

Application of Capsule Network in Text Classification Problem

Shengtao Yang

School of Cyber Security, Qilu University of Technology, Jinan, China

202201090049@stu.qlu.edu.cn

Abstract. The rise of the Internet and the big data era has led to a significant increase in text data, making text classification an essential task in natural language processing. Conventional deep learning models have advanced considerably in understanding text context dependencies; however, they still struggle with long-range dependencies and multi-scale features. In recent years, capsule networks (CapsNets) have demonstrated outstanding results in the area of image recognition due to their unique structure and dynamic routing mechanism, which has inspired researchers to explore their application in text classification tasks. This article examines the advancements in software and research related to CapsNets in the context of text classification challenges. Through its capsule units and dynamic routing algorithms, capsule networks can better express the part-whole hierarchical structure of text and effectively capture the complex relationships in text. Compared with traditional neural networks, capsule networks have shown advantages in dealing with both tasks involving text classification with either a single label or multiple labels. In addition, this paper also examines ways to optimize and enhance CapsNets for text classification, such as incorporating attention mechanisms to improve the model's capacity to recognize important information, as well as refining routing algorithms.

Keywords: Natural language processing, Capsule network, Text classification.

1. Introduction

As the Internet evolves quickly and people enter the age of big data, the amount of unstructured text data has exploded. From social media to online comments, from news reports to academic papers, massive amounts of text information not only carry human knowledge and emotions but also become an important part of the modern information society. How to effectively understand and utilize these text data has turned into a pressing issue that needs to be addressed. Text classification, as an important task in the field of Natural Language Processing (NLP), has gradually attracted people's attention.

In the past few years, advancements in deep learning technology have significantly advanced the field of natural language processing. Conventional statistical methods have slowly been replaced by deep learning models, such as recurrent neural networks (RNNs) and long short-term memory networks (LSTMs). These models achieve efficient extraction of semantic information by learning contextual dependencies in text sequences. However, as research deepens, people have found that existing models still have certain limitations in capturing complex relational structures in text, especially when dealing with long-distance dependencies and multi-scale features.

CapsNets are an emerging neural network architecture originally designed by Hinton et al. to solve challenges in computer vision. The core of the Capsule Networks is to use capsules instead of traditional neurons. Each capsule is a vector that can better express the part-whole hierarchical structure of an object and effectively transmit information through a dynamic routing. This design enables the CapsNets to acquire knowledge richer spatial hierarchical information while maintaining the invariance of the object's posture. Its unique design concept has aroused the interest of researchers, who began to explore the possibility of applying it to text processing. Compared with image data, text data exhibits a hierarchical structure, where words combine to create phrases, phrases come together to form sentences, and sentences build into paragraphs and even entire articles. As a result, the Capsule Network's ability to represent data hierarchically has demonstrated promising benefits in handling text data, particularly in tasks related to text classification. As a result, Zhao et al. [1] were the first to apply CapsNets to text classification in 2018, demonstrating the viability of Capsule Networks for this type of problem. They also suggested a process to improve efficiency by minimizing the impact of the dynamic routing process, thereby decreasing the noise in the capsule. Then Aly et al. [2] first applied CapsNets to hierarchical text classification (HMC) in 2019, using capsule networks to solve hierarchical multi-label text classification (MLTC) problems. The capsule networks were evaluated against basic shallow CapsNets to demonstrate their ability to address hierarchical MLTC challenges and to show that they surpass other neural network models. After verifying the feasibility of CapsNets in text classification problems, Numerous researchers started investigating ways to enhance and optimize CapsNets. Yu et al. [3] introduced the self-attention mechanism into CapsNets. It is used to solve the problem that the efficiency of the text classification process is reduced when modeling spatial information. The suggested new CapsNets combines the self-attention mechanism (SA), which excels at extracting global information, enhancing the model's ability to extract features by enabling capsules to hold more feature data. Jia et al. [4] introduced the multi-head attention mechanism into the CapsNets, and based on this, proposed a new CapsNets structure (AT-CapsNet). This has somewhat addressed the issue of limited capability caused by long-range dependencies and the challenge of selectively emphasizing key words in text classification. Zhun et al. [5] introduced a novel text classification model called CapsNet-GSR, which utilizes static routing grouping to extract local information from text through capsule grouping. The capsule compression and static routing techniques are employed to decrease the number of network parameters and enhance the quality of the capsules. It not only shows obvious advantages in model parameters and training time, but also makes progress in text classification accuracy. CapsNets have attracted many researchers to conduct in-depth research on them due to their unique structure and excellent effect in handling text classification problems. This paper primarily examines the present state of research on CapsNets in the field of text classification. Currently, the focus for optimizing and enhancing CapsNets involves incorporating attention mechanisms to boost the model's capacity to recognize important information, as well as refining routing algorithms.

2. Text Classification and Capsule Networks

2.1. Text Classification

Text classification technology, a key area of research in natural language processing (NLP), is essential for enhancing the efficient management and use of information. Text classification is the method of automatically sorting text data into established categories. This task is commonly applied in various areas, including sentiment analysis, news classification, spam filtering, topic modeling, etc. Traditional text classification usually assumes that each document belongs to only one category, that is, single-label classification. However, in practical applications, text often has multiple attributes and may involve multiple topics or emotions at the same time, which leads to the concept of MLTC.

MLTC allows each document to belong to multiple categories. For example, in sentiment analysis, a product review may express both love for the product and dissatisfaction with the price, so the review needs to be labeled "positive" and "negative". This complex label assignment requires the classification model to not only accurately identify each independent label, but also handle the relationship between

labels. As deep learning continues to advance, traditional neural networks, particularly convolutional neural networks (CNNs) and recurrent neural networks (RNNs), have become crucial in text classification tasks. These models are capable of effectively understanding the meaning of text by identifying advanced characteristics within it, thus enhancing the precision of classification. However, traditional neural networks also expose some limitations when dealing with MLTC tasks. Most models based on traditional neural networks do not fully consider the correlation between labels when designing, which may lead to inconsistencies when the model forecasts several labels. And because the number of labels in multi-label classification (MLC) tasks is usually large, the model is prone to dimensionality disaster, that is, as the quantity of labels grows, the model parameters expand sharply, training becomes difficult and easy to overfit.

2.2. Capsule Network

As artificial intelligence technology advances quickly, neural networks have emerged as an effective means of addressing a variety of complex issues. The effective use of CNNs, particularly in image recognition, has driven a significant transformation in the broader area of computer vision. While CNNs excel at tasks like image classification and object detection, they still have some inherent limitations, such as sensitivity to changes in object posture and lack of explicit modeling of spatial relationships between objects in images. To solve these problems, CapsNets were proposed in 2011 by Geoffrey Hinton, a pioneer in the field of deep learning, and his colleagues as an emerging neural network architecture. Aims to capture local and global information in images in a more efficient way. The "capsule" in the CapsNets is a combination of a group of neurons that together encode an instance of an object. Unlike traditional neurons that output a single value, the capsule outputs a vector that not only indicates the likelihood of the object's existence, but also includes the object's posture details, like its position, size, angle, etc. In this way, the CapsNets can more accurately represent the objects in the image and has stronger translation invariance and rotation invariance. Another important component in the CapsNets is the dynamic routing mechanism. This mechanism allows lower-level capsules to decide how to pass information based on how well their outputs match higher-level capsules. Specifically, each lower-level capsule weights its output, and the weight reflects its degree of association with the higher-level capsule. This process is like a voting mechanism, which ultimately ensures that information can flow upward along the most reasonable path. Its unique design ensures that the CapsNets demonstrate exceptional performance in tasks like image classification and object detection. Due to the distinctive design of CapsNets, people speculate that it can be utilized in the field of natural language processing. In natural language processing, text data also has a hierarchical structure, and the CapsNets can better capture the grammatical and semantic information in the text, this enhances the effectiveness of activities like text classification and sentiment analysis.

3. Application of capsule network in text classification

3.1. Effectiveness of Capsule Networks in Text Classification Problems

The use of CapsNets for text classification dates back to 2018, when Zhao et al. [1] were the first to apply CapsNets to issues in text classification. The authors proposed a model based on CapsNets and improved it for text classification tasks. The proposed model includes two variants, Capsule-A and Capsule-B, both of which are CapsNets models designed for text classification tasks. These models capture important features in text through capsule units and determine the correlation between different capsules through dynamic routing algorithms, thus enhancing the model's comprehension of text classification activities. Isolated categories are added to the original CapsNets to reduce interference. Experiments were carried out using six established benchmark datasets for text classification, as presented in Table 1. These are the six datasets used by the authors:

Table 1. Characteristics of the datasets.

Dataset	Train	Dev	Test	Classes	Classification Task
MR	8.6k	0.9k	1.1k	2	review classification
SST-2	8.6k	0.9k	1.8k	2	sentiment analysis
Subj	8.1k	0.9k	1.0k	2	opinion classification
TREC	5.4k	0.5k	0.5k	6	question categorization
CR	3.1k	0.3k	0.4k	2	review classification
AG 's news	108k	12.0k	7.6k	4	news categorization

The experimental findings indicate that CapsNets perform the best on four out of six benchmark datasets, demonstrating their effectiveness in text classification tasks. Specifically, Capsule-A and Capsule-B outperform traditional LSTM, Bi-LSTM, Tree-LSTM, CNN and other models on most datasets. Table 2 presents the outcomes of CapsNets compared to conventional methods across the six datasets.

Table 2. Capsule networks and baselines on six text classification benchmarks.

	MR	SST2	Subj	TREC	CR	AG 's
LSTM	75.9	80.6	89.3	86.8	78.4	86.1
BiLSTM	79.3	83.2	90.5	89.6	82.1	88.2
Tree-LSTM	80.7	85.7	91.3	91.8	83.2	90.1
LR-LSTM	81.5	87.5	89.9	—	82.5	—
CNN-rand	76.1	82.7	89.6	91.2	79.8	92.2
CNN-static	81.0	86.8	93.0	92.8	84.7	91.4
CNN-non-static	81.5	87.2	93.4	93.6	84.3	92.3
CL-CNN	—	—	88.4	85.7	—	92.3
VD-CNN	—	—	88.2	85.4	—	91.3
Capsule-A	81.3	86.4	93.3	91.8	83.8	92.1
Capsule-b	82.3	86.8	93.8	92.8	85.1	92.6

Through these experiments, the paper demonstrates the capability of CapsNets in text classification tasks, especially their ability to effectively handle multi-label classification problems and achieve competitive results on multiple benchmark datasets.

In 2020, Kim et al. [6] carried out an empirical study to investigate the applicability of CapsNets for text classification, and proved that CapsNets are indeed superior to traditional neural networks in some aspects. Figure 1 is the structure of the CapsNets used by Kim [6]:

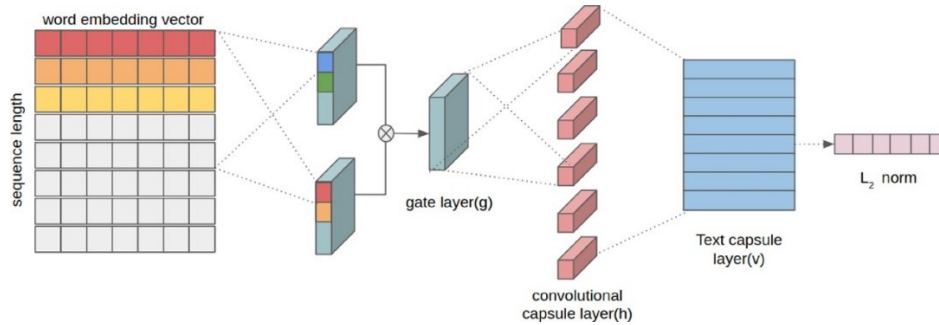


Figure 1. Capsule networks for text. Each document passes a gate layer, a convolutional capsule layer, and a text capsule layer [6]

Kim and colleagues [6] evaluated the model using seven distinct benchmark datasets, which include 20news, Reuters10, MR (2004), MR (2005), TREC-QA, MPQA, and IMDB. and proved that CapsNets can indeed be used for prediction based on seven different benchmark datasets. Classifying text using well-known benchmark datasets.

These two studies are early studies on using CapsNets to analyze text data. They have fully demonstrated that the CapsNets are effective for text classification tasks and offer benefits compared to CNNs.

3.2. *Application of capsule network in multi-label text classification problem*

Because deep neural network feature extraction and classification operations usually perform simple feature aggregation operations, such as using pooling to represent the input and outputting it as a vector. In MLTC, it has no way to capture the relationship between individual labels, which will ignore a lot of effective information. Chen et al. [7] introduced the Hyperbolic CapsNets (HYPERCAPS) to address this issue. The HYPERCAPS model is a hyperbolic version of Capsule Networks tailored specifically for MLC.

It integrates the benefits of CNNs and RNNs, with CNNs focusing on local context and RNNs addressing global context. These two components are linked in parallel to create local and global hyperbolic capsules. The hyperbolic dynamic routing (HDR) algorithm is presented, allowing for the aggregation of hyperbolic capsules in a way that is aware of labels, ensuring that the distinguishing information at the label level is maintained throughout the depth of the neural network. In order to improve the scalability of the model in large-scale MLC tasks, an adaptive routing layer is proposed to efficiently determine the candidate tags of documents. The adaptive routing mechanism selectively focuses on relevant tags, that demonstrates strong performance in MLC tasks. The CapsNets model shows strong performance in multi-label classification tasks.

Correspondingly, Wang et al. [8] proposed an algorithm called CBCN (Cascaded BiLSTM-Capsule Network for Noisy MLTC), which suggested an algorithm to enhance the CapsNets model for addressing the noisy MLTC issue. The goal of the algorithm is to enhance the model's resilience to inaccurate labels and to boost the performance of MLTC. To address the issue that the current MLTC algorithm overlooks noisy labels and fails to integrate both true and false labels, resulting in weak model robustness, an algorithm integrating cascaded BiLSTM (bidirectional long short-term memory network) and CapsNets was developed. The algorithm first trains a multi-label regression model with noise samples to distinguish the authenticity of random noise label combinations, thereby enhancing the algorithm's ability to identify pseudo labels. Then, the original MLC dataset is trained based on the model to achieve the purpose of MLC.

The CBCN algorithm was tested on two public English datasets, AAPD and RCV1-V2. It shows that the gating parameter θ can effectively constrain the feature learning between each layer of BiLSTM, the pure noise dataset LML has a significant effect on improving the CBCN algorithm, and the CapsNets Caps has a greater impact on the evaluation indicators of the CBCN algorithm. Overall, with the support of the feature extraction module, the algorithm can achieve better evaluation indicators.

The CBCN algorithm achieved an accuracy of 89.1% on the AAPD dataset and 91.1% on the RCV1-V2 dataset. indicating that the algorithm performs well in the noisy MLTC task and has strong robustness and generalization ability.

The fundamental components of CapsNets are the dynamic routing algorithm and the capsule architecture. which provide strong support for the CapsNets in dealing with MLTC problems.

3.3. *Comparison between capsule networks and traditional neural networks*

Traditional neural networks (CNNs) extract spatial features of input data through convolution kernels. For text data, this means that they can capture local patterns or n-gram features of words. This local perception allows CNNs to quickly find useful patterns in shorter sequences. However, they may encounter difficulties when processing long texts, especially when long-distance dependencies need to be captured. Because the size of the convolution kernel is usually fixed, it is not easy to capture

information across the entire document. In addition, CNNs lose a lot of data during the pooling process. In contrast, CapsNets can understand the inherent spatial connections between local features and the overall structure, allowing them to create viewpoint-invariant knowledge and automatically adapt to new perspectives. This relationship between the part and the whole is called semantic composition in natural language, and the analysis of the meaning of phrases and sentences is based on the principle of semantic composition. A capsule may consist of a collection of neurons whose activity vectors indicate the parameters for specific semantic features, allowing CapsNets to be utilized for text classification [9]. Moreover, CapsNets outperform CNNs in text classification and offer considerable benefits when transitioning from single-label text classification to multi-label text classification (MLTC).

The dynamic routing algorithm used in CapsNets can significantly minimize the impact of noise capsules, enhancing the model's performance by refining the compression function and determining the optimal number of iterations.

4. Optimization and improvement of capsule networks in text classification problems

4.1. Optimization and improvement of capsule network

The two most important aspects of CapsNets are capsules and dynamic routing. Due to the particularity of capsules and dynamic routing, the current optimization mainly focuses on optimizing and improving CapsNets models and optimizing and improving dynamic routing algorithms.

4.2. Introducing the Attention Mechanism into Capsule Networks

Wang et al. [10] introduced a model named Label-Text Bi-Attention CapsNets (LTBACN) designed for multi-label text classification tasks. This model addresses the issue of traditional MLTC methods overlooking label information and aims to enhance classification accuracy by integrating both label and text data. It employs graph convolutional networks (GCNs) to understand the relationships between labels and utilizes a bidirectional attention mechanism to learn the relevant features connecting labels and texts. The BERT model is utilized to transform text into high-dimensional vector representations, while GCN helps to incorporate label information, create a label graph structure, and facilitate information flow. The model features both a label-to-text attention mechanism and a text-to-label attention mechanism to extract text representations pertinent to labels and label representations most relevant to the text. LTBACN was tested on the AAPD and RCV1-v2 datasets and was compared with several existing methods. It demonstrated superior classification effectiveness compared to other approaches. Notably, LTBACN showed significant improvements across various metrics when compared to the most advanced methods. On the AAPD and RCV1-v2 datasets, it achieved the highest performance in terms of Micro-F1, Macro-F1, Hamming Loss, Precision@k, and nDCG@k.

Yu et al. [3] introduced a CapsNets model called SA-CapsNet that integrates a self-attention mechanism for text classification tasks. This model was designed to address the inefficiencies of traditional CNNs in handling text data and the challenges posed by spatially insensitive methods that struggle to effectively encode text structures. The capsule unit within the CapsNets captures features and spatial relationships in the text, while the self-attention mechanism enhances the model's ability to identify and represent crucial information. By using the self-attention mechanism to determine the relevance weights of various text segments and applying these weights to the feature representation, the model can automatically concentrate on the most important parts of the text. This approach aids the model in understanding deep semantic information, particularly when dealing with large text datasets. The SA-CapsNet model comprises a self-attention module that extracts features and calculates relationship weights among them, and a capsule module that constructs the text's feature vector representation. The integration of the self-attention mechanism with CapsNets boosts the classification accuracy of text data. The effectiveness of the SA-CapsNet model was tested on various text datasets, including IMDB, MPQA, and MR, achieving accuracies of 84.72%, 80.31%, and 75.38%, respectively, which were improvements of 1.08%, 0.39%, and 1.43% over other baseline algorithms.

This provides new ideas for improving the CapsNets model, which can be integrated with other models to better solve text classification problems.

4.3. Optimization of routing algorithm

Ren et al. [11] introduced a novel routing technique known as k-means routing, which assesses the connection strength between low-level and high-level capsules. This approach employs the k-means routing algorithm to enhance text classification accuracy while significantly decreasing the number of model parameters. It represents a new routing algorithm grounded in k-means clustering theory, aimed at thoroughly investigating the interactions between capsules. In contrast to the conventional dynamic routing algorithm, k-means routing utilizes cosine similarity to establish the relationship between low-level and high-level capsules. Additionally, it modifies the previous compositional coding mechanism by proposing a compositional weighted coding method to replace the traditional embedding layer. This method forms word embeddings by using the codeword vectors in all codebooks and giving them weights, thereby improving efficiency and reducing sensitivity to noise interference. Experiments were conducted on eight challenging text classification datasets to verify the effectiveness of the proposed method.

In the DBPedia dataset, the parameter count is nearly nine times lower when using the Region Emb. method compared to other approaches, while the performance gap is only 0.18%. The new routing algorithm has little impact on performance while significantly reducing model parameters. This shows that optimizing the routing algorithm is also the key to improving CapsNets in text classification tasks.

5. Conclusion

This paper provides an overview of the present use and advancements of CapsNets in text classification, highlighting their benefits compared to conventional neural networks. Although the use of CapsNets in text classification is still in the early stages, they have demonstrated certain advantages. For example, CapsNets can better capture contextual information and grammatical structure in text, which is essential for enhancing the precision of text classification. In addition, the dynamic routing feature of CapsNets enables the model to exchange information across various text samples, which improves the model's ability to generalize. CapsNets bring new ideas and technical means to text classification. Although there are still many problems to be solved, its theoretical innovation and potential in practice cannot be ignored. At present, the research on CapsNets for text classification tasks mainly focuses on improving the model of CapsNets, optimizing the routing mechanism of CapsNets, and introducing new technologies to help CapsNets better perform text classification tasks. Although the current experiments and studies show that CapsNets have certain advantages over traditional neural networks in text classification problems. But there are still many problems that have not been solved. In the future, researchers can start from the following aspects:

At present, CapsNets are applied in text classification tasks, but there are limited applications in multi-label tasks and the classification of long text data. At present, text classification tasks primarily involve straightforward classification and small datasets, while there are also numerous complex labeling tasks and larger datasets. The experimental data lacks sufficient balance. To objectively evaluate the effectiveness of CapsNets in text classification tasks, both current models and newly introduced models need to be tested on a wide range of complex text classification and lengthy text datasets.

The datasets and texts studied at this stage are mainly English datasets and English texts, and no experiments have been conducted on other languages. The existing research does not demonstrate the efficacy of CapsNets for English text classification; however, they remain effective in more intricate areas involving Chinese and Japanese. As a result, it is essential to utilize datasets in different languages to assess the performance of CapsNets.

In recent years, the transformer architecture has gained significant popularity in this area and has demonstrated superior performance compared to recurrent neural networks. In the future, the paper

hopes to combine the transformer with the CapsNets more deeply to study whether it has advantages in text classification problems.

References

- [1] Zhao, W., Ye, J. B., Yang, M., Lei, Z. Y. Zhang, S. F. and Zhao, Z. (2018). Investigating Capsule Networks with Dynamic Routing for Text Classification. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 3110–3119, Brussels, Belgium. Association for Computational Linguistics.
- [2] Aly, R., Remus, S. and Biemann, C. (2019). Hierarchical Multi-label Classification of Text with Capsule Networks. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop*, pages 323–330, Florence, Italy. Association for Computational Linguistics.
- [3] Yu X., Luo S-N., Wu Y., Cai Z., Kuan T-W. and Tseng S-P. (2024). Research on a Capsule Network Text Classification Method with a Self-Attention Mechanism. *Symmetry* 16(5):517. <https://doi.org/10.3390/sym16050517>
- [4] Jia, X, D. and W, L. (2020). Text classification model based on multi-head attention capsule network. *Journal of Tsinghua University (Science and Technology)*, 2020, 60(05): 415-421. DOI: 10.16511/j.cnki.qhdxxb.2020.26.006.
- [5] Zhu, H, J., Yu, L., Sheng Z, S., Chen, G, Q. and Wang, Z. (2021). Text classification model based on static routing group capsule network. *Journal of Sichuan University (Natural Science Edition)*, 2021, 58(06): 39-45. DOI: 10.19907/j.0490-6756.2021.062001.
- [6] Kim, J, Y., Jang, S., Park, E. and Choi, S. (2020). Text classification using capsules, *Neurocomputing*, Volume 376, Pages 214-221, ISSN 0925-2312, <https://doi.org/10.1016/j.neucom.2019.10.033>.
- [7] Chen, B, L., Huang, X., Lin, X. and Jing, L, P. (2020). Hyperbolic Capsule Networks for Multi-Label Classification. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 3115–3124, Online. Association for Computational Linguistics.
- [8] Wang, S, T., Zhang, S, J., Sun, X, F., Yang, J, J., Bi, T, T. and Zhang, Z, L. (2024). Multi-label text classification by integrating pseudo-label generation and data augmentation. *Journal of Harbin University of Science and Technology*, 1-11[2024-09-10]
- [9] He, W. and Zhu, M. (2021). A brief analysis of the current status and future of capsule neural network research, 2021, 57(03):33-43.
- [10] Wang, G., Du, Y, J., Jiang, Y, R., Liu, J., Li, X, Y., Chen, X, L., Gao, H, M., Xie, C, Z. and Lee, Y, L. (2024). Label-text bi-attention capsule networks model for multi-label text classification, *Neurocomputing*, Volume 588, 127671, ISSN 0925-2312, <https://doi.org/10.1016/j.neucom.2024.127671>.
- [11] Ren, H. and Lu, H. (2022). Compositional coding capsule network with k-means routing for text classification, *Pattern Recognition Letters*, Volume 160, Pages 1-8, ISSN 0167-8655, <https://doi.org/10.1016/j.patrec.2022.05.028>.