Research Article

Nowcasting Mobile Games Ranking Using Web Search Query Data

Yoones A. Sekhavat

Faculty of Multimedia, Tabriz Islamic Art University, Hakim Nezami Square, Azadi Blvd., East Azerbaijan, Tabriz, Iran

Correspondence should be addressed to Yoones A. Sekhavat; sekhavat@tabriziau.ac.ir

Received 26 December 2015; Accepted 23 August 2016

Academic Editor: Michael J. Katchabaw

Copyright © 2016 Yoones A. Sekhavat. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In recent years, the Internet has become embedded into the purchasing decision of consumers. The purpose of this paper is to study whether the Internet behavior of users correlates with their actual behavior in computer games market. Rather than proposing the most accurate model for computer game sales, we aim to investigate to what extent web search query data can be exploited to nowcast (contraction of “now” and “forecasting” referring to techniques used to make short-term forecasts) (predict the present status of) the ranking of mobile games in the world. Google search query data is used for this purpose, since this data can provide a real-time view on the topics of interest. Various statistical techniques are used to show the effectiveness of using web search query data to nowcast mobile games ranking.

1. Introduction

Not only does the web contain valuable data that can be integrated and exploited in different applications [1, 2], but also the activities of users when they search on the web can be useful to estimate the real-world activities in a variety of contexts. This web search data has potential to nowcast the current status of activities, since this data is available when search activity happens [3]. The big question is “whether the population’s Internet tendencies can represent their subsequent behavior.”

Research has shown that no matter whether a customer seeks knowledge on specific attributes of a product or knowledge on how a particular product compares relative to others, the Internet is the primary source of information. Each search for a product on the Internet is a valuable piece of information about an individual’s intentions to make a purchase [4]. Since online search is a measure of interest for a topic [5], terms searched by consumers on the web can provide valuable indicators of consumers’ interests. Exploiting this knowledge to make business predictions can result in a big change in business decision-making [6]. As discussed in [5], search statistics for a product can represent the interest level for that product and consequently help in predicting the sales of that product. In [7], it is shown that changes in query volumes for terms related to influenza are indicators of changes in current numbers of influenza cases. Based on this idea, Google Flu Trends, which is used to diagnose outbreaks of influenza in the United States, was developed [7]. The results showed that this tool can predict outbreaks of influenza in the United States 7 to 10 days before releasing CDC’s (Centers for Disease Control and Prevention) surveillance report [7]. Although not all of the people searching for influenza related topics, symptoms, and treatments are ill, an increase in the search for terms related to this topic is added together and shows a trend.

In this paper, we show how web search data about mobile games can be an indicator of the world ranking of mobile games. More specifically, we aim to study whether the web search data about mobile games extracted from Google Trends can contribute to the forecasting of the ranking of mobile games reported by App Annie (a business intelligence company and analyst firm that produces market reports for the apps and digital goods industry). Generally, time series regarding the ranking of mobile applications are not reported frequently (e.g., on a monthly or quarterly basis) due to the cost or convention reasons. Because of fast changes in the mobile games market, it would be desirable for developers and investors to make it possible to maintain
a current estimate of the value of these time series before it is officially reported or revised. This technique which is called “nowcasting” aims to forecast a current value instead of a future value. In the nowcasting model we propose, the past behavior of the series being modeled (mobile games ranking time series) as well as the values of more easily observed signals (web search query data) occurring in the same period of time is used to predict the current ranking of mobile games.

In a preliminary study in [9], we showed that web search data of mobile games correlates with the market demands of mobile games. We showed that this correlation can be used to nowcast the current status of mobile games where no public report is available regarding the ranking of computer games. More specifically, to find potential areas of interest for game developers in Iran, we studied the trends of web search data for Iranian mobile games. However, in this paper, we show how web search data of mobile games can be used to rank the top mobile games in the world.

The results of this study can reveal market trends for mobile games and make it possible to identify the potential areas of mobile games. Certainly, it is possible to build more sophisticated forecasting models than what we use for this purpose. However, instead of any methodological advances, we argue that the models we propose can serve as baselines to help analysts get started with a basic prediction of trends in this industry.

2. Related Work

Web search query data was used in [10] to forecast the present unemployment rate in the United States. Examining the correlation between web search data about oil price and the sales of electricity was studied in [11]. The correlation between search data about terms related to finding jobs and unemployment rate in the United States was considered in [12]. In [13], web search data was exploited to nowcast the private consumption in the United States. In [11], the authors showed that web search data can be more potential to predict indicators for subsequent consumer purchases in cases when consumers start planning purchases, a considerable time before their actual purchase.

In [14], Internet browsing data was used to inform practitioners about real-time consumer behavior in automobile purchases in Chile. In [15], using regression on Google Trends data showed that the volume of searches for cinema related terms has the potential to forecast cinema admissions models. In [16], using Google Trends data about environment-related phrases, it is shown that the public interest in environmental issues, conservation, and biodiversity has decreased since 2004.

3. Web Search Query Data on Google Trends

In comparison to surveillance reports that are generally published with delays, web search data can be available timely. Moreover, while traditional surveillance methods use sampling techniques that might affect the accuracy of the results, Internet users do not represent a random sample (according to the Internet World Stats, the penetration rate of Internet usage in 2015 in the world is 46.1%). Releasing users search queries on Google through a publicly accessible interface (Google Trends) was a big step toward using this valuable source of information for different types of prediction in various areas. This valuable information shows its importance if we take into account the fact that, according to NetMarketShare, Google owns 70.69% of search engine market share.

We employ Google Trends as a source of web search data for mobile games. Google Trends allows gaining access to information about the volume of web searches that were performed by users for different search terms (or a combination of terms), relative to the total number of searches completed by Google over time. The input into Google Trends is a search query (e.g., “Star Wars”), and the output is a time series of the relative popularity of that query over the selected period. Google logs every single query into the Google database, which is accessible through Google Trends interface (https://www.google.com/trends/). Taking into account the fact that people generally enter terms into this search engine because they are interested in something about them makes this database a valuable business intelligence source.

Web search data from Google Trends is publicly available at a weekly frequency. Techniques that aggregate related keywords are used in this tool. Increasing the volume of web queries results in increasing the average of queries related to other queries, which consequently leads to sensitivity in detection of changes in future search volume trends. Google Trends addresses this problem by normalizing the results. Instead of releasing the actual search number, the normalized data is available for users, which is based on the total number of search queries in a particular location during the time period being examined. This normalization takes into account any trends from growth in the total number of Internet users. The normalized data is rescaled to an index with a range of 0 to 100.

4. Mobile Games Market

According to Gartner research group [8], global video game sales will reach $111.1 billion by 2015, where mobile games with $22 billion are surpassing PC games. This report states that mobile games are the fastest-growing segment of the market, in comparison to PC games, handheld video games, and video game console, where the game market revenue for mobile game is doubled between 2013 and 2015. The pervasiveness of smartphones as well as the growth in mobile device technologies is driving these trends faster than before. According to this report, the majority of mobile application buyers are from the United States and Europe, where smartphone penetration is considerable and users in these areas have the means to purchase games on multiple platforms.

According to Big Fish Games [17], revenue from games and game content is now surpassing revenue from movie Box Office sales ($10 billion per year), trying to pass television viewing as well. In this report, mobile games are highlighted
4.1. Mobile Games versus PC Games. We argue that web search data is correlated with market trends in the mobile game industry. As a first study, to show the effectiveness of using web search data as an indicator of trends in the game market, the results of the search for “mobile game apps” and “PC games” are shown in Figure 1. As you can see in this figure, searches for PC games show a decreasing trend, while there is an increasing trend for “mobile game apps.” This is compatible with Gartner’s report [8] regarding the market revenue for computer games shown in Figure 2, in which although both PC games and mobile games revenues are increasing, the speed of increase for mobile games is considerably more than for PC games. This is evidence of the fact that the changes in the volume of searches for a topic can be an indicator of changes in the interest and tendency of people for that topic.

4.2. Correlation between Web Search Data and the Ranking of Mobile Games. We aim to investigate whether the volume of web search data about mobile games is correlated with the popularity and, consequently, the ranking of these games. More specifically, we study whether the volume of web search data for the name of a mobile game extracted from Google Trends can be an indicator of the ranking of these games. The results of this finding can help to nowcast the current status and the popularity of games in regions, where there is a lack of official reports.

Trends can be an indicator of the ranking of these games. The results of this finding can help to nowcast the current status and the popularity of games in regions, where there is a lack of official reports.

We selected four mobile games among the top ten mobile games (in terms of the number of downloads) from App Annie (https://www.appannie.com/), which is a company that provides market reports for mobile applications. As an indicator of the popularity of the games, we used the ranking of mobile games produced by App Annie. In particular, we studied Candy Crush Saga, Clash of Clans, My Talking Angela, and 8 Ball Pool, for which Google Trends data is also available. To show the correlation between web search data and the ranking of these mobile games, we mapped web search dates and ranking dates. As shown in Figures 3, 4, 5, and 6, there is a close relationship between the web search data for each game and the corresponding ranking history for the time periods investigated. The differences between starting dates in different diagrams are because the games have been released at different dates.

As shown in Figure 3, Google Trends search index for Candy Crush Saga and the ranking of this game in App Annie have a similar decreasing trend since the release of this game in 2013. This correlation is more explicit in the case of Clash of Clans (see Figure 4), where an increasing trend from the
beginning of 2013 to 2014 is followed with a slight decrease to the beginning of 2015 and then a sharp increase from 2015 to the end of 2015. Accordingly, Google Trends search index for *My Talking Angela* and the ranking of this game in App Annie have a similar decreasing trend (see Figure 5). In the case of 8 Ball Pool (shown in Figure 6), an increasing trend in both Google Trends data and ranking history is followed with a decrease from the beginning of 2015.

### 4.2.1. Identifying Correlations

Although correlation between Google Trends search index and ranking history from App Annie can be seen in Figures 3, 4, 5, and 6, we need a formal method to prove this claim. To this end, we use Granger causality test, which is a statistical hypothesis test to determine whether one time series can contribute to forecasting another [18]. Formally, a variable $X$ Granger-causes $Y$ if $X$ values provide statistically significant information about future values of $Y$. More specifically, $Y$ can be predicted in a better way using the histories of both $X$ and $Y$ in comparison to the use of only the history of $Y$. This technique regresses each variable on lagged values of itself and the other explanatory variable. According to regression model,

$$Y_t = \sum_{j=1}^{m} (\alpha_j Y_{t-j}) + \sum_{i=1}^{n} (\beta_i X_{t-i}) + D_t + \epsilon_t. \quad (1)$$

When the coefficients of the $X_t$ values are zero, then the series $X_t$ fails to Granger-cause $Y_t$. In this regression model, $D_t$ is a deterministic term, $\epsilon_t$ is the random error term, $\alpha_j$ is the coefficient on the lagged $Y$ values, and $\beta_i$ is the coefficient on the lagged $X$.

### Table 1: Granger causality test result for the top 20 games in App Annie ($P = 0.05$). Games in which the Google Trends search index does not provide statistically significant information about the future values of the ranking are marked with *.

<table>
<thead>
<tr>
<th>#</th>
<th>Game name</th>
<th># of examined points</th>
<th>F</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Candy Crush Saga</td>
<td>111</td>
<td>4.62951</td>
<td>0.03355</td>
</tr>
<tr>
<td>2</td>
<td>Clash of Clans</td>
<td>110</td>
<td>9.91220</td>
<td>0.00211</td>
</tr>
<tr>
<td>3</td>
<td>My Talking Angela</td>
<td>50</td>
<td>3.57057</td>
<td>0.04476</td>
</tr>
<tr>
<td>4</td>
<td>8 Ball Pool</td>
<td>77</td>
<td>13.20322</td>
<td>0.00050</td>
</tr>
<tr>
<td>5</td>
<td>Piano Tiles*</td>
<td>85</td>
<td>1.05844</td>
<td>0.85428</td>
</tr>
<tr>
<td>6</td>
<td>Subway Surfers</td>
<td>121</td>
<td>10.02277</td>
<td>0.00328</td>
</tr>
<tr>
<td>7</td>
<td>Color Switch*</td>
<td>65</td>
<td>1.61014</td>
<td>0.24453</td>
</tr>
<tr>
<td>8</td>
<td>Temple Run</td>
<td>96</td>
<td>19.90538</td>
<td>0.00213</td>
</tr>
<tr>
<td>9</td>
<td>Traffic Rider</td>
<td>94</td>
<td>17.69072</td>
<td>0.00297</td>
</tr>
<tr>
<td>10</td>
<td>Pou</td>
<td>83</td>
<td>12.90426</td>
<td>0.00166</td>
</tr>
<tr>
<td>11</td>
<td>Basketball Stars*</td>
<td>49</td>
<td>1.22456</td>
<td>0.30063</td>
</tr>
<tr>
<td>12</td>
<td>Clash Royale</td>
<td>113</td>
<td>10.23644</td>
<td>0.00285</td>
</tr>
<tr>
<td>13</td>
<td>Stack*</td>
<td>97</td>
<td>3.94716</td>
<td>0.08219</td>
</tr>
<tr>
<td>14</td>
<td>slither.io</td>
<td>64</td>
<td>6.65071</td>
<td>0.00670</td>
</tr>
<tr>
<td>15</td>
<td>Disney Crossy Road</td>
<td>89</td>
<td>4.80256</td>
<td>0.02357</td>
</tr>
<tr>
<td>16</td>
<td>My Talking Tom</td>
<td>74</td>
<td>2.79033</td>
<td>0.04819</td>
</tr>
<tr>
<td>17</td>
<td>Hungry Shark World</td>
<td>102</td>
<td>61.19054</td>
<td>0.00025</td>
</tr>
<tr>
<td>18</td>
<td>Dream League Soccer 2016</td>
<td>89</td>
<td>4.64344</td>
<td>0.01555</td>
</tr>
<tr>
<td>19</td>
<td>Candy Crush Jelly Saga</td>
<td>75</td>
<td>147.517</td>
<td>0.000085</td>
</tr>
<tr>
<td>20</td>
<td>Geometry Dash Meltdown</td>
<td>95</td>
<td>13.16245</td>
<td>0.000299</td>
</tr>
</tbody>
</table>
Ideally, all mobile games for which App Annie rankings are available must be examined in order to find whether there are correlations between the Google Trends search index and the ranking history from App Annie. However, this is not practical due to the large number of mobile games available in app stores. Because of this limitation, 20 games among the top mobile games of 2016 were selected. Using Wessa [19], we computed Granger causality for Google Trends search index \(X_t\) and App Annie ranking \(Y_t\). The results of these tests are shown in Table 1. According to the \(P\) values in this table, Google Trends search index values provide statistically significant information about the future values of the ranking of the mobile games in 80% of the cases (16 out of 20). As shown in Table 1, in the case of four games (including Piano Tiles, Color Switch, Basketball Stars, and Stack), Google Trends search index does not provide statistically significant information about the future values of the ranking. We attribute this to the general names of the games, as they are not specific names. Consequently, people looking for these names in Google may have something other than games in mind. For example, a user looking for “color switch” may have been looking for techniques to switch color in Photoshop rather than the Color Switch game. Such data impairs the amount of pure searches for the Color Switch game.

4.2.2. Identifying Lags. In the relationship between Google Trends search index time series and App Annie mobile games ranking time series, the App Annie ranking time series may be related to the past lags of Google Trends. To find whether there are time lags between these two time series, we run sample cross-correlation function (CCF). The cross-correlation function of two time series is the product-moment correlation as a function of lag (time-offset) between the series. The cross-correlation function is computed based on Cross-Covariance Function (CCVF). Formally, given two time series \(x_t\) and \(y_t\), where we can delay \(x_t\) by \(T\) samples, CCVF is defined as

\[
\sigma_{xy}(T) = \frac{1}{N-1} \sum_{t=1}^{n} (x_{t-T} - \mu_x)(y_t - \mu_y).
\]  

(2)

Based on this function and given \(\mu_x\) and \(\mu_y\) as the means of two time series \(x_t\) and \(y_t\) with \(N\) samples in each, the cross-correlation function is computed as

\[
r_{xy}(T) = \frac{\sigma_{xy}(T)}{\sqrt{\sigma_{xx}(T)\sigma_{yy}(T)}}.
\]  

(3)

Using Wessa [19], we computed the cross-correlation function for Google Trends search index \(X_t\) and App Annie ranking \(Y_t\). Through this test, we were interested to find whether there are time lags between Google Trends search index and App Annie ranking. Among 20 games shown in Table 1, we selected the top four games based on their rankings in 2016 from App Annie. These games include Candy Crush Saga, 8 Ball Pool, My Talking Angela, and Clash of Clans. For Candy Crush Saga (Figure 7), there are nearly equal maximum values between 0 and −10. The correlations in this region are negative, indicating that an above-average value of Google search index is likely to lead to a below-average value of App Annie ranking about 5 weeks later. In the case of My Talking Angela (Figure 8), the average of time lag is around 10 weeks. For Clash of Clans (Figure 9) and 8 Ball Pool (Figure 10), the lag is around one week. Although the time lags are different, the correlations for all of them are negative, representing the notion that Google search index is likely to lead to a below-average value of App Annie ranking around 5 weeks later.

4.2.3. Forecasting Mobile Games Ranking. We showed the correlation between Google Trends search index and App Annie ranking for mobile games. We also showed that the above-average value of Google Trends search index is likely to lead to a below-average value of App Annie ranking about 5 weeks later. Now, we aim to investigate to what extent Google Trends search index can be used to forecast the future rankings. To this end, we employed multiple regression technique. Generally, multiple regression, which is an extension of simple linear regression, allows predicting the value of a variable based on the values of two or more other variables. In our case, game ranking is our dependent variable, where we aim to predict its values. Google Trends
search index is our independent variable (a.k.a. the predictor or explanatory variable).

In multiple linear regression, the relationship between independent variables and a dependent variable is modeled by fitting a linear equation to observed data. In this model, values of independent variable \( x \) are associated with values of the dependent variable \( y \). Given \( p \) explanatory variables \( x_1, x_2, \ldots, x_p \), the regression line is defined as \( \mu_y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p \), where this line describes how \( \mu_y \) changes with independent variables. The model is expressed as data = fit + residual, where the “fit” term represents the expression \( \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p \) and the “residual” shows the deviations of the observed values \( y \) from their means \( \mu_y \) (shown by \( \varepsilon \)). Consequently, multiple linear regression, given \( n \) observations, is

\[
y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \varepsilon_i, \quad i = 1, 2, \ldots, n.
\]  

The results of running regression on Google Trends search index and application ranking for the top four games based on their ranking in 2016 from App Annie are shown in Figures 11, 13, 15, and 17 for Candy Crush Saga, Clash of Clans, My Talking Angela, and 8 Ball Pool, respectively. In addition to representing the actual values and interpolated values based on these data, residuals, which are deviations of the observed values from their mean, are shown in a histogram of residuals (Figures 12, 14, 16, and 18).

4.3. Nowcasting Virtual Reality Games. Another important trend in computer games market is games based on virtual reality and augmented reality. Using technologies related to virtual and augmented reality is one of the fastest growing areas in computer games that must be considered by practitioners in the game industry. Acquiring Metaio (https://techcrunch.com/2015/05/28/apple-metaio/) (specialized in augmented reality technologies) by Apple and Oculus (https://www.oculus.com/) (specialized in virtual reality solutions) by Facebook shows how big IT firms are moving in this direction. Social media games are another fast growing segment in this industry that takes into account broadband connectivity and networking features to enhance social experiences. As expected, Google Trends also verifies this claim, where tendency to search for virtual reality games has increased in recent years (see Figure 19).

Introducing virtual reality devices such as Oculus Gear VR for Samsung phones as well as many other Gear VR devices makes it possible to experience the virtual reality using mobile phones without the need for a complicated device. Google Cardboard is a simple paper version of this tool that simply allows having the same experience. Introducing these tools has resulted in a significant change in mobile games, where many new mobile games are currently developed based on this technology. Various mobile games such as KioskAR [20] are designed based on augmented
Although mobile games based on virtual reality and augmented reality are not currently among the top mobile games, the growth in this area is significant.

Using Google Trends, we extracted the data for the four top virtual reality tools including Samsung Gear VR, HTC Vive, Oculus Rift VR, and Microsoft HoloLens. Samsung Gear VR, which is powered by Oculus, allows using the Samsung Galaxy smartphone as its processor and displaying VR scenes. The Galaxy handset is put in front of the lenses (using a micro USB dock). HTC Vive is a VR headset made in collaboration with Valve at MWC. Oculus Rift is the virtual reality headset which is currently owned by Facebook. This device is plugged into PCs DVI and USB ports and tracks head movements in order to provide 3D imagery using its stereo screens. Microsoft HoloLens is a virtual/augmented reality tool that combines real world elements with virtual "holographic" images (this device is not released yet). Although some of these tools are not released yet, as shown in Figure 20, searches for them on the Internet, which represent demands and tendencies to use them, are increasing based on Google Trends.
6. Conclusion

In this paper, we showed how web search query data about mobile games (extracted from Google Trends) has correlation with market demands for mobile games and, consequently, the ranking of these games (extracted from App Annie) in the world. We used statistical techniques including cross-correlation function and Granger causality test to show this correlation for the 4 top mobile games extracted from App Annie based on the number of downloads. Finally, regression techniques were used to nowcast the current ranking of mobile games based on web search data extracted from Google Trends for these mobile games. Based on these findings, we argue that this correlation can be used to nowcast the current status of mobile games in regions where no report is published publicly about the ranking and the popularity of mobile games.

Although our model is simple, it shows the potential of using this technique to nowcast the overall status of mobile games market for developers and investors in real time. For future work, we aim to enhance our experiments and use regression techniques to quantitatively predict the present status and the ranking of mobile games, where no ranking history is available in advance.

Competing Interests

The author declares that there are no competing interests regarding the publication of this paper.

References
