Electronic Commerce Research and Applications 14 (2015) 58–74

Contents lists available at ScienceDirect

Electronic Commerce Research and Applications

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

Visualizing market structure through online product reviews: Integrate topic modeling, TOPSIS, and multi-dimensional scaling approaches

Kun Chen [a,1], Gang Kou [b,*], Jennifer Shang [c,2], Yang Chen [b,3]

[a] School of Management and Economics, University of Electronic Science and Technology of China, Chengdu, Sichuan 610054, China
[b] School of Business Administration, Southwestern University of Finance and Economics, Chengdu, Sichuan 610074, China
[c] Katz Graduate School of Business, University of Pittsburgh, Pittsburgh, PA 15260, USA

ARTICLE INFO

Article history:
Received 28 September 2013
Received in revised form 20 November 2014
Accepted 21 November 2014
Available online 12 December 2014

Keywords:
Market structure
Text mining
Topic modeling
Ranking of products
TOPSIS

ABSTRACT

Studies have shown that perceptual maps derived from online consumer-generated data are effective for depicting market structure such as demonstrating positioning of competitive brands. However, most text mining algorithms would require manual reading to merge extracted product features with synonyms. In response, Topic modeling is introduced to group synonyms together under a topic automatically, leading to convenient and accurate evaluation of brands based on consumers' online reviews. To ensure the feasibility of employing Topic modeling in assessing competitive brands, we developed a unique and novel framework named WVAP (Weights from Valid Posterior Probability) based on Scree plot technique. WVAP can filter the noises in posterior distribution obtained from Topic modeling, and improve accuracy in brand evaluation. A case study exploring online reviews of mobile phones is conducted. We extract topics to reflect the features of the cell phones with a qualified validity. In addition to perceptual maps derived by multi-dimensional scaling (MDS) for product positioning, we also rank these products by TOPSIS (Technique for Order Performance by Similarity to Ideal Solution) so as to visualize the market structure from different perspectives. Our case study of cell phones shows that the proposed framework is effective in mining online reviews and providing insights into the competitive landscape.

1. Introduction

Online product reviews offer businesses the opportunity to economically and expediently perform in-depth and comprehensive market landscape analysis. These consumer-generated data have received increasing attention from academics and practitioners in recent years (Laroche et al. 2005; Gopal et al. 2006). Although online voices of the consumer (VOC) are in the form of free-text, they have proven to reflect essential characteristics of the market (Campbell et al. 2011; Godes and Mayzlin 2004; Duan et al. 2008; Shao 2012), and they can leverage changes of traditional marketing activities (Onishi and Manchanda 2012). Nevertheless, there remains an obstacle which prevents online product reviews from reaching its best potential. Namely, the online user-generated content is enormous and qualitative in nature, making it difficult to analyze the data quantitatively and attain meaningful knowledge therein (Godes et al. 2005). Due to lack of effective methods to extract key features of these texts, businesses were unable to obtain useful information to develop a market structure map. Yet, the understanding of these viewpoints and market structure is crucial for product development, pricing, promotion/campaign, and brand positioning. Companies thus resort to product ratings as proxies for product reviews. For example, Chintagunta et al. (2010) use product ratings to examine the relationship between consumer reviews and sales in the movie industry, while Chevalier and Mayzlin (2006) use them to study the book industry.

Market structure is the depiction of relationship among brands based on various approaches such as consideration set (Urban et al. 1984), brand-switching data (Cooper and Inoue 1996) and brand associative networks (John et al 2006). With the advance of text mining techniques based on natural language processing (NLP) (Feldman et al. 2007), researchers have begun to elicit structured and quantitative information of brands in the market from online product reviews. For instance, Lee and Bradlow (2011) developed a text-mining algorithm for online product reviews. And Netzer et al. (2012) proposed a hybrid text-mining and semantic network analysis tool for market-structure surveillance. However, based on

* Corresponding author. Tel.: +86 15982498501.
E-mail addresses: chengk@uestc.edu.cn (K. Chen), kougang@yahoo.com (G. Kou), shang@katz.pitt.edu (J. Shang), francisnju@gmail.com (Y. Chen).
[1] Tel.: +86 28 87512680.
[2] Tel.: +1 412 648 1681.
[3] Tel.: +86 18828041300.

http://dx.doi.org/10.1016/j.elerap.2014.11.004
1567-4223/

the "Bag of Words" assumption, these methods require human intervention to distinguish similar product features and cannot perform the task of eliciting market structure in a completely automatic manner. For example, Lee and Bradlow (2011) use manual reading to identify 39 distinct clusters of product attributes from 99 clusters extracted by K-means. Their studies motivate us to tap into the online product reviews through text-mining techniques.

In our study, we propose a hybrid method to look into the online product reviews. The contributions of this research are multifold:

(1) We utilize topic modeling to save efforts of manual reading to identify product features with synonyms which are likely to distort the elicited market structure (Lee and Bradlow 2011), and thus enhance the automatic level of visualizing market. Traditional text classifying methods assume that a term in a consumer's review represents a unique concept or level of attribute dimensions, thus creating a superfluous number of product features when synonyms are utilized by different product reviewers. In response, Topic modeling technique is a generative probabilistic model of latent dirichlet allocation (LDA), which is capable of grouping synonyms into the same topic and derive the probability distribution of a topic across all documents simultaneously (Blei et al. 2003).
(2) To exploit the strength of topic modeling, a method named WVAP is developed based on Scree plot technique to filter noises in the outputs of Topic modeling, and to elicit valid weight matrix of alternative brands. Since noises result from Topic modeling will accumulate and obscure the hidden differences among valid posterior probabilities, they prevent Topic modeling from accurately evaluate competitive brands. Furthermore, there exist serious format discrepancy among Topic modeling, TOPSIS, and MDS in inputs and outputs. With the help of the WVAP method, these three frameworks can be better integrated to automatically perform market structure visualization. In short, to tackle the data congruence issue, we use WVAP to transform the results of topic modeling into inputs feasible for other methods so as to efficiently implement the proposed hybrid model.
(3) We adopt the TOPSIS method to evaluate products and derive product rankings (Chen and Hwang 1992). Although other MCDM (Multi Criteria Decision Making) methods such as AHP (Analytic Hierarchical Process) are also suitable (Kou et al., 2012), TOPSIS is computationally simple once the weights of each criterion are known by WVAP method. Besides developing perceptual maps of market structure, our ranking of products helps marketers better grasp products' performance in general.

The remainder of the paper is organized as follows. In Section 2, we review methods for obtaining market structure data. Section 3 proposes the WVAP model to integrate topic modeling, TOPSIS, and MDS. Section 4 presents a case study to illustrate the application of the proposed framework. We summarize the case study and generalize its managerial implications in Section 5. In Section 6, we discuss the limitation of this research and future research direction. Finally, the conclusion of this research is given in Section 7.

2. Literature review

Market structure research has played a central role in providing marketers with insights into competitive landscape and marketing opportunities, and in turn prompting marketing actions (Elrod et al. 2002). Marketing strategies, such as market segmentation, brand positioning, pricing, and new product development, are developed based on market structure research (Erdem and Keane 1996). In order to infer market structure, researchers have applied multi-dimensional scaling, conjoint analysis, and cluster analysis to analyze survey data (Elrod 1991; DeSarbo et al. 1991; Green and Srinivasan 1978). Although the validity of these methods is often supported, there are limitations. For example, due to accessibility of data and prolonged data collection process, samples collected by survey or focus group tend to be in small size and cross sectional. It is thus extremely time-consuming to infer dynamic market structure utilizing survey and focus group. Another limitation of these methods is that the product features need to be predetermined before initiating the data collection process. To address the deficiency of the conventional methods in the Internet era, new methods have been developed to analyze the market structure by automatically extracting product characteristics and brand positions from online customer reviews.

The online user-generated data provides a new opportunity to observe the market (Urban and Hauser 2004). Through text mining, researchers can discern patterns and trends hidden in the abundant product reviews (Lee and Bradlow 2011; Netzer et al. 2012). By "listening in" online voice, researchers have made significant improvement in tapping into user-generated content to extract product features and build competitive market landscape. Armed with online product reviews and text-mining techniques, dynamic market structure could be derived more conveniently due to the growing body of unsolicited user-generated online content. Also, product attributes hidden in online consumer discussion need not be predetermined. With the help of an unsupervised learning algorithm, product features which concern consumers most could emerge from the large-scale, unstructured text data automatically.

Despite breakthroughs in mining market structure from online product reviews, several limitations are present:

(i) Most text-mining algorithms, which adopt "Bag of Words" assumption, are likely to assign similar words to different categories, and require additional human reading to combine them. Specifically, these algorithms assume that every word or term is a unique concept of a product attribute. Given that synonyms are prevalent in many text documents, the assumption would lead to superfluous product features and inaccurate market structure.

Table 1
Literature on market structure most relevant to our studies.

Market structure research methods	References	Strengths	Weaknesses
Traditional approaches	Elrod (1991), DeSarbo et al. (1991), Green and Srinivasan (1978)	Reliable methods of inferring market structure from survey-based data	Unsuitable for deriving dynamic market structure; product attributes need to be predetermined
Text-mining techniques	Lee and Bradlow (2011), Netzer et al. (2012)	Deriving dynamic market structure and extracting product attributes from user-generated content	"Bag of Words" and unique term assumption; human intervention in identifying crucial product attributes; excluding rankings of products