A novel hybrid optimization algorithm of computational intelligence techniques for highway passenger volume prediction

Wu Deng [a,b,c,*], Wen Li [a], Xin-hua Yang [a]

[a] Software Institute, Dalian Jiaotong University, Dalian 116028, PR China
[b] Key Laboratory of Advanced Design and Intelligent Computing (Dalian University), Ministry of Education, Dalian, PR China
[c] Key Laboratory of Intelligent Manufacture of Hunan Province, Xiangtan University, Xiangtan, PR China

ARTICLE INFO

ABSTRACT

A novel hybrid optimization algorithm combining computational intelligence techniques is presented to solve the multifactor highway passenger volume prediction problem. In this paper, we can get and discretize a reduced decision table, which implies that the number of evaluation criteria such as travel quantity, fixed-asset investment, railway mileage, and waterway passenger volume are reduced with no information loss through rough set theory (RST) method. Particle swarm optimization (PSO) algorithm based on the random global optimization is inducted into the network training. The PSO algorithm is used for glancing study in order to confirm the initial values, and then the back propagation neural network (BPNN) is used for given accuracy to found the PSO-BPNN model. And this reduced information is used to form a classification rule set, which is regarded as an appropriate input parameter to training PSO-BPNN model. The RST-PSO-BPNN model is obtained to forecast highway passenger volume. The rules developed by RST analysis show the best prediction accuracy if a case matches any one of the rules. The keystone of this hybrid optimization algorithm is using rules developed by RST for an object that matches any one of the rules and the PSO-BPNN model for one that does not match any of them. The effectiveness of our optimization algorithm was verified by experiments comparing the traditional gray model method. For the experiment, highway passenger volumes of China during the period 1995–2009 were selected, and for the validation, the novel hybrid optimization algorithm is reliable.

© 2010 Elsevier Ltd. All rights reserved.

1. Introduction

Econometric demand models have been used for many years to provide important behavioral insights into the highway passenger businesses in many the total increase in journeys (Yin, Wang, Xu, et al., 2002). Correct prediction is the basis for a scientific decision; highway passenger volume prediction (HPVP) is a scientific analysis to make scientific judgments for HPVP according to the existing highway passenger volume data in the future (Tsai, Lee, & Wei, 2009). With the sustained, rapid and healthy growth of the economy and the gradual improvement of the denizen income and living standards, changing of consumption fashion and conversion of consumption concept, passengers have an ever-increasing demand for the highway transportation, which is also facing more and more severe competition in the transportation market (Khashei & Bijari, 2010; Pai & Hong, 2005). Therefore, correct HPVP will be of great significance for investment structure, optimization allocation of funds and management decision, etc.

In the artificial intelligence domain, computational intelligence (CI) is a fairly new category of algorithms, including evolutionary computing, fuzzy computing, rough set, artificial neural network and granular computing. Wide applications of these algorithms have proven that they are very useful in practice to solve real world problems when deterministic solutions are hard to obtain. So many different prediction model prototypes have been applied, such as exponential smoothing method, non-parametric regression, artificial neural network, linear recursive method, expert experience predicting, gray prediction model and rough set theory, and combination prediction method (Inuiguchi, 2006; Khosravi, Nahavandi, & Creighton, 2010; Li & Wang, 2004). Although comparative studies and selection models have discussed the issue of models selection, it is still controversial to say which model prototype can globally obtain the best predictive performance among alternatives. The common advantage of rough set theory (RST) and artificial neural network (ANN) is that they do not need any additional information about data like probability in statistics or grade of membership in fuzzy-set theory (Bazan, Skowron, & Synak, 1994). RST has proved to be very effective in many practical applications. However, in RST, the deterministic mechanism for the description of an error is very simple (Craven & Shavlik,

1997).Therefore, the rules generated by RST are often unstable and have low classification accuracies. So RST cannot forecast for high accuracy. ANN is considered the most powerful classifier for low classification-error rates and robustness. But ANN has two obvious shortcomings when applied to large data problems. The knowledge of ANN is buried in their structures and weights (Lu, Setiono, & Liu, 1996; Swiniarski & Hargis, 2001). Particle swarm optimization (PSO) algorithm is a global optimization evolutionary algorithm; it can optimize all kinds of complex problems. So the combination of RST, PSO and ANN is very natural for their complementary features. One typical approach is to use the PSO approach to optimize the topology structure and parameters of ANN for constructing the PSO-ANN model. RST approach is used as a pre-processing tool for the PSO-ANN model (Ahn, Cho, & Kim, 2000; Pawlak, 1982; Wang & Ziarko, 1985). RST provides useful techniques to reduce irrelevant and redundant attributes from a large database with a lot of attributes. ANN has the ability to approach any complex function and possesses a good robustness. Therefore, this study takes the advantages of the inductive conclusion and adopts the combining RST, PSO and ANN (RST-PSO-ANN) to develop the HPVP model. The effectiveness of proposed novel hybrid prediction method was verified with experiments that compared the traditional gray model method and proposed the hybrid prediction method. The experiment result shows that prediction value and actual value are very proximate. This method has instructional significance for HPVP.

2. Rough set theory

Rough set theory introduced by Pawlak in 1982 is a mathematical tool to deal with vagueness, incompletion and uncertainty of information (Zhang, Wu, & Liang, 2001). The philosophy of the method is based on the assumption that with every object some information (data, knowledge) can be associated. Objects characterized by the same information are indiscernible in view of the available information. The indiscernibility relation generated in this way is the mathematical basis for the rough set theory.

2.1. Information system

An information system can be seen as a four-tuple $S = \{U, Q, V, f\}$, where U is a finite set of objects, called the universe, Q is a finite set of attributes, $V = U_{a \in Q}$, Va is a domain of attribute a, and f: $U \times Q \to V$ is a total function such that $f(x, a) \in Va$, for every $a \in Q$, $x \in U$, called an information function. In classification problems, an information system is also seen as a decision table assuming that $Q = C \cup D$ and $C \cap D = \Phi$, where C is a set of condition attributes and D is a set of decision attributes (Xie, Li, & Zhou, 2002; Zhang & Qiu, 2006).

2.2. Indiscernibilify relation

Let $S = \{U, Q, V, f\}$ be an information system: every $P \subseteq Q$ and generates a indiscernibility relation $IND(P)$ on U, which is defined as follows (Sun, Yuan, Yu, et al., 2007):

$$IND(P) = \{(x, y) \in U \times U | \forall a \in P, \quad f(x, a) = f(y, a)\}$$

Obviously, $IND(P)$ is an equivalence relation for any P. Equivalence classes of $IND(P)$ are called P-elementary sets in S. If $P = Q$, the Q-elementary set is called atoms. The family of all equivalence classes of relation $IND(P)$ on U is denoted by $U|IND(P)$, or in short, $U|P$.

2.3. Approximation of sets

Let $P \subseteq Q$ and $X \subseteq Q$. The P-lower approximation of X (denoted by P_X) and the P-upper approximation of X (denoted by P^-X) are defined in the following expressions (Kodogiannis & Anagnostakis, 2002; Wang, 2003; Wang, Wei, Zhang et al., 2007; Zhao, 2005):

$$P_-(X) = \{X \in U | P : Y \subseteq X\} \tag{1}$$

$$P^-(X) = \{X \in U | P : Y \cap X = \Phi\} \tag{2}$$

$P_-(X)$ is the set of all objects from U which can be certainly classified as elements of X employing the set of attributes P. $P^-(X)$ is the set of all objects of U which can possibly be elements of X using the set of attribute P. The P-boundary (doubtful region) of set X is defined as:

$$Bn_p(X) = P^-(X) - P_-(X) \tag{3}$$

The set $Bn_p(X)$ is the set of objects which cannot be certainly classified to X using the set of attributes P only. Decision rules derived from a decision table can be used for recommendations concerning new objects. Specifically, matching its description to one of the decision rules can support the classification of a new object. With every set $X \subseteq Q$, we can associate an accuracy of approximation of set X and P in S, or in short, accuracy of X, defined as:

$$dP(X) = card(P_-(X)) / card(P^-(X)) \tag{4}$$

2.4. Reduction of attributes

Let $S = \{U, Q, V, f\}$, the reduction of condition attribute C means a nonempty subset $P \subset C$ satisfied the following condition (Zhang, Xiao, & Wang, 2005):

(1) $IND(P) = IND(C)$
(2) There is no subset of $P' \subset P$, which satisfied $IND(P') = IND(C)$

The process of finding a smaller set of attributes than the original one with the same classification capability as the original set is called attribute reduction. Attribute reduction is one of the most important concepts in RST. A reduction is the essential part of an information system (related to a subset of attributes) which can discern all objects discernible by the original information system.
Core is the intersection of all reductions, denoted as: CORE $(C) = \cap$ RED(C), where RED(C) is the reduction of S (Rady, Kozae, & Abd El-Monsef, 2004).

Given S, condition attributes C and decision attributes D, $Q = C \cup D$, for a given set of condition attributes $P \subset C$; we can define a positive region $POS_P(D) = \bigcup_{X \in U/D} P\_X$ (Dimitras, Slowinski, & Susmaga, 1999). The positive region POS_P contains all objects in U, which can be classified without an error into distinct classes defined by $IND(D)$ based only on information in the $IND(P)$.

2.5. Degree of dependency

Another important issue in data analysis is discovering dependencies between attributes. Let an information system $S = \{U, Q, V, f\}$, $Q = CUD$ and the dependability between D and C is defined (Ahn, Cho, & Kim, 2000; Zhu, Chen, Geng, & Liu, 2008):

$$K = \gamma_c(D) = |POS_P(D)| / |U| (0 \leqslant K \leqslant 1) \tag{5}$$

If $k = 1$ we say that D depends totally on C, and if $k < 1$, we say that D depends partially (in a degree k) on C. The coefficient k expresses the ratio of all elements of the universe, which can be properly classified to blocks of the partition U/D, employing attributes C and will be called the degree of the dependency.