Journal of Statistics & Management Systems
ISSN 0972-0510 (Print), ISSN 2169-0014 (Online)
Vol. 27 (2024), No. 2, pp. 285–294
DOI : 10.47974/JSMS-1254

A pragmatic analysis of text-based sentiment analysis models from an analytical perspective

Tabassum H. Khan *
Department of Computer Science & Engineering
G H Raisoni University
Amravati
Maharashtra
India

Sonali Ridhorkar †
Department of Computer Science & Engineering
G H Raisoni Institute of Engineering & Technology
Nagpur
Maharashtra
India

Abstract

NLP sentiment analysis involves pre-processing, feature extraction, feature selection, sentiment classification, and post-processing. Pre-processing can include PoS tagging, stop-word removal, language-specific missing work detection and correction, etc. Word2vec, N Gramme, term frequency (TF), distance measures, etc. can extract and select features. CNNs, DNNs, and other machine learning and deep learning models can be used for classification and post-processing. These models form a sentiment analysis engine, which can be assessed for accuracy, precision, recall, analysis delay, computational complexity, etc. The wide variation in algorithmic combination and performance makes it difficult for researchers to choose the best model(s) for their application. This task requires a lot of model testing and validation, which increases time-to-market and product cost. This text surveys recently proposed sentiment analysis models to reduce ambiguity and speed up system development. This survey characterises reviewed models' nuances, advantages, limitations, and future research scopes to help readers choose the best models for their application deployment. These models' delay, accuracy, precision, recall, computational complexity, and probable use are also compared in this text. This comparison lets readers choose the best model(s) for their application deployment and use them without validation. Reducing testing delay

* *E-mail:* **khantabi0786@gmail.com (Corresponding Author)**
† *E-mail:* **sonali.ridhorkar@raisoni.net**

will improve model performance for system design and deployment for multiple sentiment analysis applications.

Subject Classification: *68Q30.*

Keywords: *Sentiment, Analysis, Natural language, Processing, Delay, Precision, Complexity.*

1. Introduction

To determine a sentence's sentiment, sentiment analysis involves data collection, pre-processing, polarity calculation, sentiment score evaluation, and aggregation. WordNet is used to evaluate text polarity in most sentiment analysis models. The model collects data from various document sources and pre-processes it using PoS tagging and chunking. These operations remove all non-action words from the input sentence and extract only action words. A polarity estimation engine assigns context-sensitive polarity weights to extracted words. Calculate this polarity weight using equation 1, which uses sentence context.

$$P_i = \sum_{j=1}^{N_w} P_{DB_j} * w_j \tag{1}$$

P_i represents the polarity of the i^th word, P_DB represents the polarity of the word in the database, w_j represents the weight of the word in context, and N_w represents the number of similar words in the database. The input text sentiment score is the sum of these polarities [1, 2]. This task requires a sentiment aggregation engine to remove incorrect polarities from the sentence for sentiment generation and analysis. The model's performance is validated by comparing these sentiments to ground truth. Researchers propose many efficient sentiment analysis system models. The following section discusses these models' nuances, advantages, limitations, and drawbacks. This survey would help researchers choose the best model(s) for their application deployments. After this survey, the reviewed models are evaluated and compared for analysis delay, representation accuracy, precision, recall, computational complexity, and likely use. This comparison would help researchers choose the best sentiment analysis models for their applications and use cases. This text concludes with some intriguing observations about the reviewed models and suggests using fusion and machine learning to improve them for different use cases.

2. Literature review

Over time, researchers have proposed many context-sensitive sentiment analysis models for application-specific sentiment analysis. Few of these algorithms are context-independent, so sentiment analysis scalability can be improved. A new sentiment analysis model using comparison enhanced bidirectional long-short-term memory (BiLSTM) with multiple attention network is proposed in [1]. The BiLSTM model extracts multilevel features, while the MHA model estimates polarity. Word2Vec estimates feature sets for the BiLSTM model, which processes them using linear scaled dot-product attention model and sentence embedding with polarity-based sampling. Equation 2 combines positive and negative sample weights and biases to estimate sentence polarity.

$$S_{out} = W_s[W_n * \bigcup(S_{in} \in N) + n^b]\,[W_p * \bigcup(S_{in} \in P) + p^b] \tag{2}$$

S_out, S_in, W_s, W_n, W_p, N,P, and n^b,p^b represent output, positive, and negative sentiment dictionaries, along with their biases.

The model was tested on various textual datasets and identified sentiments with 86.9% accuracy, outperforming recurrent neural networks (RNN) at 83.5%, convolutional neural networks (CNN) at 85.8%, ordered neurons LSTM (ON-LSTM) at 85.6%, Deep LSTM (D-LSTM) at 85.3%, BiLSTM attention networks (BiLSTMA) at 85.2%, MHA at 85.15%, and BiLSTM AUC and precision values were similar across datasets, indicating high scalability and low overheads for the CE BiLSTM MHA model. In [2], researchers propose using Kernel Optimisation Function (KOF) to learn sentiment context vector for multidomain sentiment analysis. It integrates sentiment orientation with statistical data and propagates semantic information to all corpus word representations. Since each corpus entry earns a reward during document evaluation, real-time context understanding improves sentiment analysis. The proposed model has 73.06% semantic accuracy and 64.61% syntactic accuracy, which is higher than hybrid principal component analysis (HPCA) at 49.6% & 47.21%, GloVE at 72.9% & 57.06%, Word2Vec with skip Gramme (WsG) at 66.51% & 57.02%, and Word2Vec with continuous bag of words (WCBoW) at 66.51% & 57.02%. Lexicon-based analysis with aspect-based sentiment detection improves accuracy. Lexicon-enhanced attention network (LEAN) with BiLSTM is proposed in [3]. Hidden sentiment representation, content embedding, and aspect embedding are done separately using two BiLSTMs. BiLSTM results are combined to generate content-aspect hidden representation, which is combined with aspect embedding attention matrices to evaluate. The final SoftMax-based evaluation layer is visible in this dual-LSTM network.

LEAN with BiLSTM model had an accuracy of 79.1% on the SimEval 2014 dataset, outperforming LSTM (74.3%), Target Dependent (TDLSTM) (75.6%), Target Connection LSTM (TCLSTM) (76%), attention-based LSTM (ATLSTM) (76.2%), AT with autoencoder-based LSTM (ATAE LSTM) (77.2%), and interactive attention networks (This accuracy is sufficient for moderately scaled applications, but fused deep learning models or context-aware networks can improve it. In a context-aware model [4], researchers used attention mechanism to represent text using topic semantic information. NN training and sentiment evaluation use this TSI feature vector. This model uses a network structure like [3] and multiple pooling layers to improve sentiment analysis. These layers improve model feature extraction and selection, improving accuracy and precision. The TSI model with attention network (TSIAN) has an accuracy of 90.8%, higher than TSI without AN at 89.7%, BiLSTM at 87.62%, LSTM at 85.1%, CNN at 86%, FastText at 84.3%, and SVM at 83.8% on the same datasets. The model has higher precision and recall than other methods, but attention networks cause delay, which must be reduced using parallel processing and redundancy reduction. In [5], researchers used transfer learning with heterogeneous modalities (HMTL) for text-to-AV data alignment to propose a high-speed sentiment analysis model for audio and visual (AV) data. Textual features can be converted to AV features by the proposed model. Training accuracy is assessed by converting these features into heterogenous modalities. The trained classifier predicts sentiments with 64.8% accuracy without text data. Augmented deep learning models like capsule attention networks (CANs) can improve this accuracy [6], with Knowledge Guided CAN (KGCAN) for aspect-level sentiment evaluation. As suggested by [3], word2vec with BiLSTM is combined with capsule attention network for sentiment analysis. Knowledge Guidance uses Twitter, Lap14, Res15, and SpATSA datasets with positive and negative sentiment samples. On the same datasets, the proposed KGCAN model outperforms SVM (75.6%), LSTM (86.8%), Memory Network (MemNet) (85.4%), attention-over-attention (AOA) (75.8%), IAN (83.5%), Transformer Net with learning factorization (TNet-LF) (89.07%), and aspect-specific graph convolutional networks (ASGCN) (88.99%).

Researchers used barrage dictionaries and different polarity models to estimate sentiment in a novel model [7]. For barrage sentiment estimation, the BSET model uses deep learning and dictionary analysis.

On the 'An les houches accord' dataset, the BSET model had 92.4% accuracy, higher than extended sentiment dictionary (ESD) at 87.7% and double ESD (DESD) at 91.32%. However, the model was only tested for language-specific sentiment analysis and should be tested on other datasets for better performance. Additionally, hybrid indices like Compounded Aggregated Positivity Index (CAPI) should be used to estimate model performance [8]. Equation 3 represents a relational stochastic mutual information value that combines sentiment value probabilities to form the index sets.

$$MI_{x,y} = \frac{1}{N * M} \sum_{i=1}^{N} \sum_{j=1}^{M} P(s(x_i) \mid s(y_j)) * \log \left[\frac{P(s(x_i))}{P(s(y_j))} \right] \tag{3}$$

Where MI represents mutual information between sentences x,y, M,N represents number of words in each sentence, and P,s represents stochastic probability and sentiment of the word for which CAPI is needed for different use cases.

[9] suggests a reviewer credibility-based product recommendation model to boost sentiment analysis trust. Reviewer credibility, user interest, and sentiment improve recommendation (CISER) performance. Users receive credibility scores based on eCommerce purchase patterns and review correlation. A novel user interest metric estimates aesthetic review level to determine interest-pattern mining features. It outperforms content-based filtering (CBF) (85.8%), item-based collaborative filtering (IBCF) (77.9%), and user-based collaborative filtering (UBCF) (79.6%) on eCommerce datasets with 93.5% accuracy using FastText and PSO. Multiple deep learning and optimisation models slow this scalable model. In [10], researchers combined LSTM with Naïve Bayes, Logistic Regression, and SVM to improve feature augmentation in long news text analysis. The proposed model outperforms LSTM (89.1%), SVM (82.3%), NB (79.6%), and LR (83.4%) with 90.5% accuracy on the same datasets. This suggests ensemble machine learning models improve sentiment analysis accuracy. Opinion mining, emotion understanding, and sentiment analysis (OMEUSA) can determine sentence- and document-level sentiments with over 91.5% accuracy, moderate delay, and good precision [11]. BiLSTM with weighted word vectors captured context in [12]. Word vectors from TF-IDF and word2vec models are classified by a feedforward neural network (FFNN) for sentiment analysis. These computations let the model augment input features and output a context-aware feature vector. These features help the FFNN TF-IDF BiLSTM model outperform BiLSTM

(90.85%), TF-IDF (91.18%), and Word2Vec (89.06%) on various datasets with 91.5% accuracy. CNN aspect-based sentiment analysis can benefit from gated recurrent unit (GRU) [13].

To improve multidomain sentiment analysis, GRU-CNN uses word2vec for feature extraction, stop-word removal, and segmentation. GRU-CNN outperforms Text CNN (88.15%), DE CNN (83.29%), Dis LSTM (83.14%), LSTM-CNN (80.6%), and SVM (83.17%) with 89.56% accuracy on different datasets. This performance makes the model suitable for many sentiment analysis applications. Convolution networks extract many features, making training and validation time-consuming. The Genetic Algorithm (GA) model in [14] reduces this delay by minimising the polarity of similar words in a sentence for different sentiments.

NB with GA has 97.6% accuracy on IMDB, Amazon, and Yelp, suggesting movie analysis. This model must be tested on other datasets and applications to estimate large-scale scalability and performance. Timeseries analysis of stock market data revealed investor herding [15]. From stock market reports, sentence tokenization, degree modifiers, and privatives estimate herd behaviour. Although moderately accurate, deep learning models like LSTM, GRU, CNN, and RNN can improve feature augmentation and classification.

They generated aspect-specific text-feature representations in a high-efficiency entity sensitive attention and fusion Network (ESAFN) for aspect-based sentiment analysis [16]. This textual fusion layer aggregates these representations, and aspect-level visual representation & attention models prune features. From input images and aspect-based visual and textual representations, the model estimates sentiment polarities.

Multimodal fusion gives the proposed ESAFN method an accuracy of 73.38%, which is higher than Residual Target (ResTarget) (59.88%), multiple Generative Adversarial Networks (MGAN) (71.17%), recurrent attention network on memory (RANM) (70.68%), ResRANM (71.55%), TFN ResRANM (69.91%), and ResESTR (72.03%). Without image-text training datasets, classification accuracy is moderate, but augmentation and fusion models can improve it. Geological tweet sentiment analysis can benefit from feature augmentation methods like LDA and LISA [17]. The model uses daily temporal analysis to reduce feature redundancies and improve polarity estimation. This gives it 79.5% regional accuracy. OR, MMNN, and CSHGCNN are proposed in [18, 19, 20]. These models perform accurate sentiment analysis but take a long time to train and validate. The OR model has 83.2% accuracy, while the MMNN model that uses BiGRUs with fusion LSTM has 77.5%, higher than conditional random fields (CRF) at 61.5%, NNCRFP at 71.03%, LSTM-CRF at 75.3%, and MNN at 73.71% on similar datasets. The BiGRU & GCN-based CSHGCNN model reduces feature vectors with 80.55% accuracy, outperforming Tree LSTM (74.25%), Tree GCN (76.83%), and Hierarchical attention networks (HAN) (78.45%). This comprehensive review shows that sentiment analysis is well-researched, but fusion and augmentation strengthen models. Selecting the best models is needed. The next section compares these models' accuracy, precision, recall, and delays.

3. Statistical analysis

Researchers propose many sentiment analysis models, according to the review. These models include linear classification and deep learning-based automatic sentiment analysis. This section compares these models statistically by accuracy (A), application of use (App), precision (P), recall (R), and processing delay (D). This evaluation would help researchers find the best language processing models. Figure 1 compares models' real-time sentiment analysis accuracy and uses. This review shows that HMTL [5] is best for Audio & Video sentiment analysis, CISER [9] for eCommerce, LDA LISA [17] for Geographic Data, and BSET [7] for Language Specific sentiment analysis. Ensemble [10] methods also perform better in News Text sentiment analysis. An algorithm score is introduced to aid model selection. This score can be calculated using equation 4.

$$AS = \frac{A}{100} + P + R + \frac{5}{D} \quad (4)$$

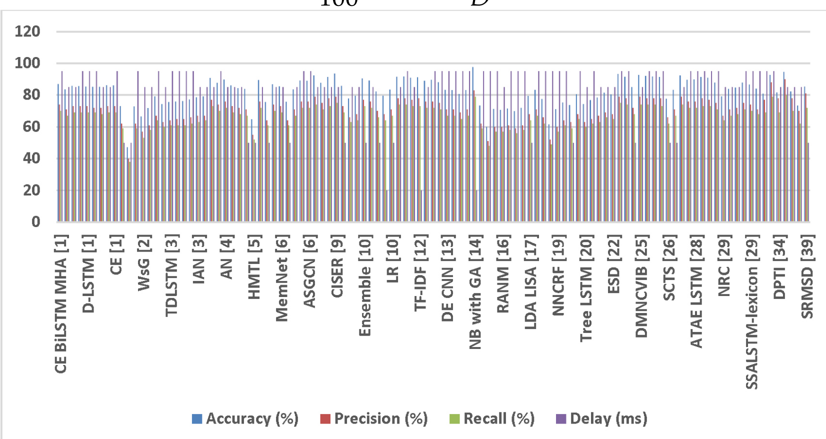

Figure 1

Comparative Analysis of Reviewed Models

This score was also evaluated for all models, and it was observed that NB with GA [14], TF-IDF [12], NB [10], Ensemble [10], and SRMSD [39] had an overall high performance for sentiment analysis across various application deployments.

4. Conclusion and future work

This text reviews a wide variety of machine learning models for context-dependent, and independent sentiment analysis. The reviewed models are evaluated in terms of accuracy, precision, recall, and end-to-end delay performance. From this performance it is observed that NB with GA, RRNN, MANSM, DMNCVIB, DByN, CEHAN, BiGRU CNN Dropout, BiGRU CNN L2, FFNN TF-IDF BiLSTM, OMEUSA, WSABSA-CNN, BiGRU CNN, and AS BiGRU outperform other models in terms of accuracy, precision, and recall performance. These models must be used in order to improve overall efficiency of sentiment analysis for any application deployment. Furthermore, the models NB with GA, TF-IDF, NB, Ensemble, and SRMSD outperformed other models when evaluated in terms of their algorithmic score, which is a combination of accuracy, precision, recall, and delay parameters. In future, researchers can perform model fusion, and identify metrics which can be improved via bioinspired modelling. Evaluation of these models must be done on larger datasets, and on wider application scenarios to validate their performance, and ensure their applicability to context-independent sentiment analysis applications.

References

[1] Y. Lin, J. Li, L. Yang, K. Xu and H. Lin, "Sentiment Analysis With Comparison Enhanced Deep Neural Network," in IEEE Access, vol. 8, pp. 78378-78384 (2020), doi: 10.1109/ACCESS.2020.2989424.

[2] Zhu, Luyao & Wei, Li & Shi, Yong & Guo, Kun. SentiVec: Learning Sentiment-Context Vector via Kernel Optimization Function for Sentiment Analysis. *IEEE Transactions on Neural Networks and Learning Systems*. PP. 1-12 (2020). 10.1109/TNNLS.2020.3006531.

[3] Z. Ren, G. Zeng, L. Chen, Q. Zhang, C. Zhang and D. Pan, "A Lexicon-Enhanced Attention Network for Aspect-Level Sentiment Analysis," in IEEE Access, vol. 8, pp. 93464-93471 (2020), doi: 10.1109/ACCESS.2020.2995211.

[4] G. Xu, Z. Yu, Z. Chen, X. Qiu and H. Yao, "Sensitive Information Topics-Based Sentiment Analysis Method for Big Data," in IEEE Access, vol. 7, pp. 96177-96190 (2019), doi: 10.1109/ACCESS.2019.2927360.

[5] Seo, S., Na, S., & Kim, J. HMTL: Heterogeneous Modality Transfer Learning for Audio-Visual Sentiment Analysis. IEEE Access, 8, 140426-140437 (2020).

[6] B. Zhang, X. Li, X. Xu, K. -C. Leung, Z. Chen and Y. Ye, "Knowledge Guided Capsule Attention Network for Aspect-Based Sentiment Analysis," in IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 28, pp. 2538-2551 (2020), doi: 10.1109/TASLP.2020.3017093.

[7] Z. Cui, Q. Qiu, C. Yin, J. Yu, Z. Wu and A. Deng, "A Barrage Sentiment Analysis Scheme Based on Expression and Tone," in IEEE Access, vol. 7, pp. 180324-180335 (2019), doi: 10.1109/ACCESS.2019.2957279.

[8] M. Rodríguez-Ibáñez, F. Gimeno-Blanes, P. M. Cuenca-Jiménez, S. Muñoz-Romero, C. Soguero and J. L. Rojo-Álvarez, "On the Statistical and Temporal Dynamics of Sentiment Analysis," in IEEE Access, vol. 8, pp. 87994-88013 (2020), doi: 10.1109/ACCESS.2020.2987207.

[9] S. Hu, A. Kumar, F. Al-Turjman, S. Gupta, S. Seth and Shubham, "Reviewer Credibility and Sentiment Analysis Based User Profile Modelling for Online Product Recommendation," in IEEE Access, vol. 8, pp. 26172-26189 (2020), doi: 10.1109/ACCESS.2020.2971087.

[10] W. Niu and L. Wu, "Sentiment Analysis and Contrastive Experiments of Long News Texts," 2019 IEEE 4th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), pp. 1331-1335 (2019), doi: 10.1109/IAEAC47372.2019.8997550.

[11] P. Sánchez-Núñez, M. J. Cobo, C. D. L. Heras-Pedrosa, J. I. Peláez and E. Herrera-Viedma, "Opinion Mining, Sentiment Analysis and Emotion Understanding in Advertising: A Bibliometric Analysis," in IEEE Access, vol. 8, pp. 134563-134576 (2020), doi: 10.1109/ACCESS.2020.3009482.

[12] G. Xu, Y. Meng, X. Qiu, Z. Yu and X. Wu, "Sentiment Analysis of Comment Texts Based on BiLSTM," in IEEE Access, vol. 7, pp. 51522-51532 (2019), doi: 10.1109/ACCESS.2019.2909919.

[13] N. Zhao, H. Gao, X. Wen and H. Li, "Combination of Convolutional Neural Network and Gated Recurrent Unit for Aspect-Based Senti-

ment Analysis," in IEEE Access, vol. 9, pp. 15561-15569 (2021), doi: 10.1109/ACCESS.2021.3052937.

[14] F. Iqbal et al., "A Hybrid Framework for Sentiment Analysis Using Genetic Algorithm Based Feature Reduction," in IEEE Access, vol. 7, pp. 14637-14652 (2019), doi: 10.1109/ACCESS.2019.2892852.

[15] R. Ren and D. Wu, "An Innovative Sentiment Analysis to Measure Herd Behavior," in IEEE Transactions on Systems, Man, and Cybernetics: Systems, vol. 50, no. 10, pp. 3841-3851, Oct. (2020), doi: 10.1109/TSMC.2018.2864942.

[16] J. Yu, J. Jiang and R. Xia, "Entity-Sensitive Attention and Fusion Network for Entity-Level Multimodal Sentiment Classification," in IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 28, pp. 429-439 (2020), doi: 10.1109/TASLP.2019.2957872.

[17] T. Hu, B. She, L. Duan, H. Yue and J. Clunis, "A Systematic Spatial and Temporal Sentiment Analysis on Geo-Tweets," in IEEE Access, vol. 8, pp. 8658-8667 (2020), doi: 10.1109/ACCESS.2019.2961100.

[18] S. E. Saad and J. Yang, "Twitter Sentiment Analysis Based on Ordinal Regression," in IEEE Access, vol. 7, pp. 163677-163685 (2019), doi: 10.1109/ACCESS.2019.2952127.

[19] Y. Bie and Y. Yang, "A multitask multiview neural network for end-to-end aspect-based sentiment analysis," in Big Data Mining and Analytics, vol. 4, no. 3, pp. 195-207, Sept. (2021), doi: 10.26599/BDMA.2021.9020003.

[20] E. Zuo, H. Zhao, B. Chen and Q. Chen, "Context-Specific Heterogeneous Graph Convolutional Network for Implicit Sentiment Analysis," in IEEE Access, vol. 8, pp. 37967-37975 (2020), doi: 10.1109/ACCESS.2020.2975244.