

# SUPORTE DE ONTOLOGIAS APLICADAS À MINERAÇÃO DE DADOS POR REGRAS DE ASSOCIAÇÃO

Eduardo de Mattos Pinto Coelho<sup>1</sup>, Marcello Peixoto Bax<sup>2</sup>, Wagner Meira Jr.<sup>3</sup>

<sup>1</sup>Prefeitura Municipal de Belo Horizonte

<sup>2</sup>Escola de Ciência da Informação. Universidade Federal de Minas Gerais (UFMG)

<sup>3</sup>Departamento de Ciência da Computação, (UFMG)

([emattos@pbh.gov.br](mailto:emattos@pbh.gov.br), [bax@eci.ufmg.br](mailto:bax@eci.ufmg.br), [Meira@dcc.ufmg.br](mailto:Meira@dcc.ufmg.br))

## **Abstract.**

*Data Mining (DM) for association rules tends to generate an unmanageable number of rules affecting the scope of its application. To solve this problem we propose the use of ontologies in the stages of pre and post-processing tasks to support the MD. In addition, the article points out that human organizations require the notions of possibility, subjectivity and interpretation, contrasting with the notions of necessity, objectivity and explanation, useful in fields of natural sciences. These requirements demand new perspective on ontologies and DM, often sheltered by soft computing.*

**Resumo.** *Mineração de dados(MD) por regras de associação tende a gerar um número intratável de regras prejudicando a abrangência de sua aplicação. Para solucionar esse problema propõe-se o uso de ontologias nas etapas de pré e pós-processamento no suporte às tarefas de MD. Além disso, o artigo ressalta que organizações humanas exigem as noções de possibilidade, subjetividade e interpretação, contrastantes com as noções de necessidade, objetividade e explicação, úteis em domínios de ciências naturais. Tais exigências demandam novas perspectiva em ontologias e MD, normalmente abrigadas pela computação suave.*

## **1. Introdução**

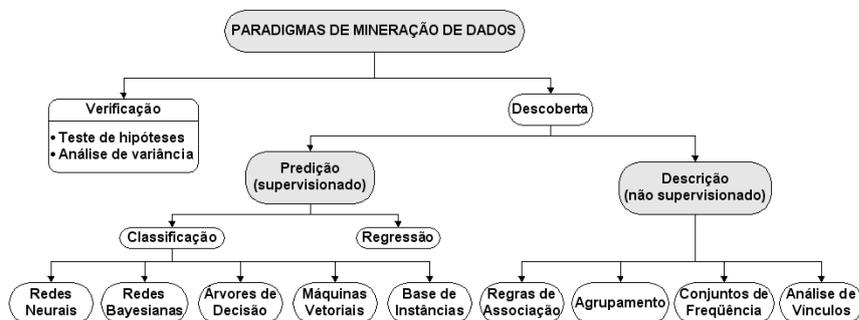
Ontologias e mineração de dados (MD) são áreas, em geral, construídas com base nas ciências naturais (CN). Lidam com as noções de necessidade, objetividade e explicação, em ciências físicas, químicas e biológicas. Entretanto, em organizações, fenômenos humanos prevalecem, demandando metodologias de ciências humanas (CH) que, por sua vez, exigem as noções de possibilidade, subjetividade e interpretação. Tais noções já são abrigadas por tecnologias humano-cêntricas na perspectiva da computação suave. Assim, com base em sistemas *difusos*, propõe-se utilizar ontologias no suporte à atividade de MD nas fases de pré e pós-processamento. Agrega-se então valor conceitual, associado ao conhecimento de domínio, e propicia-se suporte semântico no processamento dos dados. O analista é auxiliado na explicação e interpretação das regras obtidas da MD. Inicialmente o artigo introduz noções de MD. A partir da análise das tarefas de prescrição, predição, descrição, explicação e interpretação, conforme abordados em Domingues (2004), são discutidos métodos que devem ser incorporados a ontologias em contextos humano-sociais, típicos de questões surgidas em economia, administração, contabilidade, direito, etc.

## **2. Mineração de Dados**

A MD é uma das etapas da Descoberta de Conhecimento em Bases de Dados. É um processo de identificação de padrões novos, potencialmente úteis e compreensíveis, em um conjunto de dados. Esses padrões podem se constituir, dentre outros, em regras de associação, ou seqüências temporais que permitam revelar relacionamentos não aleatórios entre atributos de variáveis. Em especial, o volume excessivo de dados justifica a MD para a descoberta de padrões que possam revelar informações úteis à tomada de decisão.

A MD lidará com questões de ordem computacional e de usabilidade. Ter-se-á de lidar com a confiabilidade dos dados, a geração de resultados excessivos, autonomia e conhecimento do negócio para a análise dos resultados, desempenho e produtividade, facilidade de uso e utilidade.

Há uma variedade de paradigmas de mineração de dados. Objetivando nossa análise, consideramos mais interessante a taxonomia que realça a existência de modelos supervisionados e não-supervisionados.



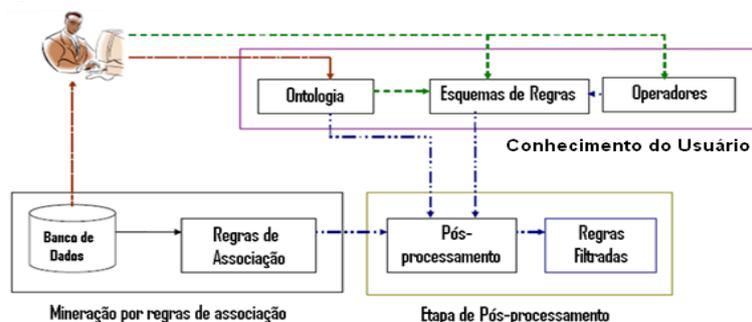
**Figura 1.**  
Taxonomia de paradigmas de mineração de dados (Maimon e Rokach, 2005, p. 7).

Em **modelos** supervisionados, como o próprio nome já diz, o processo de detecção de padrões é supervisionado. As classes à qual cada padrão possa ser pertinente é pré-definida. São **modelos prescritivos**. Além disso, são preditivos no sentido de serem mais adequados a revelar tendências.

Já em modelos não-supervisionados, não se conhecem a priori classes às quais os padrões possam ser pertinentes. São **modelos de ênfase descritiva**, no sentido de serem mais adequados a apresentarem descrições, relevando características dos dados minerados.

### 3. Ontologias no suporte à MD

A despeito dos algoritmos de MD atenderem critérios de desempenho na criação de novos dados, freqüentemente, eles são insuficientes para garantir seu uso prático. Conhecer o domínio é crucial. Vem-se desenvolvendo um paradigma MD guiado pelo conhecimento,



Cao (2010). Em reforço a esse paradigma, ontologias são utilizadas. Ontologias no pós-processamento de regras de associação aparece em Marinica, Guillet e Briand (2009) e Marinica e Guillet (2010).

**Figura 2.** MD com suporte de ontologias, esquemas de regras e operadores.

### 4. Peculiaridades das ciências humanas



(intenções, sentimentos, consciência, valores e fins visados pelos agentes humanos) (Domingues, 2004, p. 107). As CN debruçam-se sobre causas, descrição e explicação causal dos fatos. As CH estendem esse leque para buscar a descrição e explicação das motivações, razões e crenças. A descrição dever-se-á, necessariamente, lançar seus elementos à tarefa interpretativa que, em CH, raramente já estará encapsulada pela tarefa explicativa.

### **4.3. Explicação**

Com base nos elementos descritivos há distintas formas de explicação possíveis: genéticas, estruturais, funcionais, finais e, dentre outras, causais. **A explicação causal é considerada a forma de explicação por excelência** (Domingues, 2004, p. 116). Entretanto, a questão da causalidade, assim como a questão da vaguidade, é assunto de intensos debates. A causalidade aproxima-se da vaguidade. Assim como ela, a noção de causalidade tem sido útil para fins de análise e, como a vaguidade, ao invés de ser abandonada, ressurge revigorada nas últimas décadas. Uma mesma coisa pode ser causa de efeitos contrários; pode-se identificar uma causalidade recíproca ou circular; pode ocorrer dependência mútua, ação, ou influência de causa e efeito e, ainda, causalidade freqüentemente é confundida com condicional lógica. Conclui Domingues (2004, p. 119) que o importante é a análise causal depender da consideração de um contexto mais amplo, que se decide em outro nível de análise: a interpretação.

### **4.4. Interpretação (compreensão)**

Muitos pretendem que a interpretação já se decide no nível da explicação, e não é senão um de seus aspectos. Entretanto, para Domingues, é o caso de distinguir uma da outra, considerando que a explicação incide sobre os fatos, ou coisas. Já a interpretação envolve a significação, o sentido deles. Portanto, a interpretação irá introduzir as unidades significativas de análise, como as hipóteses, os modelos (tipos ideais), as postulações de sentido, e assim por diante (Domingues, p. 119-120). Em CH, a candidata de ter a primazia no método e de conduzir a análise é a tarefa interpretativa.

### **4.5. Inter-subjetividade**

A subjetividade é crucial na análise do conhecimento em domínios de CH. Anscombe considera não só a subjetividade de um indivíduo, mas a de coletividades inteiras. Desafio maior, é a inter-subjetividade; comunicação das consciências individuais. Na construção de ontologias, a inter-subjetividade é tratada. Ontologias expressam consenso.

Outro lado da questão é o uso de ontologias na etapa de decifração do sentido intersubjetivo na tarefa da interpretação. A verdade a ser considerada advirá da convicção de verdade recolhida pela análise na comunidade. Em CI e CC, a demanda pela inter-subjetividade é eventual. Mas tal demanda é sentida nas comunidades de inteligência onde a inter-subjetividade é regra. Distintos analistas são confrontados numa análise conjunta que traduza a convergência das análises individuais.

## **5. Quais fundamentos em ontologias demandam as ciências humanas?**

Peculiaridades das CH em relação as CN são: primazia da interpretação sobre a explicação; importância e necessidade de subjetividade em conjunto com a objetividade; flexibilização da noção de necessidade (lei) para a noção de possibilidade. Ao buscar agregar conhecimento de domínio, via ontologias, no suporte à MD em contextos organizacionais, impõe-se a nós duas questões: (1) As ontologias fundamentais, ou de alto nível, referências na construção de novas ontologias, lidam com peculiaridades

atinentes ao domínio de CH? (2) Como ontologias poderiam incorporar as noções de subjetividade e possibilidade, suportando a tarefa de interpretação? Não examinaremos essas questões, mas teceremos a seguir alguns comentários, indicações e avaliações iniciais que poderão nortear avaliações mais aprofundadas no futuro.

### 5.1. Ontologias Fundamentais<sup>1</sup>.

DOLCE, a SUMO e a BFO explicitam suas orientações filosóficas. Consideram universais, particulares, ou tropos, seu compromisso ontológico, entidades no tempo e no espaço, e como se relacionam. Oberle et. al (2007) avalia as citadas acima, com base em quatro pares de escolhas ontológicas. Descritiva, ou revisionária (prescritiva); multiplicativa, ou reducionista; atualista, ou possibilista; e endurante, ou perdurante. Considerando o subjetivismo e a noção de possibilidade inerentes às CH, avalia-se que escolher o descritivismo, o multiplicativismo e o possibilismo permite melhor tratar dessas peculiaridades. O descritivismo, ao considerar o que a realidade é, e não como deveria ser, está apto a tratar da linguagem natural, do senso comum, da vaguidade, de objetos não físicos, da diferenciação entre objetos e processos etc. O multiplicativismo por considerar a diferenciação de entidades a partir do não compartilhamento de propriedades fundamentais, também está mais próximo ao senso comum. O possibilismo, por levar à modalidade que faz uso da possibilidade.

Comenta-se brevemente as três ontologias. Ver detalhes em Sreejith (2008). A DOLCE adota uma metafísica descritiva, baseada em Strawson<sup>2</sup>, em oposição à metafísica prescritiva, e considera aspectos lingüísticos e de engenharia cognitiva. Procura incorporar em sua estrutura elementos que permitam lidar com artefatos cognitivos, marcas culturais e convenções sociais (um tipo de metafísica cognitiva). Inspira-se na noção de *deep background* desenvolvida por Searle (2002, *Intentionality*). A SUMO vai na mesma direção, classificando atos intencionais<sup>3</sup>, orientações, interações sociais (Sreejith, 2008, p. 66). Oberle et. al (2007), em solução híbrida (SWIntO), integra a DOLCE e a SUMO. Acomoda resistências à DOLCE, abstrata demais, e à SUMO de axiomatização árdua. A DOLCE teria uma proposta mais abrangente, já a SUMO teria uma taxonomia mais rica. Por fim a BFO distingue entre dois tipos de entidades: substanciais ou continuantes, e processuais, ou ocorrentes. Smith considera a BFO um subconjunto de DOLCE, adequada ao tratamento de instâncias, tipos e relações. Mais adequada que a SUMO, que, segundo esse autor, não apresenta um tratamento claro de relações entre instâncias *versus* relações entre tipos (Barry Smith, *Upper Level Ontologies*). Em CH, a DOLCE e a SUMO já são ao menos parcialmente aptas a lidar com as peculiaridades apresentadas anteriormente, enquanto que a BFO vai na contramão das considerações ontológicas envolvendo aspectos do senso comum, da linguagem natural, da consideração de atos intencionais, vaguidade, causação mental etc. Consideramos que, a despeito dos riscos, é inevitável, para o bem da própria objetividade científica, buscar lidar e tratar de questões que se nos impõem por sua força e penetrabilidade em amplos domínios, envolvendo aspectos cognitivos, tais como a vaguidade, a subjetividade, a intenção e a causação mental. A BFO não pretende alcançar domínios das CH. Ver debate recente (Merril, 2010a e 2010b *versus* Smith e Ceusters, 2010).

---

<sup>1</sup> *Foundational Ontologies*, por vezes traduzidas como ontologias “fundacionais”, um neologismo. Também são chamadas de ontologias de alto nível (*upper ontologies*).

<sup>2</sup> STRAWSON, P.F.. “Individuals. *An Essay in Descriptive Metaphysics*. Routledge, 1959.

<sup>3</sup> <http://virtual.cvut.cz/ksmsaWeb/browser/print/3%23IntentionalProcess> e <http://swserver.cs.vu.nl/partitioning/SUMO/>

## 5.2 A emergência da computação suave, granular e humano-cêntrica

A insuficiência das soluções focadas em detalhes técnicos como eficiência de algoritmos, vem levando à consideração da usabilidade, conhecimento do domínio e capacidade de interpretação de resultados. A computação suave e granular se associa à uma abordagem humano-cêntrica, flexível e com foco na interação do homem com a máquina. Considerando-se essas três abordagens convergentes – computação suave, granular e humano-cêntrica – destacam-se as soluções em sistemas difusos, redes neurais e algoritmos genéticos. Mitra (2002) dá uma visão da literatura disponível sobre MD. Sistemas difusos preocupam-se com a natureza amigável ao usuário (Pedrycz e Gomide, 2007, p. xvii). Sistemas difusos podem ser integrados com outras ferramentas de computação suave levando à geração de sistemas mais poderosos, com aplicações em reconhecimento de padrões, processamento de imagens, e inteligência de máquina (Mitra e Pal, 2005). Nesse contexto, confirmando a emergência dessas abordagens, vem surgindo estudos que aplicam tecnologias de sistemas difusos a MD e ontologias.

## 6. Conclusão

As ferramentas computacionais de MD, por si só, não garantem o sucesso em ambientes organizacionais. Esses ambientes sócio-humanos possuem peculiaridades e exigem abordagens distintas daquelas associadas às CN. Ao invés do caráter de lei associada à noção de necessidade, há a noção de possibilidade; ao invés da ênfase na explicação, há a ênfase na interpretação, e ao invés do expurgo da subjetividade, há a sua concomitante consideração, não em contraposição, mas em cooperação com a objetividade científica. O surgimento de novas abordagens apresentadas traz novas metodologias que permitem tratar as peculiaridades dos domínios sócio-humanos.

## 7. Bibliografia

- CAO, Longbing. Domain-Driven Data Mining: Challenges and Prospects. *Transaction on Knowledge And Data Engineering*, Vo. 22, nº 6, June, 2010, pp. 755-769.
- DOMINGUES, Ivan. *Epistemologia das Ciências Humanas. Tomo I: Positivismo e Hermenêutica – Durkheim e Weber*. Edições Loyola, 2004, 671 p.
- MAIMON, Oded e ROKACH, Lior. "Introduction To Knowledge Discovery in Databases". Chapter 1, In: *The Data Mining and Knowledge Discovery Handbook*, Springer, 2005, p. 1-17,
- MARINICA, Claudia e GUILLET, Fabrice. "Knowledge-Based Interactive Postmining of Association Rules Using Ontologies," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 6, Feb. 2010, pp. 784-797.
- MARINICA, C., GUILLET, F., & BRIAND, H. Post-Processing of Discovered Association Rules Using Ontologies. *Proc. 2008 IEEE Intern Conf. on Data Mining Workshops*, 2009, pp.126–133.
- MERRILL, G.H. "Ontological realms: Methodology or misdirection?" *Applied Ontology*, V. 5, 2010, pp. 79-108.
- MERRILL, G.H. "Realism and reference ontologies: Considerations, reflections and problems". *Applied Ontology*, Vol 5, 2010, pp. 189-221.
- MITRA, S. e PAL, S. K. Fuzzy sets in pattern recognition and machine intelligence. *Fuzzy Sets And Systems*, 156, 2005, pp. 381-386.
- OBERLE, Daniel et al. DOLCE ergo SUMO: On Foundational and Domain Models in SWIntO (SmartWeb Integrated Ontology). Junho, 2007
- PEDRYCZ, Witold e GOMIDE, Fernando. *Fuzzy Systems Engineering – Toward Human-Centric Computing*. IEEE Press, John Wiley & Sons, 2007.
- SEARLE, John R. Intencionalidade. Martins Fontes, 2002, 390 p.
- SMITH, Barry e CEUSTERS, Werner. "Ontological realism: A methodology for coordinated evolution of scientific ontologies. *Applied Ontology*, Vol 5, 2010, pp. 139-188.
- SREEJITH, A. A Project Report on Neo-Vaisesika Formal Ontology. Department Of Computer Science, Cochin University of Science & Tecnology, Kochi, 2008.