A Novel Hybrid Taguchi-Grey-Based Method for Feature Subset Selection

Hsin-Yun Chang[1] and Chung-Shan Sun[2]

[1] Department of Business Administration,
Chin-Min Institute of Technology
110 Hsueh-Fu Road, Tou-Fen, Miao-Li 305, Taiwan
Department of Industrial Technology Education,
National Kaohsiung Normal University
116 Heping 1st RD., Lingya District, Kaohsiung 802, Taiwan
ran_hsin@ms.chinmin.edu.tw
[2] Department of Industrial Technology Education,
National Kaohsiung Normal University
116 Heping 1st RD., Lingya District, Kaohsiung 802, Taiwan
x9106@ms21.hinet.net

Abstract. In this paper, a novel hybrid Taguchi-Grey-based method for feature subset selection is proposed. The two-level orthogonal array is employed in the proposed method to provide a well-organized and balanced comparison of two levels of each feature (i.e., the feature is selected for pattern classification or not) and interactions among all features in a specific classification problem. That is, this two-dimensional matrix is mainly used to reduce the feature subset evaluation efforts prior to the classification procedure. Accordingly, the grey-based nearest neighbor rule and the signal-to-noise ratio (SNR) are used to evaluate and optimize the features of the specific classification problem. In this manner, important and relevant features can be identified for pattern classification. Experiments performed on different application domains are reported to demonstrate the performance of the proposed hybrid Taguchi-Grey-based method. It can be easily seen that the proposed method yields superior performance and is helpful for improving the classification accuracy in pattern classification.

Keywords: Feature Subset Selection, Taguchi Methods, Grey-based Nearest Neighbor Rule, Pattern Classification.

1 Introduction

In recent years, different pattern classification approaches have been investigated for classifying new, unseen instances. In a pattern classification model [12], a set of training instances or examples, denoted as training set *TS*, is collected. Each instance or example is described by *p* features and a class label. Generally, all features of each instance will be taken into consideration during the classification process. Many real-world classification problems, however, involve redundant or irrelevant features that

L. Rueda, D. Mery, and J. Kittler (Eds.): CIARP 2007, LNCS 4756, pp. 457–465, 2007.

usually greatly affect the overall classification accuracy. To improve the performance, various feature selection or feature subset selection methods have been developed. These methods focus on selecting important and relevant features from the original feature set, as well as reducing the dimensionality in a particular classification problem.

Feature subset selection can be viewed as a search problem [15], where each search state in the search space specifies a possible feature subset. If each instance in a specific classification problem contains p attributes, the search space will be composed of 2^p candidate feature subsets. Obviously, exhaustive search through the entire search space (i.e. 2^p candidate feature subsets) has a very high computational cost and thus is usually unfeasible in practice, even for medium-sized p [19]. Consequently, it is difficult to select a best feature subset for pattern classification from the entire search space with respect to the tradeoff between high classification accuracy and small number of selected features.

Two well-known greedy hill-climbing approaches, sequential forward selection (SFS) [17] and sequential backward selection (SBS) [17], are commonly used for feature subset selection. As mentioned earlier, feature subset selection can be considered as a search problem, where each search state in the search space specifies a possible feature subset. In SFS, the search procedure starts with an empty feature set and then successively adds features one at a time to find the final feature subset. By contrast, in SBS, the search procedure starts with a full feature set and then successively removes features one at a time to find the final feature subset. In bidirectional feature subset selection methods [19], the search procedure may start with an empty feature set or a full feature set and then add or remove features to or from the search starting point [19] simultaneously. Accordingly, the final feature subset can be obtained. The above sequential search methods for feature subset selection are simple and easy to implement. However, local optimal final feature subsets are often obtained during the search procedure. Another similar sequential search method for feature subset selection is proposed in [11]. First, k features are added (or eliminated) to the candidate feature subset at a time. Accordingly, l features are eliminated (or added) from the candidate feature subset at a time ($k > l$). These two steps are repeated until a final feature subset is obtained. In this case, the values of k and l, which will significantly affect the final result, should be determined carefully. To avoid being trapped into local optimal results, random search [4] through the entire search space is also commonly used to find the final feature subset. This method can help the search procedure to escape from local maximums [19] (i.e., non-deterministic heuristic search). However, inconsistent final feature subsets may be derived from different runs [19].

During the search procedure, each feature or generated feature subset should be evaluated by an evaluation criterion. Generally, two kinds of evaluation criteria, independent criterion and dependent criterion [19], are adopted to evaluate each feature or generated feature subset in feature subset selection. An independent criterion [13, 16, 18] is used to evaluate the goodness of each feature or generated feature subset by considering the original characteristics of the training set. In this case, pattern classification methods are not involved in each evaluation process. As for dependent criterion, pattern classification methods are directly used to evaluate the goodness (i.e., classification ability or accuracy) of each feature or generated feature subset. By contrast, the corresponding feature subset selection methods are named as the wrapper approaches [16]. Generally, the wrapper models, which focus mainly on improving the classification accuracy of pattern classification tasks, often yield superior performance (i.e., high classification accuracy) than the filter models. However, the wrapper approaches are more computational expensive than the filter approaches [16, 19]. As a result, many pattern classification methods that have very high computational costs, such as neural networks [2] and decision trees [20], may not be suitable to be used as evaluation criteria for evaluating each feature or generated feature subset.

In this paper, a novel hybrid Taguchi-Grey-based method for feature subset selection is proposed. The two-level orthogonal array is employed in the proposed method to provide a well-organized and balanced comparison of two levels of each feature (i.e., the feature is selected for pattern classification or not) and interactions among all features in a specific classification problem. That is, this two-dimensional matrix is mainly used to reduce the feature subset evaluation efforts prior to the classification procedure. Accordingly, the grey-based nearest neighbor rule and the signal-to-noise ratio (SNR) are used to evaluate and optimize the features of the specific classification problem. In this manner, important and relevant features can be identified for pattern classification. As a result, the hybrid Taguchi-Grey-based method proposed here has wrapper nature [16] (In wrapper feature subset selection methods, each candidate feature or feature subset is evaluated according to the classification ability obtained by the pattern classification model). That is, features will be selected based on the special properties of the corresponding pattern classification model and thus the goal of feature subset selection method here is to maximize the classification accuracy. Experiments performed on different application domains are reported to demonstrate the performance of the proposed hybrid Taguchi-Grey-based method. It can be easily seen that the proposed method yields superior performance and is helpful for improving the classification accuracy in pattern classification.

The rest of this paper is organized as follows. The concepts of the Taguchi methods used in the proposed method are reviewed in Sections 2. Section 3 proposes a novel hybrid Taguchi-Grey-based method for feature subset selection. In Section 4, an example is given to illustrate the proposed method. In Section 5, experiments performed on different classification problems are reported and discussed. Finally, the conclusions are given in Section 6.

2 Taguchi Methods

In robust design [23], products, processes or equipments can be evaluated and improved by considering different related design parameters (or factors). As a well-known robust experimental design approach, the Taguchi method [22] uses two principal tools, the orthogonal array and the signal-to-noise ratio (SNR), for the above purpose of evaluation and improvement. Consider that a specific object domain (e.g. product, process or equipment) contains q design parameters (or factors). Orthogonal arrays are primarily used to reduce the experimental efforts regarding these q different design factors. An orthogonal array can be viewed as a fractional factorial matrix that provides a systematic and balanced comparison of different levels of each design factor and interactions among all design factors. In this two-dimensional matrix, each column specifies a particular design factor and each row represents a trial with a specific combination of different levels regarding all design factors. In the proposed method, the well-known two-level orthogonal array is adopted for feature subset selection. A general two-level orthogonal array can be defined as follows.

$$L_w(2^{w-1}), \tag{1}$$

where $w=2^k$ ($k \geq 1$) represents the number of experimental trials, base 2 specifies the number of levels of each design factor, and w-1 is the number of columns (i.e., the number of design factors) in the orthogonal array.

For example, an orthogonal array $L_{16}(2^{15})$ can be created for a specific object domain that contains 15 design factors with two levels (i.e., level 1 and level 2). Notably, by using the two-level orthogonal array, only 16 experimental trials are needed for the purpose of evaluation and improvement. By contrast, all possible combinations of 15 design factors (i.e., 2^{15}) should be taken into consideration in the full factorial experimental design, which is obviously often inapplicable in practice.

Once the orthogonal array is generated, the observation or the objective function of each experimental trial can be determined. Accordingly, the signal-to-noise ratio (SNR) is used to evaluate and optimize the design parameters (or factors) of the specific object domain. In general, two kinds of signal-to-noise ratios (SNRs), the smaller-the-better and the larger-the-better characteristics [23], are commonly considered for the evaluation task.

Consider that a set of k observations $\{y_1, y_2, \ldots, y_k\}$ is given. In the smaller-the-better characteristic, the signal-to-noise ratio (SNR) is calculated as follows.

$$SNR = -10\log\left(\frac{1}{k}\sum_{i=1}^{k} {y_i}^2\right), \tag{2}$$

Similarly, in the larger-the-better characteristic, the signal-to-noise ratio (SNR) is calculated as follows.

$$SNR = -10\log\left(\frac{1}{k}\sum_{i=1}^{k}\frac{1}{{y_i}^2}\right), \tag{3}$$

The signal-to-noise ratio (SNR) is used to measure the robustness of each design parameter (or factor). That is, “high quality” of a particular object domain can be achieved by considering each design parameter with a specific level having high signal-to-noise ratio (SNR).

In summary, the Taguchi method offers many advantages for robust experimental design. First, the number of experimental runs can be substantially reduced (compared with the full factorial experimental design). Meanwhile, the significance of each design parameter regarding a particular object domain can be analyzed precisely. In the proposed method, the above two useful tools, the orthogonal array and the signal-to-noise ratio (SNR), are employed for feature subset selection.

3 Hybrid Taguchi-Grey-Based Method for Feature Subset Selection

In this section, a novel hybrid Taguchi-Grey-based method for feature subset selection is proposed. Consider that a particular classification task involves a set of m labeled training instances, denoted as $V=\{v_1, v_2, \ldots, v_m\}$. Each instance is described by n attributes, denoted as $F = (f_1, f_2, \ldots, f_n)$. The detailed procedures of the proposed hybrid Taguchi-Grey-based method for feature subset selection are described as follows.

Step1. Generate the two-level orthogonal array L with respect to the n attributes, features or factors in a specific classification problem. In each experimental trial j in the two-level orthogonal array L, levels 1 or 2 in each column i mean feature i is selected in the corresponding feature set S_j for pattern classification or not, respectively.

Step2. For each feature set S_j, determine an average classification accuracy regarding the training set V (denoted by $ACC(V, S_j)$) by using the grey-based nearest neighbor rule [8, 9 ,10, 14] with leave-one-out (LOO) cross-validation method [5]. Here, $ACC(V, S_j)$ is considered as the observation or the objective function of the experimental trial j in the two-level orthogonal array L.

Step3. Calculate the corresponding signal-to-noise ratio (SNR) for each level (i.e., levels 1 or 2) of each feature or factor i according to the various observations in the two-level orthogonal array L.

Step4. Select the features whose SNR for level 1 is greater than that for level 2. These features, denoted as feature subset S, are used as the final feature subset for pattern classification.

The two-level orthogonal array is employed in the proposed method to provide a well-organized and balanced comparison of two levels of each feature (i.e., the feature is selected for pattern classification or not) and interactions among all features in a specific classification problem. In other words, this two-dimensional matrix is mainly used to reduce the feature subset evaluation efforts prior to the classification procedure. Accordingly, the grey-based nearest neighbor rule and the signal-to-noise ratio (SNR) are used to evaluate and optimize the features of the specific classification problem.

Based on the grey-based nearest neighbor rule with leave-one-out (LOO) cross-validation method [5], a classification accuracy with respect to the training set V and a particular feature set S_j (denoted as $ACC(V, S_j)$), can be obtained. Leave-one-out cross-validation implies that each instance in V is considered as the test instance once and other instances in V are considered as the corresponding training instances. In this manner, the grey-based nearest neighbor rule will be carried out m times according to m instances and n features in V. Afterwards, the average classification accuracy is calculated for evaluating the classification performance of the corresponding feature set S_j. The signal-to-noise ratio (SNR) is then used to measure the robustness of each feature of the specific classification problem. That is, high classification performance regarding the classification task can be achieved by considering each feature with a specific level having high signal-to-noise ratio (SNR). Here, the larger-the-better characteristic, as shown in Eq. (3), is selected for calculating the signal-to-noise ratio (SNR) since maximum classification accuracy is preferred in pattern classification. In the proposed method, feature i with SNR of level 1 greater than that of level 2 means that the feature is suggested to be selected in the final feature subset for pattern classification. By contrast, feature i is suggested to be removed from the original feature set F if the corresponding SNR of level 2 greater than that of level 1. (Notably, levels 1 or 2 of feature i mean the feature is selected in the corresponding feature set S_j for pattern classification or not, respectively.)

4 Illustrative Example

This section gives an example to illustrate the proposed hybrid Taguchi-Grey-based method for feature subset selection. In the Autompg classification problem [3], each instance has seven attributes, denoted by {A, B, C, D, E, F, G}. By using the proposed method, a two-level orthogonal array $L_8(2^7)$ can be generated as Table 1.

Table 1. $L_8(2^7)$ Orthogonal Array

Number of Experimental Trial	Design Factors (Features)						
	A	B	C	D	E	F	G
	Column Number						
	1	2	3	4	5	6	7
1	1	1	1	1	1	1	1
2	1	1	1	2	2	2	2
3	1	2	2	1	1	2	2
4	1	2	2	2	2	1	1
5	2	1	2	1	2	1	2
6	2	1	2	2	1	2	1
7	2	2	1	1	2	2	1
8	2	2	1	2	1	1	2

Restated, in each experimental trial j in the two-level orthogonal array $L_8(2^7)$, levels 1 or 2 of each column i mean feature i is selected in the corresponding feature set S_j for pattern classification or not, respectively. For example, in experimental trial 7, features C, D, G are selected as the final feature subset for pattern classification. By using the two-level orthogonal array $L_8(2^7)$, the experimental efforts regarding feature subset evaluation can be reduced from 128 (i.e., 2^7) trials to eight trials.

Accordingly, for each experimental trial j, the average classification accuracy regarding the training set V and the corresponding feature set S_j (denoted by $ACC(V, S_j)$) can be determined by using the grey-based nearest neighbor rule with leave-one-out (LOO) cross-validation method. Here, $ACC(V, S_j)$ is considered as the observation or the objective function of the experimental trial j in the two-level orthogonal array $L_n(2^{n-1})$. As a result, the experimental layout and signal-to-noise data of the Autompg classification problem can be summarized as Table 2. Here, the larger-the-better characteristic, as shown in Eq. (3), is selected to determine the signal-to-noise ratio (SNR) since maximum classification accuracy is preferred in pattern classification.

Table 2. Experimental layout and signal-to-noise data of the Autompg classification problem

Number of Experimental Trial	Column / Feature							Classification Accuracy (%)	Classification Accuracy SNR (dB)
	A	B	C	D	E	F	G		
1	1	1	1	1	1	1	1	69.10	36.79
2	1	1	1	2	2	2	2	66.58	36.47
3	1	2	2	1	1	2	2	72.86	37.25
4	1	2	2	2	2	1	1	63.07	36.00
5	2	1	2	1	2	1	2	81.16	38.19
6	2	1	2	2	1	2	1	73.62	37.34
7	2	2	1	1	2	2	1	71.11	37.04
8	2	2	1	2	1	1	2	66.83	36.50

Table 3. The signal-to-noise ratios of levels 1 or 2 of each feature regarding the Autompg classification problem

	A	B	C	D	E	F	G
Level 1	36.63	37.20	36.70	37.32	36.97	36.87	36.79
Level 2	37.27	36.70	37.20	36.58	36.93	37.03	37.10

Table 3 lists the signal-to-noise ratios of levels 1 or 2 of each feature regarding the Autompg classification task. As mentioned earlier, the higher the signal-to-noise ratio (SNR), the better the classification performance (i.e., classification accuracy). As a result, features B, D and E, whose SNR for level 1 is greater than that for level 2, are preferred to be selected in the final feature subset for pattern classification. By contrast, features A, C, F, and G, whose SNR for level 2 is greater than that for level 1, are preferred to be removed from the original feature set for pattern classification. Consequently, the final feature subset obtained by using the proposed method for the Autompg classification problem is {B, D, E}. The corresponding classification accuracy is 86.43%, which is significantly better than that of each experimental trial in Table 2.

5 Experimental Results

To demonstrate the performance of the proposed hybrid Taguchi-Grey-based method for feature subset selection, ten real datasets (classification tasks) [3] were used for performance comparison. Table 4 describes the main characteristics of the datasets.

Table 5 represents the classification accuracies (as mentioned earlier) of the above-mentioned grey-based nearest neighbor rule with respect to the above classification problems when the proposed hybrid Taguchi-Grey-based method for feature subset selection is performed or not (In the experiments, the cross-validation technique [21] was used for measuring the classification accuracies). The average classification accuracies regarding these classification problems can be increased from 82.35% to 85.46% when the proposed hybrid Taguchi-Grey-based method for feature subset selection is applied. That is, experimental results demonstrate that the final feature subset obtained by using the proposed method is helpful for pattern classification.

Table 4. Details of experimental classification problems [3]

Classification task	Number of instances	Number of classes	Number of features and their types
Autompg	398	3	7 (2-S, 5-C)
Breastw	699	2	9 (9-C)
Bridges	105	6	12 (9-S, 3-C)
Hcleveland	303	5	13 (8-S, 5-C)
Hepatitis	155	2	19 (13-S, 6-C)
Hhunggarian	294	2	12 (7-S, 5-C)
Tae	151	3	5 (4-S, 1-C)
Voting	435	2	16 (16-S)
Wine	178	3	13 (13-C)
Zoo	101	7	16 (16-S)

C: Continuous, S: Symbolic

Table 5. The classification accuracies (as mentioned earlier) of the above-mentioned grey-based nearest neighbor rule with respect to the above classification problems when the proposed hybrid. Taguchi-Grey-based method for feature subset selection is performed or not.

Classification problem	The proposed method is not used for feature subset selection	The proposed method is used for feature subset selection
Autompg	69.10	**78.89**
Breastw	95.85	**96.57**
Bridges	87.62	**91.43**
Hcleveland	55.78	**57.43**
Hepatitis	80.00	**83.87**
Hhunggarian	75.85	**78.91**
Tae	**66.23**	**66.23**
Voting	92.87	**94.71**
Wine	96.63	**98.87**
Zoo	96.04	**97.03**
Average	81.60	**84.39**

6 Conclusions

In this paper, a novel hybrid Taguchi-Grey-based method for feature subset selection is proposed. The two-level orthogonal array is employed in the proposed method to provide a well-organized and balanced comparison of two levels of each feature (i.e., the feature is selected for pattern classification or not) and interactions among all features in a specific classification problem. Accordingly, the grey-based nearest neighbor rule and the signal-to-noise ratio (SNR) are used to evaluate and optimize the features of the specific classification problem. Experiments performed on different application domains are reported to demonstrate the performance of the proposed hybrid Taguchi-Grey-based method. The proposed method yields superior performance and is helpful for improving the classification accuracy in pattern classification.

References

1. Aha, D.W., Kibler, D., Albert, M.K.: Instance-based Learning Algorithms. Machine Learning 6, 37–66 (1991)
2. Bishop, C.M.: Neural Networks for Pattern Recognition, Oxford (1995)
3. Blake, C.L., Merz, C.J.: UCI Repository of Machine Learning Databases Irvine, CA: University of California, Department of Information and Computer Science (1998), http://www.ics.uci.edu/ mlearn/MLRepository.html
4. Brassard, G., Bratley, P.: Fundamentals of Algorithms. Prentice Hall, New Jersey (1996)
5. Cawley, G.C., Talbot, N.L.C.: Efficient Leave-one-out Cross-validation of Kernel Fisher Discriminant Classifiers. Pattern Recognition 36(11), 2585–2592 (2003)
6. Cover, T.M., Hart, P.E.: Nearest Neighbor Pattern Classification. IEEE Trans. on Information Theory 13(1), 21–27 (1967)
7. Dasarathy, B.V.: Nearest Neighbor (NN) Norms: NN Pattern Classification Techniques. IEEE Computer Society Press, Los Alamitos, California (1990)
8. Deng, J.: The Theory and Method of Socioeconomic Grey Systems. Social Sciences in China 6, 47–60 (1984)
9. Deng, J.: Introduction to Grey System Theory. The Journal of Grey System 1, 1–24 (1989)
10. Deng, J.: Grey Information Space. The Journal of Grey System 1, 103–117 (1989)
11. Doak, J.: An Evaluation of Feature Selection Methods and Their Application to Computer Security. Technical Report, Univ. of California at Davis, Dept. Computer Science (1992)
12. Duda, R.O., Hart, P.E.: Pattern Classification and Scene Analysis. John Wiley & Sons, Chichester (1973)
13. Hall, M.A.: Correlation-based Feature Subset Selection for Machine Learning. PhD Dissertation, University of Waikato (1998)
14. Huang, C.C., Lee, H.M.: An Instance-based Learning Approach based on Grey Relational Structure. Applied Intelligence 25, 243–251 (2006)
15. Inza, I., Larrañaga, P., Sierra, B.: Feature Subset Selection by Bayesian Networks: a Comparison with Genetic and Sequential Algorithms. International Journal of Approximate Reasoning 27, 143–164 (2001)
16. Kohavi, R., John, G.H.: Wrappers for Feature Subset Selection. Artificial Intelligence 97(1-2), 273–324 (1997)
17. Liu, H., Motoda, H.: Feature Selection for Knowledge Discovery and Data Mining. Kluwer Academic, Boston (1998)
18. Liu, H., Setiono, R.: A Probabilistic Approach to Feature Selection - A Filter Solution. In: Proceedings of 13th International Conference on Machine Learning, pp. 319–327 (1996)
19. Liu, H., Yu, L.: Toward Integrating Feature Selection Algorithms for Classification and Clustering. IEEE Trans. Knowl. Data Eng. 17(4), 491–502 (2005)
20. Quinlan, J.R.: Induction of Decision Trees. Machine Learning 1, 81–106 (1986)
21. Stone, M.: Cross-validatory Choice and Assessment of Statistical Predictions. Journal of the Royal Statistical Society B 36, 111–147 (1974)
22. Taguchi, G.: Introduction to Quality Engineering. Tokyo: Asian Productivity Organization (1986)
23. Wu, Y., Wu, A., Taguchi, G.: Taguchi Methods for Robust Design. New York: ASME, pp. 7–10 (2000)