



PMME 2016

Analysis of Clustering and Classification Methods for Actionable Knowledge[★]

Arumugam P, Christy V

Department of Statistics, Manonmanium Sundar University, Tirunelveli, Tamilnadu, India

Abstract

Data Mining becomes a vital aspect in data analysis. Study on data mining is very much depends on the performance of the clustering. Clustering before classification is termed as cluster Classifier. Recently knowledge based approached has become the key forces in data classification. Here performed a four way comparison of Logistic Regression (LR), Classification and Regression Trees (CART), Random Forest (RF) and Neural Network (NN) models using a continuous and categorical dependent variable for classification. A Customer relationship management (CRM) data set is used to run these models. Measurement of different classification accuracy methods are used to compare the performance of the models. Based on the efficient method actionable knowledge is derived from the proposed methodology.

© 2017 Elsevier Ltd. All rights reserved.

Selection and Peer-review under responsibility of International Conference on Processing of Materials, Minerals and Energy (July 29th – 30th) 2016, Ongole, Andhra Pradesh, India

Keywords: Data Mining, Random Forest, Actionable Knowledge, Clustering, Classification.

[★] This is an open-access article distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike License, which permits non-commercial use, distribution, and reproduction in any medium, provided the original author and source are credited.

* Corresponding Author. +91 99 0005 8913

E-Mail: christy.eben@gmail.com

1. Introduction

Statistical methods like regression analysis, multivariate analysis and pattern recognition models have been applied to a wide range of decisions in many disciplines. Machine Learning algorithms have also been used in defining business insights. In this analysis CRM data set is used to analyze the methodology. Clustering before classification works well in defining the data classification [3]. Clustering algorithms has been surveyed and found self organizing map is the current researchers' context in the real world. The objective of the study is to define a hybrid method for clustering based on principle component analysis and neural gas combined with the self organizing map [4,7,9] and using the clusters in classification algorithms like logistic regression, classification and regression trees, neural network and random forest [5,8,6]. Although in this paper the classification method predictive accuracy is compared with CRM dataset. The main objective of this work is to provide an actionable insight to the business to obtain its productivity.

2. Related Works

Evans and Pfahringer introduced a new concept of clustering for classification[3]. This paper shows that clustering prior to classification is beneficial when using the sophisticated classifier. Ngai[5] Compared the data mining techniques like logistic regression, Linear discriminate Analysis and decision trees and found the decision trees as a appropriate model for customer relationship management. The paper [5] try to detection of outliers in multivariate data. There used various outlier techniques such as Mahalanobis distance, Cook's distance.[6] Parneet compared classification methods like Multilayer Preceptron from neural network, some tree methods and naïve bayes. [10] who extracted actionable knowledge from decision trees. Decision trees identify the features that are most discerning when it comes to identifying classes. Two well known methods are Boosting and Bagging. Breiman and Cutlers (2013) proposed random forests package using R programming, which add an additional layer of randomness to bagging. The combination of learning models increases the classification accuracy.

3. Proposed Methodology

The effectiveness of the proposed approach has been analyzed using the CRM data set. Almost 126 variables has been extracted and based on the data reduction methods like factor analysis, principle component analysis it has been derived to get some of 15 variables for the analysis [5]. The representation of the proposed work is shown in Fig.1.

متن کامل مقاله

دریافت فوری ←

ISIArticles

مرجع مقالات تخصصی ایران

- ✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
- ✓ امکان دانلود نسخه ترجمه شده مقالات
- ✓ پذیرش سفارش ترجمه تخصصی
- ✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
- ✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
- ✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
- ✓ دانلود فوری مقاله پس از پرداخت آنلاین
- ✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات