

ROSE YURI SHIMIZU

**MODELO DE INFERÊNCIA DOS PARÂMETROS CAUSAIS:
INCORPORAÇÃO DA PROBABILIDADE DE
CONHECIMENTO NAS ATRIBUIÇÕES DAS CAUSAS DOS
SUCESSOS OU FRACASSOS EM ATIVIDADES
COGNITIVAS E SUA RELAÇÃO COM AS AÇÕES DOS
ESTUDANTES**

Dissertação apresentada como requisito parcial à obtenção do grau de Mestre. Programa de Pós-Graduação em Informática, Setor de Ciências Exatas, Universidade Federal do Paraná.

Orientador: Prof. Dr. Andrey Ricardo Pimentel

CURITIBA

2015

ROSE YURI SHIMIZU

**MODELO DE INFERÊNCIA DOS PARÂMETROS CAUSAIS:
INCORPORAÇÃO DA PROBABILIDADE DE
CONHECIMENTO NAS ATRIBUIÇÕES DAS CAUSAS DOS
SUCESSOS OU FRACASSOS EM ATIVIDADES
COGNITIVAS E SUA RELAÇÃO COM AS AÇÕES DOS
ESTUDANTES**

Dissertação apresentada como requisito parcial à obtenção do grau de Mestre. Programa de Pós-Graduação em Informática, Setor de Ciências Exatas, Universidade Federal do Paraná.

Orientador: Prof. Dr. Andrey Ricardo Pimentel

CURITIBA

2015

S556m

Shimizu, Rose Yuri

Modelo de inferência dos parâmetros causais : incorporação da probabilidade de conhecimento nas atribuições das causas dos sucessos ou fracassos em atividades cognitivas e sua relação com as ações dos estudantes/ Rose Yuri Shimizu. – Curitiba, 2015.

68 f. : il. color. ; 30 cm.

Dissertação - Universidade Federal do Paraná, Setor de Ciências Exatas, Programa de Pós-graduação em Informática, 2015.

Orientador: Andrey Ricardo Pimentel .

Bibliografia: p. 64-68.

1. Aprendizagem cognitiva. 2. Atribuição (Psicologia social). 3. Sistemas tutoriais inteligentes. I. Universidade Federal do Paraná. II.Pimentel, Andrey Ricardo. III. Título.

CDD: 153.15



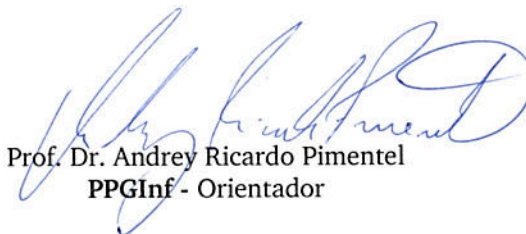
Ministério da Educação
Universidade Federal do Paraná
Programa de Pós-Graduação em Informática

PARECER

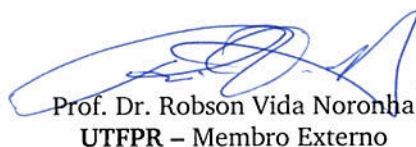
Nós, abaixo assinados, membros da Banca Examinadora da defesa de Dissertação de Mestrado em Informática, da aluna Rose Yuri Shimizu, avaliamos o trabalho intitulado, “Modelo de inferência dos parâmetros causais: incorporação da probabilidade de conhecimento nas atribuições das causas dos sucessos ou fracassos em atividades cognitivas e sua relação com as ações dos estudantes”, cuja defesa foi realizada no dia 23 de outubro de 2015, às 09:00 horas, no Departamento de Informática do Setor de Ciências Exatas da Universidade Federal do Paraná. Após a avaliação, decidimos pela:

aprovação do(a) candidato(a). **reprovação** do(a) candidato(a).

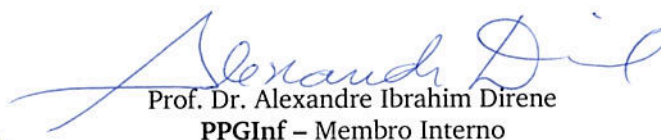
Curitiba, 23 de outubro de 2015.



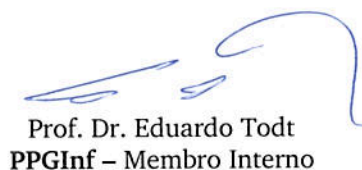
Prof. Dr. Andrey Ricardo Pimentel
PPGInf - Orientador



Prof. Dr. Robson Vida Noronha
UTFPR – Membro Externo



Prof. Dr. Alexandre Ibrahim Direne
PPGInf – Membro Interno



Prof. Dr. Eduardo Todt
PPGInf – Membro Interno



SUMÁRIO

LISTA DE FIGURAS	v
LISTA DE TABELAS	vi
RESUMO	vii
ABSTRACT	viii
1 INTRODUÇÃO	1
1.1 Objetivo geral	2
1.2 Objetivos Específicos	2
1.3 Organização do trabalho	3
2 TRABALHOS CORRELATOS	4
2.1 Análise de dados para modelar o estado afetivo	4
2.2 Análise de dados para detectar o estado cognitivo	6
2.3 Análise de dados para modelar o estado afetivo e detectar o estado cognitivo	7
2.4 Considerações finais	8
3 FUNDAMENTAÇÃO TEÓRICA	10
3.1 Teoria da Atribuição Causal	10
3.2 <i>Bayesian Knowledge Tracing</i> e suas variantes	14
3.2.1 Abordagens para ajustar o modelo BKT	19
3.2.2 Detectando Aprendizagem Momento a Momento	21
3.3 Considerações finais	26
4 MODELO DE INFERÊNCIA DOS PARÂMETROS CAUSAIS	28
4.1 Definição dos parâmetros causais	31
4.2 Cálculo dos parâmetros causais	34

4.2.1	Modelo <i>Bayesian Knowledge Tracing</i> (BKT) de Corbett e Anderson	34
4.2.2	Modelo ajustado de Baker, Goldstein e Heffernan	36
4.2.3	Cálculo da inferência dos parâmetros causais	37
4.3	Classificação e testes: análise pré-ação dos parâmetros causais com um conjunto de dados reais	44
4.3.1	Preparação dos dados	45
4.3.2	Procedimentos	47
4.3.3	Resultados	50
4.4	Considerações Finais	59
5	CONCLUSÃO E TRABALHOS FUTUROS	61
5.1	Trabalhos Futuros	63
	BIBLIOGRAFIA	68

LISTA DE FIGURAS

3.1	Diagrama da teoria da Atribuição Causal de Weiner [37]	13
3.2	Modelo de aprendizagem	14
3.3	Modelo BKT	16
3.4	Atualização de duas fases. Fonte [31]	17
4.1	Diagrama do processo geral do modelo proposto para o cálculo dos parâmetros causais.	29
4.2	Fases da atualização dos parâmetros causais	30

LISTA DE TABELAS

4.1	Validação cruzada para o conhecimento condicionado.	49
4.2	Resultados do modelo determinístico na previsão da ação do estudante através dos valores obtidos com o modelo dos parâmetros causais.	54
4.3	Decisões geradas pelo classificador e interpretações	55

RESUMO

O processo cognitivo é uma atividade que envolve vários componentes na aquisição, organização e uso da informação. A aprendizagem é uma tarefa imprecisa uma vez que o processo cognitivo envolve fatores internos dos seres humanos, como o estado afetivo e motivacional.

Neste trabalho foi desenvolvido um modelo para a inferência das probabilidades das causas dos sucessos (acertos) ou fracassos (erros) em atividades cognitivas.

As atribuições causais foram retiradas da teoria da “Atribuição Causal”. Esta teoria apresenta as percepções causais e sua relação com a emoção, motivação e expectativa.

As fórmulas foram modeladas utilizando o nível de conhecimento do estudante que é calculado de forma probabilística pelo modelo *Bayesian Knowledge Tracing* (BKT). Também é utilizada uma extensão do BKT que explora as ações futuras do estudante para determinar o nível e o momento da aprendizagem. Destaca-se que o modelo proposto faz a especialização dos parâmetros do modelo BKT em novos parâmetros visando aprofundar o diagnóstico dos fatores influenciadores do processo cognitivo dos estudantes.

Prevendo essas percepções causais pretende-se colaborar com o processo de identificação comportamental do estudante: antes da ocorrência da ação para a previsão de sua resposta, ou após a ação, atuando como um auxiliar na decisão em estratégias de ensino. Nesta dissertação foi simulado o impacto dos valores inferidos para a previsão da resposta do discente através de bases de dados dos históricos de ações de estudantes com um Sistema Tutor Inteligente (STI) em produção.

ABSTRACT

The cognitive process is an activity that involves various components in the acquisition, organization and use of information. Learning is an imprecise task since the cognitive process involves internal factors of humans as the emotional and motivational state.

In this work, we developed a model for inference of probability causes of successes or failures in cognitive activities.

The causal attributions were based on the psychology theory of “attributional theory of motivation and emotion”. This theory introduces the causal perceptions and their relationship with the emotion, motivation and expectation.

The formulas were modeled using the level of student’s knowledge which is calculated in a probabilistic manner by the model *Bayesian Knowledge Tracing* (BKT). Also we use an extension of the BKT which explores future actions of students to set the level and the moment of learning. We highlight that the proposed model uses the BKT parameters to create new parameters, more specialized, aiming to identify factors influencing the cognitive process of students.

Anticipating these causal perceptions, we intend to collaborate with the student’s behavioral identification process: before the action for students answer prediction, or after the action, acting as a subsidiary on the decision of education strategy. In this thesis it was simulated the impact of inferred values for the prediction of students action through an historical data base of students actions of a production Intelligent Tutoring System (ITS).

CAPÍTULO 1

INTRODUÇÃO

O processo cognitivo é uma atividade que envolve vários componentes na aquisição, organização e uso da informação.

Para contribuir no processo cognitivo, surgem mecanismos de apoio ao ensino por computador como os Sistemas Tutores Inteligentes (STIs). Nwana [26] descreve um STI como um programa de computador que incorpora técnicas de inteligência artificial para tentar reproduzir um comportamento, que se tivesse sido feito por um humano, seria considerado inteligente. Os STIs, segundo Nwana [26], tem sua concepção e desenvolvimento baseados na intersecção da ciência da computação, psicologia cognitiva e pesquisa educacional.

Para melhorar a colaboração dos STIs no processo cognitivo, principalmente em relação às informações referentes as características do estudante e estratégias pedagógicas, há pesquisas aplicando técnicas de aprendizagem de máquina [11, 11, 28], modelos probabilísticos [19, 4] e computação afetiva [18, 14]. Estas pesquisas visam uma interação mais “humana” com os usuários através de respostas estimuladas pelos estados cognitivos e/ou afetivos que impactam na motivação e aprendizagem do estudante.

O estado cognitivo pode ser interpretado como a capacidade do indivíduo no processamento de informações para a aquisição do conhecimento, ou seja, no processo cognitivo.

Os estados afetivos são resultantes do processo afetivo que Bandura [6] definiu como a regulação dos estados emocionais e elicitação das reações emocionais. E esses estados são refletidos no comportamento do indivíduo através da sua motivação em desenvolver atividades.

As emoções envolvidas no processo cognitivo podem ter diversas causas. O conhecimento sobre essas causas colabora no diagnóstico dos consequentes estados afetivos que geram diferentes comportamentos. A análise mais profunda das características do aluno passa a defini-lo como agente ativo no processo de aprendizagem.

O principal motivador desta dissertação é a possibilidade de contribuir no processo de construção de um STI na parte de detecção dessas causas que influenciam a aprendizagem.

1.1 Objetivo geral

O objetivo geral deste trabalho é estabelecer um modelo de inferência dos parâmetros das causas do sucesso (acertos) ou fracasso (erros) em uma atividade cognitiva.

Como os elementos dependentes dos sentimentos humanos não são mensuráveis com exatidão, o modelo utiliza em suas fórmulas a probabilidade de conhecimento e outras informações obtidas do histórico das transações dos estudantes com um STI. A probabilidade de conhecimento é uma medida interessante pois é alterada por fatores internos dos indivíduos, podendo ser calculado através de dados observáveis como a resposta correta, e é uma possível evidência resultante das emoções e motivações influenciadas pelas causas.

1.2 Objetivos Específicos

Para o desenvolvimento do modelo proposto é necessário embasar as causas que serão inferidas de acordo com um estudo psicológico. Este fundamento psicológico possibilitará, futuramente, um estudo mais detalhado de como as causas inferidas pelo modelo influenciam o processo cognitivo de um indivíduo através das possíveis consequências afetivas.

A teoria da “Atribuição Causal” de Weiner [37] será utilizada para a definição dessas atribuições causais. Esta teoria apresenta as percepções causais e sua relação com a motivação e expectativa. Prevendo essas percepções causais pretende-se colaborar com o processo de identificação comportamental do estudante: antes da ocorrência da ação para a previsão de sua resposta, ou após a ação, atuando como um auxiliar na decisão em estratégias de ensino.

Além da escolha das causas, outra decisão importante é qual método utilizar para o cálculo da probabilidade de conhecimento dos estudantes. A escolha foi do modelo *Bayesian Knowledge Tracing* (BKT) de Corbett e Anderson [19] e da variante deste modelo, desenvolvido por Baker, Goldstein e Heffernan [4], que condiciona o conhecimento

à duas respostas futuras determinar o nível e o momento da aprendizagem.

Uma aplicação prática será apresentada: previsão da ação, acerto ou erro, de um estudante na sua primeira tentativa de resolver uma questão através de um conjunto de atributos, incluindo os parâmetros causais inferidos pelo modelo proposto. Esta classificação e teste será feita com bases de dados de um STI que foi utilizado por alunos reais. O intuito desta aplicação é mostrar o impacto dos atributos causais modelados e calculados na previsão da ação, e assim analisar a possibilidade da utilização desses parâmetros na determinação do comportamento do estudante.

1.3 Organização do trabalho

No capítulo 2 é apresentada uma resenha literária das pesquisas que focam a detecção de padrões comportamentais para a inferência de estados afetivos e cognitivos.

No capítulo 3 há uma introdução às informações teóricas relevantes utilizadas no desenvolvimento do modelo proposto.

No capítulo 4 são explicadas as fórmulas de predição dos parâmetros causais, os procedimentos, testes e resultados.

E no capítulo 5 é apresentada a conclusão, com o impacto encontrado através desta pesquisa e os trabalhos futuros.

CAPÍTULO 2

TRABALHOS CORRELATOS

Inspirados no impacto da motivação no processo cognitivo, muitos pesquisadores trabalham na detecção de padrões comportamentais para a inferência de estados afetivos e cognitivos.

Os estados afetivos são resultantes do processo afetivo que Bandura [6] definiu como a regulação dos estados emocionais e elicitación das reações emocionais. O estado cognitivo pode ser interpretado como a capacidade do indivíduo no processo cognitivo.

Apesar de esta dissertação não se aprofundar no quesito afetivo, este é um assunto relacionado com a investigação das causas comportamentais. Nas próximas seções são apresentadas algumas pesquisas com foco na análise comportamental através do reconhecimento e reações às emoções, e das características dos estudantes.

2.1 Análise de dados para modelar o estado afetivo

O estado afetivo influencia a aprendizagem conforme apresentado em trabalhos como o de Craig, Graesser, Sullins e Gholson [20] e Stein, Levine, Stein, Leventhal e Trabasso [36]. Essas pesquisas exploraram a captação das ações do estudante e correlacionaram com as suas emoções. Há diversas formas de fazer essa captação, podendo ser feita por meio de sensores, imagens, tempo de interação/repouso com o sistema, questionários, padrões de ações, entre outros.

Stein et al. [36] propuseram um modelo afetivo orientado a objetivos na solução de problemas. Este modelo sustenta a ideia que o comportamento tem a finalidade de alcançar e manter uma meta, e que a aprendizagem quase sempre ocorre durante um episódio emocional. Descrevem os pensamentos que ocorrem durante esses episódios e como influenciam uns aos outros. Os dados referentes aos pensamentos vêm de informações previamente adquiridas e do monitoramento da excitação do “sistema nervoso autônomo (SNA)”. Junto

com a excitação do SNA, ocorre uma avaliação cognitiva conjunta com a reação emocional.

O trabalho de Craig et al. [20] explora o papel dos estados afetivos no processo de aprendizagem com um STI chamado *AutoTutor*. Por meio de informações comportamentais, obtidas pela interação do estudante com um sistema de diálogo e por câmera, rastreiam as emoções do estudante. As emoções estudadas foram o tédio, a confusão, a situação de “eureka” (repentina aprendizagem), a fluência e a frustração. Concluíram que o ganho de aprendizagem é positivamente correlacionado com os estados de confusão e fluência, e negativamente correlacionado com o tédio. Para frustração e “eureka” os ganhos foram insignificantes.

Outra teoria que prioriza as metas foi a de Conati [17] e, Conati e Maclaren [18] que implementaram um *framework* para modelar os estados afetivos usando redes *Bayesianas* dinâmicas informando as causas das emoções e seus efeitos no comportamento.

A construção do modelo Conati et al. [17, 18] foi baseada nos elementos da teoria psicológica “OCC” (Ortony, Clore e Collins [27]) de emoções. Esta teoria cognitiva modela a emoção cuja força depende dos eventos, agentes ou objetos do ambiente. A força foi comparada a valência, sendo positiva ou negativa. A valência de uma reação emocional depende da situação, que é definida pelos seus objetivos e preferências. A teoria OCC especifica 22 categorias de emoções, com 5 processos. Esses processos são: classificação do evento, ação ou objeto encontrado; determinação da intensidade das emoções envolvidas; interação da emoção gerada com as existentes; mapeamento do estado emocional para uma expressão emocional; e expressão do estado emocional. Nesta teoria, as emoções surgem como resultado da avaliação da situação atual em relação a seus objetivos.

As situações modeladas na rede de Conati et al. [17, 18] consistem do resultado de qualquer evento causado pela ação do usuário ou do agente em relação aos objetivos, o qual influencia os estados emocionais. Posteriormente, a rede correlaciona as reações corporais (efeitos observáveis) com as emoções. As reações corporais foram obtidas através de sensores que podem detectar as expressões corporais de interesse.

Segundo Conati et al. [18] o resultado mostrou que se os objetivos dos usuários puderem ser corretamente avaliados então o modelo pode fazer previsões com precisão razoável

sobre o estado afetivo do usuário, mas também revelou algumas fontes de imprecisão. A avaliação das metas precisa ser aperfeiçoada antes do modelo poder ser usado autonomamente com um sistema real.

2.2 Análise de dados para detectar o estado cognitivo

Outros pesquisadores utilizam a aprendizagem de máquina para aumentar a individualização das ações do sistema de acordo com as características dos estudantes, como Paiva, Santa Pinto, da Silva, Isotani e Jaques [28], Beck, Woolf e Beal [11] e, Chi, VanLehn e Litman [16].

Paiva et al. [28] propuseram uma abordagem sistemática para auxiliar a assistência aos alunos em suas necessidades pedagógicas. Denominaram o processo como *Pedagogical Recommendation Process (PRP)*. Esta abordagem é um processo cíclico e interativo, constituído por quatro passos. Cada etapa exige ações coordenadas entre os atores humanos (especialistas nos domínios pedagógicos e tecnológicos) e os atores artificiais (artefatos computacionais). Ele usa os dados da interação com os alunos como entrada, aplicando técnicas de mineração de dados para detectar e descobrir as dificuldades pedagógicas, a fim de personalizar as recomendações pedagógicas. Finalmente, o desempenho dos alunos é monitorado e avaliado. Aplicando esse processo em um estudo de caso, descobriram situações em que os alunos estavam enfrentando dificuldades, o que ajudou a gerar recomendações personalizadas com base nestes problemas específicos.

Beck et al. [11] implementaram uma arquitetura para aprendizagem de máquina que treina o sistema para aprender como ensinar os estudantes. Essa arquitetura, denominada “ADVISOR”, é formada por agentes que identificam as ações e desempenho do aluno em vários contextos de problemas, e simula o estudante com diferentes políticas de ensino. A aprendizagem por reforço [10] foi utilizada pelo agente para derivar a política de ensino para um objetivo educacional específico.

Assim como o ADVISOR [11], Chi et al. [16] propuseram uma tática tutorial induzida pela aplicação da aprendizagem por reforço nos dados dos diálogos da interação estudante-tutor. As políticas de decisão são *Elicit vs Tell* e *Skip-Justify*. Na política *Elicit vs Tell*,

o sistema decide se induz (*elicit*) ou se, simplesmente, informa (*tell*) qual é a próxima etapa; na política *Skip-Justify*, o sistema pode solicitar uma justificativa para o estudante, sobre o passo adotado ou resposta inserida.

2.3 Análise de dados para modelar o estado afetivo e detectar o estado cognitivo

A junção do estado afetivo e cognitivo em um sistema tutor inteligente foi um caminho natural no aperfeiçoamento no ensino por computador.

Bica e Vicari [14] apresentaram um agente denominado “Mediador de Autoeficácia (MAE)”, atuante no modelo do estudante, que considera as inferências das crenças de autoeficácia. A autoeficácia é a capacidade do estudante realizar uma tarefa e seu auto-julgamento dessas capacidades, sendo importante na motivação e afeto. As inferências foram realizadas através da máquina de inferência *fuzzy*.

As decisões de *feedback* no MAE foram efetuadas através de uma base de conhecimento, constantemente atualizada, e das inferências sobre os aspectos de autoeficácia do estudante que são associadas aos comportamentos, como esforço, persistência, números de tentativas, desempenho médio, entre outros. Conforme o nível de autoeficácia inferida, um tipo de *feedback* motivacional é apresentado ao estudante.

Beck e Woolf [13] desenvolveram o STI *AnimalWatch* para ensinar aritmética. O modelo do estudante deste sistema foi aprimorado com as informações sobre os desempenhos dos estudantes para cada item do domínio, ajudando na seleção do tópico, da dificuldade e do *feedback* mais adequado à situação corrente. Implementaram uma arquitetura de aprendizagem de máquina para prever a probabilidade da próxima resposta do estudante ser correta, e quanto tempo o aluno levará para gerar essa resposta. Esse agente trabalha com 4 informações de entrada: a proficiência do estudante; a dificuldade do tópico e seus componentes; a complexidade do problema; e o esforço e as dicas vistas pelo estudante.

Outro STI no campo da matemática é o “Aplusix II” de Lagud e Rodrigo [24]. O “Aplusix II” é um assistente de aprendizagem de álgebra. O sistema permite que o

estudante resolva as equações algébricas passo-a-passo, fornecendo *feedbacks* instrutivos da relação entre as etapas respondidas e possíveis erros cometidos.

Para o auxílio ao estudante foi mapeado a relação entre o afeto e o perfil. Para o perfil de aprendizagem foram analisados os históricos dos estudantes para obter o número de itens corretos resolvidos, o maior nível de dificuldade exercitado, a média de tempo e a média de etapas para resolver um item. Os estados afetivos foram determinados através de dados coletados usando métodos de observação quantitativa apresentados no trabalho Baker, Corbett, Koedinger e Wagner [1]. Essas observações foram realizadas por uma equipe de seis alunos de mestrado em educação ou ciência da computação, com experiência de ensino. Eles foram treinados para a tarefa através de uma série de discussões sobre o significado das categorias afetivas. As observações foram conduzidas de acordo com um guia que deu exemplos de ações, declarações, expressões faciais e linguagem corporal para categorizar os estados de tédio, confusão, prazer, surpresa, frustração, fluência e o estado neutro.

2.4 Considerações finais

Várias alternativas de auxílio ao estudante estão sendo modeladas e implementadas.

Há os que trabalham para a determinação dos estados afetivos, como: Stein et al. [36] que propuseram um modelo afetivo orientado a objetivos, considerando os episódios emocionais ligados ao momento da aprendizagem; Craig et al. [20] que exploraram o papel dos estados afetivos por meio de informações comportamentais, obtidas pela interação do estudante com um sistema de diálogo e por câmera; e Conati et al. [18] que construíram um modelo com dados integrados com a teoria psicológica OCC da emoção e personalidade. Uma característica semelhante dos modelo acima citados, é o uso de informações coletados de agentes externos como câmeras.

As informações históricas dos estudantes também são úteis para promover a contextualização e individualização dos retornos ao estudante, estimulando o processo cognitivo. O ADVISOR [11], por exemplo, explora o modelo do estudante e o estado do problema para definir a política de ensino simulando o estudante. Enquanto Paiva et al. [28] pro-

puseram uma abordagem sistemática para auxiliar a assistência aos alunos, detectando as dificuldades, a fim de personalizar as recomendações pedagógicas, Chi et al. [16] propuseram uma tática tutorial induzida pela aplicação da aprendizagem por reforço nos dados dos diálogos da interação estudante-tutor.

A associação das abordagens emoção-cognição foi o intuito de pesquisas como de Bica e Vicari [14] que considera as inferências das crenças de autoeficácia para escolher o *feedback* motivacional, de Beck e Woolf [13] que desenvolveram um modelo do estudante com informações sobre o desempenho para cada item no domínio, ajudando na seleção do tópico, da dificuldade e do *feedback* mais adequado à situação corrente, e de Lagud e Rodrigo [24] que personalizaram os *feedbacks* com o apoio da relação entre o afeto, e o perfil emocional e cognitivo.

O modelo desenvolvido nesta dissertação acompanha a tendência da análise dos dados para obter informações referentes ao comportamento e estados dos estudantes. O principal objetivo é encontrar padrões que expliquem a causa de um determinado comportamento. Semelhante a ideia de Conati e Maclaren [18], contempla a detecção das causas dos eventos e fundamenta-se em uma teoria psicológica. Porém, não há o uso de dados de ferramentas de captação da interação estudante-sistema (camêras, sensores), a teoria utilizada é a da “Atribuição Causal” proposta por Weiner [37] e faz o relacionamento da causa com o nível de conhecimento do estudante obtido aplicando o modelo *Bayesian Knowledge Tracing* (BKT) de Corbett e Anderson [19].

Conhecendo as causas seria possível deduzir a motivação do estudante, o que ajudaria, por exemplo, no aprimoramento dos *feedbacks*. Os métodos aplicados estão apresentados no capítulo 3 e 4.

CAPÍTULO 3

FUNDAMENTAÇÃO TEÓRICA

Neste capítulo será apresentada uma introdução sobre os modelos e técnicas que esta pesquisa fundamenta-se.

A base psicológica foi feita através da teoria da “Atribuição Causal” de Weiner [37] que é apresentada na primeira seção deste capítulo. As causas que resultam na motivação e expectativa são os objetos de interesse estudados.

O capítulo segue com informações sobre o modelo *Bayesian Knowledge Tracing* (BKT) de Corbett e Anderson [19] e suas variantes, que foram utilizadas como base técnica para o cálculo das atribuições causais inspiradas na teoria acima citada.

3.1 Teoria da Atribuição Causal

A aprendizagem é uma tarefa imprecisa uma vez que o processo cognitivo envolve fatores internos dos seres humanos. Para Bandura [5], pessoas não são nem agentes autônomos nem simplesmente transportadores mecânicos das influências ambientais. O ser humano representa um agente passivo, ao ser influenciado pelos fatores externos, e ativo, ao contribuir com sua interpretação pessoal na causa da sua própria motivação.

Na teoria da “Atribuição Causal”, Weiner [37] defendeu que as atribuições causais do sucesso ou fracasso interferem no processo cognitivo e no comportamento do estudante, uma vez que atuam na emoção e motivação. A motivação é representada como uma sequência histórica, resultado do vínculo do processo afetivo, ou seja, das reações emocionais na execução da atividade-meta. Por exemplo, não dominar um assunto do currículo escolar por falta de capacidade pode ocasionar o sentimento de vergonha no estudante, desmotivando-o e resultando no abandono do estudo.

De acordo com Barrera [9] é importante destacar que existe uma grande variação na maneira como diferentes pessoas interpretam uma mesma situação, de acordo com as

informações e com a forma pessoal de processá-las. Assim, a sequência pode representar vários cenários de uma mesma situação.

O cenário-exemplo utilizado por Weiner [37] foi: “Um jovem jogador de beisebol tem um desempenho muito ruim durante um jogo. Em vez de aparecer no próximo jogo, o rapaz fica em casa”. Neste cenário pode haver variações, como o menino praticar mais “batidas” após falhar (em vez de não ir ao jogo) ou praticar mais “batidas” após jogar bem (sucesso). O mesmo pode ser aplicado ao processo cognitivo: uma falha pode ser um motivador para mais dedicação aos estudos ou o sentimento de fracasso pode ocasionar a desistência.

Weiner [37] explica que uma sequência motivacional é iniciada por um resultado que os indivíduos interpretam como positivo (realização do objetivo) ou negativo (não-realização do objetivo). O afeto está diretamente relacionado com seu sucesso (felicidade) ou fracasso (frustração, tristeza) na realização da atividade-meta. No cenário do beisebol, o fraco desempenho do rapaz provoca sentimentos negativos. Dado a falha, a teoria tenta explicar o motivo, as causas que podem ter desencadeado essa saída.

As atribuições causais foram divididas em três dimensões ou propriedades: locus da causalidade, estabilidade e controlabilidade. As dimensões causais tem consequências psicológicas, estando relacionadas à expectativa e ao afeto [37].

O locus da causalidade influencia as consequências afetivas como a autoestima e orgulho, e pode ser classificada como interna ou externa. As atribuições internas (“motivos pessoais”) provocam maior impacto afetivo do que as atribuições externas (“fora do controle pessoal”). As atribuições “internas” são comportamentos mais consistentes que fazem parte da composição do indivíduo, como seu esforço, atenção, memória. As atribuições “externas” são as características ou ações externas ao indivíduo, sendo uma consequência ordinária de uma situação específica para o indivíduo, como sorte ou características da situação ou tarefa [37]. Por exemplo, o jovem jogador de beisebol falhou devido a uma causa interna e, portanto, ele deve estar passando por uma baixa autoestima.

A dimensão de estabilidade influencia as consequências relativas ao tempo como a expectativa de sucesso no futuro. Acontecimentos repetitivos podem ser considerados con-

sequências de causas estáveis, como, por exemplo, o grau de dificuldade de uma matéria, enquanto esforço, sorte, atenção, são causas instáveis. No cenário-exemplo, o menino antecipa sua falha (não indo ao próximo jogo) na medida em que seu fraco desempenho é percebido como uma causa estável (falta de aptidão). Ele também pode ter aumentado a expectativa de fracasso em outras atividades esportivas se a causa for percebida como global. Para Weiner [37], a estabilidade influencia aspectos temporais da expectativa, enquanto a globalidade influencia expectativas em diferentes situações cruzadas. Barrera [9] diz que a globalidade de determinadas causas está associada à possibilidade de generalização e de antecipação de resultados semelhantes em diferentes domínios.

A dimensão da controlabilidade, segundo Barrera [9], indica as causas percebidas como estando sob controle voluntário do sujeito e, portanto, passíveis de serem modificadas por ele, por exemplo, esforço e atenção, diferenciando-as de aquelas percebidas como estando fora desse controle, por exemplo, azar e dificuldade da tarefa.

A controlabilidade afeta as emoções sociais. Causas controláveis (esforço) do fracasso pessoal promovem sentimentos de culpa, enquanto as causas incontroláveis (capacidade intelectual) geram vergonha.

Os afetos resultantes das controlabilidade podem ser auto-direcionados, como o relaxamento e surpresa, e os afetos direcionados para terceiros como a raiva (falha controlável por outros), compaixão (falha incontrolável) e gratidão (causa controlável dos outros). No cenário-exemplo, é provável que o time esteja sentindo-se envergonhado e humilhado (mas não culpado), enquanto que o seu treinador ou as suas mães, sintam piedade ou pena deles (mas não raiva) [37].

Weiner [37] defende que essas atribuições causais estão relacionadas à diferentes emoções e comportamentos. A expectativa e o afeto são presumidos para determinar a ação. As ações podem ser descritas de acordo com a sua intensidade, sua latência, e assim por diante. No cenário de beisebol, o menino tem uma baixa expectativa de sucesso no futuro e está se sentindo triste, com baixa autoestima, com vergonha, e sem esperança. Estas condições promovem seu afastamento e comportamentos maléficos para a realização do objetivo desejado.

Dependendo da interpretação causal do indivíduo perante uma situação na qual está associada uma expectativa de mudança (ou não), sua motivação é influenciada positivamente (otimismo) ou negativamente (pessimismo), resultando em esforço ou desistência na tentativa de alcance da uma meta.

Barrera [9] salienta que apesar dos sucessos e fracassos sempre despertarem emoções agradáveis ou desagradáveis, ocasionam também, em função da interpretação de suas causas, uma variada gama de reações afetivas que afetarão o comportamento do sujeito. Os sucessos e fracassos atribuídos a fatores internos causam emoções mais intensas, pois afetam a autoestima da pessoa. Além disso, os afetos associados às atribuições causais influenciam a motivação, determinando a intensidade do investimento na tarefa, ou o grau de persistência frente às dificuldades encontradas.

A “Teoria da Atribuição Causal” considera o conhecimento a respeito do ambiente, de si mesmo e do outro um determinante motivacional mais importante do que o prazer. Nessa perspectiva, as emoções estão subordinadas às cognições. Porém, estas não se referem a um conhecimento necessariamente objetivo da realidade, mas à interpretação pessoal que cada um faz das causas dos acontecimentos. É essa interpretação que suscita emoções e a adoção de comportamentos [9].

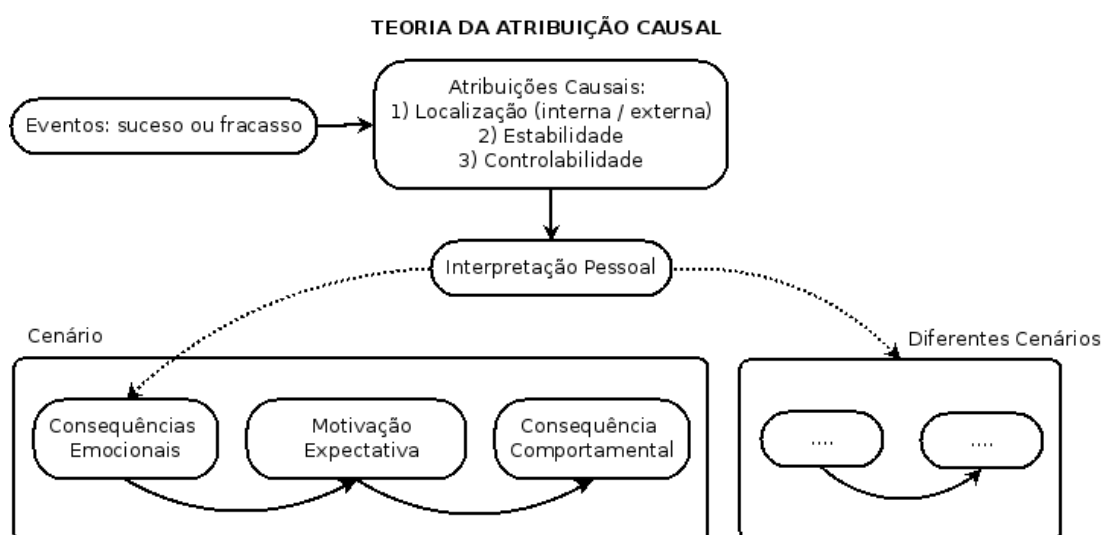


Figura 3.1: Diagrama da teoria da Atribuição Causal de Weiner [37]

A figura 3.1 mostra o esquema da teoria de Weiner [37], com as evidências de acertos e

erros como entrada para classificação de suas causas conforme as dimensões de localização, estabilidade e controlabilidade, sujeitas a interpretação pessoal resultando em diferentes cenários, com diferentes emoções, motivações e comportamentos.

Para caracterizar esta teoria os atributos estudados nesta dissertação, e detalhados no capítulo 4, foram os fatores habilidade, esforço, sorte e dificuldade.

3.2 *Bayesian Knowledge Tracing* e suas variantes

O *Bayesian Knowledge Tracing* (BKT) de Corbett e Anderson [19] é um modelo cognitivo da aquisição de um conhecimento. Ele monitora as mudanças de estado de conhecimento durante a prática de um problema (conhecido como modelo *Knowledge Tracing*).

Um problema é uma tarefa que envolve múltiplas etapas. Uma etapa é uma parte observável da solução para um problema, na qual diversas ações podem ser executadas. Cada ação é chamada de transação e pode ser uma tentativa correta ou incorreta, ou a requisição de uma dica. A transação é uma interação entre o estudante e o STI.

Cada vez que o aluno tem a oportunidade ou chance de aplicar uma componente de conhecimento, o tutor atualiza sua estimativa de domínio dessa componente, baseado na ação do estudante [19].

Definição. *Uma componente de conhecimento, também denominada como regra ou habilidade, é uma parte da informação que deve ser usada para resolver uma tarefa.*

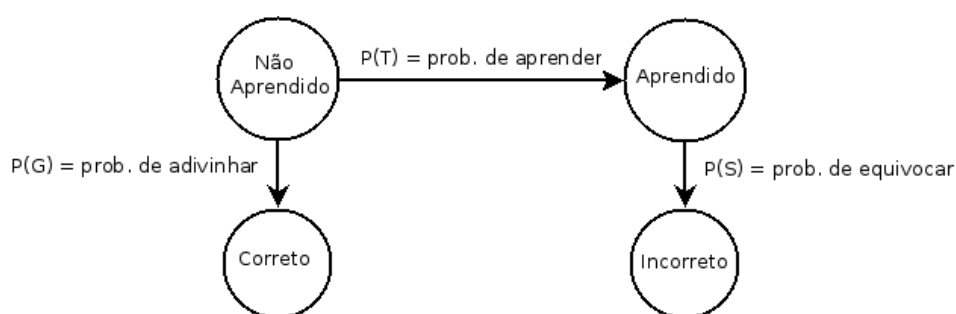


Figura 3.2: Modelo de aprendizagem

O *knowledge tracing* é composto por dois estados (ver figura 3.2) para cada regra: “aprendido” e “não aprendido”. A aprendizagem é uma medida probabilística, pois mesmo sendo considerado “aprendido”, o aluno pode equivocar-se e cometer um erro; se a regra estiver em um estado “não-aprendido”, o aluno pode adivinhar a resposta correta. A cada oportunidade de aplicar a regra, a estimativa de conhecimento é atualizada, de acordo com as ações, corretas ou não, do estudante.

Para o modelo BKT há as seguintes definições:

Definição. *O conhecimento no BKT é quando a regra encontra-se no estado “aprendido”. É representado por L .*

Definição. *A evidência é ação do aluno na resolução de uma tarefa como resposta correta ou incorreta, requisição de ajuda, etc. É representado por A .*

Definição. *O nível de conhecimento de uma regra no BKT é a probabilidade de estar no estado “aprendido” dadas as evidências, sendo representado pela probabilidade condicional $P(L|A)$.*

Definição. *A aprendizagem no BKT é uma constante que contém a probabilidade de transição do estado “não-aprendido” para o “aprendido”. É representado por $P(L_n|\neg L_{n-1}, Acao_n) = P(T)$ que mostra a probabilidade de saber na etapa n dado que não sabia na etapa $n - 1$ e a ação correta ou incorreta na etapa n .*

Definição. *A adivinhação (do inglês *Guess*) no BKT é a probabilidade de acertar uma regra estando em um estado de “não-aprendido”. A representação da probabilidade de adivinhar dado o conhecimento é $P(A = Correto|\neg L) = P(G)$.*

Definição. *O equívoco (do inglês *Slip*) no BKT é a probabilidade de errar uma regra estando em um estado de conhecimento. A representação da probabilidade de equivocar-se dado o conhecimento é $P(A = Incorreto|L) = P(S)$.*

Segundo Reye [31] o modelo BKT é equivalente a uma rede *Bayesiana* dinâmica de dois nós. Quando uma rede *Bayesiana* é usada para representar um modelo do estudante, é necessário ter uma forma, teoricamente plausível, de atualizá-lo. Essas atualizações são

baseadas em informações de duas fontes: do aluno, via suas entradas no sistema (por exemplo, pedidos de ajuda, respostas às perguntas, e as tentativas de exercícios); e do sistema, através de suas saídas (por exemplo, descrições e explicações dadas).

Em geral, quando há uma interação com o estudante, o sistema deve rever o estado de conhecimento deste aprendiz, aplicando o teorema de *Bayes* para ajustar a probabilidade de conhecimento da etapa anterior, e depois, atualizando a probabilidade de aprendizagem usando a ação da etapa atual.

O teorema de Bayes mostra a probabilidade de uma hipótese dada a observação de uma evidência e a probabilidade da evidência dada pela hipótese. A probabilidade condicional de Bayes é calculada da seguinte forma:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (3.1)$$

Sendo que na fórmula 3.1:

$P(A)$ é a probabilidade de A

$P(B)$ é a probabilidade de B

$P(A|B)$ é a probabilidade de A condicionado a B

$P(B|A)$ é a probabilidade de B condicionado a A

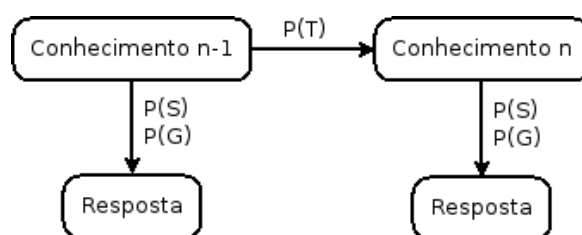


Figura 3.3: Modelo BKT

Observe na figura 3.3 que a direção da dependência é do estado atual para a evidência. Esta direção representa que o estado faz com que as evidências assumam valores particulares: a aprendizagem faz com que os alunos acertem. O processo de inferência, natural-

mente, vai na direção oposta. Esta distinção entre a direção de dependência e a direção de inferência é uma das principais vantagens das redes de *Bayesianas* [33].

Reye [31] defendeu uma abordagem em duas fases, representada na figura 3.4, para a atualização do modelo do estudante, sendo que em cada interação:

- Fase 1: a incorporação da evidência sobre o conhecimento do estudante;
- Fase 2: mudanças esperadas no conhecimento do aluno devido à tutoria.

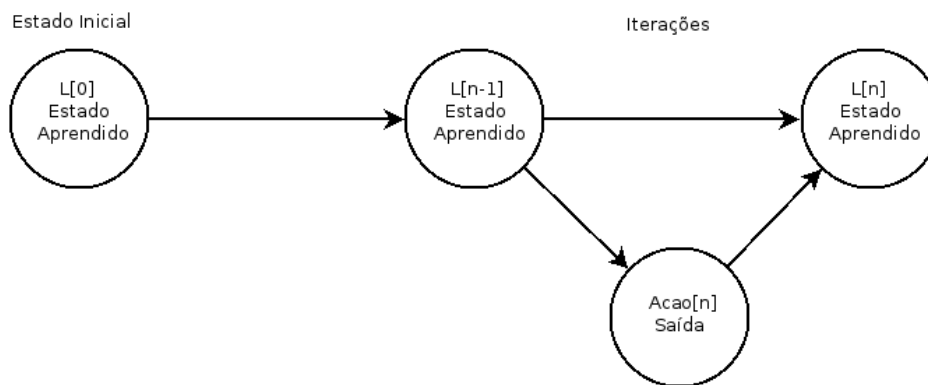


Figura 3.4: Atualização de duas fases. Fonte [31]

As equações utilizadas na fase 1 para o modelo do BKT são:

$$P(L_{n-1}|Acao_n = Correto) = \frac{P(Acao_n = Correto|L_{n-1}) * P(L_{n-1})}{P(Acao_n = Correto)} \quad (3.2)$$

$$= \frac{(1 - P(S)) * P(L_{n-1})}{P(L_{n-1}) * (1 - P(S)) + (1 - P(L_{n-1})) * P(G)}$$

$$P(L_{n-1}|Acao_n = Incorreto) = \frac{P(L_{n-1}) * P(S)}{P(L_{n-1}) * P(S) + (1 - P(L_{n-1})) * (1 - P(G))}$$

As equações utilizadas na fase 2 para o modelo do BKT são:

$$P(L_n|Acao_n) = P(L_{n-1}|Acao_n) + (1 - P(L_{n-1}|Acao_n)) * P(T) \quad (3.3)$$

Sendo que nas fórmulas 3.2 e 3.3:

$P(G)$ é a probabilidade de adivinhar a resposta

$P(S)$ é a probabilidade de equivocar-se na resposta

L_n representa o conhecimento na etapa n

L_{n-1} representa que o aluno já está no estado “aprendido”

$Acao_n$ é a ação (evidência) do aluno na oportunidade n

$P(T)$ é a probabilidade de aprendizagem

$P(L_{n-1}|Acao_n)$ é a probabilidade de aprendizado na etapa anterior condicionado a ação atual

$P(L_n|Acao_n)$ é a probabilidade de aprender condicionado a ação atual

$P(Acao_n|L_{n-1})$ é a probabilidade do resultado $Acao_n$ ocorrer quando o estudante já souber o tópico

$P(Acao_n|\neg L_{n-1})$ é a probabilidade do resultado $Acao_n$ ocorrer quando o estudante não souber o tópico

Para o modelo BKT tradicional, temos as seguintes características:

- $Acao_n$: assume apenas o valor correto e incorreto;
- $P(L_n|L_{n-1}, Acao_n) = 1$: não há esquecimento;
- Os parâmetros $P(L_0)$, $P(T)$, $P(G)$ e $P(S)$ são estimados empiricamente para cada componente de conhecimento;
- Não há probabilidades condicionais que ligam regras diferentes, ou seja, não há restrições prévias.

Dentre as variantes do modelo BKT há as que analisam o impacto da individualização na estimativa do conhecimento [29, 40], as que aprimoram os parâmetros $P(G)$ e $P(S)$ [2, 3], as que incorporam a condição da divisão da tarefa em sub-tarefas, conhecido como *scaffolding* [34], as que incorporam itens de dificuldades [30] e as que avaliam a aprendizagem com múltiplas componentes de conhecimento envolvidas [23, 39].

3.2.1 Abordagens para ajustar o modelo BKT

De acordo com Beck e Chang [12], o modelo tradicional do BKT é vulnerável ao problema estatístico da “identificabilidade”: um mesmo valor sobre o desempenho de um aluno pode ser previsto a partir de várias possíveis combinações de estimativas dos parâmetros do modelo. Por exemplo, múltiplas combinações de $P(G)$ e $P(L_0)$ podem gerar exatamente a mesma taxa de erro para $P(Acao_n = Correto)$.

A proposta de Beck e Chang [12] para resolver este problema foi avaliar as diferentes soluções para encontrar aquelas que são mais plausíveis, restringindo os parâmetros do modelo ao encontrar uma probabilidade prévia através de todas as componentes de conhecimento. Esta solução, porém, segundo Baker, Corbett e Alevén [2] introduz um novo problema, a “degeneração do modelo”, possibilitando a obtenção de parâmetros os quais induzem a um comportamento paradoxal. Este comportamento acontece quando o $P(G)$ ou o $P(S)$ são maiores que 0,5. Com o parâmetro $P(S)$ maior que 0,5, significa que um estudante que sabe uma componente de conhecimento tem maiores chances de errar do que de acertar a resposta e $P(G)$ maior que 0,5, significa que um estudante que não sabe uma componente de conhecimento tem maiores chances de acertar do que errar a resposta.

Baker, Corbett, Gowda, Wagner, MacLaren, Kauffman, Mitchell e Giguere em [3] compararam diferentes abordagens do modelo BKT. As variantes investigadas foram a versão padrão de quatro parâmetros [19], o *Contextual-Guess-e-Slip* [2] e o *Individual Difference Weights* [19].

A primeira variante foi a versão padrão de quatro parâmetros do BKT, na qual os parâmetros $P(L_0)$, $P(T)$, $P(G)$ e $P(S)$ recebem valores fixos para cada componente de conhecimento. Os valores de $P(S)$ e $P(G)$ foram limitados até 0,3 para evitar o problema da “degeneração do modelo”.

A segunda variante foi *Contextual-Guess-e-Slip* [2]. Nesta abordagem, cada componente de conhecimento tem um valor separado para $P(L_0)$ e $P(T)$. No entanto, $P(G)$ e $P(S)$ não são mais estimadas para cada componente de conhecimento; em vez disso, eles são calculados cada vez que um aluno tenta responder uma nova etapa, com base em um

modelo aprendido, treinado com 64 e aplicado com 253 componentes, através das características das ações em um contexto. Por exemplo, as respostas mais longas e pedidos de ajuda são menos propensos de serem deslizes.

Como o modelo *Contextual-Guess-and-Slip* substitui dois parâmetros em cada componente de conhecimento, tem um menor número total de parâmetros.

Na versão *Individual Difference Weights*, as estimativas de desempenho e aprendizagem variam entre as componentes. E devido as diferenças entre os estudantes, essas estimativas podem ser incorretas para os alunos abaixo ou acima da média. Segundo Corbett et al. [19] os estudantes acima da média cometem poucos erros e recebem mais exercícios de remediação do que o necessário. Os estudantes abaixo da média, apesar de receberem mais exercícios de remediação do que os alunos acima da média, recebem menos do que eles precisam, resultando desempenho pior do que o esperado nos testes. Por isso, foi incorporado ao modelo pesos individuais distintos para cada parâmetro $P(L_0)$, $P(T)$, $P(G)$ e $P(S)$.

Esta abordagem, que trabalha com quatro parâmetros para cada componente de conhecimento e quatro parâmetros para cada aluno, tem substancialmente mais parâmetros do que as outras abordagens.

Baker et al. [3] concluíram que a abordagem *Contextual-Guess-and-Slip* faz uma boa previsão de desempenho dentro de um STI, sendo comparável ou ligeiramente superior a abordagem de quatro parâmetros e do *Individual Difference Weights*. Embora, tenha falhado na previsão de desempenho pós-teste. Um dos possíveis motivos para esta falha, é que o *Contextual-Guess-and-Slip* faz o sobreajuste (modelo estatístico ajusta-se em demasiado ao conjunto de dados de teste) do desempenho do estudante.

A abordagem *Individual Difference Weights* também não foi mais preciso na previsão do pós-teste do que a abordagem padrão de quatro parâmetros. Em pesquisa anterior [19], o modelo foi vantajoso nos casos de elevado nível de $P(L_n)$.

Apesar do fracasso do *Contextual-Guess-and-Slip* para prever o desempenho no pós-teste, a estimativa *Contextual Slip* parece ser uma valiosa adição para a previsão do conhecimento e desempenho. Isto é um indicativo que algum aspecto da aprendizagem não

foi capturado pelo BKT. Contudo, $P(S)$ pode ter vários significados, incluindo indicação de conhecimento superficial ou descuido durante o uso do STI.

Baker et al. [3] destacaram o potencial do *Contextual Slip* na predição do desempenho do estudante pós-teste, porque quanto melhor esta predição for, provavelmente, melhor será a adaptação do STI.

As investigações de Baker et al. [3] sugeriram que as abordagens de força bruta para encontrar as melhores estimativas dos parâmetros - todas as combinações de parâmetros possíveis de valores (por exemplo, 0,01 0,01 0,01 0,01, 0,01 0,01 0,01 0,02, 0,01 0,01 0,01 0,03 ... 0,01 0,01 0,02 0,01 ... 0,99 0,99 0,3 0,1) - resultam em um melhor ajuste do que algoritmos anteriormente utilizados, como *Expectation Maximization* [12], *Conjugate Gradient Search* [19], e *Generalized Reduced Gradient Search* [2]. Essas mesmas investigações revelaram que a força bruta é computacionalmente tratável para o conjunto de dados e conjunto de parâmetros do BKT, uma vez que o tempo aumenta linearmente com o número de ações dos alunos, mas é constante para o número de componentes de conhecimento (uma vez que apenas uma componente é aplicada a cada ação do aluno, o número de operações matemáticas é idêntico, não importando quantas componentes estão presentes). Para a versão de quatro parâmetros [19], por exemplo, o número total a ser estimado é o número de regras cognitivas vezes quatro.

3.2.2 Detectando Aprendizagem Momento a Momento

Para aprimorar o modelo do estudante, Baker, Goldstein e Heffernan [4] propuseram um modelo que prevê a probabilidade de aprendizagem de uma componente de conhecimento em uma etapa específica na resolução de um problema. Baker et al. [4] denominaram essa probabilidade de $P(J)$ (*Just Learned*), calculada conforme a fórmula 3.4.

Definição. *O momento da aprendizagem é definido na etapa em que ocorre a passagem do estado “não-aprendido” para o estado “aprendido” dada as ações nas etapas $n + 1$ e $n + 2$.*

Especificamente, definiram a aprendizagem da componente de conhecimento entre o

instante depois que o estudante entrou com sua primeira resposta na etapa n (n -ésima oportunidade de exercitar um componente de conhecimento específico) e o instante que o estudante entra com sua primeira resposta na etapa $n + 1$.

O processo consiste de uma etapa de rotulagem e uma de treinamento. Os rótulos são obtidos pela previsão do conhecimento (modelo BKT) no passo n condicionado aos dados de correções futuras ($n + 1$, $n + 2$). O treinamento é feito com um conjunto de características que descrevem cada ação do estudante para a prever o dado rotulado em tempo de execução. A fórmula resultante foi:

$$P(J) = P(\neg L_n \wedge T | A_{+1+2}) \quad (3.4)$$

Sendo que na fórmula 3.4:

$P(J)$ é a probabilidade da aprendizagem em um momento específico

$P(\neg L_n \wedge T | A_{+1+2})$ é a probabilidade não saber porém aprender dada as ações nas etapas $n + 1$ e $n + 2$

A probabilidade das ações $P(A_{+1+2})$ nos tempos $n + 1$ e $n + 2$ é a probabilidade incondicional ou marginal obtida da probabilidade condicional e calculada conforme a fórmula 3.5.

$$\begin{aligned} P(A_{+1+2}) &= P(A_{+1+2} | L_n) * P(L_n) + \\ &P(A_{+1+2} | \neg L_n \wedge T) * P(\neg L_n) * P(T) + \\ &P(A_{+1+2} | \neg L_n \wedge \neg T) * P(\neg L_n) * P(\neg T) \end{aligned} \quad (3.5)$$

$$P(L_n | A_{+1+2}) = \frac{P(A_{+1+2} | L_n) * P(L_n)}{P(A_{+1+2})} \quad (3.6)$$

$$P(\neg L_n \wedge T | A_{+1+2}) = \frac{P(A_{+1+2} | \neg L_n \wedge T) * P(\neg L_n) * P(T)}{P(A_{+1+2})} \quad (3.7)$$

$$P(\neg L_n \wedge \neg T | A_{+1+2}) = \frac{P(A_{+1+2} | \neg L_n \wedge \neg T) * P(\neg L_n) * P(\neg T)}{P(A_{+1+2})} \quad (3.8)$$

As fórmulas 3.6, 3.7 e 3.8 são calculadas pela regra de *Bayes*. Sendo que:

$P(L_n | A_{+1+2})$ é a probabilidade conhecimento dada as ações nas etapas $n + 1$ e $n + 2$

$P(A_{+1+2} | L_n)$ é a probabilidade das ações nas etapas $n + 1$ e $n + 2$ dado o conhecimento na etapa n

$P(\neg L_n \wedge T | A_{+1+2})$ é a probabilidade de não saber porém aprender dada as ações nas etapas $n + 1$ e $n + 2$

$P(A_{+1+2} | \neg L_n \wedge T)$ é a probabilidade das ações nas etapas $n + 1$ e $n + 2$ dado o não conhecimento e aprendizagem na etapa n

$P(\neg L_n \wedge \neg T | A_{+1+2})$ é a probabilidade de não saber e não aprender dada as ações nas etapas $n + 1$ e $n + 2$

$P(A_{+1+2} | \neg L_n \wedge \neg T)$ é a probabilidade das ações nas etapas $n + 1$ e $n + 2$ dado o não conhecimento e não aprendizagem na etapa n

As probabilidades condicionadas das ações, corretas C ou incorretas $\neg C$, dependentes do conhecimento são calculadas pelas probabilidades $P(S)$, $P(G)$ e $P(T)$, e são deduzidas por Baker et al.[4] conforme apresentado abaixo:

- Quando há conhecimento, o acerto é não-equívoco e erro é equívoco:

$$P(A_{+1+2} = C, C | L) = P(\neg S) * P(\neg S)$$

$$P(A_{+1+2} = C, \neg C | L) = P(\neg S) * P(S)$$

$$P(A_{+1+2} = \neg C, C | L) = P(S) * P(\neg S)$$

$$P(A_{+1+2} = \neg C, \neg C | L) = P(S) * P(S)$$

- Quando não há conhecimento mas há aprendizagem, o acerto é não equívoco e erro é

equivoco:

$$\begin{aligned}
P(A_{+1+2} = C, C|\neg L \wedge T) &= P(\neg S) * P(\neg S) \\
P(A_{+1+2} = C, \neg C|\neg L \wedge T) &= P(\neg S) * P(S) \\
P(A_{+1+2} = \neg C, C|\neg L \wedge T) &= P(S) * P(\neg S) \\
P(A_{+1+2} = \neg C, \neg C|\neg L \wedge T) &= P(S) * P(S)
\end{aligned}$$

- Quando não há conhecimento e não ocorre aprendizagem na etapa n , o acerto é adivinhação e erro é não adivinhação, podendo aprender ou não na etapa $n + 1$:

$$\begin{aligned}
P(A_{+1+2} = C, C|\neg L \wedge \neg T) &= P(G) * P(T) * P(\neg S) + P(G) * P(\neg T) * P(G) \\
P(A_{+1+2} = C, \neg C|\neg L \wedge \neg T) &= P(G) * P(T) * P(S) + P(G) * P(\neg T) * P(\neg G) \\
P(A_{+1+2} = \neg C, C|\neg L \wedge \neg T) &= P(\neg G) * P(T) * P(\neg S) + P(\neg G) * P(\neg T) * P(G) \\
P(A_{+1+2} = \neg C, \neg C|\neg L \wedge \neg T) &= P(\neg G) * P(T) * P(S) + P(\neg G) * P(\neg T) * P(\neg G)
\end{aligned}$$

- Portanto, a probabilidade das ações quando não há conhecimento na etapa n considera tanto as situações em que ocorre a aprendizagem, quanto as que não ocorre:

$$\begin{aligned}
P(A_{+1+2} = C, C|\neg L) &= P(A_{+1+2} = C, C|\neg L \wedge T) + P(A_{+1+2} = C, C|\neg L \wedge \neg T) \\
P(A_{+1+2} = C, \neg C|\neg L) &= P(A_{+1+2} = C, \neg C|\neg L \wedge T) + P(A_{+1+2} = C, \neg C|\neg L \wedge \neg T) \\
P(A_{+1+2} = \neg C, C|\neg L) &= P(A_{+1+2} = \neg C, C|\neg L \wedge T) + P(A_{+1+2} = \neg C, C|\neg L \wedge \neg T) \\
P(A_{+1+2} = \neg C, \neg C|\neg L) &= P(A_{+1+2} = \neg C, \neg C|\neg L \wedge T) + P(A_{+1+2} = \neg C, \neg C|\neg L \wedge \neg T)
\end{aligned}$$

A probabilidade $P(A_{+1+2} = C, C|L) = P(\neg S) * P(\neg S)$ pode ser interpretada como a probabilidade das ações futuras serem corretas dado que há conhecimento, é igual a probabilidade de não haver equívoco nas duas oportunidades seguintes ($n + 1$ e $n + 2$). As outras probabilidade condicionadas são interpretadas de forma semelhante.

Após a rotulagem com as fórmulas apresentadas acima sobre uma base de dados, foi efetuada a etapa de treinamento para gerar o modelo de previsão. O treinamento foi realizado através de um conjunto de características para descrever cada ação, com

posterior classificação por regressão linear. Alguns dos atributos de entrada, definidos no trabalho de Baker et al. [4], para o algoritmo de aprendizagem de máquina foram:

- Porcentagem de erros;
- Número total de erros;
- Número de erros nos últimos 5 problemas;
- Número de erros nos últimos 8 problemas;
- Duração de resolução problema;
- Duração de resolução dos últimos 3 problemas;
- Duração de resolução das últimas 5 ações;
- Ação é uma requisição de ajuda;
- Probabilidade de aprender depois da ação corrente ($P(L_n)$);
- Probabilidade de conhecimento anterior ($P(L_{n-1})$);
- E outros (detalhes em [4]);

A base de dados utilizada era composta por 581.785 transações efetuadas em 171.987 etapas de problemas, com 253 componentes de conhecimento.

A correlação foi de 0,446 para a validação cruzada *6-fold* entre grupos de estudantes (isto é, detectores foram treinados com 5 grupos de estudantes e testado com um sexto grupo de estudantes). Esta validação, segundo os autores, aumenta a confiança de que os detectores serão mais precisos para novos grupos de estudantes.

Segundo Baker et al. [4], embora o grau de correlação alcançado tenha sido aceitável, o modelo tende a subestimar os valores de $P(J)$, particularmente aqueles que eram relativamente elevados nos rótulos originais (por exemplo $> 0,02$). As previsões de $P(J)$ para o treinamento com valores altos permaneceu maior do que as previsões de $P(J)$ para o treinamento com valores baixos. No entanto, a previsão dos valores de $P(J)$ no

treinamento com valores altos foram menores, em termos absolutos, do que os valores dos rótulos originais. Este problema poderia ser resolvido mediante uma ponderação dos altos valores mais fortemente durante a modelagem, embora esta abordagem provavelmente reduziria correlação geral.

Outro problema mencionado por Baker et al. [4] foi a incorporação de $P(L_{n-1})$ e $P(L_n)$ na construção da predição de $P(\neg L_n)$, porém argumentaram que esses termos são confiáveis na predição da aprendizagem pois $P(L_n)$ já encontra-se registrado no modelo. Para validar o modelo, este foi reajustado sem $P(L_{n-1})$ e $P(L_n)$, e o resultado foi uma correlação menor do que o modelo completo, sendo a diferença estatisticamente significativa segundo o teste pelo método de *Stouffer* ($Z = 116,51, p < 0,0001$).

3.3 Considerações finais

A teoria da “Atribuição Causal” de Weiner [37] defende que as atribuições causais dos eventos interferem no processo cognitivo uma vez que atuam na motivação. A interpretação pessoal é destacada como fonte da variação das emoções geradas em uma mesma situação.

Como a tarefa de avaliar os sentimentos, comportamentos e conhecimentos humanos são imprecisos, estes são referenciados de forma probabilística.

O modelo *Bayesian Knowledge Tracing* (BKT) de Corbett et al. [19] utiliza a probabilidade de *Bayes* para estimar o conhecimento do estudante, monitorando as mudanças de estado de conhecimento durante a prática de um exercício. Cada vez que o aluno tem a oportunidade de aplicar um conhecimento, o tutor atualiza sua estimativa de conhecimento (domínio) da regra, baseado na ação do estudante [19].

Deste modelo, surgiram variantes que incorporaram diferentes elementos como o impacto da individualização na estimativa [29, 40], as que incorporam a condição de *scaffold* [34], itens de dificuldades [30], além da avaliação do conhecimento com múltiplas componentes de conhecimento envolvidas [23, 39]. Além de outras versões para o ajuste dos parâmetros do modelo como os que aprimoram os parâmetros $P(G)$ e $P(S)$ [2, 3] e o *Individual Difference Weights* que utiliza um peso para cada estudante, calculado para

cada parâmetro, que será o mesmo para todas as componentes de conhecimento.

Uma extensão do BKT foi o trabalho de Baker et al. [4] que explorou as ações futuras do estudante para determinar o nível e o momento da aprendizagem. A ideia é a previsão do futuro e categorização do presente através dados do passado, ou seja, utilizar os históricos das ações dos estudantes com um sistema tutorial computacional para determinar seu padrão comportamental. Utiliza informações futuras relativas a determinada ação para o cálculo da probabilidade do conhecimento do estudante condicionado a duas respostas futuras. Desta forma, tenta estimar o momento exato em que ocorreu a aprendizagem.

A combinação da teoria psicológica com modelos de inferências formou a base teórica para a criação do modelo proposto nesta dissertação, explicando a importância da análise das causas dos eventos e seu peso no processo cognitivo, e obtendo as interpretações numéricas através de um explorado modelo probabilístico.

CAPÍTULO 4

MODELO DE INFERÊNCIA DOS PARÂMETROS CAUSAIS

Os capítulos 2 e 3 apresentaram pesquisas com investigações a respeito dos fatores internos do estudante que interferem no processo cognitivo e alternativas para a inferência do conhecimento.

Inspirado nos modelos e resultados dessas pesquisas, esta dissertação acompanha a tendência da análise dos dados para obter informações referentes ao comportamento e estados dos estudantes. O principal objetivo é encontrar padrões que expliquem as causas de um determinado comportamento.

Foi desenvolvido um modelo para a inferência dos parâmetros das causas dos sucessos e fracassos em atividades cognitivas, como a resolução de exercícios em um sistema tutor para a aprendizagem de uma componente de conhecimento.

As principais utilizações dos parâmetros inferidos são pré-ação, isto é, na previsão da ação atual através dos dados do passado, e pós-ação, ou seja, após a inserção da resposta, para posterior plano de ação. Essas informações podem ser utilizadas, por exemplo, para aprimorar atividades associadas ao diagnóstico dos fatores internos do estudante como emoções e motivações que colaboram para a elaboração de um esquema inteligente de ensino.

Esta pesquisa concentrou-se na elaboração do modelo de inferência dos parâmetros causais e na sua avaliação pré-ação, simulando seu impacto através de um modelo determinístico para relacionar as causas passadas com a próxima ação do estudante. O impacto no planejamento do ensino será explorado em projeto futuro.

A figura 4.1 mostra a arquitetura do fluxo da determinação das causas dos eventos do processo cognitivo e a possibilidade de diferentes saídas conforme interpretação pessoal. O fluxo inicia-se com a ocorrência dos eventos refletidos nos dados das ações dos estudantes com um STI. Esses dados servem de entrada para o cálculo da probabilidade de

conhecimento com o modelo *Bayesian Knowledge Tracing* (BKT) de Corbett et al. [19], que é aprimorado com as ações futuras, segundo modelo desenvolvido por Baker et al. [4], que por regressão linear é classificado e utilizado para prever as probabilidades das suas possíveis causas. Através dessas causas é possível ocorrer várias interpretações pessoais resultando em diferentes consequências afetivas, motivacionais e comportamentais. Essas consequências serão alvos de um estudo, mais detalhado, planejado para um trabalho futuro.

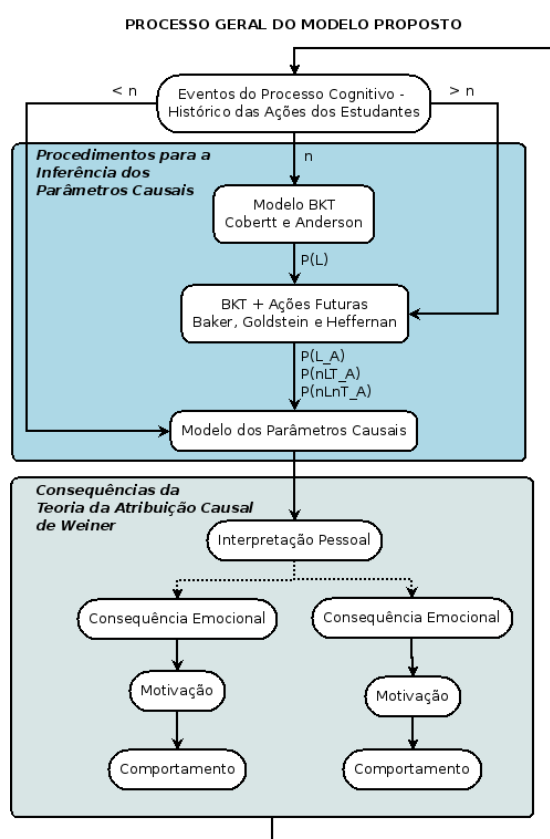


Figura 4.1: Diagrama do processo geral do modelo proposto para o cálculo dos parâmetros causais.

Como as inferências dos parâmetros causais são baseadas no modelo BKT, a monitoração das mudanças de estado de conhecimento durante a prática de um problema (conhecido como modelo *Knowledge Tracing*) também compõe a atualização dos valores das causas através da incorporação da evidência e do conhecimento inferido.

Destaca-se que o modelo proposto faz a especialização dos parâmetros do modelo BKT em novos parâmetros visando aprofundar o diagnóstico dos fatores influenciadores

do processo cognitivo dos estudantes.

O modelo está ilustrado na figura 4.2 e, mais especificamente, é apresentado as fases de atualização das probabilidades dos parâmetros causais.

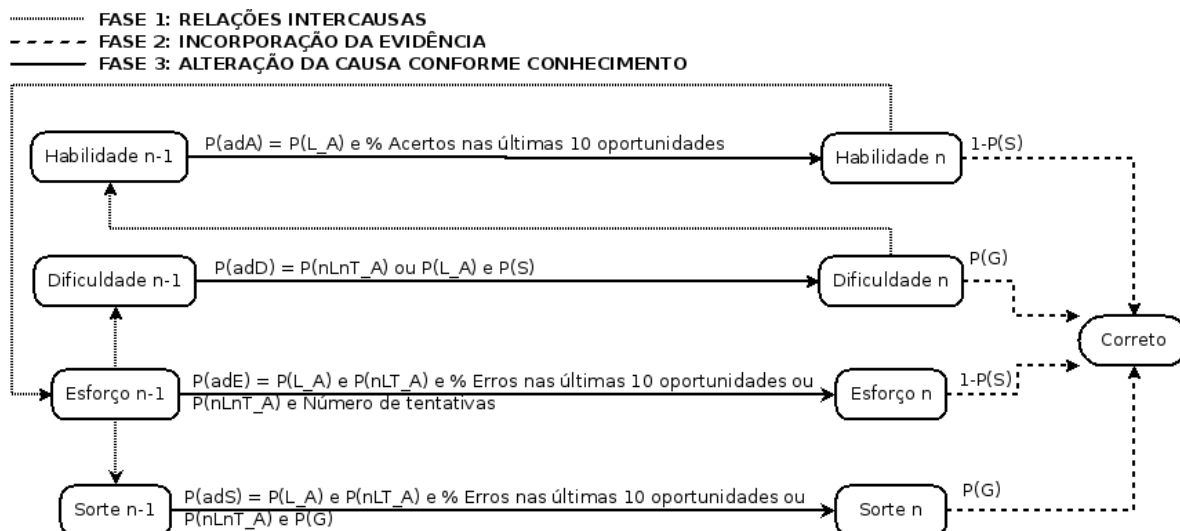


Figura 4.2: Fases da atualização dos parâmetros causais

A fórmula geral utilizada é:

$$P(Causa_n) = P(Causa_{n-1}|A_n) + P(\neg Causa_{n-1}|A_n) * Indícios \quad (4.1)$$

Sendo que na fórmula 4.1:

$P(Causa_n)$ é a probabilidade de determinada causa (habilidade ou dificuldade ou sorte ou esforço) na oportunidade n de interação do estudante com a componente de conhecimento

$P(Causa_{n-1}|A_n)$ é a probabilidade de determinada causa (habilidade ou dificuldade ou sorte ou esforço) na oportunidade $n-1$ de interação do estudante com a componente de conhecimento dado a ação atual (correta ou incorreta)

Indícios possui valores que indicam que determinada ação é consequência de determinada causa

Pela figura 4.2 e pela fórmula 4.1, observa-se que o valor da $P(Causa_{n-1})$ é reavaliada duas vezes: uma com a relação intercausas e outra com a incorporação da evidência.

As seções a seguir apresentam detalhadamente as etapas de desenvolvimento que são:

1. Definição de quais percepções causais serão inferidas;
2. Cálculo do nível de conhecimento;
3. Cálculo e modelagem dos ajustes do nível de conhecimento;
4. Cálculo da inferência dos parâmetros causais baseado no nível de conhecimento;
5. Simulação e análise do impacto dos parâmetros causais na determinação das respostas dos estudante.

4.1 Definição dos parâmetros causais

A aprendizagem é uma tarefa imprecisa uma vez que o processo cognitivo envolve fatores internos dos seres humanos. Para Bandura [5], pessoas não são nem agentes autônomos nem simplesmente transportadores mecânicos das influências ambientais. O ser humano representa um agente passivo, ao ser influenciado pelos fatores externos, e ativo, ao contribuir com sua interpretação pessoal na causa da sua própria motivação.

De acordo com Barrera [9] é importante destacar que existe uma grande variação na maneira como diferentes pessoas interpretam uma mesma situação, de acordo com as informações e com a forma pessoal de processá-las. Assim, a sequência pode representar vários cenários de uma mesma situação. Heider [22, 37] afirma que o resultado de uma ação depende das condições internas da pessoa e internas do ambiente.

Apesar desta dissertação não se aprofundar no diagnóstico dos estados afetivos gerados pelas causas, é importante mensurar probabilidades de causas que se encaixam em um estudo psicológico para um futuro mapeamento dos consequentes estados afetivos. Portanto, os parâmetros causais explorados foram baseados na “Teoria atribucional da obtenção da motivação e emoção” ou simplesmente “Teoria da Atribuição Causal” de Weiner [37]. Segundo esta teoria, detalhada no capítulo 3 seção 3.1, as causas dos eventos, de

sucesso ou fracasso, estão relacionadas à emoção; e a emoção está relacionada à motivação; e a motivação interfere nas sequências da atividade cognitiva. Por exemplo, não dominar um assunto do currículo escolar por falta de capacidade pode ocasionar o sentimento de vergonha no estudante, desmotivando-o e resultando no abandono do estudo.

A motivação é representada como uma sequência histórica, resultado do vínculo da emoção na execução da atividade-meta. Campbell e Pritchard [15, 35] sugerem que a motivação relaciona-se com um conjunto de variáveis que explicam a direção, amplitude e persistência do comportamento individual, mantendo constante os efeitos da aptidão, da habilidade e da compreensão da tarefa, e as restrições de operação no ambiente.

A interpretação pessoal das causas é destacada como fonte da variação das emoções geradas para uma mesma situação. Uma sequência motivacional é iniciada por um resultado que os indivíduos interpretam como positivo (realização do objetivo) ou negativo (não-realização do objetivo). O afeto está diretamente relacionado com seu sucesso (felicidade) ou fracasso (frustração, tristeza) na realização da atividade-meta.

Assim, Weiner [37] defendeu que as atribuições causais interferem no processo cognitivo e no comportamento do estudante pois atuam na motivação.

A teoria da “Atribuição Causal” divide as percepções causais em três dimensões ou propriedades (detalhes no capítulo 3):

- Locus (localização):
 - Interno: por exemplo, capacidade, esforço, habilidade, dificuldade individual;
 - Externo: por exemplo, características das tarefas, sorte.
- Estabilidade (muda com o tempo):
 - Estável: por exemplo, dificuldade das questões;
 - Instável: por exemplo, expectativas, habilidade, dificuldade individual, sorte.
- Controlabilidade (depende unicamente do estudante):
 - Controlável: por exemplo, esforço, atenção;

– Incontrolável: por exemplo, capacidade intelectual, habilidade, sorte.

Uma causa, dependendo de suas propriedades, afeta de forma diferente o estudante. Segundo Weiner [37], as internas provocam maior impacto afetivo do que as atribuições externas. E a dimensão de estabilidade influencia as causas relativa ao tempo, como a expectativa de sucesso no futuro. E a controlabilidade afeta as emoções sociais. Causas controláveis do fracasso pessoal promovem sentimentos de culpa, enquanto as causas incontroláveis geram vergonha.

A relação das causas com a interpretação pessoal e seus consequentes resultados apresentados por Weiner [37] foram esquematizadas figura 3.1.

As causas exploradas nesta dissertação são a dificuldade, a habilidade, o esforço e a sorte. Essas causas foram selecionadas pois suas propriedades foram abordadas no trabalho de Weiner e devido ao fato de que a resolução de problemas está intimamente ligada à dificuldade ou habilidade dos estudantes no domínio associado a estes problemas, que para atingir a habilidade é preciso esforço e que para acertar uma resposta pode ser influenciada pela sorte. Destaca-se a “resolução de problemas” como a atividade cognitiva, presentes em muitos STI, e é a fonte das informações das bases de dados utilizadas neste trabalho.

No contexto desta dissertação essas causas são definidas como:

Definição. *Dificuldade é a associação da dificuldade em executar determinada componente de conhecimento com o fator interno “conhecimento”. Ou seja, há uma transição da dificuldade da tarefa (fator externo e estável), que é estipulado de forma geral considerando as ações de todos os estudantes, para a dificuldade individual (fator interno e instável), que é passível de alterações à medida que o nível de conhecimento for consolidando-se ao longo do tempo.*

Definição. *Habilidade (fator interno e instável) é ser hábil na aplicação de um conhecimento teórico.*

Definição. *Esforço (fator interno e controlável) é o trabalho, tempo, dedicação e concentração aplicados para a consolidação de um conhecimento e adquirir habilidade.*

Definição. *Sorte (fator externo e incontrolável) é considerado como resultado do “destino” cujas causas não podem ser racionalmente explicados. A sorte também deve ser analisada como fonte causadora das ações dos estudante, podendo ser um fator decisivo para acertar a resolução de uma tarefa.*

4.2 Cálculo dos parâmetros causais

Os parâmetros causais habilidade, esforço, dificuldade e sorte foram modelados para usar os históricos de ações do estudante com um STI.

As causas são elementos que não são mensuráveis com exatidão. Assim, as equações dos parâmetros causais utilizam a probabilidade do conhecimento do estudante sobre determinado componente. Esses componentes são partes de um conhecimento sobre um tema, referenciados por Corbett et al. [19] como “componentes de conhecimento”.

A probabilidade de conhecimento é uma medida interessante pois é alterada por fatores internos dos indivíduos, podendo ser calculado através de dados observáveis como a resposta correta, e é uma possível evidência resultante das emoções e motivações influenciadas pelas causas.

4.2.1 Modelo *Bayesian Knowledge Tracing* (BKT) de Corbett e Anderson

A probabilidade para cada componente de conhecimento foi calculada pelo modelo *Bayesian Knowledge Tracing* (BKT) de Corbett et al. [19]. Os princípios do modelo BKT e as definições de componente de conhecimento, conhecimento, evidência, nível de conhecimento, aprendizagem, adivinhação e equívoco foram apresentados no capítulo 3, seção 3.2.

O BKT atualiza o conhecimento através da habilidade demonstrada na resolução de uma tarefa. O conhecimento é a compreensão técnica ou teórica de um assunto específico e a habilidade é capacidade prática em aplicar este conhecimento.

A figura 3.2 mostra os estados do BKT: do estado “não aprendido” há a probabili-

dade de adivinhar a resposta, respondendo-a corretamente, ou de aprender. No estado “aprendido” o estudante também pode errar uma questão ao equivocar-se na resolução da tarefa.

A atualização do modelo do estudante com o BKT é feita em duas fases (ilustrado na figura 3.4). Na primeira fase ocorre a incorporação da evidência sobre o conhecimento do estudante e na segunda ocorre as mudanças esperadas no conhecimento do aluno devido a tutoria. As fórmulas do BKT para a probabilidade condicionada são:

$$P(L_{n-1}|Acao_n = C) = \frac{P(L_{n-1}) * (1 - P(S))}{P(L_{n-1}) * (1 - P(S)) + (1 - P(L_{n-1})) * P(G)} \quad (4.2)$$

$$P(L_{n-1}|Acao_n = \neg C) = \frac{P(L_{n-1}) * P(S)}{P(L_{n-1}) * P(S) + (1 - P(L_{n-1})) * (1 - P(G))} \quad (4.3)$$

$$P(L_n|A_n) = P(L_{n-1}|A_n) + (1 - P(L_{n-1}|A_n)) * P(T) \quad (4.4)$$

Sendo que nas fórmulas 4.2, 4.3 e 4.4:

$P(G)$ é a probabilidade de adivinhar a resposta

$P(S)$ é a probabilidade de equivocar-se na resposta

$P(T)$ é a probabilidade de aprendizagem

L_n representa o conhecimento na etapa n

L_{n-1} representa que o aluno já está no estado “aprendido”

$Acao_n$ é a ação (evidência) do aluno na oportunidade n

$P(L_{n-1}|Acao_n)$ é a probabilidade de aprendizado na etapa anterior condicionado a ação atual

$P(L_n|Acao_n)$ é a probabilidade de aprender condicionado a ação atual

Observe que $P(L_n|A_n)$ é a soma do conhecimento apresentado anteriormente ajustado à ação atual com a probabilidade do não-conhecimento anterior porém com a ocorrência da aprendizagem.

4.2.2 Modelo ajustado de Baker, Goldstein e Heffernan

Mais precisamente, o modelo apresentado nesta dissertação utiliza uma extensão do modelo BKT que aprimora a detecção da aprendizagem do discente por meio da probabilidade condicionada aos eventos futuros e um algoritmo de classificação. Este modelo aprimorado foi proposto por Baker et al. [4] e apresentado no capítulo 3 seção 3.2.2.

$$P(L_n|A_{+1+2}) = \frac{P(A_{+1+2}|L_n) * P(L_n)}{P(A_{+1+2})} \quad (4.5)$$

$$P(\neg L_n \wedge T|A_{+1+2}) = \frac{P(A_{+1+2}|\neg L_n \wedge T) * P(\neg L_n) * P(T)}{P(A_{+1+2})} \quad (4.6)$$

$$P(\neg L_n \wedge \neg T|A_{+1+2}) = \frac{P(A_{+1+2}|\neg L_n \wedge \neg T) * P(\neg L_n) * P(\neg T)}{P(A_{+1+2})} \quad (4.7)$$

As fórmulas 4.5, 4.6 e 4.7 são calculadas pela regra de *Bayes*. Sendo que:

$P(L_n|A_{+1+2})$ é a probabilidade conhecimento dada as ações nas etapas $n + 1$ e $n + 2$

$P(A_{+1+2}|L_n)$ é a probabilidade das ações nas etapas $n + 1$ e $n + 2$ dado o conhecimento na etapa n

$P(\neg L_n \wedge T|A_{+1+2})$ é a probabilidade de não saber porém aprender dada as ações nas etapas $n + 1$ e $n + 2$

$P(A_{+1+2}|\neg L_n \wedge T)$ é a probabilidade das ações nas etapas $n + 1$ e $n + 2$ dado o não conhecimento e aprendizagem na etapa n

$P(\neg L_n \wedge \neg T|A_{+1+2})$ é a probabilidade de não saber e não aprender dada as ações nas etapas $n + 1$ e $n + 2$

$P(A_{n+1+2}|\neg L_n \wedge \neg T)$ é a probabilidade das ações nas etapas $n + 1$ e $n + 2$ dado o não conhecimento e não aprendizagem na etapa n

As probabilidades condicionadas das ações, corretas C ou incorretas $\neg C$, dependentes do conhecimento são calculadas pelas probabilidades $P(S)$, $P(G)$ e $P(T)$, e são deduzidas por Baker et al.[4] conforme apresentado no capítulo 3, seção 3.2, subseção 3.2.2.

Uma característica importante é a adequação das predições futuras através dos dados passados como porcentagens de acertos, erros, dicas, tempo de duração, além do nível de conhecimento. Portanto, para a aplicação em tempo de execução desta ideia, Baker et al. [4] fez uma classificação por regressão linear sem utilizar atributos com ações futuras, somente com eventos já observados.

Nesta dissertação os valores aprimorados resultantes da classificação são referenciados como conhecimento “ajustado” e representados como $P(L_A)$ (conhecimento), $P(nLT_A)$ (não-conhecimento e aprendizagem) e $P(nLnT_A)$ (não-conhecimento e não-aprendizagem).

4.2.3 Cálculo da inferência dos parâmetros causais

Após a definição de quais percepções causais seriam modeladas (dificuldade, habilidade, esforço e sorte - “Teoria da Atribuição Causal” de Weiner [37]), do cálculo do nível de conhecimento com o modelo BKT (Corbett et al. [19]) e a geração do modelo treinado do nível de conhecimento “ajustado” (Baker et al. [4]), o modelo de inferência dos parâmetros causais pode ser calculado.

Os parâmetros causais são atualizados em três fases a cada interação (n):

- Ajustes da $P(Causas_{n-1})$ (interação anterior):
 - Fase 1: através da relação intercausas;
 - Fase 2: incorporação da evidência.
- Cálculo da $P(Causas_n)$ (interação atual):
 - Fase 3: considerando as características (indícios) das ações.

As fases de atualização dos parâmetros causais estão ilustrados na figura 4.2. Na fase 1, as probabilidades causais calculadas são reavaliadas conforme relação intercausas, gerando as seguintes fórmulas:

$$P(Dificuldade_{n-1})' = P(Dificuldade_{n-1}) * P(Esforco_{n-1}) \quad (4.8)$$

$$P(Sorte_{n-1})' = P(Sorte_{n-1}) * P(\neg Esforco_{n-1}) \quad (4.9)$$

$$P(Habilidade_{n-1})' = P(Habilidade_{n-1}) * P(\neg Dificuldade_n) \quad (4.10)$$

$$P(Esforco_{n-1})' = P(Esforco_{n-1}) * P(Habilidade_n) \quad (4.11)$$

Essas relações foram estabelecidas de forma intuitiva uma vez que todas as causas exploradas podem ser correlacionadas.

Observe que a dificuldade e a sorte da etapa anterior, representadas nas fórmulas 4.8 e 4.9, respectivamente, são reajustadas de acordo com o esforço aplicado também na etapa anterior. Um erro por dificuldade é mais provável se houve esforço em uma situação de não-aprendizagem. Esforço em cenários negativos enfatiza a dificuldade. E se não houve esforço, uma possível resposta correta reforça a precisão do valor da sorte já calculado. Não-esforço em cenários positivos enfatiza a sorte. Desta forma, a dificuldade anterior é ajustada na proporção do esforço aplicado e a sorte, na proporção do não esforço aplicado.

A habilidade e o esforço da etapa anterior, representadas nas fórmulas 4.10 e 4.11, respectivamente, utilizam as causas da etapa atual para reajustar seus valores. Enquanto a habilidade é recalculada com a probabilidade de dificuldade na etapa atual, o esforço utiliza o valor da habilidade na etapa atual. Quanto menor a dificuldade corrente, mais correto é o valor da probabilidade de habilidade encontrada anteriormente. E a habilidade atual é consequência do esforço aplicado até o momento. Portanto, a habilidade é ajustada na proporção da não dificuldade apresentada e o esforço, na medida da habilidade adquirida.

Na fase 2 ocorre a incorporação da evidência, ou seja, da ação correta ou incorreta do estudante, no ajuste dos parâmetros causais. Essa atualização é condicionada às probabilidades de equívoco $P(S)$ e adivinhação $P(G)$.

$$P(Causa_{n-1}|A_n) = \frac{P(Causa_{n-1})' * P(A_n|Causa_{n-1})}{P(A_n)} \quad (4.12)$$

$$P(A_n) = P(Causa_{n-1})' * P(A_n|Causa_{n-1}) + \\ P(\neg Causa_{n-1})' * P(A_n|\neg Causa_{n-1}) \quad (4.13)$$

A fórmula 4.12 é a utilizada para o ajuste das causas dada a evidência na fase 2 e a fórmula 4.13 é a probabilidade de ocorrência da ação na etapa n , calculada pela probabilidade incondicional ou marginal.

Nas fórmulas 4.12 e 4.13, $Causa_{n-1}$ pode ser a habilidade, o esforço, a dificuldade ou a sorte, e A_n a evidência (ação correta ou incorreta).

A causa e o conhecimento possuem um relacionamento subjetivo, podendo ser combinados em diferentes cenários (acertar dado esforço e conhecimento, acertar dado o não-esforço e conhecimento, acertar dado a sorte e conhecimento, entre outros).

Para a incorporação da evidência alguns desses cenários foram destacados e a relação com os parâmetros de equívoco $P(S)$ e adivinhação $P(G)$ foram utilizadas para o cálculo $P(A_n|Causa_{n-1})$. Os relacionamentos estão expostos a seguir.

Como a dificuldade é condicionada à falta de conhecimento ($P(\neg L)$), as equações para seu ajuste utilizam a probabilidade de adivinhação $P(G)$ e substituindo nas fórmulas 4.12 e 4.13 geram:

$$P(A_n = Correto|Dificuldade_{n-1}) = P(G)$$

$$P(A_n = Correto|\neg Dificuldade_{n-1}) = P(\neg S)$$

$$P(A_n = Correto) = P(Dificuldade_{n-1})' * P(G) + P(\neg Dificuldade_{n-1})' * P(\neg S)$$

$$P(A_n = Incorreto) = P(Dificuldade_{n-1})' * P(\neg G) + P(\neg Dificuldade_{n-1})' * P(S)$$

$$P(Dificuldade_{n-1}|A_n) = \frac{P(Dificuldade_{n-1})' * P(A_n|Dificuldade_{n-1})}{P(A_n)}$$

A habilidade está vinculada ao conhecimento ($P(L)$), assim, as equações para seu ajuste utilizam a probabilidade de equívoco $P(S)$ e substituindo nas fórmulas 4.12 e 4.13 geram:

$$P(A_n = \text{Correto} | \text{Habilidade}_{n-1}) = P(\neg S)$$

$$P(A_n = \text{Correto} | \neg \text{Habilidade}_{n-1}) = P(G)$$

$$P(A_n = \text{Correto}) = P(\text{Habilidade}_{n-1})' * P(\neg S) + P(\neg \text{Habilidade}_{n-1})' * P(G)$$

$$P(A_n = \text{Incorreto}) = P(\text{Habilidade}_{n-1})' * P(S) + P(\neg \text{Habilidade}_{n-1})' * P(\neg G)$$

$$P(\text{Habilidade}_{n-1} | A_n) = \frac{P(\text{Habilidade}_{n-1})' * P(A_n | \text{Habilidade}_{n-1})}{P(A_n)}$$

Apesar da equação proposta para o cálculo da sorte ser associada ao conhecimento ($P(L)$) ou a aprendizagem ($P(\neg L \wedge T)$), esse suposto conhecimento é resultado de acertos ocasionados pelo acaso. Para calcular a sorte condicionada à ação utiliza-se as probabilidades de adivinhação $P(G)$ e substituindo nas fórmulas 4.12 e 4.13 geram:

$$P(A_n = \text{Correto} | \text{Sorte}_{n-1}) = P(G)$$

$$P(A_n = \text{Correto} | \neg \text{Sorte}_{n-1}) = P(\neg S)$$

$$P(A_n = \text{Correto}) = P(\text{Sorte}_{n-1})' * P(G) + P(\neg \text{Sorte}_{n-1})' * P(\neg S)$$

$$P(A_n = \text{Incorreto}) = P(\text{Sorte}_{n-1})' * P(\neg G) + P(\neg \text{Sorte}_{n-1})' * P(S)$$

$$P(\text{Sorte}_{n-1} | A_n) = \frac{P(\text{Sorte}_{n-1})' * P(A_n | \text{Sorte}_{n-1})}{P(A_n)}$$

E ao contrário da sorte, os acertos ocasionados pelo esforço são originários do conhecimento do discente, portanto, são resultados do não equívoco $P(\neg S)$ e substituindo nas

fórmulas 4.12 e 4.13 geram:

$$P(A_n = \textit{Correto} | \textit{Esforco}_{n-1}) = P(\neg S)$$

$$P(A_n = \textit{Correto} | \neg \textit{Esforco}_{n-1}) = P(G)$$

$$P(A_n = \textit{Correto}) = P(\textit{Esforco}_{n-1})' * P(\neg S) + P(\neg \textit{Esforco}_{n-1})' * P(G)$$

$$P(A_n = \textit{Incorreto}) = P(\textit{Esforco}_{n-1})' * P(S) + P(\neg \textit{Esforco}_{n-1})' * P(\neg G)$$

$$P(\textit{Esforco}_{n-1} | A_n) = \frac{P(\textit{Esforco}_{n-1})' * P(A_n | \textit{Esforco}_{n-1})}{P(A_n)}$$

Na fase 3, por fim, ocorrem as alterações das causas na etapa atual devido ao conhecimento inferido. As probabilidades do “conhecimento ajustado” ($P(L_A)$), da “aprendizagem ajustada” ($P(nLT_A)$) e da “não-aprendizagem ajustada” ($P(nLnT_A)$) foram interpretadas como indícios para determinar as causas conforme mostra figura 4.2.

Essas probabilidades podem caracterizar o aumento da dificuldade $P(adD)$, a indicação de sorte $P(adS)$, a evolução da habilidade $P(adH)$ e o aumento do esforço $P(adE)$.

O modelo utiliza as probabilidades condicionadas e considera pesos no ajuste das probabilidades. Os pesos são as porcentagens das respostas corretas e incorretas das últimas 10 oportunidades de resolver uma tarefa com uma determinada componente de conhecimento.

$$\begin{aligned} P(\textit{Dificuldade}_n) &= P(\textit{Dificuldade}_{n-1} | A_n) + \\ &P(\neg \textit{Dificuldade}_{n-1} | A_n) * P(adD) \end{aligned} \quad (4.14)$$

$$\begin{aligned} P(\textit{Sorte}_n) &= P(\textit{Sorte}_{n-1} | A_n) + \\ &P(\neg \textit{Sorte}_{n-1} | A_n) * P(adS) \end{aligned} \quad (4.15)$$

$$\begin{aligned}
P(Habilidade_n) &= P(Habilidade_{n-1}|A_n) + \\
&P(\neg Habilidade_{n-1}|A_n) * P(adA)
\end{aligned} \tag{4.16}$$

$$\begin{aligned}
P(Esforco_n) &= P(Esforco_{n-1}|A_n) + \\
&P(\neg Esforco_{n-1}|A_n) * P(adE)
\end{aligned} \tag{4.17}$$

As fórmulas 4.14, 4.15, 4.16 e 4.17 são as responsáveis pelo cálculo dos parâmetros causais da etapa atual, baseado no estado anterior da causa e seus indicativos de ocorrência.

$$P(adD) = P(nLnT_A) + P(L_A) * P(S) \tag{4.18}$$

$$\begin{aligned}
P(adS) &= (P(L_A) + P(nLT_A)) * \text{porcento_ultimas_incorretas} + \\
&P(nLnT_A) * P(G)
\end{aligned} \tag{4.19}$$

$$P(adH) = P(L_A) * \text{porcento_ultimas_corretas} \tag{4.20}$$

$$\begin{aligned}
P(adE) &= (P(L_A) + P(nLT_A)) * \text{porcento_ultimas_corretas} + \\
&P(nLnT_A) * (\text{tentativas}/10)
\end{aligned} \tag{4.21}$$

A fórmula 4.18 representa os indícios de dificuldade que são o não-conhecimento ($P(nLnT_A)$) e a probabilidade de erro apesar do conhecimento. Não ter aprendido pode indicar dificuldade na execução da tarefa, podendo ser diminuída conforme a habilidade é adquirida.

A fórmula 4.19 representa os indícios de sorte que são o conhecimento ($P(L_A)$) ou aprendizagem ($P(nLT_A)$) quando há pouca experiência e erros, e a probabilidade de adivinhar ($P(G)$) a resposta em um cenário de não-conhecimento ($P(nLnT_A)$).

Há casos em que a sorte pode ser “confundida” com habilidade, sendo um acerto considerado erroneamente como sinal de domínio e a suposta aprendizagem poderia originar-se de um momento de sorte. Porém, somente $P(nLT_A)$ e $P(L_A)$ não são bons indicativos de sorte, tendendo mais para habilidade ou esforço.

Para fazer essa diferenciação do indicativo de conhecimento por sorte ou habilidade ou esforço foi utilizado a variável *porcento_ultimas_incorretas*. Esta variável retrata a relação do “número de oportunidades” com o “número de respostas incorretas”. É a porcentagem de entradas incorretas na primeira tentativa de responder uma tarefa nas últimas 10 oportunidades de exercitar determinado conceito. Por exemplo, a porcentagem de 0,8 adverte para a ocorrência de muitos erros e supõe-se que as probabilidades $P(L_A)$ e $P(nLT_A)$ indiquem que o suposto conhecimento ou aprendizagem correspondam a eventos de sorte. Caso contrário, é maior a chance do conhecimento ou aprendizagem ser resultado da habilidade ou esforço investido.

No cálculo da sorte *porcento_ultimas_incorretas* pode indicar a experiência do estudante. O número de oportunidades de exercitar uma componente de conhecimento em tarefas distintas, tem função relevante pois impede que as respostas corretas assumam um impacto maior do que seria condizente na inferência dos parâmetros causais.

Quando não houver tentativas anteriores ao corrente, ou menor que 10, as ações não existentes serão consideradas falsas. Desta forma, *porcento_ultimas_incorretas* assume um valor maior, indicando que o estudante tem pouca experiência, convergindo para a hipótese de que as respostas corretas até o momento sejam causadas pelo acaso.

A fórmula 4.21 representa os indícios de esforço aplicado na resolução da tarefa que são o conhecimento ($P(L_A)$) ou aprendizagem ($P(nLT_A)$) na proporção do número de acertos, e o não-conhecimento mas com várias tentativas de resolver a mesma tarefa (1% de $P(nLnT_A)$ por tentativa, limitado a 10 tentativas).

Ao contrário da variável *porcento_ultimas_incorretas*, a *porcento_ultimas_corretas* retrata a relação do “número de oportunidades” com o “número de respostas corretas”. É a porcentagem de entradas corretas na primeira tentativa de responder uma tarefa nas últimas 10 oportunidades de exercitar determinado conceito.

A fórmula 4.20 representa os indícios de habilidade que são o conhecimento ($P(L_A)$) e a prática, representada por *porcento_ultimas_corretas*.

As ações das últimas oportunidades ajudam a garantir o valor estimado para a habilidade e esforço. Os acertos e erros reforçam a estimativa desses parâmetros causais. Cada vez que exercita e acerta, o fator sorte diminui e aumenta a habilidade e/ou esforço. Apesar do modelo diminuir a sorte à medida que a habilidade e esforço aumentam, isso não significa que a sorte “desapareça” e sim que a habilidade sobressai à sorte.

As probabilidades no tempo zero utilizam os valores iniciais calculados pelo modelo BKT e as fórmulas são:

$$P(Dificuldade_0) = P(\neg L_0) * P(\neg T) + P(L_0) * P(S)$$

$$P(Sorte_0) = P(L_0) + P(\neg L_0) * P(T) + P(\neg L_0) * P(\neg T) * P(G)$$

$$P(Habilidade_0) = 0$$

$$P(Esforco_0) = 0$$

A dificuldade inicial é estimada pela probabilidade de não saber e não aprender, ou pela chance de saber e errar; a sorte pela probabilidade de saber, ou não saber e aprender, ou não saber e não aprender e mesmo assim acertar. Qualquer chance de acerto no tempo zero é considerado sorte, desta forma, a habilidade e o esforço são considerados inicialmente nulos.

4.3 Classificação e testes: análise pré-ação dos parâmetros causais com um conjunto de dados reais

Para analisar o impacto dos atributos causais modelados e calculados, foi feita uma classificação para a previsão da ação, correta ou incorreta, de um estudante na sua primeira tentativa de resposta. Através deste estudo considera-se a possibilidade da aplicação desses parâmetros na antecipação do comportamento do estudante.

Apesar dos fatores internos serem conceitos imprecisos e calculados probabilística-

mente, um modelo determinístico foi aplicado com o intuito de facilitar a compreensão da relação entre os valores inferidos e a ação do estudante.

Segundo o postulado da compreensibilidade de Michalski [25], os resultados da indução por computador devem ser descrições simbólicas das entidades fornecidas, sendo semântica e estruturalmente similares àquelas que um especialista humano poderia produzir observando as mesmas entidades. Os componentes dessas descrições devem ser compreensíveis como simples pedaços de informação, diretamente interpretáveis em linguagem natural, bem como reportar conceitos quantitativos e qualitativos de maneira integrada.

Nesta seção há os detalhes sobre a preparação dos dados e os procedimentos aplicados para a geração final do modelo apresentado na subseção 4.2.3. Seguido dos resultados obtidos através dos testes efetuados.

4.3.1 Preparação dos dados

Os dados utilizados foram previamente catalogados do sistema tutor inteligente “Cognitive Tutor” desenvolvido pela “Carnegie Learning” e disponibilizados no *site* da competição da edição do ano de 2010 do *Knowledge Discover and Data Mining Cup*. Inicialmente os dados utilizados foram o da turma de 2006-2007, denominado de “*Bridge to Algebra 2006-2007*”, que compreende mais de 1.000 alunos e mais de 3 milhões de etapas, sendo que foram processados mais de 1.8 milhões de transações, e o conjunto de dados “*Algebra I 2005-2006*” que possui 575 estudantes e 607.025 transações avaliadas. A primeira base de dados foi utilizada nos treinamentos e testes dos modelos gerados, e a segunda somente para os testes, para garantir o desempenho do modelo para novos estudantes e componentes de conhecimentos. O conjunto de dados “*Algebra I 2005-2006*” será mencionado como “base de dados para teste”.

O conjunto de dados “*Bridge to Algebra 2006-2007*” foi dividido em dois arquivos: um arquivo de treinamento e um arquivo de teste.

O arquivo de treinamento foi utilizado para a classificação das probabilidades do nível de conhecimento condicionadas às ações futuras e geração dos parâmetros causais. O arquivo de teste foi utilizado para a previsão das respostas.

As linhas do arquivo de teste foram selecionadas aleatoriamente dos problemas para cada aluno. As etapas anteriores a cada linha selecionada para teste, foram inseridas no arquivo de treinamento.

Para a utilização do conjunto de dados foram necessárias algumas modificações. Existem tarefas que envolvem múltiplos componentes de conhecimento necessários para sua resolução. Neste casos esses componentes foram avaliados como uma única habilidade na determinação do nível de conhecimento, habilidade, esforço, dificuldade e sorte.

Nos conjuntos de dados existem transações sem especificação da componente de conhecimento necessária para sua execução. Como o objetivo deste trabalho é avaliar os parâmetros causais, estas transações sem a habilidade explícita, foram descartadas para facilitar a etapa de preparação dos dados.

Os atributos disponíveis para cada transação são:

- Identificação do estudante;
- Hierarquia do problema;
- Nome do problema;
- Número de vezes que o visualizou o problema;
- Nome da etapa do problema;
- Hora do início da execução da etapa;
- Hora da primeira transação;
- Hora da transação correta;
- Hora do fim da execução da etapa;
- Duração da etapa (duração de todas as transações/tentativas);
- Duração do acerto na etapa (se a primeira tentativa na etapa for correta);
- Duração do erro na etapa (se a primeira tentativa na etapa for erro ou requisição de dica);

- Acerto na primeira tentativa;
- Número de tentativas incorretas na etapa corrente;
- Número de requisições de ajuda na etapa corrente;
- Número de tentativas corretas na etapa corrente;
- Componente de conhecimento;
- Número de oportunidades de exercitar o componente de conhecimento.

4.3.2 Procedimentos

Para a aplicação do modelo proposto, depois da preparação dos dados, foram calculados os níveis de conhecimentos para cada componente de conhecimento com o modelo BKT e os rótulos das probabilidades $P(L_n|A_{n+1,n+2})$, $P(\neg L_n \wedge T|A_{n+1,n+2})$ e $P(\neg L_n \wedge \neg T|A_{n+1,n+2})$.

Em termos práticos, os dados foram armazenados utilizando a biblioteca de banco de dados SQLite, para organizar as informações, visando facilitar as constantes modificações na captura dos dados, e foi desenvolvido com a linguagem C++ e interface de programação de aplicativo OpenMP (acrônimo do inglês *Open Multi-Processing*). Para o cálculo do nível de conhecimento para treinamento, que exige mais processamento e memória, a máquina utilizada foi uma *Intel Xeon E5-2680 v2 2.80GHz (2 sockets, 10 cores por socket)* e para o processamento do nível de conhecimento com o modelo treinado e os parâmetros causais foi utilizado uma máquina *Intel Core i5-4200U 1.60GHz (1 socket, 2 cores)*.

Para encontrar as melhores estimativas dos parâmetros do modelo BKT para o cálculo do nível de conhecimento foi utilizado o método de busca por força bruta. Foram julgados todas as combinações de parâmetros possíveis de valores a uma granulometria de 0,01. A escolha por esse método foi baseado no trabalho comparativo de Baker et al.[3] que sugeriu que a abordagem de força bruta resulta em um bom ajuste para a predição dos parâmetros do modelo BKT.

Portanto, como a abordagem por força bruta é rápida de ser aplicada, resulta em boas estimativas e o objetivo foi a análise inicial do impacto do modelo desenvolvido, o quesito temporal na execução dos algoritmos implementados e a análise de outros métodos de estimativa não ocuparam muito tempo no desenvolvimento deste projeto.

Apesar disso, o OpenMP foi aplicado para diminuir o tempo de processamento do algoritmo de busca por força bruta. Os cálculos dos parâmetros BKT executaram em cerca de 145 minutos, a sua variante foi executada em cerca de 10 minutos, e o cálculo dos parâmetros causais mais a preparação dos arquivos de treinamento e teste para todas as transações e estudantes demorou cerca de 4 minutos.

Para o cálculo do conhecimento inicial foram analisadas 564 componentes de conhecimento e 1.642.020 transações processadas. Os limites estabelecidos para os valores iniciais do conhecimento $P(L_0)$, aprendizagem $P(T)$, equívoco $P(S)$ e adivinhação $P(G)$ foram respectivamente 0,99, 0,99, 0,30 e 0,30, conforme recomendação de estudos anteriores [3] (ver capítulo 3 seção 3.2.1). O critério de escolha foi a combinação com menor discrepância entre os dados e a estimativa do modelo, medido através do resíduo SSR (acrônimo do inglês *sum of squared residuals*)

Esses mesmos procedimentos foram aplicados na base de dados para teste, porém não foram utilizados nos treinamentos.

O próximo passo foi o cálculo dos ajustes do nível de conhecimento para treinamento com ações futuras. As probabilidades $P(L_n|A_{n+1,n+2})$, $P(\neg L_n \wedge T|A_{n+1,n+2})$ e $P(\neg L_n \wedge \neg T|A_{n+1,n+2})$ foram geradas a partir dos históricos dos alunos, conforme apresentado na subseção 4.2.2 deste capítulo. Estes valores (rótulos) foram utilizados para a classificação por regressão linear, com a finalidade de prever a probabilidade em tempo de execução, ou seja, sem dados futuros, usando somente ações passadas. As probabilidades resultantes da classificação são referenciadas como $P(L_A)$ (saber), $P(nLT_A)$ (não saber e aprender) e $P(nLnT_A)$ (não saber e não aprender).

Para refinar a classificação foi definido um conjunto de atributos que inclui:

- Probabilidades de conhecimento na etapa atual N ;
- Probabilidades de conhecimento na etapa anterior $N - 1$;

- Média de duração por etapa;
- Média de erros por etapa;
- Média de acertos por etapa;
- Média de requisições de ajuda por etapa;
- Porcentagem de acertos na primeira tentativa;
- Porcentagem de respostas corretas nas últimas 10 oportunidades;
- Probabilidades de conhecimento condicionado as ações (detalhes na seção 3);
- Probabilidade do “conhecimento ajustado” $P(L_{-}A)$ (utilizado no cálculo do $P(nLnT_{-}A)$).

O conjunto de atributos não foram os mesmos utilizados por Baker et al. [4] por limitação de informações como o tempo para cada transação de uma etapa específica e a informação do tipo de erro (ajuda ou inserção de resposta incorreta), gerando um modelo de regressão diferente ao do modelo original.

Os valores $P(L_{-}A)$, $P(nLT_{-}A)$ e $P(nLnT_{-}A)$, obtidos pelo modelo treinado a partir de $P(L_n|A_{n+1,n+2})$, $P(\neg L_n \wedge T|A_{n+1,n+2})$ e $P(\neg L_n \wedge \neg T|A_{n+1,n+2})$, foram inseridos nas fórmulas dos parâmetros causais expressas na subseção 4.2.3.

A validação cruzada *5-fold* foi executada para avaliar a capacidade de generalização do modelo e os resultados obtidos foram os representados na tabela 4.1.

Tabela 4.1: Validação cruzada para o conhecimento condicionado.

Variável	Coefficiente de correlação	Coef. de correlação (base de teste)
$P(L_n A_{n+1,n+2})$	0,9646	0,9777
$P(\neg L_n \wedge \neg T A_{n+1,n+2})$	0,9407	0,9455

Como os parâmetros $P(\neg L_n \wedge \neg T|A_{n+1,n+2})$ e $P(\neg L_n \wedge T|A_{n+1,n+2})$ são complementares da probabilidade $P(\neg L_n|A_{n+1,n+2}) = 1.0 - P(L_n|A_{n+1,n+2})$, somente a regressão gerada para $P(\neg L_n \wedge \neg T|A_{n+1,n+2})$ foi utilizada, sendo $P(\neg L_n \wedge T|A_{n+1,n+2})$ calculado como $P(\neg L_n|A_{n+1,n+2}) - P(\neg L_n \wedge \neg T|A_{n+1,n+2})$. Os valores obtidos foram limitados de 0,01 a 0,99 para a probabilidade ajustada $P(L_{-}A)$ e 0,01 a $(1.0 - P(L_{-}A))$ para $P(nLnT_{-}A)$.

Com o modelo gerado, as inferências dos parâmetros causais podem ser calculadas a partir das equações apresentadas na seção 4.2.3, e incorporadas nos arquivos de treinamento e teste para a análise do impacto desses novos atributos.

4.3.3 Resultados

A relevância do uso do modelo dos parâmetros causais pós-ação é uma pesquisa em aberto pois para a plena convicção de seu resultado seria necessário a aplicação em um modelo estratégico de tutoria como, por exemplo, um esquema de *feedback* inteligente introduzido em um STI e submetido ao uso dos discentes para a análise do estados emocionais e desempenhos.

Exemplo 1. *Para exemplificar o cálculo dos parâmetros causais, considere um problema, do STI Cognitive Tutor Algebra da Carnegie Learning, que possui uma componente de conhecimento denominada “Comparação de frações com denominadores distintos”. Um determinado estudante, após uma tentativa fracassada em resolver uma tarefa que envolvia a componente citada, teve sua probabilidade de habilidade estimada em 0,019, de esforço em 0,0578, de dificuldade em 0,479 e de sorte em 0,999. Em outra tentativa ele novamente erra a resposta, tendo suas estimativas de habilidade, esforço, dificuldade e sorte, atualizadas para 0,0155, 0,0522, 0,547 e 0,99, respectivamente.*

Como já mencionado, uma possível aplicação dos valores estimados é no auxílio da decisão em estratégias de ensino. Assim, com a probabilidade de dificuldade crescente do exemplo acima, o sistema tutor poderia habilitar mecanismos de intervenções pedagógicas, cujo grau de especificidade seria proporcional ao nível de dificuldade, ou podendo também, analisar o se houve esforço na resolução do exercício.

Exemplo 2. *Alguns resultados gerados pelo modelo BKT e dos parâmetros causais, para determinados componentes de conhecimento e estudantes:*

- *Componente 1:*

- *Etapa n – 2:*

* Saída: erro;

* $P(\text{Habilidade}_{n-2}) = 0,019$, $P(\text{Esforço}_{n-2}) = 0,0578$,

$P(\text{Dificuldade}_{n-2}) = 0,479$ e $P(\text{Sorte}_{n-2}) = 0,999$.

– Etapa $n - 1$:

* Saída: erro;

* $P(\text{Habilidade}_{n-1}) = 0,0155$, $P(\text{Esforço}_{n-1}) = 0,0522$,

$P(\text{Dificuldade}_{n-1}) = 0,547$ e $P(\text{Sorte}_{n-1}) = 0,99$.

– Etapa n :

* Saída: acerto;

* $P(\text{Habilidade}_n) = 0,0727$, $P(\text{Esforço}_n) = 0,134$,

$P(\text{Dificuldade}_n) = 0,456$ e $P(\text{Sorte}_n) = 0,949$;

* $P(L_n) = 0,41$.

• *Componente 2:*

– Etapa $n - 2$:

* Saída: acerto;

* $P(\text{Habilidade}_{n-2}) = 0,999$, $P(\text{Esforço}_{n-2}) = 1,0$,

$P(\text{Dificuldade}_{n-2}) = 0,16$ e $P(\text{Sorte}_{n-2}) = 0,594$.

– Etapa $n - 1$:

* Saída: acerto;

* $P(\text{Habilidade}_{n-1}) = 0,999$, $P(\text{Esforço}_{n-1}) = 1,0$,

$P(\text{Dificuldade}_{n-1}) = 0,16$ e $P(\text{Sorte}_{n-1}) = 0,495$.

– Etapa n :

* Saída: erro;

* $P(\text{Habilidade}_n) = 0,537$, $P(\text{Esforço}_n) = 0,57$,

$P(\text{Dificuldade}_n) = 0,628$ e $P(\text{Sorte}_n) = 0,495$;

* $P(L_n) = 1,0$.

- *Componente 3:*

- *Etapa $n - 2$:*

- * *Saída: erro;*

- * $P(Habilidade_{n-2}) = 0,478$, $P(Esforco_{n-2}) = 0,51$,

- $P(Dificuldade_{n-2}) = 0,622$ e $P(Sorte_{n-2}) = 0,736$.

- *Etapa $n - 1$:*

- * *Saída: acerto;*

- * $P(Habilidade_{n-1}) = 0,733$, $P(Esforco_{n-1}) = 0,789$,

- $P(Dificuldade_{n-1}) = 0,422$ e $P(Sorte_{n-1}) = 0,596$.

- *Etapa n :*

- * *Saída: erro;*

- * $P(Habilidade_n) = 0,549$, $P(Esforco_n) = 0,62$,

- $P(Dificuldade_n) = 0,678$ e $P(Sorte_n) = 0,624$;

- * $P(L_n) = 0,995$.

Analisando os valores gerados através do modelo dos parâmetros causais, observa-se que a cada acerto, as probabilidades de habilidade e esforço tendem a aumentar, enquanto dificuldade e chance de acerto ser por causa da sorte tendem a diminuir. E o contrário, a cada erro, habilidade e esforço tendem a diminuir, e dificuldade tende a aumentar. A sorte, em caso de erro, depende do nível de conhecimento: quanto mais conhecimento maior a chance do erro ser sinal de sorte no passado.

A análise do impacto pré-ação dos parâmetros causais foi com a geração de um modelo de classificação determinística, treinado com 1.817.476 instâncias e testado com 5.221 instâncias, com um conjunto de atributos.

O conjunto de atributos utilizados foram:

- Parâmetros causais da etapa anterior ($P(Habilidade_{n-1})$, $P(Esforco_{n-1})$, $P(Sorte_{n-1})$, $P(Dificuldade_{n-1})$);

- Corretas consecutivas: número de ações corretas consecutivas nas últimas 10 oportunidades;
- Duração na etapa anterior e atual: tempo do início da tarefa até a inserção da primeira resposta (segundos);
- Número de oportunidades: quantidade de vezes que o estudante já teve a oportunidade de exercitar a componente de conhecimento;
- Ação na primeira tentativa (meta).

As probabilidades das causas avaliadas pela aprendizagem de máquina são anteriores à etapa testada ($P(Habilidade_{n-1})$, $P(Esforco_{n-1})$, $P(Dificuldade_{n-1})$ e $P(Sorte_{n-1})$) pois, como a ação (resposta) é o atributo classe, não se pode utilizar as probabilidades da etapa corrente ($P(Habilidade_n)$, $P(Esforco_n)$, $P(Dificuldade_n)$ e $P(Sorte_n)$) que utiliza o atributo classe em suas fórmulas.

A técnica de classificação utilizada para o treinamento foi a árvore de decisão. Auxiliado pela ferramenta “Weka” [38], um conjunto de dados foi treinado e testado com o classificador do tipo “J48”, que implementa o algoritmo “C4.5”.

O modelo foi submetido a validação cruzada *5-fold* com o conjunto de treinamento para a captação de informações como os falsos positivos/negativos, os verdadeiros positivos/negativos e as instâncias corretamente classificadas e os resultados estão na tabela 4.2.

Os atributos “duração”, “corretas consecutivas” e “oportunidades”, explicados anteriormente, foram inseridos para atuarem como critério de interpretação para descrever melhor as características que poderiam levar ao acerto ou erro. Como a expectativa e o afeto são presumidos para determinar a ação e esta pode ser descrita de acordo com a sua intensidade, sua latência, e assim por diante, então justifica-se o uso dos parâmetros causais na previsão da ação e o uso de um conjunto de atributos para mensurar os itens de descrição de uma ação.

A maior deficiência do modelo foi na detecção dos fatores que levam ao erro, apontado pela alta taxa de falsos positivos. Esta situação pode ser explicada pelas diferenças de

Tabela 4.2: Resultados do modelo determinístico na previsão da ação do estudante através dos valores obtidos com o modelo dos parâmetros causais.

Teste	Corretos*	Incorretos*	Detalhes
Validação cruzada 5- <i>fold</i> 1.512.413 instâncias verdadeiras 305.063 instâncias falsas	1.534.182 (84,4128%)	283.294 (15,5872%)	0,785 falsos positivos 0,971 verdadeiros positivos 0,029 falsos negativos 0,215 verdadeiros negativos
Arquivo-teste 4.317 instâncias verdadeiras 904 instâncias falsas	4.366 (83,6238%)	855 (16,3762%)	0,747 falsos positivos 0,958 verdadeiros positivos 0,042 falsos negativos 0,253 verdadeiros negativos
Base de dados para teste 458.479 instâncias verdadeiras 148.546 instâncias falsas	478.684 (78,8574%)	128.341 (21,1426%)	0,669 falsos positivos 0,937 verdadeiros positivos 0,063 falsos negativos 0,331 verdadeiros negativos

* Instâncias corretamente ou incorretamente classificadas

cenários consequentes das interpretações particulares dos estudantes, que foram destacadas como fontes da variação das emoções geradas para uma mesma situação de acordo com Weiner [37].

Esta questão também envolve a percepção da autoeficácia que, segundo Bandura [6], trata-se da crença das pessoas sobre suas próprias capacidades e que a percepção da sua própria eficácia influencia os cenários que eles constroem ou reiteram. Os que tem alto senso de eficácia visualizam cenários que fornecem guias positivos para o desempenho. E quando julgam a si mesmo ineficazes, tendem a construir cenários negativos podendo resultar em eventos de falha. A investigação específica das diferentes e individuais interpretações das causas é uma pretensão de pesquisa posterior.

Além da aplicação da validação-cruzada, o modelo foi testado com as transações do arquivo-teste e uma segunda base de dados, e o resultado apresentado na tabela 4.2.

As taxas de verdadeiro positivos das análises evidenciaram que os parâmetros causais ajudaram significativamente na previsão dos acertos; as taxas de falso negativos indicaram que pouco erraram quando ocorreu a afirmação de resultados negativos; e as taxas de falso positivos e verdadeiro negativos alertaram para a deficiência na previsão da maioria dos casos de falhas na resolução das tarefas.

Interpretações sobre as consequências emocionais e comportamentais foram conjecturadas de algumas decisões (ramos e folhas), compostos pelos atributos causais, “duração”,

“corretas consecutivas” e “oportunidades”, da classificação por árvore de decisão gerada com a base de treinamento, conforme detalhado anteriormente. As probabilidades pré-ação foram interpretadas somente para ilustrar algumas das possíveis relações entre as causas, emoções e comportamentos, que podem ser obtidas pelos valores dos parâmetros causais inferidos pelo modelo proposto, e são apresentadas na tabela 4.3.

Tabela 4.3: Decisões geradas pelo classificador e interpretações

	Condições e Emoções	Comportamento
1	Habilidade $(n - 1) \leq 0,0567 \rightarrow$ vergonha Esforço $(n - 1) \leq 0,0938 \rightarrow$ culpa Dificuldade $(n - 1) \leq 0,2630 \rightarrow$ esperança Sorte $(n - 1) \leq 0,8330 \rightarrow$ esperança	Cenário negativo: - baixa motivação - baixo desempenho - erro
2	Habilidade $(n - 1) > 0,0567 \leq 0,9830 \rightarrow$ orgulho Esforço $(n - 1) \leq 0,0938 \rightarrow$ culpa Dificuldade $(n - 1) \leq 0,2630 \rightarrow$ esperança Sorte $(n - 1) \leq 0,8330 \rightarrow$ esperança	Cenário positivo: - expectativa positiva - acerto
3	Habilidade $(n - 1) \leq 0,983 \rightarrow$ orgulho Esforço $(n - 1) \leq 0,0938 \rightarrow$ culpa Dificuldade $(n - 1) > 0,132 \leq 0,134 \rightarrow$ esperança Sorte $(n - 1) > 0,833 \leq 0,927 \rightarrow$ esperança Duração atual $\leq 3 \rightarrow$ ansiedade falta de atenção Oportunidades $\leq 8 \rightarrow$ insegurança preocupação Corretas consecutivas = 0 \rightarrow tristeza raiva desapontamento relaxamento	Cenário positivo: - expectativa positiva/indiferença - acerto

Continua na próxima página

Tabela 4.3 – continuação da página anterior

	Condições e Emoções	Comportamento
4	<p>Habilidade $(n - 1) \leq 0,983 \rightarrow$ orgulho</p> <p>Esforço $(n - 1) \leq 0,0938 \rightarrow$ culpa</p> <p>Dificuldade $(n - 1) > 0,132 \leq 0,134 \rightarrow$ esperança</p> <p>Sorte $(n - 1) > 0,833 \leq 0,927 \rightarrow$ esperança</p> <p>Duração atual $> 3 \rightarrow$ atenção aumentando ansiedade aumentando</p> <p>Oportunidades $\leq 8 \rightarrow$ insegurança preocupação</p> <p>Corretas consecutivas = 0 \rightarrow tristeza raiva desapontamento</p>	<p>Cenário negativo:</p> <ul style="list-style-type: none"> - expectativa negativa - erro
5	<p>Habilidade $(n - 1) \leq 0,0344 \rightarrow$ vergonha</p> <p>Dificuldade $(n - 1) > 0,419 \leq 0,501 \rightarrow$ preocupação</p> <p>Sorte $(n - 1) \leq 0,683 \rightarrow$ esperança</p> <p>Duração atual $\leq 10 \rightarrow$ atenção aumentando ansiedade aumentando</p> <p>Oportunidades $\leq 7 \rightarrow$ insegurança preocupação</p>	<p>Cenário negativo:</p> <ul style="list-style-type: none"> - baixa motivação - expectativa negativa - erro
6	<p>Habilidade $(n - 1) \leq 0,0344 \rightarrow$ vergonha</p> <p>Dificuldade $(n - 1) > 0,361 \leq 0,446 \rightarrow$ esperança</p> <p>Sorte $(n - 1) > 0,791 \leq 0,968 \rightarrow$ esperança</p> <p>Duração atual $\leq 21 \rightarrow$ atenção</p> <p>Oportunidades $> 1 \leq 2 \rightarrow$ insegurança preocupação medo</p> <p>Corretas consecutivas = 0 \rightarrow tristeza</p>	<p>Cenário negativo:</p> <ul style="list-style-type: none"> - expectativa negativa - erro
Continua na próxima página		

Tabela 4.3 – continuação da página anterior

	Condições e Emoções	Comportamento
7	<p>Habilidade $(n - 1) \leq 0,0344 \rightarrow$ vergonha</p> <p>Dificuldade $(n - 1) > 0,263 \leq 0,446 \rightarrow$ esperança</p> <p>Sorte $(n - 1) > 0,997 \leq 0,998 \rightarrow$ esperança</p> <p>Duração atual $\leq 83 \rightarrow$ atenção</p> <p>Duração anterior $\leq 3 \rightarrow$ ansiedade</p> <p>falta de atenção</p> <p>Oportunidades $\leq 2 \rightarrow$ insegurança</p> <p>preocupação</p> <p>medo</p>	<p>Cenário positivo:</p> <ul style="list-style-type: none"> - aumento na motivação - acerto
8	<p>Habilidade $(n - 1) > 0,894 \rightarrow$ orgulho</p> <p>Esforço $(n - 1) > 0,993 \leq 0,994 \rightarrow$ orgulho</p> <p>Dificuldade $(n - 1) > 0,398 \leq 0,601 \rightarrow$ preocupação</p> <p>Sorte $(n - 1) > 0,201 \leq 0,202 \rightarrow$ preocupação</p> <p>Duração anterior $\leq 87 \rightarrow$ atenção</p> <p>Duração atual $> 12 \rightarrow$ diminuição da atenção</p> <p>Oportunidades $> 49 \leq 212 \rightarrow$ segurança</p> <p>relaxamento</p> <p>Corretas consecutivas $> 0,1 \rightarrow$ alegria</p>	<p>Cenário negativo:</p> <ul style="list-style-type: none"> - expectativa positiva gerou acomodação - erro
9	<p>Habilidade $(n - 1) > 0,894 \rightarrow$ orgulho</p> <p>Esforço $(n - 1) \leq 0,943 \rightarrow$ orgulho</p> <p>Dificuldade $(n - 1) > 0,398 \leq 0,607 \rightarrow$ preocupação</p> <p>Sorte $(n - 1) \leq 0,222 \rightarrow$ preocupação</p> <p>Duração atual $> 10 \leq 34 \rightarrow$ atenção</p> <p>Oportunidades $\leq 212 \rightarrow$ segurança</p> <p>Corretas consecutivas $\leq 0,9 \rightarrow$ alegria</p> <p>segurança</p> <p>satisfação</p>	<p>Cenário positivo:</p> <ul style="list-style-type: none"> - motivação estável - expectativa - acerto

Exemplo 3. A decisão 1 da tabela 4.3 indica que apesar de fácil (apresenta baixa dificuldade) e possivelmente ter sorte, o que gera esperança, a falta de esforço, o que gera culpa, e habilidade, o que gera vergonha, resulta em baixa motivação e baixo desempenho,

ou seja, em falha na resolução da tarefa.

Exemplo 4. *Em situação semelhante à decisão 1, o aumento da habilidade, indicada na decisão 2, foi o suficiente para alterar o cenário positivamente.*

Exemplo 5. *Na decisão 3, o erro na última oportunidade (corretas consecutivas) e o início rápido na execução da tarefa (duração atual), indicam falta de comprometimento, porém como a tarefa é fácil e, os níveis de habilidade (orgulho) e sorte são altos, compensam o pouco esforço fazendo com que acerte a resposta.*

Exemplo 6. *Na decisão 4 com o aumento no tempo de início da atividade, o que indica atenção, esperava-se que continuasse ocorrendo o acerto. Este é um caso que a interpretação pessoal, talvez por causa da culpa por não ter se esforçado, implique em um resultado diferente do esperado ao influenciar a expectativa do estudante.*

Exemplo 7. *Na decisão 5, a pouca habilidade e o médio grau de dificuldade, quando associado ao tempo para o início da tarefa e número de oportunidades relativamente baixos, resulta em um cenário negativo. Estas condições indicam vergonha, ansiedade, preocupação e insegurança.*

Exemplo 8. *A pouca habilidade e experiência, e o erro nas primeiras oportunidades, demonstrados na decisão 6, comprometem a expectativa do estudante ocasionando um erro na resolução, mesmo com condições favoráveis como a tarefa ser fácil, a provável sorte envolvida (fator externo, incontrolável) e a atenção inicial (tempo para início da tarefa).*

Exemplo 9. *Já na decisão 7, a pouca habilidade, a pouca experiência e a falta de atenção anterior (duração anterior) refletem no aumento da dedicação na etapa atual (duração atual). Esta situação aliada a facilidade da tarefa colaboram para o acerto da questão.*

Exemplo 10. *Mesmo com habilidade, esforço e experiência, o erro indicado na decisão 8, pode ser originado pela acomodação/relaxamento do estudante pelo fato de ter acertado em oportunidades anteriores.*

Exemplo 11. *A decisão 9 ilustra um padrão interessante: mesmo com um cenário semelhante ao 8, uma diminuição no esforço gera uma resposta positiva. Uma explicação é a estável dedicação indicada pela atenção na leitura inicial, sem grande variação de tempo, e pelas consecutivas respostas corretas. Outra situação plausível é que mais esforço também está associado a muitas tentativas, e não necessariamente ao conhecimento.*

4.4 Considerações Finais

Neste capítulo os elementos do capítulo 3 foram associados para compor o modelo para a inferência das probabilidades de algumas causas que interferem no comportamento do estudante.

Os parâmetros causais foram baseados da teoria da “Atribuição Causal” proposta por Weiner [37]. Esta teoria apresenta as percepções causais e sua relação com a motivação que atua no processo cognitivo. Para caracterizar esta teoria os atributos estudados foram os fatores internos habilidade e esforço, e os fatores externos sorte e dificuldade (com atualização individual).

Ressalta-se que a determinação das possíveis causas comportamentais possibilita sua combinação para a definição das consequentes emoções. Por exemplo, não ter habilidade pode resultar em um sentimento de vergonha, não ter se esforçado associa-se a culpa, uma tarefa difícil pode causar desespero, porém ter sorte, conforta através da esperança. Assim, uma expectativa seria gerada desta combinação de vergonha, culpa, desespero e esperança, sendo condicionado às probabilidades causais. A análise aprofundada deste mapeamento é um trabalho futuro.

Para a predição dos estados causais foi feita a avaliação das ações da interação do estudante com um STI, ou seja, entradas das respostas, número de tentativas, requisições de ajuda, entre outros. Dados oriundos de sensores, câmeras ou outros meios físicos não foram explorados. Por isso é independente de mecanismos extras no uso do sistema para a detecção proposta. São dados probabilísticos uma vez que a tarefa de avaliar os sentimentos humanos são imprecisos.

A valoração dos parâmetros causais foi fundamentada em um modelo bastante explo-

rado em outras pesquisas denominado *Bayesian Knowledge Tracing* (BKT) de Corbett et al. [19] que utiliza a probabilidade de *Bayes* para estimar o conhecimento do estudante e no trabalho de Baker et al. [4] que explora as ações futuras do estudante para determinar o nível e o momento da aprendizagem.

A ideia de Baker et al. [4] é a previsão do futuro e categorização do presente através dados do passado, ou seja, utilizar os históricos das ações dos estudantes com um sistema tutorial computacional para determinar seu padrão comportamental e refinar a detecção de conhecimento.

Essas previsões mais aprimoradas foram incorporadas nas fórmulas dos parâmetros causais, e em conjunto com os atributos porcentagens de acertos, erros, dicas, oportunidades e tempo de duração, foi gerado um modelo determinístico para estimar a resposta de um estudante.

Neste capítulo, a ênfase foi na aplicação dos parâmetros causais para a previsão futura das ações do estudante, servindo de análise experimental do impacto que eles podem causar na antecipação comportamental.

Por fim, os resultados da análise evidenciaram que os parâmetros causais ajudaram significativamente na previsão das respostas positivas das atividades propostas aos estudantes, que pouco erraram quando houve a previsão de resultado negativo, porém foram deficientes na previsão da maioria dos casos de falhas na resolução das tarefas.

CAPÍTULO 5

CONCLUSÃO E TRABALHOS FUTUROS

A investigação de dados para aplicação em sistemas tutores inteligentes é uma tendência de várias pesquisas que aprimoram os modelos cognitivos como Conati et al. [18], Bica et al. [14], Barnes e Stamper [8], Paiva et al. [28], Feng, Heffernan e Koedinger [21] e, Barnes e Stamper [7].

Neste trabalho foi apresentado um modelo para a inferência das percepções causais “habilidade”, “esforço”, “sorte” e “dificuldade”. Este modelo introduziu uma nova abordagem na detecção de causas do sucesso ou fracasso em uma atividade cognitiva baseado no nível de conhecimento do estudante. O diagnóstico das causas de um evento é um elemento importante na determinação da motivação na aprendizagem. E a probabilidade de conhecimento é uma medida interessante pois pode ser reflexo das emoções e motivações envolvidas na interpretação das causas.

Portanto, a combinação da teoria psicológica com modelos de inferências formou a base teórica para a criação do modelo proposto nesta dissertação, explicando a importância da análise das causas dos eventos e seu peso no processo cognitivo, e obtendo as interpretações numéricas através de um explorado modelo probabilístico.

Esta pesquisa fornece um mecanismo de apoio para o diagnóstico dos fatores internos do estudante como emoções e motivações que colaboram para a elaboração de um esquema inteligente de tutoria.

Deve-se destacar a importância dos modelos utilizados como base deste projeto. Através do histórico de ações dos estudantes com um STI, foi calculado a probabilidade de conhecimento utilizando um modelo probabilístico. O modelo aplicado foi o BKT, apresentado no consolidado trabalho de Corbett e Anderson [19], que possui diversas variantes e pesquisas correlacionadas ([29, 40, 2, 3, 34, 30, 23, 39]) como o Baker et al. [4].

Do trabalho de Baker et al. [4], esta dissertação utilizou os métodos para o atri-

moramento, através de informações futuras, das previsões do BKT para a inferência das possíveis causas de determinada ação.

Acredita-se na validade do esforço no aprimoramento da correlação dos atributos históricos com os parâmetros causais, pois esta pesquisa contribuiu em três aspectos: na especialização dos parâmetros do modelo BKT em novos parâmetros visando aprofundar o diagnóstico dos fatores influenciadores do processo cognitivo dos estudantes; no auxílio à previsão da ação corrente através dos dados passados (utilização pré-ação); e no auxílio à elaboração de um modelo complementar para a definição de estratégias tutoriais (utilização pós-ação).

A análise de impacto na pré-ação, ou seja, antes da inserção da resposta pelo usuário, gerou um modelo determinístico aceitável, baseado nas conjecturas feitas na tabela 4.3 da subseção 4.3.3, com decisões dentro do esperado para a combinação das causas e seus efeitos.

Os resultados obtidos foram satisfatórios, com altas taxas de verdadeiros positivos (cerca de 95%) e baixas taxas de falsos negativos (cerca de 4%), porém foram insuficientes na previsão da maioria dos casos de falhas, indicado pela alta taxa de falsos positivos (cerca de 74%). Este resultado mostra a deficiência na detecção dos fatores que levam ao erro, na antecipação dos possíveis cenários oriundos da interpretação das causas e da autoeficácia pelos estudantes.

Barrera [9] salienta que apesar dos sucessos e fracassos sempre despertarem emoções agradáveis ou desagradáveis, ocasionam também, em função da interpretação de suas causas, uma variada gama de reações afetivas que afetarão o comportamento do sujeito.

A expectativa é que o modelo proposto seja mais vantajoso na aplicação pós-ação, isto é, a resposta, na etapa corrente, do estudante serve de evidência para atualização das probabilidades dos parâmetros causais para serem utilizadas na construção de um plano de ação.

5.1 Trabalhos Futuros

A colaboração pós-ação é uma proposta de trabalho futuro, visando aprimorar atividades associadas ao diagnóstico dos fatores internos do estudante, como emoções e motivações que colaboram para a elaboração de um mecanismo de definição de estratégias tutoriais.

Determinando as possíveis causas comportamentais, há a possibilidade da combinação das consequentes emoções causadas pela habilidade, esforço aplicado, sorte e dificuldade da tarefa. Por exemplo, não ter habilidade pode resultar em um sentimento de vergonha, não ter se esforçado associa-se a culpa, uma tarefa difícil pode causar desespero, porém ter sorte, conforta através da esperança. Assim uma expectativa seria gerada desta combinação de vergonha, culpa, desespero e esperança. Como a teoria da “Atribuição Causal” enfatiza a interpretação pessoal das situações, esse mapeamento pode aprimorar os parâmetros causais.

Além desse mapeamento e análise da relação entre as causas e motivação, outra tentativa de refletir mais precisamente os cenários comportamentais, suprindo as carências apontadas, é considerar a incorporação de dados de fontes externas (por exemplo, câmeras e sensores), como os trabalhos de Stein et al. [36] e Conati et al. [17, 18]. E com a inserção das novas informações no modelo, uma atitude necessária será a investigação do impacto desses valores em um STI em produção, testando, entrevistando e ministrando pós-testes.

Para amenizar a deficiência do diagnóstico dos erros outro trabalho futuro é o aprimoramento do parâmetro $P(S)$, ou o uso do modelo *Contextual-Guess-and-Slip* de Baker et al. [2] no cálculo do modelo desenvolvido nesta dissertação.

Outra proposta futura é a associação da motivação com intervenções pedagógicas, com o objetivo de torná-las mais adaptativas ao estudante. O interesse nessa especialização do apoio ao estudante é reforçado por pesquisas como Beck et al. [11], Chi et al. [16] e, Roll, Alevén, McLaren, Ryu, Baker e Koedinger [32].

BIBLIOGRAFIA

- [1] Ryan Shaun Baker, Albert T Corbett, Kenneth R Koedinger, e Angela Z Wagner. Off-task behavior in the cognitive tutor classroom: when students game the system. *Proceedings of the SIGCHI conference on Human factors in computing systems*, páginas 383–390. ACM, 2004.
- [2] Ryan SJ Baker, Albert T Corbett, e Vincent Alevan. More accurate student modeling through contextual estimation of slip and guess probabilities in bayesian knowledge tracing. *Intelligent Tutoring Systems*, páginas 406–415. Springer, 2008.
- [3] Ryan SJ Baker, Albert T Corbett, Sujith M Gowda, Angela Z Wagner, Benjamin A MacLaren, Linda R Kauffman, Aaron P Mitchell, e Stephen Giguere. Contextual slip and prediction of student performance after use of an intelligent tutor. *User Modeling, Adaptation, and Personalization*, páginas 52–63. Springer, 2010.
- [4] Ryan SJD Baker, Adam B Goldstein, e Neil T Heffernan. Detecting learning moment-by-moment. *International Journal of Artificial Intelligence in Education*, 21(1):5–25, 2011.
- [5] Albert Bandura. Human agency in social cognitive theory. *American psychologist*, 44(9):1175, 1989.
- [6] Albert Bandura. Self-efficacy: The exercise of control, 1997.
- [7] Tiffany Barnes e John C Stamper. Toward the extraction of production rules for solving logic proofs. *AIED07, 13th International Conference on Artificial Intelligence in Education, Educational Data Mining Workshop*, páginas 11–20. Citeseer, 2007.
- [8] Tiffany Barnes e John C Stamper. Automatic hint generation for logic proof tutoring using historical data. *Educational Technology & Society*, 13(1):3–12, 2010.

- [9] Sylvia Domingos Barrera. Teorias cognitivas da motivação e sua relação com o desempenho escolar. *Poésis Pedagógica*, 8(2):159–175, 2011.
- [10] Andrew G Barto. *Reinforcement learning: An introduction*. MIT press, 1998.
- [11] Joseph Beck, Beverly Park Woolf, e Carole R Beal. Advisor: A machine learning architecture for intelligent tutor construction. *AAAI/IAAI*, 2000:552–557, 2000.
- [12] Joseph E Beck e Kai-min Chang. Identifiability: A fundamental problem of student modeling. *User Modeling 2007*, páginas 137–146. Springer, 2007.
- [13] Joseph E Beck e Beverly Park Woolf. High-level student modeling with machine learning. *Intelligent tutoring systems*, páginas 584–593. Springer, 2000.
- [14] F Bica e RVRM VICARI. Projeto de um agente fuzzy para inferir a auto-eficácia do aluno no contexto de sistemas tutores inteligentes. *America*, 4(6):423–428, 2006.
- [15] John P Campbell. Pritchard. rd (1976). motivation theory in industrial and organizational psychology. *Handbook of Industrial & Organizational Psychology*, páginas 63–130, 81.
- [16] Min Chi, Kurt VanLehn, e Diane Litman. Do micro-level tutorial decisions matter: Applying reinforcement learning to induce pedagogical tutorial tactics. *Intelligent Tutoring Systems*, páginas 224–234. Springer, 2010.
- [17] Cristina Conati. Probabilistic assessment of user’s emotions in educational games. *Applied Artificial Intelligence*, 16(7-8):555–575, 2002.
- [18] Cristina Conati e Heather Maclaren. Data-driven refinement of a probabilistic model of user affect. *User modeling 2005*, páginas 40–49. Springer, 2005.
- [19] Albert T Corbett e John R Anderson. Knowledge tracing: Modeling the acquisition of procedural knowledge. *User modeling and user-adapted interaction*, 4(4):253–278, 1994.

- [20] Scotty Craig, Arthur Graesser, Jeremiah Sullins, e Barry Gholson. Affect and learning: an exploratory look into the role of affect in learning with autotutor. *Journal of educational media*, 29(3):241–250, 2004.
- [21] Mingyu Feng, Neil T Heffernan, e Kenneth Koedinger. Using data mining findings to aid searching for better cognitive models. *Intelligent Tutoring Systems*, páginas 312–314. Springer, 2010.
- [22] F. Heider. *The psychology of interpersonal relations*. Wiley, 1958.
- [23] Kenneth R Koedinger, Philip I Pavlik Jr, John C Stamper, Tristan Nixon, e Steven Ritter. Avoiding problem selection thrashing with conjunctive knowledge tracing. *EDM*, páginas 91–100, 2011.
- [24] Maria Carminda V Lagud e Ma Mercedes T Rodrigo. The affective and learning profiles of students using an intelligent tutoring system for algebra. *Intelligent Tutoring Systems*, páginas 255–263. Springer, 2010.
- [25] Ryszard S Michalski. *A theory and methodology of inductive learning*. Springer, 1983.
- [26] Hyacinth S Nwana. Intelligent tutoring systems: an overview. *Artificial Intelligence Review*, 4(4):251–277, 1990.
- [27] Andrew Ortony. *The cognitive structure of emotions*. Cambridge university press, 1990.
- [28] Ranilson Oscar Araujo Paiva, Ig Ibert Bittencourt Santa Pinto, Alan Pedro da Silva, Seiji Isotani, e Patricia Jaques. A systematic approach for providing personalized pedagogical recommendations based on educational data mining. *Intelligent Tutoring Systems*, páginas 362–367. Springer, 2014.
- [29] Zachary A Pardos e Neil T Heffernan. Modeling individualization in a bayesian networks implementation of knowledge tracing. *User Modeling, Adaptation, and Personalization*, páginas 255–266. Springer, 2010.

- [30] Zachary A Pardos e Neil T Heffernan. Kt-idem: Introducing item difficulty to the knowledge tracing model. *User Modeling, Adaption and Personalization*, páginas 243–254. Springer, 2011.
- [31] Jim Reye. Student modelling based on belief networks. *International Journal of Artificial Intelligence in Education*, 14(1):63–96, 2004.
- [32] Ido Roll, Vincent Aleven, Bruce M McLaren, Eunjeong Ryu, Ryan SJ d Baker, e Kenneth R Koedinger. The help tutor: Does metacognitive feedback improve students' help-seeking actions, skills and learning? *Intelligent Tutoring Systems*, páginas 360–369. Springer, 2006.
- [33] Stuart Russell e Peter Norvig. Artificial intelligence: A modern approach. 2009.
- [34] M Sao Pedro, R Baker, e J Gobert. Incorporating scaffolding and tutor context into bayesian knowledge tracing to predict inquiry skill acquisition. *Proc. of the 6th International Conference on Educational Data Mining, Memphis, TN*, páginas 185–192, 2013.
- [35] Richard M Steers, Richard T Mowday, e Debra L Shapiro. Introduction to special topic forum: The future of work motivation theory. *The Academy of Management Review*, páginas 379–387, 2004.
- [36] Nancy L Stein, Linda J Levine, NL Stein, B Leventhal, e T Trabasso. Making sense out of emotion: The representation and use of goal-structured knowledge. *Psychological and biological approaches to emotion*, páginas 45–73, 1990.
- [37] Bernard Weiner. An attributional theory of achievement motivation and emotion. *Psychological review*, 92(4):548, 1985.
- [38] Ian H Witten e Eibe Frank. *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann, 2005.
- [39] Yanbo Xu e Jack Mostow. Using logistic regression to trace multiple sub-skills in a dynamic bayes net. *EDM*, páginas 241–246. Citeseer, 2011.

- [40] Michael V Yudelson, Kenneth R Koedinger, e Geoffrey J Gordon. Individualized bayesian knowledge tracing models. *Artificial Intelligence in Education*, páginas 171–180. Springer, 2013.