An Improved Hybrid Algorithm for Social Network Analysis

Anu Sharma†, Dr. M.K Sharma††, and Dr. R.K Dwivedi†††

†Uttrakhand Technical University, Dehradun, Uttarakhand, India
††Amrapali Institute, Haldwani, Uttarakhand, India
†††Teerthanker Mahaveer University, Moradabad, Uttar Pradesh, India

Summary
The increasing trend of on-line social networks in different fields of knowledge, social network observations recently became the center of research. This paper proposes and implements an improved hybrid technique. The data set is taken from SNA. We analyze the Performance of Classification (PoC) Accuracy in Test Data. We applied our proposed algorithm on social network dataset and found average testing error. We compared it with back-propagation and found that our hybrid average testing error is significant and less than back-propagation.
Keywords:
Neuro-fuzzy approach, Social networks, Neural network, Data mining

1. Introduction

Data mining is a method which helps in finding dissimilar patterns in the dataset. Data mining is the process of automatically discovering useful information in large repositories. Mining of information finds concealed data from substantial data bases. The recent advances in artificial intelligence and soft computing techniques have opened novel avenues for researchers to search their uses. These machine learning techniques consist of numerous intellectual computing paradigms, including "artificial neural networks (ANN), support vector machine (SVM), decision tree, neuro-fuzzy systems (NFS)", which have been successfully employ to form various real-world problems [1]. Traditional data mining uses ordered data stored in relational tables, spreadsheets, or flat files in the tabular form. Web mining and Text mining are becoming gradually more vital and popular with grow of the web and text documents. Billions of people tweet, sharing, post, and have a discussion about every day. Along with, ANFIS is a well-organized mixture of ANN and fuzzy logic for model vastly non-linear, complex, and dynamic systems [2]. Neural networks, genetic algorithms and fuzzy logic are effectual soft computing technique which can solve composite problems powerfully. Each method handles the uncertainty and ambiguity of data differently. In many applications, these technologies are combined to utilize the features of each to achieve impressive results. In this past, the fusion of neural networks and fuzzy inference systems has attracted the attention of many researchers in various scientific and engineering areas due to the growing need for intelligent decision support systems which are used to solve the real world problems. In this Paper, a new model is proposed and is suitable for the analysis of big data analysis on social networks. The proposed methodology has been validated using experiments conducted on SNAP datasets. To test the efficacy of proposed technique, we have taken datasets from the source "*www.snap.stanford.edu*". The social world is an association of connections and relations that facilitate the course and exchange of information and resources. The proposed model provides better classification accuracy than the existing systems for classifying the social network data. This paper discusses the new classification techniques.

Manuscript received November 5, 2024
Manuscript revised November 20, 2024
https://doi.org/10.22937/IJCSNS.2024.24.11.13

1.1 Hybrid Neuro-Fuzzy System

A hybrid neuro-fuzzy system is the system that uses the learning algorithm inspired by the neural network and a set of fuzzy rules. This system is initialized with a priori knowledge in the same way of fuzzy rules. The advantage of Neural Network is to learn the patterns and easy interpretation and with the fuzzy system, we can make the fuzzy rules. A node that has less or high number of links considered as weak or strong node accordingly.

1.2 Node and Friendship

This dataset consists of 'circles' (or 'friends lists'). The dataset includes node features (profiles), circles, and ego networks. Edges represent interactions between people. Nodes represent the users and edges are the mutual friendships. Edges are undirected for facebook, and directed for twitter.

1.3 Membership Function

(i) TRIMF(triangular membership function.): This function computes fuzzy membership values using a triangular membership function.

$$trimf(x,a,b,c) = \begin{Bmatrix} if\ x \le a, & 0 \\ if\ a \le x \le b, & \dfrac{x-a}{b-a} \\ if\ b \le x \le c, & \dfrac{c-x}{c-b} \\ if\ c \le x, & 0 \end{Bmatrix}$$

about arbitrary x and MF Parameters a, b, c.
(ii) TRAPMF(trapezoidal membership function): This function computes fuzzy membership values using a trapezoidal membership function.

$$trapmf(x,a,b,c,d) = \begin{Bmatrix} if\ x \le a, & 0 \\ if\ a \le x \le b, & \frac{x-a}{b-a} \\ if\ b \le x \le c, & 1 \\ if\ c \le x \le d, & \frac{d-x}{d-c} \\ if\ d \le x, & 0 \end{Bmatrix} = \max\left(\min\left(\frac{\mathrm{x-a}}{\mathrm{b-a}}, 1, \frac{\mathrm{d-x}}{\mathrm{d-c}}\right), 0\right)$$

about arbitrary x and MF Parameters a, b, c and d.
(iii) GUASSIAN (Gaussian membership function): This function computes fuzzy membership values using a Gaussian membership function.

$$gaussmf(x,\sigma,c) = e^{\frac{-(x-c)}{2\sigma^2}}$$

about arbitrary x and MF Parameters sig and c.
(iv) GUASSIAN2MF: This function computes fuzzy membership values using a combination of two Gaussian membership functions.

$$gauss2mf(x,\sigma_1,c_1,\sigma_2,c_2) = \begin{Bmatrix} left\ gaussian\ curve & e^{\frac{-(x-c_1)}{2\sigma_1{}^2}} \\ right\ gaussian\ curve & 1-e^{\frac{-(x-c_1)}{2\sigma_1{}^2}} \end{Bmatrix}$$

about arbitrary x and MF Parameters sig1 , c1, sig2, and c2. If c1 < c2, the maximum value is 1.
(v) PIMF(Pi-shaped membership function): This function computes fuzzy membership values using a spline-based pi-shaped membership function.

$$pimf(x,a,b,c,d) = \begin{Bmatrix} left\ spline\ curve & smf(x,a,b) \\ right\ spline\ curve & zmf(x,c,d) \end{Bmatrix}$$

about arbitrary x and MF Parameters a, b, c and d.
(vi) DSIGMF(Difference between two sigmoidal membership functions): This function computes fuzzy membership values using the difference between two sigmoidal membership functions.

$$sigmf(x,a,c) = \frac{1}{1+e^{-a(x-c)}}$$

about arbitrary x and MF Parameters a, and c.

2. Literature Survey

A work identifies nine themes for social media understanding and future, organized by predicted eminence [3]. Another work proposes a mechanism through which text data from social media applications can be classified [4]. Peng Liu et al. [5] devised a hybrid privacy preserving algorithm that satisfies the defined model for publishing social network data. Umman Tugba Gürsoy et al. [6] used social media mining and sentiment analysis to analyze social media data for three industry-leading companies in Turkey. The companies, chosen for analysis, were in the construction, food, and technology sectors.

Mohd Najib et al. [7] stated that ANFIS accuracy highly depends on the shape and type of membership function. The author also observed that Gaussian membership function is best fit. In past, various techniques were introduced based on neural network and fuzzy concept but they have not discussed about the sparse or dense network [2]. A work surveys positive and negative impact on fields like health, business, society, education and youth [8]. Another work on social network answers the question that how macro level structures emerged from micro level network processes. The work also answers this question with the help of theoretical model and illustrates its usefulness by fitting stochastic actor-oriented models (SAOMs) [9]. A similar work [10] defines social impact coverage ratio (SICOR) to identify the percentage of tweets and Facebook posts providing information on actual social impact, in relation to the total amount of social media data found related to specific research projects. The proposed technique, in a work, is more efficient for sparse network.

Kaur and Singh [11] presented a broad variety of approaches related in support of anomaly detection, focused on dynamic labeled anomalies. A work revisits the estimation of Exponential-Family Random Graph Models (ERGMS) for small networks and proposes Maximum Likelihood Estimation (MLE), called "ergmito" using exhaustive enumeration [12]. Raju and Sravanthi [13] discussed challenge in data sampling. Farhat Roohi [14] introduced a neurofuzzy system model that combines the neural networks and the fuzzy set theory. Anu Sharma et al. [15] presented hybrid neuro/fuzzy approach that analyzes the suitability of ANN and Fuzzy sets method in a hybrid manner for social web sites classifications. Jeyasudha and Usha [16] focused on the categories of the community deduction with the help of mind map. Informal relations and knowledge sharing are analyzed as determinants of innovation capability; results reveal that friendship relations become the base for knowledge sharing [17]. A few literatures reviewed based on the concept of Web mining for social networks analysis and a neuro-fuzzy based horizontal anomaly detection model is proposed for efficient detection of horizontal anomalies in online social networks [18].

3. Methodology and Proposed Model

3.1 System Architecture

We have assigned the weight to the nodes based on two concepts: friend circle (upto two level and three level) belongs to the node and local clustering coefficient of the node. We have considered two levels and three levels of friendship i.e. i) friend of node and friend of friend of node. ii) friend of node and friend of friend of node and friend of friend of friend of node. The procedure of assigning weight to node is by combining the above two concepts in a single step, the proposed technique assigned weight to every node in the network and normalize input vector by dividing each node weight by maximum number of weight in the matrix. Then apply our membership Function. Hybrid models of ANNs have been proposed, where different activation/transfer functions are used for the nodes in the hidden layer.

We create a weighted matrix where, instead of using 0 and 1 to indicate the absence or presence of an edge between nodes, we use the weight as an indication of the strength of the tie between the two actors. Table 1 shows the mathematical definitions of the activation functions.

Table 1: Activation Functions

Function	Definition	Range
Identity	x	(-inf, +inf)
Logistic	$\frac{1}{1-e^{-1}}$	(0,+1)
Hyperbolic	$\frac{e^x - e^{-x}}{e^x + e^{-x}}$	(-1,+1)
Exponential	e^{-x}	(0, +inf)
Softmax	$\frac{e^x}{\sum_i e^{x_i}}$	(0,+1)
Unit sum	$\frac{x}{\sum_i x_i}$	(0,+1)
Square root	$\sqrt{x}$	(0, +inf)
Sine	sin(x)	[0,+1]
Ramp	$\begin{cases} -1 & x \le -1 \\ x & -1 < x < 1 \\ +1 & x \ge 1 \end{cases}$	[-1,+1]
Step	$\begin{cases} 0 & x < 0 \\ 1 & x \ge 0 \end{cases}$	[0,+1]

오류! 참조 원본을 찾을 수 없습니다. shows the system architecture for proposed model. It consists of six foremost components i.e., “social network Data, Data collection agent, Classification module, Rule manager, Rule base, Data Manager and Administrator”. The input to the Proposed Neuro Fuzzy Classifier system is from the social network trace data.

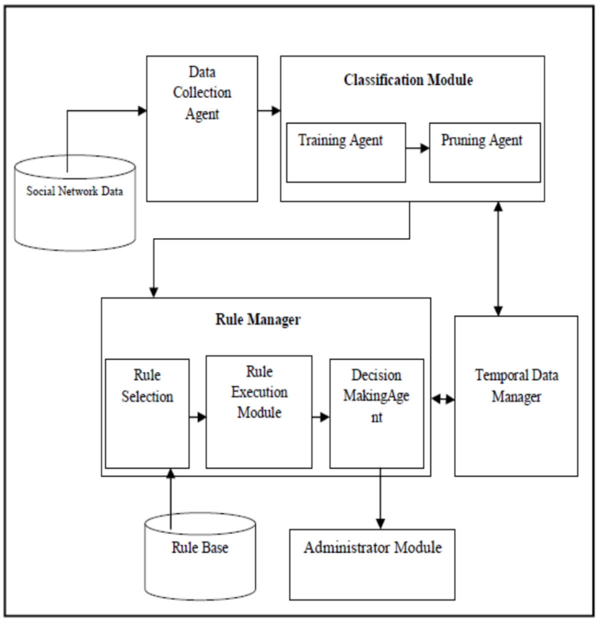

Fig. 1 System Architecture

3.2 Algorithm

The data collection agent collects social network data and gives it to classification module. The classification module with the help of training agent and pruning agent classifies the data and send to rule manager. Rules are selected from rule base. The decision making agent activates administrator module with the help of rule execution module. This time dependent rule management is stored in temporal data manager. The temporal data manager binds data to rule manager bidirectional. The temporal data manager also communicates with classification module for its inputs and outputs. The proposed algorithm is as follows:

Step 1: Read the input data d.
Step 2: Calculated weight w for each node in the graph.
Step 3: Normalize input vector v by dividing each node by maximum number in the matrix.
Step 4: Apply Membership and Activation Function
Step 5: Compare proposed hybrid algorithm with existing back-propagation (BP) algorithm

$$\delta_L = \Delta ENL * S'(nL)$$

Step 6: Compute Average Testing Error and analyze the results.

4. Experimental Consideration

The setup has a base of 64-bit operating system, Windows 10. It has Intel Core i3 1.7GHz dual core processor and a 4 GB of memory. We used SNAP Facebook Dataset. In Facebook, dataset consists of 4039 nodes and 88234 edges [19]. Table 2 shows the detailed statistics of dataset used in our work and the result of BP and HYBRID algorithm after applying membership function whereas

Table 2: Dataset Statistics

Nodes	4039
Edges	88234
Nodes in largest WCC	4039 (1.000)
Edges in largest WCC	88234 (1.000)
Nodes in largest SCC	4039 (1.000)
Edges in largest SCC	88234 (1.000)
Average clustering coefficient	0.6055
Number of triangles	1612010
Fraction of closed triangles	0.2647
Diameter (longest shortest path)	8
90-percentile effective diameter	4.7

This dataset consists of 'circles' (or 'friends lists') from Facebook. Facebook data was collected from survey participants using this Facebook app. The dataset includes node features (profiles), circles, and ego networks. Facebook data has been anonymized by replacing the Facebook-internal ids for each user with a new value. Also, while feature vectors from this dataset have been provided, the interpretation of those features has been obscured. For instance, where the original dataset may have contained a feature "political=Democratic Party", the new data would simply contain "political=anonymized feature 1". Thus, using the anonymized data it is possible to determine whether two users have the same political affiliations, but not what their individual political affiliations represent.

5. Result and Discussion

Table 3 shows the average testing error comparison for the facebook with respect to the test data set. From this table, it can be seen that the proposed algorithm outperforms the existing with respect to average testing error.

Table 3: Simulation Result

	FACEBOOK_COMBINED (FB)			
	Optimization Method			
MF TYPE	BP		HYBRID	
	EPOCH ERROR	AVERAGE TESTING ERROR	EPOCH ERROR	AVERAGE TESTING ERROR
TRIMF	2332.1092	2332.0962	279.2172	279.2169
TRAPMF	2332.1028	2332.098	369.1051	369.1026
GUASSIAN	2332.1109	2332.085	280.2015	280.2009
GUASSIAN 2MF	2332.1024	2332.0868	372.2971	372.2918
PIMF	2332.1008	2332.0826	416.8841	416.8818
DSIGMF	2332.0984	2332.1079	303.554	303.554
PSIGMF	2332.1125	2332.087	303.5572	303.5572

We also worked in the direction of finding absolute error to see the accuracy of our devised algorithm. 오류! 참조 원본을 찾을 수 없습니다. shows the comparison of absolute error for BP and HYBRID algorithm.

Table 4: Comparison of Absolute Error for existing and proposed algorithm

MF TYPE	BP	HYBRID
TRIMF	1.3E-02	3.0E-04
TRAPMF	4.8E-03	2.5E-03
GUASSIAN	2.6E-02	6.0E-04
GUASSIAN2MF	1.6E-02	5.3E-03
PIMF	1.8E-02	2.3E-03
DSIGMF	9.5E-03	0.0E+00
PSIGMF	2.6E-02	0.0E+00

The results are summarized and shown as graph in 오류! 참조 원본을 찾을 수 없습니다..

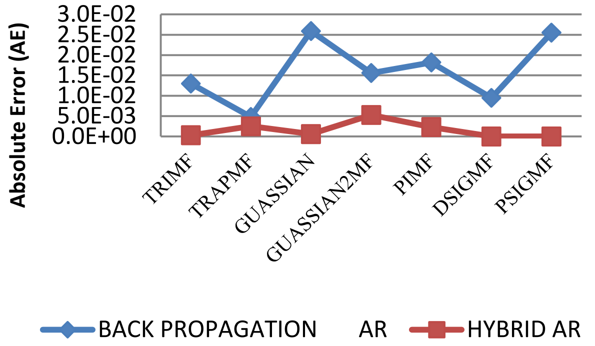

Fig. 2 The comparison of absolute errors

Fig. 2 shows that the proposed algorithm has an error of absolutely zero for TRIMF, DSIGMF and PSIGMF. The hybrid algorithm outperforms in accuracy for GAUSSIAN and PIMF. Although, the hybrid algorithm does not perform well for TRAPMF and GAUSSSIAN2MF, even the results are significant.

Table 5 shows the comparison of relative error for BP and HYBRID algorithm.

Table 5: Comparison of Relative Error for existing and proposed algorithm

MF TYPE	BP	HYBRID
TRIMF	5.6E-04	1.1E-04
TRAPMF	2.1E-04	6.8E-04
GUASSIAN	1.1E-03	2.1E-04
GUASSIAN2MF	6.7E-04	1.4E-03
PIMF	7.8E-04	5.5E-04
DSIGMF	4.1E-04	0.0E+00
PSIGMF	1.1E-03	0.0E+00

The results are summarized and shown as graph in Fig. 3.

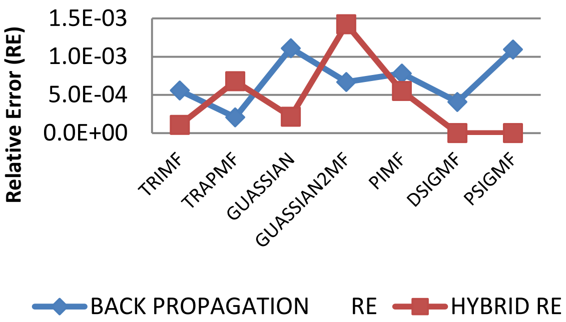

Fig. 3 The comparison of relative errors

The 오류! 참조 원본을 찾을 수 없습니다. shows the relative error for two algorithms i.e. existing BP and proposed HYBRID. The figure shows that the proposed algorithm has an error of absolutely zero and relatively better for TRIMF, DSIGMF and PSIGMF. The hybrid algorithm outperforms in accuracy for GAUSSIAN and PIMF. Although, the hybrid algorithm does not perform well for TRAPMF and GAUSSSIAN2MF, even the results are significant.

The proposed Hybrid algorithm is compared with BP algorithm by finding average testing error, absolute error and relative error. The proposed algorithm obtained results which are significant and important in this particular field of research. The hybrid algorithm not always performs better as can be seen that Relative Error (in fig 4.4) is more in Gaussian2MF and TRAPMF. The experiment shows the improved and more accurate results on analyzing the absolute and relative error.

4. Conclusion

In our work, a new technique is proposed and implemented to classify the datasets in social networks. We are going to find out the average testing error with the help of proposed algorithm and compare it with existing algorithm. The relative error is more in TRAPMF and GAUSSIAN2MF, the reason for this behavior is not clear. The error was less in all hybrid implementations, except the stated two algorithms. We observe that the proposed algorithm provides significantly better accuracy than the existing classifiers.

Our proposed algorithm is legibly suitable for effective social networks analysis. In future research works, we shall focus on how to extend this work to solve more real world problems. The comparing parameters that are used in this paper are; the accuracy of train and test, the absolute and the relative errors of social media data. The result confirms that use of hybrid gives best performance.

References

[1] H. Alzahrani, "Social Media Analytics using Data Mining," *Global Journal of Computer Science and Technology*, vol. 16, no. 4, 2016.

[2] U. Sharma, S. K. Khatri, and L. M. Patnaik, "Link Prediction in Social Networks: A Neuro-Fuzzy Approach," *Int. J. Advance Soft Compu. Appl.*, vol. 8, no. 3, 2016.

[3] G. Appel, L. Grewal, R. Hadi, and A. T. Stephen, "The future of social media in marketing. ," *Journal of the Academy of Marketing Science*, vol. 48, no. 1, pp. 79-95, 2020.

[4] N. Sharma and S. Mandloi, "Sentiment Analysis of Social Media Text Data using Back Propagation in Artificial Neural Networks," *International Journal for Research in Applied Science & Engineering Technology*, vol. 6, no. 1, pp. 1071-1077, 2018.

[5] P. Liu, L. Cui, and X. Li, "A hybrid algorithm for privacy preserving social network publication," in *International Conference on Advanced Data Mining and Applications*, Switzerland, 2014, pp. 267-278.

[6] U. T. Gursoy, D. Bulut, and C. Yigit, "Social media mining and sentiment analysis for brand management," *Global Journal of Emerging Trends in e-Business, Marketing and Consumer Psychology*, vol. 3, no. 1, pp. 497-551, 2017.

[7] M. N. M. Salleh, N. Talpur, and K. H. Talpur, "A modified neuro-fuzzy system using metaheuristic approaches for data classification," *Artificial Intelligence–Emerging Trends and Applications*, pp. 29-45, 2018.

[8] W. Akram and R. Kumar, "A study on positive and negative effects of social media on society," *International Journal of Computer Sciences and Engineering*, vol. 5, no. 10, pp. 347-354, 2017.

[9] C. Stadtfeld, K. Takács, and A. Vörös, "The emergence and stability of groups in social networks," *Social Networks*, vol. 60, pp. 129-145, 2020.

[10] C. M. Pulido, G. Redondo-Sama, T. Sordé-Martí, and R. Flecha, "Social impact in social media: A new method to evaluate the social impact of research," *PloS one*, vol. 13, no. 8, p. e0203117, 2018.

[11] R. Kaur and S. Singh, "A survey of data mining and social network analysis based anomaly detection techniques," *Egyptian informatics journal*, vol. 17, no. 2, pp. 199-216, 2016.

[12] G. G. V. Yon, A. Slaughter, and K. de la Haye, "Exponential random graph models for little networks," *Social Networks*, vol. 64, pp. 225-238, 2021.

[13] E. Raju and K. Sravanthi, "Analysis of social networks using the techniques of web mining," *International Journal of Advanced Research in Computer Science and Software Engineering*, vol. 2, no. 10, pp. 443-450, 2012.

[14] F. Roohi, "Neuro Fuzzy Approach to Data Clustering: A Framework for Analysis," *European Scientific Journal*, vol. 9, no. 9, 2013.

[15] Anu Sharma, M.K Sharma, and R.K Dwivedi, "Hybrid

Neuro-Fuzzy Classification Algorithm for Social Network," *International Journal of Engineering and Advanced Technology*, vol. 8, no. 6, 2019.

[16] J. Jeyasudha and G. Usha, "Community Spam Detection Methodologies for Recommending Nodes," *International Journal of Recent Technology and Engineering*, vol. 8, no. 2, 2019.

[17] A. Mendoza-Silva, "Innovation capability: A sociometric approach," *Social Networks*, vol. 64, pp. 72-82.

[18] A. Sharma, M. K. Sharma, and R. K. Dwivedi, "Literature Review and Challenges of Data Mining Techniques for Social Network Analysis," *Advances in Computational Sciences and Technology*, vol. 10, no. 5, pp. 1337-1354, 2017.

[19] V., Kumar, R. Sharma, W. H. Cheng, M. Atiquzzaman, K. Srinivasan, and A. Y. Zomaya, "NHAD: Neuro-fuzzy based Horizontal Anomaly Detection in online social networks," *IEEE Transactions on Knowledge and Data Engineering*, vol. 30, no. 11, pp. 2171-2184, 2018.

Anu Sharma received the B.Tech Degree from N.C College of Engineering,Israna,Panipat in 2006 and M.Tech degree for MMU, Mullna, Ambala, Haryana in 2011. Assistant professor in CCSIT, TMU, Moradabad, UP, India. She is pursuing Ph.D in CSE from Uttrakhand Technical University, Dehradaun. Her research interest includes data Mining, Neural Network, Machine Learning. She is having experience of more than 12 years. She has published more than 30 research papers in International, UGC approved, Scopus Indexed Journals/Conferences

Dr. M. K. Sharma received the Ph.D from Kumoun University.Professor in Amrapali Institute, Haldwani ,Uttarkhand, India. His research interest includes data Mining, Neural Network, Machine Learning. He is having experience of more than 20 years. He has published more than 45 research papers in international journals, UGC approved, Scopus Indexed Journals.

Dr. Rakesh K. Dwivedi received the Ph.D from IIT Roorkee in Remote Sensing .Director in CCSIT, TMU, Moradabad, UP, India. He has published more than 50 research papers in international/national conferences and more than 45 research papers in international journals. His research interest includes Remote Sensing ,Digital Image Processing, Machine Learning, Deep Learning, Sensor Network.