ESCÓLIOS SOBRE A TEORIA DOS CONJUNTOS APROXIMATIVOS

COMMENTARIES ABOUT THE ROUGH SETS THEORY

Nelson Hein e Adriana Kroenke

Universidade Regional de Blumenau – PPGCC/FURB
Rua Antônio da Veiga, 140, Sala D206, Bairro Victor Konder
Caixa Postal 1507 –Blumenau / SC – Brasil. CEP 89012-900
E-mail: hein@furb.br / didlen@terra.com.br

RESUMO

Este artigo tem como objetivo fazer a apresentação e exemplificação do uso da teoria dos conjuntos aproximativos como ferramenta Data Mining. A teoria dos conjuntos aproximativos (TCA), que vem da denominação inglesa *Rough Set Theory* (PAWLAK, 1982), vem sendo desenvolvida desde a década de 1980, cujo autor é o pesquisador polaco Zdzislaw Pawlak. Através de todo este tempo, como toda teoria, vem sendo enriquecida com novos aportes, derivados de uma maior investigação sobre seus alcances e facilidades, tanto para aplicações práticas como teóricas. Tomando em conta que é uma ferramenta Data Mining, atualmente possui aplicações em diferentes campos, sobretudo em sistemas de apoio a decisão e sistemas gerenciais de informação.

Palavras-chave: Data Mining. Rough Sets. Teoria da decisão.

ABSTRACT

This article aims to present and exemplify the use of the Rough Set Theory as a Data Mining tool. The Rough Set Theory (Teoria dos Conjuntos Aproximativos – TCA), which comes from an English denomination (Pawlak, 1982), is a theory that has been developed since the 80's whose author is the Polish researcher Zdzislaw Pawlak. Through all of this time, like any theory, it has been enriched with new contributions, derived from a larger investigation over its scopes and readiness, both to practical and theoretical applications. Taking in consideration that it is a Data Mining tool, it has currently gotten applications in different fields, especially in decision support systems and in management information systems.

Keywords: Data Mining, Rough Sets, Decision Theory.

1 INTRODUÇÃO

Ferramentas Data Mining, definidas de forma resumida, vêm a ser o conjunto de procedimentos e técnicas que buscam extrair padrões dentro de um conjunto de dados (MARAKAS, 1998). Nesse sentido, a TCA busca extrair padrões com base no conceito de "indiscernibilidade". Considerando que "indiscernir" significa não conseguir distinguir uma coisa de outra por meio dos sentidos ou da inteligência humana, o que busca a TCA é encontrar todos os objetos que produzem um mesmo tipo de informação, ou seja, que são indiscerníveis. A partir desse conceito é que foram geradas as bases matemáticas desta teoria.

A premissa central da filosofia dos conjuntos aproximativos é que o conhecimento consiste na habilidade de classificar objetos. Ao fazer isso, percebem-se algumas diferenças entre objetos, os quais formam classes de objetos que não são notavelmente diferentes. As classes de objetos indiscerníveis como os blocos básicos (conceitos) são usadas para construir conhecimento sobre um mundo real abstrato. Essa visão do conhecimento é semântica por natureza, onde a granularidade do conhecimento (indiscernibilidade de alguns objetos) é de primordial importância e pode ser usada para definir conceitos chaves da teoria: aproximação, dependência e redução. A TCA permite que, a partir de uma tabela de informação, dela se derivem regras de decisão que irão constituir o conhecimento construído.

Inicialmente, os conceitos básicos da TCA serão apresentados seguindo a linha e a nomenclatura desenvolvida por Pawlak (PAWLAK; SLOWINSKI, 1993). Posteriormente, uma aplicação irá ilustrar os conceitos desenvolvidos.

2 CONCEITOS BÁSICOS DA TEORIA DOS CONJUNTOS APROXIMATIVOS

Uma tabela de informação é uma tabela de dados, estruturada de forma que as linhas representam objetos, e as colunas, atributos. Nas entradas da tabela colocam-se os valores correspondentes v_{ij}.

Tabela 1 – Estrutura de uma tabela de informação

	atributo 1	Atributo 2	...	atributo r
Objeto 1	v_{11}	v_{12}	...	v_{1r}
Objeto 2	v_{21}	v_{22}	...	v_{2r}
...	...	...	...	...
Objeto m	v_{m1}	v_{m2}	...	v_{mr}

Fonte: elaborado pelos autores.

Considerem-se os conjuntos

U = {objeto 1, objeto 2, ..., objeto m}: conjunto universo (finito) de objetos;
Q = {atributo 1, atributo 2, …, atributo r}: conjunto (finito) de atributos.

A cada atributo $q \in Q$ está associado um conjunto de possíveis valores que qualquer objeto pode tomar, chamado domínio do atributo e denotado por V_q. Considere-se:

$$V = \bigcup_{q \in Q} V_q$$

E ainda uma função de informação:

$$f: U \times Q \rightarrow V \text{ tal que } f(x,q) \in V_q$$

Tal função associa cada par (objeto x, atributo q) ao valor correspondente v_{xq} de seu domínio V_q. A estrutura $S = \langle U,Q,V,f \rangle$ assim constituída é definida como sendo um sistema de informação.

É importante, na TCA, verificar se um subconjunto $P \subset Q$ de atributos de condição fornece conhecimento adequado a determinados propósitos, como diagnóstico, baseado

nos valores assumidos por um determinado atributo de decisão (para fins de classificação, por exemplo). Assim, dado um sistema de informação S e $P \subset Q$, diz-se que dois objetos $x, y \in U$ são *indiscerníveis* para o conjunto de atributos P se, e somente se, $f(x,q) = f(y,q)$ para todo $q \in P$; ou seja, x e y são indiscerníveis em P se apresentam os mesmos valores para todos os atributos em P. A *relação de indiscernibilidade* I_P em U é definida pela condição $(x,y) \in I_P$ se x, y são indiscerníveis para o conjunto P de atributos, ou:

$$I_P = \{(x,y) \in U \times U \mid f(x,q) = f(y,q), \forall q \in P\}.$$

A relação de indiscernibilidade I_P é uma relação de equivalência: é reflexiva, simétrica e transitiva. Portanto, efetua uma partição de U em classes de equivalência, cada uma das quais é um subconjunto dos elementos de U, que são indiscerníveis entre si. Cada uma dessas classes é um *conjunto P-elementar* em S. A família de todas essas classes é denotada por U / I_P.

$Des_P(X)$ denota a descrição do conjunto P-elementar $X \in U / I_P$ em termos dos pares (atributo,valor), isto é:

$$Des_P(X) = \{(q,v) \mid f(x,q) = v, \forall x \in X, \forall q \in P\}.$$

Seja $P \subset Q$ e $Y \subset U$. Então, a aproximação P-inferior de Y, denotada por $\underline{P}Y$, e a aproximação P-superior de Y, denotada por $\overline{P}Y$, são definidas por:

$$\underline{P}Y = \bigcup \{X \in U/I_P \mid X \subset Y\}$$

$$\overline{P}Y = \bigcup \{X \in U/I_P \mid X \cap Y \neq \varnothing\}$$

E o conjunto P-fronteira de Y, denotado por $Fr_P(Y)$, é definido por:

$$Fr_P(Y) = \overline{P}Y - \underline{P}Y .$$

O conjunto $\underline{P}Y$ é formado por todos os elementos que certamente podem ser classificados como elementos de Y, discernindo-os mediante o conjunto de atributos P. Já $\overline{P}Y$ é o conjunto de elementos de U que podem ser possivelmente classificados como elementos de Y. O conjunto P-fronteira $Fr_P(Y)$ é o conjunto de elementos que podem, possivelmente, mas não certamente, ser classificados como elementos de Y. Evidentemente, $\underline{P}Y \subset \overline{P}Y$ e $\underline{P}Y = \overline{P}Y$ se e somente se $Fr_P(Y) = \varnothing$.

A cada $Y \subset U$ associa-se uma *precisão de aproximação* do conjunto Y por P em S, definida como:

$$\alpha_P(Y) = \frac{\text{card}\left(\underline{P}Y\right)}{\text{card}\left(\overline{P}Y\right)}$$

onde card denota a cardinalidade do conjunto, satisfazendo:

$$0 \leq \alpha_P(Y) \leq 1.$$

Seja S um sistema de informação, $P \subset Q$ e seja $Y = \{Y_1, Y_2, ..., Y_n\}$ uma partição de U. O coeficiente:

$$\gamma_P(Y) = \frac{\sum_{i=1}^{n} \text{card}(\underline{P}Y_i)}{\text{card}(U)}$$

É chamado qualidade da aproximação da partição Y pelo conjunto de atributos P, ou qualidade da classificação para abreviar. Este coeficiente satisfaz:

$$0 \leq \gamma_P(Y) \leq 1.$$

Sejam R, $P \subset Q$ dois conjuntos de atributos em um sistema de informação S. Diz-se que R *depende* de P e denota-se por $P \rightarrow R$ se $I_P \subset I_R$. Descobrir dependências entre atributos é de importância primordial em TCA para a análise do conhecimento.

Outro ponto importante é a redução de atributos, de tal modo que um conjunto reduzido de atributos que forneça a mesma qualidade de classificação em relação a um conjunto original de atributos. Dado algum $P \subset Q$, o mínimo subconjunto $R \subset P$ tal que $\gamma_R(Y) = \gamma_P(Y)$ é chamado uma Y-redução de P e é denotado por $RED_Y(P)$. Um sistema de informação pode ter mais que uma Y-redução. A intersecção de todas as *Y*-reduções é chamada de Y-núcleo de P, isto é, $CORE_Y(P) = \cap RED_Y(P)$. O núcleo é a coleção dos atributos mais significativos no sistema.

Uma tabela de decisão é um sistema de informação formado por atributos de condição em C e por atributos de decisão em D, tal que $Q = C \cup D$, $C \cap D = \varnothing$. Uma tabela de decisão é determinística se $C \rightarrow D$; caso contrário, é não-determinística. Uma tabela de decisão determinística unicamente descreve as decisões a serem efetuadas quando algumas condições são satisfeitas. No caso de uma tabela de decisão não-determinística, decisões não são univocamente determinadas pelas condições.

Pode-se derivar um conjunto de regras de decisão a partir de uma tabela de decisão. Seja $U / I_C = \{X_1, X_2, ..., X_k\}$ a família de todas as classes de condição e $U / I_D = \{Y_1, Y_2, ..., Y_n\}$ a família de todas as classes de decisão. Então, $Desc_C(X_i) \Rightarrow Desc_D(Y_j)$ é chamada uma regra de decisão (C,D). As regras de decisão também podem ser expressas em declarações lógicas tipo "se ... então ...", relacionando classes de e de decisão. O conjunto de regras de decisão para cada classe de decisão Y_j (j = 1,....,n) é denotado por $\{r_{ij}\}$. Precisamente:

$$\{r_{ij}\} = \{Desc_C(X_i) \Rightarrow Desc_D(Y_j) \mid X_i \cap Y_j \neq \varnothing, i = 1,...,k\}.$$

Uma regra r_{ij} é determinística se $X_i \subset Y_j$; caso contrário, é não-determinística. Regras não-determinísticas são consequências de uma descrição aproximada de classes de decisão (categorias) em termos de classes de condição (blocos de objetos indiscerníveis por atributos de condição). Significa que, usando o conhecimento disponível, não se pode decidir se alguns objetos (da região fronteira) pertencem a uma determinada categoria ou não.

Procedimentos para a derivação de regras de decisão para tabelas de decisão são apresentados por Boryczka e Slowinski (1988), Slowinski e Stefanowski (1992), Skowron e Grzymala-Busse (1993).

3 UM EXEMPLO

O presente exemplo está baseado no modelo proposto por Pawlak. Considere a tabela de informação:

Tabela 2 – Tabela de informação

Paciente	Dor de cabeça (*C*)	Dor muscular (*M*)	Temperatura (*T*)	Gripe (*G*)
1	Não	Sim	Alta	Sim
2	Sim	Não	Alta	Sim
3	Sim	Sim	Muito Alta	Sim
4	Não	Sim	Normal	Não
5	Sim	Não	Alta	Não
6	Não	Sim	Muito Alta	Sim

Fonte: elaborado pelos autores.

Tem-se:

U = {1, 2, 3, 4, 5, 6} é o conjunto universo de objetos (pacientes);
C = {*C*,*M*,*T*} é o conjunto de atributos de condição;
D = {*G*} é o conjunto de atributos de decisão;
Q = C ∪ D = {*C*,*M*,*T*,*G*} é o conjunto de todos os atributos;

Com os seguintes domínios de atributos:

$V_D = V_M = V_G$ = {Sim, Não};
V_T = {Normal, Alta, Muito alta}.

A função de informação assume os valores

f(1,*C*) = Não; f(1,*M*) = Sim; f(1,*T*) = Alta; f(1,*G*) = Sim;
f(2,*C*) = Sim; f(2,*M*) = Não; f(2,*T*) = Alta; f(2,*G*) = Sim;

Nesse sistema de informação particiona-se U de forma a diagnosticar a gripe; portanto, de acordo com a relação de indiscernibilidade I_D sobre o atributo de decisão *G*. Obtém-se, assim, $Y = U / I_D = \{Y_1,Y_2\}$, onde Y_1 = {1,2,3,6} é o conjunto de pacientes que apresentam gripe e Y_2 = {4,5} é o conjunto de pacientes que não apresentam gripe.

Seja o conjunto de atributos P = {*M*, *T*}. Sua relação de indiscernibilidade é I_p = {(1,1),(2,2),(2,5),(3,3),(3,6),(4,4),(5,2),(5,5),(6,3),(6,6)} e a classe de conjuntos P-elementares é:

$U / I_P = \{\{1\},\{2,5\},\{3,6\},\{4\}\}$.

Os conjuntos P-elementares em U / I_P possuem as seguintes descrições:

$Des_P(\{1\}) = \{(M,\text{Sim}), (T,\text{alta})\}$;
$Des_P(\{2,5\}) = \{(M,\text{Não}), (T,\text{alta})\}$;

$Des_P(\{3,6\}) = \{(M,\text{Sim}), (T,\text{muito alta})\}$;
$Des_P(\{4\}) = \{(M,\text{Sim}), (T,\text{normal})\}$.

Ao considerar como P aproxima ao conjunto Y_1 de pacientes que apresentam gripe, tem-se:

$\underline{P}Y_1 = \{1\} \cup \{3,6\} = \{1,3,6\}$;
$\overline{P}Y_1 = \{1\} \cup \{2,5\} \cup \{3,6\} = \{1,2,3,5,6\}$;
$Fr_P(Y_1) = \overline{P}Y_1 - \underline{P}Y_1 = \{2,5\}$;
$\alpha_P(Y_1) = \frac{\text{card}(\underline{P}Y_1)}{\text{card}(\overline{P}Y_1)} = \frac{3}{5} = 0{,}6$ (precisão da aproximação).

Ao considerar como P aproxima ao conjunto Y_2 de pacientes que não apresentam gripe, tem-se:

$\underline{P}Y_1 = \{4\}$;
$\overline{P}Y_1 = \{4\} \cup \{2,5\} = \{2,4,5\}$;
$Fr_P(Y_1) = \overline{P}Y_1 - \underline{P}Y_1 = \{2,5\}$;
$\alpha_P(Y_1) = \frac{\text{card}(\underline{P}Y_1)}{\text{card}(\overline{P}Y_1)} = \frac{1}{3} = 0{,}3333$ (precisão da aproximação).

A qualidade da aproximação da partição Y pelo conjunto de atributos P é:

$$\gamma_P(Y) = \frac{\text{card}(\underline{P}Y_1) + \text{card}(\underline{P}Y_2)}{\text{card}(U)} = \frac{3+1}{6} = 0{,}667 .$$

Para descobrir as dependências e obter as reduções deve-se, inicialmente, encontrar a qualidade da aproximação sobre todos os possíveis subconjuntos P de atributos de condição. O quadro a seguir mostra os resultados.

Quadro 1 – Qualidade da aproximação

Atributos P	Qualidade Aproximação $\gamma_P(Y)$	Conjuntos P-elementares em U / I_P
$\{C,M,T\}$	0,667	{1}, {2,5}, {3}, {4}, {6}

{*M*,*T*}	0,667	{1}, {2,5}, {3,6}, {4}
{*C*,*T*}	0,667	{1}, {2,5}, {3}, {4}, {6}
{*C*,*M*}	0,167	{1,4,6}, {2,5}, {3}
{*T*}	0,500	{1,2,5}, {3,6}, {4}
{*M*}	0,000	{1,3,4,6}, {2,5}
{*C*}	0,000	{1,4,6}, {2,3,5}

Fonte: elaborado pelo autor.

4 CONCLUSÕES

Observa-se que as Y-reduções de P = {*C*,*M*,*T* } são {*M*,*T*} e {*C*,*T*}; portanto, o *Y* - núcleo de P é $CORE_{\Psi}(\{C,M,T\}) = \{C,T\} \cap \{M,T\} = \{T\}$, ou seja, *T* é o atributo mais significativo de Q, o qual não pode deixar de ser considerado, pois sua eliminação significa obter aproximações de baixa qualidade. Em relação aos atributos *C* e *M*, são mutuamente intercambiáveis. Assim, fica a critério pessoal trabalhar com {*C*,*T* } ou com {*M*,*T*}, considerando que ambos os grupos produzem a mesma qualidade de informação em relação a {*C*,*M*,*T*}.

Quanto às dependências, se P, R são dois conjuntos de atributos e se $P \subset R$, então $I_R \subset I_P$; portanto, existe a dependência $R \rightarrow P$. Assim, por exemplo, $\{M,T\} \subset \{C,M,T\}$ e, portanto, a dependência $\{C,M,T\} \rightarrow \{M,T\}$. Além destas dependências existe a dependência $\{C,T\} \rightarrow \{C,M\}$.

Da família $U / I_C = \{\{1\},\{2,5\},\{3\},\{4\},\{6\}\}$ de classes de condição e da família $U / I_C = \{\{1,2,3,6\},\{2,5\}\}$ de classes de decisão surgem as regras:

Quadro 2 – Classes de decisão

Objetos	Regra $Desc_C(X_i) \Rightarrow Desc_D(Y_j)$	Determinística?
1	{(*C*,Não),(*M*,Sim),(*T*,Alta)} ⇒ {(*G*,Sim)}	Sim
2	{(*C*,Sim),(*M*,Não),(*T*,Alta)} ⇒ {(*G*,Sim)}	Não
5	{(*C*,Sim),(*M*,Não),(*T*,Alta)} ⇒ {(*G*,Não)}	Não
3	{(*C*,Sim),(*M*,Sim),(*T*,Muito alta)}⇒{(*G*,Sim)}	Sim
4	{(*C*,Não),(*M*,Sim),(*T*,Normal)} ⇒ {(*G*,Não)}	Sim
6	{(*C*,Não),(*M*,Sim),(*T*,Muito alta)}⇒{(*G*,Sim)}	Sim

Fonte: elaborado pelo autor.

As regras simplificam-se caso se adotem a *Y*-redução {*M*,*T*} de C, sem perda da qualidade de aproximação da partição. Neste caso:

Quadro 3: Redução das regras de decisão

Objetos	regra $Desc_C(X_i) \Rightarrow Desc_D(Y_j)$	Determinística?
1	{(*M*,Sim),(*T*,Alta)} ⇒ {(*G*,Sim)}	Sim
2	{(*M*,Não),(*T*,Alta)} ⇒ {(*G*,Sim)}	Não
5	{ (*M*,Não),(*T*,Alta)} ⇒ {(*G*,Não)}	Não
3,6	{(*M*,Sim),(*T*,Muito alta)}⇒{(*G*,Sim)}	Sim
4	{ (*M*,Sim),(*T*,Normal)} ⇒ {(*G*,Não)}	Sim

Fonte: elaborado pelo autor.

REFERÊNCIAS

BORYCZKA, M.; SLOWINSKI, R. Derivation of optimal decision algorithms from decision tables using rough sets. *Bulletin of the Polish Academy of Sciences,* ser. Technical Sciences, v. 36, p. 251-260, 1988.

MARAKAS, G. *Decision support systems in the 21st century*. New York : Prentice-Hall, 1988.

PAWLAK, Z. Rough sets. *International Journal of Information & Computer Sciences*, v. 11, p. 341-356, 1982.

PAWLAK, Z.; SLOWINSKI, R. Decision analysis using rough sets. *ICS Research Report* nº 21, Institute of Computer Science, Warsaw University of Technology, Warsaw, Poland, 1993.

SLOWINSKI, R.; STEFANOWSKI, J. "Roughdas" and "roughclass" software implementations of the rough sets approach. In: SLOWINSKI, R. (Ed.). *Intelligent Decision Support. Handbook of Applications and Advances of the Rough Sets Theory*. p. 445-456. Dordrecht : Kluwer Academic Publishers, 1992. p. 445-446.

SKOWRON A.; GRZYMALA-BUSSE, J. W. From the rough set theory to the evidence theory. In: FEDRIZZI, M.; KACPRZYK, J.; YAGER, R.R. (Ed.). *Advances in the Dempster-Shafer Theory of Evidence*. New York : J. Wiley and Sons, 1993.