Uma Abordagem Baseada em Ontologias para a Predição de Ligações entre Pesquisadores em Redes de Colaboração Científica

Thiago Henrique Dias Araujo

Texto apresentado
AO
Instituto de Matemática e Estatística
DA
Universidade de São Paulo
Para
O Exame de Qualificação
DE
Mestre em Ciências

Programa: Mestrado em Ciência da Computação Orientadora: Prof^a. Dr^a. Renata Wassermann

São Paulo, março de 2016

Uma Abordagem Baseada em Ontologias para a Predição de Ligações entre Pesquisadores em Redes de Colaboração Científica

Esta é a versão original do texto elaborado pelo candidato Thiago Henrique Dias Araujo para o exame de qualificação apresentado ao Instituto de Matemática e Estatística da Universidade de São Paulo como requisito para obtenção de título de Mestre em Ciências.

Resumo

ARAUJO, T. H. D. Uma Abordagem Baseada em Ontologias para a Predição de Ligações entre Pesquisadores em Redes de Colaboração Científica. 2016. 20 f. Exame de qualificação (Mestrado) - Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2016.

A comunidade científica pode ser vista como uma rede onde cada pesquisador se relaciona com outros pesquisadores através de colaborações, como na co-autoria de um artigo científico. Alguns trabalhos utilizam ontologias para modelar o domínio das redes de colaboração, e outros aplicam técnicas de aprendizado de máquina para prever ligações entre pessoas dentro dessas redes. Entretanto, algumas limitações existem nessas metodologias por utilizarem formas pouco expressivas de representação dessas relações, ou por não aproveitarem características específicas das entidades na análise desse domínio. A proposta do presente trabalho é criar uma Ontologia capaz de indicar características próprias dessas pessoas, e descrever relações entre elas, aplicando esse conhecimento em um modelo de aprendizado capaz de descobrir e prever novas relações entre esses pesquisadores.

Palavras-chave: redes de colaboração científica, ontologia, aprendizado de máquina, predição de ligações.

Abstract

ARAUJO, T. H. D. Uma Abordagem Baseada em Ontologias para a Predição de Ligações entre Pesquisadores em Redes de Colaboração Científica. 2016. 20 f. Exame de qualificação (Mestrado) - Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2016.

Keywords: scientific collaboration networks, ontology, machine learning, link prediction.

Sumário

Lista de Abreviaturas Lista de Figuras			ix xi	
1	Inti	odução	1	
	1.1	Motivação	1	
	1.2	Objetivos	1	
	1.3	Metodologia	1	
	1.4	Contribuições	2	
	1.5	Organização do Trabalho	2	
2	Conceitos			
	2.1	Fundamentos	3	
3 Conclusões		5		
\mathbf{R}	eferê	ncias Bibliográficas	7	

Lista de Abreviaturas

SPARQL (SPARQL Protocol and RDF Query Language)

SQL (Structured Query Language) SRL (Statistical Relational Learning)

SVM (Support Vector Machine)

SWRL (Semantic Web Rule Language)

Lista de Figuras

Lista de Tabelas

Capítulo 1

Introdução

Texto texto.

1.1 Motivação

Texto texto.

1.2 Objetivos

O objetivo do presente trabalho é demonstrar que um modelo de predição de ligações em uma rede de colaboração científica, quando enriquecido com conhecimento prévio (background knowledge) extraído de uma ontologia, se torna mais eficaz em sua tarefa de predição. Para isso, espera-se que os seguintes passos sejam cumpridos:

- Modelar uma Ontologia para representar o conhecimento a respeito de uma rede de colaboração científica, contendo informações sobre pesquisadores, publicações e colaborações, e instituições.
- Construir consultas em SPARQL, que utilizam regras de inferência na Ontologia, e cujo resultado seja utilizado como conhecimento prévio para enriquecer o modelo de predição. Algumas dessas consultas são:
 - Qual a área de pesquisa de um pesquisador?
 - Quais são os pesquisadores que foram orientados por um dado pesquisador?
 - Há quantos anos um dado pesquisador frequenta uma instituição?
- Construir um modelo de predição de ligações, baseado no trabalho de Cervantes (2014).
- Enriquecer o modelo de predição com as informações extraídas da ontologia.
- Comparar a eficácia dos dois modelos.
- Propor novas consultas que possam ser utilizadas como conhecimento prévio e fazer novas comparações.

1.3 Metodologia

Texto texto.

2 INTRODUÇÃO 1.5

1.4 Contribuições

As principais contribuições deste trabalho são as seguintes:

• Item 1. Texto texto.

• Item 2. Texto texto.

1.5 Organização do Trabalho

No Capítulo 2, apresentamos os conceitos ... Finalmente, no Capítulo 3 discutimos algumas conclusões obtidas neste trabalho. Analisamos as vantagens e desvantagens do método proposto ... As sequências testadas no trabalho estão disponíveis no Apêndice

Capítulo 2

Conceitos

Texto texto

2.1 Fundamentos

Texto texto

4 CONCEITOS 2.1

Capítulo 3

Conclusões

Texto texto

 $^{^1\}mathrm{Exemplo}$ de referência para página Web: www.vision.ime.usp.br/~jmena/stuff/tese-exemplo

Referências Bibliográficas

Cervantes (2014) Everlyn Perez Cervantes. Análise de Redes de Colabração Científica: Uma Abordagem Baseada em Grafos Relacionais com Atributos, 2014. Citado na pág. 1