

Guest editorial: learning from uncertainty and its application to intelligent systems of web information

Xi-Zhao Wang · Hui Wang

Received: 28 August 2013 /

Accepted: 3 September 2013 / Published online: 24 September 2013

© Springer Science+Business Media New York 2013

With the advancement information technologies, the World Wide Web (WWW) is becoming the largest data repository ever created by mankind. Due to the nature of the Web, data on the Web are often uncertain, ambiguous, noisy, incomplete, heterogeneous and dynamic. As a result, machine learning from such data faces uncertain learning environments. How to efficiently make use of large amounts of uncertain data from the Web is becoming a great challenge to the research community, which is attracting more and more researchers to explore possible solutions.

The integration of machine learning and web-related technologies results in a new subject of study, intelligent Web Information System (WIS), with the goal of creating next generation intelligent web services. There is an obvious difference between learning from web information with and without uncertainty. If the uncertainties existing in data representation and learning process cannot be well formulated and properly handled, the performance of WIS may be greatly degraded. How to fully utilize such uncertainties in terms of their representation and processing in order to build a robust, intelligent WIS with good generalization ability is a new challenge for the research community.

This special issue is devoted to the latest progress and potential applications of WIS with uncertain learning environments. There are a total of 33 submissions. After a strict peer review process, nine papers are accepted for publication. These accepted papers cover a wide range of topics including event extraction, user authority ranking in expert finding, knowledge mining based on web data, information retrieval, and extreme learning machine. These papers are categorized and briefly described below.

- (1) Four papers discuss the problem of automatic knowledge mining based on web data. Event extraction is a challenging task in natural language understanding. To make event type recognition and event argument recognition helping each other, and to resolve the problem of unbalanced data, the paper by Ruifang He et al. considers event extraction as a sequence-labeling problem and builds an improved conditional-random-fields joint labeling model with multi-trigger embedding. The problem of expert-

X.-Z. Wang (✉)

College of Mathematics and Computer Science, Hebei University, Baoding, China
e-mail: xizhaowang@ieec.org

H. Wang

School of Computing and Mathematics, University of Ulster, Jordanstown, UK
e-mail: h.wang@ulster.ac.uk

finding which focuses on identifying experts with special skills or domain knowledge has become increasingly important with the popularity of knowledge sharing social network. In the paper by Mahmood Neshati et al. the integration of scientific-social networks, e.g. the DBLP publication network and the Twitter social network, is proposed to realize a multi environment expert finding system. The paper by Hengshu Zhu et al. presents an expert-finding framework based on the authority information in the target category as well as in the relevant categories. Along this line, a scalable method for measuring the relevance between categories through topic models, and a topical link analysis approach for ranking user authority are proposed to complete this task. Furthermore, the problems of data records extraction from semi-structured documents on the Deep Web are studied by Wachirawut Thamviset et al.

- (2) Information retrieval is one of the most feasible ways for users to efficiently obtain specific knowledge from the massive data repository on the Web. The efficiency, precision, recall and usability are important indices for evaluating an information retrieval system. In the paper by Rani Qumsiyeh et al., a web query suggestion approach WebQS, which provides the same mechanism offered by Google, Yahoo and Bing to support users in formulating keyword queries, is constructed to improve the precision and recall of search results. Meanwhile, the problem of finding similar queries is studied based on query representation analysis in the paper by Yuan Wang et al.
- (3) Extreme Learning Machine (ELM) is a new learning technique for single layer feed-forward network. Due to the extremely fast learning speed and excellent generalization capability, ELM has been extensively studied in recent years. This special issue includes three papers which focus on ELM or its applications in intelligent WIS. For example, the proposed distributed ELM with MapReduce in the paper by Junchang Xin et al. presents a possible solution for dealing with massive data on the Web. The TLS-ELM given in the paper by Jianwei Zhao, which is a learning algorithm based on ELM and total least squares (TLS) to approximate the samples with noises both in input and in output variables, provides a candidate method ideal for dealing with noisy data on the Web. Finally, the paper by Xiangguo Zhao et al. is devoted to the application of ELM to multiclass XML documents classification.

As guest editors, we hope that the diverse range of papers included in this issue of the World Wide Web Journal are helpful for readers to be informed of the advances in Learning in Uncertain Environments and Intelligent Web Information Systems. We also hope that the papers in this issue will stimulate further progress in this direction.

As guest editors, we wish to thank all contributors for submitting their works to this special issue, and all reviewers for their effort in critically reviewing the papers within limited time. We also sincerely thank the Editor-in-Chiefs, Prof. Yan-Chun Zhang and Prof. Marek Rusinkiewicz, for their support to edit this issue, and Ms Jennylyn Rosiento for her kind assistance in handling submissions and the publication process, as well as our colleagues and friends in ELM2012, ICMLC2012 for their preliminary works.

Acknowledgments The first guest editor, Prof. Xi-Zhao Wang, would like to acknowledge the support from projects NSFC (61170040 and 60903089) in editing this special issue.