

Data Mining: Estudo bibliométrico dos últimos 10 anos na área da saúde

Bruno Samways dos Santos (UTFPR) brunosantos@utfpr.edu.br
Maria Teresinha Arns Steiner (PUC-PR) maria.steiner@pucpr.br
Vagner Jorge Neckel (UNISOCIESC) vagner.neckel@sociesc.com.br

Resumo:

A análise de produção científica relacionada a uma dada área do conhecimento em determinado período, contribui para a continuidade dos estudos e evolução do tema específico. O objetivo principal desse estudo, portanto, é analisar quantitativamente o que se produziu em relação ao tema Data Mining (DM) nos últimos 10 anos, especificamente na área da saúde. O estudo foi realizado com aplicações das técnicas bibliométricas, através da consulta à base SCOPUS e seu tratamento dos dados pelos softwares Microsoft Excel® 2013, Bibexcel e VOSviewer. Para a pesquisa utilizou-se “data mining” e “healthcare” como expressões-chave e definição dos termos de busca, além de outras atividades para definição da amostra, gerando uma amostra constituída por 455 artigos publicados. Nos resultados do estudo foram apresentadas por meio de estatística descritiva, a identificação dos autores, artigos e periódicos mais referenciados nos artigos analisados. Além das principais redes de relacionamento de autores e citações; nuvem de palavras-chave e mapa de densidade de cocitação, informações que tornam este trabalho importante para o desenvolvimento e tratativa das perspectivas de abordagens sobre o DM em diversas áreas.

Palavras chave: Bibliometria, Data Mining, Saúde.

Data Mining: bibliometric study of the past 10 years in healthcare

Abstract

The analysis of scientific production related to a given area of knowledge in a given period, contributes to the further study and development of specific subject. The main objective of this study is therefore quantitatively analyze what is produced by Topic Data Mining (DM) in the last 10 years, particularly in health. The study was conducted with application of bibliometric techniques, through consultation to SCOPUS and treatment of data by Microsoft Excel® software 2013 Bibexcel and VOSviewer. For research was used "data mining" and "healthcare" as key expressions and definition of search terms, as well as other activities to define the sample, generating a sample of 455 articles published. The results of the study were presented by descriptive statistics, identifying the authors, articles and journals most referenced in the articles analyzed. Besides the main social networks of authors and citations; cloud of keywords and cocitação density map, information that make this important work for the development and dealings prospects of approaches to the DM in several areas.

Key-words: Bibliometrics, Data Mining, Healthcare.

1. Introdução

Uma grande quantidade de dados armazenados em qualquer tipo de organização, como bancos, indústrias, serviços de transporte e hospitalares, entre outros, como fonte de informação muitas vezes não percebida pelos usuários destes sistemas. As técnicas que buscam transformar estes dados em conhecimento são objetivas da área denominada Descoberta de Conhecimentos em Bases de Dados (Knowledge Discovery in Databases – KDD) (STEINER et al., 2006).

O KDD identifica padrões de dados válidos de maneira não trivial, transformando-os em dados possivelmente úteis e interpretáveis (TRONCHONI et al., 2010). É um processo importante para a tomada de decisão uma vez que permite a aplicação de técnicas para auxiliar em todo o processo de descobrimento de dados, as quais eram inicialmente aplicadas para fins comerciais apenas (CARVALHO et al., 2012).

Para Freitas (2000) o conhecimento deve satisfazer três propriedades: ser o mais correto possível; compreensível e útil, novo. Segundo Fayyad et al. (1996), o KDD é um conjunto composto, basicamente, por 5 atividades contínuas e sequenciais com relação aos dados: seleção; pré-processamento e limpeza; transformação; mineração dos dados (ou data mining; DM ou, ainda, Big Data) e interpretação dos resultados (Figura 1).

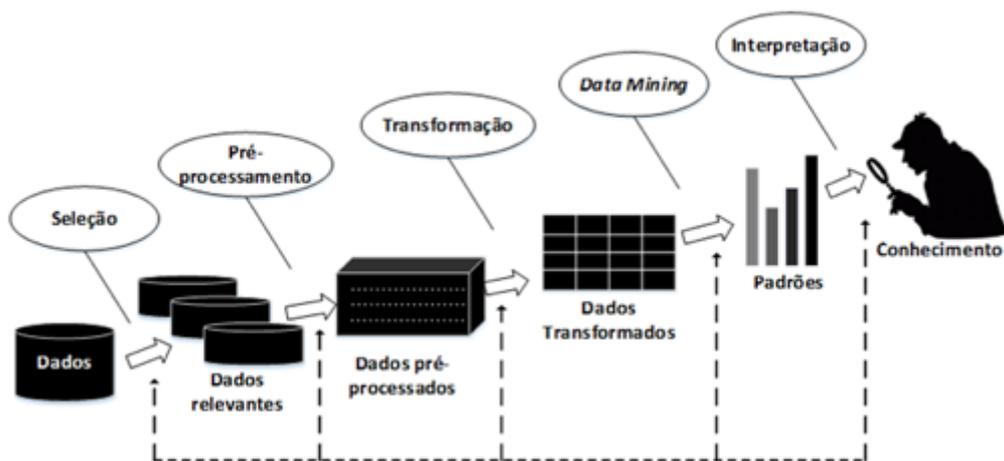


Figura 1 – Conjunto de atividades KDD

Fonte: Adaptado de Fayyad *et al.* (1996)

Para a análise de dados, vários algoritmos e procedimentos podem ser utilizados para selecionar aqueles que realmente gerarão conhecimento, dentre os quais podem ser citados: i) Testes: t-student, T^2 de Hotelling, ANOVA, MANOVA; ii) Normalização e Padronização; iii) Análise das Componentes Principais (ACP), Análise Fatorial (AF) e Transformação dos atributos; iv) Codificação dos atributos. Desta forma, é importante a seleção criteriosa dos dados para reformular o banco de dados e descartar aqueles que não agregam ou tenham sido colocados em uma planilha de forma equivocada.

O objetivo do presente artigo é apresentar um estudo bibliométrico das aplicações de DM na área da saúde nos últimos 10 anos. O trabalho está estruturado na seguinte forma: capítulo 2 possui um referencial sobre Data Mining; o capítulo 3 tem o descritivo da metodologia; o capítulo 4 explora os resultados obtidos e suas discussões; o capítulo 5 contém as considerações e; por fim, as referências bibliográficas.

2. Data Mining

O data mining é uma das etapas do processo KDD, que se tornou a ferramenta mais conhecida do mesmo, pois aplica metodologia que atende à preparação e exploração dos dados, interpretação dos resultados e a percepção para os conhecimentos minerados. Observando o crescente número de dados na área da saúde, o data mining pode ser uma ferramenta de grande importância na análise do conhecimento dos dados e auxiliar os gestores de saúde nas tomadas de decisões voltadas para estudos, prevenção e promoção da saúde (Souza, A.M.P. e Zaia, J.E. 2015). De acordo com Tan, Steinbach e Kumar (2009), DM é o processo de descoberta de informações úteis em grandes quantidades de dados. Pode-se afirmar então que não é necessário tratar manualmente dados de diferentes origens para a tomada de decisão, pois algumas técnicas matemáticas e computacionais específicas podem realizar esta tarefa com uma precisão significativa. Como complemento, DM pode ser definido como o processo para descobrir "padrões" nos dados (WITTEN; FRANK, 2005).

PhridviRaja e GuruRao (2014) ressaltam a diferença entre os dados em uma base de dados e em um armazém de dados, sendo que no primeiro, os dados estão em formato estruturado enquanto que no segundo não necessariamente.

Para a utilização das técnicas de DM, os dados devem estar previamente estruturados para processamento. A mineração de dados pode ser utilizada em diferentes situações que dependem do objetivo da aplicação: classificação e previsão; associação; análise de grupos (ou clusterização) (TAN; STEINBACH; KUMAR, 2009).

3. Metodologia

A bibliometria analisa a distribuição da bibliografia em um determinado campo do conhecimento, juntamente com o seu crescimento (PILKINGTON; MEREDITH, 2009). De acordo com Bufrem e Prates (2005), três leis básicas são importantes para uma análise bibliométrica:

- * Lei de Zipf: mensura frequência de aparições de palavras;
- * Lei de Lotka: aborda a produtividade e acordo com os autores, e;
- * Lei de Bradford: refere-se à produtividade quanto aos periódicos.

Para a condução do presente estudo bibliométrico, a etapa inicial que descreve o protocolo de busca está de acordo com a figura 2.

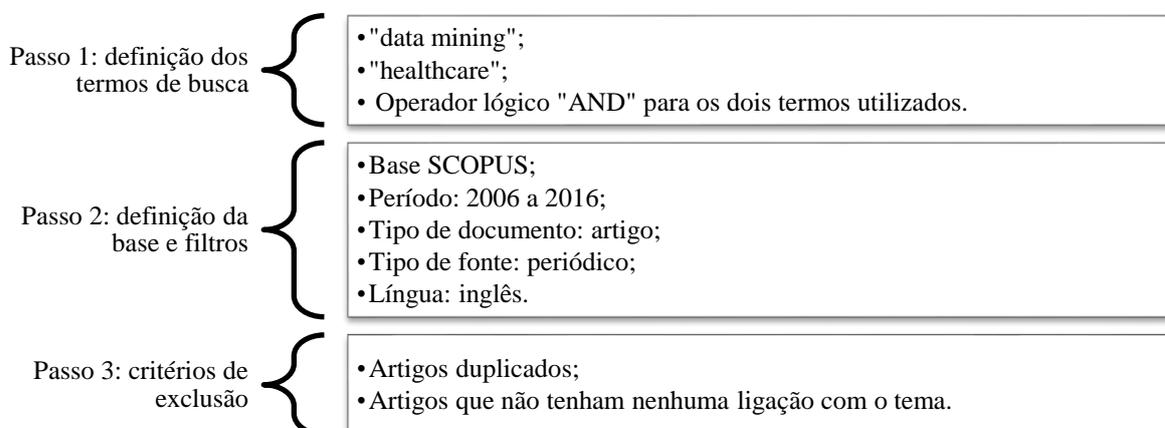


Figura 2 – Atividades sequenciais para a definição da amostra

A base SCOPUS foi utilizada por ser uma das mais utilizadas para trabalhos de estudo bibliométrico e também ter opções de seleção e salvamento de documentos em formatos específicos para os softwares utilizados no estudo bibliométrico.

Após a definição do protocolo, iniciou-se a pesquisa propriamente dita. O tamanho da amostra foi definida de acordo com a Figura 3.

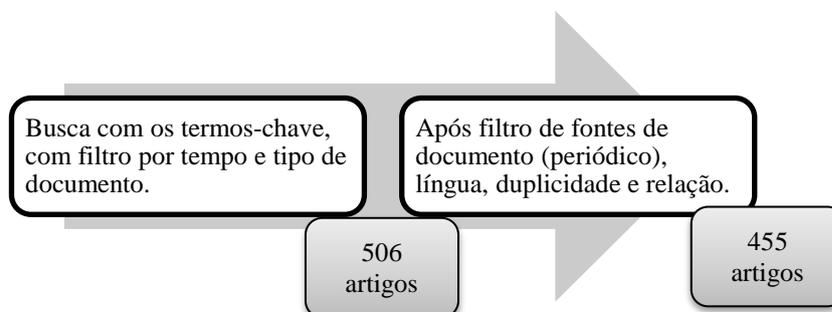


Figura 3 – Quantitativo de artigos para análise.

Após o tamanho da amostra ter sido definido, os mesmos foram tratados pelos softwares Microsoft Excel® 2013, Bibexcel e VOSviewer.

O Excel é um dos programas mais conhecidos no mundo para tratamento de dados em planilhas. O Bibexcel é um programa para análise de dados bibliográficos de qualquer dado de origem textual em um formato específico (PERSSON; DANELL; WIBORG SCHNEIDER, 2009). O VOSviewer foi criado para a montagem de mapas baseado em redes de dados, visualizar e explorar tais dados (VAN ECK; WALTMAN, 2016).

Em um primeiro momento, a base de dados foi colocada no Microsoft Excel para a análise dos artigos e organização e, em conjunto com o Bibexcel, foram montados gráficos e tabelas para geração dos seguintes resultados:

- número de citações dos 14 trabalhos mais citados;
- quantidade de publicações dos 12 autores que mais publicaram;
- quantidade de publicações dos últimos 10 anos (2006 a 2016);
- número de publicações dos 15 periódicos que mais publicaram;
- distribuição dos 10 países que mais contribuíram na área.

Em um segundo momento, utilizou-se o Microsoft Excel em conjunto com o VOSviewer para:

- geração de rede de palavras-chave;
- mapa de densidade de cocitação;

Com relação à análise de palavras-chave, também foi utilizado a ferramenta online Wordle™ para a formação de “nuvem de palavras”, muito útil na identificação dos termos individuais mais usados em todas as palavras-chave encontradas na amostra.

4. Análise dos resultados

As primeiras análises foram realizadas de forma conjunta entre Excel® e Bibexcel para analisar o crescimento ou declínio das publicações sobre data mining aplicado na área da saúde, bem como quais autores e periódicos tiveram maior influência ou número de

publicações. A figura 4 mostra o gráfico da quantidade de publicações ao longo dos últimos 10 anos.

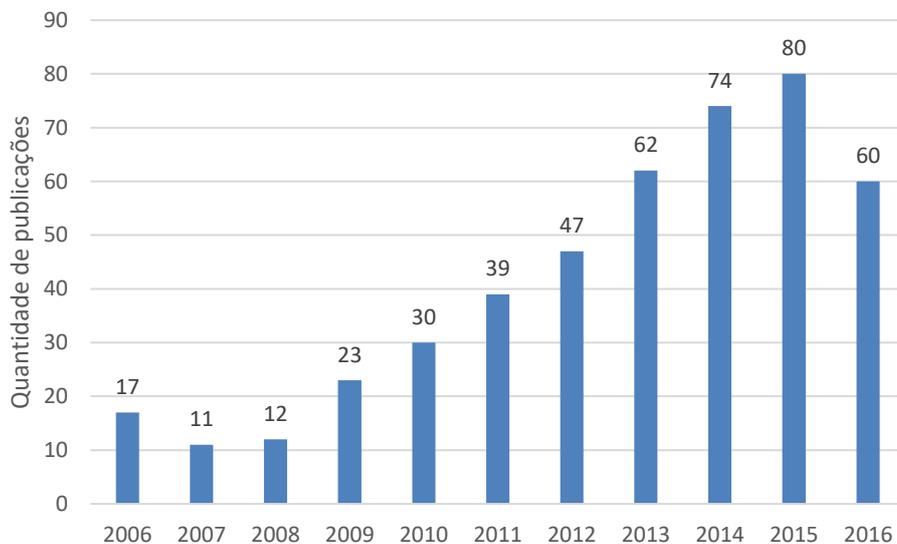


Figura 4 – Quantidade de publicações entre 2006 e 2016.

A figura 4 ilustra o crescimento da pesquisa em data mining na área de saúde a partir de 2008. Tal fato pode ser justificado pelo aumento gradativo da quantidade de dados armazenados em sistemas de saúde e também pelo surgimento de novos algoritmos para encontrar novos padrões, classificar, associar ou prever situações. Nota-se que estes termos buscados em conjunto não são novidade, aliás, a área da saúde é uma das mais estudadas com aplicações de técnicas de mineração de dados ou do KDD como um todo.

A figura 5 mostra os 12 autores que mais publicaram nos últimos 10 anos.

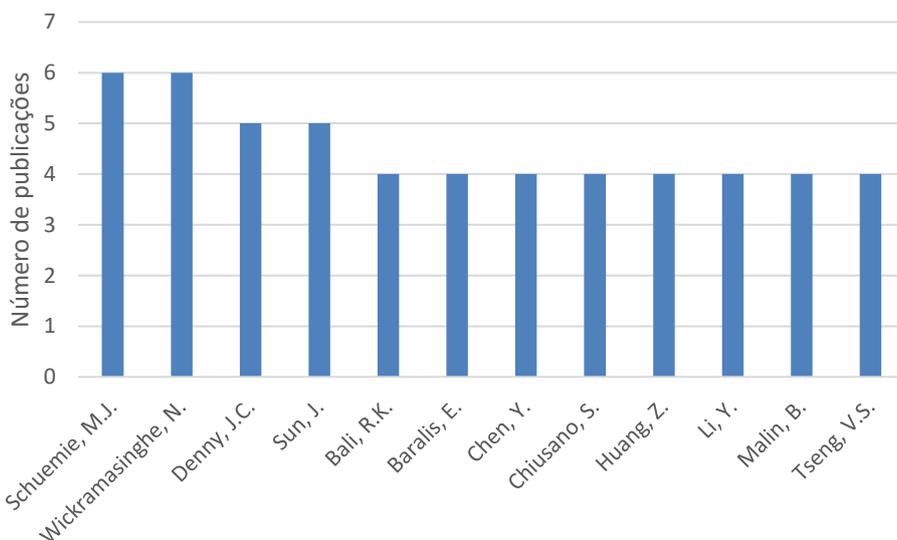


Figura 5 – Quantidade de publicações entre 2006 e 2016.

Observa-se na figura 5 uma uniformidade quanto aos autores que mais publicaram. Martijn Schuemie foi o autor que teve mais publicações, entretanto na base Scopus não consta nenhuma artigo nos termos pesquisados em que ele seja o primeiro autor. O mesmo trabalha

com saúde e computação na Erasmus University Medical Center em Rotterdam, Holanda. Quanto aos trabalhos mais citados, destacam-se os artigos da tabela 1.

Autores	Periódico	Ano	Citado por
Xu H., Stenner S.P., Doan S., Johnson K.B., Waitman L.R., Denny J.C.	Journal of the American Medical Informatics Association	2010	150
Duan L., Street W.N., Xu E.	Enterprise Information Systems	2011	102
Batet M., Sánchez D., Valls A.	Journal of Biomedical Informatics	2011	95
Afzelius L., Arnby C.H., Broo A., Carlsson L., Isaksson C., Jurva U., Kjellander B., Kolmodin K., Nilsson K., Raubacher F., Weidolf L.	Drug Metabolism Reviews	2007	94
Yoo I., Alafaireet P., Marinov M., Pena-Hernandez K., Gopidi R., Chang J.-F., Hua L.	Journal of Medical Systems	2012	94
Rebuge A., Ferreira D.R.	Information Systems	2012	92
Lu Z.	Database	2011	90
Yang W.-S., Hwang S.-Y.	Expert Systems with Applications	2006	78
Rosenbloom S.T., Denny J.C., Xu H., Lorenzi N., Stead W.W., Johnson K.B.	Journal of the American Medical Informatics Association	2011	71
Agrawal R., Johnson C.	International Journal of Medical Informatics	2007	70
Chaovalitwongse W.A., Fan Y.-J., Sachdeo R.C.	IEEE Transactions on Systems, Man, and Cybernetics Part A: Systems and Humans	2007	58
Frantzidis C.A., Bratsas C., Klados M.A., Konstantinidis E., Lithari C.D., Vivas A.B., Papadelis C.L., Kaldoudi E., Pappas C., Bamidis P.D.	IEEE Transactions on Information Technology in Biomedicine	2010	57
Adlassnig K.-P., Combi C., Das A.K., Keravnou E.T., Pozzi G.	Artificial Intelligence in Medicine	2006	50
Jin H., Chen J., He H., Williams G.J., Kelman C., O'Keefe C.M.	IEEE Transactions on Information Technology in Biomedicine	2008	48

Tabela 1 – Trabalhos mais citados da amostra

De acordo com a tabela 1, o artigo mais citado (150 citações) é o trabalho intitulado “MedEx: A medication information extraction system for clinical narratives”. Todos os autores são da Escola de Medicina da Universidade de Vanderbilt, Nashville, Estados Unidos. O trabalho foi realizado pelos seguintes grupos: Departamento de Medicina, Departamento de Pediatria e Departamento de Informática em Biomedicina. Em resumo, foi desenvolvido uma linguagem para a identificação de termos médicos específicos em uma base com dados livres, conseguindo obter resultados satisfatórios para a busca de nomes de medicamentos, frequência, pacientes, entre outros.

Outra informação extraída da tabela 1 é que o primeiro e nono trabalhos mais citados tem coautoria do pesquisador Joshua C. Denny, Professor Associado da Univerisdade de Vanderbilt em Nashville (DEPARTMENT OF BIOMEDICAL INFORMATICS, 2016). Somando apenas as duas publicações (Ambos publicados na *Journal of the American Medical Informatics Association*) da tabela 1, totaliza 221 citações.

A figura 6 mostra o gráfico com o total de publicações encontradas por periódico.

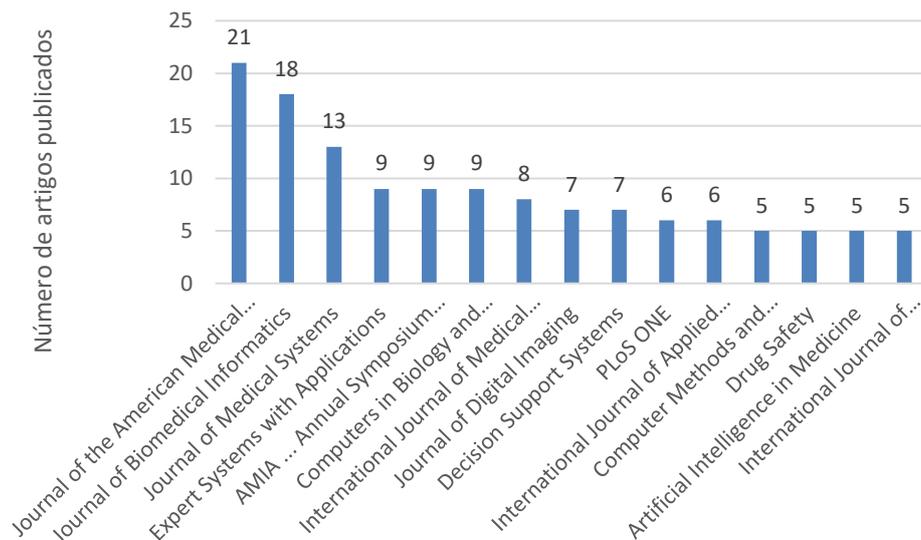


Figura 6 – Número de publicações por periódico.

Analisando-se o gráfico da figura 6 nota-se que os três periódicos que tiveram mais publicações foram o *Journal of the American Medical Informatics Association*, *Journal of Biomedical Informatics* e *Journal of Medical Systems*, com 21, 18 e 13 publicações, respectivamente. Os três periódicos têm como parte de seu escopo o tratamento de informações para a saúde ou estudo de sistemas na área de saúde, podendo assim ser relacionado com o “data mining” e “healthcare” simultaneamente. Estes periódicos possuem fator de impacto significativos, sendo eles 3,428, 2,447 e 2,213, respectivamente. De acordo com a pesquisa, o periódico com mais publicações e avaliado no webqualis da CAPES em Engenharias III é *Expert Systems with Applications*, com classificação A2 (CAPES, 2016).

Com relação ao número de documentos de acordo com os países de origem, a figura 7 ilustra a situação estudada.

Número de documentos

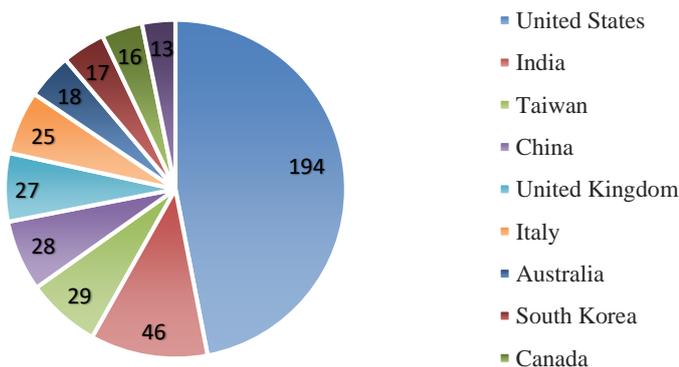


Figura 7 – Número de documentos publicados por país.

O número de documentos publicados com origem nos Estados Unidos é significativamente superior quando comparado a outros países, somando um total de 194. Este número corresponde a 43% do total de publicações, mais que o dobro de Índia, Taiwan e China (46, 29 e 28, respectivamente).

termos como “Data”, “mining”, “health”, “system”, “medical”, “care” e “analysis” apareceram com maior destaque. Após analisar os artigos e a própria base SCOPUS, o termo “article” faz parte das palavras-chave padrões da medicina.

Por fim, fez-se a análise de cocitação no software VOSviewer, como mostrado na Figura 10.

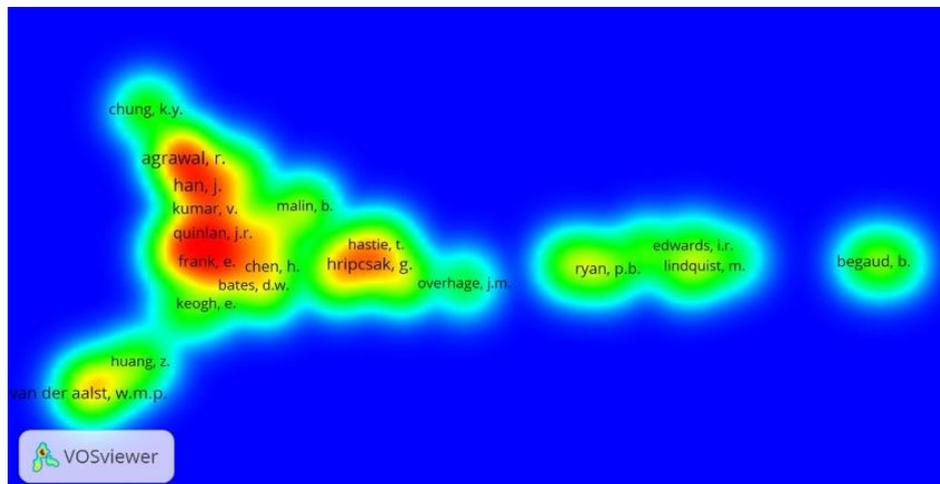


Figura 10 – Mapa de densidade de cocitação

Para a criação do mapa de densidade de cocitação, fez-se o filtro por autores que tiveram no mínimo 20 citações, totalizando assim 61 autores. O mapa da figura 10 ilustra que os autores R. Agrawal, J. Han, V. Kumar, J. R. Quinlan, E. Frank, H. Chen e D.W. Bates aparecem como referência simultaneamente com outros autores em diversos artigos. T. Hastie e G. Hripcsak também estão em destaque, porém é provável que o foco de pesquisa (técnicas, aplicações, entre outros) sejam diferentes, por isso está mais afastado do núcleo em tom vermelho. O autor B. Begaud também tem um número de cocitações significativo, indicando um possível foco de estudo distinto da concentração mais à esquerda do mapa.

5. Considerações finais

O objetivo deste trabalho foi realizar uma bibliometria sobre os estudos de data mining na área da saúde. Toda a pesquisa foi baseada na revisão sistemática da literatura, utilizando-se da base de dados SCOPUS, filtros e critérios de busca e utilização de softwares posterior análise bibliométrica.

Observou-se que a amostra obtida foi significativa (455 artigos nos últimos 10 anos), mostrando que a área da saúde, incluindo serviços, identificação de tratamentos em si, padrões de doenças, entre outros, estão sendo explorados cada vez mais nos últimos anos. Martijn Schuemie foi o pesquisador que mais publicou, enquanto que o trabalho de Xu et al. (2010) foi o trabalho mais citado nesses últimos anos. Quanto aos periódicos, o *Journal of the American Medical Informatics Association* é o que possui o maior número de artigos publicados, possuindo um fator de impacto 3,428. Os Estados Unidos é o país que contém maior número de documentos publicados na área, seguido de Índia, Taiwan e China. Quando analisou-se as palavras-chave e os termos mais utilizados individualmente nas *keywords*, destaque para “data mining”, “humans”, “health”, “healthcare”, “analysis” e “system”. Por fim, obteve-se um grupo dois grupos principais que tiveram co-citação, sendo o primeiro composto por sete autores e o segundo por dois.

A bibliometria mostrou que consegue encontrar informações de diferentes objetivos (autor, periódico, artigo, etc.). Atualmente existem diversas bases de dados compatíveis com vários

softwares programados para interpretar todas ou pelo menos grande parte sobre determinadas especificações de busca. Entretanto, é importante que a análise bibliométrica seja realizada a partir de um protocolo de busca bem definido para assim melhorar e também filtrar os dados a serem buscados. Para a base SCOPUS, os assuntos “data mining” e “healthcare” já são bem conhecidos e explorados pelos autores, mostrando que data mining não é um assunto recente assim como a sua aplicação na saúde. Pesquisas futuras poderiam trabalhar em uma estratificação para tentar encontrar lacunas a serem exploradas na saúde, como “hospitals”, “emergency systems” ou “treatment”, assim poderá ser definido uma direção para novos estudos de técnicas de data mining.

Referências

- ALTUNTAS, S.; SELIM, H.** *Facility layout using weighted association rule-based data mining algorithms: Evaluation with simulation.* Expert Systems with Applications, v. 39, p. 3-13, 2012.
- BAE, J. K.; KIM, J.** *Product development with data mining techniques: A case on design of digital camera.* Expert Systems and Applications, v. 38, p. 9274-9280, 2011.
- BARAK, S.; MODARRES, M.** Developing an approach to evaluate stocks by forecasting effective features with data mining methods. Expert Systems and Applications, v. 42, p. 1325-1339, 2015.
- BENNETT, K. P.; MANGASARIAN, O. L.** *Robust Linear Programming Discrimination of Two Linearly Inseparable Sets.* Optimization Methods and Software, London, United Kingdom, Taylor and Francis Group, v. 1, p. 23-34, 1992.
- BUFREM, L.; PRATES, Y.** *O saber científico registrado e as práticas de mensuração da informação.* Ciência da Informação, Brasília, v. 34, n. 2, p. 9-25, 2005.
- CAPES: Coordenação de Aperfeiçoamento de Pessoal de Nível Superior.** Disponível em: <<https://sucupira.capes.gov.br/sucupira/public/consultas/coleta/veiculoPublicacaoQualis/listaConsultaGeralPeriodicos.jsf>>. Acesso em: 07 set 2016.
- CARVALHO, D. R.; MOSER, A. D.; SILVA, V. A. Da; DALLAGASSA, M. R.** *Mineração de dados aplicados à fisioterapia. Fisioterapia em Movimento.* V. 25, n. 3, p. 595-605, 2012.
- CHENG, C. W.; YAO, H. Q.; WU, T. C.** *Applying data mining techniques to analyze the causes of major occupational accidents in the petrochemical industry.* Journal of Loss Prevention in the Process Industries, v. 26, p. 1269-1278, 2013.
- DEPARTMENT OF BIOMEDICAL INFORMATICS: Vanderbilt University.** Disponível em: <<https://medschool.vanderbilt.edu/dbmi/person/joshua-c-denny-md-ms>>. Acesso em: 06 set 2016.
- FAYYAD, U. M.; PIATETSKY-SHAPIRO, G.; SMYTH, P.; UTHURUSAMY, R.** *Advances in Knowledge Discovery & Data Mining.* 1 ed. American Association for Artificial Intelligence, Menlo Park, Califórnia, 1996. 611 folhas.
- FORTULAN, M. R.; FILHO, E. V. G.** *Uma proposta de aplicação de business intelligence no chão de fábrica.* Revista Gestão & Produção, v.12, n.1, p.55-66, 2005
- FREITAS, A. A.** *Uma Introdução a Data Mining.* Informática Brasileira em Análise, Centro de Estudos e Sistemas Avançados do Recife (C.E.S.A.R.), Recife, PE, ano II, n. 32, 2000.
- GOTTARDO, E.; KAESTNER.; NORONHA, R. V.** *Previsão de desempenho de estudantes em cursos EAD utilizando mineração de dados: uma estratégia baseada em séries temporais.* Simpósio Brasileiro de Informática na Educação, Rio de Janeiro, 2012.
- GUIDINI, M. P.; RIBEIRO, C. H. C.** *Utilização da biblioteca TerraLib para algoritmos de agrupamento em sistemas de informações geográficas.* VIII Brazilian Symposium on GeoInformatics, São Paulo, p.303-314, 2006.
- HAN, J.; KAMBER, M.** *Data Mining: Concepts and Techniques*, ed. 1, Nova York: Morgan Kaufmann, 2012.
- HORTA, R. A. M.; ALVES, F. J. S.** *Aplicação de técnicas de Data Mining para o entendimento da política de financiamento de empresas brasileiras.* CONGRESSO ANPCONT, Rio de Janeiro, 2012.

- KUSIAK, A.; ZENG, Y.; ZHANG, Z.** *Modeling and analysis of pumps in a wastewater treatment plant: A data-mining approach.* Engineering Applications of Artificial Intelligence, v. 26, p. 1643-1651, 2013.
- LAZZAROTTO, L. L.; OLIVEIRA, A. P.; LAZZAROTTO, J. J.** *Aspectos teóricos do Data Mining e aplicação das Redes Neurais em previsões de preços agropecuários.* Congresso da Sociedade Brasileira de Economia e Sociologia Rural, Ceará, 2006.
- LEMONS, E. P.; STEINER, M. T. A.; NIEVOLA, J. C.** *Análise de crédito bancário por meio de redes neurais e árvores de decisão: uma aplicação simples de Data Mining.* Revista de Administração, v. 40, n. 3, p. 225-234, 2005.
- MANHÃES, L. M. B.; CRUZ, S. M. S.; COSTA, R. J. M.; ZIMBRÃO, G.** *Previsão de estudantes com risco de evasão utilizando técnicas de mineração de dados.* Simpósio Brasileiro de Informática na Educação, Sergipe, 2011.
- MENDES, K. B.; FIUZA, R. M.; STEINER, M. T. A.** *Diagnosis of headache using artificial neural networks.* International Journal of Computer Science and Network Security, v. 10, n. 7, p. 172-178, 2010.
- NÄÄS, I. A.; QUEIROZ, M. P. G.; MOURA, D. J.; BRUNASSI, L. A.** *Estimativa de estro em vacas leiteiras utilizando métodos quantitativos preditivos.* Revista Ciência Rural, v.38, n.8, p.2383-2387, 2008.
- NGAI, E. W. T.; XIU, L.; CHAU, D. C. K.** *Application of data mining techniques in customer relationship management: A literature review and classification.* Expert Systems with Applications, v. 36, p. 2592-2602, 2009.
- OLIVEIRA, D. M. B de; STEINER, M. T. A.; COSTA, D. M. B. (2012);** *Técnicas da Pesquisa Operacional na avaliação de distúrbios vocais em docentes.* Revista SODEBRAS, v. 7, n. 1, 2012.
- PERSSON, O.; R. DANELL; J. WILBORG SCHNEIDER.** *How to use Bibexcel for various types of bibliometric analysis.* In Celebrating scholarly communication studies: A Festschrift for Olle Persson at his 60th Birthday, Leuven, Belgium: International Society for Scientometrics and Informetrics. ed. F. Åström, R. Danell, B. Larsen, J. Schneider, p. 9-24, 2009.
- PHRIDVIRAJ, M. S. B.; GURURAO, C. V.** *Data Mining – past, present and future – a typical survey on data streams.* Procedia Technology, v. 12, p. 255-263, 2014.
- PILKINGTON, A.; MEREDITH, J.** *The evolution of the intellectual structure of operations management – 1980-2006: a citation/co-citation analysis.* Journal of Operations Management, v. 27, p. 185-202, 2009.
- RAMOS, H. S. C.; BRÄSCHER, M.** *Aplicação de descoberta de conhecimento em textos para apoio à construção de indicadores infométricos para a área de C&T.* Ci. Inf., Distrito Federal, 2009.
- STEINER, M. T. A.; SOMA, N. Y.; SHIMIZU, T.; NIEVOLA, J. C.** *Abordagem de um problema médico por meio de processo KDD com ênfase à análise exploratória de dados.* Gestão & Produção, v. 13, n.2, p. 325-337, 2006.
- TAN, P., STEINBACH, M., KUMAR, V.** *Introdução ao Data Mining: Mineração de Dados.* 1 ed. Rio de Janeiro: Ciência Moderna Ltda, 2009.
- TRONCHONI, A. B.; ROSA, M. A. Da; PRETTO, C. O.; LEMOS, F. A. B.** *Descoberta de conhecimento em base de dados de eventos de desligamentos de empresas de distribuição.* Revista Controle & Automação, v. 21, n. 2, p.185-200, 2010.
- VAN ECK, N. J.; WALTMAN, L.** *VOSviewer Manual: Manual for VOSviewer version 1.6.4.* Universiteit Leiden, 2016.
- VIANNA, R. C. X.; MORO, C. M. C. B.; MOYSÉS, S. J.; CARVALHO, D.; NIEVOLA, J. C.** *Mineração de dados e características da mortalidade infantil.* Cad. Saúde Pública, v.26, n.3, p.535-542, 2010.
- WITTEN, I. H.; FRANK, E.** *Data Mining: Practical Machine Learning Tools and Techniques.* 2 ed. San Francisco, CA: Elsevier, 2005.
- XU, H.; STENNER, S.P.; DOAN, S; JOHNSON, K.B.; WAITMAN, L.R.; DENNY, D.C.** *MedEx: a medication information extraction system for clinical narratives.* Journal of the American Medical Informatics Association. Vol. 17, n. 1, p. 19-24, 2010.
- YAP, B. W.; ONG, S. H.; HUSAIN, N. H. M.** *Using data mining to improve assessment of credit worthiness via credit scoring models.* Expert Systems with Applications, v. 28, p. 13274-13283, 2011.