論文 / 著書情報
Article / Book Information

題目(和文)	
Title(English)	Parsing Argumentative Structure in English-as-a-Foreign-Language Learner Essays
著者(和文)	Jan Wira Gotama PUTRA
Author(English)	Jan Wira Gotama Putra
出典(和文)	学位:博士(工学), 学位授与機関:東京工業大学, 報告番号:甲第12127号, 授与年月日:2021年9月24日, 学位の種別:課程博士, 審査員:徳永 健伸,岡崎 直観,村田 剛志,宮﨑 純,齋藤 豪
Citation(English)	Degree:Doctor (Engineering), Conferring organization: Tokyo Institute of Technology, Report number:甲第12127号, Conferred date:2021/9/24, Degree Type:Course doctor, Examiner:,,,,
学位種別(和文)	博士論文
Category(English)	Doctoral Thesis
種別(和文)	論文要旨
Type(English)	Summary

（博士課程）
Doctoral Program

論 文 要 旨

THESIS SUMMARY

系・コース： Department of, Graduate major in	Computer Science Artificial Intelligence	系 コース	
学生氏名： Student's Name	Jan Wira Gotama PUTRA		

申請学位（専攻分野）： Academic Degree Requested	博士 Doctor of (Engineering)
指導教員（主）： Academic Supervisor(main)	徳永　健伸
指導教員（副）： Academic Supervisor(sub)	Simone TEUFEL

要旨（英文 800 語程度）
Thesis Summary（approx.800 English Words ）

Argument mining (AM) aims to explain how individual argumentative discourse units (e.g., sentences and clause-like segments) relate to each other and what roles they play in the overall argumentation. The automatic recognition of argumentative structure is attractive because it benefits various downstream tasks, such as text assessment, text generation, text improvement and summarisation. Existing studies focused on analysing well-written texts provided by proficient authors. However, the majority of English speakers worldwide are non-native, and their texts are frequently poorly structured, particularly if they are still in the learning phase. Yet there is no prior study that addresses the analysis of argumentative structure specifically in non-native texts. This thesis presents an argument mining study on English-as-a-foreign-language (EFL) essays of intermediate quality. It focuses on three tasks: (i) constructing a new language resource for training an AM system for EFL essays, (ii) argumentative structure parsing, and (iii) improving the quality of essays by reordering sentences.

Concerning the first task, I present the first corpus of annotated EFL essays, together with a specially designed annotation scheme. The resulting corpus, called "ICNALE-AS2R" contains 434 essays written by learners from numerous Asian countries, along with two types of manual annotation: annotation of their argumentative structure and reordering annotation. The second type of annotation indicates one way how the sentences could be reordered, resulting in an essay of overall higher quality. The annotated corpus is particularly useful for the education domain, as the argumentative structure annotation can reveal learners' argumentation-related problems, and the reordering annotation shows one way to improve the essay so that it more closely resembles a native-level production. My argument annotation scheme is demonstrably stable, achieving good inter-annotator agreement and near-perfect intra-annotator agreement. The annotated corpus comes with some additional methodological and technical contributions. First, I propose a set of novel document-level agreement metrics that can quantify structural agreement from various argumentation aspects, thus, providing a more holistic analysis of the quality of the argumentative structure annotation. The metrics are evaluated in a crowdsourced meta-evaluation experiment, achieving moderate to good correlation with human judgements. Second, all corpus annotation is performed by an external expert annotator using my newly developed web-based annotation tool TIARA. It provides versatile visualisation for structural annotation and reduces clutter in the display. The tool is easily customisable via a configuration script. Apart from its use as an annotation tool, it is also designed to support the educational use case of learning-to-write.

I also conduct a secondary evaluation using three third-party professional essay assessors to confirm whether the reordered version of essays is indeed better than the original one, in the light of the likely inherent subjectivity of the quality of sentence arrangement. The assessors exhibited low agreement with each other in their judgement. My conclusion from this experiment is that the reordered version of essays in the ICNALE-AS2R corpus cannot be treated as the single correct one. Nevertheless, the evaluation confirms that the reordering operation improves essays' quality to some degree.

Concerning the second task, I propose deep learning models to parse the argumentative structure in EFL essays in two steps: a sentence linking and a relation labelling step. The experimental results show that a biaffine model combined with sentence-BERT encoder performs best in the sentence linking task, whereas fine-tuned BERT model shows the best results in the relation labelling task. I also evaluate the parser on a cross-domain setting, where training is performed on both in-domain

(EFL essays) and out-domain (reordered essays), and evaluation is performed on the in-domain test data. I observe that the best cross-domain system achieves 94% of the in-domain system in terms of end-to-end performance. I conclude that the best training regime for my parser might mix well-written texts with less well-structured texts. I identify the sentence linking task as the main challenge; the model seems to stumble when confronted with the hierarchical nature of arguments. To improve the sentence linking performance, I extend the biaffine model using a multi-task learning setup to provide a richer supervision signal. I also propose multi-corpora training with a selective sampling strategy to increase the available amount of training data. These two strategies consistently improved the sentence linking performance on all evaluation aspects, resulting in a 15.8% increase in the F1-macro score for individual link predictions, amongst other improvements.

Concerning the third task of providing discourse level feedback for language learners, I propose a new computational task of sentence reordering. Given a sequence of sentences, presumably in sub-optimal order, the goal is to rearrange them into a well-structured text. I develop a sentence reordering system based on the results of an earlier step of argumentative structure analysis. The reordering task is formulated as a tree-traversal problem consisting of two steps: a pairwise ordering constraint classification between argumentatively related sentences, followed by a tree traversal step which generates the final output. Experimental results show that the system can perform the reordering operation selectively, that is, it reorders sentences when necessary and retains the original input order when reordering would not result in an improvement. The usefulness of argumentative structure information is confirmed in an ablation study where the system's performance on three types of input is compared: automatically generated structures, gold standard structures and random structures. I found that the factor that would boost reordering performance most would be a further improvement in the argumentative structure parsing.

Overall, this thesis contributes towards providing automated discourse-level analysis and feedback to language learners. Both the argumentative structure visualisation and reordering recommendation facilitate the learning process, particularly in analysing and revising texts.