論文 / 著書情報
Article / Book Information

題目(和文)	
Title(English)	A Study of Non-standardWord Usage on Social Media
著者(和文)	青木竜哉
Author(English)	Tatsuya Aoki
出典(和文)	学位:博士(工学), 学位授与機関:東京科学大学, 報告番号:甲第283号, 授与年月日:2025年3月26日, 学位の種別:課程博士, 審査員:奥村 学,中山 実,鈴木 賢治,篠﨑 隆宏,船越 孝太郎,高村 大也
Citation(English)	Degree:Doctor (Engineering), Conferring organization: Institute of Science Tokyo, Report number:甲第283号, Conferred date:2025/3/26, Degree Type:Course doctor, Examiner:,,,,,
学位種別(和文)	博士論文
Category(English)	Doctoral Thesis
種別(和文)	審査の要旨
Type(English)	Exam Summary

別　紙　1（論文博士）

論文審査の要旨及び審査員　（2000 字程度）

報告番号	乙 第　　号	学位申請者	青木　竜哉	
論文審査員	氏　　名	職　　名	氏　　名	職　　名
	主査　奥村 学	教授	船越孝太郎	准教授
	中山実	教授	高村大也	産業技術総合研究所 グループリーダー
	鈴木賢治	教授		
	篠崎隆宏	教授		

本論文は「A Study of Non-standard Word Usage on Social Media」（ソーシャルメディアにおける非一般的単語用法に関する研究）と題し，英文全5章より構成されている．

第1章「Introduction」（序論）では，本研究の背景と目的を述べている．ソーシャルメディアの文章には，辞書に載っている既存の単語であっても，特定のコミュニティや文化的文脈において意味や用法が変化する「非一般的な語義」が頻出すること，これらが自然言語処理システムにとって誤解析や誤翻訳の原因となり得ることを指摘している．また，既存の機械学習モデルや大規模言語モデルをそのまま用いるだけでは大規模テキストに対して計算コストが高いことや，新たなネットスラングなど素早い語義変化への対応が難しい問題があることから，より軽量かつ汎用的なアプローチの必要性を述べている．

第2章「Related Work」（関連研究）では，非一般的用法検出に近い，語義曖昧性解消や，大規模コーパス間の語義変化を含む従来研究を概観し，未知語義や，環境依存で意味が変化する単語の自動検出手法を整理している．さらに，Skip-gram による単語埋め込み学習や大規模言語モデルによる文脈理解との関連を論じることで，本研究の位置づけを明確にしている．

第3章「Word Embedding-based Non-standard Word Usage Detection」（単語埋め込みに基づく非一般的用法検出）では，まず日本語 Twitter を元に，熟練アノテータが単語ごとに「一般的な用法」，「非一般的な用法」のラベルを付与した事例を収集し，データセットを構築したことを述べている．次いで，単語埋め込みを学習する Skip-gram with Negative Sampling (SGNS) を用いて，ターゲット単語の IN ベクトルだけではなく OUT ベクトルも活用することで，ターゲット単語と周辺単語との距離やスコアに基づき非一般的用法を検出する手法を提案している．実験では，ウィンドウ内の単語への重み付けや機能語の除去により精度が向上し，日本語 Twitter を元にしたデータセットでの評価において，単純な閾値方式ながらも高い正解率を示したことを報告している．

第4章「Masked Language Model-based Non-standard Word Usage Detection」（マスク化言語モデルに基づく非一般的用法検出）では，単語単体でのウィンドウ近傍のみを参照する手法の限界を克服するため，事前学習済み言語モデルの上に，単語用法を二値分類する層を構築した新しい手法について説明している．さらに，非一般的用法の用例が不足している問題を解消するために，擬似ラベル学習を導入し，大量の無注釈コーパスから擬似的な「非一般的用例」を自動生成して学習する戦略を採用している．そして，新たに英語 Reddit のクラウドソーシングによるアノテーションデータセットを構築し，日英両言語での実験により，提案モデルがより高い精度と再現率を両立することを示している．加えて，Transformer 出力の直後に位置するトークン予測層である LM Head を微調整する意義や，SGNS 由来の単語埋め込みを併用する効果を実証している．

第5章「Conclusion」（結論）では，ソーシャルメディア上の急速な語義変化を効率的に検出するため，単語埋め込みとマスク化言語モデルを基盤とした軽量かつ汎用的なモデルを提案し，さらに擬似ラベル学習で高い精度を実現したことを総括している．

以上を要するに，本論文は，急速な語義変化や多様なユーザ層を擁するソーシャルメディア上での非一般的語義を精度良く検知し，その解析を通して自然言語処理システムの性能向上に寄与するものであり，工学的価値が大きい．よって博士（工学）の学位を授与するに十分な価値を持つものと認められる．