Proceedings of the SICE Annual Conference 2017
September 19-22, 2017, Kanazawa University, Kanazawa, Japan

A Hybrid Case Base Reasoning System for Forecasting

Chih-Ching Hsiao [1†] ,Chen-Chia Chuang [2] and Jin-Tsong Jeng [3]

[1]Department of Information Technology, Kao Yuan University, Kaohsiung, Taiwan
(E-mail: cchsiao@cc.kyu.edu.tw)
[2] Department of Electrical Engineering, National Ilan University, I-Lan, Taiwan
(E-mail: ccchuang@niu.edu.tw)
[3]Department of Computer Science & Information Engineering, National Formosa University, Taiwan
(E-mail: tsong@nfu.edu.tw)

Abstract: CBR (Case-based reasoning) is an effective reasoning mechanism that solves a new problem by remembering a previous similar knowledge and by reusing information and knowledge of that features. The case-base is a set of small case-bases. Each small case-base can be viewed as the result of granular computing. To obtain an efficient CBR system, this paper proposed a hybrid CBR system by introducing feature selection and Granular computing, it also incorporate similarity margin concept and Gaussian kernel fuzzy rough sets in case-based organization.

1. INTRODUCTION

The traditional forecasting methods have two categories such as (i) the empirical method and (ii) the dynamical equations. The first method is useful for predicting local-scale weather if recorded data are enough. The second method is based on the dynamic equations and simulations, and is referred to as computer modeling technology. The dynamical method is only useful for modeling large-scale weather phenomena and may not efficiently predict short-term weather's status. The artificial intelligence methods for weather prediction currently included model output statistics, fuzzy logic, expert systems, genetic algorithm, and particle swarm optimization [1,2]. Predicting uncertain time series appears to be serious problem in focus area, as the existing forecast of certain time series does not purely mirror the ability of predicting future decisions. Case-based reasoning (CBR) is an approach to problem solving by utilizing previous cases and experiences which are similar to the current one. It supposes that similar problems usually have similar solutions, and those problems may often take place. CBR is a successful artificial intelligence methodology currently employed in a variety of applications. Therefore, CBR can mine big data[3]. A number of CBR forecast systems have been developed [4,5]. Zhu et al.[6] proposes a hybrid CBR system by introducing reduction technique in feature selection and cluster analysis in case organization. It can effectively enhance the performance of the CBR system. In CBR, the weights of feature attributes directly affect the quality of problem solving. In [7] proposed a membrane computing-based approach to optimize the attribute weights, which can effectively improve the quality of problem solving for a CBR system. To perform the retrieval process in CBR systems, it typically exploits similarity knowledge. Kang et al.[8] proposed a novel retrieval strategy that substantially outperforms similarity-based retrieval by leveraging association knowledge, encoded via a certain form of association rules, in conjunction with similarity knowledge. Case retrieval and case revise are core parts of CBR. In [9] introduces an improved case-based reasoning method based on fuzzy c-means clustering (FCM), mutual information and support vector machine(SVM). FCM is used to divide case base to improve efficiency of the algorithm. Mutual information is introduced to calculate weights of each condition attribute and SVM is adopted to build an optical regression model for case revise. To improve the accuracy of case reuse, a two-layer random forests model is proposed [10]. First, clustering analysis is used to organize the cases in the case base, and gSOM algorithm is adopted to automatically detect the cluster number on the basis of the structure of the data. Then, a two-layer model scheme is proposed to model the mapping in every cluster.

In this paper, the forecasting problem is considered. Feature selection can be used as a purely structural method for reducing dimensionality using information contained within the dataset and preserving the meaning of the features. CBR is a computer technique, it is very effective in situations, especially in Weather Prediction, where the acquisition of the case-base and the determination of the features is straightforward compared with the task of developing the reasoning mechanism. The case-base is a set of small case-bases. Each small case-base can be viewed as the result of granular computing (GrC). Moreover, the similarity measures are usually used for case-based organization from the weather case base. However, inappropriate feature selection and case selection may not only present a dilemma in case retrieval, but also greatly increase the case base.

2. CBR BASED FORECAST MODEL

The CBR based forecast model is illustrated in Fig.1 [5].The objects manipulated by the prediction system are observational weather data. A target weather case is a set of weather elements to be predicted. A source weather case is a set of observational weather elements that describes the historical developments of the weather. Those weather cases are stored in the weather case library in a well-structured way to facilitate retrieval of cases. The life cycle of the retrieval, reuse, revision and retain of the cases may loop due to the complexity of the weather phenomena and the incompleteness of the forecast knowledge.

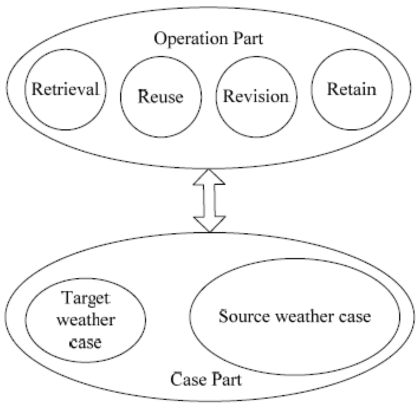

Fig. 1 A CBR based forecast model

3. HYBRID CBR SYSTEM

CBR is very effective in situations, especially in Weather Prediction, where the acquisition of the case-base and the determination of the features is straightforward compared with the task of developing the reasoning mechanism. To obtain an efficient CBR system, selection of proper features and suitable cases with appropriate case organization are very important. We propose a hybrid CBR system by introducing feature selection and Granular computing in case-based organization. The relate theories are briefly stated as follows.

3.1 Feature Selection with Multi-Granulation

A subset of a data set can be considered as a small granularity. Given a large-scale data set, the proposed algorithm[11] first selects different small granularities and then estimate on each small granularity the reduct of the original data set. Fusing all of the estimates on small granularities together, the algorithm can accomplish its feature selection.

3.2 Case organization based on Multilayer Perceptron

According to the resulting feature selection, then how to organize those cases will be considered. In this paper, a multilayer perceptron (MLP) is introduced to perform cluster analysis to the case library so the large case base can be partitioned into small groups to reduce the size of stack memory and optimize the retrieval performance. In [12] proposed MLP comes with interval-valued weights and biases, and is trained to fit data with different levels of granularity. The objective function is considered as interval-valued similarity measure in this paper.

3.3 Interval-value Similarity measure

Hedjazi [13] proposes a feature selection method for symbolic interval data based on similarity margin. In this method, classes are parameterized by an interval prototype based on an appropriate learning process. A similarity measure is defined in order to estimate the similarity between the interval feature value and each class prototype. Then, a similarity margin concept has been introduced. The heuristic search is avoided by optimizing an objective function to evaluate the importance of each interval feature in a similarity margin framework.

In [14], the authors incorporate Gaussian kernel with fuzzy rough sets and proposed a Gaussian kernel approximation based fuzzy rough set model. The basic idea is the similarity between two samples is computed with Gaussian kernel function. Therefore Gaussian kernel induces a fuzzy relation satisfying the properties of reflexivity and symmetry. Moreover, it can introduce Gaussian kernel for computing fuzzy T-equivalence relations in fuzzy rough sets and thus approximate arbitrary fuzzy subsets with kernel induced fuzzy granules.

Hence the CBR systems are divided into several small granularities. When a new problem comes, the problem feature subsets will be first represented and put together with the past cases to conduct clustering. Finally, the new problem is mapped with the previous cases and the most similar case can be obtained by case organization using similarity measurement. Feature selection and Granular computing in the case organization are crucial steps in CBR, since the retrieval efficiency and accuracy even the success of the CBR system are heavily dependent on their quality.

4. CASE STUDY

The problem with weather forecasting is the complexity that comes up of the many variables that may affect forecasting. The major goal of this paper is to introduce a forecasting method that based on a hybrid CBR system that will be used for forecasting. This model will be depend on collecting historical weather data and collect all possible attributes that are relevant to this model are: temperature: mean, max and min, cooling and growing degree days, dew point, humidity, wind speed, etc.

In the current work, we employ UCI large scale data set to test the feature selection algorithm. The dataset concern Nursery which are outlined in Tables 1. The experimental results are reported in Tables 2. shows the computational time and the classification error. The experimental results shows that also the top ranked 10 interval features yields the smallest classification error. It proves that the proposed approach can find their feature subsets for interval data in a large dataset based on multi-granulation.

Table 1 Description of data sets.

Data sets	Samples	Attributes	Classes
Nursery	12960	8	5

Table 2. Comparison of feature subsets, error and computational time

Data sets	Feature subsets	Smallest classification Error	Computational Time (S)
Nursery	1,2,3,4, 5,6,7,8	0.15	25.40

5. CONCLUSION AND FUTURE WORK

The experimental results show that can find their feature subsets in a large dataset based on multi-granulation for CBR. The cases will be organized from the resulting feature selection based on interval-valued similarity measure and be applied to forecast in the future work.

ACKNOWLEDGMENTS

This work was supported in part by Ministry of Science and Technology of Taiwan under Grant MOST 105-2221-E-244-002 -.

REFERENCES

[1] S. Esfandeh and M. Sedighizadeh, “Meteorological Data Study and Forecasting Using Particle Swarm Optimization Algorithm,” *World Academy of Science, Engineering and Technology*, vol. 59, pp. 2117-2119, 2011.

[2] XU L., GUAN Tian-yi, LI Yan-nong, ZHENG Wen-jing, GUO Jing-yi, “PM2.5 Measuring Method Using RBF Neural Network Combined With PSO Algorithm,” *27th Chinese Control and Decision Conference (CCDC)*,pp.5616-5619, 2015.

[3] Petra Perner, “Mining sparse and big data by case-based reasoning,” *Procedia Computer Science*, vol. 35. pp.19- 33, 2014.

[4] S. Ji, S. F. Yuan, and J. Yue, “A method of weather case generation based on similarity rough set”, *Proc. Int. Conf. on Management and Service Science*, pp.1-4, 2009.

[5] Hanrong Lu, Weijiang Wang and Feng Zheng , “A Case-Based Reasoning Approach to Local-Scale and Short-Term Weather Forecast,” *IEEE Conf. SMC*, pp.228-231, 2012.

[6] Guo-Niu Zhu,JieHun, JinQi,JinMa and Ying-HongPeng , “An integrated feature selection and cluster analysis techniques for case-based reasoning,” *Engineering Applications of Artificial Intelligence*, vol.39, pp.14-22, 2015.

[7] A. Yan, H.s. Shao and Z. Guo, “Weight optimization for case-based reasoning using membrane computing,” *Information Sciences*, vol.287, pp.109-120, 2014.

[8] Y. B. Kang, S. Krishnaswamy and A. Zaslavsky, “A Retrieval Strategy for Case-Based Reasoning Using Similarity and Association Knowledge,” *IEEE Trans on Cyber.* , vol. 44, no.4, pp.473-487, 2014.

[9] M. Han and Zhan Ji Cao, “An improved case-based reasoning method and its application in endpoint prediction of basic oxygen furnace,” *Neurocomputing* , vol.149, pp.1245-1252, 2015.

[10] S.S. Zhong, X. l. Xie, L. Lin, “ Two-layer random forests model for case reuse in case-based reasoning,” *Expert Systems With Applications,* vol.42, pp. 9412-9425, 2015.

[11] C. C. Hsiao, C. C. Chuang, J. T. Jeng, S. F. Su, “The Feature Selection with Multi-Granulation for Interval-Value Data,” *New Trends on System Sciences and Engineering*. *Amsterdam, Netherlands: IOS*. Jul,: vol.276, pp.89-99, 2015.

[12] M. G.C.A. Cimino, B. Lazzerini , F. Marcelloni and W. Pedrycz, “Genetic interval neural networks for granular data regression,” *Information Sciences*, vol.257, pp.313-330, 2014.

[13] Lyamine Hedjazi, Joseph Aguilar-Martin and Marie-Véronique Le Lann, “Similarity-margin based feature selection for symbolic interval data,” *Pattern Recognition Letters* , vol.32, pp.578-585, 2011.

[14] Junpeng Guoa, Wenhua Li a, Chenhua Li and Sa Gaoa, “Standardization of interval symbolic data based on the empirical descriptive statistics,” *Computational Statistics and Data Analysis*, vol.56, pp.602-610, 2012.