Information Processing and Management 000 (2016) 1–29

Contents lists available at ScienceDirect

Information Processing and Management

journal homepage: www.elsevier.com/locate/ipm

Analytical mapping of opinion mining and sentiment analysis research during 2000–2015

R. Piryani[a], D. Madhavi[b], V.K. Singh[c,*]

[a] Department of Computer Science, South Asian University, New Delhi, India
[b] Department of Computer Science & Engineering, APJAKTU, Lucknow, India
[c] Department of Computer Science, Banaras Hindu University, Varanasi, India

ARTICLE INFO

Article history:
Received 13 August 2015
Revised 1 May 2016
Accepted 5 July 2016
Available online xxx

Keywords:
Affective computing
Opinion mining
Scientometrics
Sentiment analysis

ABSTRACT

The new transformed read-write Web has resulted in a rapid growth of user generated content on the Web resulting into a huge volume of unstructured data. A substantial part of this data is unstructured text such as reviews and blogs. Opinion mining and sentiment analysis (OMSA) as a research discipline has emerged during last 15 years and provides a methodology to computationally process the unstructured data mainly to extract opinions and identify their sentiments. The relatively new but fast growing research discipline has changed a lot during these years. This paper presents a scientometric analysis of research work done on OMSA during 2000–2016. For the scientometric mapping, research publications indexed in Web of Science (WoS) database are used as input data. The publication data is analyzed computationally to identify year-wise publication pattern, rate of growth of publications, types of authorship of papers on OMSA, collaboration patterns in publications on OMSA, most productive countries, institutions, journals and authors, citation patterns and an year-wise citation reference network, and theme density plots and keyword bursts in OMSA publications during the period. A somewhat detailed manual analysis of the data is also performed to identify popular approaches (machine learning and lexicon-based) used in these publications, levels (document, sentence or aspect-level) of sentiment analysis work done and major application areas of OMSA. The paper presents a detailed analytical mapping of OMSA research work and charts the progress of discipline on various useful parameters.

1. Introduction

OMSA is a natural language processing task that uses an algorithmic formulation to identify opinionated content and categorize it as having 'positive', 'negative' or 'neutral' polarity. "What other people think" has always been an important piece of information for most of us during the decision-making process (Pang & Lee, 2008). Opinions of users not only help individuals in taking informed decisions but also help organizations in identifying customer attitudes/ opinions about products/ services. The new user-centric, participative Web allows extremely large number of users to express themselves about virtually endless topics ranging from reviews about movies, products, services to different socio-political events. However, the immense volume of data available on the Web (including various social media platforms) becomes information overload

* Corresponding author.
E-mail addresses: rajesh.piryani@gmail.com (R. Piryani), madhavidevaraj@gmail.com (D. Madhavi), vivekks12@gmail.com (V.K. Singh).

http://dx.doi.org/10.1016/j.ipm.2016.07.001
0306-4573/

Table 1
Details of dataset.

Source/index	Category	Time period	Query to extract data	No. of papers retrieved	Total no. of fields in each publication record	Date of download
Web of Science	Articles, reviews, proceeding papers, editorial material and book chapters	2000–2015	(TS = ((“Sentiment Analysis”) OR (“Sentiment Classification”) OR (“Opinion Mining”) OR (“Opinion Classification”) OR (“Affect Analysis”) OR (“Affective Computing”) OR (“Sentiwordnet”) OR (“Sentic”) OR (“mining sentiment”) OR (“mining sentiments”))) *AND* LANGUAGE: (English)	697	60	27.02.16

in absence of automated methods to extract relevant and comprehensive information. OMSA fills this gap by identifying opinionated content and producing opinion summaries. It has been this major reason that research work on OMSA has grown tremendously during the recent past.

In this paper, we present a scientometric mapping exercise to analyze and chart the progress of research work in OMSA. The primary motivation of our work has been to understand the trajectory of research work done on OMSA from the period of inception till now. We have used both computational and manual analysis for this purpose. The research publication data obtained from Web of Science (WoS) database is analyzed computationally to identify year-wise number and rate of growth of publications, types of authorship of papers on OMSA, collaboration patterns in publications on OMSA, most productive countries, institutions, journals and authors, citation patterns and an year-wise citation reference network, and theme density plots and keyword bursts in OMSA publications during the period. Thereafter a somewhat detailed manual analysis of the research publication data is performed to identify popular approaches (machine learning and lexicon-based) used in these publications, levels (document, sentence or aspect-level) of sentiment analysis work done and major application areas of OMSA. This analysis is aimed to provide an analytical account of progress of the discipline from its inception to state of the art today, major milestones in the journey, the disciplines that OMSA research has drawn inspiration from and the areas it has been applied, major approaches and methods used in the OMSA research, and a meme map of major concepts and keywords in the area. More precisely, our analytical mapping can answer research questions of the following types:

- What is the period of origin of OMSA research publications and how research work on OMSA has grown over time?
- In which countries and institutions most of the initial and subsequent research work on OMSA has been done?
- What are the top publication sources (journals) publishing research on OMSA?
- Who are most productive and most cited authors in OMSA research during the period under study?
- What is the amount of international collaboration in OMSA research?
- What kind of authorship patterns are observed in OMSA research output?
- What are the major concepts occurring in OMSA research publications and what kind of theme density plot is observed in OMSA research output?
- What are the main approaches and methods of OMSA and which of them is used in what proportion of the reported research output?
- What are the main data sources on which OMSA work is done?
- What are main application areas of OMSA research?

The paper tries to answer the questions of the type mentioned above. Knowing answer to these questions may be very useful for an understanding the origin and growth of research work in OMSA. It will help in charting the course of development of the discipline and analyze different aspects of OMSA research. The readers can trace the broader landscape of OMSA research filed and obtain a highly useful overview and understanding of the research discipline, from its origin to the current state of the art. To the best of our knowledge this work is first of its kind and is different from regular survey papers on OMSA in many respects. The rest of the paper is organized as follows: Section 2 describes the data collection and analytical methodology used. Section 3 presents analytical outcomes of the scientometric mapping of OMSA research. Section 4 presents a detailed/ manual analysis of OMSA approaches and levels, major data sources and application areas. The paper concludes in Section 5, with a short summary of the work and its usefulness.

2. Data and methodology

We have obtained research publications indexed in WoS on OMSA for a considerably large period of 16 years (2000–2015), which almost covers the entire period of origin and growth of computational OMSA research. The WoS database collection indexes documents of different types namely articles, reviews, proceeding paper, editorial material, book review etc., in various languages. We have downloaded data for articles of all types on OMSA written in English. Table 1 illustrates the query used and statistics of the data downloaded.

Please cite this article as: R. Piryani et al., Analytical mapping of opinion mining and sentiment analysis research during 2000–2015, Information Processing and Management (2016), http://dx.doi.org/10.1016/j.ipm.2016.07.001