4P-02

BERTにより生成されたタイトルパーツの並び替えに基づく論文タイトル生成について

賀来　健人[†]　大囿　忠親[‡]　新谷　虎松[‡]

名古屋工業大学情報工学科[†] 名古屋工業大学大学院情報工学専攻[‡]

1　はじめに

学生のように論文執筆に慣れていない人にとって論文タイトルの作成は困難であるという問題がある．先行研究では，深層学習と抽出型および生成型の要約手法を用いてアブストラクトから論文タイトルを生成する手法が提案された [1]．しかし，先行研究では抽出型および生成型の手法はどちらも改善の余地がある．生成型の手法は直接論文タイトルを生成するが，抽出型の手法では，キーワードを抽出しそれらを用いて論文タイトルを作成する．そのため抽出型の手法の方が見込みがあると考えられる．先行研究の抽出型の手法では，論文のアブストラクトから抽出されたキーワードを人手で論文タイトルのテンプレートに当てはめることで論文タイトルを作成した．本研究では，抽出されたキーワードから並び替えにより論文タイトルを生成する手法に焦点を当てる．具体的には，タイトルパーツの生成および並び替えによる論文タイトル候補の生成と文法の正誤確認とタイトルらしさの評価による適切な論文タイトル候補の絞り込みである．タイトルパーツとは，論文タイトルに出現すると期待される単語列である．また，先行研究ではキーワード抽出に RNN と Attention を用いた深層学習モデルを使用しているが，本研究では双方向の文脈を捉えることができる自然言語処理モデル BERT を用いた深層学習モデルを使用した．

本稿では，タイトルパーツの並び替えに基づく抽出型論文タイトル生成手法とそれに基づく論文タイトル生成システムについて述べる．

2　タイトルパーツの並び替えに基づく抽出型論文タイトル生成

論文タイトル生成システムのシステム構成図を図 1 に示す．ユーザーがクライアントにアブストラクトを入力するとアブストラクトがサーバに送られ，キーワード抽出モデルに入力される．そこでアブストラクトから抽出されたキーワードが出力され，それを用いてタイトルパーツ生成部でタイトルパーツが生成される．ここでクライアントにタイトルパーツを送信し，ユーザーにタイトルパーツを表示する．そこで，抽出されていないキーワードがある場合など必要に応じてユーザーはタイトルパーツを編集することができる．編集

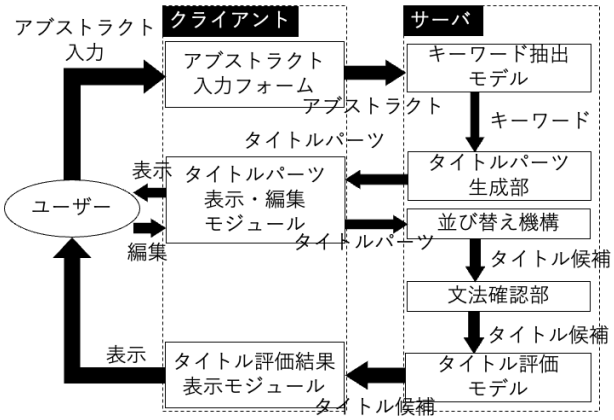

図 1: システム構成図

が終了すると再びサーバにタイトルパーツが送信され，並び替え機構においてタイトルパーツが並び替えられ，論文タイトル候補が生成される．これらの文法の正誤を文法確認部で確認し，文法的に正しいと判断された論文タイトル候補のみが出力される．出力された論文タイトル候補はタイトル評価モデルにより論文タイトルらしさがあるか判断される．タイトルらしさがあると判断された論文タイトル候補のみがクライアントに送信され，ユーザーに表示される．

本研究では，抽出されたキーワードを並び替えることによって生成された論文タイトル候補のタイトルらしさを評価することで，論文タイトル候補の中から適切なものを絞り込み，適切な論文タイトルを生成しようと考えた．しかし，この手法には 3 つの問題点がある．(1) 実行時間が長いこと，(2) 抽出できない単語があること，(3) 文法的に正しくない論文タイトル候補に対しても高評価することである．(1) の要因は，キーワードの並び替えによって生成される論文タイトル候補数が多いこととタイトル評価の実行時間が長いことである．(2) の要因は，我々が BERT により構築したアブストラクトからキーワードを抽出するモデル (図 1 のキーワード抽出モデル) が，未知語を扱えないためである．(3) の要因は，我々が BERT により構築した論文タイトルらしさを評価するモデル (図 1 のタイトル評価モデル) がタイトルらしさとして文法の正誤を学習していないためであると考えられる．

これらの問題点を解決するために，タイトルパーツと文法確認を導入する．キーワードの代わりにタイトルパーツを並び替えることで論文タイトル候補数の減少が期待され，一般語を組み合わせた専門用語にも対応しやすくなると考えられる．また，タイトルパーツ

On a Paper Title Generation Method Based on Permutations of Title Parts with BERT

[†]Kento KAKU , [‡]Tadachika OZONO and [‡]Toramatsu SHINTANI

[†]Dept. of Computer Science, Nagoya Institute of Technology.

[‡]Dept. of Computer Science, Graduate School of Engineering, Nagoya Institute of Technology.

Copyright ©2021 Information Processing Society of Japan.
All Rights Reserved.

の編集を人手によって行うことで，BERT によって抽出されなかった単語をタイトルパーツに追加することを可能にする．文法の正誤確認をすることによって文法的に誤っている候補を除外することができ，さらに論文タイトル候補を減らすことができる．以降，図 1 のキーワード抽出モデルとタイトルパーツ生成部，文法確認部，タイトル評価モデルについて説明する．

キーワード抽出モデルは，BERT を用いてアブストラクト中の単語にスコアをつけることで閾値以上の単語を論文タイトル中に出現すると期待されるキーワードとして出力する．このモデルをキーワード抽出モデルといい，学習には実際の論文のアブストラクトとタイトルを用いた．

タイトルパーツ生成部は，タイトルパーツを生成するために用いることのできる情報は論文のアブストラクトと抽出されたキーワードのみである．そこで，アブストラクト中でキーワードが連続している部分をタイトルパーツとすることとした．

文法確認部は，構文解析器によって得られた実際の論文タイトルの構文木から葉を除いたものを正しい文法として事前に記録しておく．これらを論文タイトル候補の構文木と照合し，一致するものがあれば文法的に正しいとする．

タイトル評価モデルは，BERT を用いて論文タイトル候補にスコアをつけ，そのスコアが閾値以上であれば論文タイトルとして適切であると判断する．それ以外のものは論文タイトルとして不適切であるとして論文タイトル候補から除外する．学習には，実際の論文タイトルとその語順を変えたもの，アブストラクトの 1 文目を用いた．

3 評価実験

AAAI の論文 9,229 件の論文タイトルとアブストラクトを用いてキーワード抽出モデルを 1epoch 学習し，9,232 件の論文タイトルとアブストラクトの 1 文目と語順を変えた論文タイトルを用いてタイトル評価モデルを 3epoch 学習した．これらのモデルに関して AAAI の論文 500 件を用いて評価実験を行った．また，文法確認部の正しい文法として 9,232 件の論文タイトルの構文を用い，評価実験では 3 件の論文タイトルをそれぞれ 5 つのタイトルパーツに分割し，並び替えることで生成された 360 件の論文タイトル候補を用いた．キーワード抽出モデルの評価実験では，抽出されたキーワードと実際の論文タイトル中の単語およびアブストラクトに含まれる論文タイトル中の単語との適合率および再現率を検証した．タイトル評価モデルの評価実験では，正例として実際の論文タイトルを，負例としてアブストラクトの 1 文目と語順を変えた論文タイトルを評価し，それぞれの評価結果の正解率を検証した．文法確認部の評価実験では，論文タイトル候補の文法の正誤を文法確認部と人手でそれぞれ判断し，混同行列の各要素の割合を検証した．キーワード抽出モデル，タイトル評価モデル，文法確認部の実験結果をそれぞれ表 1，表 2，図 2 に示す．

表 1 からキーワード抽出モデルの性能が低いことがわかる．要因として BERT は未知語の抽出ができないことから抽出できない単語があることが考えられる．表 2 からタイトル評価モデルは実際のタイトルとアブストラクトの 1 文目のような通常の文を正しく評価することができることがわかるが，語順を変えた論文タイトルのような文法的に正しくないものは正しく評価ができていないことがわかる．しかし，文法確認を別途行っていることから問題はないと考えられる．したがって，タイトル評価モデルは実用的であると考えられる．図 2 から文法確認部は正しい文法を持つ論文タイトル候補をほとんど除外することなく，全体の半数近い誤った文法を持つ論文タイトル候補を除外することができると考えられる．これによりタイトル評価の高速化は実現できた．しかし，未だ多くの文法的に正しくない論文タイトル候補が残っていることから文法確認部を改善する必要があると考えられる．

表 1: キーワード抽出モデルの実験結果

抽出されたキーワードとの比較対象	適合率	再現率
実際の論文タイトル中の単語	0.32	0.36
アブストラクトに含まれる論文タイトル中の単語	0.32	0.43

表 2: タイトル評価モデルの実験結果

	正解率
実際のタイトル	0.93
アブストラクトの 1 文目	0.98
語順を変えた実際のタイトル	0.72

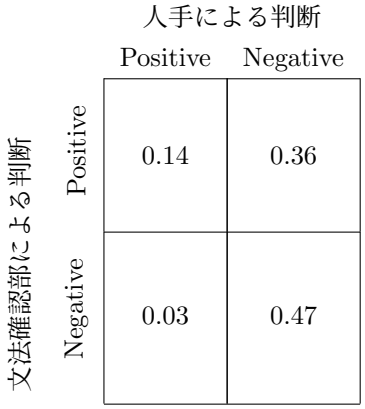

図 2: 文法確認部の実験結果

4 おわりに

本稿では，論文執筆に慣れていない人に対する論文タイトル作成支援のための抽出型論文タイトル生成システムについて述べた．本システムを用いることで適切な論文タイトル候補を提示することが可能となり，論文タイトルの作成支援が可能となった．

謝辞 本研究の一部は JSPS 科研費 19K12097, 19K12266 の助成を受けたものです．

参考文献

[1] 大部達也, 大囿忠親, 新谷虎松, “論文執筆支援エージェントのためのタイトル生成機構,” 合同エージェントワークショップ＆シンポジウム 2017, pp.232-237, 2017.

Copyright ©2021 Information Processing Society of Japan.
All Rights Reserved.