

Chapter 26

System Uncertainty Based Data-Driven Knowledge Acquisition

Jun Zhao

Chongqing University of Posts & Telecommunications, P.R. China

Guoyin Wang

Chongqing University of Posts & Telecommunications, P.R. China

ABSTRACT

In the three-layered framework for knowledge discovery, it is necessary for technique layer to develop some data-driven algorithms, whose knowledge acquiring process is characterized by and hence advantageous for the unnecessary of prior domain knowledge or external information. System uncertainty is able to conduct data-driven knowledge acquiring process. It is crucial for such a knowledge acquiring framework to measure system uncertainty reasonably and precisely. Herein, in order to find a suitable measuring method, various uncertainty measures based on rough set theory are comprehensively studied: their algebraic characteristics and quantitative relations are disclosed; their performances are compared through a series of experimental tests; consequently, the optimal measure is determined. Then, a new data-driven knowledge acquiring algorithm is developed based on the optimal uncertainty measure and the Skowron's algorithm for mining propositional default decision rules. Results of simulation experiments illustrate that the proposed algorithm obviously outperforms some other congeneric algorithms.

INTRODUCTION

Cognitive Informatics (CI) is a new transdisciplinary research field. It combines cognitive science with informatics. The combination captures both the understanding of intelligence in minds

and its implementation in computers. That characteristic makes CI develops rapidly in recent years (Kinsner, 2005; Wang, 2002; Wang, 2002; Wang, 2003; Wang, 2003; Wang, 2004; Wang & Kinsner, 2006; Wang, 2007). The general ideas of CI can be applied in studying related problems. In the field of knowledge discovery in databases, a three-layered framework can be accordingly

DOI: 10.4018/978-1-4666-0261-8.ch026

established based on those ideas (Yao, 2004). The three layers, from the inner to the outer, are philosophy layer, technique layer and application layer. Different layer deals with problems in different contexts, i.e. in mind, in computer and in application, respectively. Each layer provides services for the outer ones. According to this three-layered framework, philosophy layer conceptually provides prior domain knowledge for technique layer. However, it is very possible that prior knowledge is biased or even incorrect in some real-life applications. To make matter even worse, prior knowledge is very possibly unavailable at all in some cases which have not been well explored yet. Therefore theoretically and practically speaking, it is necessary for technique layer to develop algorithms that are independent of prior knowledge. Such algorithms are completely data-driven, namely, their knowledge acquiring process can be conducted or directed by information systems themselves, and therefore prior knowledge or external information is no longer necessary for them.

Essentially, knowledge acquiring process can be regarded as a kind of knowledge transformation. In an original information system, knowledge exists in the form of primitive data. To make the knowledge more understandable and applicable, several steps may be involved in data mining process. For example, in rough set theory, a typical processing sequence includes steps such as discretization, reduction and rule acquisition. Accordingly, knowledge is sequentially transformed from original data to discretized data, reduced forms, and then ultimate decision rules or decision trees. Conceptually speaking, knowledge needs to be kept essentially undisturbed in the transforming sequence. Thus, data-driven methods have to utilize some special features of knowledge to direct its mining process. One of the most interesting issues of data-driven knowledge acquisition is to find out such reasonable features. For knowledge in different forms, those features must be commonly existed and intrinsic, and can be

quantitatively measured and comparable. System uncertainty is such a good candidate. Due to the inherent existence of various uncertain factors, system uncertainty is an intrinsic common feature of and hence becomes an essential link between information systems and their induced knowledge systems. In fact, data-driven algorithms for acquiring decision rules have already been successfully implemented based on the minimum degree of the local certainty of information systems (Wang & He, 2002; Wang & He, 2003). Obviously, a data-driven knowledge acquiring framework based on system uncertainty requires first of all, measuring and handling system uncertainty as reasonably and precisely as possible. It is obvious that the ultimate performance may be negatively influenced if an improper uncertainty measure is applied in such an algorithm.

In rough set theory, the uncertainty concerned problem is one of the hottest research topics, and some results have already been gotten: some system uncertainty measures have been proposed (Chen, Zhu, & Ji, 2001; Düntsch & Gediga, 1998; Pawlak, Grzymala-Busse, Slowinski & Ziarko, 1995; Wang, 2001; Wang, 2001; Wang & He, 2002; Wang & He, 2003); several algorithms based on system uncertainty have been implemented (Mollestad & Skowron, 1996; Wang & He, 2002; Wang & He, 2003). All these achievements paved a concrete way for the realization of data-driven knowledge acquisition based on system uncertainty. However, to my best knowledge, up to now in literature there is no thorough study or comparison on those rough set model based system uncertainty measures. Herein, to find a suitable measuring method for directing knowledge acquisition, various system uncertainty measures based on rough set theory are comprehensively studied: their algebraic characteristics and quantitative relations are theoretically analyzed; their performance is experimentally compared; and the optimal measure is then determined based on the results of the analyses and experiments. Then a new data-driven knowledge acquiring algorithm is

13 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:
www.igi-global.com/chapter/system-uncertainty-based-data-driven/65144

Related Content

Effect of Power and Phase Synchronization in Multi-Trial Speech Imagery

Sandhya Chengaiyan, Divya Balathayil, Kavitha Anandanand Christy Bobby Thomas (2018). *International Journal of Software Science and Computational Intelligence* (pp. 44-61).

www.irma-international.org/article/effect-of-power-and-phase-synchronization-in-multi-trial-speech-imagery/223494

Data Warehousing and Decision Support in Mobile Wireless Patient Monitoring

Barin N. Nagand Mark Siegal (2012). *Machine Learning: Concepts, Methodologies, Tools and Applications* (pp. 1642-1651).

www.irma-international.org/chapter/data-warehousing-decision-support-mobile/56218

Robust Dimensionality Reduction: A Resistant Search for the Relevant Information in Complex Data

Jan Kalina (2023). *Convergence of Big Data Technologies and Computational Intelligent Techniques* (pp. 186-210).

www.irma-international.org/chapter/robust-dimensionality-reduction/314342

A New Fuzzy Rule Interpolation Approach to Terrorism Risk Assessment

Shangzhu Jin, Jike Geand Jun Peng (2017). *International Journal of Software Science and Computational Intelligence* (pp. 16-36).

www.irma-international.org/article/a-new-fuzzy-rule-interpolation-approach-to-terrorism-risk-assessment/190316

Condition Monitoring Using Computational Intelligence

Tshilidzi Marwalaand Christina Busisiwe Vilakazi (2008). *Handbook of Computational Intelligence in Manufacturing and Production Management* (pp. 106-123).

www.irma-international.org/chapter/condition-monitoring-using-computational-intelligence/19355