ATLANTIS PRESS

Advances in Social Science, Education and Humanities Research, volume 246

3rd International Conference on Politics, Economics and Law (ICPEL 2018)

New Energy Vehicles Sales Prediction Method and Empirical Research against the Backdrop of Big Data

Zhao Yuting*
School of Economics and Management
Harbin University
Harbin, Heilongjiang 150086

Zheng Shuang
School of Economics and Management
Harbin Engineering University
Harbin, Heilongjiang 150001

Xu Xinliang
School of Economics and Management
Northeast Agricultural University
Harbin, Heilongjiang 150030

***Abstract*—The sales of new energy vehicles are not only strongly related to the social economic condition, but also one of the important indicators reflecting the trend of public concern about environmental protection. This paper predicted the sales based on AR (1) model with the internet big data. The results showed that the prediction is more precise when combined consideration the internet search index, which means that the big data of Internet search engine improves the accuracy of forecasting. It also found that the Baidu index's prediction result is more precise than that of any other web index. If using integrated search index, the prediction is more accurate when the ration of the Baidu index and the 360 index is close to the ratio of their users.**

Keywords—Big data; new energy vehicles; internet search index; sales prediction

I. INTRODUCTION

According to "The 41st Statistical Report on Internet Development in China" released by China Internet Network Information Center, by the December 2017, the number of Internet users in China had reached 772 million, and the penetration rate had reached 55.8%, exceeding the global average (51.7%). 4.1 percentage points, exceeding the Asian average (46.7%) by 9.1 percentage points [1]. The development of the Internet affects consumers' purchasing habits to a certain extent. Nowadays, with the promotion of search engine, consumers often seek for the relevant information of products through the Internet when purchasing. The Internet has not only changed the consumer's information search habits, but also changed the way vendors sell. More and more vendors are using Internet search engine big data to predict the sales of products and taking the Internet as an important channel for advertising. This article based on two major Internet search engine, baidu and 360, predicts the sales volume of new energy vehicles. The prediction not only represents the selling trend of new energy vehicles, but also embodies the public's consciousness of new energy vehicles and environmental protection. It has economic significance, and also significant environmental and practical significance.

Heilongjiang Province Philosophy and Social Science Fund "Heilongjiang Province Supply Side Structural Reform Priority and Countermeasures" (18JYE675); Heilongjiang Province Philosophy and Social Science Fund "Study on Countermeasures for economic transformation and development of resource-based cities and regions in Heilongjiang" (18JYC257)

II. NETWORK SEARCH AND PREDICTION MODEL

With the emergence of Internet search engines, almost all product details can be found via the Internet. Currently, consumers are more likely to get information from the Internet before shopping, which makes it possible to predict product sales with search engine data.

Researches indicate that the consumers' behavior of using information search engines to get information is the basis for sales forecasting by means of search engine big data. Klein LR and Ford GT have studied the information search behavior of consumers and found that online and offline search behaviors are substitute to each other, which means that the traditional search will reduce as consumers prefer searching on the Internet [2]. Shim S et al. explored the relationship between consumers' willingness to purchase and their online search behavior by establishing a model to simulate consumers' willingness to purchase. And he found that consumers' online search behavior can reflect their purchase intention to some extent [3]. Pauwels K et al. studied the relationship between consumer online search behavior and their offline purchase intention, and found that the online search behavior can reflect both online and offline purchase behavior. [4]. In conclusion, there is a clear correlation between consumer online search behavior and purchase intention and behavior. To develop a step further, consumers' online search can reflect their purchase intention, and prediction can be conducted with the aid of the consumer network search data.

The search engine technology enables consumers to obtain all the information of products through web search. In recent years, with the development of big data and the sharing of big data of various search engines, the prediction of products sales volume and macro economy with the aid of search engine has

attracted the attention of many scholars. Huang Xiankai and others took Forbidden City as an example, analyzed the relationship between online search data and the number of tourists, and predicted the amount of tourists in the scenic spots. By comparing the autoregressive moving average model with and without Baidu index, they concluded that there was a correlation between the Baidu Index and the number of tourists. The introduction of the Baidu Index can improve the accuracy of the number of tourists in the tourist area [5]. Dong Qian and others used the index of relevant keywords provided by the Baidu Index as the network data, and took the real estate prices of large and medium-sized cities as the research object. The research results showed that the Baidu index can well reflect the changes in the price, and the announcement of its forecast results is also earlier than the official release[6]. Goel S and Watts D J studied the sales of video, games and music based on the search data on the Yahoo website. Through comparison, they summarized the sales forecasting model was more accurate when considering the network search index [7]. Lincoln NP et al. analyzed Google Trends data as network data and studied the sales volume of electronic products. The results showed that there was a correlation between product sales volume and web search volume and advertising. But that varies from product to product [8]. Guzman G declared that with the network search index, it is likely to predict more accurately the volatility of the stock market, and reduce the risk of stock investment when making investment decisions [9]. Kulkarni G and others adopted statistical software to study the relationship between the network search data of specific keywords and the box office, and predict the movie box office on this basis. The research results showed that the accuracy of the movie box office prediction increased with the aid of the network search data[10].

Vehicle is a kind of "high involvement" commodity. Consumers will spend a lot of energy to search for it before determining the purchase. The search for information through the Internet to predict the sales of automobiles has attracted academic community's attention in recent years. Wang Lian and others believe that it is possible to predict purchase through online search behavior which is an indicator to measure consumers' willingness to buy. Based on the sales data of 64 models in domestic market, the impact of consumer online search on vehicle sales and market share was verified with the fractional logit model [11]. Choi H and Varian H used Google's online search trend as a source of web search data, and Ford, Chevrolet and Toyota sales as sales data to study the relationship between web search and car sales. The results showed that there was a strong correlation between the two, and the prediction model with the network search data has a lower average error and better forecast results[12].

III. Confirming Web Search Index

A. Synthetic Principle of Web Search Index

The new energy vehicles relatively cover the minority of the population. Thus, this article uses direct key words synthesizer while synthesizing the search index. When we confirm its index, new energy vehicle (N) is selected as one key word, and top-three selling cars as other key words. Their names and numbers are respectively BYD AUTO (M_1), BAIC BJEV (M_2), and ZOTYE AUTO (M_3).

B. Synthetic Method of Web Search Index

According to above principle, we can determine the basic formula of the web search index is:

$$\text{index}=\log\left(s(N)+\sum_{j=1}^{3}\left(s(M_j)\right)\right) \tag{1}$$

Thereinto, $s(N)$ is the search index of the new energy vehicle. As we use month as basic unit for predicting, it is also the search index of some month's key word. Similarly, $s(M_j)$ is the key word of No.j car enterprise's search index.

On basis of above equation, we can determine Baidu web search index is:

$$\text{baidu\_index}=\log\left(baidu(N)+\sum_{j=1}^{3}\left(baidu(M_j)\right)\right) \tag{2}$$

Thereinto, $baidu(N)$ is the sum of Baidu index of new energy vehicle per day in one month, while $baidu(M_j)$ is the sum of Baidu index of No.j car enterprise. In the same way, the equation of 360 web search index is:

$$\text{360\_index}=\log\left(360(N)+\sum_{j=1}^{3}\left(360(M_j)\right)\right) \tag{3}$$

From the (2-3) formulas, the comprehensive web search index is:

$$C\_index=\lambda_1 baidu\_index+\lambda_2 360\_index \tag{4}$$
$$\lambda_1+\lambda_2=1$$

IV. Sales Forecasting Model

AR (1) model is adopted for studying how web search index influences the sales forecasting. If the effect that web search index takes on the sales volume of new energy vehicle is neglected, its quantity of sale in t period is:

$$q_t=\alpha q_{t-1}+\gamma+\varpi \tag{5}$$

Thereinto, q_t is the sales volume of new energy vehicle in t period; q_{t-1} is the sales volume in t-1 times; α is the autocorrelation coefficient between sales volumes; γ is a fixed value; ϖ is the fluctuate value of sales volume; obeys normal distributions with zero and variance to σ^2.

When we consider web search index's effect on the selling of the new energy vehicle and it is a high-involvement product, consumers would search details about their target goods. Thus, the web search index is correlated with sales volume. If we use Baidu search index to predict sales, the quantity of new energy vehicles in t period is:

$$q_t = \alpha_1 q_{t-1} + \beta_1 \text{baidu\_index}_{t-1} + \varpi \quad (6)$$

Thereinto, α_1 is the correlation coefficient between current period and the prior one under the Baidu search index, β_1 is the correlation coefficient between the sales volume of new energy vehicle in current period and the Baidu search index in prior period. ϖ is the fluctuate value of sales volume, obeys normal distributions with zero and variance to σ^2.

Similarly, if we use 360 search index to predict sales volume, the quantity of sales in t period is:

$$q_t = \alpha_2 q_{t-1} + \beta_2 360\_\text{index}_{t-1} + \varpi \quad (7)$$

If we use comprehensive web index to predict new energy vehicles' selling volume, its volume in t period is:

$$q_t = \alpha_3 q_{t-1} + \beta_3 C\_index_{t-1} + \varpi \quad (8)$$

α_2 is the correlation coefficient between current period and the prior one under the 360 search index, β_2 is the correlation coefficient between the sales volume of new energy vehicle in current period and the 360 search index in prior period. α_3 is the correlation coefficient between current period and the prior one under the comprehensive search index, β_3 is the correlation coefficient between the sales volume of new energy vehicle in current period and the comprehensive search index in prior period. ϖ is the fluctuate value of sales volume, obeys normal distributions with zero and variance to σ^2.

V. Empirical Analysis

The web search index of this article comes from Baidu index and 360 index, while the sales volume of the new energy vehicle comes from China Association of Automobile Manufacturers. Based on data's availability, accuracy and representativeness, the paper selects the sales volume of new energy vehicles and web search index from January to September in 2015(namely, t=1,2,...,9)(note: the volume of new energy vehicles is 34316 in October 2015)

TABLE I. Primary Data of New Energy Vehicle and Web Search Index

	0	1	2	3	4	5	6	7	8	9
Sales Volume of New Energy Vehicle (unit)	6011	6395	6045	14122	8320	10856	26954	16838	18900	28092
Baidu Search Index	13.95	13.86	13.76	13.08	14.05	15.39	13.25	13.89	15.98	13.89
360 Search Index	9.81	9.65	10.03	9.76	10.08	10.11	9.50	9.75	10.15	9.97

Note: t=0 represents relevant data in December 2014

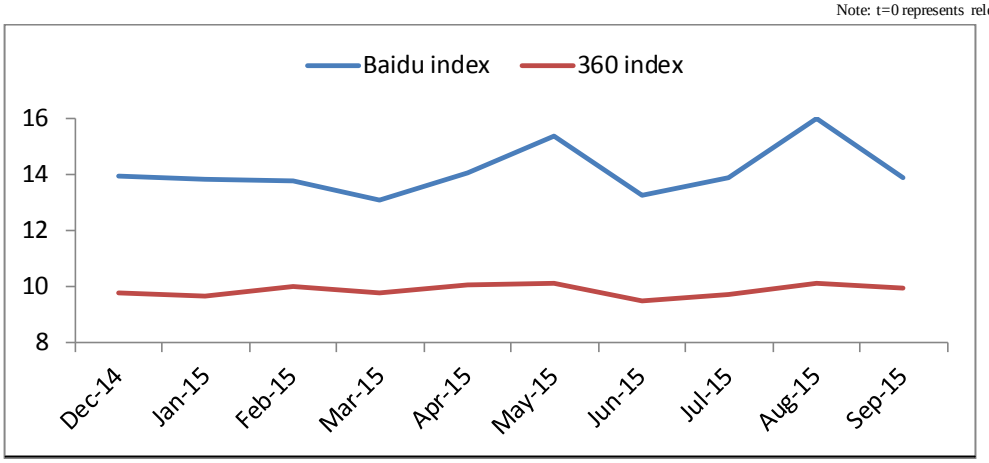

Fig. 1. Chart Compared Baidu Index and 360 Index Searching Key Word "New Energy Vehicle"

Baidu index reflects the situation that Baidu users search for the key words, while 360 index reflects their users. Take "new energy vehicle" as an example, the upper half of Cutline1 shows Baidu index's searching trend of that key word from December, 2014 to September, 2015. And the other half shows 360 index's trend in the same period. According to Cutline1, as for this key word, the two indexes show similar trend, but particular figures have obvious difference. Baidu index clearly outweigh the 360 index, which is effected by total users.

Formula (5) is the AR (1) model of new energy vehicle, in which its current sales volume is only related to prior period. Therefore, when we use SPSS22.0 to analyze it, we can replace autocorrelation analysis with simple linear regression model, and use logarithm to analyze sales data. The model is as follows:

$$\begin{aligned} & y = \alpha x + \gamma + \varpi \\ & y = \log(q_t), x = \log(q_{t-1}) \\ & s.t. \\ & \varpi \sim N(0, \sigma^2), \text{cov}(x, \varpi) = 0 \\ & \text{cov}(x, \gamma) = \text{cov}(\gamma, \varpi) = 0 \end{aligned} \quad (9)$$

Formula (9) is the simple linear regression model after using logarithm to analyze sales volume. It means that there is no relation between sales volume and demand fluctuation, sales volume and predicting constant, predicting constant and demand fluctuation.

Without considering the web search index's influence on the sales volume of new energy vehicle, we use the linear correlation analysis in spss to predict this model when we only consider the autocorrelation among sales volumes. Results are shown in TABLE II.

TABLE II. AUTOCORRELATION ANALYSIS OF THE SALES VOLUME OF NEW ENERGY VEHICLES

Model		Unstandardized Coefficients
1	(Constant)	3.630
	Sales Volume in the Prior Period	0.628

According to TABLE II, without considering web search index's effect on sales volume, the calculation method on the volume of new energy vehicles in October, 2015 is as follows:

$$0.628*\log(28092)+3.630=10.063$$

$$e^{10.063}=23495$$

Thus, it can be seen that the predicting sales volume of new energy vehicles is 23459. The deviation between true values is: $|34316-23459|/34316=0.316$ namely, the deviation is 31.6%.

Equation (6) is the AR (1) model of new energy vehicle's sales volume considering Baidu index. In this model, current sales volume is only related to the prior volume and Baidu search index. Therefore, when we use SPSS22.0 to analyze it, we can solve it with multiple linear regression models instead of autocorrelation analysis, and use logarithm to analyze sales volume. The model is as follows:

$$\begin{aligned}
&q_t=\alpha_1 q_{t-1}+\beta_1\text{baidu_index}_{t-1}+\varpi\\
&y=\alpha_1 x+\beta_1\text{baidu_index}_{t-1}+\gamma_1+\varpi\\
&y=\log(q_t),x=\log(q_{t-1})\\
&s.t.\\
&\varpi\sim N(0,\sigma^2),\text{cov}(x,\varpi)=0\\
&\text{cov}(x,\gamma_1)=\text{cov}(\gamma_1,\varpi)=0\\
&\text{cov}(\text{baidu_index}_{t-1},\gamma_1)=0\\
&\text{cov}(\text{baidu_index}_{t-1},\varpi)=0
\end{aligned}\tag{10}$$

Equation (10) is the multiple linear regression model after using logarithm to analyze sales volume, taking Baidu index into consideration. It means that there is no relation between sales volume and demand fluctuation, sales volume and predicting constant, predicting constant and demand fluctuation, Baidu index and predicting constant, Baidu index and demand fluctuation. Taking 360 index into consideration, we draw same conclusion as formula (10), so it does not appear in this article.

If we only consider Baidu search index, we predict new energy vehicle's sales volume using the linear correlation analysis with SPSS. Results are shown in TABLE III:

TABLE III. BAIDU INDEX'S INFLUENCE ON THE SALES VOLUME OF NEW ENERGY VEHICLES

Model		Unstandardized Coefficients
1	(Constant)	2.656
	Sales Volume in the Prior Period	0.543
	Baidu Index	0.152

According to TABLE III, if we only consider how Baidu search index effects the sales volume of new energy vehicles, the calculation method of new energy vehicles' sales volume in October, 2015 is as follows:

$$0.543*\log(28092)+0.152*13.89+2.656=10.329$$

$$e^{10.329}=30607$$

Thus, it can be seen that the predicting sales quantity of new energy vehicles is 30607. The deviation between true value is $|34316-30607|/34316=0.108$, namely, the deviation is 10.8%.

If we only consider 360 search index, we predict new energy vehicle's sales volume using the linear correlation analysis with SPSS. Results are shown in TABLE IV:

TABLE IV. ANALYSIS OF 360 INDEX'S INFLUENCE ON THE NEW ENERGY VEHICLES' SALES VOLUME

Model		Unstandardized Coefficients
1	(Constant)	2.983
	Sales Volume in the Prior Period	.520
	360 Index	.185

According to TABLE IV, if we only consider how 360 index affects the selling quantity of new energy vehicles, the calculation method of new energy vehicle's sales volume is as follows:

$$0.520*\log(28092)+0.185*9.97+2.983=10.154$$

$$e^{10.154}=25694$$

Therefore, it can be seen that the predicting sales quantity of new energy vehicles is 25694. The deviation between true value in October, 2015 is

$|34316-25694|/34316=0.251$, namely, the deviation is 25.1%.

If we take both Baidu and 360 search index into consideration, we predict new energy vehicle's sales volume through the linear correlation analysis with SPSS. Under this circumstance, we should first set up the coefficients combining Baidu and 360 indexes. The two coefficients should be equal, namely $\lambda_1=\lambda_2=0.5$. Then we use SPSS to predict the volume. Results are shown in TABLE V:(11.93)

TABLE V. COMPREHENSIVE WEB SEARCH INDEX'S INFLUENCE ON THE SALES VOLUME OF THE NEW ENERGY VEHICLES($\lambda_1=\lambda_2=0.5$)

Model		Unstandardized Coefficients
1	(Constant)	3.135
	Sales Volume in the Prior Period	.543
	Comprehensive Search Index	.152

According to TABLE V, when we consider how the comprehensive web search index effects the selling quantity of new energy vehicles, the calculation method of new energy vehicle's sales volume in October, 2015 is as follows:

$$0.543*\log(28092)+0.147*(13.89*0.5+9.97*0.5)$$
$$+2.894=10.210$$
$$e^{10.210}=27174$$

Therefore, it can be seen that the predicting sales quantity of new energy vehicles is 23459. The deviation between true value in October, 2015 is $|34316-27174|/34316=0.208$ namely, the deviation is 20.8%.

On the basis of TABLE V, we change the coefficient of Baidu and 360 index in the web comprehensive search index. Then, we predict new energy vehicle's sales volume through the linear correlation analysis with SPSS. Results are as follows.

TABLE VI. COMPREHENSIVE WEB SEARCH INDEX'S INFLUENCE ON THE SALES VOLUME OF NEW ENERGY VEHICLES

(λ_1=0.9, $\lambda_2=0.1$)

Model		Unstandardized Coefficients
1	(Constant)	3.135
	Sales Volume in the Prior Period	.560
	Comprehensive Search Index	0.193

According to TABLE VI, when we consider how the comprehensive web search index effects the selling quantity of new energy vehicles, the calculation method of new energy vehicle's sales volume in October, 2015 is as follows:

$$0.515*\log(28092)+0.193*(13.89*0.9+9.97*0.1)$$
$$+2.643=10.523$$
$$e^{10.523}=37160$$

Therefore, it can be seen that the predicting sales quantity of new energy vehicles is 23459. The deviation between true value in October, 2015 is $|34316-37160|/34316=0.083$, namely, the deviation is 8.3%.

We transform AR (1) model into simple and multiple linear regression models and solve relevant coefficients. Under the circumstance, we predict new energy vehicle's sales volume in October, 2015. Results are shown in TABLE VII:

TABLE VII. DEVIATION BETWEEN PREDICTED AND ACTUAL VALUE

Item	Without Index	Baidu Search Index	360 Search Index	Comprehensive Web Search Index ($\lambda_1=\lambda_2=0.5$)	Comprehensive Web Search Index (λ_1=0.9, $\lambda_2=0.1$)
Predicted Value	23489	30607	25694	27174	37160
Deviation	31.6%	10.8%	25.1%	20.8%	8.3%

As it is shown in TABLE VII, while predicting the sales volume of new energy vehicles, it is better to consider the web search index than not. If we adopt simple web index, Baidu index performs better than 360 in prediction. One reason is that Baidu owns much more users and market shares than 360. Thus, it is more accurate to use Baidu index to predict results. When we predict the sales volume through comprehensive index, it is more accurate that the number of web search index is close to its users.

VI. CONCLUSION

This paper analyzes the data of the new energy vehicle from January to October 2015 by processing the keyword data of Baidu Index and 360 Index, and studies the relationship between the network search index and the sales of new energy vehicles. It shows that the sales forecast results are better when considering the network search index, which indicates that the forecast based on the Internet search engine big data does improve the accuracy of the forecast and can back the sales prediction of new energy vehicles. At the same time, the prediction of the Baidu index is more accurate than that of any other single web index.If the comprehensive network search indexesare used, when the ratio of the Baidu index to the 360 index coefficient is close to the ratio of their number of users, the prediction is more accurate.

REFERENCES

[1] http://www.sohu.com/a/220116959_505816

[2] Klein L R, Ford G T. Consumer search for information in the digital age: An empirical study of pre−purchase search for automobiles[J]. Journal of Interactive Marketing, 2003, 17(3):29-49.

[3] Shim S, Eastlick M A, Lotz S L, et al. An online pre−purchase intentions model : The role of intention to search: Best Overall Paper Award—The Sixth Triennial AMS/ACRA Retailing Conference, 2000☆ 1[J]. Journal of Retailing, 2001, 77(3):397-416.

[4] Pauwels K, Leeflang P S H, Teerling M L, et al. Does Online Information Drive Offline Revenues? : Only for Specific Products and Consumer Segments![J]. Journal of Retailing, 2011, 87(1):1-17.

[5] Huang Xiankai, Zhang Lifeng, Ding Yusi. Research on the Relationship between Baidu Index and Tourists in Tourist Areas——Taking Beijing Forbidden City as an Example [J]. Tourism Tribune, 2013, 28(11):93-100.

[6] Dong Qian, Sun Nana, Li Wei. Real Estate Price Forecast Based on Web Search Data [J]. Statistical Research, 2014, 31(10):81-88.

[7] Goel S, Watts D J. Predicting consumer behavior with Web search.[J]. Proceedings of the National Academy of Sciences of the United States of America, 2010, 107(41):17486-90.

[8] Lincoln N P. The Relationship between Internet Marketing, Search Volume, and Product Sales [J]. Ohio State University, 2011.

[9] Guzman G. Internet Search Behavior as an Economic Forecasting Tool: The Case of Inflation Expectations [J]. Journal of Economic & Social Measurement, 2011, 36(3):4187-4199.

[10] Kulkarni G, Kannan P K, Moe W. Using online search data to forecast new product sales [J]. Decision Support Systems, 2012, 52(3):604-611.

[11] Wang Lian, Ning Yijian, Jia Jianmin. Sales and Market Share Forecast Based on Web Search: Evidence from the Chinese Auto Market [J]. Journal of Industrial Engineering and Engineering Management, 2015, 29(4):56-64.

[12] Choi H, Varian H. Predicting the Present with Google Trends [J]. Economic Record, 2012, 88(Supplement):2–9.