Toward a Sentiment Analysis Framework for Social Media

Bousselham EL HADDAOUI
ENSIAS, Mohammed V University
Rabat, Morocco
haddaoui.bousselham@gmail.com

Raddouane Chiheb
ENSIAS, Mohammed V University
Rabat, Morocco
r.chiheb@um5s.net.ma

Rdouan Faizi
ENSIAS, Mohammed V University
Rabat, Morocco
r.faizi@um5s.net.ma

Abdellatif El Afia
ENSIAS, Mohammed V University
Rabat, Morocco
a.elafia@um5s.net.ma

ABSTRACT

Nowadays, opinions and sentiments can be easily expressed through social media and have a strong social impact. Thus, the need for an automated way to analyze the generated data with less human effort and more accuracy. In this respect, sentiment analysis tasks such as; preprocessing, classification, etc. provides various techniques that achieves notable accuracy scores, but presents limitations depending on the experimental context.

Through our literature review, only few studies focused on establishing a reference framework for sentiment analysis. In this paper, we provide a literature review for common sentiment analysis tasks with discussion about future research trends, then we propose an abstraction model of a generic framework architecture for sentiment analysis in the context of social media based on previous works and enhanced with new concepts.

CCS Concepts

• Information systems→ Information retrieval • Computing methodologies→ Machine learning approaches algorithms • Computing methodologies→ Artificial intelligence.

Keywords

Sentiment Analysis, Social Media, Text Preprocessing, Machine Learning, Framework

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.

LOPAL '18, May 2–5, 2018, Rabat, Morocco
© 2018 Association for Computing Machinery.
ACM ISBN 978-1-4503-5304-5/18/05$15.00
https://doi.org/10.1145/3230905.3230919

1 INTRODUCTION

Research on the field of Sentiment Analysis (SA) have matured and keep advancing on various aspects such as: sentiment classification, document relevant attributes (opinion, holder, target, ...) extraction and other SA related tasks. Indeed, many algorithms and techniques have achieved reasonable scores. However, they still encounter some limitations and unable to resolve perfectly certain issues. In the micro level, problems such as: implicit sentiment and attributes detection, the sensitivity to context, domain and natural language processing related issues. In the macro level, an obvious finding is the absence of a reference Framework for SA that defines the work scope, gather proven techniques and suggest a generic-adaptive architecture of a SA framework in social media context.

In this paper, we will present the latest works on different known aspects of sentiment analysis and conclude with generic feature-based framework for sentiment analysis that could serve as a basis for future researches.

The rest of this paper is structured as follow. Section 2 presents a review of levels of analysis in SA and opinions extraction and representation. Section 3 provides a summary of preprocessing researches with important used techniques. Section 4 cover the classification task in SA and focus on the most used approaches. Section 5 highlight relevant research about SA frameworks for social media in which we suggest a feature-based model for a SA framework reference. Conclusions are discussed in Section 6.

2 LEVEL OF ANALYSIS, OPINION REPRESENTATION AND EXTRACTION

2.1 Levels of Analysis

Research on the field of sentiment analysis (SA) had increased activity over the last years. The ambition of extracting opinions from complex data structures, different sources and domains has led researchers to perform extensive but not yet completed work on this area. Many level of analysis were discussed, Kumar and Sebastian [1] depicted the levels of sentiment analysis into: word level that encompass dictionary and corpus based SA, sentence and document level.

2.1.1 Word Level

Back to the 1997, Hatzivassiloglou and McKeown [2] achieved 82% accuracy in determining the semantic orientation for adjectives and continue with exploring verbs and nouns as part of performing opinion mining at the word level. Esuli and Sebastiani [3] used a semi-supervised approach that uses words

representations enriched with glosses from online dictionaries. The method performed well comparing to 3 selected approaches, further research on the effect of selected glosses on the approach performance.
Salas-Zárate et al. [4] used ontologies and N-grams based method to perform semantic sentiment analysis at the aspect-level on diabetes oriented tweets. The results indicated that N-gram around gives best precision, recall and F-measure than N-gram after and before, however this requires an established domain-specific ontology and wasn't designed for a multilingual context.

2.1.2 Sentence Level

At the sentence level, features extraction and overall sentiment classification considering the context is the biggest concern of SA. Kim and Hovy [5] suggested a new approach that given a topic, detects the sentiment and the opinion holder. The authors used manually selected seed words expanded with their synonyms using WordNet to help classifying words sentiment and then the whole sentences. Adding more seed words helped improving the Recall to more than 90%. Neviarouskaya et al. [6] proposed an Affect Analysis Model (AAM) for affect analysis at the sentence level, the rule-based approach performed well in at least 2 out of 3 human annotated sentences but depends heavily on the selected database and proved limitations in sensing sentences without context. Yang and Cardie [7] used a context-constraints based approach and trained a Conditional Random Field Model (CRF) via Posterior Regularization (PR) with labeled data to build an effective method for sentiment analysis at the sentence level. The approach out performed in accuracy the existing supervised and semi-supervised models but proved limitations in performance since it's based on hard constraints.

2.1.3 Document Level

The research at this level is driven by the ambition to cover large text structure: reviews, discussions in forums, ... it aims to extract the document polarity given by a single opinion holder. Tang [8] suggested a sentiment-specific semantic representation learning framework based on 4 pillars: Word Embedding, Sentence Structure, Sentence Composition to conclude with Document Composition. The authors stated that further work should be done at every phase especially at the document composition, the method needs to leverage sentiment discourse relations.

Bhatia et al. [9] suggested a Rhetorical Structure Theory (RST) parsing based approach than can improve document level sentiment analysis through a reweighting of the discourse units. Although many RST parsers are provided, the reweighting function performance depends heavily on the training data.

Another challenge in this field is domain modeling and domain-specific sentiment classification. Denecke and Deng [10] had established a categorization of medical sentiments at different levels: health status, medical condition, diagnosis, effect of a medical event, medical procedure, medications. Researchers covered other domains to enlarge the scope of sentiment analysis at the document level.

As presented, studies have covered each level of analysis with a focus on word and sentence levels. Challenges such as word meaning variation and ambiguity through contexts still an obstacle to design efficient methodologies since it requires domain definition, word disambiguation and multilingual support.

2.2 Opinion Representation and Extraction

2.2.1 Opinion Representation

A machine-understandable representation of opinion that could be extracted from unstructured opinionated text was needed in order facilitate the sentiment analysis task. Liu [11] suggested quintuple defined by a target entity, aspect of the entity, sentiment, holder and time (Ej, Ajk, SOijkl , Hi , Tl). The sentiment can be either a polarity (positive, negative, neutral, etc. rating (staring, gamified system, etc.) or an emotion (love, joy, surprise, anger, sadness, fear, etc.).
In the context of social media, the location of the opinion holder can be also retrieved and used. Almatrafi et al. [12] performed a location-based sentiment analysis and highlighted the importance of location in the process of sentiment analysis, and to the best of our knowledge no research study has covered the impact of time and location on the expressed opinion.

2.2.2 Opinion Aspects Extraction

A challenging yet trending research area, aspects extraction become very popular since the raise of online discussions and reviews spaces such as movies websites, product reviews forums and discussion threads on social media.

Asghar et al. [13] presented a categorization of features based on a literature review as follows: morphological types (semantic, syntactic classes and lexicon structural), frequent or hot features and implicit features to which NLP based, statistical, clustering based and hybrid techniques are applied for extraction purposes. Preprocessing and feature cleaning without disturbing the aspects meaning are required to avoid unnecessary computation.

Feature extraction approaches can be divided into two categories: supervised and unsupervised methods. In the context of supervised methods, Li and Lu [14] introduced the notion of sentiment scope in which opinion feature is embedded, a graphical model based approach is applied then to extract both the entities and their sentiment information. The method achieved better results comparing to existing supervised approaches based on Conditional Random Fields (CRFs), but still have to overcome limitations related to scopes overlapping.

Although many research initiatives focused on domain-specific and supervised feature extraction, other related works have suggested alternatives to bypass current challenges. Yongmei and Hua [15] presented M-Score, an algorithm based on PMI (Pointwise Mutual Information) algorithm that support domain-independent aspects extraction. The algorithm performed better than the PMI in precision and recall, it's also a candidate solution to solve the feature extraction domain limitation. On the other hand, Federici and Dragoni [16] highlighted the need of an unsupervised approach due to annotated data sets and multi-domain opinions limitations. Indeed, they suggested two unsupervised approaches: a syntax-tree and a grammar-dependencies based approaches, results were comparable to trained system and opened for future work at this level.

Further research studies are to be expected in the following areas:

- Unsupervised, domain-independent feature extraction approaches
- Implicit features identification
- Relevant feature identification which concerns feature dimension reduction

3 DATA PREPROCESSING

The task of preprocessing extracted data for analysis is a key pillar in sentiment analysis process, it consists of reducing opinion noise and impact positively SA related tasks such as features extraction and sentiment classification without compromising the opinion meaning.

From a syntactic perspective, extracted data usually includes extra words, called "noise", which increases features space dimension. In this respect, Jianqiang and Xiaolin [17] discussed the impact of preprocessing on classification algorithms, experiments showed that stop words, repeated letters, numbers and URLs removal have a minimal impact on classifiers, and moreover, it can optimize significantly computation resources. The authors also found that replacing negation and expanding acronyms improves classifiers accuracy, whereas the preprocessing methods efficiency depends heavily on stop words data sets and acronyms dictionaries selection.

Gull et al. [18] highlighted the need to clean and transform social media data sets into a structured form, the cleaning steps included: lower case text conversion, extract change of direction indicators and removing emoticons, punctuations and URLs. Specific cleaning techniques were imposed due to the extracted data constraints, but they agreed also on removing URLs since they don't hold interesting information for analysis.

Asghar et al. [13] in his review, indicated three important preprocessing techniques:

- Part of Speech (POS) tagging, which is efficient for explicit features extraction, but lacks performance when text contains implicit ones.
- Stemming and Lemmatization: Stemming refer to the conversion of a word to its root form without context consideration, whereas Lemmatization is context-aware and rely on additional dictionary data to enhance accuracy.
- Stop Word Removal: words with high frequency in the text and do not contains relevant information for the analysis.

This provided selection doesn't contain all the known techniques, thus was the need to have a more exhaustive list. Effrosynidis et al. [19] listed 15 techniques for data preprocessing and based on experimentation they suggested the use of: stemming, repeated punctuation replacement and removing numbers where other techniques such as: removing punctuation, handling capitalized words, replacing slang and negations and spelling correction were not recommended.

Although many preprocessing techniques are available, there isn't an existing component that can perform this task effectively. From specifying which data should be removed and those to be expanded to transform data into structured and optimized sets for analysis, this framework should consider techniques combinations, domain and data independent techniques.

4 CLASSIFICATION

Classification is the main task for sentiment analysis, all the previous steps were improved to allow a better accuracy and performance in determining the polarity of text. Classifiers usually presents limitations in situations such as: sarcasm, irony, use of slang, etc. but with the significant work on preprocessing techniques, the challenges changed allowing researchers to dive into advanced classification techniques and explore new ways to perform this task.

Usually, classification techniques fall into two categories: supervised and unsupervised methods. Supervised methods rely heavily on annotated data whereas unsupervised techniques are designed to work on domain independent and unannotated data sets. These techniques use general purpose machine learning classifiers (i.e. Support Vector Machine - SVM, Naive Bayes - NB, Random Forest, Logistic Regression, etc.) to perform classification task. In this respect, researches were conducted to tackle two major problems; which techniques to use and what algorithms could achieve high accuracy in the classification task.

Yang et al. [20] compared three machine learning classifiers in radical opinion identification context, the experimentations comparing SVM, Naive Bayes and AdaBoost using annotated domain-specific features resulted in SVM outperforming the other classifiers. They also stated that adding more features will improve the classifier effectiveness. Alabbas et al. [21] confirmed also this finding in the context of detecting high-risk floods, SVM performed better than other classifiers (J48, C5.0, NNET, NB and K-NN) in term of accuracy. Furthermore, they indicated that using SVM without stemming was significantly better.

Despite the performance of classic classifiers, they are prone to lose semantic relations between features which gives us inaccurate results. You et al. [22] proposed deep learning system based on Recurrent Neural Networks model in the context of text security audit. They have used CBOW and Word2Vec for features extraction and data training, and RNN for classification. In their context the approach outperformed SVM and achieved 92.7% accuracy. Another approach combining DAN2 (Dynamic Architecture for Neural Network) and feature engineering to classify brand-related tweets was suggested by Zimbra et al. [23], the proposed technique and SVM outperformed the state-of-the-art systems (Sentiment140 and Repustate). The authors stated that their approach was more accurate for classifying mild sentiments than SVM.

The classification task can be a time consuming when it relies mostly on annotated data. Thus, more efforts were dedicated to semi-supervised and unsupervised techniques. Appel et al. [24] in his hypothesis, suggested that hybrid approaches based on unsupervised learning, fuzzy set and a strong domain knowledge can match supervised learning techniques performance. Hence, this research area still relevant and should be considered by future studies. With this perspective, Kaati et al. [25] combined data-dependent and data-independent features to perform the classification task, experimentations show that using AdaBoost (Adaptive Boosting) classifier in this context gave promising results. AdaBoost outperformed other classifiers such as SVM

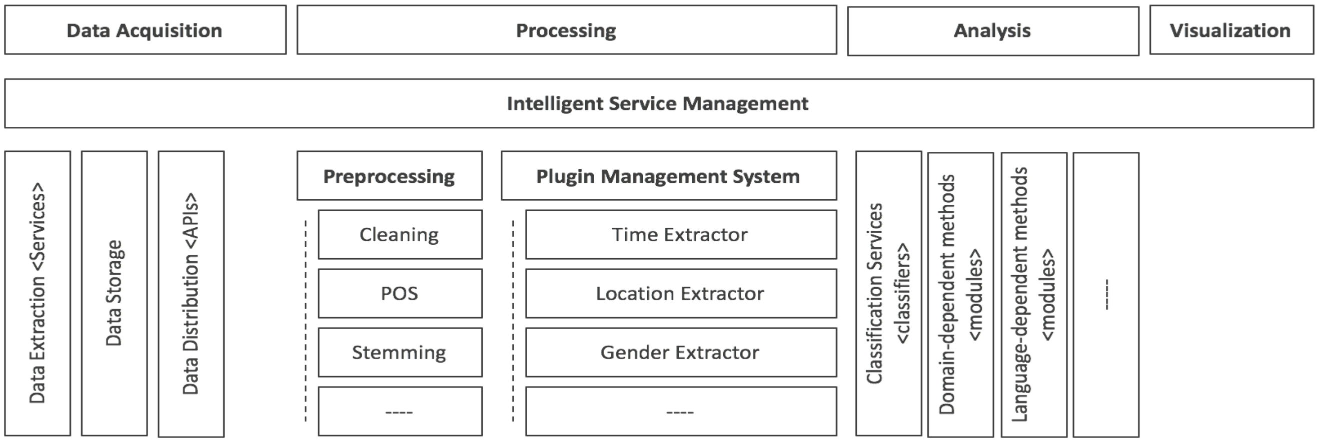

Figure 1. Feature based architecture for a SA framework

(used with default parameters). García-Pablos et al. [26] proposed a near unsupervised system (W2VLDA) based on Latent Dirichlet Allocation (LDA) that could perform sentiment analysis tasks with minimal pre-configuration and outperform similar existing systems. W2VLDA needs at least one-word seed for every domain aspect and it translation to handle multilingual environments, which tackled partially the classification task on multilingual environment since most studies focuses on English. The authors stated also the need for an automated way to handle preprocessing tasks.

Having covered supervised and unsupervised latest research findings, we noticed: First, classifiers are sensitive to preprocessing techniques which was also discussed in Jianqiang and Xiaolin [17] comparison research. They observed that some classifiers are more sensitive to preprocessing techniques than others, in their experiments: Naive Bayes and Random Forest were more sensitive to preprocessing methods than SVM and Logistic Regression. Second, hybrid methods are showing promising results and become a research trend in the area, but have many parameters (preprocessing techniques, classifier choice, etc.) to consider and problems (multilingual classification, semantic handling, etc.) to tackle. Third, a rising movement of applying techniques such as neural networks, deep learning is providing alternatives to current limitations, this can improve significantly researches on hybrid approaches and bring unsupervised techniques to a new level. Last, the need for a common framework encompassing existing state-of-the-art classification techniques and a context-aware recommender system as a layer that could suggest the optimal configuration to achieve high accuracy.

5 SENTIMENT ANALYSIS FRAMEWORK FOR SOCIAL MEDIA

Research on sentiment analysis have achieved a reasonable maturity level with a set of techniques and approaches covering all known issues, this resulted in high expectations from different stakeholders (i.e. governments, companies, organizations, etc.). Moreover, the accelerated rate of data production through social media and the frequent aspect of the task of SA have pushed the existing studies to their limits proving the need to an integrated framework that could capitalize on the state-of-the-art approaches and orchestrate the different components tasks and interactions.

On the features aspect, researches focused to achieve a generic yet extensible framework feature-based architecture that could handle sentiment analysis issues related to: multiple data sources, multilingual text, domains, etc. In this respect, Ali et al. [27] proposed a multi-channels SA framework to identify the location of disease outbreaks. The framework was defined by two major layers, a system models where service identification, task and evaluation attributes are defined, it covers the extraction, processing, analysis and visualization tasks. The top layer is service composition in which they provided a rule-based system that compose a set of services for every requested task. The framework performed well in the defined context but rely mostly on a manual and domain-dependent rules services attributes definition which is time consuming. Another framework for assessing brand's presence was suggested by Aggrawal et al. [28], the framework was designed for online marketing purposes and presented 3 components for brands presence evaluation. First, collected data from search engines were annotated (source, url, page rank) and transformed into graphs then Text Edge algorithm was applied to evaluate the nodes strength in accordance with terms, concepts and brands. Second, crawling and scraping operations were used to retrieve candidate data, then LDA was used to filter relevant data. Last, a Naive Bayes sentiment analysis-based component was used to determine brands sentiment affinity. The three components served as a basis to defined a new page rank, which is used by brands to defined their marketing strategy. The framework achieved promising results but presented certain limitations such as: inputs didn’t include data from social media sources, data cleaning used only general-purpose techniques and the sentiment analysis process didn’t consider various classifiers. Nouh et al. [29], in the context of cybersecurity, presented a framework blueprint that defined 3 layers: data-handling, analysis and a front-end. The data-handling layer retrieve, store and process data from various sources (i.e.

social media, web forums, police reports, etc.), and assign levels of uncertainty with a confidence score to each item. The analysis layer encompassed many features such as: identification and extraction, profiling, signal detection, behavior analysis, etc. The last layer which is the user frontend layer cover the exploration and visualization aspects. The framework presented a multipurpose generic architecture for a sentiment analysis system but didn't address issues related to pre-processing techniques and classifiers selection criteria.

On the infrastructure perspective, a growing movement towards new infrastructure trends such as cloud computing and big data is adopted to leverage the resources required from a SA dedicated framework. Zarra et al. [30] suggested a cloud-based sentiment analysis framework for e-learning discussion forums. The sentiment analysis component contains 3 main modules which handles data extraction and language detection, data cleaning and preprocessing, and a machine learning module that uses Naive Bayes. Woldemariam [31] extended an existing SOA (Service Oriented Architecture) and Hadoop stack based cross-media analysis framework with a sentiment analysis pipeline which covers both preprocessing and sentiment computation tasks. The author proposed a module for text cleaning using a set of techniques (i.e. POS, remove non-standard letters and repeated spaces, etc.), a module that encompassed NLP usual tasks (i.e. tokenization, stemming, etc.) then a sentiment analysis module which performs data annotation and sentiment computation.

In order to fill the research gap in terms of defining a unified sentiment analysis framework, we propose in Figure 1 a feature-based architecture based on the previous studies and enhanced with facts from the previous sections. Most studies agreed on the following components: Information Extraction, Processing, Information and Visualization as a generic overview of a framework's architecture. Thus, we presented a 3 layers model as follow:

Components Level

This layer presents the core components needed in a SA process which includes data acquisition, processing, analysis and visualization. The role of each component is explained as follows:

- Data acquisition: component focus on data extraction from various sources and in different formats, it also handles the storage and distribution operations.
- Data processing: component includes both preprocessing operations and a plugin management system that can achieve other specific tasks such as: time and location extraction.
- Data analysis: component achieves the sentiment computation task, it contains a set of proven classifiers in the field of SA and provide an extensible interface to integrate state-of-the-art complete techniques.
- Data visualization: component ensures an easy way to explore and present data to end users.

Orchestration Level

The orchestration layer should define carefully the optimal resources configuration for a specific SA task, through a semantic rule-based engine that will serve as a basis for the resources selection process. In a simple way, this layer should recommend including the repeated letters and slang handling in the preprocessing phase if the data source is a social network.

Services Level

The services layer contains the corresponding tasks for each top-level component, the service definition should contain at least the following attributes:

- Identification: a unique identifier of the service in the framework
- Description: used for documentation purposes
- Performance attributes: a set of performance indicators which can be domain or language dependent, etc.
- Selection attributes: a set of attributes that could help designing the service for a specific task
- Dependencies: a semantic rule-based definition that help identify the service requirements
- Custom meta-data: can be added to address specific cases.

We believe that the proposed feature-based framework architecture can serve a basis for future works about the subject in order to define an extensible and complete framework for sentiment analysis.

6 CONCLUSION

Sentiment analysis have reached an advanced maturity level in many aspects, which implies capitalizing the findings in a structured way to provide a common reference for researchers in this field. In our literature review, we examined an important number of studies covering various task of SA from which, we selected the most relevant documents to our subject. As a result, we provided the latest studies covering opinion modeling, extraction, data preprocessing and classification approaches. We also discussed limitations and future works at every stage and suggested enhancements in some cases: enrich opinion representation with more domain independent features such as location. Last, we provided and generic feature-based architecture for our future SA framework bringing innovative concepts such as: orchestration recommender layer that helps define in a semantic way the best services to use for a specified task, a service representation which consists of service identification, performance, selection and dependencies metadata.

In our future work, we will perform deep research to build a ready to use component at every stage and enhance the proposed framework with key new findings. Our approach will be to focus work on a specific component and use proven techniques on the others, then iterate on the presented components to deliver a candidate sentiment analysis framework for social media.

REFERENCES

[1] A. Kumar and T. M. Sebastian, "Sentiment Analysis: A Perspective on its Past, Present and Future," *Int. J. Intell. Syst. Appl.*, vol. 4, no. 10, pp. 1–14, Sep. 2012.

[2] V. Hatzivassiloglou and K. R. McKeown, "Predicting the semantic orientation of adjectives," in *Proceedings of the eighth conference on European chapter of the Association for Computational Linguistics*, 1997, pp. 174–181.

[3] A. Esuli and F. Sebastiani, "Determining the semantic orientation of terms through gloss classification," in *Proceedings of the 14th ACM international conference on Information and knowledge management*, 2005, pp. 617–624.

[4] M. del P. Salas-Zárate, J. Medina-Moreira, K. Lagos-Ortiz, H. Luna-Aveiga, M. Á. Rodríguez-García, and R. Valencia-García, "Sentiment Analysis on Tweets about Diabetes: An Aspect-Level Approach," *Comput. Math. Methods Med.*, vol. 2017, pp. 1–9, 2017.

[5] S.-M. Kim and E. Hovy, "Determining the sentiment of opinions," in *Proceedings of the 20th international conference on Computational Linguistics*, 2004, p. 1367.

[6] A. Neviarouskaya, H. Prendinger, and M. Ishizuka, "Textual affect sensing for sociable and expressive online communication," *Affect. Comput. Intell. Interact.*, pp. 218–229, 2007.

[7] B. Yang and C. Cardie, "Context-aware Learning for Sentence-level Sentiment Analysis with Posterior Regularization.," in *ACL (1)*, 2014, pp. 325–335.

[8] D. Tang, "Sentiment-Specific Representation Learning for Document-Level Sentiment Analysis," 2015, pp. 447–452.

[9] P. Bhatia, Y. Ji, and J. Eisenstein, "Better document-level sentiment analysis from rst discourse parsing," *ArXiv Prepr. ArXiv150901599*, 2015.

[10] K. Denecke and Y. Deng, "Sentiment analysis in medical settings: New opportunities and challenges," *Artif. Intell. Med.*, vol. 64, no. 1, pp. 17–27, May 2015.

[11] B. Liu, "Sentiment Analysis and Opinion Mining," *Synth. Lect. Hum. Lang. Technol.*, vol. 5, no. 1, pp. 1–167, May 2012.

[12] O. Almatrafi, S. Parack, and B. Chavan, "Application of location-based sentiment analysis using Twitter for identifying trends towards Indian general elections 2014," in *Proceedings of the 9th International Conference on Ubiquitous Information Management and Communication - IMCOM '15*, 2015, pp. 1–5.

[13] M. Z. Asghar, A. Khan, S. Ahmad, and F. M. Kundi, Eds., "A Review of Feature Extraction in Sentiment Analysis," 2014.

[14] H. Li and W. Lu, "Learning Latent Sentiment Scopes for Entity-Level Sentiment Analysis.," in *AAAI*, 2017, pp. 3482–3489.

[15] S. Yongmei and H. Hua, "Research on Domain-independent Opinion Target Extraction," *Int. J. Hybrid Inf. Technol.*, vol. 8, no. 1, pp. 237–248, Jan. 2015.

[16] M. Federici and M. Dragoni, "Towards Unsupervised Approaches For Aspects Extraction.," in *EMSA-RMed@ ESWC*, 2016.

[17] Z. Jianqiang and G. Xiaolin, "Comparison Research on Text Pre-processing Methods on Twitter Sentiment Analysis," *IEEE Access*, vol. 5, pp. 2870–2879, 2017.

[18] R. Gull, U. Shoaib, S. Rasheed, W. Abid, and B. Zahoor, "Pre Processing of Twitter's Data for Opinion Mining in Political Context," *Procedia Comput. Sci.*, vol. 96, pp. 1560–1570, 2016.

[19] D. Effrosynidis, S. Symeonidis, and A. Arampatzis, "A Comparison of Pre-processing Techniques for Twitter Sentiment Analysis," in *Research and Advanced Technology for Digital Libraries*, vol. 10450, J. Kamps, G. Tsakonas, Y. Manolopoulos, L. Iliadis, and I. Karydis, Eds. Cham: Springer International Publishing, 2017, pp. 394–406.

[20] M. Yang, M. Kiang, Y. Ku, C. Chiu, and Y. Li, "Social Media Analytics for Radical Opinion Mining in Hate Group Web Forums," *J. Homel. Secur. Emerg. Manag.*, vol. 8, no. 1, Jan. 2011.

[21] W. Alabbas, H. M. al-Khateeb, A. Mansour, G. Epiphaniou, and I. Frommholz, "Classification of colloquial Arabic tweets in real-time to detect high-risk floods," 2017, pp. 1–8.

[22] L. You, Y. Li, Y. Wang, J. Zhang, and Y. Yang, "A deep learning-based RNNs model for automatic security audit of short messages," in *Communications and Information Technologies (ISCIT), 2016 16th International Symposium on*, 2016, pp. 225–229.

[23] D. Zimbra, M. Ghiassi, and S. Lee, "Brand-Related Twitter Sentiment Analysis Using Feature Engineering and the Dynamic Architecture for Artificial Neural Networks," 2016, pp. 1930–1938.

[24] O. Appel, F. Chiclana, and J. Carter, "Main concepts, state of the art and future research questions in sentiment analysis," *Acta Polytech. Hung.*, vol. 12, no. 3, pp. 87–108, 2015.

[25] L. Kaati, E. Omer, N. Prucha, and A. Shrestha, "Detecting Multipliers of Jihadism on Twitter," 2015, pp. 954–960.

[26] A. García-Pablos, M. Cuadros, and G. Rigau, "W2VLDA: Almost unsupervised system for Aspect Based Sentiment Analysis," *Expert Syst. Appl.*, vol. 91, pp. 127–137, May 2017.

[27] K. Ali, H. Dong, A. Bouguettaya, A. Erradi, and R. Hadjidj, "Sentiment Analysis as a Service: A Social Media Based Sentiment Analysis Framework," 2017, pp. 660–667.

[28] N. Aggrawal, A. Ahluwalia, P. Khurana, and A. Arora, "Brand analysis framework for online marketing: ranking web pages and analyzing popularity of brands on social media," *Soc. Netw. Anal. Min.*, vol. 7, no. 1, Dec. 2017.

[29] M. Nouh, J. R. Nurse, and M. Goldsmith, "Towards Designing a Multipurpose Cybercrime Intelligence Framework," in *Intelligence and Security Informatics Conference (EISIC), 2016 European*, 2016, pp. 60–67.

[30] T. Zarra, R. Chiheb, R. Faizi, and A. El Afia, "Cloud computing and sentiment analysis in E-learning systems," in *Cloud Computing Technologies and Applications (CloudTech), 2016 2nd International Conference on*, 2016, pp. 171–176.

[31] Y. Woldemariam, "Sentiment analysis in a cross-media analysis framework," in *Big Data Analysis (ICBDA), 2016 IEEE International Conference on*, 2016, pp. 1–5.