IJESRT

INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY

FEATURE SELECTION

Prof. Purushottam Das*, Prof. Ankur Singh Bist
*Graphic Era Hill University, India
KIET, Ghaziabad, India

ABSTRACT

Feature selection is one of the important issues in the domain of system modelling, data mining and pattern recognition. Subset selection evaluates a subset of features as a group for suitability prior to applying a learning algorithm. Subset selection algorithms can be broken into wrapper, filter and hybrid categories. Literatures surveyed related to this are given as follows..

KEYWORDS: Feature Selection.

INTRODUCTION

Feature Selection:-

Liu and Motoda, (1998) wrote a book on feature selection. This paper give an overview of the methods developed since the 1970s. They also gave a general framework in order to examine these methods and categorize them. This book discussed the importance of feature selection algorithms with the help of various simple examples and compared those methods using different datasets. Demonstrations were given in this book using different feature selection algorithms under various circumstances.

FILTER APPROACH

Filter approaches are based on the information measures. Class discrimination capability of the feature subset is assessed using the intrinsic properties of data only. In present work emphasis is being placed on feature selection by filter based approaches and applications. Thus, the related work surveyed is being presented in this section.

Mutual information based approaches

Battiti (1994) adopted a heuristic criterion for approximating the ideal solution. Instead of calculating the joint MI between the selected feature set and the class variable, only $\mathrm{I}(\mathrm{C};\ \mathrm{f}_{\mathrm{i}})$ and $\mathrm{I}\left(\mathrm{f}_{\mathrm{i}}, \mathrm{f}_{\mathrm{j}}\right)$ are computed, where f_{i} and f_{j} are individual features, C is the class and I is the measure of mutual information. Battiti's mutual information feature selector (MIFS) selects the feature that maximizes the information about the class, corrected by subtracting a quantity proportional to the average MI with the previously selected features. Another variant of Battiti's MIFS is the min-redundancy max-relevance (mRMR) criterion.

Kwak and Choi (2002). Analysed limitations of mutual information feature selector (MIFS) and means for overcoming these limitations. They proposed two feature selection algorithms. In first approach mutual information between input attributes and output classes was used. Accuracy of the mutual information depends on the performance of a feature selection algorithm. In other approach Taguchi method was used as feature selection algorithm. This method was applied to several classification problems and compared with MIFS. Experimental observation has shown that combined algorithm performs better.

Estévez *et al.* (2009) proposed a filter method of feature selection known as normalized mutual information feature selection (NMIFS). NMIFS is an enhancement over Battiti's MIFS, MIFS-U, and mRMR methods (Battiti, 1994). NMIFS outperformed MIFS, MIFS-U, and mRMR on several artificial and benchmark data sets. They introduced the normalized MI, as a measure of redundancy, to reduce the bias of MI toward multi-valued attributes and restrict its value to an interval. Further, NMIFS was combined with a genetic algorithm to form a hybrid filter/wrapper method called GAMIFS having an initialization procedure and a mutation operator. Mutation operator was used to speed up the convergence of the genetic algorithm. GAMIFS overcomes the limitations of incremental search algorithms that were unable to find dependencies between groups of features.

Other approaches in filter method

Doak (1992) proposed an approach using the concept of sampling. Evaluation of sample is important to check which results are better, samplings before feature selection or after feature selection. Sampling was performed on highly imbalanced data. The after scenario demonstrated more stable performance than before scenario using various sampling techniques. An empirical investigation of feature selection on imbalanced data was presented. They experimented with six feature selection techniques and three data sampling methods. The before (denoted BEF) and after (denoted AFT) situations are compared for each given ranking technique and sampling method. The average over the ten runs of five-fold cross-validation outcomes was represented by each result.

Model *et al.* (2001) proposed a method for microarray based methylation analysis combined with supervised learning techniques to predict known tumour classes. The resulting filters were evaluated using an application oriented fitness criterion based on SVMs.

Yu and Liu (2003) introduced a novel concept based on correlation known as Fast Correlation Based Filter selection (FCBF). FCBF was found to be an efficient way of analyzing feature redundancy for high dimensional data and handling data of different feature types. Experiments were performed to implement, evaluate as well as compare FCBF with other feature selection algorithms. The feature selection results were compared by applying various classification algorithms.

Swiniarski and Skowron (2003) proposed an approach for feature selection which is based on Rough set method and PCA. This approach has important role in categorical clustering. The proposed approach was used with neural network. The results of principal components analysis (PCA) were used for feature projection and reduction. Experimental evaluation was made for face and mammogram recognition problem. The sequence of data mining steps was also proposed that included applications of SVD, histograms, PCA, and rough sets for feature selection.

Rogati and Yang (2003) presented a technique for text classification. This approach suggested that filter methods, which includes the statistics, were consistently better across classifiers and performance measures.

Sun *et al.* (2005) proposed an Evolutionary Gabor Filter Optimization (EGFO) approach for on road vehicle detection using filter optimization. It produced optimal problem-specific set of Gabor filters. EGFO approach aggregated filter design with filter selection by integrating genetic algorithms (GAs) with an incremental clustering approach. Improvement in the performance of on-road vehicle detection was achieved by applying a set of Gabor filters particularly optimized for the task of vehicle detection.

Deng *et al.* (2005) proposed a novel Facial Expression Recognition (FER) system based on Gabor feature and PCA + LDA. They used Gabor filter bank for feature extraction. A minimum distance classifier was designed and employed to evaluate the recognition performance in different experimental conditions.

Hall (2006) proposed a method for decision tree attributes. Decision tree based attribute filter for Naive Bayes has performance comparable with wrapper based feature subset selection for Naive Byes.

Blachnik *et al.* (2008) presented a Kolmogorov-Smirov Class Correlation Based Filter (K-S CCBF) approach that was based on fast and computationally efficient feature ranking. It is a fast redundancy removal filter approach, which utilizes class label information. It was compared with other methods, appropriate for removing redundancy, such as the simple ranking based wrapper, and Fast Correlation-Based Feature Filter (FCBF). In comparison with basic K-S CBF, results obtained do not differ significantly and were better than the results of FCBF algorithm. Initial space in wrapper-based feature selection can be significantly reduced by the proposed algorithm for high-dimensional problems. Improvement over K-S CBF is demonstrated by K-S CCBF for a few datasets such as Spam, Sonar, and Ionosphere.

Wrapper approach

Wrapper approach uses the induction algorithm as a part of the evaluation function, the same algorithm that will be used to induce the final classification model.

Kohavi and John (1997) compared the wrapper approach to induction without feature subset selection and Relief (a filter approach) to FSS. They provided a number of disadvantages of the filter approach steering research towards algorithms adopting the wrapper approach. Their approach search for an optimal feature subset adjusted to a particular learning algorithm and a particular training set.

Dash and Liu (1997) categorized several feature selection algorithms after analyzing many existing algorithms. Typical feature selection process includes four steps: generation procedure, evaluation function, stopping criterion, and validation procedure. They asserted that generation procedure can be grouped into three categories: Complete, heuristic, and random. Evaluation function can be grouped into five categories: Distance, information, dependence, consistency, and classifier error rate measures. Thirty two methods for feature selection were categorized on the basis of combinations of evaluation function and generation procedure.

Yang and Honavar, (1998) used genetic algorithm for feature subset selection in automated design of pattern classifiers. They presented a simple, inter-pattern distance based polynomial time constructive neural network algorithm. They compared this algorithm, very favourably, with computationally more expensive algorithms, in terms of generalization accuracy.

Hybrid approach

Hybrid approach is presented to overcome the weakness of filter and wrapper approaches. Many researchers combined both the methods together to improve the results. The hybrid approach is computationally more effective than wrapper approach and provides higher accuracy than filter approach.

Xing *et al*. (2001) applied feature selection methods (using a hybrid of filter and wrapper approaches) to a classification problem in molecular biology. This problem involves only 72 data points in a 7130 dimensional space. They searched for regularization methods as an alternative to feature selection. They showed that feature selection methods were preferable in the undertaken problem domain.

Das (2001) proposed a Boosting Based Hybrid for Feature Selection (BBHFS) which improves the performance of learning algorithms and performs better than wrapper methods on DNA dataset. This method included some features of wrapper methods and uses boosting. This method used boosting into a filter method and covers few advantages of wrappers such as natural stopping criterion. BBHFS is competitive with wrapper methods and much faster in selecting feature subsets. BBHFS performed better than wrapper methods on Chess dataset and on the DNA dataset.

Oh *et al.* (2004) proposed a novel hybrid GA to solve the feature selection with the goal of achieving leading-edge performance over the conventional algorithms. Local search operations are divided and embedded in hybrid GAs to fine tune the search. The hybrid GAs showed better convergence properties compared to the classical GAs. Significant improvements were observed through the proposed hybrid GAs. Hybridization offers the acquisition of subset size control. The concept of atomic operations has proven to be useful in rigorously analyzing and comparing the timing efficiencies of the algorithms.

Feature Selection Using GA

Genetic algorithms are a promising option to conventional heuristic methods. Genetic algorithms are stochastic search methods that work on the theme of natural biological evolution. GA work with a set of candidate solutions called a population and the GA obtains the optimal solution after a series of iterative computations. Literature surveyed related to this topic is given in this section.

Man *et al.* (1996) proposed a framework based on GA. Specific conditions were explained where GA was used as an optimization technique. The essential building block hypothesis and schema theory of genetic algorithms were given for the benefit of new researchers of this particular field. The advantages of GA were its working model and way of improving its evolution. In order to improve performance of GA a range of structural modifications were suggested. The problem formulation, genetic functionality of operators of GA, the inherent capability of GA was explained for solving conflicting and complex problems. Industrial application model is the basis of the described framework.

Whitley (2001) provided an overview of evolutionary algorithms covering genetic algorithms, evolutionary strategies, genetic programming and evolutionary programming. Gray codes, bit representation and real-valued representations were used for parameter optimization problems. These representations were well discussed in this paper.

Tan *et al*. (2008) proposed a framework based on genetic algorithm (GA) and used for feature subset selection. This framework has combination of various existing feature selection methods. Small subsets of features were found to build the classifier using a particular inductive learning algorithm. Experiments were performed on three data sets using three existing feature selection methods. This approach is effective and robust for finding subsets of features with smaller size and/or higher classification accuracy as compared to each individual feature selection algorithm.

CONCLUSION

This paper contains the basic approaches for feature selection. This study will be helpful for those working in the field of image processing.

REFERENCES

1. **Aczél, J. and Daróczy, Z. 1975**. *On Measures of Information and Their Characterizations*. New York: Academic.
2. **Aeberhard, S.; Coomans, D. & De Vel, O. 1992.** Comparison of classifiers in high dimensional settings. *Dept. Math.*

ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 2.114

Statist., James Cook Univ., North Queensland, Australia, Tech. Rep, pp. 92-02.

3. **Arefi, A.; Motlagh, A.M. & Teimourlou, R.F. 2011**.Wheat class identification using computer vision system and artificial neural networks. *InternationalAgrophysics*. *25*(4).
4. **Battiti, R. 1994.** Using mutual information for selecting features in supervised neural net learning. *IEEE Trans. Neural Netw.* 5(4): 537–550.
5. **Bins, J. and Draper, B. 2001.** Feature selection from huge feature sets. *In*: *Proc. Int. Conf. Comput. Vis.*, Vancouver, BC, Canada, pp. 159–165.
6. **Blachnik, M.; Duch, W.; Kachel, A. & Biesiada, J. 2008.** Feature Selection for Supervised Classification: A Komogorov Smirov Class Correlation based Filter. *In*:*AIMeth, Symposium on Methods of Artificial Intelligence, Gliwice, Poland*.
7. **Brill, F.; Brown, D. & Martin, W. 1992**. Fast genetic selection of features for neural networks classifiers. *IEEE Trans. Neural Netw.* 3(2): 324–328.
8. **Cover, T.M. and Thomas, J.A. 2006**. *Elements of Information theory*. Entropy, Relative Entropy and Mutual Information. John Wiley & Sons, Incl. Print ISBN 0-471-06259-6, online 0-471-20061-1.
9. **Dash, M. and Liu, H. 1997**. Feature selection for classification. *Intell. Data Anal.* 1(3): 131–156.
10. **Dash, M. and Liu, H. 2003.** Consistency-based search in feature selection. *Artif. Intell. J.*, 151: 155–176.