International Journal on Applications in Engineering and Technology
Volume 6 : Issue 2: August 2020, pp 1 – 6 www.aetsjournal.com ISSN (Online) : 2455 - 0523

SELECT SMALL SET OF REVIEWS USING MICRO-REVIEWS

Dr.D.SATISHKUMAR, Mr.G.JEEVANANTHAM, Mrs.R.GNANAKUMARI,
Mr.A.SIVAKUMAR, Mr. D. KALEESWARAN

***Abstract* —** As Reviews can be selected from small set of reviews. We perform reviews using the sentence with small contents. 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 formulate this objective as a combinatorial optimization problem, and show how to derive an optimal solution using Integer Linear Programming. 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 from Foursquare and Yelp.

Keywords- Micro-review, coverage, review selection.

Dr.D.SatishKumar PhD , Associate Professor , Department of Computer Science and Engineering, Nehru Institute of Engineering and Technology, Coimbatore. (Email: satishcoimbatore@gmail.com)

Mr. G. Jeevanantham M.E, Assistant Professor , Department of Computer Science and Engineering, Nehru Institute of Engineering and Technology, Coimbatore. (Email: nietjeevanantham.g@nehrucolleges.com)

Mr. R.Gnanakumari M.E, Assistant Professor , Department of Computer Science and Engineering, Nehru Institute of Engineering and Technology, Coimbatore. (Email: mailtokumari81@gmail.com)

Mr.A.Sivakumar, Assistant Professor, Department of Computer Science , Rathinam College of Arts and Science, Coimbatore. (Email: sivamgp@gmail.com)

Mr. D. Kaleeswaran M.E., Assistant Professor – Department of Computer Science and Engineering , Rathinam Technical Campus - Coimbatore.(Email: kaleeswaranme@gmail.com)

I. INTRODUCTION

Online reviews are pervasive. Today, for almost any product or service, we can find ample review content in various Web sources. For instance, Amazon.com hosts product reviews as part of an online shopping experience to assist their customers in determining which product is most suitable for their need. Yelp.com is a popular site for restaurant reviews, assisting diners to plan restaurant visits. Reviews are immensely useful in aiding decision-making, because they allow the readers to anticipate what their experience would potentially be based on the prior experiences of others, without having to make a trip to the store or the restaurant.

While useful, the deluge of online reviews also causes some issues. Readers are inundated by the numerous reviews, and it is not clear which reviews are worthy of a reader's attention. This is worsened by the length and verbosity of many reviews, whose content may not be wholly relevant to the product or service being reviewed. Reviewers often diverge, meandering around personal details that do not offer any insight about the place being reviewed.

Furthermore, it is getting increasingly more difficult to determine the authenticity of a review, whether it has been written by a genuine customer sharing her experience, or by a spammer seeking to mislead. Identifying and selecting high quality reviews to show to the users is a hard task, and it has been the focus of substantial amount of research. With the recent growth of social networking and micro-blogging services, we observe the emergence of a new type of online review content.

This new type of content, which we term micro-reviews, can be found in micro-blogging services that allow users to \check-in", indicating their current location or activity. For example, at Foursquare, users check in at local venues, such as restaurants, bars, coffee shops. AtGetGlue.com, users check in to TV shows, movies, or sports events. Check-ins are also possible within social networking sites such as Face book, or

Twitter. After checking in, a user may choose to leave a 140 character-long message about their experience, effectively a micro-review of the place or the activity. Following the Foursquare terminology, we will refer to these messages as tips.

In the case of restaurants ,these tips are frequently recommendations (e.g., what to order) or opinions (what is great or not). For example, this is a foursquare tip for a popular restaurant in New York:\Be patient. It's worth the wait. Their ramen has crack in it." Micro-reviews serve as an alternative source of content to reviews for readers interested in finding information about a place.

They have several advantages. First, due to the length restriction, micro-reviews are concise and distilled, identifying the most salient or pertinent points about the place according to the author. For example, the tip above focuses on the long wait and the quality of the ramen. Second, because some micro-reviews are written on site and in the moment right after checking in, they are spontaneous, expressing the author's visceral reaction to her experience.

II. LITERATURE SURVEY

Title: *Micro opinion Generation: An Unsupervised Approach to Generating Ultra-Concise Summaries of Opinions.-2012*

Author*: Kavita Gan esan, Cheng Xiang Zhai, Evelyne Viegas.*

This paper presents a new unsupervised approach to generating ultra-concise summaries of opinions. We formulate the problem of generating such a micro pinion summary as an optimization problem, where we seek a set of concise and non-redundant phrases that are readable and represent key opinions in text.

We measure representativeness based on a modified mutual information function and model readability with an n-gram language model. We propose some heuristic algorithms to efficiently solve this optimization problem. Evaluation results show that our unsupervised approach outperforms other state of the art summarization methods and the generated summaries are informative and readable.

Summarization of opinions is crucial in helping users digest the many opinions expressed on the web. Previous studies have primarily focused on the task of generating highly structured summaries. This could be a simple sentiment summary such as 'positive' or 'negative' on a topic of interest or a multi-aspect summary such as battery life: 1 star, screen: 3.5 stars, etc for an mp3 player While structured summaries can be useful in conveying the general sentiments about a person, product, or service, such summaries lack the level of detail that an unstructured (textual) summary could offer, often forcing users to go back to the original text to get more information. Textual summaries are thus critical in conveying key opinions and reasons for those opinions at different granularities(i.e. entity level or topic level).

In this paper, we explore the task of generating a set of very concise phrases, where each phrase (micro pinion) is a summary of a key opinion in text. The ultra-concise nature of the phrases allows for flexible adjustment of summary size according to the display constraints. Our emphasis on generating concise abstractive summaries (rather than extractive summaries), makes this a unique summarization problem which has not been previously studied. In Table 1, we how examples of envisioned micro pinion summaries. On the surface, our summarization task appears to be similar to a key phrase extraction problem.

However, since the goal is to help users digest the underlying opinions, there are some important aspects that are unique to this task. In traditional key phrase extraction, the goal is primarily to select a set of phrases to characterize documents. Thus, phrases such as battery life and screen from a set of reviews about a phone may be selected as candidate key phrases. For our task, such key phrases are meaningless without the associated opinions.

In addition, since we want readers to understand the opinions in the summary, the phrases in the summary need to be fairly well-formed and grammatically sound. Consider a phrase such as 'short battery life' in contrast to one such as 'life short battery'. Even though both phrases contain the same words, the ordering is different; changing their meaning, where the former is readable but the latter would make no sense to the reader. This readability aspect is less of a concern in traditional key phrase extraction as the phrases are only used to 'tag' documents.

Evaluation results using a set of product reviews shows that our approach is effective in generating micro pinion summaries and outperforms other summarization approaches. Further, the proposed approach is lightweight and general, requiring no linguistics or domain knowledge. The first property ensures that only a set of related words are used in the generated phrases to avoid conveying incorrect information. The assessors were not informed about which method was used to generate the summaries.

***Title**:Sentiment Analysis of Twitter Data.-2011*
***Author**:Apoorv Agarwal BoyiXie Ilia Vovsha Owen Rambow Rebecca Passonneau.*

Micro blogging websites have evolved to become a source of varied kind of information. This is due to nature of micro blogs on which people post real time messages about their opinions on a variety of topics, discuss current issues, complains, and express positive sentiment for products they use in daily life. In fact, companies manufacturing such products have started to poll these micro blogs to get a sense of general sentiment for their product. Many times the companies study user reactions and reply to users on micro blogs. One challenge is to build technology to detect and summarize an overall sentiment.

Furthermore, we also show that the tree kernel model performs roughly as well as the best feature based models, even though it does not require detailed feature engineering. We use manually annotated Twitter data for our experiments. One advantage of this data, over previously used data-sets, is that the tweets are collected in a streaming fashion and therefore represent a true sample of actual tweets in terms of language use and content. Our new data set is available to other researchers.

III. PROSPOSED REVIEW SELECTION USING MICRO-REVIEWS

The system proposes a novel mining problem. The system introduces a novel formulation of review selection. This system also proposes an Integer Linear Programming (ILP) formulation, and it provides an optimal algorithm.

The system also proposes a greedy algorithm to identity the optimal solution in coverage and efficiency. Foursquare, the currently most popular location-based social network, allows users not only to share the places (venues) they visit but also post micro-reviews (tips) about their previous experiences at specific venues as well as "like" previously posted tips.

The number of "likes" a tip receives ultimately reflects its popularity among users, providing valuable feedback to venue owners and other users. In this paper, we provide an extensive analysis of the popularity dynamics of foursquare tips using a large dataset containing over 10 million tips and 9 million likes posted by over 13, 5 million users.

Our results show that, unlike other types of online content such as news and photos, foursquare tips experience very slow popularity evolution, attracting user likes through longer periods of time. Moreover, we find that the social network of the user who posted the tip plays an important role on the tip popularity throughout its life-time, but particularly at earlier periods after posting time.

We also find that most tips experience their daily popularity peaks within the first month in the system, although most of their likes are received after the peak. Moreover, compared to other types of online content (e.g., videos), we observe a weaker presence of the rich-get-richer effect in our data, demonstrating a lower correlation between the early and late popularities. Finally, we evaluate the stability of the tip popularity ranking over time, assessing to which ex-tent the current popularity ranking of a set of tips can be used to predict their popularity ranking at a future time.

A. Problem Statements

Ideally, there would be a small number of reviews with perfect coverage and efficiency. In practice, such an ideal set rarely exists, if ever. We formulate the selection problem a optimization problem where we seek the best possible solution. However, optimizing both coverage and efficiency is a bi-criterion optimization problem, with no single optimal solution.

We need to select one of the two metrics to optimize. In most cases, perfect efficiency is not essential. There may exist a few sentences in a review that do not cover any tip on their own accord, but their presence may improve the readability of the review. It suffices to ensure that the efficiency does not fall below a certain minimum acceptable threshold. Therefore, we opt to view our problem as a maximization problem, where we constrain the efficiency, and we ask for a solution with maximum coverage.

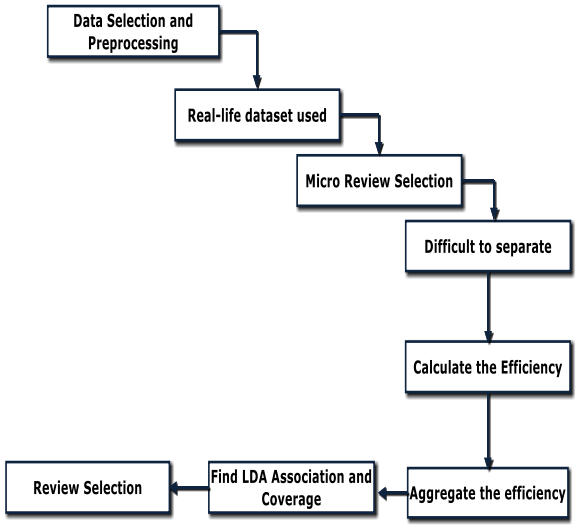

Fig.1. System Architecture

B. Proposal

Steps for selecting the reviews

Step 1: Data Selection and Preprocessing
Step 2: Micro Review Selection
Step 3: Calculate the Efficiency
Step 4: Find LDA Association and Coverage
Step 5: Review Selection

C. 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, we break a review into sentences, which are semantic units with granularity similar to that of the tips.

We view a review R as a set of sentences

$R= \{S_1,\ldots..S_{|R|}\}$

$F:U_S*T \rightarrow \{0, 1\}$

F(s, t) ={1, if s & t are similar,
0, otherwise

D. Selection Coverage

If a sentence s and a tip t are matched, then we say that scovers t. We will say that a review R covers a tip t if there isa sentence s ∈ R that is matched to the tip t. Given thcollectionof reviews R and the collection of tips T , and the matching nction F, we define for each review R the set oftips T R that are covered by at least one sentence of review R. Formally*:*

$T_R = \{t \in T : \exists s \in R, F(s,t) = 1\}$

We say that R covers the tips in T_R. We define the coverage Cov(R) of review R as the fraction of tips $|T_R| = |T|$ coveredby the review R.We can extend this definition to the case of a collection ofreviews. For a set of reviews S ≤R, we define the coverageof the set S as:

$$Cov(S) = \frac{|U_{R \in S} T_R|}{|T|}$$

that is, the fraction of tips covered by the set S.

IV. MODULE DESCRIPTION

E. Data Selection and Preprocessing:

In this module, the system will select the dataset and will preprocess the dataset. The system requires data coming from two different sources, concerning the same set of entities. The system describes the real-life dataset used in the experiment. This system gets the full set of reviews and tips of each restaurant at the time of extraction, and that these are the realistic sizes of the real-world data. It is also important to note that every restaurant is a distinct instance of the coverage problem. For reviews, the system crawlYelp.com to obtain the reviews of the top 110 restaurants with the highest number of reviews. For micro-reviews, the system crawl the popular check-in site Foursquare.com to obtain the tips of the same110 restaurants.

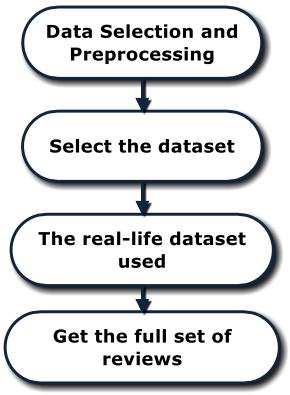

Fig.2. Data Selection and Preprocessing

F. Micro Review Selection:

In this module, the system seeks to represent micro-reviews, rather than attributes. Micro-reviews are emerging as a new type of online review content in the social media. Micro-reviews serve as an alternative source of content to reviews for readers interested in finding information about a place. Micro-reviews is a source of content that has been largely. The system first to mine micro-reviews such as foursquare tips and combine them with full-text reviews such as Yelp reviews. It is difficult to separate “reviews” from other types of content.

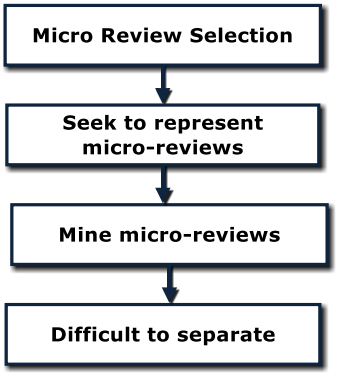

Fig.3. Micro Review Selection

G. Calculate the Efficiency:

The system introduces the notion of selection efficiency, which captures the principle that the selected set should not contain too many sentences that do not cover any tip. Some reviews may have high coverage, but at the same time they are too verbose, containing many sentences that are not relevant to any tip at all. The system would like to avoid such reviews in our selection, so the system introduces the concept of efficiency. Extending the definition of efficiency to a collection of reviews is a little more involved. The system needs a way to aggregate the efficiency of the individual reviews. For instance, by requesting that the minimum efficiency is above some threshold, we gain a guarantee that all reviews in the set obey the threshold.

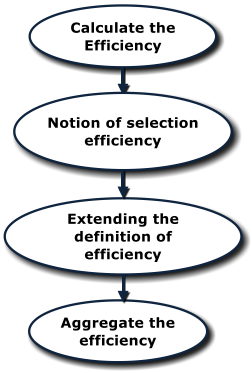

Fig. 4. Calculate the efficiency

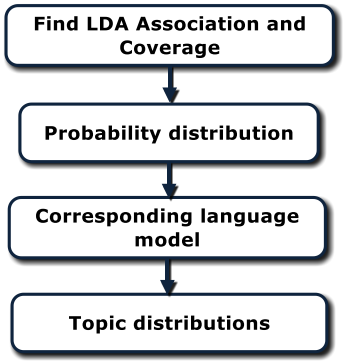

Fig. 5. LDA Association and Coverage

H. Find LDA Association and Coverage:

The system describes an approach based on the well-known Latent Dirichlet Allocation (LDA). LDA associates each tip t with a probability distribution θt over the topics, which captures which topics are most important for t. Given the topics, and the corresponding language model for each topic as it is learnt from the tips, we can estimate the topic distribution θs for each review sentence s, which captures how well a sentence s reflects the topics being discussed in the corpus of tips. To measure the semantic similarity between a review sentence and a tip, we measure the similarity of the topic distributions θs and θt.

I. Review Selection:

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. Similarly for review aggregation sites such as Google Local. To perform the selection, we need to determine when a review R∈R covers a tip t∈T. We refer to this procedure as matching reviews and tips. Given the matching, we then select a small subset of reviews that cover as many tips as possible. We refer to the number of covered tips as the selection coverage.

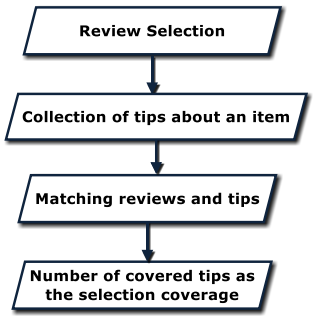

Fig.6. Review Selection

V. CONCLUSION

In this paper, we introduce the use of micro-reviews for finding an informative and efficient set of reviews. This selection paradigm is novel both in the objective of micro-review coverage, as well as in the efficiency constraint. The selection problem is shown to be NP-hard, and we design a heuristic algorithm EMaxCover, which lends itself to several definitions of aggregate efficiency. The results are evaluated over corpora of restaurants' reviews and micro-reviews. Experiments show that EMaxCover discovers review sets consisting of reviews that are compact, yet informative. Such reviews are highly valuable, as they lend themselves to quick viewing over mobile devices, which are increasingly the predominant way to consume Web content.

REFERENCES

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

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

[3] C .Cheng , H. Yang , I . King, and M. R. Lyu, "Fused matrix fact or iz a t i on w i t h g e og r a p h ic a l a nd so c i a l infl u e n c e in l o c a t i on -b a se d so c ia l ne t w or k s, "i n Proc. 26 t h A AA I Conf. Artif. Intel l. ,20 12, p. 1.

[4] K . G a n es a n, C . Z h a i, a n d J . Ha n , "Op i no si s : A g r a p h - b a s e d ap p r oa c h to a b s t r a c t ive su m m a r iz a t i on o f h ig hl y redundant opinions, " in Proc.23rdInt.Conf. Comput.Linguistics,2010,pp. 340–348.

[5] K . G a ne s an , C . Z ha i , an d E . Vi e g a s, "M ic r o p ini o n g e n e r a t ion :A n u n su p e r v ise d ap p r oa c h t o g e ne r a t i ng u lt r a - c o ncisesu m ma -r ie s of op in io n s, " i n Pro c . 21st Int. Conf. World W ide We b, 2012,pp. 869–878.

[6] H. G a o, J . T a n g , X . H u , a nd H . L iu , "E x p lo r i ng t e m p o r a le f f e c t s f or l oc a t io n r e c o mm e n d a t io n o n lo c ati o n- b a se d so c ia lne t w or k s, " i n P r o c . 7 t h AC M C o n f . R e co m m e n d e r S y st . , 2013,pp. 93–100.

[7] A. Ghose and P. G. Ipeirotis, "Designing novel review rankingsystems: Predicting the usefulness and impact of reviews," inProc. 9th Int. Conf. Electron. Commerce, 2007, pp. 303–310.

[8] M. Hu and B. Liu, "Mining and summarizing customer reviews,"in Proc. 10th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining ,2004, pp. 168–177.