

MINERAÇÃO DE REGRAS DE ASSOCIAÇÃO PARA DESCOBERTA DE PADRÕES FREQUENTES EM DOENÇA DE CHAGAS

DELAZERI, J.M.¹; SANTOS, L. I.².

¹Discente do curso superior em Ciência da Computação do IFNMG – *Campus Montes Claros*;

²Docente do IFNMG – *Campus Montes Claros*.

Introdução

A doença de Chagas (DC) é endêmica em 21 países na América Latina com cerca de 5,7 milhões de pessoas contaminadas. É uma doença que a Organização Mundial de Saúde considera negligenciada. Um estudo realizado por FERREIRA et. al (2019) aponta para uma considerável escassez de conhecimento sobre a DC entre profissionais de saúde. Essa falta de compreensão pode impactar negativamente a saúde dos pacientes, caso a doença não seja reconhecida e tratada de forma adequada. Nessa perspectiva, torna-se essencial adquirir um conhecimento mais abrangente sobre os cenários epidemiológicos da DC pois isso se torna crucial para assegurar o reconhecimento e tratamento preciso da doença, visando proteger a saúde dos indivíduos afetados.

Nesse contexto, ferramentas computacionais para análise de dados podem proporcionar percepções valiosas sobre a DC, permitindo uma compreensão mais profunda do problema. Diante da crescente complexidade dos desafios de Saúde Pública, é essencial revisar constantemente a abordagem, aproveitando a riqueza de dados disponíveis e a diversidade de métodos. A emergência de novas fontes de dados exige novas abordagens de estudo, muitas vezes coordenadas por equipes interdisciplinares (DICLEMENTE et al, 2019).

A Mineração de Regras de Associação é uma ferramenta valiosa na análise de dados de saúde pública, pois gera regras do tipo "SE - ENTÃO", que estão alinhadas com a criação de padrões compreensíveis de modelos médicos. Essas regras de associação, expressas como $X \Rightarrow Y$, indicam possíveis relacionamentos entre conjuntos de itens distintos X e Y. Isso proporciona uma abordagem de inferência que pode revelar conexões entre eventos X e Y, contribuindo para a compreensão de padrões e conceitos médicos (HAN et al., 2011).

Tendo isso em mente, o objetivo deste trabalho foi desenvolver uma abordagem para a descoberta de Regras de Associações utilizando técnicas de Mineração de Dados em um conjunto de dados sobre a Doença de Chagas.

Materiais e Métodos

O presente trabalho foi realizado utilizando o conjunto de dados SaMi-Trop, que é um estudo multicêntrico financiado pelo National Institute of Health (NIH) (grant:P50AI098461-02 and U19AI098461-06) que estabeleceu uma coorte com portadores da Doença de Chagas em 21



municípios endêmicos do estado de Minas Gerais. A primeira linha de base foi realizada entre os anos de 2013 e 2014 contando com 2160 participantes, e posteriormente em 2015 à 2016 aconteceu o primeiro segmento com a permanência de 1709 participantes. Dentre as informações contidas na base de dados, estão questões sociodemográficas, hábitos de vida, história clínica, tratamento da doença de chagas, atividade física e qualidade de vida.

As variáveis da base de dados foram limpas e filtradas através de um script da linguagem de programação Python. Neste trabalho optou-se por utilizar somente variáveis referentes à questões sociodemográficas, hábitos de vida, atividade física e qualidade de vida. Além disso, algumas variáveis que apresentavam um forte grau de correlação (acima de 0.6) foram removidas do modelo, a fim de evitar que o algoritmo gerasse regras consideradas tendenciosas e sem valor.

O algoritmo escolhido para mineração das regras foi o Apriori, pois se trata de um dos principais algoritmos utilizados nesta finalidade (FAYYAD et al., 1996) e existem vários outros trabalhos e relatos de literatura com bons resultados de qualidade das regras geradas. Para tanto, foi utilizado o pacote *Arules* do *Software R*. Foram utilizadas, ainda, duas métricas para avaliar a relevância e a qualidade das regras encontradas, são elas o suporte, que mede a frequência com que a regra ocorre no conjunto de dados e a confiança que mede a proporção de vezes em que o consequente é verdadeiro quando o antecedente é verdadeiro. O algoritmo foi configurado para obter regras com suporte mínimo de 0.01 e a confiança de 0.3 (valores estabelecidos empiricamente).

Resultados e Discussão

Um total de 18 regras foram encontradas com valores para confiança variando de 0.3 até 0.63. Na tabela 1 são apresentadas as 4 regras consideradas mais importantes por terem valores de suporte ou de confiança que se destacam dos demais. Pela tabela 1, a regra 1 indica que pacientes acometidos pela doença de chagas e que tenham sido diagnosticados como um paciente com o esôfago aumentado e ainda que tenham vasos sanguíneos visíveis no pescoço apresentam dor ao engolir alimentos. Essa possui uma confiança de 0.63 o que mostra que 63% das pessoas com esôfago aumentado e vasos sanguíneos visíveis apresentam dor ao engolir alimentos. Por mais que o suporte seja baixo (0.012), a regra 1 pode ser considerada interessante visto que não há precedentes de tais informações na literatura.

Considerações finais

Esse trabalho teve como objetivo utilizar ferramentas de mineração de regras de associação na descoberta de conhecimentos em um conjunto de dados de pacientes acometidos pela DC. Inicialmente o trabalho focou na busca por regras que não fossem consideradas óbvias e portanto um processo de filtragem de variáveis que teve como base a correlação entre elas foi utilizado. Das 18 regras encontradas somente 1 obteve confiança acima de 0.5 e essa regra mostra que abordagem adotada é promissora.

Como trabalhos futuros pode-se destacar a utilização também de variáveis de exames complementares e até mesmo a utilização de outros conjuntos de dados de pacientes acometidos pela DC.

Agradecimentos

Os autores agradecem o IFNMG — *Campus Montes Claros* e a CAPES por contemplá-los



com uma bolsa de iniciação científica através do Programa Institucional de Bolsas de Iniciação Científica (PIBIC).

Referências

DICLEMENTE, Ralph et al. Need for innovation in public health research. American journal of public health, v. 109, n. S2, p. S117-S120, 2019.

FAYYAD, Usama; PIATETSKY-SHAPIRO, Gregory; SMYTH, Padhraic. From data mining to knowledge discovery in databases. AI magazine, v. 17, n. 3, p. 37-37, 1996.

FERREIRA, Ariela Mota et al. Avaliação do conhecimento acerca do manejo clínico de portadores da doença de Chagas em região endêmica no Brasil. Revista de APS, v. 21, n. 3, 2018.

HAN, Jiawei; PEI, Jian; KAMBER, Micheline. Data mining: concepts and techniques. Elsevier, 2011.

Tabela 1. Regras encontradas

	Regra	Suporte	Confiança
1	{diagnostico_esofago_aumentado_no_raio_x & possui_vasos_sanguineos_visiveis_no_pescoco_sentado_ou_em_pe} => {dor_ao_engolir_alimentos}	0.012	0.628
2	{diagnostico_esofago_aumentado_no_raio_x} => {dor_ao_engolir_alimentos}	0.027	0.450
3	{dor_ao_engolir_alimentos & possui_problemas_tireoide} => {possui_vasos_sanguineos_visiveis_no_pescoco_sentado_ou_em_pe}	0.01	0.439
4	{dor_ao_engolir_alimentos} => {possui_vasos_sanguineos_visiveis_no_pescoco_sentado_ou_em_pe}	0.061	0.343