Mining Transformed Data Sets

Alex Burns[1], Andrew Kusiak[1], and Terry Letsche[1]

[1] Intelligent Systems Laboratory, Mechanical and Industrial Engineering,
2139 Seamans Center, The University of Iowa,
Iowa City, IA 52242 – 1527, USA
andrew-kusiak@uiowa.edu

Abstract. This research presents a method to select an ideal feature subset of original and transformed features. The feature selection method utilizes a genetic wrapper scheme that employs classification accuracy as its fitness function. The feature subset generated by the proposed approach usually contains features produced by different transformation schemes. The selection of transformed features provides new insight on the interactions and behaviors of the features. This method is especially effective with temporal data and provides knowledge about the dynamic nature of the process. This method was successfully applied to optimize efficiency of a circulating fluidized bed boiler at a local power plant. The computational results from the power plant demonstrate an improvement in classification accuracy, reduction in the number of rules, and decrease in computational time.

1 Introduction

Data transformation is an integral part of data mining and knowledge discovery. Transforming data allows for an increased understanding of the data and discovery of new and interesting relationships between features. There are numerous methods for transforming data, for example:

Arithmetic operators: (+, -, *, /) This method may involve arithmetic operators applied to a single or multiple features (e.g., X^2, X/Y^3).

Combination of features: This transformation technique combines two or more features to form a new feature.

Dicretization of features: This method encompasses taking raw feature values and grouping them with similar values.

Denoising features: Fourier transforms, wavelet transforms, moving averages, and so on.

Over-fitting is a problem of concern in data mining that may be reduced through the application of data transformation. Temporal data that contains high levels of noise may be particularly susceptible to over-fitting. Some data transformation functions may reduce or enhance these concerns. Discretization and denoising can be effective transformations for the reduction of over-fitting, whereas, the creation of

A. Burns, A. Kusiak, and T. Letsche, Mining Transformed Data Sets, Proceedings of the 8th International Conference on Knowledge-Based Intelligent Information and Engineering Systems, M.Gh. Negoita, R.J. Howlett, and L.C. Jain (Eds), Knowledge-Based Intelligent Information and Engineering Systems, LNAI 3213, Vol. I, Springer, Heidelberg, Germany, 2004, pp. 148-154.

derived features may reduce or increase over-fitting. Data that is denoised too significantly could reduce the quality of the discovered knowledge. The point where there is a loss of knowledge can be difficult to discern due to the fact the classification accuracy and the performance of the algorithm may not indicate this issue. Therefore, the selection of best transformation scheme is critical to success in data mining.

Data transformation is used often in data mining. For example, data is often normalized to improve the effectiveness of the learning algorithms (clustering, neural networks), but the effects of data transformation on classification accuracy and knowledge discovery has been limited. The significance of feature transformation was demonstrated in [1]. The research centered on "feature bundling" and was demonstrated to improve classification accuracy. An exponential smoothing transformation of stock data was considered in [2]. Different time intervals were applied to different features to capture the dynamic nature of the stock market. In [3], a genetic algorithm wrapper method was applied for feature set reduction and feature construction. The method created/transformed features through a series of arithmetic operators as part of a genetic algorithm.

This paper outlines a method for selecting the best feature transformations with a wrapper feature selection approach based on genetic algorithm (GA). The proposed method can be utilized with temporal data to improve the quality of knowledge. The GA wrapper is applied to several transformed data sets to derive the best subset of features from each set. The final feature subset not only increases the classification accuracy and knowledge generalization, but also indicates the level of sensitivity of features.

The feature selection approach presented in this paper was demonstrated on power plant data. The data is temporal, nonlinear, and complex. Two different data transformation schemes were considered. The first one involves moving averages and the second approach uses wavelet transformations. Applying these transformations increases classification accuracy of the extracted knowledge, enhances understanding of the behavior of the features, results in more generalizable rules sets (i.e., reduction of the number of rules), and decreases the computation time.

2 Techniques

This section of the paper details the transformation schemes that are critical to the proposed approach.

2.2 Wavelets

Wavelet transformations were developed to express the frequency domain and the time locality of an input function. The fact that wavelets capture the temporal nature of the data is essential to this method.

Wavelet transformations consist of a "family" of functions [4]. The Haar wavelet, which is utilized in the research, is frequently used as its computational complexity is much lower than many other wavelet families. Also, there exists a simple, less com-

A. Burns, A. Kusiak, and T. Letsche, Mining Transformed Data Sets, Proceedings of the 8th International Conference on Knowledge-Based Intelligent Information and Engineering Systems, M.Gh. Negoita, R.J. Howlett, and L.C. Jain (Eds), Knowledge-Based Intelligent Information and Engineering Systems, LNAI 3213, Vol. I, Springer, Heidelberg, Germany, 2004, pp. 148-154.

putationally complex algorithm to compute the wavelet coefficients. The algorithm begins by reading a vector of values into a one-dimensional array. It sweeps through the vector multiple times to determine the wavelet coefficients. If the wavelet's transformed value is less than the threshold, then the value is set to zero. Once this has been done, the fast in-place inverse transform is called to recompose the transformed wavelet coefficients back into denoised data. This is repeated for each feature. The threshold is defined by the user to determine the level of denoising, where higher threshold values indicate a higher degree of denoising.

2.2 Genetic Algorithm and Wrappers

A genetic algorithm is a search technique that is based on natural systems. The algorithm generates a set of solutions (chromosomes), each with corresponding fitness [5]. The three main operators in a genetic program are reproduction, crossover, and mutation. Reproduction is the process of selecting the chromosomes for breeding based on their fitness values. Chromosomes with high fitness values are more likely to be selected for reproduction. The crossover and mutation operators are utilized to generate new populations of chromosomes increase the search space of the GA.

A wrapper is a method incorporating a search algorithm and a learning classifier to define ideal feature subsets [6]. The wrapper in this model utilizes a genetic algorithm to produce possible feature subsets and a decision tree to evaluate the quality of each subset. The number of genes in each chromosome equals the number of features in the data set.

For each new population the fitness of each chromosome is computed and evaluated. The entire process is repeated until an optimal solution is determined or a termination criterion is met (e.g., an upper limit on the number of generations). Each chromosome (i.e., feature subset) is evaluated by the decision tree algorithm [7]. The algorithm was selected because it is widely used as well as it generates implicit knowledge in the form of rules. The fitness function of the wrapper algorithm is the classification accuracy of the rules generated from the chromosome. The final subset of features includes those resulting in the highest classification accuracy of the knowledge discovered by the decision tree algorithm.

2.3 Data Mining

Data mining algorithms identify valid, novel, potentially useful, and ultimately understandable patterns from data that can be used for making high confidence classifications [8]. A typical data mining algorithm generates rules that describe relationships between the input features and an outcome. Discovering hidden patterns in the data may represent valuable knowledge that might lead to discoveries. There are various data mining algorithms ranging from decision trees to clustering. The decision tree algorithm [7] utilized in this method produces rules in the following format:

IF Boiler Master <= -0.53 AND Air Master > -1.5 AND Air Fuel Ratio > 0.13 AND Biomass Feed Rate > 0.4 THEN Efficiency = 88_90

Each rule is an IF-THEN statement including the premise and the conclusion. This algorithm was chosen due to its simplicity and its production of explicit and understandable rules.

3 Feature Selection Method

This section describes a model that identifies the best feature subset. Several transformation schemes are applied to the original data set. A genetic wrapper feature selection algorithm is utilized to identify the key features from each of the transformed data sets. The selected features are then combined into a single data set. The genetic wrapper selection method is applied to the combined data set (see Figure 1).

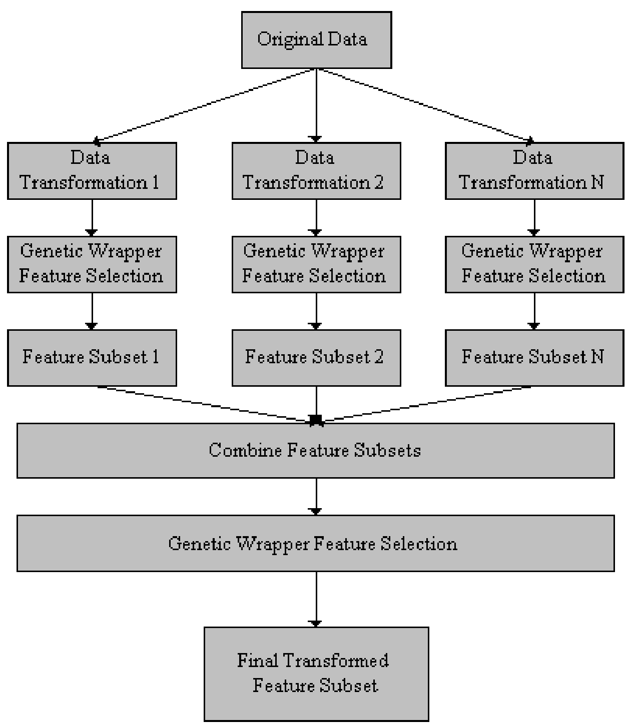

Fig. 1. Feature transformation and selection model

A. Burns, A. Kusiak, and T. Letsche, Mining Transformed Data Sets, Proceedings of the 8th International Conference on Knowledge-Based Intelligent Information and Engineering Systems, M.Gh. Negoita, R.J. Howlett, and L.C. Jain (Eds), Knowledge-Based Intelligent Information and Engineering Systems, LNAI 3213, Vol. I, Springer, Heidelberg, Germany, 2004, pp. 148-154.

Analyzing the final set of selected features will provide insight into the process and the feature interactions. A selected feature that was not transformed indicates that it is sensitive and any denoising would be detrimental to the knowledge quality, e.g., measured with classification accuracy. Conversely, selected features with high level of denoising may suggest that they are critical for outlier detection or large process shifts but are robust to small changes within the process.

A data mining algorithm is applied to the final feature subset. Comparing the results of the data mining algorithm to the results obtained from applying the same algorithm to the original data set will demonstrate the improvement the quality of the discovered knowledge.

4 Case Study

The method outlined in the previous section was demonstrated on industrial data obtained from a Circulating Fluidized Boiler (CFB). The boiler provides an excellent case study due to that fact it is a complex and temporal environment. Furthermore, there has been some research in the area that utilizes wavelets for data transformation. The applications include: partial discharge monitoring [9] and low frequency components of electrostatic discharges have been extracted using wavelets [10].

For the purposes of this case study, data on fourteen features was collected in one-minute intervals over several days. The parameters consisted of both control parameters and observed parameters. The parameters included primary and secondary air flows, boiler temperatures, pressures, and oxygen levels. The resulting data set consisted of over 12,000 observations.

The fourteen features were used to predict combustion efficiency of the boiler in the applications of the decision tree. These applications include the fitness function of the GA wrapper and the applications of the decision tree for the evaluation of the feature subsets.

Any transformation scheme can be utilized with this method, but moving average and wavelets were the focus of the case study. Both schemes capture the time behavior of data (vertical relationships) that is of importance in mining temporal data sets. The transformation were applied and examined separately. Six moving range transformations (original data, 10, 20, 30, 40, and 60 minute moving averages) were considered. Each transformation was applied to each of the features. The GA wrapper selected the most ideal feature subsets for each transformation scheme. That is, the GA wrapper selected the ideal feature subset from the set of all features that had been transformed with a 20 minute moving average. This was repeated with each moving range transformation as well as the original data. The selected features were then combined together and the GA wrapper selected the ideal subsets from the combined data set.

Four wavelet transformations were analyzed (0.3, 0.2, 0.1, and 0.01). The same procedures that were used with moving average transformation were applied to a wavelet transformation scheme.

A. Burns, A. Kusiak, and T. Letsche, Mining Transformed Data Sets, Proceedings of the 8th International Conference on Knowledge-Based Intelligent Information and Engineering Systems, M.Gh. Negoita, R.J. Howlett, and L.C. Jain (Eds), Knowledge-Based Intelligent Information and Engineering Systems, LNAI 3213, Vol. I, Springer, Heidelberg, Germany, 2004, pp. 148-154.

5 Results

The decision tree algorithm was applied to the original feature set as well as the final feature subset generated from both the moving range and wavelet transformation. The classification accuracies are a result of 10-fold cross validation [11] and it should be noted that all applications of the decision tree were completed on the same computer. The results in terms of predication accuracy, number of rules, and computation time for the moving range transformation are shown in Table 1.

Table 1. Results from the moving range transformation

Metric	Original feature set	GA wrapper selected subset
Classification accuracy (%)	92.8	93.5
Number of rules	240	228
Computational time (s)	121.4	104

It is evident that the features selected by the proposed approach improved the performance of the decision tree in all three metrics. The results from the wavelet transformations can be seen in Table 2. The difference between the original feature set metrics for the wavelet and moving average is due to the fact that the efficiency outcome was also transformed with a moving average and wavelet in the respective trials.

Table 2. Results from the wavelet transformation

Metric	Original feature set	GA wrapper selected subset
Classification accuracy (%)	67	69.9
Number of rules	385	348
Computational time (s)	170.1	169.9

The results from the wavelet transformation are not as dramatic as the moving range, but there is still marginal improvement in all metrics.

6 Conclusion

In this paper an approach for selection of the best transformed feature subset is presented. The approach utilizes a genetic algorithm wrapper and several data transformation schemes. The final feature subset contains not only the best features but also their best transformations. The feature transformation approach is well suited for temporal data as it provides new insight about the dynamics of the data and determines parameter sensitivity.

The approach was demonstrated on data from a boiler combustion process. A wavelet transformation scheme and a moving average scheme were applied to the data. The moving average scheme produced significant improvements in terms of

A. Burns, A. Kusiak, and T. Letsche, Mining Transformed Data Sets, Proceedings of the 8th International Conference on Knowledge-Based Intelligent Information and Engineering Systems, M.Gh. Negoita, R.J. Howlett, and L.C. Jain (Eds), Knowledge-Based Intelligent Information and Engineering Systems, LNAI 3213, Vol. I, Springer, Heidelberg, Germany, 2004, pp. 148-154.

classification accuracy, and the reduction in the number of rules and processing time. The approach provided more insight by repeatedly selecting the same features regardless of the type of transformation scheme. These features might be crucial to controlling the process. Furthermore there were some features that were selected for only specific transformations. These features may require only the level of control that was defined by the denoising transformation. The wavelet transformed data produced little improvement. The wavelet transformations could have denoised the data too significantly. The type of denoising transformation as well as the denoising scheme itself are critical to the quality of solutions.

References

1. Kusiak, A.: Feature Transformation Methods in Data Mining. IEEE Trans. Electron. Packag. Manuf. 24 (2001) 214–221
2. Weigend, A., Chen, F., Figlewski, S., and Waterhouse, S.R.: Discovering Technical Trades in the T-Bond Futures Market. In: Argawal, R., Stolorz, P., Piatetsky-Shapiro, G. (eds.): Proc. Fourth Int. Conf. Knowledge Discovery and Data Mining. AAAI Press, New York (1998) 354-358
3. Vafaie, H., De Jong, K.: Feature Space Transformation Using Genetic Algorithms. IEEE Intell. Systems. 13 (1998) 57–65
4. Hubbard, B.B.: The World According to Wavelets: The Story of a Mathematical Technique in the Making. 2nd edn. Peters, A.K. (eds.) Natick, MA (1998)
5. Goldberg, D. E.: Genetic Algorithms in Search, Optimization, and Machine Learning. Addison Wesley Longman Inc. New York (1989)
6. Kohavi, R. and John, G.H.: Wrappers for Feature Subset Selection. Artif. Intell. 97 (1997) 273-324
7. Quinlan, J.R.: Induction C 4.5 Programs for Machine Learning. Morgan Kaufmann, San Meteo, CA (1992)
8. Fayyad, U., Piatetsky-Shapiro, Smyth Uthurusamy, R.: Advances in Knowledge Discovery and Data Mining. AAAI/MIT Press (1995)
9. Huang, C.-M., and Huang, Y.-C.: Combined Wavelet-Based Networks and Game-Theoretical Decision Approach for Real-Time Power Dispatch. IEEE Trans. Power Syst. 17 (2002) 633-639
10. Masugi, M.: Multiresolution Analysis of Electrostatic Discharge Current From Electromagnetic Intereference Aspects. IEEE Trans. on Electromagn. Compat. 45 (2003) 393-403
11. Stone, M.: Cross-validatory Choice and Assessment of Statistical Predictions. J. R. Stat. Soc. 36 (1974) 111-147

A. Burns, A. Kusiak, and T. Letsche, Mining Transformed Data Sets, Proceedings of the 8th International Conference on Knowledge-Based Intelligent Information and Engineering Systems, M.Gh. Negoita, R.J. Howlett, and L.C. Jain (Eds), Knowledge-Based Intelligent Information and Engineering Systems, LNAI 3213, Vol. I, Springer, Heidelberg, Germany, 2004, pp. 148-154.