2014 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)
July 6-11, 2014, Beijing, China

Prediction of Online Trade Growth Using Search-ANFIS: Transactions on Taobao as Examples

Wang Jiyuan
School of Management
University of
Chinese Academy of Sciences
Beijing 100190
Email: wangjiyuan12@mails.ucas.ac.cn

Peng Geng
School of Management
University of
Chinese Academy of Sciences
Beijing 100190
Email: penggeng@ucas.ac.cn

Dai Wei
School of Management
University of
Chinese Academy of Sciences
Beijing 100190
Email: daiwei12@mails.ucas.ac.cn

***Abstract*—The growth of E-commerce which can be seen in recent years, has contributed a lot to global economy. Prediction of trade, especially in C2C market, can help decision-makers obtain the information from the online transactions and find the knowledge underlying the data. This paper facilities the traditional search index prediction system with ANFIS model. By using purchasing transactions from Taobao, a C2C company in China, this paper trains and tests the model. Results show that, compared with traditional regression analysis method, Search-ANFIS system has higher prediction accuracy in online trade prediction.**

I. Introduction

E-Commerce has contributed a lot to global economy, since it accelerated the product and service exchange rate. Especially in China, E-Commerce has help plenty of corporations and individuals benefit from the long tail economy. C2C, one type of popular commercial models, is defined as trades between consumers directly, short for Consumer to Consumer. Taobao, a C2C platform owned by Alibaba, is a typical Chinese E-Commerce company, for example, makes trading sales more than 400 billion possible. The reason why Taobao can make such progress is mainly because it helps consumers search and compare, then make choices. It's of great importance to find what people care about during certain periods. Search queries consumers left make it easier to realize. To some degree, search queries show the demands of markets because people searches for things on search engines or other websites only when he or she wants to find something or make choices[1]-[8]. The prediction itself is not just for data collection usage. The predicting process embedded in the system involves many methods, varying from traditional time series analysis to sequence mining in data mining domain. Fuzzy models, especially Takagi-Sugeno and related models, are quite popular in prediction domain[13]-[22]. But no research focused on the fuzzy-based search index prediction. This paper will introduce such a model called Search-ANFIS model, aiming at combining the new idea with fuzzy model to make predictions and reveal knowledge underlying the data. The reminder of this paper is organized as follows: Section II gives a brief survey of idea of using search index and the typical fuzzy models can be used for predicting. A general Search-ANFIS model will be given in Section III. Section IV demonstrates the results applying Taobao transactions. Some discussions will be made in Section V and Section VI presents summary and conclusions.

II. Related work

The idea of using search queries to make predictions first applied by Ginsberg et al. [1], which demonstrated the online search queries can be applied to the detection of American flu. Two weeks ahead of the prediction made by CDC (Centers for Disease Control and Prevention, America) are showed by the new method. Jurgen A. Doornik then combine the log data and with search data to improve the prediction accuracy, verifying the idea furthermore[2]. Carneiro et al. find that using google flu trend can detect regional outbreaks of influenza before CDC[8]. Brownstein et al. surveys the improvement of public healthcare facilitated with information technology, and in this paper they referred that google search frequency data has help researchers and healthcare people to discover more[6]. Researches on economic behaviors were conducted afterwards. Hyunyoung Choi and Hal Varian improves the AR model by appending searching index as a independent variable, giving the model a higher prediction accuracy[3]. Lynn Wu and Erik Brynjolfsson investigated the strong relationship between housing search index and American housing trade and price[4]. Askitas et al. set up the relationship between the search data and the unemployment rate, demonstrating a few inspiring instructions[5]. Da Zhi and his colleagues find a new proxy of investors' attention using Google search frequency, and they find that an increase in search volume index will yield higher stock prices[7]. Fuzzy theory is first proposed by Zadeh in 1965, the most important idea is that fuzzy sets consist of classes of variables with grades of membership functions[9]. Takagi and Sugeno then proposed a model which can interpret the results of operations of fuzzy variables into a crisp value using defuzzilization[10]. Jang then come up a new model based on the concept of Takagi-Sugeno model, which is ANFIS: adaptive-network-based fuzzy inference system. He combine T-S model with neural network to make fitness and prediction. Our work is mainly based on this model[11]. The three papers above are the fundamental work in fuzzy forecasting domain. Afterwards, many improvements on the construction of the model and the learning algorithms are proposed. In the application field, several researches were conducted, especially using ANFIS model. In recent years, many papers are published on related methodology topics. For example, Lin, et al. propose a mixed model combining support vector machine with fuzzy neural network to make classifications, and experimental results show that the proposed SVFNN model for pattern classification can achieve good classification performance with less number of fuzzy kernel functions[12]. W. Abdel-Hamid, A. Noureldin and N. El-Sheimy present an auxiliary fuzzy-based model for predicting the KF positioning error states during GPS signal outages[13]. Abbas Khosravi, Saeid Nahavandi, and Doug Creighton design a methodology to adapt the delta technique for the construction of PIs for outcomes of the ANFIS models, which turns out to be of better quality by using proposed optimization algorithm[14]. Mehrnoosh Davanipoor et al. propose an accelerated hybrid learning algorithm for training of fuzzy wavelet neural network, and simulation results prove that the proposed algorithm has a superior convergence speed[15]. Min Liu, Mingyu Dong, and Cheng Wu proposes a new Adaptive-Network-based Fuzzy Inference System (ANFIS)-based parameter prediction method to make parameter predictions[16]. J.P. SCatalao, H.M.I. Pousinho and V.M.F. Mendes combine wavelet transform, particle swarm optimization, and adaptive-network-based fuzzy inference system to make short-term electricity prices forecasting[17]. Moreover, M. Hanmandlu and B.K. Chauhan propose two models derived from wavelet fuzzy neural network and Choquet integral fuzzy neural network to make predictions on Indian utility data and they perform the conventional model[18]. J.P.S. Catalao et al. propose a novel hybrid approach, which combines wavelet transform, particle swarm optimization and an adaptive-network-based fuzzy inference system, to predict short-term wind power in Portugal[19]. Melek Acar Boyacioglu and Derya Avci utilize ANFIS to predict the stock market return. They predict the Istanbul stock exchange and find that ANFIS can be a useful tool for economists and practitioners dealing with the forecasting of the stock price index return[20]. Leila Naderloo and et al. apply ANFIS to the prediction of grain yield of irrigated wheat at a province in Iran, which show that the ANFIS has a good prediction accuracy on this topic[21]. Mahmut Firat and et al. successfully predict the water consumption by using ANFIS and MFIS. Many papers focus on fuzzy prediction and E-commerce prediction individually, but no one employed the ANFIS fuzzy model to make predictions on C2C trade. This paper aimed at coming up a such model to facilitate the conventional prediction system with fuzzy method and to help increase E-Commerce trade prediction accuracy and provide domain insights and knowledge.

III. Search-ANFIS prediction system

When consumers online purchase certain products, the E-Commerce system will record the transactions and write the events into log files. Figure. 2 gives the basic operations during the whole trade. When a consumer want to pay for certain products or services, the system will deal with its requests and recommend some related goods or services according to the consumer's preferences. Because the front page of an E-Commerce company cannot have too much products information on it, a tiny search engine is needed. Therefore Search queries and log file will be generated. Once the purchase finished, trading transactions will also be generated. Then people can collect the search queries for further study.

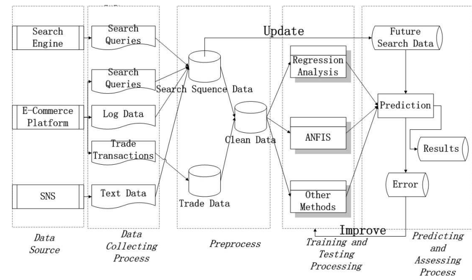

Fig. 1. Search-ANFIS Prediction system Framework

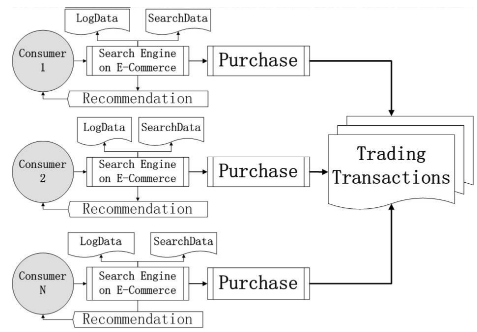

Fig. 2. Transaction Flow Diagram of E-Commerce Trades

A. System Framework

The proposed prediction system is presented in Figure. 1. The system mainly consists of 5 parts, and each of them are: Collection process part, Preprocess part, Training and Testing process part, Predicting process part and Assessment process part, respectively. The functional introduction and algorithms each part contains are described in following subsections.

B. Data Collecting Process

Generally, three types of data are popular with Internet researchers in recent: search data from search engine like google, trading and searching Data from E-Commerce websites, and text tweets from SNS. The first one is quite popular as showed in section II. Text mining is a hot topic in data mining domain, especially the text data on SNS like Tweeter, so the third one is also drawing many attention. In this paper, the second type will be chosen since it's of great value for E-Commerce companies to use its own search and trade data to make predictions. Besides, not too many papers focus on the second type. This research collects the key words which consumers typed into the search area. All the key words in the collection are well classified by Taobao. We choose three categories: 3C, clothes and makeups. Since all the key words consumers left are classified by Taobao and the number is too large, it's necessary to choose part of them from all the key words.

C. Preprocess

Preprocess mainly consists of three parts: The first one is dealing with missing data, wrong data etc. The second is to operate dimensionality reduction. The last one is to find the time difference between the search key words series and the trading series. Key words series collected with missing values and wrong values are common to see because too many uncertainties underlying the trades. There are two main reasons may lead to such phenomenon: system maintenance and human factors. We delete the series which have too many missing value or with too many zeros. This paper also choose a unified dates series begin and end. After this part of preprocess, an M by N key words matrix and an M by 1 trading vector are obtained. Commonly, making dimensionality reduction in multi-variable series applies the idea of compounding the key words series. This idea is also accepted in this paper. Generally, three compounding methods are employed as follows.

1) Direct Compounding: When investigating the relationship between key words of unemployment and unemployment rate, Askitas and Zimmermann used this method. They collected 4 categories of unemployment related key words and summed them up respectively, and the model achieved a good fitting effect. Note X_j^i as the *j*th search number of the *i*th key words series, *i*=1,2,. . . ,N and *j*=1,2,. . . ,M. N is the total number of key words and M is the duration. Then the index $Index_j$ all the key words series summed up represents the motivation of consumers, an instrument variable, in other words.

$$Index_j = \sum_{i=1}^{N} X_j^i \tag{1}$$

2) Correlated Compounding: Direct compounding method is just a basic idea of dimensionality reduction among a number of key words. Ginsberg and his colleagues adopted an other kind of method called Correlated Compounding method.

$$Index_j^* = \sum_{i=1}^{N} X_j^{i*}, j = 1, 2, \ldots, M. \tag{2}$$

Different from Direct Compounding, Correlated Compounding methods choose the key word sequence X_j^{i*} which has the largest correlation with the dependent variable in each summing iteration. Another difference between the two methods is that not all the key words are added into the index sequence in Correlated Compounding method.

3) Lag Correlated Compounding: This paper proposes a new method called Lag Correlated Compounding. The idea of this method is that each key word sequence may have different predicting ability on the sequence of dependent variable. At the beginning of compounding, correlations are first calculated for each key word sequence with the sequence of dependent variable. The key words with highest correlation with trade sequence are added to the index sequence. This rule will also be applied in the next few iterations.

$$Index_j^* = \sum_{i=1}^{N} X_{j-t(i)}^{i*}, j = 1, 2, \ldots, M. \tag{3}$$

where $X_{j-t(i)}^{i*}$ means *i*th key word with *t*th delay and the highest correlation with the sequence of dependent variable is obtained. In this paper, stepwise is accepted to implement this method.

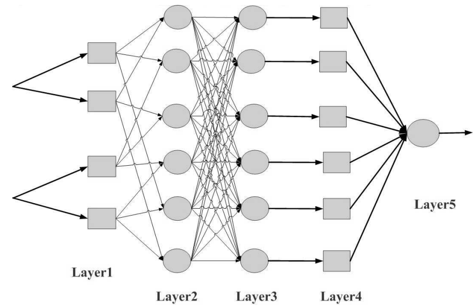

Fig. 3. A Sample Architecture of ANFIS with 2 Inputs and 6 Rules

D. Training Process

An Adaptive-Network-Based Fuzzy Inference model is employed in this paper. Before introducing the model, the paper first checks the stationary of the two series: the index sequence and the trade sequence. The statistical instrument in this paper is Adjusted Dicky-Fuller stationary test. Once stationary test failed to pass, some simple linear regression models might be useless. Then the proposed model will be employed. ANFIS was first introduced in 1993, which combines Takagi-Sugeno model with Artificial Neural Network. Generally, it has 5 layers whose nodes represent several operators. The overall structure is presented in Figure. 3. Easily to be seen, hybrid learning algorithms is more common be accepted. The functions and operations in each layer are demonstrated in following subsections.

1) The First Layer: It can be seen in Figure. 3, Layer-1 consists of nodes representing the fuzzy sets of each input variable. In this paper, the index series is chosen as the input variable. Crisp or fuzzy value O_i^1 of these nodes are defined as following formula.

$$O_i^1 = e^{(-(x_i - c_i)/a_i)^2}, i = 1, 2, \ldots, m. \tag{4}$$

$$O_i^1 = e^{(-(y_i - c_i)/a_i)^2}, i = m, m+1, \ldots, m+n. \tag{5}$$

Where x_i and y_i note the element of input sequence x and y, respectively. And A_i and B_i represent the fuzzy sets, mapping from a certain universe to the unit interval. a_i and c_i are premise parameters of the fuzzy inference system.

2) The Second Layer: Each node in the second layer contains a product operator. Two signals from Layer-1 will be processed on this node. O_i^2 represents the product of the two signals.

$$O_i^2 = O_l^1 \cdot O_k^1, i = 1, 2, \ldots, p.l, k = 1, 2, \ldots, m.l \neq k. \tag{6}$$

Where *p* is the number of nodes in Layer-2.Obviously, *m* equals *p(p-1)/2*.

3) The Third Layer: Every node in this layer functions as a normalized unit. O_i^3 denotes the *i*th node's output.

$$O_i^3 = O_i^2/(\sum_{k=1}^{p} O_k^2) \tag{7}$$

4) The Forth Layer: In this layer, product of normalized weight and consequent function will be obtained. Also in this layer, combination of premise and consequent parameters will occur in the whole architecture of the inference system.

$$O_i^4 = O_i^3 \cdot f_i \tag{8}$$

Where O_i^3 functions as normalized weight from Layer-3, and f_i is the linear function of input variables.

5) The Fifth Layer: The final layer of ANFIS architecture has only one node which is the overall output of the system. The following equation illustrates the operation.

$$O^5 = (\sum_{j=1}^{p} O_j^4)/(\sum_{j=1}^{p} O_j^3) \tag{9}$$

Details of the architecture vary among many cases, while fundamental structure keeps almost the same, however.

E. Predicting and Assessing Process

Generally, data sets are divided into 2 subsets: one for training and another for testing or predicting. In the training process, parameters of ANFIS will be changed iteratively to the direction of the minimum value of object function. Finally, parameters in premise and consequent part of rules in ANFIS will used to make prediction. Performance will be evaluated by RMSE, denoted as following.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(Actual - Predict)^2}{N}} \tag{10}$$

IV. Results

In this section, several experiment results will be demonstrated. Firstly, the structure and examples of the raw data are given. Then compounding index will be generated using Lag Correlated Compounding idea introduced in previous section. As a comparison, results of Regression Analysis will be presented. Some further explanation of the effects of search queries on prediction of trades will be discussed also. Finally, the predicting results of Search-ANFIS will be given and comparisons are also listed.

A. Data sets

Data sets used are from www.taobao.com. All the transactions are categorized as three groups which are 3C (short for standard products of Computer, Communication and Consumer Electronic), clothes and makeups, respectively. The data sets are structured as *key words consumers searched*: *frequency all the consumers searched*. The full duration is 90 days and the volume of sample is 89 because the existence of a server maintaining day. Day-on-day changes are calculated for prediction. We divide the sample into two subsets: the training set and the testing set. The training set consists of 74 historical samples and 7 samples in testing subsets are used for assessing the prediction accuracy.

TABLE I. Compounding Index: Three Categories

Date	**3C Index**	**Makeups Index**	**Clothes Index**
2011-09-21	3011	2199	1424
2011-09-22	3078	2126	1591
2011-09-23	2798	2127	1508
2011-09-24	2840	2109	1407
2011-09-25	2735	2075	1386
2011-09-26	2905	2227	1420
2011-09-27	2904	2242	1430

TABLE II. Trade Growth Ratio: Three Categories

Date	**3C growth**	**Makeups growth**	**Clothes growth**
2011-09-21	0.932555346	0.969188304	1.059289041
2011-09-22	0.949613621	0.936053359	0.884126416
2011-09-23	1.039660466	1.164335895	1.136113096
2011-09-24	0.828192399	0.812403197	0.781684352
2011-09-25	0.900191625	0.935172549	1.009738854
2011-09-26	0.925327459	0.926654383	1.200048769
2011-09-27	0.950463292	0.918136218	0.946831912

B. Preprocessing

Firstly, key words series with complete duration among the three categories will be selected, that is, some key word series with many missing values will be dropped. Those series with one or two missing values will be smoothed by two neighbored values. Then dimension reduction will be employed and the method in this paper is Lag Compounding Method. The first key word sequence accounted is the sequence has biggest correlation coefficient with trade sequence. Afterwards, new series are added to the prior index sequence, iteratively. Once the Covariance no longer appears to increase, the process should be terminated. Then algorithm marked the key words who were added. Testing sets of the Compound indexes and day-on-day ratios are demonstrated in Table. 1 and Table. 2. Moreover, selected key words are also provided in Table. 3.

C. Benchmark Model: Regression Analysis

To compare with the proposed model, a benchmark model is raised. The concrete expression of the model is shown in equation(11)to equation(13), R_i^{3c}, R_i^{cloth}, R_i^{makeup}, I_i^{3c}, I_i^{cloth}, I_i^{makeup} denote 3C Ratio, Clothes Ratio, Makeup Ratio, 3C Index, Clothes Index and Makeup Index respectively. Before the regression results, stationary test are conducted and the results are demonstrated in Table. 4.

$$R_i^{3c} = \beta + \beta_1 \Delta I_i^{3c} + \beta_2 R_{i-7}^{3c} \tag{11}$$

$$R_i^{cloth} = \beta + \beta_1 \Delta I_i^{cloth} + \beta_2 R_{i-7}^{cloth} \tag{12}$$

$$R_i^{makeup} = \beta + \beta_1 \Delta I_i^{makeup} + \beta_2 R_{i-7}^{makeup} \tag{13}$$

The extended Dickey - Fuller test method (referred to as the ADF test) are adapted in this paper, the null hypothesis is that there is at least one unit root for the series, and it means the

TABLE III. SELECTED KEY WORDS

No	3C	Makeups	Clothes
1	欧珀莱正品	笔记本	优衣库
2	mg 面膜正品	手机	香奈儿
3	化妆品正品	5830	雅莹
4	代购	ipad1 专柜	正品
5	欧泊莱正品专柜	苹果	网纱
6	美丽加芬	ipod	制服
7	按摩膏	羽博	显瘦
8	玫瑰精油	14寸	正装
9	水之印	移动电源	职业套装
10	lancome	3d 眼镜	抹胸裙
11	芙丽芳丝	3d眼镜	依迪菲
12	贝佳斯绿泥	3g	纤牌
13	丁家宜	3gs	维多利亚的秘密
14	按摩膏	相机包	小魔鱼
15	日本	相机	原单裙
16	代购	电脑包	白衬衫
17	化妆品正品	手机电池	樱桃
18	olay	g7	意大利原单
19	新生活化妆品正品	笔记本电池	演出服
20	克丽缇娜	床	原单正品
21	里美	3g无线上网卡	扎染
22	按摩膏	无线上网卡	23区
23	碧柔	c7	逸阳女裤
24	男士	defy 手机套	字母
25	贝佳斯绿泥	e63 手机套	艺域

TABLE IV. STATIONARY TEST

(ADF, p-value)	Raw Data	First Order Lag
3C index	(0.179017, 0.9695)	(-9.675709, 0.000)
3C ratio	(-8.69299, 0.0000)	
Makeups index	(0.508405, 0.9860)	(-9.645364, 0.000)
Makeups ratio	(-8.881587, 0.000)	
Clothes index	(1.125438,0.9974)	(-8.226410, 0.000)
Clothes ratio	(-9.739019, 0.000)	

series is not stable. It can be concluded that the original series of three index variables are not stable, but the first difference of them are stable at the 1% significance for rejecting the original hypothesis, indicating that they are all first order difference stationary series. And the original series of three ratio variables are stable.

The results of the model are listed in Table. 5. RMSE is employed to measure the accuracy of the model.

D. Search-ANFIS

After the benchmark model, the results of the proposed model are listed in Table. 7 and the basic parameters can be seen in Table. 6. Parameters are chosen depending on the best model performance.

E. Comparison

Compared with benchmark model, Search-ANFIS is more accurate and stable. The Clothes Category is the hardest one to predict by the benchmark model, while in the proposed model, it has been predicted in a acceptable error range. In other categories, little difference was found between two models because it's the limits of the forecasting ability of the query series. It's quite normal that in Search-ANFIS model,

TABLE V. BENCHMARK MODEL:MULTIVARIABLES REGRESSION

3C ratio	Actual	Predict	Square Error
2011-09-21	0.9326	0.9661	0.0011
2011-09-22	0.9496	0.9812	0.0010
2011-09-23	1.0397	0.8881	0.0230
2011-09-24	0.8282	0.9581	0.0169
2011-09-25	0.9002	0.9045	**0.0000**
2011-09-26	0.9253	0.9920	0.0044
2011-09-27	0.9505	0.9800	0.0009
RMSE	**0.0822**		
Makeups ratio	**Actual**	**Predict**	**Square Error**
2011-09-21	0.9692	1.0253	0.0031
2011-09-22	0.9361	0.9346	**0.0000**
2011-09-23	1.1643	0.9654	0.0396
2011-09-24	0.8124	0.9582	0.0213
2011-09-25	0.9352	0.9287	0.0000
2011-09-26	0.9267	1.0249	0.0096
2011-09-27	0.9181	0.9851	0.0045
RMSE	**0.1057**		
Clothes ratio	**Actual**	**Predict**	**Square Error**
2011-09-21	1.0593	0.9365	0.0151
2011-09-22	0.8841	1.1866	0.0914
2011-09-23	1.1361	0.9059	0.0530
2011-09-24	0.7817	0.8962	0.0131
2011-09-25	1.0097	0.9353	**0.0055**
2011-09-26	1.2000	1.0234	0.0312
2011-09-27	0.9468	1.2218	0.0756
RMSE	**0.2018**		

TABLE VI. PARAMETERS AND METHODS EMBEDDED IN THE MODEL

Category	MF	MFNo	Learning	Iterations	Training Fitness
3C ratio	Gauss	3	hybrid	280	0.0386
Makeups ratio	Gauss	3	hybrid	430	0.0582
Clothes ratio	Gauss	5	hybrid	2500	0.0431

the accuracy of training sets are higher than those in testing sets, which shows the nonexistence of overfitting. Besides, this results also illustrated the relatively stability of the system.

V. CONCLUSIONS

In this paper, a new idea was first referred about online trade prediction by introducing a few studies across various domains. Then Papers about fuzzy models were referred, and the basic concept of our model was proposed in that section too. Afterwards,the framework of the prediction model was given, and each parts of the system were also described. In the empirical study section, a benchmark model which is a traditional time series model was demonstrated. Then we presented the results of our model which turned out to be more accurate and stable than the benchmark one. However, the model proposed also has some shortcomings, for example, the scalability of the model is limited by the data source, and it may be solved by integrating more related potential sources like social network texts. Besides, the overfitting problem of the neural network also needs improving in ultrahigh dimensional modeling.

ACKNOWLEDGMENT

This research is supported by the National Natural Science Foundation of China under Grant 71202115、71172199 and

TABLE VII. SEARCH-ANFIS MODEL:TESTING SET

3C ratio	Actual	Predict	Square Error
2011-09-21	0.9326	0.9233	**0.0001**
2011-09-22	0.9496	0.9330	0.0003
2011-09-23	1.0397	0.8475	0.0369
2011-09-24	0.8282	0.8560	0.0008
2011-09-25	0.9002	0.8412	0.0035
2011-09-26	0.9253	0.8827	0.0018
2011-09-27	0.9505	0.8822	0.0046
RMSE	**0.0828**		
Makeups ratio	**Actual**	**Predict**	**Square Error**
2011-09-21	0.9692	0.9819	**0.0002**
2011-09-22	0.9361	0.9778	0.0017
2011-09-23	1.1643	0.9779	0.0347
2011-09-24	0.8124	0.9768	0.0270
2011-09-25	0.9352	0.9765	0.0017
2011-09-26	0.9266	0.9850	0.0034
2011-09-27	0.9181	0.9890	0.0050
RMSE	**0.1027**		
Clothes ratio	**Actual**	**Predict**	**Square Error**
2011-09-21	1.0591	1.0939	0.0012
2011-09-22	0.8841	0.8938	**0.0001**
2011-09-23	1.1361	1.3406	0.0418
2011-09-24	0.7817	0.9100	0.0165
2011-09-25	1.0097	1.0265	0.0003
2011-09-26	1.2000	1.0722	0.0163
2011-09-27	0.9468	0.9983	0.0027
RMSE	**0.1061**		

71203218. Bedsides, reviewers' comments are of great value and we appreciate them all.

REFERENCES

[1] J. Ginsberg, M.H. Mohebbi, R.S.Patel, L.Bramme, M.S. Smolinski and L. Brilliant,"Detecting influenza epidemics using search engine query data," *Nature*, vol. 457, pp. 1012-1014, 2009.

[2] J.A. Doornik, "Improving the Timeliness of Data on Influenza-like Illnesses using Google Search Data," *Working paper*, 2009.

[3] H. Choi, H. Varian,"Predicting the Present with Google Trends," *Economic Record*, vol. 88, pp. 2-9, 2012.

[4] W. Lynn, E. Brynjolfsson,"The Future of Prediction: How Google Searches Foreshadow Housing Prices and Sales," *Economics of Digitization*, University of Chicago Press, 2013.

[5] N. Askitas, KF.Zimmermann,"Google Econometrics and Unemployment Forecasting," *German Council for Social and Economic Data (RatSWD) Research Notes*, vol. 41, 2009.

[6] J.S. Brownstein, C.C. Freifeld and L.C. Madoff,"Digital disease detection－harnessing the Web for public health surveillance," *New England Journal of Medicine*, vol. 360(21), pp. 2153-2157, 2009.

[7] Z. Da, J. Engelberg and P. Gao,"In search of attention," *The Journal of Finance*, vol. 66(5), pp. 1461-1499, 2011.

[8] H.A. Carneiro, E. Mylonakis,"Google trends: a web-based tool for real-time surveillance of disease outbreaks," *Clinical infectious diseases*, vol. 49(10), pp. 1557-1564, 2009.

[9] L.A. Zadeh,"Fuzzy sets," *Information and control*, vol. 8(3), pp. 338-353, 1965.

[10] T. Takagi, M. Sugeno,"Fuzzy identification of systems and its applications to modeling and control," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 1, pp. 116-132, 1985.

[11] J.S.R. Jang,"ANFIS: adaptive-network-based fuzzy inference system," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 23(3), pp. 665-685, 1993.

[12] C.T. Lin, C. M. Yeh, S. F. Liang, et al,"Support-vector-based fuzzy neural network for pattern classification," *IEEE Transactions on Fuzzy Systems*, vol. 14(1), pp. 31-41, 2006.

[13] W. Abdel-Hamid, A. Noureldin and N. El-Sheimy,"Adaptive fuzzy prediction of low-cost inertial-based positioning errors," *IEEE Transactions on Fuzzy Systems*, vol. 15(3), pp. 519-529, 2007.

[14] A. Khosravi, S. Nahavandi and D. Creighton,"Prediction interval construction and optimization for adaptive neurofuzzy inference systems," *IEEE Transactions on Fuzzy Systems*, vol. 19(5), pp. 983-988, 2011.

[15] M. Davanipoor, M. Zekri, F. Sheikholeslam,"Fuzzy wavelet neural network with an accelerated hybrid learning algorithm," *IEEE Transactions on Fuzzy Systems*, vol. 20(3), pp. 463-470, 2012.

[16] M. Liu, M. Dong and C. Wu,"A new ANFIS for parameter prediction with numeric and categorical inputs," *IEEE Transactions on Automation Science and Engineering*, vol. 7(3), pp. 645-653, 2010.

[17] J.P. SCatalao, H.M.I. Pousinho and V.M.F. Mendes,"Hybrid wavelet-PSO-ANFIS approach for short-term electricity prices forecasting," *IEEE Transactions on Power Systems*, vol. 26(1), pp. 137-144, 2011.

[18] M. Hanmandlu, B.K. Chauhan,"Load forecasting using hybrid models," *IEEE Transactions on Power Systems*, vol. 26(1), pp. 20-29, 2011.

[19] J.P.S. Catalao, H.M.I. Pousinho and V.M.F. Mendes,"Hybrid wavelet-PSO-ANFIS approach for short-term wind power forecasting in Portugal," *IEEE Transactions on Sustainable Energy*, vol. 2(1), pp. 50-59, 2011.

[20] M.A. Boyacioglu, D. Avci,"An adaptive network-based fuzzy inference system (ANFIS) for the prediction of stock market return: the case of the Istanbul stock exchange," *Expert Systems with Applications*, vol. 37(12), pp. 7908-7912, 2010.

[21] L. Naderloo, R. Alimardani, M. Omid, et al.,"Application of ANFIS to predict crop yield based on different energy inputs," *Measurement*, vol. 45(6), pp. 1406-1413, 2012.

[22] M. Firat, M.E. Turan, M.A. Yurdusev,"Comparative analysis of fuzzy inference systems for water consumption time series prediction," *Journal of hydrology*, vol. 374(3), pp. 235-241, 2009.