



Descoberta de conhecimento com uso de ontologias na mineração de dados*

Application of Ontologies for Knowledge discovery in data mining

Lucélia Pinto Branquinho¹
Renata Maria Abrantes Baracho²
Maurício Barcellos Almeida³

Resumo

A quantidade de informação gerada e disponibilizada nas organizações requer estudos que acrescentem inovações na forma de representação, organização e recuperação da informação. Em domínios específicos, o desafio dos processos de recuperação da informação é criar a relação semântica entre os termos de um vocabulário especializado, gerando um modelo de representação do conhecimento eficiente por meio de instrumentos como as ontologias. O objetivo deste artigo é descrever uma proposta de uso de ontologias no processo *Knowledge Discovery in Database* (KDD), mais especificamente, na mineração de dados utilizando técnica de regras de associação. O trabalho associa o KDD às ontologias para contribuir com a recuperação da informação e, conseqüentemente, com a extração do conhecimento.

Palavras-chave: Recuperação da informação. Descoberta de conhecimento. Ontologias. Mineração de dados.

*Submetido em 07/09/2015 – Aceito em 06/10/2015

¹Mestranda em Ciência da Informação pelo Programa de Pós-graduação em Ciência da Informação da UFMG, Brasil – luceliabranquinho@gmail.com

²Professora pelo Programa de Pós-graduação em Ciência da Informação da UFMG, Brasil – renatabaracho@eci.ufmg.br

³Professor pelo Programa de Pós-graduação em Ciência da Informação da UFMG, Brasil – mba@eci.ufmg.br

Abstract

Nowadays one is able to search more and more information, which requires studies to add innovation in the traditional ways of information representation, organization and retrieval. In a specific knowledge domain, a usual challenge related to information retrieval is to create semantic relations between terms of a specialized vocabulary, generating an efficient model for knowledge representation using instruments like ontologies. The purpose of this article is to describe a proposal for the use of ontologies in Knowledge Discovery in Database process (KDD), more specifically, in data mining using association rules technique. The paper associates the KDD and ontologies to contribute to the information retrieval and, consequently, knowledge extraction.

Keywords: Information retrieval. Knowledge discovery. Ontology. Data mining.

1 INTRODUÇÃO

Segundo Dalfovo (2000), a quantidade de informações armazenadas em bancos de dados das organizações está ultrapassando a habilidade técnica e a capacidade humana na sua interpretação. Tais constatações mostram a necessidade de ferramentas que sejam capazes de analisar, automaticamente, as bases de dados para obter conhecimento e seu melhor gerenciamento, auxiliando os administradores e analistas nos processos de tomada de decisão.

A importância da utilização de métodos computacionais é grande, na medida em que o volume de informações disponíveis para apoiar a tomada de decisão cresce exponencialmente. Nesse sentido, as tecnologias voltadas para coleta, armazenamento, recuperação e disponibilização de dados vêm evoluindo por meio de técnicas, métodos e ferramentas computacionais automatizadas. Os sistemas de recuperação da informação requerem sistemas de organização do conhecimento que permitam a representação de estruturas semânticas de conhecimento de modo a facilitar a recuperação de informação, eliminando ambiguidade e controlando sinônimos. Nos últimos anos, a pesquisa em ontologia tem recebido destaque em função das possibilidades que ela oferece para a representação e organização da informação.

Com o aumento do volume e da variedade de dados sobre os clientes, seja na *Internet* ou em sistemas internos, e da velocidade das ações estratégicas dos tomadores de decisão, a realidade das empresas mudou e, conseqüentemente, o poder do KDD tornou-se imprescindível. Ao utilizar recurso de conhecimento de domínio dos dados semanticamente anotados é possível obter maior previsão e revocação dos padrões.

Para tornar a extração de conhecimento em base de dados mais direcionada e mais eficiente é imprescindível que medidas subjetivas sejam adotadas. Para isso, é importante resolver o problema da dependência do conhecimento do domínio do especialista e tornar este processo mais automatizado. Portanto, os conceitos devem ser mapeados e representados em sistemas de organização do conhecimento que possibilitem o processamento por computadores, reduzindo a dependência da presença do especialista do domínio.

Uma opção para resolver o problema da relevância dos dados obtidos é reduzir os dados a serem minerados por meio de pesquisa de representações dos conjuntos frequentes e de regras “fortes” mais coesas. A redução do conjunto das regras pode ser feita pelo mecanismo de poda. O uso do conhecimento modelado nas ontologias permite generalizar os atributos e filtrar as regras de associação, possibilitando a extração de padrões mais relevantes, promovendo a redução e, conseqüentemente, trazendo maior precisão aos resultados.

Este trabalho tem como objetivo apresentar como as ontologias têm sido aplicadas no processo de KDD para mineração de dados, utilizando técnicas de regras de associação. Foram feitas pesquisas bibliográficas nas bases: Scielo, Portal Capes, BRAPCI e Pubmed em maio de 2014, utilizando descritores como: ontologias, mineração de dados, regras de associação, KDD e similaridade semântica.

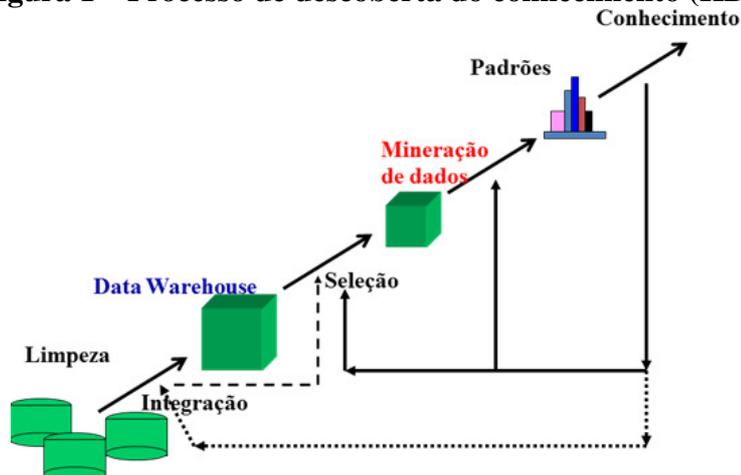
As próximas seções serão organizadas da seguinte maneira: a seção 2.1 remete a uma breve descrição sobre a mineração de dados; a seção 2.2 detalha sobre a técnica de mineração de dados por regras de associação. A seção 3 relata como as ontologias estão sendo utilizadas para obter padrões mais relevantes para a mineração por regras de associação. Finalmente, a seção 4 apresenta considerações finais e as possibilidades para pesquisas futuras.

2 MINERAÇÃO DE DADOS POR REGRAS DE ASSOCIAÇÃO

A extração de conhecimento, geralmente referenciada na literatura como *Knowledge Discovery in Database (KDD)*, é uma área multidisciplinar que incorpora técnicas utilizadas em diversas áreas como banco de dados, inteligência artificial e estatística. O processo de mineração de dados (MD) é um passo no processo de KDD que consiste na aplicação de algoritmos para extração de dados.

Segundo Fayyad et al. (1996), “Extração de Conhecimento de Base de Dados é o processo de identificação de padrões válidos, novos, potencialmente úteis e compreensíveis embutidos nos dados”. A Figura 1 representa KDD sobre a perspectiva de Han e Kamber (2006).

Figura 1 – Processo de descoberta do conhecimento (KDD)



Fonte: HAN; KAMBER, 2006

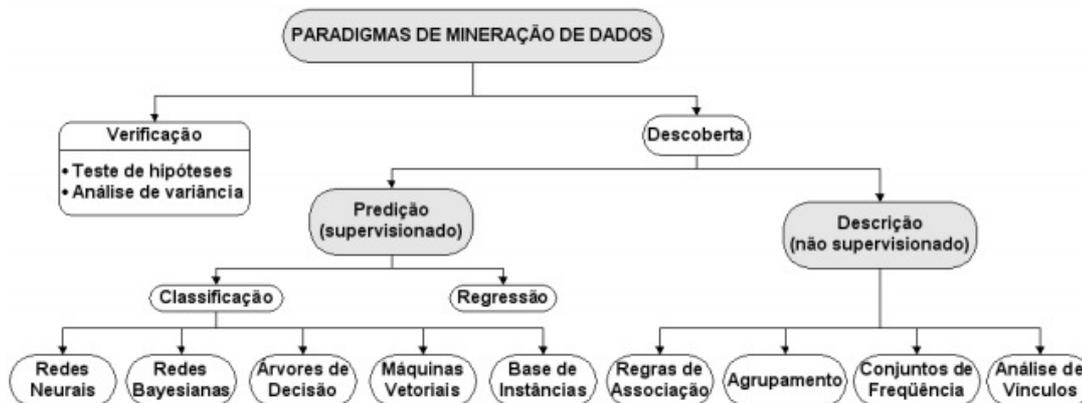
2.1 Mineração de dados

Para um melhor entendimento, o processo de KDD pode ser agrupado em três fases: pré-processamento, MD e pós-processamento. O pré-processamento compreende a captação, organização e tratamento dos dados; já a MD, os algoritmos e as técnicas para busca dos conhecimentos; o pós-processamento abrange todo o conhecimento obtido na MD e a interpretação desse conhecimento.

Conforme Figura 2, a MD possui várias técnicas que variam de acordo com o objetivo a

ser alcançado. A descoberta de padrões pode ser pré-definida, ou seja, supervisionada com características prescritivas e preditivas ou não supervisionada, onde não há conhecimento prévio, com características descritivas.

Figura 2 – Taxonomia de paradigmas de MD



Fonte: MAIMON; ROKACH apud COELHO, 2012

Na etapa de pós-processamento, dentre as medidas utilizadas para se avaliar os padrões resultantes da MD, existem as medidas de interesse objetivas (*data-drive*) e subjetivas (*user-driven*) que buscam avaliar a novidade do padrão descoberto (GONÇALVES, 2005).

Medidas objetivas envolvem a análise da regra de estrutura, o desempenho preditivo e a significância estatística. No entanto, segundo Hamani e Maamri (2009), essas medidas são insuficientes para a determinação de regras inesperadas. De fato, são necessárias medidas subjetivas.

Há duas principais medidas subjetivas que devem ser consideradas (HAMANI et al., 2014):

- a) não previsibilidade: regras interessantes são desconhecidas para o usuário ou contradizem o conhecimento prévio ou as expectativas;
- b) acionáveis: regras são interessantes se o usuário pode agir para obter alguma vantagem.

Para o sucesso na implantação do processo de KDD, há necessidade do entendimento do problema, conhecimento do domínio da aplicação, que só será possível com a participação do usuário final. Desse modo, a interação entre os usuários do processo (usuários finais, especialistas do domínio e analistas do processo) é imprescindível para o sucesso do projeto.

Em função dessa necessidade, um novo paradigma para a descoberta de conhecimento tem sido proposto para aplicação de ontologias com base na proposta de mineração guiada pelo conhecimento de domínio (*domain-driven data mining*, D3M) e pela praticabilidade de uso do conhecimento que é descoberto e entregue (*actionable knowledge discovery and delivery*, AKD) (CAO, 2010).

As ontologias estão sendo utilizadas para tornar mais relevante os padrões descobertos por meio destas técnicas de mineração, principalmente, regras de associação e redes bayesianas.

2.2 Mineração de dados por regras de associação

A técnica de mineração por regras de associação mostra a correlação entre conjuntos de itens em uma série de dados ou transações. Estabelecer regras de associação consiste em representar a condição "SE antecedente ENTÃO consequente", sendo um tipo de regra que garante com uma certa probabilidade (limiar de confiança) que, sempre que o antecedente acontece, o consequente acontecerá. Essas regras são geradas a partir de conjuntos de elementos (conjuntos de itens) que aparecem juntos com, pelo menos, alguma frequência (suporte) (FERRAZ, 2008).

As medidas objetivas, como o apoio e confiança, são frequentemente usadas para avaliar o quanto as regras de associação são interessantes.

A tarefa de mineração por regras de associação pode ser decomposta em dois passos (FERRAZ, 2008):

- a) encontrar todos os itens frequentes, ou seja, todas as combinações de itens com suporte maior que o suporte mínimo especificado pelo usuário. Esse é o passo mais demorado do processo de busca de regras de associação;
- b) gerar, a partir dos itens frequentes, todas as regras com confiança mínima maior que a especificada pelo usuário.

O suporte mínimo especificado para um *itemset* garante a significância estatística da amostra, evitando a consideração de combinações de pequena frequência. Já a confiança mínima especificada mostra que o resultado obtido não é ocasional, há coesão entre as regras obtidas (FERRAZ, 2008).

Uma vez gerado o conjunto das regras de associação, torna-se necessário reduzir sua cardinalidade, o que é feito pelo mecanismo de poda. O problema corrente passa a ser quais as regras que devem ser podadas.

Uma primeira abordagem utiliza a poda estatística, por meio de medidas objetivas de avaliação do grau de interesse das regras, ou seja, medidas que destaquem a utilidade e a inoperabilidade das regras, a outra abordagem é feita por meio de medidas subjetivas. Essa depende do conhecimento do domínio, expresso por meio da ação do usuário. Os resultados são, significativamente, melhores, mas tornam inviável a automatização do processo (FERRAZ, 2008).

2.3 Mineração por regras de associação com ontologia

Presente inicialmente na filosofia, em trabalhos de Aristóteles, a palavra ontologia tem origem grega onde onto quer dizer “ser” e logia “estudo”, assim “estudo do que é”. (ALMEIDA; BAX, 2003). Para Lowe (2007 apud Almeida, 2013,) a ontologia é considerada um

ramo da metafísica que estuda as categorias de entidades que existem e como cada uma delas está relacionada.

Segundo Kasama, Zavaglia e Barcellos (2011), ontologias envolvem os formalismos necessários para a descrição de um conhecimento, permitindo o uso da lógica e a realização de inferências a partir das informações estruturadas. Vavpetic e Lavrac (2012) referenciam o termo ‘mineração de dados semânticos’ para designar a nova mineração de dados que muda com a introdução dessas abordagens.

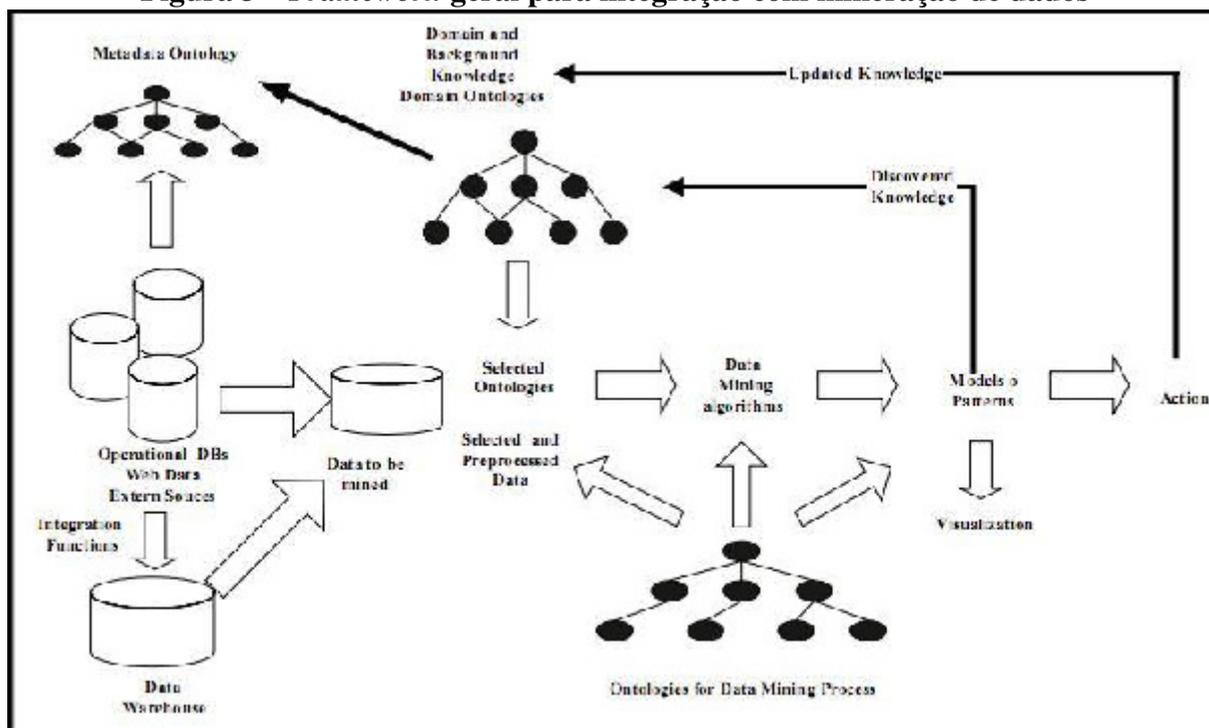
O conhecimento prévio de um domínio ou de um processo na área de mineração de dados pode ajudar a selecionar as informações apropriadas (pré-processamento), diminuir o espaço das hipóteses (de processamento), representar resultados em uma forma mais compreensível e melhorar o processamento (ou pós-processamento).

Segundo Ribeiro (2010), as ontologias podem trazer grandes benefícios, praticamente, para todas as fases do processo de KDD, seja auxiliando na escolha de algoritmos de MD, na elaboração de consultas mais eficientes ou na visualização de novas regras de associação.

Segundo Nigro, Cisaro e Xodo (2007), as ontologias foram introduzidas na KDD, pela primeira vez, no começo de 2000, sendo utilizadas, conforme Figura 3, de três modos:

- a) ontologias para a mineração de dados – incorporar conhecimento ao processo de MD com a adição de Ontologias para orientar o processo de descoberta, interpretação e validação do conhecimento extraído, ontologias para recursos e descrição do serviço;
- b) mineração de dados para as ontologias – adicionar conhecimento de domínio à informação de entrada ou saída da MD. As aplicações mais características desta abordagem são em medicina, biologia e dados espaciais, como a representação de genes, taxonomias, aplicações em geociências, aplicações médicas e, especialmente, na evolução de domínios;
- c) ontologias para metadados - armazena o conhecimento sobre os itens e os relacionamentos dos atributos extraídos da base de dados para compor uma ontologia que dá suporte à ontologia de conhecimento do domínio.

Figura 3 – Framework geral para integração com mineração de dados



Fonte: NIGRO; CISARO; XODO, 2007

Segundo Ribeiro (2010) e Coelho (2012), o uso de ontologia pode reduzir os esforços dos especialistas de domínio na definição de regras de associação e na análise de descobertas de padrões.

Nesse aspecto, a Ciência da Informação (CI), como ciência social aplicada, faz a ponte e a integração entre áreas diferentes, considerando não somente o dado, mas a inserção da informação e do conhecimento na organização, estabelecendo princípios e norteando a resolução do problema da representação do conhecimento e de como a constituição de ontologias pode auxiliar o processo de KDD.

A mineração de dados tradicional é baseada na frequência de ocorrência de casos e a co-ocorrência de itens nas transações. O significado de cada item ou instância não é levado em consideração. O conteúdo semântico extraído das ontologias permite a inserção de mais inteligência e conhecimento em mineração de dados, melhorando sua qualidade (FERRAZ, 2008).

Ao integrar métodos objetivos com abordagem semântica, é possível identificar com sucesso padrões não-triviais, distintivos, semanticamente correto e regras específicas do usuário (FERRAZ, 2008).

Alguns autores relatam, em seus trabalhos, como as ontologias estão sendo utilizadas para melhorar os resultados dos métodos de mineração de dados por regras de associação com a adoção de descritores semânticos (BELLANDI et al., 2007; FERRAZ, 2008; MARINICA, 2010; COUTO; SILVA, 2011; COELHO, 2012; VAVPETIC; LAVRAC, 2012; HAMANI et al., 2014; RIBEIRO, 2010).

Esses autores utilizam o tratamento semântico com ontologia de domínio para registro do conhecimento prévio e, assim, na fase de pré-processamento ou pós-processamento, introduzir restrições dos usuários agrupadas em dois tipos: restrições de poda destinada a filtrar itens desinteressantes, e restrições de abstração que permitam a generalização de itens em relação a conceitos de ontologias.

Segundo Ferraz (2008), a vantagem das restrições de poda é que permitem excluir, a partir do início, as informações que o usuário não está interessado. Cada regra mais geral deve ser capaz de substituir um número de regras específicas por meio de um processo de generalização. Quando isto é possível, há simultaneamente, uma melhoria semântica do conjunto de regras de associação mineradas e uma futura redução na cardinalidade do conjunto de regras.

As abordagens para poda e generalização de regras que utilizam cálculo do grau de similaridade semântica entre termos estão geralmente baseadas na estrutura hierarquia de grafos, usando propriedades de caminho (tamanho), sendo compatível com as ontologias que são representadas por grafos acíclicos orientados (DAG - *Directed Acyclic Graph*). Usualmente, esse tipo de abordagem se utiliza de relacionamentos do tipo “é-um” e “partetodo” para definir relações de subclasses e superclasses entre os conceitos presentes na hierarquia da ontologia.

As medidas de semelhança semântica estão normalmente restritas às relações de hiperonímia e meronímia, pois são estas que definem os ancestrais e os descendentes de um dado conceito (COUTO; SILVA, 2011).

Segundo Manda e Mccarthy (2013), os métodos que utilizam ontologias para calcular similaridade semântica são classificados em cinco categorias: métodos baseados em distância semântica, métodos baseados em conteúdo de informação, métodos baseados em propriedades de termos, métodos baseados na hierarquia da ontologia e métodos híbridos.

A Similaridade Resnik (RESNIK, 1999) utiliza o que é chamado de MICA (*most informative common ancestor*), ou seja, é o termo ontológico ancestral comum a dois termos que possui o maior conteúdo informacional (do inglês, *Information Content* – IC). O IC é definido como:

$$IC = -\log(p(c))(2)$$

em que $p(c)$ é a probabilidade de determinado termo ocorrer num dado cenário. O cálculo de $p(c)$ é feito somando-se todas as ocorrências do termo e dividindo-se pelo número de ocorrências totais. A similaridade entre dois termos é encontrada por meio da seguinte medida:

$$SimTermos(c1, c2) = IC(MICA)(2)$$

em que MICA refere-se ao ancestral comum mais informativo, ou seja, o ancestral calculado entre os pais de $c1$ e $c2$.

Couto e Silva (2011) ressalta, em suas pesquisas, que o conteúdo de informação que dois conceitos partilham é normalmente proporcional ao seu ancestral comum mais informativo (MICA - *Most Informative Common Ancestor*) no grafo, e o conteúdo de informação (IC - *In-*

formation Content) de um conceito é, normalmente, inversamente proporcional à sua frequência de ocorrência (numero de repetições). Por sua vez, a frequência de um conceito é também propagada para os seus ancestrais no grafo, o que faz com que o conteúdo de informação de um conceito seja proporcional à sua profundidade no grafo. Esse autor apresenta um novo método de computação de semelhança, denominado por DiShIn, que explora, de um modo eficiente, a herança múltipla presente numa ontologia.

Manda e Mccarthy (2013) apresentam uma abordagem de mineração de dados baseada na utilização de multi-ontologia em todos os níveis (MOAL), que utiliza a estrutura e os relacionamentos de uma Ontologia Genética para minerar as regras de associação. A aplicação MOAL propõe estratégias de poda para uso em diferentes aplicações que utilizam conhecimento da ontologia de domínio. O algoritmo consiste em três etapas principais:

- a) generaliza o conjunto de transações baseadas em relacionamentos transitivos;
- b) avalia e classifica as regras utilizando métricas de interesse;
- c) poda as regras altamente relacionados ou conhecidas. A ontologia categoriza as relações entre conceitos em três tipos: hiperonímia (“é um”), meronímia (“parte de”) e de regulação (restrições).

Hamani e Maamri (2009) aborda duas categorias principais de algoritmos para calcular a distância semântica entre termos, considerando estruturas hierárquicas, como as ontologias:

- a) considerando o número de arestas;
- b) baseadas no conteúdo da informação, inspirada na percepção de que pares de palavras que compartilham muitos contextos comuns são semanticamente relacionados.

A abordagem do método distância semântica, uma vez que, normalmente, produz uma medida da distância entre os dois termos, pode ser facilmente convertida em uma medida de similaridade.

Já Manda e Mccarthy (2013) apontam quatro fatores, que são, normalmente, considerados em métodos baseados em distância, como segue:

- a) densidade da ontologia: maior será a densidade quanto mais perto a distância entre os nós;
- b) profundezas de nós: quanto mais profundos os nós localizados, mais evidente a diferença;
- c) identificação dos tipos de ligações entre termos: o tipo normal “é-um” e outras relações, tais como “parte-todo”, sendo associados os pesos entre os nós;
- d) pesos de *links*: relacionar peso para as arestas que conectam a um certo nó ou com todos os nós filhos.

Segundo Ferraz (2008), as medidas de avaliação de regras de associação interessantes consideram a relação entre o antecedente e o conseqüente. Diz-se que uma instância está coberta por uma regra se o antecedente da regra for verdadeiro para a instância. Em particular, o que se pretende é utilizar relações ontológicas, do tipo “é um” (generalização e especialização) e “parte-de” (composição), como balizadoras semânticas, que permitam uma redução do número de regras de associação, geradas pela mineração de dados, de maneira mais eficaz do que as tentativas meramente sintáticas, que permitam, ao mesmo tempo, um enriquecimento semântico dos padrões obtidos após a fase de pós-processamento.

Vavpetic e Lavrac (2012) afirmam que o uso de ontologias permite que a extração de informação possa ser formulada a partir de condições que foram acordados pelos especialistas do domínio. Além disso, fornece os meios para generalizações mais eficazes dos termos extraídos, o que não seria possível usando apenas as instâncias descritas por meio de medidas objetivas.

3 CONCLUSÃO

A aplicação de métodos e tecnologias a partir da mineração de dados com embasamento semântico (ontologias) permite que informações estratégicas não explícitas dos domínios representados sejam descobertas e transformadas em conhecimento relevante para compreensão de problemas complexos e no processo de tomada de decisão

É importante destacar dois pontos de atenção sobre o uso de similaridade semântica com o uso de ontologias: primeiramente, a relevância de padrões extraídos por técnicas de identificação de similaridade semântica entre termos é altamente dependente da construção e validação das ontologias, portanto, é importante que as ontologias de domínio utilizadas tenham sido validadas por especialistas do negócio; segundo, a abordagem no processo de KDD precisa de maior aprofundamento sobre os algoritmos, não se restringindo apenas às relações ‘é-um’ e ‘parte de’, melhorando o uso da semântica formal proposta pelas ontologias na aplicação nos métodos computacionais da mineração de dados.

Conclui-se, por meio desse estudo, que o uso de ontologias no processo de KDD com regras de associação têm contribuído para o enriquecimento na extração de conhecimento inesperado e relevante para o negócio. A escolha de algoritmos adequados para cálculo da similaridade semântica de base ontológica entre os termos para obtenção de padrões sem redundância, aprimorando a revocação e precisão, torna-se o grande desafio no processo de recuperação da informação em base de dados.

Com o aumento da procura por métodos computacionais para resolver o problema da extração de conhecimento em grandes massas de dados, *Big Data*, o uso de ontologias pode facilitar a organização e recuperação das informações, ajudando a resolver os problemas de volume, variedade e veracidade. Portanto, a fusão das áreas de mineração de dados e repre-

sentação do conhecimento com o uso de ontologia de domínio é uma grande oportunidade para enfrentar o desafio de lidar com *Big Data*.

REFERÊNCIAS

- ALMEIDA, Mauricio Barcellos. Revisiting ontologies: a necessary clarification. **Journal of the American Society for Information Science and Technology**, v. 64, n. 8, p. 1682–1693, 2013.
- ALMEIDA, Mauricio B; BAX, Marcello P. Uma visão geral sobre ontologias: pesquisa sobre definições, tipos, aplicações, métodos de avaliação e de construção. *Ci.Inf.*, v. 32, n. 3, p. 7–20, 2003.
- BAEZA-YATES, R.; RIBEIRO-NETO, B. **Modern Information Retrieval**. 1. ed. New York: ACM Press, 1999.
- BELLANDI, A. et al. **Ontology-driven association rule extraction: a case study**. [S.l.: s.n.], 2007. Disponível em: <<http://ceur-ws.org/Vol-298/paper1.pdf>>. Acesso em: 10 de fev. 2014.
- CAMILO c.o. **Uma metodologia para mineração de regras de associação usando ontologias para integração de dados estruturados e não-estruturados**. 2010. Dissertação — Instituto de Informática, UFG, Goiânia., disponível em: <http://bdtd.ufg.br/tesesimplificado/tde_arquivos/6/TDE-2010-10-19T101221Z-1086/Publico/dissertacao> Acesso em: 01 de out. 2013.
- CAMILO c.o.; SILVA j.c. **Um estudo sobre a interação entre Mineração de Dados e Ontologias**. [S.l.: s.n.], 2010. Disponível em: <http://www.inf.ufg.br/sites/default/files/uploads/relatorios-tecnicos/RT-INF_002-09.pdf>. Acesso em: 01 de out. 2013.
- CAO, Longbing. **Domain-Driven Data Mining: Challenges and prospects**. [S.l.]: Transaction on Knowledge And Data Engineering, 2010. 755–769 p.
- COELHO, E.M.P. **Ontologias Difusas no Suporte à Mineração de Dados: aplicações na Secretaria de Finanças da Prefeitura Municipal de Belo Horizonte**. Belo Horizonte. 2012. Tese — Departamento de Ciência da Informação, UFMG, Belo Horizonte.
- COUTO, F.; SILVA, M. Disjunctive shared information between ontology concepts: application to gene ontology. **Journal of Biomedical Semantics**, v. 2, n. 5, 2011. Disponível em: <<http://dx.doi.org/10.1186/2041-1480-2-5>>. Acesso em: 01 de mai. 2014.
- DALFOVO, Oscar. **Quem tem informação é mais competitivo: o uso da informação pelos administradores e empreendedores que obtém vantagem competitiva**. Blumenau: Acadêmica, 2000.
- FAYYAD, Usama M et al. *Advances in knowledge discovery and data mining*. 1996.
- FERRAZ, I. **Conhecimento do mundo como instrumento enriquecedor dos resultados obtidos na mineração de dados**. 2008. Tese — UFF, Rio de Janeiro.
- GONÇALVES, E.C. Regras de associação e suas medidas de interesse objetivas e subjetivas. **INFOCOMP Journal of Computer Science**, 2005. Disponível em: <<http://www.dcc.ufla.br/infocomp/artigos/v4.1/art04.pdf>>. Acesso em: 1 de mar. 2014.
- GUARINO, N. Formal ontology and information systems. In: _____ **FORMAL ONTOLOGY IN INFORMATION SYSTEMS. Proceedings...** Amsterdam: IOS Press, 1998. p. 3–15.

HAMANI m.; MAAMRI r. **Ontology-Driven Method for Ranking Unexpected Rules**. [S.l.: s.n.], 2009. Disponível em: <<http://ceur-ws.org/Vol-547/71.pdf>>. Acesso em: 1 de mar. 2014.

HAMANI m. et al. **Unexpected rules using a conceptual distance based on fuzzy ontology**. [S.l.: s.n.], 2014. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S1319157813000141>>. Acesso em: 1 de mar. 2014.

HAN, J.; KAMBER, M. **Mineração de Dados: Conceitos e técnicas**. [S.l.: s.n.], 2006.

KASAMA d.; ZAVAGLIA c.; BARCELLOS g. **Do termo à estruturação semântica: representação ontológica do domínio da nanociência e nanotecnologia utilizando a estrutura quali**. [S.l.: s.n.], 2011. Disponível em: <http://www.erevistas.csic.es/ficha_articulo.php?url=oai:linguamatica.com:article/73&oai_iden=oai_revista909>. Acesso em: 1 de mar. 2014.

MANDA p.; MCCARTHY f. **Interestingness measures and strategies for mining multi-ontology multi-level association rules from gene ontology annotations for the discovery of new GO relationships**. [S.l.: s.n.], 2013. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/23850840>>. Acesso em: 1 de mar. 2014.

MARINICA, C. **Association Rule Interactive Post-processing using Rule Schemas and Ontologies - ARIPSO**. 2010. Tese — Department of Computer Science, Ecole polytechnique de l'Universite de Nantes, disponível em: <<http://www.claudiamarinica.com/publications.html>>. Acesso em: 1 de mar. 2014.

NAONE, E. **The new Big Data**. [S.l.: s.n.], 2014. Disponível em: <<http://www.technologyreview.com/news/425090/the-new-big-data/>>. Acesso em: 1 de mar. 2014.

NIGRO, H. O.; CISARO s. g.; XODO d. h. **Data Mining With Ontologies: Implementations, findings and frameworks, information science reference**. 2007. Imprint of — IGI Publishing, Hershey, PA.

RIBEIRO, L.S. **Uma abordagem semântica para seleção de atributos no processo de kdd**. [S.l.: s.n.], 2010. Disponível em: <<http://www.pggi.di.ufpb.br/?p=1691>>. Acesso em: 1 de out. 2013.

SILVA, D. **Estudo de funções de similaridade semântica de termos aplicadas a um domínio**. [S.l.: s.n.], 2014. Disponível em: <<http://www.cin.ufpe.br/tg/2007-2/dfs3.pdf>>. Acesso em: 1 de mai. 2014.

VAVPETIC, A.; LAVRAC, N. **Semantic Subgroup Discovery Systems and Workflows in the SDM-Toolkit**. [S.l.: s.n.], 2012. Disponível em: <<http://comjnl.oxfordjournals.org/content/early/2012/06/04/comjnl.bxs057>>. Acesso em: 1 de mai. 2014.

YOKOME, E.A. **Uma ontologia para inserir conhecimento humano em ferramentas de mineração de dados**. 2011. Dissertação — Departamento de Ciência da Computação, Unimep, Piraciaba., disponível em: <<https://www.unimep.br/phpg/bibdig/aluno/visualiza.php?cod=740>>. Acesso em: 01 de out. 2013.