Hybrid Feature Selection Method Based on a Naïve Bayes Algorithm that Enhances the Learning Speed while Maintaining a Similar Error Rate in Cyber ISR

GyeongIl Shin1, Hosang Yooun 2, DongIl Shin1 and DongKyoo Shin 1
[1] Department of Computer Engineering, Se-jong University
98 Gunja-Dong, Gwangjin-gu, Seoul 143-747, South Korea
[e-mail: sgi@gce.sejong.ac.kr, dshin@sejong.ac.kr, shindk@sejong.ac.kr]
[2] Agency for Defense Development
Songpa P.O. Box 132, Seoul, 138-600 Korea
[e-mail: yun_hosang@add.re.kr]
*Corresponding author: DongIl Shin

Abstract

Cyber intelligence, surveillance, and reconnaissance (ISR) has become more important than traditional military ISR. An agent used in cyber ISR resides in an enemy's networks and continually collects valuable information. Thus, this agent should be able to determine what is, and is not, useful in a short amount of time. Moreover, the agent should maintain a classification rate that is high enough to select useful data from the enemy's network. Traditional feature selection algorithms cannot comply with these requirements. Consequently, in this paper, we propose an effective hybrid feature selection method derived from the filter and wrapper methods. We illustrate the design of the proposed model and the experimental results of the performance comparison between the proposed model and the existing model.

---

1. **Introduction**

$\mathbf{W}$e studied various learning algorithms that can be adopted by cyber intelligence, surveillance, and reconnaissance (ISR) agents [1]. The information needed to make a command order decision is very important factor in achieving superiority in battle. Victory depends on how much of your opponent's data you possess, along with data on the general situation of the friendly side [2,3,4].

The primary aim of this investigation is to deliver information that can help a commander make decisions within a limited time period, when an agent infiltrates a network in the cyber ISR process. Because it is impossible to collect and transmit all of the information within a limited amount of time, it is necessary to collect and transmit the important information first. Our proposed cyber ISR operational concept is illustrated in **Fig. 1**.

The general ISR is a method of generating intelligence based on surveillance and reconnaissance. The purpose of the surveillance aspect is to systematically observe someone or something (e.g., the ground, the public, the sea, underwater) in various ways (e.g., visual, auditory). The purpose of reconnaissance is to obtain information on enemies, potential activities and resources. The ISR agents should also obtain specifications on the weather and the geographical characteristics of specific areas.

Cyber ISR is an intelligence surveillance reconnaissance that includes general ISR and cyberspace. Cyberspace includes public institutions, governments, commercial organizations, the ambiguity of the military and non-military division, and anonymity. Hence, there is a need for an operating system that monitors and analyzes information generated in cyberspace.

Cyber ISR is divided into Intelligence, Surveillance and Reconnaissance. Surveillance and Reconnaissance mean the task of observing targets and collecting data in various spaces. Based on the data collected by the ISR, intelligence will be processed to help commanders make decisions.
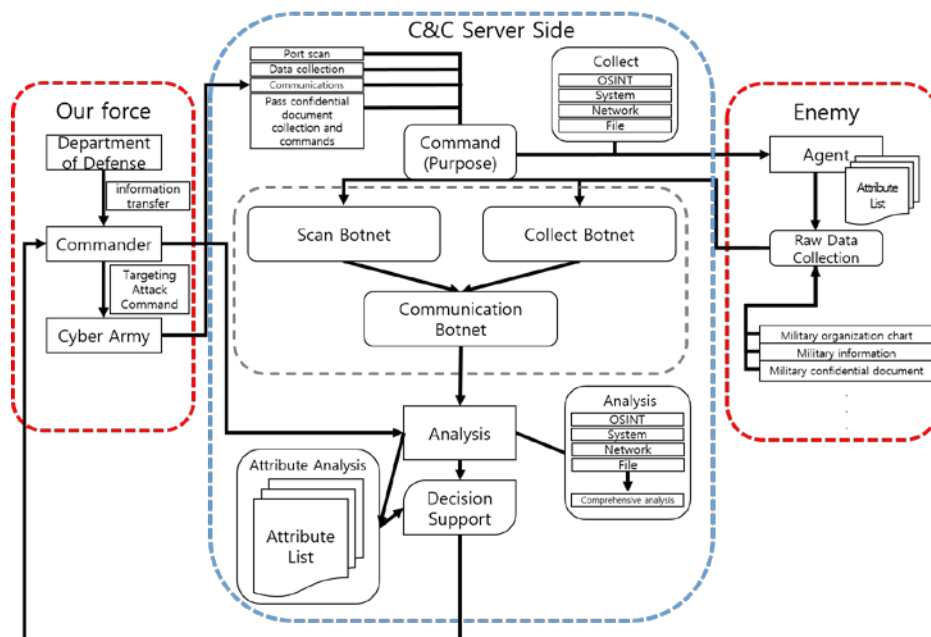

Fig. 1. Cyber surveillance and reconnaissance operation concepts

Previously collected information allows you to get a rough idea of when your computer users will be using your computer. Various types of information (e.g., the OS, service pack version, automatic lock, power saving mode of the computer, which is grasped first) can be used to roughly grasp the time that the current agent can operate the computer. Although it is possible to collect all the information, given sufficient time, if the information needs to be acquired within a short time period, a method of collecting only the information of high importance and transmitting it to the C & C server is needed [5]. Therefore, it is necessary to use a feature selection method to collect information over a limited time.

We experimented with various learning algorithms and realized that a method to enhance the learning speed is more important than decreasing the classification errors. Because a cyber ISR agent should reside in an enemy's network environment, with a limited amount of resources, the agent is required to learn quickly and return a piece of information. Moreover, the agent should maintain a classification rate that is high enough to choose the useful data. We found a very useful method for speeding up the learning process, while maintaining a similar classification rate.

Because a cyber ISR agent must be in a closed network, to prevent it from being seen by the enemy, a limited memory and fast learning process are necessary. We identified a very useful technique to increase the learning speed, while maintaining a similar rate of classification. The Filter Method is an evaluation function that uses calculated priorities, called feature evaluation values, from the training dataset. It returns a result in a very short time period, when compared to applying the Wrapper Method. However, this method has lower accuracy than the Wrapper Method.

The Wrapper Method generally has a longer execution time than the Filter Method. Unlike the Filter Method, the Wrapper Method always finds the best subset feature for the classifier. Therefore, the speed is slower than the Filter Method, but illustrates higher accuracy.

Fig. 2. Filter Method

This paper presents a hybrid feature selection method, derived from the filter and Wrapper Methods, and illustrates the model's design and initial experimental results. The experimental results show that the proposed feature selection method performs better than the Filter Method and general algorithm.

The rest of this paper is organized as follows. Section 2 describes the latest research trends in the feature selection [6,7,8,9,10,11,12]. Section 3 describes the design and methodology for the proposed model. Section 4 describes the general Filter Method and compares it with the proposed model. Section 5 concludes this paper.

Fig. 3. Wrapper Method

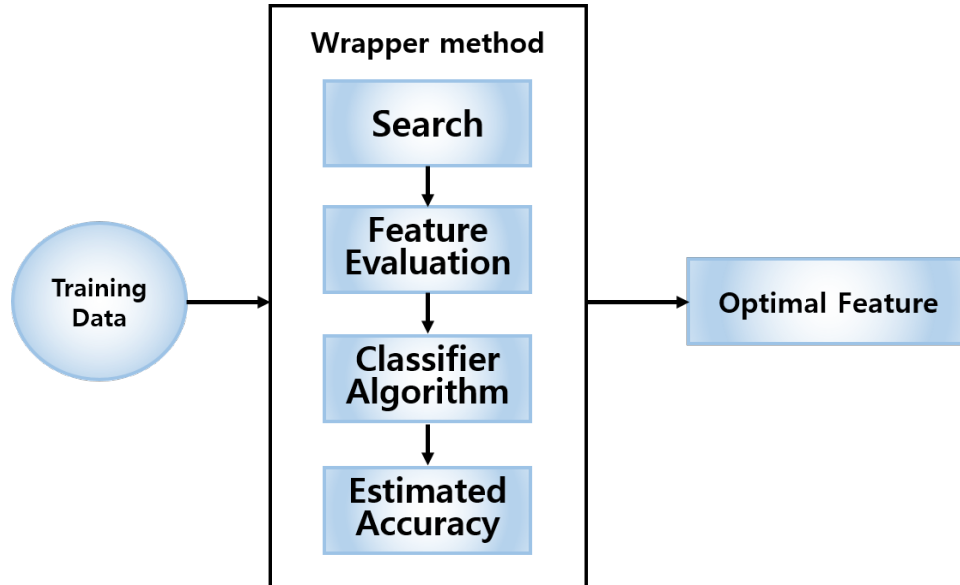## 2. Related Research

A larger amount of data makes it easier to create a high performance classifier. However, more features do not necessarily equate with higher performance. Normally, increasing the number of features increases the learning time and overfitting problems. In order to solve these problems, many papers have shown good performance by reducing the number of features using the feature selection methods. The algorithm comparisons, using the KDD 99 dataset [13], discussed the previously mentioned problems. In a previous study, the classification performance was improved, but required a large amount of time to learn.

To reduce the learning time, Song's paper compared 35 different methods using feature selection and various methods. They used the naïve Bayes algorithm as a learning model. Based on the experimental results, the correlation-based feature selection (CFS) method showed a 6.6% higher accuracy, on average, for the 35 datasets. Song suggested that the method could be used to remove irrelevant features and greatly reduce the dimensions. It had a fast learning time and low computation cost [6].

In Chandrashekar's study, they conducted a survey experiment using a feature selection method. Chandrashekar said that the feature selection could improve the computation performance, reduce the computation time in the pattern recognition applications, and improve the prediction performance. In Chandrashekar's paper, they introduced a case where the filter, wrapper, and embedded methods were applied. This paper found insight into the data, improved the classification models, and identified the unrelated variables [7].

In Hasan's study, they classified the KDD Cup 99 Dataset using the feature selection and random forest algorithms. The accuracy with the feature selection method was slightly increased, and the training time was reduced by 25%, when compared to the conventional random forest method. In addition, the false positive rate was minimized [8]. The Chandolikard paper also evaluated a model using the KDD CUP 99 Dataset. After

pre-processing using the feature selection, they evaluated the model with a rule-based algorithm [9]. Wen, Leordeanu, and Lefakis also proposed the use of this method to reduce the learning time for the feature selection [10,11,12].

3. Proposed Feature Selection Method

In this section, we propose a hybrid feature selection method that chooses an optimized feature set.

Most machine learning algorithms are designed to learn appropriate properties to make decisions. In the past, the properties that seemed to be important, in theory, were retained, while the attributes that seemed irrelevant and useless were excluded. Theoretically, in the past, scholars believed that a greater number of attributes would lead to better model performance. However, when actual experiments were conducted, a large number of attributes were found to confuse the process of generating the model, and thus, a model with good performance could not be generated. More specifically, if you assume that a class is randomly distributed among instances and is predicted to have the same value, in most cases, the opposite value in the remaining cases, then a new attribute is added to the heterogeneous class dataset.

The experimental results illustrate that these assumptions can degrade the classification accuracy [14]. This is because new attributes are chosen so that the tree can continue to divide. This has the effect of disassembling the set of available instances below the nodes. Another selection is then made, based on the more decomposed state data.

It is not good to have an unconditional number of attributes. Moreover, even if an attribute is related to the data, it is not good to have many things. A typical example of this is the overfitting problem. When an excessive number of attributes are added to a simple problem that can be sufficiently classified, the classifier is confused by this process. Consequently, the classifier is classified more complexly using many attributes. As a result, a simple problem becomes a complex problem, creating a model that overfits the data appropriately [15].

To solve these problems, it is necessary to choose an attribute that is based on a deep understanding of the learning problem. Automation techniques can be used to solve these problems efficiently. In addition, as the number of attributes decreases, the learning time is shortened, because the computation time is reduced during learning. The most important thing is that the diminution of the dimensions leads to more expressive and easily interpretable expressions. To solve this problem with an automation technique, many use feature selection and feature extraction [16,17,18].

In this paper, we use the feature selection method to solve the problem. We also propose a hybrid feature method that combines the advantages of each method by selecting the optimized features using the Wrapper Method and decreasing the time using the Filter Method to create a classifier with high performance in less time.
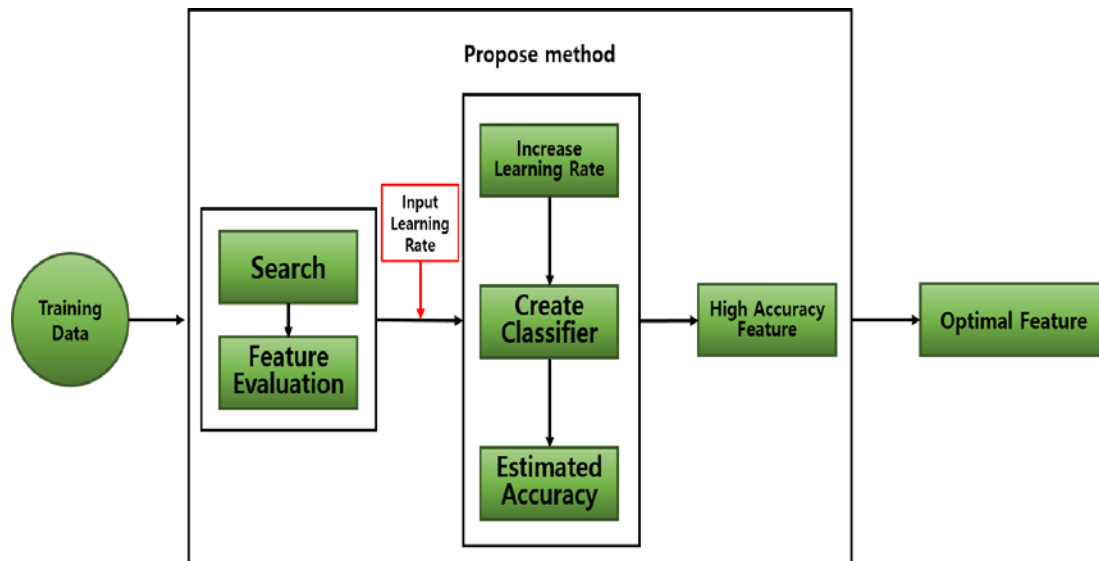
Fig. 4. Proposed method

The proposed method is illustrated in **Fig. 4.** The hybrid feature method selects the data property, search method and feature estimation method in the same manner as the general Filter Method. In this step, we use the information gain, which is a single attribute evaluator of the Filter Method, as the feature estimation method, along with a ranker for the ranking function. The information gain evaluates the attributes based on the entropy and information gain. This is obtained by subtracting the mean value of the entropy from the upper node entropy by considering the weight of the lower node in relation to the number of records in the lower node. It can be interpreted that the closer the information gain value is to 0, the better the data classification. Therefore, we evaluate the data attributes using a method to find a better data classification. The ranker ranks the individual attributes according to their evaluation results. The ranker then uses these attributes in the learning rate process after sorting them in descending order.

The next step is for the learning rate to be designated. Here, the learning rate refers to the threshold. When the initial evaluation starts, the threshold is set to 0 and increases by the value of the learning rate. The purpose of applying this method is to increase the learning speed. Based on the value derived from the information gain, all of the values that do not exceed a certain threshold are deleted. Only the remaining attributes are held together. This can reduce the unnecessary learning steps. At this time, it is possible to provide a fixed threshold value for learning, which is fast, but is likely to be low in accuracy. Therefore, by using the learning rate concept, the learning progresses by gradually decreasing the attribute, while gradually increasing the threshold based on the value of the attribute with the largest information gain value. It got a hint from the gradient descent.

The feature evaluation method used in the experiment is a single attribute evaluation method. The results obtained from Feature Evaluation illustrate the weight value of each attribute. Consequently, we have to think about where to cut the best based on this weight. This depends on the model and the data. Therefore, it requires a manual adjustment by the person. To automate this, the learning rate used in the gradient descent was applied. If you want to obtain results more quickly, set the learning rate higher. If you want to achieve high

performance, even if it takes a little longer, set the learning rate lower. These contents were found through the experiments.

By using the results of evaluating the attributes based on the information gain, and learning the attributes corresponding to the threshold in the attributes list created by using the ranker, a 10-fold cross validation using the machine learning model is applied by the user and the accuracy is stored. Because the existing Filter Method has low accuracy and the Wrapper Method has a slow execution speed, the proposed method uses the Filter Method to increase the learning speed. The Wrapper Method is used to obtain a high accuracy. It also evaluates the selected property by putting the verification step in the middle.

Thereafter, the discussed learning rate process is repeated until the threshold value becomes smaller and no features are left to be evaluated. When the learning rate process is finished, it returns the feature set with the highest accuracy during the process.

4. Experimental Results

In this section, we illustrate the results of the comparison of algorithms that do not apply any feature selection, algorithms that apply the CFS feature selection method, and algorithms that apply the hybrid feature selection method proposed in this paper. Naïve Bayes is used as the learning algorithm. The training data includes the KDD CUP 99 Dataset (KDD 99), Turkish Text Classification Dataset (TTC), Adult Dataset (Adult), Connect-4 Dataset (Connect-4), Dermatology Dataset, Contraceptive Method Choice Dataset, Chess_King-Rook vs. King-Pawn Dataset (Chess), Poker Hand Dataset (Poker Hand), Soybean (Large) Dataset (Soybean), and Statlog (Landsat Satellite) Dataset (Statlog) (**Table 1**). The machine learning algorithm uses Naive Bayes. The performance of the classifier was evaluated using a 10-fold cross validation. In addition, the performance of the classifier was further evaluated using the test dataset.

The learning rate used in the experiments was set at 0.05. CFS also used RankSearch as a search method, which sorts the attributes and ranks the most likely dependency sets using the attribute dependency evaluators.

The experiments were conducted using a multi-core CPU. The CPU was Ryzen 7 1700X, the operating system was Windows 10 Pro 1708, the memory was 64 GB, the Java version was 1.8.0_131, the editor was Eclipse, and the learning library was tested using MOA (Massive Online Analysis) version 2016.04.

Table 1. Datasets included in the experiments

Dataset	Attribute Num	Instance Num
KDD99	41	494,020
TTC	7507	3,600
Adult	14	48,842
Connect-4	42	67,557
Dermatology	34	366
CMC	9	1,473
Chess	36	3,196
Poker hand	11	1,025,010
Soybean	35	305
Statlog	36	6,435

The method used to compare the performance of the feature selection method is as follows. The TP rate refers to how much the Class A class is classified. This is the number of classes classified as real among the number of classes (True positive / Condition Positive). The FP rate indicates whether the non A class is misclassified as the A class. This is a False positive / Condition negative. The F-Measure is 2*(Precision*Recall)/(Precision+Recall). The Precision means how many of the classes are actually in the class. The Recall is equal to the TP rate. The ROC Area is closer to 1 for an Optimal Classifier. If the value is 0.5, it can be called Random Guessing. The Kappa statistic is called the Cohne's Kappa. It is a statistic that illustrates the agreement of the results when two evaluators are present. A completely different estimate is 0 and a perfect match is 1. Normally, it should be 0.8 or better, but the standard differs among scholars.

Table 2. Comparison of algorithm performances using a 10-fold cross validation

Dataset	Method	Number of Attributes Used	TP Rate	FP Rate	F-Measure	ROC Area	Kappa statistic
KDD99	Original	41	92.2%	1.0%	0.948	0.986	0.79
	CFS	8	95.3%	13.5%	0.952	0.994	0.85
	Hybrid	3	97.9%	2.1%	0.975	0.994	0.94
TTC	Original	7507	76.1%	4.8%	0.761	0.935	0.71
	CFS	54	73.5%	5.3%	0.742	0.938	0.68
	Hybrid	1345	78.1%	4.4%	0.781	0.949	0.74
Adult	Original	14	83.4%	38.2%	0.824	0.892	0.50
	CFS	5	79.9%	53.3%	0.770	0.870	0.33
	Hybrid	7	83.7%	40.7%	0.823	0.892	0.49
Connect-4	Original	42	72.1%	42.6%	0.681	0.807	0.33
	CFS	6	70.0%	50.2%	0.637	0.739	0.24
	Hybrid	35	72.3%	43.1%	0.680	0.807	0.33
Dermatology	Original	34	97.5%	0.4%	0.975	0.997	0.97
	CFS	3	54.4%	14.1%	0.442	0.765	0.41
	Hybrid	21	97.8%	0.3%	0.978	0.998	0.97
CMC	Original	9	49.3%	24.6%	0.497	0.682	0.24
	CFS	3	55.5%	24.0%	0.556	0.707	0.31

	Hybrid	3	55.5%	24.0%	0.556	0.707	0.31
Chess	Original	36	87.9%	12.3%	0.879	0.952	0.76
	CFS	7	92.0%	8.3%	0.920	0.971	0.84
	Hybrid	3	90.4%	10.2%	0.904	0.933	0.81
Poker hand	Original	11	49.1%	49.7%	0.374	0.501	0.00
	CFS	7	49.3%	49.7%	0.368	0.499	0.00
	Hybrid	3	49.9%	50.0%	0.335	0.499	0.00
Soybean	Original	35	90.6%	1.1%	0.904	0.989	0.89
	CFS	23	90.9%	1.2%	0.907	0.989	0.89
	Hybrid	34	90.9%	1.1%	0.907	0.989	0.90
Statlog	Original	36	79.5%	3.7%	0.803	0.962	0.75
	CFS	23	79.0%	3.8%	0.799	0.965	0.74
	Hybrid	31	79.6%	3.7%	0.804	0.964	0.75

The experimental results showed that the hybrid feature selection method proposed in this paper had the best performance for all of the datasets, except the chess dataset. In the chess dataset, the TP rate was higher than that of the general algorithm, and the performance was higher than that of the general algorithm.

The CFS method selected the least properties, in comparison with the other algorithms. However, the results of the CFS method were worse than those of the general algorithm in five datasets. The Dermatology dataset performed poorly. The Kappa statistic of the proposed hybrid feature selection method showed a high overall value. In particular, it illustrated a value of 0.94 for the KDD CUP problem, which was higher than the general algorithm value by 0.15. A Kappa statistic of 0.8 or more is a reliable result. Therefore, in the KDD 99 problem, the hybrid feature selection method had a very reliable result.

In the KDD 99 dataset, the FP rate (false positive rate) showed a very low result of 2.1%, which meant the corresponding class of the non-applicable class was misclassified. A lower value was better, because the cost was reduced. On the other hand, when compared with the original algorithm and the CFS algorithm, the TP rate increased to 3.1% and the FP rate increased to 12.5%.

The performance of Chandolikar's method was evaluated using the KDD 99 dataset. In Chandolikar's experiment, we used the feature selection method in front of the existing rule-based algorithm. Compared to Chandolikar's method, the proposed method offered high performance. In addition, a performance evaluation was performed using the classifier and the test dataset generated through the algorithms listed in **Table 3** using KDD 99, Adult, Poker Hand, Soybean, and Statlog, which were the test datasets among the datasets.

Table 3. Comparison of the algorithm performances using test data

Dataset	Method	Number of Attributes Used	TP Rate	FP Rate	F-Measure	ROC Area	Kappa statistic	Mean absolute error	Root mean squared error
KDD 99	Original	41	86.4%	6.2%	0.868	0.944	0.68	0.05	0.23
	CFS	8	84.4%	28.5%	0.824	0.966	0.57	0.06	0.24
	Hybrid	3	90.5%	5.3%	0.885	0.967	0.77	0.04	0.19
Adult	Original	14	83.1%	39.1%	0.822	0.891	0.48	0.18	0.37
	CFS	5	79.9%	54.0%	0.770	0.870	0.32	0.20	0.41
	Hybrid	7	83.8%	41.0%	0.825	0.892	0.49	0.17	0.36
Poker hand	Original	11	49.1%	49.8%	0.374	0.503	0.00	0.11	0.24
	CFS	7	49.5%	49.9%	0.366	0.503	0.00	0.11	0.24
	Hybrid	1	50.1%	50.1%	0.335	0.501	0.00	0.11	0.24
Soy bean	Original	35	88.0%	1.3%	0.863	0.992	0.87	0.01	0.10
	CFS	23	83.1%	2.0%	0.824	0.961	0.81	0.02	0.11
	Hybrid	34	88.0%	1.3%	0.863	0.992	0.87	0.01	0.10
Statlog	Original	36	79.6%	3.8%	0.803	0.959	0.75	0.07	0.26
	CFS	23	80.4%	3.6%	0.810	0.963	0.76	0.07	0.25
	Hybrid	31	79.7%	3.8%	0.803	0.961	0.75	0.07	0.25

The experimental results illustrated that the proposed method generally had the best performance. However, when using the CFS method, the worst performance was generally seen.
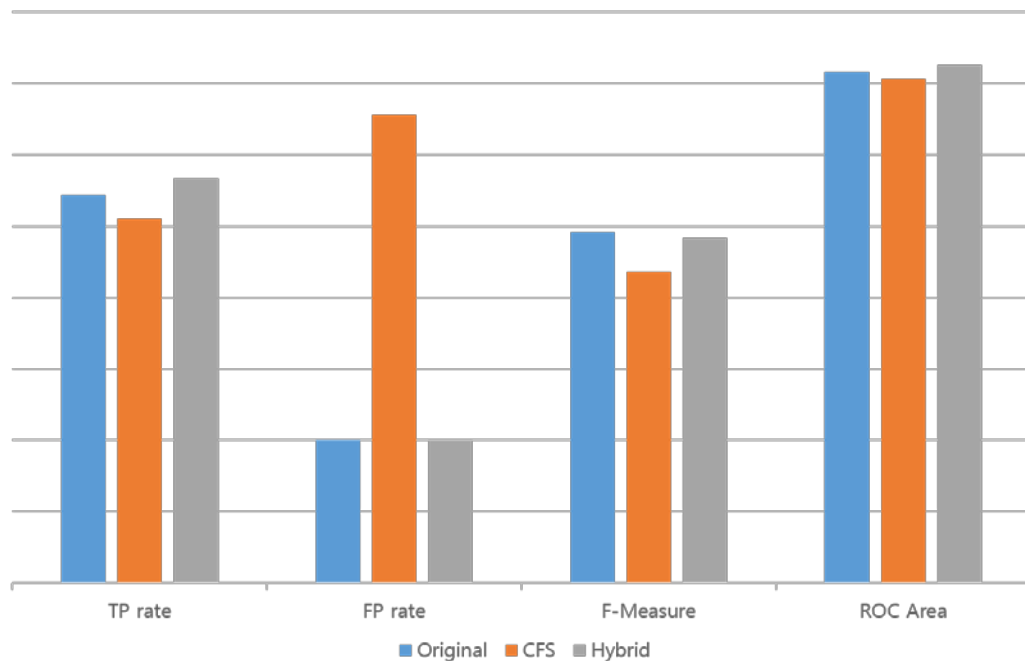

Fig. 5. Average performances for five datasets

Fig. 5 is a graph showing the average performance of the generated model, when applying the Naïve Bayes algorithm to five datasets. The definition of the y-axis means the average value of each performance index. The proposed method illustrated the highest performance, with an average TP rate of 78.4%. The CFS method had the lowest performance at 75.5%. The FP rate illustrated the best performance of 20.0%, which was the same as those of the original method (using all attributes) and the proposed method. With the exception of the F-Measure, the proposed method showed the best performance and the F-Measure illustrated a moderate performance. This reduced the number of attributes and reduced the time required for model learning. As a result, the generated model had the best performance, when compared to the other methods. As a result, the learning time and costs were reduced, while the overfitting problem was solved, to some extent.

Table 4. Comparison of times required for the attribute evaluations using the Wrapper Method and the proposed method

Dataset	Method	NB Elapsed time(s)	DT Elapsed time(s)
KDD	Wrapper	14774	time out
	Hybrid	106.271	1327.521
TTC	Wrapper	time out	time out
	Hybrid	567.716	850.831
Adult	Wrapper	77	1131
	Hybrid	3	44.184
Connect-4	Wrapper	1555	10799
	Hybrid	7.033	32.977
Dermatology	Wrapper	12	35
	Hybrid	0.484	0.506
CMC	Wrapper	1	16
	Hybrid	0.391	0.965
Chess	Wrapper	8	45
	Hybrid	0.4	0.522
Poker hand	Wrapper	13	62
	Hybrid	3.146	3.450
Soybean	Wrapper	7	23
	Hybrid	0.365	0.461
Statlog	Wrapper	81	869
	Hybrid	5.911	12.418

Table 4 compares the times required for the attribute evaluations using the Naïve Bayes algorithm, the Decision Tree algorithm for the Wrapper Method and the proposed method. The evaluation of the KDD 99 data was not completed, even after 24 h, when it was evaluated by applying the Wrapper Method and the Decision Tree algorithm. In addition, the TTC data were evaluated endlessly when both the Naïve Bayes and Decision Tree algorithms used the Wrapper Method.

As shown in **Table 4**, the proposed method significantly shortens the evaluation time, when compared to the Wrapper Method. We determined the performance difference between the Wrapper Method and the proposed method by evaluating the model performance with the test dataset using these evaluated properties. We experimented with the same five datasets, as in the previous experiment.

Table 5. Comparison of the Wrapper Method and the performance of the model generated by the proposed method (using a Decision Tree)

Dataset	Method	Algorithm	Number of attributes used	TP Rate	FP Rate	F-Measure	ROC Area	Kappa statistic
KDD 99	Original		41	73.8%	2.0%	0.787	0.882	0.53
	Wrapper method		time out					
	Hybrid		25	73.6%	2.3%	0.787	0.860	0.52
Adult	Original		14	85.8%	30.0%	0.855	0.890	0.59
	Wrapper method		14	85.8%	30.0%	0.855	0.890	0.59
	Hybrid		14	85.8%	30.0%	0.855	0.890	0.59
Poker hand	Original	Decision Tree	11	57.7%	38.0%	0.542	0.627	0.20
	Wrapper method		8	58.0%	38.5%	0.539	0.630	0.20
	Hybrid		8	57.0%	38.9%	0.532	0.622	0.18
Soy bean	Original		35	88.0%	1.3%	0.863	0.992	0.87
	Wrapper method		13	86.7%	1.8%	0.846	0.970	0.85
	Hybrid		10	82.2%	1.9%	0.794	0.965	0.80
Statlog	Original		36	85.4%	3.1%	0.853	0.927	0.82
	Wrapper method		10	85.8%	3.0%	0.858	0.939	0.83
	Hybrid		35	85.4%	3.1%	0.853	0.927	0.82

We used the test data to evaluate the generated model. The algorithm used a Decision Tree. The machine learning algorithm used in the attribute selection evaluation was the Decision Tree. The Decision Tree was applied to both the Wrapper Method and the proposed method. A 10-fold cross validation was also used.

As illustrated in **Table 5**, all three methods had similar results. The Wrapper Method and the proposed method showed that the number of attributes was reduced, because the attributes that were deemed unnecessary in the model learning were removed through the evaluator in the existing data. This reduced the amount of the computation when learning the model. When much of the data were accumulated, the learning time was significantly shortened, when compared to the general model.

In the experiment, when the attribute selection evaluation was performed using the Decision Tree in the proposed method, the learning speed increased, but the classification performance of the model did not change much. During the experiment, we found a unique point. We used the Naïve Bayes to evaluate the properties of the proposed method. The classification performance of the model was significantly increased, when the Decision Tree algorithm was applied.

Table 6. Performance comparison of the proposed method using the NB and the DT

Dataset	Method	Algorithm	Number of attributes used	TP Rate	FP Rate	F-Measure	ROC Area	Kappa statistic
KDD 99	Hybrid (NB)	Decision Tree	3	92.1%	2.3%	0.902	0.958	0.81
	Hybrid (DT)		25	73.6%	2.3%	0.787	0.860	0.52
Adult	Hybrid (NB)		8	85.4%	31.5%	0.849	0.890	0.57
	Hybrid (DT)		14	85.8%	30.0%	0.855	0.890	0.59
Poker hand	Hybrid (NB)		2	53.7%	45.2%	0.442	0.552	0.09
	Hybrid (DT)		8	57.0%	38.9%	0.532	0.622	0.18
Soy bean	Hybrid (NB)		35	88.0%	1.3%	0.863	0.992	0.87
	Hybrid (DT)		10	82.2%	1.9%	0.794	0.965	0.80
Statlog	Hybrid (NB)		31	86.4%	2.9%	0.862	0.928	0.83
	Hybrid (DT)		35	85.4%	3.1%	0.853	0.927	0.82

Table 6 compares the results of applying the Naïve Bayes and the Decision Tree with the proposed method and then applying them to the Decision Tree algorithm. As a result, the total number of attributes was 79 in the Naïve Bayes algorithm and 92 in the Decision Tree algorithm. In the case of the TP rate, the average for the Naïve Bayes algorithm was 81.1% and the average for the Decision Tree was 76.8%. In the case of the F-measure, the values were 0.784 for the Naïve Bayes algorithm and 0.764 for the Decision Tree algorithm.

In the KDD 99 dataset, the Naïve Bayes algorithm illustrated an average performance that was 4.3% higher than that of the TP rate using the Decision Tree with only 3 attributes. In addition, it showed a high performance difference of 18.3%, when compared with the original data TP rate result. This illustrated that even though only three of the 41 attributes were used, the learning speed was much faster than the learning speed, and the model performance was much higher.

Therefore, it can be seen that the attributes selected by applying the Naïve Bayes to the proposed model showed good performance when generating the model using the corresponding method, in addition to the Bayesian algorithm. In the case where an attribute evaluation is required in a short period of time to illustrate a significantly faster speed, as shown in **Table 6**, evaluating the attributes using the Naïve Bayes algorithm in the proposed model is recommended.

5. Conclusion

This paper introduced a hybrid feature selection method (HFSM) that chooses an optimized feature set. HFSM is based on a Filter Method with a fast speed and a Wrapper Method with a higher accuracy. The existing Wrapper Method illustrates a higher performance, but it is not suitable for cyber ISR, because it requires a large amount of computation time. Although the computation time is longer than that for the Filter Method, the HFSM is a more suitable algorithm for achieving a performance similar to the Wrapper Method. As shown in this paper, the computation time for the HFSM is slightly longer than that of the conventional CFS algorithm, but its accuracy is much higher than that of the CFS algorithm.

In addition, we illustrate that the proposed method is superior to the proposed method by applying the Naïve Bayes algorithm and applying the selected attributes to the tree algorithm. Consequently, the Naïve Bayes is very suitable for importing important information with a high weight in a situation where it is necessary to collect information from an enemy's network in a short time.

In this paper, the evaluation time was greatly shortened, while the performance was not lower than that of the Wrapper Method. This paper considered an attribute evaluation method that can enhance the performance of the model through various attribute combinations using the dependency property evaluation method and a search method, instead of the single attribute evaluation method used in the proposed method.

References

[1] Gyeong-Il Shin, Hosang Yooun, DongIl Shin, DongKyoo Shin, "Incremental learning method for cyber intelligence, surveillance, and reconnaissance in closed military network using converged IT techniques," *Soft Computing*, Vol 22, No. 20, pp.6835-6844, August, 2018. Article (CrossRef Link)

[2] Hurley, Matthew M. "For and from cyberspace: Conceptualizing cyber intelligence, surveillance, and reconnaissance," *Air & Space Power Journal*, Vol. 26, No.6, pp.12-33, December, 2012. Article (CrossRef Link)

[3] JH. Eom, NU. Kim and SH. Kim, "Cyber military strategy for cyberspace superiority in cyber warfare," *in Proc. of IEEE Conf. Cyber Security, Cyber Warfare and Digital Forensic (CyberSec) International Conference*, pp.295-299, July, 2012. Article (CrossRef Link)

[4] Sung Soo Choi, Tae Myeong Jeong and Jung ho Eom, "An Introduction of Cyber Warfare Attack and Security Techniques," *1th Edition, HongRung Publishing Company*, 2012.

[5] Karri Wihersaari, "Intelligence acquisition methods in cyber domain: examining the circumstantial applicability of cyber intelligence acquisition methods using a hierarchical mode," *National Defense University Course Library*, pp.1-63, April, 2015. Article (CrossRef Link)

[6] Song Qinbao, Jingjie Ni, and Guangtao Wang, "A fast clustering-based feature subset selection algorithm for high-dimensional data," *IEEE transactions on knowledge and data engineering*, Vol 25, No. 1, pp.1-14, August, 2011. Article (CrossRef Link)

[7] Chandrashekar, Girish, and Ferat Sahin, "A survey on feature selection methods," *Computers & Electrical Engineering*, Vol 40, No. 1, pp.16-28, January, 2014. Article (CrossRef Link)

[8] Hasan, Md Al Mehedi, et al., "Feature selection for intrusion detection using random forest," *Journal of Information Security*, Vol 7, No. 3, pp.129-140, July, 2016. Article (CrossRef Link)

[9] Chandolikar, N. S., and V. D. Nandavadekar, "Selection of relevant feature for intrusion attack classification by analyzing KDD Cup 99," *MIT International Journal of Computer Science & Information Technology*, Vol 2, No. 2, pp.85-90, August, 2012. Article (CrossRef Link)

[10] Lefakis Leonidas, and François Fleuret, "Jointly informative feature selection made tractable by Gaussian modeling," *Journal of Machine Learning Research*, Vol 17, No. 182, pp.1-39, 2016. Article (CrossRef Link)

[11] Leordeanu Marius, Alexandra Radu, and Rahul Sukthankar, "Features in concert: Discriminative feature selection meets unsupervised clustering," arXiv preprint arXiv:1411.7714, November, 2014. Article (CrossRef Link)

[12] Wen Xuezhi, et al., "Efficient feature selection and classification for vehicle detection," *IEEE Transactions on Circuits and Systems for Video Technology*, Vol 25, No. 3, pp.508-517, September, 2015. Article (CrossRef Link)

[13] Hyunjung Ji, Donghwa Kim, Dongil Shin, Dongkyoo Shin, "A Study on comparison of KDD CUP 99 and NSL-KDD using artificial neural network," *Lecture Notes in Electrical Engineering*, Vol. 474, pp.452-457, December, 2017. Article (CrossRef Link)

[14] Ian H. Witten, Eibe Frank, Mark A. Hall and Christopher J. Pal, *Data Mining*, 4th Edition, Morgan Kaufmann, 2017.

[15] Ron Kohavi and George H. John, "Wrappers for feature subset selection," *Artificial Intelligence*, Vol. 97, pp.273-324, December, 1997. Article (CrossRef Link)

[16] Anil Jain and Douglas Zongker, "Feature selection: evaluation, application, and small sample performance," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 19, No. 2, pp. 153-158, February 1997. Article (CrossRef Link)

[17] Huan Liu and Lei Yu, "Toward integrating feature selection algorithms for classification and clustering," *IEEE Transactions on Knowledge and Data Engineering*, Vol. 17, No. 4, pp. 491-502, March, 2005. Article (CrossRef Link)

[18] Guyon, Isabelle. and Elisseeff, "An introduction to variable and feature selection," *Journal of Machine Learning Research*, Vol. 3, pp. 1157-1182, March, 2003. Article (CrossRef Link)

GyeongIl Shin received B.S. Degree in Department of Computer Engineering from the Chosun University. He is currently pursuing the M.S degree in Department of Computer Engineering with Sejong University. His research Machine Learning and Deep Learning.

Hosang Yooun is a Senior Research Engineer at Agency for Defense Development. He received the Ph.D. degree from Kaist. His research Cyberwarfare Intrusion detection technology and Cyberwarfare active response technology.

DongIl Shin is a professor at Department of Computer Engineering in Sejong University. He received the Ph.D. degree from University of North Texas. His research Machin Learning(Deep Learning) and HCI.

DongKyoo Shin is a professor at Department of Computer Engineering in Sejong University. He received the Ph.D. degree from Texas A&M University. His research Context Awareness Middleware, Information Security and Data Mining.