THE INTERNATIONAL JOURNAL OF BUSINESS & MANAGEMENT

Can Online User Behavior Improve the Performance of Sales Prediction in E-commerce?

Dr. Hassabelrasul Yousuf AL Tom Shihabeldeen
Associate Professor, KSA, Prince Sattam Bin Abdulaziz University,
College of Science & Humanities Studies, Saudi Arabia

Abstract:
How to forecast product sales effectively and efficiency in E-commerce is a significant task for E-commerce producers to manage product inventory and design marketing strategies. However, under the uncertainty of product demand, sales prediction is a complex task. This paper presents a novel data mining framework for sales prediction based on online user behavior data. Under the framework, the relationship of sales data and online user behavior data is well modelled, and the optimal lag of online user behavior data for sales prediction is also identified. In terms of evaluation criteria, a number of books are used for sales prediction. The empirical results show the efficiency and effectiveness of the proposed framework and also revealed that among different categories of books, the forecasting performance of some categories including Finance and Exam heavily relies on online user behavior information. So, it indicates that the proposed framework can be used as a potential alternative to analyze the sales trend, and help managers in Ecommerce companies for inventory optimization and customer relationship management.

Keywords: Sales prediction, User behavior, Data mining, Ecommerce

1. Introduction

With the rapid development of E-commerce, more and more people involve in online purchase. However, with the increase of product sales in E-commerce, the fluctuation of product sales brings critical risk to E-commerce companies. On one hand, if consumers demand more of the products than producers are prepared to supply, there are not enough products to provide for online users, and then it may lead to online user churn, while on the other hand, if the product supply is more than demand, the inventory cost may increase. So, how to forecast product sales effectively and efficiency in E-commerce is a significant task for E-commerce producers to manage product inventory and design marketing strategies. However, under the uncertainty of product demand, sales prediction is a complex task.

In recent years, sales prediction attracts much attention from E-commerce companies, and researchers. A great number of methods are proposed for sales prediction. Traditional univariate time series models have been proposed for the sales prediction. For example, Kuo et al. proposed a fuzzy neural network-based method for sales prediction in 1998 [1]. Similarly, Kuo proposed a sales forecasting system based on fuzzy neural network with initial weights generated by genetic algorithm in 2001 [2]. Furthermore, Yu et al. proposed a new extreme learning machine and traditional statistical model for sales prediction [3]. Some multivariate model with factors related to sales was also suggested. For instance, Luxhoj et al. suggested a hybrid econometric and neural network model for sales forecasting, and in the proposed method, the forecasts from each of the individual sub-models are then averaged to compute the hybrid forecast [4]. Frees et al. proposed a longitudinal data model for sales prediction [5], while Guo et al. offered a multivariate intelligent decision-making model for retail sales forecasting [6].

Recently, online user behavior is regarded as a useful resource for detection and prediction, such as influenza epidemics detection [7], finance market prediction [8], the unemployment rate prediction [9], and sales prediction [10]. These models/methods employed econometrics models to analyze the relationship between the predictive variable and online user behavior time series. Different from the previous studies, a data mining framework using data mining tools has been used for prediction purpose with online user behavior data, and the experimental results show that the proposed method outperforms the traditional econometrics model [11]. Since data mining techniques can make a significant contribution to the complex prediction problems, in this paper, a data mining framework using online user behavior for the sales prediction is presented, and within the proposed framework, various data mining tools are validated and compared to examine the efficiency and effectiveness of the proposed framework. In the proposed framework, some online user behavior, such as search behavior and browse behavior are firstly considered, and the detailed index are constructed. Secondly, different data mining tools are employed to describe the relationship between the sales data and the online user behavior data. Thirdly, an optimal data mining model is selected as the predictor by using the cross-validation method. Finally, the selected predictor with proper parameter and best feature subset is used to forecast sales trend.

The rest of this paper is organized as follows. The next section describes the theoretical foundation on online customer behavior. The data mining framework using online user behavior is proposed for the sales prediction in Section 3. For illustration, the efficiency of the proposed framework and empirical analysis of sale prediction using the data mining tools are reported in Section 4. Finally, conclusions and future research directions are summarized in Section 5.

2. Theoretical Foundation

To modern consumers, the dominant problem is too many choices for them, called consumer
, which weakens the capability of smart choosing [12]. A consumer realizes he needs to buy one thing and then completes the goal through a sequence of procedures: 1) problem recognition; 2) information search; 3) appraisal of plans; 4) product selection, which is called the decision-making. And the result of the process can influence the similar decision in next time [13]. Here information search is to "search" the proper data to establish the reasonable decisions. And the amount of information using the search engine from the internet is the double of the information without searching [14]. Information about the products are from two types: internal search from one's memory and external search from the advertising, friends or the observations of others [15]. In fact, the online shopping highlights the problem of excessed choices and the volume of information. Bloch et al. has raised the conception of ongoing search and indicated the ongoing search is closely related with the product involvement [16]. And Stigler has proposed a theory of search on the hypothesis that a consumer can choose the best solution in a set of alternatives through the searching [17]. In the report released by iResearch in 2008, it shows that the proportion of the search query is 57.5%, when consumers in online shopping search information of the commodities to compare the alternative solutions, and the search engine consists of two parts: one is the search engine of internal shopping web and the other is the common search engine, such as Google or Baidu. Some previous researches have pointed out that search is an "honest signal" [18,19]. Based on those mentioned above, the consumer search behavior is strongly linked to the decision-making and purchasing behavior. In fact, consumers hardly search related information rationally. Some consumers only search several alternatives typically, especially in the limited time [20,21]. It means that once consumers begin to search the key words about the commodity or enter the page, the commodity is a huge possibility to be selected. So, in the paper, two aspects of information search should be considered: the volume of searching in internal E-commerce website and the page view of the detail of the commodity.

3. An Online User Behavior-Based Sales Prediction

In this section, a novel online user behavior-based sales prediction method is proposed. Under the framework, some useful data mining tools such as neural networks and support vector regressions are employed to construct efficient sales prediction models. The proposed method and the modelling process with details are presented in the following subsections.

3.1. The Proposed Sales Prediction Framework

The online user behavior-based data mining framework for the sales prediction is presented, and the details is shown in Fig.1. Within the proposed framework, various data mining tools are validated and compared to examine the efficiency and effectiveness of the proposed framework. In the proposed framework, some online user behavior, such as search behavior and browse behavior, are firstly considered, and the detailed index are constructed. Secondly, different data mining tools are employed to describe the relationship between the sales data and the online user behavior data. When the data mining techniques are used, the optimal time lags (here the measuring unit of lag is the day) of different kinds of commodities are confirmed, and different influences of online user behavior information to different kinds of commodities are also discussed. Thirdly, an optimal data mining model is selected as the predictor by using the cross-validation method. Finally, the selected predictor with proper parameter and best feature subset is used to forecast sales trend.

3.2. The Modelling Process

In this subsection, two most widely used data mining tools including neural networks and support vector regressions are invoked to model the relationship between the online user behavior information and the sales data. As mention above, two kinds of NNs, BPNN and RBFNN are employed, while various SVRs with different kernels are used in the paper. In the modelling process, the general relationship between product sales data and online user behavior data is modelled, and the optimal lags of online user behavior are identified.
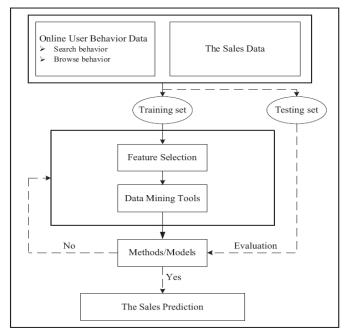
Figure 1: The Framework of the proposed method

Furthermore, sales prediction in different categories of products in is further analyzed and the effect of online user behavior on sales prediction in each category is exploited, and finally, the quantitative evaluation of the effect of online user behavior on sales prediction in different categories also discussed.

4. Empirical Analysis

4.1. Data description and evaluation criteria

All the data are real values from one of the biggest B2C websites in China, which provides the data of random 200 books from top 500 to top 1000 in the period from Feb. 2013 to May. 2013. As mentioned above, there are two types of data to be used to train and test the effectiveness of the models and the effect of the features:

- Online user behavior information: There are two factors inside the search information

o The volume of searching
o The page view of the detailed page

- Historical sales information

The historical sales data is considered in the experiments beside of the search information.

After eliminating the books with invalid data, for example, the volume of search is zero in most days, there are 65 books remained. In fact, there are 120 days in the data and the data of the former three months, which is from Feb. 2013 to Apr. 2013, is used to train the models while the data of the remained month, which is May. 2013, is used to test the models. While the daily data is used, the lag value is measured by day. In the paper, two kinds of indicators are selected as evaluation criteria in consideration of zero sales in some day. One is root-mean-square error, called RMSE, and the other is mean absolute difference, called MAE. The formulas are shown as follows:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(observed_i - predicted_i)^2}{n}} \qquad (1)$$

$$MAE = \frac{\sum_{i=1}^{n}|observed_i - predicted_i|}{n} \qquad (2)$$

Where i observed means the actual i th sales data, i predicted means i th sales prediction and n is the total number of predicting.

4.2. Experiment Results

Based on the designed experiments mentioned in Section 3, two types of models are employed to predict the sales data. For every model, there are two kinds of input: one is the historical sales data and the other adds the features, the volume of search and the page view, in addition of the sales data. Here it is worth noting that the final prediction we choose is the average of the predictions we got from the 50 times runnings of each NN model to eliminate the contingency.

1) Basic Experiments

In this part, two most widely used neural networks and two kinds of SVR are invoked to model the relationship between the search information and the sales. As mention above, two kinds of NN, BPNN and RBFNN are employed. And two kinds of SVR are used in the paper, and each kind selects three different kernel functions, which are Linear, RBF, Polynomial and Sigmoid function. Here we should notice that the SVR with RBF or Polynomial cannot fit the data well, so the results of those models are not presented. To compare the effectiveness of the features, different experiments with corresponding parameters are carried out. There are two groups of experiments. The first one is with the input of only historical sales data as marked "A". The other is with search information as mentioned in last part as marked "B". The experimental results based on the criteria MAE are shown in Table 1. As can be seen from Table 1, the results from the comparative experiments A using NNs, BPNN perform worse than RBFNN, but RBFNN needs more time to get the final prediction. Then the more important result is that the features about the search information can bring more accurate predictions in BPNN. Intuitively, BPNN with only historical sales data has worse results compared with the BPNN model with the addition of the volume of search and page view.

Number	Table Column Head		Lag						
	Name	*Detail*	*1*	*2*	*3*	*4*	*5*	*6*	*7*
A	NN	BPNN	13.0550	12.3863	13.5490	13.6213	14.3851	16.5995	17.8525
		RBFNN	11.0879	11.8428	12.6248	13.4774	13.6556	13.3258	14.2053
	$\varepsilon - SVR$	Linear	9.1605	9.8929	10.3075	11.0686	11.7482	11.6401	12.1548
		Sigmoid	11.8556	11.8259	11.7881	11.7350	11.7352	11.7330	11.7559
	$\nu - SVR$	Linear	10.3826	10.5852	10.6233	10.9237	11.4189	11.4776	12.0455
		Sigmoid	12.5856	12.6195	12.5545	12.5148	12.4879	12.4816	12.4656
B	NN	BPNN	10.0859	11.1215	12.0348	12.4816	13.1112	12.9134	13.6682
		RBFNN	13.2664	13.8016	14.8653	15.3142	16.1346	15.8687	15.5765
	$\varepsilon - SVR$	Linear	9.0092	9.8340	10.6736	11.1435	11.7831	11.6553	12.4628
		Sigmoid	12.0029	11.9358	11.9050	11.9585	11.8345	11.8915	11.9164
	$\nu - SVR$	Linear	10.1986	10.3412	10.4258	10.9532	11.5873	11.6263	12.2494
		Sigmoid	12.6029	12.6021	12.4823	12.5370	12.5003	12.5300	12.4562

Table 1: Performance of Models with the Criteria Mae

However, RBFNN with two features instead brings worse predictions. As for the SVR models, the results are analogous with NNs, except that the SVR with Sigmoid kernel function of longitudinal and horizontal dimensions has little difference. And the _ _ SVR models outperform the v_SVR models on the same condition of the same kernel function and the same lag parameter. On the whole, the _ _ SVR with Linear kernel function can perform best with the smallest MAE. In terms of different lags, the lags of 1 or 2 can bring the better predictions in Experiment A and B. The results mean that people has limited time to hesitate after they search the corresponding commodity. They may decide to buy after they search the detail in the website. The experimental results based on the criteria RMSE are shown in Table 2.

Number	Table Column Head		Lag						
	Name	*Detail*	*1*	*2*	*3*	*4*	*5*	*6*	*7*
A	NN	BPNN	23.7772	20.7681	21.2298	20.7579	23.7196	24.0512	27.1193
		RBFNN	18.4308	19.1875	19.806	20.7963	20.8881	20.4980	21.6850
	$\varepsilon - SVR$	Linear	15.1789	16.5850	17.3075	17.8258	18.8455	18.7438	19.3495
		Sigmoid	18.6290	18.5934	18.5705	18.5073	18.5039	18.5034	18.5034
	$\nu - SVR$	Linear	16.3012	16.6417	16.9351	17.0929	17.9684	18.0372	18.4885
		Sigmoid	19.2856	19.3142	19.2501	19.2117	19.1937	19.1840	19.1524
B	NN	BPNN	16.6766	17.8333	18.7476	19.0615	19.9338	19.4997	20.6375
		RBFNN	21.8220	22.0324	23.1711	24.0396	25.9970	24.7427	24.1864
	$\varepsilon - SVR$	Linear	14.9429	16.4512	17.6736	17.9537	18.7339	18.7352	19.5820
		Sigmoid	18.6495	18.5886	18.5646	18.6031	18.5129	18.5258	18.5669
	$\nu - SVR$	Linear	16.1789	16.4981	16.6982	17.1000	18.0968	18.1947	18.7359
		Sigmoid	19.2394	19.2799	19.1317	19.1618	19.1503	19.1920	19.0857

Table 2: Performance of Models with the Criteria RMSE

Number	Category Name	Lag						
		1	2	3	4	5	6	7
A	Biography	6.5761	6.1638	6.3313	6.5014	6.7515	7.0172	6.9178
	Cartoon	17.5795	17.2978	22.3503	23.0275	25.5581	25.8097	28.6002
	Management	29.3049	29.8815	30.2598	30.2210	30.2628	30.9157	32.2507
	Computer&Internet	4.3932	5.2542	5.1387	5.0401	5.0859	5.6493	5.5146
	Parenting&Families	8.6714	10.2775	10.5180	11.3308	12.1116	12.0125	12.0716
	Teaching	6.0965	6.1352	6.2548	5.6472	5.9860	5.7151	5.8910
	Finance	7.4712	7.8304	6.3622	8.3140	8.4040	6.9655	9.0187
	Fitness&Health	6.3545	5.5129	5.7224	5.7954	5.7502	5.8046	5.8905
	Economy	16.9996	15.7638	18.6841	16.9685	24.6178	24.0756	27.9092
	Exam	0.2932	0.2001	0.5722	0.2348	0.2322	0.2335	0.1773
	Motivation&Success	7.3560	8.3508	8.6431	8.5115	8.2190	8.0971	8.7285
	Travel/Map	8.7576	8.5926	8.8235	9.3464	9.9947	9.6022	9.8659
	Food/Cooking	10.6790	11.1949	10.7516	11.2806	11.4770	11.3628	11.5020
	Youth Literature	9.3233	11.1985	11.8583	12.6685	13.6322	13.1558	12.8095
	Children	25.0029	25.7403	27.5246	30.8593	29.9416	29.4973	31.0213
	Fashion/Beauty	22.2171	29.6464	32.3675	36.4560	31.7468	31.6855	31.7081
	Foreign Language	8.1409	9.4324	8.6347	9.0531	9.4484	8.7649	8.7760
	Culture	7.0950	9.3845	10.0624	10.4571	11.2998	10.9229	10.4293
	Literature	3.9796	4.2372	4.9541	5.0780	4.8364	5.1459	5.5621
	Fiction	6.3002	6.9334	8.1366	8.9073	11.9892	12.0611	12.6050
	Psychology	6.2637	6.0409	6.0806	6.0677	6.0938	6.1162	6.1066
	Art	7.9603	7.3232	8.9654	7.4264	6.6990	6.7653	6.5694
	Philosophy	7.4019	7.4830	8.0263	8.0677	7.8593	7.5654	7.8350
B	Biography	6.3967	6.4406	6.8807	6.4386	6.0584	6.4960	6.9706
	Cartoon	17.5291	18.3328	20.4099	23.7818	24.3571	25.2307	25.4439
	Management	29.5264	30.1402	32.6285	30.7271	31.4280	31.2238	35.1072
	Computer&Internet	4.3718	5.3406	5.2846	4.7523	4.4232	5.4468	5.2488
	Parenting&Families	8.3869	10.6451	11.2464	12.1993	12.2550	11.9026	12.3573
	Teaching	6.8722	6.1577	5.7526	6.0181	5.9917	6.3382	5.9422
	Finance	4.8323	6.1331	6.2599	6.8322	7.7537	6.8050	8.8509
	Fitness&Health	6.2977	5.8153	5.9269	6.0746	5.9828	5.9305	5.8553
	Economy	14.6218	14.5939	17.0723	15.9008	18.8208	20.7376	29.4469
	Exam	0.4186	0.2244	0.3690	0.2417	0.2487	0.1786	0.1555
	Motivation&Success	7.2423	8.0500	8.6186	8.6134	8.1221	8.0403	8.4483

Number	Category Name	Lag						
		1	2	3	4	5	6	7
	Travel/Map	8.5728	8.6029	8.8015	9.4985	9.4715	9.2180	9.3467
	Food/Cooking	10.3697	11.0571	10.3791	11.1179	11.3679	10.0278	11.5011
	Youth Literature	10.1217	11.9513	14.3255	12.5325	15.3772	13.6751	12.0276
	Children	25.1768	26.7748	28.6332	30.2422	29.2774	30.6389	31.6754
	Fashion/Beauty	24.2054	29.9693	32.1037	37.4490	34.3017	31.5539	31.7249
	Foreign Language	7.7297	9.5245	8.8131	9.6286	9.7672	9.5498	9.3673
	Culture	7.2275	9.3466	9.1933	9.9901	10.0049	10.0745	10.3845
	Literature	3.9228	4.0066	4.7189	4.6336	4.6058	4.6449	4.6684
	Fiction	6.2107	6.6632	7.8674	9.5350	13.1826	13.2637	14.9411
	Psychology	6.0580	5.9102	6.4632	6.0023	6.1297	6.1070	5.7507
	Art	8.0196	7.3818	8.5912	7.3080	7.1371	6.8437	6.5887
	Philosophy	6.9759	7.0165	7.1862	7.3853	7.6201	7.5140	7.8544

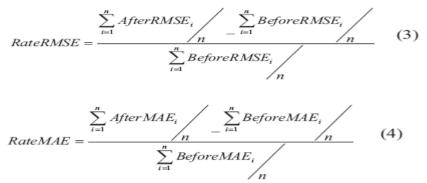*Table 3: Different Lags with Different Categories in Experiment A and B of Mae*

As can be seen from Table 2, the similar results can be observed. The _ _ SVR with Linear kernel function can bring the best predictions with the smallest RMSE. And the lags with 1 or 2 with different models except for the SVRs with Sigmoid kernel function can perform better. The results support the viewpoint that people will order in one or two days after they search their selection.

2) The optimal Lags with Different Categories
In this part, several experiments are carried out to explore the optimal lags of different categories. Aforementioned, the lag is measured by day and there are 23 kinds containing 65 books. Here the _ _ SVR with the Linear function is employed because of the best predicting performance. And as mentioned above, two kinds of experiments are carried out, A and B. The results with MAE are revealed in Table 3. As can be seen from Table 3, in experiment A, the best prediction centres are on the lag of 1 or 2 in 19 kinds of books. This mean the most historical sales data only exemplifies the value to predict the sales one or two days later. Relatively, after adding the features of search information, there are 4 categories, which are Cartoon, Teaching, Finance and Travel/Map, to be advanced the optimal lags, which means the search information is beneficial for consumers to reduce the time to hesitate. And there are 2 categories of books to be postponed the optimal lags, which are Biography and Psychology. The remarks about these two kinds of books, make the latent consumers to consider more, so the time to hesitate is respectively extended. Most categories have the optimal lag as 1 or 2 after adding the features about search information. And the respective forecast accuracy may be lowered by adding the features if the kind of book is insensitive of the remarks, which will be detailed in next part. In terms of the criteria RMSE, the experimental results are similar with MAE. The most optimal lag is identical to the result in previous Table 3, which is 1 or 2. Same as the previous, the kind- Teaching and Finance has their optimal lag advanced. And 3 kinds, Computer/Internet, Fitness and Food/Cooking, have the optimal lag postponed, which means the latent consumer will take more time to consider before they really buy. Also, the optimal lag stays the same after the features are added, which may not bring the better prediction, detailed in next subsection.

3) Enhancing Effect with Different Categories
As aforementioned, different categories react to the search information differently. To analyse the problem, the criteria is formulated as follows:

$$RateRMSE = \frac{\sum_{i=1}^{n} AfterRMSE_i / n - \sum_{i=1}^{n} BeforeRMSE_i / n}{\sum_{i=1}^{n} BeforeRMSE_i / n} \quad (3)$$

$$RateMAE = \frac{\sum_{i=1}^{n} AfterMAE_i / n - \sum_{i=1}^{n} BeforeMAE_i / n}{\sum_{i=1}^{n} BeforeMAE_i / n} \quad (4)$$

where RateRMSE and RateMAE mean the decrement rate of the RMSE and MAE after adding the features about the information search after selecting the best performance using the optimal lag. The experiments are the same with the last part, and the mean values - MAE and RMSE belonging one category are calculated. Table 4 shows the MAE and RMSE decrement rates of the different kinds.

Category	RateMAE	RateRMSE
Biography	-0.0171	-0.0190
Cartoon	0.0134	0.0146
Management	0.0076	0.0025
Computer&Internet	-0.0049	-0.0255
Parenting&Families	-0.0328	-0.0427
Teaching	0.0187	0.0597
Finance	-0.2405	-0.2379
Fitness&Health	0.0548	0.0031
Economy	-0.0742	-0.0221
Exam	-0.1226	-0.0535
Motivation&Success	-0.0154	-0.0049
Travel/Map	-0.0023	-0.0428
Food/Cooking	-0.0290	-0.0343
Youth Literature	0.0856	0.0988
Children	0.0070	0.0035
Fashion/Beauty	0.0895	0.0285
Foreign Language	-0.0505	-0.0445
Culture	0.0187	0.0119
Literature	-0.0143	-0.0069
Fiction	-0.0142	-0.0169
Psychology	-0.0480	-0.0094
Art	0.0029	-0.0033
Philosophy	-0.0576	-0.0505

Table 4: The Decrement Rate of Each Kind Book

As can be seen from Table 4, the Rate MAE of more than half of books are less than zero, which means the predicting performance becomes better after adding the features about the information search. The more typical kind with the larger ascension is Finance. The behavior that consumer search the relevant information can reflect the sales more in this category. Combined with the previous analysis, latent people tend to reduce the time to hesitate after searching books and the search information enhances the thought of buying one book in category of Finance. The Rate MAE is positive in 9 categories, Art, Children, Management, Cartoon, Culture, Teaching, Fitness &Health, Youth Literature and Fashion/Beauty, which means the latent consumers cannot be influenced by the search information, so the search information becomes the noise when predicting the sales. The Rate RMSE is almost coincident with the Rate MAE. The Finance accuracy improves most after adding the search information. There are 8 categories with their Rate RMSE positive, which are the same with the Rate MAE except for the Children category.

5. Conclusions and Future Work

This paper presents a novel data mining framework for the sales prediction using online user behavior information. Under the framework, a set of data mining methods are proposed to forecast sales trend. In the proposed method, the relationship of sales data and online user behavior data is well modelled, and the optimal lag of online user behavior data for sales prediction is also identified. In terms of evaluation criteria, a number of books are used for sales prediction. The empirical results show the efficiency and effectiveness of the proposed framework and also revealed that among different categories of books, the forecasting performance of some categories including Finance and Exam heavily relies on online user behavior information. It indicates that the proposed framework is a feasible and potential alternative for sales prediction. In addition, this study also has some research questions for further studies. Firstly, some other user behavior information, such as the user collecting volume, and also product review information, can be used to further improve the forecast performance. Secondly, the results in this paper can directly be used to optimize product inventory for cost saving. Thirdly, an online sales prediction system (SPS) can be developed to assist managers in E-commerce companies for sales trend analysis. Finally, the proposed methodology can also be applied to other research fields, especially to society hot spot, such as financial market, real estate market and labor market.

5.1. Acknowledgments

6. References

i. Kuo, Ren J., and K. C. Xue. "A decision support system for sales forecasting through fuzzy neural networks with asymmetric fuzzy weights," Decision Support Systems, vol.24, pp.105-126, 1998.

ii. Kuo, Ren Jie, C. H. Chen, and Y. C. Hwang. "An intelligent stock trading decision support system through integration of genetic algorithm based fuzzy neural network and artificial neural network," Fuzzy Sets and Systems, vol.118, pp.21-45, 2001.

iii. Yu, Yong, Tsan-Ming Choi, and Chi-Leung Hui. "An intelligent fast sales forecasting model for fashion products," Expert Systems with Applications, vol.38, pp.7373-7379,2011.

iv. Luxhǒj, James T., Jens O. Riis, and Brian Stensballe. "A hybrid econometric—neural network modeling approach for sales forecasting," International Journal of Production Economics, vol.43, pp.175-192, 1996.

v. Frees, Edward W., and Thomas W. Miller. "Sales forecasting using longitudinal data models," International Journal of Forecasting, vol.20, pp.99-114, 2004.

vi. Guo, Z. X., W. K. Wong, and Min Li. "A multivariate intelligent decision-making model for retail sales forecasting," Decision Support Systems, vol.55, pp.247-255,2013.

vii. Ginsberg, Jeremy, et al. "Detecting influenza epidemics using search engine query data," Nature, vol.457, pp.1012-1014, 2009.

viii. Da, Zhi, Joseph Engelberg, and Pengjie Gao. "In search of attention," The Journal of Finance, vol.66, pp.1461-1499, 2011.

ix. Askitas, Nikos, and Klaus F. Zimmermann. "Google econometrics and unemployment forecasting," Applied Economics Quarterly, vol.55, pp.107-120, 2009

x. Kulkarni, Gauri, P. K. Kannan, and Wendy Moe. "Using online search data to forecast new product sales," Decision Support Systems, vol.52, pp.604-611, 2012.

xi. Xu, Wei, et al. "Data mining for unemployment rate prediction using search engine query data," Service Oriented Computing and Applications, vol.7, pp.33-42,2013.

xii. Mick, David Glen, Susan M. Broniarczyk, and Jonathan Haidt. "Choose, choose, choose, choose, choose, choose, choose: Emerging and prospective research on the deleterious effects of living in consumer hyperchoice," Journal of Business Ethics, vol.52, pp.207-211,2004.

xiii. Solomon, Michael R., Rosemary Polegato, and Judith Lynne Zaichkowsky. "Consumer behavior: buying, having, and being," vol. 6, Upper Saddle River, NJ: Pearson Prentice Hall, 2009.

xiv. Laurie Peterson. "Study Places Value on Marketing at Consumer Research Stage," Marketing Daily, 2007.

xv. Strebel, Judi, Tülin Erdem, and Joffre Swait. "Consumer search in high technology markets: exploring the use of traditional information channels," Journal of Consumer Psychology, vol.14, pp.96-104,2004.

xvi. Bloch, Peter H., Daniel L. Sherrell, and Nancy M. Ridgway. "Consumer search: an extended framework," Journal of consumer research, pp.119- 126, 1986.

xvii. Stigler, George J. "The economics of information," The journal of political economy, pp.213-225, 1961.

xviii. Wu, Lynn, and Erik Brynjolfsson. "The future of prediction: How Google searches foreshadow housing prices and sales," Economics of Digitization. University of Chicago Press, 2013.

xix. Pentland, Alex Sandy. "Honest signals," MIT press, 2010.

xx. Beatty, Sharon E., and Scott M. Smith. "External search effort: An investigation across several product categories," Journal of consumer research, pp.83-95, 1987.

xxi. Moore, William L., and Donald R. Lehmann. "Individual differences in search behavior for a nondurable," Journal of consumer research, vol.7, pp.296-307