Grammatical Error Detection and Correction Using Tagger Disagreement

Anubhav Gupta
Université de Franche-Comté
anubhav.gupta@edu.univ-fcompte.fr

Abstract

This paper presents a rule-based approach for correcting grammatical errors made by non-native speakers of English. The approach relies on the differences in the outputs of two POS taggers. This paper is submitted in response to CoNLL-2014 Shared Task.

1 Introduction

A part-of-speech (POS) tagger, like any other software, has a set of inputs and outputs. The input for a POS tagger is a group of words and a tagset, and the output is a POS tag for a word (Jurafsky and Martin, 2009). Given that a software is bound to provide incorrect output for an incorrect input (garbage in, garbage out), it is quite likely that taggers trained to tag grammatically correct sentences (the expected input) would not tag grammatically incorrect sentences properly. Furthermore, it is possible that the output of two different taggers for a given incorrect input would be different.

For this shared task, the POS taggers used were the Stanford Parser, which was used to preprocess the training and test data (Ng et al., 2014) and the TreeTagger (Schmid, 1994). The Stanford Parser employs unlexicalized PCFG[1] (Klein and Manning, 2003), whereas the TreeTagger uses decision trees. The TreeTagger is freely available[2], and its performance is comparable to that of the Stanford Log-Linear Part-of-Speech Tagger (Toutanova et al., 2003). Since the preprocessed dataset was already annotated with POS tags, the Stanford Log-Linear POS Tagger was not used.

If the annotation of preprocessed data differed from that of the TreeTagger, it was assumed that the sentence might have grammatical errors. Once an error was detected it was corrected using the Nodebox English Linguistics library[3] (De Bleser et al., 2002).

2 Error Detection

The POS tag for each token in the data was compared with the tag given by the TreeTagger. Sentences were considered grammatically incorrect upon meeting the following conditions:

- The number of tags in the preprocessed dataset for a given sentence should be equal to the number of tags returned by the TreeTagger for the same sentence.
- There should be at least one token with different POS tags.

As an exception, if the taggers differed only on the first token, such that the Stanford Parser tagged it as **NNP** or **NNPS**, then the sentence was not considered for correction, as this difference can be attributed to the capitalisation of the first token, which the Stanford Parser interprets as a proper noun.

Table 1 shows the precision (P) and the recall (R) scores of this method for detecting erroneous sentences in the training and test data. The low recall score indicates that for most of the incorrect sentences, the output of the taggers was identical.

2.1 Preprocessing

The output of the TreeTagger was modified so that it had the same tag set as that used by the Stanford Parser. The differences in the output tagset is displayed in the Table 2.

2.2 Errors

Where the mismatch of tags is indicative of error, it does not offer insight into the nature of the error and thus does not aid in error correction per se. For example, the identification of a token as VBD

[1] Probabilistic Context-Free Grammar
[2] http://www.cis.uni-muenchen.de/∼schmid/tools/TreeTagger/
[3] http://nodebox.net/code/index.php/Linguistics

Proceedings of the Eighteenth Conference on Computational Natural Language Learning: Shared Task, pages 49–52,
Baltimore, Maryland, 26-27 July 2014.

Dataset	Total Erroneous Sentences	Sentences with Tag Mismatch	Erroneous Sentences Identified Correctly	P	R
Training	21860	26282	11769	44.77	53.83
Test	1176	642	391	60.90	33.24
Test (Alternative)†	1195	642	398	61.99	33.30

† consists of additional error annotations provided by the participating teams.

Table 1: Performance of Error Detection.

TreeTagger Tagset	Stanford Parser Tagset
(	-LRB-
)	-RRB-
NP	NNP
NPS	NNPS
PP	PRP
SENT	.

Table 2: Comparison of Tagsets.

(past tense) by one tagger and as VBN (past participle) another does not imply that the mistake is necessarily a verb tense (Vt) error. Table 4 lists some of the errors detected by this approach.

3 Error Correction

Since mismatched tag pairs did not consistently correspond to a particular error type, not all errors detected were corrected. Certain errors were detected using hand-crafted rules.

3.1 Subject-Verb Agreement (SVA) Errors

SVA errors were corrected with aid of dependency relationships provided in the preprocessed data. If a singular verb (VBZ) referred to a plural noun (NNS) appearing before it, then the verb was made plural. Similarly, if the singular verb (VBZ) was the root of the dependency tree and was referred to by a plural noun (NNS), then it was changed to the plural.

3.2 Verb Form (Vform) Errors

If a modal verb (MD) preceded a singular verb (VBZ), then the second verb was changed to the bare infinitive form. Also, if the preposition **to** preceded a singular verb, then the verb was changed to its bare infinitive form.

3.3 Errors Detected by POS Tag Mismatch

If a token followed by a noun is tagged as an adjective (JJ) in the preprocessed data and as an adverb (RB) by the TreeTagger, then the adverbial morpheme **-ly** was removed, resulting in the adjective. For example, **completely** is changed to **complete** in the second sentence of the fifth paragraph of the essay 837 (Dahlmeier et al., 2013). On the other hand, adverbs (RB) in the preprocessed dataset that were labelled as adjectives (JJ) by the TreeTagger were changed into their corresponding adverbs.

Dataset	P	R	$F_{\beta=0.5}$
Training	23.89	0.31	1.49
Test	70.00	1.72	7.84
Test (Alternative)	72.00	1.90	8.60

Table 3: Performance of the Approach.

A token preceded by the verb **to be**, tagged as JJ by the Stanford Parser and identified by the TreeTagger as a verb is assumed to be a verb and accordingly converted into its past participle. Finally, the tokens labelled NNS and VBZ by the Stanford Parser and the TreeTagger respectively are likely to be Mec[4] or Wform[5] errors. These tokens are replaced by plural nouns having same initial substring (this is achieved using the **get_close_matches** API of the difflib Python library).

The performance of this approach, as measured by the M2 scorer (Dahlmeier and Ng, 2012), is presented in Table 3.

4 Conclusion

The approach used in this paper is useful in detecting mainly verb form, word form and spelling errors. These errors result in ambiguous or incorrect input to the POS tagger, thus forcing it to produce incorrect output. However, it is quite likely that with a different pair of taggers, different rules

[4]Punctuation, capitalisation, spelling, typographical errors

[5]Word form

nid	829					
Sentence	This	caused	problem	like	the	appearance
Stanford Parser	DT	**VBD**	NN	IN	DT	NN
TreeTagger	DT	**VBN**	NN	IN	DT	NN
Error Type	Vt					
nid	829					
Sentence	but	also	to	reforms	the	land
Stanford Parser	CC	RB	TO	**VB**	DT	NN
TreeTagger	CC	RB	TO	**NNS**	DT	NN
Error Type	Wci					
nid	840					
Sentence	India	,	their	population	amount	to
Stanford Parser	NNP	,	PRP$	NN	**VB**	TO
TreeTagger	NNP	,	PRP$	NN	**NN**	TO
Error Type	Vform	(This was not an error in the training corpus.)				
nid	1051					
Sentence	Singapore	is	currently	a	develop	country
Stanford Parser	NNP	VBZ	RB	DT	**JJ**	NN
TreeTagger	NNP	VBZ	RB	DT	**VB**	NN
Error Type	Vform					
nid	858					
Sentence	Therefore	most	of	China	enterprisers	focus
Stanford Parser	RB	JJS	IN	NNP	**VBZ**	NN
TreeTagger	RB	RBS	IN	NNP	**NNS**	VBP
Error Type	Wform					
nid	847					
Sentence	and	social	constrains	faced	by	engineers
Stanford Parser	CC	JJ	**NNS**	VBN	IN	NNS
TreeTagger	CC	JJ	**VBZ**	VBN	IN	NNS
Error Type	Mec					

Table 4: Errors Detected.

would be required to correct these errors. Errors concerning noun number, determiners and prepositions, which constitute a large portion of errors committed by L2 learners (Chodorow et al., 2010; De Felice and Pulman, 2009; Gamon et al., 2009), were not addressed in this paper. This is the main reason for low recall.

Acknowledgments

I would like to thank Calvin Cheng for proofreading the paper and providing valuable feedback.

References

Martin Chodorow, Michael Gamon, and Joel Tetreault. 2010. The Utility of Article and Preposition Error Correction Systems for English Language Learners: Feedback and Assessment. *Language Testing* 27 (3): 419–436. doi:10.1177/0265532210364391.

Daniel Dahlmeier, Hwee Tou Ng, and Siew Mei Wu. 2013. Building a Large Annotated Corpus of Learner English: The NUS Corpus of Learner English. *Proceedings of the Eighth Workshop on Innovative Use of NLP for Building Educational Applications (BEA 2013).* 22 –31. Atlanta, Georgia, USA.

Daniel Dahlmeier and Hwee Tou Ng. 2012. Better Evaluation for Grammatical Error Correction. *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL 2012).* 568 – 572. Montreal, Canada.

Frederik De Bleser, Tom De Smedt, and Lucas Nijs 2002. NodeBox version 1.9.6 for Mac OS X.

Rachele De Felice and Stephen G Pulman. 2009. Automatic Detection of Preposition Errors in Learner Writing. *CALICO Journal* 26 (3): 512–528.

Michael Gamon, Claudia Leacock, Chris Brockett, William B. Dolan, Jianfeng Gao, Dmitriy Belenko,

and Alexandre Klementiev. 2009. Using Statistical Techniques and Web Search to Correct ESL Errors. *CALICO Journal* 26 (3): 491–511.

Daniel Jurafsky and James H. Martin. 2009. Part-of-Speech Tagging. *Speech and Language Processing: An Introduction to Natural Language Processing, Speech Recognition, and Computational Linguistics*. Prentice-Hall.

Dan Klein and Christopher D. Manning. 2003. Accurate Unlexicalized Parsing. *Proceedings of the 41st Meeting of the Association for Computational Linguistics.* 423–430.

Hwee Tou Ng, Siew Mei Wu, Ted Briscoe, Christian Hadiwinoto, Raymond Hendy Susanto, and Christopher Bryant. 2014. The CoNLL-2014 Shared Task on Grammatical Error Correction. *Proceedings of the Eighteenth Conference on Computational Natural Language Learning: Shared Task (CoNLL-2014 Shared Task)*. Baltimore, Maryland, USA.

Helmut Schmid. 1994. Probabilistic Part-of-Speech Tagging Using Decision Trees. *Proceedings of International Conference on New Methods in Language Processing.* Manchester, UK.

Kristina Toutanova, Dan Klein, Christopher Manning, and Yoram Singer. 2003. Feature-Rich Part-of-Speech Tagging with a Cyclic Dependency Network. *Proceedings of HLT-NAACL 2003.* 252–259.