Text Logical Scoring Based on Knowledge Graph

Yewei Song[1[0000−0002−6314−7515]]

University of Birmingham, Birmingham B15 2TT, UK
yxs987@student.bham.ac.uk

Abstract. This article proposes to use the scale of the knowledge graph as a criterion for judging the logic of essay, and summarizes the methods for constructing the knowledge graph from essay and the problems that may be encountered in the scoring process.

Keywords: Knowledge Graph · Text Scoring · Coreference Resolution

1 Project Category:

Theory(Computational Linguistics)

2 Project Aim:

My aim in this project is to explore a feasible method to judge whether essay is logical. Here the author proposes a hypothesis based on the concept of knowledge graph, that is, if the graph contains more items linked by logical relationships, then the corresponding essay has a clearer logical expression. It can also be said that using the construction of knowledge graph to measure the logic of essay.
The mainstream Automated Essay Scoring system cannot judge whether the logic of essay is smooth. Their judgments are based on vocabulary and grammatical perspectives. Even if the relationship between word vectors is considered, there is a lack of evaluation of the semantic network. If the relationship between knowledge graph and essay logic can be confirmed, then AES technology will have a significant accuracy improvement and it is more difficult to be deceived.
AES technology involves the field of computer science in natural language processing. The author's assumptions include pattern recognition, syntactic structure analyse, and semantic networks.

3 Related Work:

3.1 Old systems

Essay score is a very challenging natural language processing application scenario. Existing AES systems such as PEG (Project Essay Grader), IEA (Intelligent Essay Assessor), E-rater (Electronic Essay Rater), IntelliMetric and Writing Roadmap, etc[11]. However, most of the essay scoring systems are based on the error rate in Grammar, Usage, Mechanics, and Style Measures. Even if these systems use word frequency, vocabulary richness, grammatical richness and other indicators to calculate, there is a lack of measurement for high-level writing logic and expression coherence.[3]

3.2 Pre-processing

Sentence segmentation, Tokenization is the first two steps before handle text content. Sentence segmentation is to identify sentence boundaries in a given text, that is, the end of one sentence and the beginning of another sentence. Usually use "." symbol to split. Tokenization is to recognize different words, numbers and other punctuation marks. In English, a space is used to separate two words[9]. There are many effective tools available for these two steps.

3.3 Stemming or Lemmatisation

As a kind of synthetic language, English has inflection phenomenon, including past tense, present participle, plural, comparative, etc. The purpose of Stemming or Lemmatisation is to unify words with different appearances but the same meaning, which is convenient for subsequent processing and analysis. Stemming is the process of removing the prefix and suffix of a word to get the root. Lemmatisation is to transform the complex form of words into the most basic form.[1]

3.4 Part-of-speech Tagging

Part-of-speech tagging(POS tagging) is the process of marking up a word in a text (corpus) as corresponding to a particular part of speech.A simplified form in the identification of words as nouns, verbs, adjectives, adverbs, etc. Part-of-speech tagging is harder than just having a list of words and their parts of speech, because some words can represent more than one part of speech at different times, and because some parts of speech are complex or unspoken.[8]

3.5 Name Entity Recognition

Common triples extraction includes many steps, First we need to perform Name Entity Recognition(NER) on the text. A named entity refers to an entity with a specific meaning or strong reference in the text. The early recognition methods of named entities are basically based on rules.[4] Later, because statistical methods based on large-scale corpora have achieved good results in all aspects of natural language processing, a large number of machine learning methods have also appeared in named entity class recognition tasks. These methods can be divided into supervised learning, semi-supervised learning, unsupervised learning and mixed models.

3.6 Parsing and Coreference Resolution

Another important issue of semantic analysis is Coreference Resolution and Parsing. Referential resolution is to find out the real world things referred to by noun phrases in the text. Not only pronouns can refer to other things, possessives and other noun phrases are also possible, there are even a large number of nested references.[5] To get the relationship between words in essay, syntactic parsing is an important part.It is the process of analyzing the input text sentence to get the syntactic structure of the sentence. Because the purpose is to obtain dependencies, choose dependency-based parse trees.[10] The explanation on Wikipedia is as follows: *The dependency-based parse trees of dependency grammars see all nodes as terminal, which means they do not acknowledge the distinction between terminal and non-terminal categories. They are simpler on average than constituency-based parse trees because they contain fewer nodes.*

3.7 Knowledge Graph Build

The knowledge graph is a directed acyclic graph(DAG), describing the objective relationship between things, concepts and events. Each edge can be understood as subject-predicate-object triples. These relationships can be extracted from the text content of essay through NLP. We use triples whose endpoints are synonymous, as the edges of the DAG. And use the dependency-based syntactic relationship as the direction of the edges.[12]
A question worth discussing is how to perform relational reasoning on the extracted knowledge graph in essay scoring. The first is whether you need to do relational reasoning, academic writing with clear and sufficient logic, and there should be links between entities. Secondly, short essays are often difficult to provide enough context for speculation.

3.8 Tool Selection

Tools mainly include Stanford coreNLP and its toolkit like OpenIE. We also like to try nltk and other tools to compare the performance.[7]

4 Project Objectives:

- Discuss the logic deficiency of the existing AES method
- Discuss the relationship between the logic of essay and the scale of knowledge graph
- Successfully collected essay examples of different writing levels
- Solve synonym matching problem and knowledge reasoning problem
- Experiment can successfully extract knowledge graph from essay
- The knowledge graph extracted by experiment can effectively reflect the logical connection of various items in essay

- There are significant differences in the size of the knowledge graph between the various groups
- Interpret the experimental results and propose existing problems and possible solutions

5 Methodology:

Part 1: Formal. Based on literature review and qualitative analysis support, conduct supportive discussions on the subject and propose critical analysis. Each statement in a dissertation must be correct and defensible in a logical and scientific sense. Moreover, the discussions in a dissertation must satisfy the most stringent rules of logic applied to mathematics and science.
Part 2: Experimental. A well-designed experiment will start with a list of the questions that the experiment is expected to answer.[2] This experiment is based on pre-grouped data, and the score is calculated according to the algorithm. Finally, the effectiveness of the algorithm is analyzed from the score distribution and specific cases.

6 Project plan:

Week	Plans
1	Writing Research Proposal
2	Determine the project plan, collect and organize the literature, and confirm the required resources
3	Collect experimental data and read literature
4	Start writing a literature review
5	Complete literature review and design experiments
6	Discuss theoretical assumptions and write code
7	Prepare the experiment environment and experiment
8	Complete the experiment and collect data, complete the writing of the experimental part
9	Analyze experimental results and propose critical analysis
10	Complete the writing of the full text and submit as a first draft
11	Correct the insufficient parts, discuss the results and complete the conclusion
12	Proofreading and submit
13	Reserved time

7 Risks and contingency plan:

7.1 Risk 1: Synonym recognition is difficult to complete

The same entity in the text is likely to be expressed by different words, in some cases there will be special reference. Therefore, it is likely that two substantially linked knowledge networks are disconnected because they fail to resolve synonyms. This article does not focus on the problem of referential digestion, so before batch experiments, some essays will be selected for individual analysis. At the same time, consider using word vectors for equivalent treatment of synonyms. The thesaurus of synonyms can also be imported into nltk or similar tools.

7.2 Risk 2: The lack of general knowledge

Many times we rely on common sense when writing, this can reduce the redundancy in the article. If we do not cite external common sense, the knowledge graph parsed from essay is likely to be broken or fragmented. The author here proposes a concept: you can use the relevant text of the "question" of the writing task to extract triples for KG construction.

7.3 Force Majeure Risk

For computer losk or broken, I have two personal commputer, and backup all data both on iCloud and Google Drive. Also I write all text on overleaf platform. For illness and other personal risk, I will consider applying for extension or changing the plan.

8 Hardware/Software Resources:

Since most of the tools for text extraction do not require the use of GPU, the author's personal computer is used. The performance parameters are as follows:
- Intel Core i7 8809G 4C8T 3.10 GHz CPU
- 32 GB Memory
- Python with Anaconda v3.7
- Stanford CoreNLP[6]

9 Data

In order to obtain samples of different writing levels, collect IELTS test essays from the Internet before the experiment, covering 5 to 8 points. The experimental text is controlled within 4 groups, each group of texts does not exceed 20 examples.

References

1. Vimala Balakrishnan and Ethel Lloyd-Yemoh. Stemming and lemmatization: a comparison of retrieval performances. 2014.
2. Renee Elio, Jim Hoover, Ioanis Nikolaidis, Mohammad Salavatipour, Lorna Stewart, and Ken Wong. About computing science research methodology, 2011.

3. Marti A Hearst. The debate on automated essay grading. *IEEE Intelligent Systems and their Applications*, 15(5):22–37, 2000.
4. Guillaume Lample, Miguel Ballesteros, Sandeep Subramanian, Kazuya Kawakami, and Chris Dyer. Neural architectures for named entity recognition. *arXiv preprint arXiv:1603.01360*, 2016.
5. Heeyoung Lee, Yves Peirsman, Angel Chang, Nathanael Chambers, Mihai Surdeanu, and Dan Jurafsky. Stanford's multi-pass sieve coreference resolution system at the conll-2011 shared task. In *Proceedings of the fifteenth conference on computational natural language learning: Shared task*, pages 28–34. Association for Computational Linguistics, 2011.
6. Christopher D Manning, Mihai Surdeanu, John Bauer, Jenny Rose Finkel, Steven Bethard, and David McClosky. The stanford corenlp natural language processing toolkit. In *Proceedings of 52nd annual meeting of the association for computational linguistics: system demonstrations*, pages 55–60, 2014.
7. Christina Niklaus, Matthias Cetto, André Freitas, and Siegfried Handschuh. A survey on open information extraction. *arXiv preprint arXiv:1806.05599*, 2018.
8. Olutobi Owoputi, Brendan O'Connor, Chris Dyer, Kevin Gimpel, Nathan Schneider, and Noah A Smith. Improved part-of-speech tagging for online conversational text with word clusters. In *Proceedings of the 2013 conference of the North American chapter of the association for computational linguistics: human language technologies*, pages 380–390, 2013.
9. David D Palmer. Tokenisation and sentence segmentation. *Handbook of natural language processing*, pages 11–35, 2000.
10. Soujanya Poria, Basant Agarwal, Alexander Gelbukh, Amir Hussain, and Newton Howard. Dependency-based semantic parsing for concept-level text analysis. In *International Conference on Intelligent Text Processing and Computational Linguistics*, pages 113–127. Springer, 2014.
11. Salvatore Valenti, Francesca Neri, and Alessandro Cucchiarelli. An overview of current research on automated essay grading. *Journal of Information Technology Education: Research*, 2(1):319–330, 2003.
12. Zhen Wang, Jianwen Zhang, Jianlin Feng, and Zheng Chen. Knowledge graph and text jointly embedding. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pages 1591–1601, 2014.