

# Chinese Grammatical Error Diagnosis System Based on Hybrid Model

Xiupeng Wu, Peijie Huang\*, Jundong Wang, Qingwen Guo, Yuhong Xu, Chuping Chen

College of Mathematics and Informatics, South China Agricultural University,  
Guangzhou 510642, Guangdong, China

zxc2012@gmail.com, pjhuang@scau.edu.cn, mo\_xiao\_wang@163.com,  
tryven.guo@qq.com, 137610184@qq.com, 568093091@qq.com

## Abstract

This paper describes our system in the Chinese Grammatical Error Diagnosis (CGED) task for learning Chinese as a Foreign Language (CFL). Our work adopts a hybrid model by integrating rule-based method and n-gram statistical method to detect Chinese grammatical errors, identify the error type and point out the position of error in the input sentences. Tri-gram is applied to disorder mistake. And the rest of mistakes are solved by the conservation rules sets. Empirical evaluation results demonstrate the utility of our CGED system.

## 1 Introduction

Chinese as a foreign language (CFL) is booming in recent decades. The number of (CFL) learners is expected to become larger for the years to come (Xiong et al., 2014). But the flexibility and complication in Chinese morphology, pronunciations and grammar make Chinese become one of the hardest languages to learn. If you cannot make good use of the grammatical rules, maybe the many different meaning or error meaning of the sentence will be get. Empirically, there were 2 errors per student essay on average in a learners' corpus (Chen et al., 2011). From some previous research on second language acquisition, it indicated that effective provision of corrective feedback can contribute to the development of grammatical competence in second language learners (Fathman and Whalley, 1990; Ashwell, 2000; Ferris and Robers, 2001; Chandler, 2003). Therefore developing a check tool which can automatically detect and correct

Chinese grammatical errors is a very important and useful work for foreigners to learn Chinese. And it helps to detect the wrong grammar from a large number of documents.

Recently, there were a number of shared task of grammatical error correction for learners, including the CoNLL-2013 (Ng et al., 2013), the CoNLL-2014 (Ng et al., 2014), the ICCE-2014 (Yu et al., 2014), the Chinese Grammatical Error Diagnosis (CGED) task, etc. These tasks have been organized to provide a common platform for comparing and developing automatic Chinese grammatical error diagnosis system.

In NLP, grammar diagnosis is a difficult problem for sentence comprehension. In English, so much research is under way up to now and many learning assistance tools were developed by natural language processing technology to detect and correct the grammatical errors of EFL learners (Chodorow et al., 2012; Leacock et al., 2010). And the demand for automatic Chinese proofreading has driven an increase in study for this task in Chinese area. Cheng et al. (2014) and Yu and Chang (2012) designed word order error detection technology focused on the Chinese sentences in the HSK Dynamic Composition Corpus. Yuan and Felice (2013) proposed the use of phrase-based statistical machine translation to grammatical error correction. Chang et al. (2012) presented a rule-based learning algorithm combined with a log-likelihood function to identify error types in Chinese texts. In summary, all of these methods mainly focus on the statistical machine learning (SML) like n-gram language model (LM) and rule-based method, indicating that SML model and rule-based method still being useful and effective for Chinese grammatical correcting.

This paper propose a hybrid model for CGED

---

\* Corresponding author

shared task by integrating rule-based methods and n-gram statistical methods to detect Chinese grammatical errors, identify the error type and point out the position of error in the input sentences. The rule-based method provides 405 handcrafted rules (Missing rule-30, Redundant rule-75, Selection rule-300) constructed from the training set provided by organizer to identify potential rule violations in input sentences and correct the error sentences. Tri-gram is applied to disorder mistake. After the above special processes, once the candidate sentence set does not have only one sentence, we adopt two strategies: first is a general process, in which the n-gram statistical method relies on the n-gram scores of both standard (correct) corpus and four non-standard (incorrect) training corpora is used to determine the correction and the error type in

the input sentence, and second is just adjusting the priority among four special processes.

The remainder of this paper is structured as follows. Section 2 provides an overview of related work. Section 3 presents the rule we write and the tri-gram model we build. Section 4 presents our results. Section 5 is discussion of future work concludes the paper.

## 2 The Proposed System

### 2.1 System Overview

Figure 1 shows the flowchart of our CGED system. The system is mainly composed by two processes: preprocess and main process. In addition, main process contains two subprocesses: special process and general process.

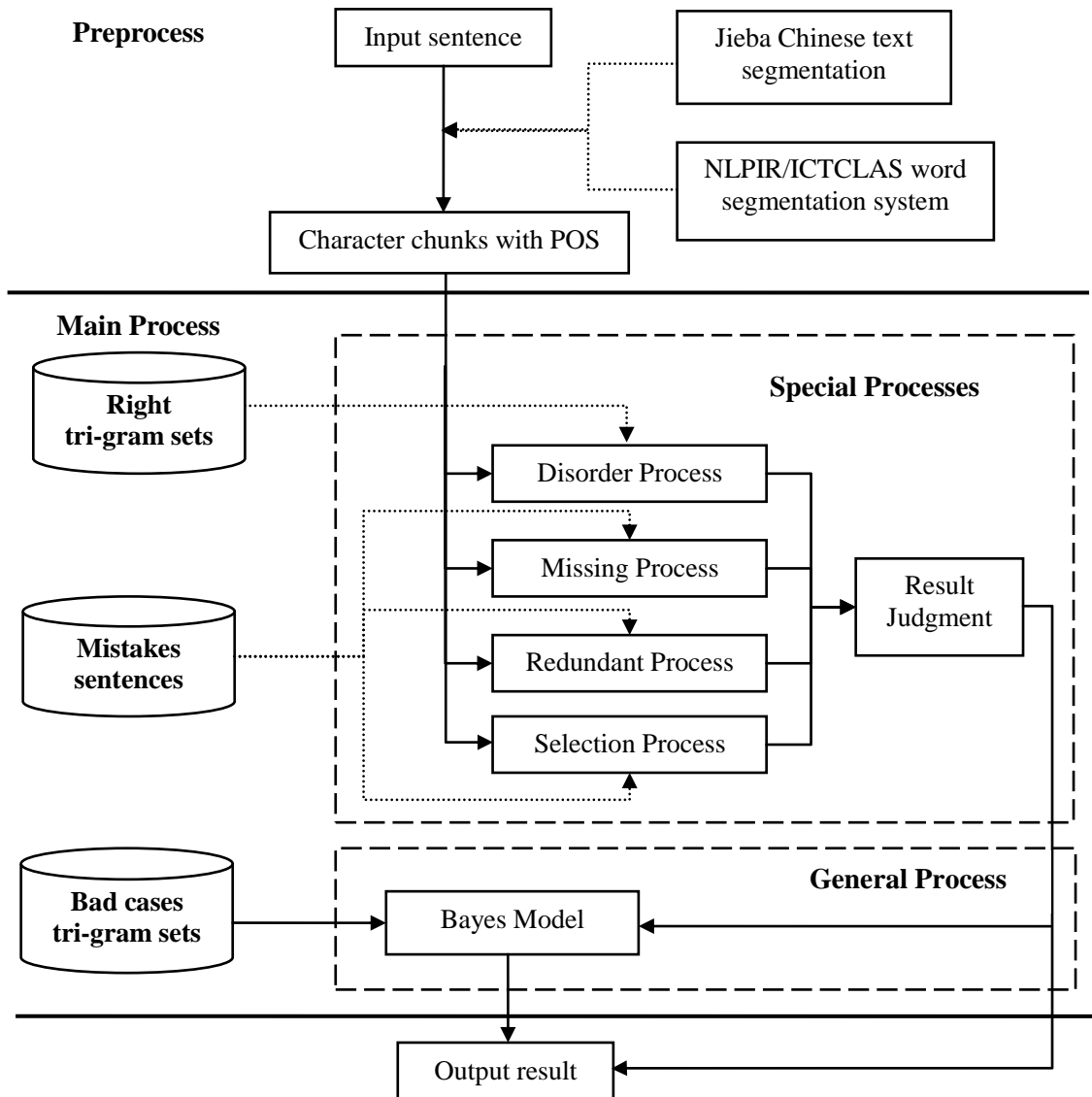


Figure 1. The flowchart of the hybrid CGED system.

It performs CGED in the following steps:

1. Given a test sentence, the CGED system gets the character chunks in the sentence with POS.

2. For each chunk in this sentence, the system will enumerate every rule in the missing, redundant and selection rules sets. In the meanwhile, we got the all permutations of the chunks. What's more we use tri-gram model (We use the corpus of CCL<sup>1</sup> to generate the frequency of tri-gram) to calculate the probabilities of each generated sentence in the all permutations and pop the highest one. We will get a candidate sentence set after this step.

3. If the candidate sentence set has only one sentence, the system will return related data based on the sentence. However, if the candidate sentence set does not have only one sentence, system will carry out the general process.

## 2.2 Preprocess

Preprocess in the system contains two modules: Chinese word segmentation and part-of-speech tagging. We uses "Jieba" Chinese text segmentation<sup>2</sup> and NLP/ ICTCLAS Chinese text segmentation<sup>3</sup> to achieve the goal. A set of chunks with POS will be generated after this process.

## 2.3 Main Process

Main process contains two subprocess: special process and general process. Special process has four different processes applied to detect and correct the sentence with correspondent grammatical error. General process decides sentences' error type through the Bayes model trained by data set from NTNU, if the sentence input has grammatical error. After that, we deal with the sentence according to its error type. If the candidate sentence set does not have only one sentence, system will carry out the general process.

## 3 Special processes and General process

### 3.1 Special processes

#### Disorder process

For the case of the missing syntax, we use the method as follow: System generates the all permutations of chunk set from preprocess. Then it calculates the probability of each sentence in

the sentence set generated by method above through tri-gram. If the highest probability one differs from the origin one, system judges that the sentence has disorder error. Result generated by the tri-gram model used in this system has the lowest degree of confidence among the special processes. As a result, disorder process has the lowest priority level.

Figure 2 shows the process of disorder error detection and correction ("O" original, "M" modified).

O: 所以我不會讓失望她  
Segmentation: 所以/c 我/r 不會/v 讓/v 失望/v 她/r  
All Permutations:  
1) 所以我不會讓失望她  
2) 所以我不會讓她失望  
(Highest probability in tri-gram)  
3) 所以我不會失望她讓  
4) 所以我讓失望她不會  
5) 所以不會讓失望她我  
6) 我不會讓失望她所以  
7) 所以我不會失望讓她  
8) 所以我讓失望不會她  
9) 所以不會讓失望我她  
...  
M: 所以我不會讓她失望

Figure 2. The process of disorder error detection and correction.

#### Missing process

For the case of the missing syntax, this paper uses rules to deal with them. Through the collection of the grammar deletion, extract the sentence features of the deletion of the grammar, and analyze the grammar and summarize the relevant rules.

We intend to extract common phrase structures from the training concentrate which has the missing syntax. The sentence structures containing the similar syntax are summed up, and the structural features of the sentence are summarized. For the missing part of syntax, we sum up about 30 rules consisting of the simplified structure of the sentence, and we give the corresponding correction rules. The common syntax missing sentence structures are similar to the rules: "m+n+v", "r+v+j+v+a+v", "c+r+d+v+m+a", we give the error correction rule that corresponds to it: "m+n+v+了", "r+v+j+v+a+的+v", "c+r+d+v+很+m+a". We

<sup>1</sup> ccl.pku.edu.cn:8080/ccl\_corpus/index.jsp?dir=xiandai

<sup>2</sup> https://github.com/fxsjy/jieba

<sup>3</sup> http://ictclas.nlpir.org/

summed up the rules of missing detection to the training set contains 60% of the missing part of the grammar of the sentence, the investigation of the sentence probability is 40%, the sentence correct probability of 15%.

Figure 3 shows the process of missing error detection and correction (“O” original, “M” modified).

<p>Case1: O: 我高興我的老師是那位小姐 Segmentation: 我/r 高興/a 我/r 的/uj 老師/n 是/v 那位/r 小姐/nr Detection rule: “r+b+r+uj+n+v+r+nr” Correction rule: “r+很+b+r+uj+n+v+r+nr” M: 我很高興我的老師是那位小姐</p> <p>Case 2: O: 他爬三個小時 Segmentation: 他/r 爬/v 三個/m 小/a 時/ng Detection rule: “r+v+m+n” Correction rule: “r+v+了+m+n” M: 他爬了三個小時</p> <p>Case 3: O: 我是你小學朋友 Segmentation: 我/r 是/v 你/r 小學/n 朋友/n Detection rule: “r+v+r+n+n” Correction rule: “r+v+r+n+的+n” M: 我是你小學的朋友</p>
---

Figure 3. The process of missing error detection and correction.

### Redundant process

Considering the redundant mistakes have strong syntax structure, this paper use rules to deal with the redundant mistakes. Previous work is done like the method of dealing with missing mistakes. Through the collection of the grammar redundant, extract the sentence features of the redundant of the grammar, and analyze the grammar and summarize the relevant rules. Then we intend to extract common phrase structures from the training concentrate which has redundant mistakes. The sentence structures containing the similar syntax are summed up, and the structural features of the sentence are summarized.

For the redundant part of syntax, we sum up about 75 rules. Rules are in form of the pattern as (wrong pattern, revised pattern, position mark) such as (“v+m+nr+ul”, ”v+m+nr”, 4), (“v+ul+n”, ”v+n”,2), (“v+n+c+d+uv+v”, “v+n+d+uv+v”, 3). The position mark takes an important part in the case of the wrong pattern

has two same tags such as “v+v+r+uj+n” in (“v+v+r+uj+n”, “v+r+uj+n”, 1). In this case, the position mark can prevent from deleting the second “v” which will turn the result into wrong direction.

We detection to the redundant part of the training set discovers that the rules has covered 118 sentences, accounting to 27.4%. In the covered cases, there are 71 sentences have been corrected rightly, accounting to 60.1% correct probability. Figure 4 shows the process of redundant error detection and correction (“O” original, “M” modified).

<p>Case1: O: 晚上五點半他們到了火車站去 Segmentation: 晚上/t 五點半/m 他們/r 到/v 了/ul 火車/n 站/v 去/v Rule: (“v+ul+n”, “v+n”, 2) M: 晚上五點半他們到火車站去</p> <p>Case 2: O: 她們一起看美國的電影 Segmentation: 她們/r 一起/m 看/v 美國/ns 的 /uj 電影/n Rule: (“ns+uj+n”, ”ns+n”, 2) M:她們一起看美國電影</p> <p>Case 3: O: 我发现了你乱丢垃圾 Segmentation: 我/r 发现/v 了/ul 你/r 乱丢垃圾/n Rule: (“v+ul+r”, “v+r”, 2) M: 我发现你乱丢垃圾</p>
--

Figure 4. The process of redundant error detection and correction.

### Selection process

In linguistics, selection denotes the ability of specific words to determine the semantic contents of their arguments. It is a semantic concept, whereas subcategorization is a syntactic one. For the cases of selection syntax, this paper takes a more empirical approach to deal with them. We intended to summarize the relevant rules by observing the semantic relations of those specific words with parse tree and dependency tree. Unfortunately, we did not have enough time and resources to do it. We had to extract the sentence features of selection syntax, analyze the grammar and summarize the rules artificially.

Through the previous 50% of the training corpus, we can collect the specific words. Then we extract the collocation of those specific words

to make the rules. In this paper, we use the NLPPIR Chinese Word Segmentation to segment the sentences. The sentences containing the same specific words and similar syntax are summed up, and then we summarize the rule, a single generative equation for these sentences. Around 300 rules are made and we also give the corresponding correction rules. We find the way to be great because some rules have good generalization ability. Most common syntax selection sentence structures are similar to the rules like: “. [部]/mq+電影”, “z+的+vn”, and we give the error correction rules that correspond to them: “. 部/mq+電影”, “z+地+vn”. However, since we extract the rules by human beings, the rules are limited and even some of them are not very reasonable.

Figure 5 shows the process of selection error detection and correction (“O” original, “M” modified).

<p>Case1:  O: 你決定那個電影  Segmentation:你/rr 決定/v 那個/rz 電影/n  Detection rule: . [部]/mq+電影  Correction rule: . 部/mq+電影  M: 你決定那部電影</p> <p>Case 2:  O: 我要清清楚楚的說明  Segmentation:我/rr 要/v 清清楚楚/z 的 /ude1 說明/vn  Detection rule: z+的+vn  Correction rule: z+地+vn  M: 我要清楚地說明</p>
---

Figure 5. The process of selection error detection and correction

### 3.2 General process

In this section, we describe an approach of general process to grammatical error diagnosis where the special process cannot well diagnose the error. Inspired by existing related work, we considered a frequency-based solution to approach the task. Therefore we use a frequency-based approach comparing n-gram frequency lists to both the standard corpus and the non-standard (error) corpus. The standard corpus above is made of correct sentences extracted from the training corpus provided by the shared task organizers and the error sentences consist of the non-standard corpus. The assumption behind this approach is that comparing a standard corpus

to a non-standard corpus using frequency-based methods levels out non-standard features present in the non-standard corpus. These features are very likely to be, in the case of this corpus, grammatical errors.

As discussed above, we can acquire the keyword lists are produced by comparing two corpora (a standard corpus and a non-standard corpus) using association metrics such as log-likelihood, chi-square or mutual information. These keywords usually reflect salient features of the learner corpus. In the case of the present comparison, it is safe to assume that a reasonable amount of salient features from the learner corpus will be in frequent distributions of words which are very likely to be errors.

For proving our approach, we pre-processed the training corpus provided by organizers. As Chinese is a logographic language we treat every character in isolation in this process. Firstly, We separate the correct sentences and the error sentences from the training corpus to construct the standard corpus and non-standard corpora (for the task have four kind of errors so we will acquire four non-standard corpora, respectively the disorder corpus, the missing corpus, the redundant corpus, the selection corpus) and then extract the n-grams keywords from all corpora including the standard and the four non-standard corpora).By respectively comparing the four non-standard corpora to the standard corpus, we can extract the four kinds of a list of ungrammatical n-grams list corresponding to the four kinds of non-standard corpora and treat them as key expressions. This calculation returned us a list of 22071 ungrammatical n-grams (disorderset-5075, missingset-1035, redundantset-7810, selectionset-8151) not present in the standard corpus. In these experiments we just used the tri-gram set extracted from the corpus .With these n-gram lists, we trained a classifiers to identify the grammatical error type. An n-gram based Multinomial Naive Bayes (MNB) classifiers to identify grammatical error sentences using the formula below:

$$p(s) = \prod_{i=1}^l p(w_i | w_{i-2} w_{i-1}) \quad (1)$$

$$p(w_i | w_{i-2} w_{i-1}) = \frac{c(w_{i-2} w_{i-1} w_i)}{\sum_{w_i} c(w_{i-2} w_{i-1} w_i)} \quad (2)$$

where p(s) is the probability of the sentence. We need to calculate the sentence’s probability in the four corpus and by comparing the probability of

the four error type, we can select the error type having the maximum probability as the candidate. If all probability is less than the threshold  $x$ , we regard the sentences as the correct sentence. After a number of tests we found that the proper optimal value.

## 4 Empirical Evaluation

### 4.1 Task

The goal of this shared task, i.e. Chinese Grammatical Error Diagnosis (CGED) task for CFL is developing the computer-assisted tools to diagnose several kinds of grammatical errors, i.e., redundant word, missing word, word disorder, and word selection. The system should indicate which kind of error type is embedded in the given sentence and it's occurred positions. Passages of CFL (Chinese as a Foreign Language) learners' essays selected from the National Taiwan Normal University (NTNU) learner corpus are used for training purpose. The training data (consisting of 2212 grammatical errors) is provided as practice. The final test data set for the evaluation consists of 1000 passages cover different grammatical errors.

### 4.2 Metrics

The criteria for judging correctness are:

(1) Detection level: Binary classification of a given sentence, i.e., correct or incorrect should be completely identical with the gold standard. All error types will be regarded as incorrect.

(2) Identification level: This level could be considered as a multi-class categorization problem. In addition to correct instances, all error types should be clearly identified, i.e., redundant, missing, disorder, and selection.

(3) Position level: Besides identifying the error types, this level also judges the positions of erroneous range. That is, the system results should be perfectly identical with the quadruples of gold standard. The following metrics are measured in both levels with the help of the confusion matrix.

In CGED task of the NLP-TEA 2015 (The 2nd Workshop on Natural Language Processing Techniques for Educational Applications), 13 metrics are measured in both levels to score the performance of a CGED system. They are False Positive Rate (FPR), Detection Accuracy (DA), Detection Precision (DP), Detection Recall (DR), Detection F-score (DF), Identification Accuracy (IA), Identification Precision (IP), Identification Recall (IR), Identification F-score (IF), Position Accuracy (PA), Position Precision (PP), Position Recall (PR) and Position F-score (PF).

### 4.3 Evaluation Results

The CGED task of CFL attracted 13 research teams. Among 13 registered teams, 6 participants submitted their testing results. For formal testing, each participant can submit at most three runs that use different models or parameter settings. Finally, there are 18 runs submitted in total.

#### Validation

We use the 30% of NLP-TEA-1 CFL Datasets (Yu et al., 2014) as validation set to test the effect and performance of the four special processes and the MNB. Table 1 shows the validation results.

Test1 validates the four special processes and gives different type of rules the same privileges when the candidate sentence set does not have only one sentence after special process. The detective accuracy of the four special processes in this test, i.e. redundant, selection, missing, and disorder are 0.934, 0.967, 0.828, and 0.491 respectively.

Thus, we perform three tests which give lower priority to disorder process when the candidate sentence set does not have only one sentence after special process:

Test2 (Redundant = Selection = Missing > Disorder): This test gives different type of rules the same privileges, and gives the minimum priority to the disorder process.

	<b>DP</b>	<b>DR</b>	<b>DF</b>	<b>IP</b>	<b>IR</b>	<b>IF</b>	<b>PP</b>	<b>PR</b>	<b>PF</b>
Test1	0.486	0.508	0.496	0.171	0.111	0.135	0.038	0.021	0.027
Test2	0.494	0.643	0.559	0.192	0.156	0.172	0.056	0.039	0.046
Test3	0.494	0.643	0.559	0.198	0.162	0.178	0.060	0.042	0.049
Test4	0.522	0.285	0.369	0.281	0.102	0.150	0.103	0.030	0.047
Test5	0.504	0.976	0.665	0.160	0.183	0.171	0.021	0.021	0.021

Table 1. Validation results of NLP-TEA-1 CFL Datasets.

Test3 (Redundant = Selection > Missing > Disorder): This test decrease the priority of missing in order to adjust the collision rate of redundant rules and missing rules.

Test4 (Redundant = Selection = Missing && Disorder = 0): This test is the result using the method of Run1 without the disorder process in order to reduce the wrong judgment rate.

And we make another test, Test5, for using MNB when candidate sentence set does not have only one sentence after special process.

We found the approach using MNB (Test5) has mostly no advantage with that only using special process (Test2, Test3, and Test4). Therefore, the three runs of our system submitted to NLP-TEA-2 CLF final test are all based on the four special processes.

#### NLP-TEA-2 CLF final test

Table 2 shows the evaluation results of the NLP-TEA-2 CFL final test. Run1, Run2 and Run3 are the three runs of our system corresponding to Test2, Test3, Test4, respectively. The “Best” indicates the high score of each metric achieved in CGED task. The “Average” represents the average of the 18 runs. As we can see from Table 2, we achieve a result close to the average level. Some typical errors of our current system will be presented in the next subsection, and the corresponding improvements are summarized in the last section.

#### 4.4 Error Analysis

Figure 6 shows some typical error examples of our system (“R” right sentence, “M” modified). This approach has several defects: if the segmentation results are wrong, and even wrong segmented place to the synthesis of the word having grammatical errors, it will lead later processing meaningless. And the tri-gram effect depends on the corpus. The Bad-case tri-gram extracted from the training corpus may not appear in the test set, which will affect the validity of the error correction. On the other hand,

**Case 1 ( Redundant)**  
R: 我碰到一个小孩  
Segmentation: 我/r 碰到/v 一个/m 小孩/n  
Rule:(“r+v+m”, “r+m”,2)  
M: 我一个小孩

**Case 2 (Missing)**  
R: 我看到一輛車的時候  
Segmentation: 我/r 看到/v 一輛/m 車/n 的/uj 時候/n  
Detection rule: “r+v+m+n”  
Correction rule: “r+v+m+n+了”  
M: 我看到一輛車了的時候

**Case 3 (Disorder)**  
R: 最近很難找到工作  
Segmentation: 最近/f 很/d 難/a 找到/v 工作 /vn  
All Permutations:  
1) 最近很難工作找到  
2) 最近很工作難找到  
3) 最近工作很難找到  
4) 工作最近很難找到  
5) 工作很难找到最近  
(Highest probability in tri-gram)  
...  
M: 工作很难找到最近

**Case 4 (Selection)**  
R: 我看到的是她很失望的臉  
Segmentation: 我/r 看到/v 的/uj 是/v 她/r 很/d 失望/v 的/uj 臉/n  
Detection rule: “看到 + r”  
Correction rule: “看見 + r”  
M: 我看見的是她很失望的臉

Figure 6. Error examples.

we use the extraction rules to correct the sentences that are syntax errors. First, training set cannot guarantee the existence of all syntax errors. Second, we extract the rules represents only a part of the grammar rules, and grammar mistakes in language is infinite, rules are not represented at all.

	FPR	DA	DP	DR	DF	IA	IP	IR	IF	PA	PP	PR	PF
<b>Run1</b>	<b>0.620</b>	<b>0.505</b>	<b>0.504</b>	<b>0.630</b>	<b>0.560</b>	<b>0.287</b>	<b>0.238</b>	<b>0.194</b>	<b>0.214</b>	<b>0.217</b>	<b>0.080</b>	<b>0.054</b>	<b>0.065</b>
Run2	0.636	0.503	0.502	0.642	0.564	0.279	0.234	0.194	0.212	0.209	0.078	0.054	0.064
Run3	0.266	0.503	0.506	0.272	0.354	0.416	0.269	0.098	0.144	0.385	0.119	0.036	0.055
Average	0.538	0.534	0.560	0.607	0.533	0.335	0.329	0.208	0.233	0.263	0.166	0.064	0.085
Best	0.082	0.607	0.7453	1.000	0.675	0.525	0.617	0.364	0.358	0.505	0.529	0.160	0.174

Table 2. Evaluation results of NLP-TEA-2 CFL final test.

In the first case about redundant, our rules (r+v+m) are in line with the sentence structure (我/r 碰到/v 一个/m), but the corrected sentence (r+m) is not correct. This sentence is a case of the missing, which has the same sentence structure of the redundant rule. The case should be due to the conflict between the redundant rules and the missing rules, which has match the same sentence structure.

In the second case about missing, we extracted the rule (r+v+m+n) from the sentence (我/r 没有/v 很多/m 时间/n) in the corpus. The corrected rule for this type of sentence is: “r+v+m+n+了”. However, this sentence is a case of the redundant, which has the same sentence structure of the missing rule. The case should be due to the conflict between the redundant rules and the missing rules, which has match the same sentence structure.

In the third case about disorder, the highest score in the tri-gram is not the result expected, because only using the POS and the sequence of words as a component of the rules has certain limitations. And tri-gram method could not distinguish long distance displacement.

In the fourth case about selection, there are two reasons for the wrong selection: incorrect usage of quantifiers and function words (“的”等虚词). Under this scenario, the artificial rules are hard to cover completely. Only using the POS and the sequence of words as the elements of the rules, it cannot resolve the problem of conflict of rules. As in the case, our rule (看到 + rr -> 看见 + rr) cannot be applied to the current test sentence, the corresponding rule to correct the sentence into a wrong sentence.

## 5 Conclusions and Future Work

This paper presents the development and preliminary evaluation of the system from team of South China Agricultural University (SCAU) that participated in the CGED shared task. We have developed hybrid model by integrating rule-based method and n-gram statistical method to detect Chinese grammatical errors, identify the error type and point out the position of error in the input sentences. Tri-gram is applied to disorder mistake. And the rest of mistakes are solved by the 405 handcrafted rules (Missing rule-30, Redundant rule-75, Selection rule-300).

It is our first attempt on Chinese grammatical error diagnosis, and our system achieves a result

close to the average level. However, we still have a long way from the state-of-arts results. Due to limitation of time and resources in this task, we have to summarize the relevant rules by extracting the sentence features of selection syntax, analyzing the grammar and summarize the rules artificially instead of observing the semantic relations of those specific words with parse tree and dependency tree. Future work will use a more effective way to capture rules to further improve the CGED. Future work will not only aim at diagnosing grammatical errors, but also explore ways to correct grammatical errors.

## Acknowledgments

This work was partially supported by National Natural Science Foundation of China under Grant No. 71472068, Science and Technology Planning Project of Guangdong Province, China under Grant No. 2013B020314013, and the Innovation Training Project for College Students of Guangdong Province under Grant No.201410564290.

## References

- Ashwell, T. (2000). Patterns of teacher response to student writing in a multi-draft composition classroom: Is content feedback followed by form feedback the best method? *Journal of Second Language Writing*, 9, 227–257.
- Chandler, J. (2003). The efficacy of various kinds of error feedback for improvement in the accuracy and fluency of L2 student writing. *Journal of Second Language Writing*, 12, 267–296.
- Chang, T. H., Sung, Y. T., & Lee, Y. T. (2012). A Chinese word segmentation and POS tagging system for readability research. In *Proceedings of the 42<sup>nd</sup> Annual Meeting of the Society for Computers in Psychology (SCiP 2012)*, Minneapolis, MN.
- Yong-Zhi Chen, Shih-Hung Wu, Ping-che Yang, et al. (2011). Improve the detection of improperly used Chinese characters in students' essays with error model. *Int. J. Cont. Engineering Education and Life-Long Learning*, vol. 21, no. 1, pp.103-116.
- Cheng, S. M., Yu, C. H., & Chen, H. H. (2014). Chinese Word Ordering Errors Detection and Correction for Non-Native Chinese Learners. In *Proceedings of the 25<sup>th</sup> International Conference on Computational Linguistics (COLING 2014)*, Dublin, Ireland, pp. 279-289.
- Chodorow, M., Dickinson, M., Israel, R., & Tetreault, J. R. (2012). Problems in Evaluating Grammatical Error Detection Systems. In *Proceedings of the*



24<sup>th</sup> International Conference on Computational Linguistics (COLING 2012), Mumbai, India, pp. 611-628

- Fathman, A. K., & Whalley, E. (1990). Teacher response to student writing: Focus on form versus content. In B. Kroll (Ed.), *Second language writing: Research insights for the classroom* Cambridge, UK: Cambridge University Press, pp. 178-190.
- Ferris, D. & Roberts, B. (2001). Error feedback in L2 writing classes: How explicit does it need to be? *Journal of Second Language Writing*, 10, pp. 161-184.
- Leacock, C., Chodorow, M., Gamon, M., et al. (2010). *Automated Grammatical Error Detection for Language Learners*. Morgan and Claypool Publishers.
- Ng, H. T., Wu, S. M., Wu, Y., Hadiwinoto, C.,Tetreault, J. (2013). The CoNLL-2013 Shared Task on Grammatical Error Correction. *In Proceedings of the 17<sup>th</sup> Conference on Computational Natural Language Learning (CoNLL-2013): Shared Task*. Sofia, Bulgaria.9 August, pp. 1-12
- Ng, H.T., Wu, S.M., Briscoe, T., et al. (2014). The CoNLL-2014 shared task on grammatical error correction. *In Proceedings of the 18<sup>th</sup> Conference on Computational Natural Language Learning (CoNLL-2014): Shared Task*. Baltimore, Maryland, USA, 26-27 June, pp. 1-14
- Xiong J. H., Zhao Q., Hou J.P., et al. (2014). Extended HMM and Ranking Models for Chinese Spelling Correction. *In Proceedings of the Third CIPS-SIGHAN Joint Conference on Chinese Language Processing (CLP 2014)*, Wuhan, China, 20-21 Oct., pp. 133-138.
- Yu, C.H., & Chen, H.H. (2012). Detecting word ordering errors in Chinese sentences for learning Chinese as a foreign language. *In Proceedings of the 24<sup>th</sup> International Conference on Computational Linguistics (COLING 2012)*, Bombay, India, pp. 3003-3018.
- Yu L.C., Lee L.H., Chang L.P. (2014).Overview of Grammatical Error Diagnosis for Learning Chinese as a Foreign Language. *In Proceedings of the 22<sup>nd</sup> International Conference on Computers in Education(ICCE 2014)*,Nara, Japan,30 Nov. - 4 Dec., pp. 42-47
- Yuan Z., Felice M. (2013). Constrained Grammatical Error Correction Using Statistical Machine Translation. *In Proceedings of the 17<sup>th</sup> Conference on Computational Natural Language Learning (CoNLL-2013): Shared Task*. Sofia, Bulgaria. 9 August, pp. 52-61.