

Estratificação de Risco em Pacientes com Infarto Agudo do Miocárdio Através de Redes Neurais Artificiais e Análise de Correspondência

Alexandre Zanini¹, Renata Teixeira Ladeira², Carlos Eduardo Pedreira¹, José Carlos Nicolau²

¹Dept. Engenharia Elétrica - PUC-RIO

²InCor-HC/FMUSP

E-mails: azanini@ele.puc-rio.br, rtlad@terra.com.br, pedreira@ele.puc-rio.br, corjnicolau@incor.usp.br

Abstract

In this article an application of Artificial Neural Networks and Correspondence Analysis, with the goal of forecasting the mortality prognosis in patients with acute myocardial infarction, it is proposed. Our database was originally composed by 150 variables associated to 522 patients with acute myocardial infarction. Correspondence Analysis was applied to find the representative variables that are statistically associated with mortality. Finally, the selected variables were used as inputs for the Kohonen's algorithm with the goal of finding patients patterns of the patients that evolve to mortality.

1 - Introdução

O infarto agudo do miocárdio (IAM) é uma das principais causas de morte da atualidade e assim há grande interesse em identificar os fatores associados ao mau prognóstico de sua evolução[1].

O objetivo central desse artigo é o desenvolvimento de um sistema de previsão de chance de óbito em pacientes com infarto agudo do miocárdio. Para este fim, estruturou-se o problema em 3 passos: 1) Identificação dos fatores relacionados com a variável óbito; 2) Identificação de padrões dos pacientes com IAM a partir das variáveis identificadas no item 2; e 3) Cálculo da probabilidade de óbito.

Como casuística¹ tem-se 522 pacientes com IAM (75% homens e média de idade igual a 62 com desvio padrão de 13 anos) incluídos de forma prospectiva e consecutiva em um banco de dados específico, abrangendo 150 variáveis, desenvolvido na Unidade Coronariana (UCo) do InCor. Estas 150 variáveis podem ser divididas nos seguintes grupos: 1) Variáveis de tempo (ex: tempo de chegada ao primeiro hospital, tempo de realização do primeiro eletrocardiograma (ECG)); 2) Dados demográficos (ex: sexo, idade); 3) Fatores de risco (ex: história de infarto prévio, história

de angina); 4) Condições de chegada ao hospital (ex: pressão sistólica, pressão diastólica); 5) Diagnóstico na admissão (ex: IAM, angina instável); 6) Alterações iniciais do ECG (ex: bloqueio de ramo esquerdo); 7) Procedimentos de recanalização (ex: uso de trombolítico); 8) Medicamentos usados nas primeiras 24 horas (ex: aspirina); 9) Procedimentos realizados no hospital (ex: cateterismo, cirurgia); 10) Evolução do paciente (ex: isquemia, reinfarto); 11) Complicações (ex: sangramentos importantes); 12) Informações da alta (ex: dias na unidade coronariana) e 13) Dados relacionados ao Cateterismo (ex: artérias com lesões).

Neste artigo, propõe-se a aplicação de Análise de Correspondência [2] e Redes de Kohonen [3] e [4] para tratar os passos 1 e 2 definidos no segundo parágrafo desta Introdução. A estratégia de modelagem proposta consiste em identificar dentre 150 variáveis aquelas associadas com a variável óbito. As variáveis selecionadas serão então usadas para classificar padrões dentre pacientes com IAM que vão a óbito.

Nas seção dois, faz-se uma breve revisão dos fundamentos teóricos do modelo usado para a metodologia proposta. Os resultados são apresentados na seção três, sendo feitas as conclusões na seção quatro.

2. Análise de Correspondência

2.1 Conceituação básica

A Análise de Correspondência (ANACOR) consiste numa extensão do teste de Pearson [1]. A análise permite não só testar a associação entre variáveis, mas também verificar quais categorias das variáveis são associadas e o sentido da associação.

A ANACOR pode ser aplicada sobre Tabelas de Contingência Bi-dimensionais. Isto é, supondo que N é o total de observações, havendo duas variáveis qualitativas (portanto bidimensional) sendo 1 variável com ℓ categorias e a outra com c categorias, tem-se a seguinte representação da tabela de contingência (figura 01):

¹ Utilizou-se um banco de dados desenvolvido pelo Instituto do Coração do Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo (InCor-HC/FMUSP).

		Colunas (variável 2)				Total	
		1	2	...	C		
Linhas (variável 1)	1	n_{11}	n_{12}	...	n_{1C}	$n_{1\bullet}$	totais marginais
	2	n_{21}	n_{22}	...	n_{2C}	$n_{2\bullet}$	
	
	ℓ	$n_{\ell 1}$	$n_{\ell 2}$...	$n_{\ell C}$	$n_{\ell\bullet}$	
Total		$n_{\bullet 1}$	$n_{\bullet 2}$...	$n_{\bullet C}$	$n_{\bullet\bullet} = N$	
		Totais marginais					

Figura 01: Exemplo de tabela bi-dimensional $\ell \times c$

Onde:

n_{ij} = frequência observada (ou contagem) da categoria i da variável de **linha** e da categoria j da variável de **coluna** (célula ij)

$$n_{i\bullet} = n_{i1} + n_{i2} + \dots + n_{iC} = \sum_{j=1}^c n_{ij}$$

$$n_{\bullet j} = n_{1j} + n_{2j} + \dots + n_{\ell j} = \sum_{i=1}^{\ell} n_{ij}$$

$$N = n_{\bullet\bullet} = \sum_{i=1}^{\ell} \sum_{j=1}^c n_{ij} = \sum_{i=1}^{\ell} n_{i\bullet} = \sum_{j=1}^c n_{\bullet j}$$

2.2 Classificações Independentes: independência x associação

Supondo: $P_{ij} \Rightarrow$ probabilidade de que uma observação retirada da população seja da **i -ésima** categoria da variável de linha e da **j -ésima** categoria da variável de coluna

$F_{ij} \Rightarrow$ frequência esperada na célula ij da tabela ao se observar N casos.

Então:
$$F_{ij} = N \cdot P_{ij}$$

(lembrando que $F_{ij} = E(n_{ij})$)

Supondo: $P_{i\bullet} \Rightarrow$ probabilidade, na população, de uma observação pertencer à categoria i da variável de linha, e

$P_{\bullet j} \Rightarrow$ probabilidade, na população, de uma observação pertencer à categoria j da variável de coluna.

Independência significa: $P_{ij} = P_{i\bullet} P_{\bullet j}$

Então:
$$F_{ij} = N P_{i\bullet} P_{\bullet j}$$

Estes valores de probabilidades **populacionais** são desconhecidos. Pode-se, no entanto, **estimar** estas probabilidades; a partir dos totais marginais observados:

$$\hat{P}_{i\bullet} = \frac{n_{i\bullet}}{N} ; \hat{P}_{\bullet j} = \frac{n_{\bullet j}}{N}$$

que são os estimadores de **máxima verossimilhança**.

Então:
$$E_{ij} = N \hat{P}_{i\bullet} \hat{P}_{\bullet j}$$

$$E_{ij} = N \frac{n_{i\bullet}}{N} \frac{n_{\bullet j}}{N} = \frac{n_{i\bullet} n_{\bullet j}}{N}$$

onde E_{ij} = valor esperado (estimado) na célula ij se

as duas variáveis são **independentes** e $n_{ij} =$ valor observado. As diferenças entre n_{ij} e E_{ij} , portanto, devem ser devidas apenas **ao acaso**, caso as 2 variáveis sejam independentes. Qualquer teste de independência entre as 2 variáveis deve se basear nas **diferenças** entre n_{ij} e E_{ij} .

2.3 O Teste Qui-Quadrado

É necessário verificar a “verdade” da hipótese:

$$H_0: P_{ij} = P_{i\bullet} P_{\bullet j}$$

Pearson (1904) sugeriu uma **estatística de teste** baseada nas **diferenças** entre os valores observados (n_{ij}) e os valores estimados (E_{ij}) das frequências, esperadas quando H_0 for verdade. A **estatística de teste** para independência entre as 2 variáveis é:

$$c^2 = \sum_{i=1}^{\ell} \sum_{j=1}^c \frac{(n_{ij} - E_{ij})^2}{E_{ij}}$$

Sob H_0 , prova-se que se as frequências observadas (n_{ij}) têm uma distribuição **multinomial** e se as frequências esperadas estimadas (E_{ij}) não são pequenas então c^2 tem uma distribuição aproximadamente **QUI-QUADRADO**.

2.4 Procedimento Gráfico para Análise de Tabelas de Contingência: Análise de Correspondência

A estatística χ^2 pode indicar que existe uma **correlação** entre as variáveis sem, no entanto, fornecer uma visão clara da natureza da relação estatística.

Para se produzir informação que seja relevante, existem procedimentos que fornecem **gráficos** construídos em espaço multidimensional.

Em linhas gerais, o procedimento calcula **scores** para as linhas e colunas e produz gráficos baseados nestes valores. As categorias que são **similares** aparecem próximas umas das outras. Desta forma é fácil identificar quais as categorias de uma variável que são **similares** ou quais categorias das duas variáveis estão relacionadas estatisticamente.

Analisando **categorias da mesma variável**, está se verificando se existem similaridades. Já ao analisar **categorias de variáveis diferentes**, está se verificando se existe associação entre as categorias das variáveis.

Este procedimento foi utilizado para analisar quais variáveis contidas no banco de dados estavam associadas à variável óbito (analisou-se também a associação entre as categorias).

3 – Resultados Numéricos

Através da ANACOR foram selecionadas 20 variáveis (ver quadro 01) que compuseram o vetor de entrada da rede de Kohonen composta de 6 neurônios.

Quadro 01: Variáveis de entrada da Rede de Kohonen

História de angioplastia prévia	Infarto sem onda Q no ECG
História de insuficiência cardíaca	Infarto de ventrículo direito
Eventos clínicos: <ul style="list-style-type: none"> • Edema pulmonar necessitando drogas • choque cardiogênico • hipotensão suficiente para ter havido intervenção • bloqueio atrio-ventricular • fibrilação ventricular • taquicardia ventricular • ruptura cardíaca • parada cardíaca inesperada • insuficiência mitral • fibrilação atrial 	Remédios na alta do hospital: <ul style="list-style-type: none"> • Inibidor da ECA • Bloqueador de cálcio • Beta bloqueador • Aspirina • Diurético • Estatina

Treinada a rede, um resultado bastante interessante encontrado foi que os óbitos ocorridos na amostra alocaram-se no mesmo grupo. A rede foi então treinada mais 4 vezes obtendo-se sempre o “mesmo”

resultado, ou seja, os óbitos sempre eram alocados no mesmo neurônio. No quadro 02 a seguir, temos a quantidade de pacientes em cada grupo:

Quadro 02: Número de pacientes por neurônio

Neurônio	Pacientes
1	103
2	62
3	94
4	134
5	45
6	84

Nota: No neurônio 6 estão localizados os 55 óbitos da amostra.

A seguir são apresentados as características dos grupos em forma gráfica. **Para cada neurônio, vê-se o percentual de incidência das variáveis.** Salienta-se que as análises serão concentradas no neurônio 6 onde localizaram-se os óbitos.

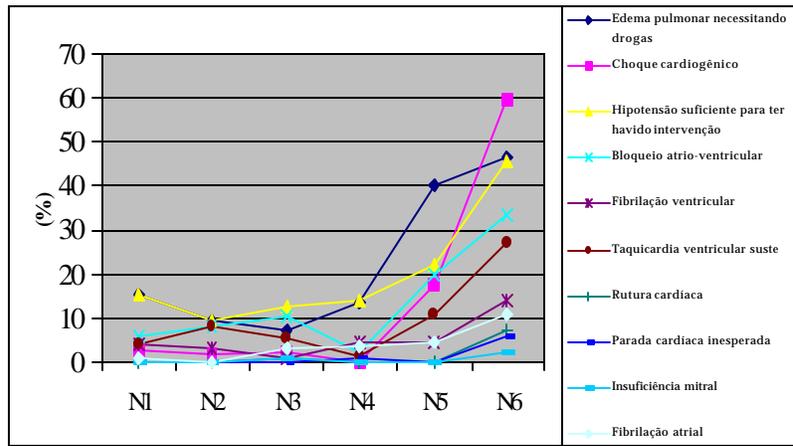


Figura 02: Eventos clínicos

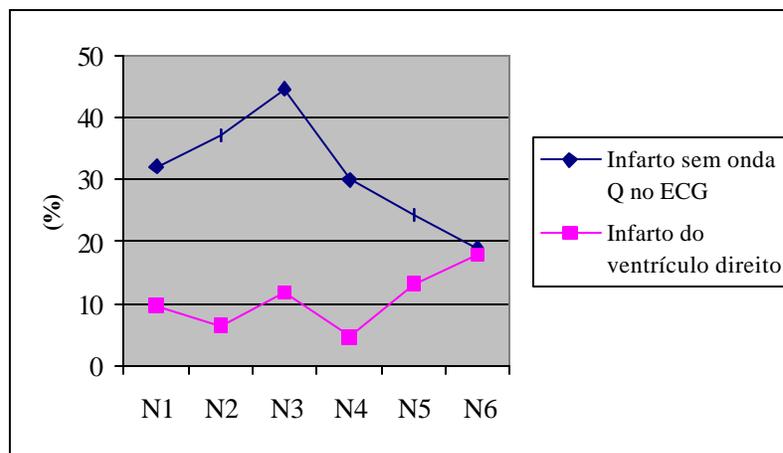


Figura 03: Infarto sem onda Q no ECG e Infarto do ventrículo direito

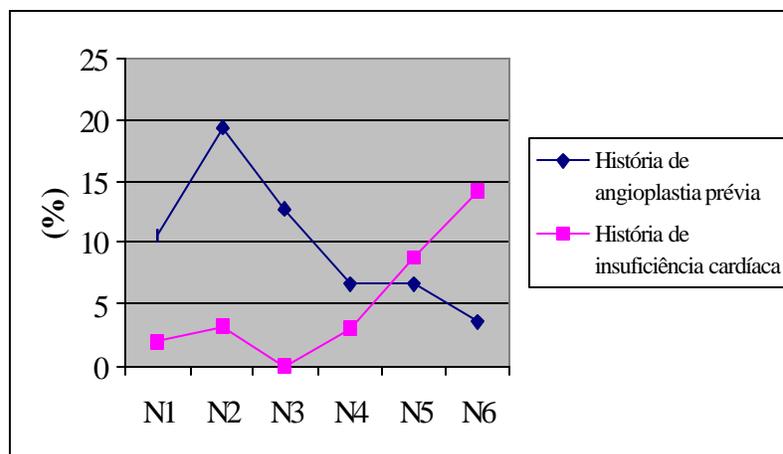


Figura 04: História de angioplastia prévia e História de insuficiência cardíaca

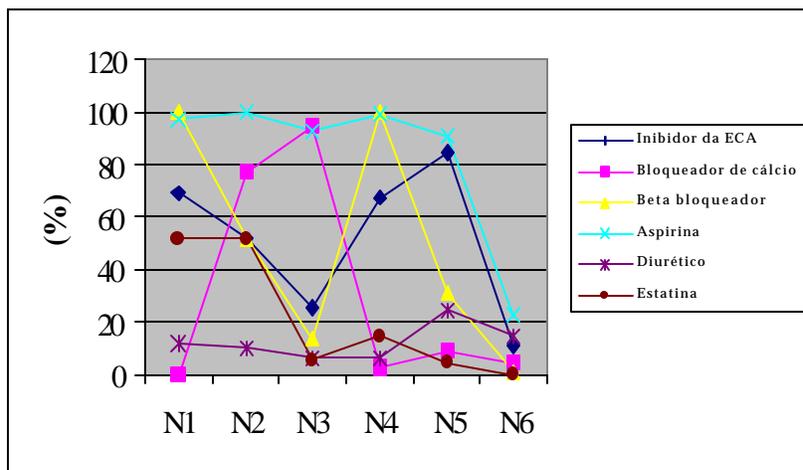


Figura 05: Remédios na alta do hospital

4 – Comentários finais

A Análise de Correspondência mostrou-se útil no sentido de identificar fatores associados estatisticamente ao óbito no IAM, reduzindo em 85% o número de variáveis a serem incluídas na rede neural artificial.

Através da Rede Neural conseguiu-se agrupar todos os óbitos em um mesmo neurônio, classificando-se portanto o perfil de um paciente com IAM que vai a óbito, ou seja, neste neurônio observou-se a existência de (ver figuras de 02 a 05):

- **Menor incidência** de angioplastia prévia, IAM sem onda Q e uso de medicamentos úteis no tratamento do IAM;
- **Maior incidência** de insuficiência cardíaca prévia, IAM de ventrículo direito, complicações hemodinâmicas e arritmias.

Atenta-se que este trabalho atingiu os objetivos de identificação dos fatores relacionados com a variável óbito e identificação de padrões dos pacientes com infarto agudo do miocárdio. Trabalhos estão sendo desenvolvidos a partir da ampliação do banco de dados (variáveis e pacientes). A partir dos resultados até então encontrados, almeja-se futuramente desenvolver um sistema de cálculo da probabilidade de óbito o que será de extrema valia para os cardiologistas.

Referências:

- [1] Tunstall-Pedoe H.; Kuulasmaa K.; Amouyel P.; Arveiler D.; Rajakangas A-M.; Pajak A.. Myocardial infarction and coronary deaths in the World Health Organisation MONICA project. Registration procedures, event rates, and case-fatality rates in 38 populations from 21 countries in four continents. (WHO MONICA project). *Circulation* 90:583-612, 1994.
- [2] Johnson, R. A. & Wichern, D. W.. *Applied multivariate statistical analysis*. 4^a ed. USA: Prentice Hall. 1998.

[3] Kohonen, T. *Self-organizing maps*. Springer Verlag. 1995

[4] Kaski, S.; Kohonen, T.. *Exploratory data analysis by the self-organizing map: structures of welfare and poverty in the world*, In: *Neural Networks in the Capital Markets*, World Scientific, 1996.