氏　　名	Mim Farjana Sultana (ミム フアージャナ サータナ)
学位の種類	博　士 (情報科学)
学位記番号	情　博　第　785　号
学位授与年月日	令和４年　９月２６日
学位授与の要件	学位規則第４条第１項該当
研究科、専攻	東北大学大学院情報科学研究科（博士課程）システム情報科学専攻
学位論文題目	Capturing Discourse Structure and Reasoning in Argumentation （論述における談話構造および論理構造の解析）
論文審査委員	（主査）東北大学教　授　乾　健太郎 東北大学教　授　伊藤彰則　　東北大学教　授　北村 喜文 東北大学教　授　鈴木　潤

論 文 内 容 の 要 旨

Chapter 1: Introduction

Argumentation can help students improve their critical thinking skills, decision making skills or writing skills. However, students often struggle to construct well-defined arguments and it is necessary to guide them by providing feedback so that they can improve their argumentation. The difficulty here is that providing feedback manually is an extremely time consuming task and requires lots of human efforts. Therefore, the importance of building an automated feedback system is enormous and assessing the quality of argumentation and capturing its underlying reasoning patterns are two of the crucial tasks to reach this ultimate goal. Assessing the quality of argumentation can help students learn how good or bad their arguments are and the capturing the reasoning patterns can help inspect the issues that makes their arguments poor.

This dissertation focuses on these two important tasks. Existing approaches for automated evaluation of argumentative texts typically rely on parsers to capture argumentative discourse. However, the performance of parsers is not always adequate, especially when they are used on noisy texts, such as student essays. To overcome this problem, in this thesis, we establish an unsupervised pre-training approach to capture argumentative discourse that does not require any parser or annotation. Our proposed unsupervised approach achieves state-of-the-art result on a task of scoring student essays.

Essay scoring systems provide feedback about the quality of argumentative essays but do not indicate the issues why the quality of an essay is good or bad. In order to improve the quality of students' argumentative writing, we need feedback systems that do not only provide a score for an argumentative text, but at the same time allow students to inspect the issues in their text. To

build such systems, deeper analysis of argumentation is necessary.

For deeper understanding, capturing writer's reasoning in argumentative texts is crucial. However, less attention has been paid to capturing reasoning patterns in argumentation and there are no studies that capture complex strategic moves in argumentation. We address this gap in this thesis and design a novel annotation scheme that capture logic patterns of strategic moves in argumentation. Our annotation study shows that human annotation for the proposed scheme is feasible and results in the creation of a corpus comprising logic patterns of strategic moves. Using our annotated corpus, we establish the task of automatic identification of logic patterns and the experimental results show moderate performance, setting a baseline for this task.

Chapter 2: Background

In general, an argument is a justifiable position consisting two components: (i) Conclusion (i.e., statement that expresses the position or belief of the arguer) and (ii) Premise (i.e., statements that provide reason for the conclusion). Example of an argument is given below where the conclusion is supported by its premise.

Conclusion: We should abolish death penalty.

Premise: Death penalty deprives the chance of rehabilitation of the criminals.

Generally, argumentation refers to the usage of arguments in a situation of disagreement or doubt (e.g., debates, student essays, online discussions). Arguments in an argumentation are connected between themselves either with a support or attack relation. For example, the following argument attacks the argument above.

Conclusion: Death penalty should not be abolished.

Premise: Rehabilitation fails in comparison with the death penalty at reducing or eliminating repeat offending

In educational context, argumentative interactions are often considered as a learning tool since through arguments, students can learn how to reason, verify, and the difference between fact and opinion which improves their critical thinking skills. The ability to form convincing arguments plays a crucial role in negotiation and decision making. However, people are often unable to develop good argumentation skill due to the lack of constructive feedback or proper guidance.

Although providing feedback is very important to improve learners' argumentation skill, providing manual feedback to each learner on their argumentation is not ideal since it is a time consuming task which requires lots of human efforts, and oftentimes difficult given the rise of massive online discussions. Hence, there has been an increasing importance of argumentative writing support systems since it can train individual learner to improve their argumentation skill by analyzing their arguments and providing feedback.

Chapter 3: Capturing Discourse Structure for Assessing the Quality of Argumentation

Automated Essay Scoring (AES), the task of both grading and evaluating written essays using machine learning techniques, is an important educational application of natural language processing (NLP). Since manual grading of student essays is extremely time consuming and requires lots of human efforts, AES systems are widely adopted for many large-scale writing assessments such as Graduate Record Examination (GRE).

In general, an essay is a discourse where sentences and paragraphs are logically connected to each other to provide comprehensive meaning. Conventionally, two types of connections have been discussed in the literature: coherence and cohesion. Coherence refers to the semantic relatedness among sentences and logical order of concepts and meanings in a text. For example, "I saw Jill on the street. She was going home." is coherent, whereas "I saw Jill on the street. She has two sisters." is incoherent. Cohesion refers to how well sentences and paragraphs in a text are linked by means of linguistic devices. Examples of these linguistic devices include conjunctions such as discourse indicators (DIs) (e.g., "because" and "for example"), coreference (e.g., "he" and "they"), substitution, ellipsis, etc.

For the precise assessment of overall essay quality or some dimensions of an essay, it is crucial to encode such discourse structure (i.e., coherence and cohesion) into an essay representation. Although discourse is one of the most important aspects of documents, less attention has been given to capturing discourse structure in an unsupervised manner. Most of the works that encapsulate discourse structure into document representation are dependent on annotation-based parsers. However, such annotations are costly, and parsers generally considers that the text is well-written which is not always true, especially in case of student essays that comprise different types of flaws (e.g., grammatical, spelling, discourse etc.).

In this thesis, We propose a novel unsupervised pre-training approach to capture long-range discourse dependencies in argumentative essays that does not require any parsers or annotations. Our proposed strategy outperforms two baseline models by a significant margin, and we achieve new state-of-the-art results for a an essay scoring task.

Chapter 4: Capturing Logic Patterns in Argumentation

Argumentation plays a central role in human communication, where refuting or attacking others' arguments is a common persuasion strategy. Attack in arguments can have different modes e.g., the counterargument can deny the conclusion (i.e., statement that expresses the position or belief of the arguer) of the attacked argument or it can deny a premise (i.e., statement that provides support or reason for the conclusion) of the attacked argument or the counterargument can deny an argumentative relation (i.e., support or attack) in the attacked argument.

Beside of having different forms, attacks in arguments often comprise complex rhetorical moves

as well e.g., one might agree with a premise while attacking the conclusion of the argument. Furthermore, arguments generally consist implicit knowledge (e.g., causal reasoning), sentiments (e.g., positive or negative feeling towards a certain concept or element), presupposition or value judgements (e.g., presupposing that some consequence has greater importance or value than another consequence) which contribute to the internal logical structure of attacks.

Prior studies in NLP that focused on attacks in arguments mainly worked on the classification of argumentative relations (e.g., support, attack, neutral), identifying attackable points in arguments, or counterargument generation. However there are no studies that capture the modes of attack (e.g., whether the counterargument denies the conclusion or the premise of the attacked argument) and the complex rhetorical moves in them (e.g., agreeing with a premise while attacking the conclusion).

To address these gaps, we introduce LPAttack (Logic Pattern of Attack), a novel annotation scheme that captures common modes of attacks and complex rhetorical moves in them as well as the implicit information and value judgments that contribute to the logical structure of attacks. We then conduct an annotation study using the proposed scheme that yields moderate agreement between two annotators indicating the feasibility of the human annotation for the scheme. Our annotation study results in the creation of a corpus of logic pattern of attacks of 250 debates.

Chapter 5: Automatic Identification of Logic Patterns in Argumentation

This chapter focuses on the formulation of the task of automatic identification of logic pattern of attacks (captured by the LPAttack scheme proposed in Chapter 4) from given arguments and counterarguments. Most of the existing argument mining tasks including reasoning patterns identification use hand-crafted features such as part-of-speech tags, lemma, n-grams, punctuation marks, discourse markers. Recently many argument mining and related tasks (e.g., stance detection and classification, classification of support, attack, or neutral argumentative relations) used pre-trained deep language representation models (e.g., BERT, GPT2, BART, T5) and achieved state-of-the-art result.

The text form of the logic patterns (annotated by the LPAttack scheme) can be seen as an abstractive summary of the given argument and counterargument. In this work, we treat the task of automatic identification of logic patterns of attacks as a logic pattern generation or summarization task and use a state-of-the-art language model which has been pre-trained on the abstractive summarization task. We demonstrate that the model yields moderate performance for the logic pattern generation task, setting a baseline for this challenging task.

Chapter 6: Conclusion

The key contributions of this thesis are summarized as follows:

Establishing an unsupervised approach to capture long-range discourse dependencies in argumentation:

We proposed a novel unsupervised pre-training approach to capture long-range discourse dependencies in argumentation that does not require any discourse parsers or annotations. We then used our unsupervised pre-training method for the quality assessment of argumentation. We demonstrated that our method is effective in capturing discourse structure of argumentation by achieving state-of-the-art performance on the assessment task.

Designing an annotation scheme to capture the reasoning patterns in argumentation:

We analyzed the internal structure of how one argument attacks or agrees with another argument which provided insights into how to represent the strategic moves in argumentation so that human annotation is plausible. Based on these insights, we designed a novel annotation scheme, defined the annotation guidelines and formulated the task of capturing the logic pattern of attacks in argumentation.

Construction of a corpus using the invented annotation scheme:

We conducted an annotation study and created a corpus comprising logic pattern of attacks using our proposed scheme. Our annotation study yielded moderate agreement between two annotators indicating the feasibility of the human annotation for the scheme.

Baseline model experiments for the automatic identification of reasoning patterns:

We considered the automatic identification of reasoning patterns as a reasoning patterns generation task and used a pre-trained language model for the generation purpose. The model achieved moderate performance, setting a baseline for this task

論文審査結果の要旨

論述は学習者の批判的思考力を養う有効な学習手段と考えられているが、十分な学習効果を得るには適切なフィードバックを提供し、それに基づいて学習者が自ら論述を改善するプロセスを確立することが必要である。個々の学習者の論述にフィードバックを提供する作業は労働集約的であり、人手だけではスケールさせるのが難しく、自動化による学習プロセスの改善が望まれている。本論文は、論述の質の自動評価を目的として、その実現に必要な要素技術の開発を論述の「全体構成の良さ」および「推論パターン」の２つのレイヤにおいて進め、それぞれタスクの設計、解決の方法論、方向性を論じたもので、全編 6 章からなる。

第 1 章は序論である。

第 2 章では、論述における基本的な概念を提示し、本研究の背景を説明している。

第 3 章では、論述における「全体構成の良さ」の評価に焦点を当て、談話依存構造、とくに長距離の依存関係の自然さを捉える評価モデルの構築方法を論じている。明示的な談話解析やアノテーションデータを必要としない新しい教師なし学習方式を提案し、提案手法が論述の全体構成の良さをモデル化するのに有効であることを、論述採点課題において世界最高の精度を達成することによって実証した。提案手法は、教師なし学習でも長距離の談話依存関係を捉えることが可能であり、全体構成の良さの評価に寄与することを初めて示したものであり、その貢献は高く評価できる。

第 4 章では、論述の背後にある「推論パターン」を捉えるための新しいアノテーションスキームの設計について論じている。従来の論述解析の研究が論述の修辞レベルの浅い構造の解析に留まっていたのに対し、本論文ではそれより一段深い「論述のロジック」に踏み込み、それを推論パターンのアノテーションとして定式化した。設計したスキームを用いたアノテーション実験では、アノテータ間の一致率で有望な結果が得られており、少なくとも人間には十分に安定したアノテーションが可能であることを示した。論述のロジックを推論パターンとして捉えることができることを示した初めての研究であり、ロジックレベルでの論述評価の実現に向けて極めて重要な基礎を与えるものである。

第 5 章では、第４章で設計した推論パターンについてこれを自動認識する課題を設定するとともに、ベースラインとなるモデルを構築し、構築したアノテーションデータを用いた解析実験を通して課題の性質を分析・議論している。論述のロジックレベルの自動解析を初めて試みたもので、高く評価できる。

第 6 章は結論である。

以上、本論文は論述自動評価に必要な２つの新しい要素技術を設計・開発・検証したものであり、情報科学の発展に寄与するところが少なくない。よって、本論文は博士（情報科学）の学位論文として合格と認める。