

Nowcast of TV Market using Google Trend Data

Seongwook Youn* and Hyun-chong Cho[†]

Abstract – Google Trends provides weekly information on keyword search frequency on the Google search engine. Search volume patterns for the search keyword can also be analyzed based on category and by the location of those making the search. Also, Google provides “Hot searches” and “Top charts” including top and rising searches that include the search keyword. All this information is kept up to date, and allows trend comparisons by providing past weekly figures. In this study, we present a predictive model for TV markets using the searched data in Google search engine (Google Trend data). Using a predictive model for the market and analysis of the Google Trend data, we obtained an efficient and meaningful result for the TV market, and also determined highly ranked countries and cities. This method can provide very useful information for TV manufacturers and others.

Keywords: ARIMA, Google trend, Nowcast, TV

1. Introduction

A range of studies have been exploring how data can be used to monitor economic trends. One approach uses search data obtained from Google, which is extremely up to date, to predict the near future based on Google Trends data. Such techniques are called Nowcasting [1]. In [2], Google Trends data was coupled with an explanation of its potential pitfalls and other caveats, and then patterns of search data from Google Trends were compared with actual statistics from local information, to demonstrate the possible utility of Google Trends.

Google Trends data are measures of the likelihood of searches, and indicate the likelihood of a random user searching for a particular term from a certain location at a certain time. Repeated queries from a single user over a short period of time can be filtered out, to prevent their impact on the overall level of interest. The Trends data are displayed on a scale of 0 to 100 after normalization. In this way, if a particular term was most searched for in the first week of the month, the annual chart of such searches would have a peak of 100 in that week and all other weeks would be displayed as proportions of the volumes of searches for the peak week. The “Search Volume Index” is displayed graphically on the screen but the underlying data can also be downloaded as a CSV file. Google Trend data are Time series data. The Time series was started at the beginning of 2004 and is now provided on a weekly basis. Results for combinations of search terms can also be generated, and the values obtained are scaled and normalized figures [2].

As the search engine becomes more pervasive, more

shoppers are using the Web to collect information about products to purchase and to narrow down the number of selections, especially for expensive products. Knowing the frequency of online searches using a search engine like Google provides a highly accurate and up-to-date but simple way to predict future business. Government data are often released after a lag of months or more, which causes a delay in assessing current economic conditions. Using Google search data and Trends can uncover sales trends before a government publishes economic data by recognizing consumers’ interests through an analysis of their online behaviors.

There are many examples showing the correlation between search and actual data. For example, the Google trend in searches of the term flu showed strong correlation between the search and real flu patient data [3]. Advances in search engine technology, such as big data analytics technologies, can provide remarkably detailed information about human behaviors. With these technologies, it is now possible to take advantage of (almost) real-time data.

A number of companies presently utilize such methods to make predictions about consumer preferences, supplies and demands for goods, as well as levels of inventory and turnover rates. In addition, many companies, including Amazon, Coca-Cola, and Volvo, are making decisions about their business strategies to achieve tremendous profits from the market.

Among such previous studies, Moe et al. showed that online behaviors can be used to indicate consumers’ interest and to predict purchase outcomes [4]. Ginsberg et al. accurately estimated the current level of weekly influenza activity in each region of the United States. There was a high correlation between the relative frequency of certain queries and the percentage of physician visits in which a patient presented with influenza-like symptoms. In their work, they could detect influenza epidemics in areas with a large population of web search users, using search

[†] Corresponding Author: Division of Electrical and Electronic Engineering, Kangwon National University, Chuncheon, 200-701 South Korea. (hyuncho@kangwon.ac.kr)

* Dept. of Software, Korea National University of Transportation, South Korea. (youn@ut.ac.kr)

Received: December 27, 2014; Accepted: September 28, 2015

queries [3]. Hand et al. investigated and showed that forecasts of cinema admissions based on seasonal patterns in the data could be improved using Google Trend data [5]. Lui et al. tested its predictive power against the US congressional elections (2008 and 2010). Based on their investigation, Google Trend data was not a good predictor in this instance, and they explained why this may be the case [6]. Choi et al. showed how Google Trend data might help predict initial claims for unemployment benefits in the United States. And they found that Google Trend data played a key role in improving forecasting accuracy [7, 8]. Wu et al. predicted housing prices and sales with Google Trend data and found that the housing search index was strongly predictive of future housing market sales and prices. Also, they suggested how those data could be used in other markets [2]. Contreras et al. provided a method to predict next-day electricity prices based on the ARIMA methodology. They showed results for Spain and California markets [9]. Catherine et al. predicted the U.S. election [6], Askitas et al. forecast the unemployment rate [10], and Preis et al. quantified trading behavior [11] using search volume activity.

With its ability to collect search queries over time, Google trends data is useful for capturing the intentions of decision makers. In particular, Google trends data provides unprecedented opportunities for making predictions in electronics markets. Google search frequencies can be used as a reliable predictor for underlying electronics market trends both in the present and in the near future, or “Nowcast”, using a very simple regression model [1].

2. Google Trends

Google is the dominant search engine in the field. People from all around the world use Google because Google’s search algorithm is superior. Google now processes more than 60% of all the online queries in the world [6, 12].

In the present study, we obtained data from Google Trends, which analyzes web searches to compute how many searches have been conducted for certain input terms, and provides weekly and monthly reports on query statistics. Online queries that have been submitted to the Google search engine since 2004 are being captured and categorized into several predefined categories. Queries not within the predefined categories are also captured systematically. It is important to note that the search index from Google Trends does not provide the absolute count of the number of queries. Instead, each Trend is calculated as the search volume for each query in a given geographical location divided by the total number of queries, so the index is always from 0 to 100. Google Trends data is easy to access and up to date.

In our investigation, we did not use a predefined category in Google Trends. We looked at a number of terms in the Google search engine, including TV, Smart TV,

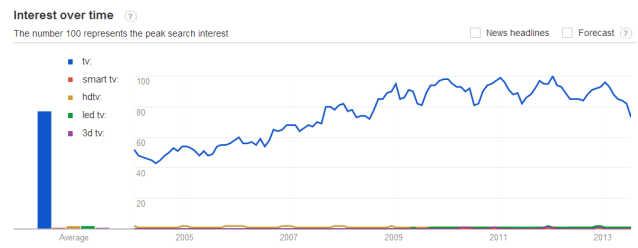


Fig. 1. Comparison of TV related search terms in Google search engine

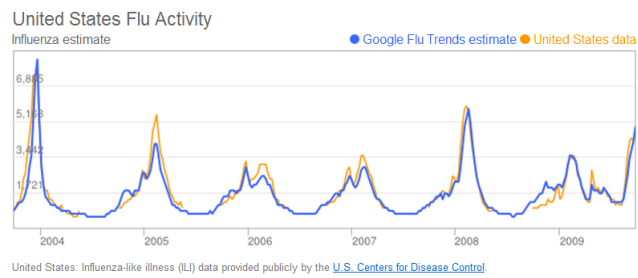


Fig. 2. United States Flu Activity

LED TV, HD TV and 3D TV. As shown in Fig. 1, many people conducted searches by just typing ‘TV’ rather than a more specific topic, such as Smart TV, LED TV, HD TV or 3D TV. Basically, since Google normalized the search frequency data, we did not. The results show that the relative frequencies of searches for Smart TV, HD TV, LED TV and 3D TV were very small in comparison to searches for the general term TV.

Based on previous studies, including the Google flu estimate, which showed that Google Trend data have the potential to predict people’s interest over time, we believe that the volume of Google search queries can be used as future economic indicators.

Google Flu Trends

As noted earlier, aggregated Google search data can be used to make weekly estimates of world influenza activity. Although not every person who searches for “flu” is actually sick, some patterns can be found. By comparing Google’s query counts and the number of people who were actually reported to have flu symptoms, we can find a close relationship [3, 13].

Fig. 2 shows annual flu activity for the United States as reported by the US Centers for Disease Control, and the Google Flu Trends estimate. There is a positive correlation between the Google flu trends estimate and the data from the US Centers for Disease Control.

In the present study, we compared TV search data taken from Google search engine. The study included data collected through September 2013. Specifically, we compared the search terms TV, Smart TV, HD TV, LED TV and 3D TV. As was seen in Fig. 1, searches for just the general term ‘TV’ were dominant. The normalized value of

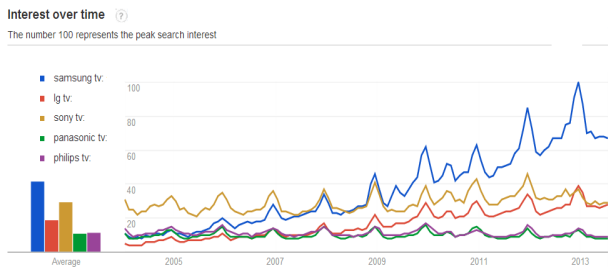


Fig. 3. Comparison of five TV manufacturers in Google search engine

all other terms, including Smart TV, HD TV, LED TV and 3D TV, were 0 or 1 in most of the period.

The study also examined search data using the names of specific TV manufacturers. As shown in Fig. 3, Samsung TV is ranked #1 from around 2009. LG and Sony are competing for 2nd place. As seen in Fig. 3, the search of each TV manufacturer has some seasonality, and shows a local peak at the end of each year. This means that end of the year is usually a heavy shopping season. The Google trend result using the keyword ‘TV’ has a highly positive correlation with the real market share.

3. Predictive Model of TV Time Series Data using ARIMA Model

We used the Auto Regressive Integrated Moving Average (ARIMA) [9] as a statistical model, and the model was implemented in R. The ARIMA model is one of the most popular prediction models for future data of a time series. The ARIMA model used in this paper is the same as the well-known model, with the difference being that the Google Trends results for commercial TV were applied to the ARIMA model. Exponential smoothing is useful for making forecasts, and does not make assumptions about the correlations between successive values of the time series data. ARIMA models include an explicit statistical model for the irregular component of a time series data, and this allows for non-zero autocorrelations in the irregular component. ARIMA models are defined for a stationary time series. Hence, if the time series data is non-stationary, to obtain a stationary time series first it is necessary to know the difference in the time series data.

The ARIMA model can make stunning predictions for a market using the Google Trend data, leveraging the trend, auto-correlation and periodicity based on historic information. ARIMA is composed of three parts: an AR (Autoregressive) model, a MA (Moving Average) model and an integrated part. Typically, the ARIMA model can be written as ARIMA (p,d,q) where p is the number of autoregressive terms, d is the order of differencing and q is the number of moving average terms. For example, ARIMA (1,1,0) is a first-order AR model with one order of differencing. In most cases, the best model turns out to be a

model that uses either only AR terms or only MA terms. The ARIMA model can be used if data is stationary. To apply the ARIMA model for the non-stationary case, non-stationary data should be transformed by an initial differentiation or logarithm.

Given a time series Y, an AR model of order p is defined as follows:

$$Y(t) = \sum_{i=1}^p \beta_i Y(t-i) + \epsilon \tag{1}$$

where Y(t) is the number of views in the tth day. The current value of Y(t) can be found from past values; β_1, \dots, β_p are the parameters of the model; and ϵ is a random shock.

An MA model of order q is defined as follows:

$$Y(t) = \sum_{i=1}^q \theta_i \epsilon_{t-i} + \epsilon_t \tag{2}$$

where $\theta_1, \dots, \theta_p$ are the parameters of the model and are again random shocks. The ARIMA model of order (p, q) can be defined with the above equations:

$$Y(t) = \sum_{i=1}^p \beta_i Y(t-i) + \sum_{i=1}^q \theta_i \epsilon_{t-i} + \epsilon_t \tag{3}$$

The random shock terms, ϵ_t , are generally assumed to be Gaussian random variables with zero mean and constant variance.

4. Experiment and Results

For the case using ‘TV’ as the keyword search on the Google search engine, local maxima are always January of each year, so January 2013, January 2012, January 2011 were the local maxima. The local minima were August 2012, June 2011, July 2010, and July 2009. Hence, we can see that the search data for ‘TV’ changes seasonally, and increases in January of each year and decreases in summer (June, July or August) of each year. We previously depicted the volume of Google search engine for ‘TV’. As can be seen, the graph is not a stationary time series, so we differenced the time series once and plotted the difference series in Fig. 4.

As shown in Fig. 4, the search result for ‘TV’ in the Google search engine is somewhat seasonal and not stationary. From Fig. 5, the differenced time series looks stationary in mean.

Based on the Auto Correlation Function (ACF) and Partial Auto Correlation Function (PACF), the Moving Average (MA) (1) model is appropriate for the data, as

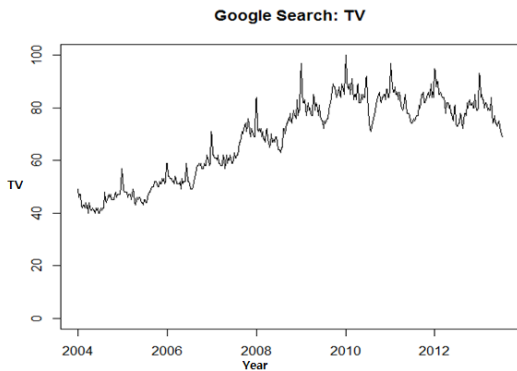


Fig. 4. TV search result in Google search engine

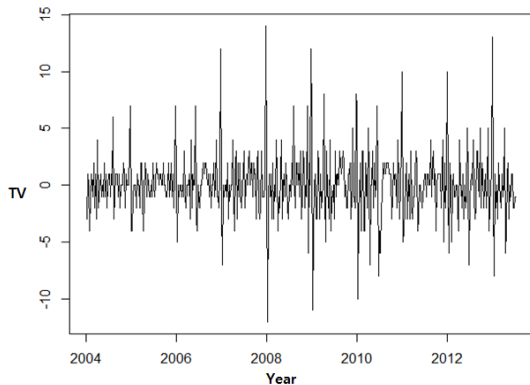


Fig. 5. Time series data of first difference

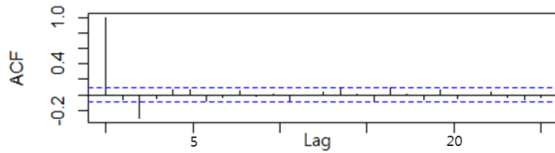


Fig. 6. Correlogram of the stationary TV time series data

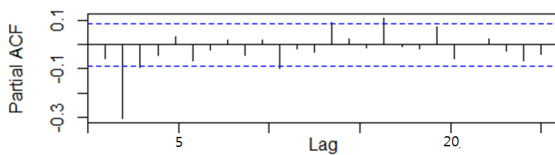


Fig. 7. Partial Correlogram of the stationary TV time series data

shown in Fig. 6 and Fig. 7. In the case of ACF, the value at lag 2 exceeded the significance bounds. In the case of PACF, the values at lag 2, 11, 14 and 17 exceeded the significance bounds.

The mean forecast error is 0.048. The time series forecast errors have a positive mean rather than a zero mean. Fig. 8 shows the results of the prediction model using an ARIMA model. The blue line represents the Nowcast using the prediction model. The red line represents the results with a 5% significance level. As can be seen in Table 1, the result is not excellent. MAE is 1.94 and RMSE is 2.82 in the experiment. These numbers

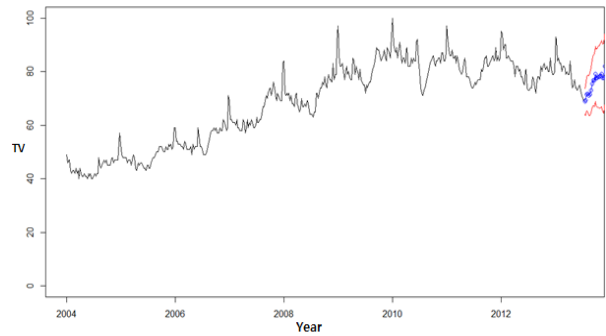


Fig. 8. Prediction model of Google trends data: 'TV'

Table 1 Results of Predictive Model

ME	RMSE	MAE	MPE	MAPE	MASE
0.0484	2.8268	1.9432	-0.0049	2.7789	0.3289

ME (Mean Error)
 RMSE (Root Mean Squared Error)
 MAE (Mean Absolute Error)
 MPE (Mean Percentage Error)
 MAPE (Mean Absolute Percentage Error)
 MASE (Mean Absolute Scaled Error)

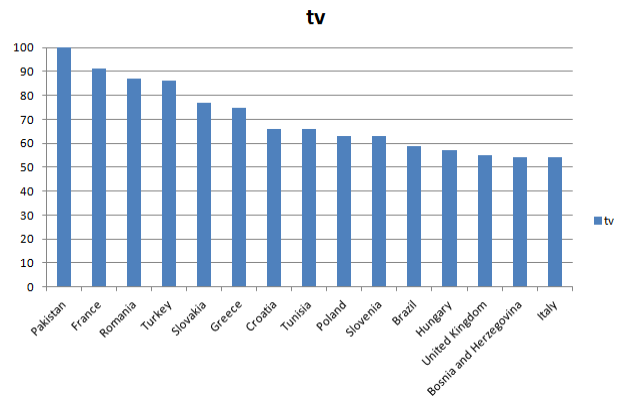


Fig. 9. Keyword 'TV' search result based on Country

should be lower for better results (less than 1 if possible). There is no accurate answer for a prediction using this kind of time series data. While low values (less than 1) are generally a good result, accuracy also depends on the characteristics of the data. In Table 1, some values are higher than 1 (i.e. RMSE, MAE and MAPE). Although it would be better if the values of RMSE and MAPE were lower, we cannot control those values. They are just the prediction results.

Fig. 9 shows the keyword 'TV' search result based on which country the search was conducted in. Pakistan, France and Romania are the top 3 'TV' search countries for this keyword on the Google search engine.

Fig. 10 shows the search result for the term 'TV' based on city. Istanbul, Paris and Ankara are the top 3 'TV' search cities on the Google search engine. According to the Google Trends data, the search for 'TV' is hot in Pakistan, France, Romania, Turkey (Country) and Istanbul, Paris,

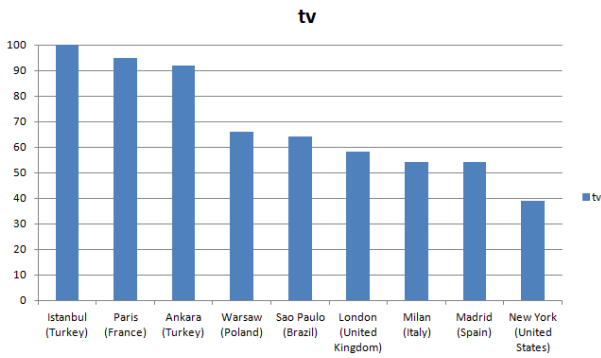


Fig. 10. Keyword ‘TV’ search result based on City

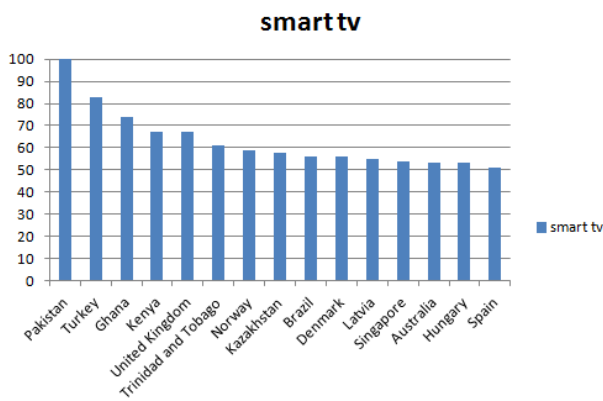


Fig. 11. ‘Smart TV’ search result based on Country

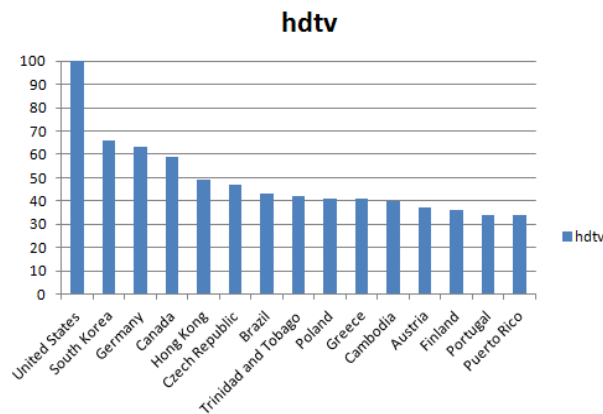


Fig. 12. ‘HD TV’ search result based on Country

Ankara and Warsaw (City). By using this information, manufacturers could intensify the marketing in these countries and cities immediately to increase revenue. We think that the searched data in Google search engine have a positive relationship with actual sales activity, similar to the case with the Google Flu Trends data. Also, as seen on the prediction model in Fig. 8, the market for ‘TV’ might be increasing smoothly.

We also investigated other ‘TV’ related terms such as Smart TV, HD TV, LED TV and 3D TV. The results are shown in Figs. 11, 12, 13 and 14. Fig. 11 shows the TV related term search results, depicting the top 15 countries

for each category. For Smart TV, Pakistan, Turkey and Ghana are top 3 in search frequency. For HD TV, United States, South Korea and Germany are the top 3. For LED TV, India, Brazil and Hungary are the top 3. For 3D TV, Brazil, United Kingdom and Hungary are the top 3. All these results were based on current Google searches, which means that if changes should occur due to some issues or policy changes in any country, the changes can be quickly

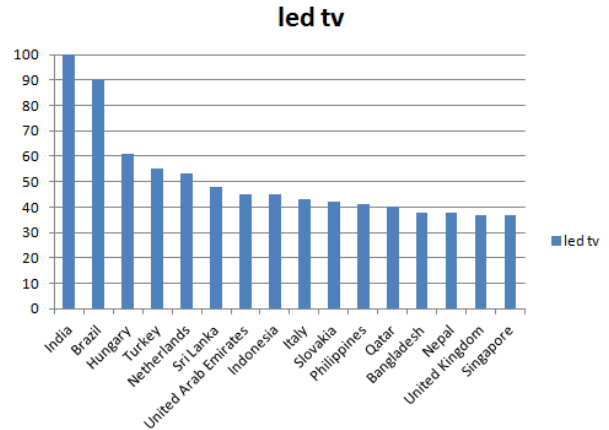


Fig. 13. ‘LED TV’ search result based on Country

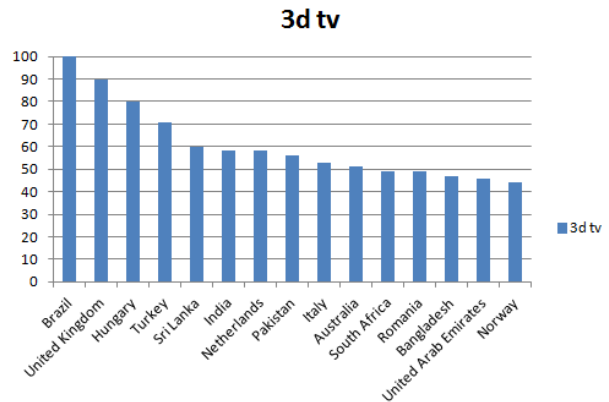


Fig. 14. ‘3D TV’ search result based on Country

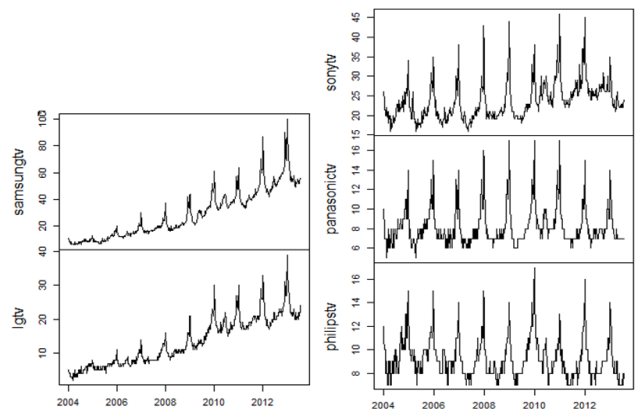


Fig. 15. Five TV search results on the Google search engine

observed. For example, BBC recently announced that they had no further plans for the 3D format after the trial period ended, which would affect the 3D search results in the UK sooner or later. However, overall, compared with the keyword ‘TV’, the search results for other terms was trivial.

The Google Trend data results for Samsung TV, LG TV, Sony TV, Panasonic TV, and Philips TV are shown in Fig. 15. The graph shows the relative interests of people in the various TV manufacturers.

5. Conclusion

In this paper, we reviewed search data for the keyword “TV” in the Google search engine and predicted the TV market based on the searched data using the ARIMA model. Also, we were able to estimate target markets based on the frequencies of specific keyword searches on the Google search engine. We began by noting that the Google search trend for the keyword “flu” was correlated with actual flu activity, and we hypothesized that TV markets could be predicted using Google search trends, because specific keyword searches might be correlated with the real market. The future will tell us if this hypothesis is accurate or not. The keyword ‘flu’ is cited by Google as an example of Google trend data. There is no specific relationship between flu and TV. Search results for the 5 TV brands are very similar to real market share, so we assume it has a relationship.

We found that we could predict the near future of the market for certain home electronic products (e.g., TVs) using Google Trends data.

Google Trends data reveals up to date trends among people, based on their keyword search activity. This method might be applied to other products (e.g., refrigerators, cell phones, etc.) or fields (e.g., home sales, diseases, etc.). Through analysis of the Google Trend data, it is also possible to determine search activity and different trends based on country and city for keywords searched in the Google search engine. As a result, it is possible to find useful market trends or economic information by watching the Google search trend. For example, the method might be used to predict daily price moves in the Dow Jones industrial average. The search results of Google Trends will change as time goes by because the searches will be different based on various factors.

The method proposed in this study based on the correlation between Google Trends and real market share may be useful in predicting future market trends, and addressing them appropriately. For example, by intensifying marketing in regions showing a high search frequency for certain products, the sales of the item (such as TVs) might be increased.

Acknowledgements

This study was supported by Korea National University of Transportation in 2015 and 2014 Research Grant from Kangwon National University.

References

- [1] Y. Carrière-Swallow and F. Labbé, “Nowcasting with Google Trends in an Emerging Market,” *Journal of Forecasting*, vol. 32, pp. 289-298, 2013.
- [2] L. W. a. E. Brynjolfsson, “The Future of Prediction: How Google Searches Fore-shadow Housing Prices and Sales,” presented at the NBER Conference Technological Progress & Productivity Measurement, 2009.
- [3] J. Ginsberg, M. H. Mohebbi, R. S. Patel, L. Brammer, M. S. Smolinski, and L. Brilliant, “Detecting influenza epidemics using search engine query data,” *Nature*, vol. 457, pp. 1012-4, Feb 19 2009.
- [4] W. W. Moe and P. S. Fader, “Dynamic Conversion Behavior at E-Commerce Sites,” *Management Science*, vol. 50, pp. 326-335, 2004.
- [5] C. Hand and G. Judge, “Searching for the picture: forecasting UK cinema admissions using Google Trends data,” *Applied Economics Letters*, vol. 19, pp. 1051-1055, 2012/07/01 2011.
- [6] P. T. M. a. E. M. Catherine Lui, “On the predictability of the U.S. Elections through Search Volume Activity,” in *e-Society Conference*, Avila, Spain.
- [7] H. Choi and H. A. L. Varian, “Predicting the Present with Google Trends,” *Economic Record*, vol. 88, pp. 2-9, 2012.
- [8] C. E. a. H. C. Hal Varian, “Predicting the Present with Google Trends,” in *Google Research Blog*, ed, 2009.
- [9] J. Contreras, R. Espinola, F. J. Nogales, and A. J. Conejo, “ARIMA models to predict next-day electricity prices,” *Power Systems, IEEE Transactions on*, vol. 18, pp. 1014-1020, 2003.
- [10] N. Askitas and K. F. Zimmermann, “Google Econometrics and Unemployment Forecasting,” *Applied Economics Quarterly*, vol. 55, pp. 107-120, 2009/04/01 2009.
- [11] T. Preis, H. S. Moat, and H. E. Stanley, “Quantifying trading behavior in financial markets using Google Trends,” *Sci Rep*, vol. 3, p. 1684, 2013.
- [12] Google, *How does Google Trends Work, Official Site, 2010*. Available: <http://www.google.com/intl/en/trends/about.html>
- [13] P. M. Polgreen, Y. Chen, D. M. Pennock, and F. D. Nelson, “Using internet searches for influenza surveillance,” *Clin Infect Dis*, vol. 47, pp. 1443-8, Dec 1 2008.



Seongwook Youn He received the B.S. degree in Computer Science from Sogang University, Seoul, Korea in 1997, and M.S. and Ph.D. degrees in Computer Science from University of Southern California, Los Angeles, CA in 2002 and 2009, respectively. Dr. Youn's current interests are Market

Data Forecast, Data Science, Personal Information Management, etc. He is currently an Assistant Professor at Department of Software, Korea National University of Transportation, South Korea.



Hyun-chong Cho He received the M.S. and Ph.D. degrees in Electrical and Computer Engineering from the University of Florida, USA in 2009. During 2010-2011, he was a Research Fellow at the University of Michigan at Ann Arbor, USA. From 2012 to 2013, he was a Chief Research Engineer in LG

Electronics, South Korea. He is currently an Assistant Professor at Kangwon National University, South Korea.