

補助事業番号 2022M-219

補助事業名 2022年度 採否の計算モデル構築と論文作成支援システムへの適用,  
教育現場への国際展開 補助事業

補助事業者名 山梨大学大学院総合研究部工学域 福本 文代

## 1 研究の概要

本事業は、英語を母国語としない執筆経験の少ない学生を対象とし、採択に向けての改善点を教示する知的学習支援システムを開発することを目的とする。学生が執筆した論文に対し、「可読性」や「新規性」などの各評価指標について、そのスコアを自動推定するモデル Peer-Review-Aspect-Score-Prediction (PRASP) を提案し、英文による論文執筆作業を支援する環境を構築した。

## 2 研究の目的と背景

近年、機械産業の急速な進展により、査読を経た多数の学術論文が次々に公開されている。論文の採否は、「可読性」や「新規性」などの評価指標が設けられ、査読者はそれらを総合的に判断し採否を決める。しかし英語を母国語としない学生にとり、論文を作成することは容易ではない。結果的に機械産業分野では欧米と比較するとアジア地域からの投稿数、並びに採択率は低く、このことは諸大学が抱える共通の課題となっている。

本研究は、英語を母国語としない執筆経験の少ない学生を対象とし、採択に向けての改善点を教示する知的学習支援システムを開発することを目的とする。「可読性」や「新規性」などの各評価指標の点数付けを行うと同時に、学生の継続性を促すために、その根拠を提示し修正を支援する。さらに日本をはじめ、中国、マレーシア、タイ、豪州の大学において、学生を対象としたフィールド調査を協働することにより、本研究がグローバルな視点で若手人材の育成に貢献できることを実証する。

## 3 研究内容 [http://cl.cs.yamanashi.ac.jp/news\\_e/jka\\_completion.html](http://cl.cs.yamanashi.ac.jp/news_e/jka_completion.html)

近年、機械産業の急速な進展により、査読を経た多数の学術論文が次々に公開されている。論文の採否は、「可読性」や「新規性」などの評価指標が設けられ、査読者はそれらを総合的に判断し採否を決める。しかし英語を母国語としない学生にとり、論文を作成することは容易ではない。論文投稿者への支援として、学会が提供している添削サービスも十分機能しているとは言い難い。結果的に機械産業分野では欧米と比較するとアジア地域からの投稿数、並びに採択率は低く、このことは諸大学が抱える共通の課題となっている。本研究は将来、学術・産業界を牽引できるような若手人材を育成するために、計算機による知的な支援という観点から、英語を母国語としない学生を対象とし、執筆論文について、各評価指標のスコアを自動推定するモデル Peer-Review-Aspect-Score-Prediction (PRASP) を提案し、英文による論文執筆作業を支援する環境を構築することを目的とする。

PRASP に関するこれまでの研究の多くは、教師あり学習、すなわちあらかじめスコアラベルが付与された訓練データから深層学習を用いスコア判定に必要となる特徴を自動的に学習し、テストデータであるスコアが付与されていない執筆論文のスコアを予測するという手法に基づいている。しかし人手によるラベル付けは労力を要することから訓練データ量は少なく、このことが精度の低下をもたらす。本研究は、この問題を解決するために、ラベルなしの大量の論文データセットを少量のラベル付き論文データと併せて利用することにより高精度で判定可能なモデルを構築する。近年、深層学習の進展により、自然言語処理において事前学習と呼ばれる手法が多数提案されている。これは、大量のラベルなしデータから、単語、あるいは文の意味表現を学習する手法であり、BERT と呼ばれる事前学習モデルは、感情分析、文書分類、文生成、質問応答など自然言語処理の様々なタスクにおいてその精度を各段に向上させている。そこで本研究では、半教師付き学習の一つである Ladder Network モデルに、事前学習モデルの一種である BERT を組み込むことにより、少ない訓練データにおいても高精度なスコア付けが可能となることを示す。本研究における PRASP のフレームワークを図 1 に示す。

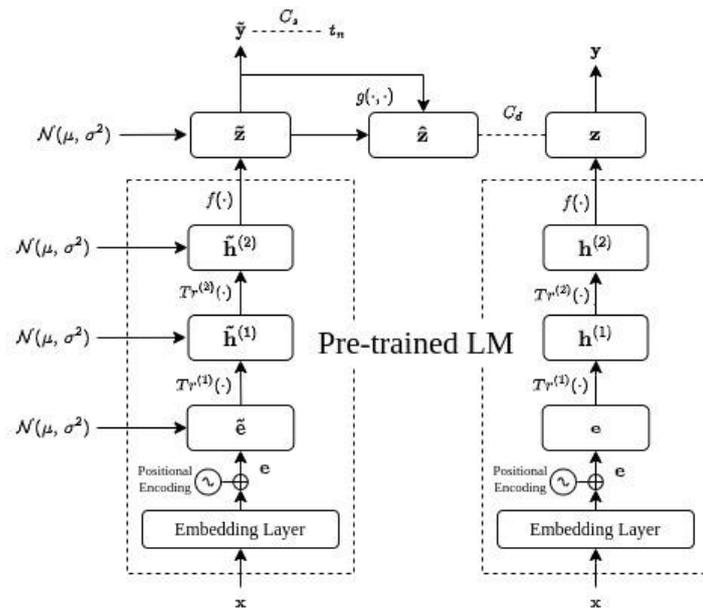


図 1 : PRASPのフレームワーク

図1においてxは入力を示し、yは評価尺度tの正解ラベルを示す。ネットワークは、2つの forward pass, clean pass, そしてcorrupted passから成る。点線枠の右は、教師付きデータ、左は教師なしデータを示し、それぞれBERTを用いて学習が行われる。最終的なコスト関数は、それぞれのコストである  $C_d$  と  $C_s$  の加算で求め、この値が最小となるよう学習がなされる。

本手法を用いて行った実験結果を他手法との比較と共に表1に示す。表1より、Accuracy では、Impact、Comparison、F1-score ではImpactを除き、本手法であるΓ-Transが関連研究、及び半教師付き学習手法と比較し、最も高い精度が得られていることが確認できる。

#### 4 本研究が実社会にどう活かされるか—展望

学術論文の公開は、研究の創発と同時に、産業競争力強化の鍵となる知識共有にも貢献する。したがって、論文執筆作業を支援する環境があれば、将来、世界レベルで学術・産業界を牽引できる若手人材の育成が可能となるため、環境構築は急務となっている。開発した支援環境は、審査を必要とするあらゆる業務支援にも応用可能であるため、社会的要請、並びに波及効果は高い。

#### 5 教歴・研究歴の流れにおける今回研究の位置づけ

研究代表者の専門は自然言語処理であり、電子化されたデータから語彙に関する知識を自動的に獲得する研究、及び自然言語処理としてテキスト分類、推薦、予測、文生成、及び要約に関する研究を実施してきた。本研究は、これまで研究代表者らが実施してきた自然言語処理応用の予測に関係し、自然言語処理の基盤技術を英文による学術論文作成の初心者学生を対象とした論文執筆作業に役立てることを目的としている。

(1) <http://www.nlp.yamanashi.ac.jp/>

(2)



学内での学生向けシステム説明

#### 6 本研究にかかわる知財・発表論文等

本研究にかかわる発表論文は以下の通りである。

[1] Semi-Supervised Learning for Aspect-based Sentiment Analysis, Hang Zheng, Jianhui Zhang, Yoshimi Suzuki\*, Fumiyo Fukumoto, and Hiromitsu Nishizaki, Proc. of 2021 International Conference on Cyberworlds, pp. 209–212, 2021.

[2] Neural Local and Global Contexts Learning for Word Sense Disambiguation, Fumiyo Fukumoto\*, Taishin Mishima, Jiyi Li, and Yoshimi Suzuki, Proc. of the 28th International Conference on Neural Information Processing, pp. 537–549, 2021.

- [3] Paraphrase Identification with Neural Elaboration Relation Learning, Sheng Xu, Fumiyo Fukumoto\*, Jiyi Li, and Yoshimi Suzuki, Proc. of the 28th International Conference on Neural Information Processing, pp. 562–573, 2021.
- [4] 局所および大域的特徴量に基づく語義の曖昧性解消, 浅川翔, 鈴木良弥, 李吉吃, \*福本文代, pp. 1797–1801, 第28回言語処理学会年次大会, 2022.
- [5] 低リソースの意見要約のための教師なしドメイン 適応, 小杉有斗, \*李吉吃, 福本文代, pp. 1986–1990, 第28回言語処理学会年次大会, 2022.
- [6] Exploiting Labeled and Unlabeled Data via Transformer Fine-tuning for Peer-Review Score Prediction, Panitan Muangkammuen, Fumiyo Fukumoto\*, Jiyi Li, and Yoshimi Suzuki, Findings of the 2022 Conference on Empirical Methods in Natural Language Processing, pp. 2233–2240, 2022.
- [7] BBSN: Bilateral-Branch Siamese Network for Imbalanced Multi-label Text Classification, Jiangjiang, Zhao, Jiyi Li\*, and Fumiyo Fukumoto, The 29th International Conference on Neural Information Processing, 2022.
- [8] Improving Peer-Review Score Prediction with Semi-Supervised Learning and Denoising Network, Muangkammuen Panitan, 福本文代\*, 李吉吃, 鈴木良弥, pp. xx-yy, 第29回言語処理学会年次大会, 2023.
- [9] Disentangling Meaning and Style for Positive Text Reframing, 徐勝, 鈴木良弥, 李吉吃, 郷健太郎, 福本文代\*, pp. xx-yy, 第29回言語処理学会年次大会, 2023.

## 7 補助事業に係る成果物

### (1) 補助事業により作成したもの

- Video: <https://s3.amazonaws.com/pf-user-files-01/u-59356/uploads/2022-11-15/ni23t9g/EMNLP%202022%20.mp4>
- Source code: [https://github.com/panitan-m/gamma\\_trans](https://github.com/panitan-m/gamma_trans)
- Poster: <https://underline.io/events/54/posters/1335/poster/5733-a-neural-local-coherence-analysis-model-for-text-clarity-scoring?tab=poster>
- Source code: [https://github.com/panitan-m/local\\_coh](https://github.com/panitan-m/local_coh)

### (2) (1) 以外で当事業において作成したもの

なし.

8 事業内容についての問い合わせ先

所属機関名： 山梨大学大学院総合研究部工学域

(ヤマナシダイガクダイガクインソウゴウケンキュウブコウガクイキ)

住 所： 〒400-8510

山梨県甲府市武田4-3-11

担 当 者： 役職名 教授 (キョウジュ) 福本文代 (フクモトフミヨ)

担 当 部 署： 福本・李研究室 (フクモト・リケンキュウシツ)

E - m a i l： fukumoto@yamanashi.ac.jp

U R L： [http://cl.cs.yamanashi.ac.jp/index\\_e.html](http://cl.cs.yamanashi.ac.jp/index_e.html)