



Aplicação de Máquinas de Vetor de Suporte na Classificação de Vozes Patológicas Utilizando o Expoente de Hurst

Jayne dos Santos Lima¹, Thamyres Tânulla C. Palitó¹, Vinícius Jefferson Dias Vieira², Silvana Cunha Costa³, Suzete E. Nóbrega Correia³, Washington César de Almeida Costa⁴

¹Graduandas do Curso Superior de Bacharelado em Engenharia Elétrica – IFPB. e-mails: {jayneslima, thamyrestamulla}@gmail.com,

²Graduando do Curso Superior de Tecnologia em Sistemas de Telecomunicações – IFPB. email: viniciusjdv@gmail.com

³Professoras da Coordenação de Engenharia Elétrica - IFPB. emails: {silvana, suzete}@ifpb.edu.br

⁴Professor da Coordenação de Ciências da Natureza, Matemática e suas Tecnologias - IFPB. e-mail: washington@ifpb.edu.br

Resumo: Técnicas de processamento digital de sinais têm sido fortemente utilizadas através da análise acústica de distúrbios provocados por patologias laringeas, devido a sua simplicidade e natureza não invasiva. No reconhecimento de padrões, uma técnica para classificação de sinais, que vem sendo utilizada recentemente, devido a sua robustez diante de dados com grande dimensão e boa capacidade de generalização, são as máquinas de vetores de suporte. Neste artigo é apresentada uma aplicação desta técnica para a classificação de sinais de vozes saudáveis e vozes afetadas por patologias na laringe, especificamente, edema de Reinke, paralisia nas pregas vocais e nódulos. Foram utilizadas quatro máquinas de vetor de suporte, uma para cada classe de sinais. A característica fornecida ao classificador, no processo de treinamento, é o parâmetro de Hurst obtido pelo método da variância no tempo. Os resultados apresentaram taxa de correta classificação superior a 99% na discriminação entre os sinais de vozes saudáveis e vozes afetadas por paralisia nas pregas vocais.

Palavras-chave: Máquinas de Vetor de Suporte (SVM), Parâmetro de Hurst, Patologias Laringeas, Processamento Digital de Sinais

1. INTRODUÇÃO

A análise acústica é uma técnica não invasiva baseada no processamento digital do sinal de voz, podendo ser empregada como uma ferramenta eficiente para o auxílio ao diagnóstico de distúrbios vocais, classificação de doenças da voz e, particularmente, sua pré-deteção. Além disso, essa técnica pode ser utilizada para determinação objetiva de alterações da função vocal, avaliações de cirurgias, tratamentos farmacológicos e de reabilitação (GODINO-LLORENTE et al, 2006).

Diferentes métodos de processamento de sinais foram propostos para serem utilizados em sistemas automáticos para a deteção de sinais de vozes patológicas. Para estas finalidades, a ideia principal de todos os sistemas propostos é o reconhecimento de padrões. Várias técnicas vêm sendo utilizadas para a classificação de sinais de vozes patológicas com o uso de diferentes parâmetros e medidas acústicas, bem como, a partir da aplicação de inúmeras ferramentas para o reconhecimento de padrões e a classificação automática dos sinais (ALMEIDA, 2010; COSTA, 2008; ESPINOSA, 2000).

Aprendizado de máquina é uma área da Inteligência Artificial que busca o desenvolvimento de sistemas baseados em procedimentos automáticos. Esta área tem como objetivo obter um ganho de desempenho na tarefa executada com a experiência adquirida com a manipulação de dados relativos ao problema tratado (ALMEIDA, 2010).

As Máquinas de Vetor de Suporte foram introduzidas como uma técnica para resolver problemas de reconhecimento de padrões sendo estendidas também para problemas de regressão e aprendizagem de máquinas (BURGES, 1998). Esta estratégia de aprendizagem foi proposta por Vapnik (VAPNIK, 1995) e tem despertado a atenção de pesquisadores devido suas principais características, que são: boa capacidade de generalização, robustez diante de objetos de dimensões elevadas, convexidade da função objetivo, ou seja, possui apenas um mínimo global, capacidade de lidar com dados ruidosos e uma base teórica bem estabelecida na matemática e estatística.

Neste trabalho as máquinas de vetor de suporte são utilizadas para realizar a classificação de sinais de vozes saudáveis e vozes afetadas por patologias laringeas, especificamente, edema de Reinke, paralisia nas dobras vocais e nódulos. A característica empregada para classificação é o expoente de Hurst (H), que é uma medida estatística de auto-similaridade, capaz de representar o



comportamento estocástico do sinal de voz (SOTERRONI, 2008; FANTINATO, 2008). O método usado para obtenção do expoente de Hurst é o da variância no tempo, por apresentar melhor potencial discriminativo entre os métodos testados em trabalhos anteriores (PALITÓ et al, 2010; LIMA et al 2011a, LIMA et al 2011b).

O artigo está organizado da seguinte forma: na seção 2 é apresentada a fundamentação teórica do expoente de Hurst e das máquinas de vetor de suporte. Os materiais e métodos são apresentados na seção 3, seguidos dos resultados e conclusão nas seções 4 e 5, respectivamente.

2. FUNDAMENTAÇÃO TEÓRICA

2.1 Expoente de Hurst

O parâmetro de Hurst, denotado por H , é uma medida que caracteriza o grau de auto-similaridade de um processo. Ao se observar um processo auto-similar em diferentes escalas de tempo, constata-se uma semelhança no padrão das amostras do processo. O valor desse expoente varia entre 0 e 1, com os valores acima de 0,5 representando correlações positivas e os valores abaixo correlações negativas. Para $H = 0,5$ o sinal ou processo é aleatório (SOTERRONI, 2008).

O método da variância no tempo considera que a série temporal representativa do sinal de voz ($X = X_i : i=1,2,\dots,N$) é dividida em $\frac{N}{m}$ blocos de tamanho m e a série agregada $X^{(m)}(k)$ é calculado como:

$$X^{(m)}(k) = \frac{1}{m} \sum_{i=(k-1)m+1}^{km} X(i), \quad k=1,2,3,\dots \quad (1)$$

A variância da amostra para a seqüência agregada é calculado como:

$$\text{var}(X^{(m)}) = \frac{1}{N/m} \sum_{k=1}^{N/m} (X^{(m)}(k))^2 - \left(\frac{1}{N/m} \sum_{k=1}^{N/m} X^{(m)}(k) \right)^2. \quad (2)$$

Este processo é repetido para diferentes valores de m . O processo decresce linearmente em escala logaritma num intervalo $-1 < \beta < 0$. A variância temporal ou agregada é obtida traçando-se um gráfico $\log(\text{var}(X^{(m)}))$ e novamente $\log(m)$, então os pontos estão situados sobre uma linha reta com $-\beta$, $0 < \beta < 1$. Valores estimados de β sugerem auto-similaridade, e uma relação para o grau de auto similaridade é dada por, $H = 1 - \frac{\beta}{2}$ (YILMAZ, 2002).

2.2 Máquinas de Vetor de Suporte

O objetivo da classificação por meio de vetor de suporte é desenvolver uma forma computacionalmente eficiente de aprender bons hiperplanos de separação em um espaço de características de alta dimensão, onde se entende por bons hiperplanos, aqueles que aperfeiçoam os limites de generalização, e por computacionalmente eficiente, algoritmos capazes de trabalhar com amostras de tamanho de ordem de 100 mil instâncias. Existem diversos limites de generalização com diferentes algoritmos: otimização da margem máxima, distribuição das margens, números de vetores de suporte, etc.

Em sua forma básica as máquinas de vetor de suporte são classificadores lineares que separam os dados em duas classes através de um hiperplano de separação. Um hiperplano ótimo separa os dados com a máxima margem possível, que é definida pela soma das distâncias entre os pontos positivos e os pontos negativos mais próximos do hiperplano. Estes pontos são chamados de vetores

de suporte e estão circulados na Figura 1. O hiperplano é construído com base em treinamento prévio em um conjunto finito de dados (VAPNIK, 1999).

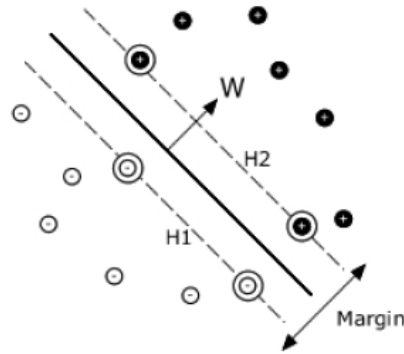


Figura 1 - Classificação de um conjunto de dados usando uma SVM linear (RODRIGUES et al 2007).

Sendo um conjunto de treinamento $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$ onde y tem valor 1 ou -1, indicando a que classe x pertence, e x é um vetor de p dimensões normalizado (com valores 0, 1 ou -1) a SVM divide o hiperplano de acordo com a seguinte equação:

$$w^T + b = 0 \quad (3)$$

em que o vetor w é um vetor perpendicular ao hiperplano, w é o vetor de entrada e b é uma variável que permite que a margem do hiperplano seja maximizada, pois sem esta variável o hiperplano obrigatoriamente passaria pela origem.

Como o interesse está na maior margem de separação, tem-se que dar atenção aos hiperplanos paralelos e ao hiperplano ótimo, mais próximo aos vetores de suporte de cada classe. Estes hiperplanos podem ser descritos pelas equações:

$$w^T + b = 1 \quad (4)$$

$$w^T + b = -1 \quad (5)$$

Sendo o conjunto de dados de aprendizado linearmente separável, é possível selecionar estes hiperplanos maximizando a distância entre eles de modo que não haja pontos no intervalo destes pontos. A geometria mostra que a distância entre estes dois hiperplanos é $2/|w|$, portanto deseja-se minimizar o valor de $|w|$, garantindo que para cada vetor de entrada x_i :

$$\begin{cases} w^T x_i + b \geq 1, & \text{se } y_i = 1 \\ w^T x_i + b \leq -1, & \text{se } y_i = -1 \end{cases} \quad (6)$$

O que pode ser reescrito como:

$$y_i [w^T x_i + b] \geq 1 \quad (7)$$

Quando é necessário realizar uma separação não linear de um conjunto de dados, aplica-se um processo de mapeamento aos exemplos contidos no mesmo. Em um espaço de maior dimensão,

também denominado espaço de características, os dados das duas classes podem ser separados por uma SVM linear com baixo erro. Esse processo de mapeamento é realizado pela introdução e uso de funções *Kernel*, que computam produtos escalares entre cada par de exemplos no espaço de características. Logo, a única modificação necessária nas SVMs para definir funções não lineares na separação dos dados é substituir todos os produtos escalares entre exemplos pela função *Kernel* (LORENA e CARVALHO, 2008).

3. MATERIAL E MÉTODOS

Os sinais de vozes utilizados neste trabalho foram gravados pelo “*Massachusetts Eye and Ear Infirmary*” (MEEI) *Voice and Speech Lab* e distribuídos por *Kay Elementrics* (KAY ELEMETRICS, 1994). Foram gravadas as vozes de 710 indivíduos (657 vozes patológicas e 53 vozes normais). As amostras contêm de 1 a 3 segundos da vogal sustentada /ah/. Foram utilizadas 167 amostras no total, sendo 53 de vozes saudáveis (SDL) e 114 de vozes patológicas: 44 sinais com edema de Reinke (EDM), 52 com paralisia nos nervos laríngeos (PRL) e 18 com nódulos vocais (NDL).

Na Figura 2 é apresentado um diagrama simplificado da metodologia empregada para o sistema proposto de classificação dos sinais de vozes, dividido em três etapas fundamentais:

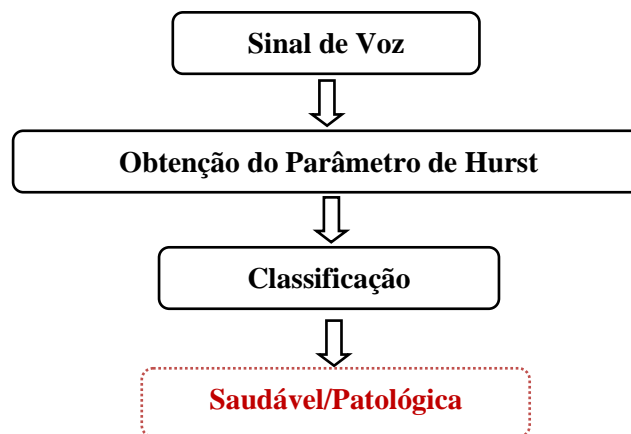


Figura 2 - Diagrama de blocos básico para o sistema proposto.

- 1) O primeiro passo é a aquisição do sinal. Neste trabalho o sinal é oriundo da base de dados citada anteriormente;
- 2) Nesta etapa ocorre a extração do expoente de Hurst, de cada sinal pelo o método da Variância Agregada que serão os dados de entrada no processo de classificação. Os Expoentes de Hurst utilizados neste presente trabalho foram extraídos e analisados em pesquisas anteriores (PALITÓ et al, 2010; LIMA et al 2011a, LIMA et al 2011b).
- 3) A última etapa é a classificação dos sinais de vozes entre saudáveis e patológicos por meio das Máquinas de Vetor de Suporte.

Os expoentes de Hurst extraídos, pelo método da variância no tempo, são utilizados como dados de entrada para as SVMs, que atuam como o método de reconhecimento de padrões. O treinamento das máquinas de vetor de suporte é realizado através do aprendizado supervisionado, pois este tipo de aprendizado detecta automaticamente os exemplos mais relevantes nas bases de dados utilizadas, os vetores de suporte e realiza o mapeamento implícito dos dados de em um espaço dimensões elevadas, através das funções *kernel*.

O processo de classificação de vozes patológicas é dividido em duas partes, o conjunto de treinamento e o conjunto de validação. Neste trabalho foram utilizados 10 subconjuntos na etapa de validação cruzada.

Foram utilizadas quatro máquinas de vetor de suporte, uma para cada classe de sinais. Em todas as máquinas utilizou-se o *kernel* polinomial com parâmetro $C = 0,8$. Este parâmetro indica a distância da margem até a linha de separação das classes. Foram realizados testes com o *kernel* RBF, porém o

polinomial se mostrou mais eficiente. Além disso, diversos valores para o parâmetro C foram empregados, porém os melhores resultados foram obtidos para $C = 0,8$.

Na Figura 3 são apresentados os expoentes de Hurst como característica de entrada fornecida a cada uma das máquinas.

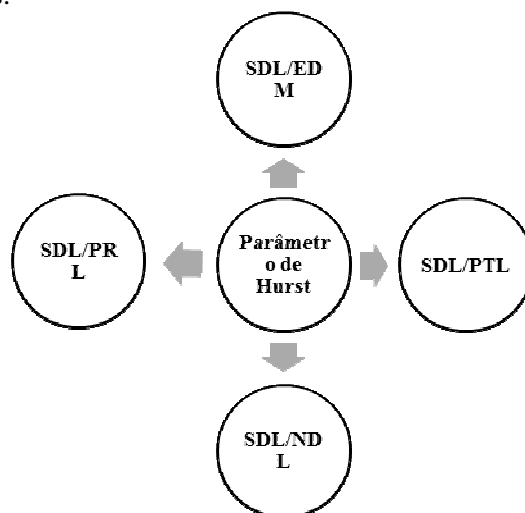


Figura 3 – Classificação através das máquinas.

4. RESULTADOS E DISCUSSÃO

As métricas de avaliação de desempenho utilizadas são: a correta classificação, que é dada pela razão entre o número de acertos pelo número total de amostras, e as taxas de verdadeiro positivo e verdadeiro negativo. O verdadeiro positivo equivale à taxa dos sinais patológicos que foram classificados corretamente pelas máquinas de vetor de suporte, ou seja, a porcentagem de sinais que apresentaram a patologia no teste e realmente eram patológicos. Verdadeiro negativo equivale à taxa de sinais que foram corretamente classificados como saudáveis, ou seja, é a porcentagem dos sinais que não apresentou a patologia no teste e realmente não a tinham.

Na Tabela 1 são apresentados os resultados obtidos para cada máquina de vetor de suporte. A classe (SDL/PRL) se destaca, com a maior taxa de correta aceitação obtida (98,1132%). Para as classes de sinais SDL/PTL e SDL/PRL, foram obtidas taxas de falso positivo inferiores a 5%. Já as classes SDL/EDM e SDL/NDL, os resultados foram considerados bons, tendo em vista que a quantidade de sinais de vozes afetadas pelas patologias edema de Reinke e nódulos vocais é menor que a quantidade de sinais saudáveis, dificultando a aprendizagem das máquinas de vetor de suporte.

Tabela 1- Taxas de classificação das máquinas de vetor de suporte.

Classe	Correta Classificação (%)	Verdadeiro Positivo (%)	Verdadeiro Negativo (%)
SDL/PTL	96,4072	96,2	96,4
SDL/PRL	99,0476	100	98,1
SDL/EDM	90,7216	96,2	84,1
SDL/NDL	94,3662	100	77,7

Nas Tabelas 2, 3, 4 e 5 são apresentadas as matrizes de confusão referentes a cada uma das quatro classes analisadas. A diagonal principal de cada matriz contém a informação de correta classificação dos sinais mostrada na Tabela 1, que é a razão da soma dos sinais da diagonal principal pela quantidade total de sinais.



Tabela 2 – Matriz de confusão da máquina de vetor de suporte para a classe SDL/PTL.

	SDL	PTL
SDL	51	2
PTL	4	110

Tabela 3- Matriz de confusão da máquina de vetor de suporte para a classe SDL/PRL.

	SDL	PRL
SDL	53	0
PRL	1	51

Tabela 4 - Matriz de confusão da máquina de vetor de suporte para a classe SDL/EDM.

	SDL	EDM
SDL	51	2
EDM	7	37

Tabela 5 - Matriz de confusão da máquina de vetor de suporte para a classe SDL/NDL.

	SDL	NDL
SDL	53	0
NDL	4	14

5. CONCLUSÕES

Neste trabalho foi avaliado o desempenho das máquinas de vetor de suporte no processo de classificação de sinais de vozes saudáveis e patológicas, a partir do parâmetro de Hurst, obtido pelo método da variância no tempo. A utilização do parâmetro de Hurst como característica de discriminação desses sinais de vozes, se mostra promissora na separação entre vozes saudáveis e patológicas, apresentando uma taxa de correta classificação superior a 99%.

Como trabalhos futuros pretende-se combinar o expoente de Hurst com outras características dos sinais de vozes, tais como entropia e coeficientes LPC. Além disso, pretende-se investigar a utilização de outras funções *Kernel* (na etapa de classificação) e filtros (na etapa de pré-processamento) para melhorar o desempenho das máquinas de vetor de suporte.

AGRADECIMENTOS

Os autores agradecem ao IFPB e ao CNPq pelo suporte técnico e financeiro, e a UFCG pela concessão da base de dados utilizada nesta pesquisa.

REFERÊNCIAS

ALMEIDA, N.C., **Sistema Inteligente para Diagnóstico de Patologia na Laringe utilizando Máquinas de Vetor de Suporte**. Dissertação de mestrado. Universidade Federal do Rio Grande do Norte UFRN. Rio Grande do Norte, 2010.

BURGES, C. J., **A Tutorial Support Vector Machines for Pattern Recognition**. Data Mining and Knowledge Discovery, vol. 2(nº2), pp.121 – 167, 1998.

COSTA, S. L. N. C. **Análise Acústica, Baseada no Modelo Linear de Produção da Fala, para Discriminação de Vozes Patológicas**. Tese de Doutorado. Universidade Federal de Campina Grande, Campina Grande, 2008.

ESPINOSA, C. H., M. F., GOMEZ, V. P. **Diagnosis of Vocal and Voice Disorders by the Speech Signal**. Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks, Vol. 4, pp. 253-258, 2000.



FANTINATO, P. C. “**Segmentação de Voz baseada na análise fractal e na transformada wavelet**”. Dissertação de Mestrado, Instituto de Física de São Carlos da USP, 2008, São Carlos, 123 p.

GODINO-LLORENTE, J. I., GÓMEZ-VILDA, P., BLANCO VELASCO, M. **Dimensionality Reduction of a Pathological Voice Quality Assessment System Based on Gaussian Mixture Models and Short-Term Cepstral Parameters**. IEEE Transactions on Biomedical Engineering, Vol. 53, No. 10, October, 2006.

HAYKIN, S. **Redes Neurais – Princípios de Práticas** (2ª ed.). Porto Alegre/RS: Editora Bookman, 2001.

KAY ELEMETRICS CORP. **DISORDERED VOICE DATABASE**, model 4337, 03 Ed, 1994.

LANDIS, J.R.; KOCH, G.G. “**The measurement of observer agreement for categorical data**”. Biometrics, v.33,n.1,p.159-174, 1977.

LIMA, J.S., PALITÓ, T.T.C., COSTA, S. C., CORREIA, S. E. N., **Avaliação do potencial discriminativo do parâmetro de Hurst para o diagnóstico de vozes patológicas**. In. 19º Simposio Internacional de Iniciação Científica da USP. São Carlos - SP, 2011a.

LIMA, J.S., PALITÓ, T.T.C., COSTA, S. C., CORREIA, S. E. N., **Avaliação Estatística do Uso do Parâmetro de Hurst na Discriminação entre Vozes Saudáveis e Vozes Patológicas**. In. VI Congresso de Pesquisa e Inovação da Rede Norte e Nordeste de Educação Tecnológica. Natal – RN, 2011b.

LORENA, A. C.; CARVALHO, A. C. P. F., “**Investigação de estratégias para a geração de Máquinas de Vetores de Suporte multiclasses**”, Instituto de Ciências Matemáticas e de Computação da Universidade de São Paulo, 2008.

TAHIM, A. P. N. **Máquinas de Vetores de Suporte (SVM)**. Florianópolis, 22 mar 2010.

VAPNIK, V. **The nature of Statistical Learning Theory**, Nova York, Springer-Vaerlag, 1995.

VAPNIK, V. N. **An overview of statistical learning theory**. IEEE Transactions on Neural Networks, 10(5):988–999, 1999.

PALITÓ, T. T. C., LIMA, J.S., VILELA, S. G., COSTA, S. C., CORREIA, S. E. N. **O uso do Parâmetro de Hurst no Diagnóstico de Patologias na Laringe**. In. 63ª Reunião Anual da SBPC. Jul, 2011, Goiania-GO.

RODRIGUES, R. B., VIANA, R., PISTORI, H., ALVAREZ, M. A. “**Máquinas de Vetores de Suporte Aplicadas à Classificação de Defeitos em Couro Bovino**”. GPEC - Universidade Católica Dom Bosco, 2007.

SOTERRONI, A. C.; DOMINGUES, M. O. and RAMOS, F. M. “**Estimativa do expoente de hurst de séries temporais caóticas por meio da transformada wavelet discreta**”, In. Brazilian Conference on Dynamics, Control and Applications, 2008.

YILMAZ, “**IP over DVB: Management of Self-similarity**,” Bogazici University, 2002.