

NEURO FUZZY LINK BASED CLASSIFIER FOR THE ANALYSIS OF BEHAVIOR MODELS IN SOCIAL NETWORKS

¹Indira Priya Ponnuvel, ²Ghosh Dalim Kumar, ³Kannan Arputharaj and ⁴Ganapathy Sannasi

¹Department of Computer Science and Engineering, Tagore Engineering College, Chennai, Tamil Nadu, India

²Department of Computer Science and Engineering, VSB Engineering College, Karur, Tamil Nadu, India

^{3,4}Department of Information Science and Technology, Anna University, Chennai, Tamil Nadu, India

Received 2013-09-01; Revised 2013-10-25; Accepted 2013-12-09

ABSTRACT

In this study, a new link based classifier using neuro fuzzy logic has been proposed for analyzing the social behavior based on Weblog dataset. In this system, data are processed using a multistage structure. This system provides a diagnosis using a neuro fuzzy link based classifier that analyses the user's behavior to specific diagnostic categories based on their cluster category in social networks. It uses random walks method to organize the labels. Since the links present in the social network graph frequently represent relationships among the users with respect to social contacts and behaviours, this work observes the links of the graph in order to identify the relationships represented in the graph between the users of the social network based on some new social network metrics and the past behaviour of the users. This work is useful to provide connection between consolidated features of users based on network data and also using the traditional metrics used in the analysis of social network users. From the experiments conducted in this research work, it is observed that the proposed work provides better classification accuracy due to the application of neuro fuzzy classification method in link analysis.

Keywords: Social Network, Neuro Fuzzy Logic, Link Classification

1. INTRODUCTION

In social data mining classification is considered as the most effective decision making techniques among all the human activities. A problem that occurs in classification is when a person or an entity has to be put into a class based on the predefined properties of the person or entity. Traditional methods available in statistical for classification including discriminate analysis were used in the past for decision making under uncertaining using baye's theorem Kajdanowicz *et al.* (2010). In this study, a new probability based model has been proposed to compute the posterior probability which is used for effective classification. The effectiveness of this proposed model is constrained by the assumptions made in this work and also based on the related work. Therefore, it is necessary to have a thorough understanding of the proposed model, application constraints and the type of data to be used.

In the past, many classification techniques have been proposed by various researchers based on the application of neural networks for effective classification (Cpalka, 2009). Neural networks provide a number of advantages in the effective classification of data. First, neural networks based classification techniques are flexible and based on weight assignment. Second, they use function approximations with respect to the activation functions. Third, they are useful to form rules for learning and decision making. Finally, training and testing for neural networks are easy to implement.

In social networks, navigation based links analysis on the past browsing pattern of users can be mined using the link structure analysis. A social network can be considered as a graphical structure consisting of links between web pages and relationships between users. Instead, a huge amount of this Weblog data can be represented using labels on links to specify the link types. Therefore, the graph will consists of nodes and links with labels

Corresponding Author: Indira Priya Ponnuvel, Department of Computer Science and Engineering, Tagore Engineering College, Chennai, Tamil Nadu, India

represented on the links such as personal characteristics including age of the person, their gender and the location in which they live currently. Moreover, other properties of a person can be put on the nodes using separate data structures. In such a scenario, other attributes of a person including marital status, income and religious belief can be represented either in nodes or on the links. This representation of knowledge pertaining to persons present in social networks are analyzed in this research work using a new neuro-fuzzy classifier. For this purpose, fuzzy rules have been formed and effective decisions are made based on inference using the rules.

There are many contributions that are made in this research work and are presented in this study. First, a new technique based on graphs for effective representation of social network data has been proposed. Second, a new classification algorithm called using neuro-fuzzy link based classifier has been proposed. Third, a new clustering algorithm called Genetic based Weighted Fuzzy C Means Clustering algorithm has been proposed in this work for effective preprocessing of data. Finally, a knowledge based decision making system has been proposed to form suitable groups in social networks. This work has been tested using a ten-fold cross-validation technique on the Weblog data set. One limitation of this work is that the decisions are made based on rules. However, it is desirable more if the decisions are made using collaborations.

The rest of this study is organized as follows: Section 2 provides a survey of related works in the area of social network analysis. Section 3 depicts the architecture of the system proposed in this research work. Section 4 explains the preprocessing carried out using the proposed clustering algorithm. Section 5 discusses the proposed classification algorithm and the results obtained from this work. Section 6 provides the result discussion. Section 7 gives conclusion on this work and suggests some possible future enhancements.

1.1. Literature Survey

There are many works on classification that are presented in the literature for various applications. For example, Gilbert and Karahalios (2009) proposed an applied link prediction technique to predict tie strength in social media. They apply link prediction in friendship applications where a person can be recommended to be a friend of another person based on mutual friends. Mar *et al.* (2012) proposed an Adaptive Neuro-Fuzzy Inference System (ANFIS) rule that was used by them for the classification of network intrusion data which tries to minimize the intrusion detection delay on the security process for networks. However, such algorithms focused more on network datasets.

Dunlavy *et al.* (2011) proposed a new link prediction model which considers temporal data and uses matrix and tensor factorizations for effective analysis. In their work, the authors focused mainly on predicting the behaviour of people using temporal data pertaining to social networks. However, their model is most suitable for link mining based on statistical features. A novel neuro-fuzzy network has been proposed for effective classification. Their classification model is useful for calculating the initial weights from the training data and to perform decision making using test data. Allali *et al.* (2011) has introduced a new approach for link prediction called internal links and weighted projection to predict links in bipartite graphs. This approach performs well in some social networks to predict links that will appear in the future however, their algorithm does not predict when links are lost.

Pujari and Kanawati (2012) proposed a Link Prediction model for Complex Networks by applying Supervised Rank Aggregation. Their approach is based on supervised rank aggregation and is motivated by the belief that each attribute can provide some unique information which can be aggregated in the end to make a better prediction of association between two unconnected entities in a network. Lu and Zhou (2012) proposed a new Link Prediction method for analysis Complex Networks: According to that, one of the major problems in studying dynamic evolution of complex networks, is the problem of link prediction. This link prediction problem aims to find new associations (edges) in a network at a given point of time t when provided with the information about the network's temporal history before time. Wang *et al.* (2013) proposed a new Markovian model based on graphs for effective social network analysis.

Comparing with all the works present in the literature, the model proposed in this study is different in many ways. First, it uses a clustering analysis to preprocess the Weblog data. Second, it performs effective classification using the proposed neuro-fuzzy model that uses rules. Third, it uses a decision manager which is capable of performing deductive inference using fuzzy rules for effective decision making. Finally, it uses a knowledge base to store the inferred knowledge in the form of facts and rules.

1.2. System Architecture

The architecture of the system proposed in this work consists of six major components namely Weblog data set, User Interface module, Clustering module, classification module, knowledge Base and Decision manager module as shown in **Fig. 1**.

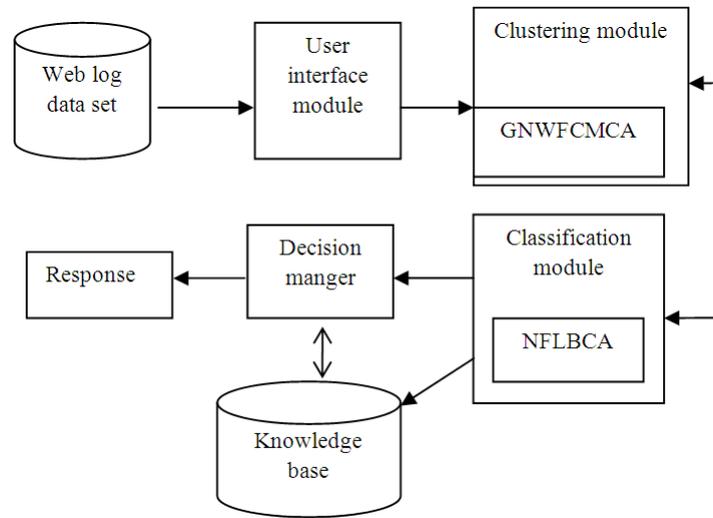


Fig. 1. System architecture

1.3. Weblog Data Set

Weblog dataset is a offline dataset collected from online database like Weblog, YouTube, BlogCatalog. This dataset has been formed by the data retrieved from internet using the default program.

1.4. User Interface Module

The user interface module collects the Weblog data from the dataset available in the internet. This dataset is used in this work to perform effective behaviour analysis that will be used for forming social network groups.

1.5. Clustering Module

This clustering module distinguishes the dissimilar from similar data using the clustering algorithm called GNWFCMCA.

1.6. Classification Module

The classification module classifies the dissimilar the given dataset using the five factor analysis which is a part of the proposed classifier. This classification module classifies the data available in the dataset using the proposed classification algorithm called NFLBCA. The classified results are used to form most related interest group.

1.7. Knowledge Base

The knowledge base module consists of facts about the target area of the application and linguistic control rules. The facts are necessary for the construction of the fuzzy system. This knowledge is frequently presented in the form of linguistic IF...THEN statements.

1.8. Decision Manager

The decision manager module consists of decision making logic and is the core of the fuzzy system. The decision is made from the fuzzy rules to interpret the given output and it decides what the output should look like.

2. MATERIALS AND METHODS

The system proposed in this work consists of three major components namely, behaviour learning module, clustering module and classification module.

2.1. Neuro Fuzzy Link Based Classification Algorithm

The classification module classifies the data set using the proposed classification algorithm called Neuro Fuzzy Link based Classifier.

2.2. Link Analysis

In many existing works on social network analysis, the social network is represented as a graph G which typically contains only one relation. In this proposed system, we propose a link analysis model called Neuro Fuzzy Link Based Classification algorithm in which we apply neuro fuzzy rules to perform link based classification. The neuro fuzzy rules are based on back propagation neural networks with weight adjustment using triangular fuzzy membership function. This neuro fuzzy modelling analysis the duration in which the connection is maintained in the graph is computed and is used to predict the relationship between the users. For

example, all links may have labelling such as “acquaintance”, “is a friend of”, “is a classmate of”, “is a colleague of”, “relative of”, “member of one club” and “member of one group”. For classifying and labeling, different link features are used in this work and a random walk method along with neuro fuzzy rules are used to perform classification accurately.

The steps of the proposed algorithm are as follows:

- Step 1: Initially, all the members are represented by a vertex of a graph which is unlabeled.
- Step 2: For each vertex, provide a list of attributes and a unique identifier.
- Step 3: Form the adjacency list for all the vertices of the graph.
- Step 4: Provide suitable labels to the edges of the graph with initial labels.
- Step 5: Form a frame system based on the hierarchy formed in the graph.
- Step 6: Form clusters based on vertex types and link types.
- Step 7: Perform a random walk, starting from a start node to all the nodes and evaluate the interest patterns.
- Step 8: Convert the graph into a frame system where each vertex is connected with related vertices based on interest.
- Step 9: Apply neuro fuzzy rules formed from training patterns and perform classification by applying the neuro fuzzy rules.

Step 10: Take a test sample and test the classification performed using neuro fuzzy rules.

Step 11: Modified the links with suitable labels.

3. RESULTS

In all three data sets, we computed the average test error, precision, recall and F1 Measure. In all cases, the link-based models outperformed the content-only models. These results are statistically significant at the 95% confidence level or above, based on a paired t-test. **Table 1** shows the summary of average accuracy, precision, recall and F1-measure using different link based models.

From this table, it can be seen that the precision, recall and F1-measure are more for the proposed NFLBCA algorithm than the existing techniques. The corresponding graph is shown in **Fig. 2**.

The conclusion about the proposed neuro fuzzy best link-based model is mixed. On all the data set, the proposed neuro fuzzy link outperforms all the models. In this work, tests have been carried out using Weblog, BlogCatalog3 and YouTube.

In the next experiment, we investigated which type of link features are more predictive with respect to in-link, out-link and co-link. **Table 2** shows the results obtained from three datasets. For binary-link using all the links(in + out + co) always give better results than using any in-link, out-link, or co-link separately and out-link seems to contribute greatest to the improved performance. The corresponding graph is shown in **Fig. 3**.

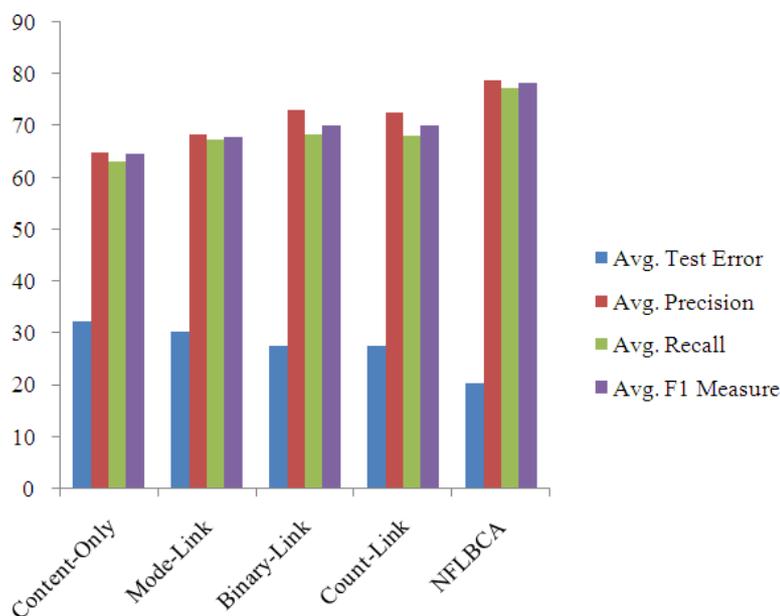


Fig. 2. Average accuracy using different link-based models

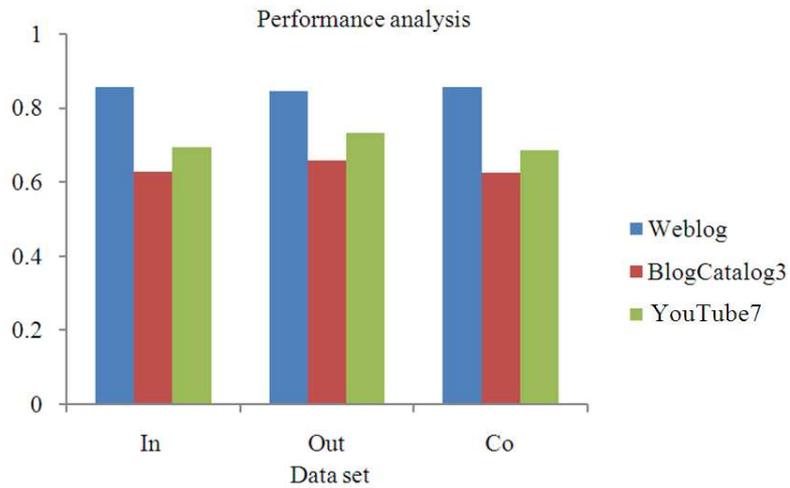


Fig. 3. The convergence rates of different iteration methods on different datasets

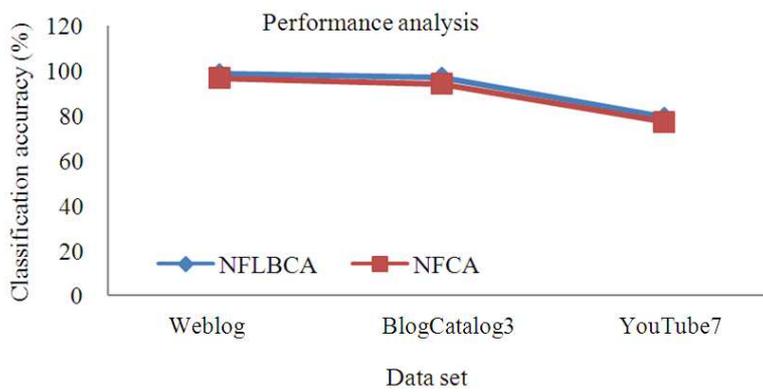


Fig. 4. Classification accuracy in training data set

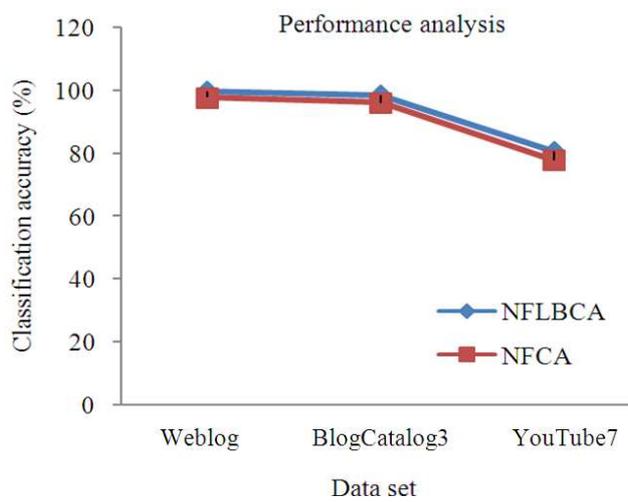


Fig. 5. Classification accuracy in testing data set

Table 1. Results on link-based models

	Content-Only	Mode-Link	Binary-Link	Count-Link	NFLBCA
Avg. Test Error	32.2	30.4	27.6	27.5	20.4
Avg. Precision	64.9	68.3	73.1	72.5	78.9
Avg. Recall	63.1	67.5	68.3	68.0	77.2
Avg. F1 Measure	64.6	67.8	70.2	70.2	78.3

Table 2. Average accuracy using in-links, out-links, co-links separately with binary-link models on BlogCatalog3, Weblog, Youtube7

	NFLBCA		
	In	Out	Co
Weblog	0.857	0.847	0.857
BlogCatalog3	0.629	0.659	0.624
YouTube7	0.695	0.732	0.686

Table 3. Performance analysis of classification accuracy in training set

	Weblog	BlogCatalog3	YouTube7
NFLBCA	98.64	97.08	79.28
NFCA	96.66	94.05	77.12

Table 4. Performance analysis of classification accuracy in test set

	Weblog	BlogCatalog3	YouTube7
NFLBCA	99.64	98.42	80.64
NFCA	97.66	96.05	77.77

Table 3 shows the performance analysis of the proposed NFLBCA with NFCA using three datasets BlogCatalog3, Weblog and YouTube7 (Tang and Liu, 2009; Tang *et al.*, 2009). In this case, all the set of samples is divided into two equal folds randomly. Every one of these two folds is alternatively used either as training set or as a test set.

From this table, it can be observed that the proposed NFLBCA provides better classification accuracy when it is compared with the existing NFCA. This is due to the fact that the link weights modeling provided in this study provides better classification accuracy. The corresponding graphical results are shown in **Fig. 4**.

Table 4 shows the classification accuracy comparison for the two classifiers namely NFLBCA and NFCA with respect to the test data set. From this table, it can be seen that the proposed NFLBCA outperforms the existing NFCA with respect to classification accuracy.

The corresponding graph is shown in **Fig. 5**. From these, it can be seen that the proposed algorithm improves the classification accuracy.

4. DISCUSSION

The proposed a new neuro fuzzy link based classification algorithm on Weblog dataset is evaluated. For this purpose, we have taken one split as a test set and the remaining two splits are used to train this proposed model: One for training and the other split is used as a validation which works with the training set to find the appropriate regularization parameter which is used for test. Using the experiments, compared several baseline models which use the link descriptions for providing an effective analysis.

In the comparison the models namely Content-Only which uses only object attributes and mode-link, binary-link and count-link model that combines a logistic regression model over the object attributes with a separate logistic regression model over the mode of the inlinks, outlinks and co-citations have been considered.

A variety of ordering schemes are compared for the iterative categorization algorithm that use in, out, co type of links where:

- In: Uses the inlink
- Out: Uses the outlink
- Co: Uses the colink

In this study, the average test error, precision, recall and F1-measure are computed. The proposed model results are statistically significant at the 95% and above confidence level which is based on the paired t-test.

5. CONCLUSION

In this study, a new classification algorithm called the neuro fuzzy link based system has been proposed and implemented to classify the weblog dataset in social networks. Since, we focused on using the link structure it helps to improve classification accuracy. Moreover, the proposed a simple framework for modeling link distribution, we proposed a graph model based on link statistics. A logistic classifier has been proposed in this work to perform effective classification. From this work, it is observed that the object attributes and link statistics out performs the binary link structure. The proposed model provides better classification accuracy than the

existing systems. One limitation of this work is the provision of decision making by a central decision manager. This limitation can be overcome in the future work by the use of intelligent agents for enhancing the classification accuracy since they can perform detective inference for effective decision making.

6. REFERENCES

- Allali, O., C. Magnien and M. Latapy, 2011. Link prediction in bipartite graphs using internal links and weighted projection. Proceedings of the IEEE Conference on Computer Communications Workshops, Apr. 10-15, IEEE Xplore Press, Shanghai, pp: 936-941. DOI: 10.1109/INFCOMW.2011.5928947
- Cpalka, K., 2009. A new method for design and reduction of neuro-fuzzy classification systems. IEEE Trans. Neural Netw., 20: 701-714. DOI: 10.1109/TNN.2009.2012425
- Dunlavy, D.M., T.G. Kolda and E. Acar, 2011. Temporal link prediction using matrix and tensor factorizations. ACM. Trans. Knowl. Discov. Data, 5: 27-27. DOI: 10.1145/1921632.1921636
- Gilbert, E. and K. Karahalios 2009. Predicting tie strength with social media. Proceedings of the Conference on Human Factors in Computing Systems, Apr. 4-9, ACM Press, New York, pp: 211-220. DOI: 10.1145/1518701.1518736
- Kajdanowicz, T., P. Kazienko and P. Doscoczek, 2010. Label-dependent feature extraction in social networks for node classification. Proceedings of the 2nd International Conference on Social Informatics, Oct. 27-29, Springer Berlin Heidelberg. Laxenburg, Austria, pp: 89-102. DOI: 10.1007/978-3-642-16567-2_7
- Lu, L. and T. Zhou, 2012. Link prediction in complex networks: A survey. Phys. A: Stat. Mech. Applic., 390: 11500-1170. DOI: 10.1016/j.physa.2010.11.027
- Mar, J., C.C. Kuo and L.S. Lou, 2012. FBFN-based pointing error correction architecture in wind-force environments. IEEE Antennas Wireless Propagat. Lett., 11: 559-563. DOI: 10.1109/LAWP.2012.2199957
- Pujari, M. and R. Kanawati, 2012. Link prediction in complex networks by supervised rank aggregation. Proceedings of the IEEE 24th International Conference on Tools with Artificial Intelligence, Nov. 7-9, IEEE Xplore Press, Athens, pp: 782-789. DOI: 10.1109/ICTAL.2012.111
- Tang, L. and H. Liu, 2009. Relational learning via latent social dimensions. Proceedings of the 15th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Jun. 28-Jul. 01, ACM Press, New York, USA., pp: 817-826. DOI: 10.1145/1557019.1557109
- Tang, L., S. Rajan and V.K. Narayanan, 2009. Large scale multi-label classification via metalabeler. Proceedings of the 18th International Conference on World Wide Web, Apr. 20-24, ACM Press, New York, USA., pp: 211-220. DOI: 10.1145/1526709.1526738
- Wang, T., H. Krim and Y. Viniotis, 2013. A generalized markov graph model: Application to social network analysis. IEEE J. Selec. Top. Sig. Proc., 7: 318-332. DOI: 10.1109/JSTSP.2013.2246767