THE TEXT AND SEQUENCE FEATURE EXTRACTION APPROACH FOR CLASSIFYING AND CATEGORIZING COMPLEX DATA

P.L.RAMESH, Vice Principal, K.B.N.College, Vijayawada, Andhra Pradesh, India
V T RAM PAVAN KUMAR HOD, Department of MCA, K.B.N.College, Vijayawada, Andhra Pradesh,

ABSTRACT: Using machine learning techniques, researchers have recently concentrated on processing multi-media data for categorizing the user's search engine queries. To create a reliable classification model that operates in a high dimensional space, a hybrid mix of a strong classifier and a deep feature extractor is used. To create a stochastic belief space policy in this study, three different types of algorithms are merged. These algorithms are belief space planning, maximum entropy reinforcement learning (ML-RL), and generative adversary modeling (GAM). The comparison between various unmodeled adversarial methods, in suggested methods is suggested to achieve robustness because the malicious behaviors were applied in the simulation. That framework is used to reduce the agent's action predictability. The Deep Learning (DL) approach, which is reinforcement-based, can be employed for multi-model categorization.

Keywords: Document Classification, Image Classification, Multi-Model Classification.

INTRODUCTION

On text and image data, the single neural network method may do categorization. Due to RL, learns the proper belief space policy (BSP) from the feature extracted data of the text and image data. The BSP is constructed using a maximum entropy calculation depending on the characteristic of the input data.The availability of multi-media in today's society—including audio, image, graphic, text, and video—on the internet creates a challenge for multi-modal analysis and has drawn increased attention from academics. However, differences in semantic gaps and a variety of modalities have created new obstacles. By using computational models in earlier research, the human brain's processing of multimodal data is simulatively created. Most user data is being acquired in the digital format and stored in the age of technology. These data are widely accessible to the general public as a result of the internet's rapid expansion, which has produced enormous amounts of multi-media data in various formats, including text, photos, audio, and video. The current analysis issues in this field are defined by the use of multi-media data with relatively high dimension and gigantic size characteristics.

Multimedia data processing thus enters a phase of rapid development. The creation of intelligent systems capable of effectively handling this kind of enormous data has emerged as a pressing issue. The number, dimensionality, and complexity of these datasets are also growing nowadays, making them more diverse and complicated. As a result, past studies had difficulty modeling and analyzing multi-media data because they are computationally inapplicable. High dimensionality issues lead to increased system and time demands as well as worsening classification task performance. As a result, the performance of techniques will produce incorrect or inaccurate findings. This is because as the number of features (dimensionality) increases, it becomes harder to distinguish the pair wise distance between points. One of the biggest problems in the fields of the science is classifying and categorizing complicated data, such as video, documents, and photographs. To address these issues, researchers are creating diverse models using DL topologies and structures. But these structures are primarily created for a specific domain or kinds of data. Deep neural networks, a significant advancement in machine learning, have up till now been directly applied to a wide range of real-world applications, including gaming, autonomous vehicles, science, and even the arts. Groundbreaking innovations in a variety of domains, including speech processing, natural language processing (NLP), and computer vision, have been made possible by DL. The majority of these studies focused on the issue of single modal DL as opposed to the difficulties associated with multimodal learning. However, combining a variety of data types in a multimedia system can considerably boost the final detection and retrieval performance, particularly when there are mistakes or missing values in one or a few modalities. The multi-model classification approach is developed by fusing generative adversary modeling, maximum entropy RL, and belief space planning. It is presented in this study to produce the stochastic belief space policy. The initial phase involves extracting the key features for classification from various dataset kinds, including text and images. The suggested RL is provided the features for classification. In planning-based approaches, RL enables the learning of a policy from a black-box simulator that simulates a difficult-to-process complex combination of behaviors.

PROPOSED METHODOLOGY

The study work implemented the multi-model classification algorithm to achieve the stochastic belief space policy. The features are initially taken from several datasets, including text and image databases. These attributes are provided to the proposed RL for training. The method for extracting the characteristics is thoroughly explained in the following subsection.

Feature Extraction and Data Pre-processing Text and Sequences Feature Extraction.

Generally, different natural language processing-based techniques—such as word embedding, TF-IDF, etc.—are used to extract the characteristics from the text files. In this study, the features from the text files are extracted using feature extraction methodologies, specifically word vectorization approaches [23] and N-gram representation [22, 24]. The vector space models are constructed from each text file using the following equation, which is based on the glove word embedding process (1). A vector-space model is a mathematical mapping of the word space can be represented using the Eq. 1.

$$= (w_{1,}, w_{2,j}, \ldots, w_{i,j}, \ldots, w_{l_j,j}\ ) \qquad (1)$$

where d_j represents the length of the document and $w_{i,j}$ indicates the GloVe word embedding vectorization of i^{th} word of i^{th} document.

Image and 3D Object Feature Extraction.

In this research, the images are either grey scale photos, colour images, or 3D objects. The feature extraction is a crucial part in the image classification process. For the MNIST dataset of grayscale images. The features of colour images are represented as h w c, while the features of grayscale images are represented as h w c, where h stands for the input image's height, w for its width, and c for the RGB-based three-dimensional colours. The n number of cloud points used to represent the 3D object's features each contain six different sets of features, including x, y, z, R,G and B. The amount of cloud points causes the one 3D object to vary from the other 3D object, which results in an unstructured representation of the 3D objects. With the aid of a straightforward instance up/down sampling technique, the unstructured 3D objects can be represented in the form of a structured format.

Proposed methodology

The Tuple (S, Aa, Ao, T, O, R, b0,) defines the active perception problem as a planning problem. The state of the word is specified as S = (So, Sp), where So is the set of observable states and Sp is the set of partially observable states. The activity of an autonomous agent is then symbolised by Aa, followed by the action of an adversary by Ao. When an action stands out, an intention can be quickly determined, and the sort of intention can then be determined based on those activities. The trasition probability is given by T: S Aa Ao S, where denotes the space of the probability distribution over the space.$O: S \times A^a \rightarrow \Delta_A a$ is the observation probability. $R: S \times A^a \times A^o \rightarrow \mathbb{R}$ is the reward function and b_0 represents the prior probability of the opponent being an adversary and the γ is the discount factor.

In this research, the following modelling presumptions are made:
1) The adversary is either a hostile foe or a civilian.
2) A civilian adversar is self-interested, and their actions can be modeled by a policy that is reactive.
$\pi^{cil}(a_t^o|s_t)$.

3) A hostile adversary is one that is predominantly goal-directed, as indicated by a recognized MDP.
4) A hostile opponent has bounded rationality, which means that it might not always be able to act in the best interest of the situation. In addition, it is likely to act dishonestly to further its objectives.
Additionally, we assume that a model of typical civilian behavior is accessible. Then, using two parameters to reflect, respectively, the level of reason and the level of deception, we build a parametric set of hostile models. A feed-forward neural network (NN) is used to represent the autonomous agent's policy. This neural network (NN) accepts as input a binary belief state that is produced by Bayesian filtering the hidden intention using an average model. A stochastic policy is what is produced. A belief dependent reward, which promotes exploratory behavior, and a state dependent reward, which ensures safety, make up the reward function.

To lessen the exploitability, a maximum entropy policy is developed using the soft Q Learning [25] algorithm. In the following subsections, we go into more detail into agent modeling, belief space rewards, and policy learning.

Opponent modeling

We use a binary variable $\lambda \in \{0,1\}$ to denote whether the opponent is a civilian or an adversary with hostile intents. Depending on λ, the opponent is expected to exhibit different behaviors,which is fully described by an opponent policy ($a^o|$). Since the action probability only depends on the current state, this model is constrained. However, we solely employ this model for policy learning, and we compare the learnt autonomous agent policy against a general history dependent opponent policy. The opponent's complete observation over the states is another implicit premise of this approach. The opponent might be modeled as a POMDP agent to disprove this presumption.

Experimental Setup

Framework

This research project is implemented by Python using the compute unified device architecture, where the device architecture is parallel computing platform and Application Programming Interface (API) model developed by Nvidia. For building the neural networks, we used the Tensor Flow and Keras libraries.

Datasets

To test and assess the performance of our technique, two types of datasets (text and image) were used. However, the model has the potential to resolve categorization issues using various types of input, such as text and images.

Text Datasets

Four distinct types of datasets, including 20Newsgroups, Reuters, WOS, and IMDB, are used to categorize the text.

A Web of Science (WOS) dataset [24] that contains the documents in three corpora as 5736, 11967, and 46985 for three distinct numbers of themes such as 11, 34, and 134 has been created from the collection and presentation of academic article abstracts.

The Reuters datasets consist of 10788 documents that are split into 90 classes, with the training dataset having 7769 documents and the testing dataset having 3119 documents. 50 000 reviews total in the IMDB dataset, with 25 000 of the most well-liked reviews for training purposes and the remaining 25 000 for testing.A 20NewsGroup dataset is defined as all 19997 papers with a maximum word count of 1000. 4000 samples are used for validation, and the remaining 15997 samples are employed in the testing procedure.

Image datasets

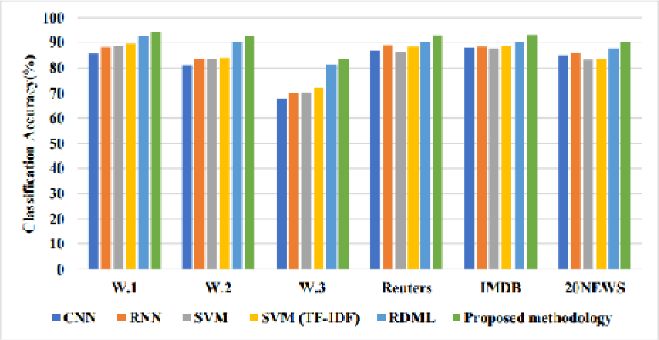
The two datasets, CIFAR and MNIST hand writing dataset, are regarded as traditional and ground truth datasets in this chapter since they are utilised for image classification.

The input feature space for MNIST is in the format 28281, and it contains handwritten numbers k 0, 1,..., 9. There are 60,000 and 10,000 data point instances in the training and test sets, respectively.

CIFAR: The dataset contains 60,000 photos in total, spread among 10 classes (bird, cars, horses, airplanes, deer, trucks, dogs, cats, and horses) and in the standard format of 32x32x3. Here, there are 50000 images in the training dataset, and the remaining images are used in the testing procedure.

Performance of the proposed algorithm in terms of document categorization

This section compares the proposed methodology to other document classification methods, including CNN [27], RNN [27], SVM [28], SVM(TF-IDF) [29], and RDML [20]. With regard to diverse document datasets, the classification accuracy of each algorithm is compared and evaluated. Table 1 displays the numerical analysis of the classification accuracy of various algorithms in relation to various datasets. The proposed methodology outperformed the current document categorization algorithms, according to an analysis of table 1.



	W.1	W.2	W.3	Reuter	IMDB	20NEWS
CNN	85.54	80.65	67.56	86.9	87.97	84.57
RNN	88.24	83.16	70.22	88.93	88.45	86
SVM	88.68	83.29	70.46	86.30	87.44	82.91
SVM (TF-IDF)	89.46	83.96	72.12	88.40	88.59	83.75
RDML	92.42	90.16	81.14	90.01	90.27	87.42
Proposed methodology	94.21	92.42	83.65	92.95	93.09	90.01

Figure 1: Comparative analysis of classification algorithms on various text datasets

Table 1: Numerical analysis on the document classification accuracy

Performance of the proposed algorithm in terms of image classification.

The effectiveness of the suggested methodology was evaluated in this section based on a comparison to several document classification algorithms, including Deep L2-SVM [30], Maxout Network [31], BinaryConnect [32], PCANet-1 [33], gcForest [34], and RDML [20]. Regarding classification error in relation to the MNIST and CIFAR-10 picture datasets, the comparison of classification algorithms is assessed. Table 2 displays the numerical analysis of the error rate of several classification algorithms in relation to various datasets. Table 1 analysis reveals that the suggested methodology outperformed the current document categorization methods.

	MNIST	**CIFAR-10**
DeepL.2-SVM [30]	0.87	11.9
Maxout Network	0.94	11.68

[31]		
BinaryConnect [32]	1.29	9.9
PCANet-1 [33]	0.62	21.33
gcForest [34]	0.74	21.33
RDML [20]	0.32	9.11
Proposed methodology	0.21	8.59

Table 2: Numerical analysis on error rate of the of various classification algorithm

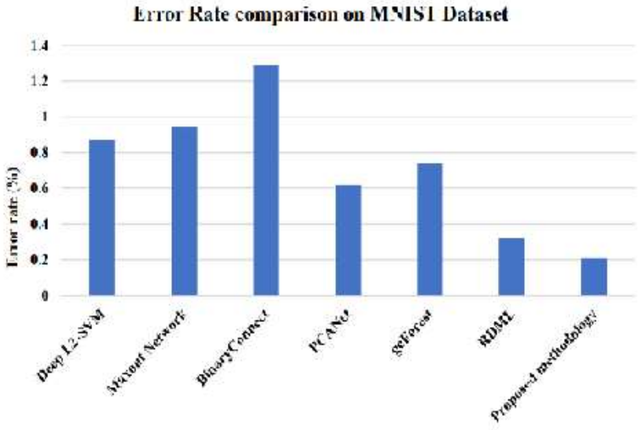


Figure 2: Comparative analysis of error rate on MNIST dataset.

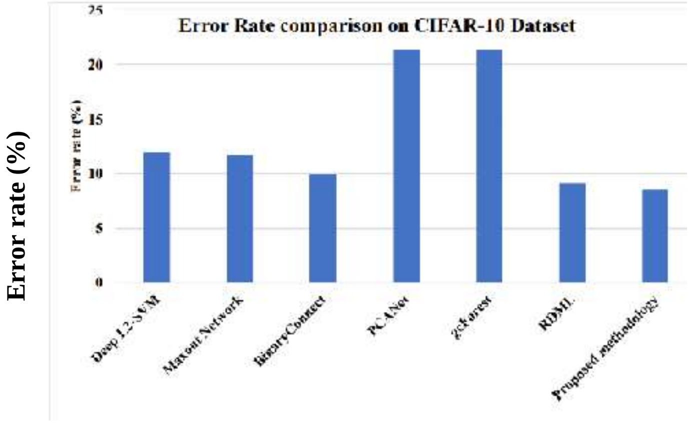


Figure 3: Comparative analysis of error rate on CIFAR-10 dataset

LITERATURE REVIVE

Kim, Edward, and Kathleen F [1] provided a multimodal DL strategy that facilitates the delivery of the intended message based on the usage of text data and pixel data. Two modalities taken from the information graphic are combined by the author. The raw pixel data is the first modality. Each image goes through a sequence of convolution and pooling layers after being scaled to a predetermined size. To get the dimension of the feature vector, a total of four convolution/pooling layers are created. Using the information graphic's text is the focus of the second mode. The author made use of the most recent Tesseract version to extract text from each of the training and testing images. In a binary representation known as a bag of words (BOW), the author encrypts the text for each image using a feature vector.whether the word is in the info graphic or not. The bag of words vector is then sent via a ReLU-activated, completely connected hidden layer.

Tian, Haiman, et al. [2] used the movies to recognise events. Initially, a lot of valuable data were initially retrieved from many modalities by using several DL techniques. Convolutional Neural

Networks (CNNs) were therefore employed for the extraction of visual and audio features and a word embedding model for the analysis of text. This framework described the two distinct data representations, such as frame-level video representation. The first step creates a joint representation for the visual and audio data by mapping the frame level classes to the video level classes using the frame-level classification results as input. The audio-visual results, video-level late fusion, and textual results were integrated in the second fusion stage to create the final video lessons. This framework also includes classes on natural disasters. The proposed approach demonstrated superior effectiveness when compared to single as well as traditional fusion procedures, which was supported by experimental findings. From single modality and fusion approaches, the accuracy was increased by over 7% to 16%.

Jiang, Yu-Gang, et al. [3] combined the helpful cues that were derived from several modalities, including motion patterns, auditory information, spatial information, and temporal dynamics information. CNN, which functioned on audio, motion, and appearance signals, was used to extract the corresponding features. The author then used a feature fusion network to create an integrated representation with the goal of capturing the relationships between the characteristics. Author employed two Long Short Term Memory networks with extracted appearance and motion features as inputs to videos in order to take advantage of the long-range temporal dynamics in videos.

The author then proposed using the contextual linkages between video meanings to improve the prediction results. The trials showed that the CNN strategy obtained 84.5% and 93.1% classification accuracy on the consumer data set and the UCF-101 dataset when compared with other techniques. The experiments were carried out on two benchmark datasets, such as the Columbia consumer videos and the UCF-101 dataset.

Kowsari, KAmran,et al. [4] discovered the optimal DL architecture for classification (RDML). These methods, which leverage a variety of inputs including video, text, symbolic data, and pictures, were used to increase the accuracy and resilience of the adata classification process. The trials were carried out on a variety of publicly accessible datasets, and the results demonstrated that the RMDL approach outperformed previous techniques on all datasets.

CONCLUSION

Given the expanding amount and size of datasets that require sophisticated classification,the classification issue is a crucial problem for machine learning. This study introduces a multi-model classification approach that combines maximum entropy RL, belief space planning, and generative adversary modelling to provide a stochastic belief space policy. By utilising numerous adversarial behaviours in the simulation framework, the prediction of the autonomous agent actions is minimised and obtained robustness when compared to unmodeled adversarial techniques. The suggested reinforcement-based deep learning algorithm may be used for multi-model categorization. Both text and image data can be classified using a single neural network technique. As a result of the RL learning the suitable belief space policy from the feature extracted data from the text and image data.The maximum entropy calculation is used to produce the belief space policy based on the feature of the input data. Experiments were conducted on various datasets, including IMDB, Reuters, WOS, CIFAR, 20NewsGroups, and MNIST, in order to compare the performance of the proposed method versus traditional techniques like a single DL or SVM model. These findings demonstrate that the suggested DL approaches can improve classification and that they give classification of datasets by majority vote flexibility. Even though the current approach produced classification accuracy with a satisfactory rate, future work will combine several efficient models to increase the accuracy across a variety of data kinds and applications.

REFERENCES

[1]. Liu, Yu, et al. "Learning visual and textual representations for multimodal matching and classification." Pattern Recognition 84 (2018): 51-67.

[2]. Abdur R, Kashif J, Haroon AB, Mehreen S (2015) Relative discrimination criterion – Anovel feature ranking method for text data. Expert Syst Appl 42(7):3670–3681

[3]. Coates A, Lee H, Ng AY (2011) An analysis of single layer networks in unsupervisedfeature learning AISTATS

[4]. Zareapoor M, Shamsolmoali P (2018) Boosting prediction performance on imbalanced dataset. Int J Inf Commun Technol 13(2):186–195

[5]. Gao L, Song J, Liu X, Shao J, Liu J, Shao J (2017) Learning in high-dimensional multimedia data: the state of the art. Multimedia Systems 23(3):303–313

[6]. Zareapoor M, Yang J (2017) A novel strategy for mining highly imbalanced data in creditcard transactions.Intell Autom Soft Comput.

[7]. Zhicheng Z, Rui X, Fei S (2018) Complex event detection via attention-based video representation and classification. Multimed Tools Appl 77(3):3209–3227

[8]. Zhu X, Jin Z, Ji R (2017) Learning high-dimensional multimedia data. Multimedia Systems 23(3):281–283

[9]. Roweis ST, Saul LK (2000) Nonlinear dimensionality reduction by locally linearembedding. Science 290: 2323–2326

[10]. Jingkuan S, Yi Y, Zi H, Heng TS, Jiebo L (2013) Effective multiple feature hashing forlarge-scale nearduplicate video retrieval. IEEE Trans Multimedia 15(8):1997–2008

[11]. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. Nature 521(7553), 436–444 (2015)

[12]. Bianco S, Cusano C, Napoletano P, Schettini R (2017) Improving CNN-Based Texture Classification by Color Balancing. J Imaging 3:33

[13]. Kim KW, Hong HG, Nam GPP, Ark KR (2017) A Study of Deep CNN-BasedClassification of Open and Closed Eyes Using a Visible Light Camera Sensor. Sensors 17:1534

[14]. Shamsolmoali, Pourya, et al. "High-dimensional multimedia classification using deepCNN and extended residual units." Multimedia Tools and Applications (2020).