Proceedings of the 2013 Industrial and Systems Engineering Research Conference
A. Krishnamurthy and W.K.V. Chan, eds.

Customers' Behavior Prediction Using Artificial Neural Network

Bichen Zheng , Keith Thompson, Sarah S. Lam, Sang Won Yoon
Department of Systems Science and Industrial Engineering
State University of New York at Binghamton
Binghamton, New York 13902

Nathan Gnanasambandam
Xerox Innovation Group
Webster, NY 14580

Abstract

In this paper, customer restaurant preference is predicted based on social media location check-ins. Historical preferences of the customer and the influence of the customer's social network are used in combination with the customer's mobility characteristics as inputs to the model. As the popularity of social media increases, more and more customer comments and feedback about products and services are available online. It not only becomes a way of sharing information among friends in the social network but also forms a new type of survey which can be utilized by business companies to improve their existing products, services, and market analysis. Approximately 121,000 foursquare restaurant check-ins in the Greater New York City area are used in this research. Artificial neural networks (ANN) and support vector machine (SVM) are developed to predict the customers' behavior regarding restaurant preferences. ANN provides 93.13% average accuracy across investigated customers, compared to only 54.00% for SVM with a sigmoid kernel function.

Keywords
Social media, data mining, neural network

1. Introduction

In recent years, several location-based social networking sites (e.g., Brightkite, Gowalla, and Foursquare) and location-based smart phone applications (e.g., Loopt and Google Latitude) have emerged. In addition to traditional social-networking capabilities, such services offer peole the ability to post short status updates identifying their current activity location by "checking in" at a venue. These check-ins can then be shared within a person's social network (often linked to a Twitter account), thus allowing users to (1) learn about new venues for socializing and (2) meet up with friends at their current location. It also opens the door to analyzing the relationships that exist between customer mobility patterns and customer behaviors. In the past, most human mobility data which has been studied comes from cell-phone tower tracking. This, however, lacks both functional information about the venues visited by the individual and social information about relationships. Also, most of customers' behavior prediction is based on the historical preferences without considering mobility characteristics and influence of social network. Thus, incorporating location-based information together with social network influence, could potentially improve customer behavior prediction.

As the popularity of the social media continues to increase dramatically, more and more customer comments and feedback about product and services are available online. It not only becomes a way of sharing information among friends in the social network but also forms a new type of survey which can be utilized by business companies to improve their existing product and service and also explore the potential marketplace. Traditionally, market analytics rely more on government data, financial reports from competitors in the marketplace and customer surveys. Social media provides another data resource for enterprises to analyze potential market trends, market size and customer behavior patterns.

Nowadays, social media has become a very important tool for customers to research and compare the product. Also, it provides a new outlet for companies to stimulate the customers' purchasing behavior. It introduces more and more factors which can affect customer behaviors and modes of decision making (e.g., friends' opinions on the social network). Clearly, social media data provide a vast source of information which can help enterprises understand customer behaviors; however, how to utilize this raw data to discovery knowledge remains a critical issue in business analytics.

In the remainder of this paper, different methodologies to predict customer behavior in the literature are summarized in Section 2. In Section 3, ANN and SVM are proposed to predict customer restaurant preferences based on social media data. Experimental results are presented in Section 4. Finally, conclusions and opportunities for future work are discussed in Section 5.

2. Literature Review

Traditionally, market analysis was conducted by utilizing time series analysis and other techniques based on historical sales data, such as customer transaction information, etc. Today, data mining technology is increasingly being utilized for customer analysis. Different models of customer behavior have been generated to learn and understand customer patterns of response to increase company revenues [1]. One of the hot issues in the literature for learning the customer behavior is to predict a customer's favored product based on previous purchasing behavior. It either can be treated as a clustering problem that clusters similar products the customer viewed and purchased previously or it can be modeled as a classification problem to distinguish products purchased by the customer and products the customer dislikes [2]. These types of research were mainly based on the features of products the customer purchased.

A multi-classifier model was used for predicting customers' purchase behavior because of the limitations of single classifier technique for classification problems [3]. In their research, the prediction task was treated as a classification problem to classify the customers' behavior into two categories: buy and not buy. The original data for this research contained 10 demographical features of the user, and 5 features of the historical purchase and review behavior based on the web data-base. By using a genetic algorithm for selection and combination of classifiers, the combined multi-classifier provided 97.8% accuracy which was higher than any individual classifier obtained by ANN. It confirmed that the proposed method can be applied for this prediction task, but it only reflected the relationship between the user's profile and user's historical purchase behavior. Other affecting factors, such as comments about the product, were not considered in the model.

Customer purchase sequences were also mined for customer purchase recommendations [4]. Since customer preferences change over time, a traditional static collaborative filter (CF) recommendation system was not sufficient to provide accurate recommendations for customers. To enhance the traditional approach, a methodology was developed in two phases: (1) building a reliable model to recognize the patterns of the historical customer purchase behavior and (2) matching a similar purchase sequence and manner for that customer to recommend a product. This model incorporated the purchase sequence and time series into consideration and provided a dynamic recommendation system for a stream of customer transaction records. The proposed methodology performed better than the traditional CF technique, but the accuracy was affected by the selection of the number of customer purchase behavior patterns.

Churn rate plays an important role for indicating the customer behavior in the wireless service industry. It costs five times as much to sign on a new subscriber as to keep an existing customer, so it is crucial to predict churn based on the activities of customers in order to maintain a longer stable relationship with existing customers. A modified ANN was utilized to predict the customer behavior for churn rate based on non-stationary data [5]. According to the stream of data in real time, the pattern of the customer was learned by neural network from a huge data set including 71 customer features, such as credit, recent activities etc. From the receiver operating characteristic (ROC) curve, the accuracy of prediction was not high but it significantly improved the accuracy from previous methodologies. Later, a prediction model was proposed via information derived from calling links which improved the accuracy [6]. It not only introduced another methodology on churn rate prediction, but also brought a critical concept into researchers' sight, which was the perspective of viewing the data. By transforming and reconstructing data, the new methodology provided more useful information than previously. A novel prediction method, improved balanced random forests (IBRF), was developed for churn rate prediction with bank client data [7]. A hybrid of the traditional balanced random forest and weighted random forest was used to take the advantage of both methods in dealing with large, imbalanced, and noisy data. The accuracy was significantly improved to 93.2% by IBRF, compared to ANN (78.1%) and decision tree (62.0%).

Analysis of location-based social media data suggests that frequency of check-ins by time of day is largely consistent across New York, Los Angeles, and Amsterdam with peaks of activity around noon and 7:00pm [8]. This suggests that people commonly are checking-in at restaurants during lunch and dinner hours. Cheng et al. [8] also examine how radius of gyration varies according to various city demographics such as population density, average household income, popularity, and social status. There is a positive correlation between the radius and each of the given demographics, thus showing that the radius is related to both population and individual characteristics.

Hubs of social activity within cities can be visualized via heat maps to get a sense of the urban social landscape. The concepts of polycentricity, fragmentation, and agglomeration have been used to characterize the social landscapes of New York, Paris, and London based on social media data [9]. The types of venue (e.g., dining, art, shopping, etc.) for various social media check-ins appear to follow a power law distribution for each of the three cities, with New York having the smallest decay rate across the various check-in locations [10].

Recently, data from location-based social media has also been used to predict human movement. A second-order Markov predictor which considers the sequence of visited locations for a given individual rather than the actual times of each visit was able to achieve a median accuracy of 49% on a set of check-ins from 1,518 people where each person had at least 50 check-ins [11]. A Gaussian spatio-temporal model with two latent states was used to predict location with 36% accuracy [12]. By adding an additional social component which accounted for the probability of checking in given that a known friend had checked in increased the accuracy to 40%.

A location-based recommendation system model was proposed to predict preferred restaurants not based on the data of previous restaurants the customers had been, but based on the weather, and the demographics of the customers such as age, nationality, mood etc. [13]. A Bayesian network was used to predict the probability distribution for each class of restaurant. The data of this research was collected manually by tracking seven volunteers in the real world, who were independent of each other, so they did not share their information and feelings with each other; however, this is not usually what customers do in reality. Generally, customers like to share their feelings and comment to their friends through different channels.

Prediction of customer behavior based on social media data is a relatively new area of research. As the number of people using social media increases, it becomes more possible to predict customer behavior based on business intelligence gathered from social media data. In this research, customer behavior is defined as the selection of a particular type of restaurant. Two approaches are proposed for predicting the choice of each customer restaurant preferences based on previous customer behavior, personal mobility characteristics and influence from the customers' social networks. Factors such as friends' opinions, the time stamp of the customer's activity, the types of restaurants the customer has previously been, etc. are used to mimic the process of human decision-making. ANN and SVM are proposed to model the relationship between these different factors and the final decision of the customer.

3. Methodology

The framework for the prediction model is shown in Figure 1. After preprocessing the data, the raw data is converted into four features: Predicting Time, Historical Preferences, Friends' Recent Impact, and Transience Status. Predicting time is the time stamp that the user wants to make a decision about the type of restaurant to choose. Historical preference is the historical choice for different types of meals (breakfast, lunch, dinner and late dinner). Historical preferences of user i are encoded as follows:

$$H^i = \left[\begin{array}{cccc|cccc|c|cccc} n^i_{11}(1) & n^i_{12}(1) & \cdots & n^i_{1M}(1) & n^i_{21}(1) & n^i_{22}(1) & \cdots & n^i_{2M}(1) & \cdots & n^i_{N1}(1) & n^i_{N2}(1) & \cdots & n^i_{NM}(1) \\ n^i_{11}(2) & n^i_{12}(2) & \cdots & n^i_{1M}(2) & n^i_{21}(2) & n^i_{22}(2) & \cdots & n^i_{2M}(2) & \cdots & n^i_{N1}(2) & n^i_{N2}(2) & \cdots & n^i_{NM}(2) \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ n^i_{11}(T) & n^i_{12}(T) & \cdots & n^i_{1M}(T) & n^i_{21}(T^i) & n^i_{22}(T) & \cdots & n^i_{2M}(T) & \cdots & n^i_{N1}(T) & n^i_{N2}(T) & \cdots & n^i_{NM}(T) \end{array}\right] \tag{1}$$

where N is number of restaurant types, M is the number of time blocks for dining each day, T is the time index of check-ins, and $n^i_{kl}(t)$ is the number of restaurant check-ins of type k during time block l for user i before time t. The friends' recent impact is the impact from all of the friends of the user in the social network based on the friends' recent activities. Finally, based on historical movement of the user, users are categorized according to the transience status of their current locations.

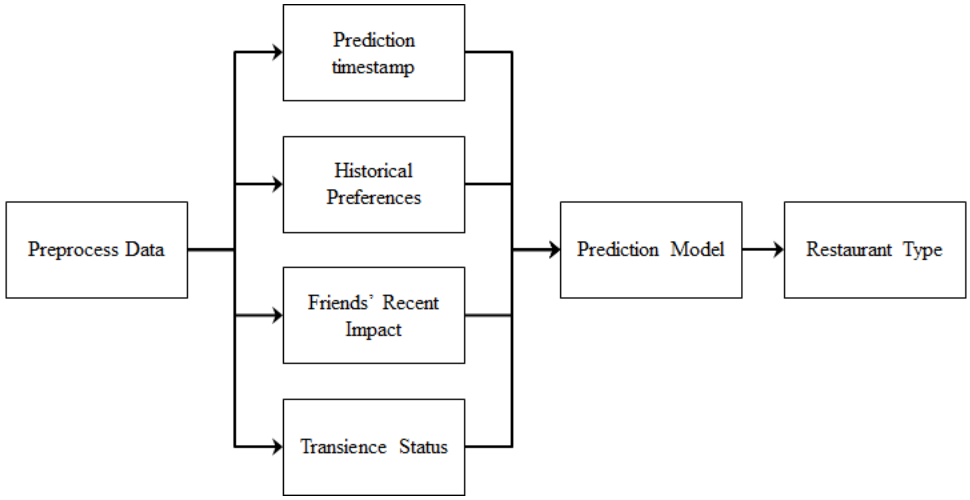


Figure 1: Prediction process framework

3.1 Data Description

In this research, data set is collected from Foursquare.com, which is a website for users to check in their locations with comments. It is unstructured data with text, time stamp and hyperlinks. 120,825 restaurant related check-ins have been collected with user ID, tweets, longitude, latitude, and time stamp as shown in Table 1.

Table 1: Sample data

User ID	Time Stamp	Longitude	Latitude	Tweet
1	07/28/2010 22:45	40.663866	-73.9845	Out with frienss. Yay NYC. (@ Anthony's Pizza) `http://4sq.com/6efEMP`
2	05/11/2010 18:45	40.726921	-73.9853	Anything plantains. (@ Caracas Arepa Bar) `http://4sq.com/1cg8Sv`
2	05/12/2010 00:55	40.747300	-74.0008	I'm at Co. (230 9th Ave, at 24th, New York) w/ 2 others. `http://4sq.com/1PW9Ft`
3	04/08/2010 01:02	40.727460	-73.9847	Fat Denny 4life! (@ Crif Dogs w/ @willmcd) `http://4sq.com/1QGNxV`

3.2 Data Preprocessing

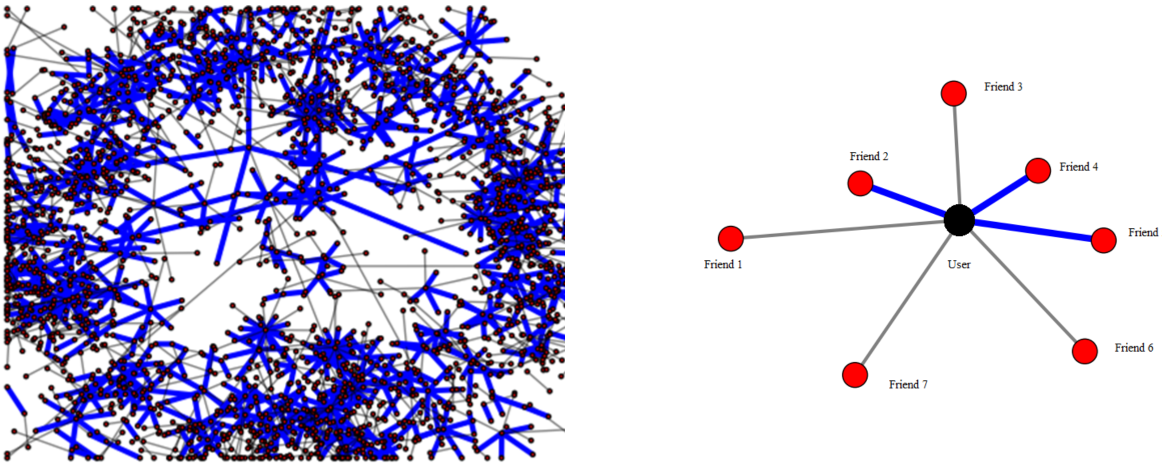
In order to transfer the original data to features of the user which can be utilized as inputs for the prediction model, data preprocessing is necessary. One important factor affecting customers' behavior is the friends' impact which is shown in Figure 3 as a weighted graph. The centroid is the user, whose preference is predicted. The surrounding nodes represent the friends of that user. The weight of the edge between users i and j represents the Influence Factor ($\theta_{ij}(t)$), which is calculated as follows:

$$\theta_{ij}(t) = 1 - \frac{1}{N}\sum_{k=1}^{N} \left| \frac{\sum_{l=1}^{M} n_{kl}^{i}(t)}{\sum_{l=1}^{M}\sum_{k=1}^{N} n_{kl}^{i}(t)} - \frac{\sum_{l=1}^{M} n_{kl}^{j}(t)}{\sum_{l=1}^{M}\sum_{k=1}^{N} n_{kl}^{j}(t)} \right| \tag{2}$$

In fact, the Influence Factor measures the similarity between the user whose preference is predicted and that user's friends based on historical check-ins. It is assumed that the more similar preferences, the more impact on the future decision. In Figure 2, the blue links indicate the similarities (weights) that are higher than 50%. Since the graph of the social network for this user evolves as time elapses, the Influence Factor calculation should be based on the predicting time t. To eliminate the effect of different scales of input, the input feature matrix, X, is normalized as follows:

$$x'^{i}_{tc} = \frac{x^{i}_{tc}}{\sigma(x^{i}_{*c})} \tag{3}$$

where x^{i}_{tc} is the feature value of record t for user i, x'^{i}_{tc} is the new feature value of record t for user i, and c is the column index of input feature matrix.



(a) Giant component of user social network

(b) Subgraph of predicted user social network

Figure 2: User's social network

The transience of a check-in location (grid) is based on an episode analysis of sequential check-ins where the grids of the first and last check-ins in the episode are the same [14]. The grid is labeled as non-transient if there are at least two episodes with a Jaccard similarity of 1 between their respective grid sets. It is labeled as semi-transient if there are at least two episodes with a Jaccard similarity between 0.5 and 1. Otherwise, the grid is labeled as transient for the individual. For this research, the transience status is used as input to the ANN model. Non-transient grids are given an input value of 1, semi-transient grids are given a value of 0.5, and transient grids are given a value of 0. Manhattan check-in transience statuses for one individual are shown in Figure 3. Here, "R" represents regular or "non-transient" grids, "S" represents "semi-transient" grids, and "T" represents transient grids.

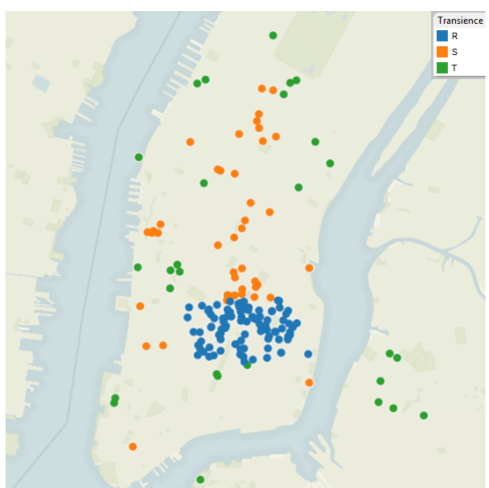


Figure 3: Transience status of user check-ins

In order to obtain the category (restaurant type) of a check-in venue, a two-tiered strategy is used. For check-ins from the website Foursquare, the category will be retrieved by scraping the category from the corresponding check-in venue web page (accessed via the URL from the check-in text). For check-ins from other sites (as well as check-ins for which the Foursquare category is null or the Foursquare web page is missing), the venue names are extracted from the text and used in conjunction with the GPS coordinates to retrieve the category using GooglePlaces API if possible. After merging the category names from Foursquare and GooglePlaces and combining similar categories, there are a total of 114 categories including the "null" category for those check-in venues which were not categorized by either

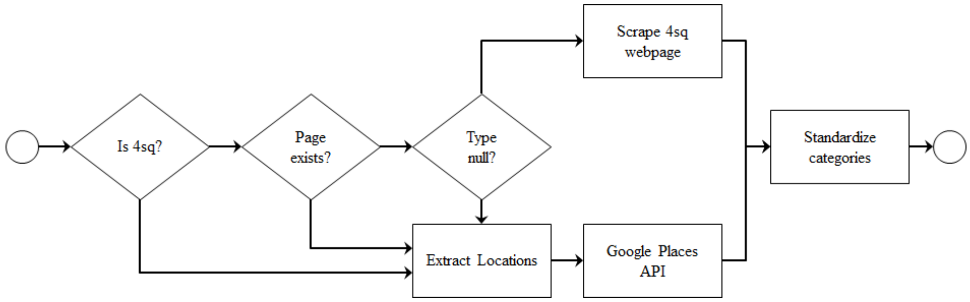


Figure 4: Restaurant type extraction flowchart

website. Thirteen of these categories are restaurant types. A flowchart for this process is shown in Figure 4. Manhattan restaurant check-in categories for one individual are shown in Figure 5.

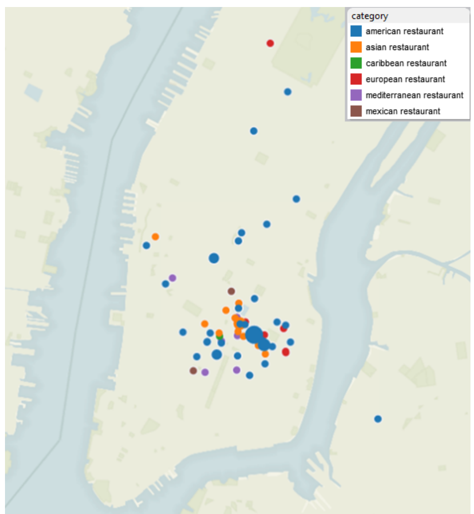


Figure 5: Restaurant check-ins for a single individual

3.3 Support Vector Machine

The first approach which can be utilized as a prediction model is SVM. SVM is a class of machine learning algorithms that can perform pattern recognition and regression based on the theory of statistical learning and the principle of structural risk minimization [15]. SVM was invented to search for the hyper-plane that separates a set of positive examples from a set of negative examples with maximum margin [16]. To obtain the classifier for different preferences based on different users, an optimization problem is solved during training as follows [16]:

$$\max_{\alpha} \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{n} \alpha_i \alpha_j y_i y_j K(x_i, x_j) \tag{4}$$

$$\text{s.t.} \sum_{i=1}^{n} \alpha_i y_j = 0, \tag{5}$$

$$0 \leq \alpha_i \leq C, i = 1, 2, \cdots, n \tag{6}$$

where x is the training vector, y is the label associated with the training vectors, α is the parameters vector of classifier hyper-plane, K is a kernel function for measuring the distance between the training vectors x_i and x_j, and C is the upper boundary of the parameters. Traditionally, SVM is used for binary classification. In this research, there are $N = 13$ restaurant types. Each time, SVM is trained to obtain the decision boundary between one and the other $N-1$ restaurant type. Three kernel functions are used to compare the effect of learning: linear function, sigmoid function, and radial basis function. Vector y is encoded by a zero-one vector to present the different preferences on restaurant.

3.4 Artificial Neural Network

The second approach used for prediction model is ANN. Neural networks have a long history for prediction problems. In particular, neural networkS have been used for predicting customer behavior [3, 5]. In this methodology, based on the data preprocessing result, traditional multilayer perceptron neural networks are trained by backpropagation to learn the relationship between the inputs and outputs. Because of the high dimensionality of the feature space, 67 are used in the input layer. There is one neuron each for time and transience status, 13 neurons for friends' impact, and 52 neurons for the customer's historical decisions (four daily meals times 13 restaurant types). Thirteen neurons, one per restaurant type, are utilized in the output layer.

Weights, w_{ij}, on the edges between neurons are initialized randomly. Then, at each iteration, the weighted inputs to a given node in the hidden layer are summed to determine the total input to that node. The output is obtained by applying a sigmoid activation function to scale this value into the interval $[0,1]$. Thus, the final output of node j in the hidden layer is given by $\frac{1}{1+\exp(-\sum_{i=1}^{67} w_{ij}x_i)}$. Similarly, the outputs of the hidden layer are fed to the output layer and scaled by another sigmoid activation function to determine the output values. During training, the correct output neuron for a given training instance is given a target value of one while the other output neurons are given target values of zero. The squared error from the target value is then used to adjust the weights between all layers by the backpropagation algorithm [17]. During the testing phase, the output neuron receiving the largest activation (closest to one) is chosen as the restaurant prediction.

Different architectures for the configurations of hidden layers has been tried to reduce the predicting error. A series of experiments (3^3 experimental designs) are conducted to evaluate the effect of different parameters settings for neural network architectures as shown in Table 2. From this pilot study, it is determined that the prediction accuracy is not significantly affected by the number of neurons in the hidden layer or the number of hidden-layers, but the training time is highly affected. The learning rate contributes most to the accuracy. Finally the neural network is set as one hidden layer with 10 neurons, and learning rate at 0.01.

Table 2: 3^3 Experiment design

Factors/Level	Low	Medium	High
Number of hidden layers	1	3	10
Number of neurons in hidden layers	1	10	100
Learning rate	0.01	0.1	0.8

4. Experimental Results

For each user, a SVM and an ANN are used to compare the predicting result. Figure 6 shows the results of the different methods on ten sample users. The accuracy (average percentage of correct prediction) is calculated based on 5-fold cross validation. From the chart, it is clear that the neural network provides a higher accuracy for most users. SVM with linear kernel function does not provide as high accuracy as other non-linear kernels. If the user feature space has a high number of dimensions which are linearly independent of each other, a linear kernel function is not able to separate the features.

In order to test the methodology on the entire data set with the users who have sufficient historical preference information, users having more than 100 check-ins related to restaurant are selected. The first 50 check-ins are considered as the basic historical preference information at time 0, and training starts from the 51st check-in. As time elapses, new check-ins join the historical information set, and the user's social network evolves, which changes the Influence Factor in real time. The overall prediction accuracy is illustrated in Table 3. Although ANN provides a more accurate prediction, the training time takes longer than SVM. Figure 7 shows the stable prediction accuracy across the entire user data set.

5. Conclusion and Future Work

In this paper, an ANN is utilized for customer behavior prediction based on social media data. With 5-fold cross-validation, the ANN provides 93.13% accuracy consistently, while support vector machine with a sigmoid function only gives 54.00% accuracy. This research proposes a novel way of using the information users receive from their social network. The high level of accuracy illustrates that this information plays an important role in the prediction

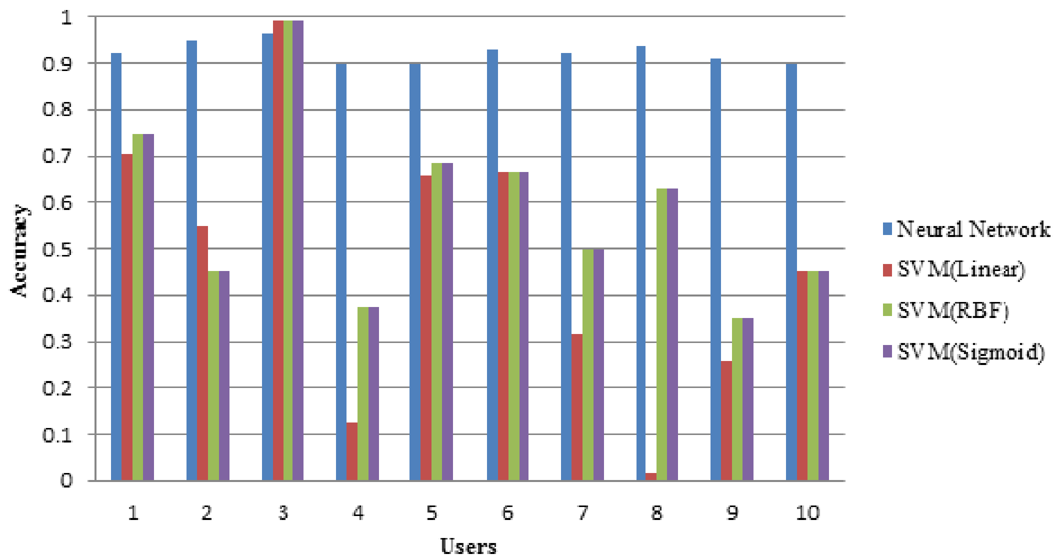


Figure 6: Prediction accuracy for ten sample users

Table 3: Prediction accuracy comparison

Methodology	Accuracy
ANN	93.13%
SVM with sigmoid function	54.00%
SVM with radial basis function	53.89%

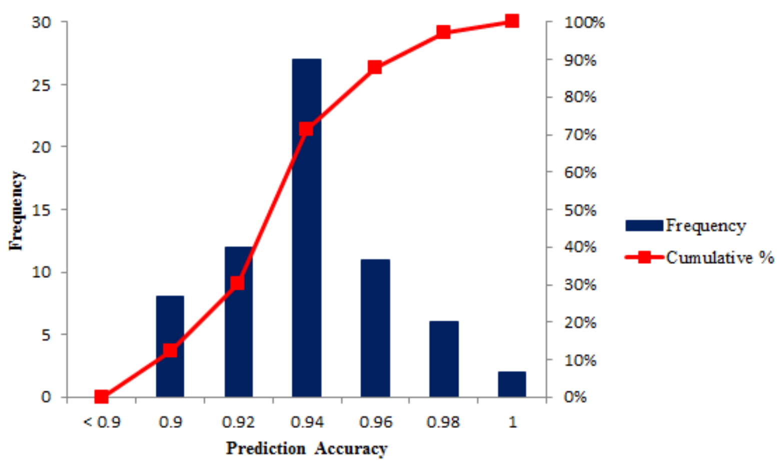


Figure 7: ANN prediction accuracy histogram

model. It has also been seen that, as the social network evolves and increases, the ANN takes longer to train and gather information from the social network. Utilizing distributed processing techniques to speed up the training is one possibility for future work. To further improve the prediction accuracy and utilize more of the available information, a sentiment analysis based on the text of check-ins is another direction for future research. Currently, the user preferences are calculated based on the number of check-ins in the proposed model. From the text of check-ins, the opinions of users can be identified, which may present their preferences more accurately. The influence factor is another path that can be further developed in the future. In the current model, all historical decisions made by a customer's immediate social circle are treated equally. By considering the elapsed time since each decision, a decay factor can be introduced to give more weight to the more recent activities. Not only is predicting customer behavior important, but also stimulating the customer purchase behavior is crucial for retailers. The giant component of the network has more power to spread information to their social network but it does not mean it will influence the customer effectively. Combining the weighted graph of the social network with the influence factor and giant component analysis could be a potential tool of marketing analysis.

References

[1] Su, X., and Shen, Z.M., 2007, "Customer Behavior Modeling in Revenue Management and Auctions: A Review and New Research Opportunities," Production and Operations Management, 16(6), 713-728.

[2] McDonald, D.W., and Ackerman, M.S., 2000, "Expertise Recommender: A Flexible Recommendation System and Architecture," Proc. of the 2000 ACM Conference on Computer Supported Cooperative Work, Philadelphia, Pennsylvania, December 2-6, 231-240.

[3] Kim, E., Kim, W., and Lee, Y., 2002, "Combination of Multiple Classifiers for the Customer's Purchase Behavior Prediction," Decision Support Systems, 34(2), 167-175.

[4] Cho, Y.B., Cho, Y.H., and Kim, S.H., 2005, "Mining Changes in Customer Buying Behavior for Collaborative Recommendations," Expert Systems with Applications, 28(2), 359-369.

[5] Yan, L., Miller, D.J., Mozer, M.C., and Wolniewicz, R., 2001, "Improving Prediction of Customer Behavior in Nonstationary Environments," Proc. of International Joint Conference on Neural Networks, July 15-19, Washington DC, 2258-2263.

[6] Yan, L., Fassino, M., and Baldasare, P., 2005, "Predicting Customer Behavior via Calling Links," Proc. of International Joint Conference Neural Networks on Components, July 31- August 4, Montreal, Quebec, 2555-2560.

[7] Xie, Y., Li, X., Ngai, E.W.T., and Ying, W., 2009, "Customer Churn Prediction Using Improved Balanced Random Forests," Expert Systems with Applications, 36(3), 5445-5449.

[8] Cheng, Z., Caverlee, J., Lee, K., and Sui, D., 2011, "Exploring Millions of Footprints in Location Sharing Services," Proc. of the Fifth International AAAI Conference on Weblogs and Social Media, July 17-21, Barcelona, Spain, 81-88.

[9] Bawa-Cavia, A., 2011, "Sensing the urban: Using Location-based Social Network Data in Urban Analysis," Proc. of The First Workshop on Pervasive Urban Applications, June 12-15, San Francisco, California, 1-7.

[10] Phithakkitnukoon, S., and Olivier, P., 2011, "Sensing Urban Social Geography using Online Social Networking Data," Proc. of Fifth International AAAI Conference on Weblogs and Social Media, July 17-21, Barcelona, Spain, 36-39.

[11] Li, N., and Chen, G., 2009, "Analysis of a Location-based Social Network," Proc. of the 2009 International Conference on Computational Science and Engineering, August 29-31, Vancouver, Canada, 263-270.

[12] Cho, E., Myers, S., and Leskovec, J., 2011, "Friendship and Mobility: User Movement in Location-based Social Networks," Proc. of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, August 21-24, San Diego, California, 1082-1090.

[13] Park, M., Hong, J., and Cho, S., 2007, "Location-based Recommendation System Using Bayesian User's Preference Model in Mobile Devices," Proc. of 4th International Conference on Ubiquitous Intelligence and Computing, Hong Kong, July 11-13, 1130-1139.

[14] Ho, I.F., Thompson, K., Gnanasambandam, N., Lam, S., and Yoon, S.W., 2012, "Mining Transient vs. Non-Transient User Movement Behavior in Urban Environments," Proc. of the Industrial and Systems Engineering Research Conference, May 18-22, Orlando, FL.

[15] Idicula-Thomas, S., Kulkarni, A., Kulkarni, B., Jayaraman, V., and Balaji, P., 2006, "A Support Vector Machine-based Method for Predicting the Propensity of A Protein to be Soluble or to Form Inclusion Body on Overexpression in Escherichia coli," Bioinformatics, 22(3), 278-284.

[16] Cortes, C., and Vapnik, V., 1995, "Support-vector Networks," Machine Learning, 20(3), 273-297.

[17] Rumelhart, D.E., Hinton, G.E., and Williams, R.J., 1986, "Learning Representations by Back-propagating Errors," Nature, 323(6088), 533-553.