Clustering of Covid-19 Time Series

Lisa Pucknat, Alexander Zorn

Lab Development and Application of Data Mining and Learning Systems:
Data Science and Big Data

24st November of 2020

UNIVERSITÄT BONN

Problem Definition

Task: Analysis of Covid-19 pandemic

- ▶ Usage of dataset with daily Covid cases
- ▶ Clustering algorithms for time-series to find clusters by countries and timespans
- ▶ Prediction of future cases using cluster analysis of the results

Dataset

European Centre for Disease Prevention and Control
An agency of the European Union

All topics: A to Z | News & events | Publications & data | Tools | About us

Home > All topics: A to Z > Coronavirus > Threats and outbreaks > COVID-19 > Situation updates on COVID-19 > COVID-19 data > Daily update of new reported cases of COVID-19 by country worldwide

COVID-19 data

Daily update of new reported cases of COVID-19 by country worldwide

Download the daily number of new reported cases of COVID-19 by country worldwide

▶ Each entry contains the number of new cases/deaths reported per day and per country.

▶ ECDC collects data available from health authorities worldwide on a daily basis
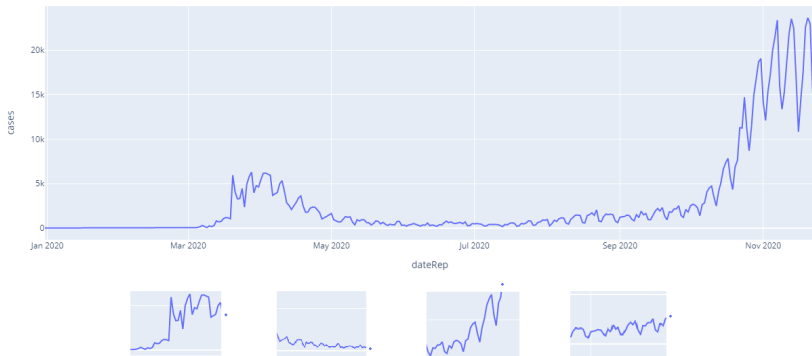
Clustering

Time-series clustering is used to define a grouped structure of similar objects in unlabeled data based on their similar features.

We distinguish between:

1. Whole time-series clustering := Group similar countries into same group
2. **Subsequence clustering := Division of time-series data into intervals using sliding windows technique**
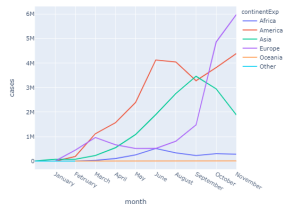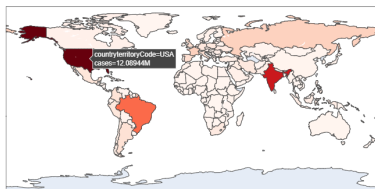
Subsquence clustering

Visualization

Usage of **Plotly | Dash** for Visualization

▶ Python framework for building web analytic applications

Next Steps

- ▶ Preparation/Refinement/Augmentation of data
- ▶ Implementation of different clustering techniques for prediction
- ▶ Search results for possible sources for mistakes
- ▶ Integrate prior knowledge
- ▶ Iterate

References I

[1] Mohammed Ali et al. "Clustering and Classification for Time Series Data in Visual Analytics: A Survey". In: *IEEE Access* 7 (2019), pp. 181314–181338.

[2] *ECDC Daily covid-19 cases.*
     https://www.ecdc.europa.eu/en/publicationsdata/download-todays-data-geographic-distribution-covid-19-cases-worldwide. Accessed: 23-11-2020.