

Comparação entre Redes Neurais Artificiais e Rough Sets para Classificação de Dados

Renato C. Vieira, Marcelo B. Tenório, Mauro Roisenberg, Paulo S. S. Borges

INE – Departamento de Informática e Estatística

UFSC – Universidade Federal de Santa Catarina

Campus Universitário – Trindade – 88040-900 – Florianópolis – SC

{renatocv, marcelot, mauro, pssb}@inf.ufsc.br

Abstract

Sometimes the data bases are grounds to the application of Artificial Intelligence (AI) techniques that can be use as importants tools to find the solution for the problems of production and extraction of knowledge and development of expert systems. In this way, become valid a work that aim to make a comparison between distincts techniques of treatment and utilization of data bases in the Data Mining Process. This paper will make a comparison of two techniques of AI driven to a medical diagnosis problem. In this paper we will explore both approaches of IA, that is, the Symbolic Artificial Intelligence, through of Rough Sets Theory and the Conexionist Artificial Intelligence, using for this the Artificial Neural Networks (ANN).

1. Introdução

A disposição dos dados em uma base de dados não deve possuir redundância, pois isso além de aumentar a complexidade computacional, oblitera os dados, de tal sorte que a extração de regras de produção torna-se uma tarefa penosa em que se desperdiça tempo e recurso computacional. Neste contexto, surge a Teoria de Rough Sets (TRS) que trata esta redundância, diminuindo a quantidade de dados não relevantes à resolução dos problemas. É importante ressaltar que a TRS depois de descartar tais dados, possui a propriedade de criar regras que serão utilizadas como mecanismos de inferências, desta forma enquadrando a TRS como um ramo da Inteligência Artificial Simbólica.

A Inteligência Artificial Conexionista é sustentada pelas Redes Neurais Artificiais (RNAs), que são amplamente difundidas nos estudos de sistemas especialistas, além de estarem se mostrando como eficientes ferramentas para a classificação de dados no processo de *Data Mining* [1]. As RNAs aprendem por exemplos, os quais podem ser tuplas de uma base de dados e desde que os exemplos tenham sido bem escolhidos, mostram-se poderosas ferramentas de interpolação, isto é, do mesmo modo que uma máquina de inferência, as RNAs também podem classificar

diferentes padrões através da interpolação entre os exemplos apresentados na base de dados.

Este artigo irá inicialmente introduzir a utilização da IA em problemas de classificação de dados, para em seguida descrever brevemente as principais características da Teoria de Rough Sets, então explicar a base de dados em questão, e finalmente apresentar a metodologia envolvida nas implementações e comentar os resultados obtidos como efeito de comparação.

2. Classificação de Dados

No processo de *Data Mining*, a etapa de classificação de dados tem como objetivo classificar atributos comuns de uma base de dados em diferentes classes, sendo o modelo capaz de prever em que classes se enquadram novos exemplos que apresentar-se-ão à base de dados. A classificação de dados vem sendo bastante difundida no que tange as áreas da estatística e de aprendizado de máquina. [2], [3] e [4]. Uma aplicação bastante usual e adequada da classificação de dados é sua utilização em problemas de diagnóstico médico, onde a base de dados consta de variados sintomas (atributos) que serão analisados por um algoritmo de *Data Mining* com fins de levantar um diagnóstico [5] e [6].

As Redes Neurais Artificiais são uma maneira prática para realizar a classificação de dados, pois os métodos baseados nas RNAs suportam diferentes funções de aprendizado que podem ser representadas por atributos contínuos, discretos ou mesmo vetoriais [7].

Convém ressaltar que as RNAs por vezes são consideradas alternativas não tão viáveis para processos de *Data Mining* em base de dados muito extensas, visto que depois de treinadas as RNAs tornam-se geralmente de difícil compreensão e também pelo fato de que muitos dos métodos de treinamento são lentos, principalmente quando são utilizados algoritmos mais comuns que lançam mão da utilização do método de otimização do gradiente descendente para o ajuste dos pesos. Entretanto, outras alternativas como aprendizado não supervisionado utilizando mapas auto-organizáveis [8] e aprendizado on-line [9] também constam na bibliografia.

Certamente para se extrair um conhecimento implícito nos exemplos para gerar padrões classificáveis

em bases de dados é de suma importância determinar quais atributos são relevantes, mesmo sendo esta tarefa complexa. Assim, surge a TRS que tenta representar o conhecimento fazendo uma minimização dos dados redundantes que se encontram obliterados nas bases de dados.

3. Teoria de Rough Sets

Buscando encontrar atributos relevantes seguindo algum critério de relevância, surgem os métodos de seleção de atributos, tendo como principal abordagem a denominada Seleção de um Subconjunto de Atributos, que objetiva encontrar um subconjunto de atributos possivelmente mínimo e ótimo [10].

Um dos métodos para seleção de atributos é a abordagem matemática de Zdzislaw Pawlak [11] chamada de Teoria de Rough Sets. A seleção de atributos é feita através dos chamados redutos que são capazes de manter as mesmas propriedades da representação de conhecimento quando esta é feita utilizando todos os atributos. Em suma, há uma redução de atributos sem alterar a representação de conhecimento.

A TRS primeiramente define as relações de não-discernimento, onde se definem os atributos redundantes ou dispensáveis.

Definição 3.1. Para cada subconjunto de atributos $B \subseteq A$ em um Sistema de Informação $A=(U,A)$, uma relação de equivalência $IND_A(B)$ é associada, chamada de Relação de Não-Discernimento, conforme definição abaixo:

$$IND_A(B) = \{(x,y) \in U^2 \mid \forall a \in B, a(x) = a(y)\}$$

Aliado a isto, nem todos os atributos podem ser necessários para discernir classes, desta forma, com o objetivo de reconhecer quais atributos são realmente importantes para determinada classificação, surgem os chamados redutos.

Definição 3.2. Um Reduto de B é um conjunto de atributos $B' \subseteq B$, tal que todos os atributos $a \in B - B'$ são dispensáveis e $IND(B') = IND(B)$. O termo $RED(B)$ é utilizado para denotar a família de redutos de B .

Assim, pode-se inferir que as informações contidas em um reduto são suficientes para distinguir os objetos de diferentes classes de um sistema de informação [10].

Posteriormente os redutos relativos são transformados em regras através da união dos valores dos atributos condicionais da classe de objetos da qual o reduto foi originado com os atributos correspondentes ao reduto. Finalmente, a partir das regras pode-se fazer inferências para se chegar a uma classificação adequada.

4. Base de Dados

Será utilizada uma base de dados extraída da Clínica *EMONT GODINE* da Universidade Católica de Lovain, na Bélgica [12].

Esta base consta na classificação de doenças reumatológicas através de um conjunto de 50 casos, com 19 atributos, a saber: idade, sexo, dor lombar,

rigidez na coluna, deformação na coluna, mobilidade, dor ao toque no sacroilíaco, artrite, rigidez matinal, bursite, tofos, sinovite, artralgia, nódulos reumáticos, HLA-B27, deformação nas juntas, alterações laboratoriais, evidências radiológicas e tomografia computadorizada.

Em função da disposição dos valores dos atributos em cada caso, os pacientes recebem três diagnósticos: Espondilite Anquilosante (EA), Artrite Gotosa (AG) e Artrite Reumatóide¹ (AR).

É importante ressaltar que para ambas implementações a base de dados foi dividida aleatoriamente em 38 casos para treinamento e 12 para testes. Uma pequena amostra é vista na tabela 1.

Tabela 1: Amostra da base de dados

	Atributos					Decisão
	S e x o	I d a d e	Dor L o m b a r	A r t r i t e	R i g i d e z m a t i n a l	Diagnóstico
Casos					X	
1	m	30	sim	não	não	EA
2	f	23	não	sim	sim	AG
3	m	30	não	sim	sim	AG
4	f	63	não	não	sim	AR
5	m	21	sim	não	não	EA
6	f	56	sim	sim	sim	AR

5. Implementação

5.1. Rough Sets

Com o intuito de aplicar a TRS no referido problema, foi utilizado o software Rosetta². Primeiramente a variável idade foi discretizada através do algoritmo de Raciocínio Booleano [13]. Em seguida os redutos foram gerados utilizando o método de redução por Algoritmo Genético [14]. Finalmente, foram geradas as regras.

Para a classificação escolhemos dois algoritmos: Votação padrão e Votação com rastreamento de objeto [13].

5.2. Redes Neurais Artificiais

Para fazer a classificação dos casos com seus respectivos diagnósticos foram utilizadas duas diferentes topologias de RNA.

É importante ressaltar que neste trabalho será dada ênfase tanto ao aprendizado supervisionado, utilizando o algoritmo *Backpropagation* fazendo com que a rede classifique os dados por interpolação depois de treinada,

¹ Na base de dados existia apenas um paciente que teve como diagnóstico "Normal", entretanto este caso foi ignorado.

² <http://www.idi.ntnu.no/~aleks/rosetta>

bem como tratar o problema com uma rede que utiliza aprendizado não supervisionado em que a classificação se dá através da “clusterização” de padrões (Rede de Kohonen).

Nos experimentos a primeira rede a ser utilizada foi uma Rede *FeedForward* com aprendizado *Backpropagation*. Esta rede foi configurada com 19 neurônios de entrada, representando cada atributo, 7 neurônios na camada intermediária que tem como função de saída a função tangente sigmoide e um neurônio na camada de saída, utilizando a função sigmoide. Convém ressaltar que a taxa de erro escolhido foi de 0,001.

Já na Rede de Kohonen também foram utilizados 19 neurônios na camada de entrada, bem como uma grade dimensional de 13x13, formando assim 169 neurônios na camada de saída. Além disso, foram necessárias 5000 interações para que a rede convergisse para a solução.

6. Resultados

A tabela 2 mostra através de matrizes de confusão os resultados obtidos no problema em questão.

Tabela 2: Resultados obtidos
Rough Sets

Votação Padrão					
Real	Predito				
		EA	AG	AR	Total(%)
	EA	4	0	0	100
	AG	0	0	3	0
	AR	0	1	4	80
	Total(%)	100	0	57,14	66,67
Votação com rastreamento de objeto					
Real	Predito				
		EA	AG	AR	Total(%)
	EA	3	1	0	75
	AG	0	2	1	66,67
	AR	0	0	5	100
	Total(%)	100	66,67	83,33	83,33
Redes Neurais Artificiais					
FeedForward					
Real	Predito				
		EA	AG	AR	Total(%)
	EA	4	0	0	100
	AG	0	2	1	66,67
	AR	0	1	4	80
	Total(%)	100	66,67	80	83,33
Kohonen					
Real	Predito				
		EA	AG	AR	Total(%)
	EA	4	0	0	100
	AG	0	3	0	100
	AR	0	1	4	80
	Total(%)	100	75	100	91,67

Analisando primeiramente o que tange os resultados obtidos pela TRS, pode-se observar que o algoritmo de

votação com rastreamento de objeto apresentou resultados melhores, ou seja, enquanto o algoritmo de votação padrão teve 4 erros, o segundo algoritmo teve apenas 2 erros.

No âmbito das RNAs observou-se que a Rede *Feedforward* teve um desempenho inferior em relação a Rede de Kohonen, pois esta apresentou apenas um erro, que foi ao indicar Artrite Gotosa ao invés de Artrite Reumatóide. Todavia, vale observar que nenhum teste feito, tanto para TRS, quanto para RNA obteve êxito ao classificar este caso clínico. Podemos aqui aventar a possibilidade de o próprio médico ter errado no diagnóstico, passando de forma errônea os dados para a base de dados em questão.

Ao comparar as duas tecnologias percebe-se que as RNAs obtiveram um desempenho superior, embora o resultado da TRS, principalmente pelo algoritmo de votação com rastreamento de objeto também tenha alcançado um resultado relativamente bom.

7. Conclusões

A utilização de métodos de Inteligência Artificial para a resolução de problemas tal qual o descrito neste trabalho vem sendo muito difundida. Entretanto, a bibliografia produzida pela comunidade de IA nem sempre se foca em produzir trabalhos comparativos entre as duas vertentes da IA, isto é, simbólica e conexionista. Portanto, este trabalho surgiu com a intenção de levantar na prática a questão comparativa entre duas tecnologias de diferentes abordagens da IA.

Embora este artigo tenha um cunho comparativo, deve-se ressaltar que ele é um estudo experimental e que os resultados obtidos são válidos apenas para o problema de diagnóstico médico aqui proposto. Sendo que para uma maior validação deste efeito comparativo, mais experimentos devem ser feitos, bem como uma análise matemática mais profunda é de suma importância.

Como futuros trabalhos, sugere-se a utilização da Teoria de Rough Sets para fins de pré-processamento antes dos dados serem apresentados à RNA, pois isto poderia diminuir a quantidade de informação apresentada a rede, conseqüentemente diminuindo a complexidade computacional envolvida. Outro trabalho relevante seria envolver Redes Neurais Rough [15].

Referências

- [1] J. M. Barreto. *Inteligência Artificial: No limiar do século XXI*. Duplic Edições, 1997.
- [2] L. Breiman, J. Friedman, R. Olshen, et al. *Classification and Regression Trees*. Pacific Groves, CA, Wadsworth, 1984.
- [3] M. Metha, R. Agrawal, J. S. Rissanen. *A Fast Scalable Classifier for Data Mining*. In *Proceedings of the 5th International Conference on Extending Database Technology (EDBT)*, pages 18-32, Avignon, 1996.
- [4] J. R. Quinlan. *C4.5: Programs for Machine Learning*. Morgan Kaufman, 1993.

- [5] C. M. Bishop. *Neural Networks for Pattern Recognition*. Oxford University Press, 1996.
- [6] P. W. T. A. Simões, S. M. Nassar, M. M. S. Pires. *Sistema de Apoio na Avaliação da Falência do Crescimento Infantil*. In: Congresso Brasileiro de Computação. Workshop de Informática Aplicada à Saúde, 2001.
- [7] M. R. Sousa. *Mineração de Dados: Uma implementação fortemente acoplada a um sistema gerenciador de banco de dados paralelo*. Dissertação. Programa de Engenharia de Sistemas e Computação – COPPE/UFRJ, 1998.
- [8] J. Vesanto. *Using SOM in Data Mining*. Licentiate's thesis. Helsinki University Of Technology. Finland, 2000.
- [9] M. W. Craven, J. W. Shavlik. *Using Neural Networks for Data Mining*. Future Generation Computer Systems, 1999.
- [10] A. D. Pila, M. C. Monard. *Seleção de Atributos Relevantes para Aprendizado de Máquina Utilizando a Abordagem de Rough Sets*. Dissertação. Instituto de Ciências Matemáticas e de Computação da Universidade de São Paulo - ICMC/USP. São Carlos, 2001.
- [11] Z. Pawlak. *Rough Sets*. International Journal of Computer and Information Sciences, pages 341-356, 1982.
- [12] M. C. A. F. Gabriel. *Análise da Utilização de Redes de Kohonen no Auxílio ao Diagnóstico de Doenças Reumatológicas*. Dissertação. Universidade Federal de Santa Catarina - UFSC. Florianópolis, 2002.
- [13] A. Ohrn. *Discernibility and Rough Sets in Medicine: Tools and Applications*. PhD thesis. Norwegian University on Science and Technology. 1999.
- [14] J. Wroblewski. *Finding minimal reducts using genetic algorithms (extended version)*. In *Proc. Second International Joint Conference on Information Sciences*, pages 186–189, September 1995.
- [15] P. Lingras. *Comparasion of neofuzzy and rough neural networks*. Information Sciences, pages 207-215, 1998.