An Approach for Sentiment Based Product-Feature Diversification of User Generated Reviews

Nasir Naveed[1], Thomas Gottron[2], Zahid Rauf[3]

[1]Department of Computer Science and Information Technology, Virtual University Islamabad, Pakistan, [2]Institute WeST, University of Koblenz, Germany, [3]Department of Electrical Engineering, Faculty of Information and Communication Technology, Balochistan University of Information Technology, Engineering and Management Sciences, Quetta, Pakistan

Abstract

Information shared online by the web users, is increasingly becoming a good source for others to learn from it. In many cases, the shared information is reflecting users' experiences and opinions about events or use of certain products. The volume of this shared information is huge. It is humanly very time consuming to read all this information and make an informed decision. The challenge is to analyze shared contents automatically, find dimensions of discussions, associated opinions and summarize them so that the user can have an extensive overview of the information for decision making. The research work presented in this paper addresses this challenge of information diversification by using probabilistic topic modeling and opinion extraction techniques. The proposed method automatically extracts dimensions of a particular discussion and combines it with the opinions presented in the discussion for information diversification purpose. Experiments on the real-world dataset indicates that our method is able to extract dimensions of a discussion and successfully associate it with the opinions expressed against these dimensions. The method presents users with both positive and negative opinions against a certain discussion to give an overview of the discussion.

keywords: Social Web; Product Reviews; Sentiment Detection; Diversity Ranking

Corresponding author's email: *nasir@vu.edu.pk*

INTRODUCTION

Interactive web provides opportunities to users for sharing information. Among others, the shared information reflects users' daily life experiences of the use of products and services. Web portals such as epinion.com and cnet.com facilitates users to write reviews of the products. These reviews discuss various features or dimensions of the products. Users share what they like and what they don't like in a particular product in the form of opinions. Users also mention which features of a product are important to them. Web provides an interactive way of online text publishing. These reviews help a new user to gain some knowledge of the people say about certain product before deciding whether or not to buy it. Collectively, these reviews provide a comprehensive coverage of the product features and associated opinions about these features, thereby, influence the reader's decision in the purchase process.

Assume a use case where a customer wants to buy a new camera. In the very first step, the user would like to gain some knowledge of the product as well as opinions of other people about this product. Online reviews provide a good starting point for this. For popular products, the number of reviews published by the portals may range from few to hundreds for each product.

It is observed that one single review does not cover all dimensions or features of a product. The extent of coverage also varies for each dimension. To gain complete knowledge of the product in question to make an informed decision, a user has to read multiple reviews, costing the user reasonable chunk of time and effort. The product review portals provide feature to readers for indicating that which reviews are worth reading. This feature is star rating of the reviews. The reader has option to give stars to a review indicating if the review is worth reading or not. But star rating mechanism does not guarantee that a review covers all features of a product and that all positive and negative concerns are addressed. Thus, the task in this situation is to recommend user, a set of reviews that maximize the feature coverage and at the same time addresses positive and negative opinions uttered towards the covered features. We hypothesize that when users are provided with a feature related ranked list of opinion-diversified reviews, it helps satisfying user's information needs in:

- Finding feature related reviews from the corpus
- Finding diversified user sentiments given a certain feature
- Decision making process of buying a certain product

We think of this task as opinion-based diversification of the contents in the domain of information retrieval. Our objective is to provide a method, which helps in discovering product features from reviews written in unstructured natural language and recommend a ranked list of opinion-diversified reviews related to each feature of the product.

Thereby, we support users in their decision-making process of purchasing a product by recommending opinion-diversified customer reviews.

In this contribution, we motivate the overall scenario, provide a formal definition of the task at hand and propose a method for solving the task along with a draft of an evaluation methodology for the proposed method.

This work focuses on two areas of research, mining product features from free text and content diversification using sentiments. In the past researchers have used feature mining and sentiment classification approaches separately or in combination for extracting product features from user reviews and determining the sentiment polarity of the reviews.

There exist wide range of manual to unsupervised approaches for product feature mining and sentiment detection. To name some recent approaches, (Guo et al., 2009) suggested an unsupervised approach for product-feature mining and categorization with multilevel latent semantic association.

Shi and Ming Yu (Shi et al., 2011) investigated a theoretical framework for mining issues related to product features from customer reviews and recommended a DFM (Data, Function, Mining) model. This model mines product feature in structured form from reviews. Zhai et al. (2010) proposed an extension in LDA and devised Constrained LDA for grouping product features in opinion mining task.

Eirinaki et al. (2011) suggested an algorithm for overall sentiment analysis of reviews. The algorithm identifies semantic orientation of the specific component of the review that leads to specific sentiment. Qui et al. (2009) recommended a self-supervised model for opinion categorization. A dictionary-based approach of opinion mining was used in their model.

For sentiment analysis, we apply a simple dictionary-based approach proposed by Tech et al., (1999) which has already been applied successfully in other social web scenarios (Naveed et al., 2011).

Ranking of product reviews based on feature coverage is also addressed by (Tsaparas et al., 2011). They also categorized it as a diversification task but stressed only on a better covering of product features. An overview of methods addressing content diversification task based and their evaluations in particular is presented by (Carterette, 2011).

The approaches mentioned above perform feature mining, sentiment analysis and some kind of diversification but do not explicitly use sentiments for review diversification. One work by (Krestel et al., 2011), which is similar to our work, uses Latent Dirichlet Allocation for finding topics in the product reviews and uses star ratings as an indicator of opinion for use in diversification problem. Their approach depends on star ratings for finding opinion at review level and lacks the ability to relate the polarity to individual aspects of the product or discussion.

However, our approach uses a different formalization and algorithm and does not rely on star ratings for determining sentiment polarity, rather we determine sentiments from the review text using a dictionary-based approach proposed by (Bradley, 1999) that is already used in other social web scenarios (Naveed et al., 2011).

Determining sentiments from the text enables us to relate polarity with individual features rather than just relating the global polarity of the review to the product. Our method can be generalized or can be adopted to scenarios where there are no explicit star ratings provided for the reviews and also where polarity is required to be related with individual aspects of the whole. One such scenario for example is political opinion mining from the online discussion threads where some politician wants to have an overview of user sentiments regarding different aspect or features of a policy decision to be made by the government.

MATERIALS AND METHODS

We now provide a formal definition for the $\textsc{FSCoverage}(k)$ task of generating a selection of k product reviews that cover the features and a diversified sentiment range.

For a product $\mathcal{P}$ (e.g. a Digital Camera), we assume a finite set $\mathcal{A}$ of features (e.g. picture quality, focus speed, ergonomics, etc.). Additionally, we define a set of opinion labels O (e.g. positive or negative valence, calm or excited arousal, etc.). Further, we have a corpus $\mathcal{C}$ of reviews related and relevant to the product $\mathcal{P}$. The reviews address features and utter sentiments about them.

In review d_i, we compute the vigor of opinion $o \in O$ with respect to feature $a \in \mathcal{A}$ with the value . $f_o(d_I, a) \in [-1, 1]$. As eventually we tend to find a set of reviews that covers for each feature the extreme points of sentiments, we further introduce functions f^+ and f^- to project the opinion strength values on the absolute positive and negative values only. Formally we define

$$f_o^+(d_I, a) = \max(0, f_o(d_I, a))$$
and
$$f_o^-(d_I, a) = |\min(0, f_o(d_I, a))|$$

This means, each review can be seen as a set of triples

$(f, o, v) \in \mathcal{S} \times O \times \mathbb{R}^+$

combining a feature f with an opinion label o and a degree v of strength of this opinion. For a given set $\mathcal{S} \subset \mathcal{C}$ of reviews, we can now define a feature-sentiment-diversity score F by:

$$F(\mathcal{S}) = \sum_{a \in \mathcal{A}} \sum_{o \in O} \left(\max_{d_i \in \mathcal{S}} f_o^+(d_i, a) + \max_{d_i \in \mathcal{S}} f_o^-(d_i, a) \right)$$

The $\textsc{FSCoverage}(k)$ task is to maximize the score $F(\mathcal{S})$ under the condition $|\mathcal{S}| \le k$. In analogy to the proof in (Tsaparas et al., 2011) it can be shown that $\textsc{FSCoverage}(k)$ is NP-hard.

Addressing the task of generating a feature related sentiment-diversified selection of product reviews actually poses several challenges:

(a) to identify the set of features $\mathcal{A}$ (if not explicitly given),

(b) to estimate the sentiment values $f_o(d_I, a)$ for a given feature in a specific review and

(c) to provide a good approximative solution for $\textsc{FSCoverage}(k)$.

We have developed an approach which uses different methods to address these three challenges. We use LDA (Blei et al. 2003) to detect topics in the product reviews. Given the strong focus of the reviews and the use case, we found that the LDA topics approximate the product features, thus, giving us with the set $\mathcal{A}$. To estimate the sentiments, we employ the ANEW sentiment dictionary (Bradley et al., 1999) to detect the global valence, arousal and dominance in a review d_i. We then estimate the values $f_o(d_i, a)$ by combining the review's global sentiments with the relation of the review to the LDA topics.

In practice we use $\tilde{f}_o(d_i, a) = o(d_i) \cdot P(a|d_i)$, where $o(d_i)$ is the global orientation of document d_i under sentiment o and $P(a|d_i)$ is the probability that d_i addresses feature a as provided by LDA. For solving $\textsc{FSCoverage}(k)$,we use a greedy algorithm. We keep track of the extent of positive and negative sentiments covered so far for each feature. Therefore, we initialize values $\mathcal{F}_o^+(a) = 0$ and $\mathcal{F}_o^-(a) = 0$ for each sentiment o and each feature a. Now, we determine the value $contrib(d_I)$ of additional positive and negative sentiments contributed by each document by d_i:

$$contrib(d_i) = \sum_{a \in \mathcal{A}} \sum_{o \in O} \left[\max(0, f_o^+(d_I, a) - \mathcal{F}_o^+(a)) + \max(0, f_o^-(d_I, a) - \mathcal{F}_o^-(a)) \right]$$

Then we add the review d_i with the highest $contrib(d_i)$ score to the set of selected reviews $\mathcal{S}$ and update the contributions by

$$\mathcal{F}_o^+(a) = \max(f_o^+(d_i, a), \mathcal{F}_o^+(a))$$
and
$$\mathcal{F}_o^-(a) = \max(f_o^-(d_i, a), \mathcal{F}_o^-(a))$$

With this updated value, we can recalculate the $contrib(d_i)$ scores of the remaining reviews and iteratively add the document that adds most coverage of features and sentiment values to the result set. This iteration is repeated until we have selected k reviews for $\mathcal{S}$.

We use probabilistic methods to discover latent features of the products from the reviews. Our method is based on Latent Dirichlet Allocation (LDA). We use dictionary based approached for determining the sentiment polarity of the post in each of the three opinion classes i.e. valence (positive vs negative), arousal (excited vs calm) and dominance (authenticity of the opinion).

Dataset

We use CNET product review website to crawl the editorial and customer reviews available for tech products in different categories like smartphone, camcorders, printers etc. For each product there exists one editorial review where editor gives his assessment for important features of the products. The website also provides a way for the customers to share their own assessment of the performance of the product in free form text. The customer is also able to provide star rating to the existing reviews whether they are helpful for other customers.

RESULTS AND DISCUSSION

In this section, we elaborate on evaluation methodologies for the proposed approach for generating a set of sentiment-diversified reviews for product features. While initial test of the above described approach provides subjectively very satisfying results, this setting will serve to demonstrate and compare the performance of the algorithm. For evaluating the approach, we will use reviews data set consisting of reviews collected from the CNET product review website.

We analyze the performance of the proposed algorithm in extracting product features and in generating feature relevant ranked list of sentiment-diversified reviews. The evaluation approach consists of an objective evaluation using state of the art metrics for measuring the performance of a search result diversification systems as well as a task based subjective evaluation for measuring the usefulness of the system for better customer satisfaction in decision making process of buying a product. Below is a brief description of both the evaluation approaches.

Baseline model: Our baseline model comprises a modified CNET review portal. CNET provides user reviews for a product in the order of the helpfulness of the review as rated by the other users by assigning between one and five stars. We select two reviews from each of the five-star categories to have a representative sample of the user comments, thus, forming an advanced baseline system.

Proposed model: is an adapted system derived from CNET review portal and based on our approach provides sentiment-diversified selection of product reviews. It generates related but ranked list of sentiment-diversified reviews for a given product against selected feature(s). Our model selects reviews that cover best extreme opinions based on the greedy algorithm presented above.

Objective evaluation

We use an objective evaluation to validate if the model is effective in producing feature related sentiment-diversified reviews of the product. As observed, each review or posting tends to cover one feature or a combination of features to different extents of coverage. So, it is safe to assume that one review is related to one or more features but with different degree of relevance. The model provides a ranked list of sentiment-diversified reviews related to a feature and assigns relevance scores to each of the review in the list.

We employ crowd sourcing for judging the relevance of a review on a graded scale of relevance ranging from irrelevant to highly relevant. We use crowd sourcing to judge the relevance of each review to the given feature. We provide each test subject with feature and ranked list of reviews related to that feature generated by the model and are then asked to judge each review and grade it on a scale of 0 – 3, where 0 means irrelevant, 1 means marginally relevant, 2 means relevant and 3 means highly relevant.

There exist metrics which measure relevance and diversity altogether in ranked retrieval evaluation. For our evaluation scenario we use (Clarke et al., 2008). It views information needs and document as sets of nuggets. discounts the value of each retrieved relevant document based on nuggets already seen and then further discounts based on the document rank. They key idea behind is to encourage diversity by means of discouraging redundancy.

Subjective evaluation

We also evaluate the model performance using task based subjective evaluation. The focus is on observing human interaction with the system. The objective of this evaluation is to judge the efficiency and effectiveness of the model in providing helpful information to the customers.

For a given task we measure for both systems

- Efficiency: measure time taken to complete a task
- Effectiveness: listing positive and negative features of the product by test subjects
- Satisfaction: questionnaires for measuring satisfaction of user in helpfulness of the system to complete task, ease of use and extent of the information provided by the system to complete the task.

From the information recorded during and after the execution of the tasks we compute the performance of both systems for their usefulness in providing helpful information.

We introduced and formalized the task of generating a selection of product reviews that cover product features and a diversified opinion range. We propose one solution to this task and used probabilistic topic modeling for mining product features from the reviews and dictionary-based approach for detecting sentiments from review. Our method combines sentiments with product features and recommend the user, a list of reviews, that maximize the product feature coverage and associated sentiments. In the future work, we will use the presented objective and subjective evaluation methods to analyze the effectiveness of our method.

REFERENCES

- Blei DM, Ng AY, Jordan MI. (2003). Latent Dirichlet Allocation. The Journal of Machine Learning Research 3:993-1022.

- Bradley MM, Lang PJ. (1999). Affective Norms of English Words (ANEW): Stimuli, instruction manual and affective ratings. Technical report C-1, Gainesville, FL. The Center for Research in Psychophysiology, University of Florida.

- Carterette B. (2009). An Analysis of NP-Completeness in Novelty and Diversity Ranking. Information Retrieval 14: 89-106.

- Clarke CL, Kolla M, Cormack GV, Vechtomova O, Ashkan A, Buttcher S, MacKinnon I. (2008). Novelty and Diversity in Information Retrieval Evaluation. In Proc. of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '08). ACM, New York, NY.

USA, pp 659-666.

- Eirinaki M, Pisal S, Singh J. (2011). Feature-Based Opinion Mining and Ranking. Journal of Computer and System Sciences 78:1175-1184.

- Guo H, Zhu H, Guo Z, Zhang X, Su Z. (2009). Product Feature Categorization with Multilevel Latent Semantic Association. In Proc. of the 18th ACM conference on Information and knowledge management (CIKM '09). ACM, New York, NY, USA, pp 1087-1096.

- Krestel R, Dokoohaki N. (2011). Diversifying Product Review Rankings: Getting the Full Picture. In Proc. of the 2011 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology 138-145:1.

- Naveed N, Gottron T, Kunegis J, Che Alhadi A. (2011). Bad News Travel Fast: A Content-Based Analysis of Interestingness on Twitter. In Proc. of the 3rd International Conference on Web Science. Koblenz, Germany. June 15 - 17, 2011.

- Qiu L, Zhang W, Hu C, Zhao K. (2009). SELC: A Self-Supervised Model for Sentiment Classification. In Proc. of the 18th ACM conference on Information and knowledge management (CIKM '09). ACM, New York, NY, USA, 929-936.

- Shi L, MingYu J, Li, Shi Qi, Mingyu Ji. (2011). A DFM Model of Mining Product Features from Customer Reviews. In Proc. of the International Conference on Control, Automation and Systems Engineering (CASE). pp 1-5.

- Tsaparas P, Ntoulas A,Terzi E. (2011). Selecting a Comprehensive Set of Reviews. In Proc. of the ACM international conference on Knowledge discovery and data mining, New York, NY, USA, 2011. ACM.

- Zhai Z, Liu B, Xu H, Jia P. (2011). Constrained LDA for Grouping Product Features in Opinion Mining. In Proc. of the 15th Pacific-Asia conference on Advances in knowledge discovery and data mining - Volume Part I (PAKDD'11), Joshua Zhexue Huang, Longbing Cao, and Jaideep Srivastava (Eds.), Vol. Part I. Springer-Verlag, Berlin, Heidelberg, pp 448-459.