Baidu Index Analysis based on Time-series Data--Take the league of legends as an example

Jun You [1], Bo Han [2, a]

[1]Department of business administration, Chongqing University of posts and telecommunications, Chongqing, China

[2]Chongqing University of posts and telecommunications, Chongqing, China

[a]648161787@qq.com

Abstract

This article main research content is aimed at "the league of legends" game, through the use of Baidu index time-series is studied, in order to detect the popularity of "the league of legends" in China as the change of time trend, by removing the extreme data of error, using ARMA（Auto-Regressive and Moving Average）， model can more accurately measure in the life cycle of the game is how to develop to the end.

Keywords

Baidu index，The league of legends，Time-series，ARMA model.

1. Introduction

Baidu index is a public website of sharing data, based on Baidu researches of a tremendous amount of website users, that shows how often a search-term is entered relative to the total search-volume across various regions of China. It is one of the most significant platform of statistics analysis on the internet during the age of big data. As largest Chinese search engine around this world, Baidu Index offer the free massive data analysis service. The internet users derive information about the scale and the ups and downs trend over time by searching a keyword (Zhongming 2007). There is horizontal axis of the main graph represents time, starting from 2006, and the vertical is how often a term is searched for relative to the total number of searches, nationally. Below the main graph, popularity is broken down by regions and cities. It is possible to refine the main graph by region and time. From the specific graph, people learn about the characteristics of Internet users who pay attention to these words and the geographical distribution of these words.

With here comes the age of the Internet, the internet search is the main method to gain the information. Baidu index is one of the application of big data.as long as users log in Baidu index website, everyone can get their required data (Liu Jiahui, Peter& Elin Dolan ,2010). Baidu index not only applied in commercial field, but also applied in some scholars such as scientific research. In order to make research results more convincing, Baidu index is an effective method to transform the qualitative analysis to quantitative analysis. The background of Baidu index is Baidu search, and users spread throughout China. It is means that data is very huge. Although these data may be take some errors about the retrieval sampling problem and approximation algorithm, it is certain that the trends of Baidu index are scientific. Therefore, as the means of analyzing and studying to relative academic field, Baidu index is being used widely in various fields.

League of Legends (abbreviated LOL) is a battle area video game on the internet, which was developed and published by Riot Games company. The game follows a freemium model，which was generally excellent received upon its release in 2009, and has grown in popularity，with an active and largest fan base (Gaudiosi,2012). In 2012, League of Legends was the most popular PC game in USA

and Europe in terms of the number of hours played. During the 2013-2014 period, around over 67 million people played League of Legends per month, 27 million per day, and over 7.5 million during peak hours.In addition, LOL organized the championship series, which widespread around the world. Similar regional competition exists in China (LPL), and then select the best game team to participate the annual World Championship, so it also attracts peoples' attention as well. The 2017 World Championship has 43 million spectators and a total prize pool about over 6 million dollars.

Xiaoyou Zhou (2008) believes that it is efficient way to get the economic information using the Baidu index; Some economists used Baidu index to forecast economic situation. For example, professor Lu Ting (2011) forecast inflation trend in China by adopting Baidu index, and the conclusion is consistent with the actual results; Some companies also use Baidu index and traditional data statistics combined with the investigation of the market situation, formulating business strategy. "Financial network" said that the bank of Merrill Lynch has cooperate with Baidu to develop "Baidu inflation index"; Feifan (2011) pointed out that Baidu index can provide guide about more keyword selection for SEO or network operators. Baidu index offer the ways to search, discover and track social hot topics and media in public, and the most important thing is to get the key index industry, monitoring how keywords data changed; Li Shan (2008) using Baidu index to analyze the popularity of network space about first batch of 5A grade scenic spot in China and distribution characteristics of the precursory effect. Overall, Baidu is applied in financial industry, tourism industry and so on.

It is common behavior that game players use the search engine to download the game and watch the e-sports game live. When online users search the LOL, certainly, users pay attention to the game. The number of searching keywords is the times of one critical word in search engine. This kind of data directly and accurately measure the degree of concern for this field.

Compared to previous study, this paper has the its values. Firstly, this is the first time to measure the popularity of individual game users by using Baidu index, and it solves the defect that the online concerns cannot be directly measured in previous studies. At the same time Baidu index verifies the feasibility as game players concerned with the agent variable. Secondly, Baidu index reflect the influence about the number of playing this game in the future, expanding its potential application value of online users.

The Baidu index is used as a visualization tool to measure the popularity of LOL. The trend of this online game presents its life span, and help game developers make the strategies of operating game. In addition, Baidu index reflects the demand of game users. Whether one of online games have strong vitality and potential depends on the demand of people. According to the new way of leisure and entertainment, people pay attention to LOL, thereby stimulating its development, which contributed to the formation and development of the network game industry chain. In the whole chain of supply and demand, the game users are the source of interest and the start of the value flow. Therefore, the most important thing of LOL development is that game company how to make the strategy of consumer product, price, marketing and market behavior, related industries and enterprises. Baidu index of LOL give the data to observe, and it is easy to give the suggestions to marketing team to attract the new users and stable existing consumer groups. Baidu index also give the data support of the game flow during the operating time. Based on analyzing these data, we can see the attention degree of network users and potential game users compared the active users in the game. Thus, it can predict the trend of development.

2. Data Description

The data is derived from Baidu index from 1/1/2012 to 27/11/2017.
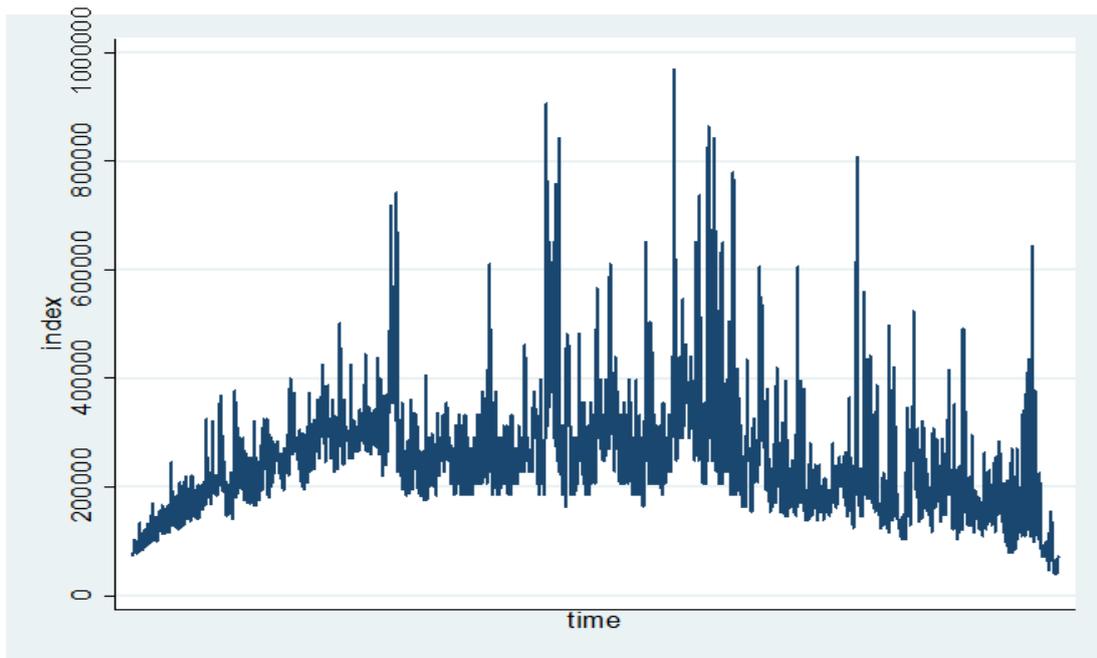
Figure 1

From figure 1, the whole line follows a general pattern which went up till 2015 and declined after then. Every electronic game has a limited life. What interest us it how this game grows or comes to an end. So, we only consider the latter problem in this article. Then we drop the data before 10/11/2015. Because the index often changes sharply day by day. We adjust the data by smoothing it, which means taking the average of five-day.
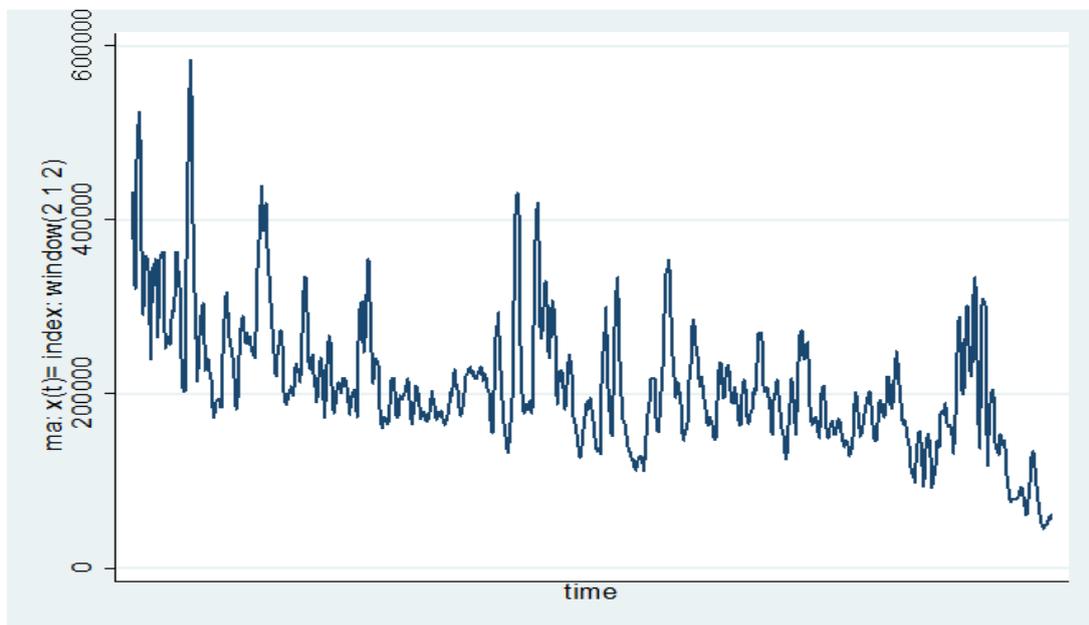
Figure 2

Figure 2 can show us a clearer pattern and avoid the influence of some extreme data. Also, we can find that the whole line is declining slowly, which indicate that we also must detrended the line.
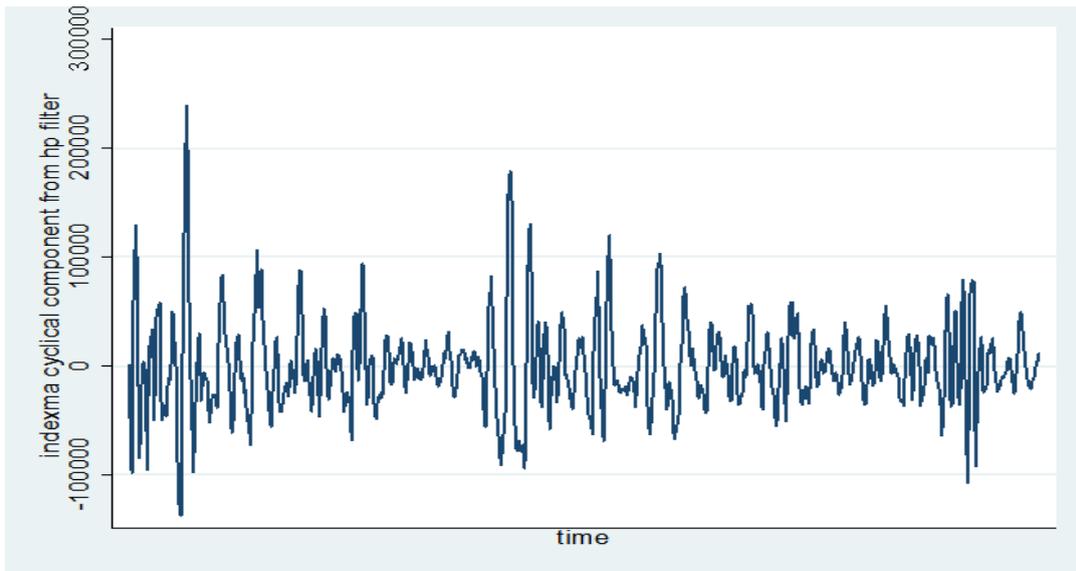
Figure 3

Table 1

|	mean	std	skewness	kurtosis	ADF
Index_mahp	2.59*e-06	40836.72	0.9763516	6.751453	0.0000 (-8.928)

(index_mahp means that data has been detrended and smoothed,)

According to table 1, the detrended line also pass the stationary test. from figure, we notice that data is not symmetric and inclined to be negative. This result could be caused by hp filter. The filter may not detrend the line well as the original line is generally declining.
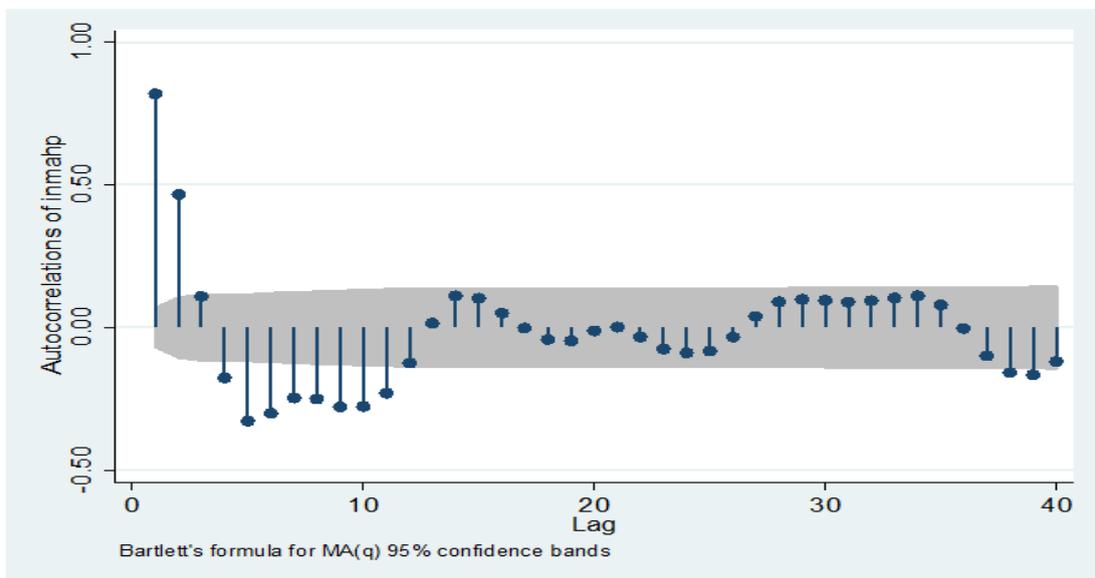
Figure 4

From figure 4, the data is not very persistent. Autocorrelation function is not significant after 10 lags and dies out slowly after. Meanwhile, the behavior of partial autocorrelation function indicates that there may exist a MA part in the real model. The PAC dies out slowly as the autocorrelation function, which can be found in appendix. So, a ARMA model is suggested to model this index.

What' more, compared to standard normal distribution, this index has a positive skewness, which means the data has a longer right tail. And its kurtosis of 5.1 imply that the distribution has a fat tail. To summary, index of lol tend to be more likely to achieve extreme large numerical value. As an

electronic sports game, every year there are many competitions hold all over the world. During those periods the Baidu index would climb to a local peak periodically.

3. Stylized fact

Index is not persistent and both autocorrelation function and partial correlation function dies out slowly. Compared to standard normal distribution, the distribution of Baidu index has a longer right tail while its tails are fatter. The index has a declining trend. The index has a very large variance, which means the data changes dramatically.

4. Model selection

From the autocorrelation function and partial correlation function, we assume a ARMA model is needed. And from information criteria(appendix), for the AR part only, lag 3 or 4 is recommended. Therefore, we separately construct ARMA (3,3), ARMA (3,4) and ARMA (4,4) to see which one fits the data best.

From the result we can find the ARMA (3,3) give us best result. All the coefficient is statistically significant. For ARMA (3,4) and ARMA (4,4), the coefficient of ma part is all not significant, which means the model is overparameterized. All the result can be found in the appendix.

Result

|	Coefficient	Standard Error	p-Value
AR			
L1	1.2482	0.0513	0.000
L2	-0.8704	0.0725	0.000
L3	0.1716	0.0456	0.000
MA			
L1	0.0502	0.0478	0.294
L2	0.6547	0.0306	0.000
L3	0.5565	0.0465	0.000

5. Diagnostic Checking
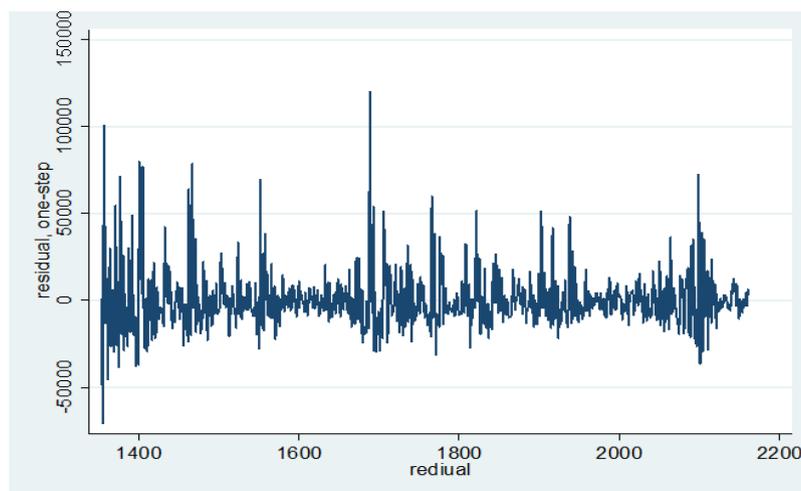
Figure 4

Finally, we check out the behavior of the residual. The autocorrelation function and partial correlation function indicate that residual almost behave like white noise, which imply that the model works well.

6. Conclusion

There are some potential problems with this article. First of all, since the Baidu corporation actually provide no access to download index data. We have to approximate the data with the help of some computer software. So, there would be estimation error about the data while the error is estimated to be less than 1%. Secondly, the life of a game does not perform like a linear trend. The remaining users could be more persistent than those who quit this game. And the forecast of this model could underestimate the life of this game. What 'more, unlike typical sports games, the electronical games have the potential to revive by publishing new version and other sources. And this kind of information do not exist in past data or explained simply by the behavior of white noise. At last, Baidu index only take into consideration the searching volume of website Baidu. In fact, there are other searching engine website such as google and 360. And we have not method to figure it out that the behavior of users of Baidu represent the whole potential uses of this game.

In this article, the model applies to Baidu index of Legend of Leagues. We use Baidu index to estimate the popularity of this game and try to forecast its developing trend based on this index. About the reliability of Baidu index has not been proven among researchers so that it is not generally used in scholar articles. Therefore it is difficult to refer to other scholars' article to implement mine.

In future research, it is well known that any product has its life cycle, and how to maximize it in a limited life cycle is a subject worthy of study. Mentioned in this article the game hero alliance, can clearly see a gradual decline in its popularity, how in the case of a gradual decline in popularity, continue to improve the benefits and research scholars can think about a topic, not just for game products, any product within the life cycle, how from mature to recession in the meantime, to maximize interests, is a very worth studying topic.

7. Appendix

ARMA results

ARMA(3,3)

Sample: 1354 - 2161   Number of obs   =   808

                              Wald chi2(6)   = 11796.12

Log likelihood = -8971.631   Prob > chi2   =   0.0000

```
------------------------------------------------------------------------------
           |              OPG
    inmahp |    Coef.   Std. Err.    z    P>|z|   [95% Conf. Interval]
-------------+----------------------------------------------------------------
inmahp     |
     _cons |  288.0697  3436.135   0.08  0.933   -6446.63   7022.77
-------------+----------------------------------------------------------------
ARMA       |
       ar  |
     L1.  |  1.248191  .0512812   24.34  0.000   1.147682   1.348701
     L2.  | -.8704493  .0725245  -12.00  0.000  -1.012595   -.728304
     L3.  |  .1715589  .0456167    3.76  0.000    .0821517   .2609661
           |
       ma  |
     L1.  | -.0502038  .0478433   -1.05  0.294   -.1439749   .0435672
     L2.  |  .6546842  .0305822   21.41  0.000    .5947442   .7146243
     L3.  |  .5565173  .0465222   11.96  0.000    .4653353   .6476992
```

```
-------------+----------------------------------------------------------------
      /sigma |   15969.7  235.9856   67.67  0.000    15507.18   16432.22
```

ARMA(3,4)

Sample: 1354 - 2161   Number of obs   =   808
                                       Wald chi2(7)   = 9515.47
Log likelihood = -8839.157   Prob > chi2   =   0.0000

```
--------------------------------------------------------------------------------
             |           OPG
      inmahp |    Coef.  Std. Err.    z   P>|z|   [95% Conf. Interval]
-------------+----------------------------------------------------------------
inmahp       |
       _cons |  300.8909  3986.956   0.08  0.940    -7513.4    8115.182
-------------+----------------------------------------------------------------
ARMA         |
          ar |
         L1. |  .5172449  .0359175  14.40  0.000    .4468479    .587642
         L2. | -.1986985  .0398767  -4.98  0.000   -.2768554  -.1205416
         L3. | -.1426806  .0340052  -4.20  0.000   -.2093296  -.0760316
             |
          ma |
         L1. |  1.011173  28.79029   0.04  0.972   -55.41676   57.43911
         L2. |  1.014182  18.54676   0.05  0.956   -35.33679   37.36516
         L3. |  1.011166  11.51301   0.09  0.930   -21.55392   23.57625
         L4. |  .9999933  36.57384   0.03  0.978   -70.68341    72.6834
-------------+----------------------------------------------------------------
      /sigma |  13445.96  245905.9   0.05  0.478          0   495412.6
```

ARMA(4.4)

Sample: 1354 - 2161   Number of obs   =   808
                                       Wald chi2(8)   = 12825.64
Log likelihood = -8828.337   Prob > chi2   =   0.0000

```
--------------------------------------------------------------------------------
             |           OPG
      inmahp |    Coef.  Std. Err.    z   P>|z|   [95% Conf. Interval]
-------------+----------------------------------------------------------------
inmahp       |
       _cons |  277.1491  3343.493   0.08  0.934   -6275.976   6830.274
-------------+----------------------------------------------------------------
ARMA         |
          ar |
```

```
   L1. |  .4952423  .0359942   13.76  0.000    .424695   .5657895
   L2. | -.2310325  .0383407   -6.03  0.000   -.3061788  -.1558861
   L3. | -.0542795  .0460742   -1.18  0.239   -.1445832   .0360242
   L4. | -.1665438  .0323745   -5.14  0.000   -.2299968  -.1030909
       |
    ma |
   L1. |  1.008316  110.6863    0.01  0.993   -215.9328   217.9495
   L2. |  1.018887  69.36153    0.01  0.988   -134.9272   136.965
   L3. |  1.016619  27.85886    0.04  0.971   -53.58575   55.61899
   L4. |  1.005132  137.5431    0.01  0.994   -268.5744   270.5846
-------------+----------------------------------------------------------------
   /sigma |  13218.92  904501.6   0.01  0.494        0     1786009
```

The behavior of residual

```
LAG     AC      PAC     Q     Prob>Q  [Autocorrelation]  [Partial Autocor]
-------------------------------------------------------------------------------
1     0.8197  0.8198  544.88  0.0000      |------           |------
2     0.4673 -0.6239  722.2   0.0000      |---           ----|
3     0.1100 -0.0043  732.03  0.0000       |                |
4    -0.1751 -0.1300  756.99  0.0000      -|               -|
5    -0.3263  0.0771  843.78  0.0000     --|                |
6    -0.3005  0.1914  917.46  0.0000     --|               |-
7    -0.2456 -0.4945  966.75  0.0000      -|             ---|
8    -0.2491 -0.1396  1017.5  0.0000      -|               -|
9    -0.2786  0.0062  1081.1  0.0000     --|                |
10   -0.2762  0.1322  1143.6  0.0000     --|               |-
11   -0.2287  0.0847  1186.6  0.0000      -|                |
12   -0.1243 -0.2447  1199.3  0.0000       |               -|
13    0.0155 -0.0595  1199.5  0.0000       |                |
14    0.1107 -0.1130  1209.6  0.0000       |                |
15    0.1035 -0.0640  1218.4  0.0000       |                |
16    0.0516  0.2159  1220.6  0.0000       |               |-
17   -0.0007 -0.1991  1220.6  0.0000       |               -|
18   -0.0409 -0.1112  1222    0.0000       |                |
19   -0.0472 -0.1235  1223.8  0.0000       |                |
20   -0.0106  0.0000  1223.9  0.0000       |                |
```

References

[1] Zhongming Ma, Gautam &Olivia. 'Interest-based personalized search.' ACM Transactions on information system,2007,25(1):1-38

[2] Jiahui Liu, Peter& Elin Dolan. 'Personalized news recommendation based on click behavior.' proceeding of the 14 ACM international conference on intelligent user interface,2010:31- 40.

[3] Gaudiosi, John. 'Riot Games' League of Legends Officially Becomes Most Played PC Game in The World'. Forbes,2012,3.

[4] Xiaoyou Zhou. 'resources of the network economy information' [J]. library and information service network using search engines, 2008 ,10: 86-89

[5] Lu Ting.' The Baidu index became a measurement of prediction of China's inflation', 2011, 06. available from http://www.sootoo.com/content/5974.shtml.

[6] Feifan.' How to use the Baidu index and propagate website ,2011,06. available from http://www.idczhuji.com/zz/news/jishuzixun/20110610/31243.html

[7] Li Shan, Qiu Rongxu, Chen Ling. 'Attention degree of cyberspace based on Baidu index: time distribution and its precursor effect.' geography and geographic information science, 2008 (6): 27-29