

# 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)	ソーシャル・コンピューティング (荒牧 英治 (教授))		
学籍番号	2311173	提出日	令和 7年 1月 17日
学生氏名	辻 航平		
論文題目	SubRegWeigh: サブワード正則化による高速アノテーション補正		
要旨			
<p>自然言語処理のデータセットには、たとえ手動でアノテーションを行ったとしてもアノテーション誤りが含まれている可能性がある。自動でアノテーション誤りの影響を軽減する手法は存在するものの、既存の手法は複数のモデルを学習する必要があり、誤りの特定に多くの時間を費やす必要がある。本論文では、複数のモデルを学習する代わりに、サブワード正則化というトークナイゼーション技術を用いることで、高速にアノテーション誤りを特定する。提案手法であるSubRegWeighは、既存手法よりも4～5倍ほど高速にアノテーション補正が行えることに加えて、文書分類および固有表現抽出タスクにおいて、精度を改善することができた。また、疑似誤りラベルを用いた実験において、提案手法は疑似誤りラベルを適切に認識できていた。</p>			