

COMBINANDO ANÁLISE BIBLIOMÉTRICA E ANÁLISE DE REDES SOCIAIS PARA A AVALIAÇÃO DE GRUPOS ACADÊMICOS

Lucas Leal Caparelli, Luciano Antonio Digiampietri

Escola de Artes, Ciências e Humanidades da Universidade de São Paulo

lucas.caparelli@usp.br

Objetivos

A pesquisa desenvolvida neste projeto objetivou caracterizar programas brasileiros de pós-graduação em Ciência da Computação de acordo com diferentes medidas bibliométricas e oriundas da análise de redes sociais. Com base nestas medidas foram utilizadas técnicas estatísticas e de inteligência artificial para identificar as medidas que são mais relevantes em relação à classificação vigente da CAPES e diferentes estratégias foram testadas para a inferência da classificação utilizando as medidas selecionadas.

Métodos e Procedimentos

Foram utilizados dados dos 59 programas brasileiros de pós-graduação em Ciência da Computação que possuíam doutorado e/ou mestrado acadêmico durante todo o quadriênio de 2013 a 2016. A partir do site da CAPES foram extraídas as listas dos orientadores de cada um destes programas. Os dados das publicações foram extraídos da Plataforma Lattes. Por fim, informações sobre a quantidade de citações recebidas pelos pesquisadores foram extraídos do perfil do Google Acadêmico dos pesquisadores (DIGIAMPIETRI, FERREIRA, 2018). Para cada programa, foram calculados atributos bibliométricos como número de publicações, número de citações recebidas pelos orientadores, etc. e atributos oriundos da análise das redes sociais de coautorias entre pesquisadores de um mesmo programa e de diferentes programas, como centralidade, densidade e diâmetro (SCOTT, 2009). Com base nas métricas calculadas foram executados algoritmos de seleção de atributos a fim de identificar as métricas mais relevantes

e estas foram utilizadas para a classificação dos programas de acordo com as notas atribuídas pela CAPES. Foram usadas as implementações disponíveis no arcabouço Weka e os testes consideraram a estratégia de validação cruzada em dez subconjuntos.

Resultados

Considerando-se os programas com conceitos de 3 a 7 utilizou-se o conjunto de atributos selecionados por *CFSSubsetEval*. Foi realizada uma classificação em dois níveis. Inicialmente separou-se os programas com conceitos 6 e 7 dos demais e depois classificaram-se os programas restantes. O classificador *RandomTree* foi capaz de classificar com 100% de acerto todos programas com conceitos 6 e 7, permitindo a classificação do subconjunto restante separadamente. Aplicando o classificador *BayesNet* a esse novo conjunto (considerando-se apenas os programas com conceitos de 3 a 5) obteve-se índice de acerto de 83,64%, totalizando uma acurácia de 86,15% para o conjunto completo de dados.

Conclusões

Analisou-se a importância de diferentes métricas em relação aos conceitos atribuídos pela CAPES em sua avaliação quadrienal, bem como na capacidade de se inferir o conceito com base nestes atributos.

Referências Bibliográficas

- DIGIAMPIETRI, L.; FERREIRA, J. (2018). Desambiguação de nomes de autores para a identificação automática de perfis acadêmicos. *Em Questão*, v. 24, p. 37-54.
- SCOTT, J. (2009). *Social network analysis: a handbook* SAGE, 2ª edição.