Asian Journal of Computer Science and Technology
ISSN: 2249-0701 Vol.7 No.3, 2018, pp. 7-12
© The Research Publication, www.trp.org.in

Prediction Algorithms: A Study

S. Santha Subbulaxmi[1] and G. Arumugam[2]
[1]Research Scholar, [2]Professor & Head (Retired), Department of Computer Science,
Madurai Kamaraj University, Madurai, Tamil Nadu, India
E-Mail: santhamd3@gmail.com, gurusamyarumugam@gmail.com

***Abstract* - Prediction algorithms make a prognosis of the future in a scientific way by analysing the data. They are being applied successfully to the problems in various fields and find good solutions. The objective of this paper is to describe about prediction algorithms and present the literature growth on it. It details the prediction process. It outlines the different types of prediction algorithms and the relevant publications on it. The paper summarizes the advantages & disadvantages of the prediction algorithms and the challenges to be addressed in the prediction field.**
Keywords: **Prediction Algorithm, Regression Algorithms, Instance Based Algorithms, Decision Tree Algorithms, Bayesian Algorithms, Clustering Algorithms, Artificial Neural Network Algorithms, Ensemble Algorithms**

I. INTRODUCTION

We are living in information age where everyone is experiencing abundant of data all around. Data science evolves with time and expands new capabilities for managing the abundant data and discovers knowledge from it. Data science is an inter-disciplinary field of information science, database systems, statistics and machine learning. The data science algorithms analyses the data and delineate what happened in the past, why it happened, what will happen, how we can make it happen. Among the descriptive, diagnostic, predictive, prescriptive algorithms, Prediction algorithms is the most popular data analysis form which extract models from the past and predict the future data trends.

The Prediction algorithms make a prognosis of the future in a scientific way by analysing the data. They are being applied successfully to the problems in various fields like medical, engineering, business and science to find good solutions with accuracy in reasonable amount of time. The outcome helps to control the uncertainty of future.

The objective of this paper is to describe about prediction algorithms and present the relevant research publications of it. The literature growth in this field shows up the future research directions of the field. This paper is organized as follows: Section II details the prediction process and the different types of prediction algorithms with relevant literature on it. Section III tabulates the advantages and disadvantages of each prediction algorithm. Section IV summarizes the future research direction of the prediction algorithms.

II. PREDICTION ALGORITHMS

Predictive process applies the methods and techniques in data mining, probability, statistics, machine learning on the historical facts to forecast the future. Prediction is required in numerous applications including fraud detection, target marketing, performance prediction, manufacturing and medical diagnosis. The Prediction process can be thought of two phases - Training and Testing.

During the data pre-processing, raw data is extracted, cleansed, integrated, transformed, reduced, discretized and then loaded as structured data. The feature selection process distilled the important and relevant features. In the training phase, the learning algorithms are applied on the relevant data to build the model. The learning algorithms are chosen based on the nature of problem. A best model is developed at the end of the training phase. The model may be a simple linear equation or it may be complex neural network which purely depends on the problem nature. In the testing phase, the best model derived from the training phase is applied on the relevant data to derive the predicted outcome. The developed model is evaluated with the performance metrics like accuracy, precision, recall, AUC, F-measure, etc.

The prediction algorithms involved in the prediction process can be categorized on their similarity. Fig. 1 shows up the seven types of prediction algorithms and few novel prediction algorithms under it.

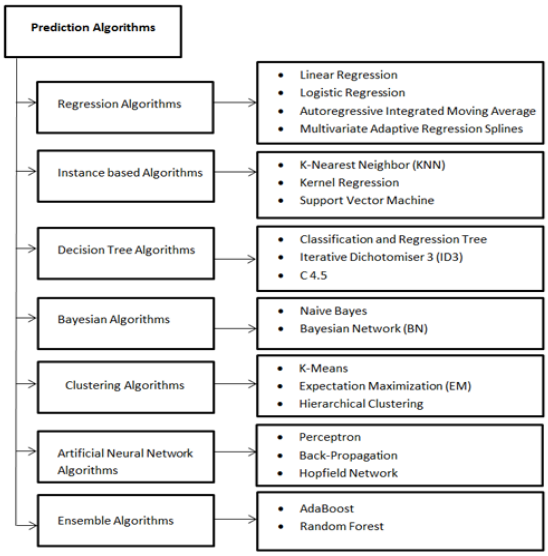

Fig. 1 Types of Prediction Algorithms

A. Regression Algorithms

Regression algorithms apply the statistical power to model the relationship between variables.

1. Linear Regression: Linear regression is a statistical method to study the relationship between two continuous type variables. The Simple linear regression assumes a linear relationship between an input variable (x) and a single output variable (y). Multiple linear regressions assume a linear relationship between more than one input variable and a single output variable. [1] To overcome the performance issue of linear regression under the parameter vector which is sparse, a new technique is adapted by using a Bayesian framework. Linear regression performs well with simple problems. But the solutions for complex problems are not error free. [2] A novel method, Gaussian mixture continuous-density hidden Markov model parameters is used to obtain minimum classification error.

2. Logistic Regression: Logistic regression is a powerful statistical way to model binomial outcomes with one or more explanatory variables. It measures the relationship between the categorical dependent variable and one or more independent variables by estimating probabilities using a logistic function, which is the cumulative logistic distribution. Variable selection is one of the important problems in logistic regression. To overcome this, a new technique logistic elastic net [3] is proposed. The technique has grouping effect which means that the strongly correlated predictors tend to be in or out of the model together. The logistic elastic net is particularly useful when the number of predictors (p) is much bigger than the number of observations. Another limitation of logistic regression is, it suffers from over fitting. It often fails in considering the structural information. A novel matrix-based logistic regression [4] is proposed to overcome this.

3. Autoregressive Integrated Moving Average (ARIMA): ARIMA adapts a "stochastic" modelling approach that can be used to calculate the probability of a future value lying between two specified limits. In the Autoregressive (AR) process, the series current values depend on its own previous values. In the Moving Average (MA) process, the current deviation from mean depends on previous deviations. ARIMA integrate both the Autoregressive and Moving Average process into one. [5] ARIMA is integrated with the complex neurofuzzy system and the new computing model is able to predict a complex-valued output, of which the real and imaginary parts can be used for two different functional mappings. ARIMA is adapted to solve problems in different application fields. [6] ARIMA with the modified genetic algorithm model was able to forecast real-time dynamic rain-induced attenuation.

4. Multivariate Adaptive Regression Splines (MARS): MARS is a non-parametric regression technique which can be seen as an extension of linear models. It automatically models non-linearity and interactions between variables. The MARS approach to regression modelling effectively uncovers important data patterns from difficult relationships. [7] A novel combination of hierarchical clustering and MARS utilizes both the hierarchical clustering and MARS approaches to solve problems. [8] The combination of MARS and response surface methodology was proved effective in addressing highly nonlinear high-dimensional problems.

B. Instance Based Algorithms

Instance based algorithms which are also called as memory based learning algorithms analyses the new instances with the stored instances and makes a prediction.

1. K-Nearest Neighbor (KNN): The KNN algorithm doesn't split the dataset as training set and test set. It uses the entire dataset as the training set. When an outcome is required for a new data instance, the KNN algorithm scans the entire dataset to find the k-nearest instances that match with the new instance or search for the k number of instances which is most similar. For a regression problem it outputs the mean of the outcomes and for a classification problem it outputs the mode. User has to define the value of k. Different distance metrics will be used to calculate the similarity between the instances. Popular distance measures used in KNN problems are Euclidean distance, Hamming distance, etc. K-NN is to prone to bias when there is little class separation and the sample sizes are unequal.[9] To minimize the classification bias, the bias is examined for the binary situation. The theory of fuzzy sets is introduced along with KNN [10] and the experiment results of the model are effective.

2. Kernel Regression: Kernel regression is similar to K-NN algorithm. In KNN, all the neighbors receive equal weight and the number of neighbors must be chosen globally. In Kernel regression, all neighbors are used as in K-NN but with different weights. The closer neighbors receive higher score. The weighting function is called a kernel which measures similarity (as opposed to distance). While handling the noisy data, estimating the order and bandwidth parameters is difficult. To solve the problem, a new method is proposed [11] with a weighted MSE risk and optimized with respect to the order and bandwidth parameters. A constrained kernel regression model is proposed [12] to solve One Dimensional problems.

3. Support Vector Machine: A Support Vector Machine (SVM) is a discriminative classifier. It transforms the input space into a new space (F feature space) using a nonlinear mapping. Fuzzy support vector machine is accepted as a significant addition in the SVM family. It absorbs the outliers strongly depends on how well the training samples are assigned fuzzy Membership Values. Usually the membership functions used for the FSVM were custom made for applications, and the membership functions used for one application cannot be used for other applications. This is an important limitation. To overcome this limitation

[13], a differential treatment to data is recommended. The novel General Purpose Membership Functions use clustering techniques to recognize possible outliers. To handle the binary data in a better way, [14], a twin support vector machine is proposed.

C. Decision Tree Algorithms

Decision tree algorithms are one among the most widely used prediction algorithms for inductive inference over supervised data.

1. Classification and Regression Trees (CART): Classification and Regression Trees (CART) is an implementation of Decision Trees. The non-terminal nodes are the root node and the internal node. The terminal nodes are the leaf nodes. A fuzzy decision tree [15] is constructed. Fuzzification is achieved by superimposing a fuzzy structure over the skeleton of a CART decision tree. The algorithm has a significantly lower error rate. The CART used to treat all the different types of variables equally. To overcome this, a novel framework [16] is developed. It uses leave one out cross validation to select the splitting variable and then it performs a regular split of CART for the selected variable.

2. Iterative Dichotomiser 3 (ID3): ID3 attempts to generate a smallest possible decision tree from a given data set by employing a top down, greedy search. It split the data set into subsets by using only those attribute whose entropy is minimum or information gain is maximum. [17] A real time fuzzy rule based ID3 type decision tree with less redundancy is proposed to solve problems. [18] An ID3 decision tree with hidden layers allows interpretation of the knowledge embedded in the generated connections and weights.

3. C 4.5: C 4.5 is an extension of ID3 algorithm. It is robust in the presence of noisy data and avoids overfitting. It deals with continuous attributes and convert trees to rules. A novel decision tree algorithm NeC4.5 [19] is obtained by combining the features of C 4.5 and Neural network. [20] A hybrid of the traditional recursive implementation of C4.5 and Rainforest have shown effective results.

D. Bayesian Algorithms

Bayesian algorithms apply Bayes theorem to build the prediction model and solve problems.

1. Naive Bayes: Naive Bayes is based on Bayes' theorem with strong (naive) independence assumptions between the features. The featured image is the equation - with P (A|B) is posterior probability, P (B|A) is likelihood, P (A) is class prior probability and P (B) is predictor prior probability. There is a tight dependency-related bound on the difference between the Bayes error and NB error in the case of two binary features and two classes. To overcome this limitation, [21] a feature dependency and an error-Dependency relationships is proposed for the Naïve Bayes Classifier with Binary Features. A novel [22] hidden naive Bayes (HNB) is proposed with a hidden parent for each attribute.

2. Bayesian Network (BN): Bayesian networks belong to the family of probabilistic graphical model (GMs). It is used to represent knowledge about an uncertain domain. It is a NP-hard problem. BN is naturally a generative model, which is not necessarily discriminative. Improving the discriminative power of BN models for continuous variables, [23] a framework (GBN) was developed. A novel framework for data fusion called [24] a deformable Bayesian network (DFBN) is proposed. It have good structural flexibility.

E. Clustering Algorithms

Clustering algorithms adapts suitable methods based on the inherent structure of the data and organize the data into a group of maximum commonality.

1. K-Means: K-means partitions the dataset into K clusters and initialize a centre for each cluster. It assigns the closest cluster to each data point. It reset the centre position of each cluster to the mean of all data points belonging to that cluster. It reassigns the closest cluster to each data point until the assignments do not change from one iteration to the next. It has high expensive computation and memory costs. To overcome this, [25] a discriminative hierarchical K-means tree framework is proposed. A better tag clustering algorithm [26] is proposed for greater application of social tags in personalized search based on K-Means.

2. Expectation Maximization (EM): The EM algorithm has been one of the most important and widely used estimation methods in estimating parameters of distributions in the presence of incomplete information. It estimates the maximum likelihood in models with hidden variables. It iteratively computes the expectations of terms in the log likelihood function under the current posterior and then solving for the maximum likelihood parameters. The [27] expectation conditional maximization algorithm is developed by replacing the M-Step of the algorithm by a series of conditional maximization steps. A novel algorithm [28] is proposed to overcome the source localization problems based on EM algorithm.

3. Hierarchical Clustering (HC): HC builds a hierarchy of clusters. The results of HC are usually presented in a dendrogram. [29] To address large-scale data, various approximate and parallel algorithms have been proposed to reduce the computational cost of HC. So [30] a new linkage method based on HC assures the exactness of maintaining the quadratic nature in time complexity

F. Artificial Neural Network Algorithms

Artificial neural network algorithms are inspired by the structure and function of biological neural networks. They are able to build models to solve complex problems.

1. Perceptron: The perceptron is commonly used as a Simplified model for the discrimination and learning capability of a biological neuron. It is a linear classifier. It makes predictions based on a linear predictor function combining a set of weights with the feature vector. Arithmetic Perceptrons [31] perform more complicated arithmetic operations like subtraction. [32] Multilayer perceptions offer great computational effort by means of a sensitivity correction factor in the formulation.

2. Back-Propagation: Back propagation is a supervised learning algorithm for training Multilayer Perceptron. It looks for the minimum value of the error function in weight space using a technique called the delta rule or gradient descent. The weights that minimize the error function is then considered to be a solution to the learning problem. [33] The evolutionary strategy (ES) algorithm is proposed and the theoretical results are studied using the MLP-based decision feedback equalizer scenarios. Resilient back propagation (RProp) algorithm is an enhanced version of BP which is developed for static networks.

3. Hopfield Network: Hopfield network is a network with associative memory. It has only one layer with each neuron connected to every other neuron where all the neurons act as input and output. Hopfield neural network (HNN) machine learning has already proven successful in solving the linear mixture model. A generalized Hopfield network [35] for solving general constrained convex optimization problems was proposed. A solution [36] to the delayed reaction-diffusion Hopfield neural networks driven by infinite dimensional Wiener processes was proposed and analysed.

G. Ensemble Algorithms

Ensemble algorithms are composed of multiple models that are independently trained and their individual predictions are combined to the overall prediction.

1. AdaBoost: The AdaBoost algorithm is a popular ensemble method that combines several weak learners to boost generalization performance. Based on accuracy of previous training, it iteratively retrains the model. The algorithm suffer from the limitation that the threshold value must be manually set. To overcome this, [37] a generic framework is proposed with structural optimization techniques. It have a robust threshold mechanism. To learn and extract binary fingerprints [38] a regularized Adaboost is developed.

2. Random Forest: Random forest algorithm creates a set of decision trees from randomly selected subset of training set. It then aggregates the votes from different decision trees to decide the final class of the test object. [39] A Safe Bayesian random forest is proposed which is proven accurate and performs in speed. [40] A modified optimal fast discrete Stockwell transform with random forests classifier-based PQ detection framework has been proposed. The classification results of the proposed PQ framework are found to be very accurate and quite insensitive to noise.

III. ADVANTAGES AND DISADVANTAGES OF PREDICTION ALGORITHMS

Table I lists the key advantages and disadvantages of prediction algorithm in each category.

TABLE I ADVANTAGES & DISADVANTAGES OF PREDICTION ALGORITHMS

S.No.	Algorithm	Advantages	Disadvantages
Regression algorithms			
1	Linear Regression	Easy to understand and implement	It is limited to linear relationships It is very sensitive to outliers It assumes that the data are independent
2	Logistic Regression	Easy to understand and implement	It requires each data point to be independent of all other data points Prone to overfitting
3	Autoregressive Integrated Moving Average (ARIMA)	Adapts statistical approach strictly Increases the forecasting accuracy in minimal parameters	High cost Unstable with respect to changes in observation and changes in model specification It works for short run
4	Multivariate Adaptive Regression Splines	Works well even with a large number of predictor variables Automatically detects interactions between variables Efficient and fast Robust to outliers	Difficult to understand Prone to overfitting Model is vulnerable to missing data
Instance based algorithms			
5	K-Nearest Neighbor (KNN)	Robust to noisy data Effective if the training data is large	Need to determine the value of parameter K Have to determine the type of distance Computation cost is high

S.No.	Algorithm	Advantages	Disadvantages
6	Kernel Regression	It is nonparametric	Prone to bias if the independent variables are not uniformly distributed
7	Support Vector	Prediction accuracy is generally high Fast evaluation of the learned target function	Long training time Difficult to understand the learned function (weights)
Decision tree algorithms			
8	Classification and Regression Trees (CART)	Easy to understand and implement Easy to handle categorical features	Prone to overfitting Possible issues with diagonal decision boundaries
9	Iterative Dichotomiser 3 (ID3)	Understandable prediction rules are created from training data Builds fastest and short tree	Prone to overfitting Predicting a continuous data may be computationally expensive
10	C 4.5	Can be used for both categorical and continuous values Deals with noise	Small variation in data can lead to different decision tree Doesn't work very well with small training data
Bayesian algorithms			
11	Naive Bayes	Easy to implement Requires small amount of data to train the model	Loss of accuracy due to class conditional independence assumption
12	Bayesian Network	Have rigorous probabilistic foundation Reasoning process is semi-transparent	Computationally expensive Performance is poor on high dimensional data
Clustering algorithms			
13	K-Means	Computationally fast Easy to implement Works well with high dimensions	Difficult to predict the number of cluster (K-value) Order of the data has an impact on final result Sensitive to scale
14	Expectation Maximization (EM)	Easy to implement	Slow linear convergence Initial estimation should be carefully chosen
15	Hierarchical Clustering	Easy to implement Easy to decide the no. of clusters by looking at the dendrogram	Not suitable for large dataset Very sensitive to outliers
Artificial neural network algorithms			
16	Perceptron	Guaranteed convergence when linearly separable Very fast on test data	Thrashes when not linearly separable
17	Back-Propagation	Computing time is reduced if the weights chosen are small at the beginning Batch update of weights provides a smoothing effect on weight correction terms	Outputs can be fuzzy or non-numeric Prone to local minima resulting in sub-optimal solutions
18	Hopfield Network	Massive parallel computation	Computational efficiency is not consistent
Ensemble algorithms			
19	AdaBoost	Easy to implement Not prone to overfitting	Sensitive to noisy data and outliers
20	Random Forest	Reduction in overfitting Less variance	More complex Hard to visualize

IV. CONCLUSION

The spread of information and speed of its delivery induce so many changes and uncertainty around the world. To par with the changes induced around the world, many advanced tools, environment and algorithms are accelerating in the data mining and machine learning fields. Prediction algorithms are a better counter-part to beat up the uncertainty. The Prediction literature growth discussed in this paper is a straightforward evidence of the prediction field growth over the decade. The adoption of prediction algorithms has grown exponential across the industries. The increases in prediction capabilities hold the potential of goodness to the society. But the challenges are not over. The data management process requires improvements. The drawbacks in the existing prediction algorithms need to be rectified. New algorithms with potential capabilities are required to handle big data. New approaches are required in handling data like privacy, security, safety, etc.

** This study was presented in the International Conference on Information, Communication & Digital Security (21 - 22 March, 2018), Alagappa University, Karaikudi, TamilNadu.

REFERENCES

[1] E. G. Larsson and Y. Selén, "Linear regression with a sparse parameter vector", *IEEE Transactions on Signal Processing*, Vol. 55, No. 2, pp. 451–460, 2007.

[2] J. Wu and Q. Huo, "A study of minimum classification error (MCE) linear regression for supervised adaptation of MCE-trained continuous-density hidden Markov models", *IEEE Transactions on Audio, Speech, and Language Processing*, Vol. 15, No. 2, pp. 478–488, 2007.

[3] J. Zhang and J. Jiang, "Rank-optimized logistic matrix regression toward improved matrix data classification", *Neural computation*, Vol. 30, No. 2, pp. 505–525, 2018.

[4] C. Li and T.-W. Chiang, "Complex neurofuzzy ARIMA forecasting—a new approach using complex fuzzy sets", *IEEE Transactions on Fuzzy Systems*, Vol. 21, No. 3, pp. 567–584, 2013.

[5] S. Gong, Y. Gao, H. Shi, and G. Zhao, "A practical MGA-ARIMA model for forecasting real-time dynamic rain-induced attenuation", *Radio Science*, Vol. 48, No. 3, pp. 208–225, 2013.

[6] C. Fu, "Business Valuation Based on Intellectual Capital: A Hierarchical Clustering-MARS Approach", in *Management and Service Science (MASS), 2011 International Conference on*, 2011, pp. 1–6.

[7] Crino S, Brown DE. "Global optimization with multivariate adaptive regression splines". *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics).* Vol. 37, No.2, pp.333-340, Apr. 2007.

[8] J. E. Goin, "Classification bias of the k-nearest neighbor algorithm", *IEEE transactions on pattern analysis and machine intelligence*, No. 3, pp. 379–381, 1984.

[9] J. M. Keller, M. R. Gray, and J. A. Givens, "A fuzzy k-nearest neighbor algorithm", *IEEE transactions on systems, man, and cybernetics*, No. 4, pp. 580–585, 1985.

[10] S. R. Krishnan, C. S. Seelamantula, and P. Chakravarti, "Spatially Adaptive Kernel Regression Using Risk Estimation", *IEEE Signal Processing Letters*, Vol. 21, No. 4, pp. 445–448, 2014.

[11] H. Zhang and Z. Jiang, "Constrained kernel regression for pose estimation", *Electronics Letters*, Vol. 50, No. 2, pp. 77–79, 2014.

[12] R. K. Sevakula and N. K. Verma, "Compounding General Purpose Membership Functions for Fuzzy Support Vector Machine Under Noisy Environment", *IEEE Transactions on Fuzzy Systems*, Vol. 25, No. 6, pp. 1446–1459, 2017.

[13] X. Ma, Q. Ye, and H. Yan, "L2P-Norm Distance Twin Support Vector Machine", *IEEE Access*, Vol. 5, pp. 23473–23483, 2017.

[14] A. Suárez and J. F. Lutsko, "Globally optimal fuzzy decision trees for classification and regression", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 21, No. 12, pp. 1297–1311, 1999.

[15] A. Painsky and S. Rosset, "Cross-validated variable selection in tree-based methods improves predictive performance", *IEEE transactions on pattern analysis and machine intelligence*, Vol. 39, No. 11, pp. 2142–2153, 2017.

[16] N. R. Pal and S. Chakraborty, "Fuzzy rule extraction from ID3-type decision trees for real data", *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, Vol. 31, No. 5, pp. 745–754, 2001.

[17] K. J. Cios and N. Liu, "A machine learning method for generation of a neural network architecture: A continuous ID3 algorithm", *IEEE Transactions on Neural Networks*, Vol. 3, No. 2, pp. 280–291, 1992.

[18] Z.-H. Zhou and Y. Jiang, "NeC4. 5: neural ensemble based C4. 5", *IEEE Transactions on Knowledge & Data Engineering*, No. 6, pp. 770–773, 2004.

[19] Y. Yang and W. Chen, "Taiga: performance optimization of the C4. 5 decision tree construction algorithm", *Tsinghua Science and Technology*, Vol. 21, No. 4, pp. 415–425, 2016.

[20] L. Kuncheva and Z. Hoare, "Error-dependency relationships for the Naïve Bayes classifier with binary features", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 30, No. 4, pp. 735–740, 2008.

[21] L. Jiang, H. Zhang, and Z. Cai, "A novel Bayes model: Hidden naive Bayes", *IEEE Transactions on knowledge and data engineering*, Vol. 21, No. 10, pp. 1361–1371, 2009.

[22] L. Zhou, L. Wang, L. Liu, P. Ogunbona, and D. Shen, "Learning discriminative Bayesian networks from high-dimensional continuous neuroimaging data", *IEEE transactions on pattern analysis and machine intelligence*, Vol. 38, No. 11, p. 2269, 2016.

[23] K. Kampa, E. Hasanbelliu, J. T. Cobb, J. C. Principe, and K. C. Slatton, "Deformable Bayesian network: A robust framework for underwater sensor fusion", *IEEE Journal of Oceanic Engineering*, Vol. 37, No. 2, pp. 166–184, 2012.

[24] S. Chen, X. Yang, and Y. Tian, "Discriminative hierarchical K-means tree for large-scale image classification", *IEEE transactions on neural networks and learning systems*, Vol. 26, No. 9, pp. 2200–2205, 2015.

[25] J. Yang and J. Wang, "Tag clustering algorithm LMMSK: improved K-means algorithm based on latent semantic analysis", *Journal of Systems Engineering and Electronics*, Vol. 28, No. 2, pp. 374–384, 2017.

[26] P. Govindasamy and R. Dillibabu, "Maximum likelihood estimation and optimisation of parameters of software reliability models using evolutionary optimisation techniques", 2018.

[27] L. Lu, H.-C. Wu, K. Yan, and S. S. Iyengar, "Robust expectation-maximization algorithm for multiple wideband acoustic source localization in the presence of nonuniform noise variances", *IEEE Sensors Journal*, Vol. 11, No. 3, pp. 536–544, 2011.

[28] S. Rajasekaran, "Efficient parallel hierarchical clustering algorithms", *IEEE transactions on parallel and distributed systems*, Vol. 16, No. 6, pp. 497–502, 2005.

[29] Y. Jeon, J. Yoo, J. Lee, and S. Yoon, "Nc-link: A new linkage method for efficient hierarchical clustering of large-scale data", *IEEE Access*, Vol. 5, pp. 5594–5608, 2017.

[30] S. A. Cannas, "Arithmetic perceptrons", *Neural Computation*, Vol. 7, No. 1, pp. 173–181, 1995.

[31] A. Navia-Vazquez and A. R. Figueiras-Vidal, "Efficient block training of multilayer perceptrons", *Neural computation*, Vol. 12, No. 6, pp. 1429–1447, 2000.

[32] S. Siu, S.-S. Yang, C.-M. Lee, and C.-L. Ho, "Improving the back-propagation algorithm using evolutionary strategy", *IEEE Transactions on Circuits and Systems II: Express Briefs*, Vol. 54, No. 2, pp. 171–175, 2007.

[33] P. A. Mastorocostas, "Resilient back propagation learning algorithm for recurrent fuzzy neural networks", *Electronics Letters*, Vol. 40, No. 1, pp. 57–58, 2004.

[34] C. Li, X. Yu, T. Huang, G. Chen, and X. He, "A generalized Hopfield network for nonsmooth constrained convex optimization: Lie derivative approach", *IEEE transactions on neural networks and learning systems*, Vol. 27, No. 2, pp. 308–321, 2016.

[35] X. Liang, L. Wang, Y. Wang, and R. Wang, "Dynamical behavior of delayed reaction–diffusion Hopfield neural networks driven by infinite dimensional Wiener processes", *IEEE transactions on neural networks and learning systems*, Vol. 27, No. 9, pp. 1816–1826, 2016.

[36] P.-B. Zhang and Z.-X. Yang, "A novel AdaBoost framework with robust threshold and structural optimization", *IEEE transactions on cybernetics*, Vol. 48, No. 1, pp. 64–76, 2018.

[37] H. Yu and P. Moulin, "Regularized Adaboost Learning for Identification of Time-Varying Content.", *IEEE Trans. Information Forensics and Security*, Vol. 9, No. 10, pp. 1606–1616, 2014.

[38] N. Quadrianto and Z. Ghahramani, "A very simple safe-Bayesian random forest", *IEEE transactions on pattern analysis and machine intelligence*, Vol. 37, No. 6, pp. 1297–1303, 2015.

[39] M. V. Reddy and R. Sodhi, "A Modified S-Transform and Random Forests-Based Power Quality Assessment Framework", *IEEE Transactions on Instrumentation and Measurement*, Vol. 67, No. 1, pp. 78–89, 2018.

[40] M. H. Sherif, R. J. Gregor, and J. Lyman, "Effects of load on myoelectric signals: the ARIMA representation", *IEEE Transactions on Biomedical Engineering*, No. 5, pp. 411–416, 1981.

[41] D. Barber, D. Saad, and P. Sollich, "Test error fluctuations in finite linear perceptrons", *Neural computation*, Vol. 7, No. 4, pp. 809–821, 1995.

[42] Li J, Li X, Huang B, Zhao L."Hopfield neural network approach for supervised nonlinear spectral unmixing", *IEEE Geoscience and Remote Sensing Letters*, Vol.13, No.7, pp. 1002-1006, Jul. 2016.