

Word2vec を用いたパラグラフ・ライティング

プロジェクトマネジメントコース 矢吹研究室 1442069 氏名 須山 武弘

1. 序論

レポートや論文を書く際には、読みやすく、論理的な文章を書くことが大切である。読みやすく、論理的な文章とは、伝えたいことが読み手に伝わり、段落内の話題が統一されている事が必要である。論理的文章は学習をしなければ書けるようにはならない。欧米では、半年～1年をかけて勉強することがある。

パラグラフ・ライティング (Paragraph writing) とは、論理的文章を書くための世界標準の書き方である [1]。冒頭にトピックとなる文を配置し、後ろに補助する文を配置するピラミッド構造的な文章の書き方である。最初にトピックとなる文を配置することにより、結論がはっきりし、内容の理解が深まることや、冒頭にトピックを記すことで、速読が可能になるなど、多数のメリットがある。

本研究では、Word2vec を利用し、単語の意味をベクトル化し、定量的結果から、パラグラフ・ライティングの補助ができるのではないかと考えられる [2]。

2. 目的

本研究では、文字列である文章をベクトルで表し、パラグラフ・ライティングの補助をすることが目的である。段落内文章の意味のベクトルを Word2vec で定量的に表す。結果として、同じようなベクトルの位置に分布すると、定量的に意味が同じであると考えられる。

3. 手法

本研究では、Word2vec を用いて文章をベクトル化し、数値的・面的から論理的文章作成を補助する。通常、文章は定量的に表すことができないが、Word2vec は、単語をベクトルとして数値へ変換できるニューラルネットのことである。文章をベクトルへ変換することによって単語の意味を個人の主観ではなく、定量的に表すことができる。

対象文章の形態素解析を Mecab で実行し、Wikipedia の情報を利用したコーパスである、東

北大学・岡崎研究室「日本語 Wikipedia エンティティベクトル」を利用し、単語のベクトル化を行い、主成分分析の結果を図示する。

4. 結果

手法にあるように、Word2vec を用いて文章のベクトル化を行った出力結果が図 1 である。過去に人間による添削済みの文章であり、同じ色のベクトルが同段落の文章のベクトルである。

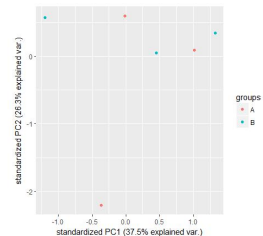


図 1 word2vec による分析結果

5. 考察

同段落にもかかわらず、ベクトルは散らばって分布している事がわかる。この内容では、それぞれの文章が違う話題を示していると考えられる。

パラグラフ・ライティングにおける考え方として、段落内の文は冒頭の一文を補助する文であることから、本研究の結果から、うまくパラグラフ・ライティングができていないという結果になったと考えられる。

6. 結論

今回の結果から、Word2vec を用いて数値によって文章を見ていくことで、個人の主観だけでなく、定量的に文章の書き方を定義することも期待される。

参考文献

- [1] 倉島保美. 論理が伝わる 世界標準の「書く技術」. 講談社, 2012.
- [2] 西尾泰和. word2vec による自然言語処理. 株式会社オライリー・ジャパン, 第 2 版, 2017.