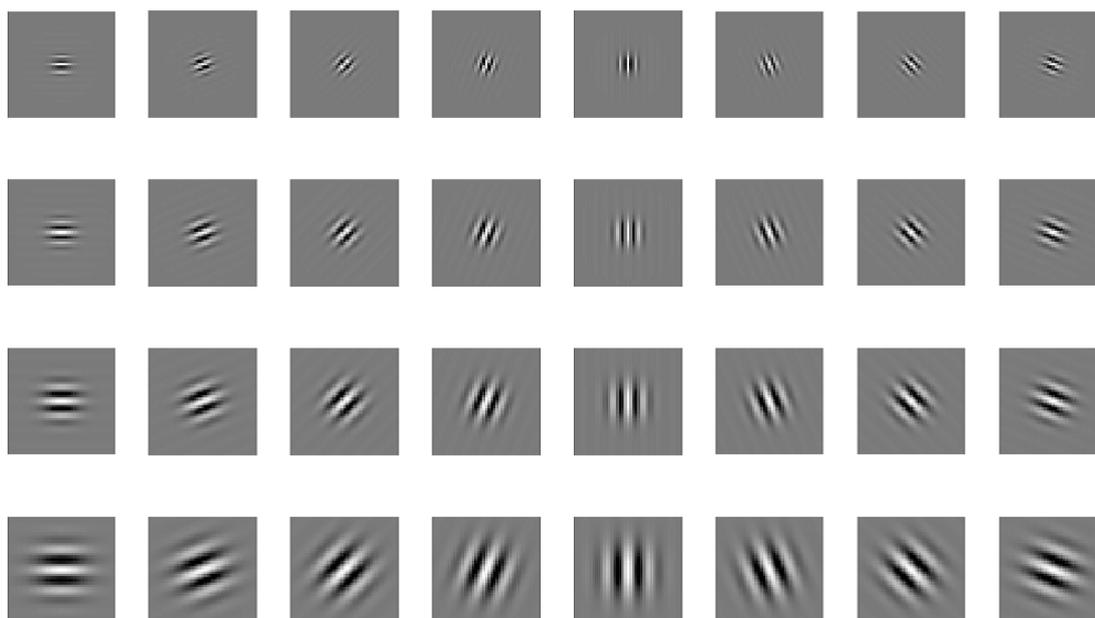


Figura 4.5: Exemplo de um banco de filtros Gabor com 4 escalas e oito orientações

dimensionalidade e seleção de atributos mais relevantes como a técnica AdaBoost (Witten et al., 2016).

4.3.3 Treinamento de classificadores para detecção de emoções

Tomando-se como base os atributos mais relevantes obtidos após a aplicação de um banco de filtros gabor, técnicas de aprendizagem de máquinas supervisionadas podem ser aplicadas. O algoritmo SVM (*Support Vector Machines*) (Witten et al., 2016) é um exemplo de técnica que pode ser utilizada para o treinamento de classificadores e construção de modelos de inferência de emoções baseadas em imagens da face.

Para a etapa de construção, treinamento, validação e testes do classificador SVM são necessários dados rotulados relativo ao domínio desejado. O banco de dados “Cohn-Kanade Expression Database” (Kanade et al., 2000) é um exemplo de banco de dados rotulados com informações afetivas que têm sido utilizado em trabalhos correlatos (Whitehill e Omlin, 2006; Gottardo e Pimentel, 2016; Mella e Bellavista, 2016; Akputu et al., 2018). Este banco de dados foi projetado pelo Grupo de Análise Afetiva da Universidade de Pittsburg ³ com o objetivo de propiciar o desenvolvimento de pesquisas envolvendo análise e síntese automática de imagens faciais.

No total o banco de dados Cohn-Kanade disponibiliza 486 imagens, podendo ser utilizado para treinamento e testes de algoritmos de aprendizagem de máquina. Este banco de dados disponibiliza imagens com seus respectivos metadados relativos às Unidades de Ações (Action Units) e rótulos classificando cada imagem nas sete emoções básicas de Ekman (1992) mais o estado neutro.

Destaca-se que na obtenção das imagens aspectos como iluminação e ângulo das imagens foram controlados, utilizando-se lâmpadas de alta intensidade e também coletando imagens em ângulos de rotação de 30 graus. Na Figura 4.6 são demonstrados exemplos de três emoções básicas sendo simulada por um participante do experimento para construção do banco de dados

³Link da página principal do grupo de pesquisa: <http://www.pitt.edu/~emotion/index.html>

Cohn-Kanade. As imagens mostradas são de um participante cuja divulgação é permitida licença de uso do banco de dados.

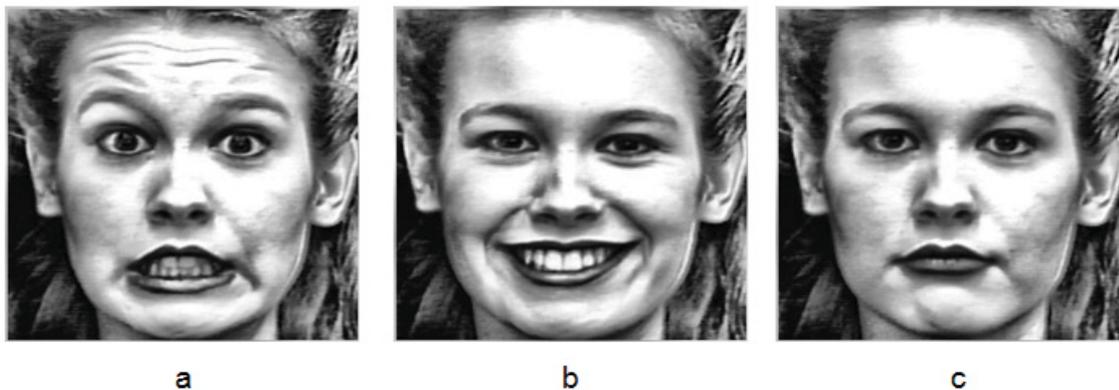


Figura 4.6: Imagens do banco de dados Cohn-Kanade em que um participante simula três emoções básicas: a) medo b) alegria c) neutro Fonte: Kanade et al. (2000)

A partir de um banco de dados rotulado como o Cohn-Kanade, por exemplo, é possível aplicar o processo de detecção de face e extração de atributos para construção e treinamento de modelos utilizando classificadores como o SVM. Este modelo pode então ser utilizado para classificar novas instâncias de imagens.

O trabalho de Gottardo e Pimentel (2016) apresenta uma discussão e um experimento prático de inferência de emoções básicas utilizando os procedimentos e o banco de dados Cohn-Kanade descritos nesta seção. Um exemplo de resultado deste processo é mostrado na Figura 4.7 em que são exibidos quadros de um vídeo feito com autor deste trabalho. Além de um retângulo destacando a região da face, nas figuras é apresentado texto com o resultado do processo de classificação. Neste caso, são exibidos resultados em que a classificação foi correta. Por outro lado, na Figura 4.8 são mostrados exemplos de classificação incorretas em que as imagens visualmente não condizem com o resultado expresso pelo classificador.



Figura 4.7: Quadros do vídeo demonstrando experimento com o autor em que as imagens foram classificadas corretamente como: a) neutro, b) surpresa e c) felicidade Fonte: Gottardo e Pimentel (2016)

Os resultados do experimento apresentado em Gottardo e Pimentel (2016) demonstraram que emoções com expressões mais destacadas como felicidade, raiva, surpresa e desgosto obtiveram índices elevados de acerto na classificação. Esses índices superam 90% de acurácia, e poderiam, se analisados de forma isolada, serem considerados excelentes. Entretanto, índices extremamente inferiores foram obtidos com as emoções tristeza, medo, chegando a nenhum

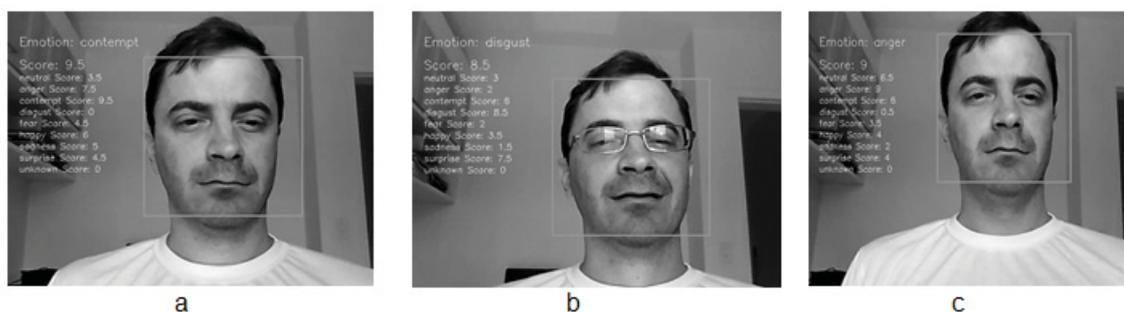


Figura 4.8: Quadros do vídeo demonstrando experimento com o autor em que as imagens foram classificadas incorretamente como: a) desprezo, b) desgosto e c) raiva Fonte: Gottardo e Pimentel (2016)

acerto para a emoção desprezo. Observou-se também neste trabalho que a emoção Tristeza foi classificada em 55% dos casos como Neutro.

Os resultados dos experimentos apresentados em Gottardo e Pimentel (2016) foram considerados importantes para a sequência da realização desta pesquisa. Observou-se que o banco de dados Cohn-Kanade apresenta severas limitações para aplicação prática, como pretendido neste trabalho. A principal limitação refere-se ao fato de que as imagens do banco de dados Cohn-Kanade são obtidas em um ambiente com iluminação, posição, qualidade das câmeras e intensidade das emoções (ver Figura 4.6) que não representam adequadamente uma situação real. Como exemplo, no experimento a ser descrito no próximo capítulo as condições encontradas nos laboratórios de informática das escola apresentam condições de iluminação, qualidade de equipamento de captura de imagens e posição dos estudantes bastante diferentes àquelas contidas no banco Cohn-Kanade.

Diante do exposto acima, identificou-se que seria necessário a busca de alternativas ao banco de dados Cohn-Kanade. Contudo não foram identificadas opções gratuitas que pudessem atender aos requisitos deste trabalho. Adicionalmente, a construção de um banco próprio para ser utilizado nesta pesquisa fica fora do escopo do trabalho e uma atividade adicional que faria o cronograma previsto para a pesquisa ser extrapolado. Desta forma, optou-se por utilizar uma ferramenta gratuita chamada de EmotionAPI ⁴ que é disponibilizada pela Microsoft e a Universidade de Oxford. Esta escolha deu-se após a realização de testes preliminares que demonstraram resultados melhores que os obtidos com o banco de dados Cohn-Kanade.

De forma semelhante ao processo descrito acima nesta seção, a EmotionAPI recebe uma imagem como entrada e detecta a face. A partir da imagem da face o processo de inferência de emoções é executado, retornando um score no intervalo entre 0 e 1 (normalizados para somarem um) para o mesmo conjunto de emoções básicas utilizadas neste trabalho que são: raiva, desgosto, medo, alegria, surpresa, tristeza, desprezo e neutro.

4.4 Fusão dos componentes cognitivo e físico

Como o modelo híbrido de inferência de emoções utiliza dois componentes distintos (i.e. cognitivo e físico), faz-se necessário a aplicação de métodos para fusão destes dois componentes para a obtenção da inferência final. Neste sentido, conforme observa Kuncheva (2004), combinar múltiplos classificadores é uma abordagem promissora para melhorar o desempenho global do processo de classificação.

⁴<https://azure.microsoft.com/pt-br/services/cognitive-services/emotion/>

Para a criação de ambientes capazes de lidar com diferentes fontes de informação, três diferentes abordagens de fusão têm sido propostas e avaliadas na literatura (Schels et al., 2013). Estas abordagens são denominadas de: i) fusão antecipada (*early fusion*), ii) fusão intermediária (*mid-level fusion*) e iii) fusão tardia (*late fusion*). Na Figura 4.9 é representado esquematicamente as três abordagens de fusão de classificadores em que os quadrados com fundo cinza representam o processo de fusão e os retângulos com fundo cinza escuro representam o resultado da classificação. Os retângulos com fundo branco representam um conjunto de atributos e os quadrados com fundo branco representam processos de transformação ou classificação.

Na Fusão antecipada, as informações são combinadas em níveis iniciais pela concatenação de atributos individuais em um vetor de maior dimensão, como mostrado à esquerda da Figura 4.9 ‘a’. Estratégia de combinação inversa é utilizada na Fusão tardia (Figura 4.9 ‘c’) em que os resultados dos processo de classificação são fundidos.

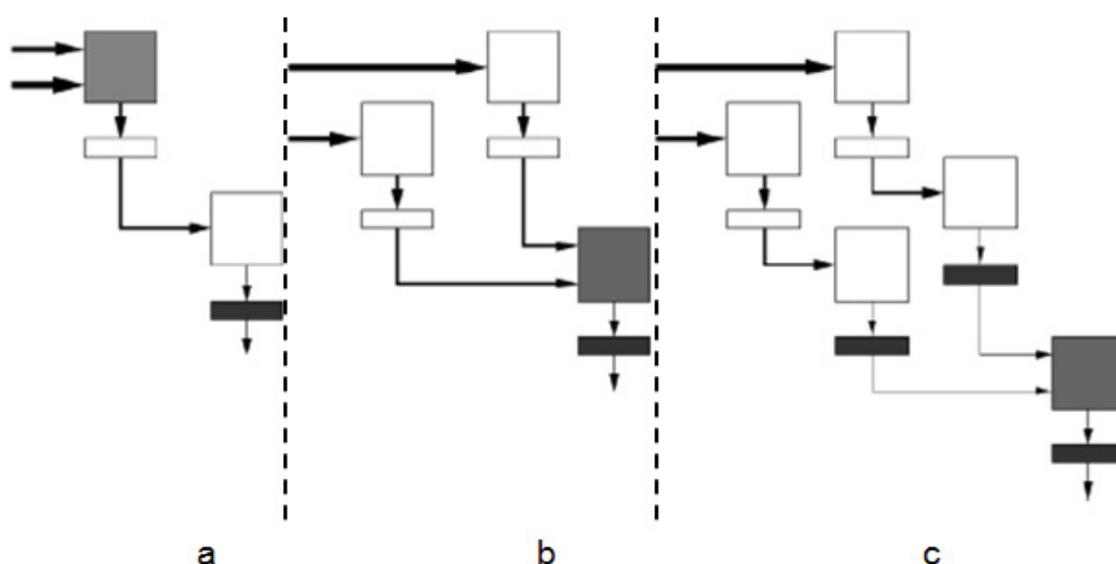


Figura 4.9: Representação esquemática das diferentes abordagens de fusão de classificadores: fusão antecipada (a), fusão intermediária(b) e fusão tardia(c). Fonte: Adaptado de Schels et al. (2013)

É importante destacar que a escolha da arquitetura de fusão é fortemente relacionada com o problema a ser tratado e a complexidade dos classificadores utilizados (Schels et al., 2013). Porém, os referidos autores observam que a estratégia de fusão antecipada e intermediária implicam no aumento da dimensionalidade e a ocorrência do efeito negativo conhecido como “maldição da dimensionalidade” (em inglês “*curse of dimensionality*”).

Além disso, em aplicações de inferência de emoções, a fusão antecipada e intermediária tornam-se pouco intuitivas tendo em vista que as diferentes origens de informações tendem a ser composta de dados altamente heterogêneos (Schels et al., 2013). Por outro lado, utilizando a abordagem de fusão tardia é possível realizar a compensação de falha de um determinado canal ou sensor (Schels et al., 2013). Este aspecto apresenta-se como um benefício importante relacionado diretamente com as características do modelo de inferência proposto. Como exemplo, o componente de inferência de expressões faciais irão falhar no caso em que o estudante mover a cabeça em um ângulo que a câmera não consiga captar. Desta maneira, no contexto da presente proposta, optou-se por utilizar a fusão tardia.

Para a implementação da fusão de informações e classificação de fontes distintas, diferentes técnicas podem ser utilizadas. Estas técnicas envolvem a combinação de múltiplos

classificadores, configurando o que é denominado de Sistemas de Múltiplos Classificadores - MCS (*Multiple Classifiers Systems*) (Witten et al., 2016). Regras fixas de combinação de classificadores podem ser utilizadas, dependendo do tipo de saída dos classificadores (Kuncheva, 2004). Dentre estas regras pode-se citar, por exemplo, soma, média, máximo, mínimo, etc. Porém, conforme observa Duin (2002), a utilização destas regras fixas somente será a melhor alternativa em condições específicas.

Diante disso, optou-se por utilizar técnicas de combinação que operam sobre as saídas dos classificadores individuais. Nesta abordagem as saídas dos classificadores individuais são tratadas como entradas para outro classificador genérico para a construção, treinamento e teste de um modelo de classificação chamado secundário que será responsável pela fusão e obtenção do resultado final (Duin, 2002; Tulyakov et al., 2008; Kuncheva, 2004).

A vantagem desta estratégia de fusão é que ela é capaz de aprender a melhor combinação dos classificadores bases, podendo automaticamente considerar os pontos fortes e fracos de cada um (Tulyakov et al., 2008). Desta maneira, esta abordagem foi a escolhida para o processo de fusão dos componentes físicos e cognitivos do ModHEmo.

Testes iniciais com a base de dados utilizada neste trabalho indicaram que algoritmos de classificação baseados em técnicas simples e rápidas como Árvores de Decisão (*decision trees*) e K-vizinhos mais próximos (KNN) atingiram os melhores resultados. Considerando estas técnicas, foram escolhidos os algoritmos RandomForest (árvore de decisão) e IBK (KNN) (Witten et al., 2016) para serem utilizados neste trabalho.

O algoritmo RandomForest (Breiman, 2001) consiste na combinação de classificadores baseados em árvore em que cada árvore é formada por um vetor aleatório obtido de forma independente e com a mesma distribuição para todas as árvores da floresta. Cada árvore realiza a classificação e o resultado final é escolhido considerando-se o resultado mais comum entre as árvores. Esta técnica tem como vantagem a redução da tendência de *overfitting*⁵ das árvores de decisão tradicionais.

O algoritmo IBK (Witten et al., 2016) é baseado na técnica conhecida como *k-nearest-neighbor classifier* ou classificador baseado em vizinhos mais próximos.

Neste trabalho foi utilizado a implementação dos algoritmos RandomForest e IBK usando a ferramenta Weka⁶. Além disso, a validação cruzada de 10 grupos (*10-fold cross validation*) foi usada para dividir a base em teste e treinamento, considerando-se que é uma técnica robusta e altamente utilizada (Witten et al., 2016).

Como o objetivo principal deste trabalho não inclui o ajuste dos parâmetros de algoritmos, utilizou-se os valores padrão dos parâmetros Weka. Para o algoritmo IBK o valor padrão é a distância Euclidiana e número de vizinhos igual a 1. Para o algoritmo RandomForest o padrão da altura máxima da árvore é definida como ilimitada, número de iterações (número de árvores) igual a 100.

4.5 Exemplo ilustrativo do processo de inferência do ModHEmo

Com o objetivo de demonstrar o processo de inferência do ModHEmo, será apresentado a seguir um exemplo ilustrativo que descreve as etapas de execução do ModHEmo. É importante reforçar que os valores utilizados no exemplo são apenas para ilustração. A Figura 4.10 apresenta

⁵classificador preciso nos dados que foram usados para treinamento, porém com desempenho ruim em um conjunto independente de dados de teste

⁶<https://www.cs.waikato.ac.nz/ml/weka>

sequencialmente as 10 etapas executadas pelo ModHEmo para inferência da emoção no instante 'T1' em que simula-se a ocorrência do evento 'Estudante comete um erro' em um software educacional. Na etapa 1 (componente cognitivo), ocorre o mapeamento do evento para a etiqueta AffectML 'BadEvent' (ver Tabela 4.2), com valores de intensidade e probabilidade igual a 0,5. Na etapa 2, o modelo ALMA realiza o processo de inferência da emoção, neste caso Angústia (ver Tabela 4.3), e calcula os valores (intensidade) das oito emoções cognitivas. Neste exemplo, considerou-se que no instante 'T1' em que evento relevante ocorreu, todas as emoções cognitivas estava zeradas (estado neutro). Desta forma, apenas a emoção Angústia tem intensidade igual a 0,5 correspondente à intensidade do evento ocorrido.

Na etapa 3, é realizado o mapeamento dos valores das emoções conforme os quadrantes cognitivos (ver Figura 3.9). Neste caso o valor da emoção Angústia foi adicionado ao quadrante Q3. Por fim, na etapa 4 do componente cognitivo é realizada a normalização dos valores dos quadrantes para o intervalo [0,1] usando a função Softmax descrita na Seção 4.1.1. Neste caso a normalização resultou em valor 1 para o quadrante Q3, pois todos os demais tem valor igual a zero.

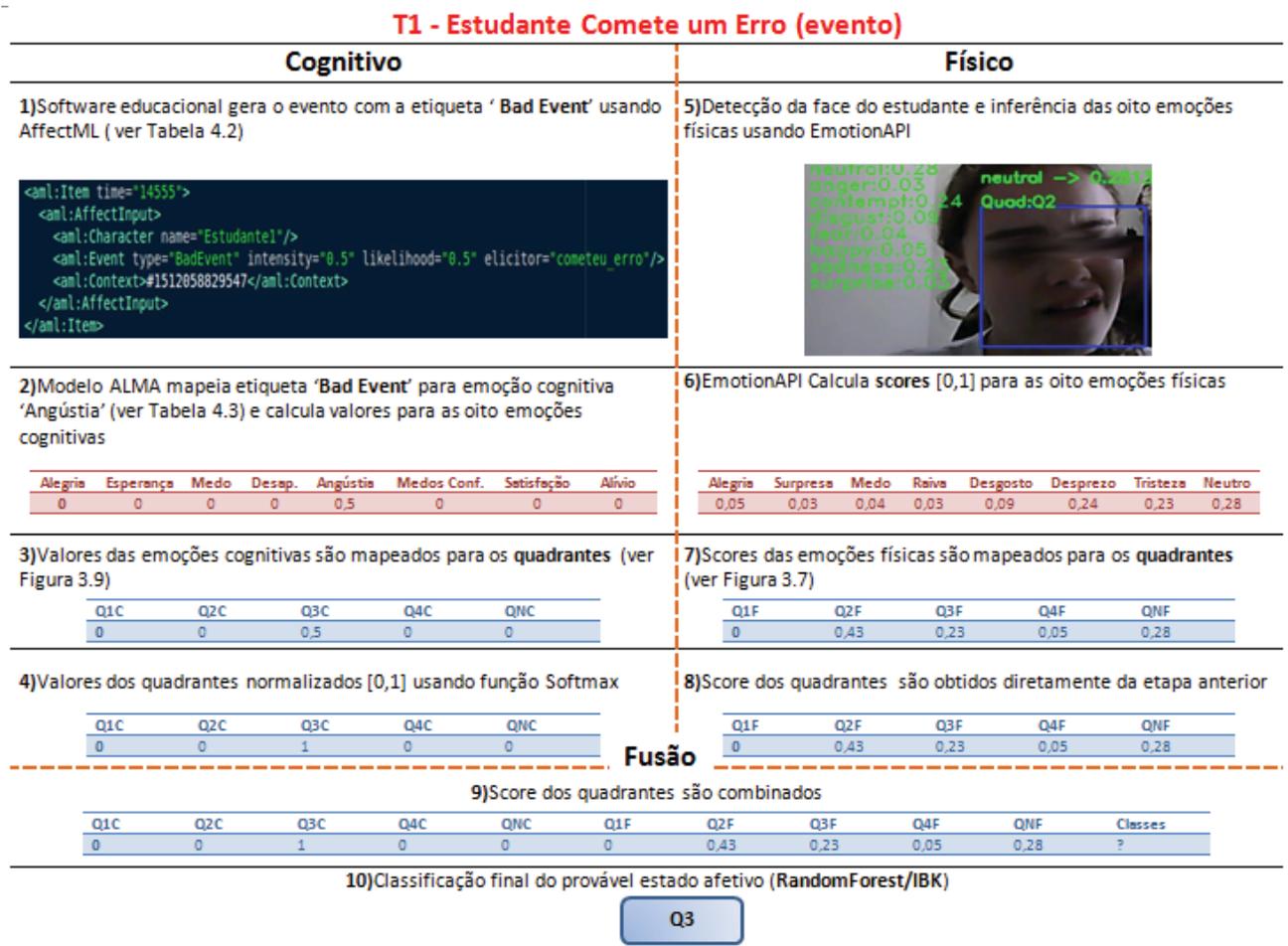


Figura 4.10: Simulação de execução do ModHEmo

Na etapa 5 (componente físico) é realizada a captura da imagem, detecção da face (retângulo azul) e inferência das emoções usando a EmotionAPI. À esquerda da figura apresentada nesta etapa (em verde), pode-se ver o *score* de cada emoção física. No centro (em verde) é destacada a emoção com maior *score* (neutro) e o quadrante com maior *score* (Q2). Na etapa

6 são calculados os scores [0,1] para cada uma das oito emoções físicas. Nas etapas 7 e 8 são mapeados os valores das emoções para os quadrantes conforme Figura 3.7.

O processo de fusão inicia-se na etapa 9 em que os quadrantes dos componentes físico e cognitivo são combinados em um único conjunto de dados. Considerando-se o conjunto de dados combinado é realizada então a etapa final (10) que consiste na inferência do provável quadrante, utilizando os algoritmos de classificação RandomForest e IBK.

É importante destacar que, **enquanto não ocorrerem novos eventos no software educacional**, o modelo ALMA irá automaticamente realizar o processo de declínio das emoções cognitivas. Este processo de declínio pode chegar até o estado neutro, em que todas as emoções cognitivas estão zeradas. Caso um novo evento ocorra antes que as emoções cognitivas estiverem zeradas, os valores serão considerados no novo processo de cálculo do modelo ALMA. Desta forma, poderão ser obtidos valores para mais de uma emoção cognitiva em um determinado instante.

No instante que ocorrer um novo evento relevante e monitorado no software educacional, o processo descrito na Figura 4.10 é repetido sucessivamente.

4.6 Conclusão

Neste capítulo foi apresentado o Modelo Híbrido de Inferência de Emoções - ModHEmo, destacando-se o embasamento conceitual e os detalhes de sua concepção e implementação. Este modelo, representa a estrutura central do processo de inferência das emoções relacionadas à aprendizagem.

Uma das principais características do ModHEmo refere-se à sua capacidade de tratamento simultâneo e fusão de informações cognitivas e físicas. É importante ressaltar que emoções cognitivas são suscitadas baseando-se na interpretação pessoal de um evento, indicando as causas potenciais de uma determinada emoção. Por outro lado, no componente físico a inferência das emoções é feita baseando-se nos efeitos observáveis de uma emoção. Desta maneira, presume-se que a integração das inferências obtidas via expressões faciais (reação física) poderão ser úteis para capturar diferentes reações individuais e integrá-las com as informações cognitivas no processo de inferência final.

Diante disso, o modelo é composto por três componentes principais especialmente projetados para realizar inferências de emoções utilizando modelos cognitivos e físicos. O modelo OCC foi escolhido para embasar as inferências do componente cognitivo utilizando-se como entrada os eventos ocorrido no ambiente computacional. Esta escolha foi feita considerando-se que atualmente este modelo é amplamente utilizado e consolidado na comunidade científica. O componente físico utiliza como base conceitual o modelo de emoções básicas de Ekman que são inferidas utilizando a expressão facial dos estudantes. Para viabilizar o processo de fusão, as inferências de ambos os componentes são mapeadas para os quadrantes das emoções apresentados no Capítulo 3.

Finalmente, o modelo baseia-se em uma estratégia de fusão que utiliza as saídas dos componentes físico e cognitivo como entrada para algoritmos de classificação. Estes algoritmos realizam o treinamento e teste de modelos de classificação secundários que são responsáveis pela fusão dos componentes e inferência final da provável emoção do estudante.

Capítulo 5

Experimentos e Resultados

Neste capítulo serão apresentados inicialmente detalhes dos experimentos realizados, destacando-se a metodologia e os materiais e métodos utilizados. Na sequência, serão apresentados e discutidos os resultados obtidos com os experimentos em ambientes reais de aprendizado, bem como as possíveis limitações destes resultados.

5.1 Metodologia

Com o objetivo de verificar a viabilidade do modelo de inferência proposto e também analisar os seus resultados, experimentos em ambientes reais de aprendizagem foram realizados com uma versão implementada do ModHEmo. Nestes experimentos participaram um total de 38 estudantes com idade variando entre 10 a 14 anos. Estes estudantes estavam cursando entre o quinto e nono ano do Ensino Fundamental na Escola Municipal de Ensino Fundamental Jaguaretê, situada no distrito de Jaguaretê, na zona rural de Erechim-RS. Foi disponibilizado pela escola um total de 1:30 hrs para a realização de cada experimento.

Um experimento com 21 estudantes foi realizado no laboratório de informática do Núcleo Tecnológico Municipal (NTM), situado na cidade de Erechim-RS nos dias 02/10/2017 e 05/10/2017. Para os demais estudantes (17), o experimento foi realizado no laboratório da escola Jaguaretê no dia 30/11/2017. Desta forma, foi possível coletar dados em dois laboratórios de informática distintos com variação de espaço, iluminação e configuração de equipamentos, por exemplo.

Cabe ressaltar que os experimentos foram aprovados e seguem os procedimentos recomendados pelo comitê de ética na pesquisa do Instituto Federal de Educação, Ciência e Tecnologia do RS - IFRS onde o autor atua como docente.

Os experimentos foram conduzidos tentando-se fazer com que a pesquisa interferisse o mínimo possível no comportamento habitual ou natural dos estudantes ao usar os laboratórios de informática e o software educacional. Com este objetivo, para a execução dos experimentos, foram planejadas e seguidas o conjunto de etapas descrito a seguir:

1. Antes da execução dos experimentos nas escolas, foram realizados estudos piloto com dois estudantes da mesma faixa etária dos estudantes que participaram dos experimentos. Desta forma, foi possível simular a realização do experimento e também verificar o funcionamento dos softwares envolvidos. Como resultado desta atividade ajustes de interface e configurações no software educacional foram realizados. Os dados coletados nestes estudos piloto não foram incluídos nos resultados descritos na sequência deste capítulo.

2. Os laboratórios de informática das escolas foram preparados com antecedência, de maneira que quando os estudantes chegaram para a atividade o ambiente já estava pronto. Esta preparação envolveu principalmente a configuração do software educacional e a instalação e configuração de webcams, pois os computadores não possuíam estes dispositivos.
3. Antes de iniciar os experimentos, foi apresentado aos estudantes uma visão geral desta pesquisa, destacando principalmente os objetivos e a metodologia do trabalho. Foram informados que as câmeras iriam gravá-los durante a atividade, mas que nenhuma informação pessoal seria divulgada. Além disso, os estudantes foram orientados a realizar a atividade naturalmente, da mesma maneira que em outras aulas no laboratório. Deixou-se claro também que nenhuma restrição sobre a posição e movimentos seriam necessárias e que a atividade não faria parte das notas escolares. Cabe ressaltar que a professora da turma acompanhou os experimentos.
4. Na sequência, foi realizada explicação detalhada sobre o funcionamento e objetivos do software educacional (jogo Tuxmath a ser descrito a seguir) a ser utilizado. Em seguida, os estudantes acessaram o software educacional, realizaram atividades para ambientarem-se ao jogo e tirarem dúvidas. Esta etapa não foi gravada e não faz parte dos dados coletados.
5. Antes de iniciar as atividades, os estudantes responderam às questões de um formulário conforme mostrado na Figura 5.1. Neste formulário são solicitadas informações pessoais, o objetivo dos estudantes com a realização da atividade e a expectativa em relação ao resultado do jogo. Além disso, devido às diferenças de altura entre os estudantes, antes de iniciar as atividades foi realizado um procedimento para ajustar a posição das Webcams de forma que a face dos estudantes ficasse mais centralizada possível.
6. Depois de finalizadas as etapas anteriores, passou-se para a execução da atividade principal que envolveu a execução do jogo e coleta dos dados. Maiores detalhes sobre esta etapa serão apresentadas na Seção 5.2.3.
7. Finalizada a execução do jogo, restava ainda realizar a rotulagem dos dados coletados. Para isso os estudantes, primeiramente, foram orientados sobre os objetivos da rotulagem e sobre o funcionamento da ferramenta de rotulagem. A ferramenta e mais detalhes do processo de rotulagem serão apresentadas na Seção 5.2.2.

Da mesma forma que no trabalho de Muñoz et al. (2010), nesta pesquisa foi utilizado um formulário (ver Figura 5.1) para obter informações relevantes sobre os estudantes. Neste formulário, destacam-se as informações relativas a objetivo e expectativa dos estudantes com a realização da atividade. Estas informações são fundamentais pois o processo de inferência do modelo OCC/ALMA é fortemente baseada nos objetivos pessoais e também na expectativa ou prospecto. Desta forma, considerando o objetivo do estudante e a expectativa em relação ao jogo, foi possível setar as variáveis desejabilidade (*desirability*) e prospecto (*prospect*) no modelo ALMA. Embora existam ferramentas padronizadas para medir a motivação/objetivos dos estudantes como MSLQ (*Motivated Strategies for Learning Questionnaire*) (Pintrich et al., 1991), decidiu-se, por questão de simplicidade e tempo do experimento, utilizar o questionário simplificado apresentado na Figura 5.1.

Somente foram incluídos no conjunto de dados deste trabalho os 35 dos 38 estudantes cuja resposta foi “Realizei a atividade buscando melhorar habilidades matemáticas” na questão sobre o objetivo dos estudantes do formulário da Figura 5.1. Esta opção deu-se pelo fato de que este foi o único objetivo considerado neste trabalho para configuração das etiquetas do

Identificação Participantes da Pesquisa

Por Favor, responda às perguntas abaixo

***Obrigatório**

Seu nome *

Sua resposta

Idade *

Sua resposta

Ano/Série *

Sua resposta

Sexo

- Masculino
- Feminino

Qual das opções abaixo melhor descreve o OBJETIVO que você considerou durante o jogo TuxMath *

- Realizei a atividade buscando melhorar habilidades matemáticas
- Realizei a atividade apenas como diversão
- Realizei a atividade apenas para pontuar mais que os colegas
- Outro:

Qual a sua expectativa em relação ao resultado do jogo *

- Estou confiante que irei vencer o jogo
- Acredito que irei perder o jogo

Figura 5.1: Formulário com questões respondidas pelos estudantes antes de iniciar o jogo

modelo ALMA/OCC no software educacional, conforme pode ser visto na Tabela 5.1. Destes 35 estudantes, 32 realizaram o processo de rotulagem da base de dados (a ser descrito na Seção

5.2.2). Portanto, o número de estudantes final considerado neste trabalho foi **32**, 15 do sexo feminino e 17 do sexo masculino.

A Figura 5.2 apresenta um cenário dos experimentos no momento que os estudantes estavam realizando as atividades no software educacional no laboratório de informática. Nesta figura pode-se visualizar a interface do software educacional além das Webcams posicionadas para capturar as expressões faciais dos estudantes.



Figura 5.2: Estudantes durante os experimentos

5.2 Materiais e procedimentos utilizados no experimento

Nesta seção serão descritos os materiais aplicados para projetar e desenvolver os experimentos em um ambiente real de aprendizado.

5.2.1 Software Educativo

Para a realização dos experimentos utilizou-se o software educacional ‘Tux, of Math Command’ ou TuxMath ¹. Esta escolha deveu-se ao fato de que o TuxMath é um software de código aberto cuja licença torna possível a customização do seu código fonte e utiliza uma abordagem educacional baseada em jogos.

O Tuxmath permite que estudantes pratiquem suas habilidades aritméticas, enquanto defendem pinguins do ataque de cometas. Os jogadores devem responder à equação matemática contida no cometa para destruí-lo, antes que ele atinja um dos iglus ou pinguins. Caso a equação não seja respondida corretamente, o cometa atinge e destrói os pinguins e iglus. Existem três tipos de cometas no Tuxmath:

- Normais: são na cor azul e não apresentam nenhuma característica especial.
- Bônus: cometas na cor vermelha que, se destruídos, permitem a reconstrução de iglus que foram danificados anteriormente.
- Super Poderes (*PowerUp*): tipo especial de cometa na cor amarela que, se destruído, o jogador ganha uma arma especial (super bomba). Diferente dos demais, este tipo de cometa percorre a tela na direção horizontal, podendo aparecer tanto na esquerda ou direita da tela.

¹<http://tux4kids.alioth.debian.org/tuxmath/index.php>

Por padrão, no TuxMath um conjunto de quatro pinguins protegidos por quatro iglus são criados na parte inferior da tela. Se um cometa atingir um dos quatro iglus pela primeira vez apenas o iglu será destruído, porém, o pinguim ainda permanecerá vivo. Caso um cometa atingir um iglu destruído, o pinguim será morto e sairá do jogo. Se todos os pinguins forem mortos, o jogo acaba com a derrota do jogador.

O TuxMath disponibiliza um conjunto de níveis de complexidade do jogo, com grau crescente de dificuldade nas equações matemáticas. Os níveis disponíveis no TuxMath são listados abaixo:

- Cadete Espacial: adição simples.
- Sentinela: adição e subtração até 10.
- Defensor: adição, subtração e divisão até 10.
- Especialista: todas as operações com números até 20, incluindo números negativos.
- Comando: todas as operações dos níveis anteriores, incluindo múltiplas operações no mesmo cometa.

Dentro de cada nível, os cometas no TuxMath são liberados em ‘ondas’ com um número crescente de cometas (2,4,6,8,10..) em cada onda. Uma nova onda inicia somente após todos os cometas da onda anterior serem destruídos ou atingirem os iglus. Na Figura 5.3 é apresentado uma tela do TuxMath, demonstrando os iglus e pinguins na parte inferior e os cometas em queda com as equações matemáticas relacionadas a cada cometa. Nesta figura um dos quatro iglus já foi destruído e o pinguim morto.

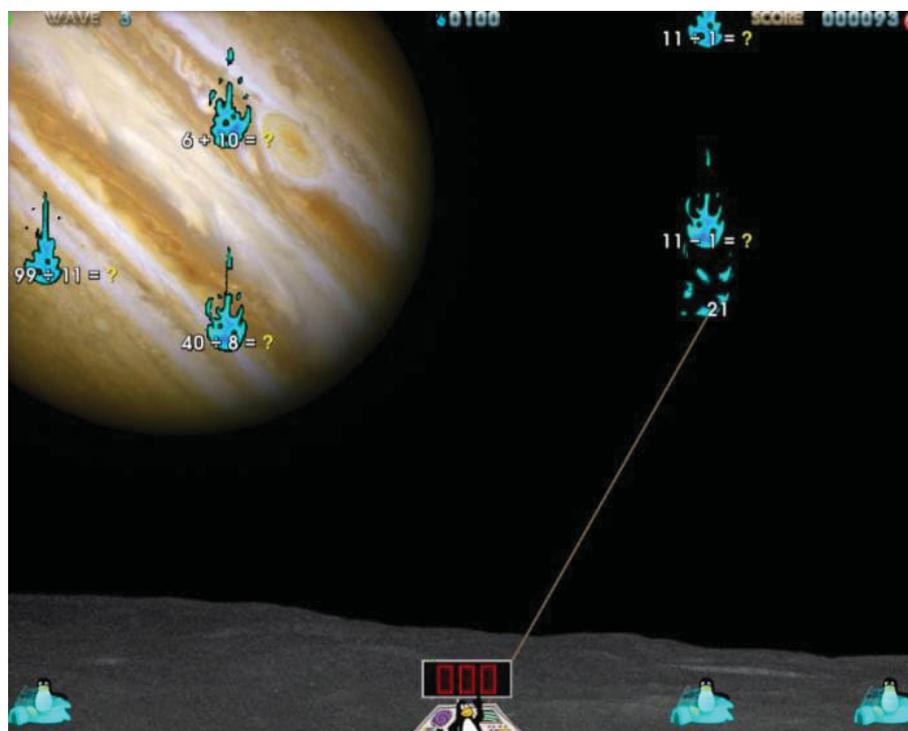


Figura 5.3: Exemplo da Interface do TuxMath.

Para a realização dos experimentos foram realizadas alterações no TuxMath padrão de forma a atender a um conjunto de necessidades da presente pesquisa. Dentre estas alterações, destacam-se:

- implementação de critério para vencer ou perder o nível atual do jogo. A fim de não prolongar o tempo do experimento, foi utilizado um máximo de 25 cometas por nível. Se após 25 cometas ainda restar pelo menos um pinguim vivo, considerou-se que o estudante venceu o jogo (nível). Caso contrário, considera-se que o estudante perdeu o jogo (nível)
- implementação de funcionalidade para verificar se o estudante venceu ou perdeu determinada onda. Neste caso, considerou-se que o estudante venceu a onda, caso tenha destruído a metade ou mais dos cometas de cada onda. Caso contrário, considera-se que a onda foi perdida.
- criação de procedimento para geração de logs para monitoramento dos eventos relevantes a serem monitorados neste trabalho (A lista completa dos eventos será apresentada a seguir na Tabela 5.1).

5.2.2 Rotulagem da base de dados dos experimentos

Após a conclusão da sessão de atividades no Tuxmath, os estudantes foram convidados a realizar a rotulagem da base de dados criada automaticamente durante a execução do jogo. Esta etapa dos experimentos é de fundamental importância, pois os dados coletados permitem verificar o desempenho das inferências feitas pelo ModHEmo.

Para a tarefa de rotulagem foi desenvolvida uma ferramenta personalizada que foi instalada em cada computador do laboratório. Esta ferramenta permite que os estudantes revejam a sessão do jogo através de uma interface que mostra de forma sincronizada um vídeo capturado pela webcam juntamente com um vídeo que mostra a gravação da tela do jogo. Esta abordagem baseia-se no protocolo de julgamento retrospectivo (Rosenberg e Ekman, 1994), que é uma técnica validada para julgamento de estados afetivos *offline*, sem interrupções durante a sessão de aprendizagem (Bosch e D’Mello, 2017).

Os vídeos são pausados automaticamente pela ferramenta de rotulagem no próximo *frame* disponível após a ocorrência de um evento listado na Tabela 5.1. Neste momento, uma imagem com cinco *emoticons* representativos (um para cada quadrante mais estado neutro) é mostrada e solicitado ao estudante para escolher o *emoticon* que melhor representa a emoção experimentada naquele instante. Após a resposta do estudante, o processo continua. A Figura 5.4 mostra uma captura de tela com um exemplo da interface da ferramenta de rotulagem da base de dados utilizada pelos estudantes. Destaca-se nesta figura quatro partes principais, identificadas com números romanos I,II,III e IV inseridos na figura:

- I) A metade superior da figura mostra vídeo do estudante capturados pela webcam no *timestamp* 2017-10-02 12:36:42.450.
- II) Na parte inferior, observa-se interface do jogo sincronizada com a parte superior (i.e no mesmo *timestamp*).
- III) Imagem com *emoticons* e emoções principais representativas de cada um dos quadrantes. Esta imagem é exibida sempre que um evento monitorado ocorre e permite ao estudante selecionar a figura que melhor representa a emoção que estava sentindo neste instante.
- IV) Descrição do evento que ocorreu no *timestamp* específico que é “Erro do Programa - Cometa apareceu no meio da tela !!!!”.

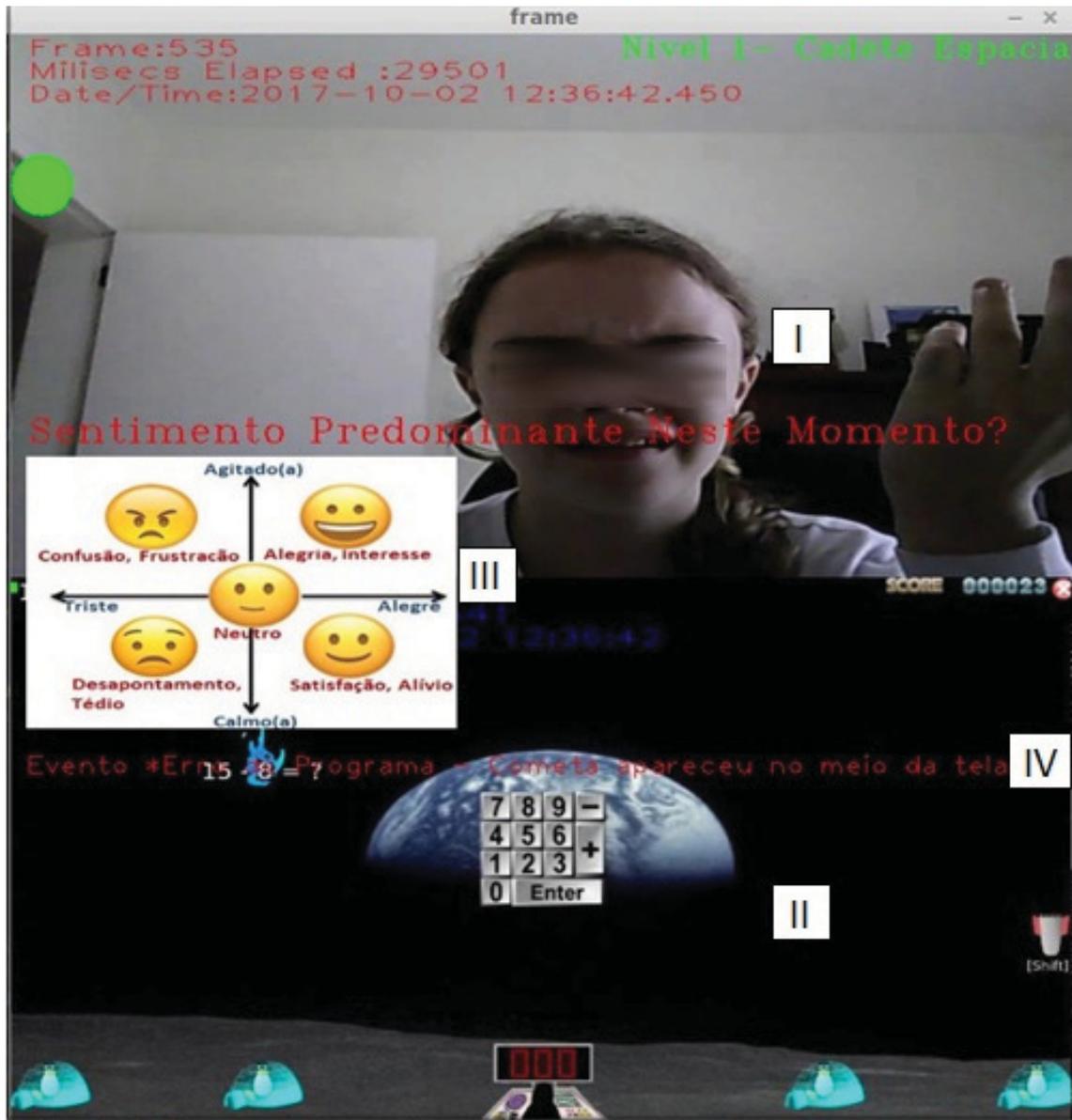


Figura 5.4: Interface da ferramenta de rotulagem da base de dados no instante da ocorrência de um evento monitorado.

A aplicação da abordagem de rotulagem de dados descrita acima teve como objetivo facilitar a atividade para o estudante, buscando-se também aumentar qualidade dos dados coletados. Entende-se que uma fator importante para isto reside no fato da ferramenta permitir uma visão retrospectiva completa (não apenas nos instantes dos eventos monitorados), sem interromper a realização da atividade de aprendizagem. Sendo assim, todas as reações físicas do estudante, sincronizada com os eventos e ações realizadas na interface do jogo são consideradas.

Um aspecto importante a ser observado no processo de rotulagem refere-se à estratégia utilizada para que os estudantes avaliem suas emoções durante o experimento com o software educacional. A abordagem clássica conhecida como *Self-Assessment Manikin - SAM* (Bradley e Lang, 1994) é uma ferramenta de avaliação de respostas afetivas que tem sido usada em diferentes áreas (Hayashi et al., 2016). Como pode ser visto na Figura 5.5, o SAM representa as emoções expressas fisicamente em termos de expressão facial e reações corporais de bonecos. A Figura

5.5 mostra um SAM para a avaliação afetiva das dimensões valência, ativação e dominância, nesta ordem.

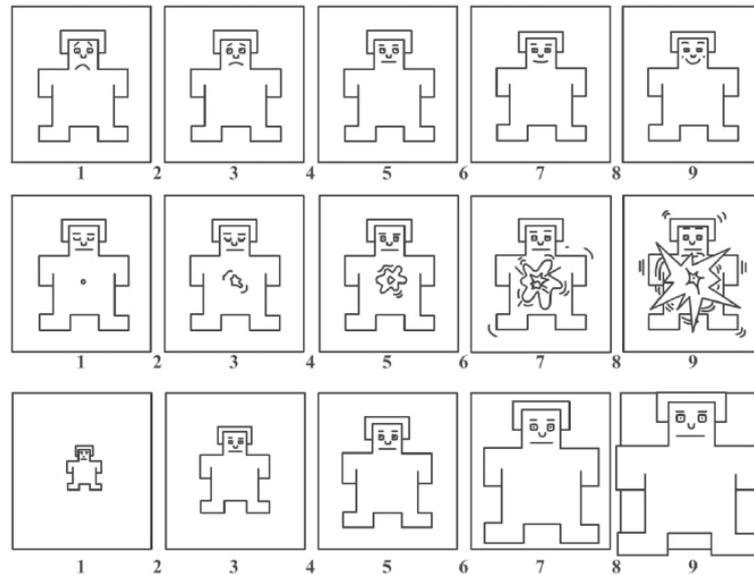


Figura 5.5: Representação Clássica da Ferramenta SAM: Fonte Bradley e Lang (1994)

Entretanto, pesquisas com as de Yusoff et al. (2013) e Hayashi et al. (2016) têm demonstrado que crianças encontram dificuldades de uso e entendimento da ferramenta SAM original, reportando não gostar da aparência e cor desta ferramenta. Neste sentido o trabalho de (Hayashi et al., 2016) propõe substituir cada figura do SAM original por um **emoji** ou *emoticon* correspondente. Os resultados apresentados neste trabalho demonstraram melhorias no entendimento e uso por parte das crianças.

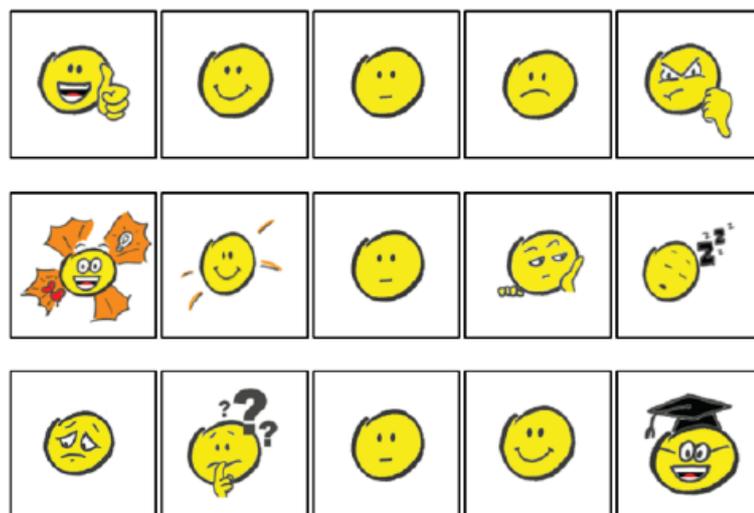


Figura 5.6: Proposta de Alterações ao SAM. Fonte: Hayashi et al. (2016)

Neste contexto, como o presente trabalho envolve crianças, considerou-se que seria importante utilizar uma ferramenta mais intuitiva que o SAM original. Desta forma, utilizando-se como referência a proposta de (Hayashi et al., 2016), desenvolveu-se uma alternativa customizada para o presente trabalho. A Figura 5.7 apresenta a abordagem de avaliação afetiva utilizada neste trabalho. Como neste trabalho utiliza-se apenas as dimensões valência e ativação a figura foi

adaptada para esta situação. Adicionalmente, optou-se por utilizar *emoticons* ou *emojis* familiares às crianças tendo em vista seu amplo uso em redes sociais e aplicativos de mensagens. A escolha dos *emoticons* para cada quadrante foi realizada com o auxílio de um especialista na área de Interação Humano-Computador.

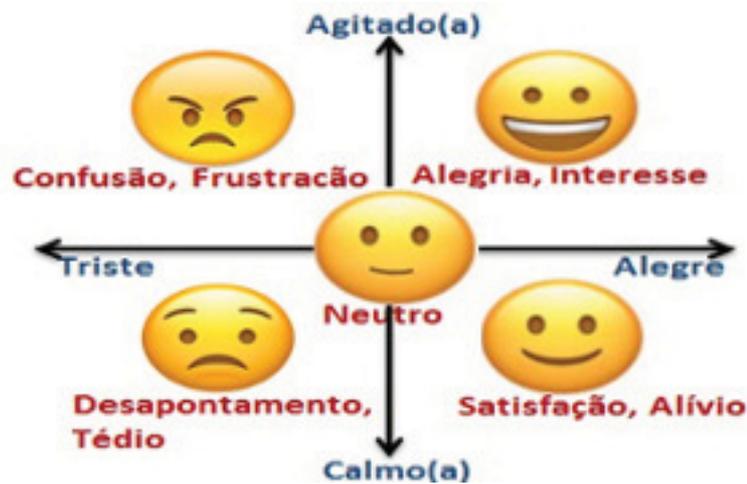


Figura 5.7: Ferramenta de Avaliação Afetiva Utilizada nesta Pesquisa

5.2.3 O Experimento

Com o auxílio da professora de matemática das turmas, o nível do jogo foi escolhido considerando-se a idade e a série de cada estudante. Orientou-se os estudantes do quinto ao sétimo ano a jogar os três níveis iniciais do TuxMath, ou seja: cadete especial, sentinela e defensor. Já os estudantes do oitavo e nono anos foram orientados a jogar os níveis sentinela, defensor e comando.

É importante observar que, por limitações de tempo dos experimentos, nem todos os estudantes jogaram exatamente três níveis. Estudantes que foram mais ágeis jogaram níveis adicionais, enquanto esperavam os colegas terminarem as atividades. Por outro lado, alguns estudantes com maior dificuldade não conseguiram completar os três níveis. Porém, todos os estudantes completaram no mínimo dois e no máximo 4 níveis. Esta estratégia foi usada buscando maximizar a quantidade de dados coletados e evitar que os estudantes que terminaram as atividades atrapalhem os demais. Entendeu-se que a diferença entre o número de níveis não teria impacto significativo nos resultados.

Enquanto os estudantes jogavam, uma lista de eventos foi monitorada pela versão personalizada do TuxMath. A Tabela 5.1 apresenta a lista completa dos eventos monitorados, juntamente com a descrição e a intensidade atribuída a cada evento. Nesta tabela também são mostradas duas colunas com as etiquetas do modelo ALMA (conforme Tabela 4.2) atribuídas para cada evento, de acordo com a expectativa e objetivos informados pelos estudantes. Buscando criar artificialmente situações que poderiam induzir emoções, desenvolveu-se um procedimento gerador de *bugs* aleatórios no Tuxmath. Sempre que um *bug* é inserido artificialmente por este procedimento, também se torna um evento monitorado (últimas quatro linhas da Tabela 5.1).

Evento Monitorado	Descrição do Evento	Etiqueta ALMA (expectativa de vencer)	Etiqueta ALMA (expectativa de perder)	Intensidade
Início do Jogo	Início de cada nível do jogo	GoodLikelyFutureEvent	BadLikelyFutureEvent	0,7

Venceu Onda	Estudante destruiu metade ou mais dos cometas da onda	EventConfirmed	EventDisConfirmed	0,5
Perdeu Onda	Estudante destruiu menos da metade dos cometas da onda	EventDisConfirmed	EventConfirmed	0,5
Venceu Nível	Após a liberação dos 25 cometas do nível, ainda restaram no jogo, ao menos, um pinguim vivo	EventConfirmed	EventDisConfirmed	0,7
Perdeu Nível	Todos os pinguins foram mortos antes de se alcançar o total de 25 cometas do nível	EventDisConfirmed	EventConfirmed	0,7
Capturou Cometa de Super Poderes	Cometa especial foi capturado (i.e resposta correta à equação do cometa especial), permitindo a utilização da arma chamada 'super bomba' que destrói de uma vez todos os cometas visíveis na tela	GoodEvent	GoodEvent	0,4
Perdeu Cometa com Super Poderes	Cometa com super poderes passou pela tela e não foi capturado (i.e não teve a equação respondida corretamente), impedindo o uso da 'super bomba'	BadEvent	BadEvent	0,4
Uso da Bomba Inteligente	Foi feito uso da 'super bomba' para a destruição simultânea de todos os cometas visíveis na tela	GoodEvent	GoodEvent	0,4
Cometa Atingiu Iglu ou Pinguim	Cometa atingiu e destruiu o iglu ou pinguim. O primeiro cometa atinge e destrói o iglu que protege o pinguim. Sem proteção, se um próximo cometa atingir a posição irá matar o pinguim	BadEvent	BadEvent	0,3

Nuvem Reconstrói Iglu	Caso o estudante tenha destruído um cometa bônus e tenha tido um iglu destruído, após o final da onda, iglu será reconstruído por uma nuvem que libera flocos de neve	GoodEvent	GoodEvent	0,3
Erro do Programa - Aumentou a Velocidade!!!	<i>Bug</i> aleatório que aumenta a velocidade de quedas dos cometas em aproximadamente 200% por 3 segundos	BadEvent	BadEvent	0,5
Erro do Programa - Cometa apareceu no meio da tela !!!!	<i>Bug</i> aleatório que faz um cometa aparecer no meio da tela ao invés de aparecer na parte superior. Com isto, o tempo para que o estudante responda à equação deste cometa, evitando que atinga um iglu ou pinguim é diminuído pela metade	BadEvent	BadEvent	0,5
Erro do Programa - Cometa foi Destruído Automaticamente !!!!	<i>Bug</i> aleatório que faz com que um cometa seja destruído sem nenhuma ação do estudante	GoodEvent	GoodEvent	0,5
Erro do Programa - Resposta era Correta !!!!	<i>Bug</i> aleatório que faz com que uma resposta correta não destrua o respectivo cometa	BadEvent	BadEvent	0,5

Tabela 5.1: Eventos monitorados no TuxMath durante os experimentos

Os eventos são utilizados como entrada para o componente cognitivo do ModHEmo. Após a ocorrência de um evento monitorado no TuxMath, a imagem do rosto do aluno foi capturada com uma webcam básica, configurada a 30 *frames* por segundo (fps), sendo esta imagem usada como entrada para o componente físico do ModHEmo. Com estas entradas, o modelo é então executado, resultando na inferência da provável emoção do estudante.

5.3 Resultados dos experimentos

Nesta seção, serão descritos os principais resultados obtidos com os experimentos descritos nas seções anteriores deste capítulo. Com o objetivo de melhor organizar esta seção, a apresentação dos resultados foi dividida em nove subseções cujos objetivos e tipos de informação são apresentadas na Tabela 5.2. O tipo de informação ‘consolidada’ representa que os valores são relativos a todos os 32 estudantes, enquanto ‘individual’ refere-se a apenas um estudante.

Tabela 5.2: Seções de resultados e seus objetivos

Seção	Objetivo	Tipo de Informação
5.3.1 Informações gerais dos experimentos	Descrever informações globais que são úteis para o entendimento dos demais resultados	Consolidada
5.3.2 Resultados detalhados de um estudante específico	Apresentar e discutir resultados detalhados das inferências para um estudante específico	Individual
5.3.3 Dinâmica de transição e persistência das emoções	Apresentar e discutir informações que auxiliam a entender a dinâmica de transição das emoções de um estudante específico	Individual
5.3.4 Descrição dos indicadores utilizados para apresentar e avaliar os resultados	Apresentar conceitualmente os indicadores que foram utilizados para avaliação dos resultados consolidados	NA
5.3.5 Resultados das inferências do ModHEmo com a fusão dos componentes	Descrever os resultados considerando a fusão dos componentes físico e cognitivo do modelo de inferência. Esta subseção representa a proposta original do ModHEmo	Consolidada
5.3.6 Resultados das inferências usando apenas o componente físico do ModHEmo	Mostrar os resultados de um cenário em que se utiliza somente o componente físico.	Consolidada
5.3.7 Resultados das inferências usando apenas o componente cognitivo do ModHEmo	Mostrar os resultados de um cenário em que se utiliza somente o componente cognitivo.	Consolidada
5.3.8 Resultado das inferências considerando apenas dimensão valência	Apresentar resultados de um cenário que utiliza apenas a dimensão valência.	Consolidada
5.3.9 Resultado das inferências considerando apenas dimensão ativação	Apresentar resultados de um cenário que utiliza apenas a dimensão valência.	Consolidada

É importante destacar que as quatro últimas subseções (5.3.6, 5.3.7, 5.3.8 e 5.3.9) representam alguns cenários hipotéticos. Estes cenários demonstram a flexibilidade de efetuar pequenas alterações nas configurações do ModHEmo para obter diferentes tipos de informações. Adicionalmente, cabe ressaltar que nas cinco últimas subseções (5.3.5, 5.3.6, 5.3.7, 5.3.8 e 5.3.9) são apresentados individualmente os resultados dos algoritmos RandomForest e IBK.

5.3.1 Informações gerais dos experimentos

Nos experimentos realizados foram gerados um total de 1.669 eventos (conforme Tabela 5.1), perfazendo uma média geral de 52,16 ($\sigma = 16,77$) eventos por estudante. A Tabela 5.3

detalha o número de ocorrência para cada um dos eventos monitorados e o percentual do total que cada evento corresponde. A tabela está ordenada de forma decrescente do número de eventos.

Tabela 5.3: Detalhamento da ocorrência de eventos monitorados nos experimentos

Descrição do Evento	Número de Ocorrências	Percentual
Cometa Atingiu Iglu ou Pinguim	698	41,82%
Venceu Onda	325	19,47%
Início do Jogo	108	6,47%
Perdeu Onda	92	5,51%
Erro do Programa - Cometa apareceu no meio da tela !!!!	76	4,55%
Erro do Programa - Aumentou a Velocidade !!!!	62	3,71%
Erro do Programa - Resposta era Correta !!!!	61	3,65%
Venceu Jogo!!!!	60	3,59%
Capturou Cometa de Super Poderes	59	3,54%
Nuvem Reconstrói Iglu	32	1,92%
Perdeu Jogo	30	1,80%
Uso da Bomba Inteligente	29	1,74%
Perdeu Cometa com Super Poderes	21	1,26%
Erro do Programa - Cometa foi Destruído Automaticamente !!!!	16	0,96%
Total	1669	100,00%

Analisando-se os resultados expressos na Tabela 5.3, pode-se observar que o evento “Cometa Atingiu Iglu ou Pinguim” teve um número de ocorrência (698) significativamente superior aos demais eventos. Este resultado era esperado, pois são exibidos em cada nível do jogo 25 cometas para cada estudante. Considerando-se que cada um dos 32 estudantes jogou em média aproximadamente 3 níveis diferentes do jogo, no total foram exibidos aproximadamente 2400 cometas. Desta forma, 698 (29%) não tiveram a equação matemática respondida corretamente, atingindo então os iglus ou pinguins.

Destaca-se também na Tabela 5.3 a relação entre os eventos “Venceu Jogo!!!!” e “Perdeu Jogo”. Pode-se observar que os níveis vencidos pelos estudantes (60) é o dobro do número de níveis perdidos (30). Adicionalmente, observa-se que a soma destes dois eventos (90) é diferente ao valor do evento “Início do Jogo” (108). Esta situação, ocorre devido ao fato de que estudantes não rotularam completamente 18 níveis do jogo.

Conforme já descrito na Seção 4.4, no processo de fusão dos componentes físico e cognitivo do ModHEmo é gerado um conjunto de dados com as inferências de cada componente. Este conjunto de dados tem um total de 1.669 registros que correspondem a cada um dos eventos monitorados que estão destacados na Tabela 5.3. A Tabela 5.4 apresenta um detalhamento deste conjunto de dados. Nesta tabela apresenta-se o nome do atributo e sua descrição, bem como medidas estatísticas que descrevem a distribuição dos valores destes atributos. Além disso, a última linha da tabela mostra o atributo que contém as classes atribuídas pelos estudantes no processo de rotulagem descrito anteriormente. O conjunto de dados apresentado na Tabela 5.4 é de fundamental importância para a inferência final das emoções, pois serve como base

para a construção, treinamento e testes dos modelos de classificação criados com os algoritmos RandomForest e IBK.

Tabela 5.4: Detalhamento do conjunto de dados obtidos no processo de fusão dos componentes físico e cognitivo do ModHEmo

Nome do Atributo	Descrição	Mínimo	Máximo	Média	Desvio Padrão
Q1C	Valor dos scores do quadrante 1 do componente cognitivo	0	1	0,209	0,316
Q2C	Valor dos scores do quadrante 2 do componente cognitivo	0	1	0,065	0,211
Q3C	Valor dos scores do quadrante 3 do componente cognitivo	0	1	0,516	0,432
Q4C	Valor dos scores do quadrante 4 do componente cognitivo	0	1	0,202	0,324
QNC	Valor dos scores da classe “neutro” do componente cognitivo	0	1	0,001	0,024
Q1F	Valor dos scores do quadrante 1 do componente físico	0	1	0,182	0,368
Q2F	Valor dos scores do quadrante 2 do componente físico	0	1	0,117	0,263
Q3F	Valor dos scores do quadrante 3 do componente físico	0	0,88	0,081	0,138
Q4F	Valor dos scores do quadrante 4 do componente físico	0	0,85	0,025	0,128
QNF	Valor dos scores da classe “neutro” do componente físico	0	1	0,593	0,437
QUADL	Classe obtida no processo de rotulagem do conjunto de dados				

As classes atribuídas pelos estudantes (atributo QUADL da Tabela 5.4) no processo de rotulagem da base de dados têm uma distribuição conforme mostrado no gráfico da Figura 5.8. Observa-se que a classe QN, ou seja, estado neutro, foi a que teve maior frequência, atingindo o valor de 513 (aproximadamente 31% dos 1.669 registros do conjunto de dados). Por outro lado, a classe Q4 (reconstrução) com 226 registros (aproximadamente 14%) obteve a menor frequência.

A frequência de ocorrência das diferentes emoções reportados pelos estudantes e apresentados na Figura 5.8 apresentam-se compatíveis com o que costuma ocorrer com estudantes realizando atividades em um software educacional. A maior ocorrência do estado neutro era esperada para o tipo de atividade proposta e mostra-se compatível com o resultado obtido no trabalho de D’Mello et al. (2007). Além disso, a maior ocorrência da classe Q2 que engloba emoções como confusão e frustração confirma que estas reações afetivas são comuns em atividades envolvendo aprendizagem (Pour et al., 2010; Baker et al., 2010) e que não significam necessariamente um estado a ser evitado, podendo estar correlacionado positivamente com aprendizagem (D’Mello et al., 2007).

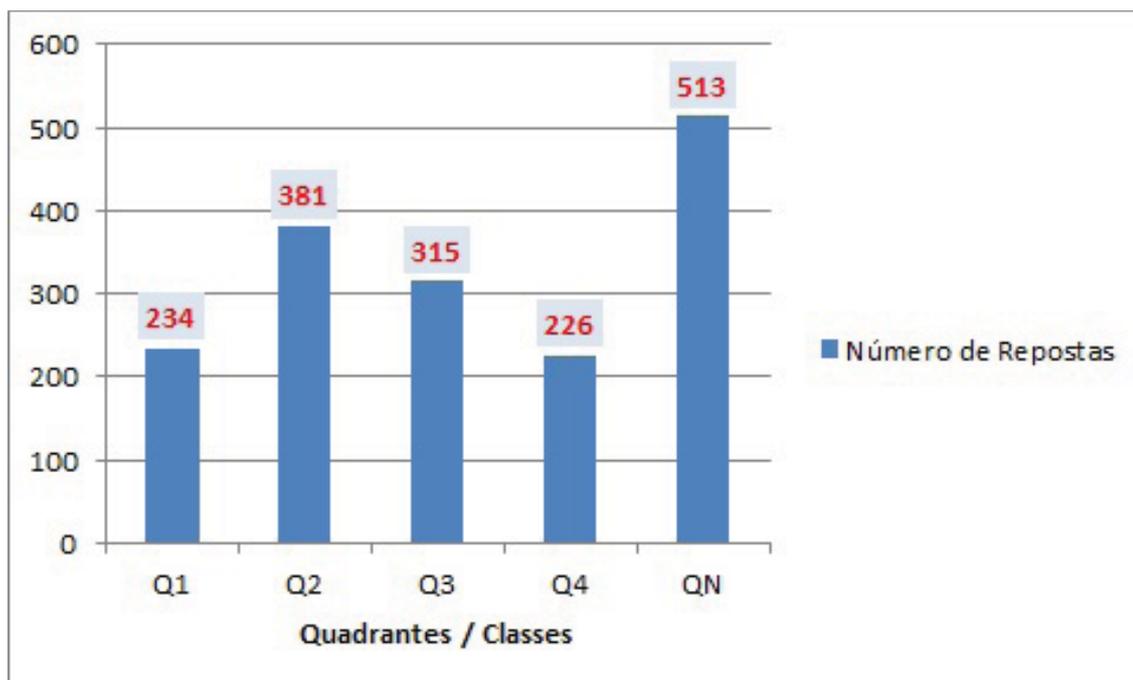


Figura 5.8: Distribuição das classes obtidas no processo de rotulagem

5.3.2 Resultados detalhados de um estudante específico

Neste tópico serão apresentados resultados detalhados do processo de inferência do ModHEmo para um estudante específico² que foi escolhido por ter um número de eventos (50) próximo da média geral (52,16). Para isso, foram gerados dois modelos de classificação, sendo um para o algoritmo RandomForest e outro para o IBK. Na sequência estes modelos foram executados usando-se os eventos e as imagens da face do estudante escolhido como parâmetros de entrada, simulando-se assim a aplicação real do modelo construído.

Desta forma, foi possível obter as inferências detalhadas do ModHEmo e compará-las com a rotulagem feita pelo estudantes. Na Figura 5.9 observa-se uma representação gráfica temporal exibindo a aderência entre os quadrantes rotulados pelo estudante (linha azul) e as inferências do ModHEmo (linha vermelha). O eixo horizontal do gráfico da Figura 5.9 representa os instantes em que os eventos ocorreram e o eixo vertical representa os quadrantes ou classes. A ordem dos quadrantes no gráfico foi organizada de forma que classe mais positiva (“Q1” valência e ativação positiva) é colocada no topo e a mais negativa (“Q3” valência e ativação negativa) na parte inferior, com o quadrante neutro (QN) no centro. Os instantes ou intervalos em que as linhas estão sobrepostas (66% do tempo) significam que as inferências do ModHEmo coincidem com as classificações reais (rotuladas pelo estudante).

Pode-se observar analisando gráfico da Figura 5.9 que, mesmo em instantes que não há coincidência exata, em determinados momentos as linhas com os valores reais e inferidos apresentam-se em quadrantes próximos. Este comportamento pode ser observado, por exemplo, entre os instantes “09:22:59” e “09:23:34” (número 1 na Figura 5.9). Neste período três erros ocorreram: i) classificar como Q1 quando o real era QN, ii) classificar como QN quando o real era Q2 e iii) classificar como Q4 quando o real era Q1. Portanto, mesmo sem coincidência exata em todos os momentos, os erros ocorridos não foram extremos. Um erro extremo poderia ser, por exemplo, classificação com valor Q1 e real ser Q3 ou vice versa.

²O nome do estudante não será revelado por questões de confidencialidade.

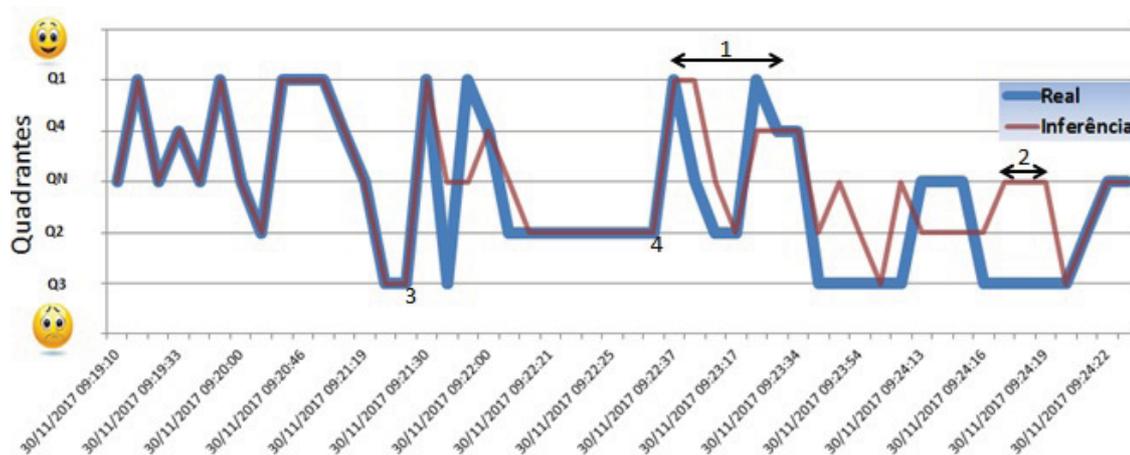


Figura 5.9: Aderência entre a rotulagem e a inferência do ModHEmo para um estudante específico

No intervalo aproximado entre 09:24:17 e 09:24:19 (número 2 na Figura 5.9) ocorreram as maiores divergências contínuas entre as linhas. Neste período o resultado da classificação foi QN quando o real seria Q3. Porém, mesmo neste intervalo, não são observadas inconsistências extremas, como, por exemplo, o quadrante inferido ser “Q1” e o real “Q3”.

A ocorrência de erros de classificação é uma condição comum em modelos de classificação, como os utilizados nesta pesquisa. Além disso, considerando-se a complexidade e a incerteza envolvidas no processo de inferência de emoções já discutidas neste trabalho, erros de classificação são esperados. Neste sentido, Picard (1997) lembra que mesmo observadores humanos apresentam dificuldades e discordam ao realizar classificação de emoções de outras pessoas. Como exemplo, no trabalho de Lagud e Rodrigo (2010) o índice Cohen Kappa (Witten et al., 2016), que mede a concordância entre observadores, foi 0,63³. O trabalho de Robson (1993) indica que um especialista humano identifica emoção com 75% de precisão.

Desta forma, buscando identificar possíveis causas dos erros de classificação do ModHEmo, foi realizada uma análise nos dados recolhidos nos experimentos realizados. Um erro recorrente (como mostrado no número 2 na Figura 5.9), foi a classificação no quadrante QN de eventos rotulados pelo estudante em outros quadrantes. Analisando-se os logs com os valores individuais destes eventos, percebeu-se que o componente físico teve valores altos (normalmente próximo a 1) no quadrante neutro, influenciando assim a decisão final.

Diante disso, uma análise da classificação das imagens da face dos estudantes realizadas pelo componente físico revelou situações que poderiam gerar erros de classificação. Um exemplo desta situação é mostrado na Figura 5.10. Neste caso, a mão oculta parte da face do estudante, incluindo a região da boca, que é um componente fundamental das expressões faciais.

Sobrepostos à imagem do estudante na Figura 5.10 pode-se verificar o resultado da inferência do componente físico (retângulo azul), cujo resultado foi Neutro (Quadrante QN). Observa-se, entretanto, que o score individual (retângulo vermelho) de 0,3 da emoção surpresa (*surprise*) é significativo, podendo eventualmente ser superior aos demais (em função da posição da boca) e mudar o resultado final.

Em outros casos, devido ao posicionamento, não foi possível reconhecer a face do estudante. Nestes casos a inferência do componente físico ficou indisponível, restando apenas os dados cognitivos para decisão do resultado final.

É importante destacar que estas dificuldades relacionadas à postura eram esperadas visto que não foram impostas restrições de posicionamento e movimentação aos estudantes.

³Cohen Kappa igual a 1 indica concordância perfeita

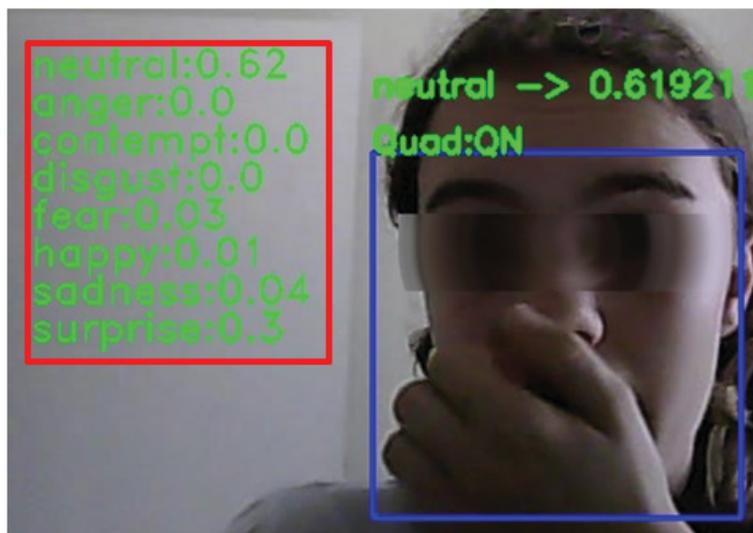


Figura 5.10: Estudante com a mão ocultando parte da face

Porém, entendeu-se que restringir movimentos e posição agregaria uma condição artificial, não representando adequadamente um ambiente real de ensino.

Observou-se ainda que o processo de inferência das emoções via expressão facial também apresenta erros que não são relacionados à posicionamento. Foi possível identificar esta situação através da análise manual das imagens da face dos estudantes comparando-as com a classificação feita pela ferramenta EmotionAPI. Por questão de confidencialidade, decidiu-se não apresentar exemplo destes erros, pois seria necessário desfocar partes da face (olhos, boca) para ocultar a identidade do estudante. Porém, verificou-se que este processo de ocultação da identidade acaba também impossibilitando a análise visual da figura a ser inserida neste documento.

Nas considerações finais deste trabalho serão apresentadas algumas propostas de trabalhos futuros que poderiam tratar as dificuldades apresentadas acima.

5.3.3 Dinâmica de transição e persistência das emoções

A dinâmica das transições, persistência e quantidade de ocorrência das emoções são a base para a inferência de estados emocionais de alto nível relacionados à aprendizagem como humor do aprendiz, ciclo vicioso e concentração engajada (ver Seção 2.6). Desta forma, com o objetivo de identificar a dinâmica das relações entre emoções, serão apresentados a seguir grafos com diferentes visões e em diferentes instantes da sessão de jogo do estudante cujos dados foram apresentados na Figura 5.9. Serão mostradas duas versões dos grafos: uma versão baseada nas classes reais (rotuladas pelos estudantes) e outra utilizando as inferências do ModHEmo. Desta maneira, pode-se comparar também a aderência das inferências do ModHEmo com as classes reais.

Os grafos foram gerados utilizando a ferramenta gratuita Gephi ⁴. Os nós do grafo representam os quadrantes com o tamanho do nó sendo definido com o valor do grau ponderado deste nó (incidência) calculado pela ferramenta Gephi. O valor do grau ponderado é mostrado em cada um dos nós. As arestas representam as transições (incluindo persistência) entre os quadrantes. A cor da aresta é herdada do nó de origem e a espessura é proporcional ao peso da aresta no grafo. Em cada aresta é mostrado numericamente o peso desta aresta. Para facilitar a

⁴<https://gephi.org/>

visualização, nas arestas com peso igual a 1 (um) foram omitidos os valores numéricos. Além disso, os nós do grafo foram posicionados de forma a refletir as emoções representadas pelos quadrantes utilizados para o processo de inferência do ModHEmo (ver Figura 3.5). Em todos os grafos apresentados nesta seção o quadrante inicial é o neutro (QN), conforme pode ser observado na Figura 5.9.

As Figuras 5.11 e 5.12 apresentam grafos considerando todo o intervalo de tempo de jogo do estudante sendo tratado nesta subseção. De forma a permitir uma comparação, a Figura 5.11 apresenta o grafo com as classes reais (rotulada pelos estudantes) e a Figura 5.12 apresenta o grafo baseado nas inferências do ModHEmo.

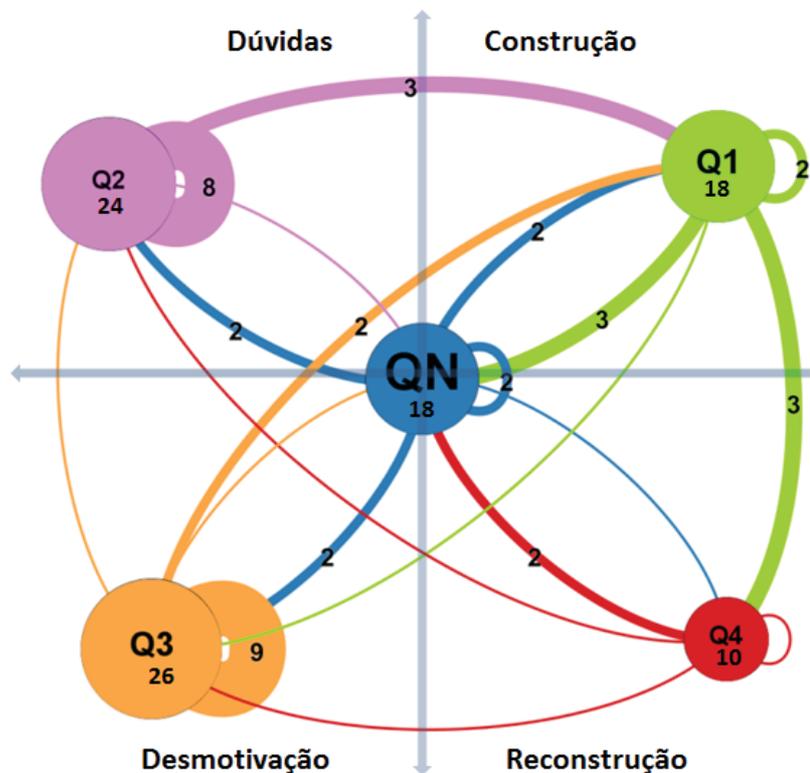


Figura 5.11: Grafo de sessão completa do estudante com as classes reais

Comparando-se os grafos inferidos e real apresentados nas Figuras 5.11 e 5.12 pode-se perceber que no grafo com os dados reais os quadrantes que mais se destacam são o Q2 e Q3 cujo grau ponderado é 24 e 26, respectivamente. No grafo com as inferências do ModHEmo pode-se perceber que os quadrantes mais destacados são o QN e Q2 com grau ponderado igual a 28 e 32, respectivamente. Em comum, o estado Q2 aparece nos dois grafos, indicando um “humor do aprendiz” no quadrante denominado como “Dúvidas”.

Um diferença relevante entre os grafos das classes reais (Figura 5.11) e classes inferidas (Figura 5.12) é em relação a reincidência da classe Q3. No grafo com as informações inferidas as arestas ligando o quadrante Q3 a ele mesmo são menores (ver espessura da aresta). Desta forma, neste caso, possíveis ciclos viciosos não seriam detectados corretamente pelo ModHEmo.

É importante destacar também a persistência que é observada nos estados Q2 e Q3, pois estas situações são indicativas de um possível ciclo vicioso. Nestes casos, intervenções automáticas poderiam ser pertinentes, buscando trazer o estudante para um estado mais favorável ao aprendizado.

Conforme observam (D’Mello et al., 2007) mesmo estados transitórios podem ser problemáticos se estes transitarem para outros estados negativos. Em nenhum dos dois grafos

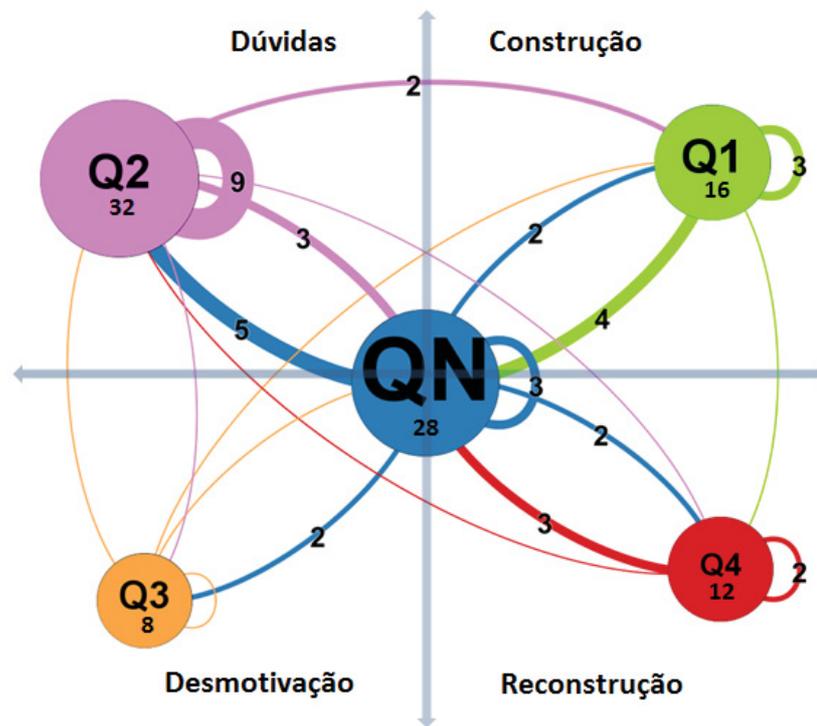


Figura 5.12: Grafo de sessão completa do estudante com as classes inferidas pelo ModHEmo

das Figuras 5.11 e 5.12 é observada esta situação visto que não existem transições recorrentes e significativas entre os estados Q2 e Q3 que também indicariam ciclos viciosos.

Os grafos considerando o tempo total da sessão do estudante apresentam uma visão estática das interações ocorridas durante a sessão inteira. Porém, com a ferramenta Gephi, é possível criar uma visão dinâmica da evolução das transições e interação entre as emoções. Nesta ferramenta, as alterações no grafo podem ser acompanhadas sequencialmente na forma de um vídeo. As Figuras 5.13 e 5.14 apresentam uma visão do gráfico no instante “30/11/2017 09:20:08” (Número 3 da Figura 5.9). Neste instante o estudante finalizou o primeiro nível do jogo.

Na primeira sessão representada nos grafos das Figuras 5.13 e 5.14 pode-se perceber uma transição constante entre os quadrantes Neutro e Q1, sem indicativos de ciclos viciosos e com poucas ocorrências de estados negativos. Adicionalmente os grafos com as classes reais e inferidas são idênticos, indicando uma perfeita correspondência das inferências do ModHEmo.

As Figuras 5.15 e 5.16 apresentam uma visão do gráfico no instante “30/11/2017 09:22:37” (Número 4 da Figura 5.9). Neste instante o estudante finalizou o segundo nível do jogo. Ambos os grafos das Figuras 5.15 e 5.16, apesar de não ter uma correspondência total entre real e inferido, refletem a reincidência dos estados Q2 e Q3 (principalmente Q2), indicando já neste instante um possível ciclo vicioso em andamento com possíveis ações de remediação sendo necessárias.

5.3.4 Descrição do indicadores utilizados para avaliar os resultados dos experimentos

Para permitir uma avaliação completa dos resultados e também para viabilizar comparações com trabalhos correlatos, serão apresentados um conjunto de medidas repetidamente utilizadas para avaliar o desempenho de modelos de classificação, incluindo:

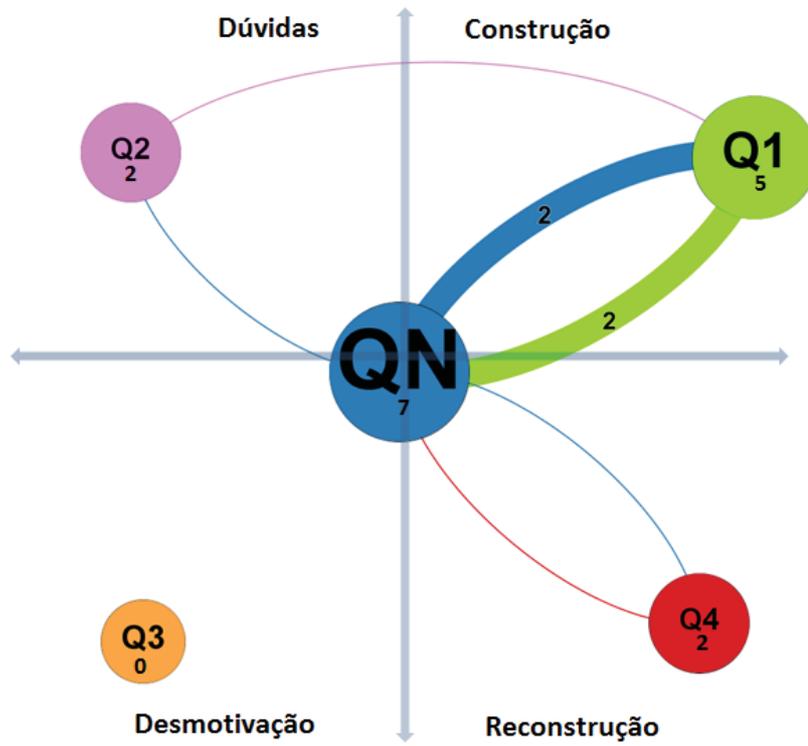


Figura 5.13: Gráfico dinâmico com as classes reais no instante "30/11/2017 09:20:08"

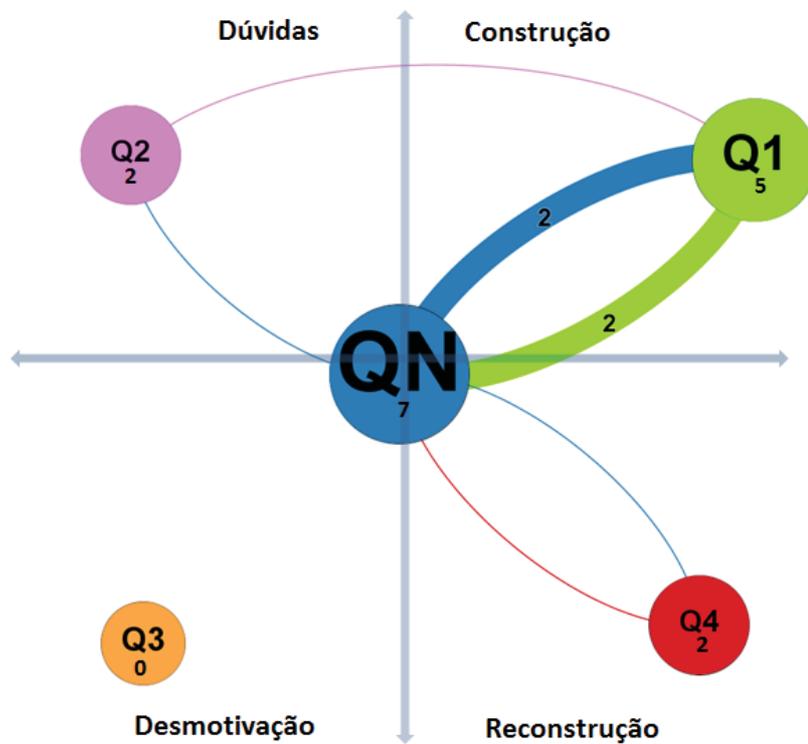


Figura 5.14: Gráfico dinâmico das classes inferidas pelo ModHEmo no instante "30/11/2017 09:20:08"

- Acurácia Global (percentual e número de instâncias) e Índice Cohen Kappa
- Matriz de Confusão

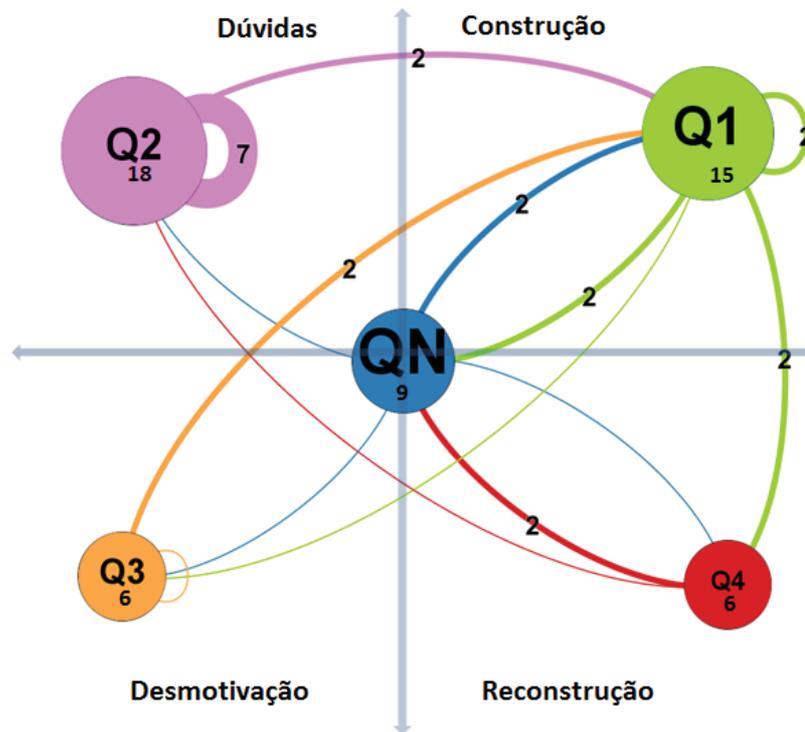


Figura 5.15: Grafo dinâmico com as classes reais no instante “30/11/2017 09:22:37”

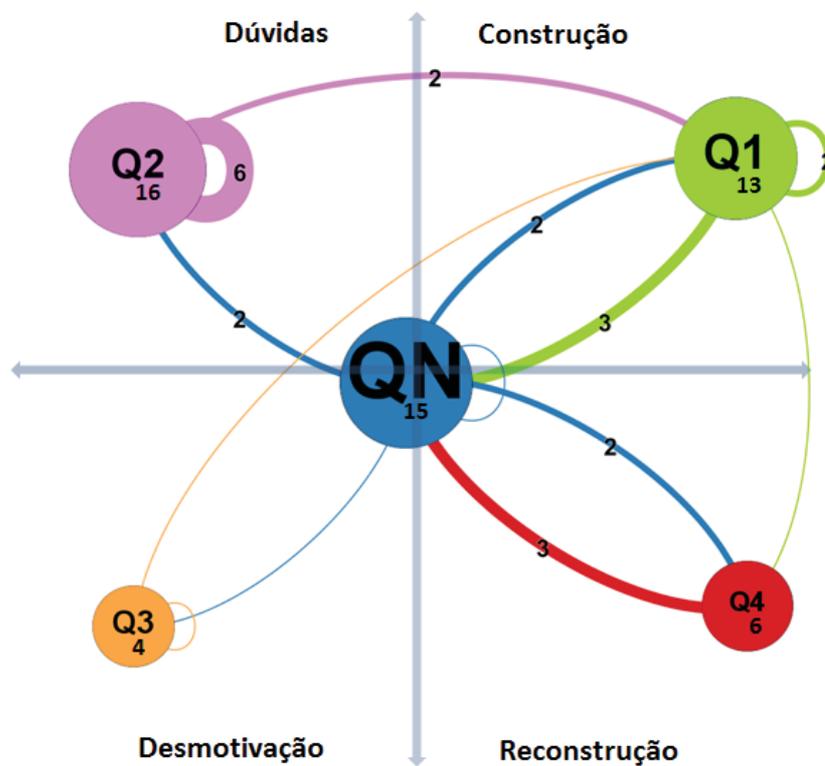


Figura 5.16: Grafo dinâmico das classes inferidas pelo ModHEmo no instante “30/11/2017 09:22:37”

- Índices de Efetividade por Classe : Verdadeiro Positivo, Falso Positivo, Precisão, Estatística F e AUC.

- Curvas ROC (*Receiver Operating Characteristics*)

A estatística Cohen Kappa é usada para medir a concordância entre inferências e a rotulagem de um conjunto de dados, corrigindo os acertos que ocorrem por acaso. O valor máximo deste índice é 1 e o valor esperado para um classificador randômico é zero (Witten et al., 2016).

Em problemas de classificação multi classe, os resultados podem ser exibidos como uma matriz de confusão bidimensional com uma linha e coluna para cada classe. Cada elemento da matriz mostra o número de instâncias da base de teste para os quais a classe real é a linha e a classe inferida é a coluna. Bons resultados correspondem a valores elevados nos elementos diagonal principal e valores baixos, idealmente zero, fora da diagonal principal (Witten et al., 2016). Para facilitar a visualização, foi destacado com fundo cinza a diagonal principal das matrizes de confusão a serem exibidas na sequência desta seção.

Os índices de efetividade por classe a serem apresentados neste trabalho permitem a avaliação do desempenho do ModHEmo especificamente para cada classe utilizada nos experimentos. A definição destes índices apresentada a seguir foi obtida em (Witten et al., 2016).

A taxa de “Verdadeiro Positivo” indica a relação entre instâncias corretamente classificadas de uma dada classe e o total de instâncias desta classe. Este índice corresponde à acurácia para cada classe. A taxa de “Falso Positivo” indica as instâncias incorretamente classificadas para uma determinada classe. A “Precisão” é a proporção de instâncias que foram corretamente classificadas em uma dada classe dividido pelo total de instâncias classificadas nesta classe. A “Estatística F” é uma medida obtida pela combinação entre precisão e recobertura (*recall*). O índice AUC (*Area Under the Curve*) relaciona-se diretamente com as curvas ROC, descritas a seguir.

Curvas ROC (*Receiver Operating Characteristics*) são um termo usado na detecção de sinal para caracterizar a compensação entre taxa de acerto e taxa de alarme falso em um canal com ruído. As curvas ROC descrevem o desempenho de um classificador sem considerar a distribuição de classes ou os custos dos erros. Estas curvas plotam a taxa de verdadeiro positivo no eixo vertical e a taxa de falso positivo no eixo horizontal. Um classificador ótimo apresentará a área sobre as curvas (AUC) ROC com valores próximos de 1, sendo 0,5 o valor esperado de um classificador randômico (similar a Cohen Kappa igual a 0) (Witten et al., 2016).

Na Figura 5.17 é mostrado um exemplo de três curvas ROC. A linha vermelha diagonal representa os valores esperados para um classificador aleatório. Valores acima desta diagonal, como a linha azul, representam classificadores que são superiores a um classificador aleatório. Por fim, um classificador perfeito é representado pela linha verde.

Na sequência desta seção serão apresentados os resultados das inferências realizadas pelo ModHEmo considerando o conjunto de indicadores descritos acima. Os valores dos indicadores dos modelos de classificação foram calculadas utilizando a ferramenta Weka.

Inicialmente, serão apresentados os resultados incluindo-se todos os atributos apresentados na Tabela 5.4. Como neste trabalho considera-se a hipótese que a fusão de dados físicos e cognitivos poderia melhorar a efetividade do processo de inferência, serão apresentados também dois cenários adicionais: em um dos cenários consideram-se apenas os atributos físicos e em outro somente os atributos cognitivos. Desta maneira, pretende-se avaliar os resultados da fusão dos componentes em comparação com os resultados individuais.

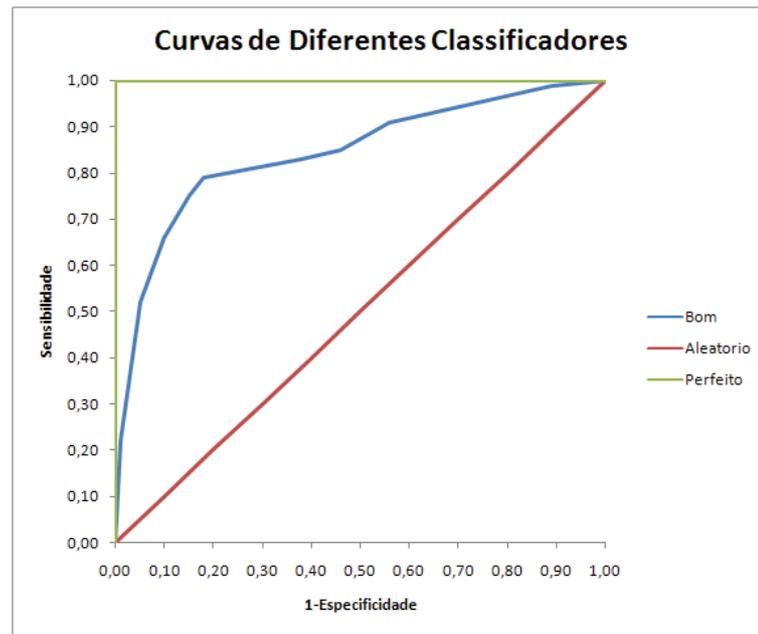


Figura 5.17: Exemplos de Curvas ROC

5.3.5 Inferências do ModHEmo com a fusão dos componentes físico e cognitivo

Considerando o conjunto completo de atributos apresentados na Tabela 5.4, o algoritmo RandomForest atingiu 1098 (65,79%) instâncias corretamente classificadas de um total de 1.669. Desta forma, 571 (34,21%) das instâncias foram classificadas de forma incorreta por este algoritmo. A estatística Kappa obtida foi de 0,5581.

A Tabela 5.5 apresenta a matriz de confusão com os resultados do algoritmo RandomForest utilizando o conjunto de dados completo. A diagonal principal da matriz de confusão, destacada com fundo cinza, representa as inferências corretas para cada uma das cinco classes consideradas nos experimentos.

Tabela 5.5: Matriz de confusão do algoritmo RandomForest usando conjunto de dados completo

Classe Correta \ Classe Prevista	QN	Q2	Q1	Q4	Q3
QN	365	37	25	35	51
Q2	63	264	17	16	21
Q1	35	21	161	9	8
Q4	46	23	21	127	9
Q3	70	41	9	14	181

Na Tabela 5.6 são apresentadas medidas de desempenho obtidas no processo de classificação dos quadrantes com o algoritmo RandomForest e o conjunto de dados completo.

Na Figura 5.18 é apresentado gráfico com a plotagem das curvas ROC para cada uma das cinco classes com os resultados do algoritmo RandomForest utilizando o conjunto de dados completo.

O algoritmo IBK alcançou 1055 (63,21%) instâncias corretamente classificadas de um total de 1.669. Portanto, 614 (36,79%) das instâncias foram classificadas de forma incorreta.

Tabela 5.6: Índices de desempenho por classe do algoritmo RandomForest utilizando o conjunto de dados completo

Classe	Verdadeiro Positivo	Falso Positivo	Precisão	Estatística F	AUC
QN	0,712	0,185	0,630	0,668	0,846
Q2	0,693	0,095	0,684	0,688	0,889
Q1	0,688	0,050	0,691	0,690	0,911
Q4	0,562	0,051	0,632	0,595	0,875
Q3	0,575	0,066	0,670	0,619	0,839
Média Geral	0,658	0,105	0,659	0,657	0,867

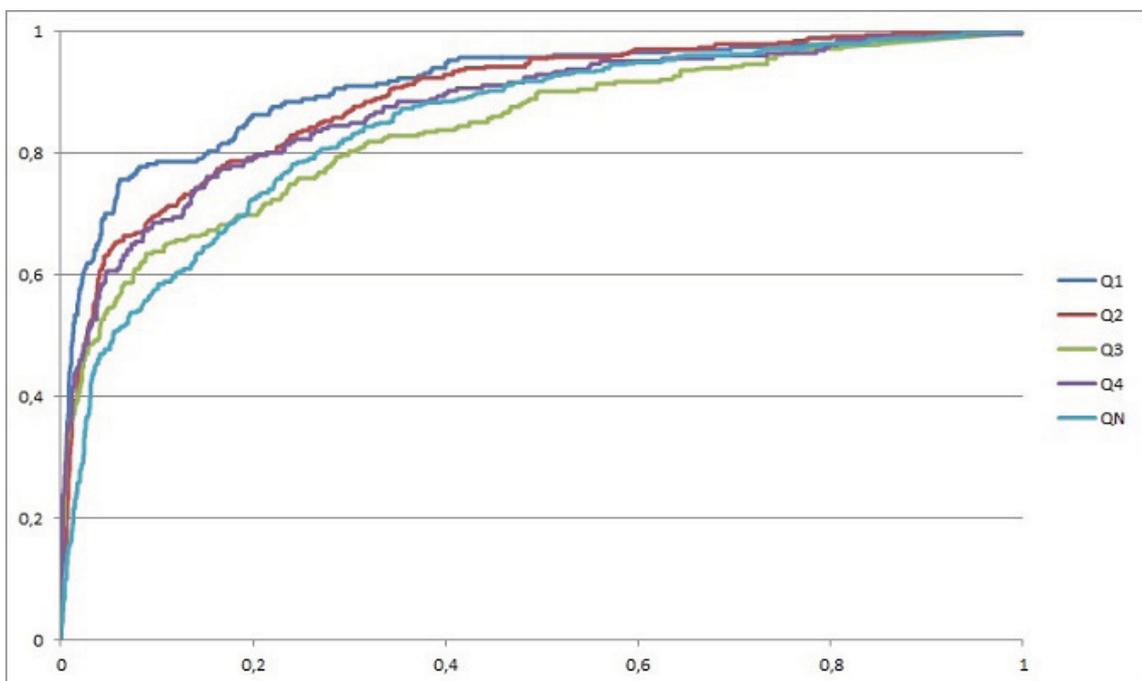


Figura 5.18: Curvas ROC por classe do algoritmo RandomForest utilizando o conjunto de dados completo

A estatística Kappa obtida foi de 0,5274. A Tabela 5.7 apresenta a matriz de confusão com os resultados do algoritmo IBK.

Tabela 5.7: Matriz de confusão do algoritmo IBK usando conjunto de dados completo

Classe Correta \ Classe Prevista	Classe Prevista				
	QN	Q2	Q1	Q4	Q3
QN	326	55	27	48	57
Q2	58	264	17	18	24
Q1	29	22	163	14	6
Q4	37	25	28	121	15
Q3	71	42	8	13	181

Na Tabela 5.8 são apresentadas medidas de desempenho obtidas no processo de classificação dos cinco quadrantes com o algoritmo IBK e o conjunto de dados completo.

Tabela 5.8: Índices de desempenho por classe do algoritmo IBK utilizando o conjunto de dados completo

Classe	Verdadeiro Positivo	Falso Positivo	Precisão	Estatística F	AUC
QN	0,635	0,169	0,626	0,631	0,781
Q2	0,693	0,112	0,647	0,669	0,830
Q1	0,697	0,056	0,671	0,683	0,851
Q4	0,535	0,064	0,565	0,550	0,744
Q3	0,575	0,075	0,640	0,605	0,800
Média Geral	0,632	0,108	0,631	0,631	0,800

Na Figura 5.19 é apresentado gráfico com a plotagem das curvas ROC para cada um das classes para o algoritmo IBK utilizando o conjunto de dados completo.

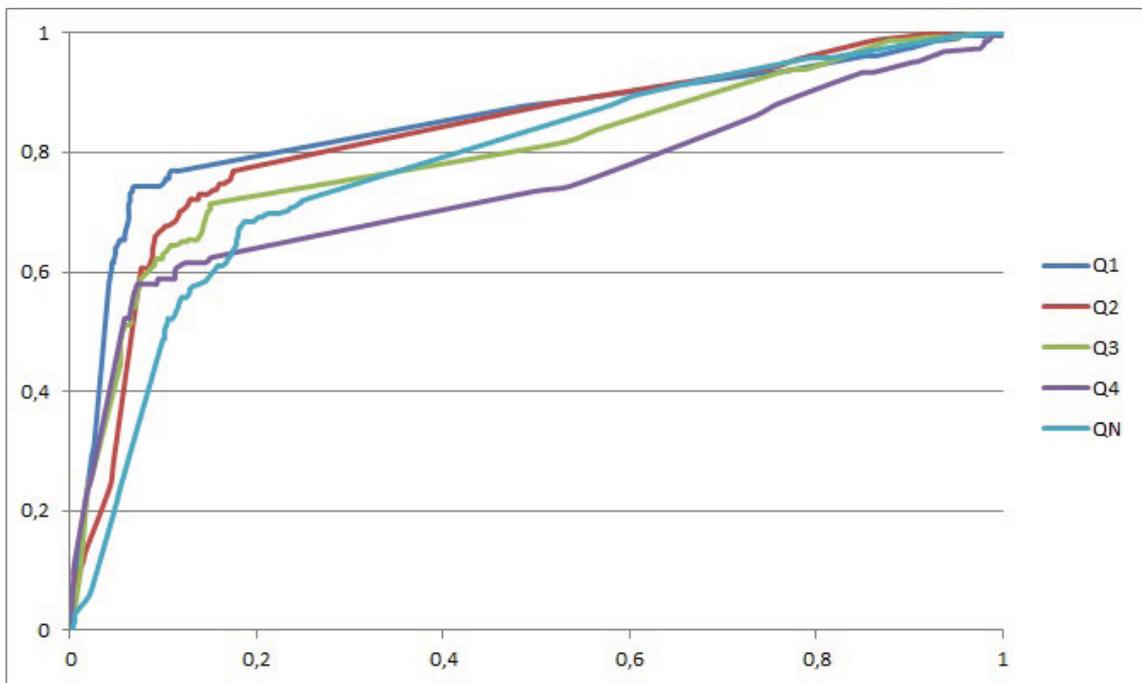


Figura 5.19: Curvas ROC por classe do algoritmo IBK utilizando o conjunto de dados completo

Conforme apresentado acima nesta seção o algoritmo RandomForest atingiu um índice de acurácia de 65,79% enquanto o algoritmo IBK obteve 63,21% neste índice. Para verificar a relevância estatística desta diferença foi realizado o teste T-Student Pareado (*Pair-Wise T-Test*) (Witten et al., 2016) com nível de significância de 0,05. O resultado deste teste demonstrou que existe diferença estatisticamente relevantes entre os resultados com o algoritmo RandomForest apresentando resultados superiores.

Considerando-se as medidas de desempenho dos dois algoritmos apresentadas nas Tabelas 5.6 e 5.8 pode-se perceber que os valores são semelhantes entre as cinco classes. Não se observa uma classe específica que tenha índices significativamente acima ou abaixo das demais e que possa distorcer a análise dos resultados globais ou por classe. Como comparação pode-se citar o trabalho de Doumbouya et al. (2018) em que as taxas de acurácia tiveram uma grande variação entre as classes de emoções, ficando entre 30% e 96%.

Como exemplo, na coluna “Precisão” para algoritmo RandomForest (ver Tabela 5.6) teve um diferença de 0,061 entre a classe Q1 com maior precisão (0,691) e a classe QN com menor precisão (0,630). Para o algoritmo IBK (ver Tabela 5.8) a diferença é maior que para o algoritmo RandomForest. No algoritmo IBK a diferença entre a classe Q1 com maior precisão (0,671) e a classe Q4 com menor precisão (0,565) foi de 0,106.

As matrizes de confusão apresentadas nas Tabelas 5.5 e 5.7 mostram que para os dois algoritmos o erro mais comum consiste em classificar as classes Q1, Q2, Q3 e Q4 como classe QN. Este comportamento pode ser explicado devido ao fato de que a classe QN apresenta uma maior ocorrência no conjunto de dados utilizados para treinamento e testes (ver Figura 5.8).

As curvas ROC apresentadas nas Figuras 5.18 e 5.19 demonstram graficamente o desempenho dos dois algoritmos. Pode-se verificar que as curvas ROC do algoritmo RandomForest são superiores e apresentam menos variação entre as classes que a do algoritmo IBK. Além disso, a estatística AUC (*Area Under the Curve*) descritas nas Tabelas 5.6 e 5.8 apresentam resultados significativamente superiores a um classificador randômico (0,5). O algoritmo RandomForest atingiu o melhor índice (0,911) para a classe Q1 e o pior índice (0,839) para a classe Q3. Para o algoritmo IBK os índices são menores, sendo 0,851 (Q1) o melhor índice e 0,744 (Q4) o índice mais baixo.

5.3.6 Inferências do componente físico do ModHEmo

Neste trabalho, considerou-se a hipótese que a combinação dos componentes físico e cognitivo poderia ser uma abordagem eficaz para melhorar os resultados da inferência. Assim, foram realizados testes para verificar o impacto no processo de inferência de usar cada um dos componentes individualmente.

Desta forma, nesta subseção serão apresentados os resultados obtidos pelo ModHEmo utilizando apenas as informações do componente físico do modelo. Desta forma, são utilizados neste ensaio apenas os atributos Q1F, Q2F, Q3F, Q4F e QNF constantes na Tabela 5.4. De maneira semelhante, na próxima subseção serão apresentados os resultados apenas com o componente cognitivo.

Neste cenário, o algoritmo RandomForest alcançou um índice Kappa de 0,3268 e uma acurácia global de 48,95%, classificando corretamente apenas 817 das 1669 instâncias do conjunto de dados. Por sua vez para o algoritmo IBK o índice Kappa foi de 0,3196. A acurácia global deste algoritmo ficou em 48,17%, classificando corretamente 804 das 1669 instâncias do conjunto de dados.

A Tabela 5.9 apresenta a matriz de confusão com os resultados do algoritmo RandomForest. Na Tabela 5.10 são apresentadas medidas de desempenho obtidas no processo de classificação dos cinco quadrantes com o algoritmo RandomForest. Na Figura 5.20 é apresentado gráfico com a plotagem das curvas ROC para cada um das classes para o algoritmo RandomForest utilizando dados do componente físico do ModHEmo.

A Tabela 5.11 apresenta a matriz de confusão com os resultados do algoritmo IBK. Na Tabela 5.12 são apresentadas medidas de desempenho obtidas no processo de classificação dos cinco quadrantes com o algoritmo IBK com os atributos do componente físico. Na Figura 5.21 é apresentado gráfico com a plotagem das curvas ROC para cada um das classes para o algoritmo IBK utilizando dados do componente físico do ModHEmo.

Comparando-se os resultados obtidos nesta subseção utilizando apenas o componente físico do ModHEmo com a seção anterior em que foram utilizados os dois componentes, pode-se perceber que todos os índices reportados foram inferiores. Utilizando-se apenas os atributos do componente físico, a acurácia do algoritmo RandomForest caiu de 65,79% para 48,95% e do

Tabela 5.9: Matriz de confusão do algoritmo RandomForest usando atributos do componente físico

Classe Correta \ Classe Prevista	Classe Prevista				
	QN	Q2	Q1	Q4	Q3
QN	357	47	43	20	46
Q2	107	203	32	18	21
Q1	93	18	97	6	20
Q4	109	26	31	42	18
Q3	111	37	37	12	118

Tabela 5.10: Índices de desempenho por classe do algoritmo RandomForest utilizando os atributos do componente físico

Classe	Verdadeiro Positivo	Falso Positivo	Precisão	Estatística F	AUC Área
QN	0,696	0,363	0,459	0,553	0,735
Q2	0,533	0,099	0,613	0,570	0,785
Q1	0,415	0,100	0,404	0,409	0,818
Q4	0,186	0,039	0,429	0,259	0,706
Q3	0,375	0,078	0,529	0,439	0,703
Média Geral	0,490	0,168	0,496	0,476	0,748

Tabela 5.11: Matriz de confusão do algoritmo IBK usando atributos do componente físico

Classe Correta \ Classe Prevista	Classe Prevista				
	QN	Q2	Q1	Q4	Q3
QN	349	51	41	36	36
Q2	103	203	32	26	17
Q1	88	18	96	13	19
Q4	93	27	33	51	22
Q3	117	34	43	16	105

Tabela 5.12: Índices de desempenho por classe do algoritmo IBK utilizando os atributos do componente físico

Classe	Verdadeiro Positivo	Falso Positivo	Precisão	Estatística F	AUC Área
QN	0,680	0,347	0,465	0,553	0,719
Q2	0,533	0,101	0,610	0,569	0,766
Q1	0,410	0,104	0,392	0,401	0,812
Q4	0,226	0,063	0,359	0,277	0,709
Q3	0,333	0,069	0,528	0,409	0,702
Média Geral	0,482	0,166	0,485	0,471	0,738

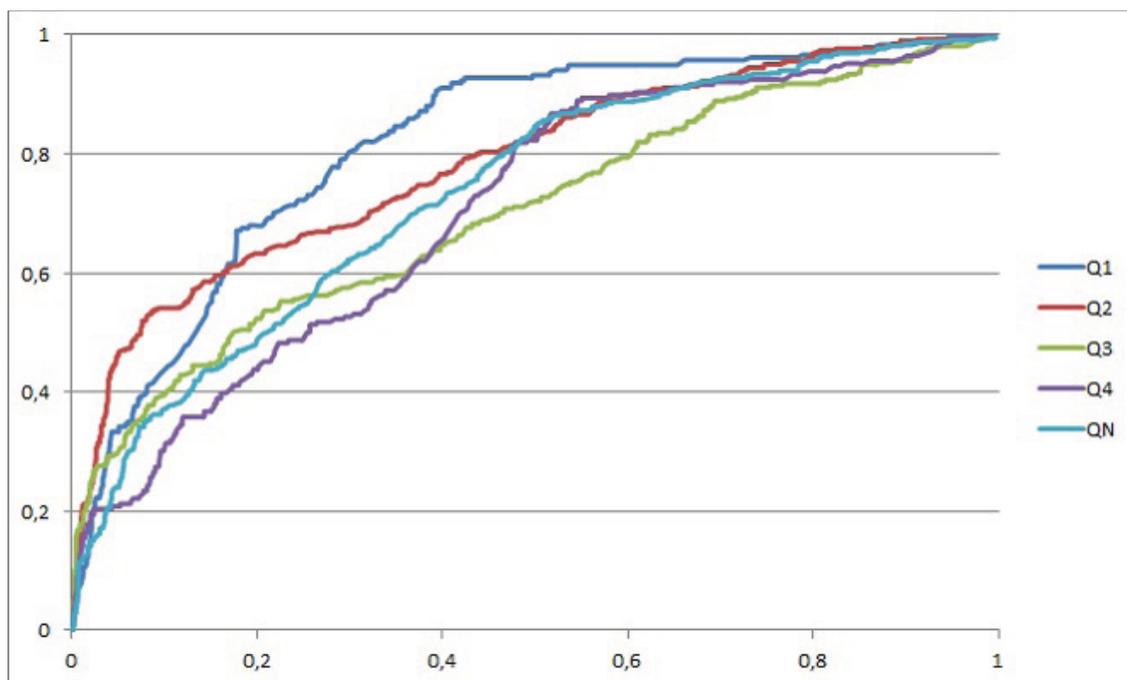


Figura 5.20: Curvas ROC por classe do algoritmo RandomForest utilizando atributos do componente físico

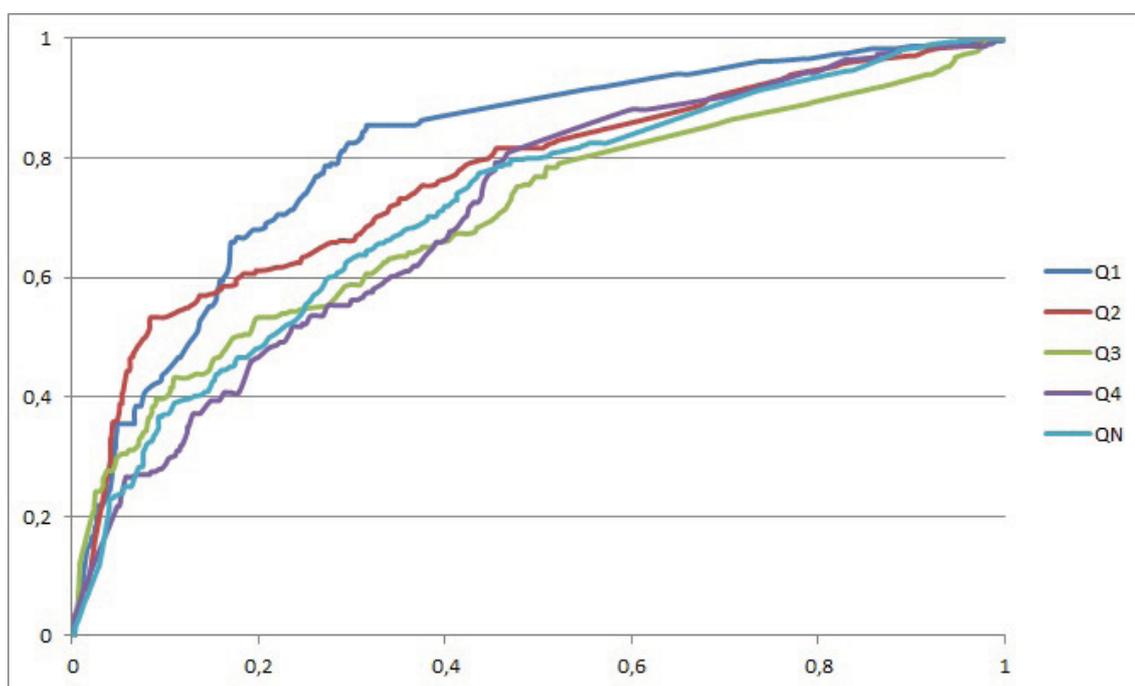


Figura 5.21: Curvas ROC por classe do algoritmo IBK utilizando atributos do componente físico

algoritmo IBK caiu de 63,21% para 48,17%. O resultado do teste T-Student pareado mostrou que estas diferenças são estatisticamente (nível de significância de 0,05).

A diferença nas taxas de precisão entre as classes também apresentou-se maior no cenário descrito nesta subseção em comparação com a seção anterior. Para o algoritmo RandomForest a diferença foi de 0,209 entre o melhor índice de 0,613 da classe Q2 e o pior índice de 0,404 da classe Q1 (ver Tabela 5.10). Esta condição pode ser verificada também para o algoritmo IBK

para o qual a diferença foi de 0,251 entre o melhor índice de 0,610 da classe Q2 e o pior índice de 0,359 da classe Q4 (ver Tabela 5.12).

5.3.7 Inferências do componente cognitivo do ModHEmo

Nesta subseção serão apresentados os resultados obtidos utilizando-se apenas as informações provenientes do componente cognitivo do ModHEmo. Para isto, serão utilizados apenas os atributos Q1C, Q2C, Q3C, Q4C e QNC já descritos na Tabela 5.4.

Considerando o cenário descrito acima, o algoritmo RandomForest atingiu o valor de 0,2233 para a estatística Kappa, enquanto que o algoritmo IBK obteve um valor de 0,2221 para esta estatística. Relativamente à acurácia global, o algoritmo RandomForest alcançou 41,76%, classificando corretamente 697 instâncias das 1669 existentes. A acurácia do algoritmo IBK foi de 41,94%, ou seja, 700 das 1669 instâncias foram corretamente classificadas.

A Tabela 5.13 apresenta a matriz de confusão com os resultados do algoritmo RandomForest. Na Tabela 5.14 são apresentadas medidas de desempenho obtidas no processo de classificação dos cinco quadrantes com o algoritmo RandomForest com os atributos do componente cognitivo. Na Figura 5.22 é apresentado gráfico com a plotagem das curvas ROC para cada um das classes para o algoritmo RandomForest utilizando dados do componente cognitivo do ModHEmo.

Tabela 5.13: Matriz de confusão do algoritmo RandomForest usando atributos do componente cognitivo

Classe Correta \ Classe Prevista	Classe Prevista				
	QN	Q2	Q1	Q4	Q3
QN	344	56	51	40	22
Q2	188	144	19	18	12
Q1	100	26	80	22	6
Q4	91	26	32	69	8
Q3	189	50	12	4	60

Tabela 5.14: Índices de desempenho por classe do algoritmo RandomForest utilizando os atributos do componente cognitivo

Classe	Verdadeiro Positivo	Falso Positivo	Precisão	Estatística F	AUC
QN	0,671	0,491	0,377	0,483	0,644
Q2	0,378	0,123	0,477	0,422	0,722
Q1	0,342	0,079	0,412	0,374	0,750
Q4	0,305	0,058	0,451	0,364	0,745
Q3	0,190	0,035	0,556	0,284	0,707
Média Geral	0,418	0,205	0,449	0,400	0,702

A Tabela 5.15 apresenta a matriz de confusão com os resultados do algoritmo IBK. Na Tabela 5.16 são apresentadas medidas de desempenho obtidas no processo de classificação dos cinco quadrantes com o algoritmo IBK com os atributos do componente cognitivo. Na Figura 5.23 é apresentado gráfico com a plotagem das curvas ROC para cada um das classes para o algoritmo IBK utilizando dados do componente cognitivo do ModHEmo.

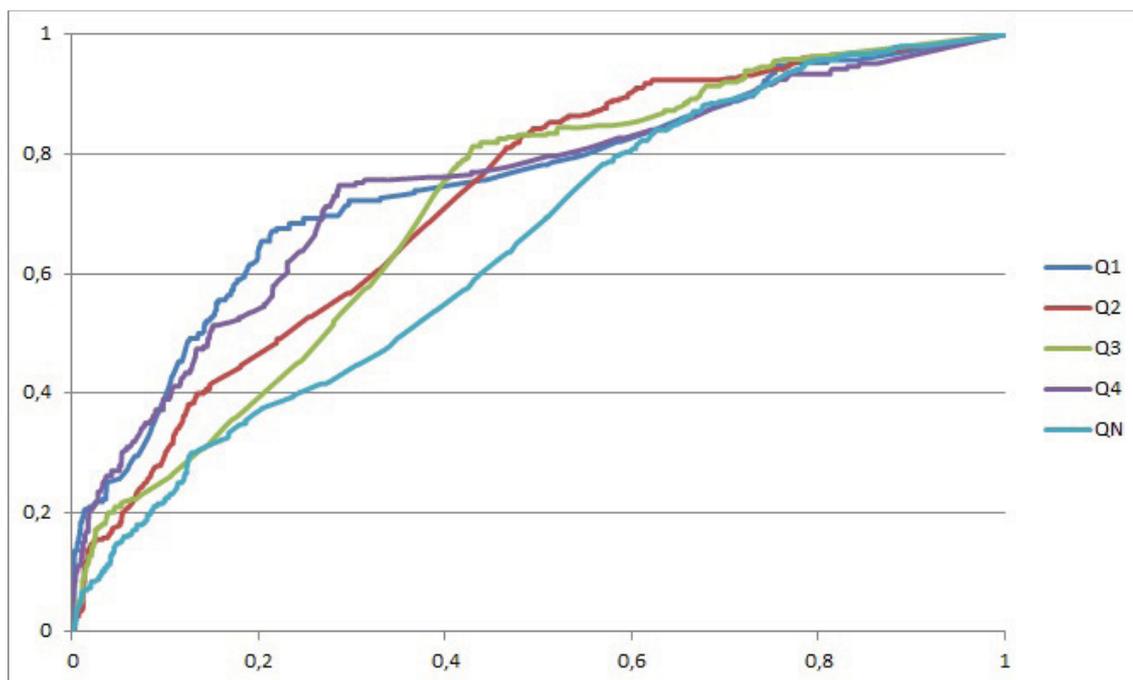


Figura 5.22: Curvas ROC por classe do algoritmo RandomForest utilizando atributos do componente cognitivo

Tabela 5.15: Matriz de confusão do algoritmo IBK usando atributos do componente cognitivo

Classe Correta \ Classe Prevista	Classe Prevista				
	QN	Q2	Q1	Q4	Q3
QN	362	51	45	34	21
Q2	194	145	14	17	11
Q1	102	24	80	22	6
Q4	100	24	34	62	6
Q3	196	50	15	3	51

Tabela 5.16: Índices de desempenho por classe do algoritmo IBK utilizando os atributos do componente cognitivo

Classe	Verdadeiro Positivo	Falso Positivo	Precisão	Estatística F	AUC
QN	0,706	0,512	0,379	0,494	0,639
Q2	0,381	0,116	0,493	0,430	0,716
Q1	0,342	0,075	0,426	0,379	0,749
Q4	0,274	0,053	0,449	0,341	0,722
Q3	0,162	0,032	0,537	0,249	0,710
Média Geral	0,419	0,208	0,451	0,396	0,697

Da mesma forma que na subseção anterior, os resultados obtidos nesta subseção, que considera apenas os atributos do componente cognitivo, foram inferiores aos obtidos com a fusão dos atributos dos componentes físico e cognitivo. A acurácia do classificador RandomForest que era de 65,79% caiu para 41,76% e do classificador IBK caiu de 63,21% para 41,94% no cenário utilizando apenas informações cognitivas. Esses resultados são ainda menores que os

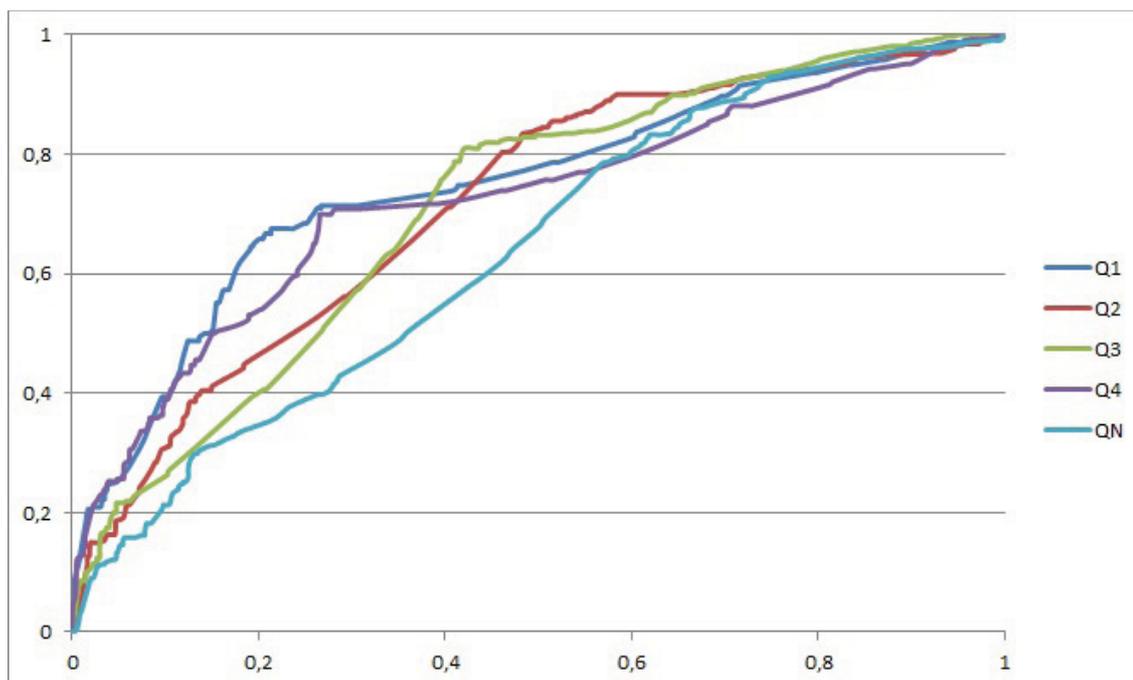


Figura 5.23: Curvas ROC por classe do algoritmo IBK utilizando atributos do componente cognitivo

obtidos quando utilizado apenas as informações do componente físico, que obteve uma acurácia de 48,95% e 48,17% para os algoritmos RandomForest e IBK, respectivamente.

Analisando-se os gráficos das Figuras 5.22 e 5.23 que apresenta as curvas ROC pode-se observar que para classe QN as curvas de ambos os classificadores aproximam-se de uma curva típica de um classificador randômico (AUC menor ou igual a 0,5).

5.3.8 Inferências usando a dimensão valência

Verificando-se os resultados apresentados na Tabela 3.1 pode-se observar que atualmente não existe consenso quanto a um conjunto predefinido de emoções relevante para o domínio educacional. Diante disso, as próximas duas subseções deste capítulo demonstram os resultados da inferência do ModHEmo considerando-se uma versão simplificada do conjunto de emoções apresentados até aqui. Desta maneira, poderá ser avaliado o comportamento da presente proposta considerando-se apenas três classes e não cinco classes como descritos até aqui neste capítulo.

A correta identificação da valência das emoções torna-se importante para inferência do estado afetivo-cognitivo chamado de “concentração engajada” (Baker et al., 2010) e/ou “humor do aprendiz” (Baker et al., 2010), já referidos na Seção 2.6. Desta forma, serão apresentados a seguir os resultados das inferências do ModHEmo levando-se em conta apenas a dimensão valência. Para isso, os quadrantes Q1 e Q4 foram agrupados em uma classes chamada “Positivo”. Por outro lado, os quadrantes Q2 e Q3 foram agrupados na classe classe “Negativo” e o quadrante neutro (QN) permanece como uma classe, aqui chamada de “Neutro”.

Considerando-se estas alterações nas classes do conjunto de dados, foram executados novamente os algoritmos RandomForest e IBK. O algoritmo RandomForest obteve um índice de acurácia de 71,42%, significando um total de 1192 das 1669 instâncias corretamente classificadas. Este mesmo algoritmo atingiu um índice Kappa de 0,56. Por sua vez, o algoritmo IBK obteve

uma acurácia de 70,10%, classificando corretamente 1170 das 1669 instâncias do conjunto de dados. O índice Kappa para o algoritmo IBK foi de 0,54.

A Tabela 5.17 apresenta a matriz de confusão com os resultados do algoritmo RandomForest. Na Tabela 5.18 são apresentadas medidas de desempenho obtidas no processo de classificação das três classes com o algoritmo RandomForest. Na Figura 5.24 é apresentado gráfico com a plotagem das curvas ROC do algoritmo RandomForest para cada um das três classes da dimensão valência.

Tabela 5.17: Matriz de confusão do algoritmo RandomForest com 3 classes da dimensão valência

Classe Correta \ Classe Prevista	Classe Prevista		
	Neutro	Negativo	Positivo
Neutro	315	128	70
Negativo	112	538	46
Positivo	52	69	339

Tabela 5.18: Índices de desempenho por classe do algoritmo RandomForest com 3 classes da dimensão valência

Classe	Verdadeiro Positivo	Falso Positivo	Precisão	Estatística F	AUC
Neutro	0,614	0,142	0,658	0,635	0,843
Negativo	0,773	0,202	0,732	0,752	0,879
Positivo	0,737	0,096	0,745	0,741	0,897
Média Geral	0,714	0,154	0,713	0,713	0,873

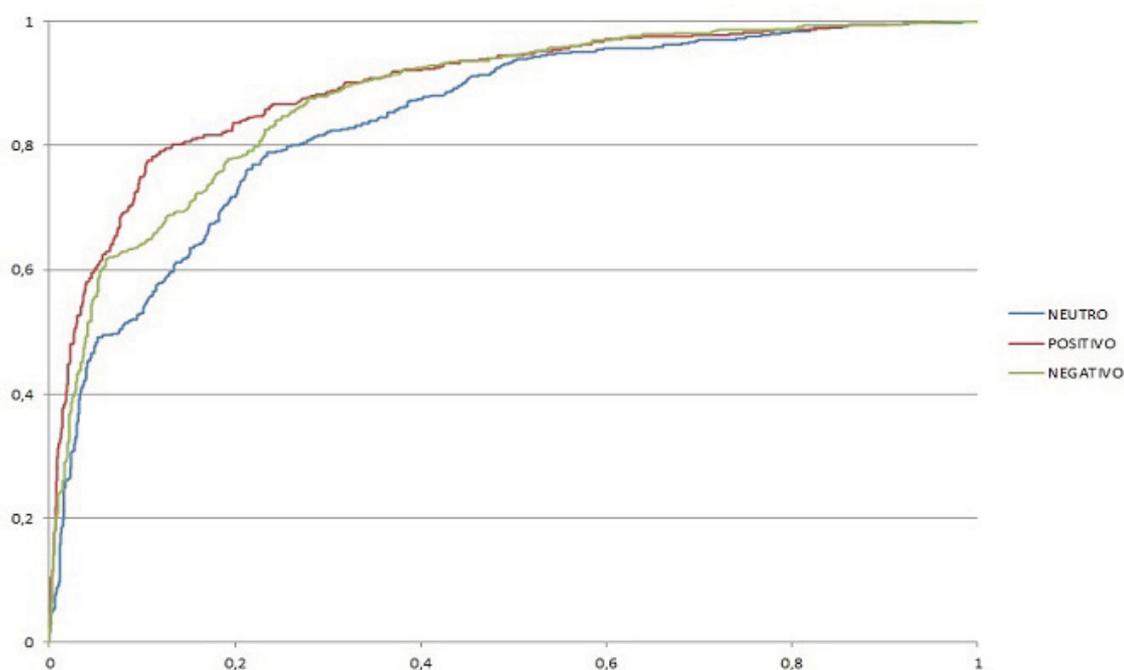


Figura 5.24: Curvas ROC por classe do algoritmo RandomForest com 3 classes da dimensão valência

A Tabela 5.19 apresenta a matriz de confusão com os resultados do algoritmo IBK. Na Tabela 5.20 são apresentadas medidas de desempenho obtidas no processo de classificação das três classes com o algoritmo IBK. Na Figura 5.25 é apresentado gráfico com a plotagem das curvas ROC para cada um das classes para o algoritmo IBK utilizando com 3 classes da dimensão valência.

Tabela 5.19: Matriz de confusão do algoritmo IBK com 3 classes da dimensão valência

Classe Correta \ Classe Prevista	Neutro	Negativo	Positivo
	Neutro	314	117
Negativo	118	528	50
Positivo	55	77	328

Tabela 5.20: Índices de desempenho por classe do algoritmo IBK com 3 classes da dimensão valência

Classe	Verdadeiro Positivo	Falso Positivo	Precisão	Estatística F	AUC
Neutro	0,612	0,150	0,645	0,628	0,783
Negativo	0,759	0,199	0,731	0,745	0,825
Positivo	0,713	0,109	0,713	0,713	0,838
Média Geral	0,701	0,159	0,700	0,700	0,816

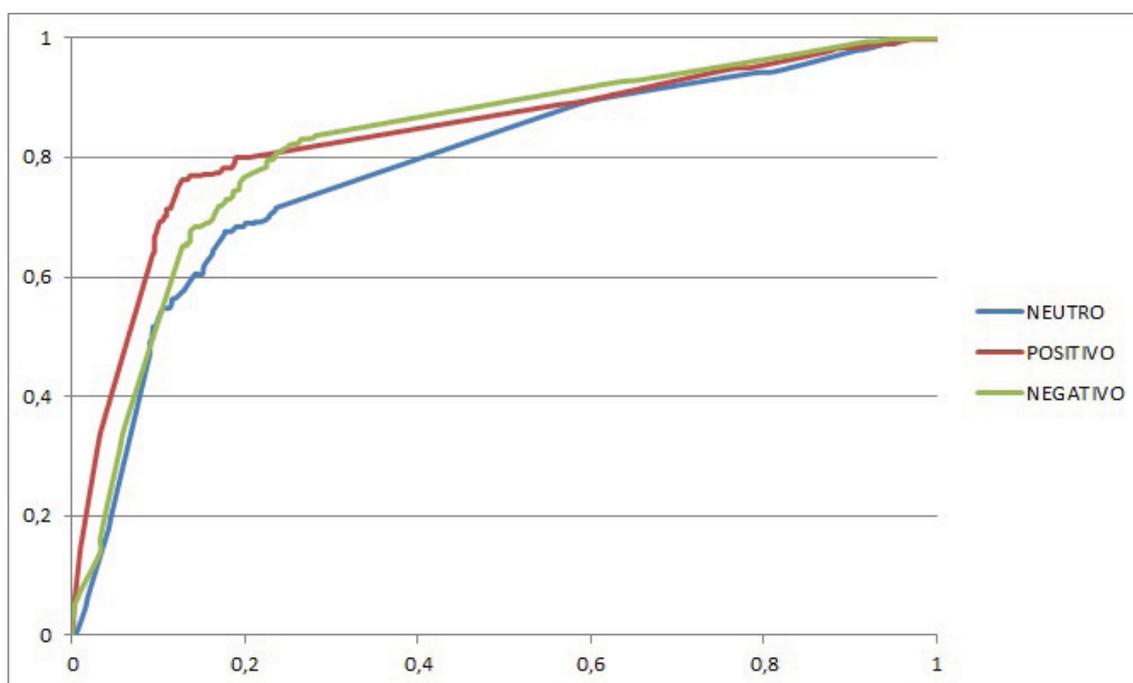


Figura 5.25: Curvas ROC por classe do algoritmo IBK com 3 classes da dimensão valência

Os resultados expressos acima considerando apenas a dimensão valência apresentam-se superiores aos resultados apresentados nos outros cenários analisados. Por exemplo, a acurácia do algoritmo RandomForest subiu de 65,79% na configuração com 5 classes e conjunto completo

de atributos (ver Subseção 5.3.5) para 71,42%, enquanto que a acurácia do algoritmo IBK subiu de 63,21% para 70,10%. O teste estatístico T-student pareado demonstrou que existe diferença significativa (nível de significância de 0,05) entre os dois algoritmos e, desta forma, o algoritmo RandomForest apresenta resultados superiores. Contudo, é importante ressaltar que esta melhoria nas taxas de acurácia eram esperadas e justificam-se pela diminuição do número de classes de 5 para 3, reduzindo assim as chances de erro no processo de classificação.

Os índices apresentados nas Tabelas 5.18 e 5.20 demonstram que os resultados são semelhantes entre as 3 classes. Como exemplo, a diferença na estatística “Precisão” para o algoritmo RandomForest foi de 0,084 entre o maior índice 0,742 (classe positivo) e o menor índice 0,658 (classe neutro). Para o algoritmo IBK a diferença foi de 0,086 entre o maior índice 0,731 (classe negativo) e o menor índice 0,645 (classe neutro). Esta situação pode ser visualizada graficamente nas Figuras 5.24 e 5.25 aonde percebe-se uma pequena distância entre as curvas das três classes.

5.3.9 Inferências usando a dimensão ativação

Da mesma forma que na subseção anterior, as informações a serem apresentadas abaixo consideram uma versão simplificada do conjunto de emoções do ModHEmo. Nesta subseção serão utilizadas três classes considerando-se a dimensão ativação. Os quadrantes com ativação positiva (Q1 e Q2) formam a classe “Agitação”. Por outro lado, Os quadrantes com ativação negativa (Q3 e Q4) formam a classe “Sonolência”. O quadrante QN continua a formar a classe “Neutro”.

Esta configuração de classes pode ser útil considerando-se o fato de que o desempenho em atividades educativas tem relação com o nível de ativação do estudante (Shen et al., 2007). De acordo com este trabalho, o desempenho em atividade de aprendizagem é menor nos níveis extremos (baixo e alto) de ativação. Em contrapartida, os melhores níveis de desempenho são obtidos com níveis intermediários de ativação.

Considerando as três classes descritas acima o algoritmo RandomForest obteve uma taxa de acurácia de 70,10%, significando que do total de 1669 instâncias do conjunto de dados 1170 foram corretamente classificadas. A estatística Kappa para este mesmo algoritmo foi de 0,5498. Por sua vez, o algoritmo IBK atingiu uma acurácia de 67,77%, sendo 0,5141 o valor da estatística Kappa.

A Tabela 5.21 apresenta a matriz de confusão com os resultados do algoritmo RandomForest. Na Tabela 5.22 são apresentadas medidas de desempenho obtidas no processo de classificação das três classes com o algoritmo RandomForest. Na Figura 5.26 é apresentado gráfico com a plotagem das curvas ROC para cada um das classes para o algoritmo RandomForest utilizando com 3 classes da dimensão ativação.

Tabela 5.21: Matriz de confusão do algoritmo RandomForest com 3 classes da dimensão ativação

Classe Correta \ Classe Prevista	Classe Prevista		
	Neutro	Agitação	Sonolência
Neutro	336	77	100
Agitação	86	472	57
Sonolência	93	86	362

A Tabela 5.23 apresenta a matriz de confusão com os resultados do algoritmo IBK. Na Tabela 5.24 são apresentadas medidas de desempenho obtidas no processo de classificação das três classes com o algoritmo IBK. Na Figura 5.27 é apresentado gráfico com a plotagem

Tabela 5.22: Índices de desempenho por classe do algoritmo RandomForest com 3 classes da dimensão ativação

Classe	Verdadeiro Positivo	Falso Positivo	Precisão	Estatística F	AUC
Neutro	0,655	0,155	0,652	0,654	0,838
Agitação	0,767	0,155	0,743	0,755	0,876
Sonolência	0,669	0,139	0,697	0,683	0,842
Média Geral	0,701	0,150	0,701	0,701	0,854

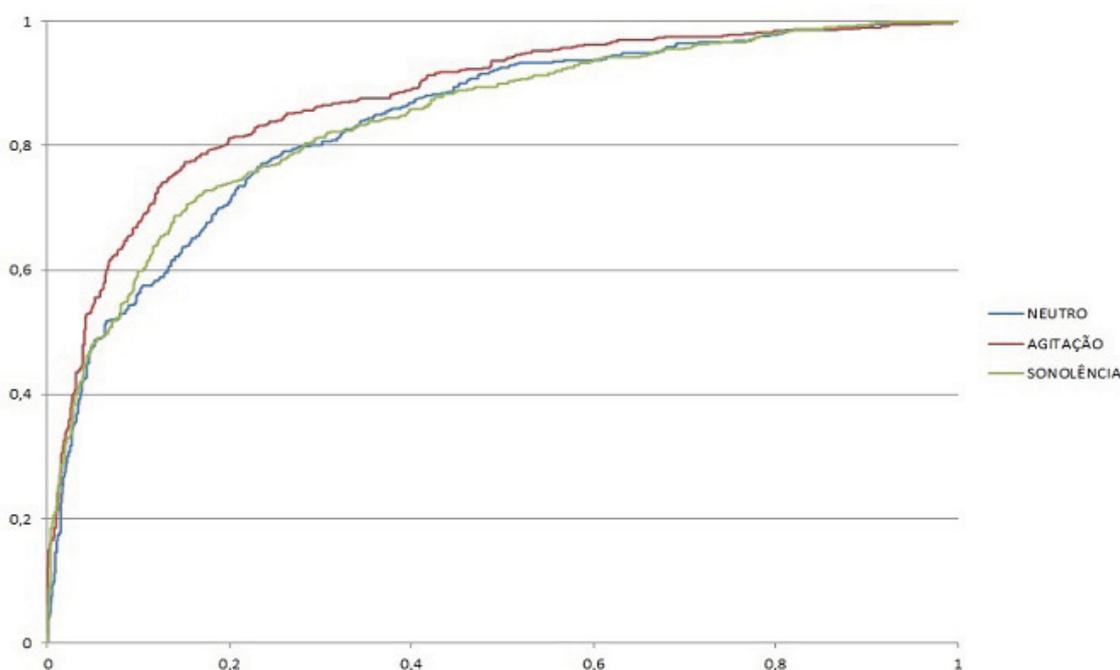


Figura 5.26: Curvas ROC por classe do algoritmo RandomForest com 3 classes da dimensão ativação

das curvas ROC para cada um das classes para o algoritmo IBK utilizando com 3 classes da dimensão ativação.

Tabela 5.23: Matriz de confusão do algoritmo IBK com 3 classes da dimensão ativação

Classe Correta \ Classe Prevista	Neutro	Agitação	Sonolência
	Neutro	313	79
Agitação	86	472	57
Sonolência	93	102	346

Em relação ao cenário com 3 classes baseadas na dimensão valência (ver Subseção 5.3.8) observa-se que os resultados nesta subseção são semelhantes. O algoritmo RandomForest obteve acurácia de 71,42% quando utilizando a dimensão valência e 70,10% com a dimensão ativação. Para o algoritmo IBK a acurácia foi de 70,10% com a dimensão valência e 67,77% com a dimensão ativação.

Tabela 5.24: Índices de desempenho por classe do algoritmo IBK com 3 classes da dimensão ativação

Classe	Verdadeiro Positivo	Falso Positivo	Precisão	Estatística F	AUC
Neutro	0,610	0,155	0,636	0,623	0,781
Agitação	0,767	0,172	0,723	0,744	0,822
Sonolência	0,640	0,158	0,660	0,650	0,776
Média Geral	0,678	0,162	0,676	0,676	0,795

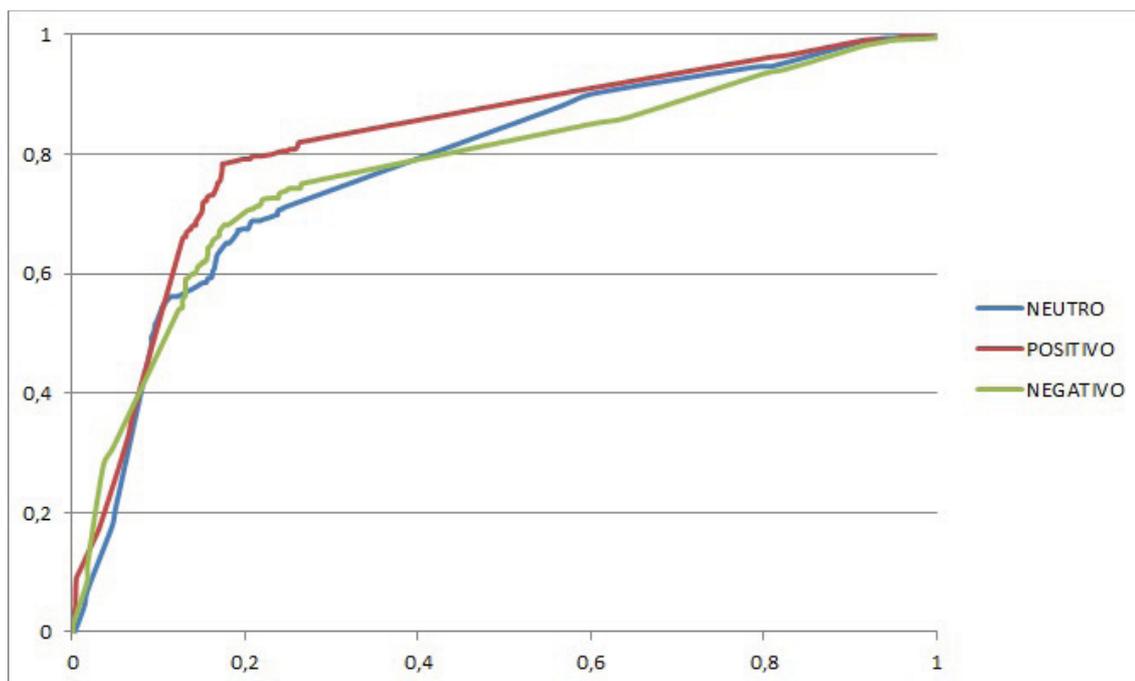


Figura 5.27: Curvas ROC por classe do algoritmo IBK com 3 classes da dimensão ativação

Os índices apresentados nas Tabelas 5.22 e 5.24 apresentam valores com pouca variação entre as três classes. Como exemplo, o índice “Precisão” do algoritmo RandomForest teve uma variação de 0,091 entre a classe “agitação”(0,743) e a classe “neutro” (0,652). Por outro lado, com o algoritmo IBK a variação foi de 0,087 entre a classe “agitação” com maior valor (0,723) e a classe “neutro” que teve menor valor (0,636).

5.4 Ameaças à validade dos resultados

Uma possível limitação dos experimentos refere-se à generalização dos resultados da pesquisa devido principalmente ao número de 38 estudantes participantes do experimento. Além disso, a quantidade de tempo em que os estudantes realizaram as atividades poderia interferir no tipo de emoções experimentadas. Um maior tempo de utilização poderia gerar mais emoções com valência negativa situadas nos quadrantes Q2 e Q3, por exemplo. Porém, experimentos adicionais são necessários para verificar se o impacto seria positivo, negativo ou mesmo nulo nos resultados finais.

Diferentes faixas etárias dos estudantes poderiam também representar mudanças no comportamento com relação à expressão das emoções, considerando-se que crianças tendem a

ser mais expressivas e menos propensas a suprimir ou disfarçar suas emoções (Ekman, 1992). Outrossim, a faixa etária dos estudantes poderia reduzir a capacidade de auto-avaliação das emoções. Neste sentido, a distribuição das emoções reportadas por adultos poderia variar e potencialmente interferir nos resultados obtidos.

Outro aspecto que pode representar uma limitação neste trabalho relaciona-se ao processo de obtenção dos objetivos dos estudantes, que é uma informação essencial para guiar o processo de inferência cognitivo do modelo OCC. Neste trabalho optou-se pela aplicação de um questionário, relativamente simples que pode apresentar algumas limitações.

No componente cognitivo do ModHEmo foram atribuídos manualmente valores para as variáveis intensidades e probabilidades dos eventos ocorridos no ambiente computacional. Este fato representa uma possível limitação dada a impossibilidade de diferenciar individualmente a intensidade e probabilidade das emoções cognitivas. Entende-se que a utilização de um STI nos experimentos poderia minimizar esta limitação, pois seria possível obter informações detalhadas sobre o estudante.

A inferência das emoções básicas para posterior mapeamento para emoções relacionadas à aprendizagem (quadrantes) pode representar também uma limitação do trabalho atual. Entende-se que a situação ideal seria inferir diretamente o conjunto de emoções relacionadas à aprendizagem. Contudo, não foram encontrados trabalhos que demonstrem a viabilidade e a implementação deste inferência direta.

5.5 Conclusão

Finalizando este capítulo, serão apresentados resumos, comparações e discussão sobre os índices que foram reportados nas seções anteriores deste capítulo. Os índices apresentados de forma resumida foram escolhidos pela sua importância na avaliação de modelos de classificação e por serem repetidamente reportados em trabalhos correlatos, permitindo assim a comparação.

Com o objetivo de facilitar a comparação das taxas de acurácias obtidas com os diferentes conjuntos de dados e algoritmos apresentados neste capítulo, a Tabela 5.25 apresenta as taxas de acurácia para cada conjunto de dados e algoritmo. Nesta tabela os conjuntos de dados foram separados nos cenários com 5 e 3 classes. Os cenários com 5 classes são subdivididos em 3 casos: fusão, físico e cognitivo. Nos cenários com 3 classes (valência e ativação) considera-se sempre a fusão das informações físicas e cognitivas.

Tabela 5.25: Comparação dos índices de acurácia dos algoritmos e conjunto de dados

Conjunto de Dados		5 Classes			3 Classes	
		Fusão	Físico	Cognitivo	Valência	Ativação
Algoritmo						
RandomForest		65,79	48,95	41,76	71,42	70,10
IBK		63,21	48,17	41,94	70,10	67,77

Os resultados expressos na Tabela 5.25 mostram que no cenário com 5 classes a fusão (coluna 2) apresenta acurácia superior quando comparada com os componentes tomados individualmente (colunas 3 e 4). Adicionalmente, as taxas de acurácia do algoritmo RandomForest foram superiores às taxas do algoritmo IBK no conjunto de dados resultado da fusão dos componentes do ModHEmo (coluna 2) e nos dois casos com três classes (coluna 5 e 6). Nestes três casos, o teste T-Student Pareado apontou diferença significativa estatisticamente (nível de significância de 0,05), confirmando os resultados superiores do algoritmo RandomForest. Nos

conjuntos de dados com atributos físico (coluna 3) e cognitivo (coluna 4) os dois algoritmos atingiram índices de acurácia semelhantes. Nestes dois casos o teste T-Student pareado apontou que não há diferença estatisticamente significativa (nível de significância de 0,05) entre os resultados dos dois algoritmos. Nos dois cenários com três classes (colunas 5 e 6) os resultados foram superiores aos casos com cinco classes.

A Tabela 5.26 apresenta o índice Cohen Kappa para os dois algoritmos e cinco diferentes conjuntos de dados apresentados. Este índice é importante pois considera o impacto que o número de classes tem sobre os modelos de classificação. Além disso, esta estatística permite a comparação com trabalhos correlatos que a reportam com frequência.

Analisando-se os resultados expressos na Tabela 5.26 pode-se perceber diferenças significativas entre os resultados obtidos com a fusão (coluna 2) comparando-se com os componentes físico e cognitivo (colunas 3 e 4). Todos os resultados do índice Cohen Kappa demonstram que o algoritmo RandomForest obteve valores absolutos superiores em relação ao algoritmo IBK. Todavia, pode-se perceber que os valores são semelhantes entre a fusão do conjunto de dados com cinco classe (coluna 2) e os conjuntos de dados com 3 classes (colunas 5 e 6). Isto indica que a eficiência do processo de classificação é semelhante e que as taxas de acurácia superiores do cenário com 3 classes reportadas na Tabela 5.25 são resultados apenas da diferença entre o número de classes.

Tabela 5.26: Comparação do índice Cohen Kappa dos algoritmos e conjunto de dados

Conjunto de Dados		5 Classes			3 Classes	
		Fusão	Físico	Cognitivo	Valência	Ativação
Algoritmo	RandomForest	0,5581	0,2233	0,3268	0,56	0,5498
Algoritmo	IBK	0,5274	0,2221	0,3196	0,54	0,5141

As curvas ROC e a Área sobre a Curva (AUC) são também indicadores importantes para avaliação de modelos de classificação e usualmente reportados em trabalhos correlatos. Na Tabela 5.27 são apresentadas as médias da AUC dos dois algoritmos em todos os conjunto de dados utilizados neste trabalho. Os valores das médias da AUC apresentadas na Tabela 5.27 mostram que o algoritmo RandomForest obteve taxas absolutas superiores em relação ao algoritmo IBK. Além disso, para os dois algoritmos, os melhores resultados foram obtidos com o conjunto de dados com três classes formadas pela dimensão valência.

Tabela 5.27: Comparação da Área Sobre a Curva (AUC) média dos algoritmos e conjunto de dados

Conjunto de Dados		5 Classes			3 Classes	
		Fusão	Físico	Cognitivo	Valência	Ativação
Algoritmo	RandomForest	0,867	0,748	0,702	0,873	0,854
Algoritmo	IBK	0,800	0,738	0,697	0,816	0,795

Comparar os resultados apresentados nesta seção com trabalhos correlatos é uma tarefa difícil e sensível. Isto se deve ao fato que diferentes aspectos precisam ser considerados para uma comparação correta e justa. Dentre estes aspectos, pode-se mencionar: i) tipos de sensores utilizados nos experimentos, ii) o experimento realizado em ambiente real de ensino ou laboratório específico, iii) número de emoções ou estados afetivos considerados, iv) tipos de

emoções (primárias ou secundárias), v) tipo de interação com o ambiente de computação (texto, voz, leitura, etc.), vi) quem faz a rotulagem dos dados (alunos, colegas, observadores externos).

Como exemplo, (Picard, 1997) observa que usando como entrada a voz é possível atingir até 91% de precisão para inferência de tristeza (*sadness*). No entanto, esta abordagem impõe limitações severas porque só pode ser usada quando a interação com software educacional inclui voz.

Relativamente ao processo de rotulagem, Baker et al. (2010) observam que os auto-relatos dos estudantes são mais sensíveis a determinadas emoções, enquanto que um conjunto diferente de emoções será particularmente relevante para observadores externos treinados. Neste sentido, o trabalho de (D’Mello et al., 2007) mostra que diferentes taxas de acurácia são obtidas quando atores distintos fazem a rotulagem. Estes autores relataram uma taxa de acurácia próxima a 62% quando juízes ou colegas externos rotularam os dados. No entanto, a precisão caiu para 52% quando o próprio estudante rotulou o conjunto de dados (caso deste trabalho).

É importante destacar também trabalhos como de Zhang et al. (2018) e Akputu et al. (2018), por exemplo, que reportam taxas de acurácia de até 83% no reconhecimento das emoções básicas. Contudo, estas taxas são obtidas considerando-se banco de dados de imagens que podem não repetir as condições de um ambiente real, como mostrado no trabalho de Gottardo e Pimentel (2016).

Considerando a revisão bibliográfica apresentada neste trabalho, não foram encontrados trabalhos prévios que usem exatamente a mesma abordagem, tipo de emoções e configuração experimental como apresentado neste trabalho. Mesmo assim, considerou-se importante realizar comparações tentando posicionar os resultados obtidos dentro do estado da arte. Desta forma, pesquisas recentes de Bosch et al. (2014) e Paquette et al. (2014) e Botelho et al. (2017) assemelham-se a este trabalho porque se concentram em emoções relacionadas a aprendizagem e usam um design experimental similar. Desta forma, serão apresentados a seguir os principais resultados destes três trabalhos e, finalmente, um comparativo de trabalhos correlatos com os resultados desta pesquisa.

Com base na expressão facial dos estudantes durante o uso de um ITS, Bosch et al. (2014) reporta o melhor índice Cohen Kappa de 0,112. Usando técnicas de mineração de texto Paquette et al. (2014) mostrou um método para inferir quatro emoções: tédio, confusão, frustração e concentração. Para estas emoções, a AUC relatada foi, respectivamente, 0,767, 0,777, 0,762, 0,738 e os melhores resultados para o índice Cohen Kappa foram de 0,486. O trabalho de Botelho et al. (2017) usou uma abordagem de aprendizagem profunda (*deep learning*), baseada em registros da interação dos alunos em um ITS, para inferir quatro emoções relacionadas ao aprendizado: confusão, concentração, aborrecimento, frustração. Neste trabalho, a melhor AUC e Cohen Kappa foi de 0,78 e 0,24, respectivamente. É importante notar que, no trabalho de Botelho et al. (2017), a rotulagem foi feita por observadores externos.

A Tabela 5.28 apresenta uma comparação de resultados obtidos em trabalhos correlatos e os resultados deste trabalhos. Pode-se observar que os resultados obtidos nos experimentos com o ModHEmo apresentados na Tabela 5.28 apontam melhorias quando comparados com (Bosch et al., 2014; Paquette et al., 2014; D’Mello et al., 2007; Botelho et al., 2017).

Tabela 5.28: Comparação de resultados do ModHEmo com trabalhos correlatos.

Trabalho	Indicador	Melhor Valor Obtido no Trabalho Correlato	Melhor Valor ModEHmo (5 Classes, fusão)
Bosch et al. (2014)	Kappa	0,112	0,5581
Paquette et al. (2014)	Kappa	0,486	0,5581
Paquette et al. (2014)	AUC	0,777	0,867
Botelho et al. (2017)	Kappa	0,24	0,5581
Botelho et al. (2017)	AUC	0,78	0,867
D’Mello et al. (2007)	Acurácia	52%	65,79%
Muñoz et al. (2010)	Acurácia	60,71%	65,79%

Capítulo 6

Considerações Finais

Neste Capítulo serão apresentadas as considerações finais desta pesquisa, envolvendo a análise dos principais aspectos discutidos neste trabalho. Além disso, serão destacadas perspectivas de trabalhos futuros e as contribuições científicas alcançadas durante o desenvolvimento desta Tese de Doutorado.

6.1 Considerações gerais sobre a pesquisa

Funcionalidades relacionadas à adaptação às necessidades individuais dos aprendizes são consideradas fundamentais para a construção de softwares educacionais que busquem influenciar positivamente o processo de aprendizagem. Considerando-se os trabalhos correlatos apresentados nesta pesquisa, observa-se que reconhecer e responder às reações emocionais dos estudantes é um requisito importante para professores em uma sala de aula tradicional. De maneira semelhante, ambientes computacionais voltados ao apoio à aprendizagem poderiam aumentar sua efetividade ao incluir funcionalidades adaptativas às emoções dos aprendizes.

Estas funcionalidades são importantes, tendo em vista que pesquisadores da área de psicologia e neurociência concordam largamente que emoções e aprendizagem são intrinsecamente relacionadas. Neste sentido, pesquisas demonstram que funções relacionadas a aprendizagem e emoções são amplamente conectadas, compartilhando a mesma região no cérebro humano.

Pesquisas na área conhecida como Computação Afetiva têm evoluído nos últimos anos no sentido de agregar a ambientes computacionais funcionalidades de reconhecimento e adaptação às reações afetivas dos seus usuários. Entretanto, o reconhecimento de emoções, que para os seres humanos é uma tarefa relativamente simples, ainda apresenta diversas barreiras para sua disseminação em ambientes computadorizados.

Mesmo com o desenvolvimento de trabalhos correlatos, pesquisadores destacam que avanços ainda são necessários até que se seja possível a construção de softwares educacionais efetivamente sensíveis às reações emocionais dos estudantes. Desta forma, o problema base tratado nesta pesquisa refere-se a como inferir as emoções de estudantes durante o uso de um software educacional. Foi considerado fundamental que o processo de inferência fosse minimamente intrusivo, de baixo custo e viável de utilização em ambientes reais de aprendizagem e não apenas em laboratórios específicos.

Para a construção deste trabalho considerou-se a hipótese de que um processo efetivo de inferência de emoções deveria incluir não apenas reações físicas, mas também informações contextuais ou cognitivas. Desta forma, a fusão de informações físicas e cognitivas foi tomada como base para estruturar um processo de inferência de emoções de aprendizes utilizando um software educacional.

Neste contexto, a presente proposta tem como objetivo principal a construção de um modelo de inferência de emoções que preencha algumas lacunas e deficiências existentes atualmente no processo de inferência das emoções. É importante ressaltar que a estruturação do processo de inferência apresentado neste trabalho caracteriza-se como um modelo aberto e expansível, permitindo que facilmente sejam adicionadas novas informações ou sensores. Para isso, bastaria obter estas novas informações e incluí-las como subcomponentes dos componentes físico ou cognitivo já existentes.

A formalização do processo de inferência foi apresentada neste trabalho através de um Modelo Híbrido de Inferência de Emoções chamado de ModHEmo. Para a construção deste modelo foi necessário o desenvolvimento de um conjunto de atividades para atingir os Objetivos Específicos apresentados no primeiro capítulo deste documento. Inicialmente, foi realizada uma ampla avaliação dos trabalhos correlatos e dos aspectos conceituais envolvidos neste processo. Este objetivo específico foi atingido com a delimitação da abordagem conceitual a ser utilizada. Desta forma, definiu-se que seria fundamental desenvolver uma abordagem de inferência baseada na fusão de informações físicas e cognitivas. Para embasar conceitualmente esta proposta entendeu-se que o Modelo OCC (Ortony et al., 1990) e o Modelo de Ekman (Ekman, 1992) são os mais adequados para tratar as informações cognitivas e físicas, respectivamente.

A escolha do conjunto de emoções a serem consideradas consistiu em uma atividade fundamental para construção desta pesquisa. Entende-se que este fato é relevante levando-se em conta que atualmente não existe um conjunto consolidado de emoções relacionadas à aprendizagem a ser considerado no desenvolvimento de modelos de inferência. Este objetivo específico foi alcançado através da proposta baseada em quadrantes que refletem situações relevantes no processo de aprendizagem. Neste sentido, um ponto de destaque do trabalho refere-se ao método de escolha das cinco classes de emoções utilizando uma abordagem baseada nos quadrantes formados pelas dimensões valência e ativação. Estes quadrantes representam situações relevantes no processo de aprendizagem e, desta forma, podem ser utilizados para a construção de ambientes computacionais de ensino adaptativos.

Foram avaliadas também as técnicas computacionais que poderiam ser utilizadas para a construção do processo de inferência de emoções. Após amplo estudo, este objetivo específico foi alcançado com a escolha do Modelo ALMA (Gebhard, 2005) para as inferências cognitivas e a EmotionAPI para inferência de emoções usando expressões faciais. Além disso, foi definida uma estratégia de fusão dos componentes para inferência do resultado final do modelo utilizando os classificadores RandomForest e KNN. Considerando-se estas técnicas computacionais, foi então implementado computacionalmente o modelo híbrido de inferência de emoções.

Para viabilizar a aplicação prática do modelo implementado, foram realizadas customizações no software educacional TuxMath. Utilizando uma versão customizada do TuxMath, experimentos foram realizados com 38 estudantes em um ambiente real de ensino. Durante a realização das atividades um conjunto de eventos no software educacional foram monitorados e as prováveis emoções dos estudantes foram inferidos nestes instantes. Para treinamento e teste dos classificadores que fazem a inferência final do modelo foi utilizada uma abordagem de rotulagem de dados em que os estudantes fizeram a avaliação retrospectiva de suas emoções. Neste processo, visando facilitar o entendimento do estudantes, foi proposto uma abordagem diferenciada do método tradicional (SAM), utilizando *emoticons*. É importante destacar que, por estar fora do escopo deste trabalho, não foi realizada avaliação sobre o método de rotulagem usando *emoticons*. No entanto, este método baseia-se em trabalhos correlatos (Hayashi et al., 2016; Yusoff et al., 2013) que demonstram a viabilidade de aplicação deste tipo de abordagem para crianças.

A realização dos experimentos com a versão executável do ModHEmo demonstrou primeiramente a viabilidade técnica da proposta atual. Este fato é relevante pois a fusão de informações físicas e cognitivas implementada nesta pesquisa é pouco explorada em trabalhos correlatos. Considerando a revisão da literatura apresentada nesta pesquisa, pode-se observar que um número limitado de trabalhos apresentam a implementação de uma proposta de fusão de informações físicas e cognitivas.

Os resultados obtidos nos experimentos podem ser considerados promissores, apesar das limitações apresentadas e discutidas ao final do capítulo anterior. Mesmo que uma comparação direta com trabalhos correlatos seja difícil, os resultados alcançados são semelhantes ou superiores ao estado da arte consultado. Como exemplo, nos experimentos considerando cinco classes e com fusão dos componentes físicos e cognitivos do ModHEmo foi possível obter uma acurácia de até 65,79%. Nos experimentos com três classes a taxa de acurácia chegou aos 71,42%.

Foram realizados também ensaios específicos com o objetivo de verificar a validade da hipótese considerada neste trabalho. Usando apenas atributos cognitivos, a acurácia foi respectivamente 39,25% e 40% para os algoritmos RandomForest e IBK. Por outro lado, apenas com atributos físicos a acurácia foi de, respectivamente, 55,29% e 52,19% para os algoritmos RandomForest e IBK. Analisando-se estes ensaios, pode-se verificar que os resultados obtidos pelos componentes físico e cognitivo individualmente foi significativamente inferior ao obtido com a fusão destes componentes. Estes resultados apontam para a validade da hipótese considerada nesta pesquisa.

Baseando-se nas inferências afetivas obtidas pode-se implementar um conjunto de adaptações no ambiente computacional, buscando-se manter o estudante engajado nas atividades de aprendizagem. Além disso, as informações afetivas poderiam ser usadas para descrever um perfil da dinâmica das emoções dos aprendizes que indicam estados afetivos/cognitivos de alto nível como ciclo vicioso, engajamento e humor do aprendiz. Estes estados tem um alto potencial de utilização para implementação de adaptações no ambiente computacional ou a aplicação de intervenções pedagógicas individualizadas.

6.2 Perspectivas de trabalhos futuros

Considerando que a abordagem híbrida proposta assemelha-se ao processo natural de inferência de emoções, entende-se que apresenta grandes oportunidades para melhorias futuras, adicionando novos dados ou sensores que forneçam informações adicionais aos componentes físico e cognitivo.

Como exemplo, em investigações futuras poder-se-ia incluir a detecção de gestos, como movimentos de cabeça, ou mãos na face que podem ser obtidos através dos vídeos coletados nos experimentos. Esta proposta potencialmente poderia atingir resultados promissores, considerando os resultados de Adams et al. (2015). No referido trabalho os autores demonstram que os movimentos da cabeça fornecem informação emocional que é complementar e não redundante com as expressões faciais. Adicionalmente, informações provenientes do teclado e do mouse poderiam ser incluídas no componente cognitivo. Caso o ambiente educacional disponibilize interações textuais, estas poderiam ser usadas juntamente com técnicas de inferência de emoção em texto.

Futuras pesquisas poderiam ainda investigar a relevância de considerar aspectos relativos a personalidade do estudante no processo de inferência do componente cognitivo do ModHEmo. A implementação desta funcionalidade seria amplamente viável e facilitada pois o modelo ALMA utilizado neste trabalho já prevê nativamente o tratamento de informações de personalidade.

Neste trabalho foram discutidas possíveis estratégias de adaptação e intervenção pedagógica, utilizando como base as inferências emocionais obtidas. Porém, a investigação do impacto destas estratégias de adaptação e intervenção não fazem parte do escopo deste trabalho, podendo ser objeto de futuras pesquisas.

Trabalhos futuros poderiam envolver a utilização de estratégias mais elaboradas para reconhecer os objetivos do estudante ao utilizar um software educacional. Como exemplo, poderiam ser avaliados a utilizado do questionário MSLQ (*Motivated Strategies for Learning Questionnaire*) (Pintrich et al., 1991) ou a proposta de avaliação dos objetivos (*goal assessment*) apresenta no trabalho de Conati (2002).

6.3 Contribuições científicas referentes a esta pesquisa

Nesta seção serão apresentadas as referências para os trabalhos publicados, todos referentes a presente tese. O último artigo refere-se a publicação em periódico e os demais em congressos.

GOTTARDO, E.; PIMENTEL, A. R. Improving Inference of Learning Related Emotion by Combining Cognitive and Physical Information In: **International Conference on Intelligent Tutoring Systems, ITS - 2018**, Montreal - Canada. Lecture Notes in Computer Science. Springer, 2018. v.10858. p.313 - 318.

GOTTARDO, E.; PIMENTEL, A. R. Reconhecimento e Adaptação à Dinâmica de Estados Afetivos Relacionados à Aprendizagem. **Anais do Simpósio Brasileiro de Informática na Educação. SBIE 2018**.

¹GOTTARDO, E.; PIMENTEL, A.R. Hybrid Model of Emotions Inference - An Approach based on Fusion of Physical and Cognitive Informations In: **20th International Conference on Enterprise Information Systems, ICEIS - 2018**, Funchal. Proceedings of the 20th International Conference on Enterprise Information Systems. , 2018. p.441 - 450.

GOTTARDO, ERNANI; PIMENTEL, ANDREY RICARDO. Affective Human-Computer Interaction in Educational Software In: Brazilian Symposium on Human Factors in Computing Systems, 2017, Joinville. **Proceedings of the XVI Brazilian Symposium on Human Factors in Computing Systems - IHC 2017**. New York: ACM Press, 2017. p.1.

GOTTARDO, ERNANI; PIMENTEL, ANDREY Modelo Híbrido de Inferência de Emoções para Ambientes de Aprendizagem: uma Proposta Baseada na Fusão de Componentes Físicos e Cognitivos In: **XXVIII Simpósio Brasileiro de Informática na Educação SBIE (Brazilian Symposium on Computers in Education), 2017**, Recife.. p.1778 - 1780

GOTTARDO, E.; PIMENTEL, A. R. Sistemas Tutores Afetivos: Estado da Arte e Desafios para o Reconhecimento de Emoções Relacionadas à Aprendizagem via Expressões Faciais In: **5º DEsafIE –Workshop de Desafios da Computação Aplicada à Educação**, 2016, Porto Alegre - RS. CSBC 2016 - XXXVI Congresso da Sociedade Brasileira de Computação., 2016. v.1.

¹Artigo selecionado para submissão de versão ampliada na revista Lecture Notes in Business Information Processing - LNBIP

Artigo aceito para publicação em Periódico

GOTTARDO, E.; PIMENTEL, A. R. Inferring Students' Emotions Using a Hybrid Approach that Combine Cognitive and Physical Data. **Lecture Notes in Business Information Processing.**, 2018.

Referências Bibliográficas

- Adams, A., Mahmoud, M., Baltrusaitis, T. e Robinson, P. (2015). Decoupling facial expressions and head motions in complex emotions. Em *2015 International Conference on Affective Computing and Intelligent Interaction (ACII)*, páginas 274–280. IEEE.
- Akputu, O. K., Seng, K. P., Lee, Y. e Ang, L.-M. (2018). Emotion recognition using multiple kernel learning toward e-learning applications. *ACM Trans. Multimedia Comput. Commun. Appl.*, 14(1):1:1–1:20.
- Alexander, S. T. V. (2008). *An affect-sensitive intelligent tutoring system with an animated pedagogical agent that adapts to student emotion like a human tutor*. Tese de doutorado, Massey University, Albany, New Zealand.
- Allen, L. K., Mills, C., Jacovina, M. E., Crossley, S., D’mello, S. e McNamara, D. S. (2016). Investigating boredom and engagement during writing using multiple sources of information: the essay, the writer, and keystrokes. Em *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge*, páginas 114–123. ACM.
- AlZoubi, O., D’Mello, S. K. e Calvo, R. A. (2012). Detecting naturalistic expressions of nonbasic affect using physiological signals. *IEEE Transactions on Affective Computing*, 3(3):298–310.
- Baker, R., D’Mello, S., Rodrigo, M. e Graesser, A. (2010). Better to be frustrated than bored: The incidence and persistence of affect during interactions with three different computer-based learning environments. *International Journal of human-computer studies*, 68(4):223–241.
- Baker, R. S., Corbett, A. T., Koedinger, K. R., Evenson, S., Roll, I., Wagner, A. Z., Naim, M., Raspat, J., Baker, D. J. e Beck, J. E. (2006). Adapting to when students game an intelligent tutoring system. Em *International Conference on Intelligent Tutoring Systems*, páginas 392–401. Springer.
- Baker, R. S., Gowda, S., Wixon, M., Kalka, J., Wagner, A., Salvi, A., Alevin, V., Kusbit, G., Ocumpaugh, J. e Rossi, L. (2012). Sensor-free automated detection of affect in a cognitive tutor for algebra. Em *Educational Data Mining 2012*.
- Baker, R. S., Ocumpaugh, J., Gowda, S. M., Kamarainen, A. M. e Metcalf, S. J. (2014). Extending log-based affect detection to a multi-user virtual environment for science. Em *International Conference on User Modeling, Adaptation, and Personalization*, páginas 290–300. Springer.
- Bartlett, M. S., Littlewort, G., Frank, M., Lainscsek, C., Fasel, I. e Movellan, J. (2006). Fully automatic facial action recognition in spontaneous behavior. Em *Proceedings of 7th International Conference on Automatic Face and Gesture Recognition, 2006. FGR 2006.*, páginas 223–230. IEEE.

- Becker-Asano, C. e Wachsmuth, I. (2010). Affective computing with primary and secondary emotions in a virtual human. *Autonomous Agents and Multi-Agent Systems*, 20(1):32.
- Bosch, N., Chen, Y. e D’Mello, S. (2014). It’s written on your face: detecting affective states from facial expressions while learning computer programming. Em *International Conference on Intelligent Tutoring Systems*, páginas 39–44. Springer.
- Bosch, N., D’Mello, S., Baker, R., Ocumpaugh, J., Shute, V., Ventura, M., Wang, L. e Zhao, W. (2015). Automatic detection of learning-centered affective states in the wild. Em *Proceedings of the 20th international conference on intelligent user interfaces*, páginas 379–388. ACM.
- Bosch, N. e D’Mello, S. (2017). The affective experience of novice computer programmers. *International Journal of Artificial Intelligence in Education*, 27(1):181–206.
- Botelho, A. F., Baker, R. S. e Heffernan, N. T. (2017). Improving sensor-free affect detection using deep learning. Em *International Conference on Artificial Intelligence in Education*, páginas 40–51. Springer.
- Bouchet, F., Harley, J. M. e Azevedo, R. (2018). Evaluating adaptive pedagogical agents’ prompting strategies effect on students’ emotions. Em *International Conference on Intelligent Tutoring Systems*, páginas 33–43. Springer.
- Bradley, M. M. e Lang, P. J. (1994). Measuring emotion: the self-assessment manikin and the semantic differential. *Journal of behavior therapy and experimental psychiatry*, 25(1):49–59.
- Bratman, M. E. (1990). What is intention. *Intentions in communication*, páginas 15–31.
- Breiman, L. (2001). Random forests. *Machine learning*, 45(1):5–32.
- Burleson, W. (2011). Advancing a multimodal real-time affective sensing research platform. Em *New perspectives on affect and learning technologies*, páginas 97–112. Springer.
- Calvo, R. A. e D’Mello, S. (2010). Affect detection: An interdisciplinary review of models, methods, and their applications. *IEEE Transactions on affective computing*, 1(1):18–37.
- Calvo, R. A. e D’Mello, S. (2012). Frontiers of affect-aware learning technologies. *IEEE Intelligent Systems*, 27(6):86–89.
- Conati, C. (2002). Probabilistic assessment of user’s emotions in educational games. *Applied artificial intelligence*, 16(7-8):555–575.
- Conati, C. (2011). Combining cognitive appraisal and sensors for affect detection in a framework for modeling user affect. Em *New perspectives on affect and learning technologies*, páginas 71–84. Springer.
- Corbett, A. T., Koedinger, K. R. e Anderson, J. R. (1997). Intelligent tutoring systems. Em *Handbook of Human-Computer Interaction (Second Edition)*, páginas 849–874. Elsevier.
- Damásio, A. R. (2009). *O erro de Descartes: emoção, razão e o cérebro humano*. Editora Companhia das Letras.
- Dantas, A. C., de Melo, S., Moura, F. e Fernandes, M. (2015). Reconhecimento dinâmico de emoções através de expressões faciais utilizando árvore de decisão. Em *Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)*, volume 26, página 1102.

- de Melo, S. L., de Moura, F. F., Macedo, K., Alves, F. S. e Fernandes, M. A. (2014). Estudo comparativo de técnicas computacionais para classificação de emoções. Em *Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)*, volume 25, página 456.
- DeFalco, J. A., Rowe, J. P., Paquette, L., Georgoulas-Sherry, V., Brawner, K., Mott, B. W., Baker, R. S. e Lester, J. C. (2018). Detecting and addressing frustration in a serious game for military training. *International Journal of Artificial Intelligence in Education*, 28(2):152–193.
- D’Mello, S., Picard, R. W. e Graesser, A. (2007). Toward an affect-sensitive autotutor. *IEEE Intelligent Systems*, 22(4).
- D’Mello, S. K., Craig, S. D., Witherspoon, A., Mcdaniel, B. e Graesser, A. (2008). Automatic detection of learner’s affect from conversational cues. *User modeling and user-adapted interaction*, 18(1):45–80.
- Doddannara, L. S., Gowda, S. M., d Baker, R. S., Gowda, S. M. e De Carvalho, A. M. (2013). Exploring the relationships between design, students’ affective states, and disengaged behaviors within an its. Em *International Conference on Artificial Intelligence in Education*, páginas 31–40. Springer.
- Doubouya, R., Benlamine, M. S., Dufresne, A. e Frasson, C. (2018). Game scenes evaluation and player’s dominant emotion prediction. Em *International Conference on Intelligent Tutoring Systems*, páginas 54–65. Springer.
- Duin, R. P. (2002). The combining classifier: to train or not to train? Em *Proceedings of 16th International Conference on Pattern Recognition*, volume 2, páginas 765–770. IEEE.
- D’Mello, S. (2012). Monitoring affective trajectories during complex learning. Em *Encyclopedia of the Sciences of Learning*, páginas 2325–2328. Springer.
- D’Mello, S., Lehman, B., Sullins, J., Daigle, R., Combs, R., Vogt, K., Perkins, L. e Graesser, A. (2010). A time for emoting: When affect-sensitivity is and isn’t effective at promoting deep learning. Em *Intelligent tutoring systems*, páginas 245–254. Springer.
- Ekman, P. (1992). An argument for basic emotions. *Cognition & emotion*, 6(3-4):169–200.
- Elliott, C. (1997). Affective reasoner personality models for automated tutoring systems. Em *Proceedings of AI in Education’97 Workshop on Pedagogical Agents*, páginas 33–39.
- Frasson, C. e Chalfoun, P. (2010). Managing learner’s affective states in intelligent tutoring systems. Em *Advances in Intelligent Tutoring Systems*, páginas 339–358. Springer.
- Gebhard, P. (2005). Alma: a layered model of affect. Em *Proceedings of the fourth international joint conference on Autonomous agents and multiagent systems*, páginas 29–36. ACM.
- Gebhard, P., Kipp, M., Klesen, M. e Rist, T. (2003). Adding the emotional dimension to scripting character dialogues. Em *International Workshop on Intelligent Virtual Agents*, páginas 48–56. Springer.
- Glodek, M., Reuter, S., Schels, M., Dietmayer, K. e Schwenker, F. (2013). Kalman filter based classifier fusion for affective state recognition. Em *International Workshop on Multiple Classifier Systems*, páginas 85–94. Springer.

- Gottardo, E. e Pimentel, A. (2016). Sistemas tutores afetivos: Estado da arte e desafios para o reconhecimento de emoções relacionadas à aprendizagem via expressões faciais. Em *Anais do XXXVI Congresso da Sociedade Brasileira de Computação*, páginas 557–566. Sociedade Brasileira de Computação - SBC.
- Grafsgaard, J. F., Wiggins, J. B., Boyer, K. E., Wiebe, E. N. e Lester, J. C. (2013a). Automatically recognizing facial indicators of frustration: a learning-centric analysis. Em *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction*, páginas 159–165. IEEE.
- Grafsgaard, J. F., Wiggins, J. B., Boyer, K. E., Wiebe, E. N. e Lester, J. C. (2013b). Embodied affect in tutorial dialogue: student gesture and posture. Em *International Conference on Artificial Intelligence in Education*, páginas 1–10. Springer.
- Gratch, J., Marsella, S., Wang, N. e Stankovic, B. (2009). Assessing the validity of appraisal-based models of emotion. Em *Affective Computing and Intelligent Interaction and Workshops, 2009. ACII 2009. 3rd International Conference on*, páginas 1–8. IEEE.
- Gunes, H., Schuller, B., Pantic, M. e Cowie, R. (2011). Emotion representation, analysis and synthesis in continuous space: A survey. Em *Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on*, páginas 827–834. IEEE.
- Haag, A., Goronzy, S., Schaich, P. e Williams, J. (2004). Emotion recognition using bio-sensors: First steps towards an automatic system. Em *Tutorial and research workshop on affective dialogue systems*, páginas 36–48. Springer.
- Hagemann, N., O’Mahony, M. P. e Smyth, B. (2018). Module advisor: Guiding students with recommendations. Em *International Conference on Intelligent Tutoring Systems*, páginas 319–325. Springer.
- Happy, S. e Routray, A. (2015). Automatic facial expression recognition using features of salient facial patches. *IEEE transactions on Affective Computing*, 6(1):1–12.
- Harley, J. M., Bouchet, F. e Azevedo, R. (2013). Aligning and comparing data on emotions experienced during learning with metatutor. Em *International Conference on Artificial Intelligence in Education*, páginas 61–70. Springer.
- Harley, J. M., Bouchet, F., Hussain, M. S., Azevedo, R. e Calvo, R. (2015). A multi-componential analysis of emotions during complex learning with an intelligent multi-agent system. *Computers in Human Behavior*, 48:615–625.
- Hayashi, E., Posada, J. E. G., Maike, V. R. e Baranauskas, M. C. C. (2016). Exploring new formats of the self-assessment manikin in the design with children. Em *Proceedings of the 15th Brazilian Symposium on Human Factors in Computer Systems*, página 27. ACM.
- Hoque, M. E., McDuff, D. J. e Picard, R. W. (2012). Exploring temporal patterns in classifying frustrated and delighted smiles. *IEEE Transactions on Affective Computing*, 3(3):323–334.
- Izard, C. E. (1990). Facial expressions and the regulation of emotions. *Journal of personality and social psychology*, 58(3):487.

- Jaques, N., Conati, C., Harley, J. M. e Azevedo, R. (2014). Predicting affect from gaze data during interaction with an intelligent tutoring system. Em *International Conference on Intelligent Tutoring Systems*, páginas 29–38. Springer.
- Jaques, P. A. (2004). Using an animated pedagogical agent to interact affectively with the student.
- Jaques, P. A., Bocca, E. e Vicari, R. M. (2003). Considering student's emotions in computational educational systems. Em *Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)*, volume 1, páginas 515–524.
- Jaques, P. A., Lehmann, M. e Jaques, K. S. F. (2008). Avaliando a efetividade de um agente pedagógico animado emocional. Em *Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)*, volume 1, páginas 145–154.
- Jaques, P. A., Nunes, M. A. S., Isotani, S. e Bittencourt, I. (2012). Computação afetiva aplicada a educação: Dotando sistemas tutores inteligentes de habilidades sociais. Em *Anais do Workshop de Desafios da Computação Aplicada à Educação*, páginas 50–59.
- Jaques, P. A., Vicari, R., Pesty, S. e Martin, J.-C. (2011). Evaluating a cognitive-based affective student model. Em *International Conference on Affective Computing and Intelligent Interaction*, páginas 599–608. Springer.
- Jaques, P. A. e Vicari, R. M. (2007). A bdi approach to infer student's emotions in an intelligent learning environment. *Computers & Education*, 49(2):360–384.
- Ji, Y., Chang, K. H. e Hung, C.-C. (2004). Efficient edge detection and object segmentation using gabor filters. Em *Proceedings of the 42nd annual Southeast regional conference - ACM*, páginas 454–459. ACM.
- Kanade, T., Cohn, J. F. e Tian, Y. (2000). Comprehensive database for facial expression analysis. Em *Proceedings of Fourth IEEE International Conference on Automatic Face and Gesture Recognition*, páginas 46–53. IEEE.
- Kapoor, A., Bursleson, W. e Picard, R. W. (2007). Automatic prediction of frustration. *International journal of human-computer studies*, 65(8):724–736.
- Khan, F. A., Graf, S., Weippl, E. R. e Tjoa, A. M. (2010). Identifying and incorporating affective states and learning styles in web-based learning management systems. *IxD&A*, 9:85–103.
- Kort, B., Reilly, R. e Picard, R. W. (2001). An affective model of interplay between emotions and learning: Reengineering educational pedagogy-building a learning companion. Em *Advanced Learning Technologies, 2001. Proceedings. IEEE International Conference on*, páginas 43–46. IEEE.
- Kuncheva, L. I. (2004). *Combining pattern classifiers: methods and algorithms*. John Wiley & Sons.
- Lagud, M. C. V. e Rodrigo, M. M. T. (2010). The affective and learning profiles of students using an intelligent tutoring system for algebra. Em *International Conference on Intelligent Tutoring Systems*, páginas 255–263. Springer.
- Lajoie, S. P. e Derry, S. J. (2013). Motivational techniques of expert human tutors: Lessons for the design of computer-based tutors. Em *Computers as cognitive tools*, páginas 83–114. Routledge.

- Lisetti, C. L. e Gmytrasiewicz, P. (2002). Can a rational agent afford to be affectless? a formal approach. *Applied Artificial Intelligence*, 16(7-8):577–609.
- Littlewort, G., Bartlett, M. S., Fasel, I., Susskind, J. e Movellan, J. (2006). Dynamics of facial expression extracted automatically from video. *Image and Vision Computing*, 24(6):615–625.
- Matsumoto, D. e Ekman, P. (2004). The relationship among expressions, labels, and descriptions of contempt. *Journal of personality and social psychology*, 87(4):529.
- McCrae, R. R. e John, O. P. (1992). An introduction to the five-factor model and its applications. *Journal of personality*, 60(2):175–215.
- McDaniel, B., D’Mello, S., King, B., Chipman, P., Tapp, K. e Graesser, A. (2007). Facial features for affective state detection in learning environments. Em *Proceedings of the Cognitive Science Society*, volume 29, páginas 467–472.
- Mella, L. e Bellavista, D. (2016). Emotime: Recognizing emotional states in faces.
- Morais, F., da Silva, J., Reis, H., Isotani, S. e Jaques, P. (2017). Computação afetiva aplicada à educação: uma revisão sistemática das pesquisas publicadas no brasil. Em *Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)*, volume 28, páginas 163–172.
- Muñoz, K., Mc Kevitt, P., Lunney, T., Noguez, J. e Neri, L. (2010). Playphysics: an emotional games learning environment for teaching physics. Em *International Conference on Knowledge Science, Engineering and Management*, páginas 400–411. Springer.
- Muñoz, K., Mc Kevitt, P., Lunney, T., Noguez, J. e Neri, L. (2011). An emotional student model for game-play adaptation. *Entertainment Computing*, 2(2):133–141.
- Nicolaou, M. A., Gunes, H. e Pantic, M. (2011). Continuous prediction of spontaneous affect from multiple cues and modalities in valence-arousal space. *IEEE Transactions on Affective Computing*, 2(2):92–105.
- Okur, E., Alyuz, N., Aslan, S., Genc, U., Tanriover, C. e Esmé, A. A. (2017). Behavioral engagement detection of students in the wild. Em *International Conference on Artificial Intelligence in Education*, páginas 250–261. Springer.
- Ortony, A., Clore, G. L. e Collins, A. (1990). *The cognitive structure of emotions*. Cambridge university press.
- Padrón-Rivera, G. e Rebolledo-Mendez, G. (2015). Identifying affective trajectories in relation to learning gains during the interaction with a tutoring system. Em *International Conference on Artificial Intelligence in Education*, páginas 756–759. Springer.
- Padrón-Rivera, G., Rebolledo-Mendez, G., Parra, P. P. e Huerta-Pacheco, N. S. (2016). Identification of action units related to affective states in a tutoring system for mathematics. *Journal of Educational Technology & Society*, 19(2).
- Paquette, L., Baker, R. S., Sao Pedro, M. A., Gobert, J. D., Rossi, L., Nakama, A. e Kauffman-Rogoff, Z. (2014). Sensor-free affect detection for a simulation-based science inquiry learning environment. Em *International Conference on Intelligent Tutoring Systems*, páginas 1–10. Springer.

- Paquette, L., Rowe, J., Baker, R., Mott, B., Lester, J., DeFalco, J., Brawner, K., Sottolare, R. e Georgoulas, V. (2016). Sensor-free or sensor-full: A comparison of data modalities in multi-channel affect detection. *International Educational Data Mining Society*.
- Pardos, Z. A., Baker, R. S., San Pedro, M. O., Gowda, S. M. e Gowda, S. M. (2013). Affective states and state tests: Investigating how affect throughout the school year predicts end of year learning outcomes. Em *Proceedings of the Third International Conference on Learning Analytics and Knowledge*, páginas 117–124. ACM.
- Perusquia-Hernandez, M., Hirokawa, M. e Suzuki, K. (2017). A wearable device for fast and subtle spontaneous smile recognition. *IEEE Transactions on Affective Computing*, 8(4):522–533.
- Pham, P. e Wang, J. (2018). Predicting learners' emotions in mobile mooc learning via a multimodal intelligent tutor. Em *International Conference on Intelligent Tutoring Systems*, páginas 150–159. Springer.
- Picard, R. W. (1997). *Affective computing*, volume 252. MIT press Cambridge.
- Picard, R. W., Fedor, S. e Ayzenberg, Y. (2016). Multiple arousal theory and daily-life electrodermal activity asymmetry. *Emotion Review*, 8(1):62–75.
- Picard, R. W., Papert, S., Bender, W., Blumberg, B., Breazeal, C., Cavallo, D., Machover, T., Resnick, M., Roy, D. e Strohecker, C. (2004). Affective learning—a manifesto. *BT technology journal*, 22(4):253–269.
- Pintrich, P. R. et al. (1991). A manual for the use of the motivated strategies for learning questionnaire (mslq).
- Plutchik, R. (1980). A general psychoevolutionary theory of emotions. *Emotion: Theory, research and experience*, 1:3–33.
- Poria, S., Cambria, E., Bajpai, R. e Hussain, A. (2017). A review of affective computing: From unimodal analysis to multimodal fusion. *Information Fusion*, 37:98 – 125.
- Posner, J., Russell, J. A. e Peterson, B. S. (2005). The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. *Development and psychopathology*, 17(03):715–734.
- Pour, P. A., Hussain, M. S., AlZoubi, O., D'Mello, S. e Calvo, R. A. (2010). The impact of system feedback on learners' affective and physiological states. Em *International Conference on Intelligent Tutoring Systems*, páginas 264–273. Springer.
- Reis, H., Alvares, D., Jaques, P. e Isotani, S. (2018a). Analysis of permanence time in emotional states: A case study using educational software. Em *International Conference on Intelligent Tutoring Systems - ITS*, páginas 180–190. Springer.
- Reis, H. M., Maillard, P. A. J. e Isotani, S. (2018b). Sistemas tutores inteligentes que detectam as emoções dos estudantes: um mapeamento sistemático. *Brazilian Journal of Computers in Education (Revista Brasileira de Informática na Educação - RBIE)*, 26(03):76–107.
- Roberts, B., Pioch, N. e Ferguson, W. (2000). Verbal coaching during a real-time task. *International Journal of Artificial Intelligence in Education (IJAIED)*, 11:377–388.

- Robson, C. (1993). Real world research: A resource for social scientists and practitioners-researchers. *Massachusetts: Blackwell Publishers.*
- Roseman, I. J., Spindel, M. S. e Jose, P. E. (1990). Appraisals of emotion-eliciting events: Testing a theory of discrete emotions. *Journal of personality and social psychology*, 59(5):899.
- Rosenberg, E. L. e Ekman, P. (1994). Coherence between expressive and experiential systems in emotion. *Cognition & Emotion*, 8(3):201–229.
- Russel, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, 39:1161–1178.
- Salmeron-Majadas, S., Santos, O. C. e Boticario, J. G. (2015). Towards multimodal affective detection in educational systems through mining emotional data sources. Em *International Conference on Artificial Intelligence in Education*, páginas 860–863. Springer.
- Sarrafzadeh, A., Alexander, S., Dadgostar, F., Fan, C. e Bigdeli, A. (2008). “how do you know that i don’t understand?” a look at the future of intelligent tutoring systems. *Computers in Human Behavior*, 24(4):1342–1363.
- Schels, M., Glodek, M., Meudt, S., Scherer, S., Schmidt, M., Layher, G., Tschechne, S., Brosch, T., Hrabal, D., Walter, S. et al. (2013). Multi-modal classifier-fusion for the recognition of emotions. *Coverbal Synchrony in Human-Machine Interaction*, páginas 73–98.
- Scherer, K. R. (2000). Psychological models of emotion. *The neuropsychology of emotion*, 137(3):137–162.
- Scherer, K. R. (2005). What are emotions? and how can they be measured? *Social science information*, 44(4):695–729.
- Scherer, K. R., Schorr, A. e Johnstone, T. (2001). *Appraisal processes in emotion: Theory, methods, research*. Oxford University Press.
- Shen, L., Leon, E., Callaghan, V. e Shen, R. (2007). Exploratory research on an affective e-learning model. Em *Proceedings of Workshop on Blended Learning*, páginas 267–278.
- Shen, L., Wang, M. e Shen, R. (2009). Affective e-learning: Using "emotional" data to improve learning in pervasive learning environment. *Journal of Educational Technology & Society*, 12(2):176.
- Steunebrink, B. R., Dastani, M. e Meyer, J.-J. C. (2009). The occ model revisited. Em *Proc. of the 4th Workshop on Emotion and Computing*.
- Tiam-Lee, T. J. e Sumi, K. (2018). Adaptive feedback based on student emotion in a system for programming practice. Em *International Conference on Intelligent Tutoring Systems*, páginas 243–255. Springer.
- Tulyakov, S., Jaeger, S., Govindaraju, V. e Doermann, D. (2008). Review of classifier combination methods. Em *Machine Learning in Document Analysis and Recognition*, páginas 361–386. Springer.
- Vail, A. K., Grafsgaard, J. F., Boyer, K. E., Wiebe, E. N. e Lester, J. C. (2016). Predicting learning from student affective response to tutor questions. Em *International Conference on Intelligent Tutoring Systems*, páginas 154–164. Springer.

- Van Den Broek, E. L. (2012). Affective computing: a reverence for a century of research. Em *Cognitive Behavioural Systems*, páginas 434–448. Springer.
- Vanlehn, K. (2006). The behavior of tutoring systems. *International journal of artificial intelligence in education*, 16(3):227–265.
- Viola, P. e Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. Em *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, volume 1, páginas 511–518. IEEE.
- Whitehill, J. e Omlin, C. W. (2006). Haar features for faces au recognition. Em *Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition*, páginas 97–101. IEEE.
- Witten, I. H., Frank, E., Hall, M. A. e Pal, C. J. (2016). *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann.
- Wixon, M., Arroyo, I., Muldner, K., Burleson, W., Rai, D. e Woolf, B. (2014). The opportunities and limitations of scaling up sensor-free affect detection. Em *Educational Data Mining 2014*. Citeseer.
- Woolf, B., Burleson, W., Arroyo, I., Dragon, T., Cooper, D. e Picard, R. (2009). Affect-aware tutors: recognising and responding to student affect. *International Journal of Learning Technology*, 4(3-4):129–164.
- Woolf, B. P., Arroyo, I., Cooper, D., Burleson, W. e Muldner, K. (2010). Affective tutors: Automatic detection of and response to student emotion. Em *Advances in Intelligent Tutoring Systems*, páginas 207–227. Springer.
- Xiao, X., Pham, P. e Wang, J. (2017). Dynamics of affective states during mooc learning. Em *International Conference on Artificial Intelligence in Education*, páginas 586–589. Springer.
- Yerkes, R. M. e Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit-formation. *Journal of comparative neurology*, 18(5):459–482.
- Yusoff, Y. M., Ruthven, I. e Landoni, M. (2013). Measuring emotion: A new evaluation tool for very young children. Em *Proc. of the 4th Int. Conf. on Computing and Informatics, ICOCI*.
- Zagalo, N., Torres, A. e Branco, V. (2005). Emotional spectrum developed by virtual storytelling. Em *International Conference on Virtual Storytelling*, páginas 105–114. Springer.
- Zatarain-Cabada, R., Barrón-Estrada, M. L., Camacho, J. O. e Reyes-García, C. A. (2013). Integrating learning styles and affect with an intelligent tutoring system. Em *2013 12th Mexican International Conference on Artificial Intelligence (MICA)*, páginas 247–253. IEEE.
- Zhang, F., Mao, Q., Shen, X., Zhan, Y. e Dong, M. (2018). Spatially coherent feature learning for pose-invariant facial expression recognition. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, 14(1s):27.