

Filtragem horizontal de dados na predição de coautorias em redes sociais acadêmicas

Vitor Rodrigues da Cunha Estima, Luciano Antonio Digiampietri

Escola de Artes Ciências e Humanidades - Universidade de São Paulo

vrcestima@usp.br, digiampietri@usp.br

Objetivos

A predição de coautorias, com base nas informações sobre a produção científica do Brasil que podem ser obtidas da Plataforma Lattes, pode favorecer a comunicação entre pesquisadores e otimizar o processo de produção científica, identificando possíveis colaboradores.

Na predição de relacionamentos, tipicamente, a base de dados utilizada sofre de desbalanceamento de dados, ou seja, em uma rede social típica haverá muitos pares de pessoas que não se relacionarão e apenas uma pequena quantidade que irá se relacionar.

O objetivo deste trabalho é estudar e desenvolver técnicas de filtragem horizontal e classificação binária em inteligência artificial a fim de lidar com o desbalanceamento de dados e otimizar a predição.

Métodos e Procedimentos

O desenvolvimento deste trabalho foi iniciado por uma revisão bibliográfica de forma a identificar quais são as principais estratégias para tratar o desbalanceamento em grandes bases de dados. A base de dados utilizada foi obtida pelo Grupo de Análise de Redes Sociais e Cientometria (GARSC) da USP.

Os algoritmos de inteligência artificial utilizados neste trabalho foram aqueles disponíveis no arcabouço Weka.

As técnicas de filtragem especificadas e que estão sendo desenvolvidos ao longo deste trabalho estão sendo implementadas na linguagem de programação Java.

Após realizada a filtragem da base de dados, é realizada novamente a classificação, e então, uma comparação com os resultados anteriores, para entender o impacto que a filtragem horizontal teve na predição de coautorias.

Resultados

As técnicas de filtragem horizontal usadas melhoraram o equilíbrio entre as classes “irão se relacionar” e “não irão se relacionar” que são utilizadas para classificar cada par de pessoas presentes na base de dados utilizada, possibilitando assim uma predição mais precisa das coautorias.

Conclusões

Os resultados parciais obtidos são positivos, aplicando técnicas simples de filtragem horizontal já é observada uma diminuição nos efeitos negativos causados pelo desbalanceamento durante o processo de classificação.

Está sendo desenvolvida uma nova técnica de filtragem baseada em algoritmos genéticos que deverá trazer resultados ainda melhores.

Bibliografia

DIGIAMPIETRI, L. A. et al. Um sistema de predição de relacionamentos em redes sociais. Anais do XI Simpósio Brasileiro de Sistemas de Informação (SBSI 2015). p.139-146, 2015.
MARUYAMA, W. T. Predição de coautorias em redes sociais acadêmicas. 2016. 154 f. Dissertação (Mestrado em Ciências) - Escola de Artes, Ciências e Humanidades, Universidade de São Paulo, São Paulo, 2016.