Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
25th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems

Forecasts of Covid-19 evolution by nearest epidemic trajectories detection

Tomasz M. Rusin

Abstract

Robust method of short-term forecast of Covid-19 epidemic in small administrative units (districts) is proposed. By identifying similar sections of epidemic evolutions in the past it is possible to obtain short-term forecast of epidemic in given district. Examples of one and two-weeks forecasts for three cities in Poland during third epidemic wave (March and April 2021) are shown. Difference between epidemic evolutions in third wave and previous waves caused by Covid B.1.1.7 UK variant is observed. Proposed algorithm allows one to manage epidemic locally by entering or releasing anti-Covid restrictions in groups of small administrative units.

© 2021 The Authors. Published by Elsevier B.V.
This is an open access article under the CC BY-NC-ND license (https://creativecommons.org/licenses/by-nc-nd/4.0)
Peer-review under responsibility of the scientific committee of KES International.

Keywords: Covid-19; Epidemic forecast; Short-time forecast; Distance between trajectories

1. Introduction

In 2020 and first half of 2021 the Covid-19 epidemic in Poland proceeded in three main waves. First wave started at the beginning of March 2020 and lasted to half of May 2020. Three-months stabilization period finished at the beginning of September 2020 and the number of infections gradually increased having its apogee at the end of October 2020. Next stabilization lasted to the beginning of March 2021 and the maximum of third epidemic wave was observed at the beginning of April 2021. Then the epidemic gradually decreased until the end of April 2021, being time of preparing of this paper.

There are two reasons making difficult to forecast Covid-19 epidemic: management of epidemic by the government, and fast virus mutations. First reason relates to restrictions introduced by the government such as: social distance requirement, closure of hotels, restaurants, shopping malls, and others. These limitations allow one to reduce virus transmission between people and effectively reduce number of new infections. Because of high social and financial costs the restrictions were cyclically introduced and released depending on epidemic phase. Second reason is that first
and second epidemic waves were caused by Wuhan virus mutation, while third wave was dominated by British virus mutation B.1.1.7 that is more transmissive than the previous one [1].

In this work we propose method for a short-time forecasting of Covid-19 epidemic which is robust in respect to above difficulties. In this method one forecasts course of the epidemic in small administrative units (districts) taking into account historical epidemic evolutions in other districts. Since all districts experience the same governmental restrictions, and inhabitants of all districts are infected by the same disease, the epidemic pattern in given district may repeat patterns occurring previously in other districts. This method is analogous to short-time forecasts in meteorology: if the atmospheric front passes through the city and gives 10 cm of rainfall, one can expect that in a neighboring city it would also give about 10 cm of rainfall, but few hours later.

In standard approach one describes epidemic course by epidemic models. In classical SIR model [2] one introduces three nonlinear differential equations relating changes of susceptible (S), infected (I) and recovered (R) populations with volumes of these populations. In SIR model there are two parameters related do virus transmission probability: the average time between contacts and the average time until removal of infection. The dimensionless combination of these parameters \( R_0 \) is called the basic reproduction number, and it is main parameter characterizing epidemic evolution in many epidemic models [2]. In more advanced models as, e.g., SEIR model one considers more “states” of population, i.e., the exposed population (E) and one introduces more parameters to the model, see [2].

It is known that the probability of getting a disease is not constant in time, which is assumed in epidemic models, but it varies in time because of governmental restrictions. As the epidemic progresses, reactions to it may change the frequencies of contacts between people and strongly modify probability of getting a disease. Introduced restrictions such as masks, social distancing and lockdown alter the contacts between people in a way to reduce the speed of the epidemic. On the other hand, some people ignore restrictions by joining illegal meetings or they do not respect social distancing. As a result, it is practically impossible to determine correct values of virus transmission probability and consequently, values of parameters entering to equations of epidemic models. One is also not able to predict future values of these parameters. As an example, in exactly solvable epidemic model in Ref. [3], in the rising part of first epidemic wave in Italy a value \( R_{0,0} = 2.65 \) was used. At the maximum of epidemic wave \( R_{0,1} = 1.45 \) was taken, while for the descending part of epidemic wave \( R_{0,2} = 0.85 \) was applied. Thus \( R_0 \) decreased three times during the first wave of the epidemic in Italy.

A possible alternative to epidemic models is use of machine learning methods. In these methods one collects large amount of data related to epidemic evolution and to the society affected by a disease. Data are collected at the lowest possible level (district, communities, or single person level) and then, by using machine learning techniques, one develops algorithms modeling or predicting epidemic course. Algorithm described in this work is also based on machine learning approach and it treats epidemic evolution line as an one-dimensional trajectory. Using standard functions measuring similarity between trajectories our algorithm searches in the past trajectories which are most similar to the selected trajectory. Then, knowing next elements of already found trajectories the algorithm forecasts most probable next elements of the selected trajectory. In proposed algorithm there are no parameters that are directly related to virus transmission probability, and therefore this algorithm works well when actual and future values of virus transmission probabilities are not known.

The work is organized as follows. In the next section we propose algorithm for preparing a short-term forecast of epidemic course in given district. Then we apply this algorithm to forecast epidemic evolution in three big cities in Poland during third wave of Covid-19 epidemic. In Section 3 we discuss obtained results. The work is concluded by Summary. Technical aspects related to data acquisition, transformations and details of distance functions between trajectories are left to the Appendix.

2. Forecasting algorithm and results

Poland is divided in \( N_{dis} = 380 \) administrative units, called districts. The main idea of the algorithm is based on two assumptions:

1. There exist pairs of districts with short-scale (usually few weeks) similarity of epidemic evolution.
Fig. 1. Example of epidemic evolutions in two districts. Black lines: number of daily infections. Epidemic evolution in District 1 is few days ahead of epidemic evolution in District 2. Blue section in District 2 line is the training section. Blue section in District 1 line indicates section of epidemic which is most similar to training section. Solid magenta sections for District 1 line can be used as forecast of epidemic in District 2 (dotted magenta sections).

2. If evolution of epidemic in first component of pair is ahead of evolution in second component of pair, then epidemic evolution in first component of pair can be used as forecast of epidemic evolution in second component of pair.

The above idea is illustrated in Figure 1, where we plot number of new daily cases in two districts. In both districts the epidemic evolves in a similar way, and the similarity is most pronounced in pair of sections indicated by blue lines. This pair includes six consecutive days. First component of pair evolves nine days earlier than second component of pair. In District 1 the number of new infections gradually increases in six-day period covered by blue line, and it continues to increase in next three days. Then it stabilizes for few days with some oscillations. Two magenta sections in District 1 line indicate further course of epidemic in District 1, which can be used as a forecast of epidemic course in District 2 indicated by two dotted magenta lines.

2.1. Preprocessing

In preprocessing one eliminates missing data, outliers, and non-representative districts, e.g., those with discontinuous data. Districts with small number of observations are also eliminated. All quantities are replaced by their 7-day averages per 100k inhabitants, see Discussion. Our algorithm can be used for any epidemic measure as, e.g., number of daily or cumulative new infections, number of deaths, active cases, convalescents etc.

2.2. Algorithm

Here we concentrate on 7-day average of new Covid-19 infections per 100k inhabitants. Consider a section of epidemic evolution in district $D_{tr}$ between dates $d_{beg}$ and $d_{end}$ and we call it the training section $S_{tr}$, see Figure 1. The duration of this section is $N_{tr} = d_{end} - d_{beg} + 1$ days. Our goal is to forecast epidemic course in district $D_{tr}$ for $N_{fc}$ days after the end of $S_{tr}$, i.e., for days $d_{end} + 1, \ldots, d_{end} + N_{fc}$. Let $d_0$ be first day of epidemic in Poland which is common for all districts. To find sections of epidemic course that are most similar to $S_{tr}$ we select all sections $S_i$ (in all $N_{Dis}$ districts) fulfilling the following conditions:
For training section $S_{tr}$ there is approximately

$$N_S \approx N_{Dis}(d_{end} - N_{fc} - d_0)$$

sections $S_i$ fulfilling above conditions. As an example, by taking $N_{Dis} = 380$ districts, $d_{end} = 2021 - 05 - 01$ (end of third epidemic wave), $d_0 = 2020 - 09 - 01$ (beginning of second epidemic wave) and $N_{fc} = 14$ days one obtains $N_S \approx 82,080$ sections $S_i$ which should be compared with $S_{tr}$.

Let $f_k(S_i, S_j)$ be $k \geq 1$ functions measuring distance between two sections of epidemic course, see Table A.1 in Appendix. Then, for each $k$ one performs the following steps:

- Calculate $N_S$ distances $f_k(S_{tr}, S_i)$ between all pairs $\{S_{tr}, S_i\}$.
- Select $N_{few} \ll N_S$ pairs with the shortest distances $f_k(S_{tr}, S_i)$. Usually there is $1 \leq N_{few} \leq \times 10^{-3} N_S$.
- Extend each of $N_{few}$ sections $S_i$ by $N_{fc}$ subsequent elements. The elements $N_{tr} + 1, \ldots, N_{tr} + N_{fc}$ of $S_i$ are filled by subsequent elements of epidemic evolution line.
- Create an empty section $S_p$ having $N_{tr} + N_{fc}$ elements.
- Fill all elements of $S_p$ by medians of $N_{few}$ sections $S_i$. As a result, elements $1, \ldots, N_{tr}$ of $S_p$ approximate epidemic evolution in $S_{tr}$, while elements $N_{tr} + 1, \ldots, N_{tr} + N_{fc}$ of $S_p$ are forecasts of epidemic evolution for $S_{tr}$.

If there is only one function $f_k(S_i, S_j)$ then the above algorithm gives the final results. Otherwise, the final forecast is the median of forecasts calculated with use of $k$ distance functions. Optionally, the obtained fits and forecasts are multiplied by number of inhabitants in training district.

The above algorithm follows the principle of causality, because the forecast is based only on data from the past. This algorithm has no parameters that are directly related to Covid-19 epidemic as, e.g., to $R_0$, but to improve accuracy of forecasts some parameters entering to the algorithm as, e.g., the length of training section, should be correlated with epidemic wave length, see Discussion.

2.3. Results

Here we present results for actual data and forecasts for Covid-19 epidemic evolution in three big cities in Poland during third epidemic wave. Note that each city is a separate district.

In Figures 2, 3 and 4 we analyze, fit and forecast 7-day average of daily new infections. In Figures 2 and 3 we plot actual number of new infections (black dots), fits for training sections (blue dots) and forecasts (magenta dots) calculated with use of algorithm in Section 2.2. First conclusion from both figures is that for training sections all fits are nearly indistinguishable from actual data. This confirms accuracy of finding the nearest sections of epidemic evolution both for 21 and 28-days training periods. However, the forecasts are not such accurate, and we discuss examples of both accurate and less accurate results.

In panel a) of Figure 2 we show a proper forecast of breaking rising epidemic trend in Warsaw with accuracy of two days. This breakdown was caused by a strong lockdown introduced in Poland in last week of March 2021. The forecasting algorithm found in the past a 35-days pattern with continuous growth of daily new infection finished by a lockdown. This pattern was repeated in panel a) of Figure 2.
Fig. 2. Actual evolution of epidemic in Warsaw, 7-day average (black dots), fit to period section (blue dots) and forecast (magenta dots) for third Covid-19 wave. a) Beginning of third wave; b) decreasing phase of third wave. Vertical axes: number of new daily infections registered in Warsaw. Horizontal axes: day of reporting of infections, see Ref [4, 5].

In panel b) of Figure 2 we analyze the descending part of third epidemic wave in Warsaw. Here our algorithm expects stabilization of daily new infections, but in fact the epidemic continues decreasing trend. The explanation is that in third epidemic wave the government changed rules existing in the past and released lockdown more slowly then for previous epidemic waves. Therefore the epidemic pattern in panel b) of Figure 2 does not exist in the past, and the algorithm gives improper forecast.

Fig. 3. The same as in Figure 2 but for: a) Wroclaw, center of third Covid-19 wave; b) Rzeszow, beginning of third Covid-19 wave.
Fig. 4. Forecast of new infections for day +7 (magenta dots) vs actual values in two cities (black dots). Actual values: seven-day averages. Vertical axes: number of new infections; horizontal axes: dates. For a given date $d$ in the upper panel the training period (7 days) includes days $d - 14$ to $d - 8$, and day $d$ is the seventh day of the forecast.

In panel a) of Figure 3 we show a center of third Covid-19 wave in Wroclaw. Our algorithm correctly fits to actual data in the training section, and correctly predicts a decrease of the epidemic, but it again predicts much slower fall than observed in actual data. The explanation is similar as for the previous case.

In panel b) of Figure 3 we show the rising part of the epidemic wave in Rzeszow. In this example our algorithm correctly predicts both the epidemic stabilization and the slope of its decay, but the algorithm expects maximum of the wave one week earlier than it occurred in practice. The explanation is that the lockdown in Rzeszow was introduced at the same time as in the rest of the country, and it was introduced later than expected from local epidemic evolution.

In Figure 4 we compare forecasts for day +7 with actual values of new Covid-19 infections at this day. The forecasts for day +$d$ are calculated from data in range $[d - 14 \ldots d - 8]$ (upper panel) or in range $[d - 21 \ldots d - 8]$ (lower panel). In both panels the forecasts qualitatively agree with actual Covid-19 evolutions. Quantitatively, there are differences between actual data and forecasts. In the rising phase of the third Covid-19 wave the forecasts underestimate the volume of new infections. Wave maxima are localized with accuracy of two days for Warsaw and about one week for Gdansk. In descending part of the wave both forecasts overestimate the number of new infections. A possible explanation of this observation is that the virus mutation responsible for the third wave of epidemic in Poland is different than the virus mutation responsible for previous waves. This explanation agrees with results of Ref. [1] stating that the third Covid-19 wave in Poland is mostly caused by the so-called British mutation $B.1.1.7$ of Covid-19.

3. Discussion

Here we briefly discuss several aspects of our method. In our algorithm, the use of several distance functions reduces over-fitting. The reason is that the algorithm selects only a small fraction of epidemic sections. It may happen that all these sections are incorrect because they fit to a one-time event or to a data error. But it is probable that other distance function does not fall into this trap and they provide correct forecasts. Median steps in algorithm are voting mechanisms eliminating over-fitting from the forecast.

Use of data standardization (see Appendix) improves accuracy of forecasts for all distance functions. It was found that there is no definitively better or worse distance function between pairs of epidemic sections, see Appendix, but it seems that the distance function based on histograms is slightly less effective than other distance functions. This
observation agrees with recent works, see Ref. [6] who showed a similar accuracy of DTW and EDR classifiers in biomedical applications. Bagnall et al. [7] compared on many training data sets the accuracy of DTW versus other modern classifiers and showed that only few classifiers are significant better than DTW. Therefore we believe that the distance functions used in this work are effective enough to our problem.

The distance function based on area between two lines (AD) is generalization of Euclidean distance between two lines. Euclidian distance is regarded in literature as less accurate than DWT, EDR and LCSS distance functions, but in our calculations the results obtained with use AD have comparable accuracy to those obtained with use of DWT, EDR and LCSS. There are two advantages of AD over other distance functions: i) AD has linear time complexity and ii) it is possible to speed-up calculation of AD between the training section and two neighboring epidemic sections having a non-empty intersection, see Eq. (A.2). The area distance allows one to find effectively segments of epidemic evolution in other districts that are close enough to the training section. Then more accurate algorithms as DWT, EDR and LCSS having square time complexity are applied only for preselected fragments of epidemic curves in other districts.

To explain reasons of improper forecasts arising from accurate fits to training section we consider a training section lasting two weeks and one-week of forecast. The algorithm searches in the past several two-weeks epidemic pattern that are most similar to pattern in the training section. However, identification of two-weeks epidemic sections being close to the training section does not guarantee that the epidemic evolution after the training period will follow the pattern found in the past. In general there are three options. First is that the epidemic continues without change of government restrictions. This occurs when the number of new infections in training section is low, or it is high locally but low in whole country. Second option is that the government repeats anti-epidemic restrictions implemented in the past, e.g., when the number of new daily infections exceeds predefined threshold. The last option is that the government changes the rules used in the past and extends or delays anti-Covid restrictions in different way than in previous epidemic waves. Our forecasting algorithm works correctly for first two cases, but for the last case it gives improper results.

From above discussion it is clear that the main reason of incorrect forecasts is unpredictability of governmental restrictions, and the accuracy of distance functions used in calculations is less important. In Poland the governmental decisions about anti-Covid restrictions are entered in one or two weeks in advance and this is a practical time-scale of accurate forecasts.

Algorithm in Section 2.2 has no parameters directly related to virus transmission probabilities. However lengths of training and forecasting periods should be related to Covid-19 epidemic dynamics. Length of the training period should not exceed the half of epidemic wave. As seen from Figures 2, 3 and Figure 5 of Ref. [3], the epidemic wave lasts about two months, so the training period should not exceed five weeks. Parameter \(d_0\), i.e., the common epidemic “zero day” should be far in the past, and its value allows one to include or eliminate some epidemic waves. Length of the forecast should not exceed length of the training period to avoid spurious forecasts. Experiments show that maximum length of reliable forecast is half of length of the training period. For Covid-19 epidemic in Poland the last condition gives approximately two weeks of reliable forecast.

The algorithm in Section 2.2 can be applied to other quantities characterizing Covid–19 epidemic as: number of deaths, number of occupied hospital beds, volume of active cases etc. Practical limitation in forecasting these quantities in Poland is lack of accurate data in long-term perspective, see Appendix. For above reason we applied our algorithm only to forecasts of new daily infection.

There are several possible applications of algorithm proposed in this work. First, one can obtain realistic short-term forecasts of epidemic in all districts. Then the forecast for Poland is sum of \(N_{\text{dis}} = 380\) forecasts for all districts. Second, forecasts in districts can serve as benchmarks for anti-Covid restrictions applied by the government. One can compare effects of actual restrictions with effects of similar restrictions applied in the past. Next, the decision-makers can obtain recommendations about proper time of entering or releasing the restrictions. Finally, our algorithm allows one to manage epidemic locally by entering or releasing anti-Covid restrictions in group of small administrative units as voivodeships or groups of districts.

4. Summary

In this work we introduced a robust algorithm forecasting Covid-19 epidemic evolution. The algorithm identifies pairs of districts with similar epidemic evolutions in few-weeks intervals and it forecasts evolution of epidemic in
first component of pair on a base of epidemic evolution on second component of pair. The maximum forecast length is on the order of two weeks. The algorithm correctly predicts breaks of epidemic trends caused by introduction or releasing the anti-Covid restrictions in Poland. Examples of forecasts in three cities confirm accuracy of the training phase, but the forecasting phase is less accurate because of unpredictable character of epidemic management by government. Obtained results for third wave of Covid-19 epidemic indicate that this wave differs from previous ones. This difference can be explained by appearance of new Covid-19 mutation in Poland, so called British mutation, which is confined by independent biological research. It is expected that our algorithm can be successfully applied to forecast possible next Covid-19 epidemic waves.

Acknowledgements

This work is element of PROME project and it is realized as part of 13/CNT/2020 contract, module M2.C.5 Project of prototypes of ML algorithms using NN models.

Appendix A.

Here we present auxiliary issues related to the main text.

A.1. Districts characteristics

Since public available data about Covid-19 epidemic in Poland are available on a district level we briefly characterize these units. Poland is divided in three levels of administrative units: 16 voivodeships, 380 districts and 2477 communities. Out of 380 districts there are 314 village-like districts and 66 cities treated as independent districts. The last group includes 66 largest cities in Poland including its capital (Warsaw). In average there is 100k inhabitants per district. Large districts correspond to main cities on Poland, and the largest district is Warsaw (1.76m inhabitants). Some agglomeration are divided into several districts as, e.g., Silesia agglomeration but the other are not. In village-like districts there is usually one larger city and many small villages. Usually, in each district there is at least one hospital, one high-school, one big shopping mall, one sport center and local or governmental administration offices.

A.2. Available data

In Poland there are several statistics of Covid-19 epidemic which are available daily at districts level: number of new infections, deaths, active cases, convalescent, tests etc. However, only first two statistics, namely number of new infections and deaths are collected from the beginning of epidemic. Stable daily reports are available from mid-January 2021, and backward updates of these reports to 2020 − 11 − 25 are done [4]. Data from 2020 − 03 − 01 to 2020 − 11 − 24 are obtained from the civic project in Ref. [5].

The specifics of current reporting methodology is that the numbers reported by Ministry of Health for weekends are five up to ten times smaller than the corresponding numbers during week-days. For this reason one should analyze only 7-day averages (or their multiplies) of epidemic data. Raw data or other averaging periods include spurious oscillations caused by week-based seasonality of epidemic data.

A.3. Distance functions

In implementation of forecasting algorithm in Section 2.2 five distance functions between two epidemic sections are used, see Table A.1. The area distance (AD) between two sections is calculated using trapezoid rule

\[
 f_{AD}(S_1, S_2) = \frac{1}{2} \sum_{n=1}^{N} (A_{12}(x_n) + A_{12}(x_{n-1})) \Delta x = \frac{\Delta x}{2} \left( A_{12}(x_0) + 2 \sum_{n=1}^{N-1} A_{12}(x_n) + A_{12}(x_N) \right),
\]

where \( A_{12}(x) = |S_1(x) - S_2(x)| \) and \( \Delta x = x_n - x_{n-1} \) is constant for all points. Let \( S_3 \) be the epidemic section of the same length as \( S_2 \) but shifted by +1 day. Then sections \( S_2 \) and \( S_3 \) have common elements except \( x_0 \) and \( x_{N+1} \).
Table A.1. Distance functions used for identification of similar sections of epidemic evolution.

| Seq | Shortcut | Description | Standardization | Time complexity | Reference |
|-----|----------|-------------|-----------------|-----------------|-----------|
| 1   | AD       | Area between curves | Yes/No | $O(n)$ | Eq. (A.1) |
| 2   | DTW      | Dynamical Time Warping | Yes/No | $O(n^2)$ | [8] |
| 3   | EDR      | Edit Distance on Real Sequences | Yes/No | $O(n^2)$ | [9] |
| 4   | LCSS     | Longest Common Subsequence | Yes | $O(n^2)$ | [10] |
| 5   | Histogram | Distance between two histograms | Yes/No | $O(n)$ | [11] |

For $f_{AD}(S_1, S_2)$ we calculate the sum in Eq. (A.1) from $x_0$ to $x_N$ while for $f_{AD}(S_1, S_2)$ we calculate this sum from $x_1$ to $x_{N+1}$. When $f_{AD}(S_1, S_2)$ in Eq. (A.1) is already computed, then $f_{AD}(S_1, S_3)$ can be obtained from $f_{AD}(S_1, S_2)$ by adding corrections arising from $x_0$ and $x_{N+1}$ elements of the series

$$f_{AD}(S_1, S_3) = f_{AD}(S_1, S_2) - \frac{\Delta x}{2} A_{12}(x_0) + \frac{\Delta x}{2} |S_1(x_N) - S_3(x_{N+1})| .$$ (A.2)

The above formula speeds up calculations of AD between the training section and many neighboring sections $S_2, S_3, \ldots$. The distance functions DTW, EDR and LCSS are implemented according to algorithms published in corresponding references. In histogram approach one treats two sections as histograms and calculates distance between them using Algorithm 1 in Ref. [11].

For all distance functions the common length of both sections is assumed, and no windowing is applied. Some distance functions require standardization of entering sections. To make standardization of section $S_i$ one first calculates for this section the mean $\mu_i$ and the standard deviation $\sigma_i$. Then for each point $x_j^i$ of section $S_i$ one calculates

$$z_j^i = \begin{cases} (x_j^i - \mu_i)/\sigma_i, & \sigma_i > 0, \\ 0, & \sigma_i = 0. \end{cases}$$ (A.3)

In practice the last case in Eq. (A.3) occurs for sections $S_i$ describing small village-like districts without new infection in given period. In implementation of our algorithm only standardized sections are compared. For EDR and LCSS algorithms a threshold value $\varepsilon \approx 0.4\sigma_i$ is used [6].

References

[1] https://www.news-medical.net/news/20210224/Mapping-the-dominance-of-the-SARS-CoV-2-UK-variant-in-Europe.aspx. (2021).
[2] https://en.wikipedia.org/wiki/Compartmental_models_in_epidemiology. (2021).
[3] Luca Dell Anna. (2020) “Solvable delay model for epidemic spreading: the case of Covid-19 in Italy”, *Scientific Reports* 10: 15763.
[4] Polish Ministry of Health. (2021) https://www.gov.pl/web/koronawirus/wykaz-zarazen-koronawirusem-sars-cov-2.
[5] Michal Rogalski. (2020) “COVID-19 w Polsce”, bit.ly/covid19-poland.
[6] Lu Sha, Mirchevska Gordana, Phatak Sayali, Li Dongmei, Luka Janos, Calderone Richard, and Fonzi William. (2017) “Dynamic time warping assessment of high-resolution melt curves provides a robust metric for fungal identification”, *PLOS ONE* 12: e0173320.
[7] Bagnall, A., Lines, J., Bostrom, A., Large J. and Keogh Eamonn. (2017) “The great time series classification bake off: a review and experimental evaluation of recent algorithmic advances”, *Data Min. Knowl. Disc. 3*: 606-660. See also: arXiv:1602.01711.
[8] https://en.wikipedia.org/wiki/Dynamic_time_warping. (2021).
[9] Chen Lei, M. Tamer Ozsu, and Vincent Oria. (2005) “Robust and Fast Similarity Search for Moving Object Trajectories”, *Proceedings of 24th ACM International Conference on Management of Data (SIGMOD ’05)*: 491–502.
[10] Michail Vlachos, George Kollios, and Dimitrios Gunopoulos. (2002) “Discovering Similar Multidimensional Trajectories”, in *Proceedings 18th International Conference on Data Engineering*: 673–684.
[11] Sung-Hyuk Chaa, Sargur N. Srihari. (2002) “On measuring the distance between histograms”, *Pattern Recognition* 35: 1355–1370.