ELSEVIER
CrossMark
Available online at www.sciencedirect.com
ScienceDirect
Procedia Computer Science 89 (2016) 506 – 513
Procedia Computer Science

Twelfth International Multi-Conference on Information Processing-2016 (IMCIP-2016)

Improving the Accuracy of Fuzzy Decision Tree by Direct Back Propagation with Adaptive Learning Rate and Momentum Factor for User Localization

Swathi Jamjala Narayanan[a,*], Rajen B. Bhatt[b] and Boominathan Perumal[a]

[a] *School of Computing Science and Engineering, VIT University, Vellore 14, India*
[b] *Robert Bosch Research and Technology Center, Pittsburgh, USA*

Abstract

Most prevailing availability of wireless networks has elevated an interest in developing a smart indoor environment by utilizing the hand held devices of the users. The user localization helps in automating the activities like automating switch on/off of the room lights, air conditioning etc., which makes the environment smart. Here, we consider locating the users as a pattern classification problem and use Fuzzy decision tree (FDT) as a knowledge discovery method to locate the users based on the wireless signal strength observed by their handheld devices. To increase the FDT accuracy and to achieve faster convergence, we came up with a novel strategy named Improved Neuro Fuzzy Decision Tree with an adaptive learning rate and momentum factor to optimize the parameters of FDT. The proposed approach can be used for any classification problem. From the results obtained, we observe that our proposed algorithm achieves better convergence and accuracy.

Peer-review under responsibility of organizing committee of the Organizing Committee of IMCIP-2016

Keywords: Adaptive Learning Rate; Fuzzy Decision Tree; Momentum Factor; User Localization; Wireless Signal Strength.

1. Introduction

The significant idea of user localization is to locate the users having handheld devices or computers with the help of wireless signal strength accessed by their devices in indoor environments. For instance, a simple wireless device can hold several wireless signals when the device is within the region of several Wi-Fi-Zones. Based on the position of Wi-Fi routers or the hotspots from the wireless devices, the observed Wi-Fi signal strength of the devices may differ. Due to this characteristic, these signal strengths can be used to locate the users in indoor environments.The existing literature for user localization has used several machine learning approaches to locate users. Alonso *et al.*[1] used a fuzzy rule base to locate robots even in small scale variations with a robust low level controller. Yu *et al.*[2] proposed methods using fingerprinting localization which is based on neural networks and ultra-bounded signals and showed significant performance compared to other techniques. Nguyen[3] proposed a scalable WiFi based localization algorithm which handles different mobile devices during deployment phase. Garcia *et al.*[4] proposed an online incremental learning model using fuzzy concepts to deal with the uncertainties and varying conditions. Pei *et al.*[5] proposed least square

*Corresponding author. Tel.: +91 -9790084851.
E-mail address: swathi.jns@gmail.com

Peer-review under responsibility of organizing committee of the Organizing Committee of IMCIP-2016
doi:10.1016/j.procs.2016.06.104

support vector machine technique and showed better accuracy compared to other techniques. Bin Abdullah *et al.*[6] proposed a novel method to reduce signal noise during data collection and later used neural network to estimate the user location. Galvan-Tejeda *et al.*[7] evaluated random forest, nearest centroid, K-nearest neighbours and artificial neural networks classifiers as cost functions for indoor user localization and showed that nearest centroid method outperforms the other classifiers in predicting the user location. Gaxiola-Pacheco and Licea[8] proposed Type-2 fuzzy inference system to predict the user zone location. YuFeng *et al.*[9] proposed SVM with bilinear median interpolation method and compared with Radial basis function neural network for indoor user localization. Zou *et al.*[10] proposed online sequential extreme learning machine algorithm and obtained good accuracy for user localization problem. Cho[11] proposed machine learning methods to recognize location both in indoor as well as in outdoor using smart phone logs.

Considering the user localization problem as pattern classification problem, these exists several machine learning methods for classification. But most of the techniques lack readability and in those cases, decision trees play a major role in developing rules which are comprehensive and interpretable. To handle the uncertainty in the data, fuzzy decision trees[12] (FDT) emerged and are used widely. FDT generates readable and understandable rules for pattern classification. The working nature of the FDT with respect to its performance depends on the primary fuzzy clustering of the input attributes. Apart from input clustering, the other two parameters that affect performance are alpha-cut standard parameter and leaf selection threshold. After these parameters are decided, we apply Fuzzy ID3 for inducing FDT. Generally, as local optimal decisions are taken during FDT induction, there is a chance to get lesser accuracy. To overcome this issue, Rajen and Gopal[13] have applied direct back propagation strategy on the FDT. As an extension of that work to improve the convergence rate, here we propose an improved Neuro-Fuzzy Decision Tree with an adaptive learning rate and momentum factor to increase the convergence of the FDT optimization algorithm and to improve the accuracy as well.

2. Pattern Classification Problem using Fuzzy Concepts

Assuming there are n training patterns, where each of the training pattern X^i consists of p input variables, and the training pattern would get classified to one of the q categories present in the decision variable y. Each input variable x_j is partitioned into c_j fuzzy sets. $\mu_{F_{jk}}(x_j^i)$ denotes the membership degree for the values of the attribute x_j to the fuzzy set F_{jk}. The class is indexed by the variable $l(1 \leq l \leq q)$ and it is considered to be as a crisp set. The fuzzy singleton values of the decision attribute (y^i) related to the l^{th} class is given as follows:

$$\mu_l(y^i) = \begin{Bmatrix} 1; & \text{if } y^i \text{ belongs to } l^{\text{th}} \text{ category} \\ 0; & \text{otherwise} \end{Bmatrix} \tag{1}$$

A normal fuzzy rule is written as

$$\text{if}(x_1 \text{ is } F_{1k}) \quad \text{and} \quad (x_2 \text{ is } F_{2k}) \ldots (x_p \text{ is } F_{pk}); \quad \text{then} \quad (y \text{ is } l) \tag{2}$$

3. Fuzzy ID3

For the generation of FDT, fuzzy classification entropy measure is used to determine the best node for inducing FDT. For each of the fuzzy partitions $\{F_{jk} | j = 1, \ldots, p; k = 1, \ldots, c_j\}$ of the attribute x_j, the certainty factor[14] of the l^{th} class is determined using Equation (3)

$$\beta_{jk}^l = \frac{\sum_{i=1}^n \min\{\mu_{F_{jk}}(x_j^i), \mu_l(y^i)\}}{\sum_{i=1}^n \mu_{F_{jk}}(x_j^i)}; \quad 0 \leq \beta_{jk}^l \leq 1 \tag{3}$$

The fuzzy classification entropy[14] measure of the fuzzy F_{jk} partition is given in Equation (4)

$$\text{Entropy}_{jk} = -\sum_{l=1}^{q} \beta_{jk}^l \times \log_2(\beta_{jk}^l) \tag{4}$$

The average fuzzy entropy[14] of the attribute x_j is determined using Equation (5)

$$E_j = \sum_{k=1}^{c_j} w_{jk} \times \text{Entropy}_{jk} \tag{5}$$

where

$$w_{jk} = \frac{\sum_{i=1}^{n} \mu_{F_{jk}}(x_j^i)}{\sum_{k=1}^{c_j} \left(\sum_{i=1}^{n} \mu_{F_{jk}}(x_j^i) \right)} \tag{6}$$

The process of FDT induction[14] is as given below:

The Prerequisites for inducing FDT are fuzzy clusters, leaf threshold and the node selection criterion.

Process of FDT induction:

While there are several candidate nodes
DO
Select one candidate node using a search strategy like fuzzy classification entropy,
Generate its child-nodes,
Then the child-nodes meeting the leaf selection threshold have to be assigned as leaf-nodes, otherwise the process continues by considering remaining child-nodes as new candidate nodes and the procedure is repeated until the stopping criterion is met.
End

The α-cut is used to reduce the fuzziness and it is determined as given in Equation (7)

$$\mu_{AA_\alpha}(a) = \begin{cases} \mu_{AA}(a); & \mu_{AA}(a) \geq \alpha \\ 0; & \mu_{AA}(a) < \alpha \end{cases} \tag{7}$$

The FDT of the user localization dataset for direct backpropagation using our proposed approach is shown in Fig. 1.

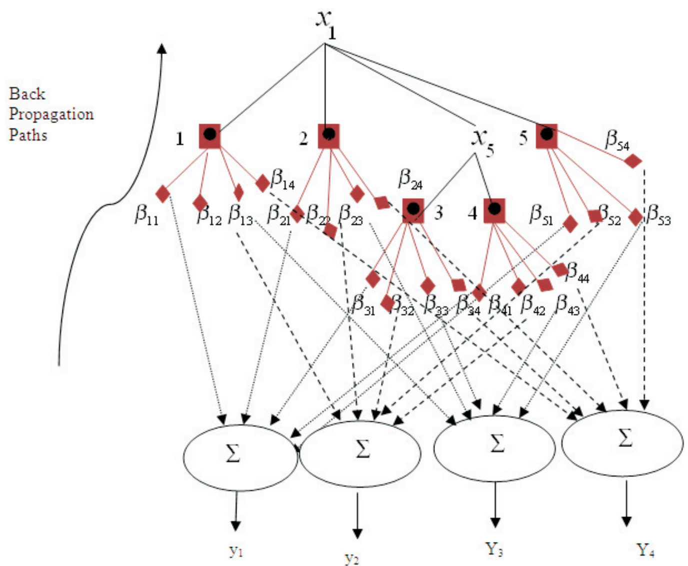

Fig. 1. FDT of User Location Dataset for Direct Back Propagation.

In the given FDT, there are five paths (m) specifying the fuzzy rules. The firing strength of the rules for the i^{th} test pattern and l^{th} class is obtained using Equation (8)

$$y_l^i = \sum_{m=1}^{M} \mu_{\text{path}_m}^i \times \beta_{ml} \tag{8}$$

β_{ml} denotes the certainty factor and it is determined using Equation (3). The aggregated certainty factor from all the leaf nodes with respect to class l is assigned to y_l^i. The unique class for a pattern p is selected using maximum membership degree

$$l^i = \max_{l=1,\ldots,q} \{y_l^i\} \tag{9}$$

The Fuzzification of the input attributes are represented as Gaussian membership function, which is a differentiable function. The membership degree of the i^{th} pattern with respect to path$_m$ is determined using Equation (10)

$$\mu^i \text{path}_m = \Pi_j \mu F_{jm}(x_j^i) = \prod_j e^{\left(-\frac{(x_j^i - c_{jm})^2}{2\sigma_{jm}^2}\right)} \tag{10}$$

where the center is denoted as c_{jm} and the width of the Gaussian membership function is denoted as σ_{jm}.

4. Improved N-FDT with Adaptive Learning Rate And Momentum Factor

The non-linear optimization problem considered for minimizing error E of FDT is given in Equation (11)

$$E = \frac{1}{2n} \sum_{l=1}^{q} \sum_{i=1}^{n} (d_l^i - y_l^i)^2 \tag{11}$$

E denotes the error of the FDT, n denotes the total number of training patterns, d_l^i denotes the anticipated class for the pattern i and y_l^i denotes the prediction certainty of the pattern i. The parameters $c_{jm}, \sigma_{jm}, \beta_{ml}$ of the FDT are to be optimized for reducing the error E.

Given a current iteration value β_{ml}^t, the new value for β_{ml}^{t+1} with momentum term[15] is obtained as

$$\beta_{ml}^{t+1} = \beta_{ml}^t - \eta^t \times \frac{\partial E}{\partial \beta_{ml}^t} + \rho(-\eta^{t-1}) \times \left[\frac{\partial E}{\partial \beta_{ml}^{t-1}}\right] \tag{12}$$

The adaptive learning rate[16] η^t for the iteration t is given by

$$\eta^t = \frac{\langle \delta^{t-1}, \delta^{t-1} \rangle}{\langle \delta^{t-1}, \psi^{t-1} \rangle} \tag{13}$$

where $\langle \bullet, \bullet \rangle$ denotes the standard inner product and

$$\delta^{t-1} = \beta_{ml}^t - \beta_{ml}^{t-1} \tag{14}$$

$$\psi^{t-1} = \frac{\partial E}{\partial \beta_{ml}^t} - \frac{\partial E}{\partial \beta_{ml}^{t-1}} \tag{15}$$

To overcome the problem of the adaptive learning rate becoming large, such that it overshoots the minimum or possibly diverge, a parameter μ called maximum growth factor[16] is introduced.

Where, the new learning rate λ^t is obtained as

$$\lambda^t = \begin{cases} \eta^t, & \left|\frac{\eta^t}{\eta^{t-1}}\right| \leq \mu \\ \mu\eta^{t-1}, & \text{otherwise} \end{cases} \tag{16}$$

The update rules derived for tuning FDT using adaptive learning rate and momentum factor are

$$\beta_{ml}^{t+1} = \beta_{ml}^t - \lambda^t \times \frac{\partial E}{\partial \beta_{ml}^t} + \rho(-\lambda^{t-1}) \times \left[\frac{\partial E}{\partial \beta_{ml}^{t-1}}\right] \tag{17}$$

where

$$\frac{\partial E}{\partial \beta_{ml}} = -\frac{1}{n}\sum_{i=1}^{n}(d_l^i - y_l^i)\mu_{\text{path}_m}^i \tag{18}$$

$$c_{jm}^{t+1} = c_{jm}^t - \lambda^t \frac{\partial E}{\partial c_{jm}^t} + \rho(-\lambda^{t-1})\left[\frac{\partial E}{\partial c_{jm}^{t-1}}\right] \tag{19}$$

where

$$\frac{\partial E}{\partial c_{jm}} = \frac{1}{n}\sum_{i=1}^{n}\left(\mu_{\text{path}_m}^i\left(\frac{x_j^i - c_{jm}^t}{(\sigma_{jm}^t)^2}\right) \times \sum_{l=1}^{q}\beta_{ml}^{t+1}(d_l^i - y_l^i)\right) \tag{20}$$

$$\sigma_{jm}^{t+1} = \sigma_{jm}^t - \lambda^t \frac{\partial E}{\partial \sigma_{jm}^t} + \rho(-\lambda^{t-1})\left[\frac{\partial E}{\partial \sigma_{jm}^{t-1}}\right] \tag{21}$$

where

$$\frac{\partial E}{\partial \sigma_{jm}} = \frac{1}{n}\sum_{i=1}^{n}\left(\mu_{\text{path}_m}^i\left(\frac{(x_j^i - c_{jm}^{t+1})^2}{(\sigma_{jm}^{\tau})^3}\right) \times \sum_{l=1}^{q}\beta_{ml}^{t+1}(d_l^i - y_l^i)\right) \tag{22}$$

The parameters are updated by moving in reverse from the bottom leaf nodes to the top root node. The traversal follows the order of $\beta_{ml} \rightarrow c_{jm} \rightarrow \sigma_{jm}$. In Equation (19) we have updated c_{jm} by using the updated parameter β_{ml} and for updating σ_{jm} in Equation (21), the new values of c_{jm} and β_{ml} are used. The process of parameter updating continues until the error is reduced or the maximum epochs are reached.

5. Computational Experiments

The experiment is conducted using Fuzzy ID3 algorithm[17,18]. The numerical values are clustered into specified number of clusters using Fuzzy c-means algorithm[19]. We can also use the other partitioning techniques like grid partitioning and subtractive clustering[20–23]. For the experimental purpose, we have chosen optimal number of clusters using Xie-Beni[24] cluster validity measure and passed the optimal cluster number to FCM algorithm for each of the continuous attribute. We performed tenfold cross validation to get the results of Fuzzy ID3 for the user localization problem. The dataset for this experiment is generated using a phone with Android OS. The wireless signal strengths detected by the phone are obtained from seven wireless routers. We collected signals from four locations namely conference room, kitchen, indoor sports room, and work areas. These locations are considered to be as classes or categories for the classification problem. The signals received at various locations are considered as the seven dimensional input attributes. From each of the location, we observed several signal strengths by voting the signal strengths at every one second interval. Then the same signal collection process was continued at another location and again we collected signal strengths. The training data is obtained base on the wireless signal strengths recorded at four indoor locations. After we train the classification model using training data, we can deploy the model to determine user's location in one of the four classes mentioned above. To report the results of our approach here we show the results of one fold, for the user localization dataset. The accuracy of FDT on training patterns is 87.4% and on test patterns, we obtained 86%. The graphs given in figures 2- 3 shows the mean square error and accuracy obtained for 500 epochs using NFDT[13], where we could observe that the accuracy increase occurs in slow rate. For implementing Improved N-FDT, we stored the certainties of the leaf nodes for each class label, the attribute numbers specified in the nodes while moving from bottom to top in the path and the membership values of all the variables in the path. The equations (17-22) are applied for each path of the FDT to reduce the error given in Equation (11). We tuned the FDT parameters by using the improved neuro fuzzy decision tree for 50 epochs with the adaptive learning rate and obtained classification accuracy as 94.18% for training data and 93.24% for test data. The Figs. 4–5 shows the decrease in MSE and increase in accuracy of Improved NFDT for user localization problem.

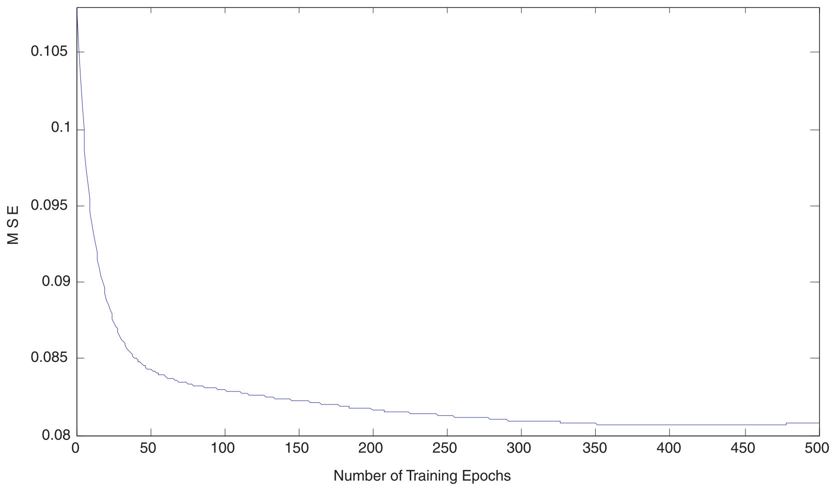

Fig. 2. Plot of MSE for User Localization Dataset using NFDT.

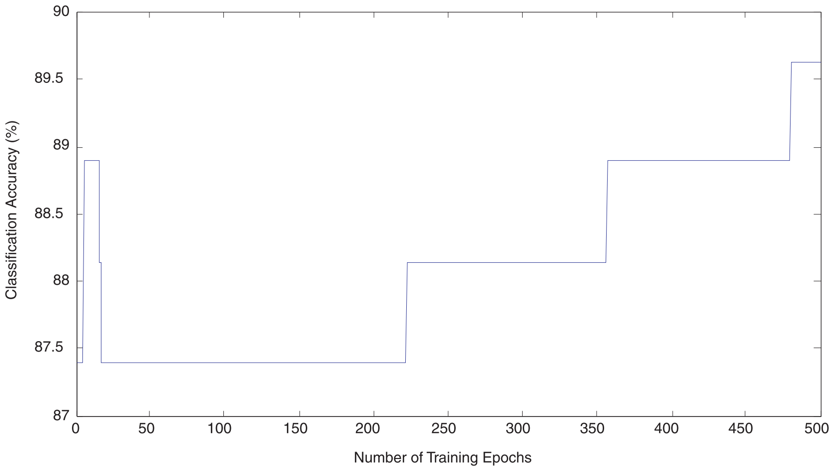

Fig. 3. Plot of Classification Accuracy for User Localization Dataset using NFDT.

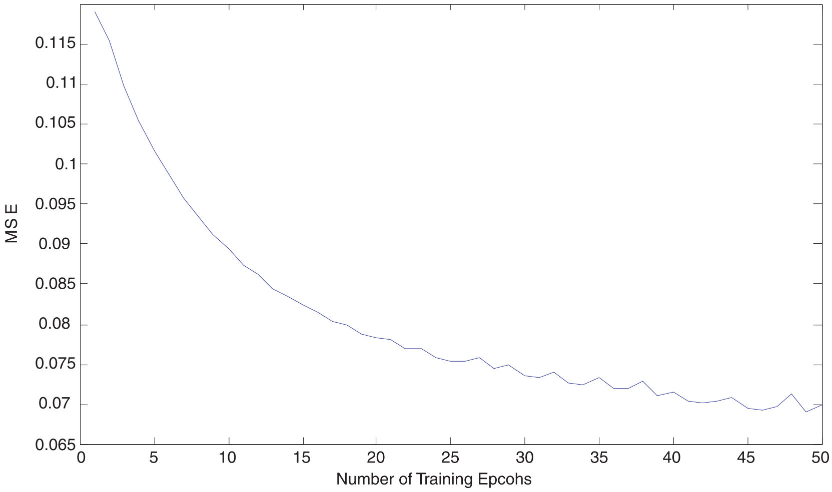

Fig. 4. Plot of MSE for User Localization Dataset using Improved NFDT.

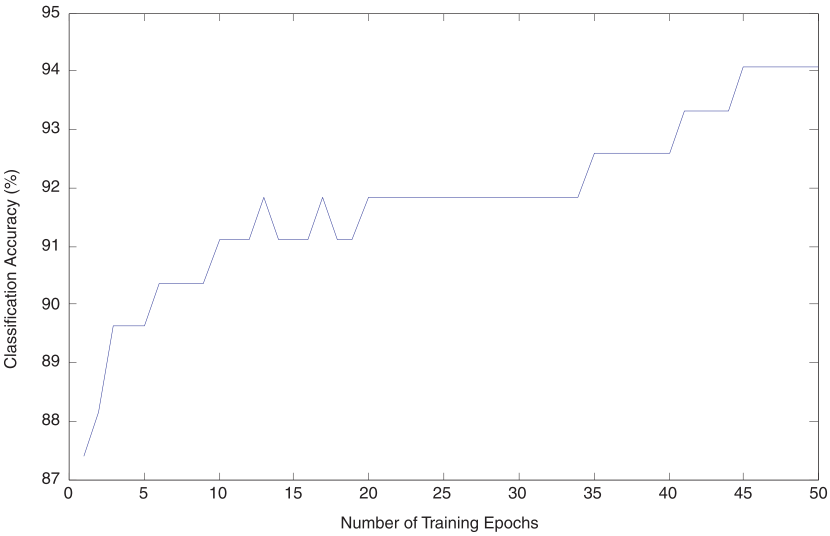

Fig. 5. Plot of Classification Accuracy for User Localization Dataset using Improved NFDT.

6. Conclusions

FDT induction algorithms are widely used to extract understandable classification rules. In this work, we consider locating the user using the wireless signal strength as a pattern classification problem. From the experimental results, we observe that the FDT generates fuzzy rules and the patterns are classified at good accuracy rate. To further improve the accuracy of FDT, an improved neuro fuzzy decision tree strategy is proposed which is based on first order gradient-descent method which includes an adaptive learning rate parameter and momentum factor added to the update equations. The proposed strategy is a general approach using which we could achieve quicker convergence and also improve the accuracy of FDT for any classification problem. The proposed strategy is fairly simple and efficient to tune the FDT parameters. In this process, the hierarchical representation of the FDT is not disturbed during training process. By updating the certainty factors, centers and sigma, the bias of initial parameters influencing the performance of FDT is reduced.

References

[1] J. M. Alonso, M. Ocana, N. Hernandez, F. Herranz, A. Llamazares, M. A. Sotelo and L. Magdalena, Enhanced WiFi Localization System Based on Soft Computing Techniques to Deal with Small-Scale Variations in Wireless Sensors, *Applied Soft Computing* vol. 11(8), pp. 4677–4691, (2011).

[2] L. Yu, M. Laaraiedh, S. Avrillon and B. Uguen, Fingerprinting Localization Based on Neural Networks and Ultra-Wideband Signals, In *Proceedings of IEEE International Symposium on Signal Processing and Information Technology*, IEEE, pp. 184–189, (2011).

[3] N. T. Nguyen, H. H. Bui, S. Venkatsh and G. West, Recognizing and Monitoring High-Level behaviors in Complex Spatial Environments, In *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, IEEE, vol. 2, p. 620, (2003).

[4] T. Garcia-Valverde, A. Garcia-Sola, A. Gomez-Skarmeta, J. Botia, H. Hagras, J. Dooley and V. Callaghan, An Adaptive Learning Fuzzy Logic System for Indoor Localisation Using WiFi in Ambient Intelligent Environments, In *Proceedings of IEEE International Conference on Fuzzy Systems*, IEEE, vol. 1–8, (2012).

[5] L. Pei, J. Liu, R. Guinness, Y. Chen, H. Kuusniemi and R. Chen, Using LS-SVM Based Motion Recognition for Smartphone Indoor Wireless Positioning, *Sensors*, vol. 12(5), pp. 6155–6175, (2012).

[6] M. F. A. Bin Abdullah, K. A. Bin Kamarul Hatta and E. Jeganathan, Adaptive WiFi Fingerprinting for Location Approximation.

[7] C.E. Galván-Tejada, J. P. García-Vázquez, E. García-Ceja, J. C. Carrasco-Jiménez and R. F. Brena, Evaluation of Four Classifiers as Cost Function for Indoor Location Systems, *Procedia Computer Science*, vol. 32, pp. 453–460, (2014).

[8] C. Gaxiola-Pacheco and G. Licea, Localization of Indoor Areas Using WiFi Signals, Type-2 Fuzzy Inference Systems and Data Mining, In *Proceedings of the World Congress on Engineering*, p. 1, (2014).

[9] YuFeng, JiangMinghua, LiangJing, QinXiao, HuMing, PengTao and HuXinrong, An Improved Indoor Localization of WiFi Based on Support Vector Machines, *International Journal of Future Generation Communication and Networking*, vol. 7(5), pp. 191–206, (2014).

[10] H. Zou, X. Lu, H. Jiang and L. Xie, A Fast and Precise Indoor Localization Algorithm Based on an Online Sequential Extreme Learning Machine, *Sensors*, vol. 15(1), pp. 1804–1824, (2015).

[11] S. B. Cho, Exploiting Machine Learning Techniques for Location Recognition and Prediction with Smartphone Logs, *Neurocomputing*, (2015).

[12] R. Weber, Fuzzy-ID3: A Class of Methods for Automatic Knowledge Acquisition, In *Proceedings of Second International Conference on Fuzzy Logic and Neural Networks*, pp. 265–268, (1992).

[13] R. B. Bhatt and M. Gopal, Neuro-Fuzzy Decision Trees, *International Journal Neural Systems*, vol. 16(1), pp. 63–78, (2006).

[14] X.-Z. Wang, D. S. Yeung and E. C. C. Tsang, A Comparative Study on Heuristic Algorithms for Generating Fuzzy Decision Trees, *IEEE Transactions on SMC-B*, vol. 31(2), pp. 215–226, (2001).

[15] Ning Qian, On the Momentum Term in Gradient Descent Learning Algorithms, *Neural Networks*, (12), pp. 145–151, (1999).

[16] V. P. Plagianakos, D. G. Sotiropoulos and M. N. Vrahatis, An Improved Backpropagation Method with Adaptive Learning Rate, *Technical Report*.

[17] Y. Yuan and M. J. Shaw, Induction of Fuzzy Decision Trees, *Fuzzy Sets and Systems*, vol. 69, pp. 125–139, (1995).

[18] R. B. Bhatt and M. Gopal, Neuro-Fuzzy Decision Trees, *International Journal Neural Systems* vol. 16(1), pp. 63–78, (2006).

[19] J. C. Bezdek, Pattern Recognition with Fuzzy Objective Function Algorithms, NY: Plenum Press; (1982).

[20] S. J. Narayanan, I. Paramasivam, R. B. Bhatt and M. A. Khalid, Study on the Approximation of Clustered Data to Parameterized Family of Fuzzy Membership Functions for the Induction of Fuzzy Decision Trees, *Cybernetics and Information Technologies*, vol. 15(2), pp. 75–96, (2015).

[21] S. J. Narayanan, R. B. Bhatt, I. Paramasivam, M. Khalid and B. K. Tripathy, Induction of Fuzzy Decision Trees and its Refinement Using Gradient Projected-Neuro-Fuzzy Decision Tree, *International Journal of Advanced Intelligence Paradigms*, vol. **6**(4), pp. 346–369, (2014).

[22] G. K. Nayak, S. J. Narayanan and I. Paramasivam, Development and Comparative Analysis of Fuzzy Inference Systems for Predicting Customer Buying behavior, *International Journal of Engineering and Technology*, vol. 5(5), pp. 4093–4108, (2013).

[23] R. B. Bhatt, S. J. Narayanan, I. Paramasivam and M. Khalid, Approximating Fuzzy Membership Functions from Clustered Raw Data. In *India Conference (INDICON)*, pp. 487–492, (2012).

[24] Wang, Weina and Yunjie Zhang, On Fuzzy Cluster Validity Indices, *Fuzzy Sets and systems*, vol. 158(19), pp. 2095–2117, (2007).