
IGI PUBLISHING **ITB15226**
701 E. Chocolate Avenue, Suite 200, Hershey PA 17033-1240, USA
Tel: 717/533-8845; Fax 717/533-8661; URL-http://www.igi-global.com
This paper appears in the publication, **Data Mining and Knowledge Discovery Technologies** edited by D. Taniar © 2008, IGI Global

Chapter IX

Domain Driven Data Mining

Longbing Cao, University of Technology, Sydney, Australia
Chengqi Zhang, University of Technology, Sydney, Australia

Abstract

Quantitative intelligence-based traditional data mining is facing grand challenges from real-world enterprise and cross-organization applications. For instance, the usual demonstration of specific algorithms cannot support business users to take actions to their advantage and needs. We think this is due to quantitative intelligence focused data-driven philosophy. It either views data mining as an autonomous data-driven, trial-and-error process, or only analyzes business issues in an isolated, case-by-case manner. Based on experience and lessons learned from real-world data mining and complex systems, this article proposes a practical data mining methodology referred to as domain-driven data mining. On top of quantitative intelligence and hidden knowledge in data, domain-driven data mining aims to meta-synthesize

Copyright © 2008, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

quantitative intelligence and qualitative intelligence in mining complex applications in which human is in the loop. It targets actionable knowledge discovery in constrained environment for satisfying user preference. Domain-driven methodology consists of key components including understanding constrained environment, business-technical questionnaire, representing and involving domain knowledge, human-mining cooperation and interaction, constructing next-generation mining infrastructure, in-depth pattern mining and postprocessing, business interestingness and actionability enhancement, and loop-closed human-cooperated iterative refinement. Domain-driven data mining complements the data-driven methodology, the metasynthesis of qualitative intelligence and quantitative intelligence has potential to discover knowledge from complex systems, and enhance knowledge actionability for practical use by industry and business.

Introduction

Traditionally data mining is presumed as an automated process. It produces automatic algorithms and tools with limited or no human involvement. As a result, they lack the capability of adapting to external environment change. Many patterns mined but few are workable in real business. On the other hand, real-world data mining must adapt to dynamic situations in the business world. It also expects actionable discovered knowledge that can afford important grounds to business decision makers for performing appropriate actions.

Unfortunately, mining actionable knowledge is not a trivial task. As pointed out by the panel discussions of SIGKDD 2002 and 2003 (Ankerst, 2002, Fayyad & Shapiro, 2003), it was highlighted as one of the grand challenges for the extant and future data mining. The weakness of existing data mining partly results from the data-driven trial-and-error methodology (Ankerst, 2002), which depreciates the roles of domain resources such as domain knowledge and humans. For instance, data mining in the real world such as crime pattern mining (Bagui, 2006) is highly constraint based (Boulicaut & Jeudy, 2005; Fayyad et al., 2003). Constraints involve technical, economical, and social aspects in the process of developing and deploying actionable knowledge. For actionable knowledge discovery from data embedded with the previous constraints, it is essential to slough off the superficial and captures the essential information from data mining.

Many data mining researchers have realized the significant roles of some domain-related aspects, for instance, domain knowledge and constraints, in data mining. They further develop specific corresponding data mining areas such as constraint-based data mining to solve issues in traditional data mining. As a result, data mining is progressing toward a more flexible, specific, and practical manner with increasing capabilities of tackling real-world emerging complexities. In particular, data min-

Copyright © 2008, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

26 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/domain-driven-data-mining/7518

Related Content

Coronavirus Pandemic (COVID-19): Emotional Toll Analysis on Twitter
Jalal S. Alowibdi, Abdulrahman A. Alshdadi, Ali Daud, Mohamed M. Dessoukyand Essa Ali Alhazmi (2022). *Research Anthology on Implementing Sentiment Analysis Across Multiple Disciplines (pp. 1761-1782).*
www.irma-international.org/chapter/coronavirus-pandemic-covid-19/308574

A Method for Generating Comparison Tables From the Semantic Web
Arnaud Giacometti, Béatrice Markhoffand Arnaud Soulet (2022). *International Journal of Data Warehousing and Mining (pp. 1-20).*
www.irma-international.org/article/a-method-for-generating-comparison-tables-from-the-semantic-web/298008

Mining Customer Knowledge for a Recommendation System in Convenience Stores
Shu-Hsien Liao, Chih-Hao Wen, Pei-Yuan Hsian, Chien-Wen Liand Che-Wei Hsu (2014). *International Journal of Data Warehousing and Mining (pp. 55-86).*
www.irma-international.org/article/mining-customer-knowledge-for-a-recommendation-system-in-convenience-stores/110386

A Survey of Collective Anomaly Detection on Sequence Dataset
Xiaodi Huang, Po Yunand Zhongfeng Hu (2023). *International Journal of Data Warehousing and Mining (pp. 1-22).*
www.irma-international.org/article/a-survey-of-collective-anomaly-detection-on-sequence-dataset/327363

Privacy-Preserving Data Mining on the Web: Foundations and Techniques
Stanley R.M. Oliveiraand Osmar R. Zaïane (2009). *Handbook of Research on Text and Web Mining Technologies (pp. 468-481).*
www.irma-international.org/chapter/privacy-preserving-data-mining-web/21741