

An explanatory Study of Sentiment Analysis and Sentiment Classification Approaches

Amit Rakshit

Assistant Professor, Department of Mechanical Engineering, Kanad Institute of Engineering and Management, Purba - Bardhaman

ABSTRACT

Recently, sentiment analysis (SA) is one of the common methods in decision making field. The SA is a procedure of automated classification/extraction sentiment from the client opinion/reviews by computation, Natural Language Processing (NLP), and text mining methods. It aims to localize the opinion, recognize the sentiment and categorize the separation of this sentiments. It is beneficial in several methods like individuals, business, marketing and governments. In this study, a broad analysis of several SA methods is made with their datasets, advantages, objectives, efficiency outcomes and underlying methods. A comprehensive relation is created for analyzing the distinct features of several methods based on language, scope, polarity, task, domain oriented and dataset. The application of SA in various fields are deliberated. This study provides the entire depiction of SA methods.

Keywords: Social media, Emotions, Sentiment analysis, Sentiment classification, Opinion Mining

I. INTRODUCTION

The SA/opinion Mining (OM) is the procedure of automated classification/extraction of sentiments from the client's opinion/review using computation, NLP and text mining methods. The 2 terms are similar and refers to similar implication. On the other hand, few scientists illustrate that these 2 words are differ from one another. The OM investigates and extracts the view regarding an entity when SA seeks the opinion, detect the sentiment stated and lastly categorizes the polarity. It detects beneficial in several domains like online shopping, government, marketing, business, and so on. Recently, the end user demonstrates their sentiment by review/comment using internet. These analyses are utilized and examined via main businesses for making decisions. Mostly, online shopping sites (for example Amazon, Flipkart) inspire the user to offer opinion as sentiment on various features of the obtained product. Hundreds to thousands of clients depend on online analysis and ninety percentage of user decides for purchasing the product on the basis of analysis in April 2013. The retailer utilizes this review for enhancing the product quality and recognize the client fulfillment. Other customers utilize this review for making decision to buy the product/not.

The dataset utilized in SA performs a major part. There are several methods for collecting SA information. The significant information is gathered from the product analyses. It is useful to the retailer as they could alter the quality of product on the basis of SA outcomes. The main reviews source is the reviews website. It is utilized to the product, it is beneficial for discusses of political parties, stock exchange, and news. From political discusses, the people opinions on election

candidates are defined. The blog and Social media are the better source of data because of data distribution. Few estimated datasets are presented for performing research. Recently, several enhancement and application on SA techniques are presented. This study gives an entire depiction of SA application and methods. In this study, classification of several SA methods is made. Several studies are performed and articles are preserves on raising gradually. The 2 elaborate studies are proposed that is concentrates on the applications and problems in SA. The methods for overcoming the problems are deliberated. Few studies are proposed and made in this survey. This study is distinct from present research in the succeeding manners. At first, the classification of several current SA methods is offered. It is beneficial to the researcher for selecting essential method to particular application. Then, numerous methods of SA are categorized with the details of every methods. It offers an entire opinion of SA domain. Next, the association of every studied methods are created on the basis of various measures. Lastly, applications and problems in the SA fields are deliberated.

II. SENTIMENT CLASSIFICATION TECHNIQUES

SA is assumed as a Sentiment Classification (SC) procedure and deliberated in this segment. The SC methods are mainly separated to three kinds: The lexicon, hybrid and ML approaches. The ML technique utilizes linguistic characteristics. The lexicon dependent approach is depending upon sentimental lexicon, a collection of identified sentimental words. It is categorized to corpus and

dictionary dependent approaches. The semantic polarity is defined by semantic/statistical approaches. The hybrid method integrates lexicon and ML methods. The classification hierarchy of SA methods. The ML methods are categorized to unsupervised and supervised methods. The supervised technique utilizes several label trained documents. The unsupervised approaches are utilized in conditions if it is difficult for identifying this label trained documents. Then, Lexicon dependent method is depending upon recognizing the review lexicon is utilized for investigating the. It is separated in 2 manners: corpus and dictionary dependent methods. The dictionary-based method is depending upon finding review seed word and detect dictionary of opposites and meanings. The corpus-based technique initiates list of sentimental words and search for another sentimental words with context certain orientation. Lastly, the corpus-based technique is separated to semantic/statistical methods.

Several NLP methods are integrated with lexicon based approaches for identifying syntactical arrangement and find semantic relationship [1]. The NLP methods are utilized in the preprocessing level preceding projected lexicon dependent SA technique [2]. It includes automatic focus detection and sentimental analyses modules that can evaluate the user's opinion of news. [3] utilized NLP in various factors. It utilizes NLP technique to detect time and tense expressions as ranking and mining approaches. It involves 2 variables that obtain time expression and detect linguistic clues from crawled information by NLP approaches. It utilizes product review from Amazon and their outcome demonstrates that the variables are more beneficial [4].

Several approaches are not categorized as ML/lexicon based approaches, for example Formal Concept Analysis (FCA). It is utilized to arrange, analyze and visualize the information by Galois association [5]. The information includes set of entities and features are named formal concept. The Fuzzy Formal Concept Analysis (FFCA) is presented to unclear and uncertain data. The FFCA and FCA are utilized in SA properties [6, 7]. A classification architecture is proposed by FFCA for presenting document to various models. [8] utilize FCA for constructing an ontology module. Ontology dependent methods are introduced for an efficient SA of twitters post by separating every tweet to a set of factors. Smart phones are utilized and the presented architecture gives a thorough study of posts opinion depending upon certain topics. pSenti is a model level SA [9] that is combined with OM and learning based approaches. It attains an improved classification accurateness in polarity classifiers related to regular lexicon based

methods. It is depending upon two practical datasets and it executes well compared to SentiStrength.

SentiNet2 is presented in [10], free source semantic and efficient resource to OM and SA for compensating the gap among word level natural language information and model level sentiment. It is established using Semantic Web and AI. It exposes that the scheme is simply combined to a practical application for effectively integrate and relate unstructured and structured information's. The Concept level SA is employed for e-health to person review analyses [11] and crowd authentication.

III. CONCLUSION

SA is a procedure of automated classification/extraction sentiment from the client opinion/reviews by computation, Natural Language Processing (NLP), and text mining methods. It aims to localize the opinion, recognize the sentiment and categorize the separation of this sentiments. It is beneficial in several methods like individuals, business, marketing and governments. In this study, a broad analysis of several SA methods is made with their datasets, advantages, objectives, efficiency outcomes and underlying methods. A comprehensive relation is created for analyzing the distinct features of several methods based on language, scope, polarity, task, domain oriented and dataset. The application of SA in various fields are deliberated. This study provides the entire depiction of SA methods.

REFERENCES

- [1] Bolshakov Igor A, Gelbukh Alexander. *Comput Linguis (Models, Resources, Applications)* 2004.
- [2] Moreo A, Romero M, Castro JL, Zurita JM. Lexicon-based comments-oriented news sentiment analyzer system. *Expert Syst Appl* 2012;39:9166–80.
- [3] Min Hye-Jin, Park Jong C. Identifying helpful reviews based on customer's mentions about experiences. *Expert Syst Appl* 2012;39:11830–8.
- [4] Zhou L, Li B, Gao W, Wei Z, Wong K. Unsupervised discovery of discourse relations for eliminating intra-sentence polarity ambiguities. In: Presented at the 2001 conference on Empirical Methods in Natural Language Processing (EMNLP'11); 2011.
- [5] Priss U. Formal concept analysis in information science. In: Presented at the annual review of information science and technology; 2006.

- [6] Li S, Tsai F. Noise control in document classification based on fuzzy formal concept analysis. In: Presented at the IEEE International Conference on Fuzzy Systems (FUZZ); 2011.
- [7] Li Sheng-Tun, Tsai Fu-Ching. A fuzzy conceptualization model for text mining with application in opinion polarity classification. *Knowl-Based Syst* 2013;39:23–33.
- [8] Kontopoulos Efstratios, Berberidis Christos, Dergiades Theologos, Bassiliades Nick. Ontology-based sentiment analysis of twitter posts. *Expert Syst Appl* 2013.
- [9] Mudinas Andrius, Zhang Dell, Levene Mark. Combining lexicon and learning based approaches for concept-level sentiment analysis. Presented at the WISDOM'12, Beijing, China; 2012.
- [10] Cambria Erik, Havasi Catherine, Hussain Amir. SenticNet 2: a semantic and affective resource for opinion mining and sentiment analysis. In: Proceedings of the twenty-fifth international florida artificial intelligence research society conference; 2012.
- [11] Cambria Erik, Benson Tim, Eckl Chris, Hussain Amir. Sentic PROMs: application of sentic computing to the development of a novel unified framework for measuring health-care quality. *Expert Syst Appl* 2012;39: 10533–43.
- [12] Alzubi, O. A. (2015, September). Performance evaluation of AG block turbo codes over fading channels using BPSK. In Proceedings of the The International Conference on Engineering & MIS 2015 (pp. 1-6).
- [13] Kavitha, M., & Palani, S. (2014). Hierarchical classifier for soft and hard exudates detection of retinal fundus images. *Journal of Intelligent & Fuzzy Systems*, 27(5), 2511-2528.
- [14] Anuradha, M., Jayasankar, T., Prakash, N. B., Sikkandar, M. Y., Hemalakshmi, G. R., Bharatiraja, C., & Britto, A. S. F. (2021). IoT enabled cancer prediction system to enhance the authentication and security using cloud computing. *Microprocessors and Microsystems*, 80, 103301.
- [15] S. Namasudra, S. Dhamodharavadhani, and R. Rathipriya, "Nonlinear neural network based forecasting model for predicting COVID-19 cases", *Neural Processing Letters*, 2021. DOI: 10.1007/s11063-021-10495-w
- [16] Elhoseny, M., & Shankar, K. (2019). Reliable data transmission model for mobile ad hoc network using signcryption technique. *IEEE Transactions on Reliability*, 69(3), 1077-1086.
- [17] Mukherjee, R., Kundu, A., Mukherjee, I., Gupta, D., Tiwari, P., Khanna, A., & Shorfuzzaman, M. (2021). IoT-cloud based healthcare model for COVID-19 detection: an enhanced k-Nearest Neighbour classifier based approach. *Computing*, 1-21.
- [18] Li, L., Sun, L., Xue, Y., Li, S., Huang, X., & Mansour, R. F. (2021). Fuzzy Multilevel Image Thresholding Based on Improved Coyote Optimization Algorithm. *IEEE Access*, 9, 33595-33607.
- [19] Alzubi, O. A. A deep learning-based frechet and dirichlet model for intrusion detection in IWSN. *Journal of Intelligent & Fuzzy Systems*, (Preprint), 1-11.
- [20] Kavitha, M., & Palani, S. (2014). Blood vessel, optical disk and damage area-based features for diabetic detection from retinal images. *Arabian Journal for Science and Engineering*, 39(10), 7059-7071.
- [21] Sangeetha J., Jayasankar T. (2019) Emotion Speech Recognition Based on Adaptive Fractional Deep Belief Network and Reinforcement Learning. In: Mallick P., Balas V., Bhoi A., Zobia A. (eds) *Cognitive Informatics and Soft Computing. Advances in Intelligent Systems and Computing*, vol 768. Springer, Singapore. https://doi.org/10.1007/978-981-13-0617-4_16
- [22] S. Kumari, R. J. Yadav, S. Namasudra, and C. H. Hsu, "Intelligent deception techniques against adversarial attack on industrial system", *International Journal of Intelligent Systems*, vol. 36, no. 5, pp. 2412-2437, 2021. DOI: 10.1002/int.22384
- [23] Uthayakumar, J., Elhoseny, M., & Shankar, K. (2020). Highly reliable and low-complexity image compression scheme using neighborhood correlation sequence algorithm in WSN. *IEEE Transactions on Reliability*, 69(4), 1398-1423.
- [24] Chavhan, S., Gupta, D., Nagaraju, C., Rammohan, A., Khanna, A., & Rodrigues, J. J. (2021). An Efficient Context-Aware Vehicle Incidents Route Service Management for Intelligent Transport System. *IEEE Systems Journal*.
- [25] Mansour, R. F., El Amraoui, A., Nouaouri, I., Díaz, V. G., Gupta, D., & Kumar, S. (2021). Artificial Intelligence and Internet of Things Enabled Disease Diagnosis Model for Smart Healthcare Systems. *IEEE Access*, 9, 45137-45146.
- [26] Alzubi, O. A. (2016). An empirical study of irregular ag block turbo codes over fading channels. *arXiv preprint arXiv:1604.00564*.
- [27] Kavitha, M., & Palani, D. S. (2012). A New Fast Curvelet Transform with Morphological Operations based method for

- Extraction of Retinal blood vessels using Graphical User Interfacel. *International Journal of Scientific & Engineering Research*, 3(6).
- [28] Ramesh, S., Yaashuwanth, C., Prathibanandhi, K., Basha, A. R., & Jayasankar, T. (2021). An optimized deep neural network based DoS attack detection in wireless video sensor network. *Journal of Ambient Intelligence and Humanized Computing*, 1-14.
- [29] P. Pavithran, S. Mathew, S. Namasudra and P. Lorenz, "A novel cryptosystem based on DNA cryptography and randomly generated Mealy machine", *Computers & Security*, vol. 104, 2021. DOI: <https://doi.org/10.1016/j.cose.2020.102160>
- [30] Le, DN., Parvathy, V.S., Gupta, D. et al. IoT enabled depthwise separable convolution neural network with deep support vector machine for COVID-19 diagnosis and classification. *Int. J. Mach. Learn. & Cyber.* (2021). <https://doi.org/10.1007/s13042-020-01248-7>
- [31] Sekaran, R., Goddumarri, S. N., Kallam, S., Ramachandran, M., Patan, R., & Gupta, D. (2021). 5G Integrated Spectrum Selection and Spectrum Access using AI-based Framework for IoT based Sensor Networks. *Computer Networks*, 186, 107649.
- [32] Zhang, Y. H., Li, Z., Zeng, T., Chen, L., Li, H., Gamarra, M., ... & Cai, Y. D. (2021). Investigating gene methylation signatures for fetal intolerance prediction. *Plos one*, 16(4), e0250032.
- [33] Alzubi, J. A. (2021). Blockchain-based Lamport Merkle Digital Signature: Authentication tool in IoT healthcare. *Computer Communications*, 170, 200-208.
- [34] Kavitha, M., & Palani, S. (2012). Retinal blood vessel segmentation algorithm for diabetic retinopathy and abnormality classification by supervised machine learning. *Int. J. Neural Netw. Appl*, 5(1), 47-53.
- [35] Jayanthi, J., Jayasankar, T., Krishnaraj, N., Prakash, N. B., Sagai Francis Britto, A., & Vinoth Kumar, K. (2021). An Intelligent Particle Swarm Optimization with Convolutional Neural Network for Diabetic Retinopathy Classification Model. *Journal of Medical Imaging and Health Informatics*, 11(3), 803-809.
- [36] S. Kumari and S. Namasudra, "System reliability evaluation using budget constrained real d-MC search", *Computer Communications*, vol. 171, 2021. DOI: <https://doi.org/10.1016/j.comcom.2021.02.004>
- [37] Shankar, K., Perumal, E. A novel hand-crafted with deep learning features based fusion model for COVID-19 diagnosis and classification using chest X-ray images. *Complex Intell. Syst.* (2020). <https://doi.org/10.1007/s40747-020-00216-6>
- [38] Mansour, R. F., & Aljehane, N. O. (2021). An optimal segmentation with deep learning based inception network model for intracranial hemorrhage diagnosis. *Neural Computing and Applications*, 1-13.
- [39] Alzubi, J. A. (2020). Bipolar fully recurrent deep structured neural learning based attack detection for securing industrial sensor networks. *Transactions on Emerging Telecommunications Technologies*, e4069.
- [40] Kavitha, M., Lavanya, G., & Janani, J. (2018). Enhanced SVM classifier for breast cancer diagnosis. *International Journal of Engineering Technologies and Management Research*, 5(3), 67-74.
- [41] Parvathy, P., Subramaniam, K., Venkatesan, G. P., Karthikaikumar, P., Varghese, J., & Jayasankar, T. (2020). Development of hand gesture recognition system using machine learning. *Journal of Ambient Intelligence and Humanized Computing*, 1-8.
- [42] S. Namasudra, "Data access control in the cloud computing environment for bioinformatics", *International Journal of Applied Research in Bioinformatics (IJARB)*, vol. 11, no. 1, pp. 40-50, 2021. DOI: 10.4018/IJARB.2021010105
- [43] K. Shankar, Y. Zhang, Y. Liu, L. Wu and C. Chen, "Hyperparameter Tuning Deep Learning for Diabetic Retinopathy Fundus Image Classification," in *IEEE Access*, vol. 8, pp. 118164-118173, 2020, doi: 10.1109/ACCESS.2020.3005152.
- [44] Mansour, R. F., & Abdelrahim, E. M. (2019). An evolutionary computing enriched RS attack resilient medical image steganography model for telemedicine applications. *Multidimensional Systems and Signal Processing*, 30(2), 791-814.
- [45] Alzubi, O. A., Alzubi, J. A., Dorgham, O., & Alsayyed, M. (2020). Cryptosystem design based on Hermitian curves for IoT security. *The Journal of Supercomputing*, 76(11), 8566-8589.
- [46] Kavitha, M., & Palani, S. (2020). A comprehensive analysis for retinal image classification methods using real-time database. *International Journal of Business Information Systems*, 34(2), 229-252.
- [47] Nair, L. R., Subramaniam, K., PrasannaVenkatesan, G. K. D., Baskar, P. S., & Jayasankar, T. (2020). Essentiality for bridging the gap between low and semantic level features in image retrieval systems: an

- overview. *Journal of Ambient Intelligence and Humanized Computing*, 1-13.
- [48] S. Namasudra, "Fast and secure data accessing by using DNA computing for the cloud environment", *IEEE Transactions on Services Computing*, 2020. DOI: 10.1109/TSC.2020.3046471
- [49] Shankar, K., Sait, A. R. W., Gupta, D., Lakshmanprabu, S. K., Khanna, A., & Pandey, H. M. (2020). Automated detection and classification of fundus diabetic retinopathy images using synergic deep learning model. *Pattern Recognition Letters*, 133, 210-216.
- [50] Abukharis, S., Alzubi, J. A., Alzubi, O. A., Alamri, S., & O'Farrell, T. (2016). Packet error rate performance of IEEE802. 11g under bluetooth interface. *arXiv preprint arXiv:1602.05556*.
- [51] Kavitha, M., & Palani, S. (2015). Hierarchical Classifier For Microaneurysm Detection. *International Journal of Applied Engineering Research*, 10(1), 1449-1458.
- [52] Anuradha, M., Ganesan, V., Oliver, S., Jayasankar, T., & Gopi, R. (2020). Hybrid firefly with differential evolution algorithm for multi agent system using clustering based personalization. *Journal of Ambient Intelligence and Humanized Computing*, 1-10.
- [53] Hnatiuc, M., Geman, O., Avram, A. G., Gupta, D., & Shankar, K. (2021). Human Signature Identification Using IoT Technology and Gait Recognition. *Electronics*, 10(7), 852.
- [54] Miled, A. B., Dhaouadi, R., & Mansour, R. F. (2020). Knowledge Deduction and Reuse Application to the Products' Design Process. *International Journal of Software Engineering and Knowledge Engineering*, 30(02), 217-237.
- [55] S. Namasudra, R. Chakraborty, A. Majumder and N. R. Moparthi, "Securing multimedia by using DNA based encryption in the cloud computing environment", *ACM Transactions on Multimedia Computing, Communications, and Applications*, vol. 16, no. 3s, 2020. DOI: <https://doi.org/10.1145/3392665>
- [56] Shankar, K., & Elhoseny, M. (2019). Trust Based Cluster Head Election of Secure Message Transmission in MANET Using Multi Secure Protocol with TDES. *J. UCS*, 25(10), 1221-1239.
- [57] Chen, T. M., Blasco, J., Alzubi, J. A., & Alzubi O. A. (2014). Intrusion detection. *IET*, 1(1), 1-9.
- [58] Kavitha, M., Syedakbar, S., Meenal, T., Kumar, R. S., & Stonier, A. A. (2021, February). Enhanced Algorithm for Bio Metric Based Secret Data Hiding. In *IOP Conference Series: Materials Science and Engineering* (Vol. 1055, No. 1, p. 012126). IOP Publishing.
- [59] Punarselvam, E., Sikkandar, M. Y., Bakouri, M., Prakash, N. B., Jayasankar, T., & Sudhakar, S. (2020). Different loading condition and angle measurement of human lumbar spine MRI image using ANSYS. *Journal of Ambient Intelligence and Humanized Computing*, 1-14.
- [60] S. Namasudra, "An improved attribute-based encryption technique towards the data security in cloud computing", *Concurrency and Computation: Practice and Exercise*, vol. 31, no. 3, 2019. DOI: 10.1002/cpe.4364
- [61] Shankar, K., Lakshmanprabu, S. K., Khanna, A., Tanwar, S., Rodrigues, J. J., & Roy, N. R. (2019). Alzheimer detection using Group Grey Wolf Optimization based features with convolutional classifier. *Computers & Electrical Engineering*, 77, 230-243.
- [62] Alrabea, A., Alzubi, O. A., & Alzubi, J. A. (2019). A task-based model for minimizing energy consumption in WSNs. *Energy Systems*, 1-18.
- [63] KAVITHA, M., GANESH, R., & RAJKUMAR, A. FACILITIES NAVIGATION ANDPATIENT MONITORING SYSTEM USING IBEACON TECHNOLOGY.
- [64] Kumar, K. V., Jayasankar, T., Eswaramoorthy, V., & Nivedhitha, V. (2020). SDARP: Security based Data Aware Routing Protocol for ad hoc sensor networks. *International Journal of Intelligent Networks*, 1, 36-42.
- [65] S. Namasudra and P. Roy, "Time saving protocol for data accessing in cloud computing", *IET Communications*, vol. 11, no. 10, pp. 1558-1565, 2017.
- [66] Shankar, K., Lakshmanprabu, S. K., Gupta, D., Khanna, A., & de Albuquerque, V. H. C. (2020). Adaptive optimal multi key based encryption for digital image security. *Concurrency and Computation: Practice and Experience*, 32(4), e5122.
- [67] Mansour, R. F. (2017). Evolutionary computing enriched ridge regression model for craniofacial reconstruction. *Multimedia Tools and Applications*, 1-18.
- [68] Pustokhina, I. V., Pustokhin, D. A., Kumar Pareek, P., Gupta, D., Khanna, A., & Shankar, K. (2021). Energy-efficient cluster-based unmanned aerial vehicle networks with deep learning-based scene classification model. *International Journal of Communication Systems*, e4786.

- [69] Sholiyi, A., Alzubi, J. A., Alzubi, O. A., Almomani, O., & O'Farrell, T. (2016). Near capacity irregular turbo code. arXiv preprint arXiv:1604.01358.
- [70] Muthumayil, K., Buvana, M., & Jayasankar, T. (2021). Energy Utilization using Artificial Bee Colony Algorithm for Network Life Time Enhancement of Homogeneous WSNs. *International Journal of Modern Agriculture*, 10(2), 1649-1656.
- [71] S. Namasudra, R. Chakraborty, S. Kadry, G. Manogaran and B. S. Rawal, "FAST: Fast accessing scheme for data transmission in cloud computing", *Peer-to-Peer Networking and Applications*, 2020. DOI: 10.1007/s12083-020-00959-6
- [72] Mansour, R. F. (2015). Using adaptive mutation to accelerate the convergence of immune algorithms for prediction of 3D molecular structure. *International Journal of Computers and Applications*, 37(3-4), 127-133.
- [73] Shankar, K., Elhoseny, M., Kumar, R. S., Lakshmanprabu, S. K., & Yuan, X. (2020). Secret image sharing scheme with encrypted shadow images using optimal homomorphic encryption technique. *Journal of Ambient Intelligence and Humanized Computing*, 11(5), 1821-1833.