

ALGORITMO GENÉRICO DE DETECÇÃO DE NOVIDADES EM SÉRIES TEMPORAIS UTILIZANDO MODELOS DE PREVISÃO

ANDRÉ P. LEMOS*, CARLOS JÚLIO T. CRIOLLO†, WALMIR M. CAMINHAS‡

**Departamento de Engenharia Elétrica, Universidade Federal de Minas Gerais
Av. Pres. Antônio Carlos, 6627, 31.270-901, Belo Horizonte, Minas Gerais*

Emails: andrepl@cpdee.ufmg.br, carjulio@cpdee.ufmg.br, caminhas@cpdee.ufmg.br

Abstract— *Novelties or anomalies* on time series can be seen as unexpected values or a sequence of unexpected values when compared to a dataset considered to be normal. A novelty detection algorithm must be capable to classify each point of the time series as normal or novelty. This work proposes a generic novelty detection algorithm based on a forecasting model, handling a novelty as a sequence of unexpected values, in contrast to most methods based on forecasting found on literature that treats a novelty as a single point.

Keywords— Novelty Detection, Time Series, Forecasting Models

Resumo— *Novidades ou anomalias* em uma série temporal podem ser vistas como valores ou uma sequência de valores inesperados quando comparados a um conjunto de valores considerados *normais*. Um algoritmo de detecção de novidades deve ser capaz de classificar os pontos da série como novidades ou normais. Esse trabalho apresenta um algoritmo genérico de detecção de novidades em séries temporais baseado em modelos de previsão, tratando uma novidade como um sequência de valores inesperados, diferentemente da maioria dos métodos baseado em previsão encontrados na literatura que tratam uma novidade de maneira pontual.

Palavras-chave— Detecção de Novidades, Séries Temporais, Modelos de Previsão

1 Introdução

Novidades ou *anomalias* em uma série temporal podem ser vistas como valores ou uma sequência de valores inesperados quando comparados a um conjunto de valores considerados *normais*. Exemplos de novidades podem ser encontrados em séries temporais financeiras, em que uma novidade pode ser vista como uma fraude no sistema, ou na análise de eletrocardiogramas em que uma novidade pode ser vista como alguma anomalia cardíaca.

Um algoritmo de detecção de novidades deve ser capaz de classificar os dados de entrada como normais ou novidades. Porém, em muitos casos práticos, apenas dados de funcionamento normal estão disponíveis para o treinamento do algoritmo. Por isso é importante desenvolver uma técnica de detecção de novidades capaz de superar essa limitação.

Existem diversas técnicas de detecção de novidades em séries temporais encontradas na literatura que utilizam apenas dados de funcionamento normal. Em geral, essas técnicas podem ser divididas em dois grupos distintos, as técnicas baseadas em um modelo de previsão da série (Koskivaara, 1996), (Oliveira and Meira, 2006), e técnicas baseadas em classificação (Ma and Perkins, 2003), (Dasgupta and Forrest, 1996).

As técnicas baseadas em previsão geralmente tratam o problema de detecção de novidades de maneira pontual, em que um modelo de previsão é construído para a série e limites para o erro de previsão são estabelecidos. Para cada ponto a ser classificado, seu valor é previsto pelo modelo e o erro de previsão é calculado. Caso o erro esteja fora dos limites o ponto é classificado como uma

novidade. Essas técnicas não são apropriadas para séries temporais em que uma novidade é caracterizada por uma sequência de pontos que não podem ser descritos pelo modelo de previsão, como por exemplo, na análise de eletrocardiogramas em que uma novidade é caracterizada por um batimento cardíaco que difere dos batimentos tidos como normais, sendo que um batimento é formado por um sequência de pontos.

Esse trabalho propõe um algoritmo genérico de detecção de novidades em séries temporais utilizando um modelo de previsão, tratando uma novidade como uma sequência de pontos que diferem dos pontos tidos como normais, ao invés de tratar uma novidade como um único ponto. O algoritmo é denominado genérico porque não define o modelo de previsão a ser utilizado.

Este artigo está dividido da seguinte forma: na seção 2 será apresentado um algoritmo genérico de detecção de novidades em séries temporais. Na seção 3 duas instanciações do algoritmo serão apresentadas, a primeira utilizando um modelo de previsão estatístico e a segunda um modelo neural. Em seguida, na seção 4, alguns experimentos serão apresentados para validar a eficácia da técnica proposta. Finalmente, na seção 5 será feita uma análise da técnica e uma discussão dos próximos passos a serem tomados.

2 Algoritmo Genérico de Detecção de Novidades

Nessa seção será apresentado um algoritmo genérico de detecção de novidades em séries temporais baseado em um modelo de previsão da série.

Inicialmente, deve-se construir um *modelo de*

previsão $M_x(t)$ responsável por prever o valor atual da série. Esse modelo pode ser um modelo físico, construído a partir do conhecimento prévio do sistema descrito pela série, ou um modelo construído apenas a partir dos dados da série.

O modelo utiliza um conjunto de dados de treinamento formado por pontos da série tidos como normais para estimar seus parâmetros e um conjunto de dados de validação, também formado por pontos normais, para validá-los.

A partir do modelo, calcula-se seu erro de previsão $e(t)$ dado o valor atual observado da série $x(t)$ e o valor previsto do modelo para o instante atual $\hat{x}(t)$:

$$e(t) = \hat{x}(t) - x(t) \quad (1)$$

Em seguida, deve-se então definir um *intervalo de tolerância*, $[\varepsilon_1, \varepsilon_2]$, para os valores do erro de previsão a partir do nível de significância $\alpha \in [0, 1]$, definido como parâmetro do algoritmo. Definido esse intervalo, para todos os pontos a serem classificados calcula-se sua *ocorrência*:

$$O(t) = \begin{cases} 0, & \text{para } e(t) \in [\varepsilon_1, \varepsilon_2] \\ 1, & \text{caso contrário} \end{cases} \quad (2)$$

Uma *surpresa* ocorre quando o valor de $e(t)$ está fora do intervalo de tolerância, ou seja, $O(t) = 1$.

Um *evento* $E_n(t)$ é definido como uma sequência de valores de ocorrência:

$$E_n(t) = [O(t) \ O(t+1) \ \cdots \ O(t+n-1)] \quad (3)$$

sendo n o tamanho do evento, definido como um parâmetro do algoritmo.

A partir dos valores de ocorrência e dado o tamanho do evento n , $m-n+1$ eventos são gerados utilizando-se uma janela deslizante, sendo m o número total de pontos da série a serem classificados. Dessa forma, o primeiro evento contém os valores de ocorrência relativos aos pontos no intervalo $[1, n]$, o segundo contém os valores relativos ao intervalo $[2, n+1]$, assim por diante, gerando a seguinte matriz:

$$\begin{bmatrix} O(1) & O(2) & \cdots & O(n) \\ O(2) & O(3) & \cdots & O(n+1) \\ \vdots & & & \\ O(m-n+1) & O(m-n+2) & \cdots & O(m) \end{bmatrix}$$

Para cada um dos eventos gerados, calcula-se a *norma de magnitude 1 do evento* $|E_n(t)|$ definida como:

$$|E_n(t)| = \sum_{i=0}^{i=n-1} O(t+i) \quad (4)$$

Essa grandeza mede o número de surpresas que ocorreram no evento $E_n(t)$. A função de probabilidade de $|E_n(t)|$ é representada como $p(|E_n(t)|)$ para $|E_n(t)| = 0 \cdots n$. Para se definir a forma de $p(|E_n(t)|)$, assume-se que $E_n(t)$ é uma sequência de variáveis aleatórias independentes de Bernoulli com a mesma probabilidade de sucesso q . Dessa maneira a função de densidade de probabilidade discreta de $|E_n(t)|$ pode ser descrita como uma distribuição Binomial:

$$p(|E_n(t)|) = \begin{cases} \binom{n}{|E_n(t)|} q^{|E_n(t)|} (1-q)^{n-|E_n(t)|} \\ \text{para } |E_n(t)| = 0 \cdots n \\ 0 \text{ caso contrário} \end{cases} \quad (5)$$

Dado o nível de significância α , um evento $E_n(t)$ é definido como um *evento de novidade* se o valor de $|E_n(t)|$ for maior que o limite do intervalo de confiança unilateral construído a partir de $p(|E_n(t)|)$. Consequentemente todos os pontos que fazem parte do evento são classificados parcialmente como novidades.

Para se calcular o intervalo de confiança unilateral de $p(|E_n(t)|)$ deve-se encontrar o menor valor inteiro γ de forma que a função de probabilidade acumulada de $|E_n(t)|$ avaliada em γ seja igual ou exceda o valor de α .

Realizada a classificação dos eventos, cada ponto do intervalo $[n, m-n+1]$ terá n classificações parciais distintas, pois cada um dos pontos desse intervalo participa de n eventos. Já os pontos do intervalo $[1, n-1]$ possuirão i classificações, sendo i a posição do ponto no intervalo, pois participam de menos eventos. Finalmente os pontos do intervalo $[m-n+2, m]$ terão j classificações, sendo $j = m-i$.

Para se chegar a classificação final dos pontos, caso o ponto possua $k\%$ classificações parciais como novidade, o ponto é definido como uma novidade, sendo $k \in (0, 100]$.

Finalizando, para se utilizar o algoritmo proposto deve-se definir:

- O modelo de previsão $M_x(t)$.
- Os intervalos de tolerância do erro de previsão $[\varepsilon_1, \varepsilon_2]$.
- O valor da probabilidade de surpresa q .
- Os parâmetros do algoritmo, o nível de significância α , o tamanho do evento n e o percentual de classificações parciais necessárias para que um ponto seja classificado como novidade k .

3 Instanciações do Algoritmo

Nessa seção duas instanciações do algoritmo são apresentadas. A primeira, baseada em um modelo

de previsão estatístico e a segunda em um modelo neural.

3.1 Modelos Autoregressivos e Intervalos de Previsão

Nessa instancição, o modelo de previsão $M_x(t)$ é representado por um modelo de previsão autorregressivo de ordem p , $AR(p)$. Dado um processo estocástico estacionário, um modelo autorregressivo linear de ordem p é o modelo em que o valor atual do processo é expresso como uma combinação linear dos p valores anteriores $[X_{t-1}, X_{t-2}, \dots, X_{t-p}]$ acrescido de um ruído branco gaussiano a_t . Assume-se que o ruído é gaussiano, independente e identicamente distribuído (i.i.d), com média zero e variância σ_a^2 . Dessa maneira o modelo autorregressivo pode ser escrito como:

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + a_t \quad (6)$$

Dada uma série temporal que possa ser descrita como um processo estocástico, para se construir um modelo de previsão para essa série utilizando a equação (6) deve-se inicialmente estimar os valores dos parâmetros (ϕ, σ_a^2) para diversos valores de p e em seguida avaliar qual é o valor mais plausível para p a partir de algum critério estatístico.

O estimador de máxima verossimilhança para os parâmetros do modelo é dado por (Box and Jenkins, 1990), (Davis and Vinter, 1985):

$$\begin{aligned} \hat{\phi} &= (X'X)^{-1}(X'Z) \\ \hat{\sigma}_a^2 &= \frac{1}{N-p} (Z - X\hat{\phi})(Z - X\hat{\phi})^{-1} \end{aligned} \quad (7)$$

Sendo:

$$Z = \begin{bmatrix} X_{p+1} \\ X_{p+2} \\ \vdots \\ X_N \end{bmatrix} \text{ e } X = \begin{bmatrix} X_p & \dots & X_1 \\ X_{p-1} & \dots & X_2 \\ \vdots & \ddots & \vdots \\ X_{N-1} & \dots & X_{N-p} \end{bmatrix}.$$

A partir do estimador definido em (7) e dos dados da série deve-se então estimar a ordem do modelo p a ser utilizado. O *Critério Bayesiano de informação de Schwarz (BIC)* (Schwarz, 1978) é comumente utilizado para selecionar a ordem modelo:

$$BIC = \log(\hat{\sigma}_a^2) + \frac{(p+1)\log(N)}{N} \quad (8)$$

sendo p a ordem do modelo autorregressivo e N o número de pontos da série.

Calculado o valor de BIC para diversos valores de p deve-se escolher o valor de p associado ao menor valor de BIC.

A previsão do valor atual da série descrita um modelo $AR(p)$ é definida como:

$$\hat{x}_t = \phi_1 x_{t-1} + \dots + \phi_p x_{t-p} \quad (9)$$

O limite de tolerância do erro de previsão é calculado a partir intervalo de previsão do modelo. Esse intervalo é calculado a partir da distribuição condicional do modelo $p(X_t|X_{t-1}, \dots, X_{t-p})$. Caso o ruído presente no sinal seja gaussiano, essa distribuição também é gaussiana e o intervalo de previsão é calculado através de um intervalo de confiança dado o nível de significância α :

$$\hat{x}(t) \pm Q^{-1} \left(\frac{1-\alpha}{2} \right) \hat{\sigma}_a \quad (10)$$

sendo $Q(x)$ a *função de distribuição cumulativa complementar* de uma variável aleatória gaussiana (Kay, 1993).

Dessa maneira o intervalo de tolerância do erro de previsão é dado por:

$$\begin{aligned} \epsilon_1 &= -Q^{-1} \left(\frac{1-\alpha}{2} \right) \hat{\sigma}_a \\ \epsilon_2 &= Q^{-1} \left(\frac{1-\alpha}{2} \right) \hat{\sigma}_a \end{aligned} \quad (11)$$

Finalmente, a probabilidade de surpresa de $p(|E_n(t)|)$, $q = 1 - \alpha$, pois essa é a probabilidade de que o erro de previsão esteja fora do intervalo de tolerância.

3.2 Redes Neurais e Intervalos Robustos de Confiança

Para se utilizar o modelo de previsão $AR(p)$ descrito na seção 3.1 uma série de restrições devem ser assumidas a partir da série (Ehlers, 1996). Em geral, a série deve possuir função de autocorrelação amostral com decaimento exponencial ou oscilatório, função de autocorrelação parcial amostral com valores nulos após o instante p e ruído descorrelacionado e gaussiano. Nessa seção será apresentada uma instancição do algoritmo proposto que pode ser utilizada quando a série temporal não atenda esses pré-requisitos.

Utiliza-se como modelo de previsão $M_x(t)$ um comitê de redes *Multi Layer Perceptron* baseados na média de *ensemble* (Haykin, 1999). Cada rede do comitê utiliza os últimos p valores observados da série para prever o valor atual da série. A figura 3.2 ilustra a topologia de cada uma das redes.

Para calcular os intervalos de tolerância do erro de previsão utiliza-se a técnica de *intervalos robustos de confiança* (Masters, 1995). Essa técnica calcula numericamente intervalos de confiança da distribuição do erro de previsão, sem que seja necessário definir a forma de sua distribuição

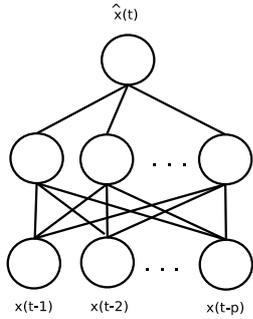


Figura 1: Topologia a rede MLP utilizada para realizar previsão da série temporal

de probabilidade, apenas assume-se que esta distribuição é simétrica. Para isso, esta utiliza a *função de distribuição amostral* do erro de previsão, $S_n(e)$. Dado um conjunto de l valores do erro de previsão ordenados, essa função retorna a fração dessa coleção menor ou igual a um dado valor do erro (Masters, 1995):

$$S_n(e) = \begin{cases} 0, & e < e(1) \\ r/n, & e(r) \leq e \leq e(r+1) \\ 1, & e(l) \leq e \end{cases} \quad (12)$$

Caso o conjunto de pontos utilizados para calcular $S_n(e)$ represente os valores de erros encontrados na prática e o tamanho desse conjunto seja suficientemente grande, $S_n(e)$ é assumido ser uma boa estimativa da verdadeira função de distribuição de probabilidade do erro de previsão $F(e)$.

Dessa maneira, para se calcular o intervalo de confiança do erro de previsão dado um nível de significância α deve-se descartar $l \times \alpha/2$ valores de cada extremo do conjunto de valores de $e(t)$ e os limites do intervalo de confiança serão os valores limites do conjunto resultante.

O valor da probabilidade de surpresa de $p(|E_n(t)|)$ é definido como $q = 1 - \alpha$, assim como na instânciação anterior.

4 Experimentos

Nessa seção serão apresentados experimentos construídos para provar a eficácia da técnica descrita.

O primeiro experimento foi gerado artificialmente de forma que a novidade a ser detectada é conhecida. Os dados foram gerados a partir do seguinte processo estocástico:

$$Z_t = \sin\left(\frac{60\pi}{N}\right) + a_t + e_t \quad (13)$$

sendo $t = 1, 2, \dots, N$, $N = 1800$ e $a_t \sim N(0, 0.1)$, isto é, um ruído branco aditivo gaussiano de média

0 e variância $\sigma^2 = 0.1$. e_t é uma *novidade* a ser detectada:

$$e_t = \begin{cases} n_t, & t \in [1540, 1590] \\ 0, & \text{caso contrário} \end{cases} \quad (14)$$

onde $n_t \sim N(0, 0.2)$.

Nesse primeiro experimento utilizou-se o método baseado no modelo $AR(p)$. Os primeiros 600 pontos de uma série gerada a partir desse processo estocástico foram utilizados para estimar os modelos de previsão e os pontos do intervalo $[601, 1200]$ foram utilizados para validá-los. Finalmente, os pontos do intervalo $[1201, 1800]$ foram utilizados com entrada no algoritmo de classificação. A figura 2 ilustra a série temporal gerada a partir de (13) nesse intervalo.

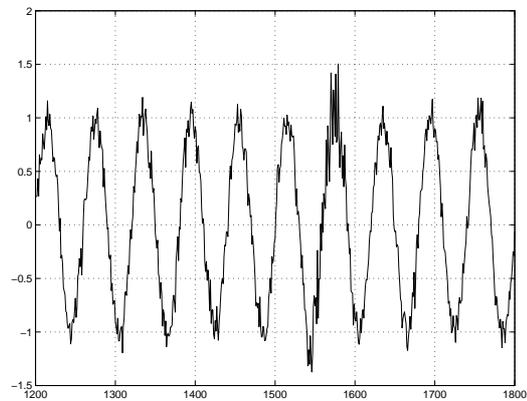


Figura 2: Série temporal com uma novidade no intervalo de tempo $[1540, 1590]$

O critério de informação BIC definiu um modelo $AR(p)$ de ordem 21 e os parâmetros do algoritmo foram ajustados para: $\alpha = 0.05$, $n = 80$ e $k = 50\%$.

A figura 3 mostra a classificação final dos pontos nesse intervalo. Pontos classificados como novidades possuem valor igual a 1. Analisando essa figura, percebe-se que o método teve êxito em detectar a novidade presente na série.

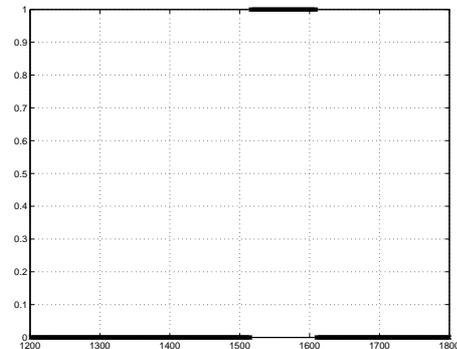


Figura 3: Classificação final dos pontos para o método baseado em um modelo $AR(p)$

A seguir, foi realizado um experimento utilizando uma série temporal real. Nesse experimento utilizou-se o método baseado em redes neurais para detectar anomalias cardíacas na série temporal resultante de um eletrocardiograma.

A figura 4 ilustra a parte do eletrocardiograma que apresenta duas contrações ventriculares prematuras. O instante de ocorrência da anomalia foi anotada por um cardiologista e é representado na figura pela anotação “V”. Pode-se notar que a anomalia cardíaca caracteriza-se como um batimento que difere dos outros batimentos anotados como normais, “•”. Essa figura e os valores da série correspondente foram obtidos a partir de (Goldberger et al., 2000).



Figura 4: Eletrocardiograma apresentando contração ventricular prematura

A série temporal relativa ao eletrocardiograma em questão possui 3000 pontos, sendo que as anomalias cardíacas estão presentes nos últimos 1200 pontos da série. Os primeiros 1000 pontos da série foram utilizados para treinar um comitê de redes MLP composto por 5 redes. Todas as redes possuem 12 entradas e 1 camada escondida composta por 5 neurônios com função de ativação sigmoideal. Os pontos do intervalo [1001, 1800] foram utilizados para estimar os limites do intervalo robusto do erro de previsão. Para aumentar o tamanho do conjunto de valores do erro de previsão foram utilizados os erros de previsão de cada uma das redes do comitê, conforme proposto em (Oliveira and Meira, 2006).

Finalmente os pontos do intervalo [1801, 3000] foram utilizados como entrada do algoritmo de classificação. Os parâmetros do algoritmo foram ajustados para: $\alpha = 0.05$, $n = 150$ e $k = 50\%$. A figura 5 ilustra a série temporal utilizada pelo algoritmo nesse intervalo.

A figura 6 ilustra a classificação final dos pontos. Pode-se notar que o algoritmo detectou as duas anomalias presentes na série.

A instanciação do algoritmo proposta na seção 3.2 pode ser vista como uma extensão do método de detecção de novidades proposto em (Oliveira and Meira, 2006) para séries temporais em que uma novidade é caracterizada por uma

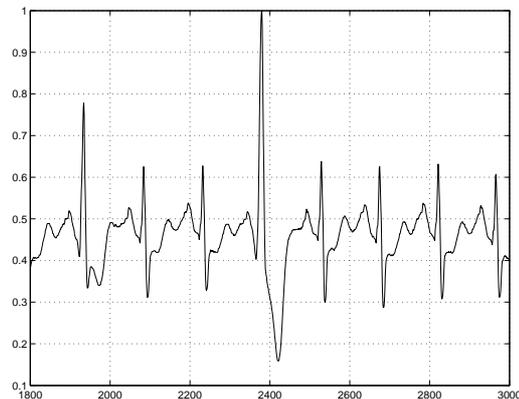


Figura 5: Série temporal utilizada pelo algoritmo para detecção das anomalias cardíacas

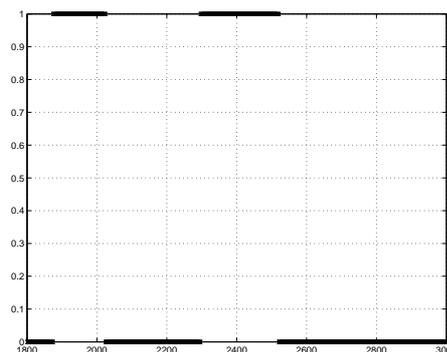


Figura 6: Classificação final dos pontos

sequência de pontos. Fazendo uma analogia entre o algoritmo genérico proposto e o método descrito em (Oliveira and Meira, 2006), esse método define que um ponto é classificado como novidade quando uma surpresa ocorre, ou seja, um ponto é definido como uma novidade caso o erro de previsão esteja fora dos intervalos de tolerância.

Dessa forma, será apresentado um experimento com o intuito de comprovar a eficiência ganha utilizando-se a extensão proposta.

Foram geradas 100 séries temporais a partir do seguinte processo estocástico:

$$Z_t = 0.2 - 3Z_{t-1} + 3\sqrt{Z_{t-1}} + a_t + e_t \quad (15)$$

sendo $t = 1, 2, \dots, m$, $m = 2000$ e $a_t \sim N(0, 0.05)$, isto é, um ruído branco aditivo gaussiano de média 0 e desvio padrão $\sigma = 0.05$ e $Z_1 = 0.3$. n_t é uma novidade a ser detectada:

$$n_t = \begin{cases} g_t, & t \in [1500, 1580] \\ 0, & \text{caso contrário} \end{cases} \quad (16)$$

onde $g_t \sim N(0, 0.1)$.

Os primeiros 500 pontos de cada série foram utilizados para treinar o modelo de previsão e os pontos do intervalo [501, 1000] foram utilizados

para validá-lo. Todas as séries foram modeladas por comitês compostos por 3 redes MLP com 1 entrada, correspondendo ao ponto da série no instante $t - 1$, e 3 neurônios na camada escondida.

A partir dos pontos a serem classificados, as probabilidades de detecção P_D e falsos alarmes P_{FA} médias das duas metodologias foram medidas para valores de α variando de 0.01 a 0.99. Em seguida, as curvas ROC (*Receiver Operating Characteristic*) ((Kay, 1993)) das duas metodologias foram geradas. A figura 7 ilustra essas curvas. Analisando essa figura, percebe-se o aumento significativo no desempenho do detector utilizando a técnica estendida, já que a P_D é maior para qualquer valor de P_{FA} .

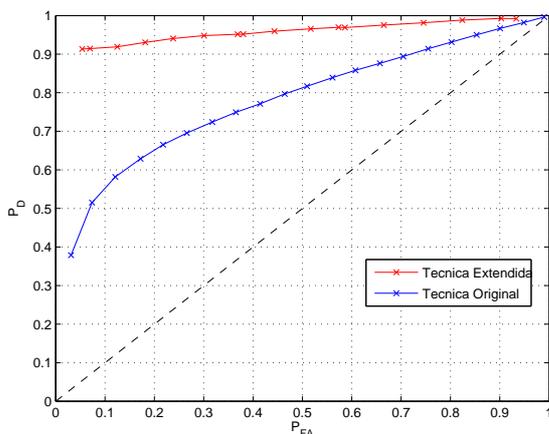


Figura 7: Curva ROC da técnica original e da técnica estendida proposta

5 Conclusões

Nesse trabalho um algoritmo genérico de detecção de novidades em séries temporais foi apresentado juntamente com duas instâncias desse algoritmo. Esse algoritmo é denominado genérico pois não define o modelo de previsão a ser utilizado.

O algoritmo proposto mostrou-se capaz de detectar novidades em dois experimentos, um artificial em que o intervalo de tempo da novidade é conhecido e um segundo experimento baseado na análise de um eletrocardiograma em que o algoritmo detectou com êxito anomalias cardíacas apontadas por um cardiologista.

Conforme descrito anteriormente, a instanciamento do algoritmo baseada em redes neurais pode ser vista como uma extensão da técnica proposta em (Oliveira and Meira, 2006). Dessa forma, um experimento foi realizado comprovando o ganho de eficiência da técnica estendida.

Como próximos passos para esse trabalho apontam-se as seguintes atividades:

- Extensão do método proposto para ser utilizado em séries temporais multivariadas.

- Desenvolvimento de um algoritmo adaptativo, que realize uma atualização dos parâmetros do modelo ao longo do tempo de forma que possa ser aplicado em séries temporais não estacionárias.

Referências

- Box, G. E. P. and Jenkins, G. (1990). *Time Series Analysis, Forecasting and Control*, Holden-Day, Incorporated.
- Dasgupta, D. and Forrest, S. (1996). Novelty detection in time series data using ideas from immunology, *Neural Information Processing Systems (NIPS) Conference*.
- Davis, M. and Vinter, R. (1985). *Stochastic Modelling and Control*, Chapman Hall.
- Dorffner, G. (1996). Neural networks for time series processing, *Neural Network World* **6**: 447–468.
- Ehlers, R. S. (1996). Análise de séries temporais, Web doc. <http://leg.est.ufpr.br/ehlers/notas>.
- Goldberger, A. L., Amaral, L. A. N., Glass, L., Hausdorff, J. M., Ivanov, P. C., Mark, R. G., Mietus, J. E., Moody, G. B., Peng, C.-K. and Stanley, H. E. (2000). PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals, *Circulation* **101**(23): e215–e220.
- Haykin, S. (1999). *Neural Networks - A Comprehensive Foundation*, 2 edn, Prentice Hall.
- Kay, S. M. (1993). *Fundamentals of statistical signal processing: detection theory*, Prentice-Hall, Inc., Upper Saddle River, NJ, USA.
- Koskivaara, E. (1996). Artificial neural network models for predicting patterns in auditing monthly balances, *Technical Report TUCS-TR-67*.
- Ma, J. and Perkins, S. (2003). Time-series novelty detection using one-class support vector machines, *Neural Networks, 2003. Proceedings of the International Joint Conference*, pp. 1741–1745.
- Masters, T. (1995). *Neural, Novel and Hybrid Algorithms for Time Series Prediction*, John Wiley & Sons, Inc., New York, NY, USA.
- Oliveira, A. L. I. and Meira, S. R. L. (2006). Detecting novelties in time series through neural networks forecasting with robust confidence intervals, *Neurocomputing* **70**: 79–92.
- Schwarz, G. (1978). Estimating the dimension of a model, *The Annals of Statistics* **6**: 461–464.