Decision Support Systems 54 (2013) 1010–1015

Contents lists available at SciVerse ScienceDirect

Decision Support Systems

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

Sentiment analysis in decision sciences research: An illustration to IT governance

Qing Cao, Mark A. Thompson *, Yang Yu

Area of Information Systems and Quantitative Sciences, Rawls College of Business, Texas Tech University, Lubbock, TX 79409, United States

ARTICLE INFO

Article history:
Received 11 December 2011
Received in revised form 31 August 2012
Accepted 7 October 2012
Available online 14 October 2012

Keywords:
Sentiment analysis
Text mining
IT governance

ABSTRACT

We propose a viable procedure to mine opinions from a large collection of texts. We demonstrate how this procedure can be applied in business research by constructing IT governance measures. Specifically, we generate a set of IT governance measures using a multi-label classification method and compare against our proposed sentiment analysis procedure. Using 10-K forms to develop our measures of IT governance, we examine the role of IT governance on firm performance. We find evidence of the five dimensions of IT governance (as proposed by the IT Governance Institute), but only one dimension significantly explains firm performance (i.e., strategic alignment) using the multi-label classification method. Using our new procedure to construct IT governance measures, we find significant evidence of four dimensions in explaining firm performance.

1. Introduction

One of the biggest challenges for current decision support systems is developing effective methods and tools to extract opinion-oriented information automatically from unstructured data [28]. While the number and types of information resources continue to grow exponentially and information technologies (e.g., search engines) are being used to manage the information overload problem, business executives are still faced with the difficulty of transforming the wealth of information to knowledge for more effective use [5]. Text mining and natural language processing are traditional methods used to help people find business intelligence from free-form data; however, these methods lack strength in detecting people's opinion [16]. Sentiment analysis (SA) has evolved over the last decade from text mining and natural language processing, but aims to determine the attitude of a speaker or a writer with respect to some specific topics [21]. More recently, SA has greatly assisted decision makers in extracting opinions from unstructured human-authored documents [28]. This type of technology reduces the need to have people read dozens or even hundreds of documents to extract business opinions on a variety of topics and for different purposes.

Recent SA studies have focused on the technical details such as either machine learning or classification algorithms rather than systematic procedures [28]. In this study, we design and establish a viable procedure for processing SA in business research. We describe the major components of mining sentiment from large collections of text including data gathering, pretreatment process, topic reorganization, sentence-level sentiment analysis, and document-level sentiment measurements. In particular, we extract sentiment polarity (positive or negative) at the sentence-level and integrate to a document-level index for the degree of sentiment (strong, middle, or weak). Moreover, we apply this procedure to an area that is lacking viable quantitative measures where sentiment analysis may be able to fill this void and provide a useful metric. That is, this new procedure is used to generate a measure of IT governance based on mining decision makers' opinion on five dimensions (as described by the IT Governance Institute) from 10-K forms. As such, we provide empirical evidence of the existence of these five dimensions from the IT Governance Institute (ITGI) and find differences in business executives' emphasis on these dimensions. This study makes two contributions to the academic literature. First, we propose a new methodology (and procedure) to extract sentiment polarity and the degree of polarity from large collections of texts. Second, we demonstrate how this new procedure can be used in business research. In our case, we apply this procedure to an area that is lacking well-developed metrics. In doing so, we provide evidence in support of the five dimensions of the IT governance framework, which may offer a road map for future research in this area. The rest of the paper proceeds as follows. In the next section, we review the previous studies on SA. Next, we describe the SA framework, and then illustrate the process of using this framework on IT governance. We conclude the paper by discussing the limitations and identifying areas for future research.

2. Background on sentiment analysis

Sentiment analysis and related research has increased considerably in the past decade due to several factors [9,27,34,35]. These factors include: 1) the rise of machine learning methods in natural language processing and information retrieval; 2) the availability of datasets for machine learning algorithms to be trained on; and 3) the realization of many commercial intelligence applications that

* Corresponding author. Tel.: +1 806 834 1535.
E-mail addresses: qing.cao@ttu.edu (Q. Cao), mark.thompson@ttu.edu (M.A. Thompson), yang.yu@ttu.edu (Y. Yu).

0167-9236/$ – see front matter
http://dx.doi.org/10.1016/j.dss.2012.10.026

the area offers. Three main elements of SA include sentiment classification, feature-based opinion extraction and comparison, and relation extraction. SA differs from text mining in several ways. For example, rather than focus on topic words in “classical” text mining, sentiment signals are the main objects in SA, which analyzes the opinions of the writer directly and avoids costly and time consuming content analysis [21,28]. In addition, rather than concentrating on word frequencies in “classical” text mining, SA considers a document or a sentence as a set of features or components, which allows researchers to explore the semantic meaning of content on a much deeper level. As such, the field of SA is well-suited to various types of intelligence applications such as business and politics [20,26].

Existing SA approaches are either based on linguistic resources (dictionary/lexicon) or on machine learning. SA based on linguistic resources is centered on lists of words with pre-determined emotional weight or polarity [13,25]. The most popular sentiment dictionaries include Dictionary of Affect of Language, WordNet-Affect and the SentiWordNet [12]. Using a linguistic-based approach, an important step is to check how many times the signal word or phrase appears. However, this approach involves a number of linguistic techniques that are not always robust and are often quite labor intensive as compared with manual coding and content analysis [23]. The other approach based on machine learning relies on a computer's ability to automatically learn the language used for expressing sentiment regardless of how “good” or “normal” the language is [3,33,37]. However, the machine needs some information to learn from (called a training corpus) and the more examples the machine has to learn from the better. In the case of SA, this is a set of examples annotated by humans [15]. Once the machine has learned from the examples, it can apply the acquired knowledge to new, unseen documents and then classify them into sentiment categories. While many different algorithms have been developed and applied in SA research, we choose to highlight and compare several common algorithms used in related studies in Table 1. The last column in Table 1 shows the accuracy of different approaches, which were evaluated by different measures including the commonly used F-measure. The average performance across the different approaches is as follows: 79.9% for naïve Bayes, 81.2% for maximum entropy, 79.2% for support vector machines, 74.7% for boosting and 76.5% for K-nearest neighbor. As such, naïve Bayes and maximum entropy approaches tend to perform better than other common approaches.

Table 1
An overview of the most popular machine learning algorithms used in SA.

	Study (Year)	Data	Accuracy (%)
Naïve Bayes	Annett and Kondrak (2008) [1]	Reviews	77.5
	Bifet and Frank (2010) [3]	Microblogs	82.5
	Chen et al. (2006) [7]	Reviews	77.5
	Dave et al. (2003) [10]	Reviews	81.9–87.0
	Gindl and Liegl (2008) [14]	Reviews	66.0
	Go et al. (2009) [15]	Microblogs	82.7
	Pang et al. (2002) [29]	Reviews	81.5
	Zhang et al. (2011) [37]	Reviews	84.5
Maximum entropy	Gindl and Liegl (2008) [14]	Reviews	83.8
	Go et al. (2009) [15]	Microblogs	83
	Shimada and Endo (2008) [33]	Reviews	77.1
	Pang et al. (2002) [29]	Reviews	81.0
Support vector machines	Annett and Kondrak (2008) [1]	Reviews	77.4
	Chen et al. (2006) [7]	Reviews	84.6
	Go et al. (2009) [15]	Microblogs	82.2
	Pang et al. (2002) [29]	Reviews	82.9
Boosting	Kudo and Matsumoto (2004) [18]	Reviews	59.6–90.2
	Cassinelli and Chen (2009) [6]	Reviews	73.0–76.0
K-nearest neighbor	Davidov et al. (2010) [11]	Microblogs	66.0–87.0

3. Automated sentiment analysis procedure

The procedural flowchart of the automated sentiment analysis we employed in this study is shown in Fig. 1. In the flowchart, each dotted line box refers to a key process and the double bordered box is designed as a support system, which includes the machine learning algorithm set, training corpus (documents), domain knowledge database (e.g., Wikipedia), and sentiment word list (e.g., Cornell movie-review datasets). There are four key processes involved in sentiment analysis: pretreatment process, sub topic (dimensional) classification, sentiment classifier builder, and construct validity. The support system is applied in the sub topic classification and sentiment classifier builder processes.

In the pretreatment process, we clean up the raw corpus (i.e., set of documents) and then store it to a computable format in the refined corpus. For example, if the raw content is in an HTML format, we will change it to a computable format (e.g., text format) by removing the HTML markups. Next, we need to decide whether all the words in the raw corpus should be normalized (e.g., capitalization and typographical corrections) before storing in the refined corpus.

A sub-topic classification process is needed when documents in the refined corpus fall into multiple categories or when there are several subtopics in documents. Then, you must decide whether to separate these subtopics and rearrange the refined corpus in a specific way. During this process, domain knowledge, training corpus and machine learning algorithm set will provide decision support. Before choosing a machine learning algorithm for sub-topic (dimensional) classification or sentiment classification, the following issues need to be addressed: 1) select a specific sentiment level for the study; 2) a training set must be constructed for which the correct classifications of the objects are known; and 3) a set of object parameters must be chosen that are powerful discriminators for classification. There are a variety of techniques for supervised machine learning algorithms that have demonstrated reasonable performance for sentiment classification including naïve Bayes [23,31], maximum entropy [2,24], support vector machines [17], boosting [32], and k-nearest neighbor [36]. In practice, naïve Bayes can be used for both binary and multiclass classification problems and affords fast, highly scalable model building and scoring. Maximum entropy can combine different kinds of statistical dependencies in one unified framework and can help to avoid data sparseness problem, which is common in language related studies. Furthermore, based on our performance comparison in Table 1, we choose to use both naïve Bayes and maximum entropy approaches for sub-topic (dimensional) classification and naïve Bayes as the sentence-level sentiment classifier (see Appendix A for more details).

The third key process involves creating the sentiment classifier, which generates sentiment polarity on the unit of analysis such as the part of speech or sentence in the document. In this process, all components in the support system (machine learning algorithm set, domain knowledge, training corpus, and sentiment word list) will provide the decision support. Additionally, performance comparison or measurement integration might also be included in this step. Once the viable classifier has been constructed, construct validity measured by the accuracy of the classifier must be tested. Construct validity is necessary for both sub-topic classification and sentiment classification.

We will demonstrate the sentiment analysis procedure in the next section.

4. An illustration to IT governance

While IT governance is not a new area of research, the lack of management models or tools to assist boards of directors with implementation of IT governance has been problematic [30]. Due to the difficulties of defining and measuring IT governance, the IT Governance Institute (ITGI) provided a conceptual framework to assist in developing a better understanding of the processes of IT governance. This conceptual model