International Journal of Scientific Research in Computer Science, Engineering and Information Technology
ISSN : 2456-3307 (www.ijsrcseit.com)
doi : https://doi.org/10.32628/IJSRCSEIT

Hierarchical Classification Using a New Hybrid Feature Selection Algorithm

A. Vinod Kumar Reddy[1], B. Hasya[2]
MCA Student[1], Assistant. Professor[2]
Mother Theresa Institute of Computer Applications, Palamaner, S. V. University, Andhra Pradesh, India

Article Info

Publication Issue :
Volume 8, Issue 6
November-December-2022

Page Number : 72-75

Article History
Accepted: 01 Nov 2022
Published: 05 Nov 2022

ABSTRACT

A common preprocessing step in the data mining industry is feature selection. Reducing the quantity of original dataset characteristics is one of its goals in order to enhance the accuracy of a prediction model. To the best of our knowledge, very few research in the literature address feature selection for the context of hierarchical classification, despite the advantages of feature selection for the classification problem. The general variable neighbourhood search metaheuristic is used to support the innovative feature selection approach that is proposed in this research. The method combines a filter step and a wrapper phase, and a global model hierarchical classifier is used to assess feature subsets. We conducted computational tests to verify the impact of the suggested approach on classification performance while employing two proposed global hierarchical classifiers, using various datasets from the proteins and pictures domains. in the written word. According to statistical testing, our feature selection strategy consistently produced prediction results that were superior to or on par with those achieved by employing all features while using fewer features, which supports its efficacy in the context of hierarchical categorization.

Keywords: Feature Selection, Hierarchical Single-Label Classification, Variable Neighborhood Search, Filter, Wrapper.

I. INTRODUCTION

The vast rise in the volume of data collected and stored in recent years has made data mining applications indispensable. The repurposing of data into useful and intelligible information raises fresh research problems. The goal of feature selection is to find as many pertinent characteristics as feasible while lowering the cost of data processing. Feature selection is frequently used as a preprocessing step in data mining projects. We will concentrate on feature selection strategies for the classification challenge in this research. As a result, we solely took labelled datasets into account. Some advantages of feature selection include raising the predicted accuracy of classifiers and cutting down on classification execution time. The scientific community has given categorization a lot of attention among data mining

jobs. In classification, the class label(s) of an example are predicted based on the problem domain that its characteristicsreflect. The literature contains categorization problems at various literary degrees of difficulty. Each dataset instance is given one or more class labels that are independent of one another in classic (flat) classification tasks. However, more difficult classification issues, known as hierarchical classification issues, do occur in many real-world applications where classes that label instances are arranged into a hierarchical structure represented by a tree or a directed acyclic graph (DAG).

II. RELATED WORKS

Hierarchical classification of G-protein-coupled receptors with data-driven selection of attributes and classifiers: Using a protein's main sequence, we attempt to solve the significant bioinformatics challenge of predicting a protein's function. We look at how G-Protein-Coupled Receptors (GPCRs), whose actions are described in a class hierarchy, are functionally categorised. We approach this problem using a unique top-down hierarchical classification system, where each node's predictor attributes and the classifier that will be used to classify the selected characteristics are determined based on data. Our novel technique greatly cut processing time without significantly reducing predicted accuracy when compared to a prior hierarchical classification system that solely selected classifiers.

Data Mining Concepts and Techniques: The concepts and methods for processing obtained data or information, which will be applied in diverse applications, are provided in Concepts and Techniques. It describes data mining in particular and the methods used to extract knowledge from the gathered data. The title of this book is "Knowledge Discovery from Data" (KDD). It focuses on the viability, applicability, efficacy, and scalability of methods for handling massive data collections. This version discusses the processes for knowing, preprocessing, processing, and warehousing data after defining data mining. Then, data warehouses, online analytical processing (OLAP), and data cube technologies are covered. The procedures for mining common patterns, connections, and correlations for huge data sets are then explained. The book discusses the ideas and techniques and describes the techniques for data classification.

Approach and Review of User-Oriented Interactive for clustering data. The subsequent chapters go on data mining trends, applications, and research frontiers as well as outlier detection.

A survey of hierarchical classification across different application domains: We talk about the task of hierarchical categorization in this survey. Since the literature in this discipline is dispersed throughout a wide range of application domains, research in one domain is frequently conducted without reference to techniques developed in other fields. We outline what constitutes a hierarchical classification task and explain why some comparable tasks shouldn't be included. Additionally, we offer a fresh viewpoint on a few of the current hierarchical categorization techniques, and on the basis of that viewpoint, we suggest a brand-new unifying framework to categorise the current approaches. Additionally, we evaluate empirical comparisons of the strategies that have been published in the literature and compare them conceptually at a high level of abstraction, outlining the benefits and drawbacks of each. Keywords Tree- structured class hierarchies, DAG-structured class hierarchies, hierarchical categorization.

Exploring attribute selection in hierarchical classification: Classes have dependence ties in the area of many classification issues, and these relationships are represented in hierarchical structures. The term "hierarchical categorization challenges" refers to these

issues. To address issues and improve prediction performance, many methods have been presented that take into account hierarchical linkages in various ways. With the aim of enhancing their individual performances, we investigate attribute selection methods in this study in combination with hierarchical classifiers from several categories. Computer simulations using 18 hierarchical datasets have shown that the classifiers used have higher predicted accuracy when the most important properties are taken into account. As a structural regularisation, we look at the hierarchical relationships in a label space tree structure. Finally, we apply two regularisation terms to a feature selection model based on sparse learning. Additionally, we apply the suggested model to a DAG situation, which broadens the applicability and robustness of our approach to a wide range of practical problems. The suggested framework for hierarchical categorization domains' usefulness is demonstrated by experimental findings on real- world datasets.Feature selection for hierarchical classification via joint semantic and structural information of labels:

In many real-world applications, when the label space is shown as a tree or a Directed Acyclic Graph (DAG), and each label contains a detailed semantic description, hierarchical classification is frequently utilised. The performance of machine learning algorithms has been shown to be enhanced by feature selection, a sort of dimension reduction strategy. Numerous feature selection techniques now in use, however, cannot be used to solve hierarchical classification issues directly because they neglect the hierarchical linkages and do not make use of the semantic data present in the label space. We suggest a unique framework for feature selection in this study that is based on the semantic and structural data of labels. In order to determine the similarity score between labels as the semantic regularisation, we first turn the label description into a mathematical representation. Second

III. Methodology

Proposed system: A global model hierarchical classifier analyses feature subsets as part of the unique feature selection approach we are implementing in the proposed system, which is based on the general variable neighbourhood search metaheuristic. With the advantage of using fewer characteristics, statistical testing revealed that our strategy for feature selection produced prediction results that were consistently better than or similar to those achieved by employing all features.

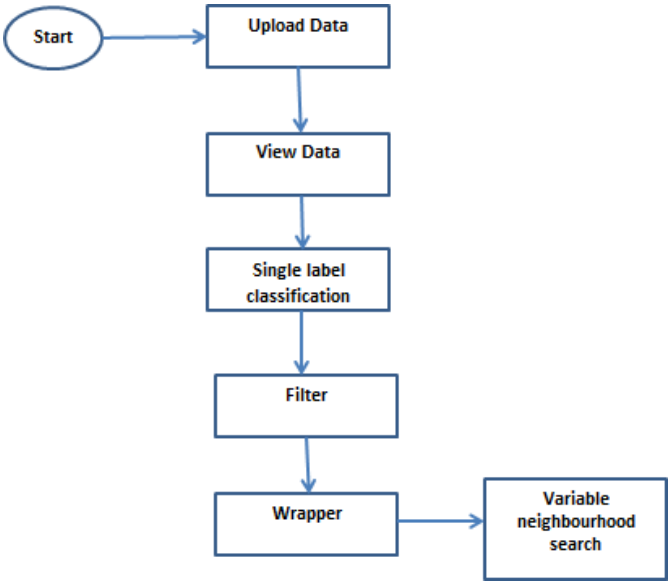

Figure 1: block diagram of project

IV. Implementation

This project is implemented by using below mentioned algorithms called Feature Selection Classification.

Feature selection has received increasing attention from researchers in recent years due to the continued rapid growth in the volume of data. Powerful as a preprocessing step, it selects a subset of predictive features to improve the performance of learning models. Data containing irrelevant or redundant features can reduce the predictive capability and increase the classification processing time of classifiers Feature selection methods can process

datasets that have previously labeled, partially labeled, and non-labeled instances, leading to the development of algorithms, respectively. A supervised feature selection algorithm determines the relevance of features by evaluating their existing correlation with the class feature. In this paper, we considered datasets with labeled instances. Therefore, we will focus on studies that proposed feature selection approaches for the supervised learning context, specifically feature selection approaches for the classification task.

V. Conclusion

We described in this research a brand-new feature selection technique designed specifically for global model hierarchical classifiers. The so-called GVNS-FSHC, which employs the SUH measure in a filter stage and the GMNB or the CLUS-HMC as the classifier in a wrapper step, is a hybrid filter-wrapper strategy we created based on the VNS metaheuristic. On twelve datasets, we compare the GVNS-FSHC technique with other feature selection methods (from proteins and images contexts).

VI. REFERENCES

[1]. J. Han, M. Kamber, and J. Pei, Data Mining: Concepts and Techniques, 3rd ed. San Francisco, CA, USA: Morgan Kaufmann, 2011.

[2]. C. N. Silla and A. A. Freitas, "A survey of hierarchical classification across different application domains," Data Mining Knowl. Discovery, vol. 22, nos. 1–2, pp. 31–72, 2011.

[3]. D. Koller and M. Sahami, "Hierarchically classifying documents using very few words," in Proc. 14th Int. Conf. Mach. Learn. San Francisco, CA, USA: Morgan Kaufmann, 1997, pp. 170–178.

[4]. A. Secker, M. N. Davies, A. A. Freitas, E. Clark, J. Timmis, and D. R. Flower, "Hierarchical classification of G-protein- coupled receptors with data-driven selection of attributes and classifiers," Int. J. Data Mining Bioinf., vol. 4, pp. 191–210, Jan. 2010.

[5]. B. C. Paes, A. Plastino, and A. A. Freitas, "Exploring attribute selection in hierarchical classification," J. Inf. Data Manage., vol. 5, no. 1, pp. 124–133, 2014.

[6]. H. Zhao, P. Zhu, P. Wang, and Q. Hu, "Hierarchical feature selection with recursive regularization," in Proc. 26th Int. Joint Conf. Artif. Intell., Melbourne, VIC, Australia, 2017, pp. 3483–3489.

[7]. Q. Tuo, H. Zhao, and Q. Hu, "Hierarchical feature selection with subtree based graph regularization," Knowl.-Based Syst., vol. 163, pp. 996–1008, Jan. 2019.

[8]. H. Huang and H. Liu, "Feature selection for hierarchical classification via joint semantic and structural information of labels," Knowl.- Based Syst., vol. 195, May 2020, Art. no. 105655.

[9]. I. Slavkov, J. Karcheska, D. Kocev, S. Kalajdziski, and S. Dzeroski, "ReliefF for hierarchical multi-label classification," in New Frontiers in Mining Complex Patterns (Lecture Notes in Computer Science). Cham, Switzerland: Springer, 2014, pp. 148–161.

[10].T. N. Dias and L. H. C. Merschmann, "Adaptation of the symmetric uncertainty measure for feature selection in single-label hierarchical classification context," (in Portuguese), in Proc. Ann. Nat. Meeting Artif. Intell. Comput., Natal, Brazil, 2015, pp. 142– 149.

Cite this article as :

A. Vinod Kumar Reddy, B. Hasya, "Hierarchical Classification Using a New Hybrid Feature Selection Algorithm", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 8 Issue 6, pp. 72-75, November-December 2022.

Journal URL : https://ijsrcseit.com/CSEIT22862