

Avaliando o Desempenho de uma Rede Neural Probabilística Baseada em Centroídes *

Patrick Marques Ciarelli¹, Elias Oliveira²

¹ Departamento de Engenharia Elétrica

² Departamento de Ciência da Informação
Universidade Federal do Espírito Santo (UFES)
Av. Fernando Ferrari s/n
29060-970 – Vitória – ES – Brazil

{pciarelli,elias}@lcad.inf.ufes.br

Abstract. *In this article is proposed a technique which uses centroids together with Probabilistic Neural Network to minimize some disadvantages of this net, such as the storage space for the neural network weights and linear time complexity order with the number of training samples. In the experiments carry out the memory usage and classification time were drastically reduced. Besides, the quality of the results was also considering improved by the a priori probability, when using it with theses centroids.*

Resumo. *Neste artigo é proposta uma técnica de centroíde junto com Rede Neural Probabilística para minimizar algumas das desvantagens desta rede, tais como o espaço de estocagem para os pesos da rede neural e a complexidade de tempo linear de acordo com o número de amostras de treinamento. Nos experimentos realizados, o uso da memória e o tempo de classificação foram grandemente reduzidos. Além disso, a qualidade dos resultados foi também melhorada pela probabilidade a priori quando usando-a com estes centroídes.*

1. Introdução

Atualmente encontramos na literatura um grande esforço, na área de Recuperação de Informação, basicamente focado na classificação e categorização de documentos textuais de conteúdos genéricos. Todavia, nos últimos anos uma atenção maior tem sido prestada à problemas de classificação de textos multi-rotulados, com aplicações práticas em diversas áreas [Li et al. 2007, Souza et al. 2008]. Junto com essas propostas são sugeridas, na literatura, diversas métricas e algoritmos para a solução dos mesmos [Zhang and Zhou 2007]. Entre estas técnicas uma que apresentou resultados relevantes foi uma versão modificada da Rede Neural Probabilística (RNP) proposta em [Oliveira et al. 2008].

Esta rede neural apresenta as vantagens de possuir uma implementação simples, necessitar de poucos parâmetros para ser configurada e possuir um treinamento rápido quando comparado ao de outras técnicas, como por exemplo máquinas de suporte vetorial e redes neurais *backpropagation*. Além disso, ela apresentou resultados relevantes quando aplicado a uma base de dados multi-rotulada com grande quantidade de categorias [Oliveira et al. 2008]. No entanto, a RNP apresenta algumas desvantagens que podem

*Este trabalho foi financiado pela Receita Federal do Brasil e pela FAPES (bolsa 41936450/2008).

ser fortemente realçadas para problemas de classificação com grandes bases de dados: necessidade de armazenamento das amostras de treinamento, desempenho reduzido na presença de amostras redundantes (menor eficácia) e tempo de classificação diretamente relacionado ao número de amostras de treinamento (menor eficiência).

Este artigo propõe uma combinação da RNP com uma técnica de centroíde que armazena a probabilidade *a priori* das categorias de forma a minimizar as desvantagens desta rede neural. Para avaliação da técnica proposta foram realizados experimentos sobre 11 bases de dados multi-rotuladas do domínio *yahoo.com* comparando os resultados advindos de uma RNP básica e a RNP com centroíde.

2. Os Algoritmos

Rede Neural Probabilística

A Rede Neural Probabilística usada neste trabalho foi a proposta em [Oliveira et al. 2008] e ela é composta de três camadas: a camada de entrada, a camada de padrões e a camada de soma. Esta rede neural precisa somente de um passo de treinamento. O treinamento consiste em associar cada amostra de treinamento w_i da categoria C_i para um neurônio da camada de padrões da categoria C_i . Dessa forma, o vetor de pesos deste neurônio é o próprio vetor de características da amostra.

Na fase de classificação é apresentada uma amostra de teste d_j à camada de entrada. Esta camada não realiza nenhuma computação, ela simplesmente passa a amostra d_j para os neurônios da camada de padrões, onde cada neurônio calcula a saída para d_j . O cálculo é mostrado na Equação 1.

$$F_{k,i}(d_j) = \frac{1}{2\pi\sigma^2} \exp\left(\frac{d_j^t w_{k,i} - 1}{\sigma^2}\right), \quad (1)$$

onde d_j é o vetor de entrada com as características do padrão, $w_{k,i}$ é a k -ésima amostra para um neurônio da categoria C_i , na qual $k = 1, \dots, N_i$, considerando que N_i é o número de neurônios de C_i . Em adicional, d_j e $w_{k,i}$ foram normalizados, tal que $d_j^t d_j = 1$ e $w_{k,i}^t w_{k,i} = 1$. O σ é o desvio padrão da Gaussiana.

O próximo passo é a camada de soma. Nesta camada são somadas as saídas dos neurônios da camada de padrões de cada categoria C_i , produzindo o valor $p_i(d_j)$ conforme a Equação 2, onde $|C|$ é o número total de categorias e h_i é a probabilidade *a priori* da categoria C_i . Se forem consideradas as probabilidades *a priori* a partir da base de treinamento, então a fração $\frac{h_i}{N_i}$ pode ser desconsiderada.

$$p_i(d_j) = \frac{h_i}{N_i} \sum_{k=1}^{N_i} F_{k,i}(d_j), \quad i = 1, 2, \dots, |C| \quad (2)$$

Finalmente, para a seleção das categorias que serão associadas pela rede neural para cada amostra de teste, consideramos aquelas que apresentam na camada de soma um valor maior do que um limiar pré-estabelecido.

A RNP proposta precisa de poucos parâmetros para ser configurada: o σ , (ver na Equação 1) e a determinação do valor do limiar. Outras vantagens das RNPs é que é fácil adicionar novas categorias, ou novas amostras de treinamento dentro de uma estrutura em

funcionamento, o que é bom para aplicações *on-line* [Duda et al. 2001]. Por outro lado, uma das desvantagens dela é o grande número de neurônios na camada de padrões quando existem uma grande quantidade de amostras de treino.

Rede Neural Probabilística com Centróide

Para minimizar a quantidade de neurônios da camada de padrões da RNP será usada uma técnica de centróide, de forma a se obter uma única amostra de treinamento por categoria. A Equação 3 ilustra o procedimento matemático para se obter o centróide para a categoria C_i , onde $w_{k,i}$ é a k -ésima amostra de treinamento da categoria C_i , N_i é o número de amostras de C_i e W_i é o centróide obtido. Para se evitar perda de informação também será obtida da base de treinamento original a probabilidade *a priori* (h_i) de cada categoria. Com esta etapa realizada, a fração $\frac{h_i}{N_i}$ da Equação 2 será reduzida a h_i , pois N_i será igual a 1 e o centróide W_i é associado ao neurônio da camada de padrões da categoria C_i .

$$W_i = \frac{1}{N_i} \sum_{k=1}^{N_i} w_{k,i}, \quad h_i = N_i \quad i = 1, 2, \dots, |C| \quad (3)$$

3. Experimentos

Em nossos experimentos usamos um conjunto de bases de dados extraídos de páginas da web do domínio *yahoo.com* [Zhang and Zhou 2007]. Este conjunto é composto por 11 bases de dados, todas multi-rotuladas, contendo em média 30 categorias cada base de dados. Todas estas bases são compostas por uma parte de treinamento e de teste com 2000 e 3000 documentos, respectivamente.

Para calibrar a RNP foram testadas três grandezas de σ^2 sobre a parte de treinamento da base de dados *Arts*: 10, 1 e 0,1. O melhor parâmetro foi 0,1. Para o limiar da rede neural foi fixado o valor de 0,5. Estes parâmetros foram usados na RNP (*RNP*), RNP com centróide e probabilidade *a priori* (*RNP_{CP}*) e, para comparação com a técnica proposta, RNP com centróide e sem a probabilidade *a priori* (*RNP_C*).

Os resultados foram analisados utilizando as métricas *Average Precision*, *One Error*, *Coverage* e *Ranking Loss* definidas em [Zhang and Zhou 2007]. As métricas *Average Precision*, *One Error* e *Ranking Loss* são definidas no intervalo entre 0 e 1. A métrica *Coverage* possui limite inferior igual a 0 e limite superior igual a $|C| - 1$. Quanto menor o valor encontrado para as métricas *One Error*, *Coverage* e *Ranking Loss* melhor o resultado, e quanto maior o *Average Precision* melhor.

As Figuras 1 (a) a (d) mostram que de forma geral a *RNP_{CP}* foi superior a *RNP_C* e com um desempenho ligeiramente superior a *RNP*. Para uma avaliação mais clara foi utilizado um teste t bicaudal pareado com 95% de intervalo de confiança e tal procedimento indicou que a *RNP_{CP}* foi estatisticamente superior aos outros algoritmos em todas as métricas. De forma semelhante a *RNP* foi superior a *RNP_C* em todas as métricas. Estes testes indicam a importância de manter a probabilidade *a priori* das categorias.

Outra verificação importante é que em ambas as versões da RNP utilizando o centróide (*RNP_{CP}* e *RNP_C*) o tempo gasto na fase de teste foi mais de 10 vezes menor em todas as bases de dados do que para a *RNP*. Além disso, o armazenamento dos pesos da RNP reduziu para cada base de dados de 2000 para aproximadamente 30 amostras,

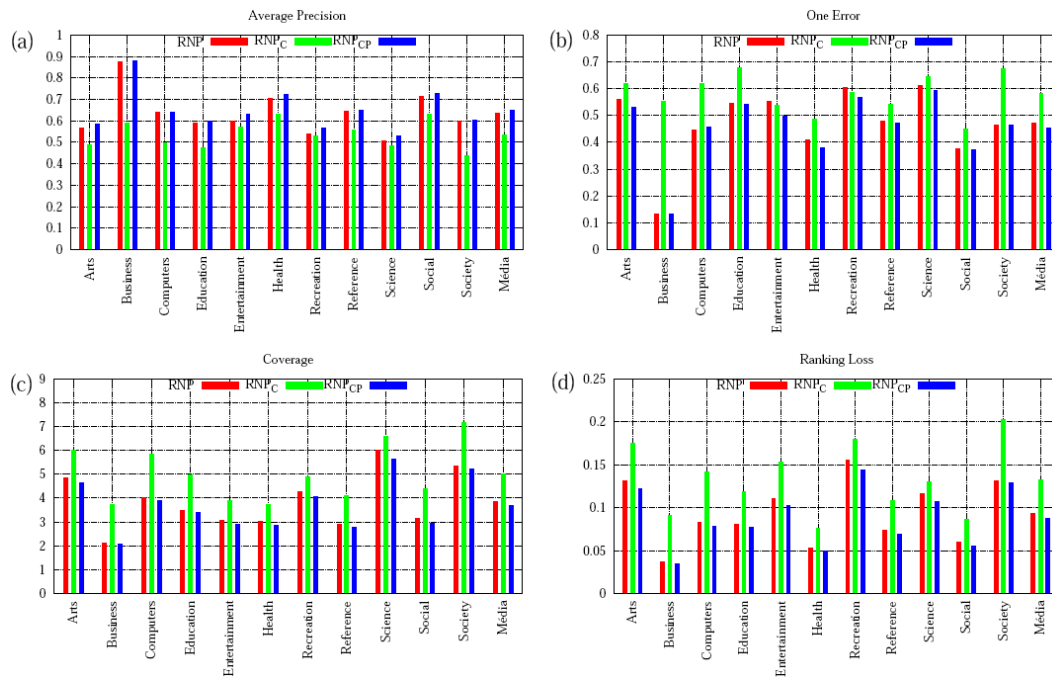


Figura 1. Resultados experimentais em termos do (a) *Average Precision*, (b) *One Error*, (c) *Coverage* e (d) *Ranking Loss*.

mais de 60 vezes menor. Tais ganhos podem ser cruciais quando se está trabalhando com bases de dados de grandes dimensões.

4. Conclusões

Este artigo apresentou uma abordagem simples baseada em centróides para minimizar algumas desvantagens da Rede Neural Probabilística. Em experimentos utilizando uma versão de centróide que armazena a probabilidade *a priori* em conjunto com a Rede Neural Probabilística foram obtidos resultados ligeiramente superiores a uma versão sem centróide além de ser obtido em um tempo menor de classificação.

Referências

- Duda, R. O., Hart, P. E., and Stork, D. G. (2001). *Pattern Classification*. Wiley-Interscience, New York, 2 edition.
- Li, X., Chen, H., Zhang, Z., and Li, J. (2007). Automatic Patent Classification using Citation Network Information: an Experimental Study in Nanotechnology. In *JCDL '07: Proceedings of the 2007 conference on Digital libraries*, pages 419–427, New York, NY, USA. ACM.
- Oliveira, E., Ciarelli, P. M., Souza, A. F. D., and Badue, C. (2008). Using a probabilistic neural network for a large multi-label problem. *Simpósio Brasileiro de Redes Neurais*.
- Souza, A. F. D., Pedroni, F., Oliveira, E., Ciarelli, P. M., Henrique, W. F., Veronese, L., and Badue, C. (2008). Automated multi-label text categorization with vg-ram weightless neural networks. *Neurocomputing*.
- Zhang, M.-L. and Zhou, Z.-H. (2007). ML-KNN: A Lazy Learning Approach to Multi-Label Learning. *Pattern Recogn.*, 40(7):2038–2048.