

Análise de Agrupamento Pelos Métodos Hierárquico e Particional em Dados de Educação a Distância.

Alini Marangoni Eyng¹, Merisandra Côrtes de Mattos Garcia¹, Graziela Fátima Giacomazzo¹, Cibele Figueiredo Freitas¹, Vilson Menegon Bristot¹, Evanio Ramos Nicoleit¹

Universidade do Extremo Sul Catarinense¹

E-mail: alinimarangoni.eyng@gmail.com, mem@unesc.net, gfg@unesc.net, cibeles@unesc.net, vro@unesc.net, elo@unesc.net

Resumo. Este artigo descreve os métodos de KDD para aplicar a descoberta de conhecimento sobre dados de uma base de dados educacional, da área de Educação a Distância. Os dados são provenientes de um Ambiente Virtual de Aprendizagem. O *data mining* é um dos métodos do KDD, que busca encontrar padrões para identificar subconjunto de dados, descobrindo informações que estão ocultas em meio a uma grande quantidade de dados brutos. Mediante as técnicas e tarefas, o *data mining* transforma dados em conhecimento útil. Para esta pesquisa, é aplicado o *data mining*, por meio de algoritmos da tarefa de agrupamento, de modo que é realizada a comparação entre o algoritmo de AGNES para o agrupamento hierárquico aglomerativo e o método Fuzzy é empregado para o agrupamento particional. A verificação dos métodos com melhor desempenho é realizada através das medidas de qualidade.

Palavras-chave: Educação a Distância, Educational Data Mining, Agrupamento, Agrupamento Hierárquico, Agrupamento Particional.

Introdução

O uso disseminado da tecnologia e dos computadores vem resultando no acúmulo de dados. Em consequência, as ferramentas tradicionais usadas para o gerenciamento de dados, tornam-se insatisfatórias para análise (TAN; STEINBACH; KUMAR, 2009). Desta forma, surgem técnicas computacionais e ferramentas, que apoiam a extração do conhecimento útil em grandes bases de dados. Essas técnicas e ferramentas são baseadas no *Knowledge Discovery in Databases* (KDD) e mineração de dados (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996, tradução nossa). O KDD aborda a metodologia geral da transformação de dados brutos em informações úteis, tendo como etapas do processo a preparação e seleção dos dados, aplicação dos algoritmos de mineração de dados, produzindo a leitura correta dos resultados (GOLDSCHMIDT; BEZERRA; PASSOS, 2015).

A mineração de dados é uma etapa do processo de KDD, entretanto é constantemente empregado como sendo o processo de descoberta de conhecimento, isso porque é essencial para o reconhecimento de padrões (HAN; PEI; KAMBER, 2011, tradução nossa).

As atividades educacionais e a utilização de banco de dados estão em fase crescente, devido à evolução da internet. Surge na área educacional a Educação a Distância (EaD), na qual é possível realizar um curso sem estar presente em uma sala de aula, bastando ao aluno ter acesso a *internet*. Os problemas da EaD são o de evasão e desempenho dos alunos, na qual o estudo e predição são difíceis, visto que o comportamento acadêmico depende de diversos fatores como variáveis pessoais (SATYANARAYANA; NUCKOWSKI, 2016, tradução nossa).

O *Education Data Mining* (EDM), uma área proveniente da mineração de dados, com foco no desenvolvimento de técnicas para coletar dados de uma base educacional. Na área do EDM a tarefa de agrupamento é aplicada para identificar grupos de alunos com características similares de aprendizado (RAMOS et al., 2017).

O problema de agrupamento está voltado para a descoberta de grupos que são significativos na extração dos dados. O primeiro objetivo deste problema é encontrar grupos que sejam semelhantes. O segundo objetivo consiste no fato dos dados serem semelhantes no mesmo grupo, porém diferentes dos outros grupos (SOARES et al., 2008). O agrupamento é dividido em dois métodos: agrupamento hierárquico e agrupamento de particionamento. Para empregar o agrupamento hierárquico, é necessário que seu início se organize formando uma decomposição hierárquica em um conjunto de dados sendo representado por uma árvore, que o divide em subconjuntos menores, até que cada subconjunto seja formado por um objeto (GOLDSCHMIDT; PASSOS; BEZERRA, 2015). O algoritmo Hierárquico Aglomerativo AGNES (Agglomerative Nesting), trabalha com o método de link único, na qual cada cluster pode ser representado por objetos de cluster. O objetivo do algoritmo é gerar repetidamente a fusão do cluster até que todo objeto se junta para formar um grupo (HUANG, 2015, tradução nossa). O agrupamento de particionamento é definido como uma divisão de conjuntos de objetos de dados em subconjuntos, que não estão ligados e cada objeto de dados fica exatamente dentro do seu subconjunto. Existem diversos algoritmos de agrupamento particional, um dos mais conhecidos é o *fuzzy c-means*, baseados na lógica *fuzzy* é usado quando os grupos não estão bem separados, o que possibilita o descobrimento de dados que podem pertencer a mais de um grupo (XU; WUNSCH, 2008, tradução nossa).

Metodologia

As etapas metodológicas empregadas no desenvolvimento, já concluídas foram as seguintes: levantamento bibliográfico, seleção da base de dados da disciplina para aplicação dos algoritmos de agrupamento, extração do banco de dados e segmentação da base de dados. A pesquisa está em fase de pré-processamento da base selecionada.

Seguindo para as etapas faltantes: aplicação do método de agrupamento particional por meio do algoritmo *Fuzzy C-Means*, aplicação do método de agrupamento hierárquico aglomerativo por meio do algoritmo AGNES, análise dos modelos obtidos por meio de medidas de qualidade para agrupamento em mineração de dados.

Desenvolvimento

A pesquisa apresentada está em desenvolvimento, de modo que se propõe aplicar a descoberta de conhecimento em uma base de dados proveniente de uma disciplina que é ministrada na modalidade à distância. A base de dados é gerada pela interação do aluno com o ambiente virtual utilizando os dados de navegação, administrativos e demográficos.

O objetivo da pesquisa é aplicar a tarefa de agrupamento e analisar a aplicação dos algoritmos AGNES para agrupamento Hierárquico Aglomerativo e *Fuzzy C-means* para agrupamento particional.

Por meio de medidas de qualidade, pretende se apresentar como resultado, qual método de agrupamento e algoritmo tem melhor qualidade para o conjunto de dados da pesquisa, identificando qual o melhor modelo de agrupamento.

Conclusão

Esta é uma pesquisa que está em andamento, porém, já são identificados alguns pontos sobre o tema abordado.

Nos últimos tempos o interesse na educação à distância, fez com que surgisse um novo conceito de mineração de dados, especialmente focado na educação e no conhecimento gerado a partir dos dados provenientes de escolas e universidades, o conceito é conhecido como *Educational Data Mining* (EDM).

Referências

- FAYYAD, Usama; PIATETSKY-SHAPIRO, Gregory; SMYTH, Padhraic. The KDD process for extracting useful knowledge from volumes of data. **Communications of the ACM**, v. 39, n. 11, p. 27-34, 1996.
- GOLDSCHMIDT, Ronaldo; BEZERRA, Eduardo; PASSOS, E. Data mining: conceitos, técnicas, algoritmos, orientações e aplicações. **Rio de Janeiro-RJ: Elsevier**, p. 56-60, 2015.
- HAN, Jiawei; PEI, Jian; KAMBER, Micheline. **Data mining: concepts and techniques**. Elsevier, 2011.
- HUANG, Wei-qing et al. an efficient cluster mining algorithm for the internal motion target path based on the enhanced AGNES. In: **Trustcom / BigDataSE / ISPA, 2015 IEEE** . IEEE, 2015. p. 1318-1323.
- RAMOS, Jorge Luis Cavalcanti et al. **Um Modelo Preditivo da Evasão dos Alunos na EAD a Partir dos Construtos da Teoria da Distância Transacional**. In: Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE). 2017. p. 1227.
- SATYANARAYANA, Ashwin; NUCKOWSKI, Mariusz. Data Mining using Ensemble Classifiers for Improved Prediction of Student Academic Performance. 2016
- TAN, Pang-Ning; STEINBACH, Michael; KUMAR, Vipin. **Introdução ao datamining: mineração de dados**. Ciência Moderna, 2009.
- XU, Rui; WUNSCH, Don. Clustering. John Wiley & Sons, 2008