Opinion Mining for Social Networking Site

K.Dhurka
Research Scholar, Department of Computer Science
Mazharul Uloom College, Ambur, Tamil Nadu, India

J.Aamir Azeez
Assistant Professor, Department of Computer Science
Mazharul Uloom College, Ambur, Tamil Nadu, India

OPEN ACCESS

Volume: 6

Issue: 2

Month: October

Year: 2018

ISSN: 2321-788X

Received: 20.10.2018

Accepted: 27.10.2018

Published: 29.10.2018

Citation:
Dhurka, K., and J. Aamir Azeez. "Opinion Mining for Social Networking Site." *Shanlax International Journal of Arts, Science and Humanities*, vol. 6, no. 2, 2018, pp. 87–92.

DOI:
https://doi.org/ 10.5281/ zenodo.1473533

Abstract
This system uses opinion mining methodology to achieve target oriented. Opinion Mining for Social Networking Site is a web application. Data mining in the proportion of opinion is a constructed we application. Hence fourth it is an effective tool in a friendly user way to post the comment of users. This options give all user in a variety of each. It helps to identify the ratings of the customers/consumers about the product or service viz good, fair, poor, etc . These comments used to generalized by global level. A data base is accumulated and discriminated both positive and negative. It helps all users, to access others comment without any restriction. It provides awareness as well a new dimension about the organization and product. It enriches reviews and new light in criticism.

Keywords: Opinion mining, sentiment analysis, opinion summarization, Social Networking Sites

Introduction

Great expansion and development of social network media (e.g., Whatsapp, Instagram) on the online everyone takes part in a discussion and posts their views. This review helps the needy who plan to purchase a product. The mammoth development of web digital powered by Web 2.0 (DiNucci, 1999) has made information available more than ever before and hence people now increasingly take their required information from one another rather than from corporations, media outlets, religion or political bodies. WWW has become the most popular social media which covers almost all form of sharing such as experiences, photos, recommendations. To do this, people get involved in social networks formed by friend lists, by the bloggers who comment/rate on a particular topic in the blog space, or by the users who write collaboratively in a wiki site (e.g., Wikipedia, Scholarpedia). People may give their opinions on the shared posts; those opinions may be positive, negative, or controversial about the product. Several research (Pang et al. 2002, Turney 2002, Agrawal et al. 2003, Dave et al. 2003, Hu and Liu 2004, Mishne and Glance 2006, Nigam and Hurst 2006, Ding et al. 2008, Gomez et al. 2008, Li et al. 2008, Tang et al. 2009) have been done for analyzing users' opinions on interest networks (i.e., user-service interaction), but based on our knowledge, no work is found in friendship networks (i.e., user-user connections). Such a user-service network are domain specific and product-feature oriented. For example, these networks may be weblogs, newsgroups, bookmarks, question/answers, movie/product review domains, as opposed to friendship networks.

Related Work

Social network analysis often focuses on macro-level research such as degree distribution, diameter, clustering coefficient, community detections, small-world effects, opted attachments, etc. (Tang et al., 2009). Recently, many researchers have analyzed social network data to find patterns of popularity or influence in various domains. Such domains include blogging (e.g., Slashdot.org) and micro-blogging (e.g., Twitter.com) domains, bookmarking domains (e.g., Digg.com), co-authorship domains (e.g., Academia.edu), movie review domains (e.g., IMDb.com), and product review domains (e.g., Amazon.com). Weblog domains define a relationship between the writer of the blog and the readers by publishing short news posts and allowing readers to comment on them. In co-authorship domains, each author is related to some specific topic; there is no random author-topic relation. Movie review domains provide ratings and brief quotes from several reviews and generate an aggregate opinion. Product platforms are designed for specific types of products. All the domains are well-structured for a specific topic whereas friendship network is more complex and heterogeneous. Moreover, the great majority of research study only features related to the network itself or simple popularity matrices of the posts (e.g., number of likes/thumbs-up, number of comments), without analyzing the correlation of these aspects with the content of the posts.

Table 1 Differences between issues handling by existing systems and proposed system

Existing Systems	Type of Network	Size of Products/Opinions	Measurements	Limitations
General IM CELF (Leskovec et al., 2007)	Social network (user-user network) (e.g., Facebook.com)	All users post about multiple products on multiple posts	A probability of users performing actions after an influential user	1. Does not consider, opinion' of users 2. Not product specific 3. Not scalable
General IM T-IK (Ahmed and Ezeife, 2013)	Social network (Trust network) (e.g., Wikipedia.com)	All users post about multiple products on multiple posts	A probability of users performing or not performing actions (+/-) by influential users	1. Does not consider 'opinion' of users 2. Positive/negative influences are explic itly given 3. Not product specific 4. Not scalable
General Opinion Mining opinion ner (Jin et al., 2009)	Domain-specific websites (user – service network) (e.g., Amazon.com)	One user posts about one product on a single post page	Comments and ratings on the product	1. Predefined product features are have explained 2. Ignore opinions about different products_

Opinion Miner takes product features as input parameters. Features are domain free, a set of options must be prepared. For example, if the system wants to extract opinion about Digital Cameras, design ideas as cover color, pixel ratio, zoom, memory, etc., and tag the reviews accordingly. The system mines opinions for reviews that have predefined product features. Moreover, the system does not consider options expressed on irrelevant product entities. For example, Samsung Galaxy page containing any review about iPhone will not be accepted as the opinion for Galaxy.

Proposed System

1. First, to consider opinions on friendship network for specific product

a) A new influence network (IN) generation model is proposed, called OBIN, Opinion Based Influence Network.

i. OBIN considers multiple posts by multiple users on a specific product
ii. OBIN aggregates all kinds of users' explicit/implicit opinions (e.g., likes/dislikes, re-shares, positive/negative comments)
iii. OBIN discovers users-users relationships

2. We propose a local search algorithm, called TPD (Topic-Post Distribution) based on network pruning strategy to find ranked list of users and opinions and to classify relevant and irrelevant users for a specific product.
3. We propose PCP-Miner (Post-Comment Polarity Miner) algorithm, to compute the popularity scores of users by extending Opinion Miner (Jin et al., 2009) with Apriori frequent pattern mining, and to calculate the influence scores of users to discover user-user relationships.
4. Experimental analysis shows that OBIN gives relevant, influential users for a product more efficiently, and the influence spread over the network is occurred more effectively than standard IM algorithms.

System Architecture

Browser

Anyone can use and access the site for the better use of the same. The proposed project must be accessed only when the user posses enabled a particular application for browse. The needy should inquire over by a search engine with relevant key words.

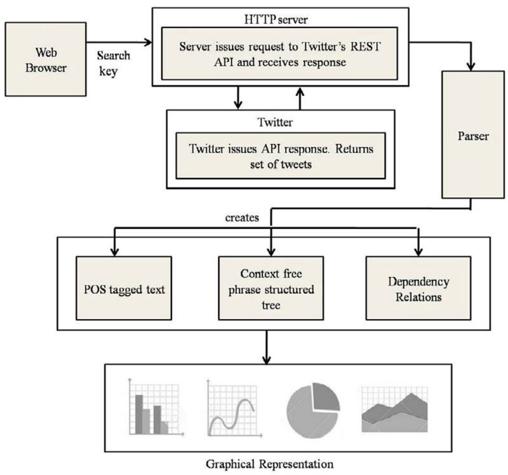

Figure 1: System architecture

Implementation

Proposed OBIN takes a social network graph G(V, E) and a product name z as input to generate an influence graph GZ(V, E) on product z from computed community preference where V is the relevant nodes extracted from G. Algorithm 1 shows the algorithm for OBIN model. OBIN has three main functions, TPD (Topic-Post Distribution), PCP-Miner (Post-Comment Polarity Miner), and influence network generator. OBIN first executes SQL queries on social network URL to extract nodes (Vs) on a product z, and then classify relevant and irrelevant nodes. This process is over by TPD method (lines A.1-A.4 in Algorithm. 1). Then PCP-Miner (lines B.1-B.2 in Algorithm. 1) takes the ranked relevant nodes, posts, and comments from TPD to identify opinion (positive or negative) comments and compute the polarity score (9Z) for each similar post. Based on the polarity score, OBIN generates an influential network that represents the community preference for the product z (line C.1-C.2 in Algorithm. 1).

Our proposed solution framework for social and opinion posts mining for community preference discovery is visualized in Figure 13. Following are the inputs to the categorized work:

1. Social network URL (e.g., Facebook.com), and topic z (e.g., iPhone).
2. Predefined threshold - Approve (A) which is the minimum number of nodes (vt) that have to be linked to the node (v) who posted the topic-post.
3. Predefined threshold - Simple response (SR) which is the minimum number of posts (w) the node (v) has to post on topic z.
4. The intermediate inputs are listed below:
5. Predefined threshold - Approve (A) which is the minimum number of nodes (vt) that have to like the topic-post (w) that is posted by node v.
6. Predefined threshold - Simple response (SR) which is the sum of the total number of unique comments (nCy) on the topic-post (w) and the total number of re-shares of the post (w) by the nodes vt.
7. Part-of-speech tag list - POS-tags from a predefined list on Table 28 - to identify a syntactic orientation of words
8. WordNet list - from (http://www.princeton.edu/wordnet/download/current-version/) to identify synonyms and antonyms

The proposed solution consists of the following four steps listed below:

Step l: At first our proposed solution framework OBIN calls TPD to extract relevant nodes v €V for a topic z and filter them according to higher influential score determined by Approve A and Simple Response SR. Lines A.1 to A.4 in Algorithm 1 shows the steps for our proposed model TPD. TPD then extracts and filters relevant posts w €W for each correct node v. Detailed steps of these processes using TPD model with algorithm and examples have pointed out in section 3.4. The resultant data kept into our transactional database for next steps.

Step 2: In this second step, our solution framework OBIN calls PCP-Miner to fetch all the opinions for each relevant post w of each appropriate node v, and apply sentence and word segmentation and some cleaning such as stemming, string matching, etc. Lines B.1 to B.2 in Algorithm 1 shows the processing steps for PCP-Miner. For each opinion sentence in the opinion text, our proposed PCP-Miner apply POS-tagging (Brill 1994) to identify adjective, adverb as opinion words and noun, noun phrase as features. Then list out the polarity of the comment, i.e., the comment expressing a positive or negative opinion. And finally, compute the popularity of the relevant post w. Detailed processes are given in section 3.5 with algorithm and examples.

Step 3: In this step, our solution framework store all the extracted and computed data into our data warehouse for further mining purpose.

Step 4: After our previous steps, we have a ranked list of mined relevant nodes EV, their corresponding popular topic-posts EW, and aggregated opinions on each post along with their polarity (positive impact or negative impact). In this fourth step (lines C.1 to C.2 in Algorithm 1), our proposed solution framework OBIN calls PoPGen model to identify the relationships among nodes vt and Vj (vitVj E V and i,j = 1,2,..., N on a topic z and how they influence to each other. Our proposed solution also identifies a global relation between nodes vital Vj for a similar topic, hence discover the community preference.

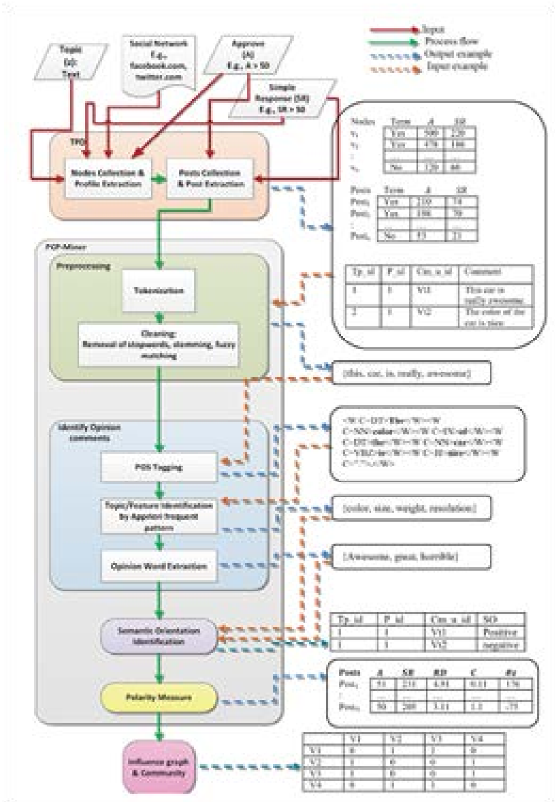

Figure 2 System Diagram of OBIN

Performance Analysis

Table 2 showed the accuracy measure of CELF and T-IM and proposed OBIN. We can see that the recall value of OBIN is 93.7%, this is because 90 opt nodes (out of 2407 relevant and irrelevant nodes) were not extracted by OBIN, and also 26 more nodes might be suitable but could not extract due to information in a language other than English. Precision is 98.24%, this is because 31 inappropriate nodes are elicited (out of 2407 nodes) by OBIN. With the same dataset, we applied CELF and T-IM and observed that OBIN is dramatically better in precision and F-score with a slight loss in recall.

	Precision	Recall	F - score
CELF	80.02%	92.7%	85.4%
T-IM	81.36%	96.09%	88.1%
OBIN	98.24%	93.71%	95.3%

Conclusion and Future Works

In this research, we proposed an effective method for discovering relevant, influential nodes from friendship network which enables more focused target marketing than existing IM algorithms. However, previous research considers opinion mining only in a user-service pool of single product page, where OBIN mines opinions from complex user-user relationship network of multiple posts, multiple products, considering both implicit and explicit ideas. Experimental results show that the proposed technique performs markedly better than the existing general IM methods. Moreover, the information extracted and computed from the friendship network further can be applied to provide recommendation systems to improve business opportunity. The resultant data stored in the data warehouse can also answer some crucial business queries such as “which relevant post is most popular?”, “who are the most influential and influenced users on the post?”, “who like the product and who do not,” “how do the users connect?”.

However, in future, we would like to apply further techniques learned in influence network generation with Google page ranking algorithm, that could result in new insights into the influence maximization problems.

References

Agrawal, R., Rajagopalan, S., Srikant, R., & Xu, Y. (2003, September). Mining newsgroups using networks arising from social behavior. In Proceedings of the 12th international conference on World Wide Web (pp. 529-535). New York, USA: ACM.

Bonchi, F., Castillo, C., Gionis, A., & Jaimes, A. (2011, April). Social network analysis and mining for business applications. ACM Transactions on Intelligent Systems and Technology (TIST), 2(3), 22.

Dave, K., Lawrence, S., & Pennock, D. (2003, May). Mining the peanut gallery: Opinion extraction and semantic classification of product reviews. In Proceedings of the 12th international conference on World Wide Web (pp. 519-528). Budapest, Hungary: ACM.

Ding, X., Liu, B., & Yu, P. (2008, May). A holistic lexicon-based approach to opinion mining. In Proceedings of the international conference on Web search and web data mining (pp. 231-240). Budapest, Hungary: ACM.

DiNucci, D. (1999). Fragmented future. Print, 53(4), 32.

Gomez, V., Kaltenbrunner, A., & Lopez, V. (2008, April). Statistical analysis of the social network and discussion threads in slashdot. In Proceedings of the 17th international conference on World Wide Web (pp. 645-654). Beijing, China: ACM.

Hu, M., & Liu, B. (2004, August). Mining and summarizing customer reviews. In Proceedings of the 10th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 168-177). Seattle, Washington, USA: ACM.

Li, X., Guo, L., & Zhao, Y. (2008, April). Tag-based social interest discovery. In Proceedings of the 17th international conference on World Wide Web (pp. 675-684). Beijing, China: ACM.

Mishne, G., & Glance, N. (2006, May). Leave a reply: An analysis of weblog comments. In Proceedings of the third annual workshop on the WWW 2006 Weblogging Ecosystem: Aggregation, Analysis and Dynamics (pp. 22-26). Edinburgh, UK: WWW06.

Nigam, K., & Hurst, M. (2006). Towards a robust metric of polarity. Computing Attitude and Affect in Text: Theory and Applications, 265-279.

Pang, B., Lee, L., & Vaithyanathan, S. (2002, July). Thumbs up?: Sentiment classification using machine learning techniques. In Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10. 10, pp. 79-86. Philadelphia: Association for Computational Linguistics.

Tang, J., Sun, J., Wang, C., & Yang, Z. (2009, June). Social influence analysis in large-scale networks. In Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 807-816). Paris, France: ACM.

Turney, P. (2002, July). Thumbs up or thumbs down?: Semantic orientation applied to unsupervised classification of reviews. In Proceedings of the 40th Annual Meeting on Association for Computational Linguistics (pp. 417-424). Philadelphia: Association for Computational Linguistics.

Web Sources

http://nevonprojects.com/opinion-mining-for-social-networking-site/

https://www.dbmi.pitt.edu/person/shyam-visweswaran-md-phd

https://research.yahoo.com/publications/6144/social-network-analysis-and-mining-business-applications

https://www.dbmi.pitt.edu/person/shyam-visweswaran-md-phd

https://dl.acm.org/citation.cfm?id=1367585

http://web.eecs.umich.edu/~twenisch/papers/wddd04.pdf

https://www.barnesandnoble.com/w/computing-attitude-and-affect-in-text-james-g-shanahan/1111362924

http://www.mdpi.com/2073-8994/8/9/89/htm