Grammatical Error Correction (GEC): Research Approaches till now

Sagar Ailani
Department of Information
Technology
K.J. Somaiya
College of Engineering
Mumbai, India

Ashwini Dalvi
Department of
Information Technology
K.J. Somaiya
College of Engineering
Mumbai, India

Irfan Siddavatam
Department of
Information Technology
K.J. Somaiya
College of Engineering
Mumbai, India

ABSTRACT
People all around the world attempts various entrance-level exams to test their proficiency in English. Most of these people are English as a Second Language (ESL) learners & some assistance from a teacher can be helpful in their progress, as they can provide continuous feedback on their writings. This process can be automated, and the process of detecting and correcting grammar errors in a text is called Grammatical Error Correction (GEC) in "Natural Language Processing" Domain. Three approaches have been used for solving GEC task, namely "Rule-based", "Classification-based", & "Machine Translation", with Machine Translation further broken down into "Statistical Machine Translation" and "Neural Machine Translation".

The paper states how the researchers have used these approaches, and what setbacks or improvements in technologies paved the way for using better and more advanced approaches to GEC. It also states the understanding on how the GEC task can be improved, and what will be the tradeoffs for achieving higher performance in the future.

General Terms
English as a Second Language (ESL), Grammatical Error Correction (GEC), Machine Translation, Natural Language

Keywords
Grammatical Error Correction (GEC), Machine Translation, Natural Language.

1. INTRODUCTION
Every year millions of people all around the world that do not speak English natively, try to learn English as a Second Language (ESL). These people often make word choices, grammar errors or syntactic errors which are influenced by their first language or mother tongue. Evidently, a proficient teacher can help an ESL Learner by giving them continuous feedback on their writing, but correcting grammatical errors manually is a rather rigorous and monotonous task. The learning process can be improved by developing an automated system that points out the grammatical mistakes in writing and further improves them. And this task is referred as Grammatical Error Correction (GEC). The automation of this task will benefit a lot of people around the world, by providing them continuous and instantaneous feedback on their writing.

A practical GEC System takes input as text, analyses the text to identify any errors within the text, & it has to keep the original meaning of the sentence intact, while correcting those errors. The task of GEC is not relatively new to Natural Language Processing area of research, and it has been continuously part of research and it was started as early as 1982. The research conducted all over these years can be divided into three approaches, Rule based Approach, Classifier based Approach, and Machine Translation Approach, with "Machine Translation" Approach can be further divided into "Statistical Machine Translation", and "Neural Machine Translation".

2. DATASETS
Learner Corpora that are used for training models for GEC can be broadly categorized into two types. Error-coded text comprises of learner text that contains error(s) which are generally coded by human annotators. And the second type of datasets are parallel datasets, which contain the original text which generally has some errors and a corrected version of the text, without explicitly mentioned errors.

Some of the publicly available datasets are NUS Corpus of Learner English (NUCLE; 59k sentence pairs) [1], Another dataset widely used is Lang-8 Corpus of Learner English, which was aggregated from an online forum, and then those sentences were corrected by members of the lang-8 community.

3. EVALUATION METRICS
Evaluation Metrics are used to assess the performance of the developed system. When evaluating a model, the system's output is compared against predefined standards by human experts. There are three widely used metrics; BLEU, GLEU, and M2 Scorer Bilingual Evaluation Understudy (BLEU) evaluates by measuring the closeness between the machine translation and human translation. BLEU requires only the corrected version of the sentence, and a numerical translation closeness metric [2].
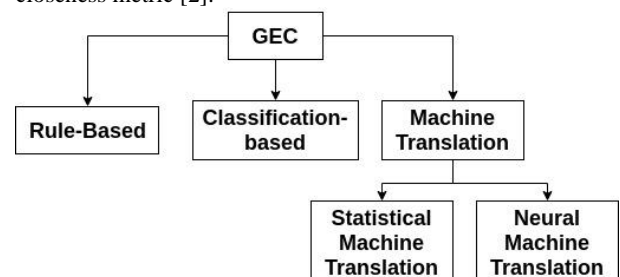

Figure 1: Approaches to GEC

Generalized Language Evaluation Understanding (GLEU) which takes into account both versions of the sentence, and it correlates much better with human ranking [3].

Max-Match (M2) Scorer states that there are multiple ways to

arrive at the same correction, and the system should be evaluated based on the set of edits that match the gold-standard as often as possible [4]. Here gold-standard implies the best possible way to correct the sentence.

4. CASE STUDY
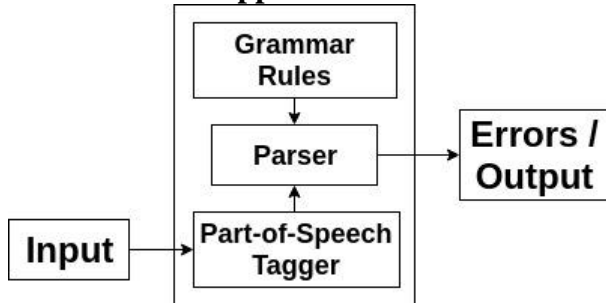
4.1 Rule based Approach

Figure 2: Traditional Rule based System Components

This approach mainly focuses on specifying hand-coded grammar rules, which the sentences must follow. It also involved simple pattern matching [5].

Syntactic Analysis was also incorporated with Rule Based System. Context Free Grammars are also used to specify grammar rules, and the parser is used to check the PoS tagged text according to the grammar rules defined [6].

Although rule-based systems are easy to implement, they are unable to detect more complex errors in writings, and also the model does not generalize well, because it becomes impossible to define rules for all combination of errors.
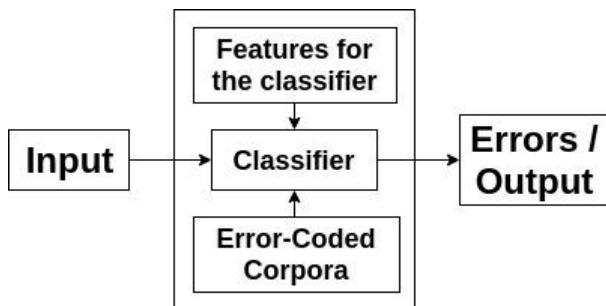
4.2 Classification based Approach

Figure 3: Components of a Classifier based System

The availability of large error-coded corpus enabled the researchers to use more data-driven approaches for GEC. Machine Learning Algorithms were used to build classifiers for correcting specific error types. Maximum entropy model was used to determine the best possible word/replacement with respect to the prior data [7].

In this classifier based approach, the possible candidates i.e. words/replacements are treated as class labels, and the surrounding n-grams, PoS tags, grammatical relations are used as features. These classifiers were used to detect article errors and achieved an accuracy of 88% [7]. Also because the features for the classifier are dependent upon the error type, a classifier can detect only a single type of error. And it assumes that the rest of the sentence is error-free and the current error is independent, which is usually not the case.

The commonly used approach to overcome this is to build multiple classifiers, each correcting one type of error and this collection of classifiers is then cascaded into a pipeline [8].

But this approach does not work well in case of dependent errors.

To solve the problem of dependent errors, researchers combined the Classification Based Approach and the Statistical Machine Translation Approach [9]. Using this method, the original sentence is decoded into multiple possible sentence corrections (also called hypothesis) with the purpose of finding the best sentence correction. This decoding process is carried out in iterations, and in each iteration the sentence from the previous iteration is improved by making incremental changes, and sentences are kept in the next iteration based on the score of current iteration. The decoder uses grammatical correctness and fluency to score sentences. The process is repeated until there are no sentences left or the max number of iterations is reached [9].

4.3 Machine Translation Approach
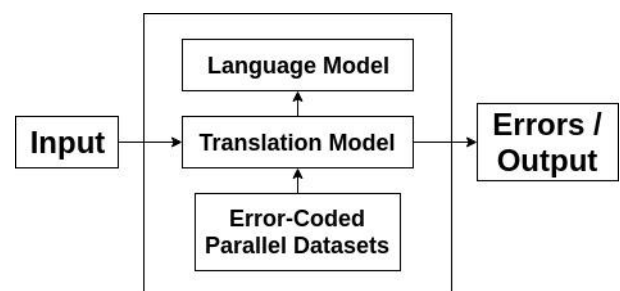
Statistical Machine Translation

Figure 4: Components of a Statistical Machine Translation System

SMT Models were first used in 2006, correcting a set of 14 countable/uncountable errors. These errors were deliberately introduced and the resulting dataset was called Chinese Learner Error Corpus (CLEC) [11]. Prior to this use of SMT in GEC research was largely held back because of lack of parallel error-annotated datasets.

In 2012, SMT models were used to solve all types of errors, researchers attempted to use a large scale learner corpus, and tried to identify the effect of learner corpus size on the phrasal-based SMT approach [10]. The results showed that the errors can be classified in two types; Errors that can be solved more efficiently by increasing corpus size, and the second type of errors were dependent on the long-range contextual information [10].

In 2014, A hybrid approach was used which incorporated a "Rule-based system" and an ad-hoc "SMT system". The rule based system generates a set of possible candidates, and then language model is used to find the probability of each of those possible candidates and selects the one with the highest probability [12].

Neural Machine Translation
NMT for GEC was first proposed in 2016 by Yuan and Briscoe [13]. NMT systems are developed using an "Encoder-Decoder" mechanism, where an encoder reads a sentence and encodes it into a vector, and the decoder outputs a translation by predicting the next word based on the encoded vector, and all previously predicted words [13]. Yuan and Briscoe used a Recursive Neural Network, which contained a "bi-directional Recursive Neural Network" as an encoder, and an "Attention-based" model as a decoder [13].

In 2018, Chollampatt and Ng [14] used a "multilayer

convolutional encoder-decoder neural network", they stated that CNN are more effective in capturing local context than RNN, whereas the multiple layers of convolution help capture wider contexts, the number of convolution layers is 7 for both encoder and decoder [14].

Also in 2018, Roman and Marcin combined SMT and NMT approaches to achieve state-of-the-art results [15]. They used a phrasal-based SMT system, and 5-gram Language Model. For NMT System. an attentional Encoder-Decoder model with a bidirectional single-layer encoder and decoder is used. Both encoder-decoder use Gated Recurrent Units as their RNN variant [15]. The output corrected by SMT system is passed as an input to NMT model. This 'pipelining' of both these models improves performance by increasing recall [15].

5. RESEARCH GAP

All the recent research that has accomplished good results have used parallel datasets for training their models. Research revealed how availability of parallel datasets let the researchers to try SMT approach for solving GEC. The performance of the models can be improved by using larger error-annotated parallel datasets. Although it is not possible to cover every possible combination of errors in these datasets, as the language is quite diverse & depends on the expressiveness of the user, but performance can still be improved by adding more parallel data to solve the problem.
NMT models have achieved better accuracy percentages than other approaches, but it has been possible by using high-end GPUs and long hours of training time. SMT's struggled to get wider contextual information about the error, and NMTs use multiple convolutional layers to find that wider context, and in turn consume more resources to produce results. So in this case, better results can be achieved, but the tradeoff is you require high-end resources.

6. CONCLUSION

The paper describes all the approaches that were used to solve GEC, stated the main reasons of that research, and why better methods were needed and what enabled the use of those methods. Research also found availability of large datasets was pivotal for the researchers to use statistical methods, and then the availability of large computational resources paved the way for using neural models.

7. REFERENCES

[1] Dahlmeier, D., Ng, H. T., and Wu, S. M., "Building a large annotated corpus of learner english: The nus corpus of learner english.." in BEA@ NAACL-HLT, pp. 22–31, 2013.

[2] Papineni, K., Roukos, S., Ward, T., and Zhu, W.-J., "Bleu: a method for automatic evaluation of machine translation," in Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics, pp. 311–318, 2002.

[3] Napoles, C., Sakaguchi, K., Post, M., and Tetreault, J., "Ground truth for grammatical error correction metrics," in Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing, Vol. 2, pp. 588–593, 2015.

[4] Dahlmeier, D., and Ng, H. T., "Better evaluation for grammatical error correction," in Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Association for Computational Linguistics). pp. 568–572, 2012

[5] Bustamante, F. R., and León, F. S., "Gramcheck: A grammar and style checker," in Proceedings of the 16th conference on Computational linguistics-Volume 1 (Association for Computational Linguistics). pp. 175–181, 1996.

[6] Ghosalkar, P., Malagi, S., Nagda, V., Mehta, Y., & Kulkarni, P,. English Grammar Checker, 2016.

[7] Han, N.-R., Chodorow, M., and Leacock, C., "Detecting errors in english article usage with a maximum entropy classifier trained on a large, diverse corpus.." in LREC, 2004.

[8] Rozovskaya, A., Chang, K.-W., Sammons, M., and Roth, D., "The university of illinois system in the conll-2013 shared task," in Proceedings of the Seventeenth Conference on Computational Natural Language Learning: Shared Task, pp. 13–19, 2013.

[9] Dahlmeier, D., and Ng, H. T., "A beam-search decoder for grammatical error correction," in Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (Association for Computational Linguistics). pp. 568–578, 2012.

[10] Mizumoto, T., Hayashibe, Y., Komachi, M., Nagata, M., and Matsumoto, Y., 2012, "The effect of learner corpus size in grammatical error correction of esl writings," Proceedings of COLING 2012: Posters, 863–872.

[11] Brockett, C., Dolan, W. B., and Gamon, M., "Correcting esl errors using phrasal smt techniques," in Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics (Association for Computational Linguistics). pp. 249–256, 2006.

[12] Felice, M., Yuan, Z., Andersen, Ø. E., Yannakoudakis, H., and Kochmar, E., "Grammatical error correction using hybrid systems and type filtering.." in CoNLL Shared Task, pp. 15–24, 2014.

[13] Yuan, Z., and Briscoe, T., "Grammatical error correction using neural machine translation" in HLT-NAACL, pp. 380–386, 2016.

[14] Chollampatt, S. and Ng, H.T., April "A multilayer convolutional encoder-decoder neural network for grammatical error correction." In Thirty-Second AAAI Conference on Artificial Intelligence, 2018.

[15] Grundkiewicz, R. and Junczys-Dowmunt, M., "Near human-level performance in grammatical error correction with hybrid machine translation." arXiv preprint arXiv:1804.05945, 2018.