JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 2, 2022, p. 844 - 850
https://publishoa.com
ISSN: 1309-3452

Image Classification for Feature Selection Using Radial Basis Function Neural Network for Classification (RBFNNC)

[1]Kumar Siddamallappa U
Research Scholar, Department of Computer Science & Engineering
Dr.A.P.J. Abdul Kalam University,
Indore -452016 (India).
Kumarkumarsm@gmail.com

[2]Nisarg Gandhewar
Assistant Professor, Department of Computer Science & Engineering
Dr.A.P.J. Abdul Kalam University,
Indore -452016 (India).
nisarg.gandhewar@gmail.com

ABSTRACT

This paper shows a new way to use a combination of filter feature selection algorithms for image classification. In pattern recognition, machine learning, and computer vision, feature selection is one of the most important problems. The main goal of feature selection is to classify images, improve how well they can be classified, and make the whole process easier to understand. The only method that is guaranteed to find the best subsets is the exhaustive search method, but it takes a lot of time to run. A new adaptive and hybrid approach to selecting features is proposed. This approach combines and uses different methods to make a more general solution. Several state-of-the-art feature selection methods are described in detail with examples of how they can be used, and a thorough evaluation is done to compare their performance to that of the proposed approach. The results show that the individual methods for selecting features perform very differently on the test cases, but the combined algorithm always gives a much better answer.

Keywords: Image classification, Feature Selection, Hybrid approach.

INTRODUCTION

Today, a huge amount of data is being collected and stored in real-world databases at a very fast rate in many different fields. As the amount of information stored grows, it doesn't mean that it's getting easier to understand and use. The users also want more advanced information. So, in data mining, feature selection helps to get the right information out of a lot of data. Out of all the features that could be used, only a subset of the relevant ones are chosen from the data being mined. By removing irrelevant and redundant features and reducing the number of dimensions, the algorithm for data mining was able to make better predictions. There are three main ways to choose which features to use: the filter, the wrapper, and the embedded. Without using a classification algorithm, the filter approach keeps as much useful information as possible from the whole set of attributes. Due to how well it works with computers, this method is quite popular, even for large datasets. On the other hand, it takes less work to use and the features that are chosen aren't as good. In the wrapper method, on the other hand, attribute selection is done by taking the classification algorithm and applying it to the selected attributes. This method chooses a subset of attributes that is best for a given algorithm, but it takes too much time and computing power for large-dimensional data. Lastly, in the embedded approach, the different evaluation criteria are used in different search phases to use the best parts of both approaches. The embedded approach can get the accuracy of a wrapper method while moving as quickly as a filter method. Image retrieval system algorithm were used to browse, search, and get relevant images from large datasets. Due to progress in the medical field, medical centers now use a lot of high-tech medical equipment to make a huge number of medical images. So, one of the most active areas of research in the medical field is content-based medical image retrieval, which helps people learn more about the field. So that, for analysis, the similarity in diagnosis methods can be synchronized by comparing the patient's current medical image with the medical database to get the medical diagnosis history.

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 2, 2022, p. 844 - 850
https://publishoa.com
ISSN: 1309-3452

The goal of a feature selection algorithm is to find the relevant set of features that gives the best rate of recognition with the least amount of computing work. As features with a high number of dimensions can make a system more complicated, this can also lead to a higher recognition rate. There needs to be a complex feature extraction algorithm that doesn't depend on other features. So, it is important to choose a subset of the best features. In the training phase, we need to come up with a learning model that lets us choose the most useful features on the fly. In this paper, instead of a very complicated model, we use a genetic algorithm that is based on the minimum description length principle (GA). The genetic algorithm is a search heuristic that is often used to find answers to problems. In a GA that tries to find better solutions, the population of candidate solutions (feature vectors) is improved step by step. The iteration process, which is also called "generation," usually starts with a group of candidate solutions that were made by chance. In each iteration, an objective function is used to judge how well each candidate solution fits the problem. This helps find the best solution. So, the next time through, the objective solution is used to make a new set of candidate solutions. This algorithm stops when the population has reached a level of fitness that is good enough or when a certain number of generations have been made.

There is a big increase in the amount of data being made because it comes from so many different places. This makes it hard to make information that is useful. Given that more and more data is being processed by apps on devices that can connect to the internet, there is a systematic need to store data. Feature selection has many uses, such as cutting down on operational time by getting rid of unnecessary or redundant attributes, improving classification accuracy, and making the classifier's organizational structure or model definition less complicated. Classification techniques are greatly affected by how well the clusters work and how much more data can be collected. In the discovery of information process, the main goal of the data pre-processing stage is to make datasets available for data mining algorithms. With the fast growth of computer technology, the amount of data keeps growing at an exponential rate. But useful information is hard to come by, so people want to know how to quickly and efficiently pull useful information and models from a large amount of data. Data mining is a method that uses databases, machine learning, pattern recognition, statistics, and other fields of study.

Figure 1.1 Hybrid Feature Selection

The technology takes a large amount of data and pulls out potential and useful information. It has become one of the most popular ways to analyze data over time. Data pre-processing is very important if you want to get the most out of data

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 2, 2022, p. 844 - 850
https://publishoa.com
ISSN: 1309-3452

mining tools. In data pre-processing, one of the most important and often used techniques is called "feature selection." It cuts down on the number of features by getting rid of redundant or useless data. This makes the data mining algorithm faster and better at making predictions. The process of feature selection is to look for the best subset of the initial feature of features. A typical feature selection process has four main steps: creating subsets, evaluating subsets, deciding when to stop, and validating the results. Firstly, the process of making subsets is mostly a heuristic search. And it depends on where you start your search and how you search. The candidate subsets in the data set with feature dimensions show that the search space is proportional to the feature dimension in an exponential way. So, to make the search process easier to compute, the search method is split into three parts: the complete search, the sequential search, and the random search.

LITERATURE REVIEW

Mera-Gaona and others (2021) Feature selection algorithms were made so that large datasets could be broken down into smaller subsets. Ensemble feature selection algorithms are a new option that can be used to help put together feature selection algorithms. Several experiments were done to show how well the framework worked. It finds relevant features by using either single feature selection (FS) algorithms or methods for selecting multiple features at once. The results of their experiment show that the ensemble feature selection worked well across the three datasets they used. Even though using feature selection methods as a pre-processing step before the classification task can be helpful, many of the feature selection techniques that are already known cannot be used directly in a hierarchical classification scenario. In the first attempts to solve the feature selection problem for the hierarchical classification problem, it was suggested to use traditional feature selection techniques and build classifiers by breaking the hierarchical classification problem into several flat classification problems. Researchers were able to use feature selection techniques and classification algorithms that are usually used in flat classification with this type of approach. It's important to note that none of them tried out global hierarchical classifiers in their experiments. Other ranked-based methods have suggested changing some popular filter feature selection algorithms to take into account the way classes are organized in a hierarchy.

Marie-Saintea et al. (2020) came up with a new feature selection method for Arabic speaker recognition systems that was based on the firefly algorithm and was inspired by nature. One way to solve nonlinear optimization problems is to use the Firefly algorithm. They showed that this method works to improve performance while making the system easier to use. Classification predicts the class label(s) of examples based on the problem domain represented by their features. In the literature, there are different levels of how hard classification problems are. In traditional (flat) classification problems, each dataset instance is labelled with one or more class labels, and the classes are not related to each other. But in many real-world applications, there are more complicated classification problems where classes that label instances are organized into a hierarchical structure shown by a tree or a directed acyclic graph (DAG). These are called hierarchical classification problems. The goal of feature selection is to find as many useful features as possible while reducing the cost of processing data. Feature selection is usually the first step in data mining tasks. In this paper, we'll talk about how to choose features for a classification task. So, we only looked at datasets with labelled examples. Some of the benefits of feature selection are that it improves the accuracy of classifiers and makes classification faster.

Hartmann and others (2019) the goal of feature selection is to find a subset of features that can be used to build a strong learning model. There is a strong link between a small number of the millions of terms and the news category that was being looked for. It solves the problem of figuring out how many features should be chosen. Choosing the best set of features is a key part of text classification that works well and quickly. In general, redundant and irrelevant features don't improve how well a learning model works. Instead, they make the model make more mistakes as it learns. Flat or hierarchical classification can be done with classification methods (using a local or global model approach). In flat classification, the methods don't care about the order of the classes. Instead, they only look at the classes of the leaf nodes when making predictions. In the local model approach, the class hierarchy is looked at from a local point of view using a mix of classifiers that look at different parts of the hierarchy separately. Based on this, we can divide local model approaches into groups based on how they use the local information of the hierarchical structure and how they build their classifiers around it. The global model approach only uses one classifier, which means that it builds a single model that takes into account the whole class hierarchy.

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 2, 2022, p. 844 - 850
https://publishoa.com
ISSN: 1309-3452

RESEARCH METHODOLOGY

The proposed method Radial basis function neural network for classification (RBFNNC) algorithm starts with a whole dataset and cuts it down using a combination of popular filter methods to get a minimal dataset with only useful features. The main benefit of this method is that you don't need to know anything about the process you're looking at ahead of time. First, a combination of filter feature selection methods is used to cut down on the number of variables. Then, the exhaustive search technique is used in a reasonable amount of time to find a subset of features that is not the best. The idea was to cut down on the feature of features and then use the exhaustive search.

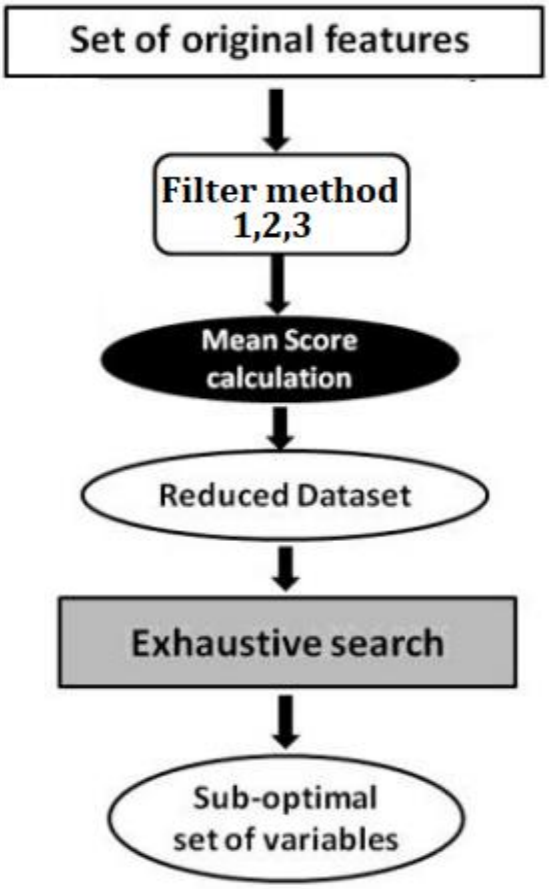

Figure 2. Flow Diagram

First, the proposed methods combine forward selection and backward elimination techniques to reduce the number of features. This makes it possible for the exhaustive search technique to be used in a reasonable amount of time. In this paper, we suggest using filter feature selection methods to narrow down the feature space before using an exhaustive search. Filter approaches are better than wrappers because they don't depend on the learning machine and are faster. This makes them much easier to use from a computational point of view. Exhaustive search is a part of the wrapper approach because it uses the performance of the classifier to choose the variables that have the most effect on the target. The proposed method is a hybrid filter-wrapper method. Each variable is given a score by each filter method. Scores are first normalized so that they have values in the range [0, 1], and then the mean of the three scores is used to combine them. A threshold decider, whose value is set to the average score, is used to pull out the interesting variable and reduce the size of the dataset. The exhaustive search is done on the dataset, which only has the variables that were chosen before. The last step is to find the combination of variables that gives the best accuracy for the classifier.

Wrapper methods treat the classifier like a black box and use its performance to choose which variables to use to get more accurate results. On the other hand, wrappers have a high chance of over fitting and require a lot of processing power. This

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 2, 2022, p. 844 - 850
https://publishoa.com
ISSN: 1309-3452

is especially true if the datasets are large or if the classifier that was built has a high computational cost. Embedded methods are cheaper to run than wrappers, but they are too specific for a given classifier to be useful. Reduced dataset is split into two parts: 75% of the data is used to train the system, and the remaining 25% is used to test the classifier (validation set). In this paper, the Bayesian classifier was used for the experimental tests. This is a classifier that can handle large datasets without taking a lot of time to run. The following indices are used to measure how well the classifier works:

- True Positive (TP) is the percentage of correctly categorized unitary samples.
- True Negative (TN) is the percentage of correctly categorized null samples.
- False Positive (FP) is the percentage of incorrectly categorized null samples as belonging to the unitary class.
- False Negative (FN) is the percentage of incorrectly categorized unitary samples as belonging to the null class.

Precision = (TP + TN) / (TP + TN + FP + FN)

The traditional definition of accuracy shows how well a classifier works by calculating the probability of correctly classifying patterns, no matter what class they belong to. But it's not a good way to measure data that isn't balanced, which happens a lot in real life, especially when the rate of imbalance is high.

RESULTS AND DISCUSSION

Several datasets from the two industries have been used to show that the proposed method works. Datasets from the public repository are chosen based on their different dimensions so that the proposed technique can be shown to work in different situations. Figure 3 shows how the combined filter method was used on the variables that were chosen, while the results were made by first using the filter method and then doing an exhaustive search on the variables that had been chosen. Lastly, the results from the exhaustive search were used to compare the results from the proposed hybrid method with the best case.

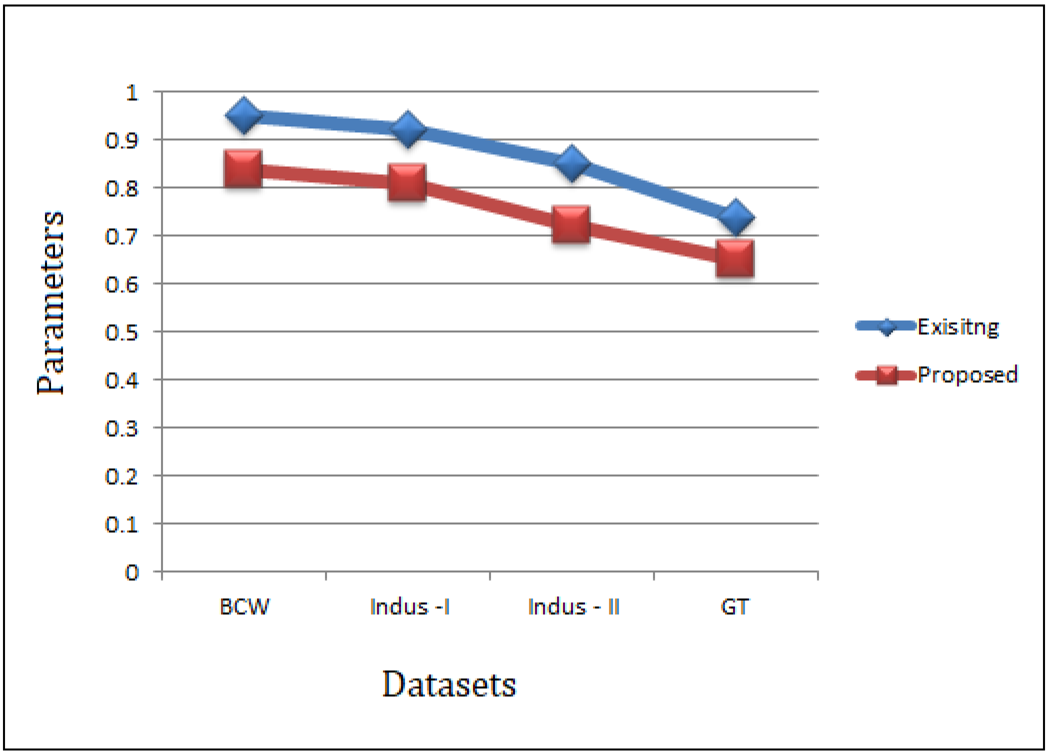

Figure 3. Comparison results

We can see that the results for the proposed hybrid method are the same for all four datasets (BCW, GT, Industrial I, and Industrial II). Also, in these situations, it is enough to use the combined filter method, which saves a lot of time. On the other hand, where the exhaustive search is used, it gives more accurate results and a better list of interesting variables. When we compare the proposed hybrid method to the exhaustive method, we can see that the performance is the same in four of the cases. In the other problems that were tested, the proposed method worked better than the exhaustive approach. However, it's important to keep in mind that the proposed method reduced the amount of work that needed to be done on the computer. Filter methods are appropriate and effective ways to deal with very large datasets. They are easy to use, fast, and don't depend on the algorithm used. Wrapper methods treat the classifier like a black box and use its performance to choose which variables to use to get more accurate results. On the other hand, wrappers have a high chance of over fitting and require a lot of processing power. This is especially true if the datasets are large or if the classifier that was built has a high computational cost. Embedded methods are cheaper to run than wrappers, but they are too specific for a given classifier to be useful.

CONCLUSION

A hybrid algorithm called RBFNN is shown for selecting features. The main idea is to use a combination of several filter methods to choose which features to use, and then, once the original dataset has been reduced, to use an exhaustive search to find the best combination of the remaining features in a reasonable amount of time. This method can be used on all datasets without making any assumptions about the data ahead of time. It can also be used on large datasets. Several datasets from a public repository and two datasets from an industrial setting have been used successfully with the proposed method. The results that were achieved show that it worked. In the future, work will be done to see how the approach can be used for other things, like clustering, making predictions, or classifying things into more than one classification. Also, the threshold will be chosen automatically to make the learning machine more accurate.

REFERENCES:

1) Mera-Gaona M, López DM, Vargas-Canas R, Neumann U. 2021." Framework for the ensemble of feature selection methods". Applied Sciences 11(17):8122–8138.
2) Marie-Saintea SL, Alalyani N. 2020. "Firefly algorithm based feature selection for Arabic text classification". Journal of King Saud University - Computer and Information Sciences 32(3):320– 328.
3) Hartmann J, Huppertz J, Schamp C, Heitman M. 2019." Comparing automated text classification methods". International Journal of Research in Marketing 36(1):20–38.
4) Cateni S., Colla V. and Vannucci M. "A Genetic Algorithm-based approach for selecting input variables and setting relevant network parameters of a SOM-based classifier". International Journal of Simulation Systems, Science & Technology. UKSim 4th European Modelling Symposium on Mathematical modelling and computer simulation, Vol.12, N°2, April 2011.
5) Cateni S., Colla V. and Vannucci M. "Variable Selection through Genetic Algorithms for classification purpose". IASTED International Conference on Artificial Intelligence and Applications, 2010, Innsbruck Austria, 15-17 February 2010.
6) S. Cateni, M. Vannucci, M. Vannocci and V. Colla. "Variable Selection and Feature Extraction through Artificial Intelligence Techniques. Multivariate Analysis in Management", Engineering and the Sciences, MSc. Leandro Freitas, Intech 2013.
7) Chinnaswamy A, Srinivasan R (2016). "Hybrid feature selection using correlation coefficient and particle swarm optimization on microarray gene expression data". In: Innovations in bio-inspired computing and applications, Springer, New York, pp 229–239
8) Najdi S, Gharbali AA, Fonseca JM (2016) "A comparison of feature ranking and rank aggregation techniques in automatic sleep stage classification based on polysomnographic signals". In: International conference on bioinformatics and biomedical engineering, Springer, New York, pp 230–241.
9) Satopaa V, Albrecht J, Irwin D, Raghavan B (2011). "Finding a kneedle in a haystack: Detecting knee points in system behaviour". In: 2011 31st international conference on distributed computing systems workshops, IEEE, pp 166–171.

10) Smetannikov I, Deyneka A, Filchenkov A (2016). "Meta learning application in rank aggregation feature selection". In: 2016 3rd international conference on soft computing and machine intelligence (ISCMI), IEEE, pp 120–123.
11) Toğaçar M, Ergen B, Cömert Z (2020). "Detection of lung cancer on chest ct images using minimum redundancy maximum relevance feature selection method with convolutional neural networks". Biocybern Biomed Eng 40(1):23–39.
12) Abualigah, L. M., Khader, A. T., & Al-Betar, M. A. (2016). "Unsupervised feature selection technique based on harmony search algorithm for improving the text clustering". 2016 7th International Conference on Computer Science and Information Technology (CSIT), 1–6.
13) Das, S., Singh, P. K., Bhowmik, S., Sarkar, R., & Nasipuri, M. (2016). "A Harmony Search Based Wrapper Feature Selection Method for Holistic Bangla Word Recognition". Computer Science, 89, 395–403.
14) Krishnaveni, V., & Arumugam, G. (2013). "Harmony search based wrapper feature selection method for 1-nearest neighbour classifier". International Conference on Pattern Recognition, Informatics and Mobile Engineering, PRIME 2013, 24– 29.