

# Sentiment Mining from Text: A Technical Review

Mita K. Dalal

Sarvajanik College of Engineering & Technology  
Surat, India

## ABSTRACT

Expression of sentiments is an integral aspect of human communication. The prevalence of Internet and social media platforms has facilitated the exchange and storage of large volumes of human communication in the form of natural language text which can be mined for various intelligence-based applications. In order to build computers that can serve humans better, there are ongoing efforts in computer science research to develop machine learning algorithms that can process the textual data and perform sentiment mining tasks such as, 'detecting the presence of emotion in text', 'selecting a model for representing emotion', 'classifying the sentiment polarity of text' and 'measuring the intensity of the expressed sentiment in text'. This paper aims to discuss the challenges posed to the various tasks associated with mining sentiment from text, review the existing methods to address these challenges in contemporary literature and identify important areas for future scope of research in this field.

## General Terms

Sentiment Mining, Text Mining, Natural Language Processing

## Keywords

Sentiment Polarity, Emotion Model, Sentiment Lexicon, Semantic Features, Linguistic Hedges

## 1. INTRODUCTION

Sentiment mining refers to the various computational aspects of identification, classification and aggregation of sentiment expressed either explicitly or implicitly in natural language text. Humans usually express sentiment through multiple modalities such as 'facial expressions', 'voice' including both linguistic expressions and non-linguistic vocalizations and 'body language' [1], hence detecting sentiment using only 'text' is in itself a challenging task. However, natural language text is one of the most prevalent means of communication on E-Commerce and social-media platforms, hence efficient mining of sentiment from blog text is relevant in many contexts [2], [3], [4], [5], [6], [7], [8], [9], [10].

Sentiment mining from text has been applied to a variety of useful and diverse applications in literature such as, understanding the sentiments of a geographically distributed team through their chat logs [2], analyzing the effect of negative opinions expressed in financial media text on potential investors [3], classifying and statistically summarizing the sentiment polarity expressed in online product reviews by customers [4], [5], discerning the emotions of students through sentiment analysis of the dialog between students and computerized tutoring system [6], mining of sentiment orientation from political blogs [7], [8], [9] etc.

Depending on the application, mining of sentiment from text involves selectively performing activities like detecting the presence of sentiment in natural language text, classification of the emotion expressed, determining the sentence-level or overall sentiment polarity of the text as positive/negative/neutral, and quantifying the intensity of the

expressed sentiment. Moreover, to effectively process web-based text, pre-processing steps are required to remove noise, detect sentence boundaries, perform parts-of-speech tagging etc. [11]. Resources such as lexical databases [12], grammar parsers [13], [14] and sentiment lexicons [15], [16] are frequently used by researchers to automate some aspects of the above mentioned tasks.

Section 2 of the paper elaborates on the major challenges faced during computational mining of sentiment from text, existing methods to handle these challenges found in literature and their limitations. The paper concludes with a brief discussion on prospective areas for further research in this field.

## 2. SENTIMENT MINING FROM TEXT: MAJOR ISSUES AND SOLUTIONS

This section examines the major issues in mining sentiment from textual data. It also explores some solutions to these issues proposed in contemporary literature and discusses their applicability and limitations.

### 2.1 Selecting an appropriate representation of Affect Categories

Scientists have proposed different models for the categorization of emotions which raises the important design issue of selecting an appropriate representation of affect categories for the sentiment mining task to be performed. The well-known emotion model proposed by Ekman [17], [18] consists of six basic emotions namely 'anger', 'fear', 'sadness', 'enjoyment', 'disgust' and 'surprise'; another widely used emotion model proposed by Plutchik [19], [20] comprises eight basic bipolar emotions namely 'joy' versus 'sorrow', 'anger' versus 'fear', 'acceptance' versus 'disgust' and 'surprise' versus 'expectancy'; while the ISEAR databank (International Survey of Emotion Antecedents and Reactions) documents seven emotion categories namely, 'disgust', 'shame', 'guilt', 'anger', 'fear', 'sadness' and 'joy' [21], [22], [23], [24]. Thus, various emotion models and categorizations exist in literature.

The emotion model chosen influences the task of computerized emotion classification of text, but this choice is dependent on the application and is also a matter of personal preference of the researcher. It is observed that for some real-world applications no emotion model is directly applicable or that the application requires focus on only a few of the affect categories. Moreover, humans express themselves using a large vocabulary of emotion indicating terms, hence a mapping is required between the terms and their emotion category. Thus, selective classification of emotion categories pertinent to the application and clustering of emotion-indicating terms to adequately represent the relevant emotion categories is an issue that needs to be resolved. One approach to resolve this issue is by generating a topic model for selective representation of an affect category. For instance, in [3] an LDA model (Latent Dirichlet Allocation) was used to generate topic-clusters for selective representation of the negative affect categories 'fear', 'anger', 'remorse' and

'contempt', wherein each topic-cluster is characterized by the most representative terms in the vocabulary of the training documents. Another approach is the use of lexical resources such as WordNet-Affect that pre-associate the usage of affective words with emotions [25].

However, these approaches have the limitation that it is difficult to accommodate incremental or evolving changes in vocabulary, since their classification is based on prior trained topic models or pre-assigned mappings between words to emotion categories.

## **2.2 Classifying the Sentiment Polarity of text**

Several applications of sentiment mining require classification of the expressed opinions as positive, negative or neutral [4], [26], [27], [28], [29], [30]. Automatic classification of the 'sentiment polarity' or 'orientation' of text is a challenging task.

Recent approaches for sentiment polarity classification are based on semantic feature set extraction [4], [28], [29]. Feature based methods for polarity classification require mining of a meaningful feature set from the opinion bearing text before tagging the sentiment polarity, which is a challenging task. The methods reported in literature to mine a feature set and their limitations are discussed next.

### *2.2.1 Mining of a semantic feature set from text*

In opinion mining the nouns and noun-phrases that frequently occur in the domain discussion are considered potential candidates for semantic features while the adjectives describing them determine orientation [30], [31], [32], [33]. For instance, websites that post reviews of electronic products such as smartphones would frequently refer to noun features such as 'touchscreen', 'camera' etc. In order to identify nouns features and their associated adjective/adverb descriptors from text, pre-processing steps like sentence boundary detection and parts-of-speech (POS) tagging [4], [28], [34], [35] are performed. Moreover, since all nouns cannot be considered as semantic features a refining/reducing step is required. The various methods in literature to extract useful features comprise Apriori approach [4], [28], [31], Seed-set expansion approach [29], [34] and Multiword based approach [35], [36]. These three approaches are briefly discussed next.

The Apriori approach for mining a feature set applies the Apriori algorithm [37] to automatically mine frequently occurring nouns and noun-phrases from text. This method has the advantage of fast, automatic compilation of a potential feature set. However, semantic features frequently consist of multiple words (for example, "front camera", "battery life"). While mining the initial feature set, the Apriori approach does not consider the sequence in which the words occur or the distance between the words, hence it additionally requires re-ordering and compactness pruning [31].

Multiwords are ordered word sequences which convey more meaning to context than the individual words constituting it [35], [38]. Multiword based feature extraction has the advantage that it mines semantic features comprising of multiple words without the overhead of pruning or re-ordering [35], [36]. However, if a multiword feature is frequent, then its substrings are also frequent. Hence a design choice has to be made regarding the length of multiword features to retain [35].

The Seed-set expansion approach requires a domain expert to specify an initial set of features, which is subsequently

expanded to generate a larger feature set through various methods like synonym look-up using a lexical resource [34], expansion through conjunction strategy [29] etc. The limitation of this method is that it requires prior domain knowledge to specify the seed feature set. Moreover some relevant features might not be mined and the final feature set could vary depending on the chosen seed list of features.

Despite the above mentioned limitations, semantic feature based extraction techniques have successfully been used in opinion mining literature for a variety of applications where sentiment is expressed explicitly through the mention of features and their descriptors.

## **2.3 Effective representation of the intensity of sentiment in text**

An important task in sentiment mining is gauging the intensity of the expressed opinion. Applications that involve opinion mining from natural language text require a quantitative measurement of the intensity of sentiment. However, achieving high accuracy at such tasks is difficult since it requires taking into consideration the adjective and adverb descriptors as well processing of linguistic hedges [39], [40], [41] if present.

The adjectives and adverbs used for expression in text are explicit indicators of sentiment orientation and intensity [4], [35], [42]. In order to acquire the quantitative representation of intensity of descriptor terms, the SentiWordNet lexicon has been frequently used in literature [5], [15], [24], [29]. For instance, the SentiWordNet score for the adjective 'excellent' is (Positivity:1 Objectivity:0 Negativity:0) while that of adjective 'nice' is (Positivity:0.75 Objectivity:0.25 Negativity:0), which indicates that the sentiment term 'excellent' conveys a stronger positive sense compared to 'nice'.

In addition to the strength of the descriptors, the intensity of a sentiment phrase is also affected by the presence of 'linguistic hedges' also referred to as 'contextual valence shifters' [39], [40], [41], [43], [44]. Linguistic hedges comprise of 'concentrators' that intensify sentiment, 'dilators' that decelerate the sentiment and 'modifiers' that invert the sentiment expressed [5], [39], [40], [41]. For instance, a product 'P' can be described as "P is good" (hedge type: none), "P is not good" (hedge type: modifier hedge not), "P is very good (hedge type: concentrator hedge very)", "P is quite good" (hedge type: dilator hedge quite). Hedges are also used to adjust granularity of sentiment polarity classification tasks. For instance, binary sentiment classification has two classes 'positive' and 'negative', which can be increased in granularity through the use of hedges to five classes labeled as 'very negative', 'negative', 'neutral', 'positive' and 'very positive'. Sentiment classes with higher granularity are made to correspond to user-assigned 5-star ratings in opinion mining systems [5], [44], [45], [46].

Opinions expressed in natural language frequently make use of hedge terms, but assimilating the exact quantitative effect of concentrators and dilators proportionate to the sentiment phrase is a challenging task. The two major approaches to linguistic hedge adjustment are Valence Points adjustment method [43], [44] and fuzzy hedge adjustment methods [5], [45]. The Valence Points method performs hedge adjustment through shifts of fixed magnitude, but such uniform adjustments do not take into account the strength of the sentiment term. This limitation can be overcome through recent approaches that quantify linguistic hedges through

fuzzy score adjustment techniques [5], [45]. However, contemporary linguistic hedge adjustment methods suffer from the limitation that their sentiment classification accuracy reduces when the granularity of the classification task is increased [5].

## **2.4 Developing Sentiment Lexicons for new languages**

Sentiment lexicons are useful in mining sentiment orientation from natural language text. However developing a sentiment lexicon for a new language is a challenging task since manual generation of sentiment lexicons is time-consuming and requires meticulous compilation from linguistic experts. Moreover, due to the evolving nature of web-based text mainly due to use of non-dictionary words and abbreviations in communication, an exhaustive vocabulary may not be available. Various approaches have been proposed in literature to facilitate automation of some aspects of sentiment lexicon generation which are briefly described next.

### *2.4.1 Use of bi-lingual dictionary*

One approach to generate a sentiment lexicon for a new language is to make use of a bi-lingual dictionary for translation and use the translated word to look up entries in an existing sentiment lexicon for another language [16], [47]. The sentiment polarity (+/-) and the subjectivity/objectivity score for the translated word is then assigned to the original word. For example, the English version of the sentiment lexicon SentiWordNet [15], [48] was used to generate SentiWordNet for Indian languages including Hindi, Bengali and Telugu through use of different bi-lingual dictionaries [16].

### *2.4.2 Expansion through synonym and antonym look-up*

After the sentiment polarity of an initial list of words for a new language has been determined via manual compilation and dictionary look-up as explained in 2.4.1, the list can be further expanded through look up of synonyms and antonyms using a linguistic resource such as the WordNet lexicon [12] [25]. The synonyms are assigned same sentiment polarity as the original word while the antonyms are assigned an opposite polarity [16].

### *2.4.3 Linguistic Resource generation via crowd-sourcing*

Non availability of linguistic resources in a particular language is a major challenge to sentiment mining. Researchers have successfully used crowd-sourcing to facilitate generation of useful linguistic resources for sentiment mining. For instance, bi-lingual dictionaries were generated through a collaborative framework called wikiBabel [49], in which the collaborators comprised of online participants selected on the basis of factors such as their language fluency in the source and target languages and domain interest. Once such a bi-lingual dictionary is available, it can be used to look-up an existing sentiment lexicon for one of the languages and generate a sentiment lexicon for the other language. As another instance of crowd-sourcing in sentiment mining, the sentiment lexicon Emolex [50] was developed for English language that associates terms with emotion labels and also tags their polarities.

### *2.4.4 Mining sentiment through emoticons*

Emoticons are frequently used in web-based communications along with text messages and are often easily interpreted. (Example happy :) sad :( angry >:( etc.) Thus, emoticons

provide useful clues regarding the emotion expressed in the communication. Enhancing sentiment analysis through interpretation of emoticons attached to text has been reported in literature in conjunction with various languages including Japanese [20], English [51], Dutch [52], Chinese [53] etc.

Thus, generation of a new sentiment lexicon requires extensive use of computational as well as human resources.

## **2.5 Preserving Privacy during sentiment mining from text**

During sentiment mining it is essential to preserve the privacy of individuals participating in activities such as posting online reviews of products, giving opinion on political matters, participating in surveys etc. Ideally it should not be possible to re-identify an anonymous blogger who has expressed sentiment, without his consent. Disclosure of an online user's identity through non-explicit identifiers could occur due to sharing of two types of data: (i) structured data such as quasi-identifiers collected in attribute-value form [54], [55] or (ii) unstructured data such as stylometric features identifiable from the published textual content itself [56], [57]. Privacy preservation methods reported in literature against inadvertent disclosure of both these types of data are discussed next.

### *2.5.1 Preserving privacy through anonymization*

Web-based forums which serve as platforms for expressing sentiment often collect and store personal data of its users in attribute-value format such as 'gender', 'date of birth', 'nationality', 'level of education', 'partial location information' etc. Furthermore, they might share this data with other government or commercial agencies for profit. The participants of the survey generally assume that their identity will remain anonymous since they have not shared any explicit identifiers such as 'name' or 'address'. However, there is a risk that the personal attributes released by the web-forum could act as quasi-identifiers which will allow the records to be linked with other publicly available information of individuals which could re-identify the person or infer his identity through multiple related queries [54]. In order to overcome this problem, the k-anonymity algorithm can be applied to anonymize information before sharing, however it suffers from the assumption that the quasi-identifiers can always be successfully identified by the data publisher [54] [55].

### *2.5.2 Preserving privacy through stylometric circumvention*

Even when a person posts opinions online in an anonymous manner, the linguistic features reflected in his writing style such as vocabulary usage, sentence structure, average sentence length, function words, n-gram occurrences, idiosyncratic features such as misspellings, grammatical usage etc. might reveal his identity [56], [57], [58].

The stylometric circumvention techniques to avoid detection documented in literature include 'obfuscation', 'imitation' and 'machine translation' [56], [57]. 'Obfuscation' requires the author to consciously monitor his writing style to prevent identification and 'imitation' requires the author to emulate the style of one or more other authors to circumvent identification. Both these methods are interesting from a linguistic perspective but are dependent on manual intervention, hence not suitable for automated real-time sentiment mining systems. The third technique i.e. 'machine translation' involves multi-step translation of source text from original language to other languages and then back to source language via automatic translation system in order to alter the

writing style of the author. For example, in [56] experiments in multi-step translations such as English to German to Japanese and then translate back to English were performed. However, this approach may not provide sufficient anonymization to authors and the language quality of the text also deteriorates due to multiple translations.

### 3. CONCLUSION

Automated mining of sentiment from text is important for several present-day applications which involve processing of user-generated opinions or comments online. Significant progress has been made in the domain of sentiment mining from natural language text especially in the generation of lexical resources for sentiment mining, algorithms for extracting semantic feature sets from text and methods for classifying sentiment polarity. However, there is a need to develop better solutions to counter some of the challenges faced in this field.

In order to efficiently mine sentiments from web-based text where the vocabulary is continuously evolving, further research is required to develop methods that enable incremental training of emotion models. There is also the need to develop efficient functions for quantization of linguistic hedges that can emulate fine-grained levels of sentiment closer to human perception. In particular, development of sentiment mining systems with this capability would have a huge impact on the online review and recommendation systems of the future. Finally, there is a dire need to develop reliable algorithms for stylometric circumvention, since existing approaches do not guarantee privacy to authors who post their opinions online.

### 4. REFERENCES

- [1] Calvo, R. A. and D'Mello, S. 2010. "Affect Detection: An Interdisciplinary Review of Models, Methods, and Their Applications," *IEEE Transactions on Affective Computing*, vol. 1, no. 1, pp. 18-37.
- [2] Brooks, M., et al. 2013. "Statistical Affect Detection in Collaborative Chat," in *Proceedings of the 2013 conference on Computer Supported Cooperative Work (CSCW '13)*, San Antonio, Texas, USA, pp. 317-328.
- [3] Moniz, A. and Franciska de Jong. 2014. "Classifying the influence of negative affect expressed by the financial media on investor behavior," in *Proceedings of the 5th Information Interaction in Context Symposium (IiX '14)*, Regensburg, Germany, pp. 275-278.
- [4] Hu, M. and Liu, B. 2004. "Mining and summarizing customer reviews," in *Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '04)*, Seattle, pp. 168-177.
- [5] Dalal, M. K. and Zaveri, M. A. 2014. "Opinion Mining from Online User Reviews using Fuzzy Linguistic Hedges," *Applied Computational Intelligence and Soft Computing*, vol. 2014, Article ID 735942.
- [6] D'Mello, S. K. and Graesser, A. 2012. "Language and Discourse Are Powerful Signals of Student Emotions during Tutoring," *IEEE Transactions on Learning Technologies*, vol. 5, no. 4, pp. 304-317.
- [7] Melville, P., Gryc W., and Lawrence, R. D. 2009. "Sentiment Analysis of Blogs by Combining Lexical Knowledge with Text Classification," in *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'09)*, Paris, France, pp. 1275-1283.
- [8] Balahur, A., Kozareva, Z. and Montoyo, A. 2009. "Determining the polarity and source of opinions expressed in political debates," in *Proceedings of the 10th International Conference on Computational Linguistics and Intelligent Text Processing*, Mexico, pp. 468-480.
- [9] Durant, K. T. and Smith, M. D. 2006. "Predicting the Political Sentiment of Web Log Posts using Supervised Machine Learning Techniques coupled with Feature Selection," in *Proceedings of the 8<sup>th</sup> International workshop on Knowledge Discovery on the Web (WebKDD 2006)*, Philadelphia, pp. 187-206.
- [10] Gill, A. J., Gergle, D., French, R. M. and Oberlander, J. 2008. "Emotion Rating from Short Blog Texts," in *Proceedings of the SIGCHI conference on Human Factors in Computing Systems (CHI 2008)*, Florence, pp. 1121-1124.
- [11] Dey, L. and Haque, S. M. 2009. "Studying the effects of noisy text on text mining applications," in *Proceedings of the 3rd workshop on Analytics for Noisy Unstructured Text Data*, Barcelona, pp. 107-114.
- [12] Miller, G. A. 1995. "WordNet: A Lexical Database for English," *Communications of the ACM*, vol. 38, no. 11, pp. 39-41.
- [13] Sleator, D. and Temperley, D. 1993. "Parsing English with a Link Grammar," in *Proceedings of the 3rd International Workshop on Parsing Technologies*, Netherlands, pp. 1-14.
- [14] Klein, D. and Manning, C. D. 2003. "Fast exact inference with a factored model for natural language parsing," *Advances in Neural Information Processing Systems (NIPS 2002)*, Cambridge, MA: MIT Press, pp. 3-10.
- [15] Baccianella, S., Esuli, A. and Sebastiani, F. 2010. "SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining," in *Proceedings of the 7th International Conference on Language Resources and Evaluation (LREC '10)*, Malta, pp. 2200-2204.
- [16] Das, A. and Bandyopadhyay, S. 2010. "SentiWordNet for Indian Languages," in *Proceedings of the 8th Workshop on Asian Language Resources (COLING 2010)*, Beijing, China, pp. 56-63.
- [17] Ekman, P. 1992. "An Argument for Basic Emotions," *Cognition and Emotion*, vol. 6, pp. 169-200.
- [18] Douiji, Y. and Mousanif, H. 2015. "I-CARE: Intelligent Context Aware system for Recognizing Emotions from text," in *Proceedings of the 10th International Conference on Intelligent Systems: Theories and Applications (SITA)*, Rabat, Morocco.
- [19] Plutchik, R. 2001. "The Nature of Emotions," *American Scientist*, vol. 89, pp. 344-350.
- [20] Yamamoto, Y., Kumamoto, T. and Nadamoto, A. 2014. "Role of Emoticons for Multidimensional Sentiment Analysis of Twitter," in *Proceedings of the 16th International Conference on Information Integration and Web-based Applications & Service (iiWAS '14)*, Hanoi, Vietnam, pp. 107-115.

- [21] Scherer, K. and Wallbott, H. 1997. The ISEAR Questionnaire and Codebook.
- [22] Balahur, A. and Hermida, J. M. 2012. "Affect Detection from Social Contexts Using Commonsense Knowledge Representations," in Proceedings of the International Conference on Privacy, Security, Risk and Trust (PASSAT 2012), Amsterdam, Netherlands, pp. 884-892.
- [23] Li, X., Pang, J., Mo, B. and Rao, Y. 2016. "Hybrid Neural Networks for Social Emotion Detection over Short Text," in Proceedings of the International Joint Conference on Neural Networks, Vancouver, pp. 537-544.
- [24] Balahur, A., Hermida, J. M. and Montoyo, A. 2012. "Building and Exploiting EmotiNet, a Knowledge Base for Emotion Detection Based on the Appraisal Theory Model," IEEE Transactions on Affective Computing, vol. 3, no. 1, pp. 88-101.
- [25] Strapparava, C. and Valitutti, A. 2004. "WordNet-Affect: An affective extension of WordNet," in Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC 2004), Lisbon, pp. 1083-1086.
- [26] Pang, B., Lee, L. and Vaithyanathan, S. 2002. "Thumbs up? Sentiment classification using machine learning techniques," in Proceedings of the International Conference on Empirical Methods in Natural Language Processing (EMNLP 2002), Philadelphia, pp. 79-86.
- [27] Turney, P. D. 2002. "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, Philadelphia, pp. 417-424.
- [28] C. -P. Wei, Y. -M. Chen, C. -S. Yang, and C. C. Yang. 2010. "Understanding what concerns consumers: a semantic approach to product feature extraction from consumer reviews," Information Systems and e-Business Management, vol. 8, no. 2, pp. 149-167.
- [29] L. Zhao and C. Li. 2009. "Ontology based opinion mining for movie reviews," in Proceedings of the 3rd International Conference on Knowledge Science, Engineering and Management, Austria, pp. 204-214.
- [30] Turney, P. D. and Littman, M. L. 2003. "Measuring Praise and Criticism: Inference of Semantic Orientation from Association," ACM Transactions on Information Systems, vol. 21, no. 4, pp. 315-346.
- [31] M. Hu and B. Liu. 2004. "Mining opinion features in customer reviews," in Proceedings of the 19th National Conference on Artificial Intelligence (AAAI '04), San Jose, USA, pp. 755-760.
- [32] Popescu, A. and Etzioni, O. 2005. "Extracting Product Features and Opinions from Reviews," in Proceedings of the Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing (HLT/EMNLP '05), Vancouver, pp. 339-346.
- [33] Hatzivassiloglou, V. and McKeown, K. 1998. "Predicting the semantic orientation of adjectives," in Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and 8th Conference of the European Chapter of the Association for Computational Linguistics (ACL '98), pp. 174-181.
- [34] L. Dey and Sk. M. Haque. 2009. "Opinion Mining from noisy text data," International Journal on Document Analysis and Recognition, Springer-Verlag, vol. 12, no. 3, pp. 205-226.
- [35] Dalal, M. K. and Zaveri, M. A. 2013. "Semisupervised Learning Based Opinion Summarization and Classification for Online Product Reviews," Applied Computational Intelligence and Soft Computing, vol. 2013, Article ID 910706.
- [36] Malandrakis, N., Potamianos, A., Iosif, E. and Narayanan, S. 2013. "Distributional Semantic Models for Affective Text Analysis," IEEE Transactions on Audio, Speech, and Language Processing, vol. 21, no. 11, pp. 2379-2392.
- [37] Agrawal, R. and Srikant, R. 1994. "Fast algorithm for mining association rules," in Proceedings of the 20th International Conference on Very Large Data Bases, Santiago, Chile, pp. 487-499.
- [38] Church, K. W. and Hanks, P. 1990. "Word association norms, mutual information and lexicography," Computational Linguistics, vol. 16, no. 1, pp. 22-29.
- [39] Zadeh, L. A. 1975. "The Concept of a Linguistic Variable and its Applications to Approximate Reasoning (Part I)," Information Sciences, vol. 8, pp. 199-249.
- [40] Zadeh, L. A. 1975. "The Concept of a Linguistic Variable and its Applications to Approximate Reasoning (Part II)," Information Sciences, vol. 8, pp. 301-357.
- [41] Zadeh, L. A. 1975. "The Concept of a Linguistic Variable and its Applications to Approximate Reasoning (Part III)," Information Sciences, vol. 9, pp. 43-80.
- [42] Hatzivassiloglou, V. and Wiebe, J. M. 2000. "Effects of Adjective Orientation and Gradability on Sentence Subjectivity," in Proceedings of the 18th conference on Computational linguistics (COLING '00), Saarbrücken, Germany, pp. 299-305.
- [43] Polanyi, L. and Zaenen, A. 2006. "Contextual Valence Shifters," Computing Attitude and Affect in Text: Theory and Applications, The Information Retrieval Series, vol. 20, pp. 1-10.
- [44] Kennedy, A. and Inkpen, D. 2006. "Sentiment classification of movie reviews using contextual valence shifters," Computational Intelligence, vol. 22, no. 2, pp. 110-125.
- [45] A. -D. Vo and C. -Y. Ock. 2012. "Sentiment Classification: A Combination of PMI, SentiWordNet and Fuzzy Function," in Proceedings of the 4th International Conference on Computational Collective Intelligence Technologies and Applications, Vietnam, pp. 373-382.
- [46] Balahur, A. and Montoyo, A. 2009. "Semantic Approaches to Fine and Coarse-Grained Feature-Based Opinion Mining," in Proceedings of the 14th International Conference on Applications of Natural Language to Information Systems, Saarbrücken, Germany, pp. 142-153.
- [47] Peirsman, Y. and Pado, S. 2011. "Semantic Relations in Bilingual Lexicons," ACM Transactions on Speech and Language Processing, vol. 8, no. 2, Article 3.
- [48] SentiWordNet. [Online]. <http://sentiwordnet.isti.cnr.it>

- [49] Kumaran, A., Saravanan, K. and Maurice, S. 2008. "wikiBABEL: Community Creation of Multilingual Data," in Proceedings of the 4th International Symposium on Wikis (WikiSym '08), Article 14, Porto, Portugal.
- [50] Mohammad, S. M. and Turney, P. D. 2013. "Crowdsourcing a Word - Emotion Association Lexicon," *Computational Intelligence*, vol. 29, no. 3, pp. 436-465.
- [51] Pak, A. and Paroubek, P. 2010. "Twitter as a Corpus for Sentiment Analysis and Opinion Mining," in Proceedings of the 7th International Conference on Language Resources and Evaluation (LREC '10), Valletta, Malta, pp. 1320-1326.
- [52] Hogenboom A., et al. 2013. "Exploiting Emoticons in Sentiment Analysis," in Proceedings of the 28th Annual ACM Symposium on Applied Computing (SAC '13), Coimbra, Portugal, pp. 703-710.
- [53] J. Zhao, L. Dong, J. Wu, and K. Xu. 2012. "MoodLens: An Emoticon-Based Sentiment Analysis System for Chinese Tweets," in Proceedings of the 18th ACM SIGKDD International Conference on Knowledge discovery and data mining (KDD'12), Beijing, China, pp. 1528-1531.
- [54] Sweeney, L. 2002. "k-Anonymity: A model for protecting privacy," *International Journal on Uncertainty, Fuzziness and Knowledge-based Systems*, vol. 10, no. 5, pp. 557-570.
- [55] B. C. M. Fung, K. Wang, R. Chen, and P. S. Yu. 2010. "Privacy-Preserving Data Publishing: A Survey of Recent Developments," *ACM Computing Surveys*, vol. 42, no. 4, Article 14.
- [56] Brennan, M., Afroz, S., and Greenstadt, R. 2012. "Adversarial Stylometry: Circumventing Authorship Recognition to Preserve Privacy and Anonymity," *ACM Transactions on Information and System Security*, vol. 15, no. 3, Article 12.
- [57] Abbasi, A. and Chen, H. 2008. "Writeprints: A stylometric approach to identity-level identification and similarity detection in cyberspace," *ACM Transactions on Information Systems*, vol. 26, no. 2, Article 7.
- [58] Almishari, M., Kaafar, D., Oguz, E. and Tsudik, G. 2014. "Stylometric linkability of tweets," in Proceedings of the 13th Workshop on Privacy in the Electronic Society, Scottsdale, USA, pp. 205-208.