

Uso de Mineração de Dados para a Análise da Relação entre Zika Vírus e Microcefalia

Rodrigo Ramos¹, Ana Maria Monteiro¹

¹Faculdade Campo Limpo Paulista (FACCAMP)

Rua Guatemala – Campo Limpo Paulista – SP – Brasil

r.amos.professor@terra.com.br, anamont@cc.faccamp.br

Abstract. *In the beginning of the year 2015 a Zika Virus epidemic occurred in several Brazilian states. At the end of the year 2015, there was a large increase in cases of microcephaly in Brazil, especially in the state of Pernambuco. There is no data yet to prove the relation between Zika Virus and microcephaly. In the year 2016, the Zika Project was designed to follow the gestation, the birth, and the initial years of children to see if this relation is true. The proposal of this work is, based on the data collected by the Zika Project, to use the existing techniques in Data Mining to help the project researchers to confirm (or not) this relation.*

Resumo. *No início de 2015 uma epidemia de Zika Vírus ocorreu em vários estados brasileiros. No final do ano de 2015, foi registrado um grande aumento dos casos de microcefalia no Brasil, em especial no estado de Pernambuco. Ainda não existem dados que realmente comprovem a relação entre o Zika Vírus e a microcefalia. No ano de 2016, foi elaborado o Projeto Zika, para averiguar se tal relação existe. A proposta deste trabalho é, a partir dos dados coletados pelo Projeto Zika, utilizar as técnicas existentes em Mineração de Dados para auxiliar os pesquisadores envolvidos no projeto na confirmação (ou não) de tal relação.*

1. Introdução

A Informática Médica é uma ciência multidisciplinar desenvolvida na interação da Ciência da Informação com a Medicina e os cuidados de saúde [Zvárová, Veselý e Vajda 2009]. Dentro da informática médica, uma das áreas relevantes é a de Inteligência Artificial (IA) aplicada à medicina. Esta é uma área na qual são propostas teorias e desenvolvidos modelos para aplicações na área da saúde [Szolovits 1982]. O desenvolvimento de IA em medicina oferece aos profissionais de saúde mecanismos que contribuem para a realização de suas tarefas cotidianas e o avanço das pesquisas médicas. Técnicas de IA foram aplicadas com sucesso em:

- Sistemas de auxílio à tomada de decisões para o diagnóstico de doenças a partir de sintomas e resultados de exames laboratoriais de pacientes.
- Sistemas de análise e reconhecimento de imagens para detectar possíveis anomalias.
- Sistemas de monitoramento de pacientes.
- Sistemas para descoberta de grupos de risco para uma doença.
- Sistemas para planejamento de terapias baseadas na condição específica de cada paciente.
- Sistemas para estudo e classificação do genoma humano a partir dos grandes volumes de dados existentes.

Alguns dos trabalhos que estão sendo desenvolvidos utilizam técnicas de Mineração de Dados (MD) que conseguem obter informações para a tomada de decisão a partir dos grandes volumes de dados gerados diariamente na prática médica.

A proposta apresentada neste artigo tem o intuito de usar técnicas de MD em dados coletados pelo Projeto Zika [Passos 2016]. No início de 2015 uma epidemia de Zika Vírus ocorreu em vários estados brasileiros e ao final do ano de 2015, foi registrado um grande aumento dos casos de microcefalia no Brasil. Ainda não existem dados que realmente comprovem a relação entre o Zika Vírus e a microcefalia. No ano de 2016, foi elaborado o Projeto Zika, para averiguar se esta relação existe. A proposta deste trabalho é, a partir dos dados coletados pelo Projeto Zika, utilizar as técnicas existentes em MD para auxiliar os pesquisadores do projeto na obtenção de respostas sobre uma relação entre Zika Vírus e microcefalia.

2. Trabalhos Relacionados

Entre os trabalhos relacionados que abordam o uso de Mineração de Dados em medicina pode-se mencionar o de Castro (2005) que propôs o uso de Redes Neurais artificiais para criar modelos de predição para as ocorrências de surtos de dengue na cidade de Goiânia.

Por sua vez Santos (2011) apresentou uma ferramenta que utilizou algoritmos de classificação e regras de associação para gerar modelos da ocorrência de dengue a partir da base de dados do Sistema de Informações de Agravos e Notificação (SINAN) e assim poder auxiliar os gestores da saúde na prevenção e diagnóstico de casos de dengue.

Li *et al.* (2015) utilizam MD para descobrir restrições temporais relacionadas com os resultados clínicos decorrentes do atendimento de pacientes, como por exemplo, exame de glicemia em pacientes com diabetes e exames laboratoriais. Para este fim são usadas regras de associação, uma das técnicas de Mineração de Dados.

Carvalho, Dallagassa e Silva (2015), a partir do uso de árvores de decisão, propõem um modelo para o diagnóstico de Diabetes Mellitus tipo 2 em beneficiários de planos de saúde.

Alves (2015) desenvolveu um modelo computacional utilizando lógica nebulosa que permite, a partir de dois biomarcadores de insuficiência cardíaca (acetona exalada e peptídeo natriurético do tipo B) a elaboração de uma terapia mais adequada para o tratamento de pacientes portadores desse problema.

Gregory e Pretto (2016) descrevem no seu artigo desde as conversas com as equipes médicas até a coleta, seleção e organização de dados que foram utilizados para agrupar pacientes levando em consideração suas características comuns. Em seguida, foi aplicado o algoritmo Apriori para a obtenção de regras por associação. Como resultado, os autores constataram uma relação entre hipertensão e estresse, principalmente em mulheres jovens.

Pires (2011) na sua tese de doutorado construiu um repositório, contendo dados de um período de 10 anos provenientes das principais bases de dados do Sistema Único de Saúde (SUS). O objetivo do trabalho foi, por um lado, extrair conhecimento relevante para melhorar a gestão em Saúde Pública e, por outro, permitir acompanhar pacientes e comparar diferentes populações.

3. Mineração de Dados

Mineração de Dados é parte de um processo denominado Descoberta de Conhecimento em Bases de Dados (*Knowledge Discovery on Databases- KDD*) [Tan *et al* 2006]. KDD é o processo de extração de informações implícitas, previamente desconhecidas e potencialmente úteis, a partir de grandes volumes de dados armazenados, geralmente, em um banco de dados [Fayyad *et al.* 1996].

Os processos que compõem a descoberta de conhecimento são: seleção dos dados, pré-processamento e limpeza, transformação, Mineração de Dados e por fim, interpretação e avaliação dos resultados.

Para a execução da mineração, são empregadas técnicas de análise de dados e algoritmos de aprendizado de máquina que extraem, a partir de dados pré-processados, conhecimento expressado através de padrões ou modelos.

Dentre as técnicas de MD a serem utilizadas para a proposta apresentada neste artigo podem ser mencionadas as de classificação e as de análise de associações:

- A classificação consiste em construir um modelo que possa ser aplicado a dados não classificados visando categorizá-los em classes. Um dado é examinado e classificado de acordo com um conjunto de classes predefinidas [Tan *et al.* 2006].
- A análise de associações objetiva descobrir regras de associação condicionadas por valores de atributos que acontecem juntos no conjunto de dados. Uma regra de associação é um padrão da forma $X \rightarrow Y$, onde X e Y são conjuntos de itens. As regras são utilizadas para a predição da ocorrência de certos itens (Y) baseada na ocorrência de outros itens (X).

4. Caracterização do Projeto

Microcefalia é uma anomalia congênita, que pode ter diversas causas, entre elas anomalias cromossômicas ou doenças maternas durante a gravidez [Marinho *et al.* 2016]. A microcefalia é medida pelo tamanho do perímetro cefálico, nestes casos, essa medida é menor que 32 cm, o valor habitualmente padrão.

Marinho *et al.* (2016) usando dados de um sistema de informações sobre nascidos vivos (Sinasc) do Ministério da Saúde, apresenta os casos de microcefalia registrados de 2000 a 2015. Apesar de não ser uma doença nova, pois já existiam casos, estes mantiveram uma certa estabilidade neste período. Somente no final de 2015 houve um aumento significativo em todas as regiões do país, com a maior incidência de casos registrada na Região Nordeste do país.

Com relação ao Zika Vírus, este foi identificado em 1947 e os primeiros casos de Zika Vírus no Brasil ocorreram no primeiro semestre de 2015. No início de novembro de 2015, o Ministério da Saúde decretou Emergência em Saúde Pública de Importância Nacional. No final de novembro, já haviam sido notificados 1.248 casos suspeitos em 14 estados brasileiros [Martelli 2015].

Diante destes fatos, foi levantada a hipótese de uma relação entre o aumento dos casos de microcefalia em recém-nascidos e o surto de Zika Vírus que acabara de ter no país, sendo essa a causa para esse repentino aumento nos casos de microcefalia.

Em 2016 foi elaborado no Hospital Universitário de Jundiaí, sob a coordenação do Professor Doutor Saulo Duarte Passos, o Projeto Zika que tem como objetivo determinar a incidência da infecção causada pelo Zika Vírus ao longo de cinco anos em gestantes e recém-nascidos, bem como analisar as associações entre as infecções pelo Zika Vírus e as malformações do sistema nervoso de recém-nascidos [Passos 2016]. O Projeto Zika pretende acompanhar as mães durante a gestação, com análises clínicas e sociais, o parto, e os primeiros anos dos bebês coletando dados sobre os casos de Zika Vírus e microcefalia.

A proposta deste trabalho é aplicar as técnicas existentes de MD, com foco nas características e sintomas apresentados pelas mães durante a gestação. E após o parto, agrupá-las de acordo com as características dos bebês no que se referem às situações possíveis, isto é, bebês que nasçam com ou sem algum dano congênito.

Após essa etapa de análise e agrupamento de casos, verificar a possível elaboração de modelos computacionais, que auxiliem as equipes médicas nos passos seguintes do projeto Zika.

5. Metodologia

Para a realização deste trabalho será adotada a seguinte metodologia de trabalho:

- Pesquisa bibliográfica dos temas referentes à Mineração de Dados aplicados a medicina e às informações relacionadas com objeto de pesquisa (Zika Vírus e Microcefalia).
- Após a disponibilização do banco de dados do Projeto Zika será feita a análise e limpeza dos dados, preparando-os para a aplicação das técnicas de Mineração de Dados.
- Após a análise dos dados serão utilizados algoritmos para discriminar atributos que influem na ocorrência de más formações nos recém-nascidos.
- Aplicação de algoritmos de MD (árvores de decisão e regras de associação) para a obtenção de modelos que permitam ajudar aos profissionais da saúde a determinar relação entre o Zika Vírus e a Microcefalia.
- Os modelos obtidos serão avaliados com a colaboração dos especialistas médicos envolvidos no projeto.
- Ao final será realizada a submissão dos resultados obtidos durante o trabalho a conferências / periódicos.

Referências

Alves, J. L. G. (2015), *Utilização de Inteligência Artificial, através de Inferência Difusa, na Correlação de Métodos de Diagnóstico de IC (Insuficiência Cardíaca)*. Dissertação de mestrado, Instituto Tecnológico de Aeronáutica (ITA), Brasil.

- Carvalho, D. R., Dallagassa, M. R., Silva, S. H. (2015), *Uso de Técnicas de Mineração de Dados para Identificação Automática de Beneficiários Propensos ao Diabetes Melitus Tipo 2*. Revista Informação & Informação. Universidade Estadual de Londrina, Brasil.
- Castro, G. G. (2005), *Suporte à Decisão para Vigilância Epidemiológica Baseado em Modelo Preditivo de Surtos de Dengue Utilizando Redes Neurais Artificiais*. Dissertação de Mestrado, Universidade Católica de Brasília, Brasil.
- Fayyad, U. Piatetsky-Shapiro, G., Smyth, P. (1996), From Data Mining to Knowledge Discovery in Databases. *AI Magazine, Volume 17, Number 3* (1996), p. 37-54.
- Gregory, G., Pretto, Fabricio. (2016), *Mineração de Dados para Descoberta de Conhecimento em Dados de Promoção à Saúde*, Revista Destaques Acadêmicos, v.8, n.4, Univates, Lajeado, Rio Grande do Sul.
- Jaenisch T., Rosenberger, K. D., Brito C., Brady, O., Brasil, P. e Marques E. (2017), *Risk of Microcephaly after Zika Virus Infection in Brazil, 2015 to 2016*. Bulletin of the World Health Organization.
- Li, X., Liu, H., Mei, J., Yu, Y, e Xie, G. (2015), Mining Temporal and Data Constraints Associated with Outcomes for Care Pathways. *Proceedings MEDINFO 2015*, p. 711-715. São Paulo, Brasil.
- Marinho, F., Araújo, V., Porto, E. M, Lopes, D., Ferreira, H. L., Coelho, M. R. S., Lecca, R. C. R., Oliveira, H. P., Almeida, I., Maranhão, M. H. N., Mendes, Y. M. M. B., Fernandes, R. M. Lima, R. B. N. e Rabello, D. (2016), *Microcefalia no Brasil: Prevalência e Caracterização dos Casos a partir do Sistema de Informações sobre Nascidos Vivos (Sinasc), 2000-2015*, Ministério da Saúde, Brasil.
- Martelli, C. M. T. (2015), *Investigação dos Casos de Microcefalia: Estudo Caso-Controle*, Projeto de Pesquisa, Recife.
- Passos, S. D. (2016), *Infecção Vertical pelo Vírus Zika e suas Repercussões na Área Materno-infantil*, Faculdade de Medicina de Jundiaí, Projeto Temático de Pesquisa, Brasil.
- Pires, F. A. (2011), *Ambiente para Extração de Informação Epidemiológica a partir da Mineração de dez Anos de Dados do Sistema Público de Saúde*. Tese de Doutorado, Faculdade de Medicina, Universidade de São Paulo, Brasil.
- Santos, M. S. N., Costa, J. C. (2011), *AMAGODIS: Algoritmos de Mineração para Apoio à Gerência de Ocorrências de Dengue a partir de Informações presentes na base dados do SINAN*. Encontro Anual de Computação, Universidade Federal de Goiás.
- Szolovits, P. (1982), Artificial Intelligence and Medicine, in *Artificial Intelligence in Medicine*, Szolovits, P. (Ed.), Chapter 1. Westview Press, Boulder, Colorado.
- Tan, P., Steinbach, M., Kumar, V. (2009), *Introdução ao Data Mining, Mineração de Dados*. Rio de Janeiro. Editora Moderna LTDA.
- Zvárová, J., Veselý A. e Vajda I. (2009), Data, Information and Knowledge, in Berka, P., Rauch, J. e Zighed, D. A (Editores) “*Data Mining and Medical Knowledge Management: Cases and Applications*”, p. 1-22, IGI Global Press, New York.