

Seminarprojekt: Advanced Analytics

Frühzeitige Erkennung von Krankheitsausbrüchen mit Google Trends in R

Autor: Tobias Böhm

Problemstellung

Kurzzeitige Krankheitsausbrüche von Infektionskrankheiten können von Krankenhäusern oft erst spät erkannt werden.

Die **Überwachung der Fallzahlen** ist oft **ineffizient**, da vorausgesetzt wird, dass infizierte Menschen sich auch in Krankenhäusern behandeln lassen.

Lösungsansatz

Es wird davon ausgegangen, dass sich Menschen bei Verdacht auf Infektion über Suchmaschinen informieren.

Unter der Annahme, dass größtenteils Betroffene nach dem Krankheitsbegriff suchen, können durch **Messung des Suchinteresses** kurzzeitige Anstiege früh und einfach erkannt werden.

Möglicher Nutzen

Rechtzeitige Einleitung von Eindämmungsmaßnahmen, wie Werbekampagnen zur Aufklärung oder die Bereitstellung von ausreichend medizinischer Ausrüstung.

GitHub

github.com/b0ehmt/gtrendsR_seminarprojekt

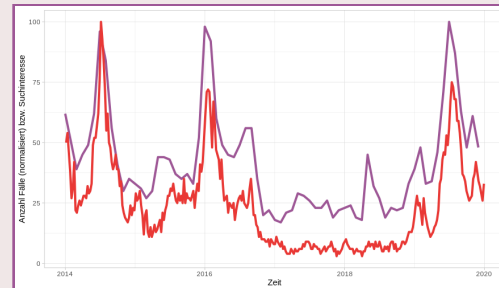
Quellen

1. [Mapping Public Attention with Google Trends](#), Vincentwx
 2. [Detecting Influenza epidemics using search query data](#), Ginsberg et al (2009). Nature
 3. [Weekly Infectious Disease Bulletin](#), Ministry of Health Singapore
 4. [COVID-19 Fälle weltweit](#), Offenes Datenportal der EU. ECDC
- 14.06.2020 | Tobias Böhm | boehmt@hochschule-trier.de

Dengue-Fieber in Singapur

Über das gtrendsR-Package wird das Suchinteresse zum Suchbegriff "dengue" in Singapur von 2014–2020 abgefragt und auf Werte von 0 bis 100 normalisiert.

Durch Einlesen der offiziellen Dengue-Fieber Fallzahlen erhält man wöchentliche Fallzahlen, die ebenfalls auf Werte von 0 bis 100 normalisiert werden.



Google Trends Daten, offizielle Fallzahlen (normalisiert)

Beide Datensätze wurden grafisch abgebildet und auf Korrelation untersucht.

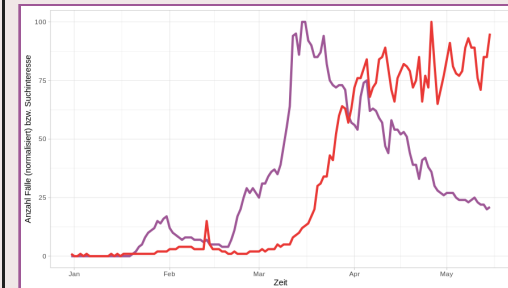
Fazit: Aufgrund einer sehr hohen Korrelation von 0,933 eignet sich Suchinteresse als Indikator hier sehr gut.

→ **Infektionskrankheiten mit geringer Ansteckungsrate und hoher Korrelation sind für diesen Lösungsansatz gut geeignet.** ✓

COVID-19 weltweit

Über das gtrendsR-Package wird das Suchinteresse zum Suchbegriff "Coronavirus" weltweit von Januar–Juni 2020 abgefragt und auf Werte von 0 bis 100 normalisiert.

Durch Einlesen der offiziellen COVID-19 Fallzahlen erhält man tägliche Fallzahlen, die ebenfalls auf Werte von 0 bis 100 normalisiert werden.



Google Trends Daten, offizielle Fallzahlen (normalisiert)

Beide Datensätze wurden grafisch abgebildet und auf Korrelation untersucht.

Fazit: Aufgrund der mäßigen Korrelation von 0,41 ist Suchinteresse als Indikator hier ungeeignet.

→ **Pandemien mit extrem hoher Ansteckungsrate und mäßiger Korrelation sind für diesen Lösungsansatz ungeeignet.** ✗