ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 4, Issue 12, December 2016

Optimal Solution Generation from Micro-Reviews using Greedy Approach

Sarika S. Hulyalkar, Prof. H. K. Khanuja

Department of Computer Engineering, Marathwada Mitra Mandal's College of Engineering, Pune, Maharshtra, India

ABSTRACT: Survey content, and the truth that reviews are exceedingly different and frequently unnecessarily verbose, clients habitually face the issue of claiming and selecting those suitable reviews on devour. Micro-reviews are rising similarly as another kind about web survey substance in the online networking. Micro-reviews would be presented towards the clients as check-in benefits. They would be brief (up to200 characters long) and profoundly focused, as opposed of the far reaching what's more verbose reviews recommend. A novel mining problem is proposed for this, which acquires together these two divergent sources of Audit content. Specifically, we utilize scope of micro-reviews concerning illustration a destination for selecting a situated of reviews that disguise proficiently those remarkable viewpoints for a substance. Our approach comprises of a two-stage process: matching audit penalties to micro-reviews, what's more selecting a little set about reviews that disguise as a number of micro-reviews concerning illustration possible, with couple penalties. We define this goal likewise a combinatorial streamlining problem, and hint at how to infer an ideal result utilizing basic straight customizing. We additionally recommend a proficient heuristic calculation that approximates the optimal results. Finally, contributing an approach for implementing a system which discovers query facets by aggregating frequent lists within the top results.

KEYWORDS: Coverage, Micro-Review, Query Facets, Review Selection, Semantic Similarity and Sentiment Similarity, Syntactic Similarity

I. INTRODUCTION

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. Users check in at local venues, such as restaurants, bars or coffee shops. After checking in, a user may choose to leave a message, up to 200 characters long, about their experience, effectively a micro-review of the place.

In the case of restaurants, tips for a popular burger joint in Pune: "*Veggie burger is too good. TheWhooper burger is a big win and the star here*" (A recommendation), "*This is by far the best fast food burger joint in the city!!*" (An opinion), and "*Ideal outlet, ample space. If u are thinking of visiting which outlet to visit then this is the one. The burgers are priced accordingly.*" (An actual tip).

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. Second, because some micro-reviews are written on site, right when the user has checked in, they are spontaneous, expressing the author's immediate and unadulterated reaction to her experience. 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, if necessary, to filter out tips without an accompanying check-in, thus, boosting the authenticity of the tips.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 4, Issue 12, December 2016

II. LITERATURE SURVEY

Sr. No.	Author, Title and Journal Name	Advantages	Disadvantage	Refer Points
1	Thanh-Son Nguyen, Hady W. Lauw, Panayiotis Tsaparas," Review Selection Using Micro-Reviews", in Proc. IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 27, NO. 4, APRIL 2015	1.Micro-reviews are concise and distilled 2.They are reactive and spontaneous 3. Gathered from the onsite check-ins 4.They are more authentic as authors are currently at the specific location 5.Provides valuable feedback	1. Finding the optimal solution is not tractable for very large problem sizes. 2. Difficult to determine whether a review has been written by a genuine customer, or by a spammer	1. The paper uses coverage of micro-reviews as an objective for selecting a set of reviews that cover efficiently the salient aspects of an entity. 2. The 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.
2	T. Lappas, M. Crovella, and E. Terzi, "Selecting a characteristic set of reviews," in Proc. 18th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining, 2012, pp. 832–840.	1. To accurately emulate the opinion distribution in the underlying corpus. 2. Improvement by previous work	1. Positive and negative comment can't generalize toarbitrary domain.	1. In this paper, the gap is filled between existing review-summarization and review selection methods by selecting a small subset of reviews that together preserve the statistical properties of the entire review corpus. 2. The proposed three heuristic algorithms for selecting a characteristic review set, which we evaluate via experiments on a wide rangeof datasets of real reviews from different domains.
3	T. Lappas and D. Gunopulos, "Efficient confident search in large review corpora," in Proc. Eur. Conf. Mach. Learn. knowl. Discovery Databases: Part II, 2010, pp. 195–210.	1. To efficiently search through a large corpus. 2. A compact set of high-quality reviews that accurately captures.	1. Problem is this system work on artificial review.	1. In this paper, formalize the Confident Search paradigm for review corpora. 2. CREST (Confident REview Search Tool), a user-friendly implementation of our framework and a valuable tool for any person dealing with large review corpora.
4	P. Tsaparas, A. Ntoulas, and E. Terzi, "Selecting a	1. Performance is statically significant.		1. In this paper we consider the review set

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 4, Issue 12, December 2016

	comprehensive set of reviews," in Proc. 17th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining, 2011, pp. 168–176.	2. High quality of the review.		selection problem where given a set of reviews for a specific item, we want to select a comprehensive subset of small size. 2. Provides authentic review using TOPQLTY algorithm sorting technique problem is based on limited review set.
5	H. Lin and J. Bilmes, "Multi-document summarization via budgeted maximization of submodular functions," in Proc. Human Lang. Technol.: Annu. Conf. North Amer. Chapter Assoc. Comput. Linguistics, 2010, pp. 912–920.	1. The best performance is achieved. 2. Submodular summarization achieves better ROUGE-1 scores.	1. The proposed system very expensive to solve.	1. In multi-document summarization, redundancy is a particularly important issue since textual units from different documents might convey the same information. 2. A high quality (small and meaningful) summary should not only be informative about the remainder but also be compact (non-redundant).
6	Y. Lu, P. Tsaparas, A. Ntoulas, and L. Polanyi, "Exploiting social context for review quality prediction," in Proc. 19th Int. Conf. World Wide Web, 2010, pp. 691–700.	1. Improve the accuracy of review quality prediction. 2. The resulting predictor is usable even when social context is unavailable.	1. A portal may lackan explicit trust network.	1. There are two methods for incorporating social context in the quality prediction: either as features, or as regularization constraints, based on a set of hypotheses. 2. The method we propose is quite generalizable and applicable for quality (or attribute) estimation of other types of user-generated content.
7	P. Sinha, S. Mehrotra, and R. Jain, "Summarization of personal photologs using multidimensional content and context," in Proc. 1st ACM Int. Conf. Multimedia Retrieval, 2011, p. 4.	1. The greedy algorithm for summarization performs better than the baselines. 2. Summaries help in effective sharing and browsing of the personal photos.	1. Computation is expensive.	1. Proposed methods to compute quality, diversity and coverage properties using multidimensional content and context data. 2. The proposed metrics which will evaluate the photo summaries based on their representation of the larger corpus and the ability to satisfy user's

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 4, Issue 12, December 2016

				information needs.
8	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.	1. It solves the optimization problem efficiently. 2. The summaries to be very readable,fairly non-redundant and informative.	1. It only uses the existing text.	1This paper presents a new unsupervised approach to generating ultra-concise summaries of opinions. 2. We measure representativeness based on a modified mutual information function and model readability with an n-gram language model.
9	E. Kouloumpis, T. Wilson, and J. Moore, "Twitter sentiment analysis: The good the bad and the omg," in Proc. 5th Int. Conf. Weblogs Social Media, 2011, pp. 538–541.	1.The best performance on the evaluation data comes from using then-grams together with the lexiconfeatures and the microblogging features. 2. Average accuracy		1. In this paper, we investigate the utility of linguistic features for detecting the sentiment of Twitter messages. 2. We take a supervised approach to the problem, but leverage existing hashtags in the Twitter data for building training data.

III. EXISTING SYSTEM APPROACH

Online review 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. 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. The Opinions expressed or the reviews written by the people about a particular products or organization or entity of concern contains important information that is very helpful to the person or the organization for their future action on improvement and development.

Disadvantages of Existing System:

1. The efficiency increases while the coverage decreases.
2. Finding the optimal solution is not tractable for very large problem sizes.
3. Reviews getting increasingly more difficult to determine whether a review has been written by a genuine customer, or by a spammer.

IV. PROPOSED SYSTEM APPROACH

Our work introduces a novel formulation of review selection, where the goal is to maximize coverage while ensuring efficiency, leading to novel coverage problem. The coverage problem we consider are of broader interest, and they could find applications to different domains. We consider approximation and heuristic algorithms, and study them experimentally, demonstrating quantitatively and qualitatively the benefits of our approach. We also propose an Integer Linear Programming (ILP) formulation, and provide an optimal algorithm. This allows us to quantify the approximation quality of the greedy heuristics.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 4, Issue 12, December 2016

The proposed a novel mining problem, which brings together these two unequal sources of review content. We use coverage of micro-reviews as an objective for selecting a set of reviews that cover efficiently the salient aspects of an entity. In this paper we are proposing k-NN Classification algorithm to find the nearest distance items for selecting the reviews which are given in the Microblogs. A more sophisticated approach, K-Nearest Neighbor (k-NN) classification, finds a group of k objects in the training set that are closest to the test object, and bases the assignment of a label on the predominance of a particular class in this neighborhood. Among them, the simple and efficient method for the implementation and understanding of non-parameterized classification was the K-Nearest Neighbor (k-NN) which has been well-received. For instance, in order to improve the classification accuracy, in addition to local information from k-NN of new non-classified data, general information about neighbors in each class is analyzed separately.

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.

Our contribution part is, when the user searches hotels in a particular area then the list of hotels are generated, we evaluate the importance of these hotels and menus, and rate them based on their importance. Routes to the selected hotels will be shown to the user using locations. Applying the top-k algorithm, which is based on our motivation that good hotels should frequently appear in the top results which we will get from rating of reviews results. The lists of hotels are extracted from more unique content of search results; and the hotels which are famous, will have higher weights. Also, the system should be recommending the hotel for particular interest of user. Collaborative filtering methods are based on collecting and analyzing a large amount of information on users' behaviors, activities or preferences and predicting what users will like based on their similarity to other users.

Advantages of Proposed System:

1. KNN classification is easy to understand and easy to implement classification technique.
2. Virtually identical to the optimal solutions in coverage and efficiency, but it is much faster computationally.
3. Find a small number of reviews that best cover or represent the tips.
4. Longer reviews are weighted more than shorter reviews in the aggregate efficiency.
5. The performance results of the proposed model were best ones reported as the state-of-the-art classifiers in terms of classification accuracy for the same data sets.
6. Use of query facets is going to help users choose hotels according to their taste and choice by assigning weights to famous hotels

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 4, Issue 12, December 2016

V. SYSTEM ARCHITECTURE

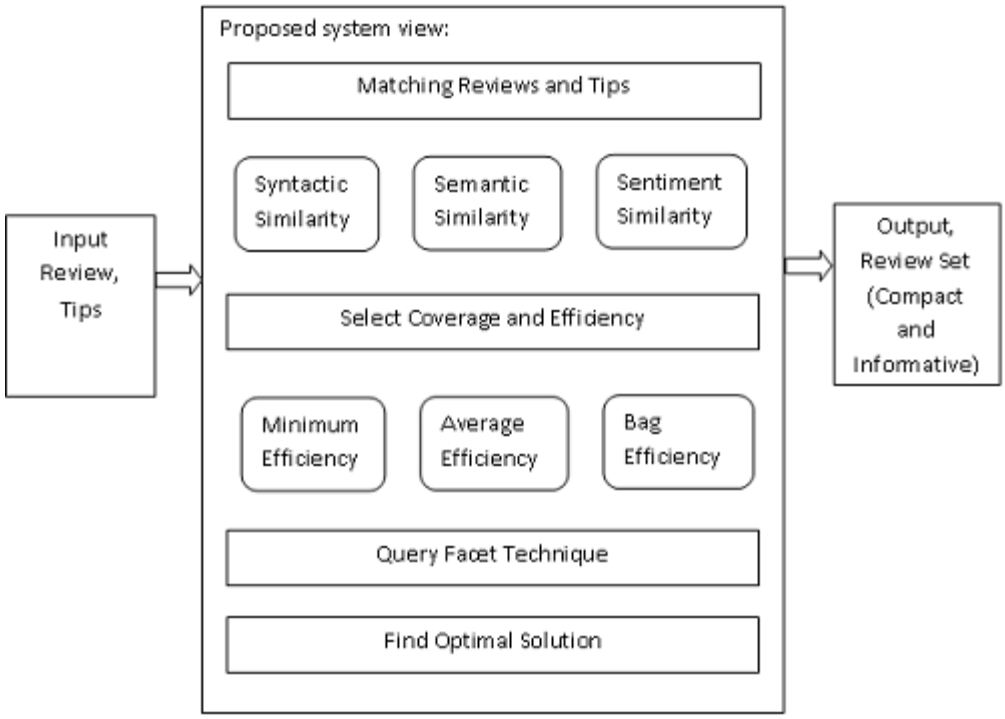

Fig.1. Block Diagram of Proposed System

VI. 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 EffMaxCover, which lends itself to several definitions of aggregate efficiency. The results are evaluated over corpora of restaurants' reviews and micro-reviews. Experiments show that EffMaxCover 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]Thanh-Son Nguyen, Hady W. Lauw, Panayiotis Tsaparas," Review Selection Using Micro-Reviews", in Proc. IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 27, NO. 4, APRIL 2015
[2] T. Lappas, M. Crovella, and E. Terzi, "Selecting a characteristic set of reviews," in Proc. 18th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining, 2012, pp. 832–840.
[3] T. Lappas and D. Gunopulos, "Efficient confident search in large review corpora," in Proc. Eur. Conf. Mach. Learn. knowl. Discovery Databases: Part II, 2010, pp. 195–210.
[4] P. Tsaparas, A. Ntoulas, and E. Terzi, "Selecting a comprehensive set of reviews," in Proc. 17th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining, 2011, pp. 168–176.
[5] H. Lin and J. Bilmes, "Multi-document summarization via budgeted maximization of submodular functions," in Proc. Human Lang. Technol.: Annu. Conf. North Amer. Chapter Assoc. Comput. Linguistics, 2010, pp. 912–920.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 4, Issue 12, December 2016

[6] Y. Lu, P. Tsaparas, A. Ntoulas, and L. Polanyi, "Exploiting social context for review quality prediction," in Proc. 19th Int. Conf. World Wide Web, 2010, pp. 691–700.
[7] P. Sinha, S. Mehrotra, and R. Jain, "Summarization of personal photologs using multidimensional content and context," in Proc. 1st ACM Int. Conf. Multimedia Retrieval, 2011, p. 4.
[8] 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.
[9] E. Kouloumpis, T. Wilson, and J. Moore, "Twitter sentiment analysis: The good the bad and the omg," in Proc. 5th Int. Conf. Weblogs Social Media, 2011, pp. 538–541.