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

横井 祥

東北大学 大学院情報科学研究科
助教

言葉が埋め込まれた空間の形と言葉の意味の接続

§1. 研究成果の概要

単語埋め込み空間の幾何学的な特徴および文の表現の構成方法に関して大きく 4 つの研究をおこなった。

[内藤, 横井, 下平; NLP 2021] では、実際のコーパスから学習された単語埋め込みが持つバイアスを学習アルゴリズム毎に特定した。また単語頻度で重みつけした重心を除去することでバイアスを大幅に軽減でき、文の表現を作るためのヒューリスティックである加法構成がより正確に成り立つことを理論・実験の両面から示した。

[横井, 下平; NLP 2021] では、単語埋め込み空間の等方性を担保するために、単語頻度を考慮しながら白色化するアルゴリズムを提案した。提案法により後段タスクでの性能が一貫して著しく向上すること、また同様の指針で測定した等方性の度合いが埋め込み空間の品質の良好な説明変数になることを示した。

[小林, 栗林, 横井, 乾; NLP 2021] では、Transformer と呼ばれる多層ニューラルネットワークが単語ベクトルを動的に更新する際、周辺文脈をどの程度「混ぜて」いるかについて定量的に評価する方法を考案した。結果、注意機構によって周辺文脈を混ぜる作用は残差結合およびレイヤー正規化によってほとんどキャンセルされることがわかった。

[Hanawa, Yokoi, Hara, Inui; ICLR 2021] では、機械学習モデルを事例ベースに説明する手法を評価するための規範的なガイドラインを提案した。実験の結果、事例を表現するための分散表現としては、単語ベクトルのように内部表現を直接用いるよりも勾配ベクトルを用いる方が説明の観点では優位であること(説明手法としての最低限の要請を満たすこと)を示した。

【代表的な原著論文情報】

- Kazuaki Hanawa, Sho Yokoi, Satoshi Hara, Kentaro Inui. Evaluation of Similarity-based Explanations. The Ninth International Conference on Learning Representations (ICLR). March 2021.