Research on TextCNN-based Evaluation of Rationality of Narrative Text Structure

Jincheng Wang
Information Engineering College, Capital Normal University, Beijing, China
wangjincheng@cnu.edu.cn

Jie Liu*
Information Engineering College, Capital Normal University, Beijing, China
liujxxxy@126.com

ABSTRACT

Automatic Essay Scoring refers to the use of computers to score composition by some technologies. This process does not require human intervention. Rational text structure analysis is an important part of automatic essay scoring. However, the study of text structure is still in its infancy, ignoring its importance to the evaluation. Existing research lacks a corpus for the evaluation of text structure. The recognition of text components mostly uses artificial experience for feature selection, and evaluation model is established based on them. To figure out these problems, this paper refers to the curriculum standards, works with experts to build text structure standard and labeling method, and formulate corresponding labeling specifications. Finally build a corpus of a certain scale. TextCNN are used to build a model for the text structure. Model treats each article as a whole for training, and realizes the use of deep learning algorithms to make the model automatically evaluate. The results in test set show that in the constructed narrative composition corpus for grades 5-9, the accuracy of the model can reach 72.4%.

CCS CONCEPTS

• **Computing methodologies**; • **Artificial intelligence**; • **Natural language processing**;

KEYWORDS

Automatic essay scoring, Rationality of text structure, TextCNN model, Corpus

ACM Reference Format:
Jincheng Wang and Jie Liu*. 2021. Research on TextCNN-based Evaluation of Rationality of Narrative Text Structure. In *The 5th International Conference on Computer Science and Application Engineering (CSAE 2021), October 19–21, 2021, Sanya, China.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3487075.3487160

1 INTRODUCTION

Writing is an important way to reflect the comprehensive ability of students. The current composition grading is mainly done manually which requires teachers and takes a long time. Artificial composition scoring is subjective and easily interfered by various factors. Automatic essay scoring refers to a system that integrates linguistics, natural language processing and other technologies to evaluates the content, structure and other aspects based on artificially dimensions or models. The research on English composition is becoming more perfect. But on Chinese composition is not sufficient, ignoring the importance of the text structure to the scoring.

Text structure is the unity of internal connection and external form between the contents of article. It reflects students' writing ability to a large extent. For narratives, the structure is particularly important, which is composed of three elements. The central idea is the soul of article, material is the blood of article. And structure is the skeleton of article which is also a means of planning the layout and a method to reflect the central idea. Practicing text structure is a training for students to improve their logical thinking and upgrading language expression. Using deep learning to realize the text structure can assist teachers in evaluating character description more objectively, which has important practical significance.

In summary, the contributions of this article are as follows:

- In view of the current Chinese composition corpus's insufficient labeling of the text structure, we will formulate the evaluation criteria for text structure of narrative composition based on the curriculum standards and field experts, and collect the narrative composition of elementary and middle school. It lays the foundation for subsequent research on technology and terminology recognition;
- Different from the traditional machine learning that training is carried out by artificially setting features, this paper proposes a deep learning model for text structure. By constructing a text structure rationality model based on TextCNN, each article is trained as a whole to design and realize the text structure rationality of Chinese character composition, the accuracy rate in test set reaches 72.4%.

2 RELATED WORK

2.1 Automatic Essay Evaluation

In the early automatic essay system, E-rater extracts text features to identify text components, and completes the scoring by calculating the number of different text components. At the beginning of this century, researchers put the focus on the research based on neural network. Fei Dong[1] et al. studied the use of recurrent neural networks to automatically learn text representation. Adriano Veloso[2] uses linear regression of MSNLP to add a multi-faceted feature analysis. Berggren S J[3] used a variety of machine learning architectures to represent various input, and applied multi-task learning to the joint prediction of composition scoring. Among domestic scholars, Qi et al.[4] studied the automatic evaluation of subjective questions and studied the model of subjective. There is

CSAE 2021, October 19–21, 2021, Sanya, China

ACM ISBN 978-1-4503-8985-3/21/10…$15.00
https://doi.org/10.1145/3487075.3487160

little research on automatic evaluation of Chinese composition. Xu et al.[5] proposed a cohesion-driven unsupervised model; Hovy[6] were based on the recurrent neural network consistency model, proposed a method of textual consistency modeling based on neural network.

2.2 Text structure

Early text structure consistency modeling studies only proposed models for simple concepts, mainly focusing on linguistic aspects such as rhetorical structure and text representation theory. In recent years, Grid-based method proposed by Barzilay[7] uses a solid grid to represent the profile document. Louis[8] based on the Hidden Markov Model, using syntactic information, proposed a textual consistency model. In order to improve the performance of discourse consistency modeling, Li and Hovy[9] explored the coherence between sentences, and output the probability through recurrent neural network. However, simply returning low overall scores to students does not clearly indicate what causes scores to be far from enough. Researchers began to study specific dimensions of quality recently, such as coherence, composition review relevance, and so on.

3 INTRODUCTION TO TECHNIQUES

3.1 Convolutional Neural Network

Convolutional neural network is a neural network that uses convolution kernels for local feature calculation. By extracting the local feature, the input information can be classified into translation invariant according to the hierarchical structure. In one-dimensional convolutional neural networks, one-dimensional arrays can be used as data input.

3.2 Word2Vec

Word2Vec is a language model researched by Mikolov. The basic idea is to simplify the language model and predict more words with similar meaning by combining the contextual semantic information of words[10]. Word2vec is suitable for word vector training, with high efficiency and fast speed. Word2Vec uses words to predict words, including two methods: skip-gram and CBOW[11]. Skip-gram uses the center word to predict surrounding words, the word vector of the input feature word, and can output the word vector of the context word. CBOW uses surrounding words to predict the central word, input the word vector of a word corresponding to the context of a certain characteristic word, and can output the word vector.

3.3 TextCNN

Convolutional neural networks are mostly used in the field of image processing, and image pixels have a certain relationship with their neighbors. Therefore, the texture information of the area is usually used, and the convolution kernel is small. In the field of natural language, N-gram model features are generally used, which requires the features of several consecutive words, the convolution kernel is large. Kim[12] proposed the TextCNN model based on the task of text sentiment classification. It has two channels, namely static and non-static word embedding layer features. The input of most natural language processing tasks generally represents sentences in vectors. The TextCNN model uses a k-dimensional vector to represent a word in a sentence.

4 THE STANDARDS OF TEXT STRUCTURE AND CORPUS CONSTRUCTION

4.1 Establish the Scoring Standard

In the corpus labeling, the research team discussed the structure of text with domain experts, and jointly refined the standard of text structure. According to the composition scoring standards, the text structure is judged according to the structure, level, order and other aspects. Field experts combine the perspective of teacher evaluation to make standards around whether the structure of the article is complete, whether the context is echoed.

After discussions with field experts, the standards were established, and the preliminary labeling specifications were formulated. Based on this specification, 10 papers are randomly selected from the corpus for trial bidding, the consistency is calculated and the different parts are analyzed. Finally, the following evaluation indicators and auxiliary evaluation criteria for the rationality of narrative text structure are determined.

Text composition standard:

Excellent: The structure is rigorous and reasonable (beginning, main body, ending part must be comprehensive), the conception is ingenious and novel, and the text must be complete. The layout is thorough, coherent, clear, and detailed.

Good: The structure is complete, with textual elements. The organization is clearer. Most of the components are mainly written around the central idea, and a small part of it deviates from the theme or appears mixed in paragraphs.

Medium: The structure is basically complete, but not reasonable enough and lacks detailed arrangements. The textual elements are mainly written around the center.

Poor: The structure is chaotic, the text is incomplete, and the text is not complete.

Auxiliary judgment:

- The narrative text have a complete chapter structure, with a beginning, event description or character description, and ending part.
- The beginning should lead to the event/topic.
- The main section describes one thing clearly, with plot design, creating waves, and vivid twists; or write a few stories around the center, use subtitles, a clue, etc. to link together into the text. Show the character's spiritual qualities, etc.
- Pay attention to whether the main section contains character description, it can be classified according to the appearance, character language, character action, etc.
- Concentrate on lyric at the end and point out the theme.

4.2 Construction of the Corpus

4.2.1 Corpus data annotation processing. This paper collects Chinese narrative essays from the fifth to ninth grades to build the text structure corpus. Among them are 689 in grade 5, 507 in grade 6, 567 in grade 7, 625 in grade 8, and 636 in grade 9.

Table 1: Annotate Test Consistent Results

Annotator	Independent consistency/%	Consistency with final marking /%
A	81.9	93.2
B		86.5

Table 2: Annotate Consistent Results

	Value	Asymp. Std.Errora	Approach Tb	Approx Sig
Kappa	.771	.012	53.593	.000

The group members are invited to participate in the scoring according to the standard. In order to ensure the consistency of the labeling results of the corpus, experts are invited to lead the group members to conduct trial bidding. After being familiar with the labeling specifications, 50 papers were randomly selected from the corpus for pre-labeling, and labeling consistency was calculated. In the pre-annotated corpus, if the results of two members are inconsistent, the third one is invited to judge, and finally two of the three are judged as the result. Annotate to form the final corpus. Calculate the consistency between the two independently annotated corpora and the consistency with the final corpus. The consistency results are shown in table 1

4.2.2 Corpus quality assessment. The Kappa coefficient formula is shown in (1), which is used for consistency testing and can also be used to measure classification accuracy. The calculation of kappa coefficient is based on a confusion matrix. P_0 is the sum of the number of samples correctly classified in each category divided by the total number of samples. P_e is as in formula (2). Assuming that the number of real samples in each category is $a_1, a_2, \ldots, a_C$,the predicted number of samples in each category are $b_1, b_2, \ldots, b_C$, and the total number of samples is n.

$$K = \frac{P_0 - P_e}{1 - P_e} \tag{1}$$

$$P_e = \frac{a_1 \times b_1 + a_2 \times b_2 + \ldots + a_C \times b_C}{n \times n} \tag{2}$$

From Table 2 we can know that the text structure score kappa=0.771(>0.75) marked by the group members indicates that the marking has high consistency and can be used for model training. In order to verify the quality of the corpus, after the above corpus labeling process, domain experts are invited to randomly select 95 articles from the corpus for quality evaluation. This paper uses F1 value as the evaluation index. The specific method is to use the expert's result as the standard answer C1. The annotation result in the corpus is C2. After calculation, the accuracy F1 value in the corpus is 83.16%, indicating that the quality of the annotation in this corpus is guaranteed to a certain extent. Through the model, the full text is used as a whole to master more information, which has strong computability.

$$P = \frac{\mathrm{C_1 \text{ and } C_2 \text{ marked consistent results}}}{\mathrm{C_2 \text{ marked result}}} \tag{3}$$

$$R = \frac{\mathrm{C_1 \text{ and } C_2 \text{ marked consistent results}}}{\mathrm{C_1 \text{ marked result}}} \tag{4}$$

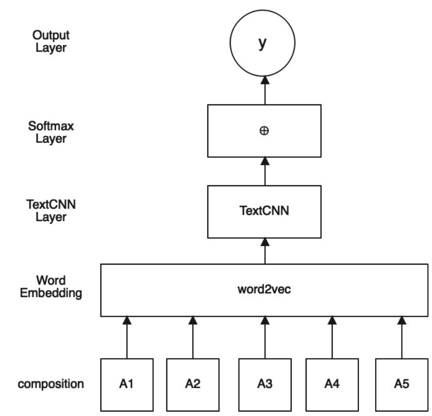

Figure 1: Rationality Model of Text Structure Based on TextCNN.

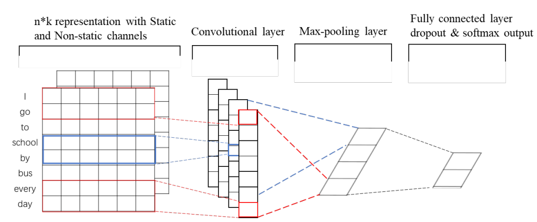

Figure 2: TextCNN Model.

$$F_1 = \frac{2 \times \mathrm{P} \times \mathrm{R}}{\mathrm{P} + \mathrm{R}} \tag{5}$$

5 MODELS

5.1 Text Structure Rationality Model

This paper randomly divides the data into 90.1% training set and 9.9% test set. In order to truly evaluate the generalization ability of the model, this article uses ten-fold cross-validation to determine the model with the best performance. The text structure rationality model based on TextCNN is shown in Figure 1, and the specific TextCNN model is shown in Figure 2

Firstly, check the integrity of each composition content in the corpus, and then perform text preprocessing. This article uses NLPIR

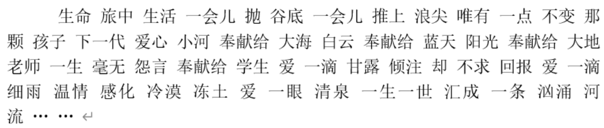
生命 旅中 生活 一会儿 抛 谷底 一会儿 推上 浪尖 唯有 一点 不变 那颗 孩子 下一代 爱心 小河 奉献给 大海 白云 奉献给 蓝天 阳光 奉献给 大地 老师 一生 毫无 怨言 奉献给 学生 爱 一滴 甘露 倾注 却 不求 回报 爱 一滴 细雨 温情 感化 冷漠 冻土 爱 一眼 清泉 一生一世 汇成 一条 汹涌 河流 …… ……

Figure 3: Text after Word Segmentation and Removing Stop Words.

Chinese word segmentation system to segment all texts in Chinese. In order to save space and improve the efficiency of the model, some words that appear frequently but not affect the actual meaning of the article are usually filtered out after the word segmentation operation is completed. This paper uses the stop word list issued by the Chinese Academy of Sciences as a reference to remove stop words in the text. Figure 3 shows the text after segmentation and removal of stop words.

Due to one-hot cannot express the relationship between words and the dimension is too large. This article uses word2vec for pre-training. The composition after word segmentation and stop words removal is expressed by word2vec vector and input to the text structure rationality model based on TextCNN.

The input layer converts a piece of text into the input required by the convolutional layer, which can be represented by $S = [x_1^d, x_2^d, \ldots, x_i^d, \ldots, x_n^d]$, where $n = length_S$,d are the dimensions of the embedded word vector, and x_i^d represents the word vector representation of word i in sentence S. Usually, $x_{i:i+j}$ is used to represent the word vector combination matrix of $x_i, x_{i+1}, \ldots, x_{i+j}$. Then use a convolution kernel $W \in \mathrm{R}^{hd}$ to extract the convolution operation of text feature c_i as shown in equation 6), where h is the size of the moving window for extracting new features.

$$c_i = f\,(W \cdot x_{i:i+h-1} + b) \tag{6}$$

$b \in R$ is a known bias, and f is a nonlinear activation function. The window text sequence of the input sentence can be used $\{x_{1:h}x_{2:h+1}, \ldots, x_{n-h+1:n}\}$ to extract the convolution feature sequence $c \in R^{n-h+1}$.

$$c = [c_1, c_2, \ldots, c_{n-h+1}] \tag{7}$$

The convolutional layer outputs the feature vector matrix c and then enters the pooling layer. Use f convolution kernels of different sizes $\{f_1, f_2, \ldots, f_F\}$ to perform the above sliding window text convolution operation to obtain multiple sets of text features.

The fully connected layer is cascaded after pooling. Use dropout to prevent overfitting, and use the activation function to classify the output. Finally combing to the softmax layer to output the final label probability value.

The pre-trained word vectors are sent to the TextCNN model for classification. This paper uses PyTorch framework for classification. The classification result and the accuracy of the test set are obtained. The complete process is shown in Figure 4

5.2 Result

This paper analyzes 3024 narrative Chinese essays in grades 5-9. Group members complete the scoring and labeling under the leadership of experts. The test set consists of 301 articles randomly

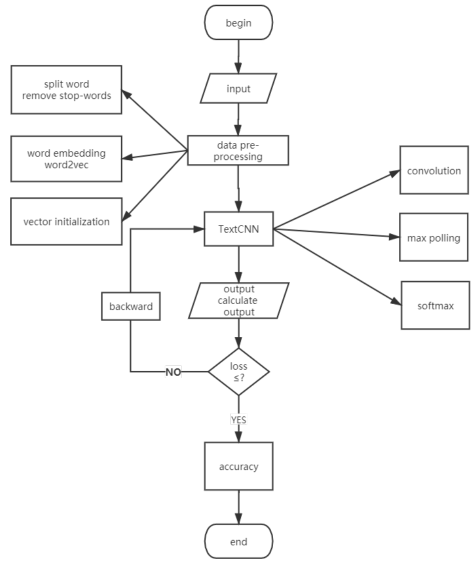

Figure 4: Flow Chart of Text Structure Rationality Model Based on TextCNN.

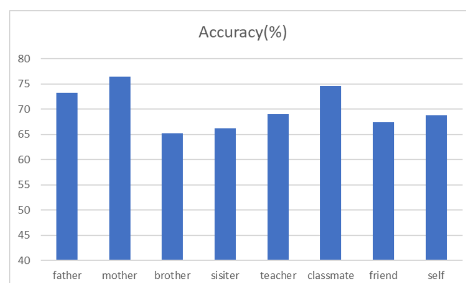

Figure 5: Accuracy Prediction Results of Different Character Classification Models.

selected from the constructed corpus. In order to get a better performance model, this paper uses ten-fold cross-validation to test the experiment, and uses the Accuracy index to evaluate the model performance.

As is shown in Table 3 and Figure 5, using the tags of characters in corpus to classify and sort the descriptions of different characters, the accuracy of the evaluation is between 65.2% and 76.4%, which has a large range of variation. That may because that although Chinese narrative compositions often describe characters and events as the main method, when there are more characters and events, and a large number of rhetorical techniques are included, the current model training cannot be more accurate for judgment. The accuracy of articles under a certain category can be improved by adjusting the model structure and parameters later.

We can learn from Table 4, the text is regarded as a whole and put into the TextCNN model for training. Use manual evaluation as the standard, the accuracy of text structure in test set obtained by classifying according to different grades is 70.9% to 73.2%, the data

Table 3: Accuracy Prediction Results of Different Character Classification Models

Classification	father	mother	brother	sister	teacher	classmate	friend	self
Accuracy(%)	73.2	76.4	65.2	66.1	69.1	74.6	67.4	68.7

Table 4: Accuracy Prediction Results of Different Grades

Grade	5	6	7	8	9
Accuracy(%)	0.727	0.709	0.732	0.719	0.718

Table 5: Accuracy Prediction Results of text Structure Rationality Model Based on TextCNN

	Deep learning based on TextCNN
Accuracy(%)	72.4

volatility is small. The relatively stable accuracy rates of different grades indicate that the model has universal applicability for the assessment of the text structure, the standards formulated for the text structure are relatively reasonable. The corpus has high consistency, which can enable the model to effectively learn the required features, thereby perform repetitive training. The two grades with lower accuracy may be caused by the lower number of articles in the test set. Thus, the number of texts in training set will also affect the accuracy of model recognition. When constructing the text structure corpus, we should pay attention to sufficient corpus for each grade, so as to reduce other influencing factors.

As can be seen from table 5, the accuracy of the text structure rationality model based on TextCNN in test set is 72.4%. The experimental results show that the deep learning algorithm, which takes the text as a whole, makes the model to automatically learn features feasible, and can effectively evaluate the structure of the composition.

6 CONCLUSION AND FUTURE WORK

The main work of the paper is divided into two parts. Firstly, according to the characteristics of the composition text structure, the evaluation criteria for the text structure of the Chinese narrative composition are formulated. Based on tagging specification, large-scale corpus tagging was carried out, and the 5-9 grade narrative Chinese composition was constructed with highly consistency. A total of 3024 articles were judged for the text structure part, which to some extent alleviated the current lack of Chinese composition corpus of the text structure part.

In addition, this paper adopts TextCNN model to treat each text as a whole and evaluate it by automatically learning features through the model. The final accuracy is close to the accuracy of machine learning that setting features manually. In future work, under the premise of high consistency, we will improve the quality of annotation by adding multi-dimensional text components, expand the size of the corpus, and further improve the corpus. We will improve the deep learning model according to the characteristics of Chinese composition to increase the accuracy.

ACKNOWLEDGMENTS

This work is supported by National Science and technology innovation 2030 major projects (2020AAA0109700), National Natural Science Foundation of China (62076167), and Beijing Municipal Education Commission-Beijing Natural Fund Joint Funding Project (KZ201910028039).

REFERENCES

[1] Dong F, Yue Zhang, Yang J (2017). Attention-based Recurrent Convolutional Neural Network for Automatic Essay Scoring[C]// Proceedings of the 21st Conference on Computational Natural Language Learning (CoNLL 2017).
[2] Amorim E, Veloso A (2017). A Multi-aspect Analysis of Automatic Essay Scoring for Brazilian Portuguese[C]// Proceedings of the ACL.
[3] Berggren S J, Rama T, Vrelid L (2019). Regression or classification? Automated Essay Scoring for Norwegian[C]// 14th Workshop on Innovative Use of NLP for Building Educational Applications.
[4] LIU Wei, QI Zi-sen (2016). Automated Assessment of Subjective Tests[J]. JOURNAL OF BEIJING UNIVERSITY OF POSTS AND TELECOM, 18(4): 108-116.
[5] XU Fan, ZHU Qiaoming, ZHOU Guodong, WANG Mingwen (2014). Cohesion-driven Discourse Coherence Modeling. 28(3): 11-21.
[6] Li J, Hovy E (2014). A Model of Coherence Based on Distributed Sentence Representation[J] //Proceedings of the EMNLP.
[7] Barzilay R. Lapata M (2008). Modeling local coherence: an entity-based approach[J]. Computation al Linguistics. 34(1):1-34.
[8] Louis A., Nenkova A (2012). A coherence model based on syntactic patterns[C]//Proceedings of the EMNLP-CNLL, 1157-1168.
[9] LiJ, Hovy E (2014). A Model of Coherence Based on Distributed Sentence Representation[J]//Proceedings of the EMNLP, 2039-2048.
[10] TANG Ming, ZHU Lei, ZOU Xianchun (2016). Document vector representation based on word2vec[J]. Computer Science, 43 (6): 214- 217.
[11] ZHOU (2015). Word2vec's working principle and application to explore[J]. Science &Technology Information Development and Economy, 25: 145-148.
[12] Kim Y (2014). Convolutional neural networks for sentence classification[C]. Proceedings of the EMNLP, 1746-1751.