Adaptive-compartmental model of coronavirus epidemic and its optimization by the methods of artificial intelligence

S. P. Levashkin, O. I. Zakharova
Artificial Intelligence Lab, PSUTI, 443010, Russia, Samara
ai_lab@psuti.ru

S. V. Kuleshov, A. A. Zaytseva
LANI-SPIIRAN, St. Petersburg Federal Research Center of the Russian Academy of Sciences, 199178, Russia, St. Petersburg
kuleshov@iias.spb.su

Abstract. In this work, we develop a systemic approach to the study of a new model of COVID-19 pandemic. The main goal is to minimize the pandemic damage to economy and society by defining the model optimal management parameters. Our approach consists of two main parts: 1) the adaptive-compartmental model of the epidemic (ACM-SEIR) – a generalization of the classical SEIR model and 2) the module to tune ACM-SEIR parameters using artificial intelligence methods (collection, storage and processing of big data from heterogeneous sources) that allow the most accurate adjustment of ACM-SEIR parameters turning it into an intelligent system for decision support called herein iACM-SEIR. We show that among iACM-SEIR parameters, the most important are individual economic, demographic and psychologic characteristics of society and the governmental actions.

1. Introduction
Today the world is faced with the COVID-19 pandemic, which is likely to have long-term consequences for the economies of states and world’s society. This situation can be unambiguously considered as a serious biogenic threat that inflicts tangible damage on socio-economic systems, state formations and society. At the same time, over-pressing scientific tasks are the correction of existing and development of new mathematical models of the epidemics spread and methods of information monitoring not only of the spread of the virus, but also of the socio-economic environment. These models and methods should reflect modern realities and describe both the biogenic threat itself and its consequences as accurately as possible. Therefore, timely monitoring of economic, psychologic and other social processes occurring within "viral environment" will allow more optimal management decisions to be made and the consequences of a pandemic to be minimized by more accurate tuning of the forecast model.

Herein, we describe a hybrid model of the spread of viral diseases, combining methods of machine learning, data mining, and theory of complex systems. Based on the well-known Susceptible-Exposed-Infectious-Removed (SEIR) epidemic spread model [1], we significantly modify it by introducing new processes and parameters that more accurately reflect modern realities (this new model we call adaptive-
compartmental or ACM-SEIR). Then this model is completed with a system of complex information monitoring of social phenomena of the heterogeneous Internet resources aimed at minimizing the damage it brings (in this way our final product is formed – an intelligent adaptive-compartmental model or iACM-SEIR to support management decisions making).

The scientific novelty of this research consists of:
1) Integration of the monitoring system and the model of the spread of viral epidemics;
2) Use of unlimited set of resources rather than a limited set of predefined Internet sources as sources to tune the parameters of unstructured data models;
3) Identification of poorly predictable reactions of society to certain events expressed by Internet content using machine-learning methods.

The rest of the paper is organized as follows. In Section 2, we describe ACM-SEIR model, its main processes and parameters. In Section 3, we study ACM-SEIR using machine learning. In Section 4, we present the information monitoring system based on associative-ontological approach. We use this system to adjust the ACM-SEIR parameters thus transforming it into intelligent one (iACM-SEIR). Section 5 contains some illustrative results obtained using iACM-SEIR model with initial population of Moscow, Wuhan or New York size (13 millions). Finally, Section 6 contains our conclusions and sketches a direction of the future research.

2. Adaptive-compartmental model of the epidemic
The epidemic spread model ACM-SEIR is based on the Susceptible-Exposed-Infectious-Removed (SEIR) model with a total population size $N$ and with two additional classes: $P$ that represents public perception of risk in relation to the number of severe and critical cases and deaths; and $C$ representing the number of cumulative cases (both reported and unreported).

Let $S$, $E$, and $I$ denote susceptible, exposed and infectious populations, and $R$ denote a removed group (i.e., recovered or deceased). The model will also determine the function $b$ – the authorities’ reaction rate or "the effect of government actions". Transmission of infection by external carriers (zoonotic transmission), denoted by $z$, is modeled as a stepwise function that goes to zero after some date from the beginning of the epidemic. Further, only the sustained transmission of COVID-19 from person to person after this date and the population mobility before the self-isolation regime introduction are modeled. Thus, the Adaptive Compartmental Model (ACM) ACM-SEIR is a system of seven strongly nonlinear ordinary differential equations (ODE) to determine the main variables $S$, $E$, $I$, $R$, $N$, $P$, $C$, $b$ as well as a set of parameters to be determined and tuned from data acquired from heterogeneous resources:

$$
\begin{align*}
\frac{dS}{dt} & = - \frac{b_0 Sz}{N} - \frac{b(t)SI}{N} - mS, \\
\frac{dE}{dt} & = \frac{b_0 Sz}{N} + \frac{b(t)SI}{N} - (d + m)E, \\
\frac{dI}{dt} & = dE - (h + m)I, \\
\frac{dR}{dt} & = hI - mR, \\
\frac{dN}{dt} & = -mN, \\
\frac{dP}{dt} & = phI - lP, \\
\frac{dC}{dt} & = dE,
\end{align*}
$$

where

$$
b(t) = b_0 (1 - a) \left(1 - \frac{p}{N}\right)^k.
$$


The main parameters of the ACM-SEIR model are: {initial population size $N_0$ (constant); initial susceptible population $S_0$ – a proportion of $N_0$ (constant); number of zoonotic cases $z$ (a stepwise function); transmission speed $b_0$ (a stepwise function); governmental action strength $a$ (a stepwise function); intensity of responds of the population $k$ (constant); level of mobility $m$ (a stepwise function); mean latent period $1/d$ (constant); mean infectious period $1/h$ (constant); proportion of severe cases $p$ (constant); mean duration of public reaction $1/l$ (constant)}. Note that, in addition to introducing new parameters and functions, ACM-SEIR model can be generalized by adding new equations to the system for other processes such as $S$, $E$, $I$, $R$, etc. Moreover, for this model the number of equations in the system and the number of its parameters are not critical.

3. Study of ACM-SEIR by machine learning approach

ACM-SEIR model, being as close as possible to the real phenomenon of epidemic spread, has neither analytical nor any easily determined numerical solution. Therefore, for a qualitative study of the model, an approach based on deep learning is used.

The general scheme of this approach is as follows. First, the solution of the ODE system is written as a sum of two parts. The first part satisfies the initial conditions and does not contain configurable parameters. The second part is structured so as not to affect the initial conditions. This part includes a feed-forward artificial neural network containing configurable parameters (weights). Thus, by construction, the initial conditions are satisfied, and the neural network is trained to satisfy the system as a whole. Moreover, when calculating the solution, a certain “loss function” is minimized step by step (the solution will be “exact” if this function is equal to 0). Note that to optimize the computing, it is possible to replace not all the right-hand sides of the equations with a neural network, but only their non-linear parts, which cause the greatest problems.

Fragment of code in Python to solve the ODE system and plot the graphics (results of the code execution see in Section 5) is given below:

```python
In [1]:
from scipy.integrate import odeint
import numpy as np
import matplotlib.pyplot as plt
%matplotlib inline
import mpld3
mpld3.enable_notebook()

In [2]:
def plotseird(t, colors, **kwargs):
    f, ax = plt.subplots(1,1,figsize=(10, 6))
    for metric_name, metric_values in kwargs.items():
        color = colors.get(metric_name, 'b')
        ax.plot(t, metric_values, color, alpha=0.8, linewidth=2.5, label=metric_name)
    ax.set_xlabel('Time (days)', labelpad = 12)
    ax.set_ylabel('Population', labelpad = 35)
    ax.yaxis.set_tick_params(length=0)
    ax.xaxis.set_tick_params(length=0)
    ax.grid(b=True, which='major', c='w', lw=2, ls='--')
    legend = ax.legend(borderpad=2.0)
    legend.get_frame().set_alpha(0.6)
    for spine in ('top', 'right', 'bottom', 'left'):
        ax.spines[spine].set_visible(False)
    plt.savefig('last.png')
    plt.show()
```

4. Defining and tuning the model parameters by associative-ontological approach to data acquisition

To adjust the parameters and functions of the model, artificial intelligence methods in digital monitoring of the Internet space [2, 3, 4] are used, in particular, monitoring the psychological state of society, assessing the phsyhographic properties of people by their messages published in the Internet. To collect the dataset for model parameters’ clarifying we take into account the modern features of storing and providing users with media content, including data retrieving and web-scraping technologies [3, 4], as
well as volumes of dynamic content generation per unit of time. We have developed the monitoring and crawled system that includes modules to realize the following functions:

- forming a list of processed pages;
- downloading and analyzing content;
- semantic filtering of content (criteria of usefulness, significance, absence of duplication);
- storing the search database (include datasets);
- organizing search in database.

To process the obtained datasets in the form of natural language texts, an associative-ontological approach (AOP) is used, which includes the theoretical foundations for representing texts in the form of loaded undirected graphs (associative environment of the text). AOP works on the principle of automatic formation of a semantic environment (or, in other words, an associative ontology) of an arbitrary text in real time during the operation of the system [5, 6]. In addition, AOP uses basic standards of the Resource Description Framework (RDF) and Web Ontology Language (OWL) technologies, but not the implementation of the Semantic Web (Linked Data, Linking Open Data) with the pre-formed and verified ontologies of subject areas [7]. AOP, unlike machine learning methods used in NLP approaches, has the property of interpretability of knowledge representation, which allows the possibility of manual editing of automatically created associative ontologies or creating rules for working with them. AOP, being an extension of the n-gram approach, allows the use of developments in probabilistic models for the needs of natural language processing. The developed monitoring system allows generating datasets for additional configuration of the ACM-SEIR model and conversion to the iACM-SEIR model, collecting and analyzing data according to the following parameters:

- indicator of traffic jams;
- indicators of self-isolation of the population (indicator of movement of people around the city);
- information about the weather, weather events that contribute to colds;
- information about the weather, weather events that facilitate the movement of people around the city;
- share of closed shopping malls / shops / catering establishments;
- number of news on the topic – the number in absolute or conventional units;
- class of the population's reaction to messages about the development of the situation with the virus, about objective indicators - a set of numbers showing belonging to the class of reactions;
- class of the population's reaction to reports of anti-epidemic restrictions;
- quantitative assessment of the reaction to official messages (number of likes, reposts, comments).

Also, an element for setting up the model parameters is sentiment analysis – a text classification tool that analyzes an incoming message and determines whether the author's reaction to the topic of the message is positive, negative or neutral [8, 9, 10].

Besides the sentiment analysis, intent analysis, which analyzes the user's intent behind a message and identifies whether it relates to opinion, news, action, complaint, suggestion, rating, or request is used.

These two analyzes are complemented by contextual semantic search, which takes thousands of messages and a concept (such as "COVID-19") as input and filters all messages that match that concept exactly.

Finally, the results obtained are formalized in the form of statistical reports and visualized in the form of graphical charts to reflect the picture of the state and mood of users on COVID-19 problem and tune the parameters of the model.

5. Results
In this section, we present some results obtained using the approaches described in the previous sections.
The test results with initial population $N_0 = 13,000,000$ are presented in the following figures.

Fig. 1. COVID-19 spread graphs with a high level of compliance with individual protection rules, but in the absence of government action.

Fig. 1 shows that with a high level of compliance with the rules of individual protection, the number of infected is about 1% of the total population at the peak of the epidemic.

If the degree of compliance with the rules of individual protection is lower, then the number of people infected with COVID-19 increases sharply as can be seen in Fig. 2.

Fig. 2. COVID-19 spread graphs with low compliance with individual protection rules and in the absence of government action.

Thus, the parameter of compliance with individual protection rules is key in this model of the spread of the epidemic.

The degree of government response is also an extremely important social factor to battle the spread of the virus. Fig. 3 shows the graphs of the COVID-19 spread with high individual responsibility of citizens as well as with active government actions.
In fig. 4 we can see that approximately on 150th day after the 1st peak of the epidemic, in the event of weakening of the government's actions and the reaction of the population, the 2nd wave of the epidemic begins (this exactly happened in Russia in October 2020 after May's peak).

It can be concluded that these two social factors have a significant impact on the number of infections and the coronavirus spread. With an optimal combination of both factors, the simulation results show a decrease in the number of cases from millions to thousands.

6. Conclusion
We have presented a complex hybrid model of an intelligent adaptive-compartmental model of the coronavirus epidemic spread and showed the possibility of optimizing its economic damage by fine-tuning the model parameters. Making decisions to prevent / mitigate threats like COVID-19 requires playing out many possible options with a relatively simple apparatus of interpretation in a short period.

Therefore, we used a combination of modern machine learning technologies, artificial intelligence and system dynamics methods. Such approach allowed us to cope with the main difficulties of this complex task, namely: 1) representing a real world phenomenon, our model contains a very large number of parameters, which, in the end, must be taken into account in order to make the model as close as possible to real processes of epidemic spread; 2) the dimension of the data space required to achieve the same goal is very high, and the amount of data is large. These two circumstances make the traditional mathematical apparatus of statistical analysis, regressions, grid computations, etc. practically inapplicable.

However, modern methods of deep learning, data mining and cloud computing allowed us to cope with the multifactoriality of the problem, the "curse of dimensionality" and the large amount of data required for realistic model tuning.
Note also that the problem under consideration is of a complex interdisciplinary nature and is associated with the development of methodology, tools and technologies for increasing the efficiency of public administration in crisis situations.
In the future research, we are going to monitor and acquire more COVID-19 data from different sources to refine instantly the model parameters.

Acknowledgments
The reported study was partly funded by RFBR, project number 20-04-60455.

References
[1] van den Driessche P 2017 Reproduction numbers of infectious disease models. J. Infect. Dis. Model 2(3) 288–303
[2] Raghavendra T S and Mohan K G 2019 Web mining and minimization framework design on sentimental analysis for social tweets using machine learning J. Proc. Comp. Sc. 152 230–5.
[3] Laliwala Z and Shaikh A 2013 Web Crawling and Data Mining with Apache Nutch (Packt Publishing) p 136
[4] Boeing G and Waddell P 2016 New Insights into Rental Housing Markets across the United States: Web Scraping and Analyzing Craigslist Rental Listings J. of Plan. Ed. and Research arXiv:1605.05397
[5] Kuleshov S V, Zaytseva A A and Markov S V 2015 J. Int Tech. On Transp. 4 40–5.
[6] Kuleshov S V, Zaytseva A A and Aksenov A J 2019 The tool for the innovation activity ontology creation and visualization Advances in Intelligent Systems and Computing vol 763 (Springer, Cham) 292–301
[7] Khan S and Safyan M 2014 Semantic matching in hierarchical ontologies J. of King Saud Un. – Comp. and Inf. Sc. 3(26) 247–57
[8] Patel N V and Chhinkaniwala H 2019 Investigating Machine Learning Techniques for User Sentiment Analysis IJDSST 11(3) 1–12.
[9] Priya C, Santhi K and Vincent P M D 2019 Provision of Efficient Sentiment Analysis for Unstructured Data Information Systems Design and Intelligent Applications (Springer, Singapore) 199–207.
[10] John C A 2012 Using Twitter to measure behavior patterns. J. Epid. 23(5) 764–5.