Contextualized Meaning Extraction: A Meta-Algorithm for Big Data Text Mining with Pragmatics

Jeffrey D. Wall, Michigan Technological University, Houghton, MI, USA
Rahul Singh, The University of North Carolina at Greensboro, Greensboro, NC, USA

ABSTRACT

Text mining is a powerful form of business intelligence that is used increasingly to inform organizational decisions. Current text mining algorithms rely heavily on the lexical, syntactic, structural, and semantic features of text to extract meaning and insight for decision making. Although semantic analysis is a useful approach to meaning extraction, pragmatics suggests that a more accurate meaning of text can be extracted by examining the context in which the text is recorded. Given that massive amounts of textual data can be drawn from multiple and diverse sources, accounting for context is increasingly important. A conceptual model is provided to explain how concepts from pragmatics can improve existing text mining algorithms to provide more accurate information for decision making. Reversing the pragmatic process of meaning expression could lead to improved text mining algorithms. The theoretical process model developed herein can provide insight into the development and refinement of text mining algorithms that draw from diverse sources.

KEYWORDS

Big Data, Business Intelligence, Context, Pragmatics, Sentiment Analysis

INTRODUCTION

For nearly a decade, managers have ranked organizational and business intelligence (BI) as one of the most important emerging technologies for organizations (Luftman & Ben-Zvi, 2011). Increasingly, organizations use BI tools to extract information from textual expressions to inform organizational decisions (Harrysson, Metayer, & Sarrazin, 2012; Hira, 2005). A textual expression refers to a statement made by an individual that is recorded in a digital, textual format. Advancements in Big Data technology allow for the rapid, real-time collection of massive amounts of textual expressions from diverse sources (e.g., email, social media, and blogs) (Chang, 2017; Watson, 2014). Information mined from textual expressions can inform organizational activities such as product and brand development, competitive benchmarking, and impression management (Harrysson et al., 2012; Lee & Bradlow, 2011). To accommodate accurate decision making, text mining algorithms must capture accurate meaning to prevent misinformation from entering organizational decision making processes. However, extracting accurate meaning from textual expressions through text mining is difficult and not well understood (Cambria, 2016; Harrysson et al., 2012; Reyes & Rosso, 2012; Rosman, 2012).
Organizations employ text mining tools to extract various forms of information from textual expressions, including: topics, events, opinions, emotions, styles, genres, vernaculars, and interactions (Abbasi & Chen, 2008). Despite the wide range of information that can be extracted from textual expressions, the current pattern of analysis used to extract different types of information from text is mostly similar. Although multiple text mining algorithms exist, most forms of textual analysis rely on lexical, syntactic, structural, and/or semantic features of the text itself to extract meaning from an expression (Abbasi & Chen, 2008). Even many new “contextualized” text mining algorithms rely only on the text itself for contextual information (Saif, He, Fernandez, & Alani, 2016). These traditional text mining approaches assume that the meaning of a textual expression can be derived solely from the text itself.

Pragmatics, however, suggests that meaning cannot be derived directly from text itself. Pragmatics is the study of the meaning of an expression in the context in which it is uttered. Pragmatics is concerned with understanding the knowledge, beliefs, expectations, and intentions of the speaker and hearer, and other contextual factors that inform an understanding of the meaning of an expression (Grice, 1957; Haugh, 2013). Pragmatics assumes that individuals communicate differently in different situations. Because organizations employ Big Data collection methods that draw different data types from multiple and diverse sources (Chang, 2017), accounting for context is increasingly important. Pragmatics has been adopted in IS research in areas such as the Pragmatic Web. However, pragmatics is not yet widely adopted across text mining approaches, such as sentiment analysis. Pragmatics may be the next stage in the evolution of textual analysis and deserves increased attention (Cambria, 2016). The failure to meaningfully account for context in textual analysis is an oversight that deserves careful thought and attention.

This paper seeks to improve text mining algorithms, particularly those concerned with the extraction of sentiment, opinions, and emotions. Sentiment analysis, also called opinion mining, is a common form of text mining that is used to analyze product reviews or to examine attitudes toward a particular company, product, or brand (Kennedy, 2012). Sentiment analysis algorithms attempt to identify the polarity (i.e., positive or negative) of attitudes and emotions expressed by individuals toward an object or features of an object (Duric & Song, 2012; Pang, Lee, & Vaithyanathan, 2002; Yang & Chao, 2014). Further, sentiment analysis algorithms attempt to capture the strength of the attitude or emotion expressed toward the object (e.g., tolerate vs. like vs. love) (Pang et al., 2002; Saif et al., 2016; Wilson, Wiebe, & Hwa, 2004).

By employing a pragmatics perspective, this paper identifies the process by which individuals express intended meaning. We argue that sentiment analysis algorithms must reverse this process to extract more accurate meaning from textual expressions. The process model is founded on literature from semantics, pragmatics, discourse analysis, communication, and other fields. Theory development is often neglected in text mining research in favor of algorithm development. Yet theory development is strongly needed in IS research to provide future guidance for bodies of research (Rivard, 2014). Our conceptual process model acts as a meta-algorithm, an algorithm that guides the use or development of other algorithms. The meta-algorithm can be implemented by researchers in a number of ways.

The model herein contributes several important insights to the research and practice of gathering organizational intelligence from textual expressions. By drawing from pragmatics, discourse analysis, and communication research, the model shows that text mining algorithms cannot rely solely on the structure of text to derive accurate meaning from textual expressions. Understanding the context in which an expression is uttered in addition to the linguistic structure of the expression is crucial to extracting accurate sentiment information from text. Pragmatics may be particularly important to consider when extracting meaning from expressions of sentiment, belief, and emotion (Kecskes, 2008). Thus, our choice to explore sentiment analysis is well grounded.

The remainder of this paper continues as follows. First, a process model is proposed to explain the way in which individuals transform intended meaning into linguistic expression, resulting in a process model consisting of three stages. Second, it is argued that text mining algorithms should be developed.
Social Media Intelligence for Business
www.igi-global.com/article/social-media-intelligence-for-business/162707?camid=4v1a

A Discrete Artificial Bees Colony Inspired Biclustering Algorithm
www.igi-global.com/article/discrete-artificial-bees-colony-inspired/67536?camid=4v1a