Meta Paths and Meta Structures: Computing Relevance in Large Heterogeneous Information Networks

ABSTRACT

A heterogeneous information network (HIN) is a graph model in which objects and edges are annotated with types. Large and complex databases, such as YAGO and DBLP, can be modeled as HINs. A fundamental problem in HINs is the computation of closeness, or relevance, between two HIN objects. Relevance measures can be used in various applications, including information retrieval, entity resolution, and product recommendation. We study the problem of discovering meta paths, a recently proposed concept used for relevance computation in HINs. A meta path is essentially a sequence of node classes and edge types between two nodes in a HIN. Current methods assume that meta paths are found by domain experts. However, in a large and complex HIN, retrieving meta paths manually can be tedious and difficult. We examine how to find meta paths automatically. Specifically, users are asked to provide example pairs of nodes that exhibit high proximity. We then investigate how to generate meta paths that can best explain the relationship between these node pairs. Since this problem is computationally intractable, we propose a greedy algorithm to select the most relevant meta paths. We also present a data structure to enable efficient execution of this algorithm. We incorporate hierarchical relationships among node classes in our solutions. Extensive experiments on YAGO and DBLP show that our approach captures important meta paths in an efficient and scalable manner.

We further generalize the notion of meta path to “meta structure”, which is a directed acyclic graph of object types with edge types connecting in between. The meta structure, which is more expressive than the meta path, can describe complex relationship between two HIN objects (e.g., two papers in DBLP share the same authors and topics). We develop three relevance measures based on meta structure. Due to the computational complexity of these measures, we design an algorithm with data structures proposed to support their evaluation. Our extensive experiments show that meta structure relevance is more effective than meta paths.

BIOGRAPHY

Dr Reynold Cheng is an Associate Professor of the Department of Computer Science in the University of Hong Kong. He was an Assistant Professor in HKU in 2008-11. He received his BEng (Computer Engineering) in 1998, and MPhil (Computer Science and Information Systems) in 2000, from the Department of Computer Science in the University of Hong Kong. He then obtained his MSc and PhD from Department of Computer Science of Purdue University in 2003 and 2005 respectively. Dr Cheng was an Assistant Professor in the Department of Computing of the Hong Kong Polytechnic University during 2005-08. He was a visiting scientist in the Institute of Parallel and Distributed Systems in the University of Stuttgart during the summer of 2006.

Dr Cheng was granted an Outstanding Young Researcher Award 2011-12 by HKU. He was the recipient of the 2010 Research Output Prize in the Department of Computer Science of HKU. He also received the U21 Fellowship in 2011. He received the Performance Reward in years 2006 and 2007 awarded by the Hong Kong Polytechnic University. He is the Chair of the Department Research Postgraduate Committee, and was the Vice Chairperson of the ACM (Hong Kong Chapter) in 2013. He is a member of the IEEE, the ACM, the Special Interest Group on Management of Data (ACM SIGMOD), and the UPE (Upsilon Pi Epsilon Honor Society). He is an editorial board member of TKDE, DAPD and IS, and was a guest editor for TKDE, DAPD, and Geoinformatica. He is an area chair of ICDE 2017, a senior PC member for DASFAA 2015, PC co-chair of APWeb 2015, area chair for CIKM 2014, area chair for Encyclopedia of Database Systems, program co-chair of SSDT 2013, and a workshop co-chair of ICDE 2014. He received an Outstanding Service Award in the CIKM 2009 conference. He has served as PC members and reviewer for top conferences (e.g., SIGMOD, VLDB, ICDE, EDBT, KDD, ICDM, and CIKM) and journals (e.g., TODS, TKDE, VLDBJ, JS, and TMC).

All are welcome!