2010

Predict Market Share with Users' Online Activities Data: An Initial Study on Market Share and Search Index of Mobile Phone

Kaiquan Xu
City University of Hong Kong, kaiquan.xu@student.cityu.edu.hk

Jin Xu
Southwest Jiaotong University

Long Liu
USTC-CityU Joint Advanced Research Centre

Jimmy SJ. Ren
City University of Hong Kong

Wei Wang
City University of Hong KongCity University of Hong Kong

See next page for additional authors

Authors
Kaiquan Xu, Jin Xu, Long Liu, Jimmy SJ. Ren, Wei Wang, Stephen Shaoyi Liao, and Yuxia Song

PREDICT MARKET SHARE WITH USERS' ONLINE ACTIVITIES DATA: AN INITIAL STUDY ON MARKET SHARE AND SEARCH INDEX OF MOBILE PHONE

Kaiquan Xu, Department of Information Systems, City University of Hong Kong, Hong Kong SAR, China, kaiquan.xu@student.cityu.edu.hk

Jin Xu, Department of Electronic Commerce and Information Management, School of Economics & Management, Southwest Jiaotong University, Chengdu, China

Long Liu, USTC-CityU Joint Advanced Research Centre, Suzhou, China

Jimmy SJ. Ren, Department of Information Systems, City University of Hong Kong, Hong Kong SAR, China

Wei Wang, Department of Information Systems, City University of Hong Kong, Hong Kong SAR, China

Stephen Shaoyi Liao, Department of Information Systems, City University of Hong Kong, Hong Kong SAR, China, issliao@cityu.edu.hk

Yuxia Song, Department of Information Systems, City University of Hong Kong, Hong Kong SAR, China

Abstract

Acquiring accurate and timely market share information is very important for producers to arrange producing plan and design marketing strategy. However the high cost and long period of collecting survey data in survey-based method make it much difficult to easily get latest market shares data. Recently, the emerging online web systems provide users with new and convenient ways of searching, learning, experiencing and buying products. The users' activities data captured by these web systems can reflect users' buying intentions and behaviours very well, and contain very valuable information for predicting market shares. In this study, the correlation between Google search index and market shares of mobile phones is analyzed with time series analysis technology. The experiment result shows the statistically significant relationships exist between search index and market shares. This indicates the easily got search index data with low cost has the power of timely forecasting market shares. This study opens a door to apply users' online activities data to accurately and timely predict market shares, which will bring many benefits to producers and customers.

Keywords: Online Search, Market Share, Predictions.

1 INTRODUCTION

Market shares statistics play core role for producers in arranging producing plans and designing marketing strategies, and provide useful information for consumers to make right buying decisions. Usually the market shares data come from some consultation corporations (such as Nielsen.com, Gartner.com etc.), who adopt survey-based method to estimate the market shares of one domain. The survey-based method need to cover large numbers of users with wide distributions in surveys, in order to learn customers' buying behaviours and intentions. So the survey-based method requires very long period and large amount of manpower and capitals. For example, Gartner usually uses three months to publish last quarterly market shares on mobile phone. However, in modern society, accurately and timely estimating market shares with low cost is very important for producers to design right strategies and make appropriate decisions, and also is very helpful for consumers to make satisfied purchase decisions. Obviously, the survey-based method is not very appropriate for this requirement.

The emerging online web systems, such as search engine, online forum, blog, social network etc, provide users with many convenient ways for searching, learning, experiencing and buying products. There are more and more users to rely on these online systems. They usually use these systems to look over product descriptions, compare similar products before making purchase decisions, and even buy products. The users' activities data captured by these online systems can reflect users' intentions to buy certain products very well, and should be very valuable information resources for learn users' purchase behaviours. So these data should have potential power in predicting market shares of products. Also there are several advantages of these data for predicting market shares: first the data is vast and latest, so the market shares can be forecasted accurately and timely; secondly it is easy to get almost freely, which means the market shares can be predicted using these data with low cost. The users' activities data include various categories, such as search index of search engine, posts in forums, blogs and social networks. For validating if the users' activities data is useful in forecasting market shares, in this study, the correlation between Google search index and market shares of mobile phones is analyzed with time series analysis technology. The experiment results indicate the search index is significantly correlated with market shares and owns the power of predicting market shares. Also some other interesting results are discovered. This work validates the feasibility of applying large amount of online activities data to accurately and timely predict market shares with low cost. This can bring many benefits to producers and customers.

The paper is organized as follows: Section 2 introduces the related work; the data collection and description are introduced in Section 3; Section 4 presents the empirical methods and results; the discussion and limitations are shown in Section 5; Section 6 summarizes the work and indicates future work.

2 RELATED WORK

Several existing studies validate online behaviors can reveal consumers' intention and can be used to predict purchase outcomes. Moe and Fader (2004) proposed a model of conversion behavior to predict each customer's probability of purchasing based on a history of visits and purchases (Moe & Fader 2004). Kuruzovich et al. (2008) studied the relationships between the online information and shopping process and suggested different business models for infomediaries (Kuruzovich et al. 2008).

Also there is some work using Google Trends (Google 2010) to analyze some phenomena, for example Ginsberg et al. (2008) analyzed flu outbreaks using Google Flu Trends (Ginsberg et al. 2009).

Wu and Brynjolfsson (2009) analyzed the correlation between Google search index, housing sales and prices, and discovered the search index was significantly correlated with housing sales and prices (Wu & Brynjolfsson 2009). Choi and Varian (2009) also correlated housing trends in the US using economic statistics (Choi & Varian 2009).

Different with the existing work, our work focuses on detecting the relationship of market shares of mobile phone and search index, and predicting market shares using the low-cost search index data, because accurate and timely market shares prediction is very important for producers and consumers.

3 DATA SET

3.1 Data Collection

In this study, two kinds of data were collected: market shares of mobile phones and the search numbers on vendors in search engines. The market shares data of mobile phones comes from Gartner (Gartner 2010), world's leading IT research company. Gartner publishes the reports on mobile terminal market shares of vendors in each quarter by combining primary surveys and secondary research. We mainly collected the quarterly market shares of Nokia and Motorola on mobile phone from 2004 Q4 to 2009 Q4. Here, one vendor's quarterly market share is the ratio of its sale to total sales of all vendors.

The search indexes from Google trends (Google 2010) are as the numbers that vendors are searched in search engine. Google Trends analyzes a portion of Google web searches to compute how many searches have been done for the concerned terms in one period. The search index can reflect the numbers of vendors being searched. Google provides two modes of scaling on search index: relative mode and fixed mode. Here, the relative mode is adopted: the data is scaled to the average search traffic for your term (represented as 1.0) during the time period you've selected (Google 2010). The search indexes of Nokia and Motorola were collected from 2004 October to 2009 December. Considering the market shares data is the global statistics, not limited to special countries, the search index data also should cover the global searches on vendors, (for example, not only on English word "NOKIA", but also Chinese "诺基亚"). Google Trends does not consider the multi-language term problem. Here, every term is translated into other 52 language versions by using Google translator, and all versions are together as the search terms being input into Google Trends to get the search indexes. For aligning with the quarterly market shares data, the search index data is processed by averaging them in each quarter, and the quarterly search indexes are acquired.

3.2 Data Description

The collected data is scatted in the following figures.

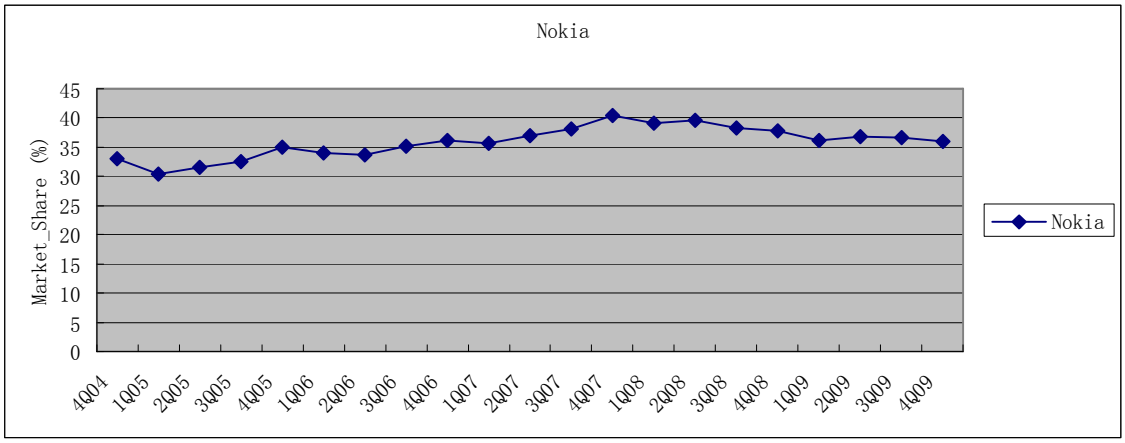

Figure 1. Market Share of Nokia.

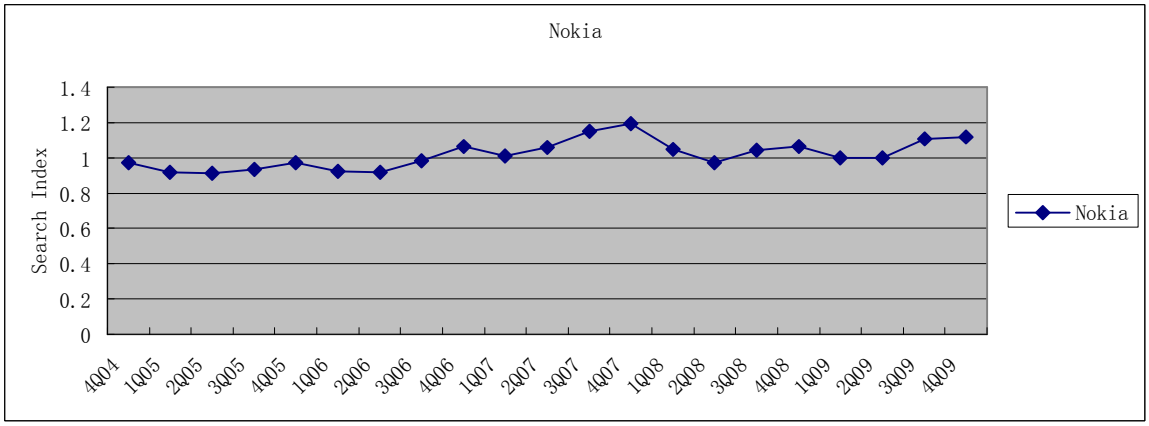

Figure 2. Search Index of Nokia

As shown in Figure 1, the market share of Nokia fluctuated from Quarter 4 of 2004 to Quarter 2 of 2007, and then reached the peak in Quarter 1 of 2008. After that, it declined gradually. In Figure 2, the search index of Nokia show the similar pattern with the market shares of Nokia, and it also peaked at Quarter 1 of 2008, and decreased gradually after that.

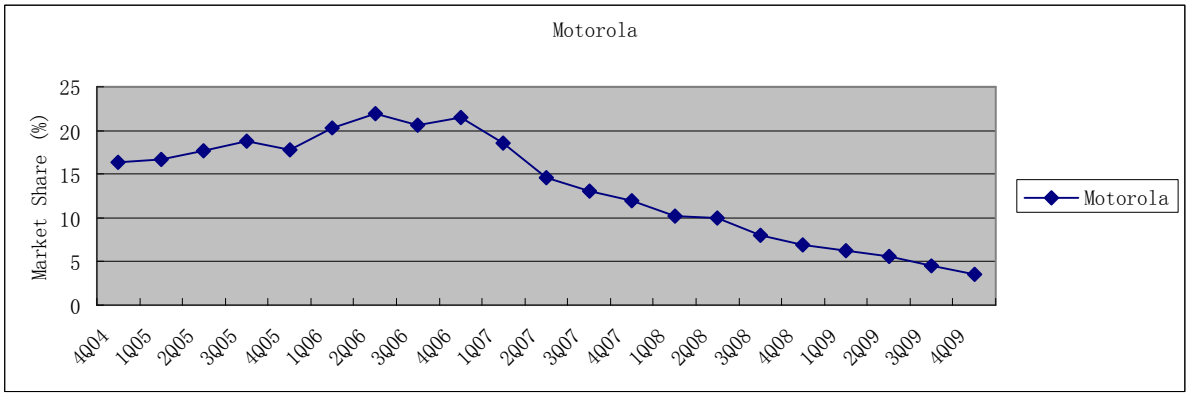

Figure 3. Market Share of Motorola

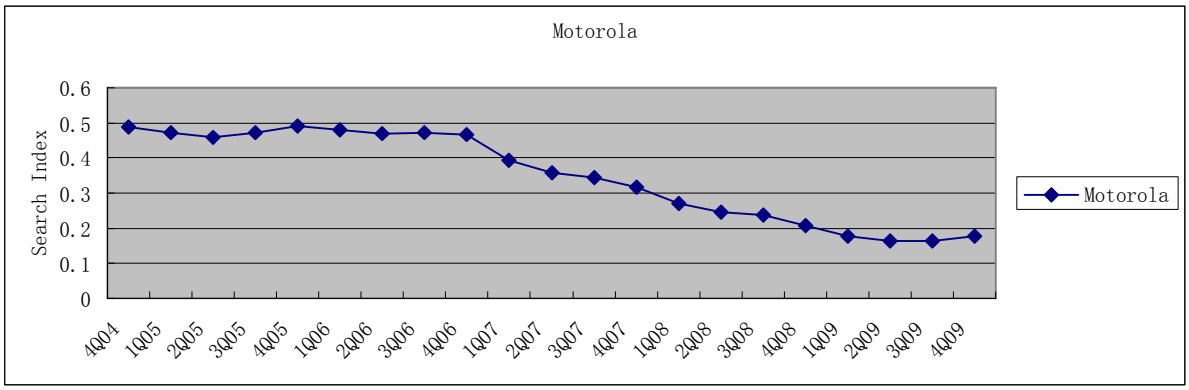

Figure 4. Search Index of Motorola

Comparing Figure 3 and 4, the line of Motorola's market share display the similar trend with the line of Motorola's search index. Both fluctuated before Quarter 1 of 2007 and went down quickly after that.

The almost synchronous changes of market shares and search indexes demonstrate some evidences that search indexes on vendor's names may be correlated with the market shares of these vendors, and should be very valuable information resources for predicting the market shares.

4 EMPIRICAL METHODS AND RESULTS

The simple Auto-Regressive (AR) models are built to detect the relationship between search index and market shares. The baseline AR model is built with the lagged market shares as explanatory variables:

$$MarketShare_t = \phi_1 MarketShare_{t-1} + ... + \phi_p MareketShare_{t-p} + \varepsilon \quad (1)$$

Here, p is set to 1 after some statistical validation.

The improved AR model adds search index and its lags as explanatory variables, to check if the search index can predict the market shares.

$$MarketShare_t = \phi_1 MarketShare_{t-1} + ... + \phi_p MareketShare_{t-p} + \varphi_0 SearchIndex_t + \varphi_1 SearchIndex_{t-1} ... + \varphi_q SearchIndex_{t-q} + \varepsilon \quad (2)$$

Here, p is set to 1, and p is set to 2, after some statistical validation.

Table 1 and 2 show the analysis results for Nokia and Motorola.

Model	market share of last quarter	search index of current quarter	search index of last quarter	search index of the quarter before last	Adjusted R2
Baseline (1)	0.654*** (0.000)				0.869
Improved (2)	0.558** (0.018)	13.148*** (0.005)	-6.003 (0.267)	11.130** (0.046)	0.860
*P<.1, **P<.05, ***P<.01.					

Table 1. Regression between market shares and search index (Nokia)

Model	market share of last quarter	search index of current quarter	Adjusted R2
Baseline (1)	1.000*** (0.000)		0.935
Improved (2)	0.541*** (0.000)	24.658*** (0.000)	0.971
*P<.1, **P<.05, ***P<.01.			

Table 2. Regression between market shares and search index (Motorola)

It can be found that the lagged market shares are strongly correlated with the contemporary market shares, which indicates the market shares for current period are related to sales from the previous period. For both Nokia and Motorola, the statistically significant relationships between the current search index and contemporary market shares can be detected. As shown in Table 1, a 1% increase of search index in current quarter is correlated with a 13.148% increase of Nokia market share in the contemporaneous quarter, and with a 24.658% increase of Motorola market share. This means contemporary search index has predictive power on market shares.

In Table 1, the search indexes of the quarter before last remain to be positively correlated with the present market shares, although the search indexes of last quarter fail to be correlated with the market shares. A 1% increase of search indexes of the quarter before last is correlated with a 11.130% increase of present market shares of Nokia. This demonstrates the usefulness of past search indexes in predicting market shares. But for Motorola, the search indexes of the quarter of last are not significantly correlated with the market shares, different from Nokia. Maybe this phenomenon is related with the price strategy of Nokia, because the prices of new Nokia products usually gradually decrease and become stable in half of year after publishing. When new products are published, some people only pay attention to them by looking over product descriptions with search engine, and then buy them after six months.

5 DISCUSSION & LIMITATIONS

The analysis in previous sections shows the evidences that the easily acquired search index data is correlated with market shares of mobile phone, and is very good predictors for market shares prediction. The timely and accurate predictions of market shares using low cost search index data can bring many benefits to both producers and customers. For producers, first this can greatly reduce the costs in estimating their market shares, because with the availability of vast search index data, the collection of data in survey-based method can be avoided in some degree. Second, quickly estimating market shares with search index data can help producers timely arrange producing plans and design marketing strategies. Third, through predicting competitors' market shares with search index data, producers can learn latest competitors' situations and set related strategies. For consumers, the accurate and timely market shares information can help them make informed decisions to buy satisfied products with low prices.

For this initial study, several issues should be resolved in the future. First the data sample is small. We can only access the quarterly market shares of mobile phones of different vendors, from Q4 2004 to Q4 2009. So the statistical analysis can only be done on this limited sample. Since of the small sample, the prediction performance of the improved model can not be compared with that of basic model,

because doing prediction will leave part of sample out for predicting. In the future, the fine-grained data, such monthly market shares, will be acquired to resolve this issue. Second, the search index from Google does not cover all activities to searching for mobile phone. Some people use other search engines, or visit certain websites directly. However, Google is the most popular search engine, so trends of Google search index can generally reflect activities of people. Also the experiment still detects the significant correlation between market shares and search indexes, and shows the power of search index data in forecasting market shares.

6 CONCLUSION & FUTURE WORK

Accurate and timely market shares prediction is very important for producers and customers. However, the survey-based method is high cost and long period. Large amount of users' activities data provided freely by various online web systems contains very valuable information for predicting market shares. In this study, we analyze the correlation of search indexes and market shares of mobile phones, and discover the existence of statistically significant relationships between them. This shows the search index data owns power of forecasting market shares, and indicates the feasibility of using low-cost search index data for timely forecasting market shares.

For our future research, large sample set about mobile phone market shares should be collected. Then the prediction performance of the improved model including search indexes will be compared with the baseline model, in order to check the effectiveness of search index data for market share prediction. Also other factors will be considered for market shares prediction, such as location, season et al. The prediction power of search index for the market shares of several competitors also will be analyzed with Vector Autoregression (VAR) model.

Reference

Choi, H. and Varian, H. (2009). Predicting the Present with Google Trends. Available at http://www.google.com/googleblogs/pdfs/google_predicting_the_present.pdf.

Gartner. (2010). Gartner. Available at http://www.gartner.com.

Ginsberg, J. Mohebbi1, M.H. Patel1, R.S. Brammer, L. Smolinski1 M.S. and Brilliant L. (2009). Detecting influenza epidemics using search engine query data. Nature, 457, 1012-1014.

Google. (2010). Google Trends. Available at http://www.google.com/trends.

Kuruzovich, J. Viswanathan, S. Agarwal, R. Gosain, S. and Weitzman S. (2008) Marketspace or Marketplace? Online Information Search and Channel Outcomes in Auto Retailing. Information Systems Research, 19(2), 182-201.

Moe, W.W. and Fader P.S. (2004). Dynamic Conversion Behavior at E-Commerce Sites. Management Science, 50(3), 326 – 335.

Wu, L. and Brynjolfsson, E. (2009). The Future of Prediction: How Google Searches Foreshadow Housing Prices and Quantities. Proceedings of the 2009 International Conference on Information Systems, 147.