



Nailia Gabdrakhmanova^{1,*} and Maria Pilgun²

- ¹ S.M. Nikol'skii Mathematical Institute, Peoples Friendship University of Russia (RUDN University), 117198 Moscow, Russia
- ² Institute of Linguistics, Russian Academy of Sciences, 125009 Moscow, Russia; mpilgun@iling-ran.ru
- * Correspondence: Gabdrakhmanova_nt@rudn.ru or gabd-nelli@yandex.ru; Tel.: +7-91-6859-1560

Abstract: The relevance of this study is determined by the need to develop technologies for effective urban systems management and resolution of urban planning conflicts. The paper presents an algorithm for analyzing urban planning conflicts. The material for the study was data from social networks, microblogging, blogs, instant messaging, forums, reviews, video hosting services, thematic portals, online media, print media and TV related to the construction of the North-Eastern Chord (NEC) in Moscow (RF). To analyze the content of social media, a multimodal approach was used. The paper presents the results of research on the development of methods and approaches for constructing mathematical and neural network models for analyzing the social media users' perceptions based on their digital footprints. Artificial neural networks, differential equations, and mathematical statistics were involved in building the models. Differential equations of dynamic systems were based on observations enabled by machine learning. Mathematical models were developed to quickly detect, prevent, and address conflicts in urban planning in order to manage urban systems efficiently. In combination with mathematical and neural network model the developed approaches, made it possible to draw a conclusion about the tense situation around the construction of the NEC, identify complaints of residents to constructors and city authorities, and propose recommendations to resolve and prevent conflicts. Research data could be of use in solving similar problems in sociology, ecology, and economics.

Keywords: cognitive systems; neural networks; differential equations; social media; speech perception

1. Introduction

This paper is an extended version of the paper presented at the 14th International Symposium INTELS '20, which will be published after the conference. The paper proposes an improved and expanded methodology that is tested on new material of urban conflicts.

Differential features of the human brain cognitive systems and artificial intelligence have a rather long history of development [1]. One of the relevant tasks is to analyze social media users' perception of certain situations, events and phenomena using new technologies. The main source of information, in this case, is digital footprints, among which unstructured linguistic data are of importance. Neural network technologies open up new possibilities for studying the specifics of perception of a situation on large volumes of verbal data. Various aspects of linguistic representations in artificial neural networks have been presented in studies over recent years [2–4], and the speech behavior of Facebook users enabled creation of new personality constructs [5]. Various aspects of perception are presented in numerous works, for example, speech perception [6], and Bayesian analysis of data from cross-linguistic studies on color and memory perception [7], etc. It should be noted that the new autoregressive language model GPT-3 with 175 billion parameters marked a new stage in the field of natural language processing technologies [8].



Citation: Gabdrakhmanova, N.; Pilgun, M. Intelligent Control Systems in Urban Planning Conflicts: Social Media Users' Perception. *Appl. Sci.* 2021, *11*, 6579. https://doi.org/ 10.3390/app11146579

Academic Editor: Askhat Diveev

Received: 26 June 2021 Accepted: 14 July 2021 Published: 17 July 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). The analysis of speech behavior and perception of social media actors can be applied in a wide variety of areas, particularly in intelligent control systems gaining popularity in recent years [9–11].

The demand for intelligent control systems explains the emergence of numerous scientific works in this area [12–16] It should also be noted that researchers show great interest in the study of intellectual control of robotic systems [17,18]. In urban projects, intelligent control systems have also taken a strong position; these systems are especially in demand in the structure of a smart city. The accumulation of knowledge about the complexity of the modern urban environment has led to doubts about the effectiveness of management approaches based on institutional, administrative and geographic isolation and has drawn attention to the search for ways of intelligent urban management [19]. Thus, building automation and advanced sensing [20], electronic solutions to reduce social isolation and reconstruction of problem areas [21] have been developed and are actively used; and the social dividends of using information technologies in urban systems were also analyzed [22].

In addition, in the 21st century, the requirements for expertise as an important element of multi-criteria decision support tools designed to guide the most complex and uncertain decisions have become more stringent. The organization and mechanisms of expert decision-making support with the help of E-Expertise and Modern Collective Intelligence obtained a special status [23]. These algorithms and the characteristics of well-reasoned decision-making turned out to be in demand, especially in the field of environmental protection and safety [24]. It is the state of the environment that becomes the most important indicator of the urban system efficiency [25].

The development of the urban structure inevitably collides with the resistance of residents since, to some extent, it affects the symbolic space, the historically formed image of the city. The transfer of a significant part of communications to the digital environment makes it relevant to study conflicts in a virtual environment, develop new techniques and approaches, in particular, those involving text mining, machine learning, and agent-based modeling [26]. In addition, network data turn out to be an indispensable empirical material for the study of social processes, behavioral characteristics, perception of various events and phenomena by residents, and allow real-time analysis on large data sets [5].

It can be assumed that the analysis of social media data and intelligent control systems technologies have great potential for timely response, detection and defusing of urban conflicts, search for compromise solutions to improve the efficiency of urban systems management.

The aims of this study were to develop and test an algorithm for analyzing urban planning conflicts. This paper presents the results of research on the development of methods and approaches to the construction of mathematical and neural network models for analyzing the perception of social networks users on the basis of their digital footprints, as well as for the rapid identification, prevention and resolution of urban planning conflicts necessary for effective urban systems management.

Our research questions are as follows: Will the results of all applied models be consistent? In addition, will the developed approach make it possible to identify conflict zones around urban development projects?

Hypothesis 1 (H1). The approach of combining mathematical and neural network models to analyze the social media users' perception based on their digital footprints allows for the effective analysis of social tension or its absence around urban development projects.

Sections of manuscript: Introduction. 2. Materials and Methods. 2.1. Text Analysis Methods. 2.2. Approaches to Mathematical Modeling. 2.3. Formalization and Mathematical Statement of Problem. 2.4. Mathematical Models. 2.4.1. Mathematical Statistics. 2.4.2. ODE and Its Parameterization. 3. Results. 3.1. Neural Network Semantic Model. 3.1.1. Rating of Digital Platforms According to Actors' Preferences. 3.1.2. Semantic Network Analysis. 3.1.3. Sentiment Analysis of the Content. 3.1.4. Analysis of the Topic Structure

and Content Analysis. 3.1.5. Analysis of the Users' Perception of the Situation. 3.2. Mathematical Modeling Results. 3.2.1. Statistical Analysis of Data. 3.2.2. Autoregressive Models ARIMA. 3.2.3. ODE 1 Models. 3.2.4. ODE 2 Model. 3.2.5. Analysis of Model Accuracy. 4. Discussion. 5. Conclusions. Author Contributions. Funding. Institutional Review Board Statement. Informed Consent Statement. Data Availability Statement. Conflicts of Interest. Appendix A. References.

2. Materials and Methods

The data for the study was obtained from social networks, microblogs, blogs, instant messaging, forums, reviews, videos dedicated to the construction of the North-Eastern Chord (NEC) in Moscow, collected for the period between 1 January 2019 to 20 September 2020, 13,827,304 messages. The dynamics of messages are shown in Figure 1.



Figure 1. Dynamics of messages on the NEC construction.

Data analysis was carried out using the Brand Analytics monitoring and analysis system for social media and mass media.

Several parallel models were used to optimize the analysis of the social media users' perception based on their digital footprints. The results were obtained using the semantic neural network model. In parallel with the neural network model, statistical analysis of experimental data was performed, and dynamic models of processes were built.

The types of models used for data analysis were: a model based on the neural network paradigm of using neural-like elements with temporal summation of signals [27], statistical analysis [28] and differential equations [29].

2.1. Text Analysis Methods

Using a model based on the neural network paradigm of using neural-like elements with temporal summation of signals (the so-called corticomorphic associative memory), the topic structure of the database and the results of summarization were analyzed, which made it possible to extract explicit information. On the other hand, such a network representation of the text made it possible to isolate and interpret the semantic network in the form of a set of interrelated concepts. With the help of the semantic network, implicatures, semantic accents that are most important for actors were analyzed and rated [27].

Sentiment analysis was performed using the Eureka Engine sentiment determination module. The technique is based on a statistical algorithm for conditional random CRF fields using sentiment dictionaries. Sequences of lexemes are used as input data, after

which the algorithm calculates the probabilities of possible sequences of tags and chooses the maximum probable one.

Content analysis was performed in accordance with [30,31] using the AutoMap text mining tool.

For visual analytics, the Tableau platform was used.

2.2. Approaches to Mathematical Modeling

Today, complex processes require mathematical and computer modeling to be managed. The approaches to mathematical modeling have seen substantial change thanks to the explosion in available data and to the advances in machine learning. Complex nonlinear dynamics arise in many fields of science, but uncovering the underlying differential equations directly from observations poses a challenging task. Paper [32] is the first to propose a method that can automatically generate symbolic equations for a nonlinear coupled dynamical system directly from time-series data. The most interesting of the recent studies are papers that dwell upon Gaussian processes [33], physics-informed neural networks, i.e., neural networks that are trained to solve supervised learning tasks while respecting any given law of physics described by general nonlinear partial differential equations [34].

By using state-of-the-art methods to derive differential equations from data, this paper investigates the applicability of such methods to sociological modeling on social media data. However, the problem covered herein has several peculiarities that set it apart. The complexity of this problem stems from its specific features, such as data format. Data is textual. Neural networks and cluster analysis are applied to convert text into numerical data. Section 2.3 addresses the solution. Another issue is that models are built upon data of the object's implicit variables. Implicit variables are defined herein as variables that are functions of the modeled object's state variables. Observational data are known to be very noisy. Models built herein are based on earlier research in sociophysics [35–38]. The problem is relevant in light of the recent events (terror attacks, strikes, etc.).

This paper presents mathematical problems in the context of addressing two main classes of problems: (1) using machine learning to build mathematical models on observational data sets; (2) behavioral analysis of the system based on such models. The developed approaches and models are proven effective by building parallel models and checking the resulting solutions for consistency with the solutions derived from visual data analysis.

2.3. Formalization and Mathematical Statement of Problem

Section 2.1 describes how text mining isolated a cluster of texts mentioning attitude to the subway line project from the entire social media data array. Text mining also split the entire cluster into subclusters based on attitude: negative, positive, or neutral. Filtered and clustered data helped word the problem.

A project is planned that may affect the interests of residents in a specific, limited area. The problem is to build a mathematical model to assess the utility of this project on social media data only without any additional research such as population surveying. The utility is defined herein as the residents' positive attitude to the project.

Let:

- 1. A be the subpopulation with a negative attitude to the project;
- 2. B be the subpopulation with a positive attitude to the project;
- 3. C be the subpopulation with a neutral attitude to the project.

Social attitude to the project could be assessed by analyzing the size of these subpopulations in dynamics. However, such data is not available. Therefore, it is proposed to estimate such subpopulations by means of digital footprint. The models are based on the assumption that the community consists of three subpopulations A, B, C, holding a negative, positive, or neutral attitude to the project. Let N(A), N(B), and N(C) be the sizes of the sets A, B, C, and D be the sum of the disjoint sets D = A + B + C. Subjects may move from one group to another as a result of communication or an event. Therefore, the percentage of each subpopulation may change over time. Figure 2 shows the flowchart of the modeled objects. Let D1 be the set of social media actors whose posts and comments contain one of the three possible attitudes to the project. Assume that the set D is mapped to the set D1,

$$F: D \rightarrow D1.$$

Assume that the mapping F is such that the percentage of negative/positive/neutral groups in D1 fully matches that in D.

The objective is to use digital data of the actors in D1 to model the dynamics of the residents' attitude to the project.



Figure 2. Diagram of the object modeled.

Mathematical Statement of Problem. Let X(t), P(t), Z(t) be the number of posts at the time t that express negative, positive, or neutral attitude to the project, respectively; let S(t) = X(t) + P(t) + Z(t), t = 1, 2, ..., n.

Consider society as a dynamic system. Let the state of the dynamic system at the time *t* be described by the vector Y(t) = (X(t), P(t), Z(t)). The objective is to use observations of X(t), P(t), Z(t) over the period t = 1, 2, ..., n to build a mathematical model to describe the behavior of the dynamic system.

The behavior dynamics of X(t), P(t), Z(t) contains important data: the ratio of these figures; for this reason, the figures are normalized for modeling. The normalization rule is as follows: x = X/S; p = P/S; z = Z/S, then the follows holds: x + p + z = 1, y = (x, p, z), $y \in D \subset R^3$.

To make a managerial decision on the basis of the built mathematical models, introduce the preferential attitude and its numerical representation. Let χ be a non-empty set. Interpret the elements of this set as a possible choice of attitude to the project: \tilde{x} , \tilde{p} , \tilde{z} (negative, neutral, positive). Numerical representation of the preferential attitude \succ is such function $V: \chi \to R$ that

$$y \succ x \Leftrightarrow V(y) > V(x).$$

The function $V(\cdot)$ can be defined in different ways, e.g., as a function of the first differences of a time series.

2.4. Mathematical Models

2.4.1. Mathematical Statistics

Models based on mathematical statistics are used herein as parallel models. Parallel models can be seen as something similar to cross-validation in machine learning.

The quality of the solution to the symbolic regression problem was assessed by means of the calculated paired coefficients.

Considering the data {x(t), p(t), z(t), t = 1, ..., n} as time series, we can build autoregressive models [39,40]. Autocorrelation functions and partial autocorrelation functions were found on the initial data of the time series x(t), p(t), z(t). Studies showed linear ARMA models could be built for these series. ARMA are dynamic statistical models. This study used ARMA models to analyze the behavior of the dynamic systems and to cross-validate the ODE models. Section 3.2.2 presents the output of the resulting ARMA models.

2.4.2. ODE and its Parameterization

Dynamic systems (DS) are conventionally modeled by differential equations. In order to use differential equations as a solution, one needs to find a differential equation from the observational data. This is an inverse problem. Inverse problems are conventionally solved by analytical or numerical methods. Artificial neural networks are on the rise as a method for finding differential equations from data. The methods are convenient to use in parametrizing differential equations; another advantage is that they handle data noise pretty well—this research utilized machine learning to build differential equations.

Consider the system in question as a dynamic system S, the variables (x(t), p(t), z(t)) being its response to some input u(t). Several events (bots, media) can be considered inputs. Let *P* be the modeled object; *U* be the managerial decisions, y = (x, p, z) be the system's response to a control input. With these definitions, state the key problems pertaining to the dynamic system (DS):

This paper investigates the third problem. The other two have been solved in part. DS behavior was described by a system of autonomous differential equations:

$$\frac{dx}{dt} = f_1(u, x, p, z)$$

$$\frac{dz}{dt} = f_2(u, x, p, z)$$

$$\frac{dp}{dt} = f_3(u, x, p, z)$$
(2)

The right-side function in (2) was found by two methods, resulting in the ODE 1 and ODE 2 models. ODE 1 models were built in the local domain. ODE 1 models were linear systems of ordinary differential equations; ODE 2 models were nonlinear systems of ordinary differential equations.

The occurrence rate of bots or posts in the media can serve as the variable u(t) in (2). The variable was not used in modeling for this research; however, the variable can be introduced if data is available.

Let us state the mathematical problem for problems written in (1) as forward and inverse ODE solution problems.

Statement of the forward problem.

where $F = (f_1, f_2, f_3), y(t) = (x(t), p(t), z(t))^T, y_0$ are the initial conditions. For this forward problem, from the given initial conditions find $y(t_1), y(t_2), y(t_3), \ldots$. The forward problem is stated to solve Problems 1, 2.

Statement of the inverse problem.

Readings $y(t_1)$, $y(t_2)$, $y(t_3)$, ... are known for some points. From these readings, find $F(\cdot)$ in (3). To solve the inverse problem means to solve the problem of identifying the system, i.e., Problem 3.

ODE 1 Model

The ODE 1 model was built by cluster analysis using the Nelder–Mead method. Figure 1 shows the non-normalized time series in dynamics. It makes clear that the initial data do not have identical time series. To describe DS behavior in terms of differential equations, the observation interval [0; n] was split into several subintervals by cluster analysis. A model was built for each cluster. This approach is similar to modeling in the low-mod domain, where the behavior of a dynamic system can be described with only a few variables [41]. As a result, we were able to choose the simplest writing for the right

side in (3) to describe the processes and to track the change in the model parameters in order to analyze the processes in the system.

DS behavior in each cluster was described by a system of autonomous differential equations:

$$\frac{dx}{dt} = a_{11}x + a_{12}p + a_{13}z
\frac{dz}{dt} = a_{21}x + a_{22}p + a_{23}z
\frac{dp}{dt} = a_{31}x + a_{32}p + a_{33}z$$
(4)

The parameters of the model (4) were estimated by the Nelder–Mead method [42]. The following function served as the optimization criterion:

$$F = \sum_{i=1}^{T} \left[\left(\frac{dx_i}{dt} - (a_{11}x_i + a_{12}z_i + a_{13}p_i) \right)^2 + (dz_i - (a_{21}x_i + a_{22}z_i + a_{23}p_i)^2 + (dp_i - (a_{31}x_i + a_{32}z_i + a_{33}p_i))^2 \right] \rightarrow min$$
(5)

Find the optimum of F in (5) for all a_{ij} , $i = \overline{1,3}$, $j = \overline{1,3}$.

When estimating the model parameters, set the threshold h. If the optimization (5) for some *i*,*j* produces estimates: $|a_{ij}| < h$, then assume $a_{ij} = 0$.

ODE 2 Model

The ODE 2 was built using symbolic regression. Control problems necessitate equations that describe the behavior of the dynamic system over the entire modeled domain. The structure $F(\cdot)$ in (3) is unknown. Let DS be describable as a nonlinear dynamic system. The nonlinear function F(y(t)) represents the dynamic constraints that define the system's motion equations. Given that epidemiological models are well-suited for such problems in sociology, the structure of the functions f_i was written in the form of second-degree polynomials. The first step is to combine data into three large matrices:

$$Y(t) = \begin{bmatrix} y^{T}(t_{1}) \\ \vdots \\ y^{T}(t_{n}) \end{bmatrix}, \quad \dot{Y}(t) = \begin{bmatrix} \dot{y}^{T}(t_{1}) \\ \vdots \\ \dot{y}^{T}(t_{n}) \end{bmatrix}, \quad \Theta(Y) = \begin{bmatrix} 1 & y(t_{1}) & P(y(t_{1})) \\ \vdots & \vdots & \vdots \\ 1 & y(t_{n}) & P(y(t_{n})) \end{bmatrix},$$

where $1 = (1 \ 1 \ 1)$, $P(y(t_1)) = (x^2 \ p^2 \ z^2 \ xp \ pz \ xz)$.

Each element in the matrices is a vector. Introduce the coefficient matrix:

$$\Xi = (\xi_1, \xi_2, \xi_3)^T, m = 11, \Xi \in \mathbb{R}^{m \times 3}$$

where

$$\xi_i = (\xi_{1i}, \xi_{2i}, \dots, \xi_{mi})^T.$$

Then the system (3) will be rewritten as:

$$\frac{dY}{dt} = \Theta(Y)\Xi.$$

The structure (3) is found by solving the optimization problem:

$$\xi_i = \underset{\xi'}{\operatorname{argmin}} ||\Theta(Y)\xi' - \dot{Y}_i||$$

where Y_i is the ith column of the matrix Y.

The coefficients ξ_i were approximated by an artificial neural network, an MLP with a single hidden layer and the training tuples $\langle y_i, \dot{y}_i \rangle$, I = 1, ..., *n*'. Where the coefficients were below the selected threshold h in absolute value, they were equated to zero. Section 3 presents the solution.

8 of 20

3. Results

3.1. Neural Network Semantic Model

3.1.1. Rating of Digital Platforms According to Actors' Preferences

At the initial stage, digital platforms were identified where the actors discussed the problems of the NEC construction most actively. After identifying conflict zones, these platforms' data were used to defuse the conflict.

To discuss topics related to the NEC construction, the actors preferred microblogging (11,864,333), social networks (826,516), videos (560,681), instant messaging (517,164), blogs (55,403), as well as Internet media (2 706), forums (378), thematic platforms (103), shops (11), print media (5), television (1) and review sites (1).

Among the digital platforms where digital content dedicated to the NEC project implementation was most actively disseminated, the social network VKontakte was leading substantially. Top positions were also taken by Facebook, Twitter, mos.ru and Instagram, respectively (Figure 3).



Figure 3. Rating of digital platforms by popularity among the actors.

3.1.2. Semantic Network Analysis

The semantic network of the consolidated dataset shows the semantic accents that prevail when covering the topic of the NEC implementation (Appendix A). A significant role is played by transport facilities, which are most closely related to the NEC construction, as well as various problems of urban planning policy and construction and technical aspects. Despite the fact that the topic structure and semantic network show the predominance of official content with specific given topics, the core also includes nominations signaling the main claims of residents: Botanical Garden (99), Elk Island (99), National Park (99), cutting down trees (98) (Figure 4). In Figure 4, color and size are shared by clusters according to the weight of the vertices.



Figure 4. Core of the semantic network.

3.1.3. Sentiment Analysis of the Content

The sentiment analysis of the general database showed that the proportion of the content with neutral and positive sentiment is larger than that of the negative content (Figures 5 and 6).



Figure 5. Messages with various sentiment types.



Figure 6. Sentiment consolidated database on various digital platforms.

It is revealing that the quantitative indicators of positive and negative communications differ on various digital resources. While positive messages are mostly found on the following platforms: VKontakte, Twitter, mos.ru, and mosday.ru, negative messages are mostly found on these: VKontakte, Facebook, Twitter, mos.ru, and Instagram (Figure 7).



Figure 7. Distribution of negative and positive messages across digital platforms.

This situation is due to the leading position of the social network VKontakte in the Russian-speaking media space. According to quantitative data, VKontakte is significantly ahead of similar resources. A large number of positive messages on Twitter, mos.ru, and mosday.ru is, to a certain extent, explained by their active use by official and biased authors and sources. Russian-speaking users traditionally use Facebook as a platform for discussing political and socially significant issues, often with negative connotations.

3.1.4. Analysis of the Topic Structure and Content Analysis

The topic structure of the consolidated dataset reflects the predominant nature of the content, which consists mostly of official messages describing the progress of the construction and potential improvements in the transport situation in the city after the NEC was put into operation. The topics of the negative cluster in accordance with the rating are as follows: traffic limitations due to the NEC construction; destruction of the Losiny Ostrov National Park for the sake of the NEC construction; violation of legislation when cutting the Losiny Ostrov park area; accusations of misuse and overspending of budgetary funds; destruction of the Kuskovo park; danger for architectural monuments (in particular, for the Church of the All-Merciful Savior in Kuskovo, (1644); deterioration of the ecological situation due to the felling of trees; danger of a techno-ecological catastrophe (Lyubertsy sewer collector); and the introduction of a paid fare through the NEC.

The positive (as well as the neutral) cluster consists mainly of official messages. Despite the quantitative predominance, the topics are represented by duplicate messages of the same type concerning the construction progress and the expected improvement of the transport situation after the NEC is put into operation. Positive cluster topics according to the rating are as follows: improving the transport situation; reconstruction of urban facilities, improvement of the environment for the movement of citizens; transport accessibility of areas through which the NEC passes; and the improved accessibility to the MCC and metro stations.

An interesting media fact should be noted: in the rating of positive and negative texts, the top position was taken by texts about the congestion of Altufevskoe highway during the NEC construction.

3.1.5. Analysis of the Users' Perception of the Situation

Despite the significant quantitative predominance of the official content, which forms the positive and negative clusters, an increase in the number of negative reactions of the users indicates an increase in the negative perception of the NEC construction by the actors.

The dynamics of references with various sentiments also show that the growth of negative messages is accompanied by the growth of positive messages. However, it should be borne in mind that the negative content is generated mainly by the users, while the positive content is generated by official sources and biased actors (Figures 8 and 9).



Figure 8. Sentiment of the actors' digital footprints.



Figure 9. Dynamics of messages with various sentiment types.

3.2. Mathematical Modeling Results

3.2.1. Statistical Analysis of Data

Statistical analysis of data $\{X(t), P(t), Z(t)\}, t = 1, 2, ..., n$ was performed in the Statistica 10 software [43]. The research carried out was further used in the development of mathematical models. Table 1 shows a fragment of the initial data.

Tal	ble	e 1.	Fragment c	of t	he	initial	data.
-----	-----	------	------------	------	----	---------	-------

t	Р	X	Ζ
1	4	2	3
2	3	0	6
3	1	0	4
4	1	1	6
5	0	2	14
6	1	1	16
7	1	0	53
8	2	1	57
9	3	0	23

The calculated sample estimates of the correlation coefficients (Table 2) indicate that the indicators under consideration have a linear relationship.

Table 2. Sample correlation coefficient matrix.

	X	Ζ	Р
Х	1.000000	0.530344	0.522465
Ζ	0.530344	1.000000	0.486965
Р	0.522465	0.486965	1.000000

We investigated the behavior of the dependence of indicators *X*, *Z* (non-normalized indicators are considered). Figure 10 shows a graph of polynomial smoothing of the dynamics of indicators *X*, *Z* in the OXZt plane.





When constructing a mathematical models, the initial data are transformed according (2.3) to the rule: x(t) = X(t)/S(t); p(t) = P(t)/S(t); z(t) = Z(t)/S(t).

According to the calculated estimates of the correlation coefficients (Table 3) between variables x and z, the correlation coefficient is high.

	x	z	p
x	1.000000	-0.676560	-0.090046
Z	-0.676560	1.000000	-0.147367
р	-0.090046	-0.147367	1.000000

Table 3. Sample correlation coefficient matrix converted data.

A negative sign in front of the correlation coefficient may indicate that the number of people with a negative attitude is increasing due to the transition of actors from the group of those with a neutral attitude towards the group of those with a negative attitude.

3.2.2. Autoregressive Models ARIMA

Observational data $\{x(t)\}, \{p(t)\}, \{z(t)\}, t = 1, ..., n$ can be considered as time series. Figure 11 shows a diagram of the dynamics of the normalized x(t) value. On the graph, the abscissa is time and the ordinate is x(t) = X(t)/S(t). For time series, there is a whole arsenal of linear stochastic models. Below, time series models are constructed using normalized data for the first indicator, i.e., according to the data for $\{x(t), t = 1, ..., n\}$. It is known that autoregressive models can be constructed only for stationary series. After identifying the trend:

$$\begin{aligned} x_{trend}(t) &= 0.0874 + 0.0016t, \\ \widetilde{x}(t) &= x(t) - x_{trend}(t) \end{aligned} \tag{6}$$



Figure 11. Diagram of series x(t) = X(t)/S(t) (blue line) and model values (red line).

The ARIMA(1,0,1) model is constructed:

$$\widetilde{x}(t) = 0.39\widetilde{x}(t-1) + \varepsilon_{t-1} - 0.12\varepsilon_{t-1}, \text{ where } \varepsilon_t \in N(0,1).$$
(7)

The autoregressive model (7) and model (6) can be used to estimate the forecast for indicator x(t) several steps ahead. The coefficient 0.0016 of the model (6) indicates an increase in the number of people in the group with a negative attitude towards the project.

The root-mean-square error of the model is: $Q = \sqrt{\frac{1}{n} \sum_{t=1}^{n} (x(t) - \hat{x}(t))^2} = 0.0037$, where $\hat{x}(t)$ —model values x(t). Figure 11 shows a diagram of the dynamics of the normalized x(t) value (blue line) and model values (red line).

After identifying the trend z(t):

$$z_{trend}(t) = 0.85 + 0.0016t,$$

$$\tilde{z}(t) = z(t) - z_{trend}(t)$$
(8)

The ARIMA(1,0,1) model is constructed:

$$\widetilde{z}(t) = 0.44\widetilde{z}(t-1) + \varepsilon_{t-1} - 0.14\varepsilon_{t-1}$$
, where $\varepsilon_t \in N(0,1)$.

Using the continuous models built to describe the dynamics of negative and neutral messages, let us find estimates of the rate of the share of negative $(\tilde{d}x)$ and neutral $(\tilde{d}z)$ messages. According to (6) and (8), we can write $\tilde{d}x > \tilde{d}z$, therefore $\check{x} \succ \check{z}$.

3.2.3. ODE 1 Models

ODE 1 models were built as systems of linear differential equations (4) was parametrized by the Nelder–Mead method. A program was developed to parametrize the ODE 1 model. Numerical experiments showed the models (4) to be valid only for local forecasting. For that reason, the time interval was split into several subintervals, each of which was modeled. Building the models of the system of linear differential equations (4) enabled convenient analysis of the stability of trajectories the dynamic system has in the phase space, and provide easy identification of bifurcation points. The systems of differential equations were tested for stability [30] by spectral analysis; testing essentially boiled down to the stability testing of a trivial solution: stationary points. Analysis showed the models built on data from different periods of observation had different types of stable points. This fact might indicate the presence of bifurcation points.

Let us take a look at the analysis of one model:

$$\begin{cases} \frac{dx}{dt} = 0.75x + 0.49z\\ \frac{dz}{dt} = -0.6x - 0.11z \end{cases} A = \begin{bmatrix} 0.75 & 0.49\\ -0.6 & -0.11 \end{bmatrix}$$

Characteristic values: $\lambda_{1,2} = 0.32 \pm 0.33i$, therefore, the stationary point, the unstable focus of the system trajectory that was arbitrarily close to the zero at the initial timepoint deviate from the ε -neighborhood of the zero as *t* grows on.

Next to a nonsingular point, all vector fields have an identical, simple structure. In order to test the behavior at singularities, use one of the built models:

Near a nonsingular point, all vector fields have a simple and identical structure. To study the behavior at the points of singularities, we use one of the constructed models:

$$A = \begin{bmatrix} 0.64 & 0.15 \\ -0.58 & -0.07 \end{bmatrix}$$

 a_{22} can be ignored in view of its smallness. Then, from the system of differential equations with the defined matrix, find:

$$dx/dz = -(1.1 + 0.25 \, z/x).$$

According to the general theory [30], obtain:

$$du/dz = -1, 1 - (u + 1/u), u = x/z$$

The dynamic system is known to be oscillating at k = 1, 1 < 2 and change the direction of movement no more than once at k = 2. This value of the coefficient can be further used to add the control parameter to the general formulation of the system of differential equations:

$$\begin{cases} \frac{dx}{dt} = (a_{11} + \mu)x + 0.49a_{12}z\\ \frac{dz}{dt} = a_{22}x \end{cases}$$

where μ is the control parameter.

Conclusions. (1) Spectral analysis shows the stationary points of the system to be unstable. (2) Negative attitude to the project is on the rise. (3) Analysis of singularities found out ways to introduce the control parameters. (4) The solution obtained by means of the ODE 1 models $\tilde{x} \succ \tilde{z}$.

3.2.4. ODE 2 Model

ODE 2 models were built as systems of nonlinear differential equations. Right-side equations and model parameter estimates were obtained by symbolic regression. The functions f on the right side of (2) were structured as second-degree polynomials. The coefficients of the polynomial were approximated by an artificial neural network. Where the coefficients were below the selected threshold h in absolute value, they were equated to zero. ANN-based modeling produced the following models:

$$\begin{cases} \dot{x} = -0.36x + 1.67z - 0.14xz - 0.32x^2 - 2z^2 \\ \dot{z} = -0.4x + 0.01z - 0.07xz + 0.46x^2 + 0.41z^2 \end{cases}$$
(9)

First-approximation stability analysis of the solutions (9) showed the solutions to be asymptotically stable.

The solution obtained by means of the ODE 2 models: $\check{x} \succ \check{z}$.

3.2.5. Analysis of Model Accuracy

Analysis shows the solutions found by the ODE 1 do not have initial data stability. Below is a comparison of the resulting solutions *ARMA* and ODE 2 in the neighborhood of the selected initial point and actual values (Table 4). Let:

 X_1 be the ODE 2 solution;

$$X_2$$
 be the ARMA solution;

$E_i(t) = (x(t) - X_i(t))^2$; MSE $= \sum_{t=1}^n E_i(t)$; NMSE $= \frac{1}{n} \sum_{t=1}^n E_i(t)$

Table 4. Analysis of model accuracy.

t	x	X_1	E ₁	X_2	<i>E</i> ₂
1	0.2500	0.2500	0.0000	0.3454	0.0091
2	0.2000	0.2511	0.0026	0.2832	0.0069
3	0.1058	0.2556	0.0225	0.0964	0.0001
4	0.1039	0.2629	0.0253	0.1395	0.0013
5	0.1882	0.2724	0.0071	0.2365	0.0023
6	0.2069	0.2835	0.0059	0.2558	0.0024
7	0.2000	0.2958	0.0092	0.2515	0.0026
8	0.2833	0.3091	0.0007	0.1022	0.0328
MSE			0.0091		0.0072
NMSE			0.0956		0.0848

3.2.6. ODE 1 vs. ODE 2

Analysis shows the solutions found by the ODE 1 do not have initial data stability. Below is a comparison of the resulting solutions by the ODE 1 and ODE 2 in the neighborhood of the selected initial point. Let:

 X_1 be the ODE 2 solution;

 X_3 be the ODE 1 solution;

absd be the absolute difference between X_1 and X_3 : *absd* = $|X_1 - X_3|$; *D* be the relative difference between X_1 and X_3 : $D = absd/|X_3|$.

The mean difference between the solutions produced by the ODE 1 and ODE 2 models is 0.02; the relative difference is 0.04, which indicates a satisfactory quality of the models (Table 5).

Table 5. Comparison of solutions the resulting by the ODE 1 and ODE 2.

t	X_3	X_1	absd	D
0	0.29	0.29	0	0
0.1	0.32	0.33	0.00	0.02
0.2	0.35	0.37	0.01	0.04
0.3	0.4	0.42	0.02	0.05
0.4	0.44	0.47	0.02	0.06
0.5	0.49	0.52	0.03	0.0
0.6	0.54	0.58	0.03	0.06
0.7	0.60	0.64	0.03	0.06
0.8	0.67	0.71	0.03	0.05

4. Discussion

The paper presents a solution to the problem of discovering what the residents think of a subway line construction project from their social media. The problem was split into subproblems: (1) filter the data; (2) build a mathematical model; and (3) make a managerial decision. These subproblems were solved by artificial neural networks. Solving the subproblems one by one is essentially an algorithm in which an AI automates solving a managerial problem. Visual analytics and statistical modeling were used to verify the proposed approaches to mathematical modeling. The authors' comparative analysis showed the solutions produced by visual analytics, mathematical statistics and differential equations matched: the actors have negative attitudes toward the project. Discrepancies in the analyses made by visual analytics and by mathematical models may stem from complex relations that simple methods cannot grasp. The system of differential equations was parametrized by Nelder–Mead optimization and by artificial neural networks.

Although the solutions by the two models were virtually identical at and near the initial conditions, we believe ANN-based models have several advantages. They are adaptive, resistant to noise, easier to use, and have high computational performance. Equally important is the fact that the structure of the ODE 2 models was selected from a class, and the class was selected on the basis of a priori data. However, the model structure was finalized by the neural network from data.

The analysis of messages shows that most of the media content covering the implementation of the NEC project consists of official sources and traditional media. The rating of texts is topped by the mayor's tweet (17 May 2017) about the completion of the construction of a NEC section from Shchelkovskoye highway to Losinoostrovskaya street (1 031 548).

The analysis of official messages shows the general representation of the media content related to the NEC construction. Meanwhile, the residents' perception of the project progress and their attitude towards the NEC construction is shown by the analysis of the content generated by users and actors' digital footprints in social networks.

The dynamics of messaging with various sentiment types show a sharp drop in the number of texts with a neutral sentiment during the analyzed period, while the number of negative and positive messages increases.

The analysis of the dynamics of messages shows that neutral messages forming a large part of the media content regarding the NEC construction, and positive references do not in any way correlate with the negative content. Thus, the official neutral and positive messages do not correlate and do not affect the perception of users who generate the negative content dedicated to the NEC implementation project, and expressed the negative perception of the project by the residents (Figures 12 and 13).



Figure 12. Dynamics of the neutral message base.

The peak of negative reactions was recorded on August 21, 2020 (Figure 13) and was associated with the trial of Moscow City Duma deputy, Oleg Sheremet, who campaigned against the NEC construction and the destruction of Kuskovo Park. Thus, the NEC construction is actively involved in the political process, associated with the activities of certain political actors, embedded in political content, and is brought beyond the urban conflict.



Figure 13. Dynamics of the negative message base.

Approaches for constructing mathematical models based on digital footprints have been developed. The peculiarity of the problem is that the data used are the data of the indirect variables of the dynamic system, and the data of the indirect variables are very noisy. To solve the problem, models were constructed in the form of systems of linear differential equations in a few-mode region. The estimates of the coefficients of the equations were found using the Nelder–Mead method, for which a Python code had been developed. Physics-based neural networks were used to describe the behavior of a dynamic system throughout the entire course of the study. The study shows that machine learning allows for the construction of adequate mathematical models for the problem considered.

5. Conclusions

The analysis of the solutions from all models showed that all solutions were consistent. The developed approach of combining mathematical and neural network models to analyze the social media users' perception based on their digital footprints led to the conclusion about the tense situation around the construction of the NEC.

The proposed approach for analyzing the social networks users' perception based on their digital footprints made it possible to quickly identify conflict zones associated with the NEC implementation project and propose recommendations to resolve and prevent conflicts.

The solutions obtained using the developed mathematical models in Section 3.2 are fully consistent with the study carried out on the issue under consideration in Section 3.1. Therefore, we can speak about the adequacy of the constructed mathematical models. The advantage of the constructed mathematical models is that they allow predicting the behavior of dynamic systems. It is possible to introduce control into the constructed models, which will make it possible to plan in advance certain measures to stabilize the situation.

Research data could be of use in solving similar problems in sociology, ecology, and economics.

Author Contributions: Conceptualization. N.G. and M.P.; methodology. N.G. and M.P.; software. N.G. and M.P.; validation. N.G. and M.P.; formal analysis. N.G.; investigation. M.P.; writing—original draft preparation. M.P.; writing—review and editing. N.G.; visualization. M.P.; supervision. N.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by RFBR. Grant 19-08-000261.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Publicly available datasets were analyzed in this study. This data can be found here: (ttps://vk.com, accessed on 26 June 2021), (https://www.livejournal.com/, accessed on 26 June 2021), (https://twitter.com, accessed on 26 June 2021), (https://www.youtube.com, accessed on 26 June 2021).

Conflicts of Interest: The authors declare no conflict of interest.



Appendix A Semantic Content Network

References

- 1. Dustdar, S.; Yu, E.; Salinesi, C.; Rieu, D.; Pant, V. (Eds.) *Advanced Information Systems Engineering*; Springer International Publishing: Cham, Switzerland, 2020; 574 p.
- Brandon, P.; Traylor, A.; Pater, J. Learning Reduplication with a Neural Network without Explicit Variables. 2019. Available online: http://works.bepress.com/joe_pater/38/ (accessed on 5 February 2021).
- 3. Linzen, T. What can linguistics and deep learning contribute to each other? Response to Pater. *Language* **2019**, *95*, e98–e108. [CrossRef]
- 4. Pater, J. Generative linguistics and neural networks at 60: Foundation, friction, and fusion. Language 2019, 95, e41–e74. [CrossRef]
- Kullkarni, V.; Kern, M.L.; Stillwel, L.D.; Kosinski, M.; Matz, S.; Ungar, L.; Skiena, S.; Schwartz, H.A. Latent Human Traits in the Language of Social Media: An Open-Vocabulary Approach. *PLoS ONE* 2018, *13*, e0201703. [CrossRef] [PubMed]
- 6. Demuth, K. Perception, production, and individual differences. *Appl. Psycholinguist.* **2018**, *39*, 735–741. [CrossRef]
- Lowry, M.; Bryant, J.B. Blue is in the eye of the beholder: A cross-linguistic study on color perception and memory. J. Psycholinguist. Res. 2019, 48, 163–179. [CrossRef] [PubMed]
- 8. Brown, T.; Tom, B.; Mann, B.; Ryder, N.; Subbiah, M.; Kaplan, J.; Dhariwal, P.; Neelakantan, A.; Shyam, P.; Sastry, G.; et al. Language models are few-shot learners. *arXiv* **2020**, arXiv:2005.14165v4.
- 9. Decker, A. Der Social-Media-Zyklus. Schritt für Schritt zum Systematischen Social-Media-Management im Unternehme; Gabler Verlag: Wiesbaden, Germany, 2021; 227 p.
- 10. Matei, S.A.; Rebillard, F.; Rochelandet, F. (Eds.) *Digital and Social Media Regulation: A Comparative Perspective of the US and Europe;* Palgrave Macmillan: London, UK, 2021; 197 p.
- Kharlamov, A.A.; Raskhodchikov, A.N.; Pilgun, M. Smart city data sensing during COVID-19: Public reaction to accelerating digital transformation. *Sensors* 2021, 21, 3965. [CrossRef]
- 12. Kravets, A.G.; Bolshakov, A.A.; Shcherbakov, M. (Eds.) *Cyber-Physical Systems: Modelling and Intelligent Control*; Springer International Publishing: Cham, Switzerland, 2021; 338 p.
- 13. Siddique, N. Intelligent Control: A Hybrid Approach Based on Fuzzy Logic, Neural Networks and Genetic Algorithms; Springer International Publishing: Cham, Switzerland, 2014; 517 p.
- 14. Pan, I.; Das, S. Intelligent Fractional Order Systems and Control: An Introduction; Springer International Publishing: Cham, Switzerland, 2013; 438 p.
- 15. Das Sharma, K.; Chatterjee, A.; Rakshit, A. Intelligent Control: A Stochastic Optimization Based Adaptive Fuzzy Approach; Springer: Singapore, 2018; 302 p.
- Silhavy, R.; Senkerik, R.; Oplatkova, Z.K.; Silhavy, P.; Prokopova, Z. (Eds.) Automation Control Theory Perspectives in Intelligent Systems. In *Proceedings of the 5th Computer Science On-line Conference 2016 (CSOC2016)*; Springer International Publishing: Cham, Switzerland, 2016; Volume 3, 438 p.
- 17. Li, Z.; Xia, Y.; Su, C.-Y. Intelligent Networked Teleoperation Control; Springer: Berlin/Heidelberg, Germany, 2015; 282 p.
- 18. Szuster, M.; Hendzel, Z. Intelligent Optimal Adaptive Control for Mechatronic Systems; Springer International Publishing: Cham, Switzerland, 2018; 382 p.
- 19. Stubbs, M.; Lemon, M.; Longhurst, P. Intelligent urban management: Learning to manage and managing to learn together for a change. *Urban Stud.* 2000, *37*, 1801–1811. [CrossRef]
- 20. Wen, J.T.; Mishra, S. (Eds.) Intelligent Building Control Systems: A Survey of Modern Building Control and Sensing Strategies; Springer International Publishing: Cham, Switzerland, 2018; 313 p.

- Turner, J.; Holmes, L.; Hodgson, F.C. Intelligent urban development: An introduction to a participatory approach. *Urban Stud.* 2000, *37*, 1723–1734. [CrossRef]
- 22. Arun, M.; Teng Yap, M. Singapore: The development of an intelligent island and social dividends of information technology. *Urban Stud.* **2000**, *37*, 1749–1756. [CrossRef]
- 23. Gubanov, D.; Korgin, N.; Novikov, D.; Raikov, A. *E-Expertise: Modern Collective Intelligence*; Springer International Publishing: Cham, Switzerland, 2014; 313 p.
- 24. Merad, M.; Trump, B.D. Expertise Under Scrutiny, 21st Century Decision Making for Environmental Health and Safety; Springer International Publishing: Cham, Switzerland, 2020; 168 p.
- 25. Acebillo, J.A. Barcelona: Towards a New Urban Planning Approach. Spatium 2006, 13–14, 55–71. [CrossRef]
- 26. Deutschmann, E.; Lorenz, J.; Nardin, L.G.; Natalini, D.; Wilhelm, A.F.X. (Eds.) *Computational Conflict Research;* Springer: Heidelberg, Germany, 2020; 264 p.
- 27. Kharlamov, A.; Pilgun, M. (Eds.) *Neuroinformatics and Semantic Representations, Theory and Applications;* Cambridge Scholars Publishing: Newcastle upon Tyne, UK, 2020; 317 p.
- 28. Rozanov, Y. Random Process; Nauka: Moscow, Russia, 1971; 156 p.
- 29. Arnold, V.I. Geometric Methods in the Theory of Ordinary Differential Equations; ICMMO: Moscow, Russia, 2012; 189 p.
- 30. White, M.D.; Marsh, E. Content analysis: A flexible methodology. Libr. Trends 2006, 1, 22–45. [CrossRef]
- 31. Krippendorff, K. Content Analysis: An Introduction to Its Methodology, 3rd ed.; SAGE Publications, Inc.: Los Angeles, CA, USA, 2012; 441 p.
- 32. Bongard, J.; Lipson, H. Automated reverse engineering of nonlinear dynamical systems. *Proc. Natl. Acad. Sci. USA* 2007, 104, 9943–9948. [CrossRef] [PubMed]
- 33. Raissi, M.; Perdikaris, P.; Karniadakis, G.E. Numerical Gaussian processes for time-dependent and non-linear partial diaferential equations. *SIAM J. Sci. Comput.* **2018**, *40*, 172–198. [CrossRef]
- Raissi, M.; Perdikaris, P.; Karniadakis, G.E. Inferring solutions of differential equations using noisy multifidelity data. J. Comput. Phys. 2017, 335, 736–746. [CrossRef]
- Castillo-Chavez, C.; Song, B. Models for the transmission dynamics of fanatic behaviors. In *Bioterrorism: Mathematical Modeling Applications in Homeland Security*; SIAM Frontiers in Applied Mathematics: Philadelphia, PA, USA, 2003; Volume 28, pp. 155–172.
- Stauffer, D.; Sahimi, M. Discrete simulation of the dynamics of spread of extreme opinions in a society. *Phys. A Stat. Mech. Appl.* 2006, 364, 537–543. [CrossRef]
- 37. Bettencourt, L.M.A.; Cintron-Arias, A.; Kaiser, D.I.; Castillo-Chavez, C. The power of a good idea: Quantitative modeling of the spread of ideas from epidemiological models. *Phys. A Stat. Mech. Appl.* **2006**, *364*, 513–536. [CrossRef]
- Mena-Lorca, J.; Hethcote, H.W. Dynamic models of infectious diseases as regulators of population sizes. J. Math. Biol. 1992, 30, 693–716. [PubMed]
- 39. DeGroot, M.H. Probability and Statistics; Addison-Wesley: Boston, MA, USA, 1986; 911 p.
- 40. Hair, J.F.; Anderson, R.E.; Tatham, R.L.; Black, W.C. Multivariate Data Analysis; Prentice-Hall: Hoboken, NJ, USA, 2006; 785 p.
- 41. Gabdrahmanova, N.T. Principy i priemy postroenija nejrosetevyh matematicheskih modelej slozhnyh stohasticheskih system. *Kavk. Nauch. Zap.* **2010**, *4*, 191–204.
- 42. Nelder, J.A.; Mead, R. A simplex method for function minimization. Comput. J. 1964, 7, 308–313. [CrossRef]
- 43. Borovikov, V.P.; Ivchenko, G.I. *Forecasting in STATISTICA in Windows Environment*; Finance and statistics: Moscow, Russia, 2006; 368p.