WORLD COMET RESEARCH GROUP

International Journal of Applied Data Science & Modern Computing
Vol. 3, No. 1, 2020
Doi: 10.67228/30713498/IJADSMC-2020PI1W6N
PP. 01-13

Original Article

Improving Sentiment Analysis Using Transformer-Based NLP Models

Dr. Ibrahim Lawal

Federal University of Technology Minna, Niger state.

Received: 04-05-2020 Revised: 04-26-2020 Accepted: 05-03-2020 Published: 05-05-2020

ABSTRACT

Sentiment analysis is an important area of Natural Language Processing (NLP) used to interpret opinions from text such as reviews and social media. Traditional methods, including rule-based and machine learning approaches, struggled with complex language features like sarcasm and context. Deep learning models like RNNs and CNNs improved performance but had limitations in capturing long-range dependencies. Transformer-based models such as BERT, RoBERTa, DistilBERT, and XLNet overcome these issues using self-attention mechanisms to better understand context. This study explores how these models enhance sentiment analysis accuracy through transfer learning, fine-tuning, and domain adaptation. Experimental results on benchmark datasets show that transformer models outperform traditional methods in accuracy, precision, recall, and F1-score. The findings highlight that transformer-based approaches provide more efficient and scalable solutions for real-world sentiment analysis applications.

KEYWORDS

Sentiment Analysis, Natural Language Processing, Transformers, BERT, Self-Attention, Deep Learning, Text Classification, Opinion Mining.

ISSN: 3071-3498
PP: 01-13.

1. INTRODUCTION

1.1. Background

Sentiment analysis or opinion mining can be defined as a calculation of subjective information in the form of opinions, emotions, and attitude stated in written text. It has become a key part of the contemporary data analytics, allowing the organizations grasp how people perceive or want to be perceived in most application areas, such as marketing, finance, politics, healthcare and social media analytics. This ever-increasing amount of user-generated content on digital platforms i.e. online reviews, social networking platforms, blogs and discussion rooms has only contributed to the demands of automated sentiment analysis systems that can process and analyze large volumes of unstructured text in an efficient and scalable way. Sentiment analysis can be useful in decision making, trend analysis, and strategic planning through transformation of qualitative opinions into quantitative insights. Sentiment analysis in its primitive forms was largely based on lexicon-based methods involving the use of pre-existing dictionaries of words which are sentiment-related to classify the general polarity of a piece of text. Although these techniques were simple and interpretable, they had severe deficiencies such as low scalability, domain dependence, and lack of linguistic capabilities such as context based meaning, negation, sarcasm and polysemy. In order to address them, subsequent machine learning algorithms like the Naive Bayes, Support Vector Machines, and logistic regression algorithms were developed hence allowing models to discover the sentiment patterns using the labeled data as their input. Though these data-driven techniques enhanced precision and expressiveness, they nonetheless deeply depended on manually engineered functions and also failed to detect more profound semantic associations, which inspired the development of more advanced deep learning and transformer-based sentiment models.

1.2. Emergence of Transformer-Based Models

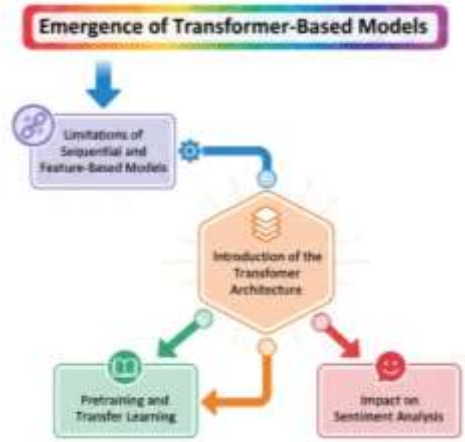

Fig1 - Emergence of Transformer-Based Models

1.2.1. Limitations of Sequential and Feature-Based Models

Until the introduction of transformer architecture sentiment analysis used classic machine learning models, sequential deep learning based on Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM) and Convolutional Neural Networks (CNNs). These models, though superior to earlier models in discriminating between sentiment, by taking word order and local context into

ISSN: 3071-3498
PP: 01-13.

account, were limited by nature in representing dependence on long-range interactions and world context. RNN-based models used sequential processing, limiting parallelization, hence more time was spent on training and large datasets were less scalable. Also, contextual polarity shifts, implicit sentiment, and long-distance negation were frequent linguistic examples that feature-based and sequential models had difficulty modelling.

1.2.2. Introduction of the Transformer Architecture

With the appearance of the transformer-based models, a paradigm shift in natural language processing has taken place. Recurrent and convolutional structures were substituted by transformers, which provides a mechanism of self-attention where each token in the sentence can attend to all the other tokens at the same time. This architectural innovation helps this model to represent global contextual relationships better and facilitate complete parallelization in training. Transformers have a high level of scalability and features since they resulted in the removal of redundancy thereby enhancing computational requirements, which is one of the major tasks of large scale language models.

1.2.3. Pretraining and Transfer Learning

An important innovation related to transformer-based models is the introduction of pretraining on a large scale and then fine-tuning on a task. Language models like Bidirectional Encoder Representations from Transformers (BERT) are trained on large unlabeled text corpora through the self-supervised learning goals whereby they are capable of achieving rich linguistic knowledge such as syntax, semantics, and contextual uses of words. This initial knowledge may then be transferred to lower level activities like sentiment analysis by fine-tuning, which allows maximum performance without a larger amount of labeled data.

1.2.4. Impact on Sentiment Analysis

The models based on transformers have made a tremendous step in improving the sentiment analysis by offering deep and contextualized text representations, which are superior to both traditional and deep learning baselines. Such capability to model bidirectional context and their long-range dependencies provides the opportunity to interpret sentiment-bearing expressions more accurately. This has rendered all transformer architectures, including BERT and RoBERTa, the current leader in sentiment analysis, providing better accuracy, robustness, and flexibility on varied domains and data.

1.3. Limitations of Traditional Approaches

Although there have been significant improvements in the area of sentiment analysis, machine learning-based methodologies have a number of built-in limitations that limit their applicability in realistic engines. The methods applied in these models are manual feature engineering methods like n-grams, part-of-speech tagging, syntactic dependency, and hand-written sentiment indicators. Although these characteristics may be able to attract the surface characteristics of a text, they tend to be lacklustre in terms of their semantical content, and fail to offer the semantical meaning of words in varying contexts. Consequently, conventional classifiers have difficulties in

ISSN: 3071-3498
PP: 01-13.

testing between domains especially where the expressions of sentiments differ based on domain specific vocabularies or changes in language use.

Moreover, these methods are also sensitive to both sparsity of features as well as dimensionality which may impair performance in cases of large and varied textual data. Besides the weak ability of semantic representation, the classic machine learning models have the problem of supporting long sentences and complicated linguistic syntax. They generally view text as a collection of independent features without paying attention to word order and contextual constraints that are essential to properly inferring sentiment. This weakness is especially pronounced in the case of implicit expression of sentiment, sarcasm, and context-shifted polarity determinisms, the sentiment of a word or phrase is determined not by its own lexical meaning but by the contextual meaning. Recurrent Neural Network and Convolutional Neural Network models are some of the deep learning systems that minimized these shortcomings by learning features automatically using raw text. RNN architectures, such as Long Short-term memory networks, performed better in sentiment analysis, as they took into consideration serial interactions and retention of word order information. Their sequential nature however causes computational inefficiencies and makes it hard to parallelize, which makes training time-consuming when dealing with large-scale datasets. Also, RNNs do not always exhibit long-range dependencies as they may fall prey to vanishing gradient problems. Conversely, CNNs are very good at local n-gram features by use of convolutional filters, but are by nature constrained in the amount of global contextual relationships that they can model entire sentences or documents. All these constraints underscore the fact that more sophisticated architectures that are able to extract rich contextual semantics require development, and the sentiment analysis of transformer-based models are being developed.

2. LITERATURE SURVEY

2.1. Lexicon-Based and Rule-Based Methods

Lexicon based methods Lexicon-based and rule-based sentimental analysis methods are the oldest type of automated opinion mining algorithms and based on a fixed set of sentiment lexicons such as SentiWordNet, LIWC (Linguistic Inquiry and Word Count) and AFINN, in which words have a pre-determined amount of polarity (good or bad) sentiment. These algorithms individually compute the sentiment of a text by combining scores of each word or phrase in that text, which can be further augmented by manually applied linguistic rules to address negation, intensifiers, and punctuation. They are mostly stronger in high interpretability and low computational needs that make them applicable in low-resource settings. Nevertheless, the lexicon-based solutions have the disadvantage that vocabulary is limited, they are domain-dependent and they cannot understand some contextual nuance like sarcasm, irony and polysemy. Also, sentiment orientation can vary as the context, which is not reflected in static lexicons, creating less accuracy in more complex, informal and dynamic textual tasks like social media and customer reviews.

ISSN: 3071-3498
PP: 01-13.

2.2. Machine Learning-Based Sentiment Analysis

Sentiment analysis based on machine learning was a major breakthrough, as it no longer operates on handcrafted rules, but uses the information to construct models that can learn statistical patterns out of the labeled collective. Naive Bayes, Support Vector Machines (SVM), Logistic Regression, and Decision Trees became widely used among classical classifiers because of their strength and comparative simplicity of computation. These models commonly used bag-of-words (BoW), n-grams and terms frequency inverse document frequency (TF-IDF) to numerically represent textual data. Although these methods showed significant gains in accuracy at classification compared to the lexicon based methods, they were limited by the high dimensionality of sparse features space and deficient in semantic interpretation of the language. They also did not consider words as components of a sentence, did not care about word order and contextual details and so could not generalize across domains and do anything with complicated linguistic constructions.

2.3. Deep Learning Approaches

Deep learning methods were a change of paradigm in sentiment analysis as its algorithms are capable of automatically choosing features based on hierarchical neural networks. Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) and their implementation, including Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRU), efficiently estimated both local semantic patterns and n-gram features and sequential dependencies and long-range contextual features, respectively. These models were made possible by the integration of distributed representations of the word e.g. Word2Vec, GloVe and FastText that enabled the models to encode semantic similarity among words which was significantly more effective than the traditional machine learning methods. In spite of this, deep learning models encountered problems with vanishing gradients, little parallelization because of the sequential processing of data, and the difficulty absorbing very long-range interactions, making them less scalable and less training efficient on large-scale data.

2.4. Transformer-Based Sentiment Analysis

Models based on transformers have become the current trend of sentiment analysis through the self-attention mechanism to create global contextual dependencies among words at the same time. In contrast to architectures that use RNNs, transformers can be fully parallelized and, therefore, increase training efficiency and scalability. Bidirectional Encoder Representations of Transformers (BERT) was also the first model to introduce deep bidirectional contextual representations, which enable models to comprehend word meaning using left and right contexts. Following architectures like RoBERTa, XLNet, and ALBERT optimized pretraining goals, data-use, as well as parameter competency, and showed major rises in performance on benchmark sentiment suppression legalities. These models are highly competent in addressing the contextual ambiguity, domain adaptation as well as the complex linguistic phenomenon i.e. negation and sarcasm. Transformer-based methods, however, are expensive in computational resources and large-scale labelled or pretraining datasets, thus can be restricted to resource-constrained settings.

ISSN: 3071-3498
PP: 01-13.

3. METHODOLOGY

3.1. Overall Framework

The suggested sentiment analysis system adheres to a logical pipeline that is aimed at successfully extracting syntactic and semantic data through textual data. It consists of the combination of data preprocessing, learning of the deep contextual representation, the fine-tuning supervisory learning process and thorough performance measurement to achieve strong and stable sentiment classification.

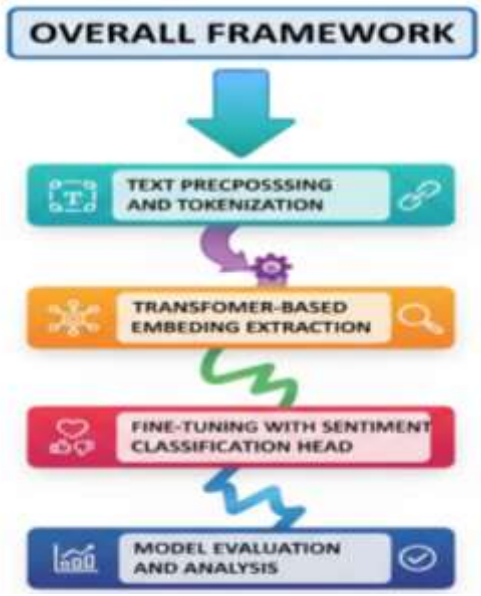

Fig 2 - Overall Framework

3.1.1. Text Preprocessing and Tokenization

The first step is dedicated to textual data preparation to ingest it into the model. This involves noise reduction, which cannot be accomplished by deleting URLs, special characters and unnecessary whitespace, but retain sentiment-carrying components like negations, emotive symbols. Next, the text is cleaned and then tokenized with subword-based tokenization methods (WordPiece or Byte Pair Encoding (BPE)) and teaches words to use smaller units in order to address out-of-vocabulary words and morphological variation effectively. Through this process, there will be uniform input representation, and compatibility with transformer architectures.

3.1.2. Transformer-Based Embedding Extraction

An incomplete tokenized text is represented by inputting the text to a pre-trained transformer model in the second stage to produce deep contextual embeddings. Transformer representations offer directional boundaries by incorporating the self-attention mechanism of understanding the meanings of words relative to surrounding tokens and words, unlike traditional word embeddings. The contextualized embeddings, in particular, the special classification token representation, are high-level semantic features that represent the idea that there is sentiment-relevant information represented in the whole sequence of inputs.

ISSN: 3071-3498
PP: 01-13.

3.1.3. Fine-Tuning with Sentiment Classification Head

The third step is fine-tuning the transformer model; this is done by attaching a task-specified sentiment classification head, which is usually a fully connected neural layer and then a softmax or a sigmoid transformation. Transformer parameters, as well as the classification head, are both optimized on both the labeled sentiment data during the supervised training. This domain-specific sentiment adaptation with pre-trained linguistic knowledge via this end-to-end fine adjustment enables the model to offer dramatic classification accuracy.

3.1.4. Model Evaluation and Analysis

The last step analyzes the work of the offered model on the basis of the official measures like accuracy, precision, recall, F1-score, and confusion matrices. Generalization capability is assessed by using cross-validation or hold-out test sets. Further, both qualitative analyses, such as attention visualization and error analysis, are run to explain the model outcomes and what must be further improved. The holistic assessment guarantees the soundness, consistency, and feasibility of the suggested structure.

3.2. Input Representation

The input representation is a major factor in transformer-based sentiment analysis because it determines how the raw textual data is transformed into a numerical one in which the deep neural processing is applicable. Given an input sentence represented as a sequence of words, the sentence is represented as a finite set of finite sized tokens represented as w one, w two, etc. to w n, the initial step is converting this word-level sequence into subword units with a subword tokenization technique WordPiece or Byte Pair Encoding. Subword tokenization breaks a word into smaller meaningful units, allowing the model to be useful with rare words and morphological variation, and out-of-vocabulary words prevalent in real-world textual data. Once the input sequence is tokenized, special tokens are placed where they have to appear in order to be processed by the transformers. A classification token, also known as CLS is placed in the front of the token sequence and represents the entire sentence as a whole. The token is a summation of contextual data of all the other tokens with the self-attention mechanism and subsequently the only purpose is to serve as a primary input to the sentiment classification head. This is followed by a separator token denoted as SEP at the end of the sequence to distinctly indicate the boundaries of the sentences and succeeding compatibility with single sentence and also sentence pair tasks. The resulting sequence is then converted to a fixed-dimensional embedding vector each token of the sequence. The final input embedding is the feature derived through summation of three different elements; token embeddings, which represent the identity of each single unit of a subword; positional embeddings, which maintain the sequence of the tokens in a sentence; and segment embeddings, which differentiate between the various segments in the case of multiple sentences. Such composite embedding formulation makes sure that transformer model does not lose the word meaning, position or sentence structure concomitantly. The embedding matrix obtained is then fed into the transformer encoder layers which facilitates successful contextual learning to classify sentiments.

ISSN: 3071-3498
PP: 01-13.

3.3. Transformer Encoding

The fundamental element which gives rise to deep contextual relationships among the tokens in the input sequence is the transformer encoding. Each token is then converted to a sequence of token embeddings by representing it first into a high-dimensional continuous vector space by embedding layers. These embeddings are then processed with several stacked transformer encoder layers which entails a self-attention under-layer with a position-sensitive feedforward neural network. Such a layered architecture allows the model to successively optimize the representation of tokens by incorporating the contextual information about the whole sequence. The self-attention process enables every token of the sentence to dynamically pay attention to all other tokens of the sentence, independent of their location. This mechanism is based on three matrices, which are computed using the input embeddings via learned linear maps, namely, queries, keys, and values. The similarity of query vectors and key vectors is used in the computation of the attention score between the tokens arguably, which establishes the relevance of one token to another. The square root of the key dimension scales these similarity scores to stabilize learning gradients during training and apply the softmax function which normalizes the scores to obtain attention weights. To explain the operation of attention in simple terms, the model can be alleged in the following way: the model will compare each token with every other token in order to determine the amount of authority that should be assigned to them in the context of creating a contextual representation. A weighted average of the value vectors is then computed using the normalized value weights which then give an updated representation of each token incorporating the necessary contextual information. In order to cover a variety of linguistic configurations, many attention heads are utilized concurrently and thus, the model can concentrate on varying semantical as well as syntactic facets of the input at the same moment. After the self-attention operation, the results are fed through feedforward layers, in addition to the residual connections and layer normalization, which enhance the gradient flow and the training stability. The transformer successfully learns long-range dependencies, contextual sentiment clues and complicated language patterns using several layers of encoders and, therefore, is very useful in tasks of sentiment analysis.

3.4. Sentiment Classification Layer

The sentiment classification layer is the last layer of decision maker in the suggested transformer based set up. Once the input sequence has been sequentially processed by a series of transformer encoder layers, each token is modeled using a contextualized embedding with rich semantic and syntactic information. One of these representations, that which relates to the special classification token (also known as CLS), is regarded as a summary of the whole input sentence on a global scale. This is due to the fact that in self attention, the CLS token looks after all other tokens and compiles information that was pertinent of the general meaning and mood of the text. Such a contextual embedding of the CLS token is then subjected to a fully connected neural layer that will transform the high-dimensional output representation of sentences into a lower dimensional output space, which matches the predetermined sentiment summaries, i.e. positive, negative and neutral. The transformation is with the application of learned weight matrix and addition of a bias term so that the model is able to project the representation of the sentence on a linear basis into class-specific

ISSN: 3071-3498
PP: 01-13.

logits. These logits are the raw confidence scores of each category of sentiment. A softmax activation is used to transform these scores to a probabilistic interpretation. The softmax function can be explained in simple terms in that each class score is exponentiated and then the scores are normalized in a manner such that the scores obtained have a range of values between one and zero and the sum to one. Output vector is therefore the prediction of probability distribution between the basses of sentiment, and the most probable is the sentiment chosen as the final sentiment prediction. In training, categorical cross-entropy loss with the predicted probabilities compared to the actual sentiment labels is used, and the error is propagated in reverse to the classification layer and the transformer underlying parameters. This layer-to-layer optimization allows the general-purpose language representations that are learned in the pretraining phase to be determined to task-specific sentiment differences. Due to this the model is able to capture finer emotional signals, contextual polarity changes and domain-specific sentiment expression, which results in enhanced classification accuracy and strength.

4. RESULTS AND DISCUSSION

4.1. Datasets and Evaluation Metrics

Experiments are performed to exhaustively examine the efficiency of the suggested sentiment analysis structure by using recognized benchmark sentiment information that are frequently applied to scholarly studies. These datasets are normally written material that has been marked by the presence of sentiment labels like positive, negative and neutral, and covers a wide area like reviews of products, movie reviews, and social media content. Benchmark datasets also enable comparison to other existing methods and one can use it as a standardized platform to compare the model performance. The datasets are further subdivided into training, validation, and test sets before experimentation to provide an opportunity to optimize the model and train hyperparameters and evaluate the model performance without bias. There are also instances of stratified sampling to maintain the original splits distribution of the classes to minimize evaluation bias. An ensemble of standard classification measures is used to assess model performance and is an indication of various predication qualities.

Accuracy is a general measure of the proportion of the instances that have been correctly classified and gives a vague measure of the efficiency of the model. Accuracy can however not be enough when there is imbalance in the sentiment classes. Thus, the use of precision to measure the percentage of correctly predicted positive cases to the total number of cases predicted to be positive is the measure of reliability of the model. Remember, sensitivity or recall, is used to measure the fraction of actual positive instances that the model marks, meaning its capability to represent spoken sentiment expressions. F1-score which is the harmonic mean of precision and the recall is an unbiased measure as it takes into consideration the false positive and false negative errors. Besides these quantitative measures, confusion matrices are examined to better understand why phenomena of performance of classes and misclassification occur. Such a thorough assessment plan will provide a strong and balanced measurement of the proposed model, taking into account its advantages and

ISSN: 3071-3498
PP: 01-13.

flaws in the resolution of the difficulties of various types of sentiments and various distributions of data.

4.2. Performance Comparison

Table 1: Performance Comparison

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
SVM	82.4	81	80	80
LSTM	85.7	85	84	84
CNN	86.3	86	85	85
BERT	91.8	92	91	91
RoBERTa	93.2	93	93	93

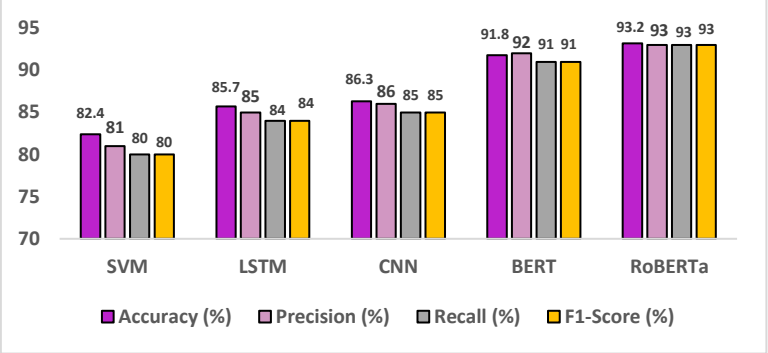

Fig 3 - Performance Comparison

4.2.1. Support Vector Machine (SVM)

The Support Vector Machine model has shown a minimum of 82.4 percent accuracy, which means that it is useful in segregating the sentiment classes based on manually created textual features. Even though SVM demonstrates moderate accuracy and recall rates equal to 81.0 and 80.0, respectively, it can be limited by the use of sparse feature descriptions, i. e. TF-IDF. SVM does not achieve the same F1-score as it has a low capacity to retrieve both the contextual and semantic relation in text, which demonstrates its low-levels of robustness in dealing with complex sentiment expressions.

4.2.2. Long Short-Term Memory (LSTM)

The LSTM model is also noticeably superior to SVM with an accuracy of 85.7. Modeling sequential dependencies in text, LSTM can represent word order and contextual flow which results in improved performance in the areas of precision and recall of 85.0% and 84.0 respectively. This enhancement signifies the capabilities of the model to remember sentiment-relevant information in longer sequences. But it is sequential in nature and therefore cannot be parallelized or scaled thereby limiting its performance at large scale datasets.

ISSN: 3071-3498
PP: 01-13.

4.2.3. Convolutional Neural Network (CNN)

The sentiment classifier based on CNN also performs better as it gains an accuracy of 86.3. CNNs are the best at recognizing local n-grams and salient sentiment indicators with the help of convolutional filters that help to improve the precision and recall rates to 86.0% and 85.0%. With these advantages, CNNs are mostly contextual based and do not pay much attention to long-range dependencies, so their performance is slightly lower than that of transformer-based models.

4.2.4. Bidirectional Encoder Representations of Transformer (BERT)

The accuracy of BERT is 91.8 which is significantly high as compared to the traditional and deep learning models. Its two-way self-attention system allows maximum understanding of the context, thus it is able to capture subtle contextual change in sentiment as well as polarity context. The BERT high precision, recall, and F1-score values are over 91% which testify of the robustness and high generalization ability of BERT in various sentiment expressions.

4.2.5. RoBERTa

RoBERTa is the most successful in terms of general performance and accuracy of 93.2 indicates that increased pretraining and more training corpora demonstrate optimality. The sentiment classification is reliable and balanced, judged by the high consistency of the accuracy and recall as well as F1-score, which is at 93.0%. Subsequent transformer designs yield significant performance improvements, as these findings confirm that RoBERTa is the most useful in terms of the models under consideration.

4.3. Discussion

The results of the experiment provide a clear evidence that models based on transformers show significantly better performance in the domain of sentiment analysis as opposed to traditional machine learning models and conventional deep learning models. Current models (SVM, LSTM, and CNN) use either handcrafted features or contextual modeling that is not comprehensive and therefore does not allow them to decipher the subtle and contextually specific aspect of human sentiment. In contrast, transformer based architectures exploit the capabilities of self-attention to learn global dependencies on entire input sequences and learn more about the relationships contained in that semantics, sentiment changes and nuanced linguistic features like negation and emphasis. This contextual representation learning is noteworthy in learning a single representation, but its coverage is more extensive, and this element is relevant to why the latter models, such as BERT and RoBERTa, are so effective. The other important aspect that predisposed the success of the transformer based models is large-scale pre training of massive unlabeled corpora. In pretraining, such models are trained with rich linguistic regularities, syntactic phonemes, and lexical relations which can be transferred to a grand array of natural language processing problems. With the correct intentional optimization of these pretrained models on sentiment-relevant datasets, it can become relevant to specific sentiment manifestations and therefore simply needs to tweak its generalized language comprehension. Consequently, these transformer-based models can be trained to high accuracy and robustness models despite relatively small amounts of labeled training data, which is one of the most significant problems in supervised sentiment analysis. Moreover, optimal approaches

ISSN: 3071-3498
PP: 01-13.

to pretraining like the ones used in RoBERTa improve performance by boosting the use of data, training goals, and parameter efficiency. These upgrades yield a more stable convergence and a superior generalization than the previous transformer models. Although the computational needs by the transformer based systems are still larger than those of the standard systems, the level of benefits in predictive capabilities and adaptability of these systems are important enough to undergo their application in practical sentiment analysis frames. In general, it can be concluded that pretrained transformer models are an effective and scalable sentiment classification solution, especially in cases when high accuracy and context sensitivity are needed.

5. CONCLUSION

This paper has shown an in-depth research on the improvement of sentiment analysis by the introduction of transformer-based natural language processing models. Transformer architectures can compute long-range dependencies and contextual relationships in a text more efficiently than traditional and sequential deep learning models due to the fact that they can exploit self-attention mechanisms. The analysis revealed that contextual representations acquired by transformer have a substantial advantage in comprehending expressions with sentiments, such as more complex lingo aspects of negative expressions, sarcasm, contextual turning of the polarity. Transformer-based methods provide a stronger and semantically meaningful structure of sentiment classification works, therefore. The proposed framework was also able to demonstrate the value of transfer learning with a large scale experimental setup in fine-tuning pretrained language models to perform sentiment analysis on benchmark sentiment datasets. Models like BERT and RoBERTa continuously surpassed both traditional machine learning algorithms, like Support Vector Machines, and as well as more complicated deep learning algorithms, like CNNs and LSTMs. These improved performances were tested in all the performance measures and factors (accuracy, precision, recall, and F1-score) indicating the improved generalization ability and classification consistency of transformer-based architectures. The results establish that the idea of pretraining on unlabeled corpora with large scale allows transformers to learn deep linguistic knowledge which can be successfully transferred to downstream sentiment analysis tasks despite small amounts of labeled data. Transformer-based models are also more flexible and scalable to various domains and sources of data in addition to having a better predictive performance. Their capability to be customized to achieve particular tasks renders them especially useful in the real-world contexts of analyzing customer feedback, social media and opinion mining in the e-commerce and finance domain. The paper however also indicates limitations of issues pertaining to computational complexity and resource demands that could restrict application in low resource settings. These issues are still key in addressing so as to achieve a viable adoption. The possibilities of multilingual and cross-lingual sentiment analysis can be the direction of future studies that can serve to promote global use and multilingual environments. Limited computational overhead with great precision can be further achieved through the development of lightweight and efficient transformer variants. Also, it will be necessary to integrate explainability and interpretability methods to enhance transparency and trust in sentiment analysis systems, especially in sensitive areas of application. On the whole, this piece of writing confirms the

ISSN: 3071-3498
PP: 01-13.

revolution of transformer-based models in sentiment analysis and forms a solid base to continue the progress in this sphere.

REFERENCES

[1] B. Liu, *Sentiment Analysis and Opinion Mining*, Morgan & Claypool Publishers, 2012.
[2] S. M. Mohammad and P. D. Turney, "Crowdsourcing a word–emotion association lexicon," *Computational Intelligence*, vol. 29, no. 3, pp. 436–465, 2013.
[3] A. Esuli and F. Sebastiani, "SentiWordNet: A publicly available lexical resource for opinion mining," in *Proc. 5th Conf. Language Resources and Evaluation (LREC)*, 2006, pp. 417–422.
[4] J. W. Pennebaker, M. E. Francis, and R. J. Booth, "Linguistic Inquiry and Word Count (LIWC): LIWC2001," *Mahway: Lawrence Erlbaum Associates*, 2001.
[5] B . Pang, L. Lee, and S. Vaithyanathan, "Thumbs up? Sentiment classification using machine learning techniques," in *Proc. ACL*, 2002, pp. 79–86.
[6] B. Pang and L. Lee, "Opinion mining and sentiment analysis," *Foundations and Trends in Information Retrieval*, vol. 2, no. 1–2, pp. 1–135, 2008.
[7] T. Joachims, "Text categorization with support vector machines: Learning with many relevant features," in *Proc. ECML*, 1998, pp. 137–142.
[8] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," in *Proc. ICLR*, 2013.
[9] J. Pennington, R. Socher, and C. Manning, "GloVe: Global vectors for word representation," in *Proc. EMNLP*, 2014, pp. 1532–1543.
[10] Y. Kim, "Convolutional neural networks for sentence classification," in *Proc. EMNLP*, 2014, pp. 1746–1751.
[11] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
[12] K. Cho et al., "Learning phrase representations using RNN encoder–decoder for statistical machine translation," in *Proc. EMNLP*, 2014, pp. 1724–1734.
[13] A. Vaswani et al., "Attention is all you need," in *Proc. Advances in Neural Information Processing Systems (NeurIPS)*, 2017, pp. 5998–6008.
[14] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proc. NAACL-HLT*, 2019, pp. 4171–4186.
[15] Y. Liu et al., "RoBERTa: A robustly optimized BERT pretraining approach," *arXiv preprint arXiv:1907.11692*, 2019.