A STUDY ON SIMILARITY JOIN IN HETEROGENEOUS INFORMATION NETWORK USING PATH BASED ALGORITHM

S. Sandhya

Archers & Elevators Publishing House
ISBN : 978-81-19385-00-3

A Study on Similarity Join in Heterogeneous Information Network Using Path Based Algorithm

S. Sandhya
Assistant Professor, Department of CSE,
GNITS,Hyderabad.

PREFACE

As a newly emerging network model, heterogeneous information networks (HINs) have received growing attention. Many data mining tasks have been explored in HINs, including clustering, classification, and similarity search. Similarity join is a fundamental operation required for many problems. It is attracting attention from various applications on network data, such as friend recommendation, link prediction, and online advertising. Although similarity join has been well studied in homogeneous networks, it has not yet been studied in heterogeneous networks. Especially, none of the existing research on similarity join takes different semantic meanings behind paths into consideration and almost all completely ignore the heterogeneity and diversity of the HINs. In this paper, we propose a Path-based Similarity join (PS-join) method to return the top k similar pairs of objects based on any user specified join path in a heterogeneous information network. We study how to prune expensive similarity computation by introducing BPLSH (Bucket Pruning based Locality Sensitive Hashing) indexing. Compared with existing Link-based Similarity join (LS-join) method, PS-join can derive various similarity semantics. Experimental results on real data sets show the efficiency and effectiveness of the proposed approach

TABLE OF CONTENTS

Sl.No.	Topic	Page No.
CHAPTER -1	Introduction	4
	1.1 Objectives	5
	1.2 Existing System	6
	1.3 Proposed System	6
CHAPTER-2	Literature Survey	7
	2.1 Introduction	7
	2.2 History	12
	2.3 Purpose	13
CHAPTER-3	Fundamental Concepts	14
	3.1 Domain Fundamentals And Description	15
	3.2 Data Mining Overview	17
	3.3 Algorithm	18
	3.4 Proposed System Algorithm	20
CHAPTER-4	System Analysis	22
	4.1 Software Models	22
	4.2 SDLC Models	23
	4.3 Modules Description	26
CHAPTER-5	Feasibility Study	27
CHAPTER-6	System Requirements	29
CHAPTER-7	System Design	31
CHAPTER-8	Implementation	37
CHAPTER-9	System testing	43
CHAPTER-10	Screenshots	46