

Seleção de Parâmetros para Classificação de Falhas do Tipo Curto-Circuito em Linhas de Transmissão

Jefferson Morais¹, Yomara Pires², Fernando dos Santos², Waldemar Neto², Aldebaro Klautau¹

¹Programa de Pós-Graduação em Ciência da Computação
Universidade Federal do Pará (UFPA)
Caixa Postal 479 – 66.075-110 – Belém – PA – Brasil

²Faculdade de Computação– Universidade Federal do Pará (UFPA)
Caixa Postal 479 – 68.746-360 – Castanhal – PA – Brasil

{jmorais, yomara, aldebaro}@ufpa.br, fernandojj2@gmail.com, netolandy@hotmail.com

Abstract. *This paper proposes a methodology to feature selection for fault classification in transmission lines. Since most work on the classification of events in power systems is concerned with an event classified according to morphology of the corresponding waveform. An important and yet most difficult problem is the classification of the cause of the event. Thus, this paper uses different techniques for pre-processing (front-ends), feature selection algorithms and machine learning techniques to classification of short-circuit faults.*

Resumo. *Este trabalho propõe uma metodologia para seleção de parâmetros para classificação de faltas do tipo curto-circuito em linhas de transmissão. Uma vez que a maioria dos trabalhos em classificação de eventos em sistemas de potência preocupa-se em classificar um evento de acordo com a morfologia da forma de onda correspondente. Um problema importante e ainda mais difícil é a classificação da causa do evento. Para tal, este artigo utiliza diferentes técnicas de pré-processamento (front-ends), seleção de parâmetros e técnicas de aprendizado de máquina para classificação de faltas do tipo curto-circuito.*

1. Introdução

A crescente exigência dos consumidores de energia elétrica acompanhado ao aumento da produção de equipamentos eletrônicos extremamente sensíveis a distúrbios e oscilações elétricas, aliado à interligação dos sistemas elétricos existentes, assim como, à privatização de concessionárias de energia são fatores que influenciam diretamente em requisitos de Qualidade de Energia Elétrica QEE ¹ cada vez mais rigorosos.

Buscando atender a tais requisitos, o setor elétrico dispõe de tecnologias avançadas para a aquisição e armazenamento de informações, as quais inclui os chamados “distúrbios” ou “eventos de interesse” [Bollen et al. 2009]. Um típico exemplo dessa tecnologia são os equipamentos de oscilografia [Luo and Kezunovic 2005] que buscam, com auxílio de algoritmos relativamente simples, anomalias na forma de onda de tensão, assim, se no momento da análise da forma de onda a variação for maior que um determinado valor predefinido, um circuito programável denominado de *trigger* (gatilho),

¹Segundo [Bollen et al. 2009], é qualquer distúrbio ou alteração nos valores de tensão, corrente ou desvio da frequência que resulte em falta ou má operação dos equipamentos dos consumidores

inicia o armazenamento dos dados além de informações adicionais como a data e a hora do ocorrido.

Apesar da tecnologia avançada, o setor elétrico ainda enfrenta muitas dificuldades para se beneficiar plenamente do processamento automático da informação, tais como algoritmos de mineração de dados. Uma das razões é a falta de base de dados padronizados e livremente disponíveis.

Classificar um determinado evento, ou distúrbio, de acordo com sua morfologia da forma de onda baseando-se, para isso, em parâmetros (tais como *Root Mean Square-RMS*) é um exercício que não apresenta grandes desafios para seus analistas. Uma vez que existem várias normas QEE que classificam os fenômenos de acordo com certas características específicas. Assim, um problema ainda mais desafiador e interessante é a classificação das causas reais deste evento, isto é, se o que provocou o evento de QEE (e.x, afundamento ou elevação de tensão) foi uma descarga atmosférica, um curto-circuito ou um chaveamento de banco de capacitores, por exemplo.

Com base nesse contexto, este trabalho tem como objetivo apresentar uma metodologia para a classificação das causas de eventos de QEE, utilizando para isso técnicas de processamento de sinais e Inteligência Artificial (IA). Para fornecer um exemplo concreto da metodologia proposta, o trabalho concentra-se na análise de faltas do tipo curto-circuito em linhas de transmissão. Esta é uma particular e importante classe de causas de eventos: estudos mostraram que essas faltas foram responsáveis por cerca de 70% dos distúrbios e *blackouts* ocorridos nos sistemas elétricos de potência (SEPs) [Bollen et al. 2009]. Daí a importância do estudo dessa classe de causas de eventos.

2. Classificação de Eventos em Sistemas de Potência

Devido a digitalização dos SEPs, uma crescente quantidade de dados se faz disponível a análises, dos quais boa parte são gráficos com informações referentes as formas de onda (tensão ou corrente) diretamente ligada a eventos de QEE. A classificação destes eventos, seguindo a morfologia de sua forma de onda, é de certa forma, uma tarefa simples. Contudo, uma problemática ainda mais relevante aos especialistas na área é a classificação das causas do distúrbio (evento).

Neste sentido, a distinção entre a “Classificação baseada na morfologia” e a “Classificação da causa do evento” é de extrema importância. Uma vez que a primeira classifica o desvio dos valores nominais da forma de onda de tensão ou corrente e a segunda, as causas reais dos desvios destes valores. A seguir serão caracterizados estes dois tipos de classificação de eventos.

2.1. Classificação de eventos baseada em morfologia

Um Sistema Elétrico de potência típico apresenta três zonas funcionais: geração, transmissão e distribuição. Estas zonas estão sujeitas a distúrbios naturais ou provocados pelo homem que resultam em perturbações que causam variações na forma de onda (morfologia do sinal), caracterizando um evento de QEE.

Em situações normais, a forma de onda de tensão $x(t)$ possui uma frequência específica que varia entre os valores de 50 ou 60 Hz. Uma vez sabido o valor nominal da amplitude de onda, por exemplo, 500 kV, é conveniente normalizar $x(t)$ por este valor e reportar a amplitude em p.u. (por unidade). Eventos de QEE como, afundamento de tensão,

elevação de tensão e interrupção são definidos em normas internacionais de medições de QEE [IEC 2002] permitindo, assim, a caracterização destes eventos por meio de seus valores de magnitude da forma de onda de tensão (através do RMS, por exemplo) e no tempo de duração do evento.

A classificação automática desses eventos é através de regras de classificação do tipo se-então. Ex: Se $RMS \leq 0:01$ p.u. e Duração < 1 minuto então o evento é uma interrupção. Em contrapartida, a classificação das causas do evento, apesar de ser mais importante, é menos explorada pela comunidade científica.

2.2. Classificação das causas dos eventos

A classificação das causas dos eventos, por sua vez, é uma tarefa bem mais complexa. Consideremos, por exemplo, que para se diferenciar os eventos causados por queimadas em linhas de transmissão dos eventos causados por descargas elétricas provenientes de raios, devem-se considerar suas “assinaturas” nas formas de onda de tensão e de corrente, visto que se assume, na maior parte dos casos, que não há sensores especiais posicionados ao longo dos sistemas de transmissão, para a captação das características específicas. Outro exemplo mais complexo consiste em distinguir quedas de árvores versus queimadas, caso ambos provoquem um curto na linha de transmissão. Assim, ao escolher as classes, deve-se levar em conta que as causas a serem distinguidas devem imprimir distintas assinaturas nas formas de onda de tensão e corrente. Caso contrário, uma classificação adequada será impraticável.

A classificação de faltas do tipo curto-circuito em linhas de transmissão é um problema de classificação de causa, pois pretende-se descobrir quais as fases estão envolvidas na geração do curto-circuito. Em um contexto mais amplo, poderia-se projetar um sistema de classificação de causas onde “curto-circuito em linha de transmissão” fosse uma classe, e de maneira hierárquica disparar-se o tipo de classificador discutido nesse trabalho para identificar quais as fases envolvidos no curto-circuito detectado pelo módulo superior.

3. Metodologia

A metodologia adotada neste trabalho utiliza a base de dados *UFPAFaults* [Morais et al. 2010]. *UFPAFaults* é uma base de dados pública e devidamente rotulada, desenvolvida no simulador *Alternative Program Program (ATP)* [EMTP 1995]. Modelos ATP têm uma longa história de boa reputação e são muito utilizados para descrever de forma concisa o comportamento atual dos SEPs. Assim, os dados gerados artificialmente para compor a base *UFPAFaults* podem ser (certamente) usados nos experimentos adotados neste trabalho.

A metodologia é dividida em duas etapas. Na primeira etapa, simulações experimentais avaliam a performance de vários classificadores de sequência. *Front ends* baseados em Wavelets são comparados com outros mais simples, baseados, por exemplo, nos valores RMS. Esses *front ends* são combinados com diversos algoritmos de aprendizagem: redes neurais artificiais (ANN), árvores de decisão (J4.8), máquinas de vetores de suporte (*Support Vector Machine - SVM*) e K-vizinho mais próximo [Witten and Frank 2005], sugerindo a adoção da arquitetura de classificação de eventos em sistema de potência denominada de Classificação de Sequência Baseada em Quadros (*Frame-Based Sequence Classification - FBSC*) definida em [Morais et al. 2010].

Na segunda etapa os parâmetros dos *Front ends* são concatenados, passando por um processo de seleção de parâmetros (*feature selection*) antes de serem submetidos aos algoritmos de aprendizado de máquina disponíveis no WEKA ² [Witten and Frank 2005] sobre a base de dados de interesse *UFPAFaults*. Nesta etapa, técnicas de seleção de parâmetros, tais como, filtros [Hall 1999] são investigadas.

4. *Front ends*

Paralelamente a forma de representação que fora adotada, uma simples amostra de dados geralmente não carrega informações suficientemente que permitam o auxílio em tomadas de decisões. Logo, um *Front end* se faz necessário para converter as amostras em parâmetros específicos gerando, assim, as sequências que serão passadas aos algoritmos de classificação.

Todavia, a escolha de um *Front end* específico, o qual apresente um melhor desempenho no processo de classificação de faltas, não é uma tarefa simples. Assim, este trabalho adotou a estratégia que consiste em concatenar as amostras dos *Front ends* em um único conjunto de sequências. A seguir serão descritos os *Front ends* adotados neste.

4.1. *Front end raw*

Um *front end* é chamado *raw* quando seus parâmetros de saída correspondem a valores das amostras originais, sem qualquer outro processamento que organize as amostras para uma matriz \mathbf{Z} de acordo com os valores escolhidos de L (número de amostras do quadro) e S (deslocamento do quadro) [Morais et al. 2010]. Por exemplo, consideremos a segmentação de uma falta (ABT) em vetores de parâmetros \mathbf{z} com 4 quadros. Neste exemplo, $L = 3$ e $S = 1$ o que fornece dois vetores \mathbf{z} , cada um de dimensão $K = 18$.

4.2. *Front end RMS*

O *Front end RMS* é o mais utilizado permitindo obter uma estimativa aproximada da amplitude da frequência fundamental de uma forma de onda. Este consiste em calcular o valor RMS janelado para cada uma das Q formas de onda. O cálculo do valor do RMS consiste no janelamento de cada uma das formas de onda Q . Considerando o tamanho da janela L e o deslocamento S , o n -ésimo valor RMS $z[n]$, $n = 1, \dots, N$ de uma forma de onda $x[t]$, $t = 1, \dots, T$ é dado por

$$z[n] = \sqrt{\frac{1}{L} \sum_{l=1}^L (x[l + (n-1)S])^2} \quad (1)$$

4.3. *Front end wavelet*

Quando se adota um *Front end* de multi-resolução tais como wavelets, cuidados especiais devem ser tomados quanto ao seu processamento, devido ao grande número de possibilidades de implementação deste. Caso contrário, a replicação dos resultados pode ser inviável.

²Software de mineração de dados de domínio público desenvolvido na Universidade de Waikato na Nova Zelândia. Disponível em <http://www.cs.waikato.ac.nz/ml/weka/>

Neste trabalho, adota-se dois tipos de *front end* wavelet: waveletconcat que concatena todos os coeficientes wavelet e organiza-os em uma matriz; e waveletenergy que calcula a energia média de cada coeficiente [Morais et al. 2010].

Determinar qual *Front end* se adequa melhor ao processo de classificação de faltas não é uma tarefa simples. Visto isso, a estratégia mais adequada seria concatenar as amostras providas pelos *Front ends* em um único conjunto de sequências.

5. Seleção de Parâmetros

Intuitivamente, quanto mais informações disponíveis ao classificador (parâmetros), melhor seu desempenho. Contudo, isto gera a maldição da dimensionalidade. A medida que cresce o número de parâmetros os dados tendem a ficar exponencialmente esparsos e com isso as amostras aparentam estar igualmente distantes, diminuindo, então, a capacidade de generalização dos classificadores.

A redução da dimensionalidade pode ser realizada, de um modo geral, através de duas abordagens distintas: extração de parâmetros (*feature extraction*), que busca gerar parâmetros mais relevantes ao classificador a partir dos parâmetros existentes. Como exemplo, temos Análise de Componentes Principais (*Principal Components Analysis*) - PCA [Castelli and Bergman 2002] e, a seleção de parâmetros (*feature selection*) que consiste em identificar e reter apenas os parâmetros, dentre o total de subconjuntos possíveis, que mais contribuem para execução de uma dada tarefa. Neste trabalho, adotaremos a abordagem *Feature selection* para redução dos parâmetros.

Os métodos de seleção de parâmetros podem ser classificados basicamente em dois grupos: filtros e *wrappers* [Hall 1999]. Os métodos do tipo filtro avaliam um dado parâmetro através do uso de heurísticas baseadas nas propriedades estatísticas dos dados sendo este independente do classificador adotado. Como consequência, métodos do tipo filtro são geralmente mais rápidos e mais práticos do que os métodos *wrappers* principalmente quando se tem um conjunto de dados de alta dimensionalidade.

6. Experimentos e Resultados

Esta seção descreve os experimentos e os resultados obtidos para a classificação de eventos em sistemas de potência usando a arquitetura FBSC e a base de dados *UFPA-Faults* [Morais et al. 2010]. É assumido ao longo das simulações o uso de um gatilho (*trigger*) [Englert and Stenzel 2002], que descarta as amostras quando a forma de onda não apresenta qualquer tipo de anomalia, e passa aos estágios subsequentes somente o seguimento onde a falta foi detectada.

Foram utilizados os *Front ends* wavelet (waveletconcat, waveletenergy), raw e RMS. A wavelet mãe utilizada foi *Daubechies* 4 com $\gamma = 3$ níveis de decomposição. Consequentemente, para cada uma das $Q = 6$ formas de onda, a decomposição wavelet gerou quatro formas de onda. Os valores de L_{min} e S_{min} foram dentre os seguintes pares: a) $L_{min} = 4$ e $S_{min} = 2$, b) $L_{min} = 9$ e $S_{min} = 4$ e c) $L_{min} = 5$ e $S_{min} = 5$. A escolha por tais valores basearam-se em experimentos anteriores.

Com relação aos *front ends* raw e RMS os valores de L e S foram os mesmos daqueles adotados para os *front ends* wavelet. São apresentados também resultados após a concatenação dos *front ends* e aplicação da seleção de parâmetros baseada em filtros. Os

algoritmos de classificação utilizados foram ANN, J4.8, KNN e SVM, todos pertencentes ao pacote de mineração de dados WEKA.

6.1. Resultados para classificação de seqüências

Esta seção apresenta os resultados obtidos para a classificação de seqüências E_s usando a arquitetura FBSC e os classificadores ANN, J4.8, KNN e SVM. Foram realizadas diversas simulações em um cenário em que os classificadores foram treinados e testados em sinais livres de ruído. Dos 52 sistemas testados, 17 apresentaram 0% de erro.

As seguintes combinações de *Front ends* e classificadores obtiveram taxa de erro $E_s = 0$ para os três pares de L_{min} e S_{min} : waveletconcat e ANN, waveletconcat e J4.8 e raw com J4.8. Também obtiveram taxas de erros $E_s = 0$ para os valores específicos de $L_{min} = 9$ e $S_{min} = 4$: ANN e waveletenergy, ANN e raw, SVM e waveletconcat, SVM e raw, SVM e RMS. As combinações: ANN e concatfrontend, J.48 e concatfrontend, KNN e raw (para $L_{min} = 4$ e $S_{min} = 2$) também atingiram taxas de erros $E_s = 0$.

Em um cenário com ruído os classificadores apresentaram desempenho variado de acordo com o *Front end* adotado. Os resultados da Figura 1 demonstram que mesmo em um cenário com ruído todos os pares de classificadores, exceto KNN, tiveram uma precisão que permitiram ao classificador de seqüencia FBSC alcançar $E_s = 0\%$.

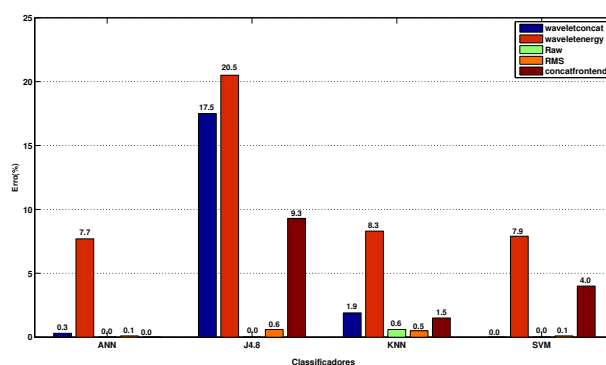


Figura 1. Taxa de erro E_s para a classificação de seqüências baseada em quadros usando os *Front ends* waveletconcat, waveletenergy, raw, RMS e a concatenação destes com $L = L_{min} = 9$ e $S = S_{min} = 4$. O conjunto de treinamento foi sem ruído e o de teste teve uma razão de sinal de ruído de 30 dB.

6.2. Resultados com seleção de parâmetros

Para os classificadores ANN e J4.8, com a seleção de 200 parâmetros mais relevantes foi possível obter o mesmo desempenho em relação ao total de parâmetros. Quanto a aplicação da heurística de avaliação, no caso do classificador ANN as menores taxas de erros são obtidas considerando o ganho de informação enquanto que para o J4.8 observa-se um equilíbrio nas taxas de erros obtidas.

Em relação aos classificadores KNN e SVM chama a atenção que com apenas 10 dos 894 parâmetros é possível obter taxas de erros mais baixas. Os resultados descritos são mostrados na Figura 2.

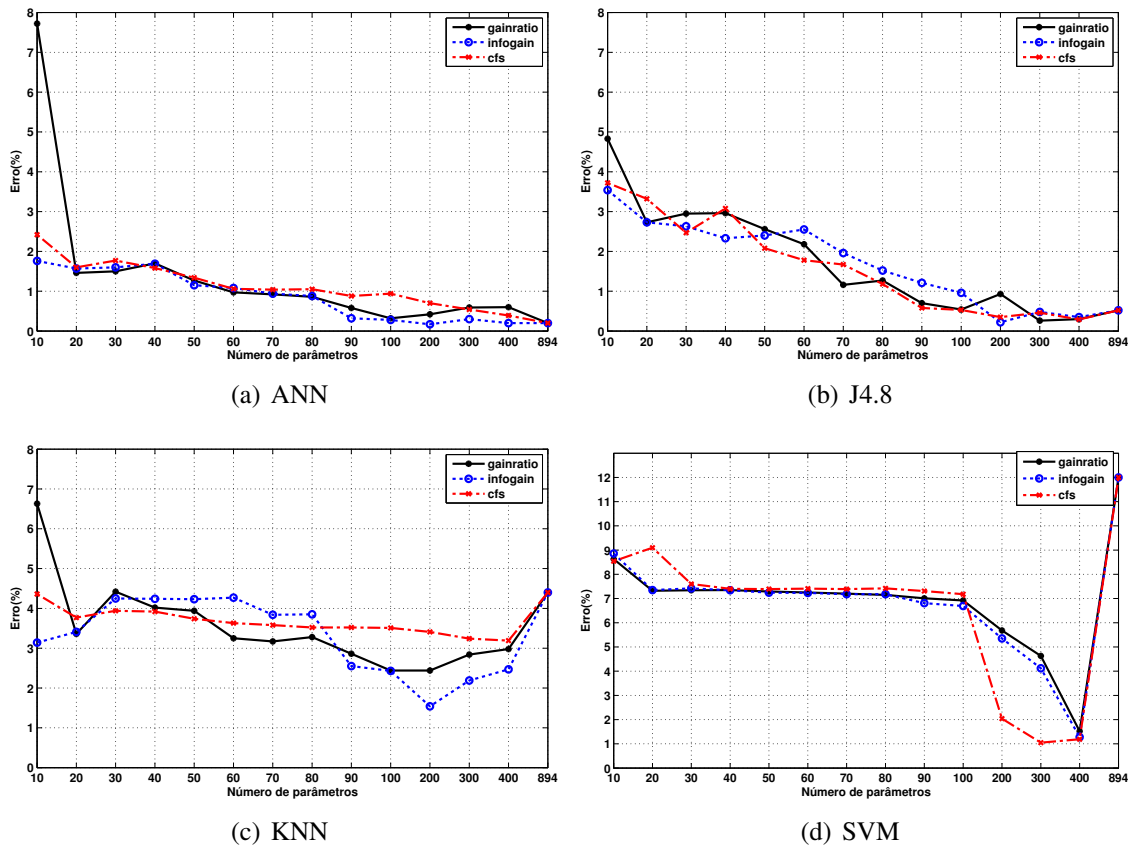


Figura 2. Taxa de erro E_f aplicando seleção de parâmetros nos *front ends* concatenados. Foi utilizado o método de seleção de parâmetros Filtro baseado nas heurísticas de avaliação ganho de informação (*infogain*), razão do ganho (*gainratio*) e correlação (*cfs*) com o número de parâmetros selecionados variando em 10, 20, 30, ..., 100, 200, 300 e 400.

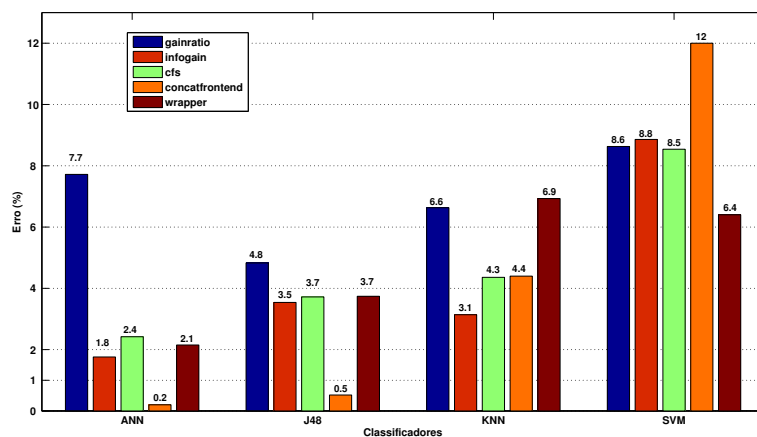


Figura 3. Taxa de erro E_f considerando a concatenação dos *front ends* e aplicando-se as técnicas de seleção de parâmetros filtros e *wrappers* fixando o número de parâmetros em 10.

Observando-se a Figura 3 pode-se inferir que os métodos do tipo filtro possuem desempenho equivalente ou superior em relação a seleção de parâmetros utilizando wrap-

per, exceto para o classificador SVM que apresenta seu melhor desempenho utilizando wrapper.

7. Conclusões

Este trabalho apresentou uma descrição dos aspectos relacionados ao projeto de módulos de classificação de eventos em sistemas de potência. Os resultados mostraram várias relações de compromisso no projeto de um classificador de sequências. A arquitetura FBSC, por exemplo, apresentou um excelente desempenho, visto que as taxas de erro atingiram valores próximos ou iguais a zero mesmo em um cenário com ruído. Outro apontamento, indica que em relação aos classificadores KNN e SVM chama a atenção que com apenas 10 dos 894 parâmetros é possível obter taxas de erros mais baixas. Já os classificadores ANN e J4.8, com a seleção de 200 parâmetros mais relevantes foi possível obter o mesmo desempenho em relação ao total de parâmetros. Por fim, os resultados obtidos neste trabalho permitiram compreender o comportamento dos classificadores na classificação de sequências representando faltas do tipo curto-circuito em linhas de transmissão.

Agradecimentos

Os autores agradecem a Eletrobras-Eletronorte pelo suporte financeiro para o desenvolvimento desta pesquisa.

Referências

- Bollen, M., Gu, I., Santoso, S., Mcgranaghan, M., Crossley, P., Ribeiro, M., and Ribeiro, P. (2009). Bridging the gap between signal and power. *Signal Processing Magazine, IEEE*, 26(4):12–31.
- Castelli, V. and Bergman, L. (2002). *Image Databases - Search and Retrieval of Digital Imagery*. John Wiley and Sons.
- EMTP (1995). *Alternative transient program (ATP) rule book*. Canadian/American EMTP User's Group.
- Englert, H. and Stenzel, J. (2002). Automated classification of power quality events using speech recognition techniques. In *14th Power Systems Computation Conference*.
- Hall, M. (1999). *Correlation-based feature selection for machine learning*. PhD thesis, University of Waikato.
- IEC (2002). Electromagnetic compatibility (EMC) - part 4 - 7: Testing and measurement techniques - general guide on harmonics and interharmonics measurements and instrumentation, for power supply systems and equipment connected thereto.
- Luo, X. and Kezunovic, M. (2005). Fault analysis based on integration of digital relay and DFR. *Power Engineering Society General Meeting*, 1:746 – 751.
- Morais, J., Pires, Y., Cardoso, C., and Klautau, A. (2010). A framework for evaluating automatic classification of underlying causes of disturbances and its application to short-circuit faults. *Power Delivery, IEEE Transactions on*, 25(4):2083–2094.
- Witten, I. and Frank, E. (2005). *Data mining: practical machine learning tools and techniques with java implementations*. Morgan Kaufmann.