

Recurrent Neural Network を用いた論文タイトル作成支援のためのキーワード抽出について

On Extracting Keywords for a Title Using a Recurrent Neural Network

大部達也¹ 大園忠親¹ 新谷虎松¹
Tatsuya Ohbe Tadachika Ozono Toramatsu Shintani

名古屋工業大学大学院工学研究科情報工学専攻¹
Department of Computer Science, Graduate School of Engineering, Nagoya Institute of Technology

1 はじめに

論文執筆において、論文のタイトルは主張点を的確に表す必要があり、タイトルの決定は重要である。論文のタイトルには規則性や定型的な言い回しがあると考えられるため、初学者が論文の内容を的確に表すタイトルを作成するためには、キーワード抽出やテンプレートの提供のような支援が好ましい。論文のアブストラクトからタイトルを生成するタスクは、文書要約の一種として考えられる。深層学習を用いた文書要約手法として、Encoder-Decoder 翻訳モデルを利用した手法がある。しかし、文書要約タスクにおいては、原文における未知語が要約に重要となるキーワードである場合が多い。Gulcehre らは辞書からの単語の生成に加えて、Attention 機構を応用して原文中の単語を指す手法を併用した、未知語に頑健な Encoder-Decoder 翻訳モデルを提案している [1]。本研究では、論文のアブストラクトからキーワードを抽出することでタイトル候補を生成し、論文執筆の支援を行うシステムの実現を目的としている。本稿では、Encoder-Decoder 翻訳モデルを応用した論文タイトル作成のためのキーワード抽出について述べる。

2 論文タイトル生成手法

文書要約は大量の文書进行处理するために用いられ、要約文の可読性も重要となる。論文タイトル生成では、人手による修正も考慮することができるので、必ずしも文法的に正しい必要はない。一方で、論文の内容を的確に示す必要があり、また、論文タイトルにはある程度の規則性が見られると考えられるため、テンプレートを利用することができる。

文書要約には抽出型の要約と生成型の要約がある。抽出型の要約は、原文中から文書単位(単語、文など)レベルで抽出することで要約文を生成する。生成型の要約は、辞書から単語を選ぶことで要約文を生成する。本研究では、抽出型の手法によりキーワードの抽出を行い、テンプレートを併用して論文タイトルを生成する。キーワード抽出は、アブストラクト中の各単語をタイトルに出現するか出現しないかの二値に分類する。図 1 に本手法のネットワークを示す。アブストラクトの各単語について Encoder によりそれまでの単語をエンコードする。また、各単語を入力した際の隠れ層を保存しておき、Attention ベクトルを作る。エンコードされたベクトルと Attention ベクトルに対して Softmax 関数を適用

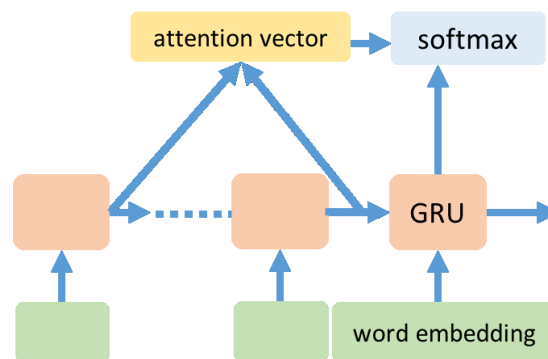


図 1 キーワード抽出ネットワーク

表 1 キーワード抽出の精度と再現率

キーワード抽出 精度	0.23
キーワード抽出 再現率	0.27

し、誤差の計算を行う。Encoder のユニットには Gated Recurrent Unit (GRU) を使い、誤差関数にはクロスエントロピーを用いる。

3 実験・考察

本研究では、Confer API¹ から取得可能な論文のタイトルとアブストラクト 3,208 件を用いて提案手法の性能を検証した。テストデータには学習データに含まれない 337 件の論文を使用した。50 エポック学習したモデルを用いてキーワード抽出の精度および再現率を検証した結果について表 1 に示す。実験では精度、再現率ともに低い値となったが、論文のタイトル生成を支援するためには再現率が重要である。そのため、今後再現率の向上が望まれる。

4 おわりに

本稿では、論文タイトル作成支援のためのキーワード抽出手法について述べた。本研究では、アブストラクト中の未知語を扱うために、辞書を必要としない抽出型の手法についてキーワード抽出の性能を検証した。提案手法では未知語を含むアブストラクトについて、論文タイトル作成が支援されると考えられる。

参考文献

- [1] Caglar Gulcehre et al. "Pointing the unknown words." In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics.

¹<http://confer.csail.mit.edu/developer>