

UNIVERSIDADE FEDERAL DO PARANÁ

THIAGO WIEZBICKI

**COMPENSAÇÃO DE DERIVA EM SENSORES DE GASES COM REDES NEURAIS
BALANCEADAS**

CURITIBA

2017

THIAGO WIEZBICKI

**COMPENSAÇÃO DE DERIVA EM SENSOES DE GASES COM REDES NEURAI
BALANCEADAS**

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia Elétrica, área de concentração de Sistemas Eletrônicos, Departamento de Energia Elétrica, Setor de Tecnologia, Universidade Federal do Paraná, como parte das exigências para a obtenção do título de Mestre em Engenharia Elétrica.

Orientador: Prof. Dr. Eduardo Parente Ribeiro

CURITIBA

2017

W652c

Wiezbicki, Thiago

Compensação de deriva em sensores de gases com redes neurais balanceadas / Thiago

Wiezbicki. – Curitiba, 2017.

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

Dissertação - Universidade Federal do Paraná, Setor de Tecnologia, Programa de Pós-Graduação em Engenharia Elétrica, 2017.

Orientador: Eduardo Parente Ribeiro.

1. Fenômeno de deriva. 2. Sensores de gases. 3. Redes neurais balanceadas. I. Universidade Federal do Paraná. II. Ribeiro, Eduardo Parente. III. Título.

CDD: 621.38



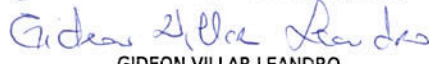
MINISTÉRIO DA EDUCAÇÃO
UNIVERSIDADE FEDERAL DO PARANÁ
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
Setor TECNOLOGIA
Programa de Pós Graduação em ENGENHARIA ELÉTRICA
Código CAPES: 40001016043P4


TERMO DE APROVAÇÃO

Os membros da Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em ENGENHARIA ELÉTRICA da Universidade Federal do Paraná foram convocados para realizar a arguição da Dissertação de Mestrado de **THIAGO WIEZBICKI**, intitulada: "**Compensação de deriva em sensores de gases com redes neurais balanceadas**", após terem inquirido o aluno e realizado a avaliação do trabalho, são de parecer pela sua APROVAÇÃO no rito de defesa. A outorga do título de mestre está sujeita à homologação pelo colegiado, ao atendimento de todas as indicações e correções solicitadas pela banca e ao pleno atendimento das demandas regimentais do Programa de Pós-Graduação.

Curitiba, 30 de Agosto de 2017.


EDUARDO PARENTE RIBEIRO
Presidente da Banca Examinadora (UFPR)


GIDEON VILLAR LEANDRO
Avaliador Interno (UFPR)


MYRIAM REGATTIERI DE BIASE DA SILVA DELGADO
Avaliador Externo (UTFPR)

AGRADECIMENTOS

Agradeço ao meu orientador, Prof. Dr. Eduardo Parente Ribeiro, pelo incentivo, contribuição e inúmeros conhecimentos repassados.

À minha futura esposa, Viviane Carvalho Reis, pelo apoio e compreensão nas horas em que estive distante para a execução desse trabalho.

Agradeço a minha família, especialmente aos meus pais, Augustinho e Olinda, pela base fornecida, sempre me incentivando e motivando nos estudos necessários.

E a todos que contribuíram de forma direta ou indiretamente para que este objetivo se concretizasse.

"O aumento do conhecimento é como uma esfera dilatando-se no espaço: quanto maior a nossa compreensão, maior o nosso contato com o desconhecido."

Blaise Pascal (1623-1662).

RESUMO

O fenômeno de deriva em sensores de gases é um dos maiores problemas encontrados em sistemas olfativos na atualidade. Fatores como contaminações por agentes externos e reações químicas na superfície sensora implicam em uma deterioração na capacidade de resposta e exatidão exigida. A ocorrência de deriva pode ser observada pelo fato da distribuição dos dados variar ao longo do tempo. A modelagem de um sistema não estacionário, que tenha como tarefa detectar essa variação contínua, imprevisível e compensando-a de maneira satisfatória, é um objetivo de grande complexidade. Neste trabalho, são aplicados e testados métodos para a compensação do fenômeno de deriva em sensores na classificação de gases com redes neurais. O primeiro consiste na ponderação das saídas de redes neurais retreinadas ao longo do tempo, onde pesos funcionam como fator de esquecimento de dados mais antigos. O segundo método caracteriza-se por evidenciar amostras mais recentes, ou seja, são replicadas entradas mais próximas temporalmente do instante atual das redes neurais na fase de treinamento. Verificou-se que a aplicação dessas técnicas na classificação de gases apresentou bons resultados comparados com estratégias mais simples como o retreino integral ou retreino parcial. No que tange a aplicação dos dois métodos usados, a replicação de amostras mostrou-se mais eficaz em lotes de dados mais antigos, sendo assim o algoritmo é mais robusto em sistemas com sensores de gases usados por longos períodos de tempo.

Palavras-chaves: Fenômeno de deriva. Sensores de gases. Redes neurais balanceadas. Auto ponderação. Ambientes não estacionários.

ABSTRACT

The drift phenomenon in gas sensors is one of the biggest problems in olphative systems nowadays. Factors as contamination by external dopants and chemical reactions in sensing surface lead to deterioration of the response and accuracy required. The drift is related to the variation of the distribution of data over time. The modeling of a non-stationary system, that has the task to detect this unpredictable continuous variation and to compensate it satisfactorily is a highly complex objective. In this work, methods are applied and tested for the compensation of the phenomenon of drift in sensors to classify gases with neural networks. The first method is the weighting of the outputs of the neural networks retrained over time, where weights are forgetting factors to older data. The second method is characterized to evidence more recent samples, therefore, the newest samples are replicated to the neural networks on training phase. It can be observed that applying the two methods on gas classification dataset showed better results than simple strategies as the integral retraining or partial retraining. Regarding the application of the two methods, the replication of the samples are more effective in batches that have older data, therefore the algorithm is more robust on systems with gas sensors used for a long time.

Key-words: Concept drift. Gas sensors. Weighted neural networks. Auto weighting. Non-stationary environments.

LISTA DE ILUSTRAÇÕES

| | |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| Figura 1 – Exemplo de sensores de gases | 14 |
| Figura 2 – Microestrutura de um sensor de gás MOX | 21 |
| Figura 3 – Antes da compensação de deriva | 23 |
| Figura 4 – Após compensação | 24 |
| Figura 5 – Tipos de deriva existentes | 25 |
| Figura 6 – Fronteira de decisão ao longo do tempo | 26 |
| Figura 7 – Modelos de deriva | 34 |
| Figura 8 – Fluxograma simplificado dos métodos de ponderação adotados . . . | 41 |
| Figura 9 – Mecanismo automatizado para coleta de dados ao longo de 36 meses | 43 |
| Figura 10 – Curvas de respostas de sensores de gases do tipo MOX. | 46 |
| Figura 11 – Análise PCA sem deriva | 47 |
| Figura 12 – Topologia generalizada de uma rede neural para todo o conjunto de gases | 48 |
| Figura 13 – Topologia de uma rede neural para a classificação de um único gás | 49 |
| Figura 14 – Quatro modelos de curvas propostas para aplicação de fatores de esquecimento nas saídas das redes neurais | 50 |
| Figura 15 – Fluxograma que contém a lógica de conjunto de dados e treinamento de cada RNA | 51 |
| Figura 16 – Fluxograma da metodologia de ponderação das saídas das RNAs . | 52 |
| Figura 17 – Modelo de condicionamento dos dados de entrada das RNAs para treinamento e classificação do lote t | 53 |
| Figura 18 – Queda na taxa de classificação correta quando se efetua o treinamento apenas com o lote 1. | 58 |
| Figura 19 – Queda na taxa de acerto com treino usando o lote 1 e também com o uso do conjunto de dados inteiro | 60 |
| Figura 20 – Comparativo entre rede neural treinada com o lote anterior, pesos quadráticos 1 e todos os lotes anteriores. | 60 |
| Figura 21 – Comparativo entre usos de pesos logarítmicos, comparadas ao uso do treinamento com o lote anterior e com todos os lotes e mesmos pesos | 61 |
| Figura 22 – Comparativo entre usos de pesos quadráticos 2 e exponenciais, comparadas ao uso do treinamento com o lote anterior e com todos os lotes e mesmos pesos | 62 |
| Figura 23 – Taxa de acerto média do uso da técnica de ponderação das entradas da RNA comparadas ao uso do treinamento com o lote anterior e com todos os lotes e mesmos pesos | 63 |

| | |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| Figura 24 – Crescimento acumulado da quantidade total de amostras ao longo do tempo | 63 |
| Figura 25 – Taxa de acerto média do uso da técnica de ponderação das entradas da RNA comparadas ao treinamento de uma RNA com todos os dados anteriores ao lote atual | 64 |
| Figura 26 – Gráfico com o resumo de todos os resultados alcançados com a topologia de múltiplos neurônios de saída | 65 |
| Figura 27 – Gráfico com o comparativo do melhor resultado alcançado (PPRA) e os três melhores resultados encontrados na literatura | 66 |
| Figura 28 – Mecanismo de utilização do algoritmo Page-Hinkley e armazenamento da detecção de deriva | 76 |
| Figura 29 – Aplicação do algoritmo Page-Hinkley no dataset concatenado | 78 |
| Figura 30 – Resultados à partir de testes com diversos classificadores usando somente o lote 1 como conjunto de dados para treinamento | 79 |

LISTA DE TABELAS

| | |
|---------------------------------------------------------------------------------------------------------------------------------------------------|----|
| Tabela 1 – Tabela comparativa de abordagens de dados e classificadores possíveis na compensação de deriva. Adaptado de: Escovedo (2015) | 35 |
| Tabela 2 – Tabela com algoritmos e valores alcançados na literatura | 39 |
| Tabela 3 – Distribuição dos dados coletados ao longo de 36 meses | 44 |
| Tabela 4 – Peso em cada conjunto de dados na entrada da rede neural | 55 |
| Tabela 5 – Tabela com o comparativo do tempo de treinamento usando diferentes algoritmos | 57 |
| Tabela 6 – Tabela com resumo de valores alcançados com todos os testes e combinações usando RNAs com múltiplos neurônios de saída | 66 |
| Tabela 7 – Tabela de seleção dos magnitude e limiar do algoritmo PHT | 77 |

LISTA DE ABREVIATURAS E SIGLAS

| | |
|-------|---------------------------------------|
| ADWIN | Adaptive Slide Window |
| AG | Algoritmo Genético |
| AM | Aprendizagem de Máquina |
| BP | Backpropagation |
| CGP | Polak-Ribière Conjugate Gradient |
| DAQ | Data Acquisition Board |
| DT | Decision Trees |
| EMA | Exponential Moving Average |
| FF | Forgetting Factors |
| IA | Inteligência Artificial |
| kNN | k-Nearest Neighbors |
| LM | Levenberg–Marquardt |
| MFC | Mass Flow Controller |
| MOX | Metal Oxide |
| MSE | Mean Squared Error |
| MLP | Multi-Layer Perceptron |
| PCA | Principal Component Analysis |
| PHT | Page-Hinkley Test |
| PPRA | Ponderação por replicação de amostras |
| RMSE | Root Mean Squared Error |
| RP | Resilient Backpropagation |
| RNA | Rede Neural Artificial |
| SCG | Scaled Conjugate Gradient |
| SVM | Support Vector Machine |

SUMÁRIO

| | | |
|----------|--------------------------------------------------------------------------|-----------|
| 1 | INTRODUÇÃO | 14 |
| 1.1 | Definição do problema | 16 |
| 1.2 | Hipóteses de pesquisa | 16 |
| 1.3 | Objetivos | 17 |
| 1.4 | Justificativa | 17 |
| 1.5 | Contribuições | 18 |
| 1.6 | Organização | 18 |
| 2 | REVISÃO | 19 |
| 2.1 | Sensores de gases | 20 |
| 2.2 | O que é o fenômeno de deriva | 21 |
| 2.3 | Compensação de deriva | 22 |
| 2.3.1 | Tipos de deriva | 24 |
| 2.3.2 | Mudança de conceito | 25 |
| 2.3.3 | Classificadores para compensação de deriva | 26 |
| 2.4 | Deteção do Fenômeno de deriva | 27 |
| 2.5 | Redes Neurais MLP | 28 |
| 2.6 | Algoritmos de treinamento para Redes Neurais | 29 |
| 2.6.1 | Levenberg–Marquardt - LM | 29 |
| 2.6.2 | Retropropagação Resiliente - RP | 30 |
| 2.6.3 | Gradiente Conjugado de Polak-Ribière-Polyak - CGP | 31 |
| 2.6.4 | Gradiente Conjugado Escalar - SCG | 32 |
| 2.7 | Comitê de classificadores | 33 |
| 2.8 | Fatores de esquecimento | 33 |
| 2.9 | Abordagem de dados e classificadores | 35 |
| 2.9.1 | Compensação de deriva em sensores de gases na literatura | 36 |
| 2.9.1.1 | Combinação de RNAs e kNN - (ADHIKARI; SAHA, 2014) | 36 |
| 2.9.1.2 | Conjunto regularizado de classificadores - (ASMITA; SHUKLA, 2015) | 37 |
| 2.9.1.3 | Conjunto de redes neurais - (KADRI <i>et al.</i>, 2013) | 38 |
| 2.9.1.4 | Demais trabalhos | 38 |
| 2.9.2 | Comparativo entre trabalhos | 38 |
| 2.10 | Considerações finais do capítulo | 39 |
| 3 | METODOLOGIA | 41 |
| 3.1 | O conjunto de dados | 42 |
| 3.1.1 | Extração de características | 45 |

| | | |
|----------|------------------------------------------------------------------------------------------------|-----------|
| 3.2 | Pré-processamento | 47 |
| 3.3 | Topologias de Redes Neurais | 48 |
| 3.4 | Fatores de esquecimento de informações | 49 |
| 3.4.1 | Fatores de esquecimento e ponderação das saídas das RNAs | 50 |
| 3.4.2 | Esquecimento de informações de entrada | 52 |
| 3.4.3 | Ponderação por replicação de amostras - PPRA | 53 |
| 4 | RESULTADOS | 56 |
| 4.1 | Treinamento | 56 |
| 4.2 | Métricas de avaliação | 57 |
| 4.3 | RNAs de saídas fragmentadas - RNAs de um neurônio de saída | 58 |
| 4.4 | Ponderação da saída da rede neural | 59 |
| 4.5 | Ponderação da entrada da rede neural | 62 |
| 4.6 | Comparativo | 64 |
| 4.7 | Comparação com melhores resultados encontrados na literatura | 66 |
| 5 | CONCLUSÃO | 69 |
| 5.1 | Contribuições | 70 |
| 5.2 | Trabalhos Futuros | 70 |
| 5.3 | Produção científica | 70 |
| | REFERÊNCIAS | 71 |
| | APÊNDICE A – DETECÇÃO DE DERIVA | 75 |
| A.0.1 | O algoritmo de Page-Hinkley | 75 |
| A.0.2 | Subdivisão dos dados por algoritmos de detecção de deriva | 76 |
| A.1 | Resultados pela aplicação do algoritmo PHT | 77 |
| | APÊNDICE B – TREINO COM APENAS O LOTE 1 | 79 |
| | APÊNDICE C – ALGORITMO - COMPENSAÇÃO DE DERIVA PELA PONDERAÇÃO DAS SAÍDAS | 80 |

1 INTRODUÇÃO

A aplicação de sistemas de medição baseados em sensores de gases é muito comum e possui amplo uso em indústrias alimentícias, químicas, controle, monitoramento, segurança pública, etc. Na Figura 1 é possível observar o exemplo de uma matriz de sensores de gases do tipo óxido metálico (MOX do inglês, *Metal Oxide*) acoplada numa câmara de passagem e medição de gases. O uso desse tipo de sensores está relacionado principalmente à detecção do tipo de substância e medição de sua concentração ao longo de todo o processo envolvido.

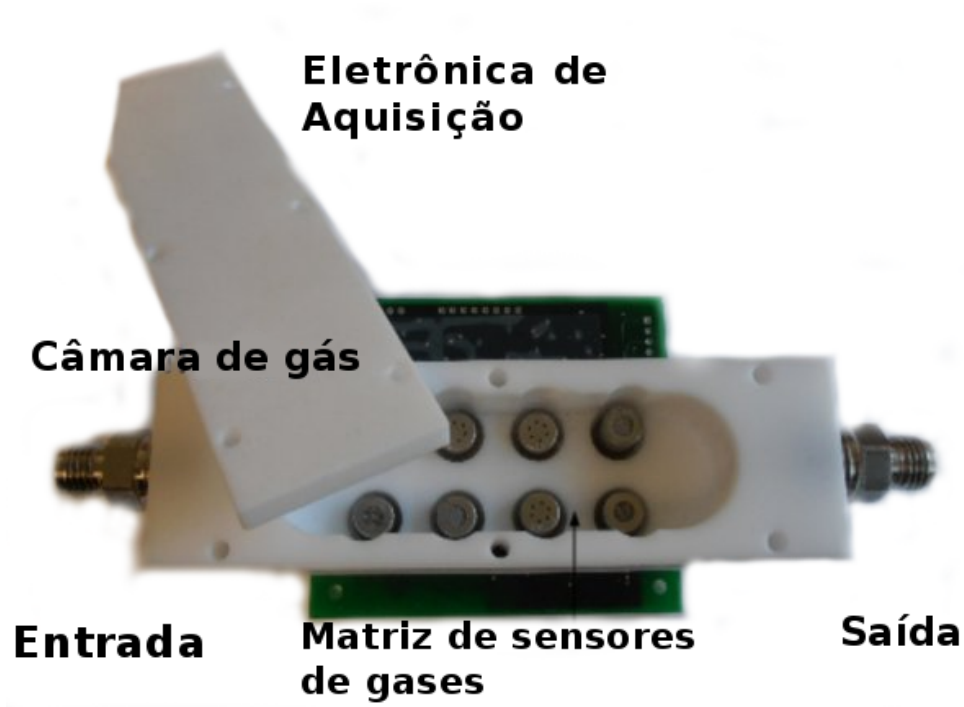


Figura 1 – Exemplo de sensores de gases

Adaptado de: Fonollosa, Vergara e Huerta (2013)

É desejável que a resposta de cada sensor para amostras de uma mesma substância seja a mesma, logo é esperado que não haja variação dessa mesma medição ao longo do tempo em que ocorre a exposição às mesmas condições anteriores. Porém em situações reais, isto não ocorre, porque sensores de gases sofrem mudanças (HOLMBERG *et al.*, 1996). O fenômeno de variação ao longo de todo esse intervalo temporal, para as mesmas condições, é conhecido como deriva (mais conhecido como *drift*, do inglês). É um dos principais problemas encontrados na área de sensoreamento voltado para detecção e mensuração de gases (ARTURSSON *et al.*, 2000).

Duas causas principais estão atreladas principalmente ao problema de deriva

em sensores de gases. A primeira se deve às interações químico e físicas em componentes que estão na microestrutura dos sensores. Os principais exemplos dessas interações são: o envelhecimento natural da superfície do sensor e a contaminação dessa estrutura por um agente externo. A segunda se deve às mudanças incontrolláveis no ambiente como por exemplo variações de temperatura e umidade (ROMAIN; NICOLAS, 2010).

Devido à natureza do problema, no qual se espera aumentar a vida útil do sensor MOX, é necessário fazer calibrações constantes e na pior das hipóteses, devido ao envelhecimento natural, efetuar a troca dos mesmos. Em ambos os casos são gerados altos custos que desencorajam essa solução simplificada. Como a atenuação de deriva não é feita, a calibração é um processo dispendioso, lento e manual. Dessa forma é necessário que sejam buscadas alternativas mais caras e complexas como por exemplo soluções ópticas (FONOLLOSA; VERGARA; HUERTA, 2013).

Algoritmos que detectam e compensam o fenômeno de deriva devem identificar pequenas mudanças, efetuar o ajuste ou compensação no sistema e serem ao mesmo tempo robustos a ruídos (GAMA, 2010).

Para reconhecer um determinado tipo de gás com o emprego de um algoritmo classificador, características de comportamento dos sensores devem ser extraídas através de um método matemático (ZLIOBAITE, 2010). Duas abordagens são comumente utilizadas para resolução desse problema: estatística, que usa probabilidade máxima *a priori*, que possui simplicidade, desempenho e resultados razoáveis (SZCZUREK; KRAWCZYK; MACIEJEWSKA, 2013) e a segunda envolvendo técnicas de inteligência artificial, nas quais algoritmos de aprendizagem de máquina podem modelar superfícies complexas com fronteiras de decisão que permitem distinguir corretamente rótulos de componentes na qual serão submetidos (GAMA, 2010).

Devido à relevância do uso de sensores de gases, além das finalidades citadas anteriormente, existem pesquisas focadas na aplicação de sensores olfativos para a detecção de doenças, um exemplo é o uso de cachorros para detecção de Mal de Parkinson (WILLIS *et al.*, 2004). Cada conjunto de moléculas em nosso corpo pode se apresentar numa maior ou menor quantidade, que pode variar em cada indivíduo, e em muitos casos a presença de uma substância em excesso pode indicar uma doença (WILLIS *et al.*, 2004). Devido a esse e outros fatores é de extrema importância o estudo de sensores de gases. Indiretamente, deseja-se obter a melhor qualidade de medição e confiabilidade na amostra ao longo do tempo, logo sempre é necessário compensar a existência de deriva nesse meio.

No que diz respeito ao desenvolvimento deste trabalho, deseja-se usar técnicas de ponderação em redes neurais com foco em aplicar fatores de esquecimento em dados que se tornam menos irrelevantes ao longo do tempo, buscando diminuir a

influência do grau de deriva ocasionado. Variados tipos de configurações de RNAs serão avaliados em conjunto. Cada subconjunto de dados utilizará uma divisão em formato de lotes, sendo que cada um apresenta uma relação temporal ao longo do processo de coleta dos dados. Fatores de ponderação e esquecimento são aplicados com a finalidade de tornar dados mais relevantes que outros de maneira dinâmica.

1.1 DEFINIÇÃO DO PROBLEMA

A construção e operação de sensores de gases estão ligadas intrinsecamente a fatores químicos, e apresentam a ocorrência de deriva. Nem sempre é possível prever com antecedência a presença desse problema pois suas características são variáveis. O alto custo, atrelado à troca precoce dos sensores de gases e o alto tempo despendido em efetuar calibrações de forma manual e muitas vezes imprecisa, torna o processo de uso desses equipamentos inviável. Sendo assim a melhora da vida útil dos sensores é altamente desejável.

Para mitigação deste problema geralmente são adotadas soluções que se baseiam em: métodos estatísticos que envolvem a monitoração e gravação dos dados ao longo de toda a vida útil do equipamento e pós análise probabilística do fenômeno. Estas abordagens podem apresentar resultados aceitáveis, porém o tempo para se obter um sistema desse tipo é longo. Outros métodos se baseiam na construção de classificadores: DTs (Árvores de decisão, do inglês *Decision Trees*), RNAs (Redes Neurais Artificiais), SVMs (Máquinas de Vetor de Suporte, do inglês *Support Vector Machines*) que são massivamente combinados em grande quantidade para que com a especialidade de cada um, atinjam o melhor resultado possível.

A última abordagem é a mais adotada, porém ela não valida ou leva em consideração o grau de importância de dados que sofreram do problema de deriva ao longo de todo o tempo. Usualmente são usados todos os dados coletados no passado ou somente num passado menos distante. Isso na maioria das vezes acaba prejudicando a taxa de acerto do sistema ao longo do tempo.

1.2 HIPÓTESES DE PESQUISA

A hipótese desse trabalho consiste em verificar se a aplicação de RNAs com fatores de ponderação e esquecimento (FF, do inglês *Forgetting Factors*) em matrizes de sensores de gases, é capaz de compensar de maneira adequada a deriva com o aumento da correta taxa de classificação de cada um dos rótulos. Duas abordagens serão avaliadas: a primeira consiste em ponderar todos os neurônios das camadas de saídas da RNAs treinadas com conjuntos de dados passados, na segunda abordagem, serão atribuídos graus de importância diretamente às amostras que treinarão a RNA

para classificação. Estima-se que possam existir melhoras na taxa de classificação comparada a métodos tradicionais. Deseja-se verificar de qual maneira de uso de FFs e tratamento de deriva é mais eficaz e produz melhorias de forma estável e consistente ao passar do tempo.

1.3 OBJETIVOS

Como objetivo geral dessa pesquisa, deseja-se verificar a eficácia de modelo de compensação de deriva em sensores na detecção de gases com uso de classificadores, sendo os objetivos específicos a serem alcançados:

- Analisar um conjunto de dados de sensores de gases disponível na literatura;
- Desenvolver pré-processamento e condicionamento do conjunto de dados;
- Confirmar a ocorrência de deriva através de testes com classificadores;
- Analisar mecanismos de ponderações baseados em fatores de esquecimentos nas saídas das RNAs;
- Desenvolver e analisar um método de ponderação através do condicionamento dos dados antes do treinamento da RNA;
- Validar a eficácia dos dois métodos propostos e testados, usando critérios de testes e avaliação de resultados obtidos, verificando oportunidades de possíveis melhorias.

Dessa forma espera-se um aumento da vida útil dos sensores de gases, redução de custos relativos à calibração constante e manual e principalmente melhoria na taxa de acerto na detecção de um variado tipo de amostra ao longo do tempo, de modo que garanta um acréscimo na qualidade do processo.

1.4 JUSTIFICATIVA

Sistemas adaptativos e evolutivos são cada vez mais usados na resolução de problemas que sofrem alguma variação em decorrência do tempo. Técnicas de inteligência artificial têm se tornado comuns em problemas onde a grande quantidade de variáveis envolvidas torna a modelagem matemática do problema complexa. Dentro do contexto de aplicação de métodos para sensores de gases é muito comum que sejam combinadas técnicas para que a resolução do problema de deriva seja mais eficaz. Deseja-se dessa forma estabelecer uma abordagem que leve em consideração o grau de importância dos dados ao longo do tempo, utilize métodos de compensação

da deriva e possa usar uma menor quantidade de classificadores para obter resultados iguais ou melhores que encontrados na literatura.

1.5 CONTRIBUIÇÕES

No desenvolvimento desse trabalho é possível destacar que são obtidas contribuições para a área de sensoriamento olfativo:

- Inicialmente na aplicação de fatores de esquecimento ao longo de todas as saídas de redes neurais, ao contrário da maioria de métodos que utiliza uma janela de tamanho fixo temporalmente, estes pesos são aplicados ao longo de todo o conjunto de dados, sem ignorar dados que possam ter importância menor na classificação presente. Não foi encontrado durante a pesquisa a aplicação de tal método na área de sensoriamento olfativo, logo é uma abordagem nova.
- A segunda contribuição, cuja inovação é apresentada no problema de compensação de deriva em sensores de gases, condiz com abordagem de dar ênfase às amostras de sensores de gases de passado mais recente no treinamento de uma rede neural. Dessa forma são aplicados fatores de esquecimento diretamente na distribuição dos dados ao longo do tempo, variando a quantidade de amostras que existem e são "criadas" pelo sistema de ponderação dinamicamente.

1.6 ORGANIZAÇÃO

Este trabalho é dividido em cinco capítulos. No segundo capítulo, denominado *Revisão*, é fornecida uma descrição geral sobre o problema de deriva em sensores de gases ao longo da literatura, avaliando o estado da arte do problema e métodos comumente empregados. São apresentados em conceitos necessários para entendimento da temática relativa ao problema, o que é a deriva e como ela está presente em sensores de gases e que ferramentas são usadas e necessárias para a sua atenuação. No capítulo *Metodologia*, são descritas técnicas, métodos, algoritmos e critérios de avaliação que são usados neste trabalho para resolução do problema e da nova abordagem proposta. Em *Resultados*, para cada tipo de algoritmo aplicado no conjunto de dados, são mostradas tabelas, gráficos e comparativos dos números alcançados, em seguida são discutidos cada um dos resultados analisando vantagens e desvantagens de cada. No capítulo *Conclusão*, são sumarizados os principais resultados obtidos e são apresentadas sugestões para trabalhos futuros.

2 REVISÃO

Sistemas baseados em narizes eletrônicos (do inglês, *E-nose system*) têm a função de imitar funções de sensibilidades olfativas biológicas (ZHANG *et al.*, 2012). São usados na maioria das vezes matrizes de sensores especializados na detecção de determinados tipos de odores, por sua vez estes são caracterizados pela passagem de gases característicos sob esse conjunto, que conforme o mecanismo de construção de cada sensor, possui menor ou maior resposta à dada substância (YAMAZOE; SHIMANOE, 2009).

É notável o grande leque de aplicações industriais que se aproveitam do uso dessas tecnologias (HU *et al.*, 2014). É possível utilizar em aplicações farmacêuticas, tanto na detecção de substâncias em máquinas de produção em série de medicamentos, quanto no processo de pesquisa e testes de componentes químicos. Indústrias alimentícias, na fabricação de seus produtos, utilizam a medição de gases para verificar a qualidade e confiabilidade de seu processo produtivo na análise de substâncias de acordo com um padrão pré estabelecido (SHEIK *et al.*, 2014). Nesse último caso o interesse está não somente na detecção do tipo de substância presente no ar, mas também em sua respectiva concentração. Demais aplicações de grande importância podem envolver por exemplo casos de segurança pública, onde em ambientes críticos podem ser empregados sensores desse tipo para a detecção de substâncias que apresentem toxicidade e possam causar danos à saúde de seres humanos e animais (LIU; TANG, 2013).

Nem sempre o uso de uma matriz de sensores de gases, cuja tarefa é a detecção das mais variadas substâncias, possui implementação direta e livre de calibrações constantes. Existe um envelhecimento natural desses sensores, causado principalmente por fatores químicos envolvidos tanto na natureza de fabricação desses como no ambiente à que suas operações são submetidas. A passagem contínua dos mais diversos agentes químicos, durante grandes intervalos de tempo, sob variações de temperatura, pressão, umidade, entre outros fatores externos, provocam um desgaste físico/químico dessa matriz de sensoriamento (LLOBET *et al.*, 1997). Dessa maneira o desgaste implica numa variação na resposta dos sensores ao longo do tempo, logo, o grau de confiabilidade na mensuração não permanece o mesmo conforme há aproximação do fim do seu ciclo de vida.

O envelhecimento natural ou desgaste infelizmente não é um fenômeno previsível, de fácil modelagem ou possível de se reproduzir da mesma forma em outros sistemas que usem os mesmos sensores. Cada aplicação possui seu grau de especificidade, variabilidade e deterioração. Devido a isto, se essa variação ocorre de forma

contínua, imprevisível e de difícil compreensão, constata-se a ocorrência do fenômeno de Deriva de Conceito (do inglês *Concept Drift*).

2.1 SENSORES DE GASES

Dentro do grupo de sensores de gases, destaca-se como o grupo de maior estudo e investigação o tipo baseado em semicondutor de óxido metálico (MOX, do inglês *Metal Oxide*). Estes destacam-se no meio científico em sensoriamento sobre condições atmosféricas devido: simplicidade de uso, grande quantidade de gases detectáveis possibilitando muitos campos de aplicação e principalmente ao baixo custo e flexibilidade atrelada à sua produção (WILLIAMS, 1999). O uso desse tipo de sensoriamento teve como início às pesquisas sobre efeitos de reações química sobre óxidos metálicos de Heiland, Bielanski *et al.* e Seiyama *et al.* e quando Taguchi efetuou a primeira venda desses sensores baseados em semicondutores (BARSAN; KOZIEJ; WEIMAR, 2007). Fabricantes na qual podem-se adquirir sensores são: *Figaro, FIS, MICS, UST, CityTech, Applied-Sensors, NewCosmos, etc.*

Um sensor de gás do tipo MOX é composto basicamente de:

- Camada sensível;
- Substrato;
- Eletrodos que medem as características elétricas;
- Aquecedor, que é totalmente separado da camada de sensibilidade e dos eletrodos por uma camada de isolamento (BARSAN; KOZIEJ; WEIMAR, 2007).

O dispositivo na maioria das vezes opera com ar na presença de umidade e outros gases residuais. Nessas situações as temperaturas de trabalho giram em torno de 200 a 400°C na superfície do óxido metálico. Nem todos os gases descritos têm influência direta sobre ou reagem com o material de superfície devido à barreiras existentes (SAHM *et al.*, 2006).

A Figura 2 mostra no nível microscópico o funcionamento de um sensor de gás do tipo MOX. Esta figura mostra apenas uma das várias microestruturas de eletrodos que são encontradas ao longo da superfície sensível.

Primeiramente com o aquecimento do substrato, há o depósito de partículas de gás na camada sensível do substrato. Logo na aplicação de uma corrente elétrica i , devido à resistência ΔR , é possível observar uma determinada tensão V sob os polos dos eletrodos. Cada substrato, eletrodos e as partículas de gás que se depositam apresentam uma característica diferente, logo serão medidas resistências elétricas diferenciadas para cada tipo de material.

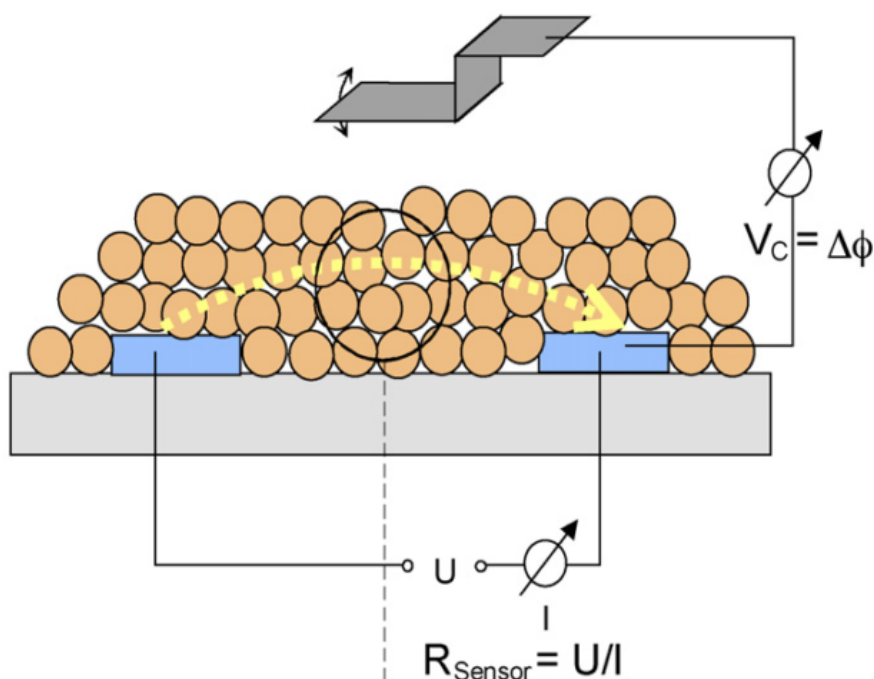


Figura 2 – Microestrutura de um sensor de gás MOX

Adaptado de: BARSAN, KOZIEJ e WEIMAR (2007)

É importante notar que a interpretação acima é para apenas um par de eletrodos em toda a sua microestrutura, para verificar em escala global o nível de contribuição $\Delta\chi$, para o sensor como um todo, há a relação

$$\Delta\chi = \Delta\Phi - \kappa_B T \ln \left(\frac{R_F}{R_I} \right) \quad (2.1)$$

onde Φ é a função de trabalho, indicando o grau de afinidade eletrônica do substrato de cada material, κ_B é o potencial de rompimento da barreira eletrônica entre eletrodos, T é a temperatura da superfície do substrato, R_F e R_I são respectivamente as resistências medidas antes e depois das alterações de ambiente em função das medições efetuadas (BARSAN; KOZIEJ; WEIMAR, 2007).

2.2 O QUE É O FENÔMENO DE DERIVA

Em muitas aplicações há a presença de ambientes dinâmicos onde há um fluxo de dados contínuo. Se um sistema não é estacionário, é dito que há uma mudança de conceito, ou na maioria dos casos, ocorrência de deriva.

Conforme a equação 2.2, se a função de densidade de probabilidade F_{x^t} de uma amostra x contida em um conjunto C_k for igual a mesma função de densidade de

probabilidade aplicada à uma amostra x no intervalo de tempo $t + \delta$ que pertence ao mesmo espaço de amostras C_k , é dito que não houve variação na distribuição média dos dados ao longo do tempo, logo não houve deriva e o ambiente é estacionário.

$$Fx^t(x|C_k) = Fx^{t+\delta}(x|C_k) \quad (2.2)$$

Porém se o caso contrário acontece, conforme a equação 2.3, tem-se o denominado problema de deriva (GAMA, 2010).

$$Fx^t(x|C_k) \neq Fx^{t+\delta}(x|C_k) \quad \forall x \in C_k \quad (2.3)$$

Conforme descrito na equação 2.3, quando há variação na distribuição das amostras ao longo do tempo, o sistema é dito não estacionário.

2.3 COMPENSAÇÃO DE DERIVA

É importante que se for constatada a presença de deriva em sensores de gases, esta seja atenuada. A Figura 3 mostra através da aplicação do método de análise por componentes principais (PCA, do inglês, *Principal Component Analysis*) numa rede de sensores de gases, características de dispersão ao longo do espaço do conjunto de amostras conforme a ocorrência da deriva se torna cada vez mais forte. A seta nesta mesma Figura, indica o gradiente da direção espacial de maior concentração das amostras. Se a compensação de deriva não for adotada, é possível concluir que ocorra o crescimento de um espalhamento multidimensional, ao longo de todos os eixos das características, a ponto de provocar uma fusão dos conjuntos, impossibilitando a diferenciação, ou seja, a classificação ao longo do tempo. Traçar uma fronteira de decisão se torna uma atividade de grande complexidade (GUTIERREZ-OSUNA, 2000).

A Figura 4, porém mostra, com a mesma aplicação de PCA, que se algum método de compensação for adotado na tentativa de retardar, ou na melhor das hipóteses, regredir a incidência de deriva, é possível obter conjuntos que definem tipos de gases distintos, sobre o mesmo subespaço de maneira mais concentrada, logo a aplicação de um classificador torna a tarefa de determinar o tipo de gás e sua avaliação mais fácil (GUTIERREZ-OSUNA, 2000). Dessa forma, compensar a deriva significa manter as classes mais separadas o possível.

É correto dizer dessa forma que a Figura 3 apresenta um grau de deriva muito mais elevado que os dados apresentados na Figura 4, cujo método de compensação foi adotado.

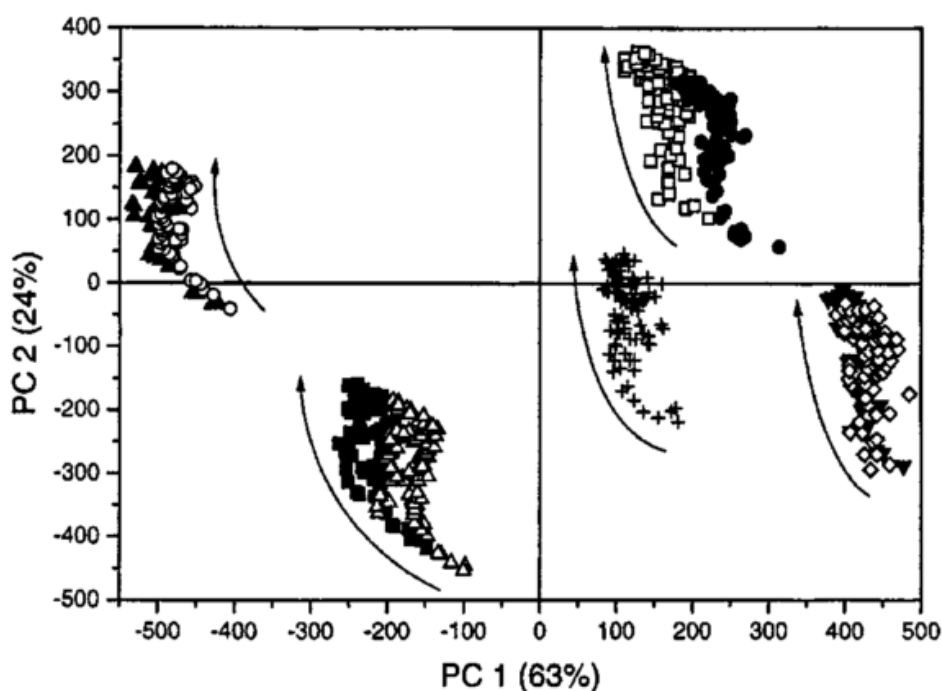


Figura 3 – Antes da compensação de deriva

Adaptado de: Gutierrez-Osuna (2000)

Se um sistema apresenta a ocorrência de deriva não compensada, ao longo do tempo t :

$$\Psi_{final} > \Psi_{inicial} \quad (2.4)$$

Onde $\Psi_{inicial}$ é o grau de deriva no instante inicial, ou nesse caso, quando os sensores de gases são novos e Ψ_{final} é o grau de mudança de conceito num momento após o início das medições.

Compensar a ocorrência de deriva conforme descrito anteriormente, significa diminuir o efeito de alteração (deriva) na função de densidade probabilidade $F_{x^t}(x)$, ou mais especificamente, atenuar a amplitude em que a média das amostras sofrem variação ao longo do tempo, estando de acordo com a equivalência 2.4.

Dentro do contexto que relata a necessidade de se compensar a deriva em sensores de gases, a natureza química e o processo natural de envelhecimento, é necessário fazer com que distribuição de amostras desses sensores tenham a mínima variação de acordo com o aumento do tempo de vida dos mesmos.

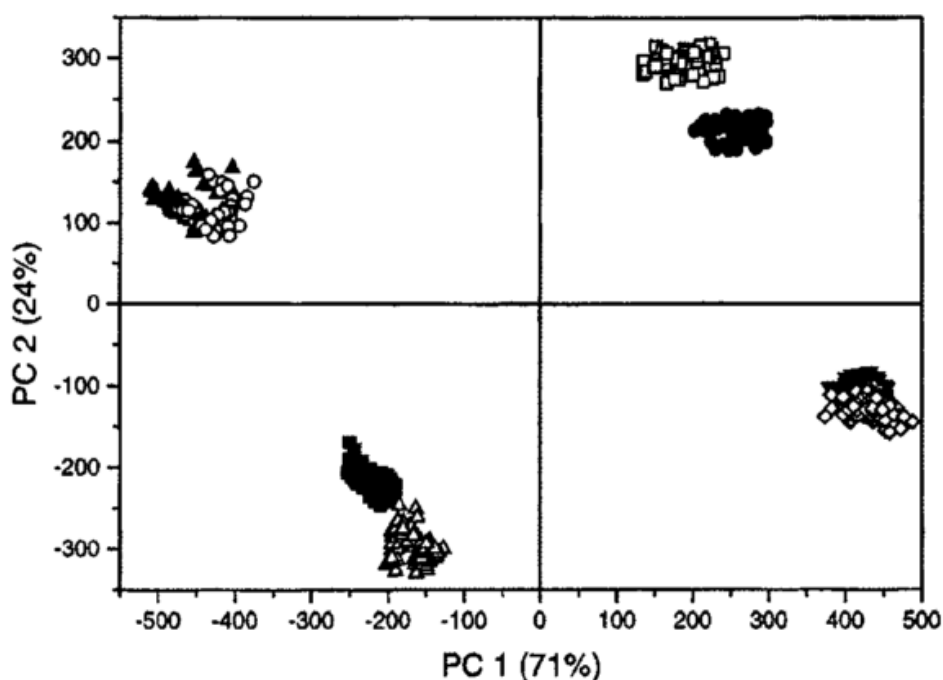


Figura 4 – Após compensação

Adaptado de: Gutierrez-Osuna (2000)

2.3.1 Tipos de deriva

Dentro do contexto que tange a ocorrência de deriva, é possível notar diferentes tipos desse fenômeno (ZLIOBAITE, 2010). A Figura 5 descreve de forma gráfica e distingue quatro tipos de ocorrências: Abrupta, Gradual, Incremental, Contexto e Recorrente.

A primeira forma, abrupta, ocorre quando dentro de uma distribuição média das amostras ao longo do tempo, há uma variação grande, de alta frequência, de maneira imediata, de forma que a nova média, é superior ou inferior à antiga, com grande destaque, porém que a distribuição dos novos dados na ocorrência da deriva abrupta será muito próxima à nova média, sem retorno ao valor antigo. Muitas das vezes esse fenômeno é conhecido como *Concept Shift* (GAMA *et al.*, 2004).

Na forma de deriva gradual, a mudança de conceito, ou seja, a ocorrência de deriva, se dá de maneira progressiva, onde gradualmente e constantemente algumas amostras apresentarão valores diferentes da média atual, porém com o passar do tempo, a quantidade de amostras que irão apresentar essa nova média irá aumentar, caracterizando uma deriva gradual.

No que diz respeito à mudança incremental, há a variação na média de forma crescente ou decrescente ao longo do tempo, porém com a característica que a variação média das amostras se dê em passos contínuos e no mesmo sentido. Se a média está

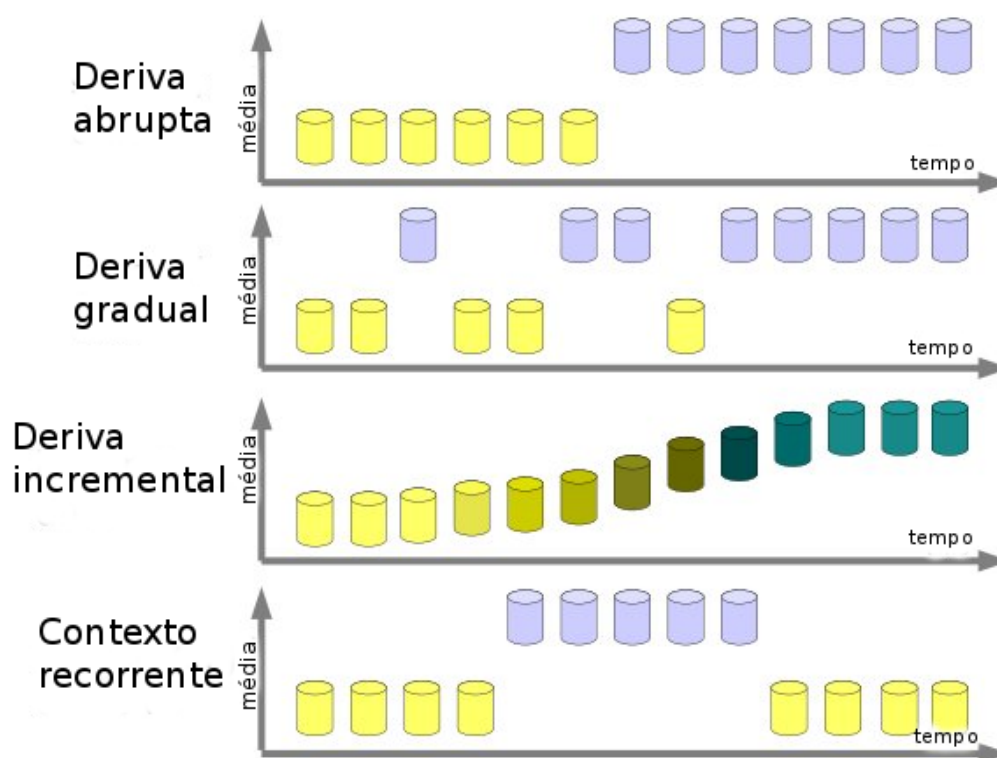


Figura 5 – Tipos de deriva existentes

Adaptado de: Zliobaite (2010)

aumentando, logo as próximas incidências de deriva irão apresentar aumento médio positivo e vice-versa. Após esse processo de variação incremental a tendência é de estabilização.

Para o tipo contexto recorrente, há variações abruptas ao longo do tempo, porém quase sempre com retorno ao contexto anterior e assim sucessivamente.

2.3.2 Mudança de conceito

A Figura 6 traz o que é conhecido como mudança de conceito. Além de haver variações na distribuição das amostras ao longo do tempo na ocorrência de deriva, ocorre a variação na fronteira de decisão que discrimina cada rótulo de gás ao longo do tempo. A Figura serve apenas para fins ilustrativos, não caracteriza-se o conjunto de dados utilizados nesse trabalho, cuja fronteira de decisão não é linearmente separável.

Dados os conjuntos de dados $C1$ e $C2$, simbolizando o conjunto de dados no instante inicial $t1$ e no intervalo futuro $t2$, onde com o passar do tempo amostras com os rótulos A e B respectivamente sofrem do fenômeno de deriva. Nessas circunstâncias, a fronteira de decisão representada de maneira bidimensional através da Figura 6 não apresenta semelhança alguma entre os intervalos de tempo onde houve um acréscimo

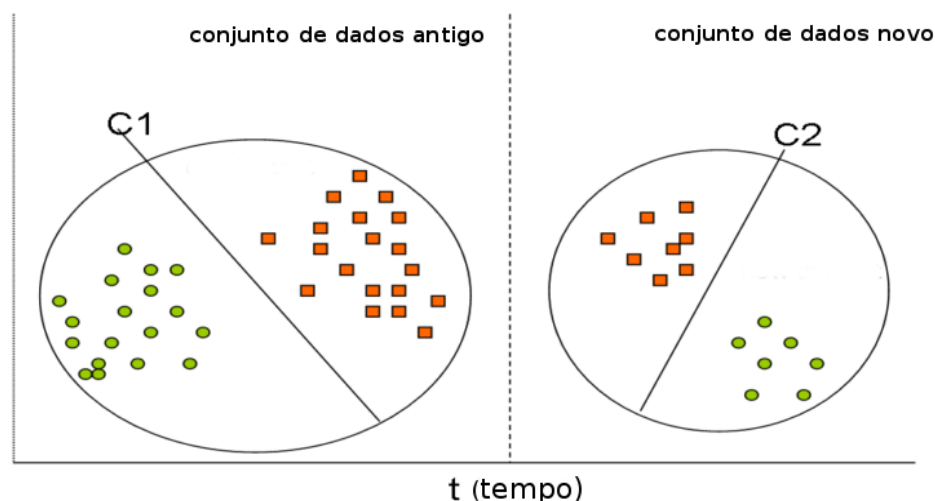


Figura 6 – Fronteira de decisão ao longo do tempo

Fonte: Wang *et al.* (2010)

ou decaimento da deriva nesse conjunto de dados. Logo, podemos dizer que para esse conjunto de dados houve uma mudança de conceito.

Ainda é importante destacar que a figura 6 traz de maneira ilustrativa uma fronteira de decisão que se altera ao longo do tempo, mais conhecida como mudança de conceito. A linearidade vista nessa figura é apenas para fins didáticos, de forma que uma fronteira de decisão do mundo real não é uma reta ou simples curva que pode ser equacionável de maneira direta. Classificadores podem modelar superfícies de decisão complexas, independentemente da quantidade de dimensões.

2.3.3 Classificadores para compensação de deriva

Abordagens comuns que podem ser adotadas na resolução do problema consistem em: estatística e uso de inteligência artificial. A primeira usa a máxima probabilidade a priori de acontecimento do fenômeno, a qual consiste em analisar toda a ocorrência temporal dos dados e traçar curvas de probabilidade. Essa abordagem é simples, apresenta tempo computacional baixo e resultados intermediários (BERMAK *et al.*, 2005). A segunda envolve o uso de inteligência artificial, mais precisamente técnicas de aprendizagem de máquina, nas quais algoritmos classificadores modelam de forma matemática, complexas superfícies de decisão, com grau de generalização visando de obter resultados precisos na classificação (GAMA *et al.*, 2004). A grande vantagem é que mecanismos de aprendizagem de máquinas permitem, com base em treinamentos, modelar de forma complexa e por aproximação, um fenômeno com grande quantidade de variáveis envolvidas, tal como o fenômeno de deriva na classificação de gases, mapeando o espaço de entrada no de saída.

Para compensação de deriva e eficaz classificação em ambientes não estacionários espera-se de um classificador (SCHLIMMER; GRANGER, 1986):

- Detectar através de monitoramento contínuo qualquer mudança na distribuição do conjunto de dados;
- Esquecer o que foi aprendido quando o conhecimento não for mais útil para a classificação de dados que sofreram mudanças;
- Demonstrar aprendizado à partir da nova presença de dados sem a obrigatoriedade de apresentar todo o conjunto de dados para o classificador;
- Alterar os seus próprios parâmetros de maneira dinâmica, compensando alterações que ocorrem na base de dados.

Grande parte dos trabalhos publicados na área de ambientes não estacionários que apresentam a ocorrência de deriva e conseqüentemente abordam o uso de classificadores, são da última década (ESCOVEDO, 2015). Na literatura o termo mais comum é *concept drift*, porém na tradução para a língua portuguesa, decidiu-se usar o termo deriva de conceito.

2.4 DETECÇÃO DO FENÔMENO DE DERIVA

Dentro da necessidade de se compensar ou melhorar os efeitos que o problema de deriva implica em uma rede de sensores de gases, inicialmente é necessário analisar o conjunto de dados e verificar se há mudança de conceito, ou seja, se é observado que os dados se comportam de maneira não estacionária (GAMA, 2010).

Para a aplicação de algoritmos que façam a detecção de deriva é adotada a abordagem de janelas de dados. Em cada etapa o algoritmo adota uma janela fixa que utiliza somente amostras nela contidas. O maior desafio, no entanto é selecionar o tamanho ideal da janela: uma janela pequena pode possibilitar rápida resposta e adaptabilidade do sistema, porém períodos mais estáveis pode prejudicar o seu desempenho. Em contraposto, janelas grandes podem possuir resultados bons e estáveis em períodos não oscilatórios, mas o sistema não consegue reagir às mudanças rápidas de conceito. Dessa forma parâmetros de performance, distribuição dos dados e propriedades dos dados devem ser monitorados o tempo todo (GAMA *et al.*, 2004).

Durante todo o processo de análise da deriva é comum adotar dois tipos de análises: de dados passados em uma janela de referência e o uso de uma janela atual com os pontos mais recentes. Quando a janela é adaptativa é necessário ajustar sua extensão de acordo com a magnitude da deriva, normalmente se a ocorrência de deriva

é detectada, a janela diminui, em contraposto, a janela aumenta (HULTEN; SPENCER; DOMINGOS, 2001).

Algoritmos conhecidos para a detecção de deriva: Janelamento Adaptativo (ADWIN, do inglês *Adaptive WINDdowing*), Teste de Page-Hinkley (PHT, do inglês *Page-Hinkley test*), Modelo de Janela Fixa Cumulativa (FCWM, do inglês *Fixed Cumulative Windows Model*), etc (GAMA *et al.*, 2004).

2.5 REDES NEURAIAS MLP

Redes Neurais artificiais, são ferramentas padrão na elaboração de superfícies de decisão complexas, não polinomiais e irregulares. Caracterizam-se como ferramentas que propiciam o estabelecimento de um mapeamento de entrada desconhecida à uma saída, cuja relação é dada através de transformações lineares que são "aprendidas" com base numa função de treinamento. Devido à esse comportamento, as redes neurais são conhecidas como aproximadores universais, cuja capacidade de aproximar qualquer função com o desejável grau de precisão.

Entre dos mais variados tipos de RNAs, destaca-se a do tipo Perceptron de Múltiplas Camadas (MLP, do inglês *Multi-Layer Perceptron*). Esse tipo de RNA foi escolhida nesse trabalho devido à sua simplicidade de implementação e boa capacidade de classificação num problema sem deriva. Como objetivo do trabalho, não há foco em criar um novo classificador na finalidade de obter-se a máxima taxa de acerto, com todos os meios possíveis. Foca-se, em estudar o fenômeno de deriva, ações podem ser adotadas além do classificador e ações possíveis de serem aplicadas no conjunto de dados para atenuar esse problema.

RNAs são úteis na aplicação de resolução de problemas do mundo real quando não há presença de um modelo matemático bem definido. Na última década é claro o aumento do uso dessas ferramentas para análise de dados provenientes de uma matriz de sensores (PARDO; SBERVEGLIERI, 2004).

Dada a Equação 2.5, é possível com a construção de RNAs de camadas do tipo MLP, com retropropagação do erro, a obtenção de um aproximador universal (HORNIK; STINCHCOMBE; WHITE, 1989) pela relação

$$\Phi = \sum_{j=1}^N \alpha_j \sigma (y_j^T x + \theta_j) \quad (2.5)$$

onde $y_j \in \mathbb{R}^n$, $\alpha_j, \theta_j \in \mathbb{R}$, são densidades no espaço de funções contínuas no cubo unitário se σ é uma função sigmoideal contínua, logo Φ é o valor da função de aproximação capaz de projetar uma fronteira de decisão para classificação de uma amostra x (HORNIK; STINCHCOMBE; WHITE, 1989).

Nesse trabalho serão usados apenas classificadores baseados em RNAs do tipo MLP, visto que não se deseja modificar a estrutura de um classificador padrão, ou na pior das hipóteses efetuar a combinação extensa de diversos classificadores com a única finalidade de aumentar a taxa de classificação sem a preocupação da natureza do fenômeno.

2.6 ALGORITMOS DE TREINAMENTO PARA REDES NEURAIS

Uma rede neural consiste basicamente em n unidades (neurônios) na camada de entrada, m unidades de processamento (neurônios da camada intermediária) que podem ser "treinadas" para aproximar um limitado tipo de função $f : R^n \rightarrow R^m$. A partir do momento que a rede neural é alimentada com novas amostras do vetor $x \in R^n$ e seu mapeamento correto $f(x)$, um algoritmo de aprendizado é aplicado para ajustar os pesos e os limiares (*biases*) através de uma função de gradiente que dá a direção que minimiza a diferença entre $f(x)$ e as k unidades (neurônios) da camada de saída da rede neural artificial (LESHNO *et al.*, 1993).

A seguir será abordada uma breve teoria à respeito das quatro funções de treinamento que foram utilizadas ou testadas ao longo desse trabalho. A escolha dessas funções é dada através de simulações de RNAs MLP treinadas com as mesmas, de forma que o número se restringiu à 4, pois foram as que possibilitaram resultados dentro do objetivo para uma rede neural com uma ampla gama de características de entrada. Para maiores detalhes de funcionamento de cada um dos algoritmos, as referências poderão ser consultadas.

2.6.1 Levenberg–Marquardt - LM

Dada a função $F : R^n \rightarrow R^m$, continuamente diferenciável, e o algoritmo de Levenberg (1944) e Marquardt (1963) é um otimizador para resolução de soluções numéricas, onde a função principal é dada por:

$$\phi(x) = \frac{1}{2} \sum_{i=1}^m f_i^2 = \frac{1}{2} \|F(x)\|^2 \quad (2.6)$$

de modo que o algoritmo é capaz de localizar os mínimos locais de uma função não linear de mínimos quadrados conforme mostrado acima (MORÉ, 1978).

O algoritmo de Levenberg–Marquardt (LM), ficou conhecido e se tornou a forma padrão de resolução de funções não lineares com mínimos quadrados. O algoritmo pode ser considerado como uma combinação do método de Descida do Gradiente em conjunto com o método de Gauss-Newton (LOURAKIS, 2005).

Dada a combinação de métodos que o algoritmo LM utiliza, quando a solução está muito distante o comportamento é dado pelo método de Descida do Gradiente, que é lento e garante a convergência. Quando o erro é pequeno o algoritmo comuta para o comportamento através do método de Gauss-Newton.

Dado a função f que mapeia um vetor de parametrização $p \in \mathbb{R}^m$ para estimar um vetor de medição $\hat{x} = f(p)$, $\hat{x} \in \mathbb{R}^n$ inicia-se o algoritmo com um parâmetro inicial estimado p_0 e medidos através do vetor x que são providos e designados à encontrar o vetor p^+ que melhor satisfaça a função f , de forma que minimize a distância quadrada $\epsilon^T \epsilon$ com $\epsilon = x - \hat{x}$ (LOURAKIS, 2005).

A base do algoritmo LM consiste numa aproximação linear f na vizinhança de p , sendo:

$$f(p + \delta_p) \approx f(p) + J\delta_p, \quad (2.7)$$

onde J é a matriz Jacobiana $\frac{\partial f(p)}{\partial p}$. Para um pequeno $\|\delta_p\|$ a aproximação é dada por uma série de Taylor. Sendo um algoritmo de aproximação, que é iterativo, são produzidos uma série de vetores $p_0, p_1, p_2 \dots$ que convergem num mínimo local p^+ para f (LOURAKIS, 2005).

Devido à natureza desse algoritmo com a finalidade de aproximação para localização de máximos e mínimos, o mesmo pode ser "adaptado" para uso na fase de treinamento de RNAs.

Apesar de o uso desse algoritmo apresentar bons resultados no início desse trabalho, o uso do método de Descida do Gradiente, cálculo da matriz Jacobiana, expansão por Séries de Taylor (usa cálculos de derivas de n ordens) atrelada à grande quantidade de neurônios, inviabilizou a fase de treinamento, deixando-o com tempos exaustivamente altos.

2.6.2 Retropropagação Resiliente - RP

O algoritmo de Retropropagação Resiliente do erro (RP, do inglês *Resilient Backpropagation*), consiste em melhorar as funções de ajustes de pesos de forma localizada de acordo com a função de erro (RIEDMILLER; BRAUN, 1993). Dessa forma busca-se ganhos de velocidade em durante a fase de treinamento da RNA.

O funcionamento desse algoritmo consiste na inserção de um valor Δ_{ij} em cada neurônio que determina isoladamente a magnitude da atualização de cada um dos pesos. Esse valor de adaptação envolve durante a fase de aprendizado a assinatura

local da função de erro E de acordo:

$$\Delta_{ij}^{(t)} = \begin{cases} \eta^+ * \Delta_{ij}^{t-1}, & \text{se } \frac{\partial E}{w_{ij}}^{(t-1)} * \frac{\partial E}{\partial w_{ij}}^{(t)} > 0 \\ \eta^- * \Delta_{ij}^{t-1}, & \text{se } \frac{\partial E}{w_{ij}}^{(t-1)} * \frac{\partial E}{\partial w_{ij}}^{(t)} < 0 \\ \Delta_{ij}^{t-1}, & \text{caso contrário} \end{cases} \quad \text{onde } 0 < \eta^- < 1 < \eta^+ \quad (2.8)$$

de forma que toda vez que a derivada parcial do peso correspondente w_{ij} muda o sinal, onde indica que a magnitude da atualização ultrapassou o limiar desejado e o algoritmo alcançou o mínimo local, o valor de atualização Δ_{ij} sofre uma subtração do fator η^- . Se não houver mudança de sinal, o valor de atualização sofre um "leve" aumento de valor com a intenção de acelerar a convergência em regiões de mínimo global.

Uma vez que o valor de atualização foi estabelecido, os pesos dos neurônios são atualizados com a seguinte regra:

$$\Delta w_{ij}^{(t)} = \begin{cases} -\Delta_{ij}^{(t)}, & \text{se } \frac{\partial E}{\partial w_{ij}}^{(t)} > 0 \\ +\Delta_{ij}^{(t)}, & \text{se } \frac{\partial E}{\partial w_{ij}}^{(t)} < 0 \\ 0, & \text{caso contrário} \end{cases} \quad (2.9)$$

de forma que $w_{ij}^{(t+1)} = w_{ij}^{(t)} + \Delta w_{ij}^{(t)}$, ou seja, se a derivada é positiva (aumenta o erro), o peso é decrementado pelo valor de atualização, se a derivada é negativa, o peso é adicionado ao valor de atualização. O processo é iterativo e se repetirá até que ocorra a convergência (RIEDMILLER; BRAUN, 1993).

2.6.3 Gradiente Conjugado de Polak-Ribière-Polyak - CGP

O algoritmo de Gradiente Conjugado de Polak-Ribière-Polyak, também conhecido com CGP (do inglês *Conjugate Gradient Polak-Ribière-Polyak*), é um algoritmo de comum uso para problemas de otimização com aplicação em larga escala devido à sua velocidade (SHI; SHEN, 2007).

Às vezes o algoritmo CGP, não apresenta convergência global, nem todas as sequências minimizadas. Apesar disso mostra-se um excelente método capaz convergir com vários erros de arredondamento, com a velocidade de busca superior à muitos algoritmos de Descida de Gradiente (POLYAK, 1969).

Dada a função $f(x)$ que é continuamente diferenciável, possui um mínimo global em R^n , o gradiente $g(x) = \nabla f(x)$, há um nível $L(x_0) = \{x \in R^n | f(x) \leq f(x_0)\}$ iniciando-se a sequência com um dado x_0 , existe um $L > 0$ que:

$$\|g(x) - g(y)\| \leq L\|x - y\|, \forall x, y \in B \quad (2.10)$$

Iniciando-se a busca, para $\mu \in (0, \frac{1}{2})$, $\rho \in (0, 1)$ e $c \in (\frac{1}{2}, 1)$, atribua $s_k = \frac{1-c}{L_k} \cdot \frac{\|g_k\|^2}{\|d_k\|^2}$ e α_k é o maior α em $\{s_k, s_k\rho, s_k\rho^2, \dots\}$ sujeito à:

$$f_k - f(x_k + \alpha d_k) \geq -\alpha \mu g_k^T d_k, \quad (2.11)$$

em seguida:

$$g(x_k + \alpha d_k)^T d(x_k + \alpha d_k) \leq -c \|g(x_k + \alpha d_k)\|^2, \quad (2.12)$$

onde:

$$d(x_k + \alpha d_k) = -g(x_k + \alpha d_k) + \frac{g(x_k + \alpha d_k)^T (g(x_k + \alpha d_k) - g_k)}{\|g_k\|^2} d_k \quad (2.13)$$

e L_k é uma aproximação para a constante de Lipschitz L de $g(x)$. Onde a convergência do algoritmo CGP é dada pelo seguinte pseudocódigo:

1. Selecione $x_0 \in R^n$ e atribua $d_0 = g_0, k := 0$.
2. Se $\|g_k\| = 0$, então pare, senão siga para o passo 2.
3. Atribua $x_{k+1} = x_k + \alpha_k d_k$.
4. Atribua $k := k + 1$ e volte ao passo 1.

Dessa forma, em virtude da não garantia de convergência do método, neste trabalho foi usado apenas para fins de comparação de velocidade e não como resultado final.

2.6.4 Gradiente Conjugado Escalar - SCG

Treinar uma RNA consiste que em cada época ocorra uma minimização da taxa de erro em relação à saída com o ajuste de pesos. A maioria dos algoritmos de treinamento de redes neurais é baseado no método de descida do gradiente, cuja velocidade de convergência é baixa em aplicações de larga escala (MØLLER, 1993).

O algoritmo de Gradiente Conjugado Escalar (SCG, do inglês *Scaled Conjugate Gradient*), consiste numa técnica que evita a busca em linha por cada iteração de treinamento usados no algoritmo de Levenberg–Marquardt (MØLLER, 1993). Dessa forma é possível obter ganhos de velocidade em larga escala.

As estratégias gerais de otimização também são usadas com o algoritmo SCG, porém a direção de busca e o tamanho do passo são mais rápidas usando a informação com a aproximação de segunda ordem:

$$E(w + y) \approx E(w) + E'(w)^T y + \frac{1}{2} y^T E''(w) y \quad (2.14)$$

Dessa forma a expansão em Séries de Taylor limita-se numa derivada de no máximo segunda ordem. Logo a velocidade de convergência é mais rápida que os métodos de descida de Gradiente convencionais. Possibilita-se dessa forma a adoção em RNAs em larga escala.

No contexto desse trabalho, devido à grande dimensionalidade de neurônios utilizados na topologia proposta, esse foi o algoritmo que apresentou melhores resultados em tempos de treinamento aceitáveis.

2.7 COMITÊ DE CLASSIFICADORES

A melhor maneira de obter-se uma combinação de classificadores é através da técnica do voto majoritário (ADHIKARI; SAHA, 2014). Pode ser definido conforme a Equação 2.15. Onde y é o rótulo final obtido baseado na combinação de classificadores, β_i é o peso do classificador e $f_i(x)$ é a saída do classificador aplicada na amostra atual.

$$y = \operatorname{argmax} \left(\sum_{i=1}^n \beta_i f_i(x) \right) \quad (2.15)$$

Outra maneira que pode ser mais usual é adotando-se uma combinação linear de classificadores $y_i = \sum_k w_k d_{ki}$, onde $w_k \geq 0$, $\sum_k w_k = 1$, onde w_k são os pesos inseridos para cada classificador e d_{ki} é a amostra sobre inspeção do classificador y_i . Essa forma pode ser adaptada para o que é conhecido como fusão de classificadores (ADHIKARI; SAHA, 2014).

Nesse estudo, toda vez em que há menção a um comitê de classificadores, é atribuída a junção de várias RNAs do tipo MLP distintas, com especialidade única de classificação de um dado gás, que são combinadas para a tomada de uma decisão conjunta.

2.8 FATORES DE ESQUECIMENTO

Dentro do contexto que tange a ocorrência de deriva e métodos de compensação, existe a hipótese conhecida como "Fatores de Esquecimento" (FF, do inglês *Forgetting Factors*). Esse conceito consiste na aplicação de técnicas para tornar de-

terminado conhecimento mais importante que outro através da diminuição do grau de importância de algum dado ao longo do tempo (LUGHOFER *et al.*, 2015).

Dado que a ocorrência de deriva é diretamente ligada à variação gradual de uma amostra ao longo do tempo, necessita-se classificá-la no instante atual t , tendo como conhecimento passado o intervalo $[1; t - 1]$. Com base na hipótese de que amostras mais recentes temporalmente, ou seja, mais próximas do intervalo t , possuem maior importância que amostras nos intervalos mais antigos, então (ESCOVEDO, 2015):

$$Fator_{t-1} < Fator_{t-2} < Fator_{t-3} < \dots < Fator_1 \quad (2.16)$$

pode ser observada pela relação 2.16 que os fatores de esquecimento mais recentes temporalmente são menores que os que se encontram mais distantes no tempo. Isto significa que um dado muito próximo temporalmente deve ser "pouco esquecido", ou seja, deve ter um peso maior na tomada de decisão num problema com deriva. Assim como um dado que é muito antigo, deve apresentar alta "taxa de esquecimento", ou seja, deve ter um peso menor na tomada de decisão.

De modo que os fatores de esquecimento podem ser aplicados através do mecanismo de pesos diretamente nos classificadores, atribuindo-se valores de ponderação maiores em dados mais recentes.

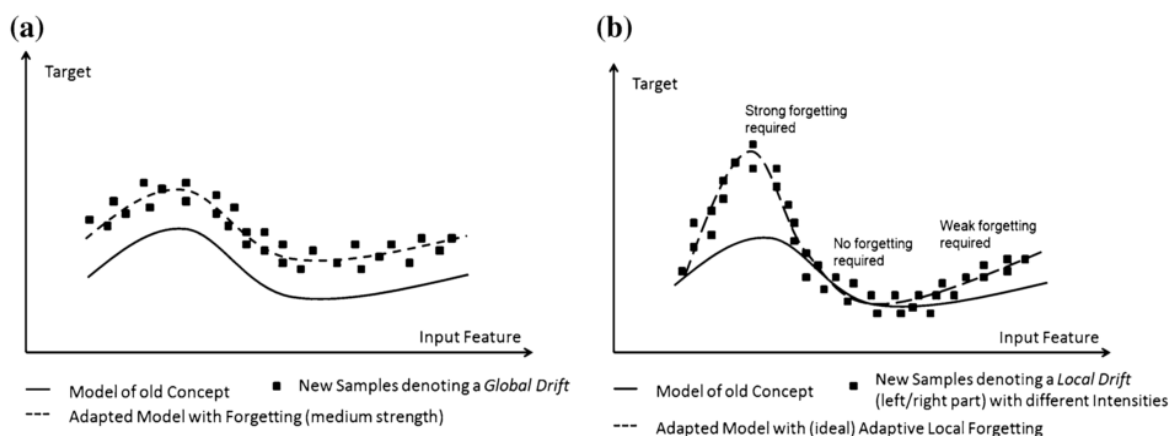


Figura 7 – Modelos de deriva

(a) Acontecimento de deriva global (b) Deriva local, exigindo um grande fator de esquecimento, Fonte: Shaker e Lughofer (2014)

A Figura 7 exemplifica duas situações distintas nas quais é necessária a aplicação de fatores de esquecimento ou pesos na ocorrência de deriva; Na Figura 7 (a) há a ocorrência de deriva com a mudança respectiva de conceito ao longo de todo o conjunto de dados, a mudança é dada de maneira global, porém com leve intensidade, logo a aplicação de fatores de esquecimento se dá de maneira adaptativa, porém

com baixa intensidade. Já na Figura 7 (b) há uma mudança de conceito localizada somente em determinado intervalo de tempo, porém quando a mudança é detectada, a intensidade é alta, logo há necessidade de se ponderar a característica de maneira intensa também, com a aplicação de pesos menores há esse classificador específico (SHAKER; LUGHOFER, 2014).

2.9 ABORDAGEM DE DADOS E CLASSIFICADORES

Para a construção de classificadores na compensação da ocorrência de deriva após a detecção, em seguida a divisão do conjunto de dados, pode-se adotar abordagens diferenciadas quanto ao tipo do método, instante de detecção (*online ou offline*), entrada dos dados, tipo de classificador e união dos mesmos conforme a Tabela 1.

No caso da abordagem ativa, é necessário sempre um algoritmo que faz a detecção de deriva em tempo de execução, na qual quando verificada a mudança, deve-se tomar alguma ação de retreino, dimensionamento da janela de amostragem, etc. Como principal desvantagem desse método, destaca-se a perda de confiabilidade no caso de mal funcionamento do mecanismo de detecção (ESCOVEDO, 2015). Enquanto isso a abordagem passiva leva em consideração um mecanismo de controle em malha aberta, presumindo-se sempre a ocorrência do problema e efetuando diretamente a compensação.

Tabela 1 – Tabela comparativa de abordagens de dados e classificadores possíveis na compensação de deriva. Adaptado de: Escovedo (2015)

| Abordagem Ativa x Passiva | |
|----------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Ativa | Possui método de detecção de deriva online, ou seja, sempre faz a tentativa de detecção de deriva atualizando o modelo somente quando é detectada. |
| Passiva | Assume a ocorrência de deriva em andamento e atualiza continuamente o modelo, se houver mudança, há aprendizado, senão o conhecimento é reforçado. É conhecida também como abordagem offline. |
| Entrada Individual x Entrada em Lotes | |
| Individual | Aprende uma amostra por vez, é mais responsivo, porém instável, sujeito a ruídos. |
| Lotes | Trabalha em blocos de amostras, porém é ineficiente com poucas amostras, utiliza mecanismo de enjanelamento para controle dos blocos. |
| Classificador específico x livre | |
| Específico | Utiliza apenas um tipo de classificador. |
| Livre | Utiliza qualquer classificador sem seguir alguma recomendação |
| Classificador único x comitê | |
| Classificador único | Um classificador apenas. |
| Comitê | Utiliza um conjunto de classificadores na tomada de decisão. |

No que se refere à abordagem em blocos ou individual, é importante destacar que é possível converter amostras individuais em blocos e vice-versa. Para o uso de um

classificador apenas ou um comitê especializado em tomada de decisão há a vantagem de utilizar o comitê na busca de uma solução mais votada (ESCOVEDO, 2015).

No contexto desse trabalho, a abordagem utilizada é passiva. O conjunto de dados é dividido em lotes de amostras não ordenadas e com distribuição não uniforme. Dessa forma a ocorrência de deriva é dada a partir da fronteira temporal que divide um lote de dado ao outro.

2.9.1 Compensação de deriva em sensores de gases na literatura

Em função de compensar e resolver o problema de deriva encontrado em larga escala em sensores de gases, vários classificadores são propostos na literatura. É possível usar técnicas de combinação massiva de classificadores distintos, tanto com o uso de comitês ou métodos de ponderações com características distintas.

Vergara *et al.* (2012) propõem o uso de uma SVM que primeiro é treinada com um período de dados pouco anterior ao atual, no qual ainda a rede de sensores é nova, em seguida o mesmo testa uma nova amostra ao longo de todo o período posterior. De forma que:

$$C_T = \sum_{t=1}^{T-1} SVM(B_t) \quad (2.17)$$

Onde C_T é o classificador usado para fins de compensação da deriva que ocorre no conjunto de dados de índice T , de forma que a classificação é dada pela união de todos os classificadores treinados por lotes contidos no intervalo $1 < t < T - 1$, ou seja, as máquinas de vetor de suporte que foram treinadas com lotes de dados B_t nesse mesmo intervalo.

Resultados bons são encontrados na combinação de diferentes SVMs para o mesmo conjunto de dados, baseados no uso de 10 SVMs ou mais. A seguir serão mostrados quatro trabalhos na literatura que utilizaram o mesmo conjunto de dados desse texto para entendimento das técnicas utilizadas e resultados obtidos.

2.9.1.1 Combinação de RNAs e kNN - (ADHIKARI; SAHA, 2014)

Adhikari e Saha (2014), usam uma combinação de RNAs do tipo MLP com classificadores de cálculo de vizinhos mais próximos (kNN, do inglês *k-Nearest Neighbors*) com dois critérios distintos de avaliação: Voto majoritário e pesos aplicados no classificador.

A topologia de RNAs construídas consistem em: 128 neurônios na camada de entrada (um para cada característica), 6 neurônios na camada de saída (um para cada gás) e duas camadas intermediárias de até 10 neurônios cada.

Para o algoritmo kNN utiliza-se a distância euclidiana padrão com a escolha de $k = 7$, este valor portanto escolhido através de testes.

A combinação dos dois classificadores é linear de modo que $y_i = \sum_j w_j d_{ij}$, onde, $w_{ij} \geq 0$, $\sum_j w_j = 1$. Para a ponderação das RNAs as saídas são ponderadas na forma de mediana:

$$y_i = \text{mediana}_j d_{ji} \quad (2.18)$$

onde d_{ji} é o voto do classificador j para a classe C_i . Para o voto do classificador usa-se a regra $d_{ji} = 1, \dots, K$, onde K é o número de saída de cada classificador.

Para o classificador kNN, a combinação se dá através de pesos uniformes com voto majoritário:

$$y_i = \text{maxarg} \sum_{i=1}^t \beta_i f_i(x) \quad (2.19)$$

onde y_i é a saída final do classificador x são as entradas, β_i são os pesos dos classificadores e $f_i(x)$ é a saída do classificador para o lote de dados atual.

2.9.1.2 Conjunto regularizado de classificadores - (ASMITA; SHUKLA, 2015)

Asmita e Shukla (2015) utilizam para classificação e compensação de deriva um conjunto de máquinas de vetor de suporte (SVM, do inglês *Support Vector Machine*), com o uso de votos majoritários na qual o conjunto de pesos é calculado com um algoritmo de regularização.

O objetivo do método proposto é calcular o peso de cada classificador para execução de um processo de fusão:

$$\text{minarg}_{\beta_1 \dots \beta_t} = \sum_{i=1}^{mt} \sum_{j=1}^t \max(0, 1 - \beta_i y_i f(x_i)) \quad (2.20)$$

onde β_i é calculado através de voto majoritário de uma combinação de um algoritmo de regularização de valor Singular de Decomposição (SVD, do inglês *Singular Value Decomposition*) e o processo de regularização de Tikhonov, este último dado por:

$$h_{t+1}(x) = \text{maxarg}_{y=1..L} \sum t : f_t(x) = y\beta_t + (\lambda_1)^2 (\|\beta\|_2)^2 \quad (2.21)$$

de forma que λ é responsável por controlar o limiar entre estabilidade e precisão do sistema de controle da regularização. Este processo depende também do índice do conjunto de dados, dado por t , e o peso inicial estimado por β_t .

Ao final do procedimento, é possível obter um vetor de pesos regularizados, que com um conjunto de SVMs treinadas, estabelece-se um peso adequado para cada classificador.

2.9.1.3 Conjunto de redes neurais - (KADRI *et al.*, 2013)

Kadri *et al.* (2013) criam um comitê de classificadores utilizando kNN e MLP com mecanismos de combinação e voto majoritário, estes baseados no crescimento do erro de classificação ao longo do tempo conforme a ocorrência de deriva. Dessa forma o autor propõem um método *online* de integração para resolução desse problema.

O cálculo dos pesos do classificador RNA é dado pelo método clássico de voto majoritário, onde $f(x) = \sum_{i=1}^m w_i f_i(x)$.

Para o uso do método de obtenção de pesos pelo algoritmo kNN, duas funções de cálculos de distâncias são usadas: distância Euclidiana e distância cossenoidal. Logo:

$$Ed_{st} = \sqrt{(x_s - y_t)(x_s - y_t)^T} \quad (2.22)$$

onde Ed_{st} é a distância Euclidiana entre os vetor x_s e y_t e $(x_s - y_t)^T$ é o vetor transposto de $(x_s - y_t)$.

Para a distância cossenoidal:

$$Cd_{st} = \left(1 - \frac{x_s y_t^T}{\sqrt{(x_s x_s^T)(y_t y_t^T)}} \right) \quad (2.23)$$

onde Cd_{st} é a distância cossenoidal entre o vetor x_s e y_t .

O peso para cada um dos classificadores treinados com a técnica de monitoramento recorrente do erro é dado por:

$$w_i^k = \frac{1}{\sum_{j=1}^N erv_{ij}^k} \quad (2.24)$$

onde M é o número de modelos base, N é o número de vizinhos próximos, erv_{ij}^m é o erro de aplicado ao índice de m na amostra do índice j no conjunto de validação.

2.9.1.4 Demais trabalhos

Hu *et al.* (2014), Fonollosa, Vergara e Huerta (2013), Liu e Tang (2013) e Rodriguez-Lujan *et al.* (2014) usam respectivamente métodos de SVM com *deep learning* na fase de treinamento, RNAs MLP com método *multikernel* RBF, SVMs com

pesos estáticos e SVMs iterativas. Para maiores detalhes de implementação de cada um dos métodos listados anteriormente, os respectivos artigos podem ser consultados.

2.9.2 Comparativo entre trabalhos

A Tabela 2 traz um resumo das principais técnicas adotadas, vantagens e desvantagens.

Tabela 2 – Tabela com algoritmos e valores alcançados na literatura

| Publicação | Classificadores | Descrição da abordagem | Vantagens | Desvantagens |
|--------------------------------------|-----------------|------------------------------------------------------------|---------------------------------------------------|----------------------------------------------------------|
| Adhikari e Saha (2014) | RNA MLP e kNN | Voto majoritário com a combinação dos dois classificadores | Simplicidade | Treino de classificadores distintos resultados razoáveis |
| Vergara <i>et al.</i> (2012) | SVM | Conjunto de SVMs usando somente o lote anterior | Único tipo de classificador | Não usa mecanismos de pesos |
| Asmita e Shukla (2015) | SVM | Conjunto de SVMs usando pesos com kernel RBF | Realimentação com um único tipo de classificador | Usa somente o conjunto t-1 e têm resultados razoáveis |
| Kadri <i>et al.</i> (2013) | RNA e kNN | RNAs com pesos calculados por kNN | Baixo tempo de treinamento | Não usa mecanismos de pesos, resultados razoáveis |
| Hu <i>et al.</i> (2014) | SVM | Conjunto de SVMs com deep learning | Convergência depois de longo tempo de treinamento | Complexidade de treinamento |
| Fonollosa, Vergara e Huerta (2013) | SVM | Método multikernel RBF | Único tipo de classificador | Não produz resultados diferentes de MLP |
| Liu e Tang (2013) | SVM | Conjunto de SVMs com pesos estáticos manuais | Único tipo de classificador | Método simplificado sem uso de todos lotes anteriores |
| Rodriguez-Lujan <i>et al.</i> (2014) | ISVM | SVM iterativa | Mais rápida que SVM convencional | Complexidade de modelagem da ISVM |

2.10 CONSIDERAÇÕES FINAIS DO CAPÍTULO

Ao longo desse capítulo foram apresentadas as principais definições acerca do fenômeno de deriva em sensores de gases, variáveis que estão envolvidas e métodos utilizados na literatura para melhoria dos sistemas envolvidos.

O fenômeno de deriva consiste na variação da distribuição das amostras ao longo do tempo, dessa forma há uma divergência entre as médias iniciais e finais do conjunto de dados com o avanço da janela temporal.

Sensores de gases possuem importância fundamental na indústria e em sistemas do dia a dia, porém sofrem do problema de deriva devido à sua construção química e à exposição sob diversos agentes contaminantes externos.

Evitar a ocorrência de deriva em sensores de gases não é possível, porém existe a possibilidade da atenuação de sua ocorrência com: Avaliação e simulação estatística com base no cálculo de probabilidades de cada uma das variáveis envolvidas. No segundo caso é comum o uso de algoritmos de inteligência artificial com classificadores que "aprendem" esse fenômeno. Dentro das abordagens de IA, há sempre na maioria dos casos a aplicação de algoritmos como SVMs, RNAs, DTs através de técnicas de criação em massa de vários classificadores combinados em grande número para atenuar a deriva. Esses últimos na maioria das vezes descartam amostras de dados muito antigas que não tenham grau de importância relevante quando há alto índice de deriva.

3 METODOLOGIA

A abordagem proposta nesse trabalho consiste na aplicação de redes neurais (RNAs), para a classificação de dados provenientes de uma matriz de sensores de gases. Através de testes e experimentos de classificação é notada a presença do fenômeno de deriva, estes são feitos apenas com o primeiro lote de dados sem alguma técnica de compensação. Com base nisso é adotada uma abordagem passiva, ou seja, de maneira *offline* há o conhecimento de deriva no conjunto de dados, sem que haja esforço e detectar qual o grau de deriva ocorrendo.

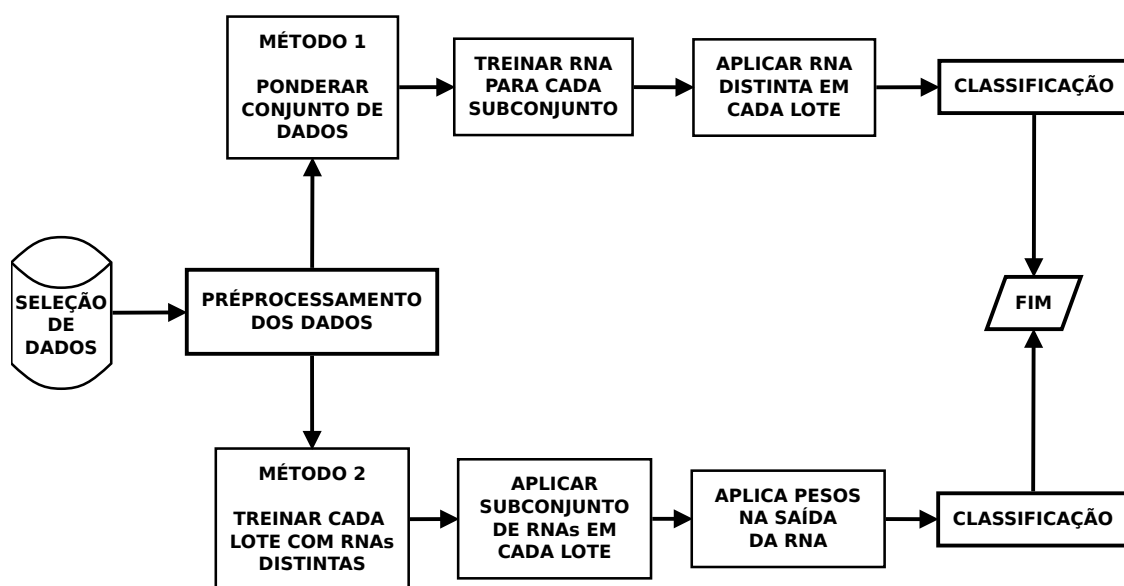


Figura 8 – Fluxograma simplificado dos métodos de ponderação adotados

A Figura 8 mostra de maneira geral e simplificada os processos que foram adotados para a compensação de deriva. Primeiramente, para fins experimentais, na fase de processamento inicial dos dados e escolha de uma abordagem contínua sem lotes, tentou-se detectar o fenômeno de deriva com o algoritmo PHT no conjunto de dados concatenados. Em seguida, é efetuado um pré-processamento dos dados através da normalização escalar global entre os intervalos $[-1;+1]$ para melhoria da qualidade do treino das RNAs. Em seguida são construídas as topologias de RNAs que serão usadas nessa pesquisa. Na fase de treino são usados algoritmos de boa performance como: SCG, RP, CGP, onde a quantidade alta de neurônios não afeta significativamente o tempo de treinamento.

A aplicação do algoritmo PHT descrito acima, devido à natureza do conjunto de dados não apresentou bons resultados, onde as fronteiras de decisão ficaram distantes temporalmente das linhas que dividem os lotes de dados. Dessa forma sua utilização

foi descartada nesse trabalho e abordagem é seguida através de lotes. Uma breve descrição desse algoritmo pode ser encontrada no apêndice A.0.1.

Ainda conforme a Figura 8 há dessa forma uma subdivisão na maneira em que são aplicados fatores de esquecimento. O primeiro método é a aplicação do conjunto de dados original através das redes neurais onde os neurônios de saída sofrem uma ponderação de acordo com seu grau de importância na classificação do dado em questão. A ponderação é dada por curvas que tentam simular o grau de esquecimento do sistema. No segundo método, o conjunto de dados sofre uma ponderação com técnicas que evidenciem amostras mais relevantes que outras, baseando-se na repetição de dados mais recentes. Após esse último passo, é treinada uma RNA especialista na resolução da deriva naquele intervalo temporal.

3.1 O CONJUNTO DE DADOS

Nesse trabalho será usado um conjunto de dados disponibilizado por Fonollosa, Rodríguez-Luján e Huerta (2015), o qual é considerado rigoroso e preciso devido aos cuidados na sua coleta. Todos os dados foram coletados num intervalo de 36 meses num ambiente controlado e adequado com o uso de 16 sensores do tipo MOX. A quantidade de amostras disponíveis é de 13910 coletas através da medição da concentração de seis tipos de gases distintos: Etanol, Etileno, Acetona, Amônia, Acetaldeído e Tolueno. Cada sensor no final apresenta 8 diferentes características que são baseadas no estado da arte de extração de dados de sensores de gases (VERGARA *et al.*, 2012). A Figura 9 apresenta todo o mecanismo de coleta e tratamento de amostras ao longo dos três anos de trabalho.

Todo o ambiente de operação do equipamento está sob temperatura controlada de 25 °C com erro máximo de $\pm 1\%$. Existem 7 cilindros de gases pressurizados, sendo 6 para cada um dos gases descritos anteriormente e 1 de ar pressurizado para efetuar a limpeza do sistema e manter a umidade relativa de mensuração em torno de 10 %.

São adotados quatro tipos de sensores diferentes: TGS2600, TGS2602, TGS2610, TGS2620 que são fabricados pela empresa *Figaro Inc.* Cada um dos quatro tipos de sensores possui uma resposta multivariada diferente para cada tipo de composto. Na estrutura total são usados quatro sensores de cada tipo, totalizando dezesseis, que são dispostos na seguinte ordem: TGS2602; TGS2602; TGS2600; TGS2600; TGS2610; TGS2610; TGS2620; TGS2620; TGS2602; TGS2602; TGS2600; TGS2600; TGS2610; TGS2610; TGS2620; TGS2620.

Os 16 sensores são acoplados em uma câmara hermeticamente fechada de 60 ml que recebe em seu fluxo máximo uma taxa de até 200 ml/min do gás amostrado. A câmara é colocada sobre uma placa de aquecimento de sensores que mantém a

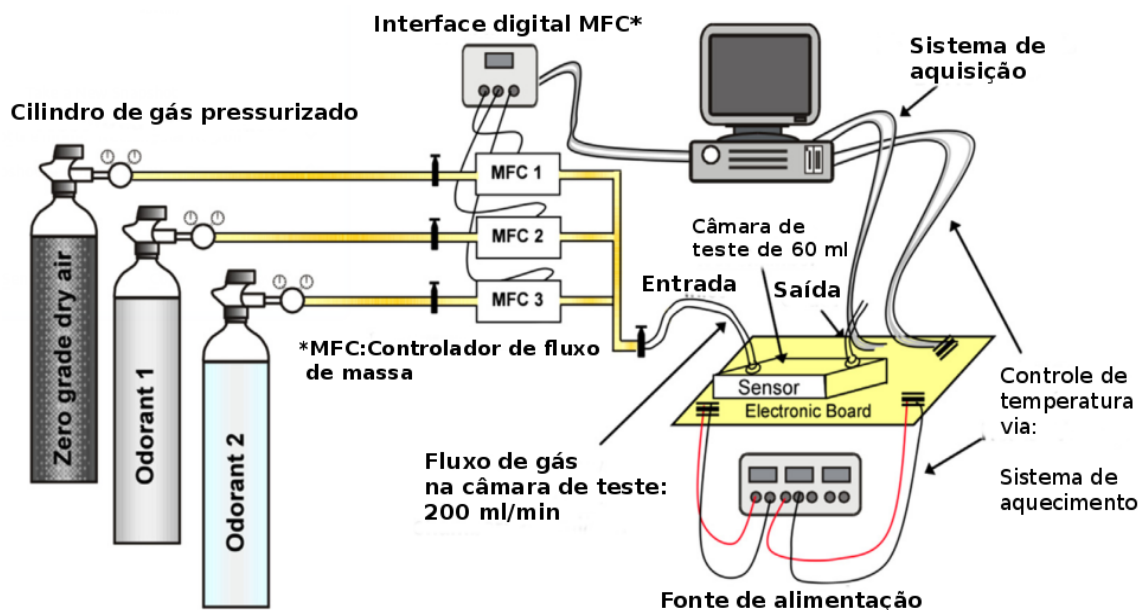


Figura 9 – Mecanismo automatizado para coleta de dados ao longo de 36 meses

Fonte: Fonollosa, Rodríguez-Luján e Huerta (2015)

temperatura da superfície sensora em um valor fixo de 400 °C. A placa de aquecimento é alimentada por uma fonte DC externa de 5 Volts.

Conforme mostrado na Figura 9 o controle de amostragem é feito com uma interface de Controle de Fluxo de Massa (MFC, do inglês *Mass Flow Controller*), que aciona válvulas e faz a liberação do fluxo de gás conforme comandos externos recebidos. Tanto as interfaces MFC quanto a placa de aquecimento, são acionados via controle *On-Off* por um computador externamente ligado à placa de aquisição de sinais e demais interfaces. O funcionamento do computador de controle consiste basicamente:

- Manter constante a temperatura da superfície que contém o substrato sensor em 400 °C através do sensor de temperatura ligado na placa de aquisição, ligando e desligando o aquecedor;
- Efetuar a leitura e aquisição dos dados da matriz sensora;
- Executar abertura e fechamento de válvulas da interface MFC.

Primeiramente o computador efetua a liberação das válvulas de cada interface MFC usando um critério aleatório de seleção de qual gás será amostrado. O tempo total de aquisição dos dados é de 300 s, sendo 100 s a fase de injeção do gás na câmara e 200 s o tempo de recuperação (fase de limpeza), na qual é injetado ar comprimido. A taxa de amostragem da placa de aquisição e gravação de dados é de 100 Hz.

O autor do conjunto de dados divide toda a sua amostragem em 10 lotes, que são subdivisões dos dados coletados ao longo de 36 meses de maneira temporalmente ordenada conforme a tabela 3. A distribuição dos dados não é uniforme, é irregular, a quantidade de amostras por tipo de gás não é a mesma em todos os lotes e existem lacunas em alguns conjuntos de dados (VERGARA *et al.*, 2012). Essa abordagem de distribuição irregular dos dados é proposital conforme argumenta o autor, onde busca-se simular um ambiente de funcionamento real, ou seja, do mesmo modo que acontece com o uso de sensores de gases, a fim de se verificar a ocorrência de deriva nos dados coletados e produzir métodos e técnicas que não sejam viciadas a uma distribuição ideal das amostras. Exemplos desse fato podem ser observados na Tabela 3, são a diminuição do grande do número de amostras nos lotes 4 e 5. Outro fator que também torna a distribuição dos dados irregular conforme relatado pelo autor, é ausência da mensuração do gás Tolueno nos lotes 3,4 e 5, onde ocorreu uma falha no equipamento de aquisição de dados.

Mais exemplos podem ser dados pela irregularidade na distribuição desse conjunto de dados: Existem 6 meses sem coleta de dados entre os lotes 9 e 10, esse tempo poderá ajudar na validação da robustez do método proposto, pois os sensores nesse intervalo de tempo podem ser contaminados por agentes externos e outras interferências que são irreversíveis na camada sensora e trazem sem via de dúvidas uma dinâmica adicional ao problema de compensação de deriva.

Tabela 3 – Distribuição dos dados coletados ao longo de 36 meses

Adaptado de: Fonollosa, Rodríguez-Luján e Huerta (2015)

| Lote | Meses | Número de amostras | | | | | |
|------|-------------|--------------------|---------|--------|-------------|---------|---------|
| | | Etanol | Etileno | Amônia | Acetaldeído | Acetona | Tolueno |
| 1 | 1,2 | 83 | 30 | 70 | 98 | 90 | 74 |
| 2 | 3,4,8,9,10 | 100 | 109 | 532 | 334 | 164 | 5 |
| 3 | 11,12,13 | 216 | 240 | 275 | 490 | 365 | 0 |
| 4 | 14,15 | 12 | 30 | 12 | 43 | 64 | 0 |
| 5 | 16 | 20 | 46 | 63 | 40 | 28 | 0 |
| 6 | 17,18,19,20 | 110 | 29 | 606 | 574 | 514 | 467 |
| 7 | 21 | 360 | 744 | 630 | 662 | 649 | 568 |
| 8 | 22,23 | 40 | 33 | 143 | 30 | 30 | 18 |
| 9 | 24,30 | 100 | 75 | 78 | 55 | 61 | 101 |
| 10 | 36 | 600 | 600 | 600 | 600 | 600 | 600 |

Como pode-se observar de acordo com a tabela 3, trata-se de um problema do mundo real, não sendo possível contar com dados distribuídos ao longo do tempo de maneira simétrica, em quantidades proporcionais, distribuição uniforme ou com continuidade temporal. A existência desses problemas, aliadas a presença inevitável da deriva no conjunto de dados, torna apenas o problema mais difícil ainda de se resolver. Compensar a deriva mostra-se essencial nesse caso como forma de melhorar

a classificação de um sistema que se altera ao longo do tempo de maneira imprevisível.

3.1.1 Extração de características

Para cada amostragem de um gás são extraídas no total 128 características, sendo 8 características distintas por sensor da matriz original. Dois métodos para extração de características são adotados: Amplitude da variação de impedância do sensor normalizada e Dinâmica de resposta no período de absorção e adsorção.

A Equação 3.1 descreve o considerado "Padrão Ouro" na extração de características de sensores de gases, conforme mostrado na Figura 10 (a), onde são levados em consideração os valores de máxima e mínima impedância durante o período de preenchimento da câmara de sensores com gases e a impedância mínima durante o período de limpeza do mesmo local (LLOBET *et al.*, 1997):

Sendo ΔR a diferença máxima na aquisição da amostra em ohms:

$$\Delta R = \max_r[k] - \min_r[k] \quad (3.1)$$

É de interesse obter-se a sua versão normalizada e em módulo conforme a Equação 3.2:

$$||\Delta R|| = \frac{\max_r[k] - \min_r[k]}{\min_r[k]} \quad (3.2)$$

Onde k é o intervalo contido entre $[0;T]$, sendo T o instante final onde a amostra é passada através da câmara de sensoreamento. Usando esse método que é o estado da arte no sensoreamento de gases são obtidas duas características: ΔR e $||\Delta R||$.

As outras 6 características são extraídas a partir da dinâmica de funcionamento de cada sensor, avaliando assim o formato da curva de resposta de cada amostra aplicada na câmara. A Equação 3.3 descreve o filtro de média móvel (EMA, do inglês, *Exponential Moving Average*), que traz as características desejadas nesse sistema de forma normalizada entre $[-1;+1]$:

$$y[k] = (1 - \alpha)y[k - 1] + \alpha(x[k] - x[k - 1]) \quad (3.3)$$

Conforme descrito anteriormente, são coletadas 6 características usando o filtro EMA, com valores $\alpha=0.1, 0.01, 0.001$. No total são 8 características utilizadas por sensor: 6 filtradas (Equação 3.3) e 2 relativas à amplitude de ativação (Equações 3.1 e 3.2). O filtro é aplicado com os valores de α durante toda a fase de absorção e adsorção do gás. A Figura 10 (b,c,d,e,f,g) exemplifica a passagem do gás Aldeído ao longo do sistema, produzindo respectivamente 3 características que representam

o ponto máximo obtido para os filtros aplicados na fase de absorção que dura 100 s. Logo são obtidas 3 características que representam o valor mínimo dos dados filtrados durante a fase de adsorção que dura em média 200 s. Como existem 8 características por sensor e 16 sensores no total, chega-se ao número de 128 características totais.

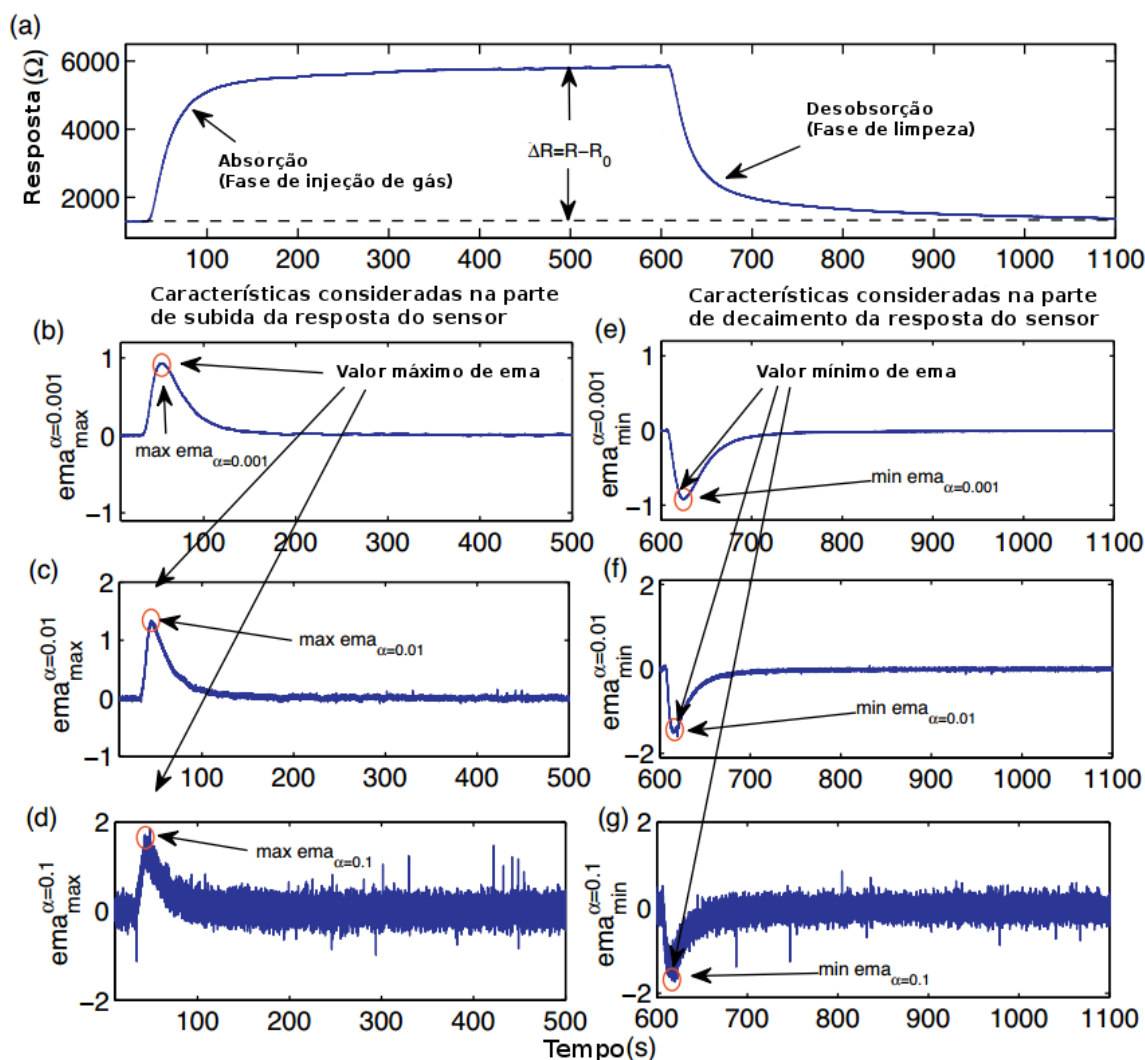


Figura 10 – Curvas de respostas de sensores de gases do tipo MOX.

(a) Resposta típica na passagem de Acetaldeído. (b) - (d) Representação dos pontos amostrados na resposta imediata dos sensores ao longo da passagem de um gás com a aplicação das três configurações do filtro de média móvel exponencial nos momentos de máxima absorção e desabsorção. Fonte: Vergara *et al.* (2013)

Uma das alternativas para verificação da qualidade desses dois métodos de extração de características é a aplicação da transformada PCA. Esta visa reduzir a dimensionalidade das 128 características para facilitação de análise, simplificação do problema e uso de características mais importantes apenas. No contexto deste trabalho, essa transformada não foi aplicada devido ao grande número de sensores usados e a importância de todas as características que são consideradas o estado da arte em sensores do tipo MOX.

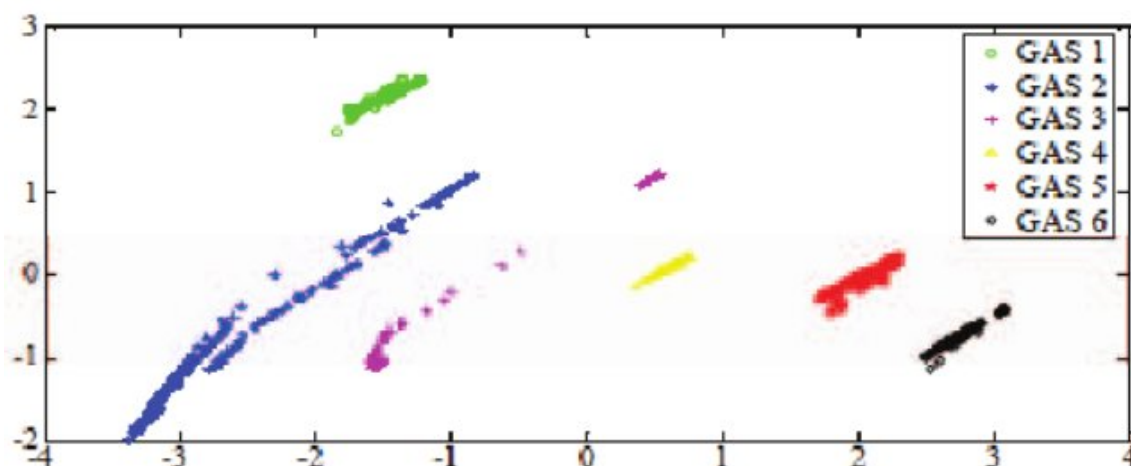


Figura 11 – Análise PCA sem deriva

Adaptado de: Adhikari e Saha (2014)

A Figura 11, mostra a aplicação da transformada PCA ao longo das 128 características reduzidas para duas dimensões que podem ser analisadas através de gráficos (ADHIKARI; SAHA, 2014). Pode-se visualizar que houve agrupamento espacial de amostras e conjuntos distintos, logo é possível traçar uma superfície de decisão que possa classificar um determinado gás através de um classificador treinado com as características obtidas pelos métodos abordados.

Apesar da menção sobre a aplicação da análise principal de componentes, tal técnica não foi utilizada neste trabalho.

3.2 PRÉ-PROCESSAMENTO

Para cada um dos 10 lotes de dados disponíveis são usados cada um dos sensores separadamente, onde é feita a extração de diferentes características para cada um dos gases distintos. O conjunto de dados apresenta uma grande amplitude numérica, desde valores decimais (menores que 1) até valores que podem ser da ordem de dezenas de milhares (maior que 10000). Aplicou-se uma normalização das entradas para que todas elas ficassem numa mesma faixa dinâmica, compatível com a função de ativação da rede neural usando a transformação:

$$y[n] = 2 \times \frac{x[n] - \text{Min}(x)}{\text{Max}(x) - \text{Min}(x)} - 1 \quad (3.4)$$

onde $y[n]$ representa a amostra normalizada, $x[n]$ é a o dado que não sofreu normalização. A variável $n = 1..T$, onde T é a quantidade total de amostras do banco de dados. As funções $\text{Min}(x)$ e $\text{Max}(x)$ representam respectivamente os valores mínimos e máximos do conjunto de dados global para cada uma das 128 características.

Com a aplicação dessa técnica de normalização, são obtidos dados na faixa $[-1;+1]$, que facilitará o treinamento das RNAs com função de ativação do tipo tangente hiperbólica, ganhando desempenho e melhores resultados.

3.3 TOPOLOGIAS DE REDES NEURAIS

O método para resolução do problema de deriva em sensores de gases nessa pesquisa é baseado no uso de RNAs com aplicação de fatores de esquecimento de dados antigos. Basicamente no que diz respeito à topologia de RNAs que são usadas no treinamento de classificadores de gases e compensação da ocorrência de deriva, são usados dois modelos: Uma RNA com várias saídas na qual o neurônio de maior ativação resultará na seleção no gás, a outra é através construção de um comitê de classificadores, na qual são formadas RNAs especialistas em cada tipo de gás, seis RNAs distintas com um neurônio de saída, onde através de uma função de agregação, será selecionado o rótulo cuja RNA possua o neurônio com maior nível de ativação.

A Figura 12 exemplifica a topologia adotada de uma RNA de múltiplas camadas sem recorrência, ou seja, não possui realimentação do neurônio de saída. Nesse caso é usada uma camada intermediária de 67 neurônios, número obtido pela média aritmética da quantidade de entradas com a quantidade de saídas. Com esse formato de rede neural, é possível com uma única passagem de dados obter um rótulo diretamente. Esse formato descrito será usado no método de ponderação das saídas das RNAs e da aplicação de pesos na entrada das RNAs confeccionadas.

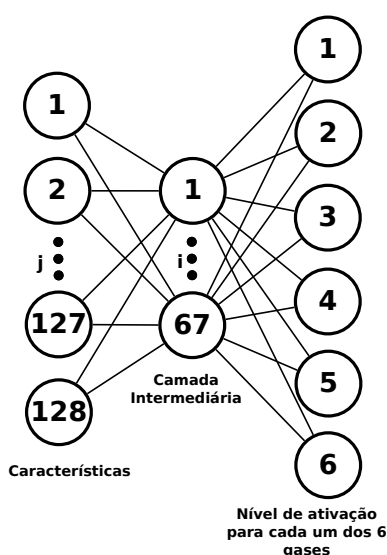


Figura 12 – Topologia generalizada de uma rede neural para todo o conjunto de gases

A Figura 13 apresenta um desenho simplificado da outra arquitetura de RNA testada. Nesse modelo há a construção de 6 RNAs distintas para cada classificação que se deseja usar. Esse modelo é conhecido como comitê de classificadores. Somente

um neurônio de ativação é usado na saída da RNA, sendo que existem 64 neurônios na camada intermediária. Este formato de construção de RNA será usado no modelo de ponderação das saídas através de pesos apenas, pois apresentaram resultados semelhantes com o modelo anterior. O rótulo para a formulação e decisão da técnica de comitê de classificadores é obtido de forma semelhante a do caso anterior, onde a saída de maior valor é escolhida.

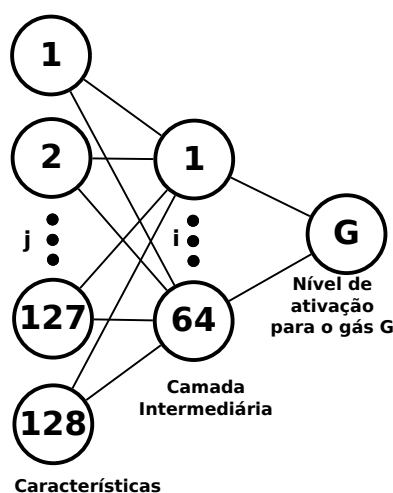


Figura 13 – Topologia de uma rede neural para a classificação de um único gás

3.4 FATORES DE ESQUECIMENTO DE INFORMAÇÕES

No total foram usadas 4 funções: uma exponencial, duas equações quadráticas e uma função logarítmica:

$$w_t = \frac{e^t}{e^c} \quad (3.5)$$

$$w_t = 0.009167t^2 - 0.02167t + 0.3 \quad (3.6)$$

$$w_t = -1.395 * 10^{-18}t^2 + 0.1t - 1.269 * 10^{-16} \quad (3.7)$$

$$w_t = \frac{\ln(t)}{e^t} \quad (3.8)$$

onde w_t representa o valor numérico do peso aplicado ao lote t . A variável c na curva exponencial representa o valor de ajuste para normalização no intervalo $[0; 1]$ e varia de acordo com o lote em que a amostra está contida. Dessa forma tenta-se relacionar de maneira direta, a baixa importância para conjuntos de dados mais antigos

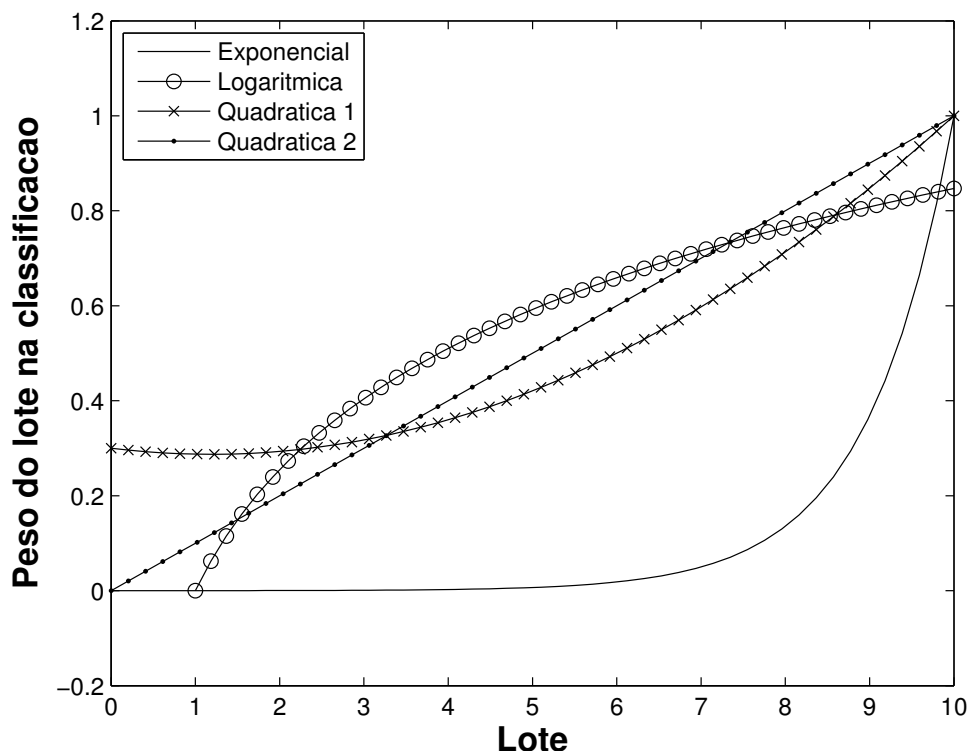


Figura 14 – Quatro modelos de curvas propostas para aplicação de fatores de esquecimento nas saídas das redes neurais

Fonte: Wiezbicki e Ribeiro (2016)

e maior relevância para dados mais novos, este relacionamento é dado baseando-se na hipótese que o conhecimento de deriva num intervalo de tempo mais próximo é mais importante que um classificador confeccionado com dados provenientes de lotes mais antigos. A Figura 14 mostra o gráfico de acordo com o desenho de cada uma das equações na forma normalizada.

3.4.1 Fatores de esquecimento e ponderação das saídas das RNAs

Essa pesquisa tem como centro de discussão a hipótese de que é possível executar a atenuação do fenômeno de deriva em sensores de gases baseado no conceito de fatores de esquecimento de informação. Dessa forma são aplicados pesos normalizados numa escala $[0; 1]$ nos neurônios de saída de cada uma das RNAs usadas na classificação de um dado contido no lote t .

Inicialmente, adota-se o treinamento de cada rede neural com seu respectivo lote conforme a Figura 15. Para cada lote no intervalo de $[1; T]$ onde T é quantidade total disponível de lotes, é treinada uma ou mais redes neurais. Se forem usados conjuntos de RNAs, são treinados qNN RNAs para cada lote e somadas suas saídas, a fim de se obter um comportamento médio.

À seguir, são feitas ponderações distintas para dados que estão contidos em

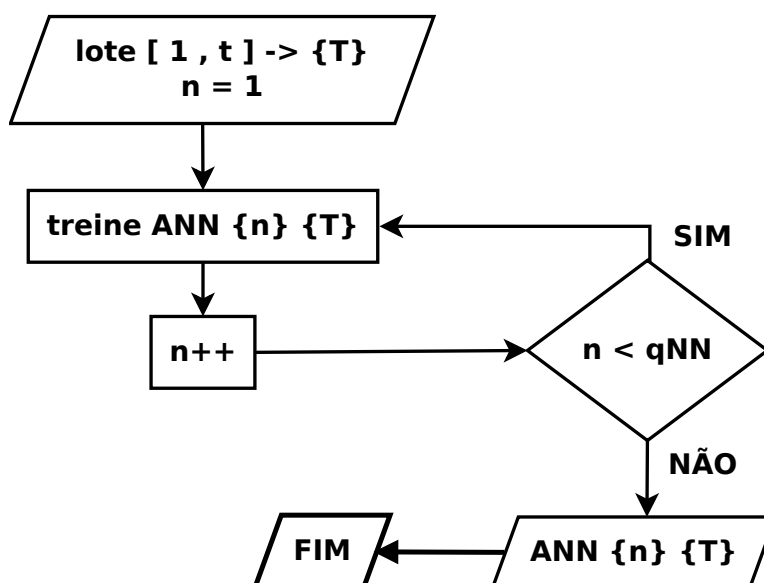


Figura 15 – Fluxograma que contém a lógica de conjunto de dados e treinamento de cada RNA

unidades temporais diferentes. A Figura 16 descreve de forma macro o método adotado para compensação de deriva em sensores de gases com esse tipo de ponderação:

Ainda conforme a Figura 16, dado o lote de dados t cuja classificação pretende-se obter, serão usados dados provenientes do conjunto de informações contidas no intervalo $[1; t - 1]$, que são decorrentes do passado e já foram obtidos de maneira que o sistema proposto seja totalmente causal. A esse último conjunto de todos os lotes de dados passados é dado o nome de E , onde serão testadas 4 curvas características de fatores de esquecimento de conhecimento irrelevante, a esse conjunto de curvas é dado o nome de C . Dessa forma, serão aplicadas todas as redes neurais sobre as n amostras contidas no lote t , que foram treinadas somente com lotes no intervalo $[0; t - 1]$ descritos pelo subconjunto $ANN\{n\}\{E\}$. Após a passagem de cada uma das RNAs são aplicados o conjunto de pesos W , de cada uma das funções descritas pelo conjunto C , de modo que $W[C]$ irá multiplicar cada uma das saídas providas por $ANN\{n\}\{E\}$. Esse processo em seguida sofre uma repetição se nesse caso é usado um conjunto de qNN classificadores destinados à classificação de um mesmo lote. Com base no fato que todas as amostras $n \in lote_t$, foram aplicadas a cada uma das RNAs balanceadas, o conhecimento provido, ou seja, a classificação que será dada com base em informações prévias onde houve ocorrência de deriva é dada pela soma de todas as saídas das RNAs. Essa última metodologia consiste no que é conhecido como Voto Majoritário.

O voto majoritário de vários classificadores é o rótulo final obtido baseado na combinação ponderada dos classificadores.

No apêndice C pode ser encontrado o pseudocódigo para que foi implementado

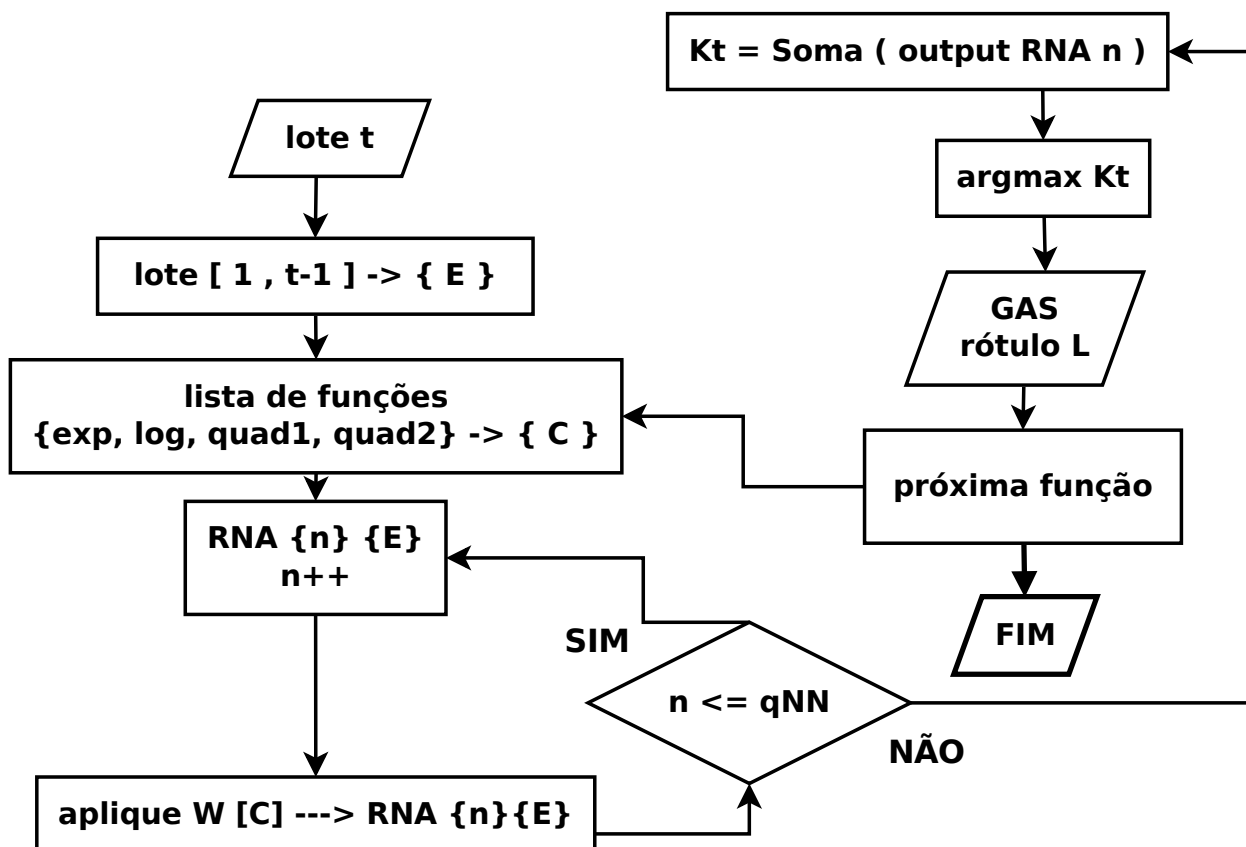


Figura 16 – Fluxograma da metodologia de ponderação das saídas das RNAs

nesse trabalho para o uso desse método.

3.4.2 Esquecimento de informações de entrada

A segunda abordagem seguida nesse trabalho será a aplicação de fatores de esquecimento através do condicionamento de dados na entrada das RNAs. A principal diferença desse método consiste no fato de que a aplicação dos pesos não será feita através da ponderação dos dados pós-processados pelas RNAs confeccionadas. A aplicação dos pesos neste caso é feita diretamente modificando-se a frequência em que cada amostra aparece no conjunto de dados de treinamento. A segunda diferença, não menos importante, é que o treino das redes neurais elaboradas não é feito por lote, mas pelo novo conjunto de dados na qual é aplicada a ponderação por replicação de amostras (PPRA). Este novo conjunto é distinto na classificação de cada lote de informação disponível.

A Figura 17 exemplifica o mecanismo geral de funcionamento desse método no conjunto de dados. Neste caso deseja-se classificar uma amostra ao longo do tempo. De modo inicial, na presença do conjunto de dados na qual se deseja classificar β_t , constrói-se um conjunto de dados oriundos de todos os lotes mais velhos denominado $\{T\}$, onde $\{T\} \supset \beta_1.. \beta_{t-1}$. Para o novo conjunto de dados $\{T\}$ é aplicada a ponderação PPRA

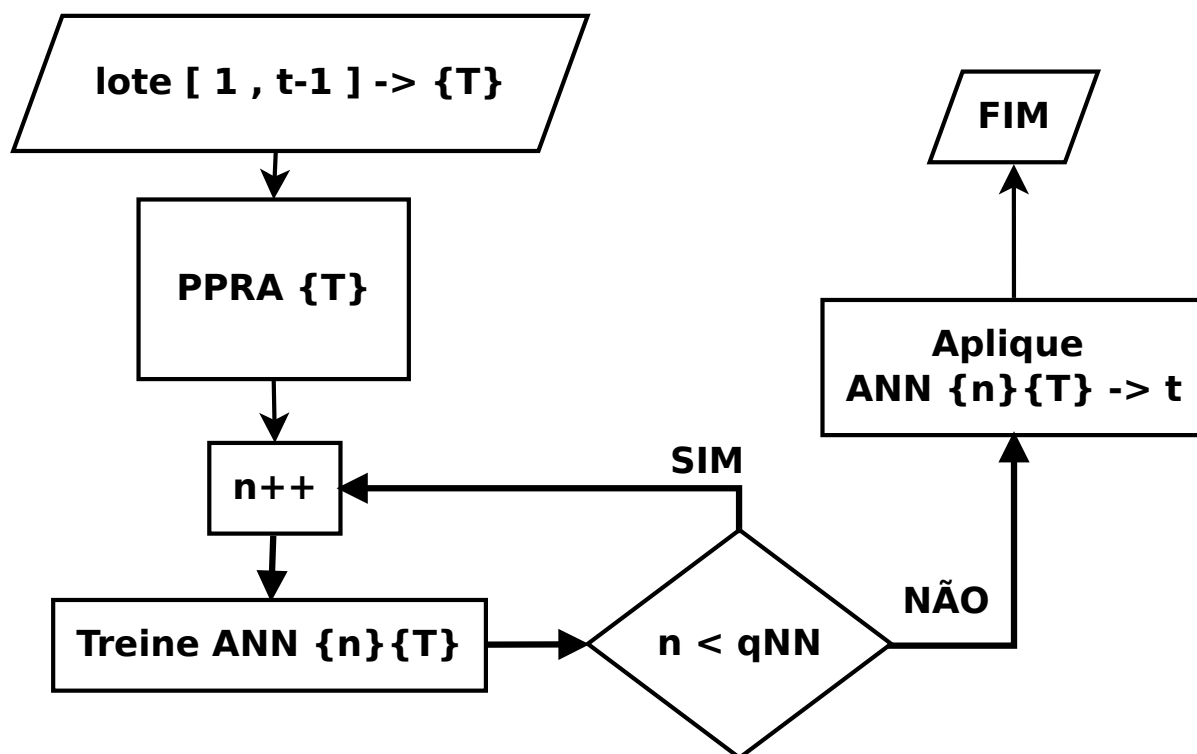


Figura 17 – Modelo de condicionamento dos dados de entrada das RNAs para treinamento e classificação do lote t

sob o mesmo, na qual o histograma original do mesmo conjunto sofre deformação em seu desenho original, com a finalidade de compensar a ocorrência de deriva em todo o histórico de dados, tendo-se como evidência que a mesma ocorre a cada mudança de lote de dados. Com o novo conjunto de dados obtido, são treinadas qNN RNAs, se nesse caso for usado um conjunto de classificadores. Em posse do novo classificador RNA treinado, o mesmo é utilizado para a correta classificação de dados provenientes do lote que está sendo amostrado no intervalo presente.

3.4.3 Ponderação por replicação de amostras - PPRA

A Equações 3.9, 3.10 e 3.11 descrevem o mecanismo de funcionamento da técnica PPRA, que realiza a ponderação através da alteração do conjunto de dados antes do treinamento de um classificador. Dessa maneira, implicando na atribuição de pesos que ocorre uma variação na distribuição das amostras, ao longo do intervalo de tempo t desejando-se obter uma classificação robusta ao acontecimento de deriva. Dado o conjunto de dados B_t , que deseja-se classificar, onde Ψ_t é a transformação de

um novo conjunto de dados a partir da aplicação de PPRA:

$$\begin{aligned}\Psi_1 &= 1 \times B_1 \\ \Psi_2 &= 1 \times B_1 \\ \Psi_3 &= 2 \times B_2 + 1 \times B_1\end{aligned}\tag{3.9}$$

Sendo Ψ_1 , Ψ_2 e Ψ_3 os novos conjuntos de dados transformados a partir da aplicação do método PPRA sobre os lotes de dados B_1 , B_2 e B_3 respectivamente. O coeficiente que multiplica cada um dos conjuntos B significa a quantidade de vezes que o mesmo será repetido ao longo do tempo na formação do novo conjunto Ψ . Pode-se expandir dessa forma:

$$\Psi_t = (t - 1) \times B_{t-1} + (t - 2) \times B_{t-2} + \dots + (1) \times B_1\tag{3.10}$$

Evidenciando que o novo conjunto de dados Ψ_t é composto pela combinação de repetições de cada um dos lotes de dados B_t , sobre o intervalo temporal passado x , onde $1 < x < t - 1$. Juntando toda a expressão é possível equacionar que:

$$\Psi_t = \sum_{j=1}^{t-1} (j \times B_j)\tag{3.11}$$

na qual Ψ_t simboliza o novo conjunto de dados para treinamento da RNA, que é obtido quando deseja-se classificar uma amostra contida no lote de dados B_t . De maneira que se o conjunto de amostras contidas em um mesmo lote é B_i , o conjunto final dos dados será dado pela repetição da amostra ao longo do lote i levando-se justamente em consideração o lote na qual a mesma pertence no estado atual. Como consequência há a concatenação de todas essas amostras em um único lote Ψ_t com a soma de todos os subconjuntos obtidos previamente.

A Tabela 4 traz como resultado da ponderação PPRA, a nova frequência em que as amostras serão contabilizadas em Ψ_t . Por exemplo na classificação de uma amostra que esteja no intervalo temporal contido no lote 5, será treinada uma RNA por dados provenientes de conjunto obtido por PPRA aplicado aos lotes no intervalo [1; 4]. Como consequência da PPRA, é obtido um novo conjunto de dados com quatro réplicas de todas as amostras contidas no lote 4, três réplicas de todas as amostras contidas no lote 3, duas réplicas de todas as amostras contidas no lote 2 e lote 1 irá conter seu conjunto de dados original.

Tabela 4 – Peso em cada conjunto de dados na entrada da rede neural

| Lote | lotes e quantidades usadas no treinamento |
|-------------|---------------------------------------------------------------|
| 1 | 1 x 1 |
| 2 | 1 x 1 |
| 3 | 2 x 2, 1 x 1 |
| 4 | 3 x 3, 2 x 2, 1 x 1 |
| 5 | 4 x 4, 3 x 3, 2 x 2, 1 x 1 |
| 6 | 5 x 5, 4 x 4, 3 x 3, 2 x 2, 1 x 1 |
| 7 | 6 x 6, 5 x 5, 4 x 4, 3 x 3, 2 x 2, 1 x 1 |
| 8 | 7 x 7, 6 x 6, 5 x 5, 4 x 4, 3 x 3, 2 x 2, 1 x 1 |
| 9 | 8 x 8, 7 x 7, 6 x 6, 5 x 5, 4 x 4, 3 x 3, 2 x 2, 1 x 1 |
| 10 | 9 x 9, 8 x 8, 7 x 7, 6 x 6, 5 x 5, 4 x 4, 3 x 3, 2 x 2, 1 x 1 |

4 RESULTADOS

Nessa seção serão apresentados os principais resultados alcançados ao longo dessa pesquisa e discussão analítica de cada um dos valores obtidos. Todo o projeto consistiu no uso das seguintes ferramentas computacionais:

- Software MatLAB 2012a - Usado como ferramenta principal com o uso do *toolbox* de Redes Neurais, foram criados *scripts* para pré-processamento dos dados, automatização de funções de treinamento, armazenamento de resultados e *plot* de gráficos.
- LibSVM 3.18 - Usados os conjuntos de ferramentas para leitura do conjunto de dados disponibilizado pelo autor original (formato padrão libSVM) e uso de funções de normalização e separação de dados (CHANG; LIN, 2011).
- Shell Script, SED, AWK - Conjunto de linguagens de *script* e programação usados para a extração de dados, formatação e padrão legível pelo *MATLAB*, divisão e unificação dos lotes, separação de dados, *plot* de gráficos, etc.
- LibreOffice Calc - Software de planilha para montagem de tabelas, visualização geral de resultados.
- SNNS e JavaNNS - *Stuttgart Neural Network Simulator* e sua versão gráfica em linguagem JAVA, são ferramentas que foram utilizadas na confecção e projeto da topologia das RNAs, com o estudo de funções de treinamento e análise de épocas.

4.1 TREINAMENTO

A pesquisa desenvolvida consistiu no uso de apenas classificadores RNA. As topologias utilizadas foram definidas na seção *Metodologia*. Em relação aos dados, 70% foram utilizados para treinamento, 15% para validação e 15% para testes para cada RNA treinada em cada lote específico. Uma RNA treinada nunca será usada na classificação de uma mesma amostra de um lote na qual foi treinada.

Devido ao fato de alguns lotes possuírem baixa quantidade de dados, foi escolhido o percentual relativamente alto (70%) a fim de obter um grau de treinamento e generalização adequado para cada RNA. O número de épocas de treinamento é obtido através de análise gráfica do erro de cada função de treinamento distinta, onde o critério de parada consiste na estabilização da queda do erro e se faz por inspeção visual no gráfico de erros x épocas da ferramenta MATLAB.

Além do mais foram utilizadas 4 funções de treinamento iniciais, para verificação de performance média das RNAs desenvolvidas, obtendo-se um tempo de treinamento distinto para cada uma de acordo com a Tabela 5:

Tabela 5 – Tabela com o comparativo do tempo de treinamento usando diferentes algoritmos

| Função de treinamento | Tempo médio de treino por RNA [min] |
|--------------------------------------------|-------------------------------------|
| Levenberg–Marquardt (LM) | 120 |
| Gradiente Conjugado Escalar (SCG) | 8 |
| Gradiente Conjugado de Polak-Ribière (CGP) | 11 |
| Retropropagação resiliente (RP) | 5 |

Usando o algoritmo de treinamento LM, a quantidade de épocas apesar de ser baixa, apresentou um tempo de convergência muito alto, inviabilizando a aplicação em larga escala.

Foram obtidas taxas de acertos muito variáveis a cada execução com o teste das funções RP e CGP, que apresentam baixo tempo de convergência. Desta maneira foi utilizada a função de treinamento SCG, pois além de obter resultados parecidos, possui baixo tempo de treinamento e alta estabilidade na classificação de uma mesma amostra, isto é, RNAs treinadas iguais para um mesmo conjunto de dados, mostram baixo desvio padrão. Uma explicação mais detalhada é dada na seção *Fundamentação Teórica* e é explicada em virtude da topologia de RNA escolhida.

Cada resultado é dado pela média dos resultados finais da classificação de 10 RNAs distintas, com inicialização de pesos aleatória em cada.

4.2 MÉTRICAS DE AVALIAÇÃO

O critério de avaliação de desempenho dos classificadores desenvolvidos na tarefa de compensar o problema de deriva em sensores de gases é baseado no percentual de acerto de acordo com:

$$y[t] = \frac{Qtd(acertos(t))}{Qtd(T)} \quad (4.1)$$

Onde $y[t]$ é o percentual de acerto médio das amostras do conjunto de dados que se deseja classificar no lote t e é dado pela divisão do número de acertos usando aquele conjunto de dados pela quantidade total de amostras no lote. Essa quantidade é sempre calculada com base na quantidade de amostras classificadas corretamente pelo total de amostras na qual tentou-se classificar.

A explicação dada pelo uso dessa simples métrica é:

1. Ter como forma de comparar resultados que utilizam a mesma base de dados na literatura com o mesmo critério de avaliação;
2. Ter uma análise que simbolize o ganho por lote. Como nesse caso cada modelo construído é aplicado na classificação de um lote distinto, não é simples estabelecer um indicador único que possa demonstrar o grau de erro na classificação e desvio.

De forma geral o uso de um indicador percentual por conjunto de dados, permite a visualização gráfica do fenômeno de deriva, onde é possível perceber sua ocorrência. A Figura 18 exemplifica a diminuição da taxa de acerto para lotes subsequentes quando o treinamento é realizado com o conjunto de dados contidos no lote 1.

Dessa forma, a queda na taxa de acerto de maneira contínua com o passar do tempo, para cada um dos dez lotes de dados disponível caracteriza a ocorrência da deriva nos sensores de gases utilizados no experimento.

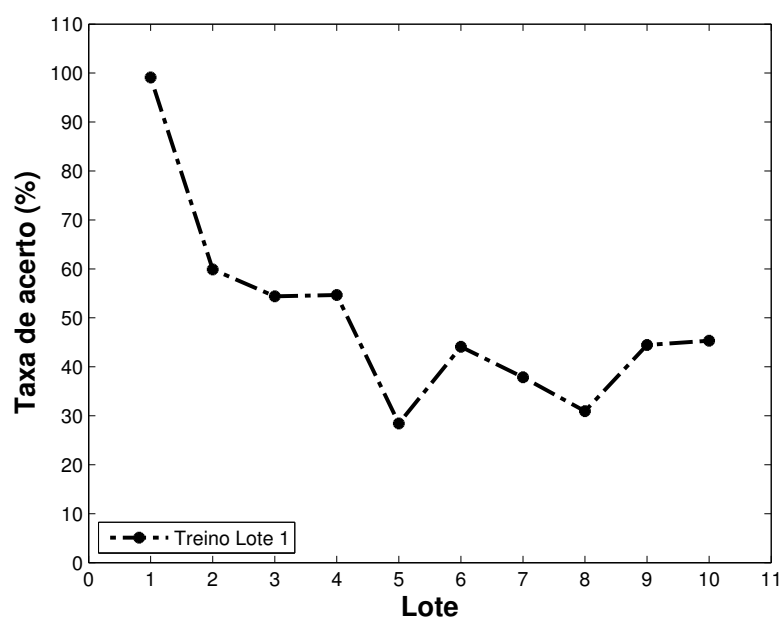


Figura 18 – Queda na taxa de classificação correta quando se efetua o treinamento apenas com o lote 1.

4.3 RNAs DE SAÍDAS FRAGMENTADAS - RNAs DE UM NEURÔNIO DE SAÍDA

Fragmentar as saídas das RNAs, consiste conforme descrito na seção de Metodologia, combinar diversas RNAs distintas com apenas um neurônio de saída, cujo objetivo é de aprimoramento dos resultados finais o qual se deseja alcançar. Com a finalidade de se classificar 6 gases distintos, são treinadas 6 RNAs para classificação de uma amostra contida em cada lote, o neurônio da RNA que tiver maior fator de ativação corresponderá ao gás classificado.

Não foi utilizada a abordagem de construção de um comitê de classificadores em conjunto da técnica de ponderação da entrada da RNA por dois motivos:

- Desempenho muito próximo ao uso de RNAs de múltiplas saídas (RNA única). Como o desenvolvimento deste trabalho consiste em uma abordagem comparativa e não absoluta, a taxa de acerto absoluta não apresenta elevado grau de importância.
- O crescimento exponencial ao longo do tempo da quantidade de amostras implica na necessidade de treino de RNAs com uma quantidade muito grande de amostras, porém multiplicado o tempo por 6, sendo uma RNA por gás, logo o uso dessa abordagem não se tornou interessante devido ao longo tempo de treinamento da RNA. A classificação e simulação dos 10 lotes de dados nesse caso poderia levar mais de 6 horas (tempo estimado levando em consideração o tempo de treino de uma RNA para o lote 10 levar 40 minutos).

4.4 PONDERAÇÃO DA SAÍDA DA REDE NEURAL

Os resultados obtidos pelo uso desse método baseiam-se na aplicação de cada uma das 4 curvas descritas pela Figura 14 nos resultados obtidos pelos neurônios de saídas das RNAs em cada uma das aplicações sobre o lote em questão. Para fins de comparação, são adotadas as comparações relativas ao teste de redes neurais treinadas com o lote anterior ($t - 1$) conforme a Figura 19 e também com todos os lotes anteriores, ou seja, o intervalo $[1 : t - 1]$, conforme a Figura 20 de forma que possa ser possível avaliar a eficácia do sistema de uso dos pesos dinâmicos propostos na compensação da deriva que ocorre ao longo do tempo em sensores de gases.

Para os testes de RNAs treinadas com o lote anterior, $t - 1$, por exemplo, na classificação de uma amostra que está contida no intervalo temporal equivalente ao lote 5, irá usar para sua classificação uma ou mais RNAs que foram treinadas com o lote 4. Por sua vez no teste que consiste no uso de todos os lotes anteriores do intervalo $[1 : t - 1]$, por exemplo na classificação de uma amostra que está no mesmo intervalo temporal contido no lote 5, irá usar a mesma quantidade de pesos e RNAs que foram treinadas com os lotes 1, 2, 3 e 4 e assim por diante.

Na Figura 20, é possível observar os resultados obtidos no uso desses testes, comparando a aplicação de pesos do lote anterior em relação à abordagem de uso dos pesos quadráticos 1. É possível notar uma queda acentuada na classificação do lote 6 em relação a outros. A correta taxa de classificação desse lote tem valores inferiores a 60% em todos os testes aplicados, de forma que há uma diferenciação enorme levando em consideração a taxa de acerto de lotes anteriores e posteriores, onde o valor esperado seria em torno de 80%, ou seja, valor próximo da média de

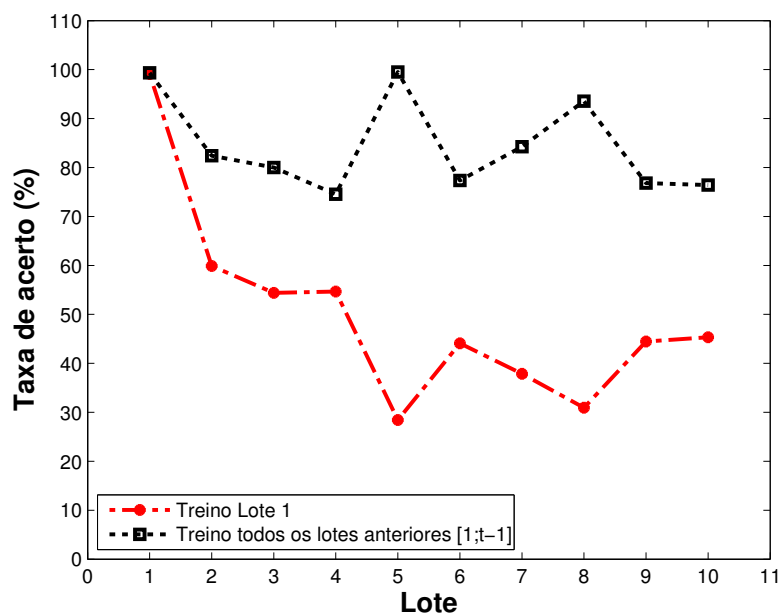


Figura 19 – Queda na taxa de acerto com treino usando o lote 1 e também com o uso do conjunto de dados inteiro

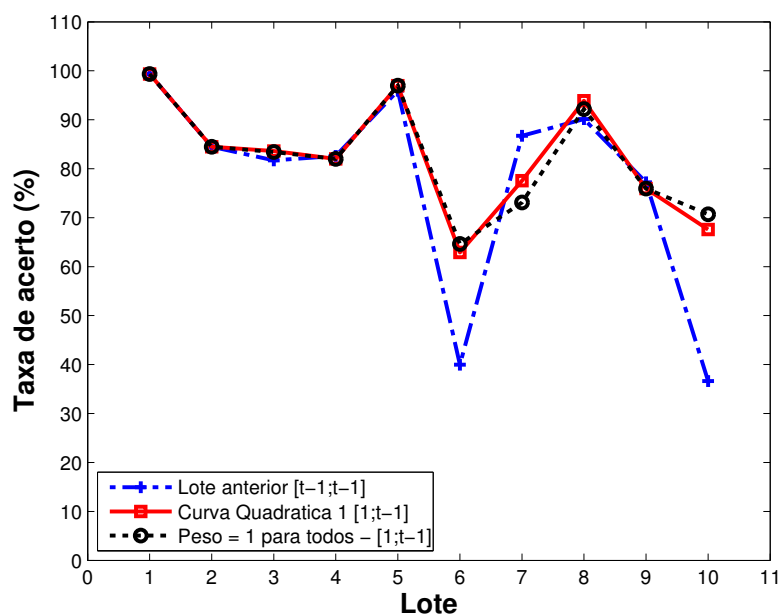


Figura 20 – Comparativo entre rede neural treinada com o lote anterior, pesos quadráticos 1 e todos os lotes anteriores.

classificação para os lotes anteriores. Duas razões podem ser levantadas em relação a esse comportamento. A primeira é relativa a possível falta de dependência no problema das características presentes no lote 6 em relação aos lotes anteriores. Outra razão é que a quantidade de amostras presentes para treinamento de RNAs do lote 6 é muito superior em relação à quantidade de amostras encontradas em todos os lotes anteriores conforme pode ser verificado na Tabela 3.

Nota-se ainda na mesma Figura 20 que melhores resultados foram alcançados

em 8 dos 10 lotes testados, inclusive no lote 6 com a aplicação de ponderação da saída e a curva quadrática 1.

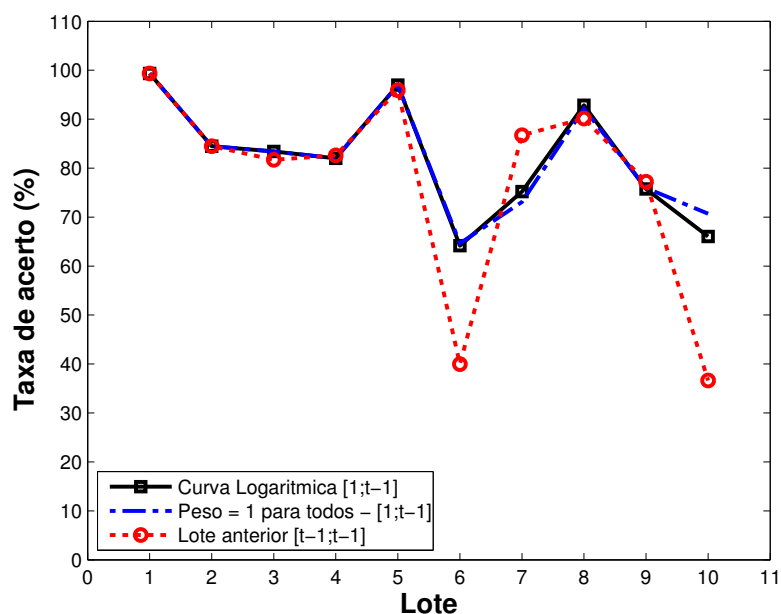


Figura 21 – Comparativo entre usos de pesos logarítmicos, comparadas ao uso do treinamento com o lote anterior e com todos os lotes e mesmos pesos

A Figura 21 traz resultados provenientes da aplicação de pesos designados pela curva logarítmica proposta na seção anterior e o uso de pesos fixos para todos os lotes usados com todas as RNAs provenientes de intervalos de tempo anteriores à t com pesos iguais para todos. Inicialmente nota-se um comportamento igual na taxa de classificação dos 6 primeiros lotes com o mesmo comportamento e argumento em relação ao lote 6 da queda acentuada na taxa de acerto. Logo, em seguida é possível notar uma leve melhora em relação à classificação de amostras contidas no intervalo temporal situados nos lotes 7,8 e 9 respectivamente, porém sem grandes ganhos na classificação dos mesmos.

Já a Figura 22 traz o comparativo da aplicação do conjunto de classificadores com o uso de ponderações no formato da segunda curva quadrática contra o aumento exponencial. Esta última projeção traz o fator de importância de amostras mais recentes, logo sem ignorar a presença de dados provenientes de lotes onde os sensores de gases eram mais novos. Inicialmente é possível notar que o uso da curva exponencial mostrou ganho aproximado de quase 90% no lote 3, sendo o melhor resultado encontrado entre todos os métodos estabelecidos. Apesar da baixa taxa de acerto no lote 6, onde a curva quadrática 2 apresentou quase 60% de acerto, resultados melhores são obtidos em todos os outros lotes, excluindo o lote 10.

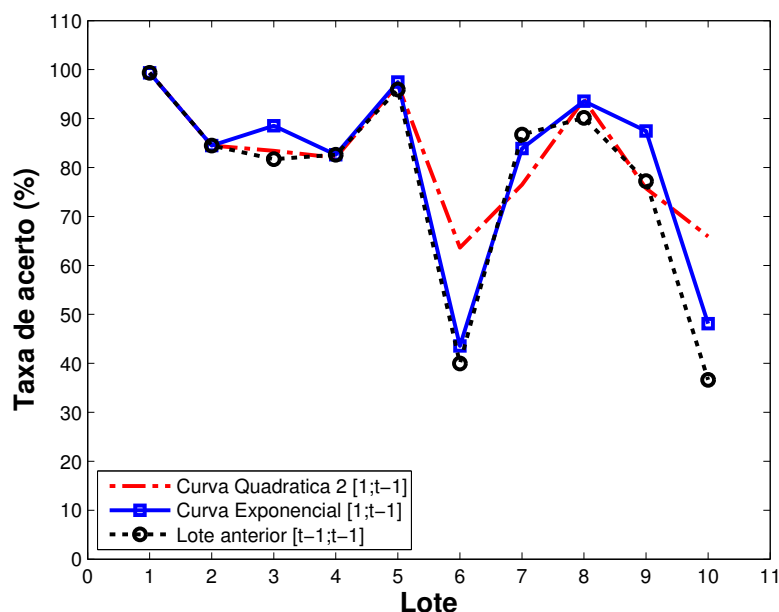


Figura 22 – Comparativo entre usos de pesos quadráticos 2 e exponenciais, comparadas ao uso do treinamento com o lote anterior e com todos os lotes e mesmos pesos

4.5 PONDERAÇÃO DA ENTRADA DA REDE NEURAL

Os resultados obtidos com a ponderação das entradas das RNAs se baseiam na aplicação da Equação 3.11. Nela aplica-se diretamente o conjunto de dados originais para obtenção de um novo conjunto ponderado. O formato dos dados obtidos para treinamento da RNA são baseados de acordo com a distribuição da Tabela 4.

A principal diferença é que não há uma RNA treinada para cada lote. O treinamento numa situação real de funcionamento é *online*, ou seja, na classificação de uma amostra que esteja temporalmente inserida num lote na qual não há RNA disponível, é treinada em tempo de execução uma nova RNA para classificação do lote t que por sua vez o treino é composto pela alteração do conjunto de dados disponíveis no passado.

A Figura 23 mostra a taxa de acerto de cada uma das RNAs especialistas treinadas com conjunto de dados ponderados. Como resultados em que é possível destacar, verifica-se uma taxa de acerto próxima de 100% para o lote 5, alcança-se 68.43% para o lote 6, valores superiores a anteriores e também em relação a outros métodos, taxas acima de 70% sempre para os lotes 7,8 e 9. Uma observação importante nesse caso é a taxa de acerto próxima de 80% para o lote 10, um resultado muito acima de outros classificadores, onde o valor máximo atingido é em torno de 60%.

A Figura 24 demonstra o crescimento do conjunto de dados ao longo do tempo de maneira acumulada. Ao ser aplicada a Equação 3.11 ao longo de todos os conjuntos de lotes de dados, e em consequência do envelhecimento dos sensores de gases,

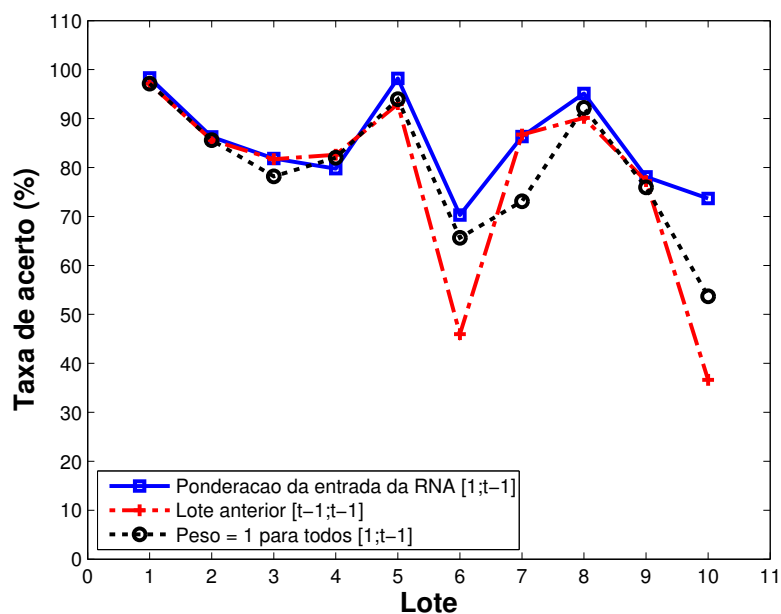


Figura 23 – Taxa de acerto média do uso da técnica de ponderação das entradas da RNA comparadas ao uso do treinamento com o lote anterior e com todos os lotes e mesmos pesos

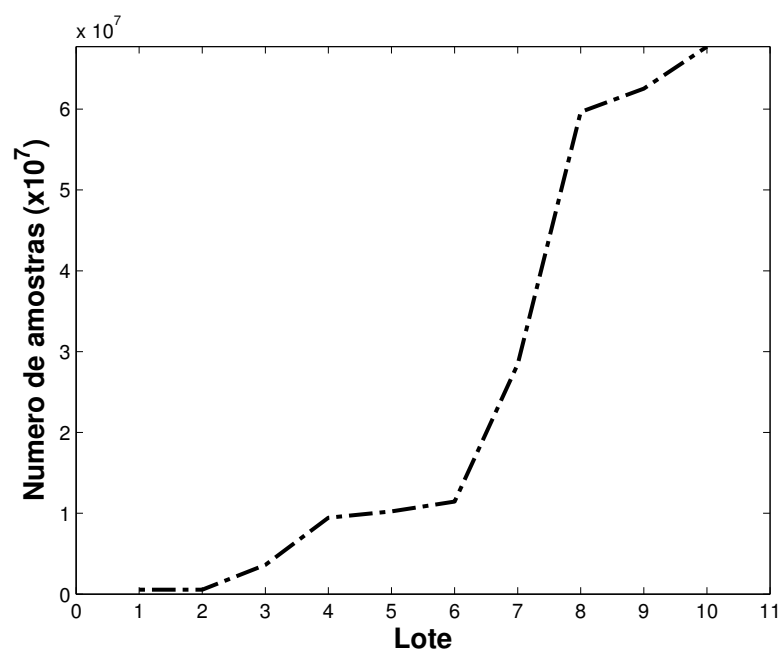


Figura 24 – Crescimento acumulado da quantidade total de amostras ao longo do tempo

maior será o conjunto de dados ponderado em relação à quantidade de amostras do conjunto inicial.

Ainda pela Figura 24 é possível visualizar um comportamento similar a um crescimento exponencial. É importante analisar nesse caso, a grande taxa de crescimento de amostras ao longo do tempo no conjunto de dados, irá implicar num aumento acentuado no tempo de treinamento da RNA. Dessa forma é possível concluir que a

partir do instante em que o sistema se torna mais velho, este apresentará um maior conjunto de dados para treinamento e auxílio na compensação de deriva.

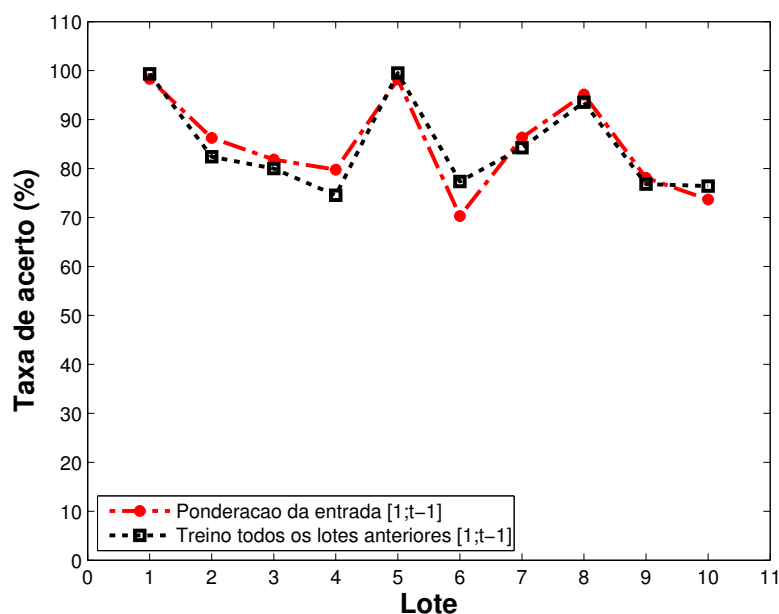


Figura 25 – Taxa de acerto média do uso da técnica de ponderação das entradas da RNA comparadas ao treinamento de uma RNA com todos os dados anteriores ao lote atual

Na Figura 25 é possível observar os resultados provenientes da comparação entre o método de ponderação da entrada da RNA com a técnica de treino de apenas uma RNA por lote, com todos os dados passados sem ponderação. O novo método aplicado para evidenciar amostras mais importantes foi superior na maioria dos casos, tendo um desempenho inferior na classificação dos lotes 6 e 10. Como destaque principal, há um ganho em relação aos outros métodos de compensação na classificação de lotes de dados com ocorrência grande de deriva, como por exemplo os lotes 9 e 10.

4.6 COMPARATIVO

Nessa seção final, é mostrado na Figura 26 o resultado comparativo de todas as curvas aplicadas nas saídas das RNAs e o método de ponderação da entrada da RNA. Devido a resultados muito próximos utilizando RNAs com múltiplas saídas, como RNAs com um neurônio na camada de saída, foi incluído no gráfico apenas os resultados da primeira topologia.

A Tabela 6 disponibiliza os números alcançados para cada lote de dados testados ao longo dos 36 meses de coleta evidenciando-se em vermelho os melhores resultados alcançados. Analisando detalhadamente, é possível visualizar que a taxa de acerto é idêntica na maioria dos métodos quando aplicados para classificação de um dado proveniente do lote 1. A explicação é que como não há dados mais

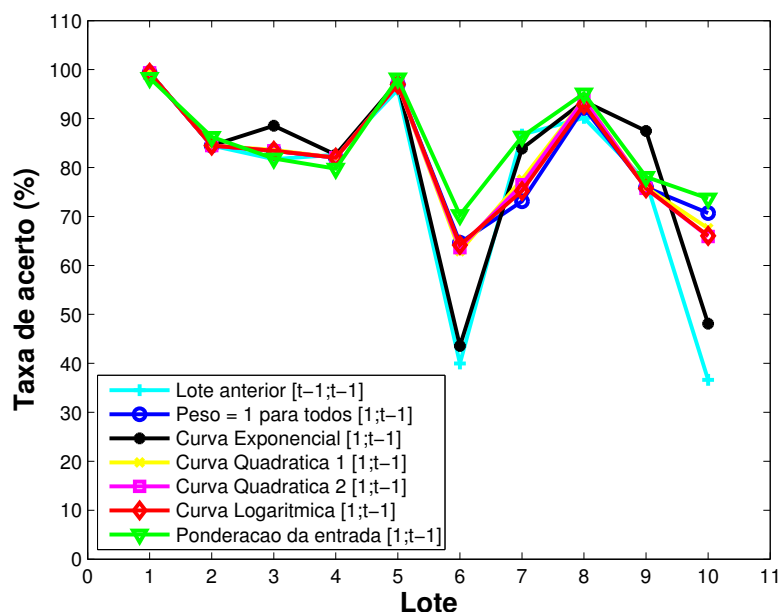


Figura 26 – Gráfico com o resumo de todos os resultados alcançados com a topologia de múltiplos neurônios de saída

antigos temporalmente do que o próprio lote, o treino das RNAs para formação dos classificadores é feito com um pedaço do mesmo conjunto de dados, mas testados com outra parte. Situações onde é necessário classificar um dado proveniente do lote 2 apresentam o mesmo sintoma anterior, é usado apenas o conjunto de dados prévio temporalmente nessa classificação, logo a aplicação de um fator de esquecimento ou ponderação é igual para qualquer curva projetada, pois não é dada ênfase em nenhuma amostra em relação à outra, logo a aplicação de fatores de esquecimento é ineficaz. Melhores resultados são obtidos com a aplicação de pesos na saída de RNAs para os primeiros lotes, enquanto o método de ponderação dos dados na entrada da RNA para treinamento e classificação apresentou resultados superiores para os últimos lotes. Uma análise que pode ser feita em relação à baixa taxa de acerto encontrada na classificação de dados contidos no lote 10 é a sua baixa dependência temporal em relação a lotes anteriores e principalmente que passaram quase 3 anos desde a coleta de dados inicial, a ocorrência de deriva é tão grande que a compensação é ineficiente, destaca-se o fato que houve uma pausa de 6 meses sem coleta de dados entre o lote 9 e 10 para fins de estudo segundo o autor do conjunto de dados.

Como resultado final nota-se que a aplicação do método de replicação de amostras nas entradas das RNAs (PPRA), foi o que apresentou melhores resultados. De todos os métodos de ponderação testados para esse conjunto de dados, tanto para entradas ou saídas, o método PPRA apresentou a melhor taxa de acerto em 5 dos 10 lotes de dados.

Tabela 6 – Tabela com resumo de valores alcançados com todos os testes e combinações usando RNAs com múltiplos neurônios de saída

| Lote | [t-1;t-1] | Peso = 1 [1;t-1] | Peso=exp [1;t-1] | Peso=Quad1 [1;t-1] | Peso=Quad2 [1;t-1] | Peso=Log [1;t-1] | PPRA |
|------|--------------|---------------------|---------------------|-----------------------|-----------------------|---------------------|--------------|
| 1 | 99.33 | 99.33 | 99.33 | 99.33 | 99.33 | 99.33 | 98.3 |
| 2 | 84.49 | 84.49 | 84.49 | 84.49 | 84.49 | 84.49 | 86.24 |
| 3 | 81.72 | 83.42 | 88.52 | 83.61 | 83.42 | 83.42 | 81.84 |
| 4 | 82.61 | 81.99 | 82.61 | 81.99 | 81.99 | 81.99 | 79.78 |
| 5 | 95.94 | 96.95 | 97.46 | 96.95 | 96.95 | 96.95 | 98.19 |
| 6 | 39.96 | 64.65 | 43.57 | 62.91 | 63.65 | 64.17 | 70.29 |
| 7 | 86.74 | 73.1 | 83.89 | 77.55 | 76.5 | 75.2 | 86.32 |
| 8 | 90.14 | 92.18 | 93.54 | 93.88 | 93.88 | 92.86 | 95.14 |
| 9 | 77.23 | 75.96 | 87.45 | 75.96 | 75.74 | 75.74 | 78.10 |
| 10 | 36.64 | 70.69 | 48.11 | 67.58 | 65.92 | 66.06 | 73.66 |

4.7 COMPARAÇÃO COM MELHORES RESULTADOS ENCONTRADOS NA LITERATURA

A Figura 27 mostra para fins de comparação, o método de melhor resultado neste trabalho (PPRA), com outros três melhores resultados encontrado na literatura:

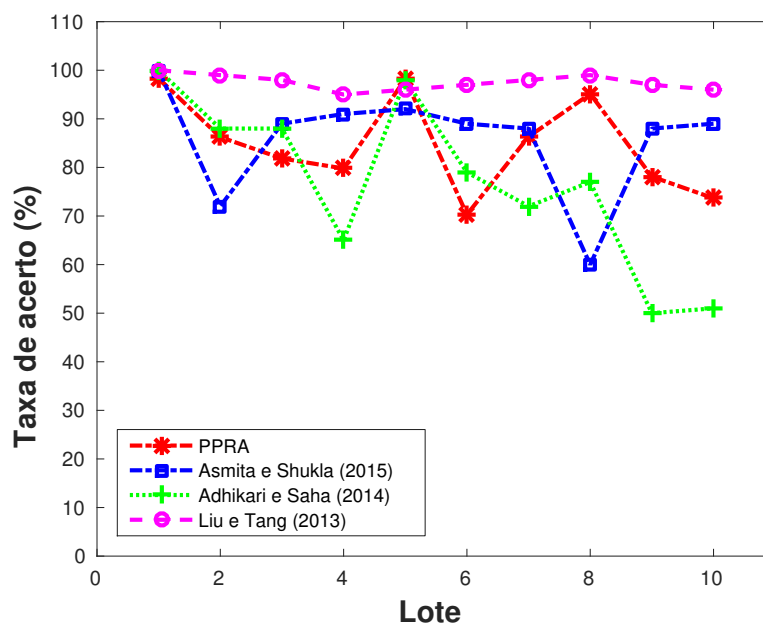


Figura 27 – Gráfico com o comparativo do melhor resultado alcançado (PPRA) e os três melhores resultados encontrados na literatura

É possível notar que o uso de pesos dinâmicos de Liu e Tang (2013) apresenta a melhor taxa de acerto. O sistema consiste no ajuste dinâmicos de pesos para um conjunto de SVMs. Praticamente todas as taxas de acerto estão próximas à 100%. Como desvantagens desse método há a necessidade de ajustar a quantidade de pesos com base numa busca transversal (do inglês *traversal search*), ou seja, são necessários encontrar pesos ótimos para cada problema. Dessa maneira não há garantia que o modelo funcione em outros conjuntos de dados. A segunda desvantagem

é a necessidade de se treinar mais de uma SVM para a classificação, implicando numa perda de desempenho.

O método de regularização de pesos proposto por Asmita e Shukla (2015) apresenta uma leve melhora na taxa de acerto comparado ao método PPRA. O funcionamento é dado pelo estabelecimento de um vetor de regularização que é previsto com base em conhecimento acumulado. Em posse do vetor de regularização, são treinadas diversas SVMs que são combinadas majoritariamente de acordo com o vetor. A grande desvantagem desse método é a necessidade de o sistema ser 100% supervisionado para o funcionamento inicial.

Com o uso de aprendizagem profunda (do inglês *Deep Learning*), Hu *et al.* (2014) utiliza um mecanismo de aprendizado recorrente baseado numa Máquina Restrita de Boltzmann. Essa técnica utiliza o fator caótico de partículas na termodinâmica para o treinamento de RNAs recorrentes. Apresenta resultados consideráveis, porém com grande complexidade a cada treinamento.

Os métodos propostos, apresentaram resultados intermediários comparado a outros na literatura. Entretanto é importante destacar as seguintes forças desse trabalho executado:

- Capacidade de modelagem de uma curva de pesos que podem ser equacionáveis;
- Simplicidade na fase de treinamento;
- Não usa combinação de classificadores de diferentes tipos com voto majoritário para melhores ganhos;
- Simplicidade de aplicação;
- Ponderação de entrada de amostras é lenta mas eficaz.

No primeiro item nota-se que é possível estender a técnica para outros conjuntos de dados com ocorrência de deriva. Existe a capacidade de aplicar as mesmas curvas para conjuntos de dados maiores ou que não sofram o problema de "degrau" de deriva ocasionado pela disponibilidade dos dados em lotes.

O treinamento tanto usando as técnicas de ponderação de entradas e saídas de RNAs é simples, não sendo necessário intervenção ou adição de passos além do cuidado na fase de generalização do treinamento.

Apenas um classificador é implementado para cada amostra que deseja-se classificar. Nesse trabalho não foi adotada a técnica de inflar o número de classificadores para aumento da taxa de acerto.

Uma vez treinado o classificador, se utilizada a técnica de ponderação de saída das RNAs, basta multiplicar os neurônios de saída por um vetor de pesos. Se a técnica utilizada é a ponderação de entrada, o classificador aprendeu a deriva com base nas amostras a qual foi treinado.

Apesar da lentidão no treinamento de RNAs e compensação de deriva por replicação de amostras de entrada, este método foi o que se mostrou mais eficaz comparado aos outros testados.

5 CONCLUSÃO

A pesquisa teve seu foco no estudo de dois métodos de ponderação aplicados usando RNA do tipo MLP para compensação do problema de deriva em sensores de gases. Ambas as aplicações com RNA se baseiam em topologias com várias saídas, e a combinação de RNA com uma única saída formando um comitê de classificadores. O primeiro método de ponderação usa uma combinação de pesos sobre diversas curvas aplicados na saída da RNA com diferentes lotes de dados de amostras obtidas no passado. O segundo método faz a replicação de dados da entrada aplicada na RNA com finalidade de dar maior importância às informações contidas em lotes mais recentes.

A melhora na taxa de classificação foi verificada usando ambos os métodos. O uso de taxas de esquecimento aplicadas na entrada, através do método PPRA, proveu melhores resultados na maioria dos lotes. Este método de dar maior importância aos dados replicando a informação da entrada no treinamento da RNA produziu melhores resultados por utilizar amostras recentes, ou seja, dos últimos lotes. É notável que ambos os métodos apresentam resultados melhores do que o simples uso do lote de dados anterior ou do uso de todos os lotes anteriores como é observada com maior frequência na literatura.

A aplicação do comitê de classificadores com RNAs de uma saída ao invés do uso de uma ou várias RNAs de múltiplos neurônios na última camada não mostrou melhores resultados que a abordagem tradicional de uma RNA. Em lotes mais antigos, onde os sensores apresentam baixa incidência da ocorrência de deriva os resultados são semelhantes para os três primeiros lotes de dados, sendo que após isso há uma queda acentuada na correta taxa de classificação em relação ao uso de uma RNA unificada.

É importante destacar a força do método de replicação de amostras na entrada da RNA. Tal abordagem não foi encontrada na literatura na área de compensação de deriva. Deve-se destacar a obtenção de bons resultados alcançados devido a simplicidade da técnica. É necessário também efetuar mais testes em conjuntos de dados distintos onde há presença de deriva para aprimoramento do método.

Nesse estudo foi verificada a interferência da deriva sobre um lote de dados de sensores de gases ao longo do tempo. Taxas de acerto mais altas que as mostradas são obtidas na literatura com o uso de algoritmos de busca para obtenção de pesos ótimos, como AGs, DTs e outros atrelados à combinação de classificadores dos mais diversos tipos. Deve-se, no entanto verificar, que há um limite máximo da capacidade

de compensação de deriva na aplicação em sensores de gases devido à natureza química do envelhecimento. Dessa forma, diferentes tipos e tecnologias de sensores devem ser desenvolvidas.

5.1 CONTRIBUIÇÕES

Devido aos resultados alcançados com a aplicação das duas técnicas de fatores de esquecimento distintas, destacam-se as contribuições abaixo:

- Desenvolvimento de um método que pode regular o grau de importância de um conjunto de dados na tomada de decisão com base na ênfase dada às amostras mais recentes, com re-treino do classificador para compensação da deriva.
- Avaliação da aplicação de fatores de esquecimento diretamente nas saídas das redes neurais de forma que sejam variados os graus de importância de cada conjunto de amostras ao longo do tempo, para compensação de deriva.

5.2 TRABALHOS FUTUROS

Com base nos resultados obtidos, abrem-se diversas oportunidades de trabalhos futuros com a intenção de não apenas obter melhores taxas de acertos, mas características de robustez e confiabilidade dos métodos empregados, entre as quais se destacam:

- Simulação dos métodos utilizados com conjunto de dados de sensores com ocorrência de deriva.
- Criação de uma abordagem semisupervisionada com detecção de deriva automática, ou seja, *online*, para parada de classificação e retreino.
- Pré-processamento do conjunto de dados com a aplicação de um filtro em amostras que não apresentem relevância para o problema. Isto pode ser aplicado conforme demonstrado na tentativa de normalização de um rótulo dentro de um lote com o deslocamento próximo à média do conjunto como um todo.
- Uso de algoritmos de *boosting* como por exemplo *ADA Boost* com a finalidade de melhoria de forma dinâmica dos pesos em tempo de execução.

5.3 PRODUÇÃO CIENTÍFICA

WIEZBICKI, T.; RIBEIRO, E. P. Sensor drift compensation using weighted neural networks. In: 2016 IEEE Conference on Evolving and Adaptive Intelligent Systems (EAIS).[S.l.: s.n.], 2016. p. 92–97.

REFERÊNCIAS

- ADHIKARI, S.; SAHA, S. Multiple classifier combination technique for sensor drift compensation using ANN and KNN. In: *Advance Computing Conference (IACC), 2014 IEEE International*. [S.l.: s.n.], 2014. p. 1184–1189.
- ARTURSSON, T.; EKLÖV, T.; LUNDSTRÖM, I.; MÅRTENSSON, P.; SJÖSTRÖM, M.; HOLMBERG, M. Drift correction for gas sensors using multivariate methods. *Journal of chemometrics*, Wiley Online Library, v. 14, n. 5-6, p. 711–723, 2000.
- ASMITA, S.; SHUKLA, K. K. Sensor Drift Compensation in Time Series Prediction through Regularized Ensemble of Classifiers. *IJARCCCE - International Journal of Advanced Research in Computer and Communication Engineering*, v. 4, n. 2, p. 63–70, feb 2015.
- BARSAN, N.; KOZIEJ, D.; WEIMAR, U. Metal oxide-based gas sensor research: How to? *Sensors and Actuators B: Chemical*, v. 121, n. 1, p. 18–35, jan 2007. ISSN 09254005.
- BERMAK, A.; BELHOUARI, S.; SHI, M.; MARTINEZ, D. Pattern recognition techniques for odor discrimination in gas sensor array. *The Encyclopedia of Sensors*, n. June 2015, 2005. ISSN 158883056X.
- CHANG, C.-C.; LIN, C.-J. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, v. 2, p. 27:1–27:27, 2011. Software available at <<http://www.csie.ntu.edu.tw/~cjlin/libsvm>>.
- ESCOVEDO, T. *Aprendizagem Neuroevolutiva e Detecção de Concept Drift em Ambientes Não Estacionários*. Tese (Doutorado) — PUC-Rio, 2015.
- FONOLLOSA, J.; RODRÍGUEZ-LUJÁN, I.; HUERTA, R. Chemical gas sensor array dataset. *Data in Brief*, Elsevier, v. 3, n. 0, p. 85–89, 2015. ISSN 2352-3409.
- FONOLLOSA, J.; VERGARA, A.; HUERTA, R. Algorithmic mitigation of sensor failure: Is sensor replacement really necessary? *Sensors and Actuators, B: Chemical*, v. 183, n. 0, p. 211–221, 2013. ISSN 09254005.
- GAMA, J. *Knowledge discovery from data streams*. [S.l.]: CRC Press, 2010.
- GAMA, J.; MEDAS, P.; CASTILLO, G.; RODRIGUES, P. Learning with Drift Detection. In: BAZZAN, A. L. C.; LABIDI, S. (Ed.). Berlin, Heidelberg: Springer Berlin Heidelberg, 2004. cap. Learning w, p. 286–295. ISBN 978-3-540-28645-5.
- GUTIERREZ-OSUNA, R. Drift reduction for metal-oxide sensor arrays using canonical correlation regression and partial least squares. In: *Proceedings of the 7th International Symposium On Olfaction & Electronic Nose*. [S.l.: s.n.], 2000. p. 1–7.
- HOLMBERG, M.; WINQUIST, F.; LUNDSTRÖM, I.; DAVIDE, F.; DINATALE, C.; D'AMICO, A. Drift counteraction for an electronic nose. *Sensors and Actuators B: Chemical*, Elsevier, v. 36, n. 1, p. 528–535, 1996.
- HORNIK, K.; STINCHCOMBE, M.; WHITE, H. Multilayer feedforward networks are universal approximators. *Neural networks*, Elsevier, v. 2, n. 5, p. 359–366, 1989.

- HU, X.; LIU, Q.; CAI, H.; LI, F. Gas Recognition Under Sensor Drift by Using Deep Learning. In: WEN, Z.; LI, T. (Ed.). *Practical Applications of Intelligent Systems*. [S.l.]: Springer Berlin Heidelberg, 2014, (Advances in Intelligent Systems and Computing, v. 279). p. 23–33. ISBN 978-3-642-54926-7.
- HULTEN, G.; SPENCER, L.; DOMINGOS, P. Mining time-changing data streams. In: ACM. *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*. [S.l.], 2001. p. 97–106.
- KADRI, C.; TIAN, F.; ZHANG, L.; LI, G.; DANG, L.; LI, G. Neural network ensembles for online gas concentration estimation using an electronic nose. *Int. J. Comput. Sci*, v. 10, n. 2013, p. 10, mar 2013.
- LESHNO, M.; LIN, V. Y.; PINKUS, A.; SCHOCKEN, S. Multilayer feedforward networks with a nonpolynomial activation function can approximate any function. *Neural networks*, Elsevier, v. 6, n. 6, p. 861–867, 1993.
- LIU, H.; TANG, Z. Metal Oxide Gas Sensor Drift Compensation Using a Dynamic Classifier Ensemble Based on Fitting. *Sensors*, Multidisciplinary Digital Publishing Institute, v. 13, n. 7, p. 9160–9173, 2013.
- LLOBET, E.; BREZMES, J.; VILANOVA, X.; SUEIRAS, J. E.; CORREIG, X. Qualitative and quantitative analysis of volatile organic compounds using transient and steady-state responses of a thick-film tin oxide gas sensor array. *Sensors and Actuators B: Chemical*, v. 41, n. 1–3, p. 13–21, 1997. ISSN 0925-4005.
- LOURAKIS, M. I. A brief description of the levenberg-marquardt algorithm implemented by levmar. *Foundation of Research and Technology*, v. 4, n. 1, 2005.
- LUGHOFFER, E.; WEIGL, E.; HEIDL, W.; EITZINGER, C.; RADAUER, T. Drift detection in data stream classification without fully labelled instances. In: *Evolving and Adaptive Intelligent Systems (EAIS), 2015 IEEE International Conference on*. [S.l.: s.n.], 2015. p. 1–8.
- MØLLER, M. F. A scaled conjugate gradient algorithm for fast supervised learning. *Neural networks*, Elsevier, v. 6, n. 4, p. 525–533, 1993.
- MORÉ, J. J. The levenberg-marquardt algorithm: implementation and theory. In: *Numerical analysis*. [S.l.]: Springer, 1978. p. 105–116.
- PARDO, M.; SBERVEGLIERI, G. Remarks on the use of multilayer perceptrons for the analysis of chemical sensor array data. *IEEE Sensors Journal*, IEEE, v. 4, n. 3, p. 355–363, 2004.
- POLYAK, B. T. The conjugate gradient method in extremal problems. *USSR Computational Mathematics and Mathematical Physics*, Elsevier, v. 9, n. 4, p. 94–112, 1969.
- RIEDMILLER, M.; BRAUN, H. A direct adaptive method for faster backpropagation learning: The rprop algorithm. In: IEEE. *Neural Networks, 1993., IEEE International Conference on*. [S.l.], 1993. p. 586–591.

- RODRIGUEZ-LUJAN, I.; FONOLLOSA, J.; VERGARA, A.; HOMER, M.; HUERTA, R. On the calibration of sensor arrays for pattern recognition using the minimal number of experiments. *Chemometrics and Intelligent Laboratory Systems*, Elsevier B.V., v. 130, n. 0, p. 123–134, 2014. ISSN 0169-7439.
- ROMAIN, A.-C. C.; NICOLAS, J. Long term stability of metal oxide-based gas sensors for e-nose environmental applications: An overview. *Sensors and Actuators B: Chemical*, Elsevier, v. 146, n. 2, p. 502–506, 2010.
- SAHM, T.; GURLO, A.; BARSAN, N.; WEIMAR, U. Basics of oxygen and NO_2 interaction; work function change and conductivity measurements. *Sensors and Actuators B: Chemical*, Elsevier, v. 118, n. 1, p. 78–83, 2006.
- SCHLIMMER, J. C.; GRANGER, R. H. Incremental learning from noisy data. *Machine learning*, Springer, v. 1, n. 3, p. 317–354, 1986.
- SEBASTIAO, R.; GAMA, J. A study on change detection methods. In: *Progress in Artificial Intelligence, 14th Portuguese Conference on Artificial Intelligence, EPIA*. [S.l.: s.n.], 2009. p. 12–15.
- SHAKER, A.; LUGHOFFER, E. Self-adaptive and local strategies for a smooth treatment of drifts in data streams. *Evolving Systems*, v. 5, n. 4, p. 239–257, 2014. ISSN 1868-6486.
- SHEIK, S.; MARCO, S.; HUERTA, R.; FONOLLOSA, J. Continuous prediction in chemoresistive gas sensors using reservoir computing. In: *Procedia Engineering*. [S.l.: s.n.], 2014. v. 87. ISSN 18777058.
- SHI, Z.-J.; SHEN, J. Convergence of the polak–ribière–polyak conjugate gradient method. *Nonlinear Analysis: Theory, Methods & Applications*, Elsevier, v. 66, n. 6, p. 1428–1441, 2007.
- SZCZUREK, A.; KRAWCZYK, B.; MACIEJEWSKA, M. VOCs classification based on the committee of classifiers coupled with single sensor signals. *Chemometrics and Intelligent Laboratory Systems*, v. 125, n. 0, p. 1–10, 2013. ISSN 0169-7439.
- VERGARA, A.; FONOLLOSA, J.; MAHIQUES, J.; TRINCAVELLI, M.; RULKOV, N.; HUERTA, R. On the performance of gas sensor arrays in open sampling systems using Inhibitory Support Vector Machines. *Sensors and Actuators, B: Chemical*, Elsevier B.V., v. 185, p. 462–477, 2013. ISSN 09254005.
- VERGARA, A.; VEMBU, S.; AYHAN, T.; RYAN, M. A.; HOMER, M. L.; HUERTA, R. Chemical gas sensor drift compensation using classifier ensembles. *Sensors and Actuators B: Chemical*, Elsevier, v. 166, p. 320–329, 2012.
- WANG, R.; SHI, L.; FOGHLU, M. O.; ROBSON, E. A meta-learning method for concept drift. 2010.
- WIEZBICKI, T.; RIBEIRO, E. P. Sensor drift compensation using weighted neural networks. In: *2016 IEEE Conference on Evolving and Adaptive Intelligent Systems (EAIS)*. [S.l.: s.n.], 2016. p. 92–97.
- WILLIAMS, D. E. Semiconducting oxides as gas-sensitive resistors. *Sensors and Actuators B: Chemical*, Elsevier, v. 57, n. 1, p. 1–16, 1999.

WILLIS, C. M.; CHURCH, S. M.; GUEST, C. M.; COOK, W. A.; MCCARTHY, N.; BRANSBURY, A. J.; CHURCH, M. R.; CHURCH, J. C. Olfactory detection of human bladder cancer by dogs: proof of principle study. *Bmj*, British Medical Journal Publishing Group, v. 329, n. 7468, p. 712, 2004.

YAMAZOE, N.; SHIMANOE, K. New perspectives of gas sensor technology. *Sensors and Actuators B: Chemical*, v. 138, n. 1, p. 100–107, apr 2009. ISSN 09254005.

ZHANG, L.; TIAN, F.; NIE, H.; DANG, L.; LI, G.; YE, Q.; KADRI, C. Classification of multiple indoor air contaminants by an electronic nose and a hybrid support vector machine. *Sensors and Actuators B: Chemical*, v. 174, n. 0, p. 114–125, 2012. ISSN 0925-4005.

ZLIOBAITE, I. Learning under concept drift: an overview. *arXiv preprint arXiv:1010.4784*, 2010.

APÊNDICE A – DETECÇÃO DE DERIVA

A.0.1 O algoritmo de Page-Hinkley

Abaixo é exemplificado o uso e funcionamento de um algoritmo para compensação do efeito de deriva. O algoritmo *Page-Hinkley* (PHT, do inglês *Page-Hinkley Test*) faz uma análise sequencial para monitorar de forma instantânea uma mudança. O ponto forte desse método é a detecção de mudanças na média de um sinal Gaussiano, possuindo eficiência na detecção de mudanças de comportamento num processo estabelecido por um modelo (LUGHOFER *et al.*, 2015). É considerado nesse teste a variável U_T que é a diferença acumulada entre os valores observados e a média no momento atual:

$$U_T = \sum_{t=1}^T (x_t - \bar{x}_T - \gamma) \quad (\text{A.1})$$

Na Equação A.1, $\bar{x}_T = 1/T \sum_{t=1}^T x_t$, que corresponde à média dada pelo sistema desde a primeira amostra até o instante atual t . A variável γ é a magnitude da mudança que é aceitável e deve ser determinada através de testes.

O Algoritmo 1 descreve passo a passo como é dada a aplicação do teste:

Algorithm 1 O algoritmo para detecção de deriva Page-Hinkley

Entradas: $x_1, x_2, \dots, x_T \implies X$

Entradas: $\gamma \implies \text{magnituda da deriva}$

Entradas: $\lambda \implies \text{limiar de detecção}$

Saídas: $t_{PH} \implies \text{TEMPO}$

```

1: procedure PAGEHINKLEY( $x, \gamma, \lambda$ )
2:   for  $t > 0$  do
3:      $\bar{x}_t = \sum_{t=1}^T x_t$ 
4:      $U_T = \sum_{t=1}^T (x_t - \bar{x}_T - \gamma)$ 
5:      $m_T = \min(U_t, t = 1..T)$ 
6:     if  $U_T - m_T \geq \lambda$  then retorne  $\implies t_{PH}$ 
7:   end if
8: end for
9: end procedure

```

▷ Para cada amostra
 ▷ a média até o instante t
 ▷ retorno t -> deriva

Aplica-se o teste para todas as amostras do intervalo $[0; T]$, onde para se detectar aumentos, é computado o mínimo valor de U_t de modo que $m_T = \min(U_t, t = 1..T)$ onde a cada passo monitora-se a diferença entre U_t e m_T , sendo $PH_T = U_T - m_T$, de modo que se a diferença PH_T é maior que o limiar λ , uma mudança de conceito é

detectada e alarmada. Se houver aumento do limiar λ , haverá diminuição dos alarmes falso positivos, porém irá aumentar o tempo necessário para detecção de deriva. A chave para bom funcionamento desse sistema consiste no bom controle do parâmetro λ para um balanço adequado entre velocidade e falso positivos (SEBASTIAO; GAMA, 2009).

A.0.2 Subdivisão dos dados por algoritmos de detecção de deriva

Com a finalidade de se comprovar que há existência de deriva, é aplicado o algoritmo PHT, que têm melhores resultados para dados cuja distribuição não é uniforme. Primeiramente através da Equação A.2 é efetuado o processo de agregação dos dados:

$$\Phi = \sum_{t=1}^T \beta_t \quad (\text{A.2})$$

Onde Φ é conjunto de dados unificado e ordenado crescentemente de acordo com o tempo, do instante onde os sensores são mais novos para o momento que os sensores são mais velhos. A variável T representa a quantidade total de lotes de dados e β_t é o lote de dados em questão que está sobre processo de concatenação.

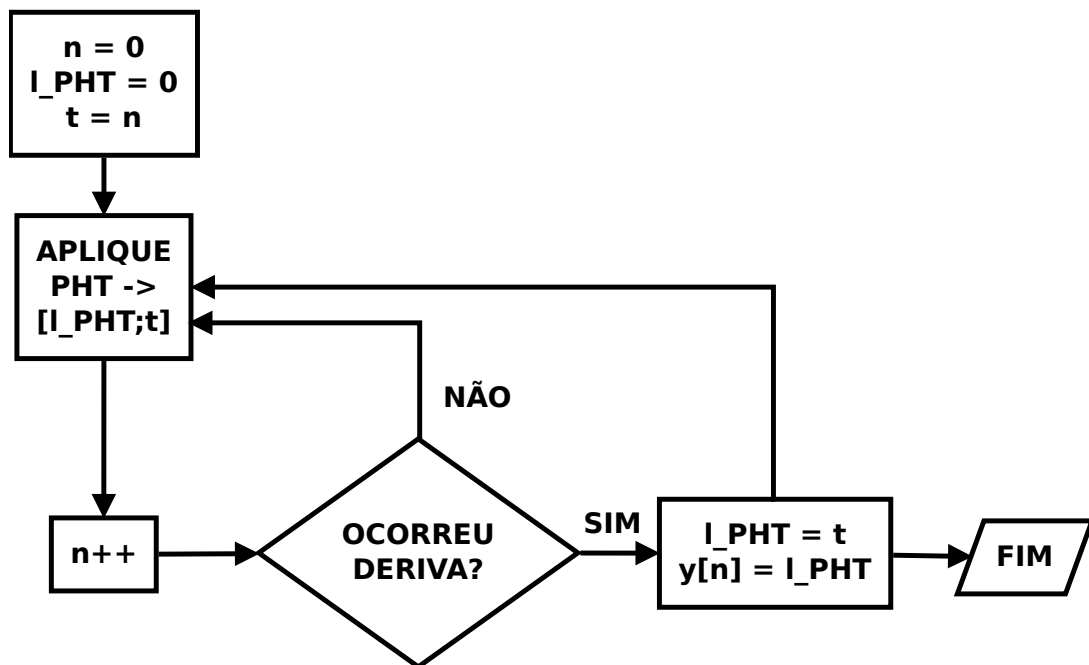


Figura 28 – Mecanismo de utilização do algoritmo Page-Hinkley e armazenamento da detecção de deriva

A seleção dos valores de magnitude e limiar é dada pela Tabela 7, onde através de tentativa e erro, com a finalidade de se aproximar a quantidade de divisões dos dados que são fornecidos pela divisão em lotes, os valores aproximados são $\lambda = 2$ e $\gamma = 0.98$. Estabelecendo-se uma divisão do conjunto de dados em 8 blocos.

Tabela 7 – Tabela de seleção dos magnitude e limiar do algoritmo PHT

| Parâmetros | | Quantidade de divisões |
|------------|----------|------------------------|
| λ | γ | |
| 0 | 0 | 0 |
| 0.1 | 0 | >10000 |
| 1 | 0 | >5000 |
| 10 | 0 | > 2000 |
| 0 | 10 | 0 |
| 0 | 5 | 0 |
| 0 | 3 | 0 |
| 0 | 2 | 0 |
| 0 | 1.5 | 1 |
| 0 | 1 | 2 |
| 0 | 0.5 | 3 |
| 1 | 0.5 | 6 |
| 2 | 0.5 | 4 |
| 2 | 1 | 6 |
| 2 | 0.9 | 7 |
| 2 | 1.1 | 15 |
| 2 | 1.05 | 13 |
| 2 | 0.98 | 8 |

A aplicação do algoritmo PHT, conforme a Figura 28, consiste basicamente em detectar-se à partir de uma magnitude γ e um limiar λ , acionar um gatilho de detecção de deriva no conjunto de dados. Conforme pode ser visto no diagrama abaixo, o estabelece-se que não há deriva, marcando-se $l_{PHT} = 0$, inicia-se a aplicação das amostras no algoritmo, onde à partir da primeira detecção, armazena-se o valor resultante em $y[n]$ marcando-se $l_{PHT} = n$ como o intervalo em que é registrada a última ocorrência. Prossegue-se o algoritmo à partir do intervalo l_{PHT} sem que haja início novamente da busca e haja sobreposição dos valores de detecção. Resumindo, o algoritmo faz a detecção de forma contínua e apresenta memória para a informação dos últimos dados que foram acusados à partir do limiar λ definido.

A.1 RESULTADOS PELA APLICAÇÃO DO ALGORITMO PHT

Na Figura 29 são mostrados os resultados obtidos à partir da aplicação do algoritmo PHT sobre a operação de concatenação de todos os lotes, de forma a somar

toda a informação em um único conjunto de dados e obter um indicador de deterioração dos sensores de gases ao longo do tempo:

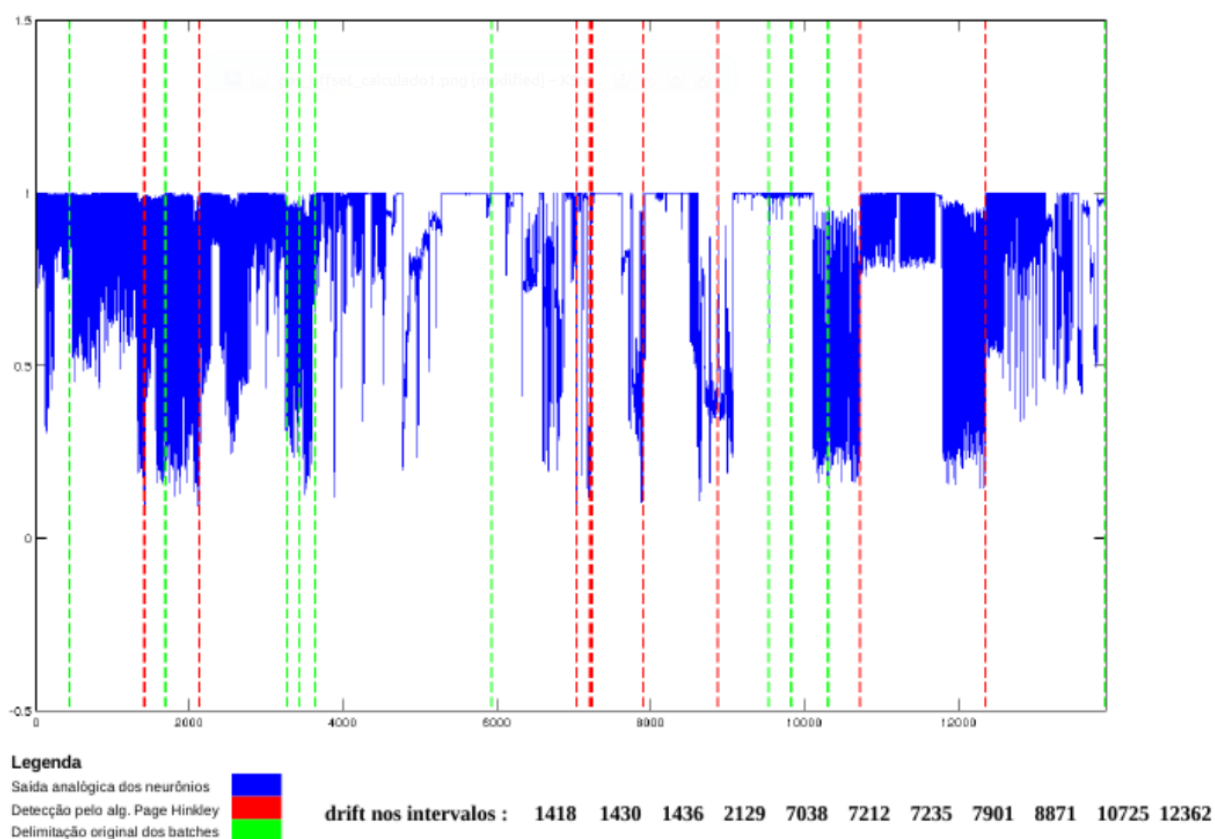


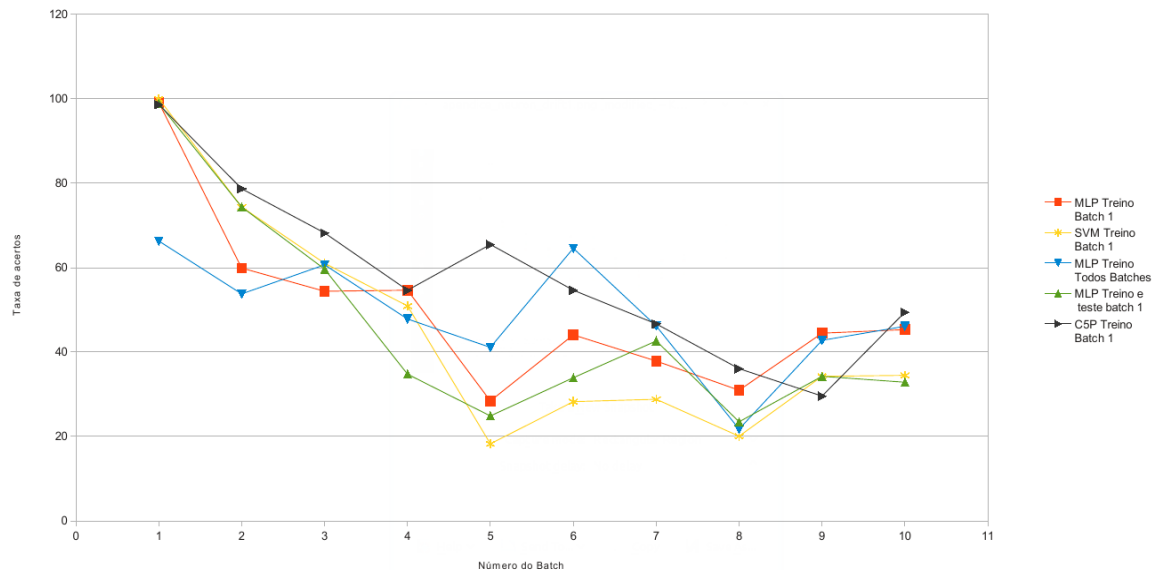
Figura 29 – Aplicação do algoritmo Page-Hinkley no dataset concatenado

Ainda na Figura 29, o eixo x representa o domínio temporal de evolução das amostras ao longo do tempo, o eixo y é uma das 128 características de entrada que compõem o conjunto de dados original e irá sofrer deriva ao longo do tempo. As linhas verticais representam em verde, a divisão temporal que cada um dos 10 lotes compõe. As linhas vermelhas são marcações aonde o algoritmo detectou a presença de deriva e assinalou uma mudança.

No uso desse algoritmo ao longo das 13910 amostras, foram adotados $\lambda = 2$ e $\gamma = 0.98$, sendo que estes foram seleccionados pela metodologia descrita na seção anterior com base na Tabela 7.

APÊNDICE B – TREINO COM APENAS O LOTE 1

Durante o início desse trabalho, foram estudados diversos classificadores com a finalidade de aplicar a classificação direta e observar qual algoritmo apresenta menor degradação.



- 1) Vermelho – Rede neural MLP treinada com o Batch 1 e testada com outros batches
- 2) Amarelo – SVM original do autor implementada e testada – Treinada e com o batch 1 e testada com outros batches
- 3) Azul – Rede neural MLP treinada e validada com todo o conjunto de dados e testada com cada um dos batches
- 4) Verde – Rede neural MLP treinada e validada com o batch 1 dividido em um lote de 2 e embaralhado. Testado com outros batches
- 5) Preto – Árvore C5P treinada com o Batch 1 e testada com outros batches

Figura 30 – Resultados à partir de testes com diversos classificadores usando somente o lote 1 como conjunto de dados para treinamento

A Figura 30 mostra o resultado direto na taxa de acerto de um classificador treinado apenas com o primeiro lote de dados e conseqüentemente aplicado no conseqüentes. Foram utilizadas RNAs MLP, SMVs e Árvores de decisão. Melhores resultados são alcançados com árvores de decisão, porém foi utilizada a RNA devida sua dinamicidade de treinamento e capacidade de projetar fronteiras de decisão complexas. Nesse último caso, deseja-se que o modelo possa ser usado em outros conjuntos de dados.

APÊNDICE C – ALGORITMO - COMPENSAÇÃO DE DERIVA PELA PONDERAÇÃO DAS SAÍDAS

O Algoritmo 2 abaixo, traz o pseudocódigo da implementação da metodologia de compensação de deriva em sensores de gases com a técnica de ponderação e aplicação de fatores de esquecimento diretamente na saída da RNA.

Algorithm 2 Treino e aplicação do algoritmo de compensação de deriva com pesos na saída

Entradas: Lista de lotes, matriz de pesos de cada função, número de RNA

Saídas: O rótulo do gás classificado com cada função

```

1: procedure DRIFTCOMPENSATION(loteLista, functionList, qNN)
2:   for  $t = 1$  to sizeof(loteLista) do                                     ▷ Cada um dos lotes
3:     for  $n = 1$  to qNN do
4:       train  $ANN_{t,n}$                                                  ▷ Treine qNN RNAs com lotet
5:     end for
6:   end for

7:   for  $t = 1$  to sizeof(loteLista) do                                     ▷ Teste em cada um dos lotes
8:     for  $c = 1$  to sizeof(functionList) do                               ▷ Para cada uma das funções da lista
9:       for  $n = 1$  to qNN do                                           ▷ Cada RNA anterior treinada com lotet
10:      for  $b = 1$  to  $t - 1$  do                                           ▷ Todos lotes mais velhos que lotet
11:        apply weight  $c_{t,b}$  to  $ANN_{b,n}$                                ▷ Aplique os pesos para classificar o
lote lotet
12:         $y_{t,n,b} = \text{teste } ANN_{b,n} \text{ com } \textit{lote}_t$ 
13:      end for
14:       $k_{t,n} = k_{t,n} + y_{t,n,b}$                                        ▷ Some as saídas de todas RNAs usadas para
classificar o lote lotet
15:    end for
16:     $output = \text{argmax}(k_{t,n})$                                        ▷ O rótulo do gás será o neurônio com maior
ativação
17:  end for
18: end for
19: end procedure

```

O funcionamento do algoritmo acima é: Dada a lista de funções a quais se deseja usar como fator de esquecimento, a quantidade de redes neurais a serem treinadas e o conjunto de dados, treina uma rede neural com o próprio conjunto de dados e armazena o resultado numa lista. Em posse da lista de RNAs treinadas, uma amostra no instante t irá utilizar para a classificação cada uma das RNAs treinadas com lotes anteriores, porém irá aplicar através do vetor de pesos, dado através das funções de ponderação para cada lote. O processo se repete para cada um dos lotes anteriores ao que deseja-se classificar. As saídas dos resultados de cada uma das

RNAs é somado. O rótulo do gás é dado pelo neurônio que possui máxima ativação para a amostra aplicada.