Expert Systems with Applications 42 (2015) 3751–3759

Contents lists available at ScienceDirect

Expert Systems with Applications

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

Linguistic features for review helpfulness prediction

Srikumar Krishnamoorthy *

Indian Institute of Management, Ahmedabad, India

A R T I C L E I N F O

Article history:
Available online 5 January 2015

Keywords:
Review helpfulness
Linguistic category features
Sentiment analysis
Machine learning
Text mining

A B S T R A C T

Online reviews play a critical role in customer's purchase decision making process on the web. The reviews are often ranked based on user helpfulness votes to minimize the review information overload problem. This paper examines the factors that contribute towards helpfulness of online reviews and builds a predictive model. The proposed predictive model extracts novel linguistic category features by analysing the textual content of reviews. In addition, the model makes use of review metadata, subjectivity and readability related features for helpfulness prediction. Our experimental analysis on two real-life review datasets reveals that a hybrid set of features deliver the best predictive accuracy. We also show that the proposed linguistic category features are better predictors of review helpfulness for experience goods such as books, music, and video games. The findings of this study can provide new insights to e-commerce retailers for better organization and ranking of online reviews and help customers in making better product choices.

1. Introduction

The advent of Web 2.0 has enabled users to share their opinions, experiences and knowledge via blogs, forums, and other social media websites. In the e-commerce context, Web 2.0 allows consumers to share their purchase and usage experiences in the form of product reviews (e.g. Amazon product reviews, CNET reviews). Such reviews contain valuable information and are often used by potential customers for making purchase decisions. However, some of the most popular products receive several hundreds or thousands of reviews resulting in a review information overload problem. Besides, the review quality across large volume of reviews exhibits wide variations (Liu, Huang, An, & Yu, 2008; Tsur & Rappoport, 2009).

In order to help potential customers in navigating through large volume of reviews, e-commerce websites provide an interactive voting feature. For example, Amazon asks its review viewers "Was this review helpful? Yes/No" to get user votes on reviews. The votes thus gathered from multiple users are then aggregated, ranked and presented, e.g. "24 of 36 people found the following review helpful". Reviews with higher share of helpful votes are ranked higher than the ones with lower helpful votes. This paper aims to study the factors that play an important role for a review to get higher helpful votes. Such an analysis is important for the following reasons: First, reviews can be effectively summarized by filtering low quality reviews. Second, websites that do not use voting feature could benefit from an automated helpfulness prediction system. Third, review ranking system could be improved with a better understanding of the underlying review helpfulness factors, avoiding early bird bias problem (Liu, Cao, Lin, Huang, & Zhou, 2007).

The review voting behaviour which influences review helpfulness can be visualized as a socio-psychological process between the reviewer and the reviewee. This process is facilitated by Web 2.0 as a communication medium. Language plays a very important role in this process between the reviewer and reviewee. In an off-line world, communication between a sender and receiver is often influenced by non-verbal cues, communication contexts and past interactions between the sender and receiver. In the absence of such external factors in the online world, language plays a crucial role. The sender's message (composed using a language) impacts the receivers cognition and influences their behaviour. As the sender's message can be composed in numerous ways, its impact on the receivers cognition and behaviour varies. Our basic intuition is that the review voting behaviour can be better understood by studying the psychological properties and propensities of the language. The Linguistic Category Model (LCM) proposed by Semin and Fiedler (1991) is a conceptual framework that models psychological properties of the language. The linguistic categories used in the LCM model and their descriptions are presented in Table 1.

The LCM model (Coenen, Hedebouw, & Semin, 2006; Semin & Fiedler, 1991) uses three broad linguistic categories, namely Adjectives (e.g. fantastic, excellent, beautiful), State verbs (e.g. love,

* Tel.: +91 79 6632 4834.
E-mail address: srikumark@iimahd.ernet.in

http://dx.doi.org/10.1016/j.eswa.2014.12.044
0957-4174/

hate, envy) and Action verbs. The action verbs are further subdivided into State Action Verbs (e.g. amaze, anger, shock), Interpretive Action Verbs (e.g. help, avoid, recommend), and Descriptive Action Verbs (e.g. call, talk, run). All of these linguistic categories are organized on a abstract-to-concrete dimension. At one extreme (ADJ) the terms are abstract, less verifiable, more disputable and least informative. While at the other extreme (DAVs), the terms are concrete, verifiable, less disputable and most informative.

Consider the following three review examples tagged with key linguistic categories:

1. A fantastic (ADJ) camera. The picture quality of this camera is wonderful (ADJ).
2. This is my first camera and I love (SV) it. The camera is excellent (ADJ).
3. I regularly take(DAV) pics with this camera. The quality of the pics has really amazed (SAV) me. Battery life is fabulous (ADJ). My only issue is that it makes (DAV) a lot of noise in autofocus mode. I strongly recommend (IAV) this camera.

Review 1 is highly abstract and subjective as it primarily uses adjectives. Review 2 uses a subjective verb 'love' indicating the emotional state of the reviewer. The last review provides a more concrete and objective description of the camera using DAVs. Besides, it also contains subjective (ADJ) opinion of the reviewer. It is evident that the review 3 with far more concrete and descriptive information is likely to be more helpful than other two reviews for purchase decision making. Therefore, our basic intuition is that the linguistic categories impact the receivers (or consumers) cognitive process, influence their voting behaviour and affect review helpfulness.

In this paper, our objective is to examine the use of such linguistic category features for predicting review helpfulness. We make a first attempt at devising a new method for extracting linguistic category features from review text and build a binary classification model. We conduct a detailed experimental analysis on two real-life review datasets to demonstrate the utility of the proposed linguistic features. Furthermore, we study the effect of product type on review helpfulness and show that the proposed linguistic features are better predictors of review helpfulness for experience goods.

The rest of the paper is organized as follows. Section 2 describes the related work on review helpfulness. Section 3 elucidates the proposed novel features used in the model. Subsequently, Section 4 presents detailed experimental analysis, results and discussions. Section 5 highlights the implications of this research to theory and practice. Finally, Section 6 provides concluding remarks and outlines directions for future research work.

Table 1
Linguistic categories.

Category	Description
ADJ[1]	Qualifies a noun; Highly subjective and abstract
SV[2]	Refers to mental or emotional state
SAV[3]	Describes the emotional consequences of an action; high positive or negative connotation
IAV[4]	Multitude of actions that have the same meaning; have a positive or negative connotation
DAV[5]	Objective description of a specific action; no positive/negative connotation

[1] Adjectives.
[2] State verbs.
[3] State Action Verbs.
[4] Interpretive Action Verbs.
[5] Descriptive Action Verbs.

2. Related literature

Zhang and Varadarajan (2006) build a regression model for predicting the utility of product reviews. They use lexical similarity, syntactic terms based on Part-Of-Speech (POS), and lexical subjectivity as features. Mudambi and Schuff (2010) formulated a linear regression model for determining factors that contribute towards review helpfulness. Their work was replicated by Huang and Yen (2013) and achieved just 15% explanatory power. The authors conclude that the review helpfulness prediction problem is considerably hard.

Lee and Choeh (2014) build a multilayer perceptron neural network model and make use of product, review metadata, reviewer and review characteristics as features. The key contribution of their work is the use of neural network model to improve helpfulness prediction accuracy. The authors demonstrate that their model works better than other linear regression models used in the literature.

Ngo-Ye and Sinha (2014) use reviewer engagement related features to predict review helpfulness. While prior studies have examined reviewer characteristics, the authors introduce a new concept of reviewer's RFM (Recency, Frequency, Monetary value) to improve the prediction performance. They demonstrate that a hybrid model combining features from textual characteristics and reviewer's RFM provide the best predictive results based on their evaluation on Yelp and Amazon reviews. The authors primarily use a simple bag-of-words model as part of textual features. They do not consider other rich set of features such as readability, subjectivity and metadata that are empirically proven to be better predictors of review helpfulness (Ghose & Ipeirotis, 2011; Kim, Pantel, Chklovski, & Pennachiotti, 2006; Liu et al., 2007).

Liu and Park (2015) present a helpfulness prediction model for travel product websites. They employ a combination of reviewer and review characteristics to predict helpfulness. More specifically, the authors use features such as reviewer's identity, reputation, expertise, valence of reviews, readability and build a text regression model to predict review helpfulness.

A non-linear regression model based on radial basis function for predicting helpfulness of movie reviews is presented by Liu et al. (2008). They utilize reviewer expertise, writing style of reviews and timeliness of reviews as features for the prediction problem. Other works in the literature that use regression model include Cao, Duban, and Gan (2011), Chua and Banerjee (2014), Ghose and Ipeirotis (2011), Korfiatis, Garcia-Bariocanal, and Sanchez-Alonso (2012), Pan and Zhang (2011). These works study various textual and non-textual characteristics of reviews to determine the factors that contribute towards helpfulness of online reviews.

REVRANK is an algorithm for ranking helpfulness of online book reviews (Tsur & Rappoport, 2009). It is an unsupervised algorithm that constructs a lexicon of dominant terms across reviews and builds a virtual core review. The similarity of each review is then assessed against the virtual core review to determine overall helpfulness ranking.

Closely related to the work on helpfulness of reviews is the work proposed by Liu et al. (2007) for detecting low quality reviews. The authors employ features related to informativeness, subjectiveness and readability to classify reviews as high or low quality ones.

Lexical, structural, syntactic, semantic and meta-data related features were used by Kim et al. (2006) for automatic helpfulness assessment. They demonstrate that the use of length of reviews, valence of reviews and unigrams achieves the best results. In one of the more recent works, Hong, Lu, Yao, Zhu, and Zhou (2012) develop a binary helpfulness based review classification system. Their system uses a set of novel features based on needs fulfillment, information reliability and sentiment divergence measure.