

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 visa a 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 pretende-se utilizar técnicas estatísticas e de inteligência artificial para identificar as medidas que são mais relevantes em relação à classificação vigente da CAPES e utilizar diferentes estratégias para tentar inferir a 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 triênio de 2010 a 2012. A partir do site da CAPES foram extraídas as listas dos docentes de cada um destes programas. Os dados das publicações foram extraídos do repositório DBLP. Por fim, informações sobre a quantidade de citações recebidas pelas publicações foram obtidas do site *Microsoft Academic Search* (Digiampietri *et al*, 2016). Para cada programa, foram calculados atributos como número de publicações, número de citações por artigo, entre outros. Com estes dados são calculadas outras métricas, tanto bibliométricas (como índice h), quanto da análise de redes sociais (como centralidade) (SCOTT, 2009), utilizando a rede de coautorias.

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 e validações foram realizados considerando a estratégia de validação cruzada em dez subconjuntos.

Resultados

O seletor de atributos “*CFS Subset Eval*” foi utilizado para selecionar as métricas devido a sua capacidade de manter apenas métricas com alta correlação com a classificação e baixa correlação entre si. A classificação foi capaz de acertar a nota CAPES de cerca de 70% dos programas utilizando um modelo baseado em redes bayesianas (estratégia que se apoia numa rede de dependências de características e distribuições condicionais). Adicionalmente, para os programas em que a nota certa não foi atribuída, o sistema errou a nota por apenas um ponto (por exemplo, atribuiu nota 4 a um programa nota 5).

Conclusões

Os resultados iniciais foram bastante promissores, acertando a classificação de cerca de 70% dos dados e, no caso dos erros, a margem é de apenas um ponto. Como trabalho futuro, o sistema será refinado, adicionando-se novos atributos (como número de orientações) e utilizando uma estratégia de classificação hierárquica.

Referências Bibliográficas

- DIGIAMPIETRI, L., LINDEN, R.; BARBOSA, L. (2016). Caracterizando Departamentos e Programas de Computação Utilizando Análise de Redes Sociais e Bibliometria. *V Brazilian Workshop on Social Network Analysis and Mining*, p. 67-78.
- SCOTT, J. (2009). *Social network analysis: a handbook* SAGE, 2ª edição.