（様式４－１）

No. 1

学 位 論 文 要 旨

2022年 03 月 23 日

ソフトウェア情報学研究科(博士 課程)
学籍番号 2362019001
氏 名 Sen Rikta
研究室名 Pattern Recognition and Machine Learning

１ 題目（日本語及び英語で記載すること）

Study on Stability Analysis of Feature Selection Algorithm and Proposal towards Stable Feature Selection Method

特徴選択アルゴリズムの安定性解析と堅牢な特徴選択法の提案に関する研究

２ 要旨

Feature selection is one of the essential preprocessing tasks in machine learning and pattern recognition problems for reducing the dimensionality of the data. It removes irrelevant and redundant features leading to simplified classification process and improved accuracy. Several feature selection algorithms have been proposed so far but for any particular problem, the quality of the selected feature subset varies from algorithm to algorithm. Usually, the quality of the feature selection algorithm is evaluated by reduction of cardinality of the selected feature subset, improvement of classification accuracy or the reduction of algorithm complexity (computational cost). But stability of feature selection algorithm is another important characteristic which needs to be considered for evaluation of any feature selection algorithm. Stability refers to the robustness of the selected feature subset to small changes in the training set or set of various parameters of the algorithm. A stable feature selection algorithm is supposed to select the same subset of features for a particular problem irrespective of any changes in the training set of samples or parameters of the algorithm. Selection of stable feature subset is especially required when the physical meaning of the features are important.

（様式４－１）

（学位論文要旨）

No. 2

学籍番号	2362019001	氏　名	Sen Rikta

Various metrics have been developed so far for measuring the stability of a feature selection algorithm. In this work, an extensive analysis of stability of various types of feature selection algorithms (filter ranked based, filter subset based, and wrapper based algorithms) has been done with various stability measures. It has been found that filter rank based feature selection algorithms possess better stability than others, Jeffries-Matusita (JM) distance based feature selection being the best. JM distance is then verified as an efficient feature selection tool by using the simulation experiment for binary classification problems. A multiclass extension of JM distance has also been proposed as a feature selection algorithm which is found to perform better compared to the previous multiclass extensions of JM distance and other rank based filter approaches. Finally the critical analysis of different stability metrics has been done in which the desired properties of stability metrics are analyzed to determine which stability metrics follow which properties. The limitations of various similarity-based stability metrics are analyzed based on their desired properties. A correction, as well as a novel extension of similarity-based stability metric, Lustgarten measure, an extension of the most popular Kuncheva index, is proposed. The proposed new stability metric fulfills all the desired properties of stability metrics and removes the limitations of other metrics. The proposed stability metric has also been verified and found to be the best among the existing stability metrics by simulation experiments with different bench mark data sets.