

# Clusterização Bayesiana para classificação de risco de pacientes

Haroldo G. B. Filho<sup>1</sup>, Márcia Ito<sup>2</sup>

<sup>1</sup>Laboratório de Sistemas Inteligentes  
Universidade Federal do Maranhão (UFMA) - São Luís, MA – Brasil

<sup>2</sup>IBM Research  
São Paulo, SP – Brasil

haroldogomes86@gmail.com, marciaito@br.ibm.com

**Abstract.** *A classification of risk groups is characterized as a significant step for decision making in a multidisciplinary health team, however in some cases the classification becomes erroneous and time consuming. This article proposes a model of Bayesian classification to optimize such processes across clusters.*

**Resumo.** *Uma classificação de grupos de risco caracteriza-se como uma etapa significativa para tomadas de decisão de uma equipe multiprofissional em saúde, porém em alguns casos a classificação torna-se errônea e demorada. Este artigo propõe um modelo de classificação bayesiana visando otimizar tais processos através de clusters.*

## 1. Introdução

A Classificação de Risco pode ser compreendida como um processo dinâmico para auxílio de tomada de decisão de uma equipe multiprofissional de saúde, com o objetivo central em realizar uma triagem ou seleção de um cenário clínico de um paciente, baseado em sua complexidade. Ao longo da história, sempre houve métodos de eleger um objeto e sua relevância, alguns baseados meramente na anamnese do paciente e outros em riscos epidêmicos, na maioria dos casos cada país possui seu método de triagem, porém tais formas suscitam precedentes externos e superficiais. Baseado nisso, em 1997 padronizou-se o Protocolo de Classificação de Riscos de Manchester (MTS, do inglês *Manchester Triage System*), onde propõe-se que cada grau progressivo de um cenário seja codificado por uma cor em relação ao tempo de resposta pela equipe multiprofissional. Basicamente, divide-se em quatro níveis: Vermelho, caracteriza-se por um estado emergencial e necessita de atendimento prioritário; Amarelo, o tempo de espera do paciente deve ser no máximo 15 minutos; Verde, o caso exige uma urgência mediana podendo esperar em até 30 minutos; e o Azul, que é utilizado para casos menos prioritários podendo aguardar em média de uma a três horas. É comum tal classificação ser tangível através de adesivos ou pulseiras.

Ainda assim, para [Gilboy 2005] a tomada de decisão na “escolha” do paciente a ser atendido ou a classificação de riscos e prioridades, torna-se uma incumbência de alta complexidade, visto que, envolvem alguns domínios, subdomínios e variáveis, tais como: Qual a Doença e intensidade para com o paciente? Quais Causas? Que tratamento se adequa a um menor prazo? E a realização de uma anamnese, sendo um

grande desafio e vicissitude tal consonância, visto a urgência e heterogeneidade de cada caso. Outro agravante se dá através de decisões unilaterais e exclusivamente baseadas em fluxogramas padronizados pelo MTS, para [Merhy 1997] um fluxograma, consiste num diagrama, em que se desenha um certo modo de organizar os processos de trabalho que se vinculam entre si e em torno de uma certa cadeia de produção, porém não há como englobar  $N$  variáveis em um caso e nem entender questões subjetivas e históricas do paciente em um curto espaço de tempo, implicando em falhas.

Diante disto, o presente trabalho objetiva a implementação de *Clusters Bayesianos* para especificação de domínios e subdomínios através de um axioma de classificação ótima, no qual se realiza simulações em uma população de 100 pacientes e estabelecimento de critérios e pesos de acordo com o domínio, auxiliando na tomada de decisão com minimização de erros, resultando é uma amostragem probabilística relacionada ao tempo de atendimento e sua real classificação de risco.

## 2. Redes de Bayes para classificação

Segundo [Dutra 2002], Redes bayesianas são grafos acíclicos dirigidos que representam dependências entre variáveis em um modelo probabilístico. Esta abordagem representa uma boa estratégia para lidar com problemas que tratam incertezas, onde conclusões não podem ser construídas apenas do conhecimento prévio a respeito do problema. Uma vez que, há mínima confiabilidade do agravamento dos sintomas, da incerteza da causa da enfermidade e da anamnese superficial, em um atendimento urgente, propõe-se uma simulação para testar casos e seus respectivos riscos, utilizando um algoritmo denominado Classificador Bayesiano Ótimo, representado pelo axioma em (1):

$$\frac{\operatorname{argmax}}{V_j \in V} \sum_{h_i \in H} P(V_j / h_i) P(h_i / D) \quad (1)$$

Onde aplicando em um caso simulado entende-se que: *argmax*, é o número de critérios utilizados no Cenário(Doenças, Anamnese, Causas, Tratamentos) em relação a  $V_j \in V$  que é a distância entre a menor e o maior variável(temperatura do corpo, por exemplo);  $\sum$  representa a população atendida;  $D$  simboliza a incerteza máxima(caracterizando-se como *pesos* em uma situação),  $P$  o número de interações testadas,  $h_i \in H$ , representam o tempo mínimo e máximo na situação. A simulação é baseada no protocolo formalizado pela MTS (*disponível em: <http://goo.gl/waK7nr>*), onde em testes realizados encontram-se na *subseção 2.1* deste presente artigo.

### 2.1. Clusterização Bayesiana

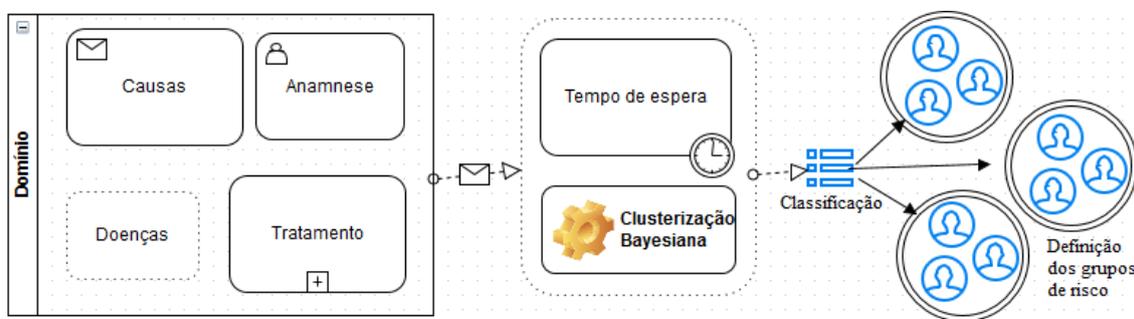
Independente da aplicação, *Clusters* são “agregações” de objetos, dados ou padrões, para determinação e especificação de domínios e subdomínios, com o objetivo de otimizar as tomadas de decisão por etapas ou estágios. Para [Jain et al., 1999] clusterização é a classificação não-supervisionada de dados, formando agrupamentos ou clusters. Ela representa uma das principais etapas de processos de análise de dados, denominada análise de *Clusters*. Na presente simulação, a Clusterização tem como objetivo “dividir” grupos através da classificação de seu respectivo risco, no auxílio a decisão da equipe multiprofissional em saúde, sendo uma resultante dos subdomínios de

risco encontrada através do Classificador Bayesiano Ótimo(ver tabela 1), em um teste proposto temos:

**Tabela 1. Simulação de interações em um casos**

Risco	P	$\Sigma$	D	hi	H(min)	Vj	V	argmax
Vermelho	1...100	100	0...10	0	0	39	45	1...5
Amarelo	1...100	100	0...10	0	15	33	38	1...5
Verde	1...100	100	0...10	0	30	30	32	1...5
Azul	1...100	100	0...10	0	180	-	-	1...5

Sendo assim, propõe-se uma Clusterização para os grupos de riscos através de subdomínios(ver figura 1):



**Figura 1. Clusterização Bayesiana para definição de grupos de risco**

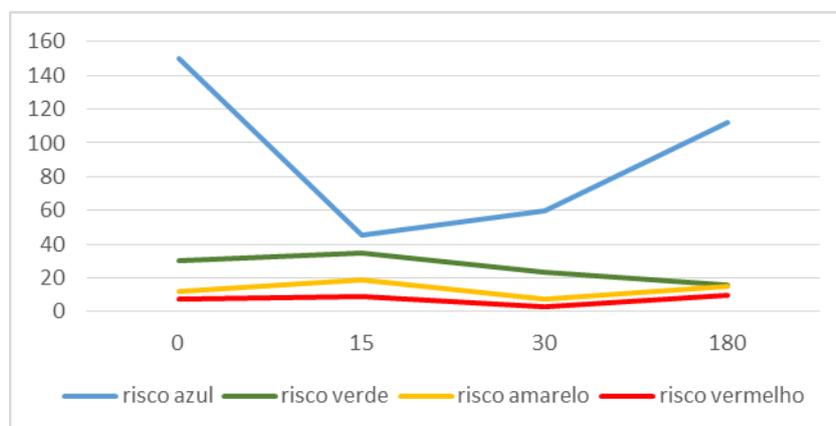
A Classificação para atendimento prioritário baseado na definição dos grupos de riscos por clusterização, mostra-se uma técnica viável para otimização de processos e tomadas de decisão visto a probabilidade de incertezas, se faz necessário entender a complexidade de um domínio e sua população, para uma resposta coesa e apressurada.

### 3. Resultados e discussões

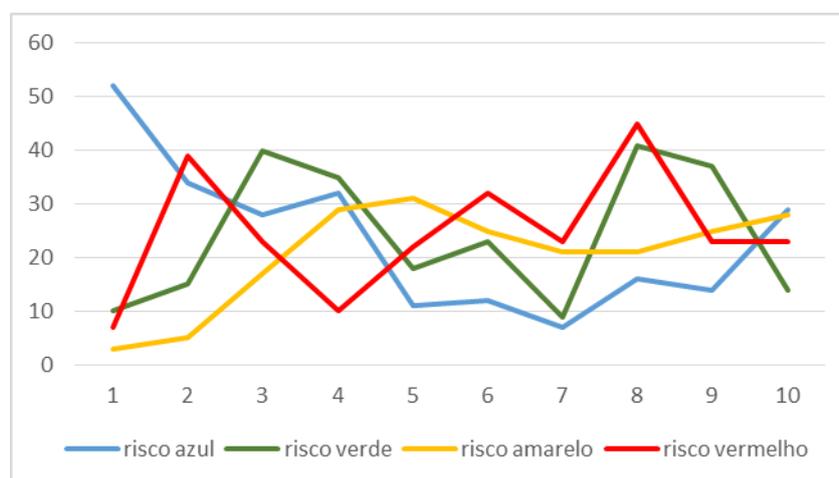
O Teste aplicou 100 interações( $P$ ) em uma população( $\Sigma$ ) de 100 pacientes, nos quatro níveis de risco, utilizando uma técnica de clusterização afim de validar os dados simulados, o primeiro teste apresentou resultados satisfatórios em relação ao tempo de resposta do atendimento em urgência padronizado no protocolo da MTS de acordo com cada nível de risco. Um segundo teste é aplicado para avaliação da coerência da classificação(se realmente um paciente  $x$ , que de forma empírica foi classificado com risco alto, realmente está em seu grupo de risco correto) baseado na relação entre o peso( $D$ ) em uma situação e seu critério de aplicação. A Validação dos resultados oriundos da utilização de técnicas bayesianas é frizada por [Jain et al 1999] onde diz que para validar a saída produzida por um processo de clusterização, geralmente se recorre a critérios de otimalidade, muitas vezes definidos de forma subjetiva.

Os resultados encontrados e validados pelo classificador podem ser representados através de dois gráficos, que demonstram cada população total em relação ao seu de risco, o gráfico 1(ver gráfico 1) apresenta os resultados do tempo otimizado pelo algoritmo classificador em função ao tempo determinado pela MTS, o gráfico 2(ver gráfico 2) apresenta a coerência da classificação utilizando o peso e em função da

resultante das variáveis em um domínio, ambos em 100 interações testadas e em uma população de 100 pacientes.



**Gráfico 1. Mensuração de resultados o tempo de atendimento padronizado pela MTS pelo algoritmo classificador Bayesiano.**



**Gráfico 2. Número de acertos na quantidade de pacientes com situação coerente a seu grupo de risco.**

O Teste 1 apresentou uma significativa resposta da classificação bayesiana em relação ao tempo exigido pelo protocolo da MTS, demonstrando um desempenho otimizado nos riscos azul, verde e amarelo, apresentando alguma perda de performance no risco vermelho, onde a simulação deu tempo de resposta em até dez minutos. O Teste 2 apresentou grande irregularidade na mensuração da coerência da classificação dos grupos de risco, inferindo-se a eficácia de uma técnica bayesiana, visto que a classificação de risco meramente por fluxogramas e protocolos, levam a uma gama de tomadas de decisão equivocadas. .

#### 4. Referências

Gilboy, N. (2005) Emergency Severity Index, Version 4: “Implementation Handbook. Agency for Healthcare Research and Quality”.

Merphy, E.E.. (Orgs.); Agir em Saúde: um desafio para o público; São Paulo, Hucitec, 1997.

Jain, A.K., Murty, M.N. & Flynn,(1999) P.J. “Data Clustering: A Review”, vol. 31 p.264