

先端科学技術研究科 修士論文要旨

所属研究室 (主指導教員)	自然言語処理学 (渡辺 太郎 (教授))					
学籍番号	2411139	提出日	令和 8年 1月 19日			
学生氏名	志子田 直輝					
論文題目	CLIPと不均衡最適輸送を用いた画像-文章間の類似性評価および局所アライメントの可視化					
要旨						
<p>Contrastive Language-Image Pre-training (CLIP) は、画像と文章を同一空間に射影して学習することで、モダリティを超えた共通の埋め込み表現を獲得した。これにより、表現間のコサイン類似度を用いた画像と文章の整合性評価が可能となった。</p> <p>本研究では、CLIPに基づく画像と文章の整合性の評価を行うとともに、両者の局所的な対応関係を可視化する手法を提案する。</p> <p>可視化手法の提案にあたり、既存手法では達成が難しい(1)追加学習が不要とすること、(2)設計がモデルの内部構造に依存しないこと、(3)画像と文章の局所間の関係が抽出できること、の3点を満たすことを目指した。</p> <p>具体的には、学習済みモデルの入力に局所的な変更を加え、それに起因する出力の差分から局所情報を推定することで、内部構造が不明なモデルに対しても適用可能とした。さらに、画像-文章間の局所対応を輸送問題として捉えることで最適輸送で整合性を算出した。</p> <p>評価実験により、提案手法は従来手法とは異なる新たなアプローチでありながら、ベースラインと同程度に人手評価と相関する整合性評価が可能であることを示した。加えて、既存手法では達成が難しい上記3点を満たしながら、画像-文章間の局所対応を解釈可能な形で可視化できることを示した。</p>						