

(続紙 1)

京都大学	博士（情報学）	氏名	Roberto Espinoza Chamorro
論文題目	Implementation of ICT for Improvement of Participation for Blood Donation in Japan: SNS Data Processing and Analysis through Topic Modeling(日本における献血参加勧奨へのICTの導入：SNSデータのトピックモデリングによる分析)		
<p>(論文内容の要旨)</p> <p>Although blood donation promotion is widely performed and evaluated through conventional method such as surveys and interviews, there are no evaluation methods for feedback collection enables blood collection authority to improve satisfactions of citizens by following citizen's voice trend. In this thesis, the author tries to develop a complete process to collect sentiments of group of people related to blood donation, such as donors, non-donners, people who differed from blood donation due to certain reasons, from unstructured SNS data using non-supervised machine learning methods.</p> <p>This thesis consists of the seven chapters.</p> <p>The first chapter provides background of this research that derived from the author's master thesis project to develop mobile application to promote blood donation for people who differed from blood donation due to certain reasons to overcome lack of donors around the world. The reaction collected via the prototype make the author realize the importance of the sentiment analysis from social public through social network service such as Twitter.</p> <p>The second chapter reviews previous studies on Twitter data analysis and pointed that BERT (Bidirectional Encoder Representation from Transformers) and that BERTopic appears as the most suitable method for document classification and topic modelling respectively.</p> <p>The third chapter explains the creation of Japanese blood donation dataset and benchmark of classification. In this step, the author utilized the BERT-model which is pre-trained using Japanese corpus available in HuggingFace. The author classified the roles of citizens regarding blood donation, namely Donor, Non-Donor and Undermined, and classified Undetermined into Potential and Deferred. And developed BDT-UC (Blood Donation Tweet – User Classification) dataset by applying the pre-trained BERT model. The accuracy of the best performing model was 0.795.</p> <p>The author performs the analysis from the fourth to fifth chapters. In the fourth chapter, the author performed the initial topic modeling to identify the best performing option and found that using MeCab instead of default OCTIS and BERT-MiniLM instead of BERT-NMF provided best performance. In the fifth chapter, the authors introduced Campaign and Informative category as subcategories of Undetermined category and introduced Countervectorizer layer. The detailed analysis of the categorized result tells that the obtained results clearly show the response to the campaign performed by the Japanese Red Cross Society.</p> <p>In the sixth chapter, the author summarized the analysis process proposed in the previous chapters to establish the complete process of Twitter data analysis including sentiment classification of the group of people. The detailed topic modeling and sentiment analysis provides some hidden topics and recurrent topics that are causes of uncertainty, fear, or hesitation, such as Intercourse Regulation.</p> <p>In the seventh chapter, the author concludes the main contribution of this thesis and limitations. This author claims this research clarifies potential of natural language processing of data collected via social networking services to follow the citizen's sentiment and concerns on social activity such as blood donation.</p>			