

数理・情報のフロンティア
2020 年度採択研究代表者

2020 年度 年次報告書

高瀬 翔

東京工業大学 情報理工学院
助教

ランダムベクトルを用いた軽量な埋め込み表現の構築

§ 1. 研究成果の概要

近年の自然言語処理における、事前学習モデルの発展において、学習コーパスの量やモデルのパラメータ数と性能が対数比例するという報告[Brown et al., 2020]があるように、性能向上のためにパラメータ数が著しく増加している。このような状況を受け、パラメータ数に対し効率の良い手法を考えることが本研究の目的である。自然言語処理においては、各単語の意味をベクトルで表現した埋め込み表現が最も巨大なパラメータ行列となることが多いため、本研究では特に、埋め込み表現について、性能を維持しつつ、パラメータ数を抑えた手法を探る。本研究で目指す手法は、ランダムベクトルを用いて埋め込み表現を構築することである。具体的には、全単語で共有のランダムベクトルを複数用意し、これの組み合わせによって各単語の埋め込みを構築する。この際、組み合わせが各単語によって固有であれば、各単語に固有のベクトルが構築されるため、従来の埋め込み表現の代替となると考えられる。2020年度は最も素朴な実装として、各単語からのランダムベクトルへの割り当てをランダムとし、埋め込みを構築する手法を開発した。本手法を機械翻訳と要約という、自然言語処理における主要な応用タスクに適用し、実験を行った。実験を通して、本手法は、従来の埋め込み表現よりも少ないパラメータ数で同等の性能を達成可能なことを示した。また、既存の、少量のパラメータを用いた埋め込み表現と比べると高い性能を達成可能なことを示した。