

様式 3

論文審査の結果の要旨

フリガナ名 セン リクタ  
氏名 SEN RIKTA  
学位の種類 博士 (ソフトウェア情報学)  
学位記番号 甲第 65 号  
学位授与年月日 令和 4 年 3 月 18 日  
学位授与の根拠 岩手県立大学学位規則第 3 条第 3 項  
学位論文題目 特徴選択アルゴリズムの安定性解析と堅牢な特徴的選択法の提案に関する研究  
論文審査委員 主査 バサビ・チャクラボルティ  
副査 土井 章男 プリマ・オキ・ディッキ

審査結果の要旨

博士論文の概要：

Feature selection is one of the essential preprocessing tasks in machine learning and pattern recognition problems for reducing the dimensionality of the data. 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. One of the important quality of an algorithm is its stability.

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 benchmark data sets.

本審査の研究内容に関して質問および回答：

Question 1: What type of dataset do you use in your simulation experiment of the proposed multiclass JM distance for feature selection? Do you use economic or high dimensional dataset here?

Reply: In the simulation experiment, different dimensional dataset is used. The motive of the experiment is to verify our proposed multiclass JM distance as a feature selection tool and compare with other previous multiclass extensions.

Question 2: Do you know what type of dataset is more suitable for your experiment of JM distance measure?

Reply: We do not correlate datasets characteristics with JM distance measure. Actually, this is not our objective. Our objective is to verify that the proposed multiclass JM distance measure works as a feature selection method like others. The type of dataset is not important for verifying the multiclass JM distance measure.

Question 3: How do you tell that your proposed corrected Lustgarten measure gives better stability value than other metrics?

Reply: We verify our proposed corrected Lustgarten measure with toy experiments and also experiment with benchmark dataset. In toy experiment, we create the conditions according to the theoretical analysis where other stability metrics give incorrect result and verify that our proposed corrected Lustgarten measure gives the correct result at those conditions. In experiment with benchmark datasets, our proposed Lustgarten measure also gives the correct result than other metrics when these different conditions are occurred.

The candidate defended his thesis well and all the examiners are satisfied.

修了試験結果：

ソフトウェア情報学の基礎知識について、学位を与える十分であることを学位審査会にて確認できましたので、合格とする。