T4C: A Framework For Time-Series Clustering-As-A-Service

Code & Paper

Alessandro Falcetta, and Manuel Roveri
{alessandro.falcetta,manuel.roveri}@polimi.it

POLITECNICO MILANO 1863

TL;DR

Time-series clustering is hard; T4C is a Python framework which can automatically ingest and cluster time-series datasets, using various models/tools, presenting the results on a web dashboard.

Background

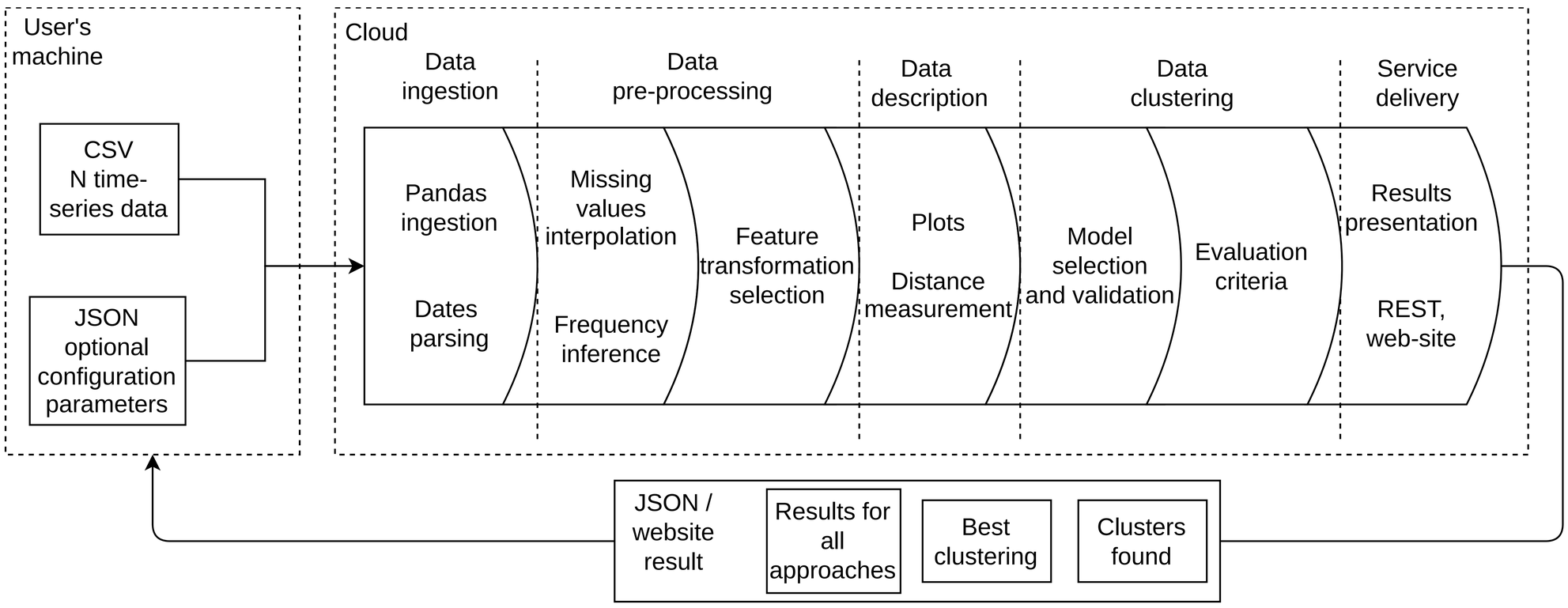
- Time-series clustering is an unsupervised data mining technique whose goal is to organize time-series into groups based on their similarity;
- Different methods for time-series clustering exist, according to the way in which they consider the input data (i.e., *shape* or *observation* based, *feature* based, *model-based);*
- Several distance measures can be used to give a quantifiable indicator on the similarity of two time-series;
- T4C includes all these aspects in a single framework.

Introduction

- T4C is a Python framework being a general and user-friendly solution to offer “time-series clustering as-a-service”.
- T4C can be used providing a JSON file containing a set of parameters and a time-series dataset. T4C will automatically cluster the time-series dataset, showing the results either through a web dashboard or a REST endpoint;
- T4C can also be served interactively, through a website.

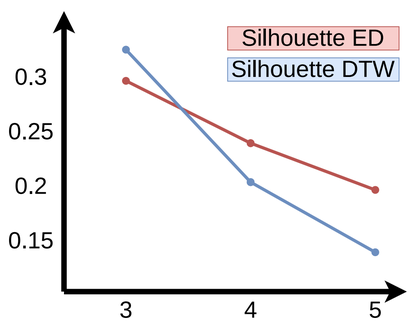
T4C, which is released to the scientific community as an open-source project, has been applied with promising results to the COVID-19 pandemic spread in Italy.

Architecture

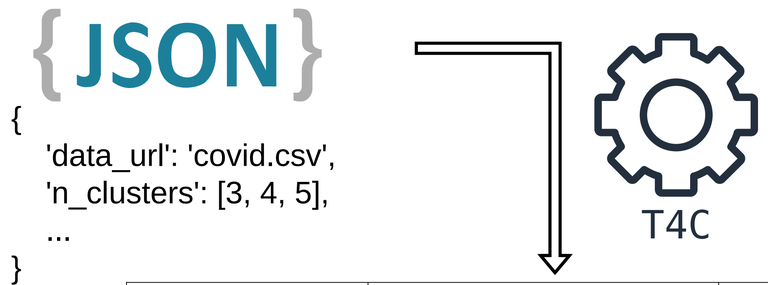


Results

1. Data on the daily "New cases" of Covid-19 have been taken for a large number of countries;
2. The time-series are clustered using T4C;
3. The results show how different countries can be grouped, according to the behaviour of the Covid-19 spread



The number of clusters impacts the clustering metrics!



Cluster 1	Cluster 2	Cluster 3
Argentina	Afghanistan	Albania
Brazil	Andorra	Algeria
France	Angola	Armenia
Germany	Antigua and Barbuda	Australia
India	Bahamas	Austria
Italy	Barbados	Azerbaijan
...	...	...

Future works

The future works will encompass the use of new clustering models, the application on T4C on industrial use-cases, and the extension on different time-series tasks (e.g., classification).

Acknowledgments

The authors would like to thank Dr. Uriel Guadarrama Ramirez for his valuable help.