Review Selection Using Micro-Reviews

N. Rajeswari[1], A. Geethika[2], D. Navya[3], A. Hussain[4], A. Rambabu[5]

[1]Associate Professor, Department of CSE, Gudlavalleru Engineering College, India

[2,3,4,5]Department of CSE, Gudlavalleru Engineering College, India

E-mail: rajeswari.gec@gmail.com[1], geethika.arja@gmail.com[2], navya11dhadi@gmail.com[3] , chinnahussain.akumalla@gmail.com[4], rrambabu115@gmail.com[5]

ABSTRACT

Given the proliferation of review content, and the fact that reviews are highly diverse and often unnecessarily verbose, users frequently face the problem of selecting the appropriate reviews to consume. Micro-reviews are emerging as a new type of online review content in the social media. Micro-reviews are posted by users of check-in services such as Foursquare. They are concise (up to 200 characters long) and highly focused, in contrast to the comprehensive and verbose reviews. In this paper, we propose a novel mining problem, which brings together these two disparate sources of review content. Specifically, we use coverage of micro-reviews as an objective for selecting a set of reviews that cover efficiently the salient aspects of an entity. Our approach consists of a two-step process: matching review sentences to micro-reviews, and selecting a small set of reviews that cover as many micro-reviews as possible, with few sentences. We also propose an efficient heuristic algorithm that approximates the optimal solution. Finally, we perform a detailed evaluation of all the steps of our methodology using data collected.

INTRODUCTION

A **review** is a report in the media in which someone gives their opinion of something such as a new book or film.Now-a-days, every social media contains reviews. Some reviews are lengthy and some other reviews are small. These reviews are highly miscellaneous and indirect. Readers may not show interest in reading these types of reviews because some reviews are irrelevant to the product or service. It is difficult to determine whether a review is given by a genuine user or by aspammer. A user can write a review minimum of 150 words long, which include their experience, interest etc.

By writing your own product reviews, you can share information from your own experiences that will contribute to the users. If customer owns a computer or a smartphone, chances are that he has done some form of online shopping before. Online shopping has many pros and cons as you have probably experienced. On one hand, he can shop in the comfort of his home or on the go. On the other, customers can never really be sure of the quality of the product until he/she receives it. Because customers can't see or test the product before buying, it means they willalways read product reviews first.All of these online reviews help a person choose the perfect product without testing it in real life. It also helps companies grow their following and brand awareness. For this reason, it's vital for your products to have an ample number of reviews, especially if it's new. A high number of positive reviews will help people discover it more quickly.With the rapid growth of social networking, we notice a new form of reviews. This content which we called *micro-reviews,* can be found in micro-blogging services that allow users. In the case of restaurants, tips are frequently recommendations (e.g., what to order), opinions (what is great or not), or actual "tips".

Micro-reviews serve as an alternative source of content to reviews for readers interested in finding information about it. They have several advantages. *First*, due to the length restriction, micro-reviews are *short and refine*, identifying the most noticeable or apropos points about the restaurant. *Second*, because some micro-reviews are written on site, right when the user has checked in, they are *ease*, author express straight. *Third*, because most authors check-in by mobile apps, these authors are likely at the place when leaving the tips, which makes the tips more likely to be *authentic*. Micro-blogging sites also have the ability filter out tips without an accompanying check- in, thus, boosting the authenticity of the tips.

Micro-reviews and reviews nicely complement each other. While reviews are lengthy and verbose, tips are short and concise, focusing on specific aspects of an item. At the same time, these aspects cannot be properly explored within 200 characters. This is accomplished in full-blown reviews which elaborate and contemplate on the intricacies of a specific characteristic. Marrying these two different reviewing approaches can yield something greater than the sum of their parts: detailed reviews that focus on aspects of a venue that are of true importance to users.

We consider the following problem. Given a collection of reviews, and a collection of tips about an item, we want to select a small number of reviews that best *cover* the content of the tips. This problem is of interest to any online site or mobile application that wishes to showcase a small number of reviews. For example, review sites such as Yelp, which recently introduced tips as part of their mobile application, would benefit from such a review selection mechanism. The need for concise and comprehensive content becomes especially more pronounced for the mobile applications of such sites, where the screen real-estate is limited, and the user attention span is shorter.

RELATED WORK

[1]The selection of "good" set of reviews is the line of work that deals with the reviews which are selected. The main objective is to select the set of reviews which covers the all attributes. The attributes to be covered are the given set of attributes. The objective of micro reviewing is which have the both positive and negative aspects of each and every attributes. The main work is to maintain the underlying distributive of both positive and negative comments in the reviews.

[2]The main Objective is to cover the modulated opinion clusters. Here consider the problem of selecting a good set of photos based on modulation, quality and coverage which is related to review selection. Here our work is done on the same baselines which distinct ways and these two ways, first is the formulation which represents micro-reviews than attributes .secondly the efficiency requirement to the coverage formulation and here we mainly focus on coverage but not efficiency. Here the max coverage algorithm is considered as a baseline, here the cost and count apply are the two constraints combined called as budgeted max coverage which is a variant of max coverage. [3]Our max coverage formulation contains both count apply and cost .Another coverage formulation is red-blue set cover which takes collection of sets and here every set may contains a mixture of blue and red elements as main objectives. The main objective of this formulation is to collect a sub-collection of all the blue elements and as well as few red elements as possible. Here we should interpret hat red element that a red element is an inappropriate and irrelevant sentence in a review ,so the main objective is to get a set that contains less number of irrelevant sentences .Here in our work efficiency is considered as a threshold constraint ,our efficiency definitions are expressed in terms of fractions instead of in terms of count this results in the substantially margined and different formulation.

[4]Our main notion of finding a "good" set of reviews is the main problem for determining the quality of each individual review. Some public sites allow its users to rate the review by its usefulness and its helpfulness this process is called review ranking here reviews are ranked according to their useful information and helpfulness to users. This review ranking works based on a supervised regression or classification approach in machine learning by considering the useful reviews as the target class. Target class can be formulated by producing set of reviews to first rank all the reviews based on the individual merit here merits are usefulness an d helpfulness of the individual review.

[5]The main drawback of this formulation is it ignores the potential similarities between the top rated reviews so the top rated.

PROBLEM STATEMENT

There will be a small number of reviews which consists perfect coverage and efficiency called as an Ideal set. When coming to the practice, such an ideal set rarely exists. Here we express the selection problem which seeks the best possible solution which gives optimized solution for the optimization problem. The main optimization problem is to optimize both coverage and efficiency with no single optimal solution. From the two metrics we need to select one metric to optimize. In many cases it is not mandatory to have perfect efficiency .there may be few sentences in a review that do not cover any tip it only helps to increase the readability of the review .There are some threshold values where the efficiency should not fall below the minimum acceptable range. So here we limit the efficiency and opt to view our problem and we ask for the solution with maximum coverage.

PROBLEM 1

Maximum coverage (EFFMAXCOVERAGE)

Here we take set of reviews R, a set of tips T, a Matching function MF it is a function which shows the matching between review sentences and tips and we also take parameters a and K

Here we limit the number of reviews in the k in the set S by following maximum coverage.

PROBLEM 2

Efficient coverage (EFFSETCOVER)

Here we take set of reviews R, a set of tips T, the matching function MF here we select a set S which belongs to R which covers the all the tips in the T such that the efficiency of the set is high and the size of the set S is minimized. This EFFSETCOVER is used to reduce the space and time resources which are limited here.so we need to select the small number of reviews to show to the user which requires less space and time which helps user for better usage and understand. Here we should understand the trade-off between coverage and efficiency.

METHODOLOGY

ALGORITHM: Greedy-EffMaxCover algorithm.

INPUTS: set of reviews R , set of tips T

Efficiency function Eff; Integer budget Value K,

Parameters a, b.

OUTPUT:

A Set of reviews S is a subset of R of size K
Initially, S is an empty set.

Matching Reviews and Tips :

Reviews and tips are of different granularity. A tip is short and concise, usually making a single point, while a review is longer and multi-faceted, discussing various aspects of an entity. Intuitively, a review covers a tip, if the point made by the tip appears within the text of the review. To make this more precise, break a review into sentences, which are semantic units with granularity similar to that of thetips.

Matching function MF is used to determine the similarity between the Tip and Review

MF(s,t)=1; if s and t are similar;

MF(s,t)=0; if s and t are not similar;

Selection Coverage :

We say that s covers t if and only if s and t are matched in a sentence. And the main objective that should be satisfied is the sentence s should belongs to set of Reviews R and set of tips T should be covered by at least one sentence of review R.
We define the coverage as Cov(R) of review R

$\text{Cov}(S) = |T_R| / |T|$ fraction of tips covered by the set S ;

Selection Efficiency:
Some sentences are not relevant to any tips in the set where as some Reviews may have high coverage is covers the most of the tips. we need to avoid the sentences which are not relevant and highly covered.so here we introduce efficiency which is called Selection Efficiency where we select the efficient Review.

$\text{Eff}(R) = |R^r| / |R|$; Rr is relevant sentences

which covers at leastone tip ;

Extending the definition of efficiency to a collection of reviews is a little more involved. It needs a way to aggregate the efficiency of the individual reviews. Here, it propose three possible definitions.

Minimum Efficiency:
In this case, the efficiency of set of reviews S is defined as the minimum efficiency of any review in the set.

$$Eff_{\min}(S) = \min_{R \in S} Eff(R)$$

Average Efficiency:
In this case the efficiency of set of reviews S is average efficiency of each review in the set.

$$\text{Eff}_{\text{avg}}(S) = \frac{\sum_{R \in S} \text{Eff}(R)}{|S|}$$

Here we calculate the Selection Efficiency and Selection Coverage of each and every Review. This helps for the increase in the efficiency.

Eff avg(S) = Eff(R) / | S |

Bag efficiency:

In this case, view a collection of reviews S as a single review R_S consisting of the union of the sentences of the reviews. Then define the efficiency of the collection as the efficiency of R_S. Formally, have $R_S = \cup_{R \in S}$ R, and

$\mathrm{Eff}_{bag}(S) = \mathrm{Eff}(R_S)$.

DATASET

A collection of related sets of information that is composed of separate elements Here we mainly use two sets they are set of Reviews and set of micro-reviews. micro-reviews are nothing but set of tips and these are from the same set of physical entity called review or positive or negative opinion given by the user. Here we pick the different and prominent domains which are highly used by the common users and where we have active platforms for Reviews and micro-reviews. Here our dataset contains data about the reviews and tips of different top restaurants. To obtain the tips of the same 110 restaurants our datasets contains the adequate restaurants information that may be tips and micro-reviews and some restaurants contains at least 50 tips. Restaurants information which covers less than 50 tips are excluded because they does not cover the minimum threshold level of tips.

CONCLUSION

We introduce the use of micro-reviews for finding an efficient set of reviews, which is novel in the objective of micro-review coverage, as well as in the efficiency constraint. We describe an optimal algorithm based on Integer Linear Programming. Since the problem is NP-hard, we also propose a greedy algorithm, which is virtually identical to the optimal solutions in coverage and efficiency, but it is much faster computationally. Evaluation over a corpora of restaurants' reviews and micro-reviews shows that our approach outperforms the baselines in discovering review sets consisting of compact, yet informative reviews.

REFERENCES

[1] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet allocation," J. Mach. Learn. Res., vol. 3, pp. 993–1022, 2003.

[2] R. D. Carr, S. Doddi, G. Konjevod, and M. V. Marathe, "On the red-blue set cover problem," in Proc. 11th Annu. ACM-SIAM Symp. Discrete Algorithm, 2000, pp. 345–353.

[3] C. Cheng, H. Yang, I. King, and M. R. Lyu, "Fused matrix factorization with geographical and social influence in locationbased social networks," in Proc. 26th AAAI Conf. Artif. Intell., 2012, p. 1.

[4] K. Ganesan, C. Zhai, and J. Han, "Opinosis: A graph-based approach to abstractive summarization of highly redundant opinions," in Proc. 23rd Int. Conf. Comput. Linguistics, 2010, pp. 340–348.

[5] K. Ganesan, C. Zhai, and E. Viegas, "Micropinion generation: An unsupervised approach to generating ultra-concise summaries of opinions," in Proc. 21st Int. Conf. World Wide Web, 2012, pp. 869–878.