Representing Scientific Knowledge: The Role of Uncertainty

Book · November 2017

DOI: 10.1007/978-3-319-62543-0

2 authors:

Chaomei Chen
College of Computing and Informatics, Drexel University

Min Song
Yonsei University

Representing Scientific Knowledge

Chaomei Chen · Min Song

Representing Scientific Knowledge

The Role of Uncertainty

Springer

Chaomei Chen
College of Computing and Informatics
Drexel University
Philadelphia, PA
USA

Min Song
Department of Library and Information Science
Yonsei University
Seoul
South Korea

ISBN 978-3-319-62541-6 ISBN 978-3-319-62543-0 (eBook)
https://doi.org/10.1007/978-3-319-62543-0

Library of Congress Control Number: 2017957689

Printed on acid-free paper

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

Preface

The 2014 Ebola outbreak in West Africa raised many urgent concerns about public health and safety, as well as legal and administrative implications. The high mortality rate of the Ebola virus heightened the tension between the public, healthcare providers, patients, and local authorities. In the United States, the White House expressed concerns about possible unintended consequences of quarantine policies enforced on doctors and nurses returning from Ebola-stricken countries. Governors of some states defended their quarantine policies, while the White House worried that the policies might not be grounded in science. Some contractors were deeply concerned about the safety of handling Ebola patients' medical wastes and whether they should discard expensive instruments just because they were used to analyze Ebola patients' blood. Some people firmly believed that people without symptoms of Ebola would not transmit the disease. However, the Center for Disease Control and Prevention (CDC) revising its own guidelines was enough reason for others to take extra prudent measures to minimize the risk.

Charles Haas, an environmental engineering professor at Drexel University, specializes in water treatment and risk assessment. He started his comprehensive search in the literature for any information on how long the Ebola virus might be able to survive in water. He did not find a clear answer in the literature. Instead, he found reports of nonzero probabilities of infection after 21 days, which was the basis for the recommended 21-day quarantine. Similarly, a group of researchers did a deep search in the literature but did not find a clear picture either. The implications of these findings on public health policies, public understanding of science, and information science are striking.

Semantic MEDLINE is a great resource for developing a good understanding of scientific knowledge in terms of semantic predications as well as their original unstructured texts. The complexity of scientific writing is strikingly high. It is common to see long and complex sentences. Studying semantic predications and the contexts in which they appear has revealed how frequently uncertainties go hand in hand with the very knowledge one aims to achieve. Knowledge that is free from uncertainty probably has no value in a research field. Understanding the epistemic status of scientific knowledge is so important that we want to claim that expertise is the knowledge of uncertainty!

The profound and integral role of uncertainty in science, especially in research fronts of a scientific field, has become the core interest of our research. In April and December 2016, two workshops in association with the National Center for Science and Engineering Statistics (NCSES) of the NSF catalyzed the focus on uncertainty further. At the April workshop with the NCSES, Chen presented some of the initial ideas and preliminary results of uncertainties associated with scientific publications in a white paper on the fidelity of visualizing scientific uncertainty.

The preparation and launch of a new open access journal, *Frontiers in Research Metrics and Analytics* (RMA), in midsummer of 2016 provided another boost to the idea. While many have pointed out the shortcomings of overly relying on simplistic and often single metrics of research productivity and quality, evaluators and policymakers are currently limited to only a few options. As a result, it is difficult to compensate the lack of semantic, diagnostic, and analytic reasoning due to over-simplifications of scientific inquiry as a complex adaptive system. The mission of RMA is therefore to bridge the currently loosely coupled research communities. The theme of improving the clarity of the epistemic status of science emerges again in the five grand challenges for accessing and communicating scientific knowledge more efficiently and effectively.

The idea of creating a Visual Analytic Observatory of Scientific Knowledge (VAO) becomes a unifying framework to stimulate and accommodate tools, resources, and applications toward meeting the five grand challenges and beyond. The research project led by Chen is supported by the NSF Science of Science and Innovation Policy (SciSIP) program (Award Number 1633286). The VAO aims to enable researchers to find the epistemic status of scientific knowledge and its provenance of evolution efficiently and effectively. With the worldwide user community of our CiteSpace tool, we believe that the VAO will substantially advance the state of the art. This book introduces the theoretical foundations of how scientific fields develop, which the reader can then use as a referential framework to guide subsequent explorations of scientific knowledge. We also introduce science mapping tools and demonstrate how these tools can help us develop a better understanding of the history and the state of the art of a scientific domain. More importantly, we want to share our methods and principles, both theoretical and practical, with our reader so that we can empower ourselves with computational techniques and analytic reasoning. In particular, creativity comes from competing, contradictory, and controversial views. Reconciliations of existing discrepancies may lead to creative solutions at a higher level. We hope our reader can benefit from the analytic and methodological value of the materials presented in the book.

Chen spent his sabbatical leave at Yonsei University in the Spring semester of 2017 and taught two courses on Yonsei University's beautiful campus on visual exploration of scientific literature. Students from these two classes eagerly and diligently explored and applied the science mapping tools we introduced in this book, namely, CiteSpace, VOSviewer, and CitNetExplorer.

In our previous work, we emphasized the pitfalls and biases of mental models in our reasoning and decision-making. In this book, we aim to demonstrate that uncertainty plays a fundamental role in representing and communicating scientific knowledge.

We are truly grateful for the encouragement and support from many people at various stages of our research and the production of the book. Chen would like to take this opportunity to thank our coauthor Min Song, researchers in his Text and Social Media Mining Lab (TSMM), and students at Yonsei University for collaborative research and the hospitality during Chen's sabbatical in Seoul. Chen is also grateful to Jiangen He and Qing Ping as graduate research assistants at Drexel University, Sergei V. Kalinin at Oak Range National Laboratory for exploring applications of science mapping in material sciences and for organizing a tutorial in Boston, Maryann Feldman for encouragements and guidance, Gali Halevi, Henk Moed, and Mike Taylor for their valuable contributions toward research on tracking emerging trends, Caroline S. Wagner for organizing one of the workshops with NCSES and serving as a guest editor for a Research Topic with RMA, and Jie Li at Shanghai Maritime University for his extensive efforts in disseminating science mapping tools in China. We would like to say thank you to Beverley Ford at Springer for her initiative, encouragement, persistence, and patience.

As always, to the members of Chen's loving family, Baohuan Zhang, Calvin Chen, and Steven Chen, thank you for everything.

Acknowledgements

Chaomei Chen wishes to acknowledge the support of the NSF SciSIP Program (Award Number 1633286) and industrial sponsorship in the past from Elsevier and IMS Health.

Min Song acknowledges the support of the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2015S1A3A2046711).

Philadelphia, USA — Chaomei Chen
Seoul, South Korea — Min Song
August 2017

Endorsement

Chaomei Chen and Min Song have written an important book that opens up a new area in the study in scientometrics and informetrics as well as information visualization, namely the study and measurement of uncertainty of scientific knowledge and how uncertainty is expressed in scientific texts. At the same time the book is a tutorial and review of relevant methods in natural language processing and gives step by step instructions on how they can be implemented. What I like most about the book, however, is how it integrates this new approach with existing theories in the history and sociology of science. In my view, uncertainty is key to understanding the development of scientific knowledge.

Henry Small
Senior Scientist
SciTech Strategies

Contents

About the Authors

Chaomei Chen is Professor in the College of Computing and Informatics at Drexel University and Professor in the Department of Library and Information Science at Yonsei University. He is the Editor in Chief of Information Visualization and Chief Specialty Editor of Frontiers in Research Metrics and Analytics. His research interests include mapping scientific frontiers, information visualization, visual analytics, and scientometrics. He has designed and developed the widely used CiteSpace visual analytic tool for analyzing patterns and trends in scientific literature. He is the author of several books such as Mapping Scientific Frontiers (Springer), Turning Points (Springer), and The Fitness of Information (Wiley).

Min Song is Underwood Distinguished Professor at Yonsei University. He has extensive experience in research and teaching in text mining and big data analytics at both undergraduate and graduate levels. Min has a particular interest in literature-based knowledge discovery in biomedical domains and its extensions to a broader context such as the social media. He is also interested in developing open source text mining software in Java, notably creating the PKDE4J system to support entity and relation extraction for public knowledge discovery.

List of Figures

List of Tables