EAI/Springer Innovations in Communication and Computing

Honghao Gao · Jung Yoon Kim
Walayat Hussain · Muddesar Iqbal
Yucong Duan *Editors*

Intelligent Processing Practices and Tools for E-Commerce Data, Information, and Knowledge

EAI
RESEARCH MEETS INNOVATION

Springer

EAI/Springer Innovations in Communication and Computing

Series Editor

Imrich Chlamtac, European Alliance for Innovation, Ghent, Belgium

Editor's Note

The impact of information technologies is creating a new world yet not fully understood. The extent and speed of economic, life style and social changes already perceived in everyday life is hard to estimate without understanding the technological driving forces behind it. This series presents contributed volumes featuring the latest research and development in the various information engineering technologies that play a key role in this process.

The range of topics, focusing primarily on communications and computing engineering include, but are not limited to, wireless networks; mobile communication; design and learning; gaming; interaction; e-health and pervasive healthcare; energy management; smart grids; internet of things; cognitive radio networks; computation; cloud computing; ubiquitous connectivity, and in mode general smart living, smart cities, Internet of Things and more. The series publishes a combination of expanded papers selected from hosted and sponsored European Alliance for Innovation (EAI) conferences that present cutting edge, global research as well as provide new perspectives on traditional related engineering fields. This content, complemented with open calls for contribution of book titles and individual chapters, together maintain Springer's and EAI's high standards of academic excellence. The audience for the books consists of researchers, industry professionals, advanced level students as well as practitioners in related fields of activity include information and communication specialists, security experts, economists, urban planners, doctors, and in general representatives in all those walks of life affected ad contributing to the information revolution.

Indexing: This series is indexed in Scopus, Ei Compendex, and zbMATH.

About EAI

EAI is a grassroots member organization initiated through cooperation between businesses, public, private and government organizations to address the global challenges of Europe's future competitiveness and link the European Research community with its counterparts around the globe. EAI reaches out to hundreds of thousands of individual subscribers on all continents and collaborates with an institutional member base including Fortune 500 companies, government organizations, and educational institutions, provide a free research and innovation platform.

Through its open free membership model EAI promotes a new research and innovation culture based on collaboration, connectivity and recognition of excellence by community.

More information about this series at http://www.springer.com/series/15427

Honghao Gao • Jung Yoon Kim • Walayat Hussain
Muddesar Iqbal • Yucong Duan
Editors

Intelligent Processing Practices and Tools for E-Commerce Data, Information, and Knowledge

Springer

EAI
RESEARCH MEETS INNOVATION

Editors
Honghao Gao
Room D319
Shanghai University
Shanghai, China

Walayat Hussain
School of Information, Systems
and Modelling, Faculty of Engineering
and IT
University of Technology Sydney
Ultimo, NSW, Australia

Yucong Duan
Computer Science
Hainan University
Haikou, Hainan, China

Jung Yoon Kim
Gachon University
Seongnam-si, Republic of Korea

Muddesar Iqbal
Computer Science and Informatics Division
School of Engineering
London South Bank University
London, UK

ISSN 2522-8595 ISSN 2522-8609 (electronic)
EAI/Springer Innovations in Communication and Computing
ISBN 978-3-030-78302-0 ISBN 978-3-030-78303-7 (eBook)
https://doi.org/10.1007/978-3-030-78303-7

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 rapid development of 5G and mobile devices, there are many new e-commerce systems used in our daily life, such as wireless payment and recommender systems. Along with this, a large number of Data, Information, and Knowledge have emerged, with features of complexity, diversity, and crossover. How to use Intelligent Processing to handle E-Commerce Data, Information, and Knowledge is a new opportunity and challenge. AI for e-commerce, such as machine learning and deep learning, can be used in information modeling, data mining, knowledge description, and system verification.

The book, including six chapters made up of contributions from the academic and industry, aims to present advances in artificial intelligence to collect, process, and mining Data, Information, and Knowledge.

The first chapter, titled Distant Supervision for E-Commerce Query Segmentation via Attention Network, introduces that the booming online e-commerce platforms demand highly accurate approaches to segment queries that carry the product requirements of consumers. And the supervised methods, especially those based on deep learning, are attractive for achieving better performance on the problem of query segmentation. However, the lack of labeled data is still a big challenge for training a deep segmentation network, and the problem of Out-of-Vocabulary (OoV) also adversely impacts the performance of query segmentation. To deal with the two challenges, the authors employ the idea of distant supervision and design a novel method to find contexts in external documents and extract features from these contexts. They propose a BiLSTM-CRF-based model with an attention module to encode external features, such that external context information can be utilized naturally and effectively to help query segmentation.

The second chapter, titled Volunteer Task Recommender in Humanitarian Supply Chain for Effective Disaster Management, discusses that due to heterogeneous profiles and versatile experience of volunteers offering services for relief operations, different challenges are faced in coordinating relief activities. Moreover, prioritization of Humanitarian Supply Chain (HSC) activities to disaster damages is another concern for organizations. Lastly, while carrying out relief operations in certain calamities, HSC task recommendation to volunteers is also a significant

problem. The authors propose an optimized volunteer task recommender based on the Systems Dynamics (SD) approach that improves the productivity of teams participating in relief operations. And they also consider a few parameters by the recommender to assess the expertise of the workforce. The results are promising enough with optimized task recommendations to resources in effective disaster management with potential for application in real-time situations.

The third chapter, titled Purpose Computation Oriented Modeling and Transformation on DIKW Architecture, points out that the DIKW models are increasingly acknowledged as an important approach to address the problems related to semantic understanding beyond various question and answering systems. However, there continues to be no unified understanding of the meaning of the DIKW concepts. Data, Information, Knowledge, and Wisdom as a whole concept of DIKW are also missing a cohesive understanding of the relationships among them. Thus, the third chapter, titled Purpose Computation Oriented Modeling and Transformation on DIKW Architecture, explores the internal relationships among resources, and constructs the resource processing model, and illustrates the specific transformation process of D_{DIK}, I_{DIK}, and P_{DIK} with examples and diagrams.

The fourth chapter, titled Toward a Blockchain-Based Rural Supply Chain Management Platform for Targeted Poverty Alleviation in China, is about targeted poverty alleviation and supply chain management. In China, targeted poverty alleviation is a very important objective. By supporting farmers in producing their goods, most of the poor individuals have been lifted out of poverty. However, there are still some challenges that must be faced. The authors present insights into rural supply chain management development. A generalized blockchain-based supply chain management platform for targeted poverty alleviation is proposed. By taking advantage of blockchain, this platform can effectively help establish trust among participants, turn the supply chain into a trusted supply chain, and enhance the sustainability of poverty alleviation.

The fifth chapter, titled Centralised Quality of Experience and Service Framework Using PROMETHEE-II for Cloud Provider Selection, presents that all stakeholders need a centralized Quality of Experience (QoE) and Quality of Service (QoS) repository for enabling a sustainable trusted relationship, as well as forming practicable successful service level agreements (SLAs). Due to the elastic nature of a cloud and lack of proper resource management, the service provider is usually caught in service violation, leading to violation penalties both in terms of trust and money. Existing literature lacks studies on a centralized repository to assist cloud providers in resource management and cloud consumer service selection. To address the issue, the authors discuss the idea of a Centralised Quality of Experience and Service (CQoES) repository framework. The approach uses the PROMETHEE-II method where each alternative is assessed based on the consumer's custom weighted QoS attributes. The framework ensures the cloud marketplace's economic growth and helps the interacting parties build a durable and long-term trusted relationship.

The sixth chapter, titled Analysis of e-Consumer Behavior During the COVID-19 Pandemic, introduces the results of the initial phase of a study of changes in consumer behavior caused by the COVID-19 pandemic. The study aims to examine

specific changes in B2C interactions of Czech and Slovak consumers during the first lockdown in 2020. The starting point for changing consumer preferences is the fact that the dominant part of consumer interactions shifted from brick-and-mortar to virtual environments, where e-commerce is a safe alternative to traditional forms of trading. The results suggest that both the supply and demand sides of the market can adapt to the nonstandard situation in a relatively short time. From the point of view of customer behavior, an increase in B2C interactions is recorded in both monitored markets. The dominant part of the interactions shifts to the time of the standard working week; the weekend decline in interactions is significantly below the average on both sides of the market.

The topic of Intelligent Processing Practices and Tools of E-Commerce Data, Information and Knowledge is still a hot and emerging research. Thus, we wish these chapters can inspire blooming studies on the related topics of E-Commerce Data, Information, and Knowledge.

Shanghai, China — Honghao Gao

Seongnam-si, Korea — Jung Yoon Kim

Sydney, NSW, Australia — Walayat Hussain

London, UK — Muddesar Iqbal

Haikou, China — Yucong Duan

Acknowledgments

We are thankful to our reviewers for their effort in reviewing the chapters. We also thank the Editor-in-Chief, Dr. Imrich Chlamtac, for his supportive guidance during the entire process. Thank you for the support from Eliska at EAI Office.

Contents

Part I
Intelligent Processing Practices and Tools for E-Commerce Data, Information, and Knowledge

Distant Supervision for E-commerce Query Segmentation via Attention Network

Zhao Li, Donghui Ding, Pengcheng Zou, Yu Gong, Xi Chen, Ji Zhang, Jianliang Gao, Youxi Wu, and Yucong Duan

1 Introduction

Query segmentation is an important task in information retrieval (IR). A query is a sequence of words (in English) or characters (in Chinese) which carries the information requirement of a user. Query segmentation task is to cut a query into several continuous subsequences called *segments* that are normally frequently used phases. Compared to the independent words or characters in the query, these meaningful segments are more significant to the search engine. For instance, a user types a "short sleeve long dress" query in the search engine, where "short sleeve" and "long dress" are the two segments which indicate a long dress with a short sleeve. If the query is processed based on independent words, many irrelevant short dresses or clothes with long sleeve may be returned. The quality of query segmentation is very important for the downstream IR task.

Z. Li · D. Ding (✉) · P. Zou · Y. Gong · X. Chen
Alibaba Group, Hangzhou, China
e-mail: lizhao.lz@alibaba-inc.com; donghui.ddh@alibaba-inc.com; xuanwei.zpc@alibaba-inc.com; gongyu.gy@alibaba-inc.com; gongda.cx@taobao.com

J. Zhang
The University of Southern Queensland, Toowoomba, QLD, Australia
e-mail: Ji.Zhang@usq.edu.au

J. Gao
Central South University, Changsha, China
e-mail: gaojianliang@csu.edu.cn

Y. Wu
Hebei University of Technology, Tianjin, China

Y. Duan
Hainan University, Haikou, China

H. Gao et al. (eds.), *Intelligent Processing Practices and Tools for E-Commerce Data, Information, and Knowledge*, EAI/Springer Innovations in Communication and Computing, https://doi.org/10.1007/978-3-030-78303-7_1

Query segmentation task has been studied extensively in research community. The existing methods can be mainly divided into three categories: unsupervised [6, 7, 18, 21, 23, 28], feature-based supervised [5, 35], and deep learning methods [13, 14]. Unsupervised methods score each segmentation combination of a query by some kinds of statistical indexes like mutual information [23]. Feature-based and deep learning methods are supervised and rely on a large number of gold segmented queries.

Supervised methods, especially deep learning, are attractive for achieving a better performance and are the focus in our work. However, a lack of labeled data is one of the big challenges for training deep neural networks. In this work, we employ the idea of long-distance supervision [17] to automatically create large amounts of gold standard data. In the e-commerce field, since queries are related to products, we build a dictionary by crawling brand names, product names, attribute names, and attribute values from the product detail pages of online shopping platforms. Then, a simple max-matching algorithm can be used to segment queries by matching subsequences in queries with the words in the dictionary.

Another challenge of the query segmentation task is the so-called out-of-vocabulary (OOV) segments [10, 31]. OOV segments are those segments that are used in the test queries but do not appear in the training queries. OOV issue has been studied widely in Chinese Word Segmentation (CWS) tasks [9, 12, 34, 37] that are very similar to query segmentation. The authors of [9] argue that OOV is the key problem of the CWS task. We indicate that OOV also impacts the performance of query segmentation to a large extent. So that we try to alleviate OOV issue by incorporating contexts from external documents. If the OOV segments can be found in external documents, we can extract some valuable features to help recognize them.

We treat query segmentation as a sequence labeling problem. The tagging scheme consists of "B" and "I." "B" (Begin) means the current character is the head of its segment, while "I" (Inner) means the current character belongs to the previous segment. An example is shown in Fig. 1. Under this tagging scheme, the target is to predict one label of "B" and "I" for each character in a query. Our approach contains two steps. The first step is to find the contexts for each character in queries and extract features from these contexts. The second step is to train our neural networks model by using these extracted features. As for each character in queries, we can obtain its left and right bi-grams. For example, the left and right bi-grams of the character "衣" in the query "高腰连衣裙白色" are "连衣" and "衣裙", respectively.

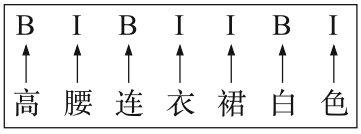

Fig. 1 The segmentation result is "高腰 (high-waisted) / 连衣裙 (dress) / 白色 (white)". Therefore, the label sequence of this query is "B/I/B/I/I/B/I"