

# An Incremental Feature Subset Selection Strategy Based on Convergency

Jian Shen<sup>1\*</sup>, Jingbo Xia<sup>2</sup>, Haiou Shen<sup>1</sup>

<sup>1</sup>Institute of Information and Navigation, Air Force Engineering University, Xi'an, Shaanxi, China

<sup>2</sup>Tan KahKee College, Xiamen University, Xiamen, Fujian, China

\*Corresponding author

**Abstract**—Traffic classification is currently a significant challenge for network monitoring and management. Feature selection based on machine learning is an effective method to realize dimension reduction and decrease redundant information. To classify traffic flows with better performance, we put forward the incremental strategy of convergence. The strategy gathers all the features that have been selected and adds an extra round of selection on the base of the original algorithm to discover the value of relationship among all the selected features. The performances are examined by experiment. Our theoretical analysis and experimental observations reveal that the proposed incremental strategy of convergence makes a further improvement on the classification accuracy.

**Keywords**—network traffic classification; feature subset selection; incremental strategy

## I. INTRODUCTION

As the foundation of network cognition, management and optimizing, the classification of network traffic is making a significant difference in resource scheduling, safety analysis and future tendency prediction [1]. With the developing of machine learning methods, the problem of classification can be solved more accurately and efficiently. Feature subset selection (FSS) plays an important role in classification problems based on machine learning, especially dealing with high-dimensional data like network traffic flows [2]. Most of the input traffic features contain irrelevant or redundant information which may degrade the classification accuracy and increase the computational complexity. The purpose of FSS method is to extract the exact feature subset that can represent the whole feature dataset with fewer variables.

The development of FSS approach has two major directions, which are filters and wrappers [3]. The filters evaluate the features by their intrinsic properties. As filters are independent from the specific classifier, they have the advantages of general and highly computational efficient. However, the relationships between the features are hardly to be discovered, which may lead to imperfect classification result. Wrappers employ the information of the classifier to find the best feature subset with much more classification accuracy [4]. They can explore the relationship among features, but at the price of expensive computation on the evaluation of feature space. Moreover, the classification performance of the feature subset selected by wrappers is strictly related to the specific classifier, and it is hard to get similar performance with other classifiers.

In this paper, we aim to improve the efficiency and accuracy of the existing FSS methods. On the base of the original FSS algorithm, we come up with the incremental strategy of convergence to add another round of selection to update the feature subset of better classification performance. This strategy collects all the features that have been selected and evaluates the individual and mutual information to pick out a better feature subset.

The rest of the paper is organized as follows. The following section presents a set of FSS methods in detail. We describe the incremental strategy of convergence in section 3. The 4th section contains the experiment and the corresponding analysis. In section 5 we provide a summary of the paper and future work.

## II. PREVIOUS WORK

### A. Hybrid Filter-wrapper Feature Subset Selection

Hybrid method is a more recent approach and a promising direction in the feature selection field [5, 6]. It uses the ranking information obtained using filter methods to guide the search in the optimization algorithms used by wrapper methods.

Recently the literatures have contained numerous references to the use of hybrid selection algorithms. Xie included a feature ranking in a sequential forward search method with the application of the F-score measure to rank the features [7], while Peng added a random sampling method to choose features from the ranking [8]. Zhang [9] and Bonilla-Huerta [10] proposed the similar methods including a Relief estimation based ranking, which were also applied to compress the searching space.

Except for the good classification performance of hybrid methods on some aspects, the drawback is also obvious. The complexity of the wrapper search in hybrid methods is still intractable because of the large consumption of wrapper evaluations.

### B. Development of IWSS

Incremental wrapper-based feature subset selection (IWSS) was presented by Ruiz R as a canonical method [11]. The features that satisfy the relevance criterion would be selected into the feature subset. The advantage of this method is the low complexity of  $O(n)$  in the number of wrapper evaluations. However, the problem is that the features are kept into the

subset once they are selected, even those which have poor performance of correlation with the latter selected ones. As a result, the size of selected feature subset cannot be restricted in a small range.

The algorithm of IWSS with replacement (IWSSr) was proposed to alleviate this problem [12]. During the selection of a feature, the addition of the feature into the subset is not the only matter that is concerned, but also the interchange with one of the included features. As a result, the feature that has already been selected would be drove away from the subset when the performance of its combination with the latter selected features gets worse. We can find that the selected subset is more representative, but it is also obvious that all the features are disposed once at a time. Therefore, the process is inefficient to deal with high-dimensional problems.

Bermejo P put forward the method of IWSS with re-ranking (IWSSr<sup>R</sup>) by the introduced concept of block [13]. The proposed algorithm can select features at the block level and improve the efficiency of the algorithm greatly. The size of block is initialized at the beginning and cannot be changed during the procedure of feature selecting. The size of block is also required to be large enough to give some freedom to the wrapper algorithm, which may increase extra evaluations.

The algorithm IWSS based on sliding block (IWSS-SB) composes of two stages, which are ranking and iterated feature selection. The stage of ranking requires  $O(n)$  filter evaluations. In this study, Information Gain (IG) which is one of the fastest attribute ranking methods is used to evaluate the predictive attributes. The stage of feature selection starts from the first block of features. The features selected from each round of selection by the wrapper method are added into the feature subset  $S$ . Then, the wrapper method runs again over the next block of features but taking into account the features already included in the feature subset  $S$ . That means the algorithm does not only select new features into the subset, but also interchange with the ones that already included in  $S$  with redundant information. If the classification accuracy of the newly updated feature subset cannot exceed the former one, the feature subset will remain unchanged. After every round of selection, sliding block will decide whether to attenuate according to the rules as below. Sliding block is defined as the number of features disposed at a time which is dynamically changing according to the result of feature selection. It is initiated at the beginning of the algorithm like IWSSr<sup>R</sup>. Under certain circumstances as follows, the size of sliding block attenuates to the half.

- 1) The density of features selected in the blocks decreases.

$$D_n < D_{n-1}, D_n = \frac{|S_n| - |S_{n-1}|}{B_n} \quad (1)$$

where  $D_n$  is the density of features selected,  $B_n$  and  $S_n$  refer to the size of block and the number of all the selected features at the  $n$  round of feature selection.

- 2) The classification accuracy of the feature subset decreases.

It is clear that the attenuation of sliding block is the penalty strategy to save evaluations when the density of features or classification performance goes down. Then, the block moves to the next part of the rank. This process keeps iterating until no more feature is added into the feature subset or sliding block attenuates to zero.

### III. INCREMENTAL STRATEGY OF CONVERGENCE

From the analysis of a set of hybrid filter-wrapper feature subset selection methods, it can be seen that the features in the final subset are updated by several rounds of iterations. The features that have been selected into the feature subset are the ones with more information as the advantage. During the following rounds of iteration, some of them are interchanged by the others. However, the displaced features may play a positive role in classification correlated with the ones in the feature subset.

Our goal is to collect the advantages of all the selected features including the displaced ones and make full use of them. So we propose the incremental strategy of convergence. All the features that have been selected into the feature subset are also copied into the subset of convergence  $S_c$ , including the ones that have been displaced. After all the iterations of feature selection, the final feature subset is not determined at the last round of selection, but selected again from the subset of convergence  $S_c$  with the wrapper methods. The added step of feature selection may cost more evaluations. However, compared with the evaluations carried out by the hybrid filter-wrapper methods, the number of extra evaluations is limited in a low level. Moreover, the improved performance of the classification can compensate the extra consumed computation.

### IV. EXPERIMENT

This paper applies traffic flow dataset of Moore to the experiment [14]. The dataset contains 12 traffic categories, which are WWW, MAIL, FTP-CONTROL, FTP-PASV, ATTACK, P2P, DATABASE, FTP-DATA, MULTIMEDIA, SERVICES, INTERACTIVES and GAMES. There are 24863 flows with 248 features in the dataset.

The experiment adopts Naivebayes as the classifier. The 10-fold cross-validation method is used to estimate the data. We take the average of the experiment which is conducted ten times as the result.

For the sake of clarity, the algorithm with the incremental strategy of convergence proposed in this work is referred as X-C, where X stands for certain feature selection algorithm. Fig 1 shows the classification accuracy of different algorithms. It can be seen that the classification performance of the algorithms with and without convergence strategy are the same, when the initialized size of block is set to be 10, 30 and 40. That is because the convergence strategy selects the exact same feature subset as the original method under such circumstances. In all other cases, we can see the obvious improvement on classification accuracy with the strategy of convergence on the base of the original FSS methods.

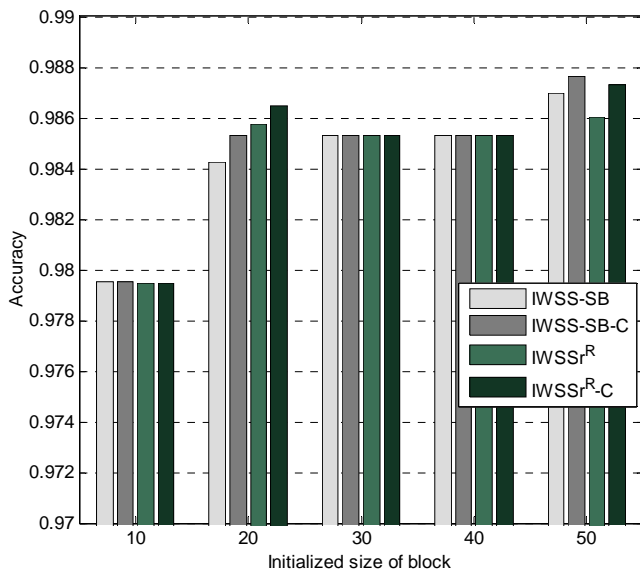


FIGURE I. CLASSIFICATION ACCURACY OF DIFFERENT ALGORITHMS

However, the strategy of convergence performs well in the classification with the price of extra evaluations. This strategy adds another round of selection on the base of the original algorithm as mentioned in section 3. The evaluations of different methods are recorded in table 1. We can find that the strategy of convergence consumes 6% to 8% more evaluations than the original method on average.

TABLE 1. EVALUATIONS OF ALGORITHMS WITH DIFFERENT INITIALIZED SIZE OF BLOCK

	IWSSr <sup>R</sup>	IWSSr <sup>R-C</sup>	IWSS-SB	IWSS-SB-C
<b>B<sub>i</sub>=10</b>	1397	1467	690	770
<b>B<sub>i</sub>=20</b>	4963	5344	1509	1688
<b>B<sub>i</sub>=30</b>	2797	2921	1741	1865
<b>B<sub>i</sub>=40</b>	2455	2580	1800	1925
<b>B<sub>i</sub>=50</b>	4665	4975	3463	3692
<b>mean</b>	<b>3255.4</b>	<b>3457.4</b>	<b>1840.6</b>	<b>1988</b>
<b>ratio</b>		<b>1.08</b>		<b>1.06</b>

To make the impact of the convergence strategy clear, we compare the classification accuracy of the two methods with and without convergence strategy as the number of consuming evaluations increases. The initialized size of block is still set to be 50. From figure 2, we can see that the classification accuracy increases quickly at the beginning and slows down in the following iterations. It shows that the classification performance is hard to enhance when it gets to the certain level. However, the strategy of convergence can make further improvement on the base of the original algorithms after the last round of selection. Above all, we can come to the conclusion that this proposed strategy may consume more evaluations, but the improvement it brings about on classification performance shows great efficiency.

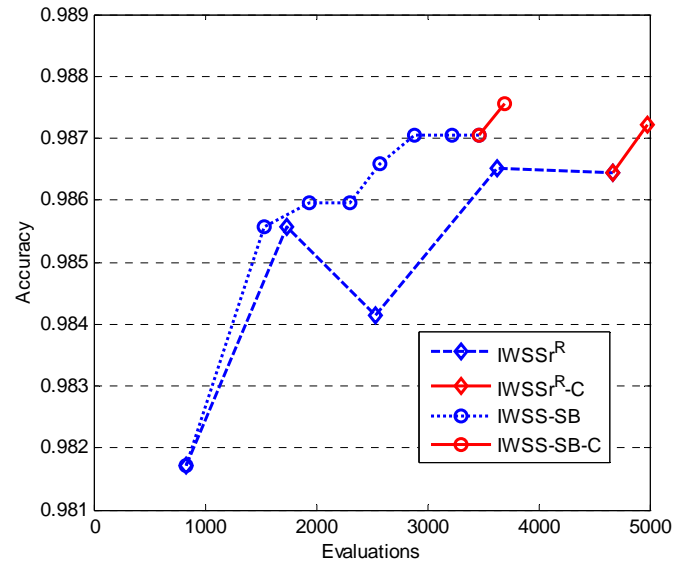


FIGURE II. CLASSIFICATION ACCURACY WHEN  $B_i=50$

## V. CONCLUSIONS

To achieve efficient classification of network traffic flows, we have proposed an incremental strategy of convergence which adds another round of selection to update the feature subset of better classification performance. The experiment showed that the strategy of convergence was presented to make further improvement on classification accuracy. We also found that the strategy could enhance the classification ability of the feature subset selected by different hybrid FSS methods. From the description and analysis, we can see that the size of block is an important factor to the performance of the proposed method. As future work, our research will be focused on the research of its impact.

## ACKNOWLEDGMENT

We would like to express our gratitude to all those who gave kind encouragement and useful instructions all through the writing. A special acknowledgement should be extended to the library assistants who supplied with reference materials of great value. We would like to thank the anonymous reviewers for their very helpful comments and feedbacks to improve the manuscript.

## REFERENCES

- [1] Khayari, R. R. Sadre, and B. R. Haverkort. "A validation of the pseudo self-similar traffic model." *International Conference on Dependable Systems & Networks* IEEE Computer Society, 2002:727.
- [2] Yang, Peng, and Q. Zhu. "Finding key attribute subset in dataset for outlier detection." *Knowledge-Based Systems* 24.2(2011):269-274.
- [3] Amiri, Fatemeh, et al. "Mutual information-based feature selection for intrusion detection systems." *Journal of Network & Computer Applications* 34.4(2011):1184-1199.
- [4] Kohavi, Ron, and G. H. John. "Wrappers for feature subset selection." *Artificial Intelligence* 97.1-2(1997):273-324.
- [5] Guyon, Isabelle, et al. "Feature Extraction: Foundations and Applications (Studies in Fuzziness and Soft Computing)." Springer-Verlag New York, Inc. 2006.

- [6] Fialho, André S., et al. "Predicting Outcomes of Septic Shock Patients Using Feature Selection Based on Soft Computing Techniques." *Information Processing and Management of Uncertainty in Knowledge-Based Systems. Applications -*, International Conference, Ipmu 2010, Dortmund, Germany, June 28 - July 2, 2010. *Proceedings DBLP*, 2010:65-74.
- [7] Xie, Juanying, and C. Wang. "Using support vector machines with a novel hybrid feature selection method for diagnosis of erythematosquamous diseases." *Expert Systems with Applications An International Journal* 38.5(2011):5809-5815.
- [8] Peng, Y., Z. Wu, and J. Jiang. "A novel feature selection approach for biomedical data classification. " *Journal of Biomedical Informatics* 43.1(2010):15-23.
- [9] Zhang, Li Xin, et al. "A novel hybrid feature selection algorithm: using ReliefF estimation for GA-Wrapper search." *International Conference on Machine Learning and Cybernetics IEEE*, 2003:380-384 Vol.1.
- [10] Bonillahuerta, Edmundo, et al. "Hybrid Filter-Wrapper with a Specialized Random Multi-Parent Crossover Operator for Gene Selection and Classification Problems." *Bio-Inspired Computing and Applications -*, International Conference on Intelligent Computing, Icic 2011, Zhengzhou, china, August 11-14. 2011, Revised Selected Papers *DBLP*, 2011:453-461.
- [11] Ruiz, Roberto, J. C. Riquelme, and J. S. Aguilar-Ruiz. "Incremental wrapper-based gene selection from microarray data for cancer classification." *Pattern Recognition* 39.12(2006):2383-2392.
- [12] Bermejo, Pablo, J. A. Gamez, and J. M. Puerta. "Incremental Wrapper-based subset Selection with replacement: An advantageous alternative to sequential forward selection." *Computational Intelligence and Data Mining*, 2009. *CIDM '09. IEEE Symposium on IEEE*, 2009:367-374.
- [13] Bermejo, Pablo, et al. "Fast wrapper feature subset selection in high-dimensional datasets by means of filter re-ranking." *Knowledge-Based Systems* 25.1(2012):35-44.
- [14] 'A. W. Moore, Dataset', <http://www.cl.cam.ac.uk/research/srg/netos/nprobe/data/papers/sigmetrics/index.html>, accessed Aug 2013