Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 3, Issue 7 (2025)

Text Classification with Deep Learning and Transfer Learning: A Survey

[1]Rabia Rehman, [2]Talha Farooq Khan, [3]Mubasher Malik , [4]Muhammad Sabir, [5]Usama Shahid

Article Details

Rabia Rehman
Department of computer science, University of southern Punjab, Multan
rabiabaloch912@gmail.com
Talha Farooq Khan
Department of computer science, University of Southern Punjab, Multan
Corresponding Author Email: talhafarooqkhan@gmail.com
Mubasher Malik
Department of Computer Science, University of southern Punjab, Multan
hodcs@usp.edu.pk
Muhammad Sabir
Department of Computer Science, University of Southern Punjab, Multan
muhammadsabir@isp.edu.pk
Usama Shahid
Department of Computer Science, University of Southern Punjab, Multan
usamashahid852@gmail.com

ABSTRACT

This paper review 2 Section briefly discusses the evolution of the text classification algorithms including rule-based and traditional machine learning approaches and deep learning models including pre-trained models. Being one of the significant areas in the NLP [1] text classification is useful for many uses including search and analysis of opinions within texts. The AI tools like ImageAI CNN, RNN and pre-trained models like BERT [2] In performing classification on text is emphasized and how each play a vital role in dealing with large and complex texts accordingly. The problem and directions for future work for the text classifications are also outlined advice to researchers and some practitioners in the study area.

DOI: https://doi.org/10.63075/mvp6rq97

INTRODUCTION

Given text understanding which is a part of natural language processing Text classification [3] entails putting a piece of text into a given set of categories. Originally the field mainly relies on artificially created rules in nature such as the keyword matching approach and has been developed into using statistical learning methods like Naive Bayes and SVM etc., and the automation and efficiency of the classification performance has also been enhanced. We have deep learning technologies such as deep convolutional neural networks, recurrent neural networks and others in the 21st century exploit at a subtler level of textual semantics but are designed with specific knowledge in mind. At the end, there are Transformer-based pre-trained models such as BERT & GPT have appeared thus to raise the bar by training effective language representations. Despite significant It has been a great progress however, there are some restrictions such as the low-resourced language processing, the ways to enhance the model. As identified in the article flexibility generalization and low resource use are the three fundamental elements of sustainable design.

There has been an increasing concern with the incorporation of features that would make the text classification models [4], [5] more accessible and flexible in the recent past. Even the latest models like BERT and GPT have shown enhanced results their execution wants large computational resources and large data sets which is making it difficult for easy adaptation. More attention is paid to the development of the so-called lite-versions of these models–less complex architectures that introduce a similar level of performance loss for example DistilBERT or TinyBERT. Smaller models are greatly helped by these models especially for low resource languages and applications in low compute environments which makes NLP technology more inclusive.

It is also thus possible to have the follow-up as an essential area of investigation where researchers aim at proposing solutions that combine the transfer learning approach and zero-shot learning to enhance the computational models generalization ability for text classification tasks. By this, transfer learning is very effective in reusing the trained model for other entirely unrelated task or on another data set of different domain without much learning. Likewise the nascent area of zero-shot learning has also been presented to address the problem of classifying without needing samples of particular classes. Through these techniques, researchers hope to close the gap between the Resource Rich Environment and the Resource Starved Environment so that text classification systems can be easily applicable to a variety of situations.

Furthermore, as it will be shown interpretability is perhaps the single most critical characteristic of classification in the context of text classification. Similar to traditional ML decision making deep learning and particularly Transformer based models, as they become more advanced are opaque in their decision making process hence affecting trust. There is research on building the XAI approaches that should offer information on how such models are able to make such predictions. This is even more so in the fields of healthcare law and finance most of which are very sensitive and require high levels of accountability. Some of the methods can therefore be adopted by researchers to enhance interpretability which will keep the more complex text classification models more accurate while at the same time allowing for easy and better understanding.

Last but not least, the issues of sustainable development of text classification systems and their deployment are currently among the top priorities of the NLP community. In recent years concerns about power consumption have arisen as models become larger and more complex. Stakeholders are also studying approaches toward training and inference to reduce carbon impacts without loss of efficiency. Other methods like quantization pruning of deep models and using knowledge distillation techniques are already under research for better computation and better performance. Solving these problems will be crucial for achieving the improvement of the next generation of text classification models in terms of both discussing sustainable design principles and increasing efficiency.

FUNDAMENTAL KNOWLEDGE

APPROACHES TO TEXT CLASSIFICATION

To begin with, text categorizing used rule-based methods that categorized text by the laid down rules accurate but expensive and rigid in application. Supervised learning techniques like Naïve Bayes, SVMs and so on are used scientific methods in this kind like decision trees and cluster analysis as well as some other unsupervised learning like automated and knowledge-based as well as learning and adaptation are highly data-intensive processes. CNNs in particular are the state of the art in deep learning. They have outperformed in learning the complex features and capturing the deep semantic information such as, though it demonstrates semantics especially enhancing the text classification on the pre-training and the fine-tuning framework.

BASICS OF TEXT CLASSIFICATION

In text classification, the first and unique step is transforming raw text to a convenient format that can be manipulated by the model. The pre-processing techniques, which included cleaning, tokenization, elimination of stop words and stemming. Feature extraction is to done after preprocessing of text and we have ways of converting text to feature vectors like BoW, TF-IDF and Word embeddings like Word2Vec. While the model training and model evaluation activities use the labeled data set, the performance of these models is assessed. Evaluation metrics were used namely the measures including accuracy, recall, and F1 scores, which have farther been acritical into and developed hyperparameter tuning and cross-validation.

TYPE OF TEXT CLASSIFICATION MODELS

Here some of the most popular and comprehensive models in machine learning are discussed including Naive Bayes, SVM, CNN RNNs, and transformer models. The issues and their realization concerning development background theoretical foundations and main components of each model are discussed proving the efficiency of achievements classification challenges across the identified types of applications.

NAIVE BAYES CLASSIFIER

Speaking of the concrete field of NLP especially text classification one identifies one of the most important positions in Naive Bayes classifier. This inherent straightforward probabilistic classifier using Bayes has turned into effective methods solution of problems for example of spam-mail recognition and document categorization because of the simplistic mathematical rule involved as well as its effectiveness implementation. Although, using the Naive Bayes classifier implies a model of independence of the feature where the probability density of one feature does not depend on the value of any other feature which is often not true in reality the given model admittedly is quite simple demonstrates high rate of effective classification depending on certain datasets. The theoretical basis of Naive Bayes classifier can be traced with Bayes theorem although this is a very simple but effective probabilistic coefficient which still holds significance today in theory of probabilities.

SVM-BASED CLASSIFICATION METHOD

The Support Vector Machine (SVM) as presented by Cortes et al, is a robust machine. Recurrent learning model commonly employed in text categorization problem its basic premise is on the ability to establish the most suitable boundaries is defined using the maximum-margin hyper plane to split decision space into two different categories of data points. SVM stands out in the task of generalizations therefore, it scores a perfect score in a large number of high-dimensional data Classification problems such as text classification.

DECISION TREE CLASSIFICATION METHOD

In text classification the use of the decision tree models seeks to determine which given category the texts belong to through learning the decision rules from the text features. One is to start first, by identifying the features of the text that needs to be mined information that are counts for example of words and terms of documents and other papers. The decision tree model used in this paper is a technique falling under supervised machine learning top down division of a tree structure. What's the most important objective of its task? Diagnose the training set to develop set of decision rules and hence generation of the perfect model that could predict the category of target variables. Last of all, after the construction of the entire decision tree many of the branches are removed or say used to the model to eradicate some adverse consequences such as over fitting and noisy input data on the classification performance.

The decision tree is an indirect algorithm of tree like structure constructed stepwise from data table in the form of supervised learning to find the category of the target variables training from learning data through decision rules. This model is established in top-down approach only unlike in other models of learning once the decision tree is fully developed one is able to utilize methods of pruning in order to reduce the noise and also the level of over fitting. Furthering the optimization of the model in computing the probabilities of new points.

DEEP LEARNING FOR TEXT CLASSIFICATION

Convolutional neural networks and recurrent neural networks used in image recognition and machine translation correspondingly have undergone a radical change by learning. Inter-connectedness of features involved these highly indicate. This section presents the basic types of networks such as FNNs, RNNs and CNNs in text classification benefits of specific architecture and further improvements of the model optimization [6] in use of the text embedding methods.

BASIC NETWORK-BASED TEXT CLASSIFICATION

Based on automatically discovered features by DL technologies all the hidden feature patterns were minimized the necessity of the feature engineering clarify the superiority and the remarkable performance on text classification of feed forward neural network and Recurrent neural network and convolutional.

Feed forward Neural Networks (FNN) [7] the foundation of deep learning in text as a fundamental model. Classification system with inputs embedding net layers multiple laterals of a neural network Input layer, hidden layers, and output layers. The FastText [8] models do this by incorporating character level n-grams into the features this enhance the training and performance of rare words. Nevertheless, it remains vulnerable to abolishing long distance sequences and semantic structures.

To further analyze the deep semantic content of texts as well as to resolve that appears in longer pieces of text, a renewed way for addressing these context dependency issues has been sought in the RNN. Therefore, to model of the sequence of the text a methodology known as Recurrent Neural Network (RNN) [9] Is used. However, almost all standard RNNs have poor performance for long sequences owing to the vanishing gradient problem. This has a tendency of making it difficult to come up with a model that requires capturing of long distance relations. To overcome this challenge, enhancement of RNNs like Gated Recurrent Neural Networks and Long Short-Term Recurrent Neural networks including; Memory networks (LSTM) (Hochreiter). Get Current popular for using text classification research LSTMs are in charge of the regulation of information through a memory cell and three particular gates an input gate a forget gate. In this way stands open the so-called output gate and thus maintains long-term dependencies information.

Moreover, RNN [10] outperforms in dealing with temporal mannerisms of the text while there is also showing that Convolutional Neural Networks (CNN) [11] is the primary approach to phrase or keyword extraction by TextCNN model using multi sizes of filters to capture the local features of the text. TextCNN uses max pooling for building representations necessary for forming higher levels depending on fine to detect local semantics but it lacks long distance relations.

Moreover, the advancement of the technologies in text embeddings has added to the improvement of model performance of it at the character level, word level, and sentence level each solving a separate problem. Vocabulary based problems learning grammatical and semantic structures and memory the possible relations of consecutive sentences.

In general, specifically comparing to conventional machine learning techniques the basic deep learning network structures have demonstrated large gains in ordering information and semantic depth extraction. Nevertheless, these basic models have their own weaknesses as well mainly including the simplicity of the model structure on the one hand and the ability to enter particular sorts of information on the other hand.

GRAPH NEURAL NETWORK-BASED CLASSIFICATION

Graph Neural Networks (GNNs) [12] suggest improving the effectiveness of class identification by extracting the structural characteristics of texts discrete internal representations in the languages levels prior knowledge of texts at the syntactic and semantic levels of structuring. All the GNNs apply the information dissemination approach thereby constructing node representations iteratively erasing and reconstructing information obtained from neighboring nodes and then self-node representation.

Different aggregation and combination functions have brought up the heterogeneity of GNN structures suitable for various circumstances in the application of systems. A versatility and high performance of GNNs are observed in various text classification studies at varying scales from words to documents. While TextRank was one of the first approaches to using graph structures for text classification to create graph models with nodes and edges as text units and the link between them that can accommodate for a number of text processing needs from sentiment analysis to subject categorization.

At a higher document level, Peng and others proposed a graph-based CNN model in 2018. This model translates text into graphs of words whereas a convolution features the text, word, graphs for document classification. This method is indeed able to parse up disjoint and long distance dependencies within texts due to the CNNs exploited semantic learning capability. Yao and his team attempt another approach in 2019 by constructing a heterogeneous graph construction consisting of the word and document nodes for the categorization of text data under the act of using a newly introduced innovation of document granularity. In this graph, the word strength is used to measure the words relation while the between them are based on the semantic relation of words Using Graph for this structure and to make GCNs transport and escalate information between Documents in order to learn the embedding for new Words & Documents this strategy boosted the performance over the various benchmarks used where average accuracy of the model was about 84.5%.

For the sentence-level classification jobs in 2019, Zhang and his group used the Dependency Tree (DT). How to first isolate the structure within the verbs and then apply the Text Graphs and then use the GNNs for a sentence classification. This method has clearly provided a much better approach of getting a higher level of accuracy of the classification model as compared to the previous methodology. In light of this, how it effectively blends long distance dependencies between the words. In that vein, the present research also extends this line of inquiry. In year 2020, Zhang & others developed independent graph representation towards the specific document of GNNs to learn features of words on the document. This approach gives a capability emphasize for contextual embeddings of out of vocabulary words in documents it has performed exceptionally in multiple standard datasets.

TRANSFER LEARNING-BASED CLASSIFICATION

Technological transfer [13] which employs comparable knowledge enhances the model's Skills upgrade in so far as performance on new tasks and reduction of the necessity of the labels. It is particularly useful in cross-domains or the multi-level text classification tasks common in fake news detection and topic identification transfer. In this paper, the authors have proved that WG learning has an edge over conventional deep learning approaches with limitation called models and has been useful in different fields such as sentiment analysis as well as computer vision.

Similarly, in 2018 when developing a new fine-tuning [14] method called universal language model fine-tuning (ULMFiT) [8], language models where transfer learning was utilized in natural language processing NLP tasks. It can be used to many text classification problems without the support of more data resources and improved efficiency at estimating the population mean in small sample data conditions and decreased error rates. The consolidation of state change was first systematically reviewed by Banerjee and colleagues in 2019 and the HTrans, a hierarchical transfer learning approach was proposed to address the multi-label text classification issues. They have demonstrated high improvement in performance when they are able to map classifiers from the upper structure of the hierarchy to that of the lower structure.

Later in the same year Houlsby Et Al. tackled the question of parameter efficiency when fine-tuning pertained models with an efficient adapter module approach that can provide satisfactory performance after tweaking a relatively small number of parameters. In 2020, Raffel and others proposed a framework that can categorized all text based tasks into a single As another text-to-text application of NLP transfer learning it offers new opportunities for application in a variety of industries. In 2021, Cao and The others proposed the new cross-domain sentiment classification fine-grained cross Deep Transfer Learning generative mechanism to promisingly advance the utilization of unlabeled data for the target field. This approach requires a domain proposed adaptation model to help minimize the disparity of feature distribution between the considered source and target domains.

Transfer learning can reduce the occurrence of data and annotations in text classification to some extent but it only pays attention to problems such as cross domain knowledge transfer how to assess the cross domain transfer ability of knowledge and the problem of content privacy. Specifically, the work shows that context-dependency of word meanings may result to negative transfer from a weakly related source domain to the target domain, hence the call for formulation of new approaches with minimized such impacts. This paper focused on a classification of the methods based on transfer learning models in the context of text classification tasks.

ULMFiT

Originally introduced by (Howard), ULMFiT is the first transfer learning methodology that has been specifically designed for use in NLP tasks. This method involves taking standard text-translation models and training them for specified uses like text categorization the error rates are reduced considerably even when there is a minimal availability of the labeled dataset. This fine tuning phase refines the description of the model to the specifics of the target application without a need for significant data increase. This is the reason why ULMFiT is effective after all the last step of the process fine-tunes the model hierarchically while keeping the learned representations intact in order to make it suitable to the task at hand in other words ULMFiT is a very flexible tool that can be used in a variety of text classification problems.

HTrans

HTrans proposed by Banerjee et al., in 2019 is a hierarchical transfer learning for solving the multi-label text classification issue. As this concept puts classifiers in a hierarchical structure at the highest level this strategy allows one to fine-tune classifiers from general to specific levels. Compared to the single-label classification setup HTrans incorporates the dependencies between labels using hierarchy and hence always provides enhanced classification results compared to the basic method. This method shows its ability and flexibility in trying to handle multiple label classification scenarios that are difficult for flat classification methods to try.

ADAPTER MODULES FOR PARAMETER EFFICIENCY

Specifically, to tackle the computational inefficiency of fine-tuning large pre-trained models Houlsby et al.mArch proposed the adapter module strategy in 2019. Unlike, model Sz where all of the model parameters are updated adapter modules comprise of small trainable parameters that are inserted into the pre-trained model. Specifically, only these adapter modules are optimized during fine-tuning while the original model parameters are fixed [15]. This approach significantly cuts down the number of computations and resource utilization needed for fine-tuning making this a practical solution for practicing transfer learning in text discrimination tasks across different platforms.

UNIFIED TEXT-TO-TEXT FRAMEWORK

For instance, Raffel et al., in 2020 offered a unified text-to-text transfer training approach that makes all NLP tasks to be in the same format. This framework employs a single model architecture to solve different tasks like classification, translation, and many more as a sequence generation task. When the task structure is unified by the described approach the application of transfer learning becomes less complicated and the area of its applicability is expanded. This versatility increases the utility of transfer learning in a vast array of text-related problems lowering the difficulty level of model SWITCH in each case.

FINE-GRAINED CROSS-DOMAIN SENTIMENT CLASSIFICATION

Cao et al proposed a novel deep transfer learning mechanism specifically for the polarity classification problem in cross domain in 2021. Essentially, this method involves a domain adaptation model that helps reduce the feature space disparity that is always witnessed when transferring knowledge from one domain to another. This approach elaborates a way of improving classification accuracy and eradicating the issues of domain shifts through efficient utilization of unlabeled data drawn from the target domain. This is particularly an advantageous position when there is little labeled data in the target domain for sentiment classification.

PRE-TRAINED MODEL CLASSIFICATION

First and simple models created with large data sets of different sort generalize characteristics of data and turn into potent but it will be the starting point for the definite NLP tasks. Moreover, this approach allows tasks to adjust with the foundation level and the performance improves and reduces training data. In general, it can be noted that rich inherent generalization in pre-trained [11] models plays an important role in terms of model effectiveness information that in turn facilitates fast convergence low dependency on data, high performance and endless reduced risk of over fitting.

General characteristics are tasted for pre-trained models in large datasets and all of them are presented in an optimized shape point as a certain amount for specialized uses such as text classification. This strategy is helpful in that there is less demand for training data accelerates the rate of model fitting optimizes its efficiency and reduces the chances of having a model that is too-good-fitting due to the available literature of this type of research has proven that global imbalances and extensive information contained within Pre-trained models use unsupervised learning techniques so frequently due to their ability to that are trained on a large amount of data that do not contain labels. Self-supervised learning is a category that is derived from the unsupervised learning discipline and supervised methods acquires knowledge about what signals to look for in unlabeled data in order to make learning possible. This makes it possible for models to learn feature extraction without a parametric annotation and this greatly reduces the costs of data preparation. Interestingly, self-supervised learning has been revealed to be helpful in large-scale models in language as in the instance of Due to these improvements the natural language processing technology has been enriched by BERT [16] and GPT [17].

The initial existing pre-trained language model is Word2Vec[18] which trained by skip-gram and convey integration continuous bag of words strategies and contains features for text classifying. However, it produces Conventional word embedding presets the fixed words which do not contemplate about the semantic aspect of the different meanings of the word "apple". Environments which might restrict model efficiency thus Word2Vec (Wu T) has a static word representation of the word "apple," which can limit the abilities of the model when facing tasks such as text classification as for the problem of modelling polysemy in the frameworks of such models like Word2Vec researchers speaking for example the procedure known as Masked Language Model (MLM) which is used in BERT.

These two ways allow the model to learn context related word embeddings without requiring manual annotations MLM practices the process of random masking of the words in the input text using this to train the model to predict these words based That requires bidirectional context which in turn can help the model equip the system with higher sensitivity to semantic changes and greater potential for generalization. During the text classification with BERT there is a so-called [CLS] in front of the string TEXT, input sequence. In the architectural design of BERT this token has a particular meaning in manufacturing a fixed-size dense vector for the classification of the whole input sequence.

Specifically, the MLM training strategy has equipped BERT with manifold potentialities linguistic functions needed for the question is whether social bots truly have comprehensive language understanding capabilities. Once pre-trained BERT is then ready for fine-tuning to make adaption to many NLP problems like text and feature classification identification of named entities and creation of question and answer facilities. In these tasks understanding of bidirectional context of text is very much helpful and in presence of BERT this is exactly the case enhances the model performance.

For example, in the sentiment analysis tasks such as predicting the sentiment polarity of words MLM can learn a lot of contextual information enables BERT better sentiment analysis in the texts. In question answering [19] systems for instance, The cut–in-context ascertainment of bidirectional context makes it easier for BERT [20] to fine-tune the position of answers With the help of the new MLM training method for the BERT model not only has the door to the new approach been unlocked direction for the resolution of the pre-trained language models but has also contributed significantly to the advancement of the field in the same process of NLP. About the bidirectional contextual dependency I think this is one of the biggest advantages of this model which makes it outstanding performance level when it comes to many tasks. Though, there are some issues for the BERT model including the tremendous recourse demand and the inability to process large texts. Despite these difficulties it can be concluded that the effective implementation of BERT [21] and its derivatives in the NLP discipline is beyond doubt illustrates the large amount of potential and beneficence of deep learning pre- trained models.

EXPLAINABILITY IN TEXT CLASSIFICATION MODELS

Explainability [22] is now standard in contemporary artificial intelligence systems as a crucial component since the degree and kind of trust demanded when using AI are different from that of traditional mechanical systems where risk levels in legal, financial, and most especially medical systems are high. In text classification, explainable models enable the people who will make the decisions to have some level of independence and be in a position to countercheck their models. To address the facet of interpretability these models methods like attention mechanism SHAP [23] and LIME are being used. For instance, attention mechanisms can help spot out dates and other words or phrases a model considers vital for categorization. While SHAP and LIME explain what the feature's input contributes to the output which gives the model's working explanation. This is especially true when it comes to medical applications for instance determining diseases from records since a wrong decision can cost the patient their lives. Likewise, in the case of business sentiment analysis explanation offers organizations ways of comprehending commentaries from consumers. By making the explain ability aspect the focus of their research it will be possible to achieve not only high accuracy of the text classification models that are employed but also their ethical quality.

TEXT CLASSIFICATION ALGORITHMS: COMPARISON AND ANALYSIS

EVALUATION TECHNIQUES

Usually, the assessment of the merits and demerits of the classification algorithms takes different aspects also such as accuracy, capabilities of generalizing, computational complexity of a model and the model itself interpretability. In order to fairly compare classification algorithms researchers apply a number of evaluation metrics and methods IT systems and technologies business value communications tools and processes. Of course, accuracy is purely among the most leading evaluation criteria it is actually the identification of correctly directory SYNONYMIZED samples to total samples in every class. The observation of classified samples to the total number of samples makes it possible to refer to the samples of each category as SYNONYMIZED samples.

COMPARISON OF PERFORMANCE

ACROSS SEVERAL COMMONLY USED DATASETS

IMDb Movie Review Dataset: When put to use in sentiment analysis it has a large number of movies and such like particularly in profiling the different reviews with the sentiment polarities being either positive or negative.

20 Newsgroups: A co399 dataset of news group messages of documents in 20 topics. It is very useful when used for classification particularly when it is subjected to category determination.

Reuters-21578: Suggested for use as a news article topic classification news wire dataset.

AG's News Corpus: A multiple category news article dataset for the explicit objective of classifying news.

DIFFERENT TEXT CLASSIFICATION MODELS HAVE SHOWN THEIR RESPECTIVE STRENGTHS AND LIMITATIONS

FNN: FNNs were observed to perform well in some simple text classification tasks and are not very suitable for representing dependencies for the texts where long-distance relations hold and thereby limiting their applicability for complex manipulations.

LSTM and RNN: Due to the recurrent nature of both models these could handle sequential data. Indeed, from the viewpoint of capability of capture long-term dependencies it is possible to state that in many cases FNNs can be outperformed by RNNs. However, this few instances suggest that while LSTMs and RNNs may experience problems with gradient vanishing or exploding for too large sequences.

CNN: As it is known, CNNs perform well by utilizing local relevance and are especially effective in capturing important local patterns in text and are particularly effective for things like activity performance and other assignments that would entail bouncing localized text patterns.

Word2vec: Among all the pre-trained word embedding techniques word2vec can also fairly deconstruct the semantics of word relations so far as it is highly capable depended on downstream model.

BERT: As with any other Transformer-based pre-trained model BERT [24] also operates in a number of linguistic phenomena during. This paper first pre-train and then obtain the new SOTA on many text classification tasks by only using fine-tuning techniques. This is because the algorithm contains bidirectional meaning which circles around a broad accommodation global cognition of language context.

The experiment proves that the models derived out of deep learning (like CNN, LSTM, BERT) usually have a higher accuracy compared to conventional machine learning models (for example, FNN in basis of word2vec) [25] in text classification tasks. Especially BERT, due to its high language comprehension capacity yields very good results in a situation where the dataset introduced contains many semantically associated samples.

It is crucial to understand that deep learning models depends on a rich training data and computational power a lot That is why the differences in the performance of algorithms are largely conditioned by their architecture features for instance, CNNs are efficient in many local features detection while LSTMs or RNNs are more appropriate addressing LTCs and AWD-Transformer's bidirectional BERT pre- trained architecture allows it to seem to lack the ability to comprehend detailed aspects of a given text context. According to Coppers and Henao classification models to be used are chosen depending on the subject if certain conditions or requirements are fulfilled.

To sum up it is indicated that the latest advances focus mostly on the integrated deep learning approach and other methods for the pre-trained text classification language models. Through retraining of a retrain trained models or in creating appropriate input cues for certain uses investigators can greatly enhance the performance of the model performance in different application conditions. As the enhancements of these technologies go on thus it is anticipated that there will be fast development in the text classification field. This article will be useful for the stakeholders in the research field as well as for practicing society as it offers the synthesis of the Working on the problem of text classification multiple authors have worked on the performance of certain text classification models on different datasets. These research findings have purpose and relevance in choosing the text classifier designing the classifier and enhancing its performance apart from making Contributed to the making of text classification useful in choosing the proper classifier designing the classifier and improving the performance of the classifier. Furniture for text classification technology to be further developed and innovated.

ANALYSIS OF DATASET LIMITATIONS

The datasets are essential for developing and testing text classification algorithms although many of the most popular datasets like IMDb, 20 Newsgroups or Reuters have several problem that prevent them from being similar to real-world situations. Most of these datasets degenerate in terms of language variation and topical coverage and are thus not easily extendable to worldwide settings. Moreover, most databases contain outliers because of restricted sampling or labeling errors which in turn affect the reliability of models. For instance, a dataset that has a lot of data in the English language will not capture the issues with multilingual text classification. Secondly, it is challenging to obtain large amounts of labelled data for deep learning in smaller datasets which degrades its efficiency. To these challenges we are witnessing a rising demand for construction of varied and more balanced datasets and improved measurement criteria to assess the effectiveness of text classification models. Such efforts are going to guarantee better models which will not be restricted to a single language domain or context only.

FUTURE TRENDS AND CHALLENGES

The obstacles to the progress of this technology and the increasing datum challenge the text classification field as reflected in the case of text data from a deluge of unstructured text data the power to infer implicit meanings dependency, semantics, and long-distance dependencies. The future development is multimodal classification surpassing the text fusion cross-lingual classification that is learning language restrictions and soaring the performance with deep learning technologies like BERT and GPT [26]. Issues concern and revolve around the treatment of textual input and specifically how natural language will be interpreted. It contains questions and answers on how to physically map the old models on the new vocabulary and how different people tried resolving the issue of multimodality and other issues of cross language processing of text. The extension of deep learning methods demonstrates new opportunities for text classification with an increased level of difficulty the new demonstration proves that progressive outcomes enhance the effectiveness of the text data classification plus more targeted identification of technological innovation and optimization strategies the text classification will be more precise and efficient.

NEW TRENDS IN MULTIMODAL TEXT CLASSIFICATION

There is a shift in the text classification field towards text-video and text-ANY other form of interaction and not simply text-data for richer model creation. The latest multipurpose frameworks such as CLIP and Flamingo takes the textual as well as the vision input, to enhance performance in the area of meme evaluation and captioning of videos. For example, the use of images and text in the same document is possible including text in memes through including both the text and context in the multimodal [27] text classification as needed. In the same way as in video analysis the use of the audio transcript of a sequence with elements of synchronizing it with visual cues for further analysis helps. With the large benefits also come some drawbacks the basic ones among them being the problem of synchronizing different modalities during the training phase and the fact that models with multiple modalities are computationally more demanding. Nonetheless, the opportunities for future development of the concept of multimodal text classification [28] to complement industries including entertainment, education, and surveillance are tremendous. The schemas of these models are yet to improve and when they do so they can revolutionize the boundary and application of text classification.

OPTIMIZING RESOURCES IN TEXT CLASSIFICATION

As deep learning models for text are enhancing it is becoming an important task to enhance these models to save their memory footprint for using them in for example, mobile devices or Internet things. As we have established, most of the models including deep learning models such as BERT and GPT require a very large amount of computational resources thus making it tricky to be implemented in the real world application. To mitigate the issues above pruning quantization and knowledge distillation have been invented. There are two primary techniques namely pruning that essentially downsizes the neural networks eliminating the weights known as inconsequential and quantization which narrows down the precision of model parameters and frequently converts the parameters to 8 bits integers' thereby saving memory and calculating power. Knowledge distillation means learning a smaller model (student) to replicate the behavior of a larger model (teacher) without a comparable level of accuracy and computational capability. For instance, there are smaller versions of BERT [29] like DistilBERT [30] which have found their way to practical use in features like customer service Chabot's and mobile sentiment analysis. By choosing to concentrate resources where they are likely to have the greatest impact researchers may bring accurate and sophisticated text classification systems into general use.

CONCLUSION

This article is a review of the evolution of text classification methodologies rule based and statistical physical methodologies to different deep learning models include CNN, RNN, LSTM, BERT and GPT and has informed them that each of them has added its own value addition in improving the probabilities of raising the bar of accuracy and speed of classification. The technology of text classification can the proposed text classification technology can had been used in various fields such as sentiment analysis and spam detection help strong Index Of business about the successful extraction of valuable information. Among the problems in this work were processing of free text and the further complexity of the semantic Some of them Practical further studies will still continue to investigate the sophistication of the model architecture interpretability cross-lingual and multimodal classification. The aim is thus to develop additional precise unequivocal and for the most part comprehensible text categorization technologies for practical purposes as well as to contribute to the further advancement of knowledge in the field of NLP as a branch of computer science.

REFERENCES

[1] K. Saifullah, M. I. Khan, S. Jamal, and I. H. Sarker, "Cyberbullying Text Identification: A Deep Learning and Transformer-based Language Modeling Approach," EAI Endorsed Trans. Ind. Networks Intell. Syst., vol. 11, no. 1, pp. 1–12, 2024, doi: 10.4108/EETINIS.V11I1.4703.

[2] A. Palanivinayagam, S. S. Gopal, S. Bhattacharya, N. Anumbe, E. Ibeke, and C. Biamba, "An Optimized Machine Learning and Big Data Approach to Crime Detection," Wirel. Commun. Mob. Comput., vol. 2021, 2021, doi: 10.1155/2021/5291528.

[3] M. E. Maron, "Automatic indexing: an experimental inquiry," J. ACM, vol. 8, no. 3, pp. 404–417, 1961.

[4] T. Joachims, "Text categorization with support vector machines: Learning with many relevant features," in Machine Learning: ECML-98, 10th European Conference on Machine Learning, in Lecture Notes in Computer Science, vol. 1398. Chemnitz, Germany, 1998, pp. 137–142.

[5] I. J. Unanue, G. Haffari, and M. Piccardi, "T3l: Translate-and-test transfer learning for cross-lingual text classification," arXiv Prepr. arXiv2306.04996, 2023.

[6] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," arXiv Prepr. arXiv1412.6980, 2015.

[7] A. Joulin, E. Grave, P. Bojanowski, and T. Mikolov, "Bag of tricks for efficient text classification," in Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics, Volume 2: Short Papers, Valencia, Spain, 2017, pp. 427–431.

[8] S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural Comput., vol. 9, no. 8, pp. 1735–1780, 1997.

[9] F. Scarselli, M. Gori, A. C. Tsoi, M. Hagenbuchner, and G. Monfardini, "The graph neural network model," IEEE Trans. Neural Networks, vol. 20, no. 1, pp. 61–80, 2008.

[10] Y. Wu et al., "Google's neural machine translation system: Bridging the gap between human and machine translation," arXiv Prepr. arXiv1609.08144, 2016.

[11] Q. Li et al., "A Survey on Text Classification: From Traditional to Deep Learning," ACM

Trans. Intell. Syst. Technol., vol. 13, no. 2, 2022, doi: 10.1145/3495162.

[12] F. Sebastiani, "Machine learning in automated text categorization," ACM Comput. Surv., vol. 34, no. 1, pp. 1–47, 2002.

[13] F. Zhuang et al., "A comprehensive survey on transfer learning," Proc. IEEE, vol. 109, no. 1, pp. 43–76, 2020.

[14] C. Sun, X. Qiu, Y. Xu, and X. Huang, "How to fine-tune BERT for text classification?," arXiv Prepr. arXiv1905.05583, 2019.

[15] Z. Yang, Z. Dai, Y. Yang, J. Carbonell, R. Salakhutdinov, and Q. V Le, "XLNet: Generalized autoregressive pretraining for language understanding," in Advances in neural information processing systems, 2019, pp. 5753–5763.

[16] J. Howard and S. Ruder, "Universal language model fine-tuning for text classification," in Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Melbourne, Australia, 2018, pp. 328–339.

[17] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Improving Language Understanding by Generative Pre-Training," arXiv Prepr. arXiv1801.06146, 2018.

[18] T. Mikolov, I. Sutskever, K. Chen, G. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," Adv. Neural Inf. Process. Syst., vol. 26, 2013.

[19] Z. Hu, D. S. McNamara, A. Graesser, and Z. Cai, "Open-domain question answering with pre-trained language models," in Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2020.

[20] A. Vaswani et al., "Attention is all you need," in Advances in neural information processing systems, 2017, pp. 5998–6008.

[21] H. Yu, Z. Yang, K. Pelrine, J. F. Godbout, and R. Rabbany, "Open , Closed , or Small Language Models for Text Classification ?," 2023.

[22] M. T. Ribeiro, S. Singh, and C. Guestrin, "'Why should I trust you?' Explaining the predictions of any classifier," in Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining, 2016, pp. 1135–1144.

[23] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in Advances in neural information processing systems, 2017, pp. 4765–4774.

[24] M. E. Peters et al., "Deep contextualized word representations," in Proceedings of NAACL-HLT, 2018, pp. 2227–2237.

[25] F. Majeed, M. W. Asif, M. A. Hassan, S. A. Abbas, and M. I. Lali, "Social Media News

Classification in Healthcare Communication," J. Med. Imaging Heal. Informatics, vol. 9, no. 6, pp. 1215–1223, Jun. 2019, doi: 10.1166/jmihi.2019.2735.

[26] T. Brown et al., "Language models are few-shot learners," in Advances in neural information processing systems, 2020, pp. 1877–1901.

[27] A. Radford et al., "Learning transferable visual models from natural language supervision," in International Conference on Machine Learning, 2021, pp. 8748–8763.

[28] J.-B. Alayrac et al., "Flamingo: a visual language model for few-shot learning," arXiv Prepr. arXiv2204.14198, 2022.

[29] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), Minneapolis, USA, 2019, pp. 4171–4186.

[30] Z. Lan, M. Chen, S. Goodman, K. Gimpel, P. Sharma, and R. Soricut, "ALBERT: A lite BERT for self-supervised learning of language representations," arXiv Prepr. arXiv1909.11942, 2019.