

# Investigation into English Grammar

Martha Bellows

Antoinette Bongiorno

Brandon DiGiulio

May 7, 2017

## **1 Abstract**

Summary of paper.

## 2 Introduction

Discuss previous paper and how we feel native English speakers grading sentences is not a good metric (reproducibility, subjectivity, credentials, etc.). How we aim to solve this issue in our paper and brief overview of our process.

### 2.1 Motivation

"We invited 3 English native speakers to do this evaluation." – original paper

- Subjective
- Unreplicable
- Inconsistent

We recognize that there is a need for a more objective method of scoring the grammaticality of sentences in order to further research into the growing field of Automatic Text Summarization. Grammaticality evaluation will help rank competing systems as well as provide developers of new solutions with a means to measure the effectiveness of their systems.

Some examples of citing formatting.

Many hospitals face significant budgetary concerns and, due to Medicare regulations, excessive readmission rates reduce Medicare payments. In the United States, readmissions cost hospitals an average of 17.4 billion dollars per year [1]. These attempts have varied in success with results averaging around 60 percent accuracy by using raw data [3].

### 2.2 NLP Community

- What is the community standard?
  - ROUGE metric
  - Human Judgement
- How is human judgement used?

- Subjective scoring on 1-5 scale
- grammaticality, non-redundancy, clarity, coherence
- Often random participants, native English speakers
- DUC conference uses a set of questions for assessors
- Why hasn't a more objective method been used before?
  - There are commercial solutions to grammar checking
  - Aside from that, the NLP community has not adopted a more objective method
  - Very few instances of previous research, Quantitative evaluation of grammaticality of summaries (Ravikiran Vadlapudi and Rahul Katragadda, 2010)

## 2.3 Grammar

What is grammar?

The system of rules and syntax that defines how things should be written, spoken.

Gives communication an understood, defined meaning between two or more parties.

## 3 Parts of Speech Tagging

Our idea is to develop a way to score the grammaticality of an input sentence based on the comparison of that sentence's POS tag sequence to a generated grammar rule set.

This grammaticality scoring method could someday replace human judgement in Automatic Text Summarization evaluation.

### 3.1 Natural Language Toolkit

Natural language processing with Python.

Excellent documentation at <http://www.nltk.org/book/>

Helps to tokenize and tag sentences.

Based on Penn Treebank POS tags.

## 3.2 NLTK - Tagging Sentences

Initial Sentence: "I went on a walk. It was nice outside."

Tokenize Sentence: ['I went on a walk.', 'It was nice outside.']

Tokenize Words: [['I', 'went', 'on', 'a', 'walk', '.'], ['It', 'was', 'nice', 'outside', '.']]

Tag Words in Sentence: [('I', 'PRP'), ('went', 'VBD'), ('on', 'IN'), ('a', 'DT'), ('walk', 'NN'), ('.', '.')] ]

## 3.3 Project Gutenberg

Used literature from Gutenberg as sentences to tag. Pride and Prejudice by Jane Austen Alice's Adventures in Wonderland by Lewis Carroll Moby Dick by Herman Melville The Picture of Dorian Gray by Oscar Wilde Assuming correct grammar.

# 4 Algorithm

## 4.1 Generating Rule Set

1. Import list of grammatically correct sentences
2. Tokenize sentences
3. Tag sentences - generate POS tags
4. Extract POS tag sequences into txt file

## 4.2 Grading Against Rule Set

To score an input sentence or sentences:

1. Read sentence(s)/summary
2. Tokenize sentence(s)
3. Tag sentence(s)
4. Extract POS tag sequences

5. Compare tag sequence generated from input sentence to existing rule set
6. Score 1 = good grammar (found in rule set), Score 0 = bad grammar (not found)

## 5 Results

What results are we expecting?

0 for bad grammatical sentence

1 for good grammatical sentence

What have we seen?

Works for sentences/structures we know are in the text file.

Does not work on all grammatically correct sentences...yet

## 6 Limitations

Infiniteness of language

Text file becoming too big

Limitations on books used

## 7 Conclusion

Conclusion

## 8 Future Work

Future work

## References

- [1] A. Catlin, C. Cowan, M. Hartman, and S. Heffler, “National health spending in 2006: A year of change for prescription drugs,” *Health Affairs*, vol. 27, no. 1, pp. 14–29, 2008.
- [2] A. Hosseinzadeh, M. Izadi, A. Verma, D. Precup, and D. Buckeridge, “Assessing the predictability of hospital readmission using machine learning,” in *Proceedings of the Twenty-Fifth Innovative Applications of Artificial Intelligence Conference*, pp. 1532–1538, AAAI, 2013.
- [3] D. Kansagara *et al.*, “Risk prediction models for hospital readmission: A systematic review,” *Journal of the American Medical Association*, vol. 306, no. 15, p. 16881698, 2011.