Computers and Electronics in Agriculture 140 (2017) 443–451

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag

Original papers

Particle Swarm Optimization based incremental classifier design for rice disease prediction

Shampa Sengupta [a,*], Asit K. Das [b]

[a] Department of Information Technology, MCKV Institute of Engineering, Liluah, Howrah 711204, West Bengal, India
[b] Department of Computer Science and Technology, Indian Institute of Engineering Science and Technology, Shibpur, Howrah 711103, West Bengal, India

ARTICLE INFO

Article history:
Received 28 January 2017
Received in revised form 22 June 2017
Accepted 23 June 2017
Available online 29 June 2017

Keywords:
Data mining
Particle Swarm Optimization
Incremental data
Dynamic classifier
Rice disease

ABSTRACT

Increase of huge amount of data in every application demands an incremental learning technique for data analysis. One of such data analysis task in dynamic environment is to design an incremental classifier for decision making and consequently updating the knowledge base of the overall system. Classifier construction depicts extraction of interesting patterns from the large repository of data and predicts the future trends based on the existing patterns. The time complexity of the classification system increases gradually and the system becomes inefficient while it is learned repeatedly for adding new group of data with the existing one in a certain interval of time. Without learning the same classifier for the whole data, if the knowledge of old data extracted by the classifier is used together with the new group of data to design the updated classifier, called incremental classifier, then time complexity reduces drastically. In the paper, the concepts of Particle Swarm Optimization technique and Association Rule Mining are used to design an incremental rule based classification system. The incremental classifier is suitable to apply on rice disease dataset for disease prediction as the characteristics of rice diseases change in time due to change of climate, biological, and geographical factors. The proposed method has been applied on both simulated rice disease dataset and benchmark datasets and the classification accuracy is measured and compared with various state of the art classification algorithms. The method is also evaluated based on some statistical measures and statistical test is done to establish its significance and effectiveness.

1. Introduction

As size of the data sets changes with time, it is very time consuming or even infeasible to run repeatedly a static knowledge acquisition algorithm for the entire dataset. Researchers are working on machine intelligence (Alpaydin, 2010) to solve this kind of problem for various real life applications. In case of incremental learning in supervised environment, new group of data arise at later stages at random basis, so restricting the environment for the specific data will be of no use, where incremental learning technique will open up the learning. Like social network and other research fields, the technique is equally useful in the field of agriculture. There are different types of rice plants diseases with different symptoms which were classified manually in early days by observing the symptoms but due to changes of climate, biological and geographical factors, characteristics of the rice diseases change with time. So, it is very difficult to diagnose the diseases manually observing the symptoms and at the same time static classification model is not an efficient solution for predicting the diseases. As a result, an automated expert incremental learning system is needed to design a classifier for rice diseases prediction. The paper (Ziarko and Shan, 1993) proposes a deterministic method for incremental modification of classification rules where the concept of decision matrix, based on Rough set Theory, is used. Ulas et al. (2009) developed an incremental classifier construction strategy and discriminant ensembles of the classifiers for some classes only. Ozawa et al. (2008) proposes an incremental learning technique on chunk data for online pattern classification problem and Chen et al. (2007) have proposed the incremental learning methodology for text data classification. A hybrid intelligent system (Seera and Lim, 2014) is developed for medical data classification using Fuzzy Min–Max neural network, Classification and Regression Tree and Random Forest for the incremental data. An incremental learning technique of hierarchical appearance model is proposed in (Wenzel and Forstner, 2008), to detect objects occurring in the

* Corresponding author.
E-mail address: shampa2512@yahoo.co.in (S. Sengupta).

http://dx.doi.org/10.1016/j.compag.2017.06.024
0168-1699/

images in hierarchy where the concept of incremental detection and classifying the new images with existing instances is used. A research paper (Ozawa et al., 2005) on online face recognition using incremental learning technique is used for adaptive learning of new features and classifiers. G. Bakırlı et al. (2011) have developed an incremental Genetic Algorithm (GA) approach for generation of dynamic classifier for the incremental datasets.

Many researchers are working for the development of some sustainable farming practice to control the crop diseases efficiently in time. Many developed systems suffer from biasness to classify the crop diseases due to consideration of user observations about the infected plants. As detection of crop diseases in time at the fields is critical for precision on-farm disease management (Schaad and Frederick, 2002), rapid development of the modern digital equipment, various computational tools and techniques attract researchers for automatic detection of crop diseases (Yang et al., 2012; Yao et al., 2009; Bravo et al., 2003; Patil and Kumar, 2011). Remote sensing technique is also used to identify the infected plants through the quantitative analysis of spectral differences. Many researchers developed expert systems (Aakif and Khan, 2015; Chaudhary et al., 2016a, 2016b; Sannakki et al., 2011; Sarma et al., 2010) for classification of diseases in the field of agriculture. Paper (Phadikar et al., 2013) automatically extracts the disease symptoms as features of the diseases and a classifier is developed based on the extracted features to predict the diseases. But the main problem with this algorithm is that it cannot be applied in incremental dataset efficiently, rather the same classifier is learned again from the beginning with the old data and newly added group of data.

The main objective of our paper is to develop a rule based classifier of high accuracy using less number of features for the incremental datasets. In a rule-based classifier, efficiency of a rule is measured based on its accuracy on test or unseen data and the number of rule components present in it. Here the concept of Particle Swarm Optimization (PSO) Algorithm (Correa et al., 2006; Freitas, 2002; Kennedy and Eberhart, 1997; Trelea, 2003; Van den Bergh, 2002; Zhao et al., 2012) and Association rule mining (Agrawal and Srikant, 1994) is used to develop an incremental classifier. The algorithm handles incremental data effectively for upgrading the existing classifier by modifying the existing rule sets whenever new set of data is added with the previous dataset. At first, PSO based training process is performed on the existing dataset to find out the initial optimal set of classification rules. When new data arrives, if static PSO based training process is re-run on the whole dataset (consisting of both existing and new data) for developing the modified classifier, then not only the efficiency of the system degrades with increased data volume but also the previous classifier already trained by the existing dataset is totally unusable that increases the overhead of the overall system. As the volume of the dataset increases with time, learning of whole data in dynamic environment rapidly increases the training time and makes the classification system inefficient. So it is desirable to upgrade the classifier with the help of new group of data and existing knowledge extracted from the previous dataset. Here, we have proposed an incremental PSO (IPSO) algorithm, which analyzes the new dataset and updates the previous classification rule sets dynamically with a reduced training time. We have used few heuristics in calculation of the fitness function for our proposed IPSO. The details of the heuristics and their importance are provided in subsequent section. Fig. 1 gives the schematic diagram of the proposed incremental classifier design technique for incremental datasets using IPSO.

Fig. 1 shows the sequence of steps in which the incremental classifier is designed and evaluated in the time interval $(t, t+\Delta t)$. The dataset available at time t is considered as existing dataset and the classification model is trained on this dataset using the concept of PSO which provides initial rule sets for dataset at instant of time t. Now after Δt time, new data, called incremental data are stored as new dataset in time interval $(t, t+\Delta t)$. In the proposed IPSO technique, initial rules and new dataset are feed in training process and the modified classification model, called new model, is generated for whole dataset available at instant of time $t+\Delta t$. This incremental classification model is evaluated by test dataset for performance measurement. At next instant of time, the new model at $(t+\Delta t)$ time is considered as old model and new group of data is taken into account for constructing further modified classifier. This process is continued after every interval of time when a group of data enters into the system. Thus, a rule based dynamic classifier is designed for incremental dataset in efficient way. Proposed method has been applied on rice disease dataset (Phadikar et al., 2013) to predict the different rice diseases to take the precautionary measures at an early stage to protect the crop as well as to provide help to the farmers.

The rest of the paper is organized as follows: Basic Concepts of PSO related to the proposed method are described in Section 2. tion 3 demonstrates the proposed dynamic rule based classification system for the incremental dataset. Section 4 shows the experimental results and comparisons with some existing state of the art methods together with the applicability of our method in rice disease prediction and finally conclusion of the paper is stated in Section 5.

2. Basics of Particle Swarm Optimization

In the proposed method, the concepts of Particle Swarm Optimization (PSO) (Correa et al., 2006) technique are used to design a rule-base classifier for incremental dataset. PSO is an evolutionary optimization algorithm proposed by Kennedy and Eberhart in 1995. In PSO, a population, called a swarm, of candidate solutions are encoded as particles in the search space. PSO starts with the random initialization of a population of particles. The whole swarm moves in the search space to search for the best solution by updating the position of each particle based on the experience of its own and its neighboring particles. In PSO, a potential solution to a problem is represented by a particle $X(i) = (x_{(i,1)}, x_{(i,2)}, \ldots, x_{(i,n)})$ in an n-dimensional search space. The coordinate $x_{(i,d)}$ of the particle $X(i)$ have a rate of change of position i.e. the velocity $v_{(i,d)}$ where $d = 1, 2, \ldots, n$. Every particle keeps a record of the best position that it has ever visited. Such a record is called the particle's previous best position and denoted by B_i. The global best position G attained by any particle so far is also recorded. Iteration comprises evaluation of each particle with the adjustment of $v_{(i,d)}$ in the direction of particle $X(i)$'s previous best position and the previous best position of any particle in it's neighborhood. The set of phases that govern PSO are: evaluate, compare and evolve. The evaluation phase measures how well each particle or candidate solution solves the problem. The comparison phase identifies the best particles and the evolve phase produces new particles based on some of the best particles previously found. These three phases are repeated until a given stopping criterion is matched. The objective of the method is to find the best particle, which gives the optimal solution of the problem. Important concepts in PSO are velocity and neighborhood topology. Each particle $X(i)$ is associated with a velocity vector. This velocity vector is updated at every generation. The updated velocity vector is then used to generate a new particle $X(i)$. The neighborhood topology defines how other particles in the swarm, such as $B(i)$ and G, interact with $X(i)$ to modify its respective