



Cintia de Faria Ferreira Carraro

**Sistema Híbrido de Extração de Regras Aplicado a
Diagnóstico de Transformadores**

Dissertação de Mestrado

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia Elétrica da PUC-Rio como requisito parcial para obtenção do título de Mestre em Engenharia Elétrica.

Orientadora: Profa. Marley Maria Bernades Rebuszi Vellasco
Co-orientador: Prof. Ricardo Tanscheit

Rio de Janeiro
Setembro de 2012



Cintia de Faria Ferreira Carraro

Sistema Híbrido de Extração de Regras Aplicado a Diagnóstico de Transformadores

Dissertação apresentada como requisito parcial para obtenção do grau de Mestre pelo Programa de Pós-Graduação em Engenharia Elétrica do Departamento de Engenharia Elétrica do Centro Técnico Científico da PUC-Rio. Aprovada pela Comissão Examinadora abaixo assinada.

Profa. Marley Maria Bernades Rebuszi Vellasco
Orientador
Departamento de Engenharia Elétrica – PUC-Rio

Prof. Ricardo Tanscheit
Co-Orientador
Departamento de Engenharia Elétrica – PUC-Rio

Prof. André Vargas Abs da Cruz
Departamento de Engenharia Elétrica – PUC-Rio

Prof. Jorge Luís Machado do Amaral
UERJ

Prof. Helvio Jailson Azevedo Martins
Centro de Pesquisa de Energia Elétrica

Prof. José Eugenio Leal
Coordenador Setorial do Centro
Técnico Científico

Rio de Janeiro, 10 de setembro de 2012

Todos os direitos reservados. É proibida a reprodução total ou parcial do trabalho sem autorização da universidade, da autora e da orientadora.

Cintia de Faria Ferreira Carraro

Graduada em Engenharia Elétrica com ênfase em Sistemas de Potência pela Universidade Severino Sombra, em 2007. Atuando desde então, em pesquisas voltadas à avaliação dos estados operativos dos equipamentos elétricos de alta tensão, contribuindo para o desenvolvimento de métodos e técnicas.

Ficha Catalográfica

Carraro, Cintia de Faria Ferreira

Sistema híbrido de extração de regras aplicado a diagnóstico de transformadores / Cintia de Faria Ferreira Carraro; orientadora: Marley Marla Bernardes Rebuzzi Vellasco ; co-orientador: Ricardo Tanscheit. – 2012.

95 f. ; 30 cm

Dissertação (mestrado) – Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Engenharia Elétrica, 2012.

Inclui bibliografia

1. Engenharia elétrica – Teses. 2. SVM. 3. Classificação. 4. Algoritmo genético. 5. Análise de resposta em frequência. 6. Impedância terminal. 7. Extração de regras linguísticas. 8. Regras linguísticas. I. Vellasco, Marley Maroa Bernardes Rebuzzi. II. Tanscheit, Ricardo. III. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Engenharia Elétrica. IV. Título.

CDD: 621.3

Para o meu filho João Pedro, meu maior
PRESENTE, pela ternura de cada abraço...

Agradecimentos

Primeiro a **Deus**, já que Ele colocou pessoas tão especiais a meu lado, sem as quais certamente não teria dado conta!

A meus pais, **Rogério e Fátima**, meus exemplos de vida, força, dedicação, amor e honestidade. Que sempre me acharam A MELHOR de todas, mesmo eu não sendo. Por eles, não quis ser A MELHOR, mas fazer o melhor de mim. Obrigada pelo amor incondicional.

Ao meu esposo, **Luiz Gustavo**, pela compreensão de tantas ausências e dos intermináveis dias dedicados a este trabalho. Muito obrigada pelo apoio inabalável e por tantas vezes me mostrar que era possível, sempre me fazendo acreditar mais que imagino. Te amo!

Ao pequeno **João Pedro**, que, no último ano, estive tão próximo (literalmente) de mim, que foi tão presente no desenvolvimento deste trabalho e que agora, me inspira a querer ser mais que fui até hoje!

Aos meus irmãos e melhores amigos: **Alexandre, Leonardo e Júlia**, pelo carinho e admiração. Por estarem presentes, mesmo quando ausentes.

A minha gigantesca família, em especial minha avó, **Mercês**, exemplo de vida e intensa dedicação à família, valores que não se aprende na escola.

Aos amigos, **Camila, Nivaldo e Roberta**, sempre presentes.

Aos amigos da PUC, **Felipe Baldner e Flávia Fontanet**, dos quais recebi muita ajuda no início do mestrado.

Aos orientadores **Marley e Ricardo Tanscheit**, por toda dedicação, sabedoria, motivação e amizade em todos os momentos. Agradeço pelos valiosos ensinamentos que sempre levarei comigo, pelo carinho e atenção desde o início do mestrado. *"Se eu vi mais longe, foi por estar de pé sobre ombros de gigantes."*

À **Adriana**, pelos esclarecimentos sobre Máquinas de Vetores Suporte e por ter colaborado no desenvolvimento de parte importante deste trabalho.

A **Helvio**, pelo grande apoio técnico e por sua amizade e compreensão sem os quais este trabalho não seria possível. Agradeço ainda pela revisão, sugestões e correção nos estágios finais desta dissertação.

Aos **amigos do Labdig**, pelas constantes discussões sobre Resposta em Frequência e força tarefa em construir um banco de dados. Agradeço ainda, ao **Rogério Azevedo** pelas simulações.

Aos **membros da banca examinadora**, pelas valiosas contribuições.

E, por fim, a todos aqueles que por um lapso não mencionei, mas que colaboraram para esta pesquisa.

Ninguém vence sozinho: OBRIGADA A TODOS!

Resumo

Carraro, Cintia de Faria Ferreira; Vellasco, Marley Maria Bernardes Rebuzzi (Orientadora); Tanscheit, Ricardo (Co-orientador). **Sistema Híbrido de Extração de Regras Aplicado a Diagnóstico de Transformadores**. Rio de Janeiro, 2012. 95p. Dissertação de Mestrado - Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro.

Este trabalho tem como objetivo construir um classificador baseado em regras de inferência fuzzy, as quais são extraídas a partir de máquinas de vetor suporte (SVMs) e ajustadas com o auxílio de um algoritmo genético. O classificador construído visa a diagnosticar transformadores de potência. As SVMs são sistemas de aprendizado baseados na teoria do aprendizado estatístico e apresentam boa habilidade de generalização em conjuntos de dados reais. SVMs, da mesma forma que redes neurais (RN), geram um "modelo caixa preta", isto é, um modelo que não explica o processo pelo qual sua saída é obtida. Entretanto, para alguns problemas, o conhecimento sobre como a classificação foi obtida é tão importante quanto a classificação propriamente dita. Alguns métodos propostos para reduzir ou eliminar essa limitação já foram desenvolvidos, embora sejam restritos à extração de regras simbólicas, isto é, contêm funções ou intervalos nos antecedentes das regras. No entanto, a interpretabilidade de regras simbólicas ainda é reduzida. De forma a aumentar a interpretabilidade das regras, o modelo FREx_SVM foi desenvolvido. Neste modelo as regras fuzzy são extraídas a partir de SVMs treinadas. O modelo FREx_SVM pode ser aplicado a problemas de classificação com n classes, não sendo restrito a classificações binárias. Entretanto, apesar do bom desempenho do modelo FREx_SVM na extração de regras linguísticas, o desempenho de classificação do sistema de inferência fuzzy obtido é ainda inferior ao da SVM, uma vez que as partições (conjuntos fuzzy) das variáveis de entrada são definidas a priori, permanecendo fixas durante o processo de aprendizado das regras. O objetivo desta dissertação é, portanto, estender o modelo FREx_SVM, de forma a permitir o ajuste automático das funções de pertinência das variáveis de entrada através de algoritmos genéticos. Para avaliar o desempenho do modelo estendido, foram realizados estudos de caso em dois bancos de dados: Iris, como uma base benchmark, e a análise de resposta em frequência. A análise de resposta em frequência é uma técnica não invasiva e não destrutiva, pois preserva as características dos equipamentos. No entanto, o diagnóstico é feito de modo visual comparativo e requer o auxílio de um especialista. Muitas vezes, este diagnóstico é subjetivo e inconclusivo. O ajuste automático das funções de pertinência correspondentes aos conjuntos fuzzy associados às variáveis de entrada reduziu o erro de classificação em até 13,38% em relação à configuração sem este ajuste. Em alguns casos, o desempenho da configuração com ajuste das funções de pertinência supera até mesmo aquele obtido pela própria SVM.

Palavras-chave

SVM; classificação; algoritmo genético; análise de resposta em frequência; impedância terminal; extração de regras linguísticas; regras linguísticas.

Abstract

Carraro, Cintia de Faria Ferreira; Vellasco, Marley Maria Bernardes Rebuzzi (Advisor); Tanscheit, Ricardo (Co-advisor). **Hybrid System for Rule Extraction Applied to Diagnosis of Power Transformers**. Rio de Janeiro, 2012. 95p. MSc Dissertation - Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro.

This work aims to develop a classifier model based on fuzzy inference rules, which are extracted from support vector machines (SVMs) and optimized by a genetic algorithm. The classifier built aims to diagnose power transformers. The SVMs are learning systems based on statistical learning theory and have provided good generalization performance in real data sets. SVMs, as artificial neural networks (NN), generate a black box model, that is, a model that does not explain the process by which its output is obtained. However, for some applications, the knowledge about how the classification was obtained is as important as the classification itself. Some proposed methods to reduce or eliminate this limitation have already been developed, although they are restricted to the extraction of symbolic rules, i.e. contain functions or ranges in the rules' antecedents. Nevertheless, the interpretability of symbolic rules is still reduced. In order to increase the interpretability of the rules, the FREx_SVM model was developed. In this model the fuzzy rules are extracted from trained SVMs. The FREx_SVM model can be applied to classification problems with n classes, not being restricted to binary classifications. However, despite the good performance of the FREx_SVM model in extracting linguistic rules, the classification performance of fuzzy classification system obtained is still lower than the SVM, since the partitions (fuzzy sets) of the input variables are pre-defined at the beginning of the process, and are fixed during the rule extraction process. The goal of this dissertation is, therefore, to extend the FREx_SVM model, so as to enable the automatic adjustment of the membership functions of the input variables through genetic algorithms. To assess the performance of the extended model, case studies were carried out in two databases: iris benchmark and frequency response analysis. The frequency response analysis is a non-invasive and non-destructive technique, because it preserves the characteristics of the equipment. However, the diagnosis is carried out by visual comparison and requires the assistance of an expert. Often, this diagnosis is subjective and inconclusive. The automatic adjustment of the membership functions associated with input variables reduced the error up to 13.38% when compared to the configuration without this optimization. In some cases, the classification performance with membership functions optimization exceeds even those obtained by SVM.

Keywords

SVM; classification; genetic algorithm; frequency response analysis; terminal impedance; extraction of linguistic rules; linguistic rules.

Sumário

1. Introdução	13
1.1. Motivação	13
1.2. Objetivos	14
1.3. Contribuições da Dissertação	15
1.4. Descrição da Dissertação	15
1.5. Estrutura da Dissertação	17
2. Máquinas de Vetores Suporte Empregadas à Classificação	18
2.1. Visão Geral	18
2.2. Máquina de Vetores Suporte	19
2.2.1. Introdução	19
2.2.2. Fundamentação Teórica	20
2.2.2.1. Minimização do Risco Empírico	20
2.2.2.2. Dimensão VC	22
2.2.2.3. Minimização do Risco Estrutural	22
2.2.3. Classificação Binária	23
2.2.3.1. SVMs com Margens Maximais	24
2.2.3.2. SVMs com Margens Suaves	26
2.2.3.3. SVMs Não Lineares	28
2.2.3.4. Mapeamento Utilizando as Funções de Kernel	29
2.2.4. Classificação em Múltiplas Classes	31
2.2.4.1. Decomposição um por Classe	32
2.2.4.2. Separação das Classes Duas a Duas	33
3. Modelo para Extração de Regras Fuzzy a partir de Máquinas de Vetores Suporte – FREx_SVM	34
3.1. Introdução	34
3.2. FREx_SVM – Caso Binário	34
3.2.1. Obtenção das Projeções dos Vetores Suporte	35
3.2.2. Definição dos Conjuntos Fuzzy	36
3.2.3. Extração de Regras Fuzzy	37
3.2.4. Acurácia Fuzzy	40
3.3. FREx_SVM – Caso Múltiplas Classes	42
4. FREx_SVM com Otimização de Parâmetros	44
4.1. Introdução	44
4.2. Estrutura e Metodologia	44
4.2.1. Módulo de Fuzzificação	45
4.2.2. Módulo Extração de Regras	45
4.2.3. Módulo de Otimização de Parâmetros	46
4.2.3.1. Intervalos de Desempenho	49
4.2.4. Classificação	50
4.3. Banco de Dados – IRIS	51
4.3.1. Resultados Obtidos	52
5. DIREF (Diagnóstico Inteligente por Resposta em Frequência)	58
5.1. Introdução	58
5.2. Análise de Resposta em Frequência	61

5.3. Medições e Simulações – DIREF	65
5.3.1. Medições Realizadas	65
5.3.1.1. Medição – Transferência de Tensão	65
5.3.1.2. Medição – Impedância Terminal	67
5.3.2. Simulações de Defeitos	68
5.3.2.1. Defeitos Estudados	72
5.4. Banco de Dados – IMPEDÂNCIA TERMINAL	77
5.4.1. Tratamento dos Dados	78
5.4.2. Resultados Obtidos	79
6. Conclusões e Trabalhos Futuros	89
6.1. Conclusões	89
6.2. Trabalhos Futuros	90
7. Referências bibliográficas	91

Lista de Figuras

Figura 1	Separações possíveis por uma função linear para 3 pontos.	22
Figura 2	Separação de 4 pontos, não podem ser separados por uma função linear.	22
Figura 3	Hiperplano de separação ótimo (vermelho).	24
Figura 4	Hiperplano de separação para um problema de classificação linear.	25
Figura 5	Variáveis Soltas.	27
Figura 6	Hiperplano de separação ótimo generalizado	27
Figura 7	Elevação do espaço de características	28
Figura 8	Mapeamento de funções não lineares	29
Figura 9	Modelo de extração de conhecimento explícito de SVM	35
Figura 10	Projeção do vetor suporte s	36
Figura 11	Definição dos conjuntos fuzzy	36
Figura 12	Conjuntos fuzzy com maior grau de pertinência	37
Figura 13	Conjuntos fuzzy para os atributos peso e altura	39
Figura 14	Modelo FREx_SVM para múltiplas classes	43
Figura 15	Módulos do FREx_SVM com otimização de conjuntos	44
Figura 16	Configuração Inicial para variável com 3 conjuntos fuzzy	46
Figura 17	Representação com variável fuzzy de 3 conjuntos	47
Figura 18	Parâmetros de um conjunto fuzzy triangular	49
Figura 19	Intervalos de desempenho de um conjunto fuzzy triangular	50
Figura 20	Configuração Inicial dos conjuntos fuzzy para as 4 variáveis de entrada - ÍRIS	57
Figura 21	Configuração das variáveis de entrada após a otimização – ÍRIS	58
Figura 22	Sistema Elétrico de Potência Típico	60
Figura 23	Vista em corte de um transformador típico	61
Figura 24	(a) circuito elétrico de um transformador ideal (b) vista em corte da parte ativa do transformador mostrando as capacitâncias parasitas	62
Figura 25	Arranjo típico para medição de admitância própria e de transferência (IEEE, 2005).	63
Figura 26	Medição no Domínio da Frequência	64
Figura 27	Procedimento para realização da Análise de Resposta em Frequência	64
Figura 28	Diagrama simplificado de medição de resposta em frequência	66
Figura 29	Resposta típica para um transformador trifásico, tensões aplicadas nos enrolamentos H e medidas nos enrolamentos X correspondentes.	67
Figura 30	Transformador trifásico, medições terminais H e X	68
Figura 31	Autotransformador – módulo da impedância terminal	68
Figura 32	Correspondência entre arranjos RLC e curva de impedância	69
Figura 33	Componentes reativas na medição	70
Figura 34	Modelo representando o enrolamento no domínio da frequência	70

Figura 35	Correspondência entre o circuito simulado e as ressonâncias da curva de impedância	71
Figura 36	Simulação de impedância terminal padrão e sem óleo	73
Figura 37	Modificação do canal de óleo após o 'espiralamento' (tilting)	74
Figura 38	Sobreaquecimento local	74
Figura 39	Curto-circuito entre espiras	75
Figura 40	Espiramento (a) causado em teste de curto-circuito e detalhes de saída (b) nitidamente deformadas	75
Figura 41	Medição de impedância terminal antes e após o curto-circuito	76
Figura 42	Enrolamento resultante de deslocamento axial	77
Figura 43	Medição de impedância terminal antes e após o deslocamento axial	77
Figura 44	Configuração Inicial para os conjuntos fuzzy das 4 variáveis de entrada - IMPEDÂNCIA	88
Figura 45	Configuração após otimização de cada variável de entrada - IMPEDÂNCIA	88

Lista de Tabelas

Tabela 1	Exemplos de kernel	31
Tabela 2	Exemplo ilustrativo do funcionamento do FReX_SVM	38
Tabela 3	Graus de pertinência para o atributo peso	39
Tabela 4	Graus de pertinência para o atributo altura	39
Tabela 5	Método decomposição um por classe – Íris (3 conjuntos) – ERROS	52
Tabela 6	Método decomposição um por classe – Íris (3 conjuntos) – NÃO CLASSIFICADOS	53
Tabela 7	Método decomposição um por classe – Íris (5 conjuntos) – ERROS	53
Tabela 8	Método decomposição um por classe – Íris (5 conjuntos) – NÃO CLASSIFICADOS	54
Tabela 9	Método separação das classes duas a duas – Íris (3 conjuntos) – ERROS	55
Tabela 10	Método separação das classes duas a duas – Íris 3 conjuntos – NÃO CLASSIFICADOS	55
Tabela 11	Método separação das classes duas a duas – Íris (5 conjuntos) – ERROS	56
Tabela 12	Método separação das classes duas a duas – Íris (5 conjuntos) – NÃO CLASSIFICADOS	56
Tabela 13	Método decomposição um por classe – Impedância - ERROS	79
Tabela 14	Método decomposição um por classe – Impedância – NÃO CLASSIFICADOS	80
Tabela 15	Método decomposição um por classe – Impedância – REGRAS	80
Tabela 16	Método decomposição um por classe – Impedância – TAXAS DE ACERTO	81
Tabela 17	Método separação das classes duas a duas – Impedância – ERROS	81
Tabela 18	Método separação das classes duas a duas – Impedância – NÃO CLASSIFICADOS	82
Tabela 19	Método separação das classes duas a duas – Impedância – REGRAS	82
Tabela 20	Método separação das classes duas a duas – Impedância – TAXAS DE ACERTO	83

1

Introdução

1.1

Motivação

Em virtude do custo devido à falha de equipamentos de alta tensão, em especial transformadores de potência, torna-se bastante relevante avaliar o estado operativo desses equipamentos. Porém, diagnosticar esses equipamentos gera custos e interrupção nos serviços em subestações, uma vez que, para o diagnóstico ser confiável, deve-se parar o equipamento para inspeção ou recorrer ao diagnóstico de um especialista. Este, no entanto, é realizado de forma subjetiva e muitas vezes é inconclusivo. Assim, é fundamental dispor-se de técnicas para evitar interrupções equivocadas de serviços em subestações para retirada de equipamentos em serviço.

Uma das técnicas consagradas de diagnóstico de transformadores de potência é analisar sua resposta em frequência [1], [2], [3], [4], [5]. Transformadores de potência podem falhar por diversas razões, tais como ocorrências associadas ao transporte do equipamento, conseqüências mecânicas associadas a forças eletromagnéticas estabelecidas por correntes de curto-circuito, envelhecimento e superação de equipamentos em operação. Estas falhas são resultado de diferentes causas e condições e afetam a confiabilidade e a expectativa de vida dos equipamentos.

A análise de resposta em frequência (*Frequency Response Analysis – FRA*) é um método de diagnóstico não destrutivo e não invasivo utilizado para identificar potenciais problemas na parte ativa de um transformador (enrolamento e núcleo) através da comparação com dados de referência da mesma unidade ou resultados similares.

FRA consiste na medição da função de transferência constituída de elementos indutivos e capacitivos dos enrolamentos do transformador [6]. Esta medição é feita sobre uma larga faixa de frequência e os resultados são comparados com uma referência conhecida como “assinatura” ou “impressão digital” do enrolamento do transformador [4]. A “impressão digital” é uma medida de referência geralmente realizada em fábrica.

O diagnóstico é realizado atualmente de modo visual e comparativo por especialistas. É geralmente demorado e bastante custoso, uma vez que o equipamento deve ser retirado de operação, além de exigir a dedicação de um técnico especialista por um tempo não desprezível. Assim, é evidente a necessidade de se desenvolver um sistema automático capaz de realizar o diagnóstico sem a necessidade de um especialista e sem a retirada do equipamento de serviço por longo período de tempo.

Neste trabalho, desenvolve-se um sistema automático de diagnóstico que toma por base um classificador denominado FREx_SVM [6]. Neste classificador, que extrai regras fuzzy de SVMs treinadas, as funções de pertinência associadas aos conjuntos fuzzy das variáveis de entrada são definidas a priori e permanecem fixas ao longo de todo o processo de construção das regras. Embora, neste modelo, o objetivo primeiro seja obter uma explicação (por meio da extração de regras), seu desempenho em termos de classificação é em geral bastante inferior ao daquele obtido pela utilização pura e simples de SVMs (sem explicação linguística).

Com a finalidade de melhorar o desempenho do modelo FREx_SVM, desenvolve e implementa-se, neste trabalho, o ajuste dos suportes das funções de pertinência por meio de algoritmos genéticos. Isto possibilita uma melhor representação dos dados de entrada e gera regras mais precisas.

1.2

Objetivos

- Estender o modelo FREx_SVM incorporando o ajuste automático das funções de pertinência, de forma a melhorar o desempenho do sistema fuzzy de classificação, sem prejudicar a interpretabilidade;
- Aplicar o sistema estendido ao diagnóstico de transformadores de potência, de forma a se prescindir de especialistas e evitar paradas desnecessárias de equipamentos.

1.3

Contribuições da Dissertação

A principal contribuição deste trabalho é:

Extensão do modelo FREx_SVM com a inserção de um módulo de ajuste das funções de pertinência associadas às variáveis de entrada. O modelo proposto utiliza os vetores suporte originados a partir da SVM treinada para definir os conjuntos fuzzy aos quais cada atributo de entrada pertence. Cada vetor suporte gera uma regra fuzzy [6]. O ajuste das funções de pertinência é efetuado por meio de algoritmos genéticos e tem como objetivo melhorar o desempenho do FREx_SVM original, no qual as funções de pertinência são fixas e definidas a priori.

Outras contribuições dessa dissertação são:

- Utilização da métrica apresentada em [7] para avaliar as regras fuzzy: acurácia fuzzy -- que mede o quanto uma regra descreve os dados.
- Análise Comparativa do desempenho dos métodos de classificação para múltiplas classes – testes com diferentes bancos de dados: Iris e Medições de FRA (*Frequency Response Analysis*). Análise do desempenho dos métodos de classificação: separação das classes duas a duas [8], [9], [10], [11], [12] e decomposição um por classe [8], [9], [13].

1.4

Descrição da Dissertação

As etapas para o desenvolvimento do modelo FREx_SVM Estendido são:

1. Estudo de SVMs binárias
2. Estudo dos métodos de decomposição um por classe (*one against all*) e separação das classes duas a duas, para classificação em múltiplas classes
3. Estudo dos Sistemas de Inferência Fuzzy
4. Estudo do modelo FREx_SVM original

5. Estudo dos algoritmos genéticos para ajustar graus de pertinência dos conjuntos fuzzy
6. Desenvolvimento de ajuste de conjuntos fuzzy com o objetivo de melhorar o desempenho do modelo FREx_SVM, reduzindo o erro de classificação
7. Estudo de casos

Neste trabalho, toma-se por base SVMs [8], [14], [15], [16], [17], [18], [19] para a criação do modelo de classificação. Como as SVMs foram originalmente desenvolvidas para classificação binária, é realizado primeiramente um estudo sobre SVMs binárias, incluindo a metodologia de aprendizado, os diferentes *kernels* utilizados e os conceitos sobre teoria da generalização.

Como o foco desse trabalho é a classificação em múltiplas classes, apresenta-se um estudo bibliográfico dos métodos de classificação de decomposição um por classe (um-contra-todos) e de separação das classes duas a duas, por exigirem menor esforço computacional e apresentarem melhor desempenho geral em [6].

Em seguida, é realizado um estudo detalhado do modelo FREx_SVM [6] original, para entendimento de seus algoritmos e respectivos parâmetros de aprendizado.

Finalmente, realiza-se um estudo sobre a utilização de algoritmos genéticos para o ajuste de funções de pertinência, envolvendo também a análise da ferramenta GACom (desenvolvida no ICA - PUC - Rio) [20]. O ajuste baseado em intervalo de desempenho [21] é implementado no FREx_SVM, com o intuito de melhorar o seu desempenho quanto à taxa de erros de classificação e pontos não classificados.

Para que este trabalho fosse tecnicamente viável, foi utilizado um banco de dados misto (dados reais e simulados) com dados de análise de resposta em frequência em transformadores, técnica que foi estudada e será detalhada ao longo da dissertação.

1.5

Estrutura da Dissertação

Este trabalho é constituído dos seguintes capítulos:

No capítulo 2 são descritos os conceitos básicos das máquinas de vetores suporte aplicadas à classificação binária e os métodos decomposição um por classe e separação das classes duas a duas, de classificação em múltiplas classes.

O capítulo 3 apresenta o modelo FREx_SVM, que é a base e motivação deste trabalho. Além do modelo FREx_SVM, este capítulo descreve a métrica utilizada para avaliação das regras fuzzy construídas a partir do modelo FREx_SVM, denominada acurácia fuzzy.

No capítulo 4 apresenta-se o modelo estendido do FREx_SVM, com a inserção do ajuste automático das funções de pertinência, e os resultados obtidos com sua aplicação ao banco de dados Iris.

No capítulo 5 é realizada uma revisão bibliográfica sobre o método de Análise de Resposta em Frequência e são discutidos os resultados obtidos com a aplicação do sistema proposto ao banco de dados de resposta em frequência.

As conclusões e trabalhos futuros sugeridos constituem o capítulo 6.

2

Máquinas de Vetores Suporte Empregadas à Classificação

2.1

Visão Geral

A Inteligência Computacional destaca-se por ser uma área da ciência que busca, através de técnicas inspiradas na natureza, o desenvolvimento de sistemas que imitam aspectos do comportamento humano, tais como aprendizado, percepção e adaptação [22].

Nesta seção será descrita a técnica de inteligência computacional em que se baseia o estudo desta dissertação: Máquina de Vetores Suporte (SVM). O desenvolvimento desta dissertação envolveu também o uso de Lógica Fuzzy e Algoritmos Genéticos (GAs – *Genetic Algorithms*). Informações detalhadas sobre a técnica de Lógica Fuzzy podem ser encontradas em [23], [24], [25], [26], [27]; já sobre algoritmos genéticos, as seguintes referências fornecem maiores detalhes [28], [29], [30], [31].

A técnica de Máquinas de Vetores Suporte (SVMs) é um tipo de algoritmo de aprendizado supervisionado que infere, a partir de um conjunto de dados com classificação conhecida, uma função que possa classificar com bom desempenho amostras não observadas durante o treinamento (um classificador que generaliza bem). De forma sucinta, cria-se um hiperplano que separa os dados de forma a colocar o maior número de dados de uma dada classe do mesmo lado, maximizando a distância de cada classe ao hiperplano que as separa. O hiperplano gerado pela SVM é determinado por um subconjunto dos pontos das duas classes, chamados de vetores suporte [6].

A maneira como o hiperplano de separação é criado é descrita em detalhe nas seções a seguir.

2.2

Máquina de Vetores Suporte

2.2.1

Introdução

As Máquinas de Vetores Suporte (SVMs, do inglês *Support Vector Machines*), propostas em 1979 a partir da generalização do algoritmo *Generalized Portrait* desenvolvido por Vapnik [32], são consideradas uma das maiores inovações na área de reconhecimento de padrões nos últimos anos [19].

O princípio básico desenvolvido por Vapnik tem embasamento na Minimização do Risco Estrutural (SRM – *Structural Risk Minimization*). O SRM possui um desempenho de generalização superior ao tradicional Risco Empírico (ERM – *empirical risk minimization*) utilizado em redes neurais convencionais. SRM baseia-se no fato de que a taxa de erro de generalização é determinada pela soma da taxa de erro de treinamento e um termo que depende da dimensão de Vapnik-Chervonenkis (chamada dimensão VC) [15], [8], [33], [34]. A dimensão VC é uma medida da capacidade ou poder de expressão de um conjunto de funções. Em se tratando de padrões separáveis, a máquina tem valor zero para erro de treinamento e minimiza a dimensão VC. Desta forma, as SVMs possuem boa generalização em problemas de classificação [6].

SVMs têm sido utilizadas em grande escala para solucionar problemas de classificação, apresentando bons resultados quando empregadas aos mais variados problemas de reconhecimento de padrões [35], tais como categorização de textos [36], categorização de SPAM [37] e reconhecimento de caracteres manuscritos [16], [17].

Nas SVMs, funções não lineares mapeiam os vetores de entrada em um espaço de características de dimensão maior, onde é construído um hiperplano de separação ótimo. Este hiperplano separa os dados de forma a colocar o maior número de pontos da mesma classe do mesmo lado, enquanto a distância de cada classe a esse hiperplano é maximizada com o objetivo de melhorar a generalização. A distância de uma classe a um hiperplano é a menor distância entre ele e os pontos dessa classe e é chamada de margem de separação [6].

Os vetores que delimitam a margem de separação são denominados os vetores suporte.

2.2.2

Fundamentação Teórica

Com a finalidade de proporcionar um embasamento teórico para o problema de aprendizado em SVM, são introduzidas a seguir algumas definições e suposições.

2.2.2.1

Minimização do Risco Empírico

Para problemas de classificação envolvendo duas classes, denominados problemas de classificação binária, a tarefa de aprendizagem a partir de amostras de treinamento pode ser formulada como segue.

Dada uma classe de funções de decisão:

$$\{f_\lambda : \lambda \in \Lambda\} \quad (1)$$

com

$$f_\lambda : \mathfrak{R}^n \rightarrow \{-1, +1\}$$

onde Λ é um conjunto de parâmetros que faz com que corresponda ao espaço de hipóteses H_Z em um conjunto de exemplos:

$(x_1, y_1), \dots, (x_n, y_n)$, com

$$x_i \in \mathfrak{R}^n, y_i \in \{-1, +1\}, i = 1, \dots, N,$$

obtidos de uma densidade de probabilidade $p(x, y)$, deseja-se encontrar uma função f_λ que forneça o menor valor possível para o funcional de risco:

$$R(\lambda) = \int |f_\lambda(x) - y| p(x, y) dx dy \quad (2)$$

O risco esperado é, portanto, uma medida de quão bem uma hipótese candidata prediz o valor correto de y para uma amostra x .

Uma vez que $p(x, y)$ é desconhecida, não há como calcular o funcional de risco $R(\lambda)$, de modo que se deve recorrer a uma estimativa obtida de dados de treinamento, ou seja, o funcional de risco empírico, que é definido pelo número médio dos erros de f :

$$R_{emp}(\lambda) = \frac{1}{N} \sum_{i=1}^N |f_{\lambda}(x_i) - y_i| \quad (3)$$

Princípio da minimização do Risco Empírico - utilizar os exemplos de treinamento para aproximar a integral (funcional de risco) pela soma finita (risco empírico) e encontrar uma função que minimize esta soma.

Para classificações binárias, o risco é igual à probabilidade de erro de classificação e o risco esperado é o erro de treinamento. Isto quer dizer que, mesmo que uma função classifique corretamente os dados de treinamento, ela pode não generalizar bem, ou seja, classificar corretamente dados não apresentados durante o treinamento.

Desta forma, minimizar somente os erros de treinamento (risco empírico) não garante um erro de teste pequeno. A teoria do aprendizado estatístico indica que é preciso restringir o conjunto de funções de estimativa a um que tenha capacidade compatível com os dados de treinamento disponíveis [38]. Daí surge a necessidade de se utilizar a minimização do risco estrutural, que se baseia no princípio de que a taxa de erro de uma máquina de aprendizado nos dados de teste (taxa de erro de generalização) está limitada pela soma da taxa dos erros de treinamento (risco empírico) e um termo que depende da dimensão VC, definida na próxima seção.

2.2.2.2

Dimensão VC

A estratégia das SVMs é solidamente fundamentada na teoria de aprendizado estatístico, ou teoria VC (Vapnik-Chervonenkis) [8], [14], [15], [16], [17], [18] visando à proposição de técnicas de aprendizado de máquina que buscam maximizar a capacidade de generalização.

A dimensão VC é um índice escalar que mede a complexidade intrínseca de uma classe de funções. Na Figura 1, é apresentada uma forma da dimensão VC para funções lineares no \mathfrak{R}^2 . Conclui-se que a dimensão VC no \mathfrak{R}^2 é 3, pois este é o número máximo de amostras que podem ser classificadas por uma reta e os pontos não podem ser colineares. Para quatro ou mais amostras, observe a Figura 2, existem padrões de rotulação que não possibilitam uma classificação correta por intermédio de uma reta.

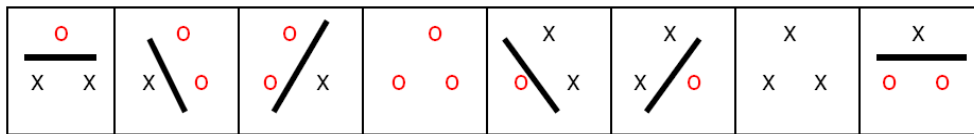


Figura 1 – Separações possíveis por uma função linear para 3 pontos.

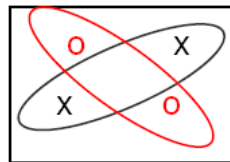


Figura 2 – Separação de 4 pontos, não podem ser separados por uma função linear

Em outras palavras, a dimensão VC é o maior N , onde N é o número máximo de exemplos de treinamento que podem ser separados em duas classes, nas 2^N maneiras de combiná-los.

De forma genérica, para funções lineares no \mathfrak{R}^n , com $n \geq 2$, a dimensão VC será $n+1$.

2.2.2.3

Minimização do Risco Estrutural

A minimização do risco estrutural (SRM - *Structural Risk Minimization*) é uma tentativa de tratamento do problema de escolha de uma dimensão VC apropriada. Na seção 2.2.2.1 observou-se que um valor pequeno para o risco empírico não implica necessariamente um valor pequeno para o risco esperado. O princípio do SRM está baseado na observação de que tanto a dimensão VC

quanto o risco empírico deveriam ser minimizados simultaneamente. Para implementar o princípio SRM é necessário propor uma estrutura aninhada para o espaço de hipóteses, tal que:

$$H_1 \subset H_2 \subset \dots \subset H_k \subset \dots,$$

com a propriedade de que $h(k) \leq h(k+1)$, onde $h(k)$ é a dimensão VC de H_k , resultando no seguinte problema de otimização:

$$\min_{H_k} \left(R_{emp}(\lambda) + \sqrt{\frac{h(k)}{N}} \right) \quad (4)$$

O princípio de SRM é bem fundamentado matematicamente, mas pode ser de difícil interpretação pelas seguintes razões:

- pode ser difícil calcular a dimensão VC de H_k , além de que existe apenas um pequeno número de classes de funções para as quais se sabe calcular a dimensão VC; [8], [15], [33], [34];
- mesmo admitindo a obtenção VC de H_k , o problema da minimização da expressão (4) pode ser de difícil solução [38].

Embora não seja trivial controlar a dimensão VC da técnica de aprendizado durante a fase de treinamento, isto é adequadamente realizado pela abordagem SVM, que busca minimizar simultaneamente a dimensão VC e o risco empírico.

2.2.3

Classificação Binária

Uma SVM é um algoritmo de aprendizado que infere de um conjunto de exemplos rotulados (cuja classe é conhecida) uma função capaz de prever os rótulos de novos exemplos desconhecidos.

A máquina de vetores suporte constrói um conjunto de hiperplanos cujos limites da dimensão VC possam ser computados e usa, então, o princípio de SRM para identificar o hiperplano de separação ótimo que maximize a margem dos elementos mais próximos [15], [16]. Isto equivale a minimizar o limite da dimensão VC.

Considere o exemplo da Figura 3. Observe que existem, para este caso, vários classificadores lineares possíveis para realizar a separação das amostras sem nenhum erro, porém somente um maximiza a margem (maximiza a distância entre o classificador linear e a amostra mais próxima de cada classe). Este classificador é chamado de *hiperplano de separação ótimo* ou *hiperplano de margem maximal*, devido ao fato de que, ao menos intuitivamente, espera-se que este hiperplano generalize melhor do que os demais quando classificar amostras não apresentadas durante a fase de treinamento.

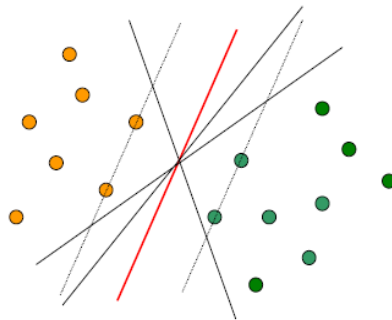


Figura 3 – Hiperplano de separação ótimo (vermelho).

Analisando a Figura 3, observa-se que o hiperplano de separação ótimo foi construído tendo como suporte dois outros hiperplanos (linhas tracejadas), que passam sobre alguns pontos para ambas as classes. Estes pontos são denominados *vetores-suporte*.

2.2.3.1

SVMs com Margens Maximais

No caso de padrões linearmente separáveis, solucionar o problema de treinamento de SVMs consiste na busca do hiperplano que separe perfeitamente os pontos de cada classe, maximizando a margem de separação.

Encontrar o hiperplano de separação ótimo de um conjunto de dados torna-se um problema de *otimização*: maximizar ou minimizar uma função objetivo sob um determinado conjunto de restrições.

Um hiperplano da forma

$$(w \cdot x) + b = 0 \quad w \in \mathcal{R}^N, b \in \mathcal{R} \quad (5)$$

representa a fronteira discriminante, de modo que o lado caracterizado por $(w \cdot x) + b > 0$ representa uma classe, e o lado caracterizado por $(w \cdot x) + b < 0$ representa a outra classe.

Este hiperplano corresponde à função de decisão:

$$f(x) = \text{sign}((w \cdot x) + b) \quad (6)$$

Onde sign é a função sinal, esta retorna o sinal do número real, tal que: $\text{sign}(y) \{ +1 \text{ para } y \geq 0 \text{ e } \text{sign}(y) \{ -1 \text{ para } y < 0$.

A Figura 4 apresenta um modelo de classificação linear. O hiperplano ótimo (linha contínua em preto) é ortogonal ao menor segmento que conecta os dois polígonos convexos de cada classe (linha tracejada em verde), interceptando-o na metade do segmento.

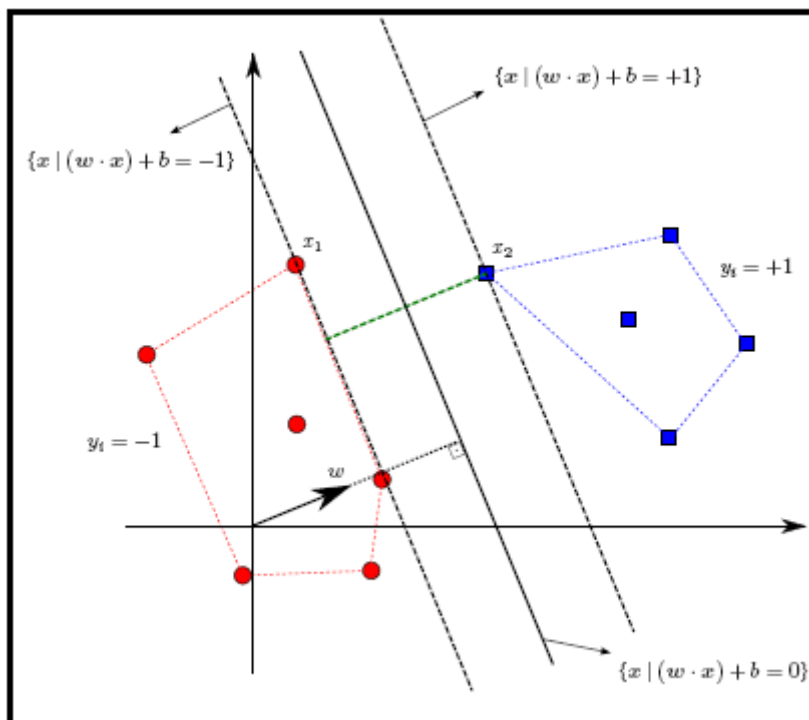


Figura 4 – Hiperplano de separação para um problema de classificação linear.

Onde w é o vetor de pesos e o escalar b é o bias. Tal que $y_i((w \cdot x_i) + b) > 0$.

Pode-se reescalonar w e b de forma que os pontos mais próximos do hiperplano (x_1 e x_2) satisfaçam a $|(w \cdot x_i) + b| = 1$, obtendo assim uma forma (w, b) do hiperplano com $y_i((w \cdot x_i) + b) \geq 1$. A margem medida perpendicularmente ao hiperplano é igual a $\frac{2}{\|w\|}$. Assim, para maximizar a margem deve-se minimizar $\|w\|$ sujeito à seguinte restrição:

$$y_i((w \cdot x_i) + b) \geq 1 \quad (7)$$

A minimização de $\|w\|$ é um problema de otimização quadrática, cuja solução possui uma expansão $w = \sum v_i x_i$ em termos de um subconjunto de exemplos de treinamento que estão sobre a margem. Na Figura 4 estes exemplos de treinamento, que estão sobre a margem, são representados por x_1 e x_2 . Tais exemplos de treinamento, chamados de vetores suporte (SV), carregam toda informação relevante sobre o problema de classificação.

Uma vez que, $w = \sum v_i x_i$, deve-se enfatizar que a função de decisão final torna-se dependente unicamente do produto interno entre os padrões, conforme apresenta a equação (8).

$$f(x) = \text{sign} \left(\sum_i v_i \cdot (x \cdot x_i) + b \right) \quad (8)$$

2.2.3.2

SVMs com Margens Suaves

Nem todos os conjuntos de dados constituem um problema linearmente separável. Para dados que não são linearmente separáveis, é impossível construir um hiperplano de separação que os classifique todos corretamente. Para melhor desempenho os dados devem ser “quase” linearmente separáveis.

Para tanto existe uma alternativa à seção anterior que admite erros de classificação através da introdução das variáveis soltas. Estas variáveis permitem a violação da equação (7). Assim o vetor x_i é classificado

corretamente como da classe y_i , $y_i \in \{-1,1\}$, quando a seguinte expressão é verdadeira:

$$y_i((w \cdot x_i) + b) + \xi_i \geq 1 \quad \text{para } i=1, \dots, N, \quad (9)$$

Onde w é o vetor de pesos \mathfrak{R}^n , o escalar b é o bias e ξ_i são as variáveis soltas não negativas associadas a cada vetor de treinamento x_i .

A Figura 5 mostra a variável solta para dois pontos classificados incorretamente pelo hiperplano de separação. A Figura 6 mostra o hiperplano de separação ótimo generalizado, admitindo erros de classificação.

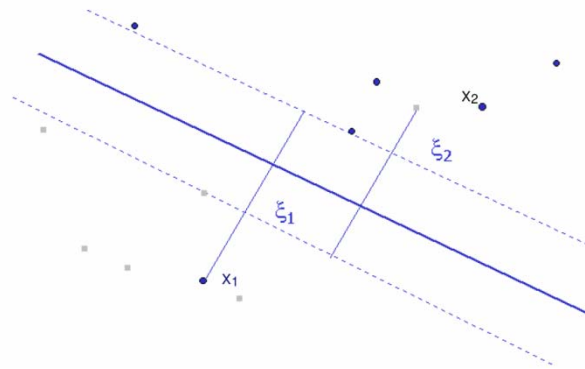


Figura 5 – Variáveis Soltas.

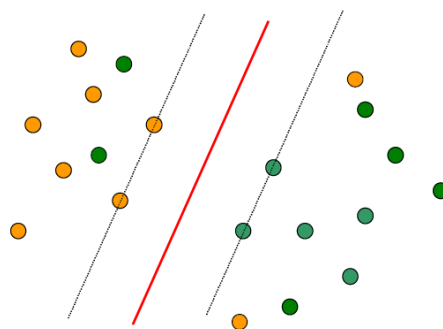


Figura 6 – Hiperplano de separação ótimo generalizado

Deste conceito de variáveis soltas surge a constante de regularização C : parâmetro de treinamento que estabelece o equilíbrio entre a complexidade do modelo e o erro de treinamento. Este parâmetro controla o peso do número de

erros, que é limitado pelo somatório das variáveis soltas, e do tamanho da margem, que é inversamente proporcional à norma de w [8], [14], [33]. Quanto maior o valor de C , maior é o peso atribuído ao número de erros (permite poucos erros) e menor peso é atribuído à margem do hiperplano (gera uma margem pequena) [6].

$$\Phi(w^i, \xi^i) = \frac{1}{2}(w^i \cdot w^i) + C \sum_{i=1}^N (\xi^i)_i \quad (10)$$

2.2.3.3

SVMs Não Lineares

Mesmo com as “variáveis de folga” (ou soltas) existem problemas que não são possíveis de separar linearmente ou ‘quase’ linearmente. A solução é utilizar um mapeamento não linear. Esse mapeamento é feito por meio de máquinas não lineares que projetam os dados em um espaço de características maior, ou seja, projetam-se os dados que estão em um espaço não linearmente separável em um no qual seja possível separá-los linearmente. Este princípio baseia-se no teorema de Cover [34], o qual afirma que os dados são mais propensos a serem separados linearmente em altas dimensões, desde que duas condições sejam satisfeitas: a transformação seja não linear e a dimensão do espaço de características seja suficientemente grande. Assim, é possível construir um hiperplano de separação ótimo nesse espaço de características. Na Figura 7 é mostrada a elevação do espaço de características do \mathbb{R}^2 para o \mathbb{R}^3 , possibilitando que os dados sejam separados por um hiperplano ótimo.

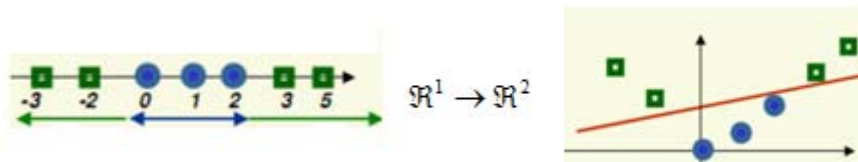


Figura 7 – Elevação do espaço de características

Da mesma forma que no caso de SVMs lineares, surgirão os vetores suporte.

Para realizar essa elevação dos espaços são utilizados os kernels, conforme descrito na próxima seção.

2.2.3.4

Mapeamento Utilizando as Funções de Kernel

Conforme descrito nas seções anteriores, a idéia básica da SVM é mapear não linearmente os vetores (exemplos de treinamento) do espaço de entrada em um espaço de características F de maior dimensão. A Figura 8 apresenta o mapeamento não linear Φ dos vetores de entrada para o espaço de características F .

$$\Phi = \mathcal{R}^N \rightarrow F \quad (11)$$

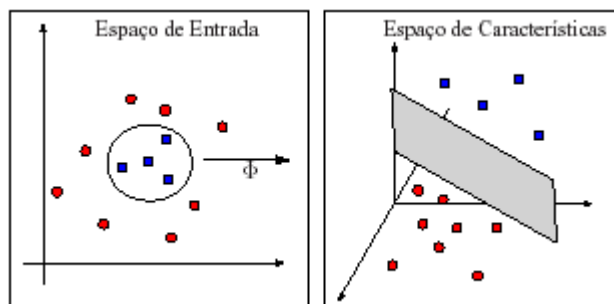


Figura 8 – Mapeamento de funções não lineares

A partir da equação da função de decisão (equação (9)), em que os padrões não são linearmente separáveis. Os termos x e x_i , são mapeados para o espaço de características F antes da realização do produto entre eles, como mostra a equação (11) a seguir.

$$k(x, x_i) = (\Phi(x) \cdot \Phi(x_i)) \quad (12)$$

Uma vez que os vetores são mapeados para espaços de maiores dimensões, o lado direito da equação (11) acima se torna computacionalmente complexo. Para contornar esta complexidade computacional, existem *kernels*

que tornam tal cálculo eficiente. A utilização das funções de kernel permite o cálculo do hiperplano sem necessariamente mapeá-lo no espaço de características [14].

Com a finalidade de exemplificar a complexidade deste mapeamento e o produto interno de vetores na equação (11), a seguir são utilizados dois vetores de duas dimensões $x = (x_1, x_2)$, $y = (y_1, y_2)$ e o *kernel polinomial* dado pela equação (12).

$$k(x, y) = (x \cdot y)^d \quad (13)$$

Para o *kernel polinomial* da equação (12), em que x e $y \in \mathfrak{R}^2$ e $d=2$, vem:

$$\begin{aligned} k(x, y) &= (x \cdot y)^2 \\ k(x, y) &= ((x_1 y_1) + (x_2 y_2))^2 \\ k(x, y) &= x_1^2 y_1^2 + 2x_1 x_2 y_1 y_2 + x_2^2 y_2^2 \\ k(x, y) &= (x_1^2, \sqrt{2}x_1 x_2, x_2^2) \cdot (y_1^2, \sqrt{2}y_1 y_2, y_2^2) \\ k(x, y) &= \Phi(x) \cdot \Phi(y) \end{aligned} \quad (14)$$

O conjunto de equações (13) acima apresentadas mostra que, através de kernels, torna-se desnecessário mapear os vetores de entrada em espaços de maior dimensão e realizar o produto interno entre estes, pois isto é feito implicitamente pelos kernels no próprio espaço de entrada. Nestas equações, o produto interno, de vetores x e y , é mapeado não linearmente em \mathfrak{R}^3 através do kernel polinomial.

Para construir classificadores não-lineares (Figuras 7 e 8), substituem-se na equação (7) cada vetor de treinamento x_i por $\Phi(x_i)$ para obter o hiperplano de separação ótimo no espaço de características F . Uma vez que a utilização dos kernels é mais eficiente computacionalmente, a função de decisão possui a seguinte forma:

$$f(x) = \text{sign} \left(\sum_i v_i \cdot (x \cdot x_i) + b \right) \quad (15)$$

$$f(x) = \text{sign} \left(\sum_i v_i \cdot (\Phi(x) \cdot \Phi(x_i)) + b \right) \quad (16)$$

$$f(x) = \text{sign}\left(\sum_i v_i \cdot k(x, x_i) + b\right) \quad (17)$$

Os parâmetros v_i são calculados como um problema de programação quadrática. O hiperplano de separação ótimo obtido em F corresponde a uma função não linear e determina-se sua forma através do kernel escolhido. Destaca-se que, apesar do hiperplano ser linear no espaço de características, no espaço de entrada este corresponde a uma função de decisão não linear, cuja forma é determinada pelo kernel utilizado. A seguir, na Tabela 1, são mostrados exemplos dos kernel mais utilizados na literatura.

Tabela 1 – Exemplos de kernel

Kernel	Expressão	Parâmetros
Polinomial	$k(x_i, x_j) = ((x_i \cdot x_j) + a)^d$	a, d
RBF	$k(x_i, x_j) = \exp\left(-\frac{1}{2\delta^2} \ x_i - x_j\ ^2\right)$	δ^2
Perceptron	$k(x_i, x_j) = \tanh(\beta_0(x_i \cdot x_j) + \beta_1)$	β_0, β_1

2.2.4

Classificação em Múltiplas Classes

As SVMs foram desenvolvidas originalmente para operar de maneira binária, ou seja, para duas classes. Porém, a grande maioria dos problemas reais possui mais de duas classes [9], [11], [39], [40], [41], [42], [43], [44], [45], [46], [47], [48]. O problema de múltiplas classes pode ser tratado de duas formas. Uma é reduzir o problema de múltiplas classes a um conjunto de problemas binários. Os métodos que utilizam esta abordagem são:

decomposição um por classe (“one-against-all”) e separação das classes duas a duas (“one-against-one”). Outra abordagem é a generalização de SVMs binárias para mais de duas classes (método de Crammer e Singer [6], [8], [49]).

O método de decomposição um por classe é construído a partir de SVMs binárias. São criadas k SVMs binárias e cada SVM separa uma classe de todas as outras. Em seguida, os resultados de todas SVMs são agrupados e realiza-se a classificação desejada nas k classes.

O método de separação das classes duas a duas também é construído a partir das SVMs binárias. São criadas k SVMs binárias e cada SVM distingue um par de classes. Constroem-se $k(k-1)/2$ SVMs. Da mesma forma que no caso de decomposição um por classe, os resultados são agrupados e então a classificação é realizada.

Os detalhes de cada um desses métodos são apresentados nas seções a seguir.

2.2.4.1

Decomposição um por Classe

Neste método, o número de classificadores, SVMs construídas, é igual ao número de classes existentes no problema. O procedimento é realizado de modo a separar uma classe de todas as outras (de modo binário), ou seja, utilizam-se todos os padrões de treinamento rotulando a classe que se quer separar com saída $y=1$ e as demais classes com $y=-1$.

O problema a ser solucionado é encontrar os valores ótimos para o vetor de pesos w^i e o bias b^i , considerando o conjunto de treinamento $\{(x_1, y_1), \dots, (x_N, y_N)\}$, para que as restrições:

$$(w^i \cdot \varphi(x_t)) + b^i + (\xi^i)_t \geq 1, \text{ se } y_t = i \quad (18)$$

$$(w^i \cdot \varphi(x_t)) + b^i + (\xi^i)_t \leq -1, \text{ se } y_t \neq i \quad (19)$$

$$(\xi^i)_t \geq 0, t = 1, \dots, N,$$

sejam satisfeitas quanto à minimização da função

$$\Phi(w^i, \xi^i) = \frac{1}{2}(w^i \cdot w^i) + C \sum_{t=1}^N (\xi^i)_t \quad (20)$$

A saída da função de decisão é dada por:

$$(w^i \cdot \varphi(x)) + b^i \quad (21)$$

A classe de um dado ponto é encontrada por meio da combinação das k saídas das SVMs, podendo ser feito de várias maneiras. Um exemplo é a combinação linear das k saídas, podendo também ser utilizada a saída de maior valor da função de decisão.

2.2.4.2

Separação das Classes Duas a Duas

Este método, da mesma forma que o método de decomposição um por classe, utiliza cada SVM com dados de duas classes. São construídas $k(k-1)/2$ SVMs binárias. As restrições e a função que se pretende minimizar são análogas ao apresentado na seção 2.2.3.2, conforme mostrado a seguir:

$$(w^{ij} \cdot \varphi(x_t)) + b^{ij} + (\xi^{ij})_t \geq 1, \text{ se } y_t = i \quad (22)$$

$$(w^{ij} \cdot \varphi(x_t)) + b^{ij} + (\xi^{ij})_t \leq -1, \text{ se } y_t = j \quad (23)$$

$$(\xi^{ij})_t \geq 0, t = 1, \dots, N,$$

$$\Phi(w^{ij}, \xi^{ij}) = \frac{1}{2}(w^{ij} \cdot w^{ij}) + C \sum_{t=1}^N (\xi^{ij})_t \quad (24)$$

Há duas maneiras de se combinar as saídas das SVMs binárias: estratégia de voto ou uso de um grafo acíclico dirigido.

A decisão da classe por estratégia de voto [9] é feita somando-se um voto à classe i se o sinal de $(w^{ij} \cdot \varphi(x)) + b^{ij}$ for positivo; caso contrário soma-se um voto à classe j . O método com uso do grafo acíclico dirigido é chamado DAGSVM (*Directed Acyclic Graph SVM*). A decisão da classe é feita por meio de um grafo com um nó externo (raiz), $k(k-1)/2$ nós internos e k folhas[6], [9].

3

Modelo para Extração de Regras Fuzzy a partir de Máquinas de Vetores Suporte – FREx_SVM

3.1

Introdução

Em problemas de classificação, pode ser importante ter-se um conhecimento explícito de como o resultado é obtido. No caso de SVMs, que são classificadores do "tipo caixa preta", os métodos de extração de conhecimento desenvolvidos [50], [51] e [52] forneceram como resultado regras do tipo simbólicas Para revelar um conhecimento linguístico foi criado o modelo de extração de regras FREx_SVM [6], capaz de extrair regras fuzzy para classificação binária ou em múltiplas classes fazendo uso das projeções dos vetores suporte de uma SVM treinada.

A opção pelo uso dos vetores suporte para a extração de regras fuzzy deve-se ao fato de estes vetores definirem o hiperplano ótimo de separação entre as classes existentes no problema. Logo, é possível dizer que um vetor suporte é representativo de uma classe.

3.2

FREx_SVM – Caso Binário

O método de extração de regras para o caso de classificação binária baseia-se nos seguintes passos:

- Treina-se uma SVM e obtêm-se os vetores suporte;
- Para cada vetor suporte encontrado durante o treinamento, verifica-se a sua projeção nos eixos coordenados;
- Define-se um número de conjuntos fuzzy para cada coordenada;
- A partir dos conjuntos fuzzy e dos vetores suporte, as regras são geradas.

A Figura 9 sintetiza o funcionamento do modelo FREx_SVM.

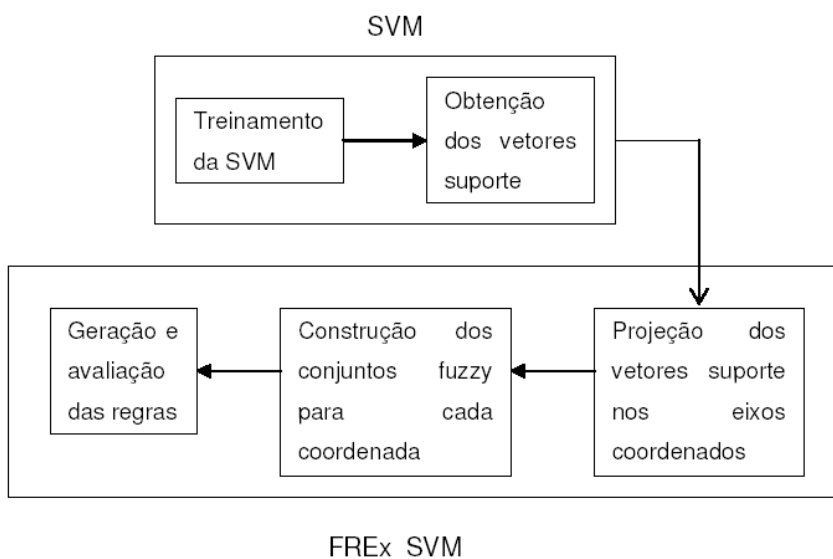


Figura 9 – Modelo de extração de conhecimento explícito de SVM

As etapas acima citadas são detalhadas nas seções a seguir.

3.2.1

Obtenção das Projeções dos Vetores Suporte

A primeira etapa consiste em obter a projeção dos vetores suporte gerados pela SVM binária nas coordenadas cartesianas. O número de projeções é igual à dimensão do espaço de entradas (número de atributos de entrada). Considere-se o caso bidimensional e seja $s = (0,61; 0,88)$ um vetor suporte de certa classe. A projeção deste vetor suporte no eixo horizontal é 0,61 e no eixo vertical, 0,88 [6]. Estas projeções são mostradas na Figura 10.

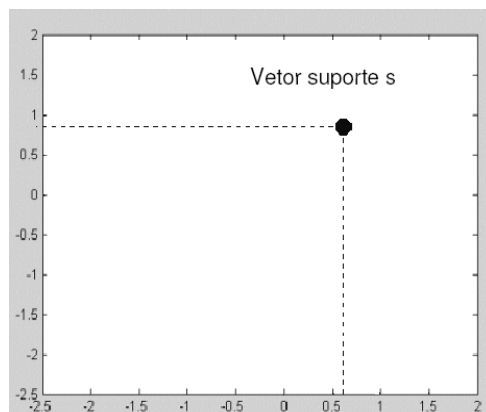


Figura 10 – Projeção do vetor suporte s

3.2.2

Definição dos Conjuntos Fuzzy

Na segunda etapa, são criados conjuntos fuzzy triangulares de suporte fixo definidos a priori. Na Figura 11, é mostrada a construção de 5 conjuntos fuzzy para um espaço de entrada de dimensão 2.

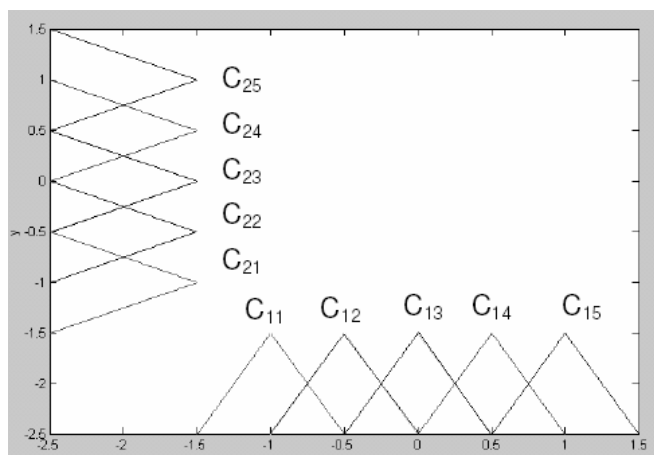


Figura 11 – Definição dos conjuntos fuzzy

Após a construção dos conjuntos e a projeção de cada vetor suporte em uma dada coordenada, encontra-se o grau de pertinência associado a cada conjunto fuzzy. Atribui-se o ponto pertencente ao conjunto que possui o maior grau de pertinência.

O procedimento é o seguinte: Seja x_i a projeção do vetor suporte na i -ésima coordenada e $\mu_{C_{ij}}(x_i)$ o grau de pertinência de x_i ao conjunto C_{ij} . Calcula-

se, $\max_{j \in \{1, \dots, t\}} \{\mu_{C_{ij}}(x_i)\}$ e determina o conjunto C_{ij} em que o grau de pertinência é máximo [6].

A Figura 12 mostra a projeção dos vetores suporte s do exemplo anterior, no eixo horizontal ($x_1=0,61$) e no eixo vertical ($x_2=0,88$). Os graus de pertinência para este vetor suporte são: para o eixo x_1 , $\mu_{C_{14}}(x_1) = 0,78$ e $\mu_{C_{15}}(x_1)=0,22$ e para o eixo x_2 , $\mu_{C_{24}}(x_2) = 0,24$ e $\mu_{C_{25}}(x_2)=0,76$ [6]. Atribui-se, então, a variável x_1 ao conjunto C_{14} e a variável x_2 ao conjunto C_{25} , que possuem maior grau de pertinência. Os conjuntos que possuem grau de pertinência 0 para este exemplo não são mostrados na Figura 12.

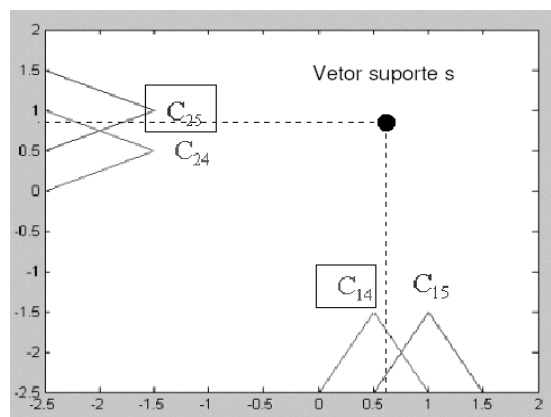


Figura 12 – Conjuntos fuzzy com maior grau de pertinência

3.2.3

Extração de Regras Fuzzy

Para cada vetor suporte encontrado durante a fase de treinamento, gera-se uma regra fuzzy [6]. O procedimento para extrair as regras é explicado a seguir.

Para cada vetor suporte x , seja C_{ij} o conjunto fuzzy da i -ésima coordenada no qual o grau de pertinência x_i é máximo, como definido na etapa anterior, onde $j \in \{1, \dots, t\}$. Seja $p = (p_1, \dots, p_n)$ um ponto de \mathcal{R}^n . A regra gerada pelo vetor suporte x será:

Regra: Se $\{p_1 \in C_{1j}1, \dots, p_k \in C_{nj}n\}$, então p é da classe do vetor suporte x .

Na Figura 12, a regra gerada é a seguinte:

Se $\{p_1 \in C_{14}$ e $p_2 \in C_{25}\}$, então p é a classe do vetor suporte s .

Pode ocorrer que dois vetores suporte de classes diferentes gerem regras conflitantes (mesmo antecedente e consequentes diferentes); neste caso calcula-se a acurácia fuzzy (seção 3.2.4) e em seguida é selecionada a regra que possuir o maior valor.

Para ilustrar o funcionamento do algoritmo FREx_SVM, considera-se um exemplo simples, mostrado na Tabela 2, que consiste em seis padrões (P1,..., P6) de dois atributos (peso e altura) definidos em duas classes (-1 e 1). Após o treinamento da SVM, determinam-se P1 e P3, da classe -1, e P2, da classe 1, como os vetores suporte (em negrito, na Tabela 2) [6].

Tabela 2 – Exemplo ilustrativo do funcionamento do FREx_SVM

Padrões	Atributos		Classe
	Peso	Altura	
P1	60	1,75	-1
P2	65	1,60	1
P3	70	1,80	-1
P4	76	1,85	-1
P5	80	1,65	1
P6	85	1,68	1

Para cada coordenada, dois conjuntos fuzzy de formato triangular e complementares são construídos, isto é, define-se para o atributo peso os conjuntos C_{11} e C_{12} e para o atributo altura, os conjuntos C_{21} e C_{22} , representados na Figura 13.

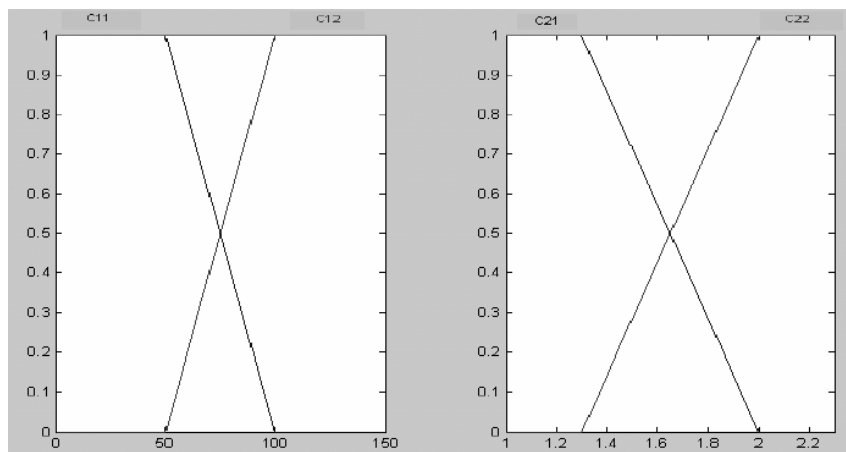


Figura 13 – Conjuntos fuzzy para os atributos peso e altura

Os graus de pertinência gerados a partir da projeção dos vetores suporte para cada atributo são apresentados nas Tabelas 3 e 4.

Tabela 3 – Graus de pertinência para o atributo peso

Padrão	Peso	μ_{C11}	μ_{C12}
P1	60	0,8	0,2
P2	65	0,7	0,3
P3	70	0,6	0,4
P4	76	0,48	0,52
P5	80	0,4	0,6
P6	85	0,3	0,7

Tabela 4 – Graus de pertinência para o atributo altura

Padrão	Altura	μ_{C21}	μ_{C22}
P1	1,75	0,42	0,58
P2	1,60	0,67	0,33
P3	1,80	0,33	0,67
P4	1,85	0,25	0,75
P5	1,65	0,58	0,42
P6	1,68	0,46	0,54

Para definir a qual conjunto fuzzy pertence cada padrão, basta observar a projeção deste na Figura 13 e verificar o que possui maior grau de pertinência associado. No exemplo apresentado, para o padrão P1, a primeira coordenada pertence ao conjunto C_{11} e a segunda coordenada C_{22} . Desta forma, a regra gerada para o ponto P1 será:

Regra 1 (P1): Se $\{x_1 \text{ é } C_{11} \text{ e } x_2 \text{ é } C_{22}\}$, então $x = (x_1, x_2)$ é da classe -1.

Analogamente, para os padrões P2 e P3, as regras geradas são:

Regra 2 (P2): Se $\{x_1 \text{ é } C_{11} \text{ e } x_2 \text{ é } C_{21}\}$, então $x = (x_1, x_2)$ é a classe 1.

Regra 3 (P3): Se $\{x_1 \text{ é } C_{11} \text{ e } x_2 \text{ é } C_{22}\}$, então $x = (x_1, x_2)$ é a classe -1.

Os vetores suporte P1 e P3 da classe -1 geram a mesma regra. Apesar de existirem seis padrões de entrada, somente três destes são vetores suporte; das três regras geradas, duas são iguais.

Na seção 3.2.4, a seguir, é apresentada a definição de acurácia fuzzy, métrica utilizada para avaliação das regras.

3.2.4

Acurácia Fuzzy

A acurácia de uma regra mede o quão bem ela descreve os dados [18]. Este conceito foi estendido para regras fuzzy, a fim de determinar o quão bem uma determinada regra descreve uma específica classe c . Assim, a acurácia fuzzy da regra r associada à classe c , $c \in \{1, 2, \dots, k\}$ é dada por:

$$FA_c^r = \frac{\sum_{i=1}^{m_c} \mu_r(x_i)}{\sum_{j=1}^m \mu_r(x_j)} \quad (25)$$

Onde:

μ_r - produto dos graus de pertinência do padrão de dados para cada conjunto presente no antecedente da regra r ;

m_c - número de padrões da classe c ;

m - número total de padrões

Para cada regra r , a soma das acurácias será igual a 1.

Se um banco de dados tem um número diferente de padrões por classe, um fator de correção CF é definido para compensar a distribuição não uniforme do padrão [53]:

$$CF_c^r = \frac{1}{m_c \left(\sum_{j=1}^k \frac{FA_j^r}{m_j} \right)}. \quad (26)$$

Neste caso a acurácia fuzzy será:

$$(FA_c^r)^* = FA_c^r \cdot CF_c^r \quad (27)$$

Para ilustrar o algoritmo de FReX_SVM, considerou-se um banco de dados simples, conforme mostrado na Tabela 2. Este banco consiste em seis padrões (P1,...,P6) e dois atributos (peso e altura), divididos em duas classes {1,-1}. Neste exemplo, depois do treinamento da SVM, P1 e P3, da classe -1 e P2 da classe 1, foram eleitos como vetores suporte (em negrito na Tabela 2).

Para cada coordenada foram construídos dois conjuntos fuzzy complementares. Usando a notação adotada anteriormente, o atributo peso tem conjuntos de C11 e C12, e o atributo altura tem conjuntos C21 e C22. Os graus de pertinência resultantes, para cada atributo de entrada no banco de dados, são mostrados nas Tabelas 3 e 4.

Os conjuntos com maior grau de pertinência associado determinam as regras, conforme regras descritas na seção anterior.

Nota-se que, conforme descrito na seção anterior, embora a entrada seja composta por seis padrões, somente três destes são os vetores suporte (que poderiam produzir três regras), no entanto duas regras (diferentes) são geradas.

Para as regras geradas este exemplo, as acurácias fuzzy são [54]:

Regra 1:

$$FA_{-1}^1 = \frac{0,8 * 0,58 + 0,6 * 0,67 + 0,48 * 0,75}{0,8 * 0,58 + 0,7 * 0,33 + 0,6 * 0,67 + 0,48 * 0,75 + 0,4 * 0,42 + 0,3 * 0,54} = 0,69.$$

Regra 2:

$$FA_1^2 = \frac{0,7 * 0,67 + 0,4 * 0,58 + 0,3 * 0,46}{0,8 * 0,42 + 0,7 * 0,67 + 0,6 * 0,33 + 0,48 * 0,25 + 0,4 * 0,58 + 0,3 * 0,46} = 0,56$$

Como observado na seção anterior, neste exemplo, as Regras 1 e 3 são iguais; assim calcula-se somente a acurácia para uma delas.

Os dados no exemplo são igualmente distribuídos em duas classes; Portanto, o valor Fo fator de correção é 1 para ambas as classes. Para os valores obtidos acima, pode-se concluir que a regra 1 fornece acurácia de 69% de todos os padrões que pertencem à classe -1, isto é, a regra 1 sozinha classifica corretamente 69% dos padrões pertencentes à classe -1, enquanto a regra 2 produz uma acurácia (menor) de 57% de todos os padrões pertencentes à classe 1.

3.3

FREx_SVM – Caso de Múltiplas Classes

Uma aplicação é dita como de múltiplas classes quando apresenta um número de classes maior do que dois. O primeiro passo para extrair regras de casos de múltiplas classes é determinar dentre os métodos de classificação apresentados na seção 2.2.4 qual será utilizado. Escolhido o método de classificação, a SVM é treinada e obtêm-se os vetores suporte.

A Figura 14 mostra um diagrama do modelo FREx_SVM para múltiplas classes.

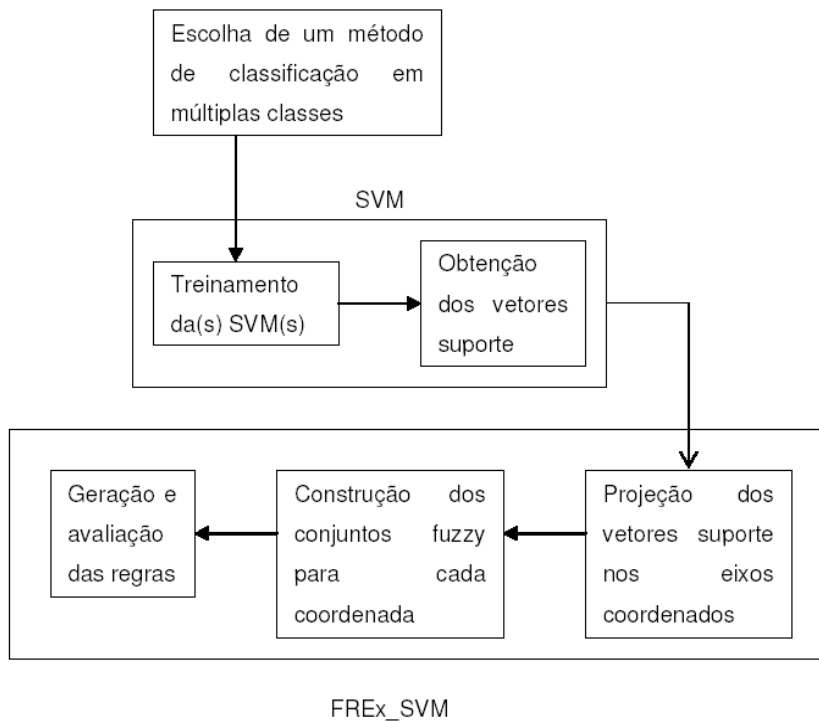


Figura 14 – Modelo FREx_SVM para múltiplas classes

Cada método de classificação gera um conjunto de regras; portanto deve-se avaliar qual método é mais adequado a um determinado conjunto de dados.

Conforme mencionado anteriormente, o método de decomposição um por classe constrói k SVMs para separar uma classe das demais. Portanto, para este método, somente os vetores suporte da classe “separada” das demais definem aquela classe. Já que os vetores suporte das demais classes não definem outra classe, eles servem somente para excluir a classe que se deseja [6].

O método de separação das classes duas a duas utiliza todos os vetores suporte para definir as classes e gerar as regras fuzzy [8].

4

FREx_SVM com Otimização de Parâmetros

4.1

Introdução

Este capítulo trata da inserção, no modelo FREx_SVM, de uma capacidade de ajuste automático das funções de pertinência.

A definição do número de conjuntos fuzzy que compõem as variáveis de entrada e saída, assim como seus formatos e suportes, tem impacto direto no desempenho do sistema de inferência fuzzy [55] [56]. O objetivo de implementar o ajuste das funções de pertinência é minimizar o erro de classificação do modelo FREx_SVM, uma vez que na sua versão original, os conjuntos fuzzy são definidos a priori e suas funções de pertinência não são ajustáveis.

Nas seções a seguir são descritas a estrutura do modelo e metodologia.

4.2

Estrutura e Metodologia

O FREx_SVM com ajuste das funções de pertinência está estruturado em módulos, conforme a Figura 15 abaixo.

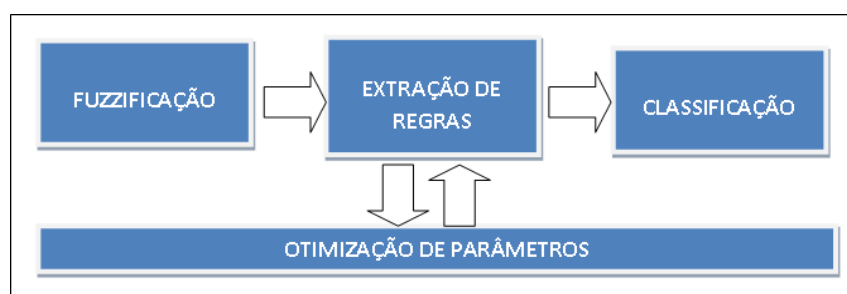


Figura 15 – Módulos do FREx_SVM com otimização de conjuntos

No módulo de Fuzzificação, a variável fuzzy de entrada é configurada, definindo-se as funções de pertinência (trapezoidal e triangular) e seus respectivos parâmetros.

No módulo de extração de regras, estas são geradas automaticamente a partir da definição dos vetores suporte de cada classe.

O módulo de laboratório permite testar configurações para as SVMs com diferentes kernels (linear, RBF e polinomial) e seus parâmetros, além de valores diferentes para a constante de regularização C .

No módulo de otimização de parâmetros é realizado o ajuste das funções de pertinência, visando à redução no erro de classificação.

Finalmente, no último módulo, a classificação é realizada.

4.2.1

Módulo de Fuzzificação

Inicialmente a variável fuzzy é configurada pelo usuário, construindo triângulos de mesma medida e com espaçamento uniforme. Neste caso foram utilizados conjuntos fuzzy triangulares e trapezoidais.

Para a geração automática de funções de pertinência, a partir de um número desejado de conjuntos, foi utilizado o conceito de intervalo de desempenho descrito na seção 4.2.3. Inicialmente as funções de pertinência apresentam uma superposição de 50%; a partir de então o módulo de otimização efetua uma busca por uma melhor configuração, sempre respeitando as restrições.

4.2.2

Módulo Extração de Regras

As regras são extraídas automaticamente, considerando-se um conjunto de treinamento definido a priori. A partir deste conjunto de treinamento definem-se os vetores suporte de cada classe. O procedimento de extração das regras é processado conforme descrito na seção 3.2.3.

4.2.3

Módulo de Otimização de Parâmetros

O módulo de otimização de parâmetros possibilita o ajuste dos suportes das funções de pertinência e, para isto, faz-se uso de algoritmos genéticos.

Definindo-se uma população de 50 indivíduos, são geradas 50 possibilidades de configurações para as funções de pertinência a cada geração. A população é da seguinte forma: um indivíduo contém a configuração do FReX_SVM (conforme Figura 16) e os outros quarenta e nove são gerados heurísticamente, respeitando as restrições descritas neste capítulo. Após a obtenção das regras para cada indivíduo, estas são avaliadas e os indivíduos com melhor desempenho são preservados para próxima geração. Ao final da execução, escolhe-se a configuração que apresenta a maior acurácia para o conjunto de regras gerado. Desta configuração são extraídas as regras.

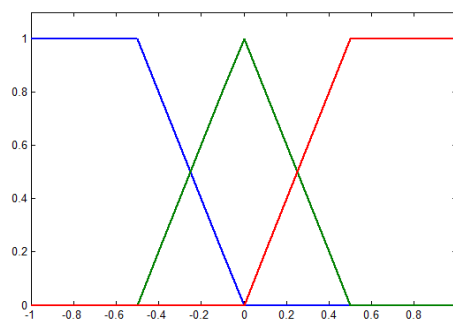


Figura 16 – Configuração Inicial para variável com 3 conjuntos fuzzy

A seguir são descritos os principais componentes da modelagem do algoritmo genético.

- **Representação** – foi utilizada a representação real, sendo que o tamanho do cromossomo é definido pelo número de conjuntos fuzzy. Cada gene representa um parâmetro do conjunto fuzzy triangular [55]. Portanto:

$$\text{tamanho do cromossomo} = 3 * \text{numero_conjuntos}$$

Assim, uma variável fuzzy com 3 conjuntos terá um indivíduo com 9 genes a serem otimizados, conforme mostrado na Figura 17.

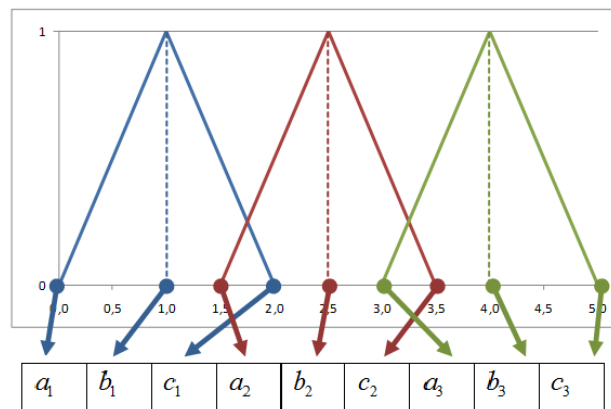


Figura 17 – Representação com variável fuzzy de 3 conjuntos (indivíduo)

- **Decodificação** – a partir do cromossoma pode-se encontrar a solução, considerando que cada conjunto de 3 genes [28] em sequência representa os parâmetros a_i , b_i e c_i do conjunto fuzzy triangular.
- **Avaliação** – o *fitness* (função de avaliação) [28] utilizado foi a acurácia fuzzy, conforme descrito na seção 3.2.4. Optou-se por esta métrica por ela permitir avaliar o quão bem uma regra gerada descreve os dados, na expectativa de que as regras geradas cobrissem e classificassem corretamente a maior parte possível dos dados disponíveis na base de dados em questão.
- **Seleção e Reprodução** – o método da roleta foi implementado para selecionar os genitores; para a reprodução utilizou-se a troca parcial da população (*steady state*) [28]. A parcela da população que será trocada a cada geração (*gap*) deve ser definida a priori. Além disso, foi semeado o conjunto fuzzy utilizado no modelo com os parâmetros fixos (sem otimização) na 1ª geração.
- **Operadores** – como se trata de um problema de representação real, foram utilizados os operadores cruzamento aritmético e mutação *creep* [28], por acreditar que estes se adequam melhor ao problema. Utilizar o

crossover aritmético garante que a partir de pais válidos sejam gerados filhos também válidos.

No crossover aritmético o filho é gerado a partir da combinação linear de seus genitores, conforme o conjunto de equações (27).

$$\begin{aligned} F_1 &= \alpha P_1 * (1 - \alpha) P_2 \\ F_2 &= \alpha P_2 * (1 - \alpha) P_1 \end{aligned} \quad (28)$$

Onde:

F_1 e F_2 filhos gerados

P_1 e P_2 genitores

α
valor aleatório no intervalo [0,1]

A mutação *creep* possui um poder maior de dispersão e pode gerar indivíduos inconsistentes; para evitar que isto ocorra, foram utilizadas pequenas variações ao redor do valor original. A mutação *creep* obedece à equação (28) abaixo:

$$X^{t+1} = \begin{cases} X^t + \Delta(\max - X^t) \\ X^t + \Delta(X^t - \min) \end{cases} \quad (29)$$

Onde:

$\Delta(s) = s * rand$

número aleatório no intervalo [0,p] $p \leq 1$

- **Parâmetros de Evolução** – devem ser definidos ao se iniciar o algoritmo: número de gerações, tamanho da população, taxas de crossover [28] e mutação.

4.2.3.1

Intervalos de Desempenho

Com a definição dos valores iniciais para os parâmetros dos conjuntos fuzzy, são calculados os intervalos de desempenho como descrito a seguir. A partir daí, a cada geração, surgem indivíduos para cada parâmetro, respeitados os limites do intervalo de desempenho. Estes garantem que da evolução sempre constarão indivíduos válidos.

O refinamento (processo de otimização das funções de pertinência), parte dos conjuntos com valores rígidos, apresentados na Figura 16. Considera-se que o suporte das funções de pertinência pode excursionar dentro de um intervalo de desempenho, tanto para esquerda quanto para a direita. Dado um conjunto fuzzy triangular i (dentre os n conjuntos associados à variável fuzzy), definido por três parâmetros a_i , b_i e c_i (Figura 18), os intervalos de desempenho garantem sempre que:

$$a_i \leq b_i \leq c_i, \forall i,$$

$$a_i \leq a_2 \leq \dots \leq a_n$$

$$b_i \leq b_2 \leq \dots \leq b_n$$

$$c_i \leq c_2 \leq \dots \leq c_n$$

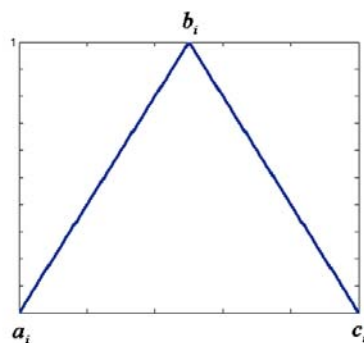


Figura 18 – Parâmetros de um conjunto fuzzy triangular

As equações (29), (30) e (31), a seguir, demonstram o cálculo dos intervalos de desempenho para os parâmetros a_i , b_i e c_i . Considere-se L como o limite do intervalo à esquerda, e R , à direita.

$$[a_i^L, a_i^R] = \left[a_i - \frac{b_i - a_i}{2}, a_i + \frac{b_i - a_i}{2} \right] \quad (30)$$

$$[b_i^L, b_i^R] = \left[b_i - \frac{b_i - a_i}{2}, b_i + \frac{c_i - b_i}{2} \right] \quad (31)$$

$$[c_i^L, c_i^R] = \left[c_i - \frac{c_i - b_i}{2}, c_i + \frac{c_i - b_i}{2} \right] \quad (32)$$

Os intervalos de desempenho são mostrados na Figura 19:

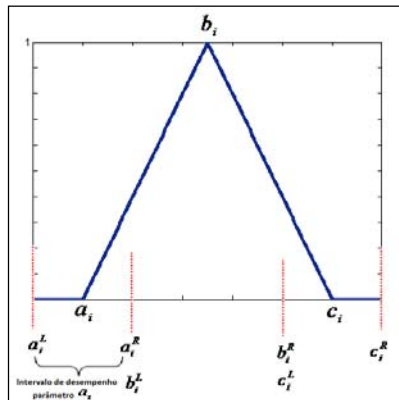


Figura 19 – Intervalos de desempenho de um conjunto fuzzy triangular

4.2.4

Classificação

Por fim, realiza-se a classificação propriamente dita. De posse do melhor indivíduo, que constitui a melhor configuração para as funções de pertinência, e das regras geradas, submete-se o conjunto de dados e realiza-se a classificação.

4.3

Banco de Dados – IRIS

Com o objetivo de validar e testar a eficiência da aplicação do algoritmo descrito neste trabalho, em especial os módulos de extração de regras e otimização de parâmetros dos conjuntos fuzzy por algoritmos genéticos, foi primeiramente utilizado o banco de dados da planta Iris.

A escolha deste banco de dados se justifica por ser ele vastamente conhecido e utilizado no campo de reconhecimento de padrões. A planta Iris se divide em três subespécies: Iris Setosa, Iris Versicolor e Iris Virgínica. Este conjunto possui 150 exemplos, sendo 50 de cada subespécie. O banco de dados usado para esta classificação está disponível no “UCI Machine Learning Repository” [57].

Existem quatro variáveis de entrada: comprimento e largura da sépala e comprimento e largura da pétala, todas medidas em centímetros.

Com objetivo de treinar e testar o algoritmo foi utilizada a validação cruzada, ou seja, os exemplos foram divididos em dois conjuntos com 75 padrões cada. Foram então realizados dois experimentos: o primeiro utilizou os 75 primeiros padrões para treinamento e os outros 75 para validação; no segundo caso, inverteram-se os conjuntos.

Foram utilizados os kernels linear e RBF (com valores de δ^2 iguais a 1, 5, 10 e 50) e valores da constante de regularização $C(0,1; 1 e 10)$, conforme [6], para uma posterior comparação de resultados. Foram geradas 15 SVMs para cada método de treinamento.

Para o treinamento das SVMs no método de decomposição um por classe foi utilizado o software de Gavin C. Cawley¹. Para o método de separação das classes duas a duas com decisão por estratégia de voto foi utilizado o software LIBSVM².

¹ <http://theoval.sys.uea.ac.uk/%7Egcc/svm/toolbox>

² <http://www.csie.ntu.edu.tw/~cjlin/libsvm>

4.3.1

Resultados Obtidos

Na extração de regras foram utilizadas as informações obtidas pelas SVMs da seção anterior, ou seja, os vetores suporte de cada uma das configurações de kernels aliados à constante de regularização C .

As Tabelas 5 e 6 apresentam, respectivamente, as taxas médias de validação de pontos classificados incorretamente e pontos não classificados, para o método de decomposição um por classe com a configuração com 3 conjuntos fuzzy.

Nesta seção, comparam-se as taxas de erro encontradas para a SVM pura (SVM) e para o modelo FReX_SVM com os conjuntos rígidos (SO) e otimizados (OTIMIZ). Os valores que obtiveram zero como resultado são representados nas Tabelas por “---”.

Tabela 5 – Método decomposição um por classe – Íris (3 conjuntos) - ERROS

DECOMPOSIÇÃO UM POR CLASSE - 3 CONJUNTOS – ERROS									
KERNEL		NSV	TREINO			TESTE			
			SVM	SO	OTIMIZ	SVM	SO	OTIMIZ	
LINEAR	C=0.1	55	2.67%	9.33%	1.33%	3.33%	9.33%	4.00%	
	C=1	48.5	3.33%	9.33%	2.67%	3.33%	9.33%	2.67%	
	C=10	46.5	4.00%	9.33%	4.00%	5.33%	9.33%	6.67%	
RBF	$\delta^2=1$	C=0.1	72.5	0.67%	9.33%	3.33%	---	9.33%	6.67%
		C=1	42	2.00%	9.33%	---	3.33%	9.33%	9.00%
		C=10	31.5	0.67%	9.33%	2.67%	5.33%	9.33%	9.00%
	$\delta^2=5$	C=0.1	73	---	9.33%	3.33%	---	9.33%	6.67%
		C=1	44	4.00%	9.33%	4.00%	5.33%	9.33%	5.33%
		C=10	24.5	1.33%	9.33%	8.00%	2.67%	9.33%	9.00%
	$\delta^2=10$	C=0.1	75	---	9.33%	3.33%	---	9.33%	6.67%
		C=1	51.5	6.00%	9.33%	4.00%	6.67%	9.33%	6.67%
		C=10	28.5	1.33%	9.33%	6.67%	2.00%	9.33%	8.00%
	$\delta^2=50$	C=0.1	75	---	9.33%	3.33%	---	9.33%	6.67%
		C=1	65.5	4.00%	9.33%	5.33%	5.33%	9.33%	8.67%
		C=10	50	2.67%	9.33%	2.67%	3.33%	9.33%	8.67%

--- = 0 (zero)

Tabela 6 – Método decomposição um por classe – Íris (3 conjuntos) – NÃO CLASSIFICADOS

DECOMPOSIÇÃO UM POR CLASSE - 3 CONJUNTOS - NÃO CLASSIFICADOS									
KERNEL		NSV	TREINO			TESTE			
			SVM	SO	OTIMIZ	SVM	SO	OTIMIZ	
LINEAR	C=0.1	55	26.67%	4.00%	<u>2.67%</u>	28.67%	11.33%	<u>8.00%</u>	
	C=1	48.5	18.67%	6.67%	<u>4.00%</u>	19.33%	12.67%	<u>5.33%</u>	
	C=10	46.5	14.00%	8.00%	---	16.67%	14.00%	<u>6.67%</u>	
RBF	$\delta^2=1$	C=0.1	72.5	34.67%	---	---	39.33%	9.33%	<u>5.33%</u>
		C=1	42	---	4.00%	3.33%	1.33%	13.33%	8.00%
		C=10	31.5	---	6.67%	5.33%	0.67%	16.00%	8.00%
	$\delta^2=5$	C=0.1	73	52.00%	---	---	55.33%	9.33%	<u>5.33%</u>
		C=1	44	---	8.00%	2.67%	---	14.67%	5.33%
		C=10	24.5	0.67%	23.33%	10.67%	1.33%	26.67%	14.67%
	$\delta^2=10$	C=0.1	75	66.67%	---	---	65.33%	9.33%	<u>5.33%</u>
		C=1	51.5	3.33%	1.33%	<u>1.33%</u>	6.00%	9.33%	6.67%
		C=10	28.5	0.67%	21.33%	10.67%	0.67%	24.00%	5.33%
	$\delta^2=50$	C=0.1	75	100%	---	---	100%	9.33%	<u>5.33%</u>
		C=1	65.5	35.53%	0.67%	<u>0.67%</u>	33.33%	10.67%	<u>5.33%</u>
		C=10	50	10.67%	4.00%	---	12.67%	11.33%	<u>10.67%</u>

--- = 0 (zero)

Para a configuração de 5 conjuntos fuzzy, são apresentadas nas abelas 7 e 8, respectivamente, as taxas médias encontradas para pontos classificados incorretamente e pontos não classificados.

Tabela 7 – Método decomposição um por classe – Íris (5 conjuntos) - ERROS

DECOMPOSIÇÃO UM POR CLASSE - 5 CONJUNTOS ERROS									
KERNEL		NSV	TREINO			TESTE			
			SVM	SO	OTIMIZ	SVM	SO	OTIMIZ	
LINEAR	C=0.1	55	2.67%	2.00%	<u>1.33%</u>	3.33%	4.00%	3.33%	
	C=1	48.5	3.33%	2.00%	<u>1.33%</u>	3.33%	4.00%	3.33%	
	C=10	46.5	4.00%	2.00%	<u>1.33%</u>	5.33%	3.33%	<u>3.33%</u>	
RBF	$\delta^2=1$	C=0.1	72.5	0.67%	2.00%	0.67%	---	4.00%	1.33%
		C=1	42	2.00%	2.00%	<u>1.33%</u>	3.33%	4.00%	<u>2.67%</u>
		C=10	31.5	0.67%	2.00%	1.33%	5.33%	4.00%	<u>3.33%</u>
	$\delta^2=5$	C=0.1	73	---	2.00%	0.67%	---	4.00%	1.33%
		C=1	44	4.00%	2.00%	<u>0.67%</u>	5.33%	4.00%	<u>3.33%</u>
		C=10	24.5	1.33%	2.00%	<u>0.67%</u>	2.67%	4.00%	3.33%
	$\delta^2=10$	C=0.1	75	---	2.00%	0.67%	---	4.00%	1.33%
		C=1	51.5	6.00%	2.00%	<u>0.67%</u>	6.67%	4.00%	<u>2.67%</u>
		C=10	28.5	1.33%	2.00%	1.33%	2.00%	4.00%	2.67%
	$\delta^2=50$	C=0.1	75	---	2.00%	0.67%	---	4.00%	1.33%
		C=1	65.5	4.00%	2.00%	<u>0.67%</u>	5.33%	4.00%	<u>2.67%</u>
		C=10	50	2.67%	2.00%	<u>0.67%</u>	3.33%	4.00%	<u>2.67%</u>

--- = 0 (zero)

Tabela 8 – Método decomposição um por classe – Íris (5 conjuntos) – NÃO CLASSIFICADOS

DECOMPOSIÇÃO UM POR CLASSE - 5 CONJUNTOS - NÃO CLASSIFICADOS									
KERNEL		NSV	TREINO			TESTE			
			SVM	SO	OTIMIZ	SVM	SO	OTIMIZ	
LINEAR	C=0.1	55	26.67%	14.67%	<u>11.33%</u>	28.67%	32.00%	<u>19.33%</u>	
	C=1	48.5	18.67%	16.00%	<u>13.33%</u>	19.33%	32.00%	19.33%	
	C=10	46.5	14.00%	13.33%	<u>10.67%</u>	16.67%	34.00%	22.67%	
RBF	$\delta^2=1$	C=0.1	72.5	34.67%	---	---	39.33%	17.33%	<u>12.00%</u>
		C=1	42	---	10.67%	8.67%	1.33%	32.00%	22.67%
		C=10	31.5	---	18.67%	14.67%	0.67%	38.00%	24.67%
	$\delta^2=5$	C=0.1	73	52.00%	---	---	55.33%	17.33%	<u>12.00%</u>
		C=1	44	---	9.33%	6.00%	---	26.00%	19.33%
		C=10	24.5	0.67%	32.00%	22.67%	1.33%	50.67%	29.33%
	$\delta^2=10$	C=0.1	75	66.67%	---	---	65.33%	17.33%	<u>12.00%</u>
		C=1	51.5	3.33%	8.00%	5.33%	6.00%	26.67%	18.67%
		C=10	28.5	0.67%	32.67%	24.67%	0.67%	48.67%	28.00%
	$\delta^2=50$	C=0.1	75	100%	---	---	100%	17.33%	<u>12.00%</u>
		C=1	65.5	35.53%	2.67%	<u>1.33%</u>	33.33%	19.33%	<u>10.67%</u>
		C=10	50	10.67%	11.33%	<u>7.33%</u>	12.67%	28.00%	19.33%

--- = 0 (zero)

Conforme visto nas Tabelas 5 a 8, para o caso da decomposição um por classe, há pontos não classificados. Isto se deve ao fato deste método ser derivado de diversas classificações binárias a partir da construção de tantas SVMs quantas forem as classes existentes no problema, onde cada SVM separa a classe pretendida das demais. Desta forma, quando os resultados das SVMs são agrupados, podem surgir pontos não classificados.

Nas Tabelas 9 a 12 encontram-se os resultados do método de separação das classes duas a duas de decisão por voto.

Tabela 9 – Método separação das classes duas a duas – Íris (3 conjuntos) – ERROS

SEPARAÇÃO DAS CLASSES - 3 CONJUNTOS – ERROS – ÍRIS									
KERNEL		NSV	TREINO			TESTE			
			SVM	SO	OTIMIZ	SVM	SO	OTIMIZ	
LINEAR	C=0.1	47.5	3.33%	9.33%	5.33%	5.33%	9.33%	8.00%	
	C=1	19.5	2.00%	9.33%	5.33%	2.67%	8.67%	8.00%	
	C=10	12	1.33%	9.33%	6.67%	4.00%	8.67%	8.00%	
RBF	$\delta^2=1$	C=0.1	70	5.33%	9.33%	<u>2.67%</u>	12.00%	9.33%	<u>6.67%</u>
		C=1	38	2.00%	9.33%	<u>1.33%</u>	4.00%	9.33%	8.67%
		C=10	30.5	0.67%	9.33%	2.67%	4.00%	9.33%	6.67%
	$\delta^2=5$	C=0.1	72.5	11.33%	9.33%	<u>2.67%</u>	12.67%	9.33%	<u>6.67%</u>
		C=1	38.5	2.67%	9.33%	4.00%	5.33%	9.33%	5.33%
		C=10	23	2.00%	9.33%	2.67%	4.00%	9.33%	5.33%
	$\Delta^2=10$	C=0.1	75	11.33%	9.33%	<u>2.67%</u>	12.67%	9.33%	<u>6.67%</u>
		C=1	45.5	4.00%	9.33%	<u>3.33%</u>	6.00%	9.33%	<u>5.33%</u>
		C=10	23	1.33%	9.33%	4.00%	2.67%	8.67%	8.00%
	$\Delta^2=50$	C=0.1	70	56.00%	9.33%	<u>2.67%</u>	56.00%	9.33%	<u>6.67%</u>
		C=1	65.5	10.67%	9.33%	<u>3.33%</u>	10.00%	9.33%	<u>5.33%</u>
		C=10	34.5	1.33%	9.33%	6.67%	3.33%	9.33%	6.67%

Tabela 10 – Método separação das classes duas a duas – Íris 3 conjuntos –NÃO CLASSIFICADOS

SEPARAÇÃO DAS CLASSES - 3 CONJUNTOS - NÃO CLASSIFICADOS - ÍRIS									
KERNEL		NSV	TREINO			TESTE			
			SVM	SO	OTIMIZ	SVM	SO	OTIMIZ	
LINEAR	C=0.1	47.5	---	8.00%	6.67%	---	13.33%	10.67%	
	C=1	19.5	---	25.33%	21.33%	---	26.67%	24.00%	
	C=10	12	---	29.33%	25.33%	---	30.67%	26.67%	
RBF	$\delta^2=1$	C=0.1	70	---	---	---	9.33%	5.33%	
		C=1	38	---	5.33%	4.00%	---	13.33%	8.00%
		C=10	30.5	---	10.67%	9.33%	---	18.67%	12.00%
	$\delta^2=5$	C=0.1	72.5	---	---	---	---	9.33%	5.33%
		C=1	38.5	---	8.00%	2.67%	---	15.00%	13.33%
		C=10	23	---	18.67%	10.00%	---	24.00%	18.67%
	$\delta^2=10$	C=0.1	75	---	---	---	---	9.33%	5.33%
		C=1	45.5	---	3.33%	4.00%	---	10.67%	8.67%
		C=10	23	---	22.67%	10.67%	---	26.00%	17.33%
	$\delta^2=50$	C=0.1	70	---	---	---	---	9.33%	5.33%
		C=1	65.5	---	0.67%	0.67%	---	10.67%	6.00%
		C=10	34.5	---	13.33%	2.67%	---	14.67%	12.67%

--- = 0 (zero)

Tabela 11 – Método separação das classes duas a duas – Íris (5 conjuntos) - ERROS

SEPARAÇÃO DAS CLASSES - 5 CONJUNTOS – ERROS – ÍRIS									
KERNEL		NSV	TREINO			TESTE			
			SVM	SO	OTIMIZ	SVM	SO	OTIMIZ	
LINEAR		C=0.1	47.5	3.33%	2.00%	<u>1.33%</u>	5.33%	4.67%	<u>3.33%</u>
		C=1	19.5	2.00%	2.00%	<u>1.33%</u>	2.67%	4.67%	4.00%
		C=10	12	1.33%	2.00%	1.33%	4.00%	4.00%	4.00%
RBF	$\delta^2=1$	C=0.1	70	5.33%	2.00%	<u>0.67%</u>	12.00%	4.67%	<u>2.67%</u>
		C=1	38	2.00%	2.00%	<u>0.67%</u>	4.00%	4.67%	<u>2.67%</u>
		C=10	30.5	0.67%	2.00%	1.33%	4.00%	4.67%	<u>3.33%</u>
	$\delta^2=5$	C=0.1	72.5	11.33%	2.00%	<u>1.33%</u>	12.67%	4.67%	<u>2.67%</u>
		C=1	38.5	2.67%	2.00%	<u>1.33%</u>	5.33%	4.67%	<u>4.00%</u>
		C=10	23	2.00%	2.00%	<u>1.33%</u>	4.00%	4.67%	4.00%
	$\delta^2=10$	C=0.1	75	11.33%	2.00%	<u>0.67%</u>	12.67%	4.67%	<u>2.67%</u>
		C=1	45.5	4.00%	2.00%	<u>1.33%</u>	6.00%	4.67%	<u>2.67%</u>
		C=10	23	1.33%	2.00%	2.00%	2.67%	4.67%	3.33%
	$\delta^2=50$	C=0.1	70	56.00%	2.00%	<u>1.33%</u>	56.00%	4.67%	<u>2.67%</u>
		C=1	65.5	10.67%	2.00%	<u>0.67%</u>	10.00%	4.67%	<u>3.33%</u>
		C=10	34.5	1.33%	2.00%	1.33%	3.33%	4.67%	4.00%

Tabela 12 – Método separação das classes duas a duas – Íris (5 conjuntos) –NÃO CLASSIFICADOS

SEPARAÇÃO DAS CLASSES - 5 CONJUNTOS - NÃO CLASSIFICADOS – ÍRIS									
KERNEL		NSV	TREINO			TESTE			
			SVM	SO	OTIMIZ	SVM	SO	OTIMIZ	
LINEAR		C=0.1	47.5	---	16.00%	12.00%	---	28.67%	15.33%
		C=1	19.5	---	50.67%	26.67%	---	56.67%	29.33%
		C=10	12	---	64.00%	29.33%	---	70.67%	33.33%
RBF	$\delta^2=1$	C=0.1	70	---	---	---	---	20.00%	10.67%
		C=1	38	---	8.00%	4.67%	---	26.67%	15.33%
		C=10	30.5	---	15.33%	7.33%	---	33.33%	22.67%
	$\delta^2=5$	C=0.1	72.5	---	0.67%	---	---	20.00%	10.67%
		C=1	38.5	---	13.33%	8.67%	---	28.00%	19.33%
		C=10	23	---	38.00%	28.00%	---	44.67%	28.67%
	$\delta^2=10$	C=0.1	75	---	---	---	---	20.00%	10.67%
		C=1	45.5	---	10.67%	6.67%	---	26.00%	15.33%
		C=10	23	---	39.33%	22.67%	---	46.00%	28.00%
	$\delta^2=50$	C=0.1	70	---	---	---	---	20.00%	10.67%
		C=1	65.5	---	6.00%	4.00%	---	23.33%	19.33%
		C=10	34.5	---	20.00%	15.33%	---	32.67%	23.33%

--- = 0 (zero)

Nota-se que no caso da configuração com cinco conjuntos fuzzy, tanto para o método decomposição um por classe (Tabelas 5 a 8) quanto para o método de separação das classes duas a duas (Tabelas 9 a 12), ocorre um aumento significativo dos dados não classificados em relação à configuração

com três conjuntos fuzzy. Esta alteração se deve ao maior particionamento dos dados em razão da pequena base de dados disponível. O método de classificação de decomposição um por classe apresentou, em geral, o melhor desempenho, devido ao fato de este gerar um maior número de vetores suporte do que no caso da separação das classes duas a duas. Este fato torna-se bastante importante, uma vez que cada vetor suporte gera uma regra.

A configuração que apresentou melhor resultado em termos de pontos classificados foi a separação das classes duas a duas (Tabelas 11 e 12), com kernel RBF com largura $\delta^2=50$ e constante de regularização $C= 0.1$, apresentando uma taxa de pontos não classificados igual a 5.63%. Esta configuração resultou em uma média de 28.5 regras, sendo 29 regras para o primeiro conjunto como teste e o segundo como treinamento, e 28 regras invertendo-se os conjuntos. A seguir, nas Figuras 20 e 21, são apresentadas as configurações, pré e pós-otimização, que apresentaram o melhor resultado.

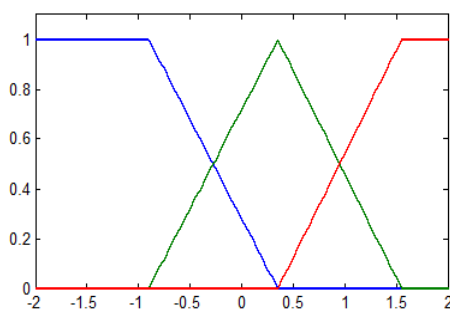


Figura 20 – Configuração Inicial dos conjuntos fuzzy para as 4 variáveis de entrada – IRIS

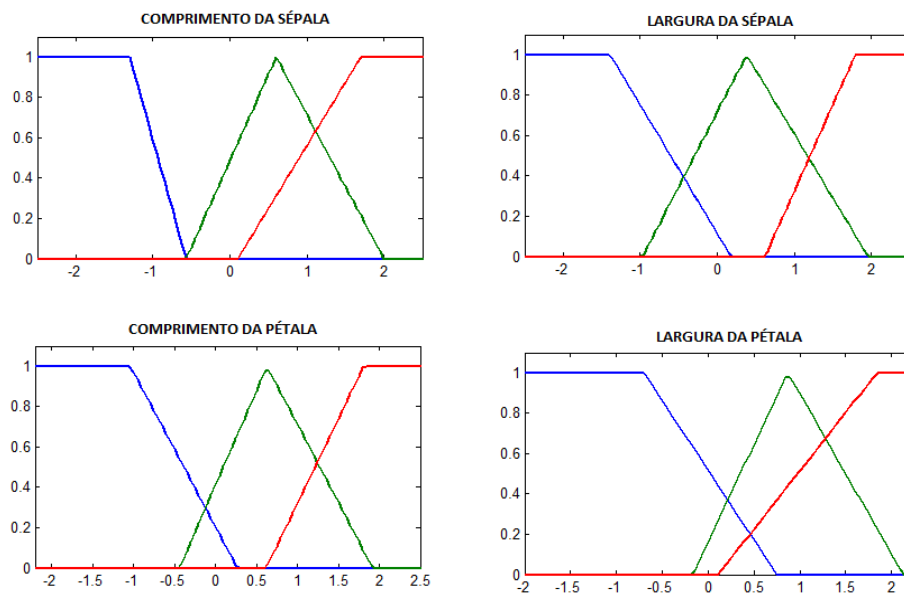


Figura 21 – Configuração das variáveis de entrada após a otimização – ÍRIS

5

DIREF (Diagnóstico Inteligente por Resposta em Frequência)

5.1

Introdução

A dependência da energia elétrica atualmente é tamanha que mesmo uma interrupção temporária pode causar grandes prejuízos. A preocupação com a continuidade do fornecimento de energia se torna cada vez maior, uma vez que a energia elétrica passou à iniciativa privada em vários países (em particular no Brasil) e é vista como um negócio.

Os transformadores de potência são equipamentos fundamentais do Sistema Elétrico de Potência (SEP) e, assim, o estudo de ferramentas de monitoramento e diagnóstico que evitem a retirada não planejada destes equipamentos de operação em decorrência de falhas é de grande importância para manter a confiabilidade do sistema de energia elétrica. As falhas e paradas não programadas – danos ao equipamento em si e consequências operacionais dependentes da intensidade da falha ocorrida, demora na reposição do equipamento e interrupção do fornecimento de energia – ocasionam grandes prejuízos às empresas do SEP.

Observando a Figura 22 a seguir, percebe-se a importância dos transformadores ao longo da transmissão e distribuição do SEP.

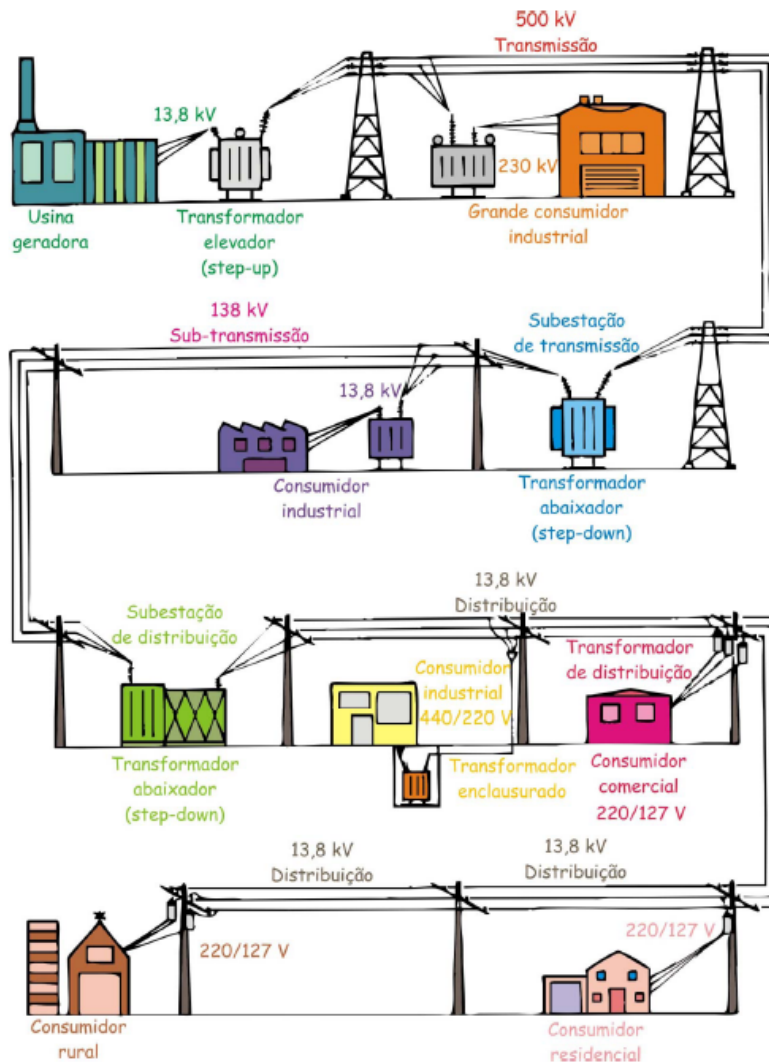


Figura 22 – Sistema Elétrico de Potência Típico

Em geral, os transformadores são equipamentos com grande confiabilidade e têm vida útil entre 20 e 35 anos. Com programas adequados de manutenção, esta vida útil pode se estender até os 60 anos. Entretanto, com o passar do tempo, e devido às solicitações a que o equipamento é exposto, as condições internas se modificam, acarretando degradação da isolação sólida e aumentando o risco de falhas.

Na Figura 23 é mostrada a vista em corte de um transformador de potência típico.

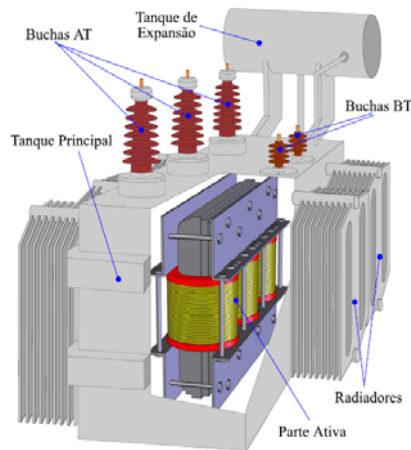


Figura 23 – Vista em corte de um transformador típico

Transformadores de potência são equipamentos bastante sensíveis e qualquer alteração na sua forma construtiva, ainda que mínima, pode ser uma causa de falha ou defeito. Neste contexto, destaca-se a Análise de Resposta em Frequência (FRA) como uma metodologia de diagnóstico [1] [2] [3] [4] [5]. Apesar de ser uma técnica consagrada e utilizada em larga escala para monitoramento e diagnóstico de transformadores, não existe uma normatização específica quanto à metodologia para realização dos ensaios ou interpretação dos resultados, que geralmente é feita de modo visual e comparativo e requer a análise de um especialista. Além disso, muitas vezes esta análise não é conclusiva.

5.2

Análise de Resposta em Frequência

De forma geral, o diagnóstico baseado em FRA é uma técnica que se fundamenta na comparação de medições em duas ocasiões distintas. A técnica consiste em apresentar o comportamento do sistema analisado frente a uma ampla faixa de frequências. É, portanto, muito sensível a alterações nas características elétricas dos enrolamentos oriundas de vários tipos de esforços mecânicos e/ou elétricos (transporte, terremotos, perda de força de compressão, curtos-circuitos etc.). O teste é não-invasivo e não destrutivo, podendo ser usado sozinho para identificar danos em enrolamentos ou como complemento a outros métodos tais como fator de potência [1]. Apesar do grande poder de detecção de

alterações nos enrolamentos de transformadores, sua interpretação ainda é feita de forma muito subjetiva e dependente de um especialista [1], [2], [3], [4] e [5].

Embora este ensaio tenha sido concebido para equipamentos novos ou reformados que estejam entrando em serviço, ele vem sendo utilizado também em fases distintas, ao longo da vida do transformador, sendo capaz de identificar possíveis alterações geométricas, causadas por esforços eletromecânicos aos quais o equipamento é exposto. Estas alterações geométricas são cumulativas, e podem ocorrer ao longo da vida útil do equipamento sem que seja caracterizado um defeito. Assim, o transformador continua em operação até que esta condição ocasione uma falha no seu dielétrico, ocasionando grandes perdas econômicas [58].

Em operação normal, o transformador de potência opera em uma frequência única (50 ou 60 Hz) determinada pelo sistema elétrico ao qual ele está ligado [59]. Desta forma, habitualmente considera-se que seu circuito equivalente é constituído por componentes resistivas e indutivas (Figura 24(a)), pois estas são as predominantes nesta frequência. Entretanto, como qualquer equipamento elétrico, o transformador possui componentes capacitivas (vide Figura 24(b)), chamadas de *capacitâncias parasitas*.

As componentes reativas (indutivas e capacitivas) do transformador são fortemente dependentes da geometria da parte ativa. Portanto, as medições destas componentes podem indicar a ocorrência de defeitos e/ou falhas, e neste caso estas componentes se tornam importantes e não devem ser desprezadas.

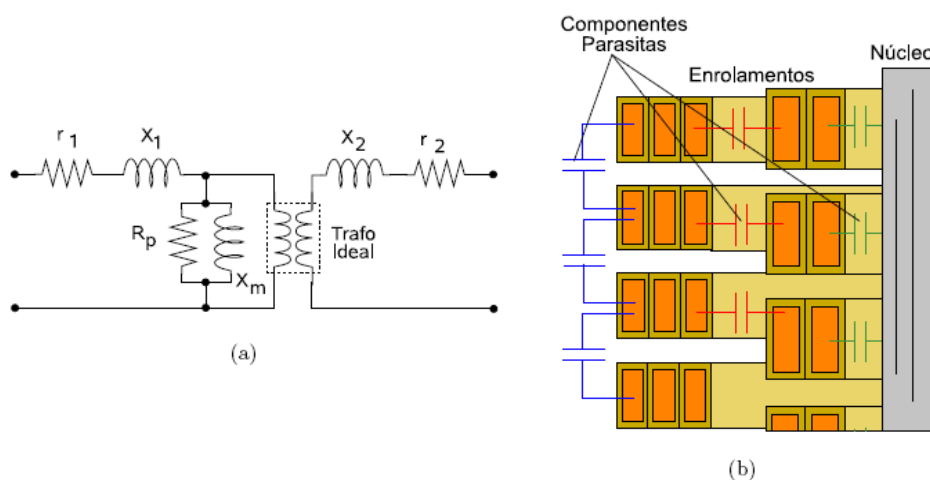


Figura 24 – (a) circuito elétrico de um transformador ideal (b) vista em corte da parte ativa do transformador mostrando as capacitâncias parasitas

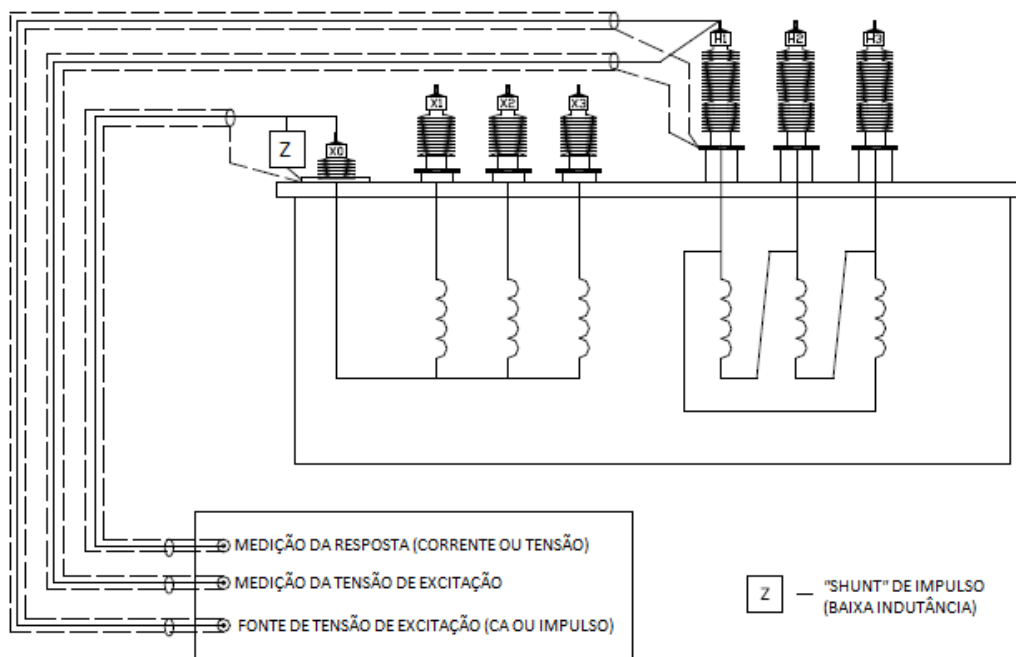


Figura 25 - Arranjo típico para medição de admitância própria e de transferência (IEEE, 2005).

Conforme ilustrado na Figura 25, a técnica consiste em aplicar um sinal de tensão numa das extremidades de um enrolamento e medir o efeito desta excitação num outro ponto qualquer acessível do transformador, podendo ser na outra extremidade deste mesmo enrolamento ou de outro qualquer. O sinal medido poderá ser em forma de tensão ou corrente, sendo que na medição de corrente é possível obter tanto a impedância quanto a admitância entre esses pontos [60].

Na Figura 26 é mostrado o arranjo com a medição no domínio da frequência aplicando sinais senoidais na faixa de frequência desejada. Este processo é repetido n vezes alterando-se a frequência do sinal aplicado; no final deste processo é obtido um conjunto de valores relacionais que se tornam a assinatura do transformador.

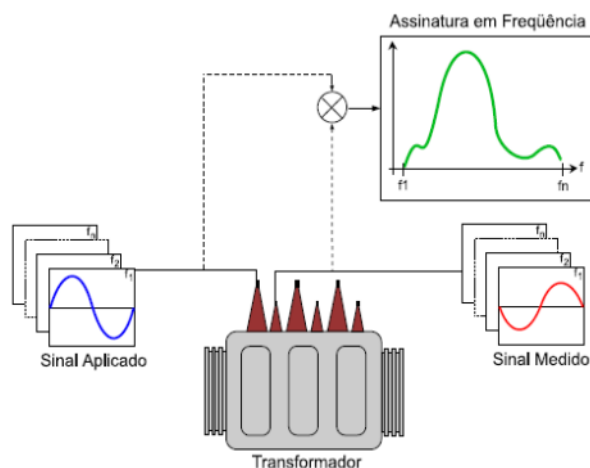


Figura 26 – Medição no Domínio da Frequência

Os desvios identificados são tomados em relação a um resultado obtido como referência, conhecido como 'assinatura'. Na falta de uma assinatura, essas comparações podem ser realizadas por meio de medições de transformadores que possuam as mesmas características elétricas ou mecânicas, por unidades irmãs (mesmo modelo), ou ainda por comparação entre as fases (no caso de transformadores trifásicos). O procedimento de diagnóstico é ilustrado na Figura 27.

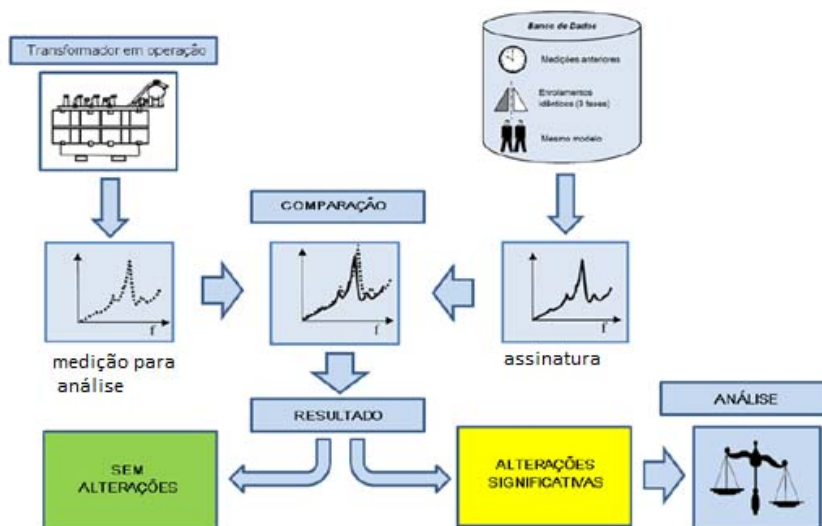


Figura 27 – Procedimento para realização da Análise de Resposta em Frequência

A análise feita a partir das medições de análise de resposta em frequência, conforme dito anteriormente é dependente de especialistas. Portanto é importante a utilização de um sistema automático que seja capaz de realizar o diagnóstico de forma objetiva e que tenha precisão além da percepção visual. Dentro desta ótica, o modelo FREx_SVM estendido (com ajuste de conjuntos fuzzy) desenvolvido nesta dissertação foi aplicado ao problema de diagnóstico de transformadores a partir das medições de impedância terminal.

As seções a seguir apresentam a aplicação do FREx_SVM estendido, detalhando a construção do banco de dados (composto por dados medidos e simulados), os tipos de defeitos/falhas estudados e sua aplicação na ferramenta desenvolvida.

5.3

Medições e Simulações - DIREF

Uma vez que se dispunha de uma quantidade insuficiente de dados reais para treinar e testar o algoritmo construído, o banco de dados foi complementado com dados simulados através do software ATP (Alternative Transient Program). Os procedimentos de medição e simulação são descritos nas seções a seguir.

5.3.1

Medições Realizadas

A seguir, são apresentados os procedimentos de medição realizados para o caso de análise de resposta em frequência.

5.3.1.1

Medição – Transferência de Tensão

A medição de resposta em frequência é realizada aplicando-se um sinal senoidal (geralmente entre 5 Vef e 10 Vef), com frequência variável na faixa de 10 Hz a 10 MHz, a um dos enrolamentos do equipamento, e medindo-se a transferência deste sinal para outro enrolamento, caracterizando uma medição

da relação de transformação em frequências distintas de 60 Hz. Na medição é considerada a maneira como os enrolamentos são conectados internamente, ou seja, a medição é realizada entre enrolamentos correspondentes. Desta forma, o número de medições depende do modelo do transformador [61].

Na Figura 28 é apresentado um diagrama simplificado de medição para um transformador trifásico com ligação delta e estrela, com a aplicação do sinal no lado de alta tensão do transformador (ligação estrela) e a medição do lado de baixa tensão (ligação delta).

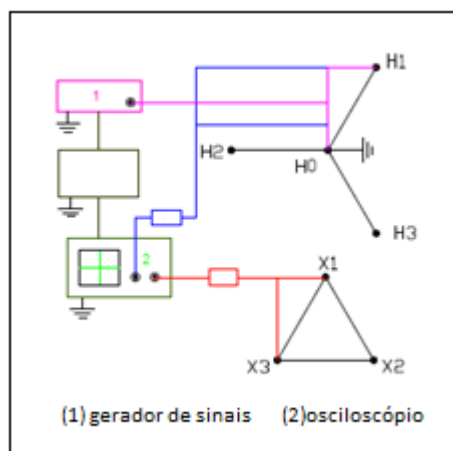


Figura 28 – Diagrama simplificado de medição de resposta em frequência

Nos gráficos de transferência de tensão, é apresentada a relação entre os valores dos sinais de tensão de saída e entrada ($V_{saída}/V_{entrada}$, no eixo das ordenadas), em função da frequência (no eixo das abscissas). A relação $V_{saída}/V_{entrada}$ é geralmente normalizada tendo como base a relação nominal do transformador na frequência característica.

Na Figura 29 é mostrado um exemplo de medição de transferência de tensão.

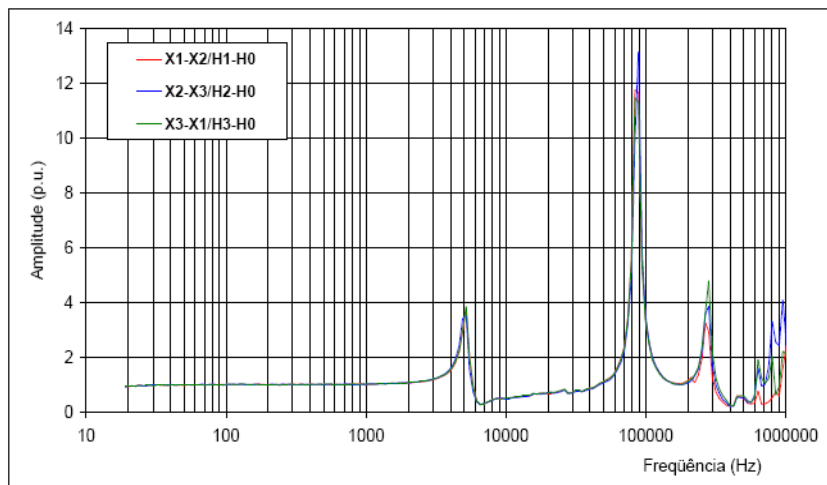


Figura 29 – Resposta típica para um transformador trifásico, tensões aplicadas nos enrolamentos H e medidas nos enrolamentos X correspondentes.

5.3.1.2

Medição – Impedância Terminal

As medições de impedância são efetuadas por meio de pontes tipo RLC e têm como princípio a injeção de um sinal de tensão através de um oscilador e a consequente medição da corrente circulante resultante. Da relação entre a tensão aplicada e a corrente resultante obtém-se a medição de impedância. Esta medida, portanto é uma medida indireta [1].

Conecta-se a ponte de impedância aos terminais externos do transformador referentes ao enrolamento que se pretende analisar e mede-se a impedância variando-se a frequência em uma faixa de interesse (em geral 10 Hz a 1MHz). O número de medições que serão realizadas é determinado pela quantidade de enrolamentos do transformador. No caso de um transformador trifásico com ligações delta e estrela devem-se realizar seis medições. A Figura 30 mostra um exemplo típico de medição de um autotransformador.

Para este trabalho optou-se por utilizar as curvas de impedância terminal, visto que, no caso das simulações, elas proporcionam uma maior facilidade de simulação do que as curvas de transferência de tensão.

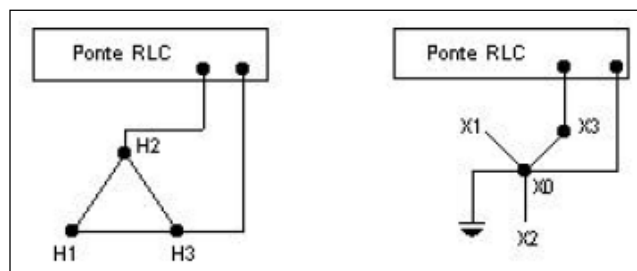


Figura 30 – Transformador trifásico, medições terminais H e X

A Figura 31 mostra uma medição típica do módulo da impedância de uma fase de autotransformador.

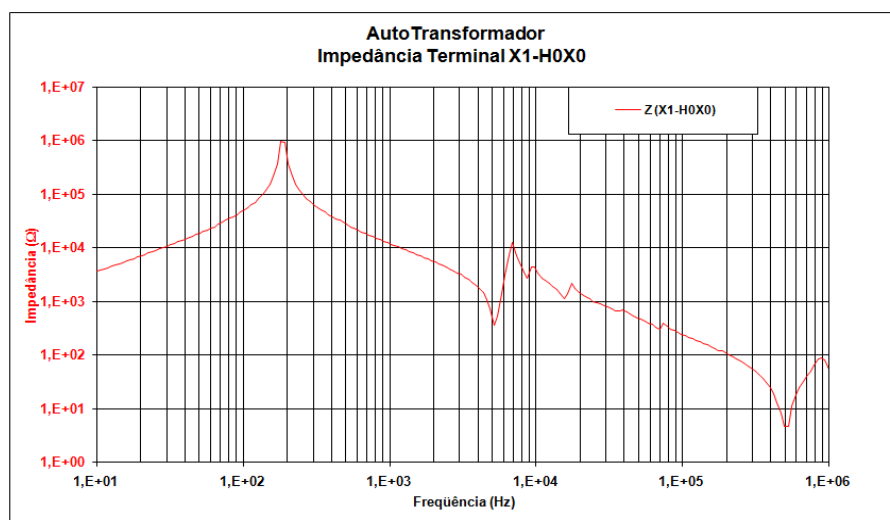


Figura 31 – Autotransformador – módulo da impedância terminal

5.3.2

Simulações de Defeitos

Simulações fizeram-se necessárias pela indisponibilidade de um número suficiente de medições reais. Além disso, simulações, não estão expostas a uma série de interferências que ocorrem no campo, com conseqüente reflexão nas curvas de medição.

A teoria na qual os ensaios FRA são baseados é fundamentada no fato de que a impedância de qualquer rede complexa RLC, ou inversamente sua admitância, é variante com a frequência de excitação. Além disso, mudanças

ocorridas em elementos da rede se refletem em mudanças nas respostas obtidas nas diferentes frequências.

Qualquer transformador de potência pode ser representado por seu circuito equivalente na forma de uma rede complexa de impedâncias, constituída de resistências, capacitâncias e indutâncias em série e paralelo, ou seja, uma rede RLC. Tal modelagem tem a vantagem de permitir a associação de parâmetros elétricos a grandezas físicas do transformador.

O que determina se ocorrerá uma amplificação (ressonância) ou uma atenuação (anti-ressonância) na medição de impedância é a configuração das componentes RLC do circuito. Um arranjo RLC paralelo produz uma ressonância, enquanto um arranjo RLC série produz uma anti-ressonância, conforme Figura 32.

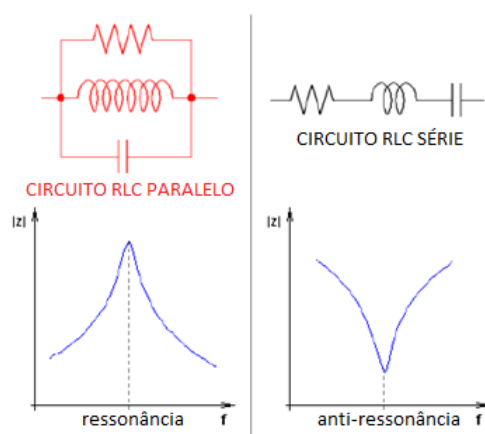


Figura 32 – Correspondência entre arranjos RLC e curva de impedância

Observando-se a Figura 33, pode-se afirmar que, antes da primeira ressonância, o enrolamento tem o comportamento tipicamente indutivo, pois a impedância aumenta com a frequência e tem ângulo próximo a 90° . Isto é esperado, por se tratar de baixas frequências. Após a primeira ressonância ocorre o contrário, quando o enrolamento passa a ter um comportamento capacitivo. A primeira ressonância ocorre justamente quando as componentes indutivas e capacitivas se anulam. Chegando-se ao segundo ponto de ressonância, ocorre uma nova inversão no comportamento do enrolamento.

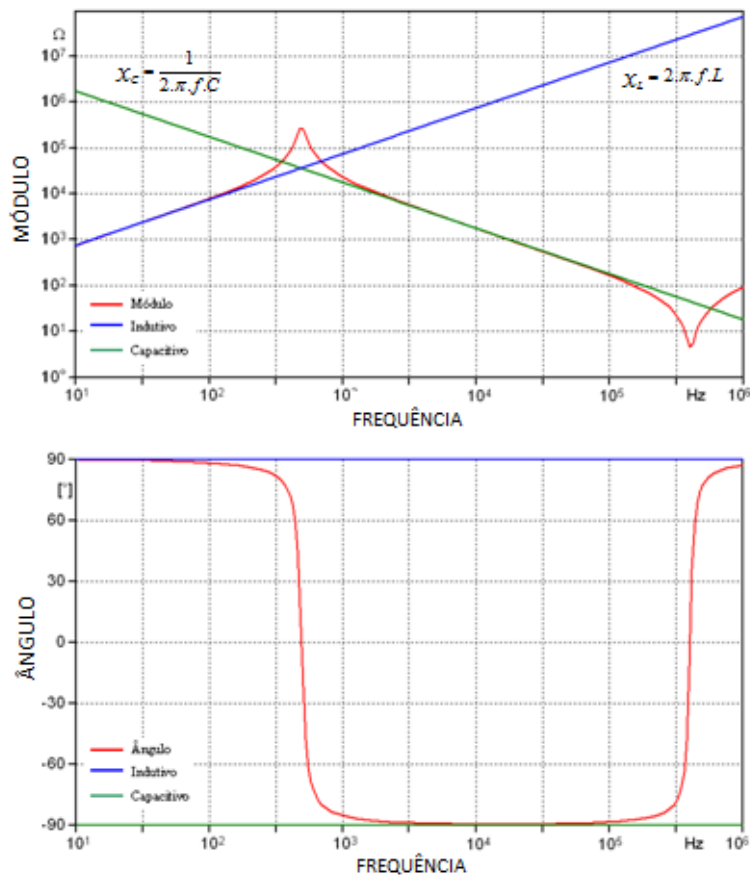


Figura 33 – Componentes reativas na medição

O modelo proposto para o enrolamento de um transformador é uma associação dos arranjos série e paralelo, mostrado na Figura 34. O indutor L_1 e o capacitor C_2 definem o primeiro ponto de ressonância. O segundo ponto é definido por C_2 e L_2 . Os resistores R_1 e R_2 têm a função de estabelecer o valor máximo e mínimo, respectivamente, da impedância. O resistor R_0 estabelece o valor inicial, ou seja, a frequência inicial da impedância [62].

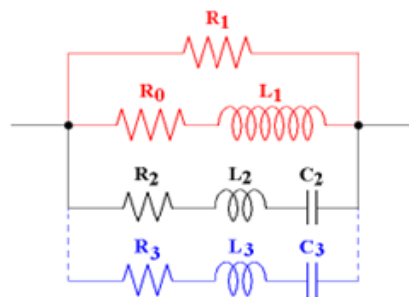


Figura 34 – Modelo representando o enrolamento no domínio da frequência

Na Figura 35, é mostrado como se realiza a simulação do circuito a partir do gráfico de medição de impedância terminal. Para simulação dos defeitos são alterados parâmetros dos circuitos que correspondem a regiões de defeito/falha nas curvas, conforme [3] [4].

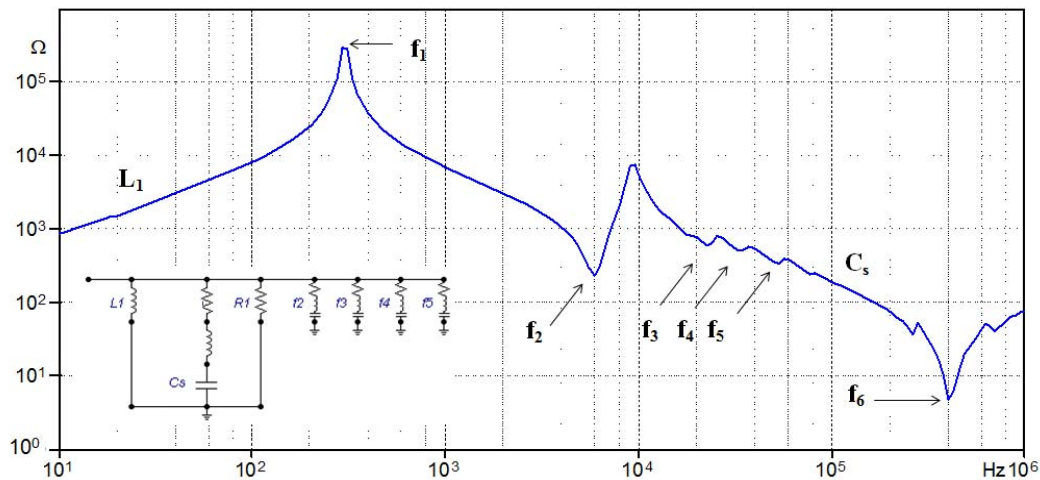


Figura 35 - Correspondência entre o circuito simulado e as ressonâncias da curva de impedância

Devido às dimensões do transformador, seria esperado um modelo de parâmetros distribuídos ao longo da extensão do enrolamento, conforme ilustrado na Figura 24(b). Na representação por meio de parâmetros distribuídos são consideradas as dimensões e as características geométricas do equipamento, além disso, deve-se levar em conta que os parâmetros mudam ao longo das dimensões espaciais. Entretanto, durante o projeto de um transformador é calculada uma distribuição linear dos campos elétricos e magnéticos no enrolamento [62], e o modelo de parâmetros distribuídos pode ser simplificado para um modelo de parâmetros concentrados, conforme o circuito apresentado na Figura 35. A representação por meio de parâmetros concentrados não tem significado físico direto em termos de equivalência, mas é justificada quando se deseja modelar a tendência do comportamento de um sistema, com a vantagem da redução dos custos computacionais em relação aos parâmetros distribuídos [62].

As alterações aplicadas às curvas assim como sua classificação correspondente foram baseadas em [3], [4], [5] e [63].

5.3.2.1

Defeitos Estudados

A resposta em frequência é sensível a alterações que ocorram nos parâmetros do núcleo, isolamento elétrico, dos enrolamentos, ligações internas ao tanque do equipamento e buchas. Portanto, alterações na amplitude da curva e/ou deslocamentos dos picos de ressonância e anti-ressonância podem ser um indicativo de defeito.

É possível, ainda, relacionar a natureza dos parâmetros com a região do espectro de frequência que sofreu alteração. Assim a região de baixas frequências compreende os defeitos de origem indutiva e caracteriza os curtos-circuitos; no entanto, a região de frequências mais altas corresponde aos defeitos de origem capacitiva que são os deslocamentos geométricos. [64]

Alteração nos Materiais Isolantes (Permissividade do Óleo)

A vida útil de um transformador está intimamente ligada ao estado de seu isolamento elétrico. O envelhecimento de um transformador compromete a parte ativa que pode evoluir para falhas dielétricas nos enrolamentos.

As variações da permissividade do óleo refletem-se nas curvas de resposta em frequência e constitui uma alternativa para avaliação do estado dos materiais. A permissividade do óleo é representada nos cálculos pela capacitância geométrica.

O cálculo das capacitâncias requer o conhecimento da permissividade dielétrica dos materiais isolante, sendo na sua forma complexa escrita como: [1]
[2]:

$$\xi = \xi_r \cdot \xi_0 \cdot (1 - j \tan \delta) \quad (33)$$

Onde:

ξ_r – permissividade relativa do material

ξ_0 – permissividade relativa no vácuo ($8.85434 \cdot 10^{-12}$ F/m)

$\tan \delta$ – tangente do ângulo de perdas

Na Figura 36 é mostrado um gráfico de simulação da impedância de um transformador com e sem óleo.

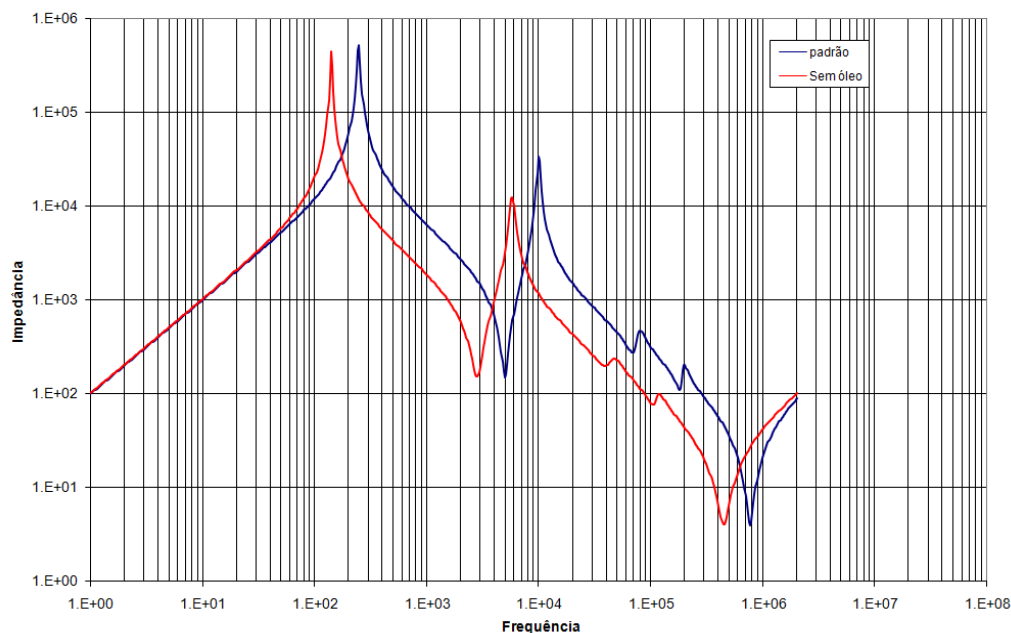


Figura 36 – Simulação de impedância terminal padrão e sem óleo

Este defeito foi somente simulado, pois na prática nem sempre é possível medir o transformador preenchido com óleo isolante, por exemplo, durante o transporte, onde comumente encontra-se preenchido com ar sintético ou nitrogênio. Como forma de comissionar o equipamento, atualmente na prática mede-se o transformador em fábrica e após o transporte (quando chega a campo); ambas as medições são realizadas sem o preenchimento do equipamento com óleo isolante. Desta forma, simular o efeito com e sem óleo torna-se bastante relevante.

Conforme ilustrado na Figura 37, é possível observar que, além da alteração da permissividade do óleo, pode ocorrer uma alteração nos canais de refrigeração (canais de passagem de óleo) ao longo do transformador, podendo ocasionar sobreaquecimento.

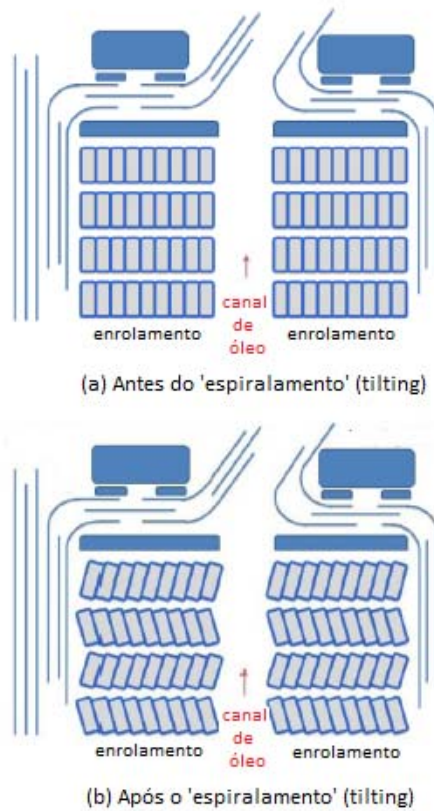


Figura 37 – Modificação do canal de óleo após o 'espiralamento' (tilting)

A seguir, na Figura 38, é mostrado um enrolamento que sofreu aquecimento. Uma das possíveis causas é falha na isolação.



Figura 38 – Sobreaquecimento local

Curto-circuito

Este tipo de defeito proporciona a retirada de operação do transformador pela atuação da proteção. Como deve ser identificado antes da falha em si, é evidente a importância do monitoramento e diagnóstico. Mesmo com a atuação do sistema de proteção, os danos causados são grandes e algumas vezes estendidos a equipamentos adjacentes a partir de explosões seguidas de incêndios. Simula-se este defeito a partir da modificação da relutância. Nas Figuras 39 e 40 são mostrados defeitos originados a partir de curtos-circuitos de origem interna ou externa.



Figura 39 – Curto-circuito entre espiras

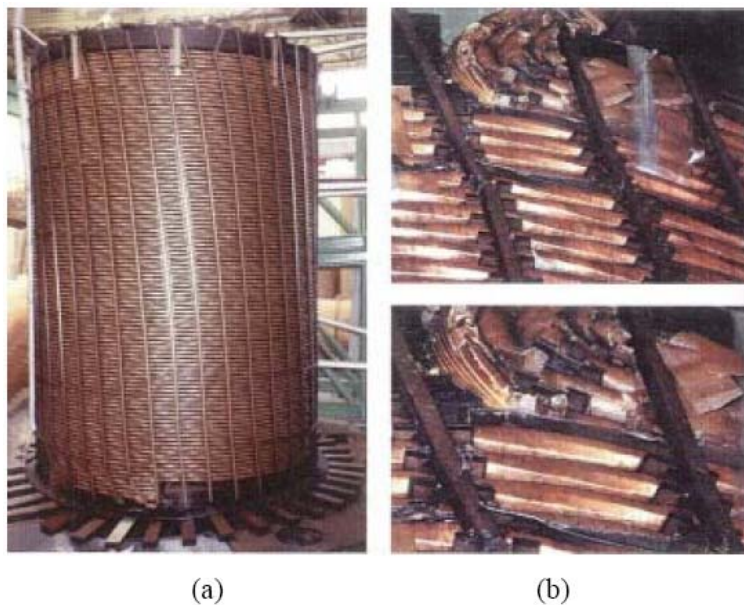


Figura 40 – Espiralamento (a) causado em teste de curto-circuito e detalhes de saída (b) nitidamente deformadas

A Figura 41 mostra medições de impedância terminal de um autotransformador monofásico antes e após o curto-circuito.

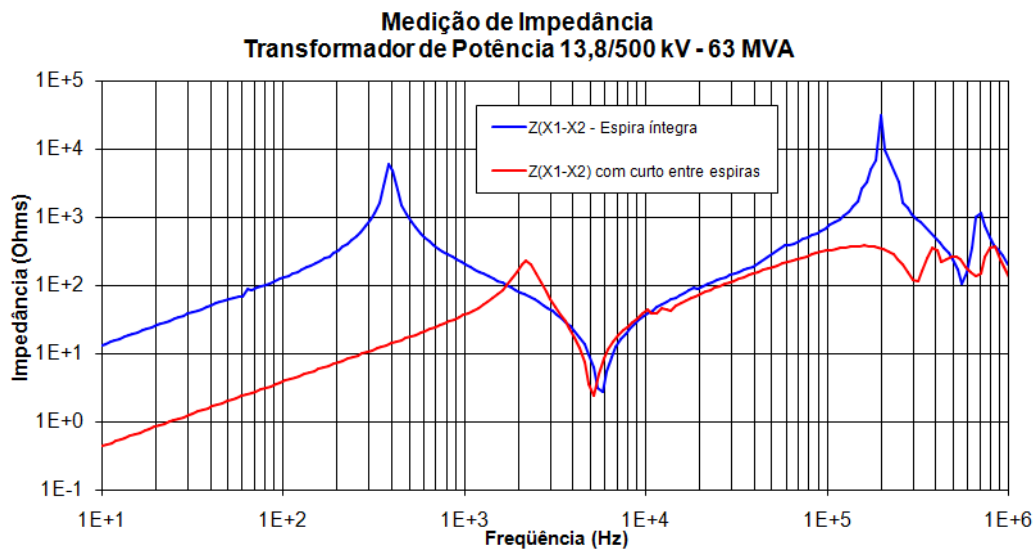


Figura 41 – Medição de impedância terminal antes e após o curto-circuito

Observa-se um deslocamento da curva de impedância ao longo da frequência e amplitude.

Deslocamento Axial

O deslocamento axial é caracterizado pela modificação da capacitância série ao longo do enrolamento, resultante da: alteração das espessuras referentes ao condutor que compõe a bobina, distância entre bobinas, distância entre bobina e anel estático, largura dos calços e largura do condutor que compõe a bobina. Nas Figuras 42 e 43 são mostrados, respectivamente, um enrolamento resultante de um deslocamento axial e um gráfico de medição de impedância terminal antes e após um deslocamento axial.



Figura 42 – Enrolamento resultante de deslocamento axial

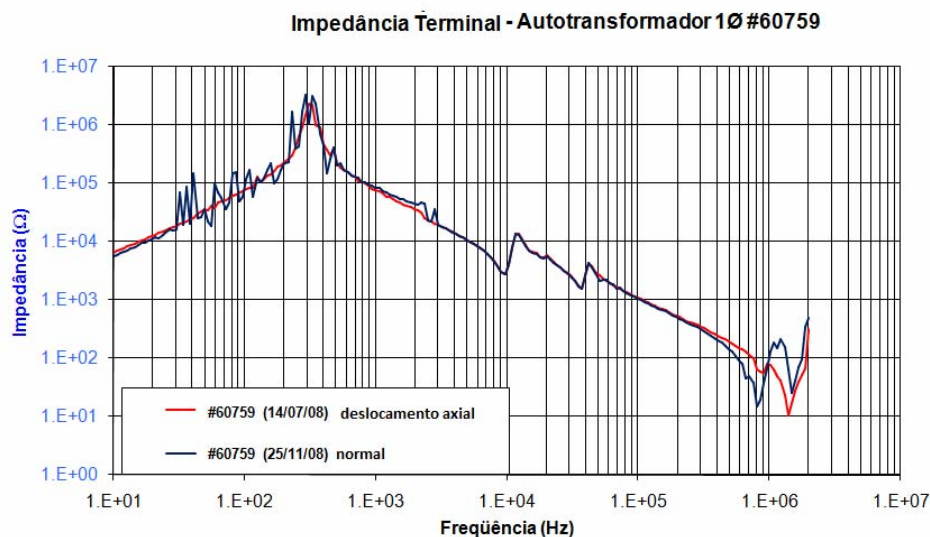


Figura 43 - Medição de impedância terminal antes e após o deslocamento axial

5.4

Banco de Dados – IMPEDÂNCIA TERMINAL

O banco de dados de análise de resposta em frequência, (medição da Impedância Terminal) contendo 142 exemplos, foi construído da seguinte forma: para as condições normais foram usados 60 exemplos reais, para curto-circuito, 16 exemplos reais e 22 simulados, para deslocamento axial, 2 exemplos reais e 20 simulados, e para o caso de defeito na isolamento, 22 exemplos simulados (em virtude da dificuldade de realizar medições nas duas condições).

O modelo FREx_SVM estendido foi adaptado ao caso de análise de resposta em frequência por meio da utilização das curvas de medição de impedância terminal, constituindo o DIREF (Diagnóstico Inteligente por Resposta em Frequência). Observou-se a consistência das regras geradas e a sua relação com os métodos de análise existentes.

A exemplo de [6], foram utilizados, kernels linear e RBF e constantes de regularização C iguais a 0,1; 1; 10 e 100.

Da mesma forma que no caso do banco de dados Iris (ver Seção 4.3), para o treinamento das SVMs no método decomposição um por classe foi utilizado o software de Gavin C. Cawley. Para o método de separação das classes duas a duas com decisão por estratégia de voto foi utilizado o software LIBSVM.

5.4.1

Tratamento dos Dados

Para as curvas de resposta em frequência do banco de dados, efetuou-se o cálculo da diferença entre as curvas [65], com o objetivo de explicitar as regiões onde ocorre a maior diferença entre as curvas de assinatura do transformador e a medição que se deseja analisar. Após o cálculo da diferença, dividiu-se a curva em três regiões de frequência, conforme orientação da Norma Chinesa [66]. A primeira região compreendendo de 1 a 100 kHz, a segunda, de 100 a 600 kHz, e a terceira, de 600 kHz a 1MHz. Calculou-se a média das diferenças em cada região para obter um número crisp que pudesse ser fuzzificado. Além disso, encontrou-se o valor de frequência onde ocorreu a maior diferença entre as curvas. Todas essas informações foram utilizadas como variáveis de entrada da máquina de vetor suporte. Desta forma, foram configuradas quatro variáveis de entrada: a média da diferença das curvas nas três regiões e a frequência onde ocorreu a maior diferença.

Em seguida, os dados foram processados de forma que cada conjunto de variáveis de entrada tivesse média zero e desvio padrão igual a um, configurando uma distribuição normal, de forma a assegurar que nenhum atributo tivesse maior influência na resolução do problema.

5.4.2

Resultados Obtidos

O tamanho do banco de dados impôs um limite à partição das variáveis de entrada. Assim, foi utilizada uma configuração com 3 conjuntos fuzzy associados a cada variável.

Os resultados obtidos para os métodos de decomposição um por classe encontram-se nas Tabelas 13 e 14, onde são mostradas as taxas de pontos classificados incorretamente e pontos não classificados, respectivamente.

Na Tabela 15, é mostrado o número de regras gerado para cada configuração, e na Tabela 16 as taxas de acerto para cada caso.

Analogamente são mostrados nas Tabelas 17 a 20 os resultados para o método separação das classes duas a duas. Assim como no caso do banco de dado Íris, em todas as tabelas são apresentados os resultados utilizando SVM pura (SVM), FReX_SVM sem otimização de parâmetros (SO) e o FReX_SVM com otimização dos conjuntos fuzzy (OTIMIZ).

Tabela 13 – Método decomposição um por classe – Impedância - ERROS

DECOMPOSIÇÃO UM POR CLASSE - 3 CONJUNTOS - ERROS									
KERNEL		NSV	TREINO			TESTE			
			SVM	SO	OTIMIZ	SVM	SO	OTIMIZ	
LINEAR	C=0.1	53.5	---	32.39%	22.53%	3.00%	45.07%	29.57%	
	C=1	41.5	---	37.32%	30.98%	---	45.07%	42.95%	
	C=10	34	---	37.32%	31.69%	---	45.07%	40.84%	
	C=100	38.5	---	30.28%	21.83%	---	43.66%	35.91%	
RBF	$\delta^2=1$	C=0.1	71	---	30.98%	16.90%	---	45.07%	31.69%
		C=1	48	---	30.98%	27.45%	---	45.07%	35.21%
		C=10	30	---	30.98%	30.98%	---	44.36%	44.36%
		C=100	27.5	---	30.98%	30.98%	---	44.36%	44.36%
	$\delta^2=5$	C=0.1	71	---	30.98%	16.90%	---	45.07%	31.69%
		C=1	61	---	30.98%	26.76%	2.82%	45.07%	33.09%
		C=10	39	2.82%	30.98%	26.05%	2.82%	41.55%	35.21%
		C=100	29	---	30.98%	30.98%	---	44.36%	44.36%
	$\delta^2=10$	C=0.1	71	---	30.98%	16.90%	---	45.07%	31.69%
		C=1	69	1.41%	30.98%	21.12%	2.82%	45.07%	43.66%
		C=10	41.5	---	30.98%	26.05%	0.71%	42.25%	40.14%
		C=100	29.5	2.82%	30.98%	30.98%	3.52%	41.55%	41.55%
	$\delta^2=50$	C=0.1	71	---	30.98%	16.90%	---	45.07%	31.69%
		C=1	71	---	30.98%	16.90%	---	45.07%	31.69%
		C=10	60	---	30.98%	30.98%	2.82%	45.07%	45.07%
		C=100	38	---	30.98%	30.98%	---	42.25%	42.25%

--- = 0 (zero)

Tabela 14 – Método decomposição um por classe – Impedância – NÃO CLASSIFICADOS

DECOMPOSIÇÃO UM POR CLASSE - 3 CONJUNTOS - NÃO CLASSIFICADOS									
KERNEL		NSV	TREINO			TESTE			
			SVM	SO	OTIMIZ	SVM	SO	OTIMIZ	
LINEAR	C=0.1	53.5	35.21%	4.30%	<u>2.11%</u>	35.00%	11.97%	<u>11.97%</u>	
	C=1	41.5	30.98%	11.97%	<u>9.15%</u>	34.51%	15.49%	<u>12.67%</u>	
	C=10	34	29.58%	13.38%	<u>5.63%</u>	33.09%	19.02%	<u>16.19%</u>	
	C=100	38.5	29.58%	11.26%	<u>10.56%</u>	33.09%	18.30%	<u>15.49%</u>	
RBF	$\Delta^2=1$	C=0.1	71	65.49%	---	---	69.02%	7.75%	<u>5.63%</u>
		C=1	48	6.34%	0.70%	<u>0.70%</u>	11.27%	7.75%	<u>4.93%</u>
		C=10	30	---	3.52%	2.81%	2.12%	12.67%	9.16%
		C=100	27.5	---	2.81%	2.81%	2.12%	9.16%	9.16%
	$\Delta^2=5$	C=0.1	71	100%	---	---	100%	7.75%	<u>5.63%</u>
		C=1	61	33.10%	1.41%	<u>0.70%</u>	32.39%	8.45%	<u>8.45%</u>
		C=10	39	7.04%	2.11%	<u>2.11%</u>	15.49%	11.97%	<u>9.86%</u>
		C=100	29	---	4.93%	4.93%	2.12%	11.38%	11.38%
	$\delta^2=10$	C=0.1	71	100%	---	---	100%	7.75%	<u>5.63%</u>
		C=1	69	33.80%	0.70%	---	35.21%	7.75%	<u>4.22%</u>
		C=10	41.5	31.69%	2.11%	<u>2.11%</u>	36.62%	11.27%	<u>9.15%</u>
		C=100	29.5	2.12%	4.23%	4.23%	7.75%	15.49%	15.49%
	$\delta^2=50$	C=0.1	71	97.18%	---	---	97.18%	7.75%	<u>5.63%</u>
		C=1	71	97.18%	---	---	97.18%	7.75%	<u>5.63%</u>
		C=10	60	35.21%	2.82%	<u>2.11%</u>	35.21%	10.56%	<u>4.93%</u>
		C=100	38	30.98%	4.23%	<u>4.23%</u>	33.80%	14.08%	<u>11.38%</u>

--- = 0 (zero)

Tabela 15 – Método decomposição um por classe – Impedância – REGRAS

KERNEL		NSV	REGRAS		
			SO	OTIMIZ	
LINEAR	C=0.1	53.5	11	23.5	
	C=1	41.5	8.5	18	
	C=10	34	8.5	18.5	
	C=100	38.5	9.5	17.5	
RBF	$\delta^2=1$	C=0.1	71	14	25.5
		C=1	48	12.5	21
		C=10	30	11	11
		C=100	27.5	11	11
	$\delta^2=5$	C=0.1	71	14	25.5
		C=1	61	12.5	21
		C=10	39	11.5	20.5
		C=100	29	10	10
	$\delta^2=10$	C=0.1	71	14	25.5
		C=1	69	12.5	23.5
		C=10	41.5	11.5	19
		C=100	29.5	10.5	10.5
	$\delta^2=50$	C=0.1	71	14	25.5
		C=1	71	14	25.5
		C=10	60	12.5	20
		C=100	38	11.5	17

Tabela 16 – Método decomposição um por classe – Impedância – TAXAS DE ACERTO

DECOMPOSIÇÃO UM POR CLASSE - 3 CONJUNTOS - TAXAS DE ACERTO									
KERNEL		NSV	TREINO			TESTE			
			SVM	SO	OTIMIZ	SVM	SO	OTIMIZ	
LINEAR	C=0.1	53.5	64.79%	63.31%	<u>75.36%</u>	62.00%	42.96%	58.46%	
	C=1	41.5	69.02%	50.71%	59.87%	65.49%	39.44%	44.38%	
	C=10	34	70.42%	49.30%	62.68%	66.91%	35.91%	42.97%	
	C=100	38.5	70.42%	58.46%	67.61%	66.91%	38.04%	48.60%	
RBF	$\delta^2=1$	C=0.1	71	34.51%	69.02%	<u>83.10%</u>	30.98%	47.18%	<u>62.68%</u>
		C=1	48	93.66%	68.32%	71.85%	88.73%	47.18%	59.86%
		C=10	30	100%	65.50%	66.21%	97.88%	42.97%	46.48%
		C=100	27.5	100%	66.21%	66.21%	97.88%	46.48%	46.48%
	$\delta^2=5$	C=0.1	71	---	69.02%	<u>83.10%</u>	0.00%	47.18%	<u>62.68%</u>
		C=1	61	66.90%	67.61%	<u>72.54%</u>	64.79%	46.48%	58.46%
		C=10	39	90.14%	66.91%	71.84%	81.69%	46.48%	54.93%
		C=100	29	100%	64.09%	64.09%	97.88%	44.26%	44.26%
	$\delta^2=10$	C=0.1	71	---	69.02%	<u>83.10%</u>	---	47.18%	<u>62.68%</u>
		C=1	69	64.79%	68.32%	<u>78.88%</u>	61.97%	47.18%	52.12%
		C=10	41.5	68.31%	66.91%	<u>71.84%</u>	62.67%	46.48%	50.71%
		C=100	29.5	95.06%	64.79%	64.79%	88.73%	42.96%	42.96%
	$\delta^2=50$	C=0.1	71	2.82%	69.02%	<u>83.10%</u>	2.82%	47.18%	<u>62.68%</u>
		C=1	71	2.82%	69.02%	<u>83.10%</u>	2.82%	47.18%	<u>62.68%</u>
		C=10	60	64.79%	66.20%	<u>66.91%</u>	61.97%	44.37%	50.00%
		C=100	38	69.02%	64.79%	64.79%	66.20%	43.67%	46.37%

--- = 0 (zero)

Tabela 17 – Método separação das classes duas a duas – Impedância - ERROS

SEPARAÇÃO DAS CLASSES DUAS A DUAS - 3 CONJUNTOS – ERROS									
KERNEL		NSV	TREINO			TESTE			
			SVM	SO	OTIMIZ	SVM	SO	OTIMIZ	
LINEAR	C=0.1	46.5	17.61%	30.98%	20.42%	18.30%	42.25%	33.09%	
	C=1	41	17.61%	30.98%	20.42%	18.30%	42.25%	28.17%	
	C=10	41	17.61%	30.98%	20.42%	18.30%	42.25%	28.17%	
	C=100	41	17.61%	30.98%	20.42%	18.30%	42.25%	28.17%	
RBF	$\delta^2=1$	C=0.1	69.5	19.71%	30.98%	<u>19.01%</u>	19.71%	45.07%	40.14%
		C=1	46	19.01%	30.98%	28.16%	19.71%	45.07%	39.43%
		C=10	29.5	17.60%	30.98%	27.46%	18.30%	45.07%	40.14%
		C=100	26	16.19%	30.98%	26.76%	18.30%	44.36%	34.50%
	$\delta^2=5$	C=0.1	70	19.71%	30.98%	<u>16.17%</u>	19.71%	45.07%	33.09%
		C=1	59.5	19.71%	30.98%	27.46%	19.71%	45.07%	42.95%
		C=10	36	16.91%	30.98%	26.05%	17.61%	45.07%	38.73%
		C=100	28.5	14.08%	30.98%	28.87%	17.61%	44.36%	40.84%
	$\delta^2=10$	C=0.1	70	19.71%	30.98%	<u>16.17%</u>	19.71%	45.07%	33.09%
		C=1	65.5	19.71%	30.98%	28.16%	19.71%	45.07%	33.09%
		C=10	39.5	18.30%	30.98%	28.87%	19.71%	45.07%	41.54%
		C=100	29.5	14.08%	30.98%	28.87%	16.91%	44.36%	40.84%
	$\delta^2=50$	C=0.1	69.5	19.71%	30.98%	<u>16.90%</u>	19.71%	45.07%	34.50%
		C=1	69.5	19.71%	30.98%	<u>16.90%</u>	19.71%	45.07%	34.50%
		C=10	60	19.71%	30.98%	16.90%	19.71%	45.07%	34.50%
		C=100	31.5	16.19%	30.98%	28.87%	19.71%	42.25%	40.84%

Tabela 18 – Método separação das classes duas a duas – Impedância– NÃO CLASSIFICADOS

SEPARAÇÃO DAS CLASSES DUAS A DUAS - 3 CONJUNTOS - NÃO CLASSIFICADOS									
KERNEL		NSV	TREINO			TESTE			
			SVM	SO	OTIMIZ	SVM	SO	OTIMIZ	
LINEAR	C=0.1	46.5	---	4.93%	2.82%	---	14.08%	9.86%	
	C=1	41	---	7.04%	5.63%	---	17.60%	16.90%	
	C=10	41	---	7.04%	5.63%	---	17.60%	16.90%	
	C=100	41	---	7.04%	5.63%	---	17.60%	16.90%	
RBF	$\delta^2=1$	C=0.1	69.5	---	---	---	7.75%	5.63%	
		C=1	46	---	1.41%	---	8.45%	7.75%	
		C=10	29.5	---	2.11%	1.41%	8.45%	7.75%	
		C=100	26	---	2.82%	2.11%	9.15%	8.45%	
	$\delta^2=5$	C=0.1	70	---	0.70%	---	---	7.75%	7.04%
		C=1	59.5	---	---	---	---	7.75%	4.23%
		C=10	36	---	4.23%	2.81%	---	12.67%	9.86%
		C=100	28.5	---	4.93%	4.93%	---	13.38%	12.67%
	$\delta^2=10$	C=0.1	70	---	0.70%	---	---	7.75%	7.04%
		C=1	65.5	---	0.70%	---	---	7.75%	2.82%
		C=10	39.5	---	3.52%	3.52%	---	12.68%	11.27%
		C=100	29.5	---	4.93%	4.93%	---	13.38%	12.67%
	$\delta^2=50$	C=0.1	69.5	---	0.70%	---	---	7.75%	7.75%
		C=1	69.5	---	0.70%	---	---	7.75%	7.75%
		C=10	60	---	0.70%	---	---	7.75%	7.75%
		C=100	31.5	---	4.93%	4.93%	---	15.49%	12.67%

--- = 0 (zero)

Tabela 19 – Método separação das classes duas a duas – Impedância – REGRAS

KERNEL		NSV	REGRAS		
			SO	OTIMIZ	
LINEAR	C=0.1	46.5	11.5	21.5	
	C=1	41	10.5	21.5	
	C=10	41	10.5	21.5	
	C=100	41	10.5	21.5	
RBF	$\delta^2=1$	C=0.1	69.5	13	25
		C=1	46	12	20.5
		C=10	29.5	11.5	20
		C=100	26	11	20
	$\delta^2=5$	C=0.1	70	13	24.5
		C=1	59.5	13	24
		C=10	36	10.5	19
		C=100	28.5	10	18
	$\delta^2=10$	C=0.1	70	13	24.5
		C=1	65.5	12.5	22
		C=10	39.5	11	18
		C=100	29.5	10	18
	$\delta^2=50$	C=0.1	69.5	13	25
		C=1	69.5	13	25
		C=10	60	12.5	25
		C=100	31.5	10	18

Tabela 20 – Método separação das classes duas a duas – Impedância – TAXAS DE ACERTO

SEPARAÇÃO DAS CLASSES DUAS A DUAS - 3 CONJUNTOS - TAXAS DE ACERTO									
KERNEL		NSV	TREINO			TESTE			
			SVM	SO	OTIMIZ	SVM	SO	OTIMIZ	
LINEAR	C=0.1	46.5	82.39%	64.09%	76.76%	81.70%	43.67%	57.05%	
	C=1	41	82.39%	61.98%	73.95%	81.70%	40.15%	54.93%	
	C=10	41	82.39%	61.98%	73.95%	81.70%	40.15%	54.93%	
	C=100	41	82.39%	61.98%	73.95%	81.70%	40.15%	54.93%	
RBF	$\delta^2=1$	C=0.1	69.5	80.29%	69.02%	80.99%	80.29%	47.18%	54.23%
		C=1	46	80.99%	67.61%	71.84%	80.29%	46.48%	52.82%
		C=10	29.5	82.40%	66.91%	71.13%	81.70%	46.48%	52.11%
		C=100	26	83.81%	66.20%	71.13%	81.70%	46.49%	57.05%
	$\delta^2=5$	C=0.1	70	80.29%	68.32%	83.83%	80.29%	47.18%	59.87%
		C=1	59.5	80.29%	69.02%	72.54%	80.29%	47.18%	52.82%
		C=10	36	83.09%	64.79%	71.14%	82.39%	42.26%	51.41%
		C=100	28.5	85.92%	64.09%	66.20%	82.39%	42.26%	46.49%
	$\delta^2=10$	C=0.1	70	80.29%	68.32%	83.83%	80.29%	47.18%	59.87%
		C=1	65.5	80.29%	68.32%	71.84%	80.29%	47.18%	64.09%
		C=10	39.5	81.70%	65.50%	67.61%	80.29%	42.25%	47.19%
		C=100	29.5	85.92%	64.09%	66.20%	83.09%	42.26%	46.49%
	$\delta^2=50$	C=0.1	69.5	80.29%	68.32%	83.10%	80.29%	47.18%	57.75%
		C=1	69.5	80.29%	68.32%	83.10%	80.29%	47.18%	57.75%
		C=10	60	80.29%	68.32%	83.10%	80.29%	47.18%	57.75%
		C=100	31.5	83.81%	64.09%	66.20%	80.29%	42.26%	46.49%

O melhor desempenho para os dados de teste em termos de pontos classificados foi obtido pelo método decomposição um por classe (Tabelas 13 e 14) para o kernel RBF com largura $\delta^2=1$ e constante de regularização $C=0,1$, apresentando uma taxa de pontos não classificados igual a 5.63%. Esta configuração apresentou em média 25.5 regras, 23 delas geradas com o primeiro conjunto para treinamento e o segundo para teste, e 28 invertendo-se os conjuntos.

Observando as Tabelas 16 a 20, que tratam das taxas de acerto (total de exemplos menos os pontos classificados incorretamente e não classificados), percebe-se a eficácia do método entre as configurações sem e com otimização dos conjuntos. Esta, em alguns casos, supera o desempenho da própria SVM.

Abaixo são descritas as regras para esta configuração, que apresentou o melhor resultado em se tratando dos dados de teste não classificados (kernel RBF com largura $\delta^2=1$ e constante de regularização $C=0,1$).

As regras para ambas as simulações foram unidas e suprimiram-se as que estavam presentes nos dois casos, resultando num total de 34, conforme apresentado abaixo.

1. Se a média da diferença entre as curvas for PEQUENA na faixa 1 e PEQUENA na faixa 2 e PEQUENA na faixa 3 e a maior diferença entre as curvas ocorrer na região de BAIXA frequência, então o transformador está **NORMAL**.
2. Se a média da diferença entre as curvas for MÉDIA na faixa 1 e MÉDIA na faixa 2 e PEQUENA na faixa 3 e a maior diferença entre as curvas ocorrer na região de BAIXA frequência, então o transformador está **NORMAL**.
3. Se a média da diferença entre as curvas for MÉDIA na faixa 1 e GRANDE na faixa 2 e MÉDIA na faixa 3 e a maior diferença entre as curvas ocorrer na região de BAIXA frequência, então o transformador está **NORMAL**.
4. Se a média da diferença entre as curvas for GRANDE na faixa 1 e MÉDIA na faixa 2 e MÉDIA na faixa 3 e a maior diferença entre as curvas ocorrer na região de BAIXA frequência, então o transformador está **NORMAL**.
5. Se a média da diferença entre as curvas for GRANDE na faixa 1 e GRANDE na faixa 2 e GRANDE na faixa 3 e a maior diferença entre as curvas ocorrer na região de BAIXA frequência, então o transformador está **NORMAL**.
6. Se a média da diferença entre as curvas for MÉDIA na faixa 1 e GRANDE na faixa 2 e MÉDIA na faixa 3 e a maior diferença entre as curvas ocorrer na região de ALTA frequência, então o transformador está **NORMAL**.
7. Se a média da diferença entre as curvas for PEQUENA na faixa 1 e MÉDIA na faixa 2 e PEQUENA na faixa 3 e a maior diferença entre as curvas ocorrer na região de BAIXA frequência, então o transformador está com **CURTO-CIRCUITO**.
8. Se a média da diferença entre as curvas for PEQUENA na faixa 1 e GRANDE na faixa 2 e MÉDIA na faixa 3 e a maior diferença entre as curvas ocorrer na região de BAIXA frequência, então o transformador está com **CURTO-CIRCUITO**.
9. Se a média da diferença entre as curvas for GRANDE na faixa 1 e GRANDE na faixa 2 e MÉDIA na faixa 3 e a maior diferença entre as curvas ocorrer na região de BAIXA frequência, então o transformador está com **CURTO-CIRCUITO**.
10. Se a média da diferença entre as curvas for PEQUENA na faixa 1 e GRANDE na faixa 2 e GRANDE na faixa 3 e a maior diferença entre as curvas ocorrer na região de BAIXA frequência, então o transformador está com **CURTO-CIRCUITO**.

11. Se a média da diferença entre as curvas for GRANDE na faixa 1 e PEQUENA na faixa 2 e PEQUENA na faixa 3 e a maior diferença entre as curvas ocorrer na região de BAIXA frequência, então o transformador está com **CURTO-CIRCUITO**.
12. Se a média da diferença entre as curvas for GRANDE na faixa 1 e GRANDE na faixa 2 e PEQUENA na faixa 3 e a maior diferença entre as curvas ocorrer na região de BAIXA frequência, então o transformador está com **CURTO-CIRCUITO**.
13. Se a média da diferença entre as curvas for MÉDIA na faixa 1 e PEQUENA na faixa 2 e MÉDIA na faixa 3 e a maior diferença entre as curvas ocorrer na região de BAIXA frequência, então o transformador está com **DESLOCAMENTO AXIAL**.
14. Se a média da diferença entre as curvas for MÉDIA na faixa 1 e MÉDIA na faixa 2 e MÉDIA na faixa 3 e a maior diferença entre as curvas ocorrer na região de BAIXA frequência, então o transformador está com **DESLOCAMENTO AXIAL**.
15. Se a média da diferença entre as curvas for MÉDIA na faixa 1 e GRANDE na faixa 2 e GRANDE na faixa 3 e a maior diferença entre as curvas ocorrer na região de BAIXA frequência, então o transformador está com **DESLOCAMENTO AXIAL**.
16. Se a média da diferença entre as curvas for GRANDE na faixa 1 e MÉDIA na faixa 2 e GRANDE na faixa 3 e a maior diferença entre as curvas ocorrer na região de BAIXA frequência, então o transformador está com **DESLOCAMENTO AXIAL**.
17. Se a média da diferença entre as curvas for GRANDE na faixa 1 e GRANDE na faixa 2 e PEQUENA na faixa 3 e a maior diferença entre as curvas ocorrer na região de MÉDIA frequência, então o transformador está com **DESLOCAMENTO AXIAL**.
18. Se a média da diferença entre as curvas for MÉDIA na faixa 1 e GRANDE na faixa 2 e MÉDIA na faixa 3 e a maior diferença entre as curvas ocorrer na região de MÉDIA frequência, então o transformador está com **DESLOCAMENTO AXIAL**.
19. Se a média da diferença entre as curvas for GRANDE na faixa 1 e GRANDE na faixa 2 e MÉDIA na faixa 3 e a maior diferença entre as curvas ocorrer na região de MÉDIA frequência, então o transformador está com **DESLOCAMENTO AXIAL**.

20. Se a média da diferença entre as curvas for GRANDE na faixa 1 e GRANDE na faixa 2 e GRANDE na faixa 3 e a maior diferença entre as curvas ocorrer na região de MÉDIA frequência, então o transformador está com **DESLOCAMENTO AXIAL**.
21. Se a média da diferença entre as curvas for GRANDE na faixa 1 e GRANDE na faixa 2 e PEQUENA na faixa 3 e a maior diferença entre as curvas ocorrer na região de ALTA frequência, então o transformador está com **DESLOCAMENTO AXIAL**.
22. Se a média da diferença entre as curvas for MÉDIA na faixa 1 e MÉDIA na faixa 2 e GRANDE na faixa 3 e a maior diferença entre as curvas ocorrer na região de ALTA frequência, então o transformador está com **DESLOCAMENTO AXIAL**.
23. Se a média da diferença entre as curvas for MÉDIA na faixa 1 e GRANDE na faixa 2 e GRANDE na faixa 3 e a maior diferença entre as curvas ocorrer na região de ALTA frequência, então o transformador está com **DESLOCAMENTO AXIAL**.
24. Se a média da diferença entre as curvas for GRANDE na faixa 1 e MÉDIA na faixa 2 e GRANDE na faixa 3 e a maior diferença entre as curvas ocorrer na região de ALTA frequência, então o transformador está com **DESLOCAMENTO AXIAL**.
25. Se a média da diferença entre as curvas for PEQUENA na faixa 1 e GRANDE na faixa 2 e PEQUENA na faixa 3 e a maior diferença entre as curvas ocorrer na região de BAIXA frequência, então o transformador está com **PERMISSIVIDADE DO ÓLEO ALTERADA**.
26. Se a média da diferença entre as curvas for MÉDIA na faixa 1 e PEQUENA na faixa 2 e PEQUENA na faixa 3 e a maior diferença entre as curvas ocorrer na região de BAIXA frequência, então o transformador está com **PERMISSIVIDADE DO ÓLEO ALTERADA**.
27. Se a média da diferença entre as curvas for MÉDIA na faixa 1 e GRANDE na faixa 2 e PEQUENA na faixa 3 e a maior diferença entre as curvas ocorrer na região de BAIXA frequência, então o transformador está com **PERMISSIVIDADE DO ÓLEO ALTERADA**.
28. Se a média da diferença entre as curvas for GRANDE na faixa 1 e MÉDIA na faixa 2 e PEQUENA na faixa 3 e a maior diferença entre as curvas ocorrer na região de BAIXA frequência, então o transformador está com **PERMISSIVIDADE DO ÓLEO ALTERADA**.

29. Se a média da diferença entre as curvas for PEQUENA na faixa 1 e PEQUENA na faixa 2 e MÉDIA na faixa 3 e a maior diferença entre as curvas ocorrer na região de BAIXA frequência, então o transformador está com **PERMISSIVIDADE DO ÓLEO ALTERADA**.
30. Se a média da diferença entre as curvas for PEQUENA na faixa 1 e MÉDIA na faixa 2 e MÉDIA na faixa 3 e a maior diferença entre as curvas ocorrer na região de BAIXA frequência, então o transformador está com **PERMISSIVIDADE DO ÓLEO ALTERADA**.
31. Se a média da diferença entre as curvas for PEQUENA na faixa 1 e PEQUENA na faixa 2 e GRANDE na faixa 3 e a maior diferença entre as curvas ocorrer na região de BAIXA frequência, então o transformador está com **PERMISSIVIDADE DO ÓLEO ALTERADA**.
32. Se a média da diferença entre as curvas for PEQUENA na faixa 1 e MÉDIA na faixa 2 e GRANDE na faixa 3 e a maior diferença entre as curvas ocorrer na região de BAIXA frequência, então o transformador está com **PERMISSIVIDADE DO ÓLEO ALTERADA**.
33. Se a média da diferença entre as curvas for MÉDIA na faixa 1 e PEQUENA na faixa 2 e GRANDE na faixa 3 e a maior diferença entre as curvas ocorrer na região de BAIXA frequência, então o transformador está com **PERMISSIVIDADE DO ÓLEO ALTERADA**.
34. Se a média da diferença entre as curvas for MÉDIA na faixa 1 e MÉDIA na faixa 2 e GRANDE na faixa 3 e a maior diferença entre as curvas ocorrer na região de BAIXA frequência, então o transformador está com **PERMISSIVIDADE DO ÓLEO ALTERADA**.

As regras acima, geradas automaticamente, são em sua grande maioria coerentes quando comparadas a diagnósticos realizados por especialistas.

Outra observação interessante é a diferença das regras geradas por dados reais e simulados. No caso dos dados de medições reais, em algumas regras, outras regiões (além de variar a região de defeito/ falha indicada por especialistas) também apresentaram informações relevantes, no caso dos dados simulados somente a região indicada por especialistas variou. A junção de dados reais e simulados no banco é interessante, uma vez que, medições reais podem carregar consigo interferências e ruídos que não fazem parte do sinal desejado.

As regras geradas pelo DIREF são inéditas na área de análise de resposta em frequência.

A seguir, as Figuras 44 e 45 apresentam a configuração dos conjuntos fuzzy antes e após a otimização dos conjuntos, para o caso do kernel RBF com largura $\delta^2=1$ e constante de regularização $C= 0.1$.

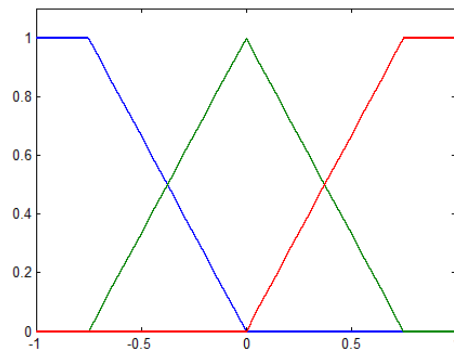


Figura 44 – Configuração Inicial para os conjuntos fuzzy das 4 variáveis de entrada – IMPEDÂNCIA

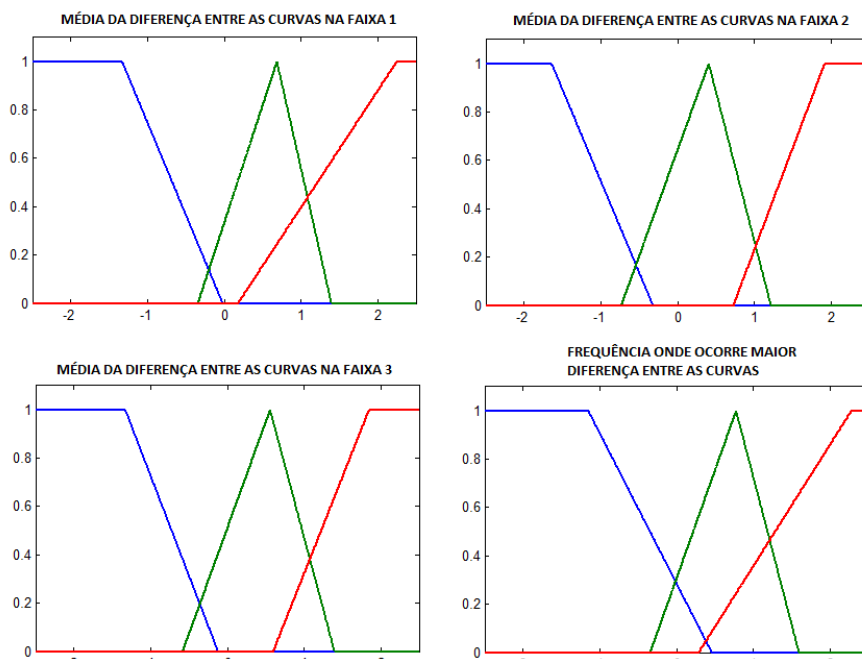


Figura 45 – Configuração após otimização de cada variável de entrada - IMPEDÂNCIA

6

Conclusões e Trabalhos Futuros

6.1

Conclusões

As Máquinas de Vetores Suporte têm obtido sucesso na solução de inúmeros problemas, em especial reconhecimento de padrões e problemas de classificação. O treinamento de SVMs é feito através da resolução de um problema quadrático com restrições lineares. Na solução deste problema são selecionados, entre todos os vetores de entrada, os vetores mais 'importantes', conhecidos como vetores suporte [6].

Como visto ao longo desse documento, o principal objetivo dessa dissertação foi estender o modelo FReX_SVM com a otimização dos conjuntos fuzzy, com o objetivo de melhorar os resultados apresentados em [6] e observar o teor e a interpretabilidade das regras linguísticas geradas. O modelo estendido foi aplicado ao banco de dados íris, conforme descrito no capítulo 4, apresentando um bom desempenho em relação ao modelo FReX_SVM original e, em alguns casos, superando os resultados da própria SVM.

No capítulo 5, foi descrita a aplicação do modelo FReX_SVM estendido a um banco de dados de análise de resposta em frequência, composto em parte por medições reais e também por dados resultantes de simulações. Esta aplicação deu origem a uma ferramenta de diagnóstico inteligente e automático a partir de dados de resposta em frequência. O diagnóstico atual é feito de forma subjetiva (visual e comparativa), depende de especialistas e não contempla todas as diferenças entre as curvas comparadas em questão (assinatura e a medição que se pretende analisar).

Os resultados deste trabalho mostram que é possível construir uma ferramenta de diagnóstico para um problema considerado por especialistas como de difícil automatização, devido às particularidades do procedimento de análise da resposta em frequência ao fato de não se ter um conhecimento difundido sobre a relação da alteração das curvas e a severidade e região do defeito. Ressalte-se, no entanto, que uma generalização das regras aqui

apresentadas para todos os tipos de transformadores existentes no mercado ainda não é possível, já que a base de dados disponível era pequena.

6.2

Trabalhos Futuros

Pode-se ainda refinar a ferramenta desenvolvida, variando-se, além dos suportes das funções de pertinência, o número de conjuntos associados a cada variável de entrada.

Quanto ao DIREF, pode-se estender a análise para outros bancos de dados além de medições de impedância terminal, tais como: medições de transferência de tensão, utilizar além das amplitudes os ângulos da medição de impedância e transferência de tensão. Alternativa interessante seria associar o diagnóstico de resposta em frequência a outras técnicas como: cromatografia de gases, emissão acústica. Pode-se ainda, utilizar outras métricas além da diferença entre as curvas, tais como: correlação, desvio padrão etc.

Uma base de dados reais maior, obtida via parcerias com fabricante de transformadores possibilitaria realizar testes mais completos.

Outra opção seria desenvolver um protótipo para simulação em ambiente controlado (laboratório), onde pudesse ser gerada uma extensiva base de dados com possíveis defeitos na tentativa de construir um banco de dados mais robusto e confiável. Com este protótipo seria possível, por exemplo, variar o torque das bobinas de diversas maneiras e observar a alteração nas curvas de medição e sua intensidade, possibilitando uma relação entre a curva de medição e os componentes internos do equipamento em análise. Ao observar a relação entre a curva de medição e o parâmetro físico do transformador, seria possível reduzir as desvantagens da utilização do modelo de parâmetros concentrados.

Referências bibliográficas

- 1 MARTINS, H.J.A. **Diagnóstico de Transformadores de Potência Através de Metodologias no Domínio da Frequência**. Rio de Janeiro, 2007. 224p. Tese de Doutorado - COPPE / UFRJ.
- 2 MENDES, J.C. **Redução de Falhas em Grandes Transformadores de Alta Tensão**. São Paulo, 1995. Tese de Doutorado - Escola Politécnica da Universidade de São Paulo.
- 3 **IEEE FRA Task Force C.57.149**. Transformer Frequency Analysis, 2005.
- 4 **CIGRE WG 33.02** – “Guidelines for representation of network elements when calculating transients”, 1990.
- 5 Sofian, D.M. **Transformers FRA Interpretations for Detection of Winding Movement**. Manchester University, July 2007.
- 6 CHAVES, A.C.F. **Extração de Regras Fuzzy para Máquinas de Vetores Suporte (SVM) para Classificação em Múltiplas Classes**. Rio de Janeiro, 2006. 225p. Tese de Doutorado – Departamento de Engenharia Elétrica, PUC - Rio.
- 7 LANAS, A.I. **Sistemas Neuro-Fuzzy Hierárquicos BSP para previsão e extração de Regras Fuzzy em Aplicações de Mineração de Dados**. Tese de Mestrado – Departamento de Engenharia Elétrica, PUC-Rio.
- 8 SCHÖLKOPF, B. & SMOLA, A.J. **Learning with Kernels**. The MIT Press, 2002.
- 9 HSU, C.-W. & LIN, C.-J. A Comparison on Methods Multi-class Support Vector Machines. In **IEEE Transactions on neural Networks**, vol. 13(2), 2002, pp. 415-425.
- 10 WESTON, J. & WATKINS, C. **Multi-class Support Vector Machines**. Technical report CSD-TR-98-04, Royal Holloway, 1998.
- 11 ABE, S. & INOUE, T. **Fuzzy Support Vector Machines for Multiclass Problems**. In ESANN proceedings, pp 113-118, 2002.
- 12 KRESSEL, U., H.-G. Pairwise classification and support vectors machines. In B.Schölkopf, C.J.C. Burges, & A.J. Smola (Eds), **Advances in kernel methods: Support vector learning**, Cambridge, MA: MIT Press, 1999, pp. 225-268.
- 13 RIFKIN, R.; KLAUTAU, A. In Defense of One-Vs-All Classification. **Journal of Machine Learning Research**. Vol. 5, 2004, pp 101-141.

- 14 CRISTIANINI, N.; SHAW-TAYLOR, J. **An Introduction to Support Vector Machines (and other kernel-based learning methods)**. Cambridge University Press, 2000.
- 15 VAPNIK, V.N. An Overview of Statistical Learning Theory. In **IEEE Trans. On Neural Networks**, vol.10 (5), pp. 988-999, 1999.
- 16 VAPNIK, V.N. **Statistical Learning Theory**. John Wiley & Sons, 1998.
- 17 CORTES, C.; VAPNIK, V. Support-vector networks. **Machine Learning**, 20 (3), 1995, pp. 273-297.
- 18 CHERKASSKY, V. e MULIER F. **Learning from data: concepts, theory, and methods**. John Wiley & Sons, Inc.1998.
- 19 LIMA, C.A.M. **Comitê de Máquinas: Uma Abordagem Unificada Empregando Máquina de Vetores Suporte**. Campinas, 2004. 342p. Tese de Doutorado – Faculdade de Engenharia Elétrica e de Computação, Universidade Estadual de Campinas.
- 20 Manual de Utilização do GACOM. PUC-Rio. Rio de Janeiro
- 21 CORTES, O.A.C. **Integração entre Lógica Nebulosa e Algoritmos Evolutivos**. Departamento Acadêmico de Informática, Centro Federal de Educação Tecnológica do Maranhão – CEFET.
- 22 PACHECO, M.A. **Algoritmos Genéticos: Princípios e Aplicações**. PUC - Rio. Rio de Janeiro, 1999.
- 23 MENDEL, J.M. **Fuzzy Logic Systems for Engineering: A tutorial**. 1995.
- 24 COX, E. **The Fuzzy Systems Handbook: A Practitioner's Guide to Building, Using, and Maintaining Fuzzy Systems**. AP Professional, 1994.
- 25 KLIR, G.J.; YUAN, B. **Fuzzy Sets and Fuzzy Logic – Theory and Applications**. Prentice Hall PTR, 1995.
- 26 ZIMMERMANN, H.-J. **Fuzzy Set Theory and Its Applications**. Kluwer Academic Publishers, 1994.
- 27 TANSCHKEIT, R. **Sistema Fuzzy**. PUC - Rio. Rio de Janeiro.
- 28 DAVIS, L. **Handbook of Genetic Algorithms**. VNR Comp. Library, 1990.
- 29 GOLDBERG, D. **Genetic Algorithms in Search, Optimization and Machine Learning**. Addison-Wesley, 1999.
- 30 MICHALEWICZ, Z. **Algorithms+Data Structures=Evolution Programs**. Springer- Verlag, 1994.
- 31 KOZA, J. **Genetic Programming: On the Programming of Computers by Means of Natural Selection**. MIT Press, 1992.

- 32 VAPNIK, V.N.; LERNER, A. **Pattern recognition using generalized portrait method. Automation and Remote Control.** Vol. 24, 1963.
- 33 SCHÖLKOPF, B.; BURGESS, C.J.; SMOLA, A.J. **Advances in Kernel Methods – Support Vector Machines Learning.** The MIT Press, Cambridge, MA, 1999.
- 34 HAYKIN, S. **Neural Networks A Comprehensive Foundation.** Macmillan College Publishing Company, 1999.
- 35 HSU, C.-W., CHANG, C.-C & LIN, C.-J. **A practical guide to SVM classification.** Technical Report, Department of Computer Science and Information Technology, National Taiwan University, 2003. Disponível em <http://www.scie.ntu.edu.tw/~cjlin/papers/guide.pdf>
- 36 JOACHIMS, T. Text categorization with support vector machines: Learning with many relevant features. In **Proceedings of the European Conference on Machine Learning**, Springer, 1998, pp.137-142.
- 37 DRUCKER, H.; WU, D.H.; VAPNIK, V.N. Support vector machines for spam categorization. In **IEEE Trans. On Neural Networks.** Vol. 10(5), 1999, pp.1048-1054.
- 38 GUNN, S. **Support Vector Machines for Classification and Regression.** ISIS Technical Report, 1998. <http://www.isis.ecs.soton.ac.uk/research/svm/>
- 39 ALLWEIN, E.L.; SCHAPIRE R.E. & SINGER, Y. Reducing Multiclass to Binary: A Unifying Approach for Margin Classifiers. In **Journal of Machine Learning Research**, Vol. 1(2), 2000, pp. 113-141.
- 40 ANGULO, C.; PARRA, X. & CATALÀ, A. **An Unified Frame work for ‘All data at once’ Multi-Class Support Vector Machines.** In ESANN, 2002, pp. 161-166.
- 41 CRAMMER, K. & SINGER, Y. On the Algorithmic Implementation Multiclass Kernel-based Vector Machines. In **Journal of Machine Learning Research**, n. 2, 2001, pp. 265-292.
- 42 KLAUTAU, A.; JEVTIC', N. & ORLISTSKY, A. On Nearest – Neighbor Error – Correcting Output Code with Application to All-Pairs Multiclass Support Vector Machines. In **Journal of Machine Learning Research**, n.4, 2003, pp.1-15.
- 43 TSUJINISHI, D. & ABE, S. Fuzzy least square support vector machines for multiclass problems. In **Neural Networks**, Vol. 16, 2003, pp.785-792.
- 44 ANGUITA, D.; RIDELLA, S. e STERPI, D. A New Method for Multiclass Support Vector Machines. **Proc of the IEEE Int. Joint Conf, on Neural Networks**, (IJCMNN'2004), Budapest, Hungary, July 2004.
- 45 FRANC, V.; HLAVAC, V. Multi-class Support Vector Machine. In **Proc. Of the 16th IEEE Int. Conf. on Pattern Recognition**, 2002, pp. 236-239.

- 46 SHAWE – TAYLOR, J. & CRISTIANINI, N. Margin Distribution and Soft Margin, in **Advances in Large Margin Classifiers**, The MIT Press, 2000.
- 47 SUYKENS, J.A. K. & VANDERWALLE, J. Multiclass Least Squares Support Vector Machines. In **Proc. Of the International Joint Conference on Neural Networks (IJCNN'99)**, Washington DC, USA, CDROM. July, 1999.
- 48 SUYKENS, J.A.K. & VANDERWALLE, J. Least squares support vector machine classifiers. In **Neural Processing Letters**, Vol. 9(3), pp. 293-300.
- 49 CRAMMER, K. & SINGER, Y. On the learnability and desing of output code for multiclass problems. In **Computational Learning Theory**, 2000, pp. 35-46.
- 50 FU, X. et al. Extracting the Knowledge Embedded in Support Vector Machines. In **International Joint Conference on Neural Networks (IJCNN'04)**, CDROM, Budapest, 2004, July 25-29.
- 51 NÚÑEZ, H.; ANGULO, C.; CATALÀ, A. **Rule Extraction from Support Vectors Machines**, In ESANN, 2002, pp. 107-112.
- 52 NÚÑEZ, H.; ANGULO, C.; CATALÀ, A. **Hybrid Architecture Based on Support Vectors Machines**, In IWANN, 2003, pp. 646-653.
- 53 GONÇALVES, L.; VELLASCO, M.; PACHECO, M.A.; DE SOUZA, F. Inverted hierarchical neuro-fuzzy BSP system: a novel neuro-fuzzy model for pattern classification and rule extraction in databases. **IEEE Transact Sist Man Cybern Part C. Appl Ver 36 -2**; 236-248, 2006.
- 54 CHAVES, A.; VELLASCO, M. & TANSCHKEIT, R. **Fuzzy rules extraction from support vector machines from multi-class classification**. Neural Comput & Applic, 2012.
- 55 BRITO, V.B. **FuzzyFuture; Ferramenta de previsão de séries temporais baseada em sistema híbrido Fuzzy-Genético**. Rio de Janeiro, 2011. 151p. Dissertação de Mestrado – Departamento de Engenharia Elétrica, PUC-Rio.
- 56 BRITO, V.; VELLASCO, M. & TANSCHKEIT, R. **FuzzyFuture: Ferramenta de Previsão de Séries Temporais Baseada em Sistema Híbrido Fuzzy-Genético**. X Congresso Brasileiro de Inteligência Computacional, 2011.
- 57 UCI Machine Learning Repository (<http://www.ics.edu/~mlearn/MLREpository.html>)
- 58 LAPWORTH, J.A. **Mechanical Condition Assessment of Power Transformers Using Frequency Response Analysis**. Doble International Conference, 1995.
- 59 KOSOW, I. **Máquinas Elétricas e Transformadores**. Globo, 1996.
- 60 VASQUES, C.M.R. **Automação do Sistema de Medição em Transformadores de Potência Através da Técnica de Resposta em**

- Frequência.** Monografia - Automação Industrial dos Sistemas de Produção, Refino e Transporte de Petróleo. SENAI, 2011.
- 61 AZEVEDO, R.M. **Elaboração de Modelo para Representação da Impedância Terminal do Autotransformador 550/ $\sqrt{3}$ / 246,83/ $\sqrt{3}$ – 14,81 kV da SE Vila do Conde.** NOTA TÉCNICA CEPEL 2012
- 62 ZAMBRANO, G. **Identificação de Parâmetros de um Circuito Equivalente de Transformadores de Potência.** Rio de Janeiro, 2009. Tese de Doutorado – COPPE, UFRJ.
- 63 TENBOHLEN S.; RYDER S. A. **Making Frequency Response Analysis Measurements: A Comparison of the Swept Frequency and Low Voltage Impulse Methods,** XIII ISH, Netherlands 2003.
- 64 SYED, M.S. **Detection of Shorted Turns and Winding Movements in Large Power Transformers Using Frequency Response Analysis.** IEEE Transactions on Power Delivery, v.3, pp. 23-27, January.
- 65 CONTIN, A.; RABACH, G.; BORGHETTO, J.; NIGRIS, M.; PASSAGLIA, R. & RIZZI, G. **Frequency-responses Analysis of Power Transformers by Means of Fuzzy Tools.** IEEE Transactions on Dielectrics and Electrical Insulation, vol. 18, n. 3, June 2011.
- 66 COMISSÃO NACIONAL DE DESENVOLVIMENTO E REFORMA DA REPÚBLICA POPULAR DA CHINA. **Análise de Resposta em Frequência de Deformação em Enrolamentos do Transformador de Potência.** In: Norma do Setor de Energia Elétrica da República Popular da China – DL/T 911-2004, China