ICIC Express Letters
Part B: Applications
Volume **8**, Number **9**, September **2017**

pp. 1253–1261

ANALYZING DISCOURSE HIERARCHICAL STRUCTURE BASED ON NAÏVE BAYES MODEL

Maosheng Zhong[1,2] and Lan Tao[2]

[1]School of Computer and Information Engineering
Jiangxi Normal University
No. 99, Ziyang Avenue, Nanchang 330022, P. R. China
zhongmaosheng@sina.com

[2]School of Information Engineering
East China Jiaotong University
No. 808, Shuanggang East Avenue, Nanchang 330013, P. R. China
taolan1992@sina.com

Received April 2017; accepted June 2017

Abstract. *Automatically analyzing discourse hierarchical structure will play an important role in discourse understanding, automatic summary and text inference. Analyzing the discourse hierarchical structure needs to understand the semantics of the discourse, but it is difficult to automatically understand the semantics of a discourse in itself. The task of analyzing the hierarchical structure of the discourse automatically and accurately is a challenging task. In this paper, our steps for analyzing discourse hierarchical structure are, beginning from the idea of discourse organization mode, and assuming that the same types of discourses have the same or similar organization mode, so marking their hierarchy structures for these discourses. Then, use these marked discourses as training corpus, and learn the mode of organizational structure from the corpus by applying machine learning method. Last, automatically analyze the hierarchical structure of other discourses with the same type by the paragraphs node merging algorithm based on Naïve Bayes model. The experimental results show that this approach has better performance than the discourse structure analysis method based on sequence alignment algorithm, and the average F value increased by nearly 30%, therefore, obtaining more effective analysis results.*
Keywords: Discourse hierarchy structure analysis, Discourse structure tree, Naïve Bayes model, Paragraph-topic mark

1. **Introduction.** Generally, a discourse (article) always describes a topic (main idea), and different paragraphs or a group of paragraphs in the discourse certainly describe the different aspects of the topic, each aspect also may be further parsed, so the whole contents of this discourse can be built a hierarchy structure tree by parsing, the root of which denotes a topic or main idea, intermediate nodes represent the sub-topic of discourse segment corresponding to the sub-tree, while the leaf nodes probably are the smallest granularity topic corresponding to a paragraph or a group of sentences. The task of discourse hierarchy structure analysis is to construct an analytical model, and look the orderly initial paragraphs of discourse as some leaf nodes, then based on the analytical model, bottom-up and recursively merge these leaf nodes into a Discourse Structure Tree (DST) based on the semantic relevance, which consists of a root node, a plurality of layers' intermediate nodes and leaf nodes that is corresponding to the initial paragraphs of this discourse. The significance of discourse hierarchy structure analysis is that the built DST is clearly able to show the hierarchy of the topic of the full discourse, easier to grasp the main idea, systematic, the harmony of components of the discourse on macroscopic view. For the task of text automatic summary, the result of text summarization based

on the discourse hierarchy structure analysis can overcome the sidedness result from the traditional method.

The literature on hierarchical structure analysis for discourse is relatively sparse. Salton et al. proposed the Relationship Map (RM) method for constructing the hierarchical topic structure of a discourse [1-3]. Obviously, RM generated from continuous paragraphs become an unordered structure, so the leaf nodes of the tree are unordered, namely, these paragraphs of the initial discourse are linearity and order, but the leaf nodes of the tree, which are corresponding to these paragraphs, are unordered. Yarri used the Hierarchical Agglomerative Clustering (HAC) method to build the hierarchical structure of a discourse [4]. However, the HAC method has to exploit the similarity measurement of adjacent fragments; in the case of the short fragment, the similarity measurement of adjacent fragments is inaccurate. Hsueh et al. described a supervised approach that trains separate classifiers for the topic and sub-topic segmentation on multiparty dialogue [5]. Jacob presented an unsupervised method for hierarchical topic segmentation by using the multi-scale lexical cohesion [6]. Moreover, Lin et al. introduced the latent semantic indexing to analyze the hierarchical structure of text [7], Zhang et al. used the similarity between adjacent paragraphs and recursively merging adjacent paragraphs to generate the level topic/subtopic structure of Web page document [8], and Zhong used the sequence alignment algorithm for analyzing the text hierarchical structure [9]. However, there may exist a problem in Lin's method, namely, when adding a level node for the tree, it is difficult to decide which text block need be ulteriorly divided. The similarity measurement of Zhang's method relies solely on the co-occurrence number of word. The method of [9] did not obtain an ideal performance.

There is one orthogonal but the related approach to obtaining nonlinear discourse structures from discourse. Rhetorical structure theory posits a hierarchical structure of discourse relations between spans of discourse [10]. This structure is richer than our hierarchical topic structure analysis, and the base level of analysis is typically more fine-grained – at the level of individual clauses. Our goal of discourse hierarchical structure analysis is a more macroscopic view and we do not expect to succeed at this level of fine-grained granularity.

In this paper, we described a new method based on Naïve Bayes model to analyze the hierarchical structure of discourse. The basic idea of our method is to assume that the same type (referring to the same topic or purpose of use) discourses have the same or similar organizational structure mode. If we can let the computer learn these modes from the training corpus, the computer also can analyze the hierarchical structure for the test corpus. The rest of this paper is structured as follows. In Section 2, we formally describe the conception of Discourse Structure Tree (DST). In Section 3, we present in detail our steps to analyze the discourse hierarchy structure based on Naïve Bayes model. Section 4 gives experimental corpus, evaluation method and the experimental results of our method, and Section 5 concludes with a discussion of our results and possible directions for future research.

2. **Discourse Hierarchical Structures.** The hierarchy structure of discourse can be represented by the $DST = (V, E)$ shown as Figure 1, where V is the set of nodes, E is a set of edges (branches set) to connect nodes. Let $V = V_1 \cup V_2$, where $V_1 = \{V_{1i} | 0 \leq i \leq m\}$ is the intermediate nodes set of DST and $V_2 = \{V_{2j} | 1 \leq j \leq n\}$ is the leaf nodes set of DST. If $V_{2j} \in V_2$ then V_{2j} denotes a sentence, a group of sentences or a paragraph.

The root node V_{10} of DST represents the topic of the current discourse, while the intermediate node V_{1i} $(1 \leq i \leq m)$ represents the different aspect of the topic (also known as subtopic). Because the different topic class has different granularity, perhaps there exist many layer intermediate nodes in the DST and each intermediate node has a corresponding topic which extracts from the discourse fragment indexed by the sub-tree

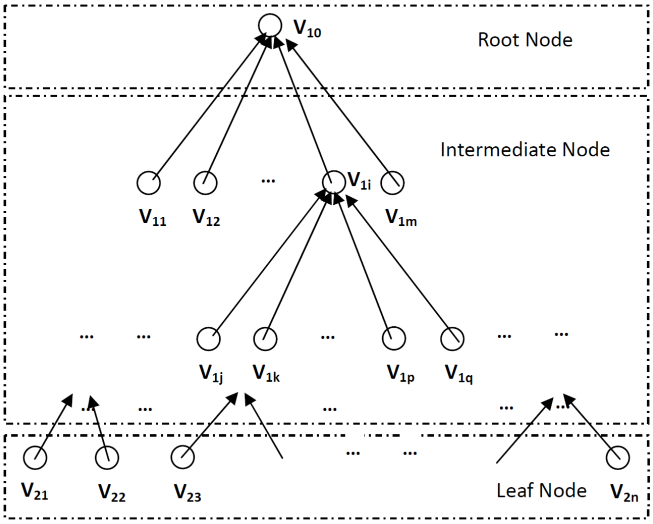

FIGURE 1. Schematic figure of discourse structure tree

corresponding to the intermediate node. The task of discourse structure analysis is to automatically build a DST shown as Figure 1 for the input discourse text by analyzing the relation between topics of the different fragment in this discourse.

3. Hierarchy Structure Analysis for Discourse Based on Naïve Bayes Model.

3.1. Main steps. Our method to analyze the discourse structure includes two stages, that is, training and testing. The main steps are shown as in Figure 2. During the training, the Naïve Bayes model must learn classification parameters and some probability parameters from training corpus, which are respectively applied to marking the topic label of nodes and to merging node for test corpus. In the testing, for a test data, if each paragraph of the data is regarded as a basic analysis unit, there are two operations, namely, topic label marking and nodes merging. Topic label marking is to mark the topic label of the initial nodes or intermediate nodes, and nodes merging is to merge several child nodes into a parent node with Naïve Bayes model. The whole process of discourse structure analysis is to mark topic label and to merge nodes repeatedly, and until a root node is generated.

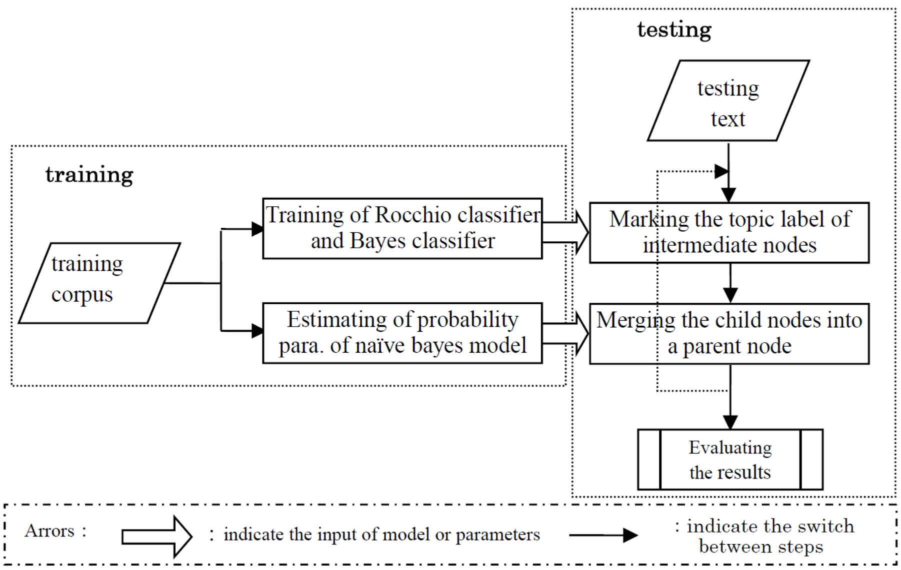

FIGURE 2. Training and testing stage of text structure analysis

3.2. **Marking topic label for nodes.** For a discourse structure tree, the leaf node (corresponding to certain initial paragraph of the discourse) and intermediate node (generated by merging operation) must be marked the topic label in the light of discourse fragment indicated by the node.

3.2.1. *Marking topic label for leaf nodes.* For a given test discourse, the topic label of each leaf node (corresponding to certain initial paragraph of the test discourse) must be marked by using a text classification method. Because the length of different paragraphs in a discourse is usually different, in order to improve the precision of the classifier, we mark the topic label for leaf node by adopting combination method on multiple classifiers, that is, marking the topic for leaf node by respectively using Naïve Bayes classifier and Rocchio classifier, then combining the results of two classifiers as final result by using some combination rules. The following is a combination method of multiple classifiers.

There are three types of methods on multiple classifiers combination [13]. The first combination is at decision level, i.e., each classifier outputs a certain class; the second combination is at sorting level, i.e., each classifier outputs a sorted list of the various possible label for a given test sample; the last is at measurement level, i.e., each classifier outputs a posterior probability of class label for a given test sample. In this paper, we use the second combination method, and combining these results of two classifiers by applying some combination rules.

For a paragraph node V_{2j}, let $CNB = \{C_1^{NB}, C_2^{NB}, C_3^{NB}\}$ denote the former three results by using Naïve Bayes classifier, and $CR = \{C_1^R, C_2^R, C_3^R\}$ denote the former three results by using Rocchio classifier, (the subscript of the element represents priority of class labels), then $C_{NB\wedge R}$ denotes the final result by combining the result of Naïve Bayes and Rocchio classifier. In our experiment, we adopt the following rules for combining the result of two classifiers.

(1) If $C_1^{NB} \in C_R$ then $C_{NB\wedge R} = C_1^{NB}$
(2) If $(C_1^{NB} \notin C_R$ and $C_1^R \in C_{NB})$ then $C_{NB\wedge R} = C_1^R$
(3) If $(C_1^{NB} \notin C_R$ and $C_1^R \notin C_{NB})$ then $C_{NB\wedge R} = C_1^{NB}$

As can be seen from the above rules, the final result is actually to take the outcome from Naïve Bayes classifier as the preferred outcome. The main reason is that we found the accuracy of Naïve Bayes classifier is higher than that of Rocchio classifier shown as the experiment and results of Section 4.3.1.

3.2.2. *Marking topic label for intermediate nodes.* At the same time of merging node, we mark topic label for the generated intermediate node. As shown in Figure 3, assuming that $v_1, v_2, \ldots, v_{q-1}$ and v_q can be merged into a node v', $v_i \in \{v_1, v_2, \ldots, v_{q-1}, v_q\} \subseteq (V_1 \cup V_2)$ and v_i has been marked the topic label, then the task of marking topic label for intermediate nodes is that how to mark the topic label of v'. In this paper, we adopt the Naïve Bayes model to mark the topic label of v'. That is, we can use the Naïve Bayes model to generate a maximum probability node with certain topic label when the topic label of $v_1, v_2, \ldots, v_{q-1}, v_q$ have been marked, as shown in Formula (1).

$$\begin{aligned} v' &= \max_{v_l' \in \{v_1 \ldots v_m\}} \{P(v_l' | v_1, v_2, \ldots, v_{q-1}, v_q)\} = \max_{v_l' \in \{v_1 \ldots v_m\}} \left\{ \frac{P(v_l', v_1, v_2, \ldots, v_{q-1}, v_q)}{P(v_1, v_2, \ldots, v_{q-1}, v_q)} \right\} \\ &= \max_{v_l' \in \{v_1 \ldots v_m\}} \{P(v_l', v_1, v_2, \ldots, v_{q-1}, v_q)\} \end{aligned} \tag{1}$$

In the above Formula (1), $P(v_l'|v_1, v_2, \ldots, v_{q-1}, v_q)$ denotes a probability that the topic of parent node is v_l' ($v_l' \in \{v_1 \ldots v_m\}$, see Section 2 for m) when the topic label of child node is $v_1, v_2, \ldots, v_{q-1}, v_q$.

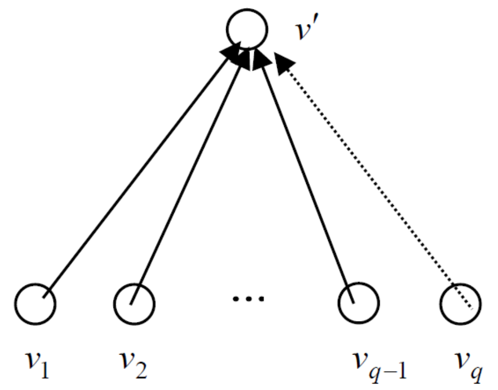

Figure 3. Node merging schematic diagram of discourse structure tree

3.3. **Merging child node.** After the topic label of leaf nodes is marked, we can bottom-up and recursively merge the intermediate nodes and mark the topic label for the merged-generated node. We adopt Naïve Bayes model to merge the intermediate nodes, and the detailed method is as follows.

As shown in Figure 3, assume that the probability that the nodes $v_1, v_2, \ldots, v_{q-1}$ can be merged into the node v' is $P(\{v_1, v_2, \ldots, v_{q-1}\} \rightarrow v')$, so Naïve Bayes model needs to solve the question whether the nodes $v_1, v_2, \ldots, v_{q-1}$ can be merged with the node v_q into the node v'. That is to calculate the probability $P(\{v_1, v_2, \ldots, v_{q-1}, v_q\} \rightarrow v')$. If $P(\{v_1, v_2, \ldots, v_{q-1}\} \rightarrow v')$ is regarded as a conditional probability, it will become into $P(v'|v_1, v_2, \ldots, v_{q-1})$ and the $P(\{v_1, v_2, \ldots, v_{q-1}, v_q\} \rightarrow v')$ will become into $P(v'|v_1, v_2, \ldots, v_{q-1}, v_q)$, and then according to the Bayes formula, we can get:

$$P(v'|v_1, v_2, \ldots, v_{n-1}, v_n) = \frac{P(v', v_1, v_2, \ldots, v_{n-1}, v_n)}{P(v_1, v_2, \ldots, v_{n-1}, v_n)} \tag{2}$$

In fact, perhaps there can get several different topic nodes by merging the $v_1, v_2, \ldots, v_{q-1}$ then we need a topic node v'_{C1} which makes $P(v'_l|v_1, v_2, \ldots, v_{q-1})$ with maximum probability ($1 \le l \le m$, assuming that we can get m prior level topic nodes at the most), that is $v'_{C1} = \max\limits_{v'_l \in \{v_1 \ldots v_m\}} \{P(v'_l|v_1, v_2, \ldots, v_{q-1})\}$. Similarly, a new node v_q and the nodes $v_1, v_2, \ldots, v_{q-1}$ can also get several different topic nodes by merging operation, so we can obtain a node v'_{C2} which makes $P(v'|v_1, v_2, \ldots, v_{q-1}, v_q)$ with maximum condition probability, that is $v'_{C2} = \max\limits_{v'_l \in \{v_1 \ldots v_m\}} \{P(v'_l|v_1, v_2, \ldots, v_{q-1}, v_q)\}$. Therefore, the condition whether the node v_q is merged with the nodes $v_1, v_2, \ldots, v_{q-1}$ or not is that: 1) $v'_{C1} = v'_{C2}$; or 2) $P(v'_{C2}|v_1, v_2, \ldots, v_{q-1}, v_q) > P(v'_{C1}|v_1, v_2, \ldots, v_{q-1})$.

According to the Bayes formula, we can get the following equation to solve for the prior level node v'_{C2} that is merged from the nodes $v_1, v_2, \ldots, v_{q-1}, v_q$ namely:

$$\begin{aligned} v'_{C2} &= \max_{v'_l \in \{v_1 \ldots v_m\}} \{P(v'_l|v_1, v_2, \ldots, v_{q-1}, v_q)\} = \max_{v'_l \in \{v_1 \ldots v_m\}} \left\{ \frac{P(v'_l, v_1, v_2, \ldots, v_{q-1}, v_q)}{P(v_1, v_2, \ldots, v_{q-1}, v_q)} \right\} \\ &= \max_{v'_l \in \{v_1 \ldots v_m\}} \{P(v'_l, v_1, v_2, \ldots, v_{q-1}, v_q)\} \end{aligned} \tag{3}$$

In Formula (3), $P(v'_l, v_1, v_2, \ldots, v_{q-1}, v_q)$ is the probability of the sub-tree which is composed of a node of a discourse and its children nodes. The probability can be estimated by the training corpus, namely:

$$P(v'_l, v_1, v_2, \ldots, v_{q-1}, v_q) = \frac{N(subtree(v'_l, v_1, v_2, \ldots, v_{q-1}, v_q))}{N(subtree(trainText))} \tag{4}$$

In Formula (4), $N(subtree(v'_l, v_1, v_2, \ldots, v_{q-1}, v_q))$ is the number of sub-trees, in which the parent node is v'_l and the child nodes are $v_1, v_2, \ldots, v_{q-1}, v_q$. $N(subtree(trainTe\text{-}xt))$ is the total number of sub-trees of the training corpus.

4. **Experiments and Results.**

4.1. **Experimental corpus and corpus annotation.**

In this paper, we downloaded 40 government work reports of Chinese State Council from 1954 to 2008 and 153 local government work reports in the year of 2003, 2004, 2005, 2007, 2008 as the experimental corpus, amounting to 193. The reason of choosing the government work reports as the experimental corpus is that the government work report is longer, more normative in writing mode, more distinct hierarchy (it makes that non-professional person can easily analyze the structure of the discourse) and more general or standard in using vocabulary. The length of each discourse is about 1.5 to 2.0 million words and generally include 70 to 90 paragraphs.

For the discourse corpus, we manually marked topic label of all paragraphs and the hierarchical structure. In the process of marking, each discourse is marked initially by one person, and then checked by another, and if there exists disagreement in the initial labelling results, the final result is labelled through negotiation.

4.2. **Experiment evaluating method.**

In this paper, we use the method of [9] to evaluate the accuracy of the discourse structure tree, that is, the precision, the recall and the value of $F_{\beta=1}$ as the following formula

$$precision(P) = \frac{\sum_{1\le i\le m}\left(\min\left(n^f_{il}, n^h_{il}\right) + \min\left(n^f_{ir}, n^h_{ir}\right)\right)}{\sum_{1\le i\le m}\left(n^f_{il} + n^f_{ir}\right)} \times 100\% \tag{5}$$

$$recall(R) = \frac{\sum_{1\le i\le m}\left(\min\left(n^f_{il}, n^h_{il}\right) + \min\left(n^f_{ir} n^h_{ir}\right)\right)}{\sum_{1\le i\le m}\left(n^h_{il} + n^h_{ir}\right)} \times 100\% \tag{6}$$

$$F_{\beta=1} = \frac{2 \times P \times R}{P + R} \tag{7}$$

4.3. **Experimental results and analysis.**

4.3.1. *Comparative experiments of marking accuracy between single classifier and multi-classifier fusion.*

In the experiments, we use 20 discourses as training data and 10 as test data. The average number of paragraphs of discourse is 65 (namely, equate to 1300 training samples). We respectively use Naïve Bayes classifier, Rocchio classifier and the multiple classifier fusion as describing in the above Section 3.2.1, to mark the topic label of paragraphs. The experimental result is shown in Figure 4. In Figure 4, Rocchio classifier uses 20 features to mark the topic label of paragraph.

It can be seen from Figure 4 that the performance of multi-classifier fusion is better than the single classifier during marking topic label for paragraphs, which is basically close to about 0.9. In addition, it can be seen from the figure that the performance of Naïve Bayes classifier is better than that of Rocchio classifier.

4.3.2. *Influence of the result of marking topic label for paragraphs by using different classifiers on the accuracy of discourse structure analysis.*

When Naïve Bayes model is used to analyze the discourse structure, the result of marking topic label for the initial paragraph has a great influence on the accuracy of discourse structure analysis. In this paper, we compare the accuracy of discourse structure analysis based on different result of marking topic label by using the Rocchio classifier, the Naïve Bayes classifier and the multi-classifier

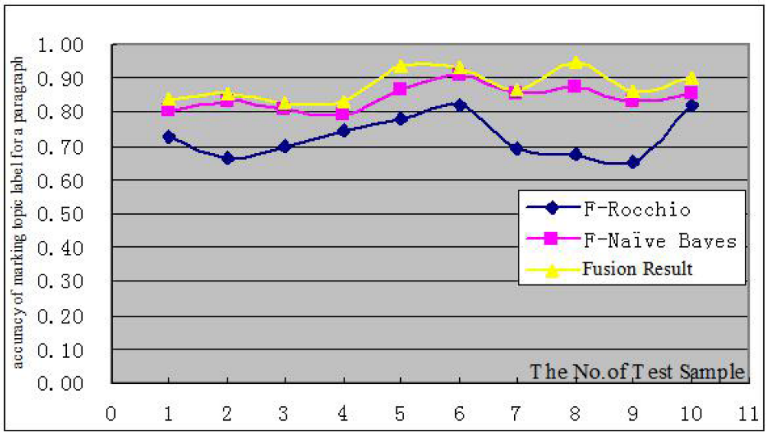

Figure 4. The comparative experiments of marking accuracy between the single classifier and the multi-classifier fusion (20 features)

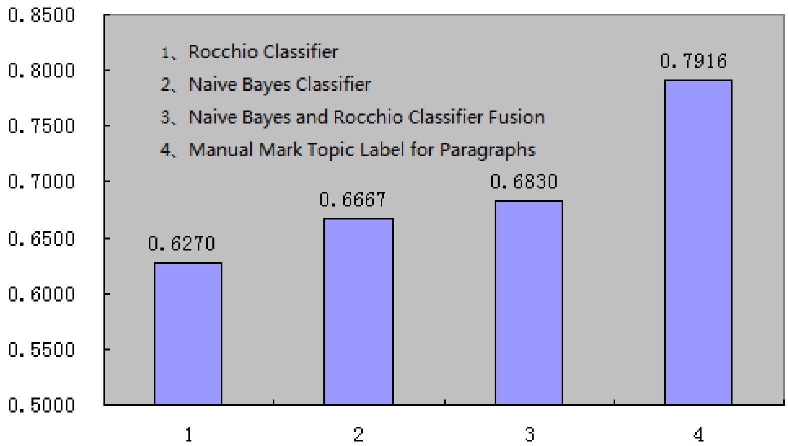

Figure 5. Influence of different methods of marking topic label for paragraphs on the accuracy of the discourse structure analysis

fusion in Figure 5, where $F_{\beta=1}$ in Formula (7) is used to evaluate the accuracy of discourse structure analysis.

As can be seen from Figure 5, the performance of different classifiers, which mark topic label for paragraphs, has a great influence on the accuracy of the discourse structure analysis.

4.3.3. *Comparative experiments of accuracy of the discourse structure analysis based on Naïve Bayes model and sequence alignment algorithm.* Lastly, we randomly chose 30 discourses from all discourses as test data and the rest discourses as training data for the comparative experiments. We did the experiments of discourse structure analysis using the method based on Naïve Bayes model proposed in this paper and using the method of sequence alignment algorithm in [9] as baseline. We adopted the multi-classifier fusion based on Naïve Bayes and Rocchio classifier to mark the topic label of paragraphs in a discourse. 30 test data are divided into three groups and 10 data each group, and $F_{\beta=1}$ mentioned in the above Section 4.2 is used to evaluate the performance of different discourse structure analysis methods. The experiment results are shown in Table 1. As can be seen from the table, the performance of the discourse structure analysis based on Naïve Bayes model is much better than that of the method based on sequence alignment algorithms.

TABLE 1. Comparison of accuracy of the discourse structure analysis based on Naïve Bayes model and sequence alignment algorithm

No. of Test group	Accuracy $F_{\beta=1}$ of Text Structure Analysis based on Naïve Bayes model	Accuracy $F_{\beta=1}$ of Text Structure Analysis based on sequence alignment algorithm
01	0.6871	0.5502
02	0.7169	0.5295
03	0.6928	0.5250
Average value	0.6989	0.5349

4.3.4. *Analysis of experiment results.* The above results of experience show that the performance of the method proposed in this paper is better than that based on the sequence alignment algorithm. If we ignored the influence of the accuracy of marking topic label for paragraphs, we think that the average accuracy of node merging algorithm based on Naïve Bayes model reaches $F_{\beta=1} = 0.6989 \div 0.9 = 0.7766$ (where 0.9 is the average accuracy of marking topic label for paragraphs) and the maximum $F_{\beta=1} = 0.8069 \div 0.9 = 0.8966$ (where 0.8069 is the best performance achieved in experiments). Therefore, the performance is relatively satisfactory.

5. **Conclusions.** We present a hierarchical structure analysis method based on Naïve Bayes model in this paper. The method assumes that the same type discourse should have the same or similar organizational structure mode, so if we marked their organizational structure for some discourses with the same type in advance and used these marked discourses as training data, after the machine learned the organizational structure mode from training data, then it can automatically analyze the hierarchical structure for other same type discourses. The method highly depends on the accuracy of the classifier for marking topic label of each paragraph in a test data, and is also affected by the number of training data with the same type. Therefore, in order to obtain better performance, we need to further research how to improve the accuracy of the classifier for marking topic label, and need increase the number of the training data. Moreover, how to make full use of the semantic information of discourse to get more accurately the hierarchical structure the discourse is also our future research work.

Acknowledgment. This work is supported by the Natural Science Foundation of China (No. 61462027).

REFERENCES

[1] G. Salton and J. Allen, Automatic text decomposition and structuring, *Proc. of Intelligent Multimedia Information Retrieval Systems and Management*, vol.1, pp.6-20, 1994.
[2] G. Salton, J. Allen and C. Buckley, Automatic structuring and retrieval of large text files, *Communications of the ACM*, vol.37, no.2, pp.97-108, 1994.
[3] G. Salton, J. Allen, C. Buckley and A. Singhal, Automatic analysis theme generation and summarization of machine-readable texts, *Science*, vol.264, no.3, pp.1421-1426, 1994.
[4] Y. Yarri, Segmentation of expository texts by hierarchical agglomerative clustering, *Proc. of Recent Advances in Natural Language Processing*, Bulgaria, 1997.
[5] P. Y. Hsueh, J. Moore and S. Renals, Automatic segmentation of multiparty dialogue, *Proceedings of EACL*, 2006.
[6] E. Jacob, Hierarchical text segmentation from multi-scale lexical cohesion, *Proc. of the Association for Computational Linguistics-Human Language Technologies (NAACL-HLT 2009) Conference*, 2009.
[7] H. Lin, X. Zhan and T. Yao, Text structure analysis method based on latent semantic indexing, *Pattern Recognition and Artificial Intelligence*, vol.13, no.1, pp.47-51, 2000.
[8] Y. Zhang, L. Gong and Y. Wang, Hierarchical subtopic segmentation of web document, *Wuhan University Journal of Natural Sciences*, vol.11, no.1, pp.47-50, 2006.

[9] M. Zhong, Using sequence alignment algorithm for analyzing text hierarchical structure, *Journal of Computational Information Systems*, vol.9, no.6, pp.2269-2276, 2013.

[10] W. C. Mann and S. A. Thompson, Rhetorical structure theory: Toward a functional theory of text organization, *Text & Talk*, no.8, pp.243-281, 1988.

[11] F. Yuan and J. Yuan, Naive bayes Chinese text classification based on core words of class, *Journal of Shandong University (Natural Science)*, vol.41, no.3, pp.46-49, 2006.

[12] L. Zeng, An improved clustering-based text classification algorithm using SVM, *Software Guide*, vol.7, no.6, pp.37-39, 2008.

[13] Z. Kou and C. Zhang, Multi-agent based classifier combination, *Chinese Journal of Computers*, vol.26, no.2, pp.174-179, 2003.

[14] Y. Li, W. Feng and G. Zhou, Elementary discourse unit in Chinese discourse structure analysis, *Chinese Conference on Chinese Lexical Semantics*, vol.7717, pp.186-198, 2012.

[15] S. Xu and P. Li, Recognizing chinese elementary discourse unit on comma, *International Conference on Asian Language Processing*, pp.3-6, 2013.

[16] I. Keskes, F. B. Zitoune and L. H. Belguith, Splitting arabic texts into elementary discourse units, *ACM Trans. Asian Language Information Processing*, vol.13, no.2, p.23, 2014.

[17] F. Kong, H. Wang and G. Zhou, A CDT-styled end-to-end Chinese discourse parser, *International Conference on Computer Processing of Oriental Languages & National CCF Conference on Natural Language Processing and Chinese Computing*, 2016.

[18] X. Kang, H. Li, L. Zhou, J. Zhang and C. Zong, An end-to-end Chinese discourse parser with adaptation to explicit and non-explicit relation recognition, *Proc. of the 20th Conference on Computational Natural Language Learning: Shared Task (CoNLL)*, pp.27-32, 2016.

[19] G. Lv, N. Su, R. Li, Z. Wang and Q. Chai, Frame-based discourse structure modeling and relation recognition for Chinese sentence, *Journal of Chinese Information Processing*, 2015.

[20] X. Chu, Z. Wang, Q. Zhu and G. Zhou, Recognizing nuclearity between Chinese discourse units, *2015 International Conference on Asian Language Processing (IALP)*, pp.197-200, 2016.