Lecture Notes in Computer Science 12647

Founding Editors

Gerhard Goos
Karlsruhe Institute of Technology, Karlsruhe, Germany
Juris Hartmanis
Cornell University, Ithaca, NY, USA

Editorial Board Members

Elisa Bertino
Purdue University, West Lafayette, IN, USA
Wen Gao
Peking University, Beijing, China
Bernhard Steffen
TU Dortmund University, Dortmund, Germany
Gerhard Woeginger
RWTH Aachen, Aachen, Germany
Moti Yung
Columbia University, New York, NY, USA

More information about this subseries at http://www.palgrave.com/gp/series/7409

Yan Jia · Zhaoquan Gu ·
Aiping Li (Eds.)

MDATA: A New Knowledge Representation Model

Theory, Methods and Applications

Springer

Editors
Yan Jia
Harbin Institute of Technology, Shenzhen
Shenzhen, China

Zhaoquan Gu
Guangzhou University
Guangzhou, China

Aiping Li
National University of Defense Technology
Changsha, China

ISSN 0302-9743 ISSN 1611-3349 (electronic)
Lecture Notes in Computer Science
ISBN 978-3-030-71589-2 ISBN 978-3-030-71590-8 (eBook)
https://doi.org/10.1007/978-3-030-71590-8

LNCS Sublibrary: SL3 – Information Systems and Applications, incl. Internet/Web, and HCI

© Springer Nature Switzerland AG 2021
This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

With the fast development of information technologies, many researchers are dedicated to knowledge representation, a high-level understanding and abstraction of massive data. Although many knowledge representation models have been proposed, such as symbolic logic models, semantic networks, expert systems, semantic webs, and knowledge graphs, these models cannot represent procedural knowledge, i.e. dynamic knowledge that changes with temporal and spatial factors.

In this edited book, we introduce a new knowledge representation model called MDATA (Multi-dimensional Data Association and in Telligent Analysis). After investigating a large number of related works in knowledge representation and cognitive models, we propose the MDATA model on the basis of the widely adopted knowledge graphs. By modifying the representation of relations and properties in knowledge graphs, dynamic knowledge can be efficiently described with temporal and spatial characteristics. The MDATA model can be regarded as a high-level temporal and spatial knowledge graph model which has strong capabilities for knowledge representation.

Similar as knowledge graphs, there are many key technologies in the MDATA model, such as entity recognition, relation extraction, property extraction, etc. The MDATA model pays particular attention to temporal and spatial characteristics, and the recognition of such temporal and spatial characteristics is important. These characteristics could be time period, location, or even IP address. In order to improve the knowledge graph, knowledge reasoning and completion are necessary, such as entity reasoning and relation completion. In the MDATA model, the dynamic knowledge should also be updated by reasoning, which combines the temporal and spatial factors. As many existing technologies have been proposed in constructing knowledge graphs, we need to extend them to the MDATA model.

The MDATA model can represent knowledge from different dimensions and it can be applied in many critical applications. The outbreak of the coronavirus pandemic has caused serious consequences. By representing personal knowledge such as dynamic travel information by the MDATA model, we can identify close contacts of infected persons in an efficient way. Considering network attacks that might happen every day around the world, the MDATA model can combine the analysis of massive data from log or network traffic to detect attack events. For social network analysis, in which information updates quickly and people's opinion might change, the MDATA model can represent such dynamic knowledge for better analysis, such as computing individual influence, community detection and sentiment analysis.

To promote the MDATA model, we have brought together research teams from many universities, research institutes, and companies, such as the Harbin Institute of Technology, Shenzhen (China), Guangzhou University (China), National University of Defense Technology (China), Zhejiang University of Technology (China), Hangzhou Dianzi University (China), The University of Queensland (Australia), The University

of New South Wales (Australia), etc. We have applied the MDATA model in many practical applications: for example, we have developed a cyberspace security situation awareness system that can detect various network attacks in a timely way. We have also organized many seminars and workshops to exchange the latest research, such as the MDATA workshop in the IEEE International Conference on Data Science in Cyberspace. We hope the MDATA model will be widely adopted in various areas and can promote the study of knowledge representation significantly.

The edited book is broken into 14 chapters. The first two chapters introduce the background and the representation method of the MDATA model. Some key technologies are then introduced from Chapter 3 to Chapter 9, including entity recognition, entity alignment, knowledge reasoning, etc. In the remaining five chapters, we introduce applications of the MDATA model, including network attack detection, social network analysis, etc.

This book can be treated as a handbook of knowledge representation models. This book introduces some key technologies in the MDATA model and it also covers applications in which the MDATA model has a valuable role to play. These applications are just a small sample of many potential applications that can benefit from the model. This book offers in particular an in-depth introduction of the model and the related technologies. The MDATA model should be of interest to readers from other research fields such as databases, cyberspace security, and social networks as the need for knowledge representation arises naturally in many practical scenarios.

Should you have any questions or suggestions, please contact the editors via e-mail to jiayanjy@vip.sina.com, zqgu@gzhu.edu.cn, or liaiping@nudt.edu.cn.

January 2021

Yan Jia
Zhaoquan Gu
Aiping Li

Organization

Editorial Team

Yan Jia	Harbin Institute of Technology, Shenzhen, China
Zhaoquan Gu	Guangzhou University, China
Aiping Li	National University of Defense Technology, China

Program Committee

Zhihong Tian	Guangzhou University, China
Bin Zhou	National University of Defense Technology, China
Qi Xuan	Zhejiang University of Technology, China
Yizhi Ren	Hangzhou Dianzi University, China
Weihong Han	Guangzhou University, China
Le Wang	Guangzhou University, China
Ye Wang	National University of Defense Technology, China
Mohan Li	Guangzhou University, China
Shanqing Yu	Zhejiang University of Technology, China
Jinyin Chen	Zhejiang University of Technology, China
Hongkui Tu	National University of Defense Technology, China
Rong Jiang	National University of Defense Technology, China
Xiang Wang	National University of Defense Technology, China
Qingpeng Zhang	City University of Hong Kong, Hong Kong, China

Contents

Introduction to the MDATA Model 1
Yan Jia, Zhaoquan Gu, Aiping Li, and Weihong Han

The Framework of the MDATA Computing Model 19
Yan Jia, Aiping Li, Zhaoquan Gu, Bin Zhou, Ye Wang, Xiang Wang, and Hongkui Tu

Spatiotemporal Data Cleaning and Knowledge Fusion 32
Huchen Zhou, Mohan Li, Zhaoquan Gu, and Zhihong Tian

Chinese Named Entity Recognition: Applications and Challenges 51
Qisen Xi, Yizhi Ren, Siyu Yao, Guohua Wu, Gongxun Miao, and Zhen Zhang

Joint Extraction of Entities and Relations: An Advanced BERT-based Decomposition Method 82
Changhai Wang and Aiping Li

Entity Alignment: Optimization by Seed Selection 99
Xiaolong Chen, Le Wang, Yunyi Tang, Weihong Han, Zhihong Tian, and Zhaoquan Gu

Knowledge Extraction: Automatic Classification of Matching Rules 117
Yunyi Tang, Le Wang, Xiaolong Chen, Zhaoquan Gu, and Zhihong Tian

Network Embedding Attack: An Euclidean Distance Based Method 131
Shanqing Yu, Jun Zheng, Yongqi Wang, Jinyin Chen, Qi Xuan, and Qingpeng Zhang

Few-Shot Knowledge Reasoning: An Attention Mechanism Based Method 152
Haocheng Xie, Aiping Li, and Yan Jia

Applications of Knowledge Representation Learning 165
Chenchen Li, Aiping Li, Ye Wang, and Hongkui Tu

Detection and Defense Methods of Cyber Attacks 185
Kai Xing, Aiping Li, Rong Jiang, and Yan Jia

A Distributed Framework for APT Attack Analysis 199
Yulu Qi, Rong Jiang, Aiping Li, Zhaoquan Gu, and Yan Jia

Social Unrest Events Prediction by Contextual Gated Graph Convolutional Networks 220
Haiyang Wang, Bin Zhou, Zhipin Gu, and Yan Jia

Information Cascading in Social Networks 234
Liqun Gao, Bin Zhou, Yan Jia, Hongkui Tu, and Ye Wang

Author Index 255