

Deep Learning-Based Exploration of Text Meaning

Junlin Huang ^{1, *}, Xuan Liu ²

¹ Science Department, McMaster University, Hamilton, Canada

² Software Department, Liaoning Technical University, Xingcheng City, China

* Corresponding Author Email: hjl17308185822@gmail.com

Abstract. In the field of Natural Language Processing (NLP), Recognizing of Textual Implications (RTE) belongs to one of the core tasks, aiming at recognizing semantic relations among texts. With the continuous improvement of related models, the application of Transformer has pushed the development of this field, but the existing models are still insufficient in numerical reasoning and multilingual scenarios. Therefore, this paper proposes and analyzes an improved hybrid transformer-QRNN model, exploring its potential to optimize the above shortcomings as well as its limitations in textual entailment tasks. Although this architecture outperforms traditional approaches on generalized datasets and significantly improves semantic analysis, it still faces shortcomings when dealing with scenarios involving numerical reasoning and multilingualism, and the structural complexity affects the efficiency when dealing with long texts. In the future, we need to optimize the numerical symbolization module, extend the multilingual support, and simplify the model structure. In addition, this hybrid architecture combining Transformer and QRNN proposed in this paper provides more efficient technical support for real-world scenarios such as legal contract contradiction detection and multilingual customer service systems.

Keywords: Recognizing of Textual Implications; Transformer-QRNN hybrid model; digital reasoning; multilingual scenarios; local dependency modeling.

1. Introduction

Natural Language Processing (NLP), as an important branch of artificial intelligence, involves the understanding and generation of natural language, aiming to enable machines to understand and generate human language. One of its core tasks is Recognizing Textual Entailment (RTE), which is to identify the semantic relationship (entailment, contradiction, neutrality) between texts through reasoning and logical judgment. Textual entailment is widely used in machine translation, manual customer service, sentiment analysis, and text summarization. It was first proposed in the PASCAL RTE Challenges, which aims to evaluate the application effect of natural language technology in practice [1-2].

Early research on textual entailment focused on text similarity calculation and simple logical reasoning methods. However, these methods have limitations when faced with complex semantic relationships and are unable to fully capture the nuances of context, such as the logical problems involving quantifiers and comparatives in the FraCaS dataset. With the release of high-quality manually annotated datasets such as Stanford Natural Language Inference (SNLI) and Multi-Genre Natural Language Inference (Multi-NLI) and the continuous breakthroughs in deep learning technology, the field of textual entailment research has gradually shifted to using neural network models to handle complex semantic reasoning tasks [3-5]. This shift has not only promoted the development of natural language understanding technology but also provided new research ideas and methods for solving deeper semantic problems. In their research, Bowman et al. first introduced Recurrent Neural Network (RNN) into the textual entailment task. Although this method can effectively model sentence semantics, it is prone to gradient vanishing or exploding problems due to the complexity of the model [1]. To solve this problem, researchers proposed the Long Short-Term Memory (LSTM) architecture, which significantly enhances the ability to handle long-distance dependencies. Although these models have made some progress in general scenarios, they face certain

challenges in digital reasoning (such as numerical calculation), professional fields (such as legal contract contradiction detection), and low-resource languages (such as Persian).

This paper will propose an improved Transformer architecture. It enhances the local language modeling capability and combines the knowledge injection strategy to improve its performance in digital processing and cross-language reasoning. The purpose of this study is to explore the collaborative optimization path of local dependency and global attention mechanism at the theoretical level. At the practical level, it provides adaptation support for the Recognizing Textual Entailment (RTE) task in professional fields and low-resource languages. Through the above research, it is expected to provide some theoretical support and practical guidance for the development of the field of textual entailment.

2. The Current Status of Research on Textual Implication

2.1. Traditional Methods

In the field of text entailment Recognizing, traditional methods mainly rely on manual features and rules. For example, there are vocabulary-based methods, rule-based methods, graph matching methods, and semantic methods. Although traditional methods have achieved certain results in the task of identifying text entailment, they still have many limitations. Like the limitations of vocabulary-based methods, Rahimi & ShamsFard et al. pointed out that these methods only work at the string input level and are not universal. And it cannot effectively capture the deep semantic information in the sentence. Similarly, other traditional methods face similar defects as vocabulary methods, especially when dealing with semantic depth, syntactic complexity and contextual dependence [6]. Later, with the introduction of Recurrent Neural Network (RNN) and the introduction of the Long Short-Term Memory (LSTM) architecture, researchers also proposed a method that combines Convolutional Neural Networks (CNN) with Long Short-Term Memory (LSTM). These methods have achieved significant improvements in the Recognition of text entailment.

2.2. Transformer Architecture

The Transformer architecture is a deep learning model structure for Natural Language Processing (NLP)[7]. It mainly solves the problem of low efficiency of traditional Recurrent Neural Networks (RNN) and Long Short-Term Memory (LSTM) when processing long sequences, and achieves significant performance optimization. The Transformer architecture consists of two main parts: encoder and decoder, both of which use the Multi-Head Attention layer[2]. The self-attention mechanism is the core of the Transformer. When processing sequences, it can calculate the relationship between words and enable the model to adjust the representation of each word according to contextual information. This mechanism helps the model effectively process long-distance dependencies and overcomes the shortcomings of step-by-step calculation in the traditional Recurrent Neural Network (RNN) model[8].

2.3. Transformer and QRNN

The Transformer architecture achieves significant performance improvements in text entailment tasks through the self-attention mechanism, but its quadratic computational complexity leads to excessive resource consumption when processing long texts. Although this architecture can model global context, it still has problems processing local language features. In response to the dual needs of text entailment tasks in terms of computational efficiency and feature capture, researchers developed the QRNN model, which combines the loop structure with parallel computing characteristics to improve computational efficiency while maintaining sequence processing capabilities.

3. Methods

This paper aims to analyze and explore whether the improved Transformer model can optimize the Recognizing ability of textual implications. The specific methods are as follows:

3.1. Model Architecture Design

This paper introduces an improved Transformer architecture, integrating QRNN into the traditional Transformer structure. QRNN processes the local dependency features of text sequences through convolution operations, improving local modeling capabilities while ensuring computational efficiency. The entire model is divided into encoding and decoding layers, each of which combines QRNN and multi-head self-attention mechanisms to better understand the semantic structure of the input text [9-11].

3.2. Implementation of QRNN

The core of Quasi-Recurrent Neural Network (QRNN) is to capture local features by dividing the input sequence into multiple sub-blocks and using masked convolution operations. In the specific implementation process, the model first performs word vectorization preprocessing on the text sequence. Then, the convolution layer is used to extract features from each sub-block to form a semantic representation of the local short sequence. These local representations are further combined with global context information in the deep network to build a more complete semantic expression system. Previous studies such as the Recurrent Neural Network (RNN) introduced by Bowman's team [1] and the Bidirectional Long Short-Term Memory (Bi-LSTM) model proposed by researchers such as Graves [4], although they effectively solve the problem of long-term dependency modeling, are still insufficient in capturing complex semantic relationships. To address this shortcoming, Quasi-Recurrent Neural Network (QRNN) has improved its performance by strengthening the modeling ability of local dependencies. Studies have shown that this model improves the perception of local information of text sequences by constructing local windows through the QRNN model [3].

3.3. Integration of Attention Mechanism

In the design of the encoding and decoding architecture, the multi-head self-attention framework is integrated to model long-range associations across sequences. This design enables the model to dynamically assign information weights to different semantic positions, thereby enhancing the ability to parse contextual associations. Specifically, the system focuses on the differential feature dimensions of the input sequence through parallel calculation of attention heads to form a complementary semantic representation space. It is worth noting that the Densely Interactive Inference Network (DIIN) model developed by Gong et al. [5] achieves fine-grained semantic matching through hierarchical attention stacking, while the Embeddings from Language Models (ELMo) dynamic word vector proposed by the Peters team verifies the key role of context-sensitive representation in semantic ambiguity resolution[6]. The current study constructs a dual-channel feature interaction system by combining the local window modeling advantages of QRNN with the global association characteristics of the attention mechanism, effectively breaking through the performance bottleneck of traditional methods in processing deep text logic reasoning tasks [12-14].

3.4. Training and Optimization

The training phase needs to be carried out based on a large-scale textual content dataset, with the cross entropy loss function as the core objective function for parameter optimization. The Adam optimizer is introduced in the parameter update process to adjust the gradient direction: Adam dynamically adjusts the learning rate of each parameter, combining the exponential moving average of the first-order and second-order moments to stabilize the parameter update, and balances the parameter update amplitudes of different dimensions through the adaptive learning rate mechanism, thereby improving the convergence efficiency and stability of the training process. Existing literature points out that the advantages of the Adam algorithm in the optimization of non-stationary objective

functions have been widely verified, and its adaptive characteristics can effectively alleviate the problem of gradient direction deviation[15,16].

In order to prevent the model from overfitting the training data, the early stopping strategy and dropout regularization technology are implemented simultaneously. The former dynamically terminates the training process by monitoring the performance of the validation set, and the latter randomly blocks neuron connections during the feedforward calculation process to enhance the robustness of the network. Studies have shown that dropout forces the model to learn more generalizable feature representations by randomly inactivating neuron units during forward propagation. This mechanism has been proven to have significant engineering practice value for deep neural network regularization. In addition, to address the zero value problem of probability estimation, Laplace smoothing ensures that the model can effectively handle unregistered word events by adding a small constant term to the probability distribution. Existing literature has fully demonstrated the role of such a combination of technologies in improving the generalization ability of the model. In particular, in natural language processing tasks, the coordinated use of adaptive optimizers and random regularization methods has been regarded as a standard practice[7].

3.5. Evaluation Metrics

In order to evaluate the effect of the model, a comprehensive evaluation is performed using indicators such as precision, recall and F1 score. [3] By comparing with the existing mainstream text entailment model, the performance of the improved model in different semantic contexts is verified to ensure its effectiveness and superiority. This paper hopes to achieve higher accuracy and better semantic understanding in the text entailment Recognizing task, and provide new perspectives and technical support for research in related fields.

4. Discussions

The improved Transformer model proposed in this paper is integrated with QRNN to enhance the modeling ability of local dependencies of text sequences, but there are still some limitations in practical applications. First, the numerical reasoning ability is lacking. The current model has limited effect on processing the implied relationship of calculations and logical reasoning involved in the text. In particular, it is difficult to accurately discover the logical relationship between numerical values for texts containing time, quantity or mathematical comparisons. Secondly, multilingual generalization is limited. The model is mainly based on English data sets, and there is a lack of experience in other languages, whether in grammatical structure or semantic expression. For example, the language difference between Chinese and English may affect the reasoning ability of the model. In addition, computational efficiency also faces bottlenecks. QRNN and multi-head attention mechanisms have improved the semantic modeling ability to a certain extent but also increased the complexity of calculations. In long texts or some specific scenarios, the training and reasoning time are greatly extended [17-19].

Because of the limitations mentioned above, future research may optimize the algorithm design from the following aspects. First, enhance the model's ability to process numbers, which requires the design of special functional modules to convert numbers into a form that can be directly calculated by computers and reason through simple mathematical rules. Second, the model's support for different languages can be expanded. Today, the model is mainly trained based on English data, while the structure and expression of other languages in the world, such as Chinese and Persian, are quite different. Perhaps other models that already support multiple languages can be used as a blueprint, and then some minor adjustments can be made to ensure that the model can adapt to the logical characteristics of different languages. At the same time, cases implied in texts containing multiple languages can be collected to help the model find the rules between different languages and reduce misjudgments caused by language differences. Finally, for optimizing the operating efficiency of the model, the current method improves accuracy but makes the model complex, resulting in reduced

efficiency. Therefore, the model structure can be simplified to avoid excessive processing of some details.

Through these improvements, future models will be better at handling numerical reasoning, supporting multilingual scenarios, and more efficient in actual use.

5. Conclusion

This paper studies the text implication Recognizing technology based on deep learning and proposes a hybrid architecture of QRNN and Transformer. By integrating the local feature extraction capability of QRNN and the global attention mechanism of Transformer, and introducing the knowledge enhancement strategy, it aims to optimize the performance of the model in digital reasoning, professional fields, and low-resource language scenarios. The study uses the cross entropy loss function and adaptive optimizer for training, combines the early stopping strategy and regularization technology to control overfitting, and evaluates the model performance through indicators such as accuracy and recall.

Existing models have obvious defects in processing textual implications tasks involving numerical calculations and multilingual scenarios. The improved Transformer-QRNN architecture improves the granularity of semantic analysis through local window modeling, but the efficiency of long text processing is reduced due to the parallel computing requirements of multi-head attention and QRNN. Although text performs better than traditional methods on general datasets, it still faces challenges in professional fields and low-resource languages.

Future research can be further optimized from three aspects: first, design a text digital symbolization module and combine it with a rule engine to enhance numerical reasoning capabilities; second, make the model support multilingual scenarios and use cross-language alignment technology to optimize generalization; third, simplify the model structure and focus the model on key points. The significance of this study is to provide adaptation solutions for practical scenarios such as legal contract contradiction detection and multilingual customer service systems, and promote the better integration of text implication technology into people's lives.

Authors Contribution

All the authors contributed equally and their names were listed in alphabetical order.

References

- [1] S.R. Bowman, L. Vilnis, O. Vinyals, A.M. Dai, R. Jozefowicz, S. Bengio, Generating sentences from a continuous space, arXiv.org, 2016. <https://arxiv.org/abs/1511.06349>
- [2] A.G. Gillioz, J. Casas, E.R. Mugellini, O.A. Khaled, Overview of the Transformer-based models for NLP tasks, IEEE Xplore, 2020. <https://ieeexplore.ieee.org/abstract/document/9222960>
- [3] X. Zhu, Z. Yan, W. Wang, S. Hu, X. Wang, C. Liu, QRNN-Transformer: Recognizing textual entailment, 2024 IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), 2024, pp. 1–7. <https://doi.org/10.1109/AVSS61716.2024.10672593>
- [4] A. Graves, Navdeep, A.R. Mohamed, Hybrid speech recognition with deep bidirectional LSTM, 2013 IEEE Workshop on Automatic Speech Recognition and Understanding, 2013, pp. 273–278. <https://doi.org/10.1109/ASRU.2013.6707742>
- [5] Y. Gong, H. Luo, J. Zhang, Natural language inference over interaction space, arXiv preprint, 2018. <https://arxiv.org/abs/1709.04348>
- [6] M.E. Peters, M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee, L. Zettlemoyer, Deep contextualized word representations, arXiv preprint, 2018. <https://arxiv.org/abs/1802.05365>
- [7] A.K. Shahade, Multi-lingual opinion mining for social media discourses: An approach using deep learning-based hybrid fine-tuned Smith algorithm with Adam optimizer, Int. J. Inf. Manag. Data Insights 3(1) (2023) 100029. <https://www.sciencedirect.com/science/article/pii/S2667096823000290>

- [8] N. Lan, E. Chemla, R. Katzir, Bridging the empirical-theoretical gap in neural network formal language learning using minimum description length, arXiv, 2024. <https://arxiv.org/abs/2402.10013>
- [9] Research Gate, Unveiling the environmental implications of automatic text generation and the role of detection systems, n.d. https://www.researchgate.net/publication/373174176_Unveiling_the_Environmental_Implications_of_Automatic_Text_Generation_and_the_Role_of_Detection_Systems
- [10] S. Raza, A. Omar, J. Alabi, J. Alistair, J.A. Name, T. Bolukbasi, Y. Cai, A. Caliskan, J.P. Chiu, H. Dawkins, S. Dev, L. Ding, T. Eftimov, ... Y. Liu, Nbias: A natural language processing framework for bias identification in text, *Expert Syst. Appl.*, 2023. <https://www.sciencedirect.com/science/article/abs/pii/S0957417423020444?via%3Dihub>
- [11] Y. Zha, Y. Yang, R. Li, Z. Hu, Text alignment is an efficient unified model for massive NLP tasks, arXiv, 2023. <https://arxiv.org/abs/2307.02729>
- [12] H. Ma, N. Afzal, H. Yang, G. Yu, S.A. Oliveira, R. Rajendran, Y.W. Chen, C.H. Lin, H. Wang, M. Xu, J. Liu, Z. Xu, S. Qu, Heart failure prognosis prediction: Let's start with the MDL-HFP model, *Inf. Syst.*, 2024. <https://www.sciencedirect.com/science/article/abs/pii/S0306437924000668>
- [13] A. Paramasivam, A. Answering, A survey on textual entailment-based question answering, *J. King Saud Univ. - Comput. Inf. Sci.*, 2021. <https://www.sciencedirect.com/science/article/pii/S1319157821003311>
- [14] I.M.S. Putra, Recognizing textual entailment: A review of resources, approaches, applications, and challenges, *ICT Express* 9(3) (2023) 343–354. <https://www.sciencedirect.com/science/article/pii/S2405959523001145>
- [15] V.S. Silva, A. Freitas, S. Handschuh, XTE: Explainable text entailment, arXiv, 2020. <https://arxiv.org/abs/2009.12431>
- [16] A. Poliak, A survey on recognizing textual entailment as an NLP evaluation, arXiv, 2020. <https://arxiv.org/abs/2010.03060>
- [17] M. Shajalal, Md. Atabuzzaman, M.B. Baby, Md.R. Karim, A. Boden, Textual entailment recognition with semantic features from empirical text representation, SpringerLink, n.d. https://link.springer.com/chapter/10.1007/978-3-031-33231-9_12
- [18] I. Dagan, D. Roth, M. Sammons, *Recognizing textual entailment: Models and applications*, Springer, 2013.
- [19] Research Square, A knowledge-based approach for recognizing textual entailments with a focus on causality and contradiction, 2024. <https://www.researchsquare.com/article/rs-3826973/v1>