Volume 4, Issue 11, November 2014 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

An Extrinsic Approach for Detaining the Subjective Relevant Events Based on User's Interestingness Measure

[1]R. Sridevi, [2]V. Saranya
[1]PG Scholar, [2]Asst. Professor
Dept of CSE, Sri Vidya College of Engineering and Technology,
Viruthunagar, TN, India

Abstract-Social media is an emerging trend for communication between people worldwide and it has attracted millions of users. A typical characteristic of such sites is that they allow anyone to post or like anything they like on any subject. Here the user's feedback and satisfaction is playing a vital role. In some cases the user may not give the feedback or reviews which they have viewed or downloaded directly. Instead of that the user is just searching the available things based on their interest. So there is no possibility for capturing the domain interest behavior of the user explicitly. We investigate the problem of predicting user's domain interest in social Medias, where we attempt to predict whether a user will be satisfied with domain suggestion explicitly. The domains may be sports, motivation, nature, positive quotes, cooking, healthcare, gardening, general knowledge, job recruitment, pencil arts, shopping, comedies, books, school memory, data mining, networks, cloud computing, grid computing, image processing, communication networks, mobile computing, and mechanical engineering. We find the maximum number of like in each domain and we recommend events based on that domain. We complement our results with a thorough investigation of the interactions and information patterns in their likes and shares that correlate with domain interest of a user.

Keywords: Data mining, social media, search engine , recommendation system , Events

I. INTRODUCTION

The social media has gained its popularity from the emerging new technology. An accumulation of technologies and techniques appeal to and make possible social interaction drive between humans. The user can create their own public profiles and can share the likes about the various domains in the social media around the world. The people can also create their own logs and have a discussion about the current affairs. The people around the world can communicate with one another by using the internet. The user can share information, likes and views to another. The search engine is the software program used to excerpt needed information. The search engine has the index for each and every keyword used by the user. The search engines maintain a large database for the web pages. The ranking by the search engine is based on the weight age of the keyword in the query. There are also tools used for the most effective and specific searches. The example of familiar search engine used is Google. It uses software spiders to extract the requested information. The search engine is used to hunt for the documents of the query. The query in the search engine should have keywords and the related information is excerpted. The information got from the search engine should be evaluated for correction. The user review is the specific instance for providing information. The user review has the major impact on the social media. The user review can be for software, hardware and other services. The user review can also be for the product or organization. The positive and negative reviews from the user are used for the improvement of the sites and their organizations. The user also gets benefit from the reviews in the social media. The reviews from the user now decide quality the services provided. The negative reviews are used for the betterment of the product.

The example of the product review can be form online shopping site. The review is got from the user by email and free calls. In this paper, the user review is excerpted and tidings about the user like and what they share and download are noted. In social media, users perform various actions within their social networks such as expressing their interests to a particular website in Facebook or retweeting a comment made by a friend on Twitter. The manners in which different users express their interests to extraction is to be very secular because the user likes and dislikes changes with time to time.. The domain is caught by the intrinsic and extrinsic behavior. The intrinsic behavior is domain dependent tiding excerption. As compared with extrinsic the intrinsic extraction is more convenient. But this paper is based on the extrinsic excerption of domain. The history of the user review is assessed. The history can be based on the particular range and the user reviews are excerpted.

The user history are analyzed and given as input to the Recommendation System. Recommendation systems or recommendation engines filter information to present items like consumer products, movies, music, books, news and images to you that are the most likely to be interesting. The recommender system will use details of the registered user's profile and opinions and habits of their whole community of users and compare the information to reference characteristics to present the recommendations. Recommender Systems (RS) typically apply techniques and methodologies from other neighboring areas – such as Human Computer Interaction (HCI) or Information Retrieval (IR).

However, most of these systems bear on their core an algorithm that can be understood as a particular instance of a Data Mining (DM) technique. In this paper the RS is based on the excerpted history from the user reviews. The pattern matching is also used for the meticulous verification of the suggested events. By using the history, the user likes and dislikes are excerpted and the events are recommended to the user. The events recommended are always intrinsic to the user because events are based on the meticulous observance form the user history. The events suggested to the user are very nifty.

II. BACKGROUND RESEARCH

OPINION MINING

Opinions and sentiments[2] expressed in text reviews can be generally analyzed at the document, sentence, or even phrase (word) levels. The purpose of document-level (sentence-level) opinion mining is to classify the overall Subjectivity or sentiment expressed in an individual review document (sentence).

Hatzivassiloglou and Wiebe [3] studied the effects of dynamic adjectives, semantically oriented adjectives, and gradable adjectives on predicting subjectivity; they proposed a supervised classification method to predict sentence subjectivity. Pang et al. a particular resource vary across social media. The user review has a high impact on both the social media and search engine. From the user review, the domain is extracted. The domain extraction is based on the weight age of keywords used in the user review and score are also calculated for the keywords. The domain[4] proposed three machine learning methods, naive Bayes, maximum entropy, and support vector machines, to classify the whole movie reviews into positive or negative sentiments. They found that standard machine learning techniques produced good results in comparison to human-generated baselines. Moreover, machine learning methods did not perform as well on sentiment classification as on a traditional topic based categorization. To prevent a sentiment classifier from considering irrelevant or even potentially misleading text, Pang and Lee [5] proposed to first employ a sentence-level subjectivity detector to identify the sentences in a document as either subjective or objective, and subsequently discarding the objective ones. They then applied the sentiment classifier to the resulting subjectivity extract, with improved results.

McDonald et al. [15] investigated the use of a global structured model that learns to predict sentiments on different levels of granularity for a textual review. The primary advantage of the proposed model is that it allows classification decisions from one level in the text to influence decisions at another. A regression method based on the bag of opinions model was proposed for review rating prediction from sparse text patterns [6]. Review rating estimation is a much more complicated problem compared to the binary sentiment classification. Generally, sentiments are expressed differently in different domains. The sentiment classification methods discussed above can be tuned to work very well in a given domain; however, they may fail in classifying sentiments in a different domain. Bollegala et al. [7] proposed a cross-domain sentiment classifier using an automatically extracted sentiment thesaurus. An unsupervised learning method was proposed to classify review documents as thumbs up (positive) or thumbs down (negative) in [8]. The sentiment of each review document is predicted by the average sentiment orientations of phrases in the review.

Domain-dependent Contextual information is also considered for better estimation of the phrase sentiments. One limitation of this work is its reliance on an external search engine. Zhang et al. [9] proposed a rule-based semantic analysis approach to classify sentiments for text reviews. They used ord dependence structures to classify the sentiment of a sentence, and predicted document-level sentiments via aggregating the sentence sentiments. Rule-based approaches like this typically suffer from poor coverage due to the lack of comprehensiveness in their rules. In addition, Maas et al. [10] presented an approach to document-level and sentence-level sentiment classification tasks, which uses a mix of unsupervised and supervised techniques to learn word vectors by capturing semantic term-document information as well as rich sentiment content. Differently, sentiment analysis at the phrase (word) level mainly focuses on classifying sentiment polarities of opinion phrases (words). Generally, the sentiment polarity of an opinion word is usually context-dependent as well as domain-specific. Wilson et al. [11] presented an approach to predicting contextual sentiments at the phrase level by applying machine learning techniques on a variety of feature factors.

Yessenalina and Cardie [12] presented a compositional matrix-space model for phrase-level sentiment analysis. One of the benefits of the proposed approach is that by learning matrices for words, the model can handle unseen word compositions (e.g., unseen bigrams) as long as the component unigrams have been learned. A two-level affective reasoning method was proposed to mimic the integration of conscious and unconscious reasoning to address word-level sentiment analysis tasks [13]. Note that opinion mining at the document, sentence, or phrase (word) level does not discover what exactly people liked and disliked in reviews. In other words, it fails to associate the identified sentiments to the corresponding features commented on in the reviews. Clearly, an extracted opinion without the corresponding feature (opinionated target) is of limited value in reality [1]. Next, we survey existing work on extracting opinion features.3. Recommendations in social networks there are findings in the sociological and psychological disciplines that point to the relevance of a person's social network in determining their tastes, preferences, and activities. The principle of Homophile, for instance, is well established in the Social Networks field. McPherson et al. reported how "similarity breeds connection". They discovered that "people's personal net- works are homogeneous with regard to many sociodemographic, behavioral, and intrapersonal characteristics". In other words, we share many attributes with the people close to us.

Reversing this principle suggests that, if we have information about the connections in a person's network, we can infer some of the person's attributes. It is possible that at least some of the similarities within a network are caused by the influence and interactions of the people in the network. People tend to remember information that was concretely given to them (that is, in personal interactions) better than abstract information (like statistical base rates). For example, Hogarth states that when considering to buy a certain car model we will likely give more thought to the direct advice of a friend than to each of the 100 respondents to a survey in a specialized magazine. More specifically, Leskovec et al. discuss the phenomenon of information cascades, in which individuals adopt a new action or idea due to influence by others. In the most extreme cases, knowledge about a full network's behavior determines the behavior of its members – making a "top hits" list available in a mu- sic downloading website affects the popularity of the songs, and several different networks, kept in isolation of each other, prefer completely different songs, to the point that it is impossible to predict which will be the most popular songs for a network without observing the behavior of the users in the network.

TRADITIONAL USER FEEDBACK MECHANISMS

Software vendors have to take into account that fast evolving technologies often imply frequent changes in requirements [4]. User feedback commonly plays an important role in requirements validation and verification. Some of the most common traditional user feedback gathering mechanisms met in our study include: email, survey, forum, feedback form, usability testing, and review meetings. Traditional usability testing for feedback gathering has many restrictions related to the scope of the functionality being evaluated, the size of the evaluation group, and the fact that users are not in there typical work environments during the usability tests [9]. During beta testing, users often fail to report bugs, thinking that they are part of the design [9]. Also, a developer's notion of usage expectations might not necessarily match the actual user behavior. While this form of end-user feedback solicitation might be useful for improving software usability, the collected feedback is not a collaborative activity.

SOCIAL MEDIA AND USER FEEDBACK

Social networks are being formed across the web through social media channels. With this form of online social interaction, knowledge sharing over the Internet has become a two-way street: people can extract the data over the web, as well as contribute their own content and collaborative share information with others. Examples of social media include: blogs (e.g., Word press, Blogger), forums (e.g., Yahoo Answers), wikis (eg, PBWorks), shared online

videos (e.g.,YouTube, Video), social networking (e.g., Facebook, LinkedIn), social bookmarking (e.g., Delicious, Google Reader), photography sharing (e.g., Flickr, Picasa), business reviews (e.g., Yelp), and virtual worlds (e.g., Second Life). Business needs are no longer the only force behind technology development - consumer voices are also being heard. In this paper we focus on how any of the above social media channels can be leveraged to gather user feedback.

ONLINE SOCIAL NETWORKS AND FEEDBACK

The communication flow within a community can either be one-to-one, one-to-many, or many-to-many [5]. In a one-to-one environment, conversation is visible only to those involved in it, e.g., email or a chat room. One-to-many cardinality describes hub-based communities, where all questions are aimed at a single participant (expert). In many-to-many communities all members have the ability to communicate with each other. In this paper, we consider a user feedback community to be one that uses either one-to-many or many-to-many communication flow types. Users can communicate with businesses either through a customer- or company-controlled environment [3]. An example of the former would be a third-party managed public forum where users freely exchange ideas; the business is able to join the community and collaborate with the users, but has no control over the environment. Our research focuses on company controlled user feedback gathering tools and their use.

ONLINE USER INNOVATION COMMUNITIES

Companies are taking advantage of Web 2.0 technologies by adapting open business models and engaging with their customers via User Innovation Communities (UIC's). One such example is Dell's Idea Storm [6]. Just like with any other online community, maintaining a successful UIC means having to overcome some major challenges such as: understanding the ideas posted, identifying the best ideas, balancing the needs of transparency, and sustaining the community. In [6], authors examine the challenges Dell's Idea Storm community faced during the first 18 months and conclude with suggestions for overcoming those such as understanding and identifying the best ideas, sustaining the community, and avoiding disclosure to competitors.

SUMMARIZING SOCIAL MEDIA CONTENT

Web 2.0 technologies have enabled users to generate online content by uploading text, images or video through various social web applications. User-generated content can be further classified as tags, bookmarks, comments, ratings, etc. Tags or folksonomies [7] can be thought of as labels assigned to digital information such as a video, an article, or a book review. Agichtein et al. [1] analyze these types of community content generation and propose a methodology for separating high quality content from the rest by building a graph-based model of contributors based on their usage behavior. In this paper, we analyze some barriers of parsing user feedback, and study how interactions from system administrators impact the development of online user communities.

OUR DOMAIN: BOARDGAMEGEEK

BoardGameGeek (BGG) is the most popular community for people that play board and card games. It hosts a database of more than thirty thousand games, along with their corresponding reviews, photos, rules Clarifications, ratings, player aids, and other user- generated content. Over forty thousand registered users have rated games in this website, and many of them have recorded some of their personal information in their profiles and made explicit their friendship links ('Geek Buddies', in BGG parlance) with other users. The relative openness of the BGG database (the web-site provides an XML API that gives access to some of its data), as well as its moderate size made it an appropriate choice to try our social-based recommendations algorithm.

DATA COLLECTION

We intended to obtain our data directly from the BGG system administrators. We e-mailed them twice, but received no response from them. Therefore, we decided to write a set of crawlers to gather the data. At this point, we have finished collecting the data from the BGG website.
We crawled for the following data:

- Games that any user has rated (about 30K)
- Users that have rated any games (about 40K)
- Ratings given by all users to any games
- Games in all users' wish lists, as well as the weight given to wish list items by each user
- Links between users ("Geek Buddies")

get the user information and ratings we used an API provided by the BGG website. However, getting the game and Geek Buddy data was more problematic, as it was only available to logged-in users. To overcome this our crawler had to "trick" the website, passing as a logged-in user. We sent requests to the BGG website every 2 seconds; a complete pass to crawl the data we need takes slightly more than 3 days if run sequentially. We saved the collected data in a database, and wrote scripts to generate flat text files with sparse representations of information to be used by the machine learning algorithm.

III. PROBLEM STATEMENT

The advertisements in the social media are sometimes irrelevant to the user .Sometimes the user may not give proper feedback. In search engine, there is no proper measure of interest.

Removing the unrelated events with the user based events by proposing recommendation system. The like/dislike feature are added to each URL based on link analysis & web crawling algorithms. This can be solved by interestingness measure which is completely depends on extrinsic approach.

IV. METHODOLOGY

PROPOSED ARCHITECTURE

User gives a mailID and password in a login page of social media or search engine. Check whether the mailID and password is correct and that page is moved to the user's home page of the social media or search engine. In social media, user gets like, share, view the information on any subject. In search engine, user gets view, like, download the information on any subject. Now we consider only likes in social media or downloads in search engine. After that we calculate the how many likes are there different domain in social media or how many downloads are in different domain in search engine. Find the maximum like in social media or downloads in search engine. We have to the maximum like in social media or the maximum download in search engine on the subject is domain of the user. Now recommendation system is used. The domain related events are recommended to the user in social media

Architecture

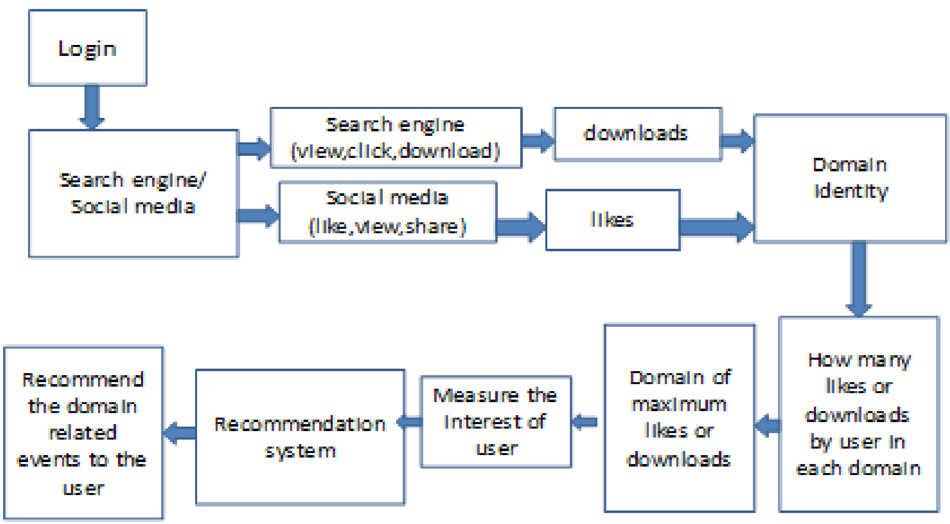

Fig 1 Proposed Architecture diagram

The domain related events are sent to the user's mailID in search engine.

OPINION FEATURE IDENTIFICATION

Domain relevance et al [2], characterizes how much a term is related to a particular like (i.e., a domain) based on two kinds of statistics, namely, dispersion and deviation.

Dispersion quantifies how significantly a term is mentioned across all documents by measuring the distributional significance of the term across different documents in the entire like (horizontal significance). Deviation reflects how frequently a term is mentioned in a particular document by measuring its distributional significance in the document (vertical significance).

Each like L_a has a like frequency (LF – IDF).

LF-Like Frequency

-Inverse Domain Frequency

Each like L_a has a like frequency LF_{ab} in a domain D_b and global domain frequency DFi.

The weight W_{ab} of like L_a in domain in domain D_b the calculated as follows:

$$W_{ab}=\begin{cases} (1+\log\ \mathrm{LFab}) * \log(\frac{N}{DFi}) & \text{if} LF_{ab}>0 \\ o & otherwise \end{cases}$$

a=1……M for a total number of M likes.

b=1…..N for a total number of N domain in the corpus.

The standard variance SV_a for like L_a is calculated as follows.

$$SV_a= \sqrt{((\textstyle\sum_{b=1}^{N}(\mathrm{Wab}-wa'))}/N$$

The average weight W_a of like L_a across all domain is calculated by

$$W_a = 1/N\ (\textstyle\sum_{b=1}^{N} wab$$

The dispers on disp of each like L in the corpus is defined as follows

$$Disp_a = W_a'/SV_a$$

The derivation devi of like L in domain D is given by

$$Devi_{ab}=W_{ab}-W_b'$$

Wr the average weight in the domain calculated over all M terms as follows.

$$W_b'=1/M \textstyle\sum_{a=1}^{M} Wab$$

The domain relevance dr for like L in the corpus is finally defined as follows.

$$Dr_a=disp_a*\textstyle\sum_{b=1}^{N} deviab$$

clearly,the domain relevance dv_a incorpoates horizontal and vertical distributional significance of term LF_i in the likes.

EXTRINSIC ALGORITHM

Input: A domain specific/independent corpus C

Output: Domain relevance scores(IDR or EDR)

```
For (candidate feature LFi)
     for (document Dl in D)
          find  wab then
          cal SVa
           cal dispa
       for(document Dj in C)
          cal (deviab)
Return(list)
```

OPINION FEATURES ALGORITHM

Input: Domain review corpus R and domain-independent corpus D

Output: A validated list of opinion features of the user domain related events.

```
Login
    search(S.M/S.E)
        extract(info)
For (word)
   cal Tf
   count domain
    find MAX
     measure interest
If(info=match (events))
     recommend;
Else
return
```

WEB CRAWLER ALGORITHM

web crawler can copy all the pagers the pages they visit for later processing by a search engine that indexes the downloads pages so that users can search them much more quickly

Input: set of popular URLs S
Output: Repository of visited web pages R

```
L1: If (s==NULL)
        then(P->S & P==NULL)
    get (P*)
    if(P*==R) then
        return 1;
else
        add (P*->R)&&(P* !=R,S)&(P*->s)
loop  L1
```

V. EXPERIMENT DESIGN

We conducted various experiments to comprehensively evaluate the EDR performance on more hours.we compared the proposed result and existing result

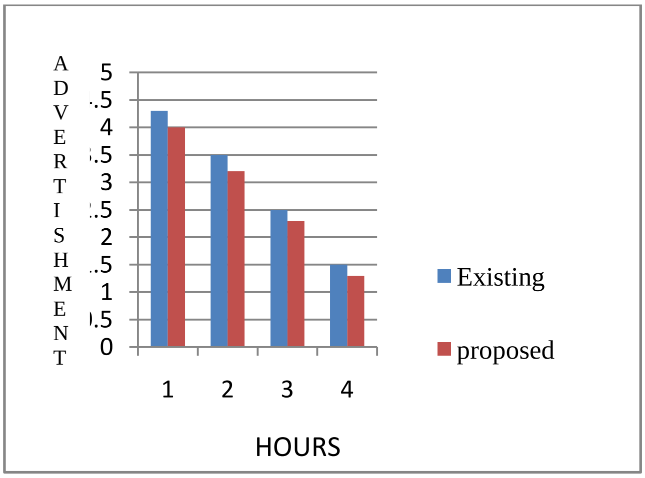

Fig 2 Advertisement versus hours

The advertisements are more in existing system compared to proposed system. The events are more in proposed system compared to existing system.

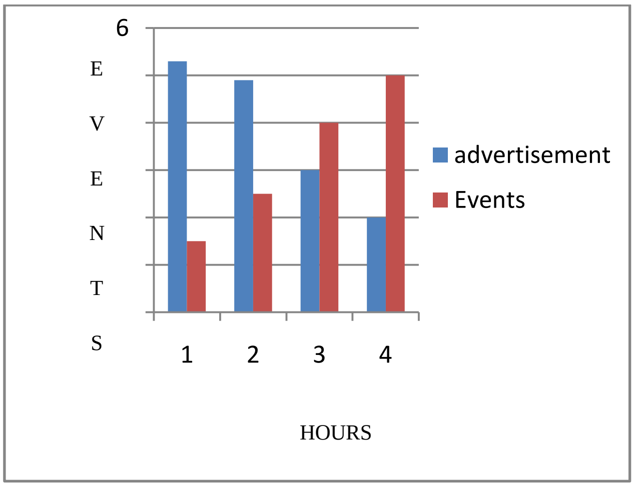

Fig 3 Events versus Advertisement

VI. CONCLUSIONS

Overall we produce an extrinsic approach, which produces more events and reduce advertisement in social media and search engine to the user. Experimental results demonstrate that the EDR but also outperform main stream method, namely, web crawler algorithm**.**

REFERENCES

[1] B. Liu, “Sentiment Analysis and Opinion Mining,” SynthesisLectures on Human Language Technologies, vol. 5, no. 1, pp. 1-167, May 203

[2] Identifying Features in Opinion Mining via Intrinsic and Extrinsic Domain Relevance,Zhen Hai, Kuiyu Chang, Jung-Jae Kim, and Christopher C. Yang.

[3] V. Hatzivassiloglou and J.M. Wiebe, “Effects of Adjective Orientation and Gradability on Sentence Subjectivity,” Proc. 18th Conf. Computational Linguistics, pp. 299-305, 2000.

[4] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up?: Sentiment Classification Using Machine Learning Techniques," Proc. Conf. Empirical Methods in Natural Language Processing, pp. 79-86, 2002.

[5] B. Pang and L. Lee, "A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts," Proc. 42nd Ann. Meeting on Assoc. for Computational Linguistics, 2004.

[6] L. Qu, G. Ifrim, and G. Weikum, "The Bag-of-Opinions Method for Review Rating Prediction from Sparse Text Patterns," Proc. 23rd Int'l Conf. Computational Linguistics, pp. 913-921, 2010.

[7] D. Bollegala, D. Weir, and J. Carroll, "Cross-Domain Sentiment Classification Using a Sentiment Sensitive Thesaurus," IEEE Trans. Knowledge and Data Eng., vol. 25, no. 8, pp. 1719-1731, Aug. 2013.

[8] P.D. Turney, "Thumbs Up or Thumbs Down?: Semantic Orientation Applied to Unsupervised Classification of Reviews," Proc. 40th Ann. Meeting on Assoc. for Computational Linguistics, pp. 417- 424, 2002.

[9] C. Zhang, D. Zeng, J. Li, F.-Y. Wang, and W. Zuo, "Sentiment Analysis of Chinese Documents: From Sentence to Document Level," J. Am. Soc. Information Science and Technology, vol. 60, no. 3, pp. 2474-2487, Dec. 2009.

[10] A.L. Maas, R.E. Daly, P.T. Pham, D. Huang, A.Y. Ng, and C. Potts, "Learning Word Vectors for Sentiment Analysis," Proc. 49th Ann. Meeting of the Assoc. for Computational linguistics: Human Language Technologies, pp. 142-150, 2011.

[11] T. Wilson, J. Wiebe, and P. Hoffmann, "Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis," Proc. Conf. Human Language Technology and Empirical Methods in Natural Language Processing, pp. 347-354, 2005.

Yessenalina and C. Cardie, "Compositional Matrix-Space Models for Sentiment Analysis," Proc. Conf. Empirical Methods in Natural Language Processing, pp. 172-182, 2011.

[12] E. Cambria, D. Olsher, and K. Kwok, "Sentic Activation: A Two- Level Affective Common Sense Reasoning Framework," Proc. 26th AAAI Conf. Artificial Intelligence, pp. 186-192, 203.

[13] F. Li, C. Han, M. Huang, X. Zhu, Y.-J. Xia, S. Zhang, and H. Yu, "Structure-Aware Review Mining and Summarization," Proc. 23rd Int'l Conf. Computational Linguistics, pp. 653-661, 2010.

[14] S.J. Pan and Q. Yang, "A Survey on Transfer Learning," IEEE Trans. Knowledge and Data Eng., vol. 22, no. 10, pp. 1345-1359, Oct. 2010.

