



VI SBQEE

21 a 24 de agosto de 2005

Belém – Pará – Brasil



Código: BEL 13 7679

Tópico: Sistemas de Monitoramento e Tratamento de dados

REDES NEURAIS NA CLASSIFICAÇÃO DE EVENTOS DE QUALIDADE DE ENERGIA ELÉTRICA: UMA ABORDAGEM COMPARATIVA

AUGUSTO S.
CERQUEIRA
ENG. ELÉTRICA

CARLOS A. DUQUE
ENG. ELÉTRICA

ROGÉRIO M.
TRINDADE
ENG. ELÉTRICA

MOISÉS V. RIBEIRO
ENG. ELÉTRICA

UFJF

RESUMO

Este trabalho apresenta uma análise comparativa da performance de três métodos utilizados para classificação de eventos de qualidade de energia. O primeiro método utiliza uma rede neuronal de múltiplas camadas, o segundo método utiliza máquinas de vetores de suporte (SVM) e o terceiro método é baseado na extração de parâmetros utilizando mapeamento tempo-frequência. A teoria dos três métodos é resumidamente apresentada, juntamente com a comparação de seus desempenhos e esforço computacional.

PALAVRAS-CHAVE

Qualidade de Energia, Classificação de Distúrbios, SVM, Redes Neurais.

1.0 INTRODUÇÃO

O aumento do número de cargas de natureza não linear (sistemas microprocessadores, fontes chaveadas, inversores de frequência, etc.) em instalações residenciais, comerciais e industriais tem afetado consideravelmente a qualidade da energia entregue pela concessionária de energia, ao mesmo tempo em que a perda da qualidade de energia tem provocado danos aos processos controlados por dispositivos sensíveis. Neste contexto, a aplicação de técnicas de reconhecimento de padrões e de extração de parâmetros [1] (Theodoridis&Koutroumbas, 1999) para monitoramento da QE, com destaque para a classificação de distúrbios e a identificação de fontes geradoras de tais distúrbios [2](Jacques),

são relevantes e têm, recentemente, recebido atenção de pesquisadores do mundo todo.

Dentre as diversas áreas de estudo relacionadas com a QE, a de classificação de distúrbios tem recebido diversas contribuições nos últimos 5 anos. Isto se deve ao fato que para se tomar atitudes corretivas e/ou punitivas nos sistemas de energia é preciso ter à mão ferramentas que identifiquem e quantifiquem os diversos distúrbios que ocorrem em um dado ponto da rede.

A Fig. 1 mostra o diagrama de blocos de um classificador genérico. O bloco 1 é o bloco de entrada de dados que supostamente contém o distúrbio a ser classificado. O vetor de saída \mathbf{x} contém N amostras do distúrbio. O bloco 2 corresponde a fase de pré-processamento do vetor de dados, cujo objetivo é extrair do vetor de dados um conjunto de parâmetros representativos do evento \mathbf{p}_x , ou simplesmente mapear o vetor \mathbf{x} em um novo vetor de dados, \mathbf{u} , em um novo espaço. Finalmente o bloco 3 implementa o classificador propriamente dito que em geral é um classificador não-linear.



Fig. 1. Diagrama de blocos de um classificador genérico.

Dentre as técnicas aplicadas ao bloco 2, destacam-se a transformada de wavelet, transformada FFT, S-transform, estatísticas de ordem superior (EOS), extração do sinal do erro,

entre outras. Tanto as redes neurais multicamadas perceptron (MLP) quanto às máquinas de vetores de suporte (SVM) podem ser vistas como ferramentas não-lineares que mapeiam o vetor de entrada em um novo espaço não-linear, de forma a possibilitar a classificação de classes não-linearmente separáveis.

A implementação do bloco 3 emprega técnicas de inteligência computacional para realizar a classificação dos eventos. Em geral, são utilizados redes neurais ou classificadores fuzzy [3]-[4]. Pode-se usar também SVM para realizar a classificação dos eventos [5], onde a escolha da classe a qual pertence o evento pode ser realizada por um simples método de votação [6].

Este trabalho discute três técnicas de classificação e apresenta resultados comparativos em termos de eficiência. Duas grandes contribuições são apresentadas neste trabalho: a primeira diz respeito à comparação entre técnicas de classificação. De fato, como o tema de classificação de eventos de QE é recente, as pesquisas têm se concentradas no desenvolvimento de novas metodologias de classificação e pouca importância se tem dado na comparação com outros métodos. A segunda contribuição é que pela primeira vez o método SVM é aplicado como classificador de distúrbios elétricos.

As seguintes técnicas são comparadas neste trabalho: (a) Redes Neurais MLP aplicada a componente do erro MLP_e; (b) Método SVM também aplicado a componente do erro (c) Método da Representação ótima Tempo-Frequência (OTFR – *Optimal Time-Frequency Representation*).

Na próxima seção, o método MLP_e é descrito. A seção 3.0 apresenta a técnica SVM. Na seção 4.0, é discutido o método OTFR para que em seguida sejam apresentados os resultados comparativos entre as técnicas. Finalmente, na Seção 6.0, as conclusões são apresentadas.

2.0 REDES NEURAIS MLP APLICADAS À COMPONENTE DO ERRO (MLP_e)

O método MLP_e encontra-se bem documentado em [7]. A Fig. 2 mostra o seu diagrama de blocos.

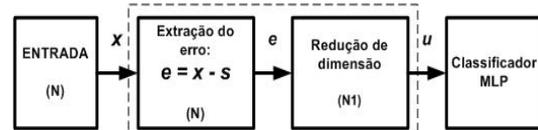


Fig. 2. Diagrama de blocos do método MLP_e.

No método MLP_e, o pré-processamento do sinal é dividido em duas etapas conforme apresentado na Figura 2. A primeira etapa consiste em extrair do vetor de entrada \mathbf{x} , de dimensão N , a componente senoidal que representa o estado permanente da rede, \mathbf{s} , gerando o vetor de erro \mathbf{e} , de mesma dimensão que o vetor original \mathbf{x} . O passo seguinte consiste do alinhamento dos eventos pelo início e pela redução da dimensão do vetor de erro, gerando o vetor \mathbf{e}_s de dimensão $N1$. No modelo usado para gerar os resultados deste trabalho, foi utilizado $N = 1024$ e $N1 = 64$.

A técnica da extração da componente fundamental do sinal de entrada tem sido utilizada pelos autores [8] para analisar a existência de eventos e classificá-los. Esta técnica extrai do sinal a componente de maior energia (a componente fundamental) entregando aos algoritmos de detecção e classificação apenas a informação necessária, o que aumenta o desempenho dos detectores e classificadores de eventos. Outra característica do método MLP_e é o baixo esforço computacional comparativamente com outros processos de classificação, sendo portanto indicado para classificação em tempo real.

Após o pré-processamento do sinal de tensão, o vetor \mathbf{e}_s é apresentado a rede MLP de classificação que utiliza uma única camada escondida. Pode-se interpretar esta camada escondida, com N_h neurônios, como um mapeamento não-linear do vetor de entrada \mathbf{e}_s em um novo espaço com dimensão N_h , sendo então a classificação realizada na camada de saída da rede, composta por N_s neurônios, onde N_s é igual ao número de classes a serem separadas.

A rede foi treinada para classificar os seguintes eventos: afundamentos e sobre-tensão, *notchs*, *spikes*, harmônicos e chaveamentos capacitivos. Neste trabalho, usou-se $N_h = 8$ e $N_s = 5$.

A Fig. 3 mostra as janelas típicas de eventos e a Fig. 4 as janelas de erro com dimensão reduzida (64 amostras) prontas para serem classificadas pela rede MLP. De cima para baixo temos: (a) harmônico, (b) afundamento (c) chaveamento capacitivo (d) *notch* (e) *spike*.

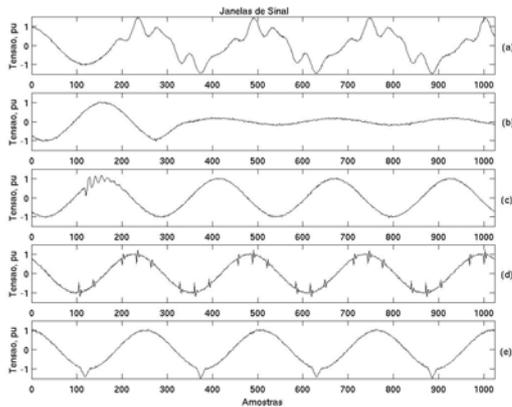


Fig. 3. Janelas de sinal com os eventos de qualidade de energia.

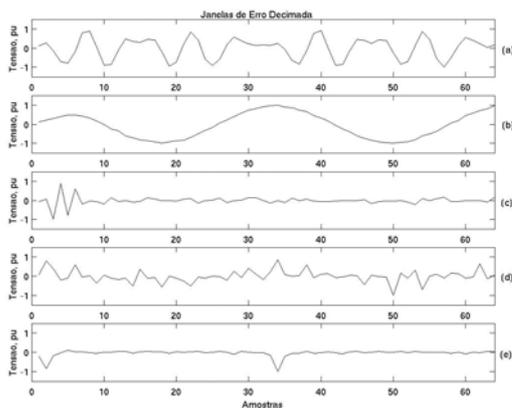


Fig. 4. Janelas de erro decimada, pronta para ser aplicada a rede MLP.

3.0 MÁQUINA DE VETORES DE SUPORTE (SVM)

O método SVM é conhecido como uma excelente ferramenta de classificação que, diferentemente das redes neurais, conduz a um problema de otimização quadrática com restrições lineares [5]. Ele realiza um pré-processamento não-linear que consiste no mapeamento do vetor de entrada, de dimensão N , num espaço de mesma dimensão através de uma função não-linear $\varphi(\cdot)$. Desta forma, um problema de separação de classes não-linear, pode tornar-se linear neste novo espaço, conforme ilustrado na Figura 3.

3.1 Problema de otimização

Considerando um problema de separação linear de duas classes, a equação do hiperplano de separação pode ser expressa como

$$d_i(\mathbf{w}_o^T \mathbf{x}_i + b_o) \geq 1$$

para $i=1,2,\dots,N$

onde \mathbf{w}_o é a matriz de pesos ótimos, b_o a constante de polarização ótima, d_i é a saída desejada (1 ou -1) para o evento projetado no plano e \mathbf{x}_i o vetor de entrada.

Agora se as classes não forem linearmente separáveis, aplica-se uma transformação não-linear no vetor de entrada de forma a que as classes passem a ser linearmente separáveis, conforme:

$$d_i(\mathbf{w}_o^T \varphi(\mathbf{x}_i) + b_o) \geq 1$$

para $i=1,2,\dots,N$

Portanto, dado um conjunto de treinamento $\{(\mathbf{x}_i, d_i)\}_{i=1}^N$, deseja-se achar os valores ótimos do vetor de peso \mathbf{w} e da polarização b , sujeito as condições descritas em (1). A função de custo a qual o vetor de pesos deve minimizar pode ser vista abaixo:

$$\Phi(\mathbf{w}) = \frac{1}{2} \mathbf{w}^T \mathbf{w}$$

Este é um problema de otimização quadrática que pode ser resolvido pelo método dos multiplicadores de Lagrange [9]. Dado o conjunto de treinamento $\{(\mathbf{x}_i, d_i)\}_{i=1}^N$, achar os multiplicadores de Lagrange que maximizam a função objetiva

$$Q(\alpha) = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j d_i d_j K(\mathbf{x}_i, \mathbf{x}_j)$$

sujeito às restrições:

$$(1) \quad \sum_{i=1}^N \alpha_i d_i = 0$$

$$(2) \quad 0 \leq \alpha_i \leq C \quad \text{para } i=1,2,\dots,N$$

onde C é uma constante positiva especificada pelo usuário e

$$K(\mathbf{x}_i, \mathbf{x}_j) = \varphi^T(\mathbf{x}_i)\varphi(\mathbf{x}_j), \quad (3)$$

é o produto interno kernel. A Equação (3) é um caso especial do Teorema de Mercer [10].

3.2 Classificação

Uma vez determinados os coeficientes de Lagrange em (2), a classificação entre duas classes pode ser feita utilizando

$$\sum_{i=1}^M \alpha_i d_i K(\mathbf{x}, \mathbf{x}_i) \begin{cases} \geq 0 & \in \text{classe1} \\ < 0 & \in \text{classe2} \end{cases} \quad (4)$$

Portanto, não é necessário voltar ao espaço original dos dados e nem conhecer a função não-linear $\varphi(\cdot)$.

A Equação (4) permite avaliar se o vetor \mathbf{x} pertence a classe 1 ou a classe 2. Quando temos um problema de separação com mais de duas classes deve-se dividir o problema em separação de classes duas à duas. Portanto, o número de máquinas de vetores de suporte necessárias para a separação de C classes, pode ser calculado através de

$$\frac{C!}{2 \cdot (C-2)!} \quad (5)$$

3.3 Sistema Implementado

Para uma comparação direta com o método MLP_e, foi aplicado o método SVM no vetor de erro com dimensão reduzida, \mathbf{e}_s . Olhando a Fig. 2, basta substituir a rede MLP por um sistema de classificação SVM.

Como se deseja separar 5 classes, precisamos de 10 máquinas (conforme Equação (5)) de vetores de suporte em paralelo para realizar a classificação. Cada uma dessas 10 saídas indica se é uma determinada classe (ex: *flicker*) ou se é outra classe (ex: *notch*). Portanto, uma classe pode ser identificada em no máximo 4 das 10 saídas do sistema. Neste trabalho, o critério adotado para classificação foi: a classe vencedora é aquela indicada em 4 das 10 saídas do sistema.

A função base radial (Equação (6)) foi utilizada como produto interno *kernel*, uma vez que esta função atende ao teorema de Mercer.

$$K(\mathbf{x}, \mathbf{x}_i) = \exp\left(-\frac{1}{2\sigma^2} \|\mathbf{x} - \mathbf{x}_i\|^2\right) \quad (6)$$

4.0 MÉTODO OTFR

Em [11]-[12] foi introduzido o uso da representação tempo-frequência (*time frequency representation* - TFR) para a extração de um conjunto reduzido de parâmetros visando a classificação de alguns distúrbios que ocorrem isoladamente em sistemas de potência. A vantagem desta técnica advém do fato de que conjuntos representativos de vetores de parâmetros associados aos vários distúrbios em sistemas de potência podem ser extraídos a partir do plano tempo-frequência dos distúrbios. A versão discreta da representação tempo-frequência utilizada em [12]-[13] é dada por [14]

$$\begin{aligned} TFR[n, k] &= F_{\eta \rightarrow n}^{-1} \{ F_{\tau \rightarrow k} \{ A[\eta, \tau] \varphi[\eta, \tau] \} \} \\ &= \frac{1}{N} \sum_{\eta=0}^{N-1} \sum_{\tau=0}^{N-1} A[\eta, \tau] \varphi[\eta, \tau] \exp\left(\frac{2\pi k \tau}{N}\right) \exp\left(\frac{2\pi \eta n}{N}\right) \end{aligned}$$

para a extração de parâmetros, sendo

$$A[\eta, \tau] = F_{n \rightarrow \eta} \{ R[n, \tau] \} = \sum_{n=0}^{N-1} R[n, \tau] \exp\left(\frac{2\pi \eta n}{N}\right) \quad (7)$$

o plano de ambigüidade,

$$R[n, \tau] = x^* [n] x[\text{mod}(n + \tau, N)]$$

a função de autocorrelação do vetor \mathbf{x} , $F\{\}$ e $F^{-1}\{\}$ a DFT e a IDFT, τ o atraso no tempo discreto, η o deslocamento discretos na frequência, n a n -ésima amostras no tempo discreto, k a k -ésima denota amostras na frequência e $\text{mod}(p_1, p_2)$ é resto da divisão de p_1 por p_2 . o *kernel* $\varphi_i[\eta, \tau]$ tem a mesma dimensão do plano de ambigüidade. Em [11]-[12] foi adotado

$$A[\eta, \tau] \varphi_i[\eta, \tau] = \begin{cases} A[\eta, \tau], & \text{se } \varphi_i[\eta, \tau] = 1 \\ 0, & \text{se } \varphi_i[\eta, \tau] = 0 \end{cases}$$

ou seja, o *kernel* $\varphi_i[\eta, \tau]$ é definido com uma matriz binária.

De acordo com [11]-[13], $N-1$ *kernels* são necessários para o projeto de um sistema de classificação de N distúrbios. Um *kernel* K_s pode ser usado para separar uma única classe ou um grupo de classes de distúrbios. Para separar uma única classe, o *kernel* K_i é responsável por separar a classe i das classes $\{i+1, \dots, C\}$. Para o caso da separação de um grupo de classes, o

kernel K_i realiza a separação entre as classes $\{i, i+1, \dots, m\}$ das classes restantes $\{m+i+1, m+i+2, \dots, C\}$. A seguir, novos kernels são necessários para unicamente identificar a classe i do grupo de classes $\{i+1, \dots, m\}$. O critério para escolha dos kernels é o discriminante de Fisher [11],[13].

A Fig. 6 mostra o diagrama de blocos do classificador, baseado na TFR, introduzido por Wang et. al [11]-[12] para a classificação dos seguintes distúrbios: harmônicas, *sag/swell*, chaveamento capacitivo, *notches* e *spikes*. Nessa figura, os blocos kernel H, SS, C, N e S implementam os kernels associados aos respectivos distúrbios; os blocos Classe SS, C, N e S fazem uso das redes neurais MLP para aproximar as superfícies de separação entre as classes de distúrbios e o bloco $\rho(f-t)$ implementa o classificador linear de Heaviside dado por (7)

$$\rho(f-t) = \begin{cases} f \in H, & \text{se } f-t \geq 0 \\ f \notin H, & \text{se } f-t < 0 \end{cases}$$

onde H representa a classe harmônica, $f = |A[\eta_h, \tau_h] \varphi_i[\eta_h, \tau_h]|$ e $t = (t_{\min} + t_{\max}) / 2$ sendo $t_{\min} = \min\{|A[\eta_h, \tau_h] \varphi_i[\eta_h, \tau_h]|\}$ e $t_{\max} = \max\{|A[\eta_h, \tau_h] \varphi_i[\eta_h, \tau_h]|\}$.

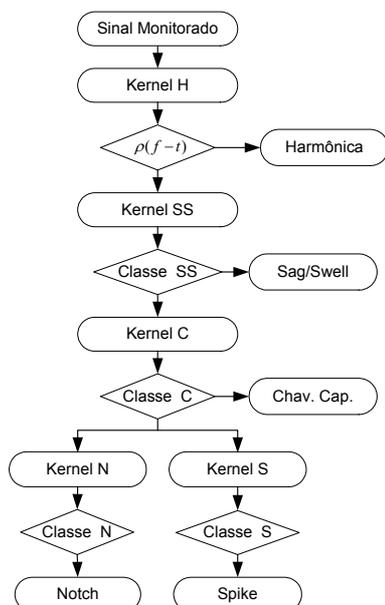


Fig. 6. Diagrama de blocos do classificador proposto em [11]-[12].

5.0 RESULTADOS

Para comparação dos três métodos descritos, o mesmo banco de dados foi utilizado. Neste banco de dados, cada classe possui 500 eventos, resultando num total de 2500 eventos. Cada evento foi amostrado com uma frequência de amostragem de $f_s = 15360$ Hz e possui um comprimento total de 1024 amostras, o que corresponde a 4 períodos da componente fundamental do sinal.

Na Tabela 1, vemos o resultado da eficiência na classificação dos três métodos implementados. A última linha da tabela apresenta o valor da eficiência global, definida como o produto das eficiências de cada classe. Pode-se ver que o método MLP_e e o OTFR apresentam performances similares, com ligeira vantagem para o MLP_e. Já o método SVM apresentou uma performance bastante superior aos demais.

Tabela 1. Eficiência na classificação dos três métodos implementados.

	Eficiência no Teste (%)		
	MLP_e	SVM	OTFR
Harm.	97,3	99,8	97,7
Sag/Swell	99,7	100	98,5
Ch. Cap.	95,3	100	96,2
Notch	98,3	99,6	96,8
Spike	99,7	100	97,1
Efic. G.	90,6	99,4	87,0

A performance superior do método SVM já era esperada, pois neste método, temos a garantia de chegar aos valores ótimos dos pesos, enquanto que em redes MLP esta garantia não existe. Outro fator a ser levado em consideração, é que no método SVM o problema é separado em classes duas à duas, o que também favorece a uma melhor performance, com o custo de se aumentar a complexidade computacional.

O método MLP_e apresenta uma complexidade computacional bem inferior ao SVM e ao OTFR. O método OTFR apresenta uma complexidade elevada para extração dos parâmetros, enquanto que para classificação do evento sua complexidade é bastante baixa. No caso do método SVM, a complexidade advém principalmente do número de vetores de suporte utilizado e pela separação do problema em classes duas à duas.

6.0 CONCLUSÕES

Neste trabalho, foram comparados três métodos utilizados para classificação de eventos

em qualidade de energia. O MLP_e, o SVM e o OFTR.

O método SVM apresentou a melhor performance dos três, aliado a um alto custo computacional. Desta forma, sua implementação para aplicações de classificação *online* não é indicada. Entretanto este método pode servir para aplicações *offline* e para avaliação dos resultados obtidos por outros métodos.

Do ponto de vista da performance, os métodos MLP_e e OFTR foram equivalentes, porém, como a complexidade computacional do MLP_e é bem menor, este leva vantagem quando se visa aplicações *online*.

7.0 REFERÊNCIAS

- [1] S. Theodoridis e K. Koutroubas, "Pattern recognition," San Diego: Academic Press, 1999.
- [2] S. Santoso, W. M. Grady, E. J. Powers, J. Lamore e S. C. Bhatt, "Characterization of distribution power quality events with Fourier and wavelet transforms," IEEE Trans. on Power Delivery, 15 (1), 247-254
- [3] E. P. Santos e F. J. V. Zuben, "Efficient second-order learning algorithm for discrete-time recurrent neural networks," Em L. R. Medsker & L. Jain (Eds.), (p. 47-75), Boca Raton, FL: CRC Press.
- [4] E. Styvaktakis, M. H. J. Bollen, e I. Y. H. Gu, "Expert system for classification and analysis of power system events," IEEE Trans. on Power Delivery, 17 (2), 423-428, Abril, 2002
- [5]
- [6] V. N. Vapnik, "Statistical Learning Theory," John Wiley & Sons, 1998.
- [7] M. Moreira e E. Mayoraz, "Improved pairwise coupling classification with correcting classifiers," Proc. 10th European Conf. on Machine Learning, Chemnitz, Alemanha, pp. 160-171, Abril 1998.
- [8] A. S. Cerqueira, C. A. Duque, R. M. Trindade e M. V. Ribeiro, "Digital system for detection and classification of electrical events," ISCAS, 2005.
- [9] C. A. Duque, M. V. Ribeiro, F. R. Ramos, J. Szczupak, " Power Quality Event Detection Based on the Principle of Divide to Conquer and Innovation Concept," IEEE Tans. on Power Delivery, 2004 (aceito para publicação).
- [10] D. P. Bertsekas, "Nonlinear programming", Belmont, MA: Athenas Scientific, 1995.
- [11] R. Courant e D. Hilbert, "Methods of mathematical physics," Vol. I e Vol. II, New York: Wiley Interscience, 1970.
- [12] W. Wang e A. V. Marnishev, "Classification of power quality events using optimal time-frequency representations – Part 1: Theory," IEEE Trans. on Power Delivery, vol. 19, no. 3, pg. 1488-1495, Jul. 2004.
- [13] W. Wang, G. I. Rowe e A. V. Marnishev, "Classification of power quality events using optimal time-frequency representations – Part 2: Application," IEEE Trans. on Power Delivery, vol. 19, no. 3, pg. 1496-1503, Jul. 2004.
- [14] B. W. Gillespie e L. Atlas, "Optimizing the time-frequency kernels for classification," IEEE Trans. on Signal Processing, vol. 49, no. 3, pg. 485-496, Mar. 2001.