

## 論文内容要旨

|   |  |     |          |
|---|--|-----|----------|
| 報告番号  | 甲 先 第 463 号  | 氏 名 | 瑪依努尔 孜帕尔 |
| 学位論文題目  | A Study on Childcare Information Extraction from Twitter Text<br>(Twitter テキストからの育児情報抽出に関する研究) |     |          |
| <p>内容要旨</p> <p>It is difficult to obtain necessary information accurately from Social Networking Service (SNS) while raising children, and it is thought that there is a certain demand for the development of a system that presents appropriate information to users according to the child's developmental stage.</p> <p>There are still few examples of research on knowledge extraction that focuses on childcare. This research aims to develop a system that extracts and presents useful knowledge for people who are actually raising children, using texts about childcare posted on Twitter. In many systems, numbers in text data are just strings like words and are normalized to zero or simply ignored.</p> <p>In this paper, we created a set of tweet texts and a set of profiles created according to the developmental stages of infants from "0-year-old child" to "6-year-old child". For each set, we used ML algorithms such as NB (Naive Bayes), LR (Logistic Regression), ANN (Approximate Nearest Neighbor algorithms search), XGboost, RF (random forest), decision trees, and SVM (Support Vector Machine) to compare with BERT (Bidirectional Encoder Representations from Transformers), a neural language model, to construct a classification model that predicts numbers from "0" to "6" from sentences. The accuracy rate predicted by the BERT classifier was slightly higher than that of the NB, LR, and ANN, XGboost, and RF, decision trees and SVM classifiers, indicating that the BERT classification method was better.</p> |  |     |          |