

MODELO PARA SUPORTE À DESCOBERTA DE CONHECIMENTO EM BASE DE DADOS (KDD): APLICAÇÃO EM ESTRATÉGIAS NO MERCADO DE MEDICINA DIAGNÓSTICA¹

Lucelia Pinto Branquinho²
Renata Maria Abrantes Baracho³
Mauricio Barcellos Almeida⁴

RESUMO:

O objetivo deste trabalho é descrever um modelo que faz uso de ontologias biomédicas de doenças e testes laboratoriais para aprimorar o processo de *Knowledge Discovery in Database* (KDD) tornando mais eficiente a recuperação da informação sobre o comportamento de prescrição de testes laboratoriais, no caso deste experimento, relacionados às hepatites virais. O modelo desenvolvido instancia uma ontologia de domínio de testes complementares das hepatites virais para generalizar os atributos, na fase de pré-processamento, e posteriormente, classificar as regras de associação obtidas considerando a similaridade semântica entre o antecedente e consequente.

Palavras-chave: Regras de associação. Ontologia. Medicina Diagnóstica.

ABSTRACT:

The objective of this study is to describe a proposal for use of biomedical ontologies disease and laboratory tests on the process with Knowledge Discovery in Database (KDD) to make more effective information retrieval on the prescribing behavior of laboratory tests in the case of this experiment related to viral hepatitis. The model developed instantiates an ontology of further testing of viral hepatitis to generalize the attributes, in pre-processing phase, and post-processing phase classify the association rules obtained considering the semantic similarity between the antecedent and consequent. The domain ontologies are used to advance knowledge and introduce restrictions on users allowing pruning and generalization rules, filtering interesting items.

Keywords: Association Rules. Ontologies. Medical diagnosis.

¹ Artigo publicado sobre o assunto em conferência em Proceedings of the 19th World Multiconference on Systemics, Cybernetics and Informatics (WMSCI 2015), Orlando, US, conforme citado nas referências.

² Mestranda em Ciência da Informação, pelo Programa de Pós-graduação em Ciência da Informação da Universidade Federal de Minas Gerais, Brasil.

³ Professor pelo Programa de Pós-graduação em Ciência da Informação da Universidade Federal de Minas Gerais, Brasil.

⁴ Professor pelo Programa de Pós-graduação em Ciência da Informação da Universidade Federal de Minas Gerais, Brasil.

1 INTRODUÇÃO

As organizações precisam obter informações úteis aos seus objetivos para criar estratégias adequadas e promover a inovação agregando valor à tomada de decisão. Uma das técnicas para extração do conhecimento referenciada na literatura é o *Knowledge Discovery in Database* (KDD)

Neste sentido, as tecnologias voltadas para coleta, armazenamento, recuperação e disponibilização de dados vêm evoluindo através de técnicas, métodos e ferramentas computacionais automatizadas.

A extração de conhecimento com adoção de ontologias biomédicas permite incorporar à mineração de dados medidas subjetivas para generalização e classificação com base na hierarquia e relações que estes termos compartilham trazendo maior efetividade e relevância na classificação dos padrões de regras de associação.

A aplicação de métodos e tecnologias para mineração de dados semântica 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. Ao integrar métodos estatísticos com abordagem semântica é possível identificar com sucesso padrões não triviais, semanticamente corretos, e mais relevantes e específicos (FERRAZ, 2008).

As próximas seções serão organizadas da seguinte maneira: a seção 2 detalha o problema de pesquisa; a seção 3 remete a uma breve descrição sobre a mineração de dados por regra de associação e o uso de ontologias neste processo; já na seção 4 detalha o procedimento metodológico do modelo proposto. Já na seção 4, relata os resultados do modelo que usa as ontologias para obter padrões mais relevantes para a mineração por regras de associação. Finalmente, a seção 6 apresenta considerações finais e as possibilidades de pesquisas futuras.

2 DESCRIÇÃO DO PROBLEMA

A comunidade de mineração de dados se confronta com o desafio de explorar o grande volume de dados utilizando recurso de conhecimento de domínio dos dados semanticamente anotados para obter maior precisão e revocação na descoberta de conhecimento.

Uma opção para resolver o problema da relevância dos dados obtidos é reduzir os

dados a serem minerados através de pesquisa de representações dos conjuntos freqüentes e regras “fortes” mais coesas. A redução do conjunto das regras pode ser feita pelo mecanismo de poda. A questão é: Quais as regras que devem ser podadas?

Segundo Ferraz (2008), há duas abordagens para poda das regras:

- 1) Medidas objetivas com uso de técnicas estatística;
- 2) Medidas subjetivas com dependência da ação do usuário, pois depende do conhecimento do domínio.

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 de forma que possibilitem o processamento por computadores reduzindo a dependência da presença do especialista do domínio.

Nesse aspecto, a Ciência da Informação estrutura a aquisição da informação e do conhecimento na organização estabelecendo princípios e norteando a resolução do problema da representação do conhecimento com a adoção de ontologias para auxiliar o processo de KDD.

Para o mercado de medicina diagnóstica é importante entender o comportamento de prescrição dos médicos no diagnóstico das doenças para antecipar tendências e assim realizar ações de marketing e vendas direcionadas ao mercado.

Considerando a necessidade do mercado de medicina diagnóstica, este estudo tem como objetivo aprimorar o processo de mineração de dados através da introdução do conhecimento do domínio na fase de pré-processamento e pós-processamento à mineração de dados.

3 KDD COM O USO DE ONTOLOGIAS

3.1 KDD

Segundo Fayyad et al (1996) “Extração de Conhecimento de Base de Dados (KDD) é o processo de identificação de padrões válidos, novos, potencialmente úteis e compreensíveis embutidos nos dados”.

Para um melhor entendimento o processo de KDD pode ser agrupado em três fases: pré-processamento, mineração de dados (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 deste conhecimento.

Em geral, o conhecimento descoberto através de processos de KDD é expresso na forma de regras e padrões. Na busca de padrões são escolhidas técnicas e algoritmos. A escolha da técnica de MD apropriada depende do tipo de tarefa de KDD a ser realizada. Dentre os diferentes tipos de técnicas que podem ser utilizadas para extração em bases de dados encontram-se as regras de associação, classificação entre outros.

3.2 MINERAÇÃO DE DADOS POR REGRAS DE ASSOCIAÇÃO

A mineração por regras de associação é uma das técnicas mais utilizadas sendo a tarefa de associação realizada por meio de algoritmos que geram padrões que caracterizam o quanto a presença de um conjunto de itens, nos registros de uma base de dados, implica na presença de outro conjunto distinto de itens, nos mesmos registros (FERRAZ, 2008).

A partir de conjuntos de elementos que aparecem juntos com pelo menos alguma frequência (suporte), as regras de associação representam a condição "SE antecedente ENTÃO consequente" com garantia probabilística (confiança), que sempre que o antecedente ocorrer o consequente também estará presente (FERRAZ, 2008).

O suporte de uma regra $P \rightarrow Q$ é definido como o percentual de transações da base de dados em que o antecedente e consequente da regra aparecem na mesma transação:

$$\text{Suporte } (P \rightarrow Q) = \frac{\text{total de transações com ocorrências } (P \cup Q)}{\text{total de transações}}$$

A confiança de uma regra $P \rightarrow Q$ é definida como o percentual de transações, dentre as que possuem o antecedente da regra, em que antecedente e consequente aparecem conjuntamente na mesma transação.

$$\text{Confiança } (P \rightarrow Q) = \frac{\text{total de transações com ocorrências } (P \cup Q)}{\text{total de transações com ocorrências } P}$$

Para mensurar a dependência entre os itens há métricas como o *Lift* (GONÇALVES, 2011). Esta métrica é utilizada para avaliar dependências entre antecedente e consequente, quanto maior o valor do *lift*, mais interessante a regra, pois A aumentou ("lifted") B numa maior taxa.

Seja D uma base de dados de transações. Seja $A \Rightarrow B$ uma regra de associação obtida

a partir de D. Dada uma regra de associação $A \Rightarrow B$, esta medida indica o quanto mais frequente torna-se B quando A ocorre. O valor do *lift* para $A \Rightarrow B$ é computado por:

$$\text{Lift}(A \rightarrow B) = \frac{\text{Confiança}(A \rightarrow B)}{\text{Suporte}(B)}$$

O exemplo clássico de regra de associação é o de “cestas de compras” na área de marketing com a intenção de estimular a compra de itens que são comprados juntos.

3.3 KDD COM USO DE ONTOLOGIAS

Segundo Nigro et al. (2007), um dos problemas mais importantes e desafiadores do processo de KDD é a definição do conhecimento prévio; este pode ser originado a partir do processo ou do domínio.

As ontologias são utilizadas para mapear o conhecimento do mundo possibilitando o uso e manipulação por máquinas permitindo, por exemplo, inferir novos conhecimentos sobre o domínio pesquisa do resultando assim em uma melhor automação da Descoberta de Conhecimento, seja ela em qualquer uma das fases que constitui o processo de KDD (HAMANI, 2014).

A mineração de dados por regras de associação tradicional utiliza medidas objetivas estatísticas para obter o conjunto de itens frequentes, ou seja, desconsidera o significado de cada item ou instância. O conteúdo semântico extraído das ontologias permite a inserção de mais inteligência e conhecimento ao processo mineração de dados melhorando sua precisão (FERRAZ, 2008).

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. O conhecimento prévio de um domínio ou de um processo na área de mineração de dados pode ajudar a selecionar informações mais apropriadas (pré-processamento), diminuir o conjunto de dados a serem mineradas (processamento) e representar resultados de uma forma mais compreensível (pós-processamento) (HAMANI, 2014).

Segundo Nigro et al. (2007), através da representação do conhecimento por ontologia é possível transformar a MD em “mineração de conhecimento”.

4 PROCEDIMENTOS METODOLÓGICOS

A abordagem proposta consiste na hipótese de que os dados contidos nas bases de dados possuem uma relação com os conceitos envolvidos no negócio e os relacionamentos existentes entre eles o que conduzirá a descoberta de semântica implícita conduzindo à seleção dos subconjuntos de dados que devem ser minerados.

Baseada na metodologia de pesquisa científica de Lakatos e Marconi (1991) e Gil (1994), a pesquisa pode ser classificada como pesquisa aplicada, com abordagem qualitativa e quantitativa, que tem como objetivo explicar um experimento que busca comprovar a melhoria na recuperação de informação.

4.1 CONTEXTO

A empresa estudada é uma organização nacional fundada nos anos 50 que atua no segmento de medicina diagnóstica. A medicina diagnóstica contempla diferentes especialidades direcionadas à realização de exames complementares no auxílio ao diagnóstico médico.

Em função da inexistência de divulgação de dados consolidados sobre o mercado de medicina diagnóstica atendida pela saúde suplementar (ANS), a empresa em questão busca aprimorar seu monitoramento ambiental para melhor entender a real demanda por serviços neste segmento direcionando suas ações comerciais e de marketing. A recuperação de informação sobre o mercado utilizando análise da base dados da empresa é uma excelente oportunidade para extrair conhecimento.

Ao obter informações mais interessantes o especialista KDD poderá ajudar na geração de conhecimento pelas equipes de Inteligência competitiva, comercial, produtos, novos negócios e de marketing na segmentação dos produtos no mercado.

Neste estudo optou-se por analisar a prescrição médica dos testes laboratoriais relacionados ao diagnóstico de hepatites virais.

4.2 MODELAGEM ONTOLÓGICA

A ontologia, criada e disponibilizada como Ontologia de testes laboratoriais para Hepatite Viral (HVO), propõe-se a representar um domínio específico para exames complementares relacionados ao diagnóstico de hepatites virais. Foi desenvolvida usando a

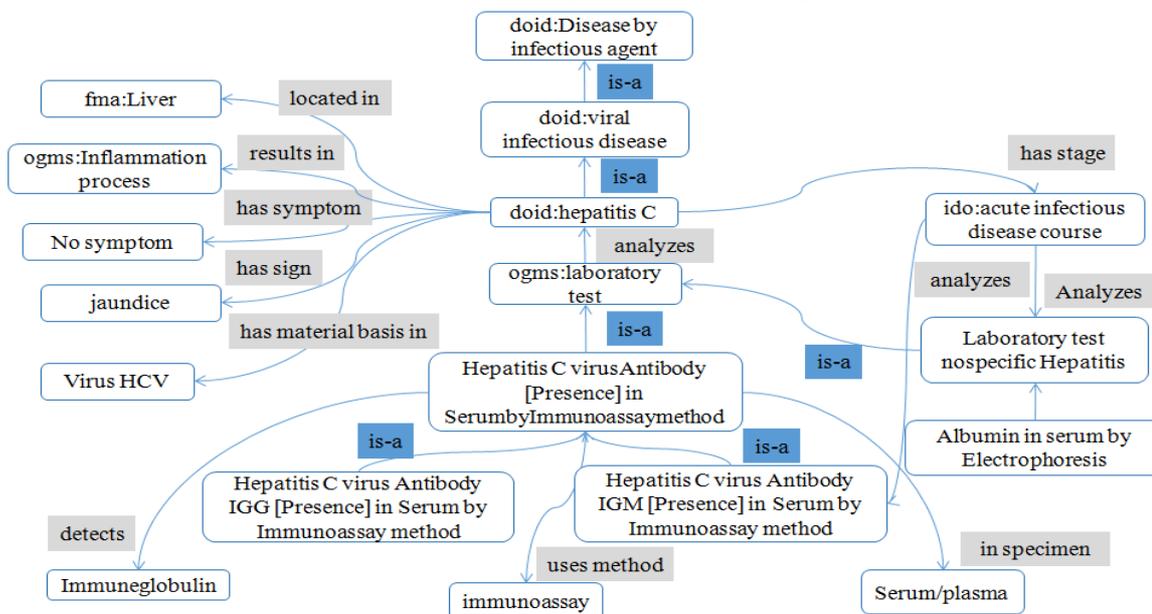
linguagem OWL, especificamente a OWL-DL (BAADER et al., 2003).

Seguindo o modelo proposto por Coelho (COELHO; ALMEIDA, 2012), a aquisição de conhecimento sobre o domínio foi conduzida através de um levantamento inicial com a pesquisa de dados secundários. Em um segundo momento foi realizado o roteiro da entrevista com os especialistas que objetivou a coleta de informações para se obter respostas para validar o conhecimento existente e esclarecer dúvidas e mal entendidos.

As fontes de conhecimento utilizadas, os termos empregados e suas respectivas descrições são provenientes de publicações relacionadas ao diagnóstico de hepatites virais (livros, periódicos nacionais/internacionais, LOINC, SNOMED-CT, e do Ministério da Saúde brasileiro, entre outros). Vale ressaltar que grande parte das entidades descritas na HVO são provenientes das ontologias OGMS, IDO5, DOID, OBI6 e FMA.

Para cada exame complementar do LOINC foi criada uma subclasse associada à classe de testes complementares (OGMS: laboratory test). Na propriedade de anotação das classes criadas foram relacionadas às informações referentes à identificação do exame complementar, informações adicionais provenientes da relação com a doença, os sintomas, sinais, transmissão e a relação referente ao órgão/sistema o qual a doença predominante ataca, conforme exemplificado na Figura 1.

Figura 1 - Exemplo classificação hepatite C



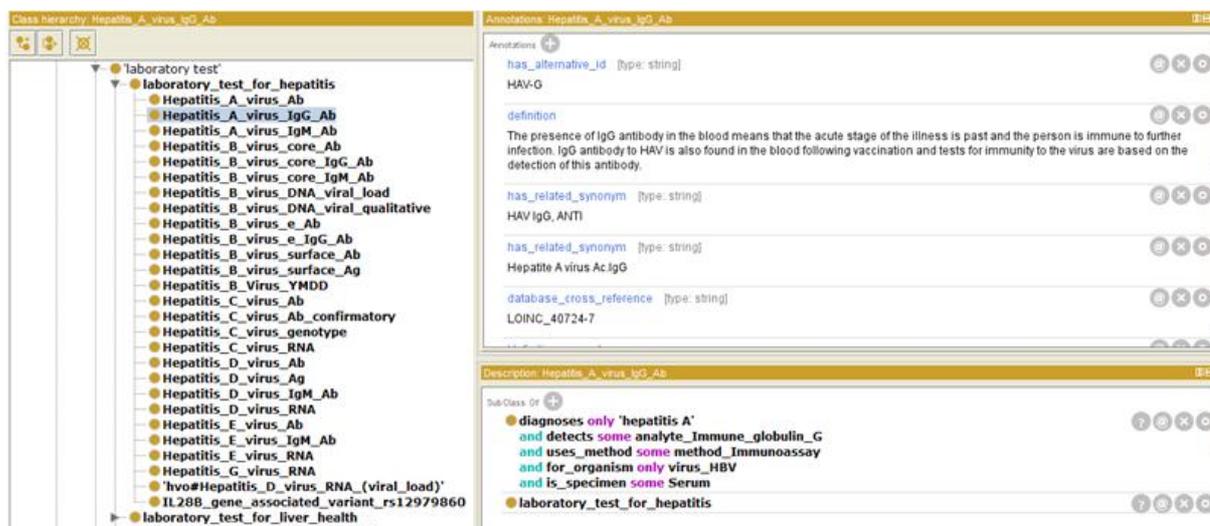
Fonte: Adaptado de EILBECK et al., 2013.

⁵ Disponível em: <http://infectiousdiseaseontology.org/page/Main_Page>. Acesso em: 1 mai. 2014.

⁶ Disponível em: <http://obi-ontology.org/page/Main_Page>. Acesso em: 1 mai. 2014.

Nos axiomas de equivalência (descrito por meio de restrições existenciais) é detalhada uma parte do processo de diagnóstico da doença sendo composta por, pelo menos, um pedido médico (nesse caso classes de *hvo: laboratory_testing_encounter*) o qual é composto de exames complementares (*ogms: laboratory test*) para diagnóstico da doença (*doid: hepatitis A*), conforme exemplificado na Figura 1 e Figura 2.

Figura 2 – Relação de exames complementares – Anotações e relações em Protegé v.5⁷



Fonte: Elaborado pelo próprio autor.

4.3 KDD COM ONTOLOGIAS

No levantamento dos dados foram extraídos da base de dados da empresa de medicina diagnóstica os pedidos e seus respectivos testes laboratoriais. O conjunto de dados foi armazenado no banco de dados PostgreSQL 9.4 contendo quatro tabelas: pedido, exame, pedido_exame, pedido_inferencia e exames_não_inferidos.

Após as fases de compreensão do domínio e entendimento dos dados, foi realizada a fase de preparação dos dados. O objetivo principal dessa fase foi a seleção, seguida da preparação dos dados armazenados no sistema de gerenciamento de banco de dados, para um formato adequado a aplicação do algoritmo *Apriori* (AGRAWAL et al., 1993), utilizando a linguagem R. No Quadro 1 estão representados alguns pedidos.

⁷ Disponível em: <protegewiki.stanford.edu>. Acesso em: 1 .mar. 2014.

Quadro 1 – Relação de pedidos de exames – Unidades

Ped	DtPed	Idade	Sexo	Exames
8000001	01/01/2015	37 ^a	F	UR
8000028	01/01/2015	42 ^a	F	AU CBI CHLA-G CHLA-M E2 FSH GS-HTLV1
8000101	02/01/2015	49 ^a	M	HCV

Fonte: Elaborado pelo próprio autor.

Um aplicativo Java utilizando o *framework* Jena⁸ e o *plugin* Pellet⁹ foi construído para importar o modelo da ontologia e realizar a inferência dos pedidos utilizando o arquivo HVO.OWL (representa a ontologia) salvo em RDF.

Esta aplicação lê os pedidos e exames complementares do banco de dados Postgree e cria as instâncias considerando a classe ‘*hvo.laboratory test*’ e suas subclasses e interpreta os axiomas de equivalência das classes de diagnóstico de hepatites virais gerados pelo motor de inferência Pellet o qual identifica os termos que podem ser generalizados, conforme exemplificado no Quadro 2.

Quadro 2 – Relação de pedidos com generalização de exames

Pedido	Exame 1	Exame2	Exame N	Exames generalizados
8000022	ELISAG	ELISAM	HTLV1	Exames Hepatite A
8000101	25-VD3	ELFC	CRE	Exames Hepatite G

Fonte: Elaborado pelo próprio autor.

Após generalização dos atributos os padrões foram extraídos com a execução da mineração de dados por regra de associação através do pacote Arules¹⁰ que adota o algoritmo Apriori. Esta técnica utiliza duas informações para a inferência de regras de associação: suporte e confiança. Contudo, nem sempre essas métricas são suficientes para determinar se padrões encontrados nos dados são significativos. Portanto, além de suporte e confiança foi utilizada a métrica *Lift*. A Tabela 1 apresenta um exemplo dos padrões obtidos na fase de pós-processamento.

⁸Disponível em: <<https://jena.apache.org/>>. Acesso em: 10.mai.2014.

⁹Disponível em: <<https://github.com/complexible/pellet>> Acesso em: 10.mai.2014.

¹⁰ Disponível em: <<https://cran.r-project.org/web/packages/arulesViz/index.html>>. Acesso em: 10.mai.2014.

Tabela 1 - Exemplo de regras de associação – Unidade Terceirizados

Regras	Suporte	Confiança	Lift
{exame31=ALUM,exame367=PTH} => {exame461=HEPATITE B}	0.014	1	1,10
{exame398=T4-RIE,exame417=TSH-B} => {exame461=HEPATITE B}	0.010	0.925	1,02

Fonte: Elaborado pelo próprio autor.

Após o pós-processamento, foi utilizado o cálculo de similaridade semântica entre o antecedente e o conseqüente para as regras de associação obtidas de forma a classificar os padrões mais relevantes utilizando o modelo de Tversky (1977) através de uma aplicação Java (Harispe, 2013).

Ao *lift* da regra foi somado o valor obtido com o cálculo da similaridade entre os exames complementares do antecedente e conseqüente para classificar os padrões de regra de associação obtidos.

5 RESULTADOS

O desenvolvimento da ontologia com os respectivos mecanismos de inferência para atuar na pré – mineração reduziu o número de atributos a serem minerados pelo algoritmo de regras de associação Apriori, melhorando a desempenho do minerador, devido redução das combinações e, conseqüentemente, do número de resultados no pós-processamento facilitando a análise dos resultados e classificação, conforme apresentado na tabela 2.

Tabela 2 - Atendimentos e regras de associação

Base	Atendimentos em que constam exames hepatites virais			Qtde regras sem ontologia	Qtde regras com ontologias
	Jan	Fev	Mar		
Unidades	927	819	13	573	258
Terceirizados	7652	7698	17331	221	115

Fonte: Elaborada pelo próprio autor.

Uma segunda contribuição é que a generalização dos termos possibilita um maior significado aos resultados capaz de orientar o processo de análise na fase de pós-mineração.

Nesse aspecto, este método exerce uma função de métrica subjetiva dos resultados obtidos.

A partir do conjunto final de dados, já tratados, foram determinadas as regras de associação mais relevantes, conforme exemplificado na tabela 3, através da classificação pelo resultado do cálculo de similaridade semântica entre os termos.

Tabela 3 - Regras de associação classificadas - Lift + SSM

Antecedente	Consequente	Lift	SSM	Resultado
Glicose, HIV 1 e 2 Ab Ag, TSH-B	Hepatitis C	1,24	0,238	1,478
Glicose, Urina rotina, VDRL	Hepatitis B	1,04	0,285	1,325

Fonte: Elaborado pelo próprio autor.

6 CONCLUSÃO

HVO foi desenvolvida empregando a linguagem de ontologias OWL, especificamente a sublíngua OWL-DL, o que a faz ter decidibilidade computacional, e definições de linguagem natural mais concisas.

As ontologias permitem descrever a estrutura complexa do domínio biomédico possibilitando que as dificuldades oriundas do aumento na quantidade de dados em nível exponencial e da ausência de ferramentas para descoberta de conhecimento em domínios especializados sejam minimizadas. O uso da ontologia de domínio no processo de KDD possibilitou reduzir o número de regras e aumentar a precisão e relevância dos padrões obtidos.

A utilização de ontologias nas fases de pré e pós- processamento à MD por regras de associação agregou o conhecimento de domínio e propiciou suporte semântico auxiliando o analista a explicar e interpretar as regras obtidas; possibilitando categorizar e classificar as regras geradas através da generalização e poda.

Com os resultados obtidos é possível mostrar como as diversas tecnologias ligadas ao processo de KDD e organização do conhecimento pode apoiar a tomada de decisão. A inclusão de embasamento semântico (ontologias) permite que informações estratégicas sejam descobertas e transformadas em conhecimento relevante para compreensão de forma a entender, por exemplo, o comportamento médico quando da solicitação de testes complementares ao diagnóstico médico.

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 uma grande oportunidade de aprimorar o processo de recuperação da informação em base de dados.

REFERÊNCIAS

- AGRAWAL, R.; IMIELINSKI, T.; SWAMI, A. N. Mining association rules between sets of items in large databases. In: 20th VLDB CONFERENCE, 1993, Santiago. Proceedings...Ed. P. Buneman e S. Jajodia, Santiago: ACM Press Santiago, 1993. p. 207–216. Disponível em: <<http://rakesh.agrawal-family.com/papers/vldb94apriori.pdf>> Acesso em: 1 mar. 2014.
- ALMEIDA, M. B.; BAX, M. P. Uma visão geral sobre ontologias: pesquisa sobre definições, tipos, aplicações, métodos de avaliação e de construção. Ci.Inf., Brasília, v. 32, n. 3, p. 7-20, 2003.
- ALMEIDA, M. B.; TEIXEIRA, L. M. D.; COELHO, K. K.; SOUZA, R. R. Aportes linguísticos no apoio a construção de ontologias. Liinc em Revista, v.6, n.2, p.384-410, set. 2010.
- BARACHO, R. M. A.; BRANQUINHO, L.; ALMEIDA, M.B.; SOUZA, R.R. (2015). Using Ontologies and Inference Engine in Association Rule of Data Mining: Application in Sales Strategies in Medical Laboratory Diagnostic Market. Publicado em Proceedings of the 19th World Multiconference on Systemics, Cybernetics and Informatics(WMSCI 2015), Orlando, US.
- BARACHO, R. M. A.; BRANQUINHO, L.; ALMEIDA, M.B. (2015). Ontologies in support of data-mining based on associated rules: a case study in a medical diagnosis company. Publicado poster em Anais do 7º Ontobras, São Paulo, BR.
- 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, 2012.
- FAYYAD, U. M.; PIATETSKY-SHAPIRO, G.; SMYTH, P.; UTHURUSAMY, R.. Advances in Knowledge Discovery and Data Mining. 1996
- FERRAZ, I. Conhecimento do mundo como instrumento enriquecedor dos resultados obtidos na mineração de dados. Tese - UFF, Rio de Janeiro, 2008
- GAN, M., DOU, X., JIANG, R. From ontology to semantic similarity: calculation of ontology-based semantic similarity. Disponível em: <<http://www.hindawi.com/journals/tswj/2013/793091/>>. Acesso em: 10 mai. 2014.
- GONÇALVES, E.C. Data mining com a ferramenta WEKA. III Fórum de Software Livre de Duque de Caxias (III FSLDC). Escola Nacional de Ciências Estatísticas IBGE/ENCE). 2011. Disponível em: <http://forumsoftwarelivre.com.br/2011/arquivos/palestras/DataMining__Weka.pdf>. Acesso em 01 out. 2013.
- HAMANI, M., MAAMRI, R., KISSOUM, Y., SEDRATI, M., Unexpected rules using a conceptual distance based on fuzzy ontology, 2014. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S1319157813000141>> Acesso em: 1 mar. 2014

HARISPE, Sébastien. A framework for unifying ontology-based semantic similarity measures: a study in the biomedical domain. *Journal of biomedical informatics*, v. 48, p. 38-53, 2013. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S1532046413001834>>. Acesso em: 1 fev. 2015.

NIGRO, H. O.; CISARO S. G.; XODO, D. H. *Data Mining With Ontologies: Implementations, Findings and Frameworks*, Information Science Reference - Imprint of: IGI Publishing, Hershey, PA, 2007.

RIBEIRO, L.S. Uma abordagem semântica para seleção de atributos no processo de kdd. Disponível em: <<http://www.ppgi.di.ufpb.br/?p=1691>> Acesso em: 01 out. 2013.

VICKERY, B.C. Ontologies. *Journal of Information Science*, v. 23. n. 4, p. 227-86, 1997.