氏　　名	ERONEN JUUSO KALEVI KRISTIAN
授与学位	博士（工学）
学位記番号	博甲第203号
学位授与年月日	令和4年9月6日
学位授与の要件	学位規則第4条第1項
学位論文題目	Improving Multilingual Automatic Cyberbullying Detection With Feature Density And Cross-lingual Zero-shot Transfer （素性密度及びクロスリンガルゼロショット転送学習による多言語のネットいじめ自動検出の改良に関する研究）
論文審査委員	主査　准教授　プタシンスキ・ミハウ・エドムンド 教　授　桝井　文人 教　授　奥村　貴史 教　授　前田　康成 教　授　阿部　良夫

学位論文内容の要旨

In this thesis, I study two different methods for improving multilingual automatic cyberbullying detection. First, I study the effectiveness of Feature Density (FD) using different linguistically-backed feature preprocessing methods in order to estimate dataset complexity, which in turn is used to comparatively estimate the potential performance of machine learning (ML) classifiers prior to any training. I hypothesize that estimating dataset complexity allows for the reduction of the number of required experiments iterations, making it possible optimize the resource-intensive training of ML models which is becoming a serious issue due to the increases in available dataset sizes and the ever-rising popularity of models based on Deep Neural Networks (DNN). The problem of constantly increasing needs for more powerful computational resources is also affecting the environment due to alarmingly-growing amount of CO2 emissions caused by training of large-scale ML models. I use cyberbullying datasets collected for multiple languages, namely English, Japanese and Polish. The difference in linguistic complexity of datasets allows us to additionally discuss the efficacy of linguistically-backed word preprocessing.

Second, I study the selection of transfer languages for automatic abusive language detection. I demonstrate the effectiveness of cross-lingual transfer learning for zero-shot abusive language detection. This way it is possible use existing data from higher-resource languages to build better detection systems for low-resource languages. The datasets are from eight different languages from three language families. I measure the distance between the languages using several language similarity measures, especially by quantifying the World Atlas of Language Structures. I show that there is a correlation between linguistic similarity and classifier performance, making it possible to choose an optimal transfer language for zero shot abusive language detection.

Next, I demonstrate that this method is also generally applicable to multiple Natural Language Processing tasks, specifically sentiment analysis, named entity recognition and dependency parsing. I show that there is also a correlation between linguistic similarity and zero-shot cross-lingual transfer performance for these tasks, allowing us to select an ideal transfer language in order to aid with the problem of dealing with low-resource languages. Lastly, I explore what kind of linguistic features are the most important from the point of view of cross-lingual transfer and how they affect the cross-lingual transfer process. Using this information, I also show that the World Atlas of Language Structures can be quantified into an effective linguistic similarity method.

審査結果の要旨

近年、ソーシャル・ネットワーキング・サービス（SNS）の普及により、ネット上のいじめやヘイトスピーチの問題がますます蔓延し、過激になりつつあり、数多くの国と言語のユーザが危険にさらされている。この多大な社会問題を解くためには、機械学習を用いた有害情報の自動検出分類手法における研究がされてきた。しかし、SNS の種類、SNS が用いられている国、言語の種類など、それぞれの組み合わせが数多く、それらのために異なる手法を開発しなければならない。さらに、手法を開発するためのデータが少ない言語もあり有害情報の自動検出分類手法を効率よく開発することが困難である。この問題を解決するためには、著者は２つの取り組みを行なった。まず、開発プロセスで用いられるパラメーターの組み合わせを調査し、普段高い性能となる組み合わせを 68 中 12 に削減でき、手法の開発プロセスに必要な時間を 5 分の 1 に下げることができた。さらに、多言語モデルを学習し、多資源の起点言語の知識情報を応用する移転学習の手法を用いて訓練データがない目的言語でも有害情報検出手法の開発を可能にした。さらに、どの目的言語にどの起点言語が最適であるかを明らかにした。

結論として、著者は、今後の有害情報検出手法への根本的な貢献をした他、自然言語処理分野において昨今課題となりつつある移転学習に関連して最適起点言語選択の課題について示唆を与える新知見を得たと判断される。なお、自然言語処理、人工知能など複数の分野に跨る深淵な課題に対して貢献するところ大なるものがある。よって著者は、博士(工学)の学位を授与される資格があるものと認める。