Revue d'Intelligence Artificielle
Vol. 36, No. 3, June, 2022, pp. 439-444
Journal homepage: http://iieta.org/journals/ria

Ensemble Assisted Multi-Feature Learnt Social Media Link Prediction Model Using Machine Learning Techniques

Ugranada Channabasava*, Bevoor Krishnappa Raghavendra

Department of Information Science & Engg, Don Bosco Institute of Technology, Bengaluru, Karnataka, Affiliated to Visvesvaraya Technological University, Belagavi 590018, Karnataka, India

Corresponding Author Email: channasan11@gmail.com

https://doi.org/10.18280/ria.360311

Received: 7 May 2022
Accepted: 4 June 2022

Keywords:
social media link prediction, multiple features learning based link-prediction, ensemble learning, feature selection

ABSTRACT

In this paper a robust consensus-based ensemble assisted multi-feature learnt social media link prediction model is developed. Unlike classical methods, a multi-level enhancement paradigm was considered where at first the focus was made on extracting maximum possible features depicting inter-node relationship for high accuracy of prediction. Considering robustness of the different feature sets, we extracted local, behavioural as well as topological features including Jaccard coefficient, cosine similarity, number of followers, intermediate followers, ADAR. The use of these all features as link-signifier strengthened the proposed link-prediction model to train over a large data and to ensure higher accuracy. Undeniably, the use of aforesaid multiple features-based approach could yield higher accuracy and reliability; however, at the cost of increased computation. To avoid it, different feature selection methods like rank sum test, cross-correlation, principal component analysis were applied. The use of these feature selection methods had dual intends; first to assess which type of features can have higher accuracy and second to reduce unwanted computation. This research revealed that cosine similarity-based features don't have significant impact on eventual classification. On the contrary, cross-correlation and PCA based features had exhibited relatively higher accuracy (up to 97%). Once retrieving the set of suitable features, unlike standalone classifier based (two-class) prediction, we designed a novel consensus based ensemble learning model by using logistic regression, decision tree algorithm, deep-neuro computing algorithms (ANN-GD and ANN-LM with different hidden layers), which classified each node-pair as Linked or Not-Linked. Our proposed link-prediction model has exhibited link-prediction accuracy (98%), precision (0.93), recall (0.99), and F-Measure (0.97), which is higher than the other state-of-art machine learning methods.

1. INTRODUCTION

The Internet has extended the range of applications. Social media platforms such as Facebook, Twitter, Instagram, LinkedIn, Myspace, and YouTube receive the most attention among the major applications. According to the report, about 3.6 billion people use social media platforms globally in 2020, with that number predicted to rise to 4.41 billion by 2025. On the other hand, the number of people who add to social media platforms has risen over time. Facebook (2603 million users), Twitter (326 million users), WhatsApp (2000 million users), Instagram (1082 million users), YouTube (2000 million users), and so on These figures are useful for social data prediction and analysis, as well as one-to-many (O2TM) communication. This method has been applied to the majority of current commercial applications [1].

2. RELATED WORK

Almansoori et al. [2] recognized the link prediction and found the edges between the two nodes and these edges are most prominent link prediction result.

Wang et al. [3] extracted user's behavioural attributes & preferences from social media to comprehend its social evolution. To define a user's online social behaviour, the authors developed dual representations that included a latent representation of the behavioural element as well as historical behavioural characteristics. They exploited behavioural and topological features to do link-prediction, obtaining separate vectors representing each behavioural aspect across each instant.

Wu et al. [4] for social media link prediction, devised a Rough Set Algorithm (RSA). To illustrate the connection or connectivity between the multiple users, they created an association force model. To perform link prediction, the author applied the user's topological or structural knowledge to predict the different conditions acting between them. A dual bound RSA technique was used to measure the connection between two users.

Kundu and Pal [5] For link prediction, utilised a 3 hidden Bayesian model with diverse factors bearing an impact on social-links. The authors used text-modeling based on latent Dirichlet allocation (LDA) for the behavioural model. Using Gaussian weight based LDA to retrieve the user's implicit interest population, the negative influence on relationships was first investigated. Similarly, a hidden Nave Bayesian model was used to estimate the association between users in order to do link-prediction utilising the users' common neighbour relationships. Xiao et al. [6] used common influence set to perform link-prediction.

Mallick et al. [7] utilised random walk enabled Topo2Vec embedding to forecast social media links using network information. In addition, the symmetric paired sample image of a node pair was estimated using the random forest (RF) classifier. A paired kernel support vector machine (SVM) classifier was used to get the feature-vector and conduct possible link-prediction. the several semantic-rich relationships-based embeddings for assessing social network content evolution. Authors employed a recurrent neural network approach to do connection prediction.

Zhiyuli et al. [8] used hierarchical and structural node-embedding notion for the social-media link-prediction. Markov Logic Networks (MLNs) to anticipate links using the knowledge reasoning approach. MLNs were utilised to recover undirected connections with cycles and long-range reliance between users, as opposed to probabilistic graph models. Author conducted social-media link prediction after getting the structure and attribute dependence.

Chen et al. [9] used the deep dynamic network embedding approach to extract user's network dynamic transitional topologies and historical facts for link prediction.

3. PROPOSED WORK

This section primarily discusses the overall proposed system and its implementation. Being a multi-phased analytics problem, our proposed social media link-prediction model encompasses the following key phases.

Step-1 Data Collection and Pre-processing

Step-2 Multi-constraints or Heterogenous Feature Extraction

Step-3 Feature Selection

Step-4 Consensus based Ensemble Learning for Link-prediction.

The next parts go through the suggested concept in detail as well as its sequential implementation.

3.1 Data collection and pre-processing

Considering contemporary demands and social-media complexity, it is imperative to designing a link-prediction model which could learn over a large user-base and could predict corresponding links swiftly. Thus, retrieving the complete set of users and plotting respective network-graph a total of 9437519 node-edges were taken into consideration. Once obtaining the dataset, the different features containing local features, behavioural as well as topological features were obtained.

In synch with the overall research intend and real-world demand, in our proposed method we have tried to use as maximum features as possible to perform potential link prediction. As indicated above, we have applied different features, say multi-constraints features including local, behavioural and topological features. To understand feature extraction and feature sensitive link-prediction problem, the following example can be considered. Let the set of data instances be $V=v_{(i,i=1,\ldots,n)}$ which is defined in the form of a social media graph G= (V,E), , where states the set of observed links. Now, with this mapped graph, in social media link prediction we intend to predict how closely or probably as unobserved or unknown link exists in between a random node-pair (Vi,Vj) within the network. Majority of the existing methods allocate a node-edge value also called connection weight score(x,y). In fact, this score value signifies the similarity or the measure of proximity between the nodes x and y. For instance, the smallest distance between two nodes and y as very small or negligible would signify that the aforesaid nodes are linked or connected.

We applied the following key features:

a) Jaccard coefficient of the followers

b) Jaccard coefficient of the followee

c) Cosine similarity of the followers

d) Cosine similarity of the followee

e) Number of followers (source as well as destination nodes)

f) Number of followee (source as well as destination nodes)

g) Intermediate followers (say, common Neighbour or friends)

h) Intermediate followee

i) ADAR index.

3.2 Feature selection

Once estimating above derived features and realizing the fact that the overall cumulative feature set can be significantly huge that as a result can impose huge computation, we performed different feature selection methods. Here, our prime motive was to retain only significant features for further classification purpose and to assess which feature set-based model can perform superior. In other words, we have tried to segment a feature selection model which could reduce computational overheads without compromising with link-prediction accuracy and allied performance. In our proposed model, we examined three different feature selection methods, Rank Sum Test, Cross Correlation Test, Principal Component Analysis (PCA).

3.3 Consensus based ensemble for link-prediction

As our proposed link-prediction problem signifies a two-class classification problem (classifying each node pair as connected or not-connected), we have applied machine learning algorithms. Recalling the fact that the different machine learning algorithms exhibits different performance over the same dataset, generalizing the result by one machine learning algorithm is suspicious and questionable. To alleviate such problem in this paper we proposed a consensus-based ensemble learning (CEL) model for social media link-prediction. Unlike classical machine learning algorithms, our proposed CEL embodies multiple base classifiers from the different operating principles. As CEL solution, we used 6 different base-classifiers CELBC, given as follows.

CEL_{BC1}: Linear Regression

$CEL_{BC2:}$ Decision Tree

$CEL_{BC3:}$ SVM

$CEL_{BC4:}$ ANN GD with 1 hidden layer

$CEL_{BC5:}$ ANN GD with 2 hidden layers

Noticeably, in our proposed classification model to perform learning over “deep-features”, we applied ANN variants (here, ANN with Gradient Descent (GD) with the different hidden layers. Hypothesizing the fact, that increase in the hidden layer often impacts classification accuracy (however at the cost of increased computation). Here, we focused on enhancing classification accuracy and hence ANN variants (ANN-GD and ANN-LM) were applied with the different hidden layers as the base classifiers.

Being a consensus-based ensemble learning model (CEL), we implement above stated base classifiers in such manner that each classifier predicts link between peer nodes distinctly and eventually obtaining individual prediction result (Linked or Not-Linked) the consensus model predicts each user-pair as Linked or Not-Linked. Here, consensus model applies maximum voting criteria, also called maximum voting ensemble to perform final link-prediction for each participating node pairs.

A snippet of the machine learning algorithms used as base classifier is given as follows.

3.3.1 Logistic regression

It's one of the most used regression approaches, and it's frequently used for text classification and mining. Logistic regression is used to apply regression over the nodes and their corresponding attributes in the link-prediction issue at hand, where node-features are the independent variable and link-probability is the regression coefficient. As a result, the regression yields two results: Linked and Not-Linked. Mathematically, we apply (1) to perform linear regression over the input features.

$$\operatorname{logit}[\pi(x)\ ]=\beta_0+\beta_1\ X_1+\beta_2\ X_2+\cdots\ldots.+\beta_m\ X_m \tag{1}$$

In above Eq. (1),

The dependent variable is represented by logit(x), and the independent variable is represented by x i. This method uses a logit function to convert the dichotomous outputs, resulting in (x)varying from 0 to 1 to - to +. According to (2) the value m denotes the total number of independent factors, while represents the likelihood of a link between node pairs. The probability result or dependent variable as output is obtained in this manner (2).

$$\pi(x)=e\hat{}(\beta_0+\beta_1X_1+\beta_2X_2+\ldots.+\beta_m\ X_m\ )/(1+e\hat{}(\beta_0+\beta_1\ X_1+\beta_2\ X_2+\ldots.+\beta_m\ X_m\ )\ ) \tag{2}$$

3.3.2 Decision tree (DT)

The decision tree is a well-known and widely used approach for data mining, pattern recognition, and text classification. Its effectiveness and robustness have undergone several changes in order to meet modern expectations. For example, to do data mining and classification over more complex data or features, multiple variations such as IDE, CART, DT C4.5, and DT C5.0 were established. Starting at the root of the tree, the input features of the node's data were divided into many branches at each node of the tree using an association rule between the split-condition. In the following step, it performed two class classifications using the Information Gain Ratio (IGR) value for each branch. As a result, each node pair was labelled as Linked or Not-Linked in our suggested C5.0 decision tree algorithm.

3.3.3 Support vector machine (SVM)

One of the most often used machine learning approaches for pattern categorization is the Support Vector Machine (SVM). SVM's computational power and resilience make it ideal for a variety of classification tasks, such as classification tasks and picture processing. SVM learns over input patterns and operates as a non-probabilistic binary classifier since it is a supervised learning concept. It uses a structural risk mitigation idea to lower the generalization error over unseen cases in order to forecast a solution or categorize the inputs. The support vector provides a subset of the training set that returns the hyper-place border values between two classes with separate features or patterns. In our proposed model, we applied (3) to perform link-prediction.

$$Y\hat{}'=w*\phi(x)+b \tag{3}$$

3.3.4 Deep neuro-computing

Amongst the major machine learning algorithms, neural network often called artificial neural network (ANN) has been applied extensively towards data learning and classification purposes. The robustness of ANN makes it efficient to be used in diverse classification problems, though based on computational complexities and adaptive computation ANN has evolved through different phases. Exploring in depth it can be found that the performance of ANN is directly related to the corresponding learning method. Thus, based on learning method, ANN has been evolved as ANN with steepest gradient (SD), ANN with gradient descent (GD), ANN with RBF (ANN-RBF), ANN with Levenberg Marquardt (ANN-LM) etc. However, in synch with non-linear heterogenous data classification ANN-LM and ANN-GD has been found more effective. Unlike ANN-SD, ANN-GD avoids local minima and convergence issue, even with large non-linear feature set. Similarly, ANN-LM possesses higher robustness than ANN-SD and ANN-GD, individually. Moreover, ANN-LM can be configured to possess feature of ANN-SD as well as ANN-GD and therefore has better performance stability even with large, non-linear and heterogenous data.

4. RESULTS AND DISCUSSIONS

Here, our key intend was to achieve a minimum voluminous set of features which could yield optimal prediction accuracy with minimum possible computation. Moreover, we intended to identify the best suitable feature selection method for link-prediction, especially over large user-base and allied feature size. Thus, obtaining the features from each feature selection algorithms, we fed them to the consensus-based ensemble learning (CEL) model. Our proposed CEL model encompassed machine learning classifiers from the different categories such as regression, decision tree, pattern mining, neuro-computing and hence can be stated as heterogenous ensemble learning model. As base classifiers we applied logistic regression, decision tree (C5.0), SVM with polynomial kernel and ANN variants. Considering superiority of ANN-GD over classical ANN-SD for non-linear large feature learning we applied ANN-GD with different hidden layers (here, we applied ANN-GD with 1, 2 hidden layers, constituting a deep-neuro-computing environment). Similarly, we applied ANN-LM with multiple hidden layers to constitute deep-neuro computing environment. Thus, a total of 6 base-classifiers (in addition to the proposed.

CEL-MVE ensemble classifier) were applied as base classifier to perform social media link prediction. Being a consensus-based learning and prediction, we applied maximum voting concept, often called MVE to predict inter-node or inter-user link probability. Noticeably, each base classifier predicts each node-pair as Linked or Not-Linked and thus labels them as “1” or “0”, respectively. Thus, employing the labels of each node-pair CEL-MVE model predicts each node-pair as Linked or Not-linked. To examine performance, we obtained confusion matrix variables. To achieve it, we obtained True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN) values. Obtaining the above stated matrix values, we derived performance parameters accuracy, precision, recall and F-Measure, as per Table 1.

4.1 Feature sensitiveness analysis

Recalling the above discussion, we applied different feature selection methods (including rank sum test, cross correlation, PCA over the retrieved features to retain only significant features which could ensure optimal performance. To examine performance of the proposed social media link-prediction model, at first, we have assessed suitability of the feature selection method towards the eventual classification. To achieve it, we obtained Accuracy and F-Measure values for the different feature sets. Table 2 and Table 3 present the accuracy and F-Measure outputs for the different features with varied machine learning models.

Table 1. Performance parameters

Parameter	Mathematical Expression	Definition
Accuracy	$\frac{(TN+TP)}{(TN+FN+FP+TP)}$	Signifies the proportion of predicted fault prone modules that are inspected out of all modules.
Precision	$\frac{TP}{(TP+FP)}$	States the degree to which the repeated measurements under unchanged conditions show the same results.
Recall	$TP/((TP+FN))$	It indicates how many of the relevant items are to be identified.
F-measure	$2.(Recall.Precision)/(Recall + Precision)$	It combines the precision and recall numeric value to give a single score, which is defined as the harmonic mean of the recall and precision.

Table 2. Accuracy performance

Feature Selection	Machine Learning (Base Classifiers)					Ensemble
Techniques	LOGR	DT	ANNGD-1H	ANNGD-2H	SVM	CEL-MVE
AM	91.9	**95.5**	86.0	88.5	90.8	**97.1**
RST	89.5	**96.2**	87.1	89.2	92.4	**98.4**
CC	92.2	**95.4**	81.9	87.4	86.5	**95.1**
PCA	78.9	81.1	81.7	**86.4**	84.5	**96.1**

Table 3. Precision performance

Feature Selection	Machine Learning (Base Classifiers)					Ensemble
Techniques	LOGR	DT	ANNGD-1H	ANNGD-2H	SVM	CEL-MVE
AM	0.91	0.95	0.79	0.81	0.91	**0.93**
RST	0.84	**0.95**	0.80	0.83	0.88	0.93
CC	0.88	0.94	0.73	**0.91**	0.79	**0.93**
PCA	0.76	0.87	0.81	0.83	**0.86**	0.85

Table 4. Recall performance

Feature Selection	Machine Learning (Base Classifiers)					Ensemble
Techniques	LOGR	DT	ANNGD-1H	ANNGD-2H	SVM	CEL-MVE
AM	0.93	0.95	0.98	**0.99**	0.98	**0.99**
RST	0.96	0.97	0.98	0.98	0.97	**0.99**
CC	0.97	0.96	**0.99**	0.96	0.98	**0.99**
PCA	0.83	0.89	0.83	0.90	0.82	**0.98**

Table 5. F-Measure performance

Feature Selection	Machine Learning (Base Classifiers)					Ensemble
Techniques	LOGR	DT	ANNGD-1H	ANNGD-2H	SVM	CEL-MVE
AM	0.92	0.95	0.90	0.91	0.90	**0.97**
RST	0.91	**0.94**	0.87	0.90	0.91	0.97
CC	0.92	**0.92**	0.89	0.89	0.88	0.97
PCA	0.83	**0.85**	0.77	0.74	0.81	0.89

Observing the results (Table 2), it can be found that amongst the maximum accuracy with all matrix(AM) with (Accuracy 97.1%), rank sum test (RST) based feature was obtained with our proposed CEL-MVE model (Accuracy 98.4%), while cross-correlation (CC) feature enabled CEL-MVE to achieve the highest accuracy of 95.1%. Similarly, PCA based features could achieve the maximum accuracy of 96.1% with CEL-MVE. The results confirm that our proposed CEL-MVE based link-prediction model outperforms other state-of-art techniques as given in Table 2; however, ANN-LM with higher hidden layers can also be stated as a potential approach towards social media link prediction. Considering feature sensitiveness, it can easily be observed that rank-sum test (RST) and All Matrix (i.e., combined features) are the most suitable features to perform link-prediction.

Table 3 presents the precision performance by the different algorithms. Observing the results, it can be found that our proposed CEL-MVE classifier model with CC features retains maximum precision of 0.95, and with AM matrix (say, All Matrix), our proposed model exhibits precision of 0.93, which signifies a satisfactory precision towards at-hand link-prediction problem. Table 4 presents the recall performance by the different machine learning algorithms with the different features. Interestingly, our proposed CEL-MVE ensemble classifier with AM, RST as well as CC features exhibits recall of 0.99, which is undeniably a significant performance indicator. It affirms robustness and efficiency of our proposed link prediction model. Moreover, observing overall result for recall, ANN-LM with higher number of hidden layers performed relatively better than other base-classifiers. It indicates that ANN-LM can be considered as a potential machine learning classifier for link-prediction task. Observing F-Measure performance (Table 5), with the different (selected) features and machine learning algorithms, we find that the features AM, RST and CC enabled retrieving F-Measure of 0.97 by CEL-MVE model), while ANN-M with two hidden layers achieved maximum F-Measure of 0.95. Undeniably, the robustness of ANN-LM with higher number of hidden layers is proven.

The overall results (Table 2 to Table 5) confirms that AM and CC features possess higher ability to yield better accuracy and reliable link-prediction. Similarly, the relative performance assessment confirms that the proposed CEL-MAE model also maintains higher accuracy (98.4%) and F-Measure (0.97), signifying superior performance over state-of-art base classifiers. Being consensus-based link-prediction, the reliability of CEL-MVE is higher and acceptable than the standalone classifier.

4.2 Machine learning performance assessment

Undeniably, our proposed consensus based multi-feature learnt ensemble model exhibited satisfactory than the classical state-of-art machine learning models (as standalone classifier); however, to assess relative performance with other existing approaches, we have performed qualitative method. In this approach, we compared the performance by our proposed algorithm with other methods. Recently, Li et al. [10] developed a prediction model for trading interactions in an online marketplace where different features such as location and online behaviour were taken into consideration. Noticeably, this approach could achieve maximum accuracy of 92.5% with supervised learning method, while it retained 97% accuracy with unsupervised learning model. Comparing the performance by our own proposed model, it can be found that the proposed CEL-MVE concept with AM feature set achieves higher accuracy (98.4%) as well as satisfactory F-measure performance. It affirms robustness of our proposed model over existing methods. In the study [11], Eberhard et al. applied deep belief network (DBN) based link-prediction model, where their best performance was obtained (with Wikipedia data) as F-score or F-Measure of 0.8577. In comparison to our proposed model, the existing approach performs inferior and hence applauds our proposed model for realistic link-prediction task. Liu et al. [12] applied topological features to perform social media link-prediction, where accuracy performance with Facebook data exhibited the accuracy of approximate 82%. Noticeably, tested different machine learning algorithms such as decision tree C4.5, Naïve Bayes, SVM, ANN, and ensemble algorithms (AdaBoost, Bagging, and Random Forest). Observing overall results in terms of prediction accuracy, it can be confirmed that our proposed model achieves better performance even with large datasets and different features. In contrast to the study [13], where authors applied merely topological information to perform link-prediction, we in our proposed model exploited topological, local as well as behavioural features to perform link-prediction. It shows robustness and higher reliability of our proposed link-prediction technique. Dasari and Devarakonda [14] developed Fuzzy based SVM for social media link-prediction. Though, authors applied Flicker social media data, the maximum accuracy by their proposed model was 98.4%, while the F-Measure performance was obtained as 0.97. Interestingly, authors had applied different features including many that we applied; however, authors lacked justifying their results and novelty. Dasari et al. [15] applied BIOBASE and DBLP databases for link prediction. In their model, Dasari and Devarakonda [16] applied shortest distance features, ADAR, Jaccard, Sum of keyword counts, clustering index etc., while as classifier authors applied decision tree, SVM (linear and RBF), K-NN, Naïve Bayes, multi-layer perceptron, ANN-RBF and Bagging (ensemble) algorithms. Interestingly, the maximum accuracy was obtained using bagging which obtained the highest accuracy of 90.87% and F-Measure of 0.9123 (with DBLP dataset). In reference to our achieved performance, it can be stated that our proposed model achieves better performance than any other existing approaches. Additionally, we have compared our proposed model with many other approaches applying behavioural as well as topological features for social media link-prediction. Interestingly, our proposed approach has performed better than other existing approaches. Due to the space constraints, we couldn't mention performance comparison with other existing methods. The novelties and robustness in our proposed method can be because of higher feature sets, significant feature selection followed by consensus-based ensemble learning for link-prediction.

5. CONCLUSIONS

To achieve multi-feature learning, in this paper different features encompassing local features, topological features as well as behavioural features were obtained from the node edge information. Thus, the use of the different features such as Jaccard coefficient, cosine similarity, number of followers, intermediate followers, ADA for each node and allied node-pair a significantly large but significant feature set was obtained, which could ensure maximum possible prediction accuracy. However, realizing large feature volume different feature selection algorithms such as rank sum test, cross-correlation, principal component analysis were applied. Noticeably, the assessment of these feature selection methods was performed individually as well as with combined feature sets that eventually helped in identifying the best feature selection method towards at hand social-media link prediction problem. Realizing the fact that the different machine learning classifiers exhibit different performance over the same input data, and therefore to introduce diversity of performance by different learning concepts, in this paper varied algorithms such as logistic regression, decision tree, SVM, deep neuro-computing with ANN-GD and ANN-LM were applied. Noticeably, as deep-neuro computing concept, ANN-GD as well as ANN-LM with three different hidden layers (1,2 hidden layers) were applied as the base classifier. Thus, a total of 5 base classifiers were used to enable consensus-based ensemble learning. As consensus formation, maximum voting ensemble (MVE) concept was applied. To be noted, in the proposed model, the individual performance assessment was done for both base classifiers as well as CEL (MVE) ensemble learning model. The use of consensus model enabled optimal classification with higher reliability. Thus, the proposed link-prediction model, being a two-class classification problem enabled each base classifier as well as ensemble learning model to classify node-pair as Linked or Not-Linked. CEL which acts based on MVE concepts employed classification or the prediction output by each base classifier to perform eventual classification or link-prediction. MATLAB 2019b based simulation with Facebook social media data encompassing a total of 1862220 users and allied 9437519 node-edges confirmed that base classifier ANN-LM with three hidden layers performs the highest link-prediction accuracy (95.6%), recall (0.98), precision (0.91), and F-Measure (0.94), while the proposed CEL-MVE model exhibits the maximum cumulative accuracy of 98.4%, precision (0.93), recall, (0.99) and F-measure of 0.97. It confirms the robustness and the suitability of the proposed model for real-world social media link prediction purposes, which can help business houses to identify or segment the connected users to propagate their target business information or service or product information to gain better and competitive market share. Since the proposed model applied merely classical regression, decision tree and machine learning method, in future other advanced algorithms like Extreme Learning Machine (ELM), Least Square SVM (LSSVM) can be assessed for link-prediction.

ACKNOWLEDGMENT

I would like to thank Visvesvaraya Technological University (VTU), GnanaSangama, Belagavi - 590018, Karnataka, India for the monetary support extended to my research work.

REFERENCES

[1] Derani, N.E.S., Naidu, P. (2016). The impact of utilizing social media as a communication platform during a crisis within the oil industry. Procedia Economics and Finance, 35: 650-658. http://dx.doi.org/10.1016/S2212-5671(16)00080-0

[2] Almansoori, W., Gao, S., Jarada, T.N., et al. (2012). Link prediction and classification in social networks and its application in healthcare and systems biology. Network Modeling Analysis in Health Informatics and Bioinformatics, 1(1): 27-36. http://dx.doi.org/10.1007/s13721-012-0005-7

[3] Wang, H., Hu, W., Qiu, Z., Du, B. (2017). Nodes' evolution diversity and link prediction in social networks. IEEE Transactions on Knowledge and Data Engineering, 29(10): 2263-2274. http://dx.doi.org/10.1109/TKDE.2017.2728527

[4] Wu, L., Ge, Y., Liu, Q., Chen, E., Hong, R., Du, J., Wang, M. (2017). Modeling the evolution of users' preferences and social links in social networking services. IEEE Transactions on Knowledge and Data Engineering, 29(6): 1240-1253. http://dx.doi.org/10.1109/TKDE.2017.2663422

[5] Kundu, S., Pal, S.K. (2018). Double bounded rough set, tension measure, and social link prediction. IEEE Transactions on Computational Social Systems, 5(3): 841-853. http://dx.doi.org/10.1109/TCSS.2018.2861215

[6] Xiao, Y., Li, X., Wang, H., Xu, M., Liu, Y. (2018). 3-HBP: A three-level hidden Bayesian link prediction model in social networks. IEEE Transactions on Computational Social Systems, 5(2): 430-443. http://dx.doi.org/10.1109/TCSS.2018.2812721

[7] Mallick, K., Bandyopadhyay, S., Chakraborty, S., Choudhuri, R., Bose, S. (2019). Topo2vec: A novel node embedding generation based on network topology for link prediction. IEEE Transactions on Computational Social Systems, 6(6): 1306-1317. http://dx.doi.org/10.1109/TCSS.2019.2950589

[8] Zhiyuli, A., Liang, X., Chen, Y., Du, X. (2018). Modeling large-scale dynamic social networks via node embeddings. IEEE Transactions on Knowledge and Data Engineering, 31(10): 1994-2007. http://dx.doi.org/10.1109/TKDE.2018.2872602

[9] Chen, H., Ku, W.S., Wang, H., Tang, L., Sun, M.T. (2016). Scaling up Markov logic probabilistic inference for social graphs. IEEE Transactions on Knowledge and Data Engineering, 29(2): 433-445. http://dx.doi.org/10.1109/TKDE.2016.2625251

[10] Li, T., Zhang, J., Philip, S. Y., Zhang, Y., Yan, Y. (2018). Deep dynamic network embedding for link prediction. IEEE Access, 6: 29219-29230. http://dx.doi.org/10.1109/ACCESS.2018.2839770

[11] Eberhard, L., Trattner, C., Atzmueller, M. (2019). Predicting trading interactions in an online marketplace through location-based and online social networks. Information Retrieval Journal, 22(1): 55-92. http://dx.doi.org/10.1007/s10791-018-9336-z

[12] Liu, F., Liu, B., Sun, C., Liu, M., Wang, X. (2015). Deep belief network-based approaches for link prediction in signed social networks. Entropy, 17(4): 2140-2169. http://dx.doi.org/10.3390/e17042140

[13] Fire, M., Tenenboim, L., Lesser, O., Puzis, R., Rokach, L., Elovici, Y. (2011). Link prediction in social networks using computationally efficient topological features. In 2011 IEEE Third International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing, pp. 73-80. http://dx.doi.org/10.1109/PASSAT/SocialCom.2011.20

[14] Dasari, K.B., Devarakonda, N. (2021). Detection of different DDoS attacks using machine learning classification algorithms. Ingénierie des Systèmes d'Information, 26(5): 461-468. https://https://doi.org/10.18280/isi.260505

[15] Dasari, K.B., Devarakonda, N. (2022). TCP/UDP-based exploitation DDoS attacks detection using AI classification algorithms with common uncorrelated feature subset selected by Pearson, Spearman and Kendall correlation methods. Revue d'Intelligence Artificielle, 36(1): 61-71. https://doi.org/10.18280/ria.360107

[16] Dasari, K.B., Devarakonda, N. (2022). Detection of TCP-based DDoS attacks with SVM classification with different kernel functions using common uncorrelated feature subsets. International Journal of Safety and Security Engineering, 12(2): 239-249. https://doi.org/10.18280/ijsse.120213