International Journal For Recent Developments in Science & Technology
2581-4575
Crossref
A Peer Reviewed Research Journal

GRAPH CONVOLUTIONAL NETWORK BASED ON MULTI HEAD POOLING FOR SHORT TEXT CLASSIFICATION

Mr. M. Uday Kumar, Nandhini Mamidi

Associate professor Department of Computer science and Engineering.

M.Tech, Department of Computer science and Engineering (21JJ1D5806), nandhumamidi08@gmail.com

ABSTRACT: Short texts, few features, and not having enough training data, among other things, are still the main problems that make it hard for standard text classification methods to work well. To solve these issues, we created the MP-GCN for classifying short texts with some help from a human. We also offer its three forms, which teach node representation in 1-order heterogeneous networks, 1-order isomorphic graphs, and 1&2 order isomorphic graphs. It checks the structure of the text network without using training word embedding as the first node trait. The multi-head technique provides the pooling representation subspaces without trainable parameters, whereas a self-attention based graph pooling methodology identifies and assesses the most significant nodes. Without using pre-training embedding, the results of the experiments showed that MP-GCN did better than the best models on five standard datasets.

Keywords: *Graph convolutional network, artificial intelligence, text classification, natural language processing.*

1. INTRODUCTION

One of the earliest NLP challenges is text categorization. The purpose of NLP is to tag words, queries, paragraphs, and documents. Researchers have developed DL models that categorize text better than ML approaches in recent years. RNN, CNN, transformer, and capsule net models are examples. Few academics have investigated semisupervised graph convolutional networks (GCNs) for text recognition in recent years [1, 2]. The major reason is that it works in a lot of real-life situations. To begin, it works better with short texts by making more connections between word nodes. It can also be used with scenes that have few or unclear meanings and no background information [2]. Second, it works well in situations where there isn't a lot of named training

2581-4575
International Journal For Recent Developments in Science & Technology
Crossref
A Peer Reviewed Research Journal

data, which is something that makes traditional neural networks not work very well [3]. Thus, we must immediately explore semi-supervised GCNs for text categorization. Also, semi-supervised GCNs are difficult to utilize. Due to varied situations, the pre-training word vector may not enhance text categorization. Instead, it may hinder graph creation. Second, it generates a network for the collection to add additional node connections (often using algorithms). Graph creation and feature extraction must consider memory and processing use.

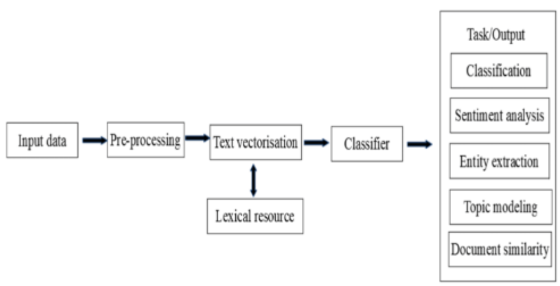

Fig.1: Example figure

Spectral and spatial graph convolutional networks (GCNs) are the major forms [4]. Spectral approaches have gained popularity in recent years. GCNs were initially discussed by Bruna et al. [4] in 2014. Convolution of a spectrum graph in space and time is challenging. To get neighborhood node characteristics, Deferrard et al. [5] used k-order Chebyshev polynomials as the convolution kernel. This made calculations faster. The Chebyshev network was made simpler by Kipf et al. [1], who also suggested a straightforward, efficient graph convolutional network (GCN) with one-order message routing. Spatial techniques also expanded quickly.

Hamilton et al. [6] proposed GraphSAGE. GraphSAGE trained in small batches with limited samples. This allowed GCNs become inductive learning networks and handle big data. According to Velikovic et al. [7], a graph attention network (GAT) would weight each connected node differently and group the most similar nodes.

2. LITERATURE REVIEW

Semi-Supervised Classification with Graph Convolutional Networks:

This work demonstrates scalable semi-supervised graph-structured data learning. It leverages sophisticated graph-based convolutional neural networks. Convolutional architecture based on localized first-order spectral graph estimation was used. Hidden layer representations of network local structure and node properties are learned by our method. It handles linearly more graph edges. We show our technique outperforms similar ones on citation networks and knowledge graph datasets.

Graph convolutional networks for text classification:

Natural language processing text categorization is well-known and significant. CNN (convolution on a grid) have been used to group items in many research. However, several research have examined GCN (convolution on non-grid, e.g., random graph) for the purpose. We recommend graph neural networks for text sorting in this study. Based on how frequently and how documents use words, a corpus' text graph is created. We train a Text GCN on the data. Our Text GCN starts with one-hot words and documents. Word and document embeddings are learned together using document class names. We found that a simple Text GCN without word embeddings or knowledge classifies text better than the best methods on various prominent datasets. Text GCN learns document and word prediction. Text GCN outperforms the best comparison algorithms when the proportion of training data is reduced. This shows that Text GCN can categorize text with less training data.

Heterogeneous graph attention networks for semi-supervised short text classification:

Short text labeling in news and tweet tags helps consumers discover information. We must explore semi-supervised short text classification immediately since labeled training material is scarce in many real-world scenarios. Most studies have only examined lengthy texts since short texts are difficult to evaluate due to a lack of labeled or dispersed material. We provide a heterogeneous graph neural network-based semi-supervised short text categorization algorithm. Spreading information around the graph maximizes tiny quantities of labeled and big amounts of unknown data. We start with a customizable HIN framework for short text modeling. This framework may add any supplementary information and illustrate how short texts relate to semantic sparsity. Use Heterogeneous Graph ATtention networks (HGAT) to integrate HIN with short text categorization. HGAT pays attention to nodes and types. Attention may learn how essential various nodes are to neighboring nodes and how significant different kinds of nodes are to the presently being examined node. After many tests, our model outperforms the top approaches on six common datasets.

Spectral networks and locally connected networks on graphs:

Signal classes are locally translationally invariant throughout their region, therefore Convolutional Neural Networks can recognize pictures and sounds well. This research investigates how CNNs might process data from more topics without a translation group. We recommend two designs. One uses a hierarchical domain grouping, while the other uses the graph Laplacian spectrum. We utilize research to demonstrate that convolutional layers can be learned with parameters independent of input size for low-dimensional graphs. This creates effective deep designs.

Convolutional neural networks on graphs with fast localized spectral filtering:

We intend to make convolutional neural networks (CNNs) function in graph-based higher-dimensional irregular domains including social networks, brain connectomes, and word embedding. In low-dimensional regular grids, CNNs store image, video, and audio. CNNs may be seen using spectral graph theory. This provides the mathematical basis and effective approaches to utilize numbers to create quick graph localized convolutional filters. The new approach is noteworthy since it works with any graph topology and has linear computation and constant learning complexity like CNNs. This novel DL system can learn local, stationary, and structural graph characteristics, according to MNIST and 20NEWS trials.

Inductive Representation Learning on Large Graphs:

From content prediction to protein functions, low-dimensional embeddings of nodes in huge networks have proved useful. Most current approaches need all graph nodes to be present during embedding training. Due to their transductive nature, earlier approaches don't function with undiscovered nodes. GraphSAGE is a wide, inductive architecture that quickly embeds new data utilizing node feature information like text attributes. We train a function that adds properties from its surroundings instead of embeddings for each node. Our approach beats strong baselines on three inductive

2581-4575
International Journal For Recent Developments in Science & Technology
Crossref
A Peer Reviewed Research Journal

node-classification tests: citations and Reddit posts sort unseen nodes in growing information graphs, and protein-protein interactions prove our algorithm works on entirely unseen networks.

3. IMPLEMENTATION

Also, semi-supervised GCNs are difficult to utilize. Due to varied situations, the pre-training word vector may not enhance text categorization. Instead, it may hinder graph creation. Second, it generates a network for the collection to add additional node connections (often using algorithms). Graph creation and feature extraction must consider memory and processing use.

Disadvantages:

1. Feature extraction need.
2. Memory consumption
3. Short text, sparse characteristics, and little training data

This study classifies short texts without instructions using an MP-GCN(Multi-head-Pooling-based Graph Convolutional Network). MP-GCN may pick critical nodes from diverse perspectives via multi-head sharing. Data classification is also excellent while utilizing minimal computational resources.

Advantages:

1. Without pre-training embedding, MP-GCN does better than the best models on the market.
2. The multi-head method gives you more than one representation region for pooling without adding any trainable factors.

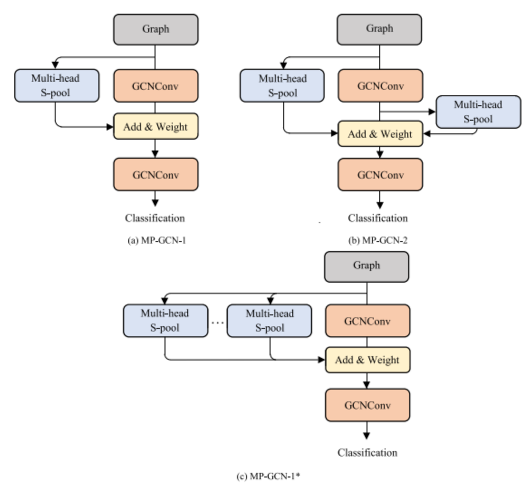

Fig.2: System architecture

Modules:

- Data exploration: Data will be imported using this module.
- Processing: This module reads and processes data.

International Journal For Recent Developments in Science & Technology
2581-4575
Crossref
A Peer Reviewed Research Journal

- Splitting data into train & test: This module will separate data into train and test.
- Model generation: Building the model - BERT + MP-GCN - GRU - LSTM - CNN - Bi-LSTM. Algorithms accuracy calculated.
- User signup & login: You may register and log in using this module.
- User input: This module aids forecasting.
- Prediction: Last forecast

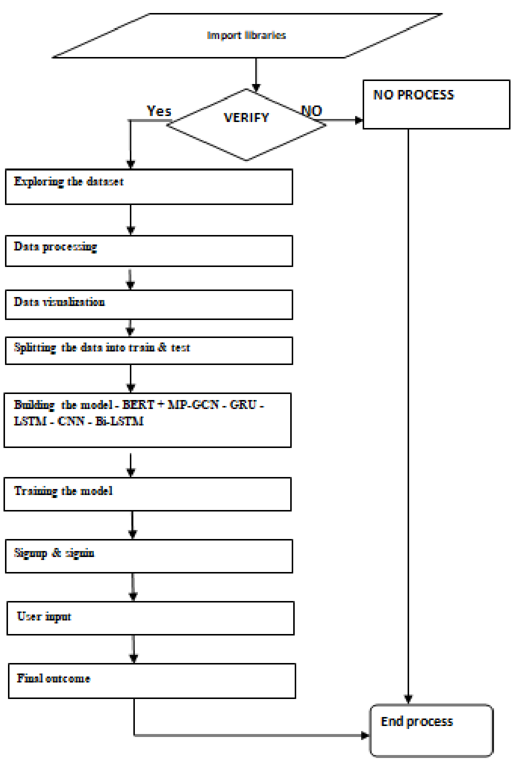

Fig.3: Dataflow diagram

TEXT CLASSIFICATION

Text categorization using ML organizes articles into predetermined categories. Organizing and using massive volumes of raw text information is crucial. Text classification research in language processing and text mining is extensive.

Semantic text classification methods use semantic word connections to figure out how similar two papers are in a broad sense. This method looks at the underlying connections between words and, by extension, between texts using the semantic approach (Altinel & Ganiz, 2018). There are several reasons why semantic text classification is better than traditional text classification.

• Determining the links between words, whether explicit or implicit.

• Discovering and applying hidden word-document relationships.

• Ability to illustrate current classes using keywords.

• Semantic text knowledge facilitates classification.

International Journal For Recent Developments in Science & Technology
2581-4575
Crossref
A Peer Reviewed Research Journal

• Support for polysemy and synonymy, unlike word semantic link-based text categorization methods.

Different semantic-based text categorization approaches incorporate word semantics. There are five main types of these techniques: methods that use subject knowledge (ontology-based), methods that use corpora, methods that use DL, methods that improve words or characters, and methods that enhance language (Altinel & Ganiz, 2018).

- Domain knowledge systems organize concepts in texts using ontologies and thesauri, which are language-dependent. Information bases include dictionaries, thesauri, and encyclopedias. Wiktionary, WordNet, and Wikipedia are popular information sites. WordNet is the most popular database.
- Corpus-based techniques utilize arithmetic computations to uncover hidden word relationships in learning texts (Zhang, Gentile, & Ciravegna, 2012). Latent semantics analysis (LSA) is a popular library-based approach (Deerwester et al., 1990).
- DL–based methods: The field of semantic text analysis has recently paid more attention to DL.
- • Enhanced word/character sequence methods: Word/character sequence–boosted systems use string-matching to extract words or characters from texts.
- Linguistic-enriched methods: These are linguistic-enriched methods, which use lexical and syntactic rules to pull out word phrases, entities, and concepts from a text and make a copy of it (Altinel & Ganiz, 2018).

Word classifiers are used in a lot of different types of tools that deal with a lot of text data. The way that email software sorts text tells it whether to send new messages to the inbox or the trash folder. On message boards, text analysis is used to choose which comments should be marked as wrong. Topic classifying a written work into one of a set of topics that have already been chosen is shown in these two examples. Many jobs that ask you to sort topics into groups use buzzwords in the text as the main way to do this.

4. ALGORITHMS

BERT + MP-GCN:

The BERT system is free and open source for ML that works with NLP. BERT is meant to help computers figure out what unclear language in text means by using the text around it to set the scene.

A Graph Convolutional Network, or GCN, is a way to learn on graph-structured data with some help from a teacher. It is built on a type of convolutional neural networks that work well and directly on graphs.

GRU:

Gated recurrent units (GRUs) are a way to control how recurrent neural networks work. They were first described by Kyunghyun Cho et al. in 2014. An LSTM with a forget gate is like a GRU. However, the GRU has fewer factors than an LSTM because it doesn't have an output gate. In some situations, the GRU is better than LSTM. It is a type of RNN. If you use datasets with longer patterns, LSTM is more accurate than GRU. GRU is faster and uses less memory.

LSTM:

LSTM is a DL architecture based on an artificial recurrent neural network. Time series and pattern issues can be solved using LSTMs. Recurrent neural networks like Long Short-Term Memory (LSTM) can estimate sequences by order. This is essential in problem-filled fields like voice recognition, machine translation, and more. LSTMs are hard in DL.

CNN:

CNNs are DL network designs. It processes pixel data to identify pictures and other tasks. CNNs are the greatest DL neural networks for discovering and identifying items. CNN learned spatial feature structures organically and adaptably via backpropagation. Fully linked layers, convolution layers, and pooling layers are used to achieve this.

Bi-LSTM:

Bidirectional LSTM layers understand how time steps in a time series or chain are connected across time in both directions. These associations may help the network learn from the complete time series at each time step. We can make any neural network recall sequences from the past or future. Bidirectional long-short-term memory.

5. EXPERIMENTAL RESULTS

2581-4575
International Journal For Recent Developments in Science & Technology
Crossref
A Peer Reviewed Research Journal

Fig.4: Home screen

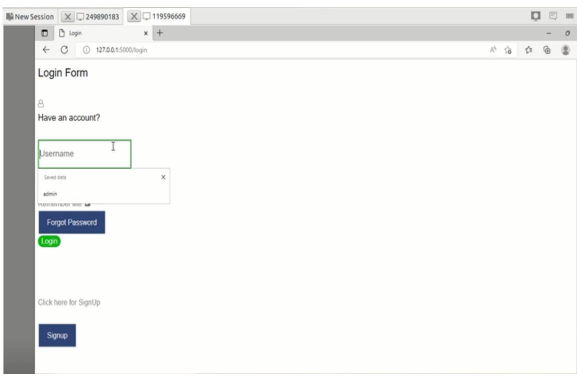

Fig.5: User login

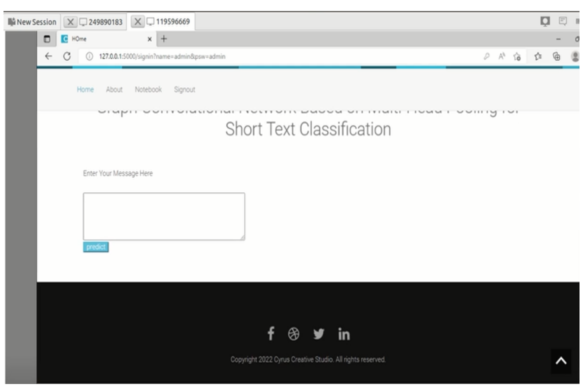

Fig.6: Main page

Fig.7: User input

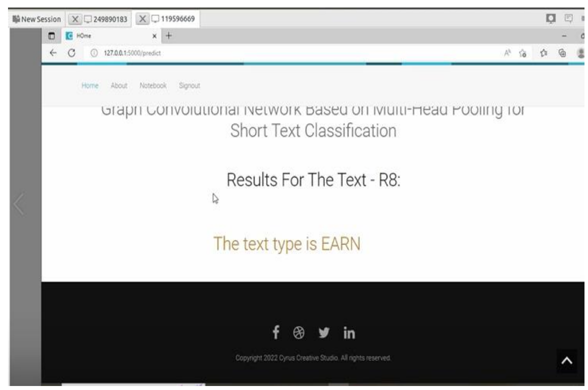

Fig.8: Prediction result

5. CONCLUSION

MP-GCN is suggested for sorting short texts in this study. Multi-head sharing improves crucial node representation learning in this network. We show three MP-GCN designs that teach node modeling in mixed graphs of one order, two orders, and one order. The tests show that MP-GCN does better than the best models on five common datasets that were not pre-trained.

6. FUTURE SCOPE

As time goes on, it will become easier to sort short texts into specific groups, like summaries of papers, news stories, very short texts, or texts from social networks. Short texts in specialized areas will have more difficult problems that need more specific models and better ways to handle text. The GNN design should also be more adaptable. It will be hard to make a GNN that is either more general or more specific in the future.

REFERENCES

[1] T. N. Kipf and M. Welling, ‘‘Semi-supervised classification with graph convolutional networks,’’ 2016, arXiv:1609.02907.

[2] L. Yao, C. Mao, and Y. Luo, ‘‘Graph convolutional networks for text classification,’’ in Proc. Conf. Assoc. Adv. Artif. Intell. (AAAI), Sep. 2019, pp. 7370–7377.

[3] H. Linmei, T. Yang, C. Shi, H. Ji, and X. Li, ‘‘Heterogeneous graph attention networks for semi-supervised short text classification,’’ in Proc. Conf. Empirical Methods Natural Lang. Process. 9th Int. Joint Conf. Natural Lang. Process. (EMNLP-IJCNLP), 2019, pp. 4821–4830.

[4] J. Bruna, W. Zaremba, A. Szlam, and Y. LeCun, ‘‘Spectral networks and locally connected networks on graphs,’’ 2013, arXiv:1312.6203.

[5] M. Defferrard, X. Bresson, and P. Vandergheynst, ‘‘Convolutional neural networks on graphs with fast localized spectral filtering,’’ in Proc. Conf. Neural Inf. Process. Syst. (NeurIPS), Jun. 2016, pp. 3844–3852.

[6] W. Hamilton, Z. T. Ying, and J. Leskovec, ‘‘Inductive representation learning on large graphs,’’ in Proc. Conf. Neural Inf. Process. Syst. (NeurIPS), Dec. 2017, pp. 1024–1034.

[7] P. Velickovic, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio, ‘‘Graph attention networks,’’ 2017, arXiv:1710.10903.

[8] J. Zhou, G. Cui, S. Hu, Z. Zhang, C. Yang, Z. Liu, L. Wang, C. Li, and M. Sun, ‘‘Graph neural networks: A review of methods and applications,’’ 2018, arXiv:1812.08434.

[9] L. Huang, D. Ma, S. Li, X. Zhang, and H. Wang, ‘‘Text level graph neural network for text classification,’’ in Proc. Conf. Empirical Methods Natural Lang. Process. 9th Int. Joint Conf. Natural Lang. Process. (EMNLPIJCNLP), 2019, pp. 3444–3450.

[10] C. Zhang, Q. Li, and D. Song, ‘‘Aspect-based sentiment classification with aspect-specific graph convolutional networks,’’ in Proc. Conf. Empirical Methods Natural Lang. Process. 9th Int. Joint Conf. Natural Lang. Process. (EMNLP-IJCNLP), Nov. 2019, pp. 4568–4578.

[11] X. Liu, X. You, X. Zhang, J. Wu, and P. Lv, ‘‘Tensor graph convolutional networks for text classification,’’ in Proc.

Conf. Artif. Intell. (AAAI), Apr. 2020, pp. 8409–8416.

[12] H. Peng, J. Li, Y. He, Y. Liu, and M. Bao, ‘‘Large-scale hierarchical text classification with recursively regularized deep graph-CNN,’’ in Proc. Conf. World Wide Web (WWW), Apr. 2018, pp. 1063–1072.

[13] Y. Kim, ‘‘Convolutional neural networks for sentence classification,’’ in Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP), Oct. 2014, pp. 1746–1751.

[14] T. Liu, X. Zhang, W. Zhou, and W. Jia, ‘‘Neural relation extraction via inner-sentence noise reduction and transfer learning,’’ in Proc. Conf. Empirical Methods Natural Lang. Process., Oct. 2018, pp. 2195–2204.

[15] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, ‘‘BERT: Pre-training of deep bidirectional transformers for language understanding,’’ 2018, arXiv:1810.04805.