Knowledge-based Identification of Emotional Status on Social Networks

Julio Vizcarra[1], Kouji Kozaki[1,*],Miguel Torres Ruiz[2],Rolando Quintero[2]

[1] The Institute of Scientific and Industrial Research (ISIR) Osaka University
Mihogaoka 8-1, Ibaraki, Osaka 567-0047. Japan.
[2] Centro de Investigación en Computación CIC , Instituto Politécnico Nacional,
UPALM-Zacatenco, CIC. Building, 07738, Mexico City, Mexico.

Abstract. A knowledge based methodology is proposed for the content understanding and sentiment identification of the shared comments in social networks. The goal of this work is to retrieve the sentiment information associated to an opinion and classify it by its polarity and sentiment by means of a semantic analysis. Our approach implements knowledge graphs, similarity measures, graph theory algorithms and disambiguation processes. The results obtained were compared with data retrieved from Twitter and users' reviews in Amazon. We measured the efficiency of our contribution with precision, recall and F-measure comparing it with the traditional method of just looking up concepts in sentiment dictionaries which usually assigns averages. Moreover an analysis was carried out in order to find the best performance for the classification by using polarity, sentiment and a polarity-sentiment hybrid . A study is presented for remarking the advantage of using a disambiguation process in knowledge processing.

Keywords: sentiment analysis, knowledge engineering, conceptual similarity

1 Introduction

Nowadays the huge information transmitted on social networks has become a rich source of information for the human understanding as well as a way of expression where the users share their sentiment status and personal opinions through comments. The sentiment identification can classify comments as positive or negative(polarity) and unveil emotions such as anger, trust, sadness ,etc., on certain topics or users. Moreover the sentiments presented in the opinions can be relevant in the design of custom services, social plans for public health, marketing, e-commerce,etc.

On the hand sentiment analysis has become one of the fastest growing research areas in computer science due the outbreak of computer-based sentiment

* Corresponding author.
E-mail address: kozaki@ei.sanken.osaka-u.ac.jp.

studies with the availability of subjective texts on the Web [16]. Furthermore the sentiment analysis has gained attention over the years in the general public as it is currently shown in Google trends [10].

Based on the previous motivation the present work aims in the identification of sentiment information in opinions on social networks. Our approach explores a content-based and semantic processing of the knowledge implicit in the comments. For each opinion we created a formal representation which it is associated with a sentiment and polarity.

2 Background

This section lists some relevant works related with the proposed methodology presenting their key features. As summary we present a discussion where we remark the main contributions of our work.

Describing briefly some similar works related with sentiment analysis are: Anja Rudat[20] explored the criteria influencing selection for retweeting in Twitter. Trying to discover relations on social networks Yuan Wang[24] proposed a methodology that inferred social relationships in microblogs based on physical interactions using user's location records. The work of Garcia-Pablos [7] proposed an unsupervised system for the aspect-based sentiment analysis. One of the limitation of this work was the necessary to define manually seed concepts and domains as input of the methodology. The work of Divya Sehgal et al., [21] proposed a real-time sentiment analysis using dictionaries but mostly focused on big data techniques that prioritize the velocity instead of a deeper analysis. Theodore Georgiou [8] proposed a community detection algorithm utilizing social characteristics and geographic locations.

Regarding the semantic processing the work of Shivam Srivastava [22] developed an algorithm to cluster places not only based on their locations but also their semantics in social networks, the contributions of this work was the geo-social clustering from check-in data. The work of Shuai Wang et al.[23] applied a semantics-based learning technique for a set of concepts previously labeled by grouping the target-related words in order to extract the semantics among words.

On the other side some researches related to social networks analysis are for instance the work of Shuiguang Deng[5] that proposed a recommendation service for the social networks with a trust enhancement method. Considering the influence on social networks the work of Meng Jiang[14] studied the interpersonal influence, the approach explains the importance of this factor for behavior prediction. Additionally Huang Liwei[12] explored the user preference, social and geographical influence in order to recommend proper POIs (Point-of-interest). The machine learning implementation of Souvick Ghosh[9] processed the media text in order to determine the polarity and sentiment using manually labeled Facebook posts.

Reviewing the state-of-the-art, most of the researches worked with key social attributes that in general dismissed the semantics focusing in the lexical processing, keywords or explicit reactions in the social media. About the methodologies that implemented machine learning techniques they were based on a high quality large training datasets on a specific domain. On the hand our work handles the comments as excerpt of the knowledge, in this gap we prioritized the semantic level, sense and meaning of the whole comment. The proposal computed semantic similarity measures, conceptual expansion, graph theory algorithms and disambiguation using on a multi domain knowledge base. The methodology is flexible which implies that the domains can be adjusted by just modifying knowledge base.

3 Methodology

This section describes the methodology in three main stages. The first stage "social networks discovery" retrieves opinions from events or public profiles by reading comments in photos, posts, videos, etc. The stage of "knowledge processing" constructs the formal representation for each comment. This module carries out processes of automatic knowledge graph construction enhanced by disambiguation. Finally the stage of "sentiment analysis" estimates the total polarity and main sentiment in the comments .

3.1 Social network discovery stage

In the stage the comments are retrieved from public events or user profiles on social network. This process obtains users, comments and the social graph's structure.

3.2 Knowledge processing stage

In this stage a content-based formal representation is constructed for each comment in the social network. This stage is composed by "lexical preprocessing" ,"knowledge graph expansion", "similarity measure" and "disambiguation".

Lexical pre-processing. In the step the concepts in a comment are processed in order provide term matching with the knowledge base. The processes considered are: stop words elimination, tokenizer, stemming, and removal of unknown concepts in the knowledge graph.

Knowledge graph expansion. In this step the set of concepts obtained in the lexical processing are expanded on the knowledge graph until finding a common root for all their senses.

Let us define $G(C, R)$ as a knowledge graph with the set of concepts C and the set of relationships R; the knowledge base expansion (Ge)(equations 1, 2) for a concept c $\in$ C is the iterative process (α iteration) of discovering new concepts in knowledge graph (G) using semantic relations (ρ)(equation 4) that connect a origin concept c to the other destination concepts $C\alpha$(equation 3).

$$Ge_0^\rho(c, G) = G_0(C_0, R_0) = G_0(\{c\}, \emptyset) \tag{1}$$

$$Ge_\alpha^\rho(c, G(C, R)) = G_\alpha(C_\alpha^\rho, R_\alpha^\rho) \tag{2}$$

$$C_\alpha^\rho = \begin{cases} \alpha = 0 & \{c\} \\ \alpha > 0 & C_{\alpha-1} \cup \{y \in C : x \in C_{\alpha-1}, \rho(x, y) \in R\} \end{cases} \tag{3}$$

$$R_\alpha^\rho = \begin{cases} \alpha = 0 & \emptyset \\ \alpha > 0 & \{\rho(x, y) \in R : x, y \in C, x \in C_{\alpha-1}\} \end{cases} \tag{4}$$

Similarity Measurement. Once the concepts were expanded and an excerpt of knowledge was constructed from the previous stage, the next step is to establish similarity measures among all concepts. In order to accomplish this task two different approaches were implemented:

1) Automatically. It was implemented the similarity measure of conceptual distance DIS-C[19] that automatically establishes the similarity among concepts following the idea of visibility in the knowledge graph.

2) Manually. For each semantic relationship in the knowledge graph we established a weight in the range [0,1].

Disambiguation. In this stage a strongly connected graph $G_D(C, R)$ is created which is disambiguated and reduced (number of nodes and relationships) by a steiner tree algorithm. In the methodology we implemented the SketchLs algorithm[11] due the capability of handling large graphs. The disambiguation process starts counting the number of occurrences(senses)(Figure 1). If a concept has only one occurrence it implies that it has only one sense and it will participate in the disambiguation of the other concepts. On the other hand if a concept has more than one occurrence this concept has to be disambiguated.

During the disambiguation if the comment has only one concept and it has several senses then a dictionary of polysemy has to be consulted for finding most probable sense. On the other hand if the comment has more than a concept then the disambiguation will be computed.

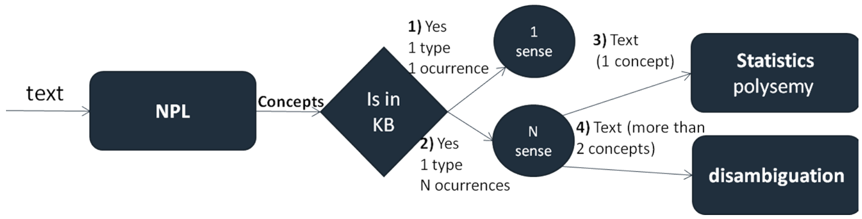

Fig. 1. Disambiguation

3.3 Sentiment analysis stage

Polarity calculation. In this step the polarity for comment is calculated $Polarity(Coment_x)$ taking into account the individual polarity of each concept $Po\left(C_P\right)$. The process starts dividing the concepts in subsets C_x considering their positive or negative polarity Po(C_x)(see equations 5-6). In order to calculate the polarity Pot(X_g) for a set of concepts X_g the arithmetic mean is computed (equation 7). The total polarity of a comment $Polarity(Coment_x)$ is calculated by the sum of positive plus negative polarities X_P and X_N respectively(see equation 8).

$$X_P = \{C_x \mid Po\left(C_x\right) > 0; C_x \epsilon X_P\} \tag{5}$$

$$X_N = \{C_x \mid Po\left(C_x\right) < 0; C_x \epsilon X_N\} \tag{6}$$

$$\overline{Pot\left(X_g\right)} = \frac{\sum_{i=0}^{n} Po\left(C_x\right)}{n}; C_x \epsilon X_g \tag{7}$$

$$Polarity(Comment_x) = \overline{Pot\left(X_P\right)} + \overline{Pot\left(X_N\right)};\ X_N, X_N \subseteq Comment_x \tag{8}$$

Sentiment identification. In this step the sentiment status is identified in a comment $Sentiment(Coment_x)$. For each concept $C_i \in Coment_x$, C_i it is expanded in the knowledge graph until finding one or more concepts linked to a sentiment S_x. The next process is to find the the closest sentiment S_x to C_i by computing a shortest path algorithm and semantic similarities. Consecutively a pre-defined numerical weight $Ws(C_x)$ is assigned for the sentimentS_x which is located between the range [-1,-1] (equation 9). Once the weight of the sentiment was obtained the next step is to calculate the sentiment value $Sen(C)x)$ for the concept C_x by multiplying the sentiment weight $Ws(C_x)$ by its polarity $Po(C_x)$ (equation 10). Finally the sentiment status with the highest sentiment value $Sen\left(C_x\right)$ is assigned to the comment $Coment_x$ (equation 11).

$$Ws\left(C_x\right) = w\left(S_x\right); C_x \rightarrow S_x, w\left(S_x\right) \in [-1, 1] \tag{9}$$

$$Sen\left(C_x\right) = Po\left(C_x\right) Ws\left(C_x\right) \tag{10}$$

$$Sentiment(Comment_x) = max\left(\{Sen\left(C_i\right) \mid C_i \in Coment_x\}\right) \tag{11}$$

The figure 2 presents the iterative process of expansion for finding the sentiment associated to a concept C_x in the knowledge base. When one or more concepts are located and they are linked to a sentiment then the Dijkstra algorithm with Fibonacci heap [6] is executed in order to select only one concept.

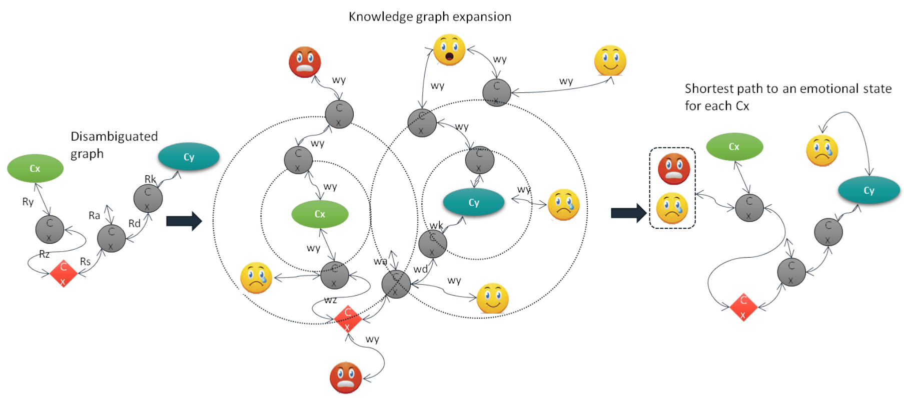

Fig. 2. Sentiment identification

4 Implementation

This section presents the results after implementing the described methodology. It is divided in two subsections: “knowledge bases” and “sentiment analysis.

4.1 knowledge bases

In this section we describe the knowledge base's structure which is composed by: general knowledge graphs for common language understanding on several domains and sentiment dictionaries mapped into the knowledge graph.

General knowledge bases

– WordNet[1] (version 3.1) is a large lexical database of English. Nouns, verbs, adjectives and adverbs are grouped into sets of cognitive synonyms (synsets).
– The Japanese WordNet[3,13] is similar to Wordnet for processing the Japanese language.
– Open Multilingual Wordnet [4][3] provides access to wordNets in a variety of 34 languages merged into English WordNet.

Sentiment dictionaries

– SentiWordnet [2] is a lexical resource that assigns polarity values to concepts in English WordNet.
– NRC_emotion_lexicon [18,17] is a list of English words associated with eight basic emotions (anger, fear, anticipation, trust, surprise, sadness, joy, and disgust).

4.2 Sentiment Analysis

In order to explain the results obtained in the sentiment analysis an example was processed from Twitter in the CNN News account. The comment considered is : “a number of people feared dead after a dam bursts in kenya with hundreds left homeless officials say”. The table 1 presents the closest sentiment and a polarity value assigned by our methodology to each concept.

Id Wordnet-Concept	Sentiment with polarity
WN:107449542-n ("flare","burst")	Sentiment:NRC_fear_NRC_anger|:Polarity:-0.25 ,
WN:107964900-n (homeless)	Sentiment:NRC_anticipation_disgust_anger|Polarity:-0.125 ,
WN:107534492-n (fear)	Sentiment:NRC_fear,sadness,anger,surprise|Polarity:-0.875 ,
WN:114509110-n (say)	NRC_surprise_anticipation|Polarity:0.5

Table 1. Sentiment-Polarity assigned to concepts

Finally the methodology estimates the total polarity and main sentiment presented in the comment(table 2).

Sentiment	Polarity	Comment
NRC_Anger	-0.1875	a number of people feared dead after a dam bursts in kenya with hundreds left homeless officials say.

Table 2. Sentiment-Polarity assigned to comment

Other relevant examples from the CNN news account are presented in table 3. We noticed a better classification using the basic sentiments instead of polarity.

Sentiment	Polarity	Comment
trust	0.2916667	This couple found a buried safe containing $52,000 worth of money, gold and jewelry in their backyard, but didn’t keep it
trust	-0.15	In an effort to keep conversations and search results on topic, Twitter announced it will use new "behavioral signals" to push down more tweets that "distort and detract"
anger	0.04166667	A massive poaching ring in Oregon and Washington is accused of killing more than 200 animals including deer, bears, cougars, bobcats and a squirrel
anger	0.041666687	An estimated 239,000 girls under the age of five die in India each year due to neglect linked to gender discrimination, a new study finds
sadness	0.25	@CNN Her father had a heart surgery and cant walk so
sadness	-0.25	Teen develops ’wet lung’ after vaping for just 3 weeks
joy	0.125	I am proud to be a woman and a feminist. The politics of Meghan Markle

Table 3. Other examples processed in twitter

5 Evaluation

This section measured the performance of our methodology comparing it with labeled data with sentimental information. We considered as a manual processing Twitter posts that we manually labeled and as automatic processing comments ranked by the users in amazon reviews. As traditional method (baseline) we proposed the process of only looking up concepts with polarity in dictionaries.

5.1 Sentiments evaluation on Amazon Reviews

We evaluated our work with precision, recall and F-measure over 10 000 comments using the dataset Amazon reviews provided by the Stanford Network Analysis Project (SNAP)[15] and shared by Xiang Zhang [25]. In this dataset an user gives scores for products in the range of one to five starts. We associated the scores with negative sentiments(anger,disgust, sadness,fear) and positive sentiments(joy, trust, anticipation, surprise) and a polarity value. The figure 3) presents the evaluation using polarity and sentiment with automatic and manual similarity measures during the semantic processing (polaritySemRelAuto, polaritySemRelManual, SSRelAuto and SSRelManual) and PolarityLexical(base line).

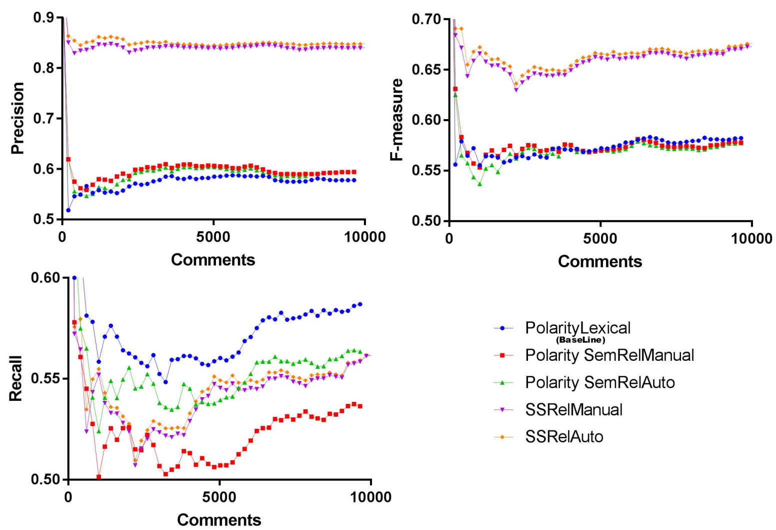

Fig. 3. Evaluation in amazon reviews

Additionally the figure 4 presents the evaluation with precision for the disambiguation process using polarity with automatic and manual similarity measures (polarityAuto, polarityManual). The results were compared to polarity lexical(baseline) with random sense selection (PolarityLexicalR1-R10).

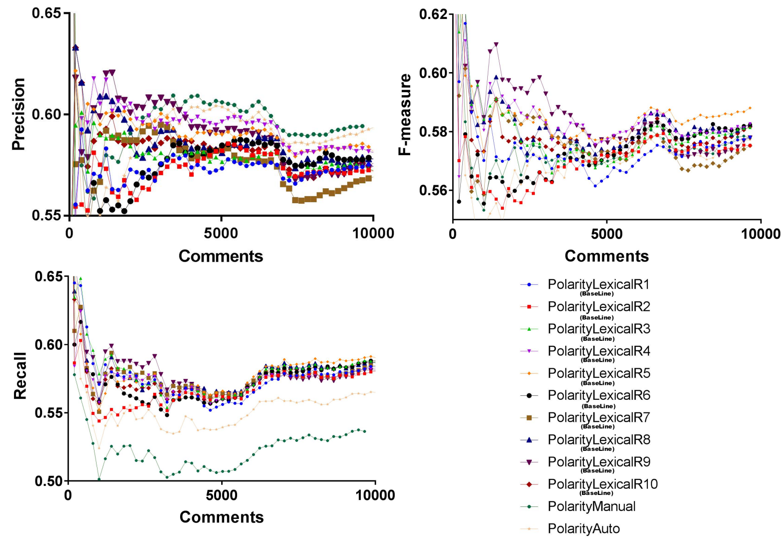

Fig. 4. Evaluation of disambiguation

5.2 Sentiments evaluation on Twitter

For this evaluation some comments were retrieved from Twitter and manually associated with a sentiment and polarity. The figure 5 presents the results only considering precision. The PrecisionLex (baseline) was calculated using only polarity. On the other hand PrecisionSS considered sentiment and computed a semantic analysis and a disambiguation process. In this experiment the PrecisionSS presented better results.

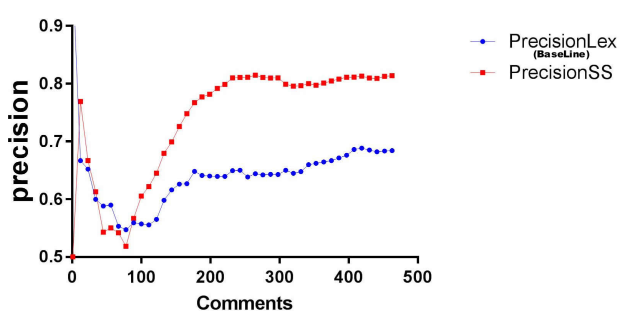

Fig. 5. Evaluation Twitter

During the experiments we noticed that the methodology provides different results for specific sentiments (figure 6). For instance the sentiment anger or disgust performed better precision because usually the comments are more explicit.

On the other hand the joy was more complicated to identify because the usage of sarcasm or more implicit sentiments in the comments.

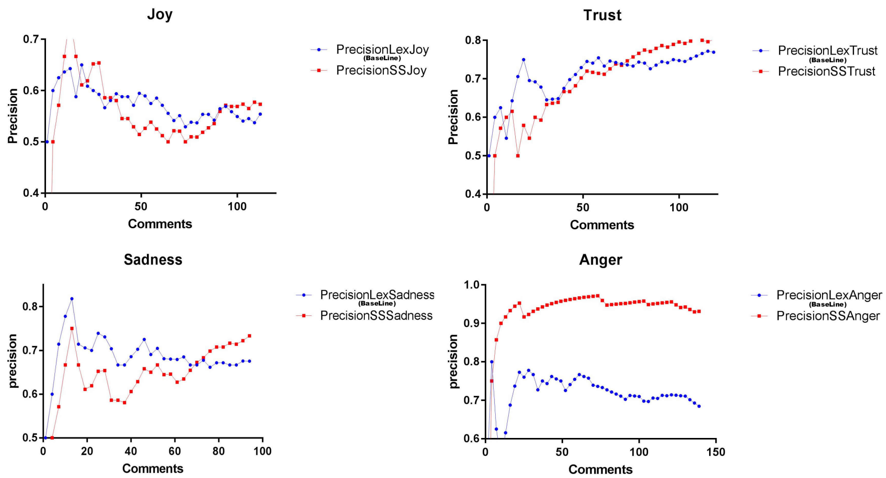

Fig. 6. Evaluation four sentiments

6 Conclusions

In this paper a content-based methodology was proposed for the polarity calculation and sentiment status identification. The novelty of the presented work is the capability of handling the comments as excerpts of knowledge. We provided a mechanism of semantic processing using knowledge graphs, graph theory algorithms, semantic similarities and disambiguation. For the sentiment identification our work explored three different approaches (polarity, sentiment, sentiment-polarity hybrid) where the sentiment-polarity processing presented the best results.

We performed several experiments in order to compared our contribution with the traditional method of just looking up concepts in dictionaries(baseline) that usually counts polarity or concepts related with sentimental information and assigns averages.

Based on the experimental analysis the best relation precision and computing consumption was presented by the combination of sentiment, manual weights in semantic processing and disambiguation (SSRelManual). On the other the highest precision was obtained with automatic weights (SSRelAuto) costing a significant increment in the usage of computing resources. Despite of the disambiguation presented a slightly better precision it provided the best combination of concepts for the construction of formal representations and thus better sentiment identification. The results obtained in the present work can be consulted at the github site: https://github.com/samscarlet/SBA.

7 Acknowledgments

This work was supported by CONACYT and JSPS KAKENHI Grant Number JP17H01789.

References

1. Princeton university "about wordnet." wordnet. princeton university (2010), <http://wordnet.princeton.edu>
2. Baccianella, S., Esuli, A., Sebastiani, F.: Sentiwordnet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining. In: LREC. vol. 10, pp. 2200–2204 (2010)
3. Bond, F., Baldwin, T., Fothergill, R., Uchimoto, K.: Japanese semcor: A sense-tagged corpus of japanese. In: Proceedings of the 6th Global WordNet Conference (GWC 2012). pp. 56–63 (2012)
4. Bond, F., Foster, R.: Linking and extending an open multilingual wordnet. In: Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). vol. 1, pp. 1352–1362 (2013)
5. Deng, S., Huang, L., Xu, G., Wu, X., Wu, Z.: On deep learning for trust-aware recommendations in social networks. IEEE transactions on neural networks and learning systems **28**(5), 1164–1177 (2017)
6. Fredman, M.L., Tarjan, R.E.: Fibonacci heaps and their uses in improved network optimization algorithms. Journal of the ACM (JACM) **34**(3), 596–615 (1987)
7. García-Pablos, A., Cuadros, M., Rigau, G.: W2vlda: almost unsupervised system for aspect based sentiment analysis. Expert Systems with Applications **91**, 127–137 (2018)
8. Georgiou, T., El Abbadi, A., Yan, X.: Extracting topics with focused communities for social content recommendation. In: Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing (2017)
9. Ghosh, S., Ghosh, S., Das, D.: Sentiment identification in code-mixed social media text. arXiv preprint arXiv:1707.01184 (2017)
10. Google: Google trends.url: https://trends.google.com/trends/?geo=us.
11. Gubichev, A., Neumann, T.: Fast approximation of steiner trees in large graphs. In: Proceedings of the 21st ACM international conference on Information and knowledge management. pp. 1497–1501. ACM (2012)
12. Huang, L., Ma, Y., Liu, Y., Sangaiah, A.K.: Multi-modal bayesian embedding for point-of-interest recommendation on location-based cyber-physical-social networks. Future Generation Computer Systems (2017)
13. Isahara, H., Bond, F., Uchimoto, K., Utiyama, M., Kanzaki, K.: Development of the japanese wordnet. (2008)
14. Jiang, M., Cui, P., Wang, F., Zhu, W., Yang, S.: Scalable recommendation with social contextual information. IEEE Transactions on Knowledge and Data Engineering **26**(11), 2789–2802 (2014)
15. Leskovec, J.: Snap: Stanford network analysis project (2014)
16. Mäntylä, M.V., Graziotin, D., Kuutila, M.: The evolution of sentiment analysisï£¡a review of research topics, venues, and top cited papers. Computer Science Review **27**, 16–32 (2018)

17. Mohammad, S.M., Turney, P.D.: Emotions evoked by common words and phrases: Using mechanical turk to create an emotion lexicon. In: Proceedings of the NAACL HLT 2010 workshop on computational approaches to analysis and generation of emotion in text. pp. 26–34. Association for Computational Linguistics (2010)
18. Mohammad, S.M., Turney, P.D.: Crowdsourcing a word–emotion association lexicon. Computational Intelligence **29**(3), 436–465 (2013)
19. Rodíguez Franco, H.: Cálculo de la visibilidad de conceptos en ontologías. Ph.D. thesis, Instituto Politécnico Nacional. Centro de Investigación en Computación (2011)
20. Rudat, A., Buder, J.: Making retweeting social: The influence of content and context information on sharing news in twitter. Computers in Human Behavior **46**, 75–84 (2015)
21. Sehgal, D., Agarwal, A.K.: Real-time sentiment analysis of big data applications using twitter data with hadoop framework. In: Soft Computing: Theories and Applications, pp. 765–772. Springer (2018)
22. Srivastava, S., Pande, S., Ranu, S.: Geo-social clustering of places from check-in data. In: Data Mining (ICDM), 2015 IEEE International Conference on. pp. 985–990. IEEE (2015)
23. Wang, S., Zhou, M., Mazumder, S., Liu, B., Chang, Y.: Disentangling aspect and opinion words in target-based sentiment analysis using lifelong learning. arXiv preprint arXiv:1802.05818 (2018)
24. Wang, Y., Xiao, Y., Ma, C., Xiao, Z.: Improving users' demographic prediction via the videos they talk about. In: Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. pp. 1359–1368 (2016)
25. Zhang, X., LeCun, Y.: Text understanding from scratch. arXiv preprint arXiv:1502.01710 (2015)