



HAL
open science

Developing Smart Supply Chain Management Systems Using Google Trend's Search Data: A Case Study

Ramin Sabbagh, Dragan Djurdjanovic

► **To cite this version:**

Ramin Sabbagh, Dragan Djurdjanovic. Developing Smart Supply Chain Management Systems Using Google Trend's Search Data: A Case Study. IFIP International Conference on Advances in Production Management Systems (APMS), Sep 2019, Austin, TX, United States. pp.591-599, 10.1007/978-3-030-29996-5_68 . hal-02460473

HAL Id: hal-02460473

<https://inria.hal.science/hal-02460473v1>

Submitted on 30 Jan 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Developing Smart Supply Chain Management Systems Using Google Trend's Search Data: A Case Study

Ramin Sabbagh and Dragan Djurdjanovic

The University of Texas at Austin, Austin, TX 78712, USA
sabbagh@utexas.edu , dragand@me.utexas.edu

Abstract. Future manufacturing companies require smarter solutions to compete in the economy. Smart supply chain management systems are one of the most effective solutions. Use of previous information can help companies to predict the demands of the market and react in an agile manner to sudden changes. Google receives over 63,000 searches per second on any given day. This huge amount of data provides us with the opportunities to investigate researches in multiple subjects and extract useful information from the raw data that is available through Google Trend. In this research, we investigate the possible relationships between searches that are made in Google for two manufacturing capability terms, namely, Precision Machining (PM) and Electric Discharge Machining (EDM). Time-series oriented research is conducted on these two datasets in order to find the dynamics characteristics as well as interesting hidden relationships between these two search items to help us build a smarter supply chain management system. Two different methods namely ARMA and ARMAV models are applied to fit a representative model to these datasets. The order of the both models are evaluated based on AIC statistic. In addition, multiple seasonal trends are detected in the datasets. Finally, Using ARMA model, we predict the datasets for one-step ahead in order to validate our models. Recognition of seasonalities and correlations between two datasets could lead to better prediction and smarter supply chain creation and management.

Keywords: Supply Chain, Time-Series Analysis, Knowledge Management.

1 Introduction

In order to remain competitive in today's volatile economy, manufacturing companies need to be provided with smart supply chain management system that enables them to manufacture products more efficiently, less expensively, and more quickly. Recent applications of Machine learning and Artificial Intelligence provided us with powerful tools and methods to make smarter decisions [1, 2, 3, 4]. To react quickly to the sudden changes and be aware of the future demands and situations, companies need to focus more on previous information and trends. Google created algorithms to help people find their way around the ever-growing amount of online content. Today, Google is a powerhouse that continues to innovate and improve the virtual world. Its ongoing success story is a result of its dedication to keep getting better, which is why

it's the go-to search engine and arguably the most trusted source of information out there. Based on recent report on Search Engine Land website [5], Google has 90.46% of the search engine market share worldwide. However, 15% of all searches have never been searched before on Google. It is reported in the Engine Land website that Google receives over 63,000 searches per second on any given day and has a market value of \$739 billion.

1.1 Google Trend

Google Trends is a website by Google that analyzes the popularity of top search queries in Google Search across various regions and languages. The website uses graphs to compare the search volume of different queries over time [6]. Google Trends also allows the user to compare the relative search volume of searches between two or more terms. The quota limits for trend search is based on the number of search attempts available per user/IP/device. Details of quota limits have not yet been provided, but it may depend on geographical location or browser privacy settings. It has been reported in some cases that this quota is reached very quickly if one is not logged into a Google account before trying to access the trends service [7].

1.2 Precision Machining versus Electric Discharge Machining

Based on what we discussed in the previous section, Google Trend is great source of data for different fields and is customizable based on different categories such as countries, time, search type, etc. In this research, aim to explore the possible relationships between searches that are made in Google for two manufacturing capability terms, namely, Precision Machining (PM) and Electric Discharge Machining (EDM). In first impression, one can think of some correlation and relationships between PM and EDM as concept. However, we perform time-series oriented research on these two datasets in order to find those interesting hidden relationships between these two manufacturing concepts.

2 Methodology

In this section, autoregressive moving average models are utilized to fit the best model to the datasets according to a specific statistic. In addition, vectorial ARMA model (ARMAV model) is introduced and the results will be compared with the regular ARMA model. Moreover, the stochastic seasonalities associated with the datasets are evaluated. Finally, we use the datasets to predict the future data points based on previous information.

2.1 Auto Regressive Moving Average (ARMA) Models

In order to fit a model to the data, ARMAX model is utilized in this paper [8]. Several ARMA ($2n$, $2n-1$) models are fitted to the model and using a procedure introduced by Pandit and Wu, we could find the appropriate order for the model that represents the

data sufficiently [9]. As a significance detection function, Akaike Information Criterion (AIC) [10] method is used to find the optimum ARMA model by passing the white noise test [11]. Using ARMA model does not give us this opportunities to get information from different dataset as an extra input variable. However, vectorial ARMA models could solve this challenge so that the model be more realistic and reliable.

2.2 Vectorial ARMA Models (ARMAV)

To investigate better options of fitting the dataset into appropriate models, ARMAV model is utilized in this section. An ARMAV model is provided for one dataset based on both datasets (using the second dataset as extra input). The results including the model order as well as residual sum of squares (RSS) is provided and compared between two different approaches at the end of this section.

2.3 Auto Regressive Characteristic Polynomial Roots

In both ARMA and ARMAV models, the roots associated with AR part of the model are mapped and based on being inside, on, or outside of the unit circle, datasets would be stationary, marginally stationary and non-stationary respectively. Finally, possible stochastic trends as well as seasonal trends are explored and detected using parsimonious models where the roots close to 1 are being pushed to be exactly one and removed from the model.

2.4 Prediction

The last step of the methodology is to predict the future of the datasets based on previous data points. For this purpose, 75 percent of the data is used as training part of the data and 25% is used as the test data. Once the model is evaluated, we predict the one-step-ahead, two-steps-ahead,...,N-steps-ahead predictions based on training dataset. In this report, the prediction results for ARMA models and ARMAV models are compared and ARMAV model is used for the prediction due to better performance.

3 Result

In this section, results are provided for different parts of the methodology. First, we provide the results for ARMA models as well as roots and seasonalities. Then, the results for ARMAV models and seasonalities are illustrated.

3.1 Dataset introduction

Two datasets will be presented. Figure 1 illustrates the dataset which is the weekly sampled PM and EDM searches for the last 5 years. It is shown that the search for EDM search has higher interest over time compared to PM search. In addition, some seasonalities and correlations between two datasets are observable. One of the main purposes of this research is to find those seasonalities in the two datasets.

The data shows the interest over time for PM and EDM searches in United States. Interest over time expresses the popularity of that term over a specified time range. Google Trends scores are based on the absolute search volume for a term, relative to the number of searches received by Google. Where there is sufficient data available, Google Trends awards a score of between 0 and 100 to inputted search terms on a month-by-month or week-by-week or day-by-day basis and on a geographical basis.

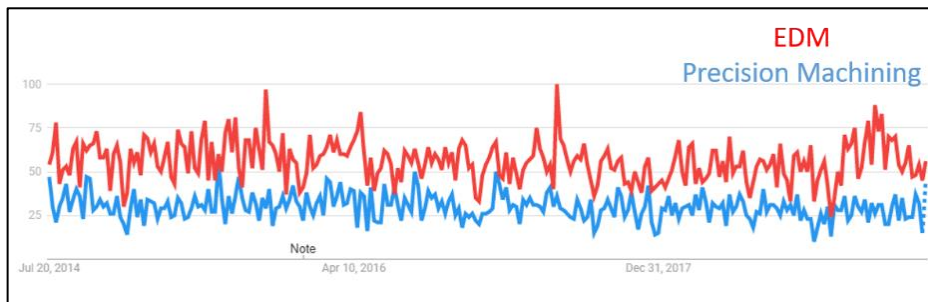


Fig. 1. Interest over time for precision machining and EDM searches

Before presenting more detailed description of the methodology, we provide some extra information regarding the search data provided by Google Trend. Top five areas with highest interest for PM are Illinois, Kansas, Oregon, California, and Wisconsin. In addition, the top five areas with highest interest for EDM are New Hampshire, Utah, Idaho, Kentucky, and South Carolina. This measure provided by Google Trend can show the real interest or demand from aforementioned states for PM and EDM respectively. Moreover, most popular related queries when searching for PM and EDM are also provided by Google Trend website.

3.2 ARMA Model Results

Figure 2 shows the autocorrelations resulted by ARMA model. For PM search data, ARMA (12,11) and for EDM data, ARMA (1,0) models were detected to be adequate models by AIC statistic. The resulting models illustrate that the PM search data is more complicated than EDM search data. It is also shown from Figure 2 that both models are confirmed to be adequate model by RSS of $9.367809e+03$ for PM search data and RSS of $2.319061e+04$ for the EDM search data.

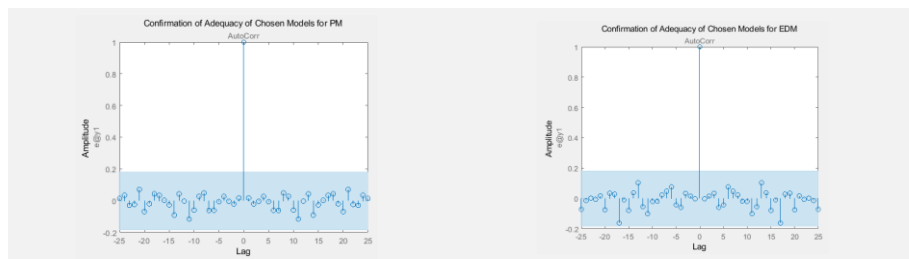


Fig. 2. Fitting ARMA Model to data - Left figure is the autocorrelation plot for Precision Machining (PM) search data and the right Figure is the autocorrelation plot for EDM search data).

3.3 Stochastic Trends and Seasonalities for ARMA Models

Figure 3 shows the roots associate with AR parts of the ARMA models for PM search and EDM search data separately. In PM data, five roots are inside the unit circle and so the PM search is a stationary dataset with one real root and four complex roots. In addition, none of the four roots is close to 1 and thus, there is no stochastic and seasonal trends in the PM data. The EDM search dataset has only one root that is real. The AR root associated with EDM data are also inside the unit circle and the EDM data is also stable dataset. After further process of EDM and PM datasets and looking for seasonal trends using parsimonious models, it turned out that none of them have seasonal trends.

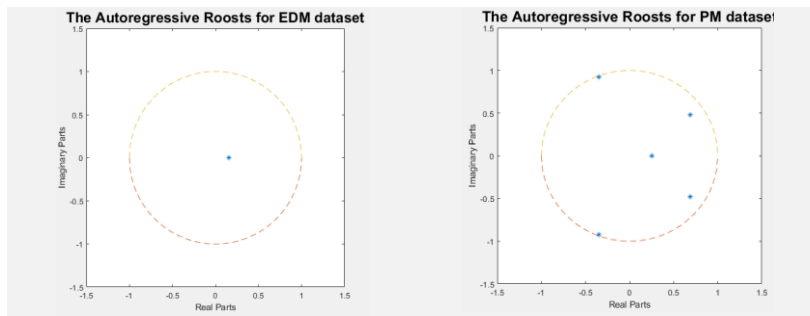


Figure 3. Auto Regressive Roots of the datasets mapping along with the unit circle.

3.4 ARMAV Model Results

In this section, the results associated with fitting an ARMAV model are illustrated. Figure 5 shows the autocorrelations resulted by ARMAV model. For PM search data driven by EDM data, ARMA (19,18) and for EDM data driven by PM data, ARMA (12,11) models were detected to be adequate by AIC statistic. The resulting models illustrate that the PM search data driven by EDM search data is more complicated than EDM search data driven by PM search data. It is also concluded from Figure 4 that both models are confirmed to be adequate models by RSS of $4.387391e+03$ for PM search data driven by EDM and RSS of $1.366753e+04$ for the EDM search data driven by PM.

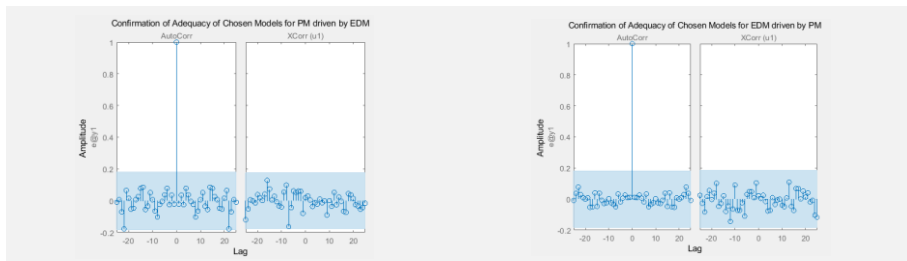


Fig. 4. Fitting ARMAV Model to data - Left figure is the autocorrelation plot for PM search data driven by EDM search data as input and Right Figure is the autocorrelation plot for EDM search data driven by PM search data as input.

3.5 Stochastic Trends and Seasonalities for ARMAV Models

Figure 5 shows the roots associate with AR parts of the ARMA models for PM search and EDM search data separately. In EDM search data driven by PM data model, all 12 roots are inside the unit circle and so the EDM search is a stationary dataset with zero real root and 12 complex roots. In addition, none of the seven roots are close to 1 and thus, there is no stochastic and seasonal trend in the PM data.

The PM search data driven by EDM search data has 19 roots with three real root and 16 complex roots. However, there are four roots in EDM data that are almost on the unit circle. These four roots could be the sources of the possible seasonality in EDM data. Moreover, the roots which are on the unit circle has the multiplicity of one. Thus, the PM data is a marginally stable dataset.

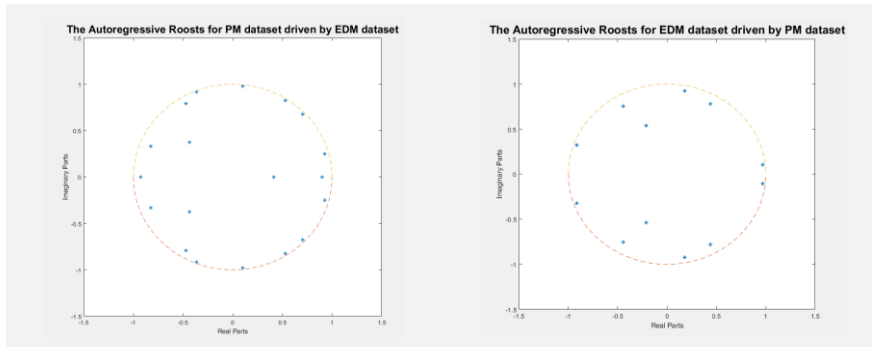


Fig. 5. Auto Regressive Roots of the datasets mapping with the unit circle. (Left plot is the PM search data driven by EDM search data; and right plot is the EDM data driven by PM data)

After further process of the PM dataset by forcing the roots close to zero to be exactly zero, the parsimonious model of the ARMAV model associated with PM dataset driven by EDM search dataset will be created in order to evaluate the seasonal trends of the EDM search data. It turned out that EDM search data has two different seasonal trends 13, and 52.

- Seasonality of 13: The ARMAV model of the PM data has the seasonality of 13. This seasonality looked unusual in the first place. However, considering the sampling rate of one week, the 13 weeks means exactly three months (91 days). It is an interesting that precision machining search data have quarterly seasonality.
- Seasonality of 52: The next detected seasonality is for 52 weeks which is exactly one year and similar to what we found based on ARMA model results.

3.6 Performance Evaluation for ARMA and ARMAV Models

Now that the adequate models are detected and roots are observed and stochastic and seasonal trends are detected both for ARMA and ARMAV models, we can compare the performance of those two models. One of the main criteria for performance of a model is the residual sum of squares (RSS). ARMAV model performed better in both

EDM and PM search data in terms of fitting the better model to the data. In order to compare the fittings, we could compare the RSS associated with two models. RSS of ARMAV model for PM driven by EDM is $4.387391e+03$. However, RSS of ARMA model for PM is $9.367809e+03$. This means a great reduction (53.17%) in RSS after using ARMAV model. On the other hand, the RSS of ARMAV model for EDM search dataset driven by PM search data is $1.366753e+04$. However, RSS of ARMA model for EDM search data is $2.319061e+04$. In this dataset we have reduction of 81.08 % in RSS after using ARMAV model.

3.7 Prediction Results for the PM Search Dataset:

In order to validate our model, we conducted a multiple-steps ahead prediction based on ARMA and ARMAV model to predict the PM search dataset. Figure 6 shows the result of the prediction. We updated the dataset by further predicting the next data point and adding the predicted points to be used as an input for the next prediction. The result is promising noticing that we have not done the prediction using ARMAV model which of course makes the prediction stronger (by using the PM search data as well to predict the EDM search data). Using ARMAV model predictions in one of the main future works for this research.

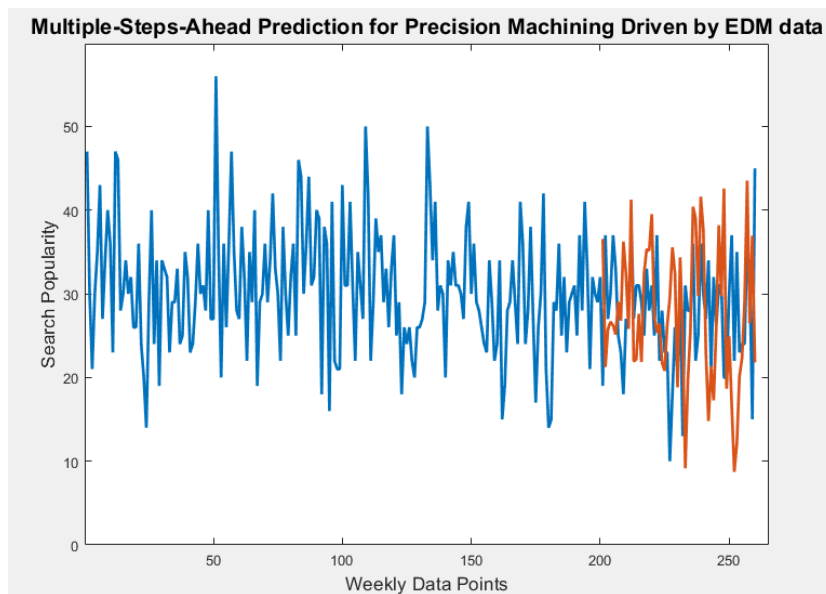


Fig. 6. Multiple-steps ahead prediction Using ARMAV model for Precision Machining (PM) dataset driven by EDM dataset.

4 Conclusion and Future Work

In this research, we explored the possible relationships between searches that are made in Google for two manufacturing capability terms, namely, Precision Machining

and Electric Discharge Machining. The purpose of this investigation is to illustrate a possible method to build a smart supply chain management system based on online data provided by Google. Time-series oriented research were conducted on these two datasets in order to find the dynamics characteristics of these two datasets as well as interesting hidden relationships between these two search items. Two different methods namely ARMA and ARMAV models are investigated in order to fit a representative model to these datasets. The order of both models were evaluated based on AIC statistic. For EDM search and PM search data, ARMAV model outperformed the ARMA model by RSS reduction of over 50 percent. Two different seasonal trends were detected in EDM search dataset. It is concluded that there are quarterly, and yearly seasonal trends in the EDM search data. However, no stochastics and seasonal trends is found in PM data. Finally, Using ARMAV model, we can predict the PM until next 60 weeks using multiple-steps ahead method with high fidelity. For future work, more complicated predictions will be considered such as using ARMAV models for more than two datasets. Google Trend could be a source in order to perform general predictions especially for the concepts in which gathering the data is difficult.

References

1. Majstorovic, V., Zivkovic, S., Djurdjanovic, D., Sabbagh, R., Kvrjic, V., Gligorijevic, N.: Building of internet of things model for cyber-physical manufacturing metrology model (CPM3). *Procedia CIRP*. **81**, 862-867 (2019)
2. Sabbagh, R., Ameri, F.: Thesaurus-guided text analytics technique for capability-based classification of manufacturing suppliers. *ASME J. Comput. Inf. Sci. Eng.* **18**(3), 031009 (2018)
3. Sabbagh, R., Ameri, F.: Supplier clustering based on unstructured manufacturing capability data. In: *Proceedings of ASME 2018 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, pp. V01BT02A036--V01BT02A036. American Society of Mechanical Engineers, Quebec City, Canada (2018)
4. Sabbagh, R.: *Semantic text analytics technique for classification of manufacturing suppliers*. Texas State University, San Marcos, Texas, USA (2018)
5. Danny Sullivan, Search engine land, <https://searchengineland.com/google-now-handles-2-999-trillion-searches-per-year-250247>
6. Carneiro, H. A., Mylonakis, E.: Google trends: a web-based tool for real-time surveillance of disease outbreaks, *Clin. Infect. Dis.*, 49(10), 1557–1564 (2009)
7. InsideGoogle: <http://google.blognewschannel.com/archives/2007/07/30/success-google-trends-updated/> (2007)
8. Bierens, H. J.: *Armax models*. (1988)
9. Pandit, S. M., Wu, S.-M., and others, 1983, *Time Series and System Analysis with Applications*, Wiley New York.
10. Sakamoto, Y., Ishiguro, M., and Kitagawa, G.: *Akaike information criterion statistics*. Dordrecht, Netherlands D. Reidel, 81 (1986)
11. Stoica, P.: A test for whiteness. *iee trans. automat. Contr.*, **22**(6), 992–993 (1977)