Predictive models for infectious diseases spread

BDVA - Are we using data in the best way to manage the COVID-19 Pandemic?

26-03-2020
# Motivation

Predictive modelling for infectious diseases spread

<table>
<thead>
<tr>
<th>Preventive policies</th>
<th>Allocation of medical resources</th>
</tr>
</thead>
<tbody>
<tr>
<td>Decision-making</td>
<td>Raise awareness</td>
</tr>
<tr>
<td>Arrangement of production activities</td>
<td>Economical impact</td>
</tr>
<tr>
<td>Seasonal prediction</td>
<td></td>
</tr>
</tbody>
</table>
Disease prediction models

**Risk assessment**

**Event prediction**

**Spatial / dynamical models**

**Event detection**
Mathematical modelling

- SIR model (Susceptible, Infectious, Recovered)
  - Compartment model
  - Human to human transmission
  - Based on differential equations
  - Improvements and elaborations
    - SIS, carriers, SEIR (exposed), etc
Mathematical modelling

- Reproduction number (R0)

- Duration of contagiousness
- Likelihood of infection per contact
- Contact rate

Reproduction number

AI for spread prediction

SIR parameters estimation

- Machine Learning (MLP, CNN, LSTM) can be used to estimate SIR parameters from time series data as a supervised problem.
- ML models learn faster than Approximate Bayesian Computation (ABC).
Big Data for spread prediction

- Spatiotemporal spread prediction using data
  - Traffic Origin Destination (OD) matrix can be used to predict spread considering population flows.
  - OD matrix can be obtained from real data.
  - Models show that there is a positive correlation between the population input and the number of confirmed cases.
  - [https://lexparsimon.github.io/coronavirus/](https://lexparsimon.github.io/coronavirus/)
AI & BigData for spread prediction

- BlueDot spread prediction
  - Published in 2020/01/27
  - Using 2019 data from the International Air Transport Association (IATA)
  - Infectious disease vulnerability index (IDVI) for each receiving country
  - 11 of the cities at the top of their list were the first places to see COVID-19 cases.
  - Journal of Travel Medicine, Volume 27, Issue 2, March 2020, taaa011
AI for spread prediction

▶ Real-time forecasting using auto-encoders
  - Modified auto-encoders forecast number of accumulative and new confirmed cases.
  - Clustering for provinces and cities is done using k-means.
## COVID-19 datasets (global)

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Frequency</th>
<th>Information</th>
<th>Level of detail</th>
<th>Format</th>
</tr>
</thead>
<tbody>
<tr>
<td>John Hopkins University</td>
<td>Daily</td>
<td>Confirmed, deaths, recovered, active</td>
<td>Per country and region (US)</td>
<td>Daily reports and Time-series (csv)</td>
</tr>
<tr>
<td>WHO situation reports</td>
<td>Daily</td>
<td>Confirmed, deaths, transmission classification</td>
<td>Per country</td>
<td>Situation report (pdf) and API</td>
</tr>
<tr>
<td>European Centre for Disease Prevention and Control</td>
<td>Daily</td>
<td>New cases and new deaths</td>
<td>Per country</td>
<td>Situation report (pdf) and Time-series (csv)</td>
</tr>
</tbody>
</table>
## COVID-19 datasets (national)

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Frequency</th>
<th>Information</th>
<th>Level of detail</th>
<th>Format</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spain</td>
<td>Daily</td>
<td>Cases, hospitalized, ICU, deaths</td>
<td>By province</td>
<td>CSV</td>
</tr>
<tr>
<td>South Korea</td>
<td>Daily</td>
<td>Cases, patients info, age, gender, etc</td>
<td>By case</td>
<td>CSV</td>
</tr>
<tr>
<td>Italy</td>
<td>Daily</td>
<td>Positive cases</td>
<td>By province &amp; region</td>
<td>CSV</td>
</tr>
<tr>
<td>Brazil</td>
<td>Daily</td>
<td>Suspects, refuses, cases, deaths</td>
<td>By state</td>
<td>CSV</td>
</tr>
<tr>
<td>Indonesia</td>
<td>Daily</td>
<td>Tested, confirmed, negative, isolated, released, patient metadata</td>
<td>By case</td>
<td>CSV</td>
</tr>
<tr>
<td>India</td>
<td>Daily</td>
<td>Confirmed, cured, deaths, patient metadata</td>
<td>By case</td>
<td>CSV</td>
</tr>
</tbody>
</table>
Challenges

- Access to quality, accurate, and complete data
- Small amount of available data
- Effects of quarantine and actions
- Variability over the time
- Variability between regions
- Compliance with privacy regulations
Challenges

“Right now the quality of the data is so uncertain that we don't know how good the models are going to be in projecting this kind of outbreak”

Marc Lipsitch, epidemiologist at the Harvard T.H. Chan School of Public Health
Thank you

For more information please contact:
daniel.calvo@atos.net
adrian.arroyo@atos.net