Neural Networks 66 (2015) 36–45

Contents lists available at ScienceDirect

Neural Networks

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

Diversifying customer review rankings

Ralf Krestel [a,*], Nima Dokoohaki [b]

[a] Department of Computer Science, University of California, Irvine, USA

[b] Decisions, Networks and Analytics (DNA), Swedish Institute of Computer Science (SICS), Stockholm, Sweden

ARTICLE INFO

Article history:
Received 31 August 2013
Received in revised form 27 December 2014
Accepted 18 February 2015
Available online 26 February 2015

Keywords:
Ranking
Topic models
Summarization
Diversification
Review recommendation

ABSTRACT

E-commerce Web sites owe much of their popularity to consumer reviews accompanying product descriptions. On-line customers spend hours and hours going through heaps of textual reviews to decide which products to buy. At the same time, each popular product has thousands of user-generated reviews, making it impossible for a buyer to read everything. Current approaches to display reviews to users or recommend an individual review for a product are based on the recency or helpfulness of each review.

In this paper, we present a framework to rank product reviews by optimizing the coverage of the ranking with respect to sentiment or aspects, or by summarizing all reviews with the top-K reviews in the ranking. To accomplish this, we make use of the assigned star rating for a product as an indicator for a review's sentiment polarity and compare bag-of-words (language model) with topic models (latent Dirichlet allocation) as a mean to represent aspects. Our evaluation on manually annotated review data from a commercial review Web site demonstrates the effectiveness of our approach, outperforming plain recency ranking by 30% and obtaining best results by combining language and topic model representations.

1. Introduction

It has become a routine among on-line and off-line consumers to inform themselves on review platforms before purchasing a certain product. This has given rise to a considerable amount of customer reviews on e-commerce Web sites. With this in mind, potential customers usually browse through a lot of on-line reviews in order to build confidence in a particular item prior to purchasing it. While reviews have become an important factor in helping Web crowds to further assess the quality of products online, the increase in volume of review data has led to an information overload. Popular products have thousands of reviews. While excess of reviews is a growing problem, recommending unbiased and helpful texts is a growing research field. The quality of reviews may vary drastically and might mislead potential buyers. And the humongous amounts of information not only distracts the confidence seeker, it might also hinder the original goal of users in the first place: They will give up buying a certain product. To deal with these problems, review recommendation techniques are proposed. Review recommendation involves implementing machine-learning techniques for analyzing product reviews based on their lexico-semantic features in order to classify the reviews and offer a balanced and useful view of the reviews to the reader.

While review recommender systems aim at automatic classification of reviews, some commercial Web sites such as Amazon and TripAdvisor[1] approach this problem by allowing users to rate the reviews using star ratings to improve the rankings (e.g. *this review was helpful vs. not helpful*). There are two inherent problems to these rankings based on user feedback: First, good objective reviews contain quite likely redundant information and ranking them based on the helpfulness score will not cover all aspects. Second, these Web sites do not take into account the personal bias. Not all reviews are helpful to everybody. Due to the fact that different users put different emphasis on different aspects, (e.g. *I do not care about battery life, but really need lots of memory*), helpfulness can only be used to filter out very badly written reviews. Therefore, researchers are increasingly distinguishing between the task of review recommendation (Aciar, Zhang, Simoff, & Debenham, 2007), and review ranking (Ghose & Ipeirotis, 2007). To improve existing review recommendation techniques and at the same time improve the ranking of reviews, we propose a novel approach to model and rank reviews. The two main components of

* Corresponding author.
E-mail address: krestel@L3S.de (R. Krestel).

1 http://www.amazon.com and http://www.tripadvisor.com.

http://dx.doi.org/10.1016/j.neunet.2015.02.008
0893-6080/

our system rely on latent Dirichlet allocation to model the reviews and on Kullback–Leibler divergence to generate an adequate ranking. We make use of the assigned star rating for the product as an indicator of the polarity expressed in the review. Our framework covers different ranking strategies based on users' needs and can adapt to various user scenarios. We currently support three strategies: summarizing all existing reviews; focusing on a particular latent topic; or focusing on positive, negative or neutral aspects. We evaluated the system using manually annotated review data gathered from a popular review Web site.

The main contributions of this paper are: (1) Introducing an algorithm to model reviews using latent topics and user-assigned product ratings. (2) Ranking of reviews to summarize all reviews for a product within the top-K results. (3) Diversification of review rankings based on star ratings and/or latent topics. The remaining of the paper is organized as follows: We present related work in Section 2; Section 3 gives an overview of our framework. Section 4 describes the modeling approach, while Section 5 describes the ranking approach. We present the evaluation in Section 6 and close with conclusions and future work.

2. Overview of the field

State-of-the-art in product review mining can be categorized into two major themes: summarization and ranking.

Summarization of reviews. Existing literature on review summarization techniques have a strong focus on review classification and recommendation (Dave, Lawrence, & Pennock, 2003). While reviews in general have been the focus of the majority of works in this field, a breed of new work focuses on opinion mining while taking online reviews as case studies. Review summaries include structured summaries of review text that provide an organized breakdown by aspects or topics, and various formats of sentiment and sentence summaries. Various summary formats complement each other by providing a different level of understanding. For instance, sentiment prediction on reviews for a product can provide a very generic picture of how users feel about a certain product. While users requiring more specific details can turn to topic-based or sentence/sentiment summaries instead. Two state-of-the-art studies about opinion summarization by Liu (2010) and Pang and Lee (2008) give a broad overview of the field. Both surveys cover previous as well as current work but their focuses vary. According to these surveys, review summarization falls under subjective classification, sentiment analysis, or under traditional text summarization. While researchers differentiate between review summarization methods and classic text summarization techniques (Zhuang, Jing, & Zhu, 2006), the connection is obvious. Both aim at identifying salient information: terms, sentences, or paragraphs. Sentiment analysis techniques try to produce a summarized sentiment consisting of sentences from a source document, a single paragraph (Beineke, Hastie, Manning, & Vaithyanathan, 2003), a structured sentence (Hu & Liu, 2004), attribute-value pairs, or just a sentiment score. To build summaries of sentence list structures, Hu and Liu (2004) introduced a method utilizing word attributes such as frequency of occurrence, part-of-speech tagging and WordNet synsets. Following this approach features are extracted, combined with their contextually close words, and finally used to generate a summary by selecting and re-structuring the sentences following the extracted features. Another approach called Opine (Popescu & Etzioni, 2005) uses relaxation labeling to find the lexico-semantic orientation of words, whereas Pulse (Gamon, Aue, Corston-Oliver, & Ringger, 2005) uses bootstrapping to train a sentiment classifier using features extracted by labeling sentence clusters with respect to their key terms. SumView (Wang, Zhu, & Li, 2013) is a semi-supervised Web application capable of review crawling along with automatic product feature extraction. Users can query features of their interest, which are processed in turn using a sentence selection along with the proposed feature-based weighted non-negative matrix factorization (NNMF) algorithm. Finally, the most characteristic set of sentences are selected to summarize the nominal features of each product. Cambria, Schuller, Xia, and Havasi (2013) shed light on new avenues on sentiment analysis and opinion mining by weighing on the notion of concept level analysis in comparison to topic level analysis. Kim, Ganesan, Sondhi, and Zhai (2011), classify existing approaches under two main categories: aspect oriented summarization and non-aspect oriented summarization. The most common category of opinion summarization technique is aspect-based opinion summarization, which involves generating opinion summaries containing a set of topics (also known as aspects or features). Aspect-based summarization involves three steps: feature identification, sentiment prediction, and summary generation. Non-aspect oriented summarization includes: sentiment summarization (Chaovalit & Zhou, 2005), basic and advanced text summarization (Kim & Zhai, 2009) and entity-based summarization (Stoyanov & Cardie, 2008). Recently hybrid models are also proposed which aim at combining aspect and sentiment models in what authors refer to as joint aspect/sentiment models (Lin & He, 2009; Moghaddam & Ester, 2011).

In our work we mine and summarize reviews by choosing complementing reviews and ranking them according to different strategies. The product ratings serve as an indication of sentiment, and the extracted latent topics ensure topical coverage of relevant aspects.

Diversification of reviews and their rankings. The problem of personalized ordering of results has been subject to research in both classic retrieval of documents as well as within recommender systems research. A first approach by Carbonell and Goldstein (1998) based on maximum marginal relevance (MMR) was used as a ranking metric which balances relevance as the similarity between query and search results with diversity as the dissimilarity among search results. Ziegler, McNee, Konstan, and Lausen (2005) take into account a user's full range of interests through diversifying generated recommendation lists and by doing so they minimize redundancy among the recommended items.

While most existing work focuses on the task of diversification of search results, there is also some recent work on review mining. Yu, Zha, Wang, and Chua (2011) look at ranking aspects of reviews. The aspect ranking algorithm identifies important aspects by taking into account the aspect frequency and influence of consumers opinions given to each aspect. When evaluating sentiment classification and aspect rating, they report better Kullback–Leibler divergence (KLD) compared to Hu and Liu (2004). Similar to our work, Xu, Meng, and Cheng (2011) state that two requirements should be taken into account while generating a good summary: representativeness and diversity, in addition to aspect-relevance and sentiment intensity. They present an aspect-based summarization method for online reviews, that incorporates an aspect-sensitive Markov random walk model to satisfy the representativeness requirement, as well as a greedy redundancy to meet the diversity requirement.

We propose a greedy algorithm to minimize the Kullback–Leibler divergence between the topic models of the top-K ranked reviews and all reviews for a product. In addition, we diversify review rankings based on latent topics and language models (LM) to get an optimal coverage for all topics within the top-K results.

3. How to rank reviews?

In contrast to Web search results, reviews for a product cannot be ranked based mainly on relevance since all reviews are supposed to be equally relevant for the product the review is