GUJARAT TECHNOLOGICAL UNIVERSITY
(Established by Government of Gujarat under Gujarat Act No. : 20 of 2007)
ગુજરાત ટેકનોલોજીકલ યુનિવર્સિટી
(ગુજરાત સરકારના ગુજરાત અધિનિયમ ક્રમાંક :૨૦/૨૦૦૭ દ્વારા સ્થાપિત)

Abstract of the Thesis

Name: SHAH KOMAL KISHORKUMAR
Enrollment No: 169999913014
Faculty: Engineering and Technology
Discipline/Branch: Computer/IT Engineering
Title of the Thesis: PERFORMANCE ENHANCEMENT FOR MULTICLASS RELATIONAL DATA CLASSIFIER USING MULTIPLE VIEW APPROACH

Abstract

Classification is a widely used data mining task with a numerous practical applications. Various algorithms have been proposed to develop accurate classifiers such as Decision Tree, Naive Bayes, Neural Network, Support Vector Machine, k-Nearest Neighbour, etc. However, traditional classification algorithms are often ineffective for many real-world applications where data is stored in a relational format across multiple tables. The primary objective of multirelational classification is to discover valuable patterns within multiple related tables in a relational database.

Many of existing multirelational classification algorithms either flattens multiple tables into a single file to be learned by traditional classifiers or upgrade traditional classifiers to work with relational data. However, these approaches have their own limitations. Flattening approach can lead to the loss of essential semantic information and generate a table with numerous missing values after spending significant time and effort on data transformation. The upgrade approach faces scalability issues and requires more time for learning. To overcome these challenges, multiple view approach is used which utilizes traditional classifiers directly on relational data to generate independent views from each related table. The models or views generated from each relation are then combined to generate a relational classification model. Multiple view approach preserves the original data representation and utilizes traditional methods, resulting in scalability and low pre-processing time, thereby gaining the advantages of both flattening and upgrading approaches.

Implementation results of multiple view approach based relational classifiers shows that they are more accurate compared to flattening and upgrading based algorithms for binary class

classification problems. However, when applied to multiclass relational data, their accuracy decreases resulting in misclassification of instances belonging to minority classes. Furthermore during experimental analysis it was also observed that relational classification models generated using existing framework loses important information about nominal attributes in relational database. To overcome these problems, we propose the Relational Multiclass Classification (RMC) algorithm that performs binarization using All–vs–All method to reduce misclassification error. We also inserted a stage in existing multiview framework to convert nominal attributes into binary to retain nominal attribute information. The experimental results obtained from real world datasets indicate that the proposed algorithm, RMC, gives promising results by improving accuracy and reducing the misclassification error of minority classes for multiclass classification problems as compared to existing multiview based algorithms.

This PhD Thesis provides fundamental data mining algorithm designed specifically for multiclass classification scenario where information is distributed across multiple related tables. The algorithm proposed in thesis is useful for datamining tasks in diverse domains, including but not limited to finance, health care and e-commerce, where data is structured within relational database management system.

List of Publication(s):

1) A comparative study and performance analysis of multirelational classification algorithms, International Journal of Business Intelligence and Data Mining, Volume 20:2, pp-121-145, February 2022.

2) Empirical Analysis Of Multiclass Classifiers Based On Dataset Characteristics, Journal for Basic Sciences, Volume 23, Issue 4, pp 776-788, April 2023.