





## Artificial neural networks applied to analysis of professional profiles

### **Abstract**

This work is aimed at exploring the utilization of artificial self organizing neural networks, Kohonen's model, for classification of professional profiles, with the finality of assisting the process of triaging candidates in Human Resources (HR) areas. The practical result of this implementation may be viable to areas in HR, knowing that the neural networks have the ability to get along and adapt rapidly to imprecise data, incomplete or totally new information.

### **Key words**

*Kohonen. Neural networks. Professional profile.*

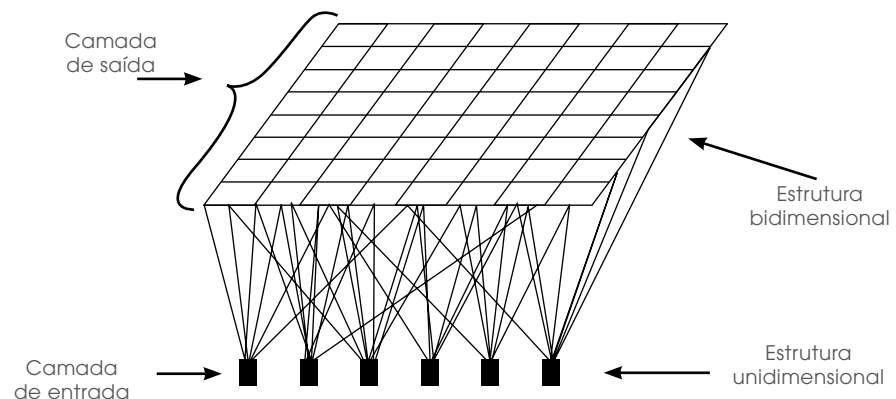


Atualmente, as redes neurais artificiais são utilizadas nas mais diversas áreas do conhecimento, tais como Engenharia, Economia, Agronomia e Recursos Humanos (RH), geralmente em problemas de reconhecimento e classificação de padrões. Há diversos modelos de RNAs, e cada um desses modelos é mais adequado para resolução de uma determinada classe de problemas. Algumas características que variam entre os modelos são: algoritmo de aprendizado, arquitetura da rede e a forma como são feitas as conexões entre as camadas.

O modelo explorado neste trabalho é um tipo de rede neural de aprendizado não supervisionado, conhecido como modelo de Kohonen, ou mapa auto-organizável – detalhado no item 1 –, que será implementado nesta pesquisa para classificar diferentes perfis profissionais, visando auxiliar a área de RH de empresas, na atividade de triagem de currículos.

## 1. Modelo de Kohonen

O modelo de Kohonen, cujo aprendizado é do tipo não supervisionado, possui apenas duas camadas (entrada e saída). Esse tipo de rede pode apresentar diferentes formatos geométricos em sua camada de saída (CABRAL JUNIOR; TIMOSZCZUK, 2003; HAYKIN, 2001); no entanto, o formato mais utilizado é o retangular. A Figura 2 ilustra um exemplo de arquitetura do modelo.



**Figura 2 – Exemplo de arquitetura do modelo de Kohonen.**

Fonte: Elaboração própria.

A rede de Kohonen é caracterizada pela formação de um mapa topográfico dos padrões de entrada, no qual as localizações espaciais dos neurônios da grade de saída são indicativas das características estatísticas intrínsecas contidas nos padrões de entrada, o que justifica o nome ‘mapa auto-organizável’ (KOHONEN, 1990). Esse modelo de rede neural foi inspirado no córtex sensitivo, em que regiões vizinhas respondem por partes sensitivas localizadas próximas do corpo humano (RITTER; KOHONEN, 1989). Como suas principais características, podemos citar:

- 1) Competição – para cada padrão de entrada apresentado à rede, os neurônios da camada de saída competem entre si para decidir quem será o vencedor. Esse neurônio vencedor representará o conjunto de características descrito pelo padrão de entrada em questão;
- 2) Cooperação – o neurônio vencedor determina a localização espacial de uma vizinhança topológica de neurônios que devem ser excitados, fornecendo, assim, a base para a cooperação entre os vizinhos;
- 3) Generalização – capacidade de a rede neural fornecer respostas adequadas a um padrão de entrada que não foi aprendido previamente.

### 1.1. Treinamento do modelo de Kohonen

Na fase de treinamento, os neurônios da camada de saída competem entre si para que haja um vencedor a cada nova entrada do conjunto de treinamento, ou seja, sempre que é apresentado um padrão do conjunto de treinamento à rede neural, existe uma competição entre os neurônios da camada de saída para representar a entrada naquele momento (BRAGA, 2000). Uma particularidade do modelo de Kohonen é que os pesos dos neurônios vizinhos do vencedor também são atualizados, configurando o processo de cooperação. No início do treinamento, as vizinhanças de atualização são grandes e podem abranger toda a grade de saída. Durante o treinamento, a redução do tamanho da vizinhança favorece uma seletividade entre os padrões de entrada e a formação de regiões de maior atividade em torno do neurônio vencedor. Esse é o processo responsável pela formação de mapas que conservam a topologia (KOHONEN, 1990). O sucesso no processo de aprendizagem da rede depende não somente do número de iterações, mas também de uma melhor representação dos padrões de entrada e dos parâmetros do algoritmo, tais como a taxa de aprendizagem e a função de vizinhança. É importante ressaltar que esses parâmetros são obtidos de forma empírica.

Para um melhor entendimento, mostramos, a seguir, os seis passos necessários à implementação do algoritmo de treinamento do modelo de Kohonen (CABRAL; TIMOSZCZUK, 2003; HAYKIN, 2001):

- 1) Antes de qualquer operação na rede, inicializam-se todos os pesos ( $\vec{w}_{ij}$ ) das sinapses com valores aleatórios pequenos em relação aos valores de entrada ( $\vec{x}_j$ );
- 2) Apresenta-se uma nova entrada ( $\vec{x}_j$ ) (do conjunto de treinamento);
- 3) Calculam-se as distâncias entre os neurônios da camada competitiva e os da camada de entrada. Para cada neurônio da camada competitiva, calcular a distância  $d_i$  entre o neurônio  $\vec{w}_{ij}$  da camada competitiva e cada neurônio  $\vec{x}_j$  da camada de entrada.

$$d_i(t) = \sqrt{\sum_{j=1}^J (x_j(t) - w_{ij}(t))^2}$$

onde:

$j = 1, \dots, J$  – Neurônios de entrada

$i = 1, \dots, N$  – Neurônios de saída

$x_j(t)$  = valor do neurônio de entrada  $j$  no tempo  $t$ .

$w_{ij}(t)$  = peso sináptico no tempo  $t$ .

- 4) Seleciona-se o neurônio vencedor: seleção do neurônio  $\vec{w}_{ij}$ , cuja distância euclidiana  $d_i$  seja a menor possível, ou seleção do neurônio  $i$ , cujo vetor de pesos sinápticos  $\vec{w}_{ij}$  é o mais parecido com o vetor de entrada  $\vec{x}_j$ , sendo:

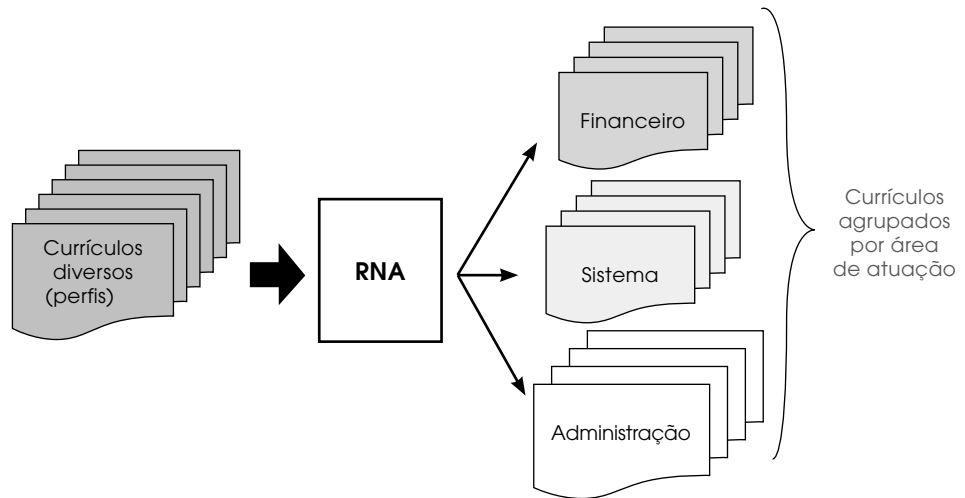
$$d_{i^*} < d_i \quad i = 1, \dots, N; i \neq i^*$$

- 5) Atualizam-se, no próprio neurônio vencedor, seus vetores de pesos e os de sua vizinhança (numa área que abrange  $\Delta(t)$  neurônios à sua volta), obedecendo ao seguinte cálculo:

$$w_{ij}(t + 1) = w_{ij}(t) + n(t) \cdot (x_j(t) - w_{ij}(t))$$



cadastrada, ou o restringem a apenas uma delas. Por esse motivo, resolvemos implementar um algoritmo de RNA na tentativa de minimizar esse tipo de problema. Essa implementação tem como objetivo auxiliar a área de RH na tarefa de classificação e triagem de currículos. A idéia de como deve ocorrer essa classificação é ilustrada na Figura 3.



**Figura 3 - Classificação de currículos por área de atuação.**

Fonte: Elaboração própria.

### 3. Modelagem

Para que as informações referentes aos perfis profissionais pudessem servir de entrada para treinamento da RNA, realizamos os seguintes procedimentos:

Escolhemos 15 perfis profissionais entre as áreas administrativas de saúde, tecnologia, direito e letras. Esses perfis foram analisados e tiveram seus atributos quantificados para compor o conjunto de treinamento da rede. Após diversos experimentos e com base nos estudos de Araújo (2002), descartamos os atributos de menor desvio padrão e selecionamos os sete principais de maior desvio padrão, fundamentais para definir a função de um profissional. Esses atributos foram valorados com pesos diferentes, de acordo com seu grau de importância. Fizemos ainda uma normalização dos valores de entrada, deixando-os na faixa entre 0 e 1, por se tratar de uma rotina de praxe em RNAs.





## 4. Implementação e parametrização da rede

A rede neural para classificação dos perfis profissionais foi implementada utilizando a linguagem de programação Delphi 5.0. Desenvolveu-se uma aplicação genérica para classificação de padrões, aqui utilizada para reconhecimento e classificação de perfis profissionais. Essa aplicação é composta, basicamente, de três módulos: 1. Módulo de parametrização da rede neural, 2. Módulo de aquisição dos padrões de treinamento e 3. Módulo de treinamento e visualização do mapa de saída.

Nesta aplicação, utilizamos os seguintes parâmetros:

- 1) Arquitetura da rede: sete neurônios na camada de entrada e 169 neurônios na camada de saída, dispostos numa grade retangular (13 x 13);
- 2) Taxa de aprendizagem:  $\eta(t) = 0,2$ ;
- 3) Equação de aprendizagem: define a taxa real de aprendizagem do vetor de pesos do neurônio, de acordo com sua posição na grade de saída. Quanto mais longe do vencedor, menor é a taxa. Usamos as seguintes equações:

$$EQ1 = \frac{n(t)}{1 + \left(\frac{\Delta}{5}\right)}$$

$$EQ2 = n(t) \cdot \left(0,9 + \left(\frac{1}{2^{\Delta}}\right) \cdot \left(\frac{1}{10}\right)\right), \text{ onde:}$$

$n(t)$  = Valor da taxa de aprendizagem utilizada no treinamento.

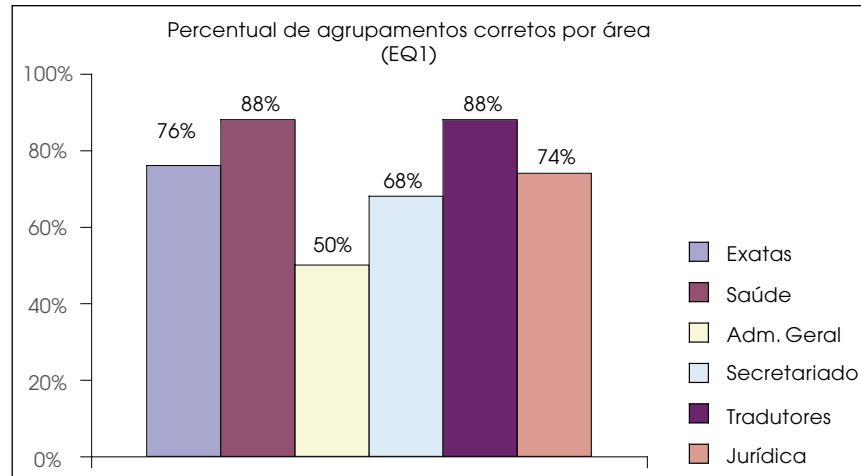
$\Delta$  = Raio da vizinhança a ser atualizado.

- Quantidade de iterações: 6 mil.

## 5. Resultados obtidos

Depois de incluído o conjunto de padrões e parametrizada a rede, inicia-se o processo de treinamento. Nesta fase, os neurônios da camada de saída devem competir entre si, representando o neurônio vencedor de cada iteração um perfil profissional do conjunto de treinamento. Perfis



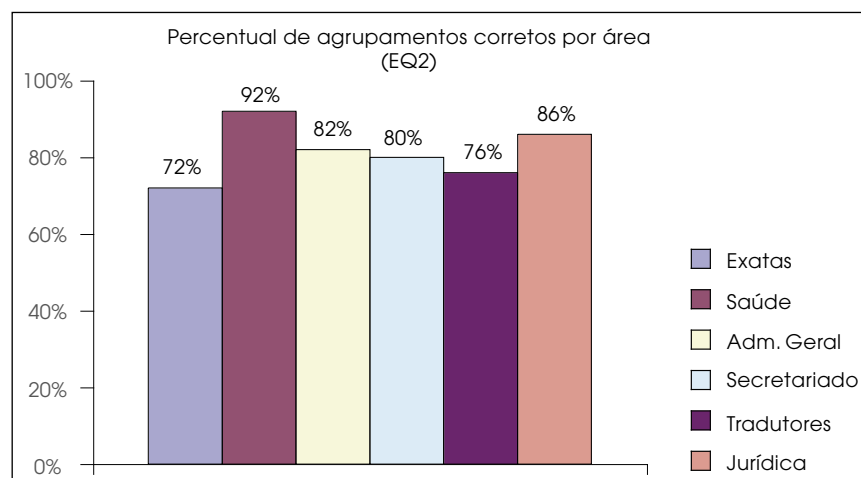


**Figura 5** – Gráfico dos percentuais de agrupamentos por área de atuação em 50 treinamentos, considerando a equação de aprendizagem EQ1.

Fonte: Elaboração própria.

Observa-se, na da Figura 5, que a maior parte dos agrupamentos ficou abaixo de 80%. Assim, consideramos insatisfatório esse resultado e concluímos que a equação para atualização de pesos EQ1, utilizada com os demais parâmetros da rede, mostrou-se ineficiente.

Procedemos, então, a mais 50 treinamentos, utilizando a 2<sup>a</sup> equação de aprendizagem (EQ2), mantendo-se inalterados os demais parâmetros da rede, e analisamos a quantidade de agrupamentos corretos, na camada de saída, em cada um dos treinamentos. A Figura 6 ilustra, de forma gráfica, um resumo estatístico desses treinamentos.



**Figura 6** – Gráfico dos percentuais de agrupamentos por área de atuação em 50 treinamentos, considerando a equação de aprendizagem EQ2.

Fonte: Elaboração própria.



