

ABCI 3.0開発加速利用（2024年度）成果概要（公開用）

課題名： 最先端AIモデルを用いた高度な言語理解・言語生成技術の研究	実施時期：2024年1月20日～3月20日 所属機関名：東京科学大学 代表者氏名：荒瀬 由紀
成果概要： 二つの文の間で意味が一致する単語ペアを推定する単語アラインメントは、文の意味的な類似度推定の基礎となる重要な基盤技術である。本研究では最適輸送を応用した単語アラインメント推定技術について、特に多対多のアラインメント推定にフォーカスし、実験と検証を行った。	
成果のポイント： 二つの文の間で意味が一致する単語ペアを推定する単語アラインメントは、文の意味的な類似度推定の基礎となる重要な基盤技術である。本研究では最適輸送を応用した単語アラインメント推定技術の研究を行っている。提案手法は単語表層だけでなく意味的な類似に基づく文間類似度推定を可能とするものであり、情報検索や近年LLMに適用されるようになったRetrieval Augmented Generationの発展にも寄与するものである。 本開発加速利用期間では、約2か月という短期間の利用であったため、アプローチの有効性を検証する実験を集中的に行った。多対多のアラインメント推定にフォーカスし、最適輸送の単語アラインメントにおける有効性を複数のデータセットによって評価を行った。比較的シンプルな一対一のアラインメントに対し、多対多のアラインメントは句としての意味の構成、複単語表現の考慮を必要とし、難易度が高い。我々の実験の結果、多対多のアラインメントにおいても最適輸送を用いるアプローチの有効性が支持される結果が得られた。今後さらに手法の改善と評価実験を継続して実施する計画である。	
成果についてより詳細な情報を提供しているWebページ、発表論文などの情報： 現在継続して検証を実施しており、成果は論文として公表する予定である。	