

I Workshop de Dissertações do Programa de Pós-graduação em Sistemas de Informação (PPgSI)

Predição de coautorias em redes sociais acadêmicas

Aluno: William Takahiro Maruyama e Orientador: Dr. Luciano Antonio Digiampietri

Contexto

As redes sociais são formadas por pessoas e seus diferentes relacionamentos. Um tipo de rede social é a acadêmica, formada tipicamente por relacionamentos de coautoria entre pesquisadores, professores e alunos em publicações científicas. A Plataforma Lattes é uma importante fonte de informação para criação de redes sociais acadêmicas. A predição de relacionamentos em redes sociais acadêmicas pode ser aplicada para prever as parcerias de coautoria. Essa predição pode ser útil para potencializar colaborações por meio da indicação de quais seriam as parcerias mais promissoras. No entanto isto não é uma tarefa simples, pois envolve desafios como: identificação de atributos, combinação dos atributos e uso correto do fator temporal.

Objetivo

Pretende-se identificar diferentes atributos relacionados tanto às características específicas de cada pesquisador como da rede social de coautorias na qual estes pesquisadores estão inseridos e, com base nestes atributos, desenvolver uma técnica de predição de relacionamentos que combine estas informações de forma a tratar tanto o problema de predição de novas coautorias como o problema geral de predição de coautorias.

Método

Inicialmente foi realizada a seleção e armazenamento da amostra para estudo. Para o primeiro experimento se obteve 657 docentes permanentes que atuaram em ambos os triênios 2004-2006 e 2007-2009 nos programas de pós-graduação em Ciência da Computação no Brasil avaliados pela CAPES. Em seguida foram identificados, selecionados e desenvolvidos os atributos e métricas estruturais de redes sociais. Além disso, aplicou-se uma filtragem horizontal e vertical nos dados antes de submetê-los para os algoritmos de classificação binária presentes no arcabouço Weka.

Resultados

O experimento inicial teve foco em coautorias inéditas. A classificação atingiu acurácia de 95,956% e destaca-se o valor preditivo positivo de 98,958% . No entanto, apenas 21,49% (sensibilidade) das novas relações de coautoria foram identificadas. No caso geral de predição obtivemos acurácia de 94,7435%.

Conclusões

O experimento apresentou resultados satisfatórios apesar do desbalanceamento das classes e esparsidade dos dados. Os mesmos sugerem que o conjunto total de atributos utilizados pode ser aplicado tanto para a predição de novas coautorias quanto para a predição geral de coautorias.