iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** I **Month of publication:** January 2023

DOI: https://doi.org/10.22214/ijraset.2023.48884

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue I Jan 2023- Available at www.ijraset.com

An Improvised Approach of Deep Learning Neural Networks in NLP Applications

Rishil Vaghasia
Department of Computer Engineering, Devang Patel Institute of Advance Technology and Research, CHARUSAT, Changa-388421, India

Abstract: ***In recent years, natural language processing (NLP) has drawn a lot of interest for its ability to computationally represent and analyze human language. Its uses have expanded to include machine translation, email spam detection, information extraction, summarization, medical diagnosis, and question answering, among other areas. The purpose of this research is to investigate how deep learning and neural networks are used to analyze the syntax of natural language. This research first investigates a feed-forward neural network-based classifier for a transfer-based dependent syntax analyzer. This study presents a long-term memory neural network-based dependent syntactic analysis paradigm. This model, which will serve as a feature extractor, is based on the feed-forward neural network model mentioned before. After the feature extractor is learned, we train a recursive neural network classifier that is optimized by sentences using a long short-term memory neural network as a classifier of the transfer action and the characteristics retrieved by the syntactic analyzer as its input. Syntactic analysis replaces the method of modeling independent analysis with one that models the analysis of the entire sentence as a whole. The experimental findings demonstrate that the model has improved its performance more than the benchmark techniques.***

I. INTRODUCTION

A language is a system of rules or a collection of symbols that are integrated and used to express ideas or disseminate information. Natural Language Processing (NLP) serves users who lack the time to learn new languages or become proficient in their current ones because not all users have a strong background in the machine-specific language. In actuality, NLP is a branch of linguistics and artificial intelligence whose goal is to enable computers to comprehend assertions and words spoken in human languages [1]. It was developed to make the user's job easier and to fulfil their desire to speak to a machine in ordinary language. It may be divided into two categories: "natural language generation and natural language understanding", which progresses the task of comprehending and producing the text [2].

The study of language encompasses phonology, which deals with sound, morphology, which deals with word formation, which deals with sentence structure, semantics, which deals with syntax, and pragmatics, which deals with comprehension. Since the author redefined the study of syntax, Noah Chomsky, one of the earliest linguists of the 12th century to develop syntactic ideas, held a special place in the field of theoretical linguistics [3]. The process of creating meaningful words, sentences, and paragraphs from an internal representation is known as natural language generation, or NLG.

In computer linguistics, the term "grammar" denotes the study of particular linguistic structures and norms, such as determining the rules for word order in sentences and categorizing words [4]. Those languages' linear laws could be articulated using techniques like part-of-speech tagging and language models. Syntactic parsing [5] has long been a prominent field in the realm of natural language processing research and has major research relevance and application value. It is one of the important techniques in numerous natural language application activities.

Researchers first used the term "neural network" to describe biological information processing systems in the 1940s [6]. Deep neural networks can be trained on a massive scale thanks to ongoing improvements in computer performance. As a result, the Deep Learning approach has significantly advanced the study of numerous machine learning domains. Deep learning uses massive amounts of information to learn complex structural representations. Such learning is accomplished by altering the network parameters via back- propagation and error- driven optimization methods between several layers of artificial neural networks [7].

II. LITERATURE REVIEW

There are various studies, developments and improvements made by several researchers and academia regarding NLP, Neural networks, and deep networks which are abstracted in the present section. The brief development if NLP is depicted in figure 1 in the form of walkthrough graph for better and brief comprehension.

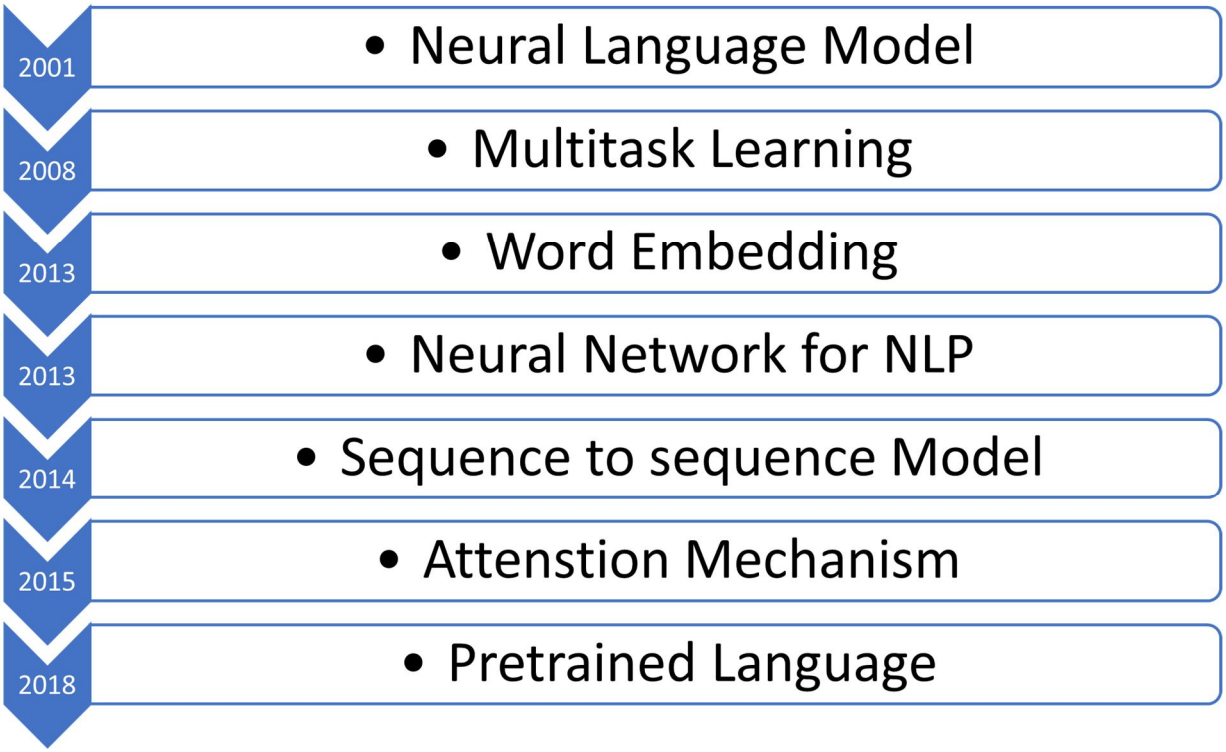


Figure 1: NLP Development

Neural language modelling, which estimates the likelihood of the next word (or token) given the previous n words, was introduced in the early 2000s. The idea of a forward neural network and lookup table that reflects the n preceding words in the sequence was put up by Bendigo et al. in their paper from 2012. The use of multitask learning in the field of NLP was suggested by the author [8], who employed two convolutional models with max pooling to perform named entity recognition and parts-of-speech tagging. In their word embedding approach [9] tackled dense vector representation of text. They also discuss difficulties with the conventional sparse bag-of-words form. The development of word embedding led to the introduction of neural networks in the NLP space, which take variable-length input for additional processing.

The convolution neural network was introduced to the challenge of phrase classification in natural language processing by the author [10]. In this study, features from sentences are extracted using a convolution neural network with two channels, and the features are then classified. The outcomes of the experiment demonstrate a considerable impact of the convolution neural network on the extraction of natural language features. Similarly, to this, [11] have critically examined and assessed the application of deep learning to Natural Language Processing (NLP), and they have summarized the models, methods, and tools that have been employed thus far. Additionally, [12] cover the use of deep neural networks in NLP.

Recurrent neural networks (RNNs), also known as neural networks that are recurrent because they carry out the same function for every data set, have also been used in natural language processing (NLP) and have been discovered to be ideal for sequential data including text, time series, financial data, speech, audio, and video among others by the author [13]. Annotation and the usage of transformers, along with attention mechanisms [14] that propose a network learns whatever to spend attention based on the current hidden state, have also significantly advanced NLP [15].

Investigating and classifying different emotional states from voice, gestures, facial expressions, and text is known as emotion detection. The author [16] examined Hinglish conversations to identify PoS usage patterns. Hinglish is an amalgam of English and Hindi. Their efforts were based on the POS tagging of the mixed script and language identification. They attempted to identify emotions in the mixed script by fusing human and machine learning. In order to assist users in prioritizing their messages based on the emotions associated with the messages, they have divided sentences into Six categories according to emotions and applied the TLBO technique [26].

Giving a statement of a semantic role is how Semantic Role Labeling (SRL) functions. For instance, in the PropBank [17] formalism, roles are assigned to terms in the phrase that acts as verb arguments. The specific attributes rely on the verb-frame, and if a phrase has more than one verb, it may also have more than one tag. The creation of a parse tree, determining which parse tree nodes represent the arguments of a particular verb, and lastly categorizing these nodes to compute the relevant SRL tags are all steps in modern SRL systems. A tree-like long- and short-term memory neural network was proposed by the author [18]. This linear model might lose considerable information since classic recurrent neural networks are typically employed to process linear sequences, particularly for data types with intrinsic structures like natural language. To get good outcomes in sentiment analysis, this model employs long and short-term memory neural networks in the analysis tree.

III. METHODOLOGY

Recursive neural networks are used to convert input sequences into output sequences, such as in problems involving sequence identification or sequence forecasting. But many of the real-world problems show how challenging it is to train recursive neural networks. Sequences in these problems frequently cover a wider time span. Recursive neural networks that wish to learn a long-distance memory find it more challenging because their gradient will eventually vanish. The author [19] suggested Long Short-Term Memory (LSTM), as a solution to this issue. This model includes the idea of a "door" that allows the network can decide when to "Forget" and add new "memory."

The long-term memory neural network, a variation of the recursive neural network, is intended to address the gradient disappearance of conventional recursive neural networks. The typical recursive neural network calculates a new hidden layer state h_1 while reading an input vector x_t from a vector sequence (X_1, X_2, …, X_n), but due to the issue of gradient disappearance, the typical recursive neural network cannot be used to describe long-distance dependence. In order to govern when to pick "memory" and when to choose "forget," long short-term memory neural networks invented "Memory Cell" and three "Control Gates."

Particularly, an "input, a forget, and an output gate is utilized by the LSTM neural network. The portion of the current input that could access the memory unit is one of them, and the forget gate regulates how much of the present memory should be erased. For instance, the neural network for long- and short-term memory is updated as follows at time t: Calculate the input gate i_t, forget gate, and candidate memory C_1 values at time t using the following formula, given input X_i [20-24]:

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i) \quad (1)$$

$$\tilde{C}_t = \tanh(W_c x_t + U_c h_{t-1} + b_c) \quad (2)$$

$$f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f) \quad (3)$$

The new memory cell's value and output value are provided simultaneously as follows:

$$C_t = i_t \odot \tilde{C}_t + f_t \odot C_{t-1} \quad (4)$$

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + V_o C_t + b_o) \quad (5)$$

$$h_t = o_t \odot \tanh(C_t) \quad (6)$$

Where $\odot$ is the component-wise product and σ is the component-wise logistic function.

The selection of the mask approach for training since batch training is necessary and the analysis sequence lengths of sentences of various lengths are not the same. However, some sentences in the batch have already been processed while they wait for the long sentences to appear since some of the sentences in the batch are too long. Sentences longer than 30 words should be eliminated from the training procedure in order to train the model more quickly. It was expected that since these sentences contain a total of 54 sentences, the effect of the final model won't be affected. The actual data utilized is displayed in Table 1 after a portion of the training data and verification data were removed.

Table 1: Statistical Data

Data Set	Total sentences	Projectable sentences
Training set	29834	29786
Development set	1710	1707
Test set	1710	1707

Accuracy, recall, and F_l value are frequently used as evaluation criteria in phrase structure analysis. In phrase structure analysis, the accuracy rate (P) is the ratio of the number of right phrases to the total number of phrases in the analysis result. In contrast, the recall rate (R) is the proportion of the correct words in the analysis result to all of the phrases in the test set. The F_1 value can be demonstrated according to equation (7).

$$F_1 = \frac{2PR}{P+R} \quad (7)$$

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue I Jan 2023- Available at www.ijraset.com

IV. RESULTS AND DISCUSSION

To be contrasted to the baseline approach, this issue is also contrasted with the Malt Parser and MST Parser, two more well-known dependency parsers. In the current work, the stackproj and nivreeager training options for Malt Parser have been used. These options correspond to the arc-standard analysis method and the arc-eager analysis algorithm, respectively. Also, give the outcomes for MST Parser in [125]. Table 2 displays the test results. The table shows that the long and short-term memory neural network-based dependency syntax analyzer has produced specific results in modelling the analysis sequence of phrases.

On the Penn Tree Bank development set, this model scored 90.50 % UAS accuracy and 90.00 % LAS accuracy, an improvement of roughly 0.60 % over the baseline method's greedy neural network dependency parser. On the test set, the model presented in the current study outperformed the baseline method's greedy neural network-dependent syntax analyzer by roughly 0.55 %, achieving UAS accuracy rates of 91.20 % and LAS accuracy rates of 90.40 %.

Table 2: WSJ Results

Analyzer	Development Set		Test Set	
	UAS	LAS	UAS	LAS
Present Model	90.5	90	91.2	90.4
Greedy feature extractor	89.6	89.2	90.3	90.2
Malt: eager	89.9	89.6	90.2	90.1
MST parser	90.1	89.8	90.4	90.1
Malt: standard	89.5	88.9	89.9	89.8
Baseline method	89.3	88.9	89.8	89.6

The experimental findings demonstrate that the dependency syntax analysis model based on the long and short-term memory neural network outperforms the greedy feed-forward neural network. In contrast to the greedy model, this model models the complete sentence using long and short-term memory neural networks, and it can classify analysis activities using historical analysis data and historical pattern data. This improves the performance of the dependent syntax analyzer. Table 3 displays the test findings on the Pennsylvania Tree Bank. This article uses a column search method during testing, and the appropriate beam size is 12.

Table 3: WSJ23 Results

Model	Accuracy	Recall rate	Fl value	Effective output	Word count does not match	Output structure error
Single follower	0.530	0.565	0.618	899	613	88
Double follow	0.725	0.734	0.795	1247	315	16

The statistics in the table demonstrate that the dual attention technique can significantly decrease the number of errors in the output results. The Fl value of the model reached 0.795 in the final output.

V. CONCLUSION

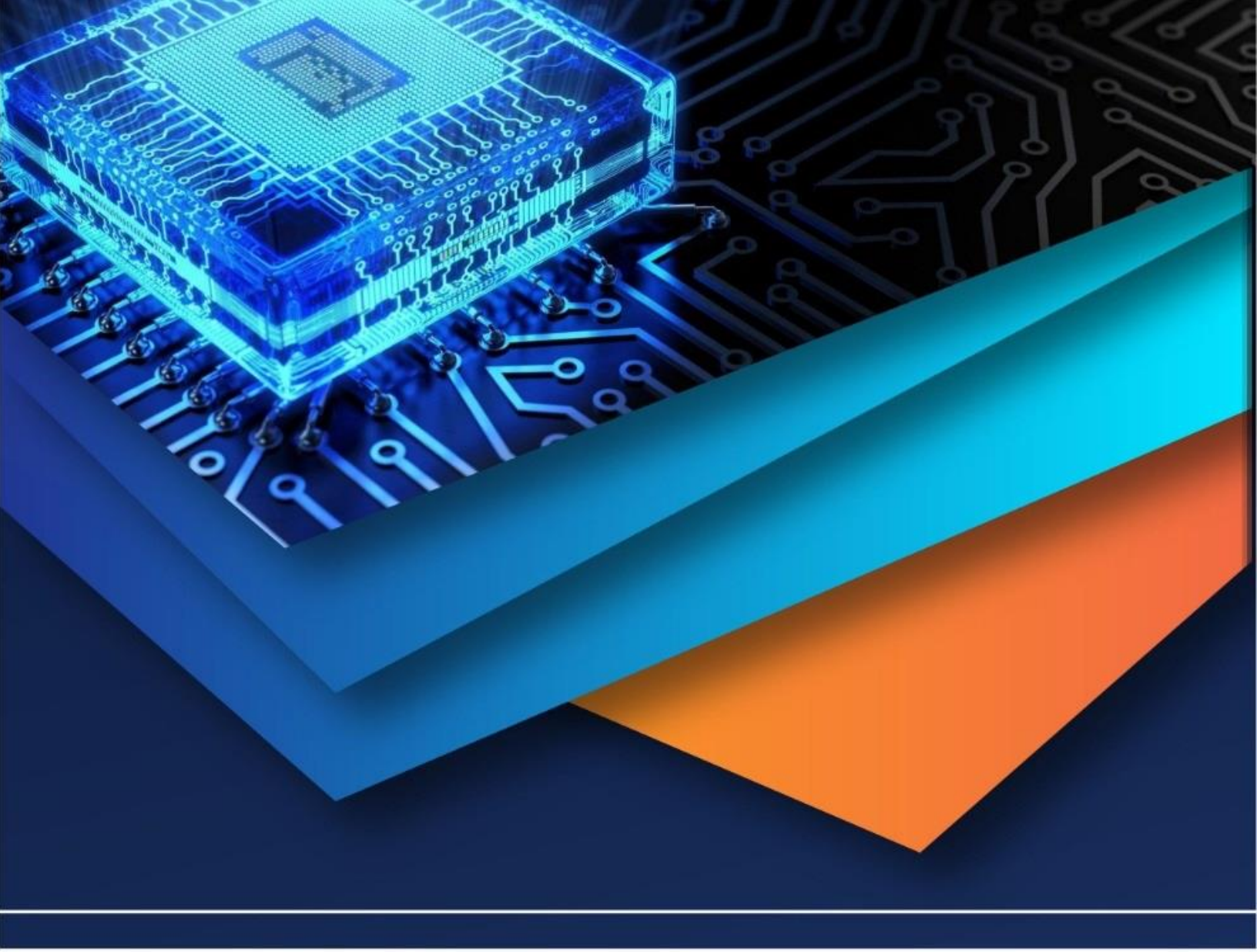
This research investigates a transfer learning-based neural network model of dependency syn- tactic analysis. The study of a feed-forward neural network as a classifier in the dependent syntax analyzer is demonstrated in the current paper, which then modifies its parameters after assessing the model to produce improved results. This study suggests a long-term memory neural network-based dependent syntactic analysis paradigm. The author comes to the conclusions listed below:

1) The baseline method's greedy neural network dependency parser was outperformed by the present model's 90.50% UAS accuracy and 90.00% LAS accuracy on the Penn Tree Bank development set, a difference of about 0.60%.
2) On the test set, the model reported in this paper performed around 0.55% better than the baseline method's greedy neural network-dependent syntax analyzer, achieving accuracy rates for UAS and LAS of 91.20% and 90.40%, respectively.

3) According to the experimental findings, the model's effectiveness increases by 0.1 to 0.15% after improvement.
4) According to the experimental findings, the model achieves an improvement of about 0.50% over the baseline technique.
5) The dual attention approach, it is found, can greatly reduce the number of errors in the output results.

REFERENCES

[1] Chi EC, Lyman MS, Sager N, Friedman C, Macleod C (1985) A database of computer-structured narrative: methods of computing complex relations. In proceedings of the annual symposium on computer application in medical care (p. 221). Am Med Inform Assoc
[2] Chomsky N (1965) Aspects of the theory of syntax. MIT Press, Cambridge, Massachusetts
[3] Choudhary N (2021) LDC-IL: the Indian repository ofresources for language technology. Lang Resources & Evaluation 55:855-867. https://doi.org/10. l 007/sl 0579-020-09523-3
[4] Chouikhi H, Chniter H, Jarray F (2021) Arabic sentiment analysis using BERT model. In international conference on computational collective intelligence (pp. 621-632). Springer, Cham
[5] Chung J, Gulcehre C, Cho K, Bengio Y, (2014) Empirical evaluation of gated recurrent neural networks on sequence modeling. arXiv preprint arXiv:1412.3555
[6] Cohen WW (1996) Learning rules that classify e-mail. In AAAI spring symposium on machine learning in information access (Vol. 18, p. 25)
[7] Collobert R, Weston J (2008) A unified architecture for natural language processing. In proceedings of the 25th international conference on machine learning (pp. 160-167)
[8] Dai Z, Yang Z, Yang Y, Carbonell J, Le QV, Salakhutdinov R, (2019) Transformer-xi: attentive language models beyond a fixed-length context. arXiv preprint arXiv: 1901.02860
[9] Davis E, Marcus G (2015) Commonsense reasoning and commonsense knowledge in artificial intelli- gence. Commun ACM 58(9):92-103
[10] Desai NP, Dabhi VK (2022) Resources and components for Gujarati NLP systems: a survey. Artif Intell Rev:1-19
[11] Devlin J, Chang MW, Lee K, Toutanova K, (2018) Bert: pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805
[12] Diab M, Hacioglu K, Jurafsky D (2004) Automatic tagging of Arabic text: From raw text to base phrase chunks. In Proceedings ofHLT-NAACL 2004: Short papers (pp. 149-152). Assoc Computat Linguist
[13] Doddington G (2002) Automatic evaluation of machine translation quality using n-gram co-occurrence statistics. In proceedings of the second international conference on human language technology research (pp. 138-145). Morgan Kaufmann publishers Inc
[14] Drucker H, Wu D, Vapnik VN (1999) Support vector machines for spam categorization. IEEE Trans Neural Netw 10(5):1048-1054
[15] Dunlavy DM, O'Leary DP, Conroy JM, Schlesinger JD (2007) QCS: A system for querying, clustering and summarizing documents. Inf Process Manag 43(6):1588-1605
[16] Müller, M.; Ewert, S. Chroma Toolbox: MATLAB implementations for extracting variants of chroma-based audio features. In Proceedings of the 12th International Conference on Music Information Retrieval (ISMIR), Miami, FL, USA, 24–28 October 2011.
[17] Fuentes, B.; Liutkus, A.; Badeau, R.; Richard, G. Probabilistic model for main melody extraction using constant-Q transform. In Proceedings of the 2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Kyoto, Japan, 25–30 March 2012; pp. 5357–5360.
[18] Durand, S.; Bello, J.P.; David, B.; Richard, G. Robust downbeat tracking using an ensemble of convolutional networks. IEEE/ACM Trans. Audio Speech Lang. Process. 2016, 25, 76–89.
[19] Di Giorgi, B.; Mauch, M.; Levy, M. Downbeat tracking with tempo-invariant convolutional neural networks. arXiv 2021, arXiv:2102.02282.
[20] Hung, Y.N.; Wang, J.C.; Song, X.; Lu, W.T.; Won, M. Modeling beats and downbeats with a time-frequency Transformer. In Proceedings of the ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Singapore, 23–27 May 2022; pp. 401–405.
[21] Desblancs, D.; Hennequin, R.; Lostanlen, V. Zero-Note Samba: Self-Supervised Beat Tracking; hal-03669865. 2022.
[22] Zonoozi, A.; Kim, J.j.; Li, X.L.; Cong, G. Periodic-CRN: A convolutional recurrent model for crowd density prediction with recurring periodic patterns. In Proceedings of the IJCAI, Stockholm, Sweden, 13–19 July 2018; pp. 3732–3738.
[23] Chen, C.; Li, K.; Teo, S.G.; Zou, X.; Wang, K.; Wang, J.; Zeng, Z. Gated residual recurrent graph neural networks for traffic prediction. In Proceedings of the AAAI conference on artificial intelligence, Honolulu, HI, USA, 27 January–1 February 2019; Volume 33, pp. 485–492.
[24] He, Z.; Chow, C.Y.; Zhang, J.D. STCNN: A spatio-temporal convolutional neural network for long-term traffic prediction. In Proceedings of the 2019 20th IEEE International Conference on Mobile Data Management (MDM), Hong Kong, 10–13 June 2019; pp. 226–233.
[25] Karim, M.E.; Maswood, M.M.S.; Das, S.; Alharbi, A.G. BHyPreC: A novel Bi-LSTM based hybrid recurrent neural network model to predict the CPU workload of cloud virtual machine. IEEE Access 2021, 9, 131476–131495.
[26] Wu, H.; Ma, Y.; Xiang, Z.; Yang, C.; He, K. A spatial–temporal graph neural network framework for automated software bug triaging. Knowl.-Based Syst. 2022, 241, 108308.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY