Google Trends as a method for prediction of suicide rates

Vitaliy Bezsheiko
Journal editor

The new study was intended to determine the validity of this method.

Note

This article was exposed to the machine translation.

Background

About every third person in the world uses the Internet, so it is not surprising that the data on its use can provide important statistical information. Given that about 5% of all queries in search engines are related to medicine and health care, these data can be used to analyze health status.

This idea is not new. Even the Google corporation (now Alphabet) had used search data to predict influenza outbreaks. Service «Google Flu Trends» had been lasted from 2005 to 2015 and is now only available as an archive of data (see 1).

One of the new areas of «Big Data» use lies in psychiatry. It is a prediction of suicide rates to assess the effectiveness of prevention interventions at the population level. The first work on the subject was published in 2010 and now this method and its validity is intensively studied.

Prediction of suicide rates

New work on prediction of suicide was published in Aug. 16 in «PLoS ONE» 2. The authors - members of the University of Vienna, Austria. The main aim of study was to evaluate the validity of such prediction, in other words how well we can predict the level of suicides, based on analysis of data from search engines.

The study had relatively simple design. First were analyzed the real statistics on suicides, taken from the USA, Germany and Austria Databases for the years 2004 -2010. These are then compared with statistics of using terms related to suicide on Google during that period. Were analysed such queries as "suicide," "depression", "how to kill myself," "suicide online", etc. in multiple languages. Analysis of data from the use of such searches performed using «Google Trends».

To assess the relationship between the data from search queries and real picture cross-correlation analysis was used. Total number of statistically significant cross-correlation coefficients, ie predicted and real data had matched, by country was as follows: United States - 9.96%, Germany - 2.29% Austria - 11.43%, Switzerland - 2.86%.

This means that the predictive capacity of suicide rates in case of Google Trends was rather small. The average significant cross-correlation coefficient was 8.34%. This is really only slightly above 5% - a level of 1st type error. In other words, only slightly more accurate prediction than from random.
To ensure objectivity in Table 1 are presented the results of other studies that have confirmed the validity of this prognostic method.

<table>
<thead>
<tr>
<th>Research</th>
<th>Country</th>
<th>Method</th>
<th>Effect size, r *</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ma-Kellams et al., 2016 3</td>
<td>US</td>
<td>correlation and linear regression analysis</td>
<td>large (0.49-0.63)</td>
</tr>
<tr>
<td>Gunn and Lester 2014 4</td>
<td>US</td>
<td>correlation analysis</td>
<td>medium to large (0.31 and 0.61)</td>
</tr>
<tr>
<td>McCarthy, 2010 5</td>
<td>US</td>
<td>correlation analysis</td>
<td>large (0.70 and 0.50)</td>
</tr>
<tr>
<td>Sueki, 2011 6</td>
<td>Japan's</td>
<td>cross-correlation analysis</td>
<td>medium to large (0.25-0.43)</td>
</tr>
<tr>
<td>Yang et al, 2011 7</td>
<td>Taiwan</td>
<td>cross-correlation and linear regression analysis</td>
<td>medium to large (0.27-0.48)</td>
</tr>
</tbody>
</table>

Table 1. The studies, which confirmed the validity for forecasting suicide rates with search queries 2. * The more index is closer to 1, the more precise was prediction.

References