

A Genetic Programming based Algorithm for Predicting Exchanges in Electronic Trade using Social Networks' Data

Shokooh Sheikh Abooli Poor

Computer Engineering Department
Municipality Training Centre of Applied Science and
Technology
Ahvaz, Iran

Mohammad Ebrahim Shiri

Department of Mathematics and Computer Science
Amirkabir University of Technology
Tehran, Iran

Abstract—Purpose of this paper is to use Facebook dataset for predicting Exchanges in Electronic business. For this purpose, first a dataset is collected from Facebook users and this dataset is divided into two training and test datasets. First, an advertisement post is sent for training data users and feedback from each user is recorded. Then, a learning machine is designed and trained based on these feedbacks and users' profiles. In order to design this learning machine, genetic programming is used. Next, test dataset is used to test the learning machine. The efficiency of the proposed method is evaluated in terms of Precision, Accuracy, Recall and F-Measure. Experiment results showed that the proposed method outperforms basic algorithm (based on J48) and random selection method in selecting objective users for sending advertisements. The proposed method has obtained Accuracy=74% and 73% earning ration in classifying users.

Keywords—Electronic business; Social networks; prediction; machine learning; genetic programming; Facebook network

I. INTRODUCTION

Electronic business [1] means production, marketing, sale and delivery of goods using electronic tools. Although electronic business is still in its infancy, it has played an important role in our daily life, such that it cannot be easily avoided. Since succeeding in electronic business requires having information and analysing marketing environment correctly, it is clear that a major part of this information should be obtained in cyberspace. Additionally, correct information analysis requires knowledge of cyberspace. Since a major part of communications in electronic business is established through available tools including social networks, analysing information obtained from these tools can help the electronic business succeed significantly. The social network in social sciences investigates relations among humans, human groups and organizations. These networks consist of organizational groups which are connected through one or multiple dependencies [2-5].

One of the novel methods for predicting the behaviour of statistical society is to use social networks [6, 7]. The main problem in this paper is that how can information of social networks be used for predictions and improving electronic businesses? In order to answer this question, a learning machine based on genetic programming is proposed to be used in social networks for predicting exchanges in electronic trade.

The approach proposed in this paper might be an important step towards improving electronic business trade.

This template, modified in MS Word 2007 and saved as a "Word 97-2003 Document" for the PC, provides authors with most of the formatting specifications needed for preparing electronic versions of their papers. All standard paper components have been specified for three reasons: (1) ease of use when formatting individual papers, (2) automatic compliance to electronic requirements that facilitates the concurrent or later production of electronic products, and (3) conformity of style throughout a conference proceeding. Margins, column widths, line spacing, and type styles are built-in; examples of the type styles are provided throughout this document and are identified in italic type, within parentheses, following the example. Some components, such as multi-levelled equations, graphics, and tables are not prescribed, although the various table text styles are provided. The formatter will need to create these components, incorporating the applicable criteria that follow.

II. RELATED WORK

Studies conducted in the context of electronic business can be categorized into four general classes:

A. Approaches based on Brand

In general, these approaches focus on the contribution of consumer, sale objectives, Brand loyalty. Pentina et al. [8] have studied the impact of relations among consumers with Facebook and Twitter brands. De Vries et al. [9] have investigated impacts of transmitted messages on brands' pages including clarity, interaction, information content, amusing issues and location of the message. Labrecque [10] has studied whether interactions and social relations of a brand, makes the user offer information and be loyal to that brand?

B. Approaches based on Modeling Information Broadcasting

Researches done in this group is mostly model information broadcasting (including financial information) in social network level. Tsur and Rappoport [11] have predicted activities and performances in social networks using contents and topologies. Bonchi et al. [12] have extracted and searched business plans through learning population structure and network dynamic. Saito et al. [13] have proposed a

probabilistic model based on information broadcasting for prediction.

C. Generic Approaches

Studied in this groups are based on not identifying network structure which is a more difficult level of finding behavioural patterns of users. Rodriguez et al. [14] have designed a generic model for tracking users' path. These researchers have improved their model through concave optimization [15]. Duong et al. [16] have resolved this problem using two approaches: The first learns graphical model potentials for a given network structure, compensating for missing edges through induced correlations among node states. The second learns the missing connections directly.

D. Approaches based on Users' Behavior

In an analysis of social networks, studying users' behaviour based on various hypotheses about the user, has attributed a lot of information. Zhang et al. [17] have identified strong users using their friends' comments. Anagnostopoulos et al. [18] have also identified users with high influence among information broadcast by users through social networks.

III. GENETIC PROGRAMMING

Genetic programming (GP) is an evolutionary computation (EC) technique that automatically solves problems without having to tell the computer explicitly how to do it. At the most abstract level GP is a systematic, domain-independent method for getting computers to automatically solve problems starting from a high-level statement of what needs to be done. Algorithmically, GP comprises the steps shown in Algorithm 1. The main genetic operations involved in GP (line 5 of Algorithm 1) are the following [19, 20]:

- Crossover: the creation of one or two offspring programs by recombining randomly chosen parts from two selected programs.
- Mutation: the creation of one new offspring program by randomly altering a randomly chosen part of one selected program.

Algorithm 1 Abstract GP algorithm

- 1: Randomly create an initial population of programs from the available primitives (see Sect. 2.2).
- 2: repeat
- 3: Execute each program and ascertain its fitness.
- 4: Select one or two program(s) from the population with a probability based on fitness to participate in genetic operations (see Sect. 2.3).
- 5: Create new individual program(s) by applying genetic operations with specified probabilities (see Sect. 2.4).
- 6: until an acceptable solution is found or some other stopping condition is met (for example, reaching a maximum number of generations).
- 7: return the best-so-far individual.

IV. THE PROPOSED METHOD

The main purpose of the proposed method is to design a learning machine with prediction ability to find potential users in social networks for business objectives. For this purpose, first, a dataset including profile information and its links in social networks is collected. Then this dataset is used to send advertisement links to users and their feedbacks are investigated. Users are marked based on opening the link or not

opening the link. Thus, a general dataset is employed to train a predictor learning machine based on genetic programming. Objective variable in genetic programming training is output vector label of "yes" or "no" which indicated whether the link is opened or not. Additionally, data in this dataset is mapped to a numerical space to establish feature vector. After training the learning machine, this machine is used to select target users for commercial operations in social networks.

Flowchart of the proposed method is shown in Figure 1. The general framework of the proposed method is comprised of two main steps:

- 1) Designing Learning Machine for Prediction Process
- 2) Evaluating performance of the designed machine for selecting potential target users in electronic business

In the first step, a learning machine is trained whose input data is the dataset collected from social networks and its output is "yes" or "no" label. The aim of this machine is to create a regression function which maps input data to output labels well.

In the second step, test data is used to test the designed machine. In order to test and validate the performance of this machine, target users for advertisement are selected randomly and using the designed machine. Finally, quality of selected users in these two methods is compared. It should be noted that opening or not opening an advertisement link shows performance quality. In the following, steps of the proposed method are described.

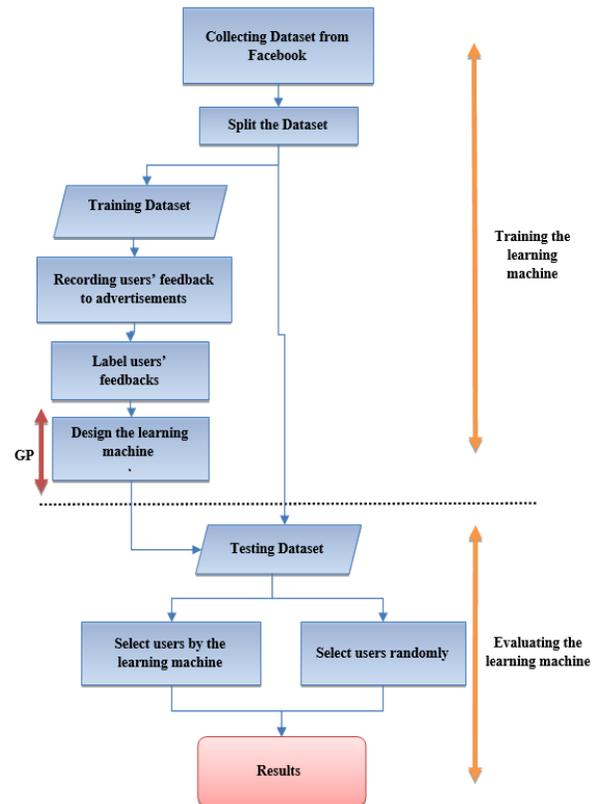


Fig. 1. Flowchart of the proposed method

A. Collecting Dataset

Dataset of this research is adopted from Facebook which is collected by Stanford University [21]. The focus of this network is on users' data. This social network creates an API programming through which users' information can be received as a web service. Main indices extracted from this dataset are shown in Table 1.

In this network, main elements are users. Except for features like name, age and etc., users are specified through their operations in groups and different events of the social network. Figure 2 shows a general scheme of users' activities on this network.

TABLE I. THE MAIN INDICES IN THE NETWORK GRAPH

indices	value
Members	6039
Total number of vertices	88324
Density	0.009
Clustering coefficient	0.6055
Number of Triangles	1612010
Average friends number	7.5
Diameter	8

In a dataset of Stanford University for Facebook, not only users' profile information is available, but also network level information is available. For instance, family relations of a person with other people are also shown. Figure 3 shows friendship circle and relation of a specific user (node v_i) with another user of the network (node u_i).

As can be seen in Figure 3, a user is in many different circles where some of these loops may overlap. This important data in this dataset is a clear characteristic which helps the learning machine to find users with common relations using this friendship circles with higher accuracy.

Each user's data can be considered as a unit record which can be represented by an adjective vector and statistically independent. This dataset has various variables which are both numerical and classified. Each record is in fact a node in the graph of the social network, which specifies a specific user with a unit index.

This dataset has four files which are described as follows:

- Edges file: edges of each node in the network. In Facebook, edges are non-directional.
- Circles file: including circles of a series of nodes.
- Feat file: including features of each node.
- Feat Names file: name of each feature is in this file. Features of users are initialized as 1 and the features which are not initialized for each user are specified with 0.

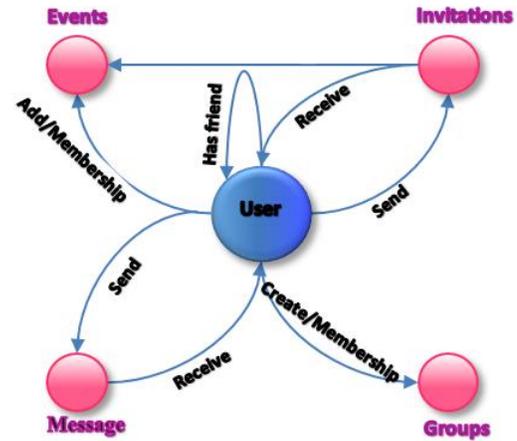


Fig. 2. General scheme of users' activities in social network

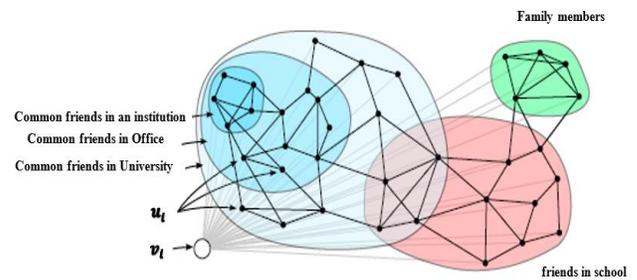


Fig. 3. Facebook Graph with Labels of a user

B. Dataset Splitting

As can be seen in Flowchart of Figure 1, used dataset (6039 records) is divided into two datasets. In the first dataset, called training dataset, 3000 records, and in the second dataset, 3039 records are selected randomly. As mentioned before, a training dataset is used to build the learning machine.

C. Recording Users' Feedback

In this step, a message containing an advertisement is sent to evaluate feedback from users about the advertisement link for users of the training dataset. At this step, 990 users have offered a positive answer which indicated that they had read the advertisement link and others have not responded. Thus, in the next step, training dataset for building the learning machine is a training dataset including 990 positive answers and 2100 negative answers.

D. Designing The Learning Machine

In the proposed method, genetic programming is employed to create the learning machine for selecting potential users. Inspired from the standard algorithm proposed for genetic programming in section 3, a modified version of the algorithm is employed in the proposed method. Since the output of the learning machine in the proposed method is either 0 or 1, a binary version of genetic programming is used. Features considered for each user in the social network is an n-dimensional feature vector as in Equation (1).

$$X = (x_1, x_2, \dots, x_n) \tag{1}$$

Each vector belongs to a class 0 or 1 which determines whether the user has opened the link or not. Therefore, there would be two different output classes c1 and c2:

$$C = (c_1, c_2) \tag{2}$$

And dataset would be:

$$dataset = (X_1, X_2, \dots, X_t) \quad t = userno \times 2 \tag{3}$$

where, *userno* is the number of people who have participated in the test. The purpose of distance learning algorithm is to find a set of metric functions *F* as below:

$$F = (f_1, f_2) \tag{4}$$

$$f_i(X) = \begin{cases} 1 & X \in c_i \\ 0 & X \in c_j \text{ and } i \neq j \end{cases} \tag{5}$$

Such that:
 $\forall_{1 \leq i \leq 2} f_i: R^n \rightarrow \{0, 1\}$

According to the definition of characteristic functions of each class, *f* functions are binary, this metric function *f*, maps each *n*-member input vector *X* to a unit vector in 2D space. Thus, cosine distance between these unit vectors is zero. In other words, desired metric function should have the following characteristic:

$$\forall (X_{li} \in c_i, X_{kj} \in c_j, i \neq j): f_i(X_{li}) = f_i(X_{kj}) = 1 \text{ and } \text{cosinSim}(F(X_{li}), F(X_{kj})) = 0 \tag{6}$$

Cosine function obtained from metric function being zero for two characteristic vectors (*if* $X_{li} \in c_i, \forall j \neq i: f_i(X_{li}) \neq f_j(X_{li})$) means that characteristic functions *f* can be good classifiers for the problem. However, finding these functions is the main challenge of the problem.

Since function *f* is binary, the general framework of the functions in Equation (5) is defined as below:

$$f_i(X) = \sum_{j=1 \dots n} \prod_{l=1 \dots n, l \neq j} (operand_1 \ o \ operand_2) \text{ where } \tag{7}$$

$$operand_1 \leftarrow x_j, operand_2 \leftarrow_{l \neq j} (x_l | const_j), o \in RO$$

$$const_j \in [\min(x_j), \max(x_j)]$$

$$RO = \{<, \leq, >, \geq, =, \neq\}$$

In fact, function *f* is the association of polymers in which each polymer is obtained through connecting monomers operand1 and operand2 with the logic operator and. In other words,

$$f_i(X) = polymer_1 \vee polymer_2 \vee \dots \vee polymer_p \tag{8}$$

$$polymer_i \tag{9}$$

$$= monomer_1 \wedge monomer_2 \wedge \dots \wedge monomer_q$$

$$monomer_i = (operand_1 \ opr \ operand_2) \tag{10}$$

In which indicated and operation and indicates or operation. Thus, function *f* can be changed into a tree recursively. If there is no repetitive monomer in the function of monomer 1 is common among several polymers, these polymers combine and create a tree as in Figure 4.

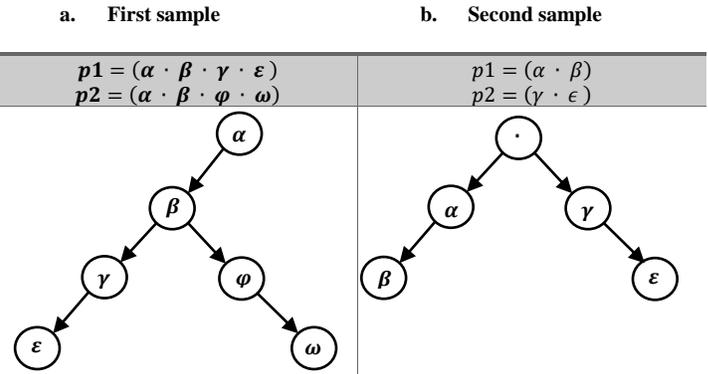


Fig. 4. Converting an algebraic term into a binary tree in genetic programming

E. Evaluating Performance of the Learning Machine

Finally, output function obtained from learning machine is used to identify potential target users who respond to advertisement links positively. The random selection algorithm is also used to evaluate the performance of the learning machine. In this step, total test dataset, 3037 users, are classified by the learning machine. In the next section, detailed results of the test are described.

V. EVALUATION RESULTS

In this section, the proposed method is evaluated and the results obtained from random selection and algorithm based on J48 [22] are compared. Moreover, all implementations and evaluations are performed in MATLAB on a PC with the Intel-i5 processor and 4GB RAM.

In addition, to train and test the proposed method, *K*-fold (*K*=10) method is employed. In this type of test, data are classified into *K* subsets. From these *K* subsets, a subset is used for test and *K*-1 subsets are used for training. This procedure is repeated *K*-times and all data are once used for test and once for training. Finally, an average of these *K* times test is selected as the final estimation. In the *K*-fold method, the ratio of each class in each subset and in the main set is the same.

A. Evaluation Measures

One of the common tools used for evaluating classification algorithms is to employ disturbance matrix. As can be seen in Table 2, disturbance matrix includes results of predictions of classifier algorithm in 4 different classes including True Positive, False Negative, False Positive and True Negative.

TABLE II. CATEGORIES OF CLASSIFICATION

Classified		True	False
Observed	True	TP	FN
	False	FP	TN

Considering the confusion matrix, following measures can be defined and evaluated:

- **True Positive** are those who were properly identified by the algorithm as interested in our product. The cost of advertising is unreservedly covered by the income.
- **True Negative** are those who were properly classified as not interested in our product. There is neither cost, nor income.
- **False Positive** is the group of recipients who were identified as interested in our product while, in reality, they were not. This group creates a cost because we have lost money invested in sending an advertisement. However, as we will see, it is not the worst classification.
- **False Negative** is the most expensive misclassification as we have lost those who would buy our product. Although, we have saved the money not invested into the campaign, those economies are incomparably small to our loss.
- **Precision** is the fraction of retrieved instances that are relevant:

$$\frac{TP}{TP + FP} \quad (11)$$

- **Accuracy** is the proportion of true results (both true positives and true negatives) among the total number of cases examined:

$$\frac{TP + TN}{TP + TN + FP + FN} \quad (12)$$

- **Recall** is the fraction of relevant instances that are retrieved:

$$\frac{TP}{TP + FN} \quad (13)$$

- **F-Measure** combines precision and recall (harmonic mean):

$$\frac{2 \times Recall \times Precision}{Recall + Precision} \quad (14)$$

$$= \frac{2 \times TP}{2 \times TP + FP + FN}$$

We will take an assumption that each positive response gives us z units of revenue while a cost of sending one advertisement is estimated as 0,01z. With these assumptions and referring to classification categories, we can evaluate revenues, costs and profits of classification for each of four previously defined groups of customers. A summary is presented in Table 3.

TABLE III. COST - REVENUE - PROFIT SUMMARY

	Revenue	Cost	Profit
TP	z	0,01z	0,99z
TN	0	0	0
FP	0	0,01z	-0,01z
FN	0	z	-z

B. First Experiment: Effect of features selection on convergence speed of the proposed method

First, the learning machine was trained using all available features. Then, features which have more important role in training this learning machine were extracted through several tests and were used to train the machine. Choosing less and more important features increases training speed of the machine using genetic programming. Among features set, 30 important features were selected among which some features are listed in Table 4.

TABLE IV. SOME IMPORTANT FEATURES OF SOCIAL NETWORKS USERS

Feature	Description
Changes Date	Date of last changes
Creation Date	Date in which the profile is created
Branch	Market branch
Birth decade	Age according to decade classification
Title	Level of Education
Profile visibility	Number of people who can visit the profile
Number of links	Number of linked groups
Number of invitations	Total number of times being invited and having invited

In addition, Figure 5 shows convergence diagram of genetic programming in different iterations for cases of not selecting important features and selecting important features. As can be seen in the results, when features are reduced, genetic programming algorithm converges faster in reducing classification error compared to the case where all features are included.

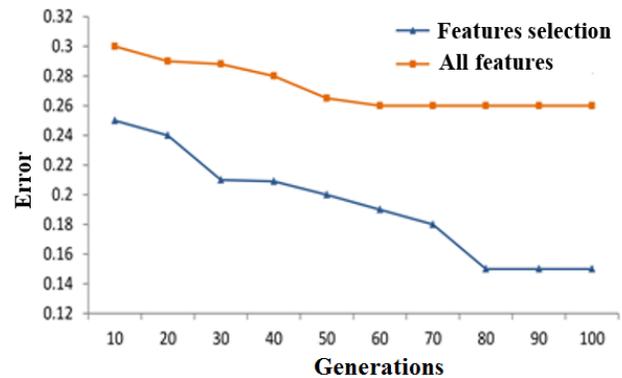


Fig. 5. Comparing convergence speed of genetic programming for using/not using features selection

C. Second Experiment: Calculating Profit

After training and testing the proposed learning machine, obtained results in terms of TP, FP, TN and FN are presented for test and training datasets in Table 5.

For clarity of evaluations, cost and revenue measures are referred (Table 3). Profit gained from test and training of the learning machine is given in Table 6. As mentioned before, maximum loss which occurs in classifying the learning machine is related to the false negative class that is the learning machine has not selected a user as a potential man in the classification process. Thus false negative should be decreased and true positive should be increased.

D. Third Experiment: Evaluating Performance of positive feedbacks using Lift diagram

After extracting important features and creating the proposed learning machine, Lift diagram, Figure 6 is used to evaluate the output of the learning machine in terms of positive feedbacks from test dataset users. In this diagram, the rate of positive answers received from users is compared to two cases of a random selection of users and selecting the users by the proposed learning machine. In this diagram, the rate of positive answers for random selection of users is considered as the base-line. Experiment results show that as a number of selected users in the test dataset increase, Lift rate decreases. Maximum Lift rate is obtained when 30% of users are selected to send them the advertisement. In this case, the proposed learning machine has performed 2 times better. That is, the proposed learning machine could have selected appropriate users to send them the advertisement.

E. Fourth Experiment: Evaluating Common Classification Measures

In this experiment, the performance of the proposed method in terms of Precision, Accuracy, F-measure and Recall in test stage is evaluated. Figure 7 shows results obtained from this test. Evaluation results showed that Precision, accuracy, recall and F-measure are 0.66, 0.72, 0.80 and 0.74, respectively.

Moreover, the performance of the proposed in test and training stage is compared in terms of accuracy and profit, where the results are given in Table 7 and Table 8, respectively and the results are compared with [22]. Comparison results show that the proposed method with 80% accuracy in training stage and 74% accuracy in the test stage, perform better than the method proposed in [22]. This shows the acceptable performance of the proposed method in selecting potential users to send them the advertisement. Table 8 has also compared profit to revenue ratio of the proposed method with that of [22] Comparison results show that the proposed method outperforms the other method.

TABLE V. CLASSIFICATION RESULTS FOR TRAINING AND TEST DATASETS

Sum	FN	TN	FP	TP	Dataset
3000	150	450	1410	990	Training
3039	259	520	1250	1010	Test

TABLE VI. PROFIT OBTAINED FROM TEST AND TRAINING DATASETS BY USING THE PROPOSED METHOD

Profit					revenue	Dataset
	Sum	FN	TN	TP		
834.65z	155.35z	145z	0.45z	9.90z	990z	Training
740.38z	269.62z	259z	0.52z	10.10z	1010z	Test

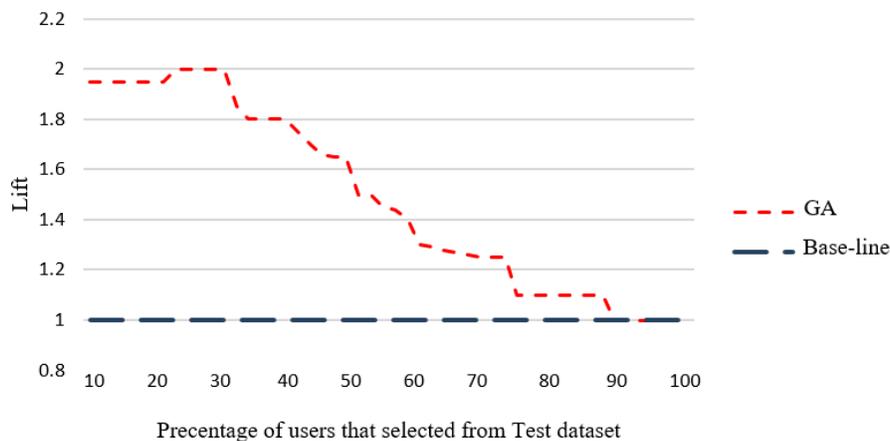


Fig. 6. LIFT diagram of the proposed learning machine compared to random selection method

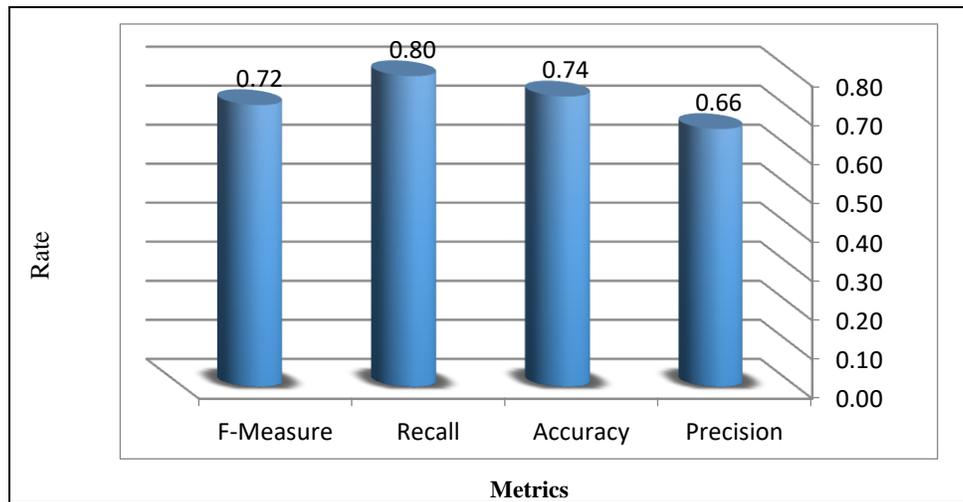


Fig. 7. Evaluating the proposed method in terms of common classification measures

TABLE VII. COMPARING PERFORMANCE OF THE PROPOSED METHOD WITH J48 [22] IN TERMS OF ACCURACY

	Test (Accuracy)	Training (Accuracy ()
Proposed Method	74%	80%
J48 [22]	65%	70%

TABLE VIII. COMPARING PROFIT TO REVENUE RATIO OF THE PROPOSED METHOD WITH J48 [22]

	Test (Accuracy)	Training (Accuracy)
Proposed Method	73%	84%
J48 [22]	35%	44%

VI. CONCLUSION

In this paper, a learning method based on genetic programming is proposed for business predictions in social networks. The main purpose of this method is to select users of the social network who give appropriate feedbacks to the advertisements, they receive. For this purpose, a dataset of users' information from Facebook was collected and studied. This dataset was divided into two test and training datasets. First, a learning machine was trained through sending an advertisement to existing users and receiving their feedbacks. The main purpose of designing the proposed learning machine is to train it to learn how to select users who may give positive answers with higher probability.

After designing and training the proposed learning machine, its performance was evaluated using the test dataset. Experiment results showed that proposed method classifies users with 74% accuracy. Additionally, LIFT test was used to compare the performance of the proposed learning machine with random selection method, and the results showed that for selection of $10\% \leq x < 90\%$ users, the proposed learning machine outperforms random selection method in selecting potential target users. Moreover, for selecting $\geq 90\%$, both methods perform the same. In the future work, we try to

improve the performance of the proposed method by combining meta learning algorithms such as decorate, bagging, and boosting with genetic algorithm.

REFERENCES

- [1] Efraim Turban, H.M.C., Jae Kyu Lee, Michael Chung, *Electronic Commerce: A Managerial Perspective*. 2000: Prentice Hall.
- [2] Liu, L., C.M.K. Cheung, and M.K.O. Lee, *An empirical investigation of information sharing behavior on social commerce sites*. International Journal of Information Management, 2016. **36**(5): p. 686-699.
- [3] Wang, Y., A. Chen, and X. Wang. A Survey Analysis of the Employment Situation and Intentions of E-Commerce Graduates. in 2009 First International Conference on Information Science and Engineering. 2009.
- [4] Brzozowska, A. and D. Bubel, *E-business as a New Trend in the Economy*. Procedia Computer Science, 2015. **65**: p. 1095-1104.
- [5] Yang, Y. and G. Deng, Behavior Study on Consumer Driven e-Commerce, in Cross-Cultural Design: 6th International Conference, CCD 2014, Held as Part of HCI International 2014, Heraklion, Crete, Greece, June 22-27, 2014. Proceedings, P.L.P. Rau, Editor. 2014, Springer International Publishing.
- [6] Fang, B., et al., *A survey of social network and information dissemination analysis*. Chinese Science Bulletin, 2014. **59**(32): p. 4163-4172.
- [7] Zhang, K.Z.K. and M. Benyoucef, *Consumer behavior in social commerce: A literature review*. Decision Support Systems, 2016. **86**: p. 95-108.

- [8] Pentina, I., et al., *Drivers and Outcomes of Brand Relationship Quality in the Context of Online Social Networks*. International Journal of Electronic Commerce, 2013. **17**(3): p. 63-86.
- [9] de Vries, L., S. Gensler, and P.S.H. Leeflang, *Popularity of Brand Posts on Brand Fan Pages: An Investigation of the Effects of Social Media Marketing*. Journal of Interactive Marketing, 2012. **26**(2): p. 83-91.
- [10] Labrecque, L.I., *Fostering Consumer-Brand Relationships in Social Media Environments: The Role of Parasocial Interaction*. Journal of Interactive Marketing, 2014. **28**(2): p. 134-148.
- [11] Tsur, O. and A. Rappoport, *What's in a hashtag?: content based prediction of the spread of ideas in microblogging communities*, in Proceedings of the fifth ACM international conference on Web search and data mining. 2012, ACM: Seattle, Washington, USA. p. 643-652.
- [12] Bonchi, F., et al., *Social Network Analysis and Mining for Business Applications*. ACM Trans. Intell. Syst. Technol., 2011. **2**(3): p. 1-37.
- [13] Saito, K., et al., *Learning Diffusion Probability Based on Node Attributes in Social Networks*, in Foundations of Intelligent Systems: 19th International Symposium, ISMIS 2011, Warsaw, Poland, June 28-30, 2011. Proceedings, M. Kryszkiewicz, et al., Editors. 2011, Springer Berlin Heidelberg: Berlin, Heidelberg. p. 153-162.
- [14] Rodriguez, M.G., J. Leskovec, and A. Krause, *Inferring networks of diffusion and influence*, in Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining. 2010, ACM: Washington, DC, USA. p. 1019-1028.
- [15] Gomez-Rodriguez, M., et al., *Influence Estimation and Maximization in Continuous-Time Diffusion Networks*. ACM Trans. Inf. Syst., 2016. **34**(2): p. 1-33.
- [16] Duong, Q., M.P. Wellman, and S. Singh. *Modeling Information Diffusion in Networks with Unobserved Links*. in Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third International Conference on Social Computing (SocialCom), 2011 IEEE Third International Conference on. 2011.
- [17] Zhang, Z., J.J. Salerno, and P.S. Yu, *Applying data mining in investigating money laundering crimes*, in Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining. 2003, ACM: Washington, D.C. p. 747-752.
- [18] Anagnostopoulos, A., G. Brova, and E. Terzi, *Peer and authority pressure in information-propagation models*, in Machine Learning and Knowledge Discovery in Databases. 2011, Springer. p. 76-91.
- [19] Koza, J.R., *Genetic programming: on the programming of computers by means of natural selection*. 1992: MIT Press. 680.
- [20] Poli, R., et al., *A field guide to genetic programming*. 2008: Lulu. com.
- [21] McAuley, J.J. and J. Leskovec. *Learning to Discover Social Circles in Ego Networks*. in NIPS. 2012.
- [22] J. Surma, A. Furmanek, *Improving marketing response by data mining in social network*, 2010 International Conference on Advances in Social Networks Analysis and Mining, p. 446-451.

AUTHOR PROFILES



systems.

Shokooh Sheikh Abooli Poor received the B.S. degree in Computer Engineering from the Islamic Azad University, Ramhormoz, Iran, in 2013, and M.S. degree in Computer Engineering from Islamic Azad University, Broujerd, Iran, in 2017. Her research interests include wireless networks, learning systems, security, data mining, and recommender



intelligent

Mohammad Ebrahim Shiri is an assistant professor in the department of computer sciences at the Amirkabir University of Technology of Tehran, Iran. He received his Ph.D. from the department of computer sciences at the University of Montreal, Canada in 1999. His current research interests include artificial intelligence, multi-agent systems, tutoring systems and distributed systems.