Tense and Aspect Error Correction for ESL Learners Using Global Context

Toshikazu Tajiri **Mamoru Komachi** **Yuji Matsumoto**
Graduate School of Information Science
Nara Institute of Science and Technology
8916-5 Takayama, Ikoma, Nara, 630-0192, Japan
{toshikazu-t, komachi, matsu}@is.naist.jp

Abstract

As the number of learners of English is constantly growing, automatic error correction of ESL learners' writing is an increasingly active area of research. However, most research has mainly focused on errors concerning articles and prepositions even though tense/aspect errors are also important. One of the main reasons why tense/aspect error correction is difficult is that the choice of tense/aspect is highly dependent on global context. Previous research on grammatical error correction typically uses pointwise prediction that performs classification on each word independently, and thus fails to capture the information of neighboring labels. In order to take global information into account, we regard the task as sequence labeling: each verb phrase in a document is labeled with tense/aspect depending on surrounding labels. Our experiments show that the global context makes a moderate contribution to tense/aspect error correction.

1 Introduction

Because of the growing number of learners of English, there is an increasing demand to help learners of English. It is highly effective for learners to receive feedback on their essays from a human tutor (Nagata and Nakatani, 2010). However, manual feedback needs a lot of work and time, and it also requires much grammatical knowledge. Thus, a variety of automatic methods for helping English learning and education have been proposed.

The mainstream of English error detection and correction has focused on article errors (Knight and Chander, 1994; Brockett et al., 2006) and preposition errors (Chodorow et al., 2007; Rozovskaya and Roth, 2011), that commonly occur in essays by ESL learners. On the other hand, tense and aspect errors have been little studied, even though they are also commonly found in learners' essays (Lee and Seneff, 2006; Bitchener et al., 2005). For instance, Lee (2008) corrects English verb inflection errors, but they do not deal with tense/aspect errors because the choice of tense and aspect highly depends on global context, which makes correction difficult. Consider the following sentences taken from a corpus of a Japanese learner of English.

(1) I had a good time this Summer Vacation.
First, I *go to KAIYUKAN [1] with my friends.

In this example, *go* in the second sentence should be written as *went*. It is difficult to correct this type of error because there are two choices for correction, namely *went* and *will go*. In this case, we can exploit global context to determine which correction is appropriate: the first sentence describes a past event, and the second sentence refers the first sentence. Thus, the verb should be changed to past tense. This deduction is easy for humans, but is difficult for machines.

One way to incorporate such global context into tense/aspect error correction is to use a machine learning-based sequence labeling approach. Therefore, we regard the task as sequence labeling: each verb phrase in the document is labeled with tense/aspect depending on surrounding labels. This model naturally takes global context into account. Our experiments show that global context makes a moderate contribution to tense/aspect correction.

[1] Kaiyukan is an aquarium in Osaka, Japan.

Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics, pages 198–202,
Jeju, Republic of Korea, 8-14 July 2012.

2 Tense/Aspect Error Corpus

Developing a high-quality tense and aspect error correction system requires a large corpus annotated with tense/aspect errors. However, existing annotated corpora are limited in size,[2] which precludes the possibility of machine learning-based approach. Therefore, we constructed a large-scale tense/aspect corpus from *Lang-8*,[3] a social networking service for learners of foreign languages. ESL learners post their writing to be collaboratively corrected by native speakers. We leverage these corrections in creating our tense/aspect annotation. Lang-8 has 300,000 users from 180 countries worldwide, with more than 580,000 entries, approximately 170,000 of them in English.[4] After cleaning the data, the corpus consists of approximately 120,000 English entries containing 2,000,000 verb phrases with 750,000 verb phrases having corrections.[5] The annotated tense/aspect labels include 12 combinations of tense (past, present, future) and aspect (nothing, perfect, progressive, perfect progressive).

3 Error Correction Using Global Context

As we described in Section 1, using only local information about the target verb phrase may lead to inaccurate correction of tense/aspect errors. Thus, we take into account global context: the relation between target and preceding/following verb phrases. In this paper, we formulate the task as sequence labeling, and use Conditional Random Fields (Lafferty, 2001), which provides state-of-the-art performance in sequence labeling while allowing flexible feature design for combining local and global feature sets.

3.1 Local Features

Table 1 shows the local features used to train the error correction model.

Table 1: Local features for a verb phrase

name	description
t-learn	tense/aspect written by the learner (surface tense/aspect)
bare	the verb lemma
L	the word to the left
R	the word to the right
nsubj	nominal subject
dobj	direct object
aux	auxiliary verb
pobj	object of a preposition
p-tmod	temporal adverb
norm-p-tmod	normalized temporal adverb
advmod	other adverb
conj	subordinating conjunction
main-clause	true if the target VP is in main clause
sub-clause	true if the target VP is in subordinate clause

We use dependency relations such as **nsubj**, **dobj**, **aux**, **pobj**, and **advmod** for syntactic features. If a sentence including a target verb phrase is a complex sentence, we use the **conj** feature and add either the **main-clause** or the **sub-clause** feature depending on whether the target verb is in the main clause or in a subordinate clause. For example, the following two sentences have the same features although they have different structures.

(2) It pours when it rains.

(3) When it rains it pours.

In both sentences, we use the feature **main-clause** for the verb phrase *pours*, and **sub-clause** for the verb phrase *rains* along with the feature **conj:when** for both verb phrases.

Regarding **p-tmod**, we extract a noun phrase including a word labeled **tmod** (temporal adverb). For instance, consider the following sentence containing a temporal adverb:

(4) I had a good time last night.

In (4), the word *night* is the head of the noun phrase *last night* and is a temporal noun,[6] so we add the feature **p-tmod:last night** for the verb phrase *had*.

Additionally, **norm-p-tmod** is a normalized form of **p-tmod**. Table 2 shows the value of the feature **norm-p-tmod** and the corresponding temporal keywords. We use **norm-p-tmod** when **p-tmod**

[2] Konan-JIEM Learner Corpus Second Edition (`http://gsk.or.jp/catalog/GSK2011-B/catalog.html`) contains 170 essays, and Cambridge English First Certificate in English (`http://www.cambridgeesol.org/exams/fce/index.html`) contains 1244 essays.

[3] `http://lang-8.com/`

[4] As of January, 2012. More details about the Lang-8 corpus can be found in (Mizumoto et al., 2011).

[5] Note that not all the 750,000 verb phrases were corrected due to the misuse of tense/aspect.

[6] We made our own temporal noun list.

Table 2: The value of the feature **norm-p-tmod** and corresponding temporal keywords

temporal keywords	value
yesterday or *last*	past
now	present
tomorrow or *next*	future
today or *this*	this

Table 3: Feature templates

Local Feature Templates
<head> <head, t-learn> <head, L, R> <L> <L, head> <L, t-learn> <R> <R, head> <R, t-learn> <nsubj> <nsubj, t-learn> <aux> <aux, head> <aux, t-learn> <pobj> <pobj, t-learn> <norm-p-tmod> <norm-p-tmod, t-learn> <advmod> <advmod, t-learn> <tmod> <tmod, t-learn> <conj> <conj, t-learn> <main-clause> <main-clause, t-learn> <sub-clause> <sub-clause, t-learn> <conj, main-clause> <conj, sub-clause>
Global Context Feature Templates
<p-tmod′> <p-tmod′, t-learn> <p-tmod′, t-learn′> <p-tmod′, t-learn′, t-learn> <norm-p-tmod′> <norm-p-tmod′, t-learn> <norm-p-tmod′, t-learn′> <norm-p-tmod′, t-learn′, t-learn>

includes any temporal keywords. For instance, in the sentence (4), we identify *last night* as temporal adverb representing past, and thus create a feature **time:past** for the verb phrase *had*.

3.2 Feature Template

Table 3 shows feature templates. <*a*> represents a singleton feature and <*a, b*> represents a combination of features *a* and *b*. Also, *a*′ means the feature *a* of the preceding verb phrase. A local feature template is a feature function combining features in the target verb phrase, and a global context feature template is a feature function including features from a non-target verb phrase. Suppose we have following learner's sentences:

(5) I **went** to Kyoto yesterday.
I ***eat** yatsuhashi[7] and **drank** green tea.

In (5), the verb before *eat* is *went*, and **p-tmod:yesterday** and **norm-p-tmod:past** are added to the feature set of verb *went*. Accordingly, the global context features **p-tmod′:yesterday** and **norm-p-tmod′:past** are added to the verb *eat*.

[7] Yatsuhashi is a Japanese snack.

Table 4: Example of global context feature functions generated by feature templates

<p-tmod′:yesterday>
<p-tmod′:yesterday, t-learn′:simple past>
<p-tmod′:yesterday, t-learn:simple present>
<p-tmod′:yesterday, t-learn′:simple past, t-learn:simple past>
<norm-p-tmod′:past>
<norm-p-tmod′:past, t-learn′:simple past>
<norm-p-tmod′:past, t-learn:simple present>
<norm-p-tmod′:past, t-learn′:simple past, t-learn:simple present>

Table 4 lists all the global context features for the verb *eat* generated by the feature templates.

3.3 Trade-off between Precision and Recall

Use of surface tense/aspect forms of target verbs improves precision but harms recall. This is because in most cases the surface tense/aspect and the correct tense/aspect form of a verb are the same. It is, of course, desirable to achieve high precision, but very low recall leads to the system making no corrections. In order to control the trade-off between precision and recall, we re-estimate the best output label $\hat{y}$ based on the originally estimated label y as follows:

$$\hat{y} = \arg\max_{y} s(y)$$

$$s(y) = \begin{cases} \alpha c(y), & \text{if y is the same as learner's tense/aspect} \\ c(y) & \text{otherwise.} \end{cases}$$

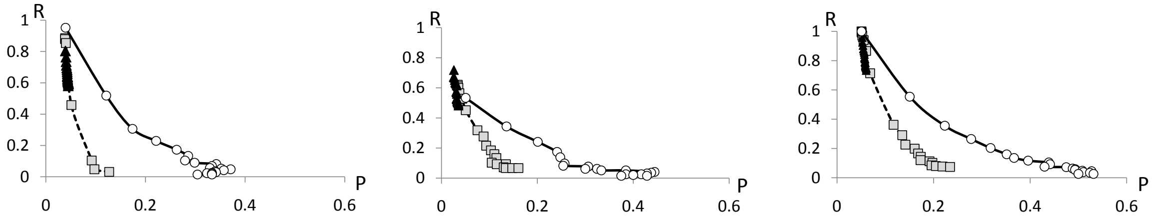
where $c(y)$ is the confidence value of y estimated by the originally trained model (explained in 4.3), and α ($0 \leq \alpha < 1$) is the weight of the surface tense/aspect.

We first calculate $c(y)$ of all the labels, and discount only the label that is the same as learner's tense/aspect, and finally we choose the best output label. This process leads to an increase of recall. We call this method **T-correction**.

4 Experiments

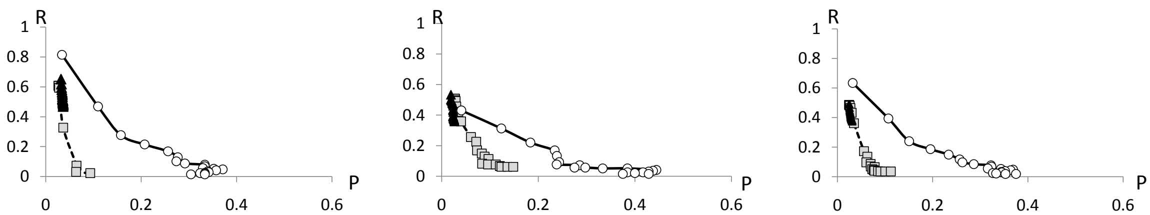
4.1 Data and Feature Extraction

We used the Lang-8 tense/aspect corpus described in Section 2. We randomly selected 100,000 entries for training and 1,000 entries for testing. The test



(a) tense (b) aspect (c) tense/aspect

Figure 1: Precision-Recall curve for error detection



(a) tense (b) aspect (c) tense/aspect

Figure 2: Precision-Recall curve for error correction

data includes 16,308 verb phrases, of which 1,072 (6.6%) contain tense/aspect errors. We used Stanford Parser 1.6.9 [8] for generating syntactic features and tense/aspect tagging.

4.2 Classifiers

Because we want to know the effect of using global context information with CRF, we trained a one-versus-rest multiclass SVM and a maximum entropy classifier (MAXENT) as baselines.

We built a SVM model with LIBLINEAR 1.8[9] and a CRF and a MAXENT model with CRF++ 0.54.[10] We use the default parameters for each toolkit.

In every method, we use the same features and feature described in Section 3, and use T-correction for choosing the final output. The confidence measure of the SVM is the distance to the separating hyperplane, and that of the MAXENT and the CRF is the marginal probability of the estimated label.

[8] http://nlp.stanford.edu/software/lex-parser.shtml

[9] http://www.csie.ntu.edu.tw/~cjlin/liblinear/

[10] http://crfpp.sourceforge.net/

5 Results

Figures 1 and 2 show the Precision-Recall curves of the error detection and correction performance of each model. The figures are grouped by error types: tense, aspect, and both tense and aspect. All figures indicate that the CRF model achieves better performance than SVM and MAXENT.

6 Analysis

We analysed the results of experiments with the α parameter of the CRF model set to 0.1. The most frequent type of error in the corpus is using simple present tense instread of simple past, with 211 instances. Of these our system detected 61 and successfully corrected 52 instances. However, of the second most frequent error type (using simple past instead of simple present), with 94 instances in the corpus, our system only detected 9 instances. One reason why the proposed method achieves high performance in the first type of errors is that tense errors with action verbs written as simple present are relatively easy to detect.

References

John Bitchener, Stuart Young, and Denise Cameron. 2005. The Effect of Different Types of Corrective Feedback on ESL Student Writing. *Journal of Second Language Writing*, 14(3):191–205.

Chris Brockett, William B. Dolan, and Michael Gamon. 2006. Correcting ESL Errors Using Phrasal SMT Techniques. In *Proceedings of COLING-ACL*, pages 249–256.

Martin Chodorow, Joel R. Tetreault, and Na-Rae Han. 2007. Detection of Grammatical Errors Involving Prepositions. In *Proceedings of ACL-SIGSEM*, pages 25–30.

Kevin Knight and Ishwar Chander. 1994. Automated Postediting of Documents. In *Proceedings of the AAAI'94*, pages 779–784.

John Lafferty. 2001. Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data. In *Proceedings of ICML*, pages 282–289.

John Lee and Stephanie Seneff. 2006. Automatic Grammar Correction for Second-Language Learners. In *Proceedings of the 9th ICSLP*, pages 1978–1981.

John Lee and Stephanie Seneff. 2008. Correcting Misuse of Verb Forms. In *Proceedings of the 46th ACL:HLT*, pages 174–182.

Tomoya Mizumoto, Mamoru Komachi, Masaaki Nagata, and Yuji Matsumoto. 2011. Mining Revision Log of Language Learning SNS for Automated Japanese Error Correction of Second Language Learners. In *Proceedings of 5th IJCNLP*, pages 147–155.

Ryo Nagata and Kazuhide Nakatani. 2010. Evaluating Performance of Grammatical Error Detection to Maximize Learning Effect. In *Proceedings of COLING*, pages 894–900.

Alla Rozovskaya and Dan Roth. 2011. Algorithm Selection and Model Adaptation for ESL Correction Tasks. In *Proceedings of the 49th ACL:HLT*, pages 924–933.