Scientific Visualization, 2024, volume 16, number 4, pages 43 - 58, DOI: 10.26583/sv.16.4.05

Event Analysis: Application in Social Forecasting

T. V. Korenkova[1], A. A. Artamonov[2], M. S. Ulizko[3]

National Research Nuclear University MEPhI, Moscow, Russia

[1] ORCID: 0009-0000-6257-8500, korenkova.tanya@mail.ru
[2] ORCID: 0000-0002-9140-5526, aartamonov@kaf65.ru
[3] ORCID: 0000-0003-2608-8330, mulizko@kaf65.ru

Abstract

Monitoring the interrelationships between social events and phenomena and forecasting the dynamics of their changes are necessary in the conditions of instability of the modern world. There are many separate methods of analysis for social forecasting, however, for this research, the method of event analysis has been chosen, which is insufficiently considered in the scientific literature within the framework of this task, but has high potential. The purpose of the article is to adapt the event analysis methodology for its use as a social forecasting tool. The main data for the study was collected from the Russian information and news resource in the period 2020-2023. Based on the classical methodology of event analysis, the classifiers presented in this paper in the form of social spheres are defined in the research. As part of the analytical comparison stage, a graph analysis was carried out (graphs of relationships between categories were constructed, central nodes-categories were identified); time series analysis was performed (segmentation of time series by the PELT algorithm, clustering of time series by the k-means algorithm); key terms for press events were defined. The final product is an analytical dashboard with filters, statistics and interactive graphs. The analytical dashboard makes it possible to compare data in a static and dynamic state, to draw conclusions about the past and future states of objects of social forecasting. The main result of the research is the event analysis methodology developed by the author, which can be used for a comprehensive analysis of news streams, adapted to the necessary categories representing a certain entity or sphere, and applied in various social organizations or monitoring services.

Keywords: social forecasting, event analysis, news analysis, graph analysis, centrality calculation, time series analysis, time series segmentation, time series clustering, PELT.

1. Introduction

In a rapidly changing world, effective forecasting of social processes is becoming one of the key tasks for researchers and practitioners [1]. Event analysis, as a method offering a systematic approach to the study of significant events and their consequences, is a tool for identifying patterns and trends in social development.

A. V. Maltseva [1] proposed an algorithm for applying event analysis using the example of studying processes in the labor market, consisting of the stages of data collection, selection of classifiers, calculation of results, comparison of the obtained values, and verification of results. At the same time, it was revealed that there is no unambiguous methodology for conducting this type of analysis, so it becomes possible to modify some stages. M. V. Novoselov [2] described some additional statistical, mathematical, and graphical methods that can be used to conduct extended event analysis, in particular, time series analysis and cluster analysis.

The number of both Russian-language and foreign scientific works at the intersection of social forecasting and event analysis is insignificant. A. A. Azarov [3] studied the term social computing — an interdisciplinary field of research that includes the study of social behavior by means of computing systems. One of the important tasks of social computing is modeling, analysis and forecasting of social behavior of actors using such methods as intent analysis, event analysis, psychographic analysis. V. I. Kudryavtseva [4] described the main approaches to conducting social forecasting: general scientific, intuitive, formalized and complex forecasting methods, and among them the use of event analysis is mentioned.

Much more research is devoted to the study of the dynamics of the development of social entities in the news and their clustering, which is associated with specific event analysis tools. J. L. Ortega [5] proposed and implemented an approach to constructing network graphs based on the frequency with which media jointly mention research papers. I. Bloch and V. Alexandrov [6] constructed time series that display the distribution of the popularity of clusters representing certain social phenomena over time.

Thus, the topics of social forecasting and event analysis are studied quite extensively in scientific publications separately, but the number of papers describing their joint application is limited. Based on the literature review, it was determined that the analysis of graphs and time series of events can be good tools for studying the relationships between social entities and tracking the dynamics of their changes. It is worth noting that these tools are rarely described by other authors at the stages of event analysis. This indicates the novelty and relevance of the research and the need to study event analysis and social forecasting in conjunction.

2. Statement of the problem

The aim of the research is to adapt the event analysis methodology for use as a social forecasting tool. The study examines key aspects of social forecasting, identifies its main areas of application, and describes the stages of the classical event analysis methodology. Particular attention is paid to the development of modifications that will improve this methodology, as well as its practical application based on real data.

In the course of the research, both the theoretical foundations of event analysis and its practical significance in various social contexts were studied. The results of the study are aimed at creating an effective tool that will not only increase the accuracy of the analysis, but will also allow for more reliable prediction of social changes.

3. Theory

Social forecasting is a process of studying the development prospects of social objects in order to improve the efficiency of their management, based on working with a variety of alternatives, combining various methods and the abstract nature of possible solutions. The main features of social forecasting are the lack of clear goals and directions of forecasting, the complexity of formalizing the social sphere and the need to combine qualitative and quantitative methods [7]. Social forecasting, unlike social science forecasting, is forecasting only in the field of sociology. However, sociology is increasingly interfering with other relations from the side of their social organization every year: economic, national, moral and ethical, etc. Based on these trends, some authors highlight a list of areas and spheres in which social forecasting can be carried out, see Table 1 [7].

Table 1. Applications of social forecasting

Sphere	Application areas of social forecasting
Science	Prospects for the development of scientific personnel, research institutions, and funding of scientific discoveries
Technologies	Prospects for the development of new technologies, informatization of society, maintaining confidentiality
Economy	Prospects for the development of social organization of labor, the fight against unemployment and inflation
Politics	Prospects for the development of state and international relations, monitoring the attitude of the people to the authorities

Legislation	Prospects for the implementation of new laws in the social sphere, measures to support the population, and the preservation of human rights
Population	Prospects for changes in the structure of society, migration processes
Education	Prospects for the development of various educational institutions, advanced training and incentives for personnel
Healthcare	Prospects for the development of medical institutions, discoveries in the field of medicine, healthy lifestyle
Culture	Prospects for the development of cultural goods, tourism, media influence, preservation of cultural heritage
Ecology	Prospects for the exploration of the Earth and space, environmental conservation, urban development and transport
Social structure	Prospects for the development of social-industrial, professional, educational and gender-age structures
Social life	Prospects for the development of public order, social needs, the fight against antisocial phenomena, inequality and poverty

According to estimates by domestic and foreign scientists, there are 150-200 different methods of scientific forecasting today. However, the number of methods that can be called basic and most common in the practice of social forecasting is significantly less and reaches 15-20. Many of these methods are more likely to be techniques and methods of forecasting that take into account the nuances of the dynamics of the development of objects. As a rule, either expert or factual methods of social forecasting are used separately. Not many methods meet the criteria of complexity, flexibility, universality and study in scientific papers: in particular, explicative ones [3]. The main advantage of these methods is that they are initially designed for use in the social sphere, and are not borrowed from more exact sciences. Some of these methods mostly contain descriptive analysis, but at the final stage, it is assumed that recommendations and social forecasts are developed by experts. These methods take into account the limitations and complexity of formalizing data from the social sphere, are directly related to the detection of relationships and patterns in data, and also involve the use of factual methods in combination with expert methods. In practice, this occurs when experts are provided with factual information about an object in advance or are introduced to previously made factual forecasts, or, conversely, in the process of extrapolation modeling of development trends of an object, along with factual data, expert assessment data are taken into account. Thus, based on the above conclusions and analysis of scientific papers, a hypothesis has been put forward that the explicative method of event analysis has sufficient flexibility and a set of tools for its application in the task of social forecasting.

Event analysis as a method of political science originated in the 1960s in the scientific papers of Charles McClelland [8]. Event analysis is a quantitative method for studying political reality that focuses on the systematic analysis of reports of events. Its «relative» is content analysis, both methods perform quantitative analysis of texts, but in different ways. The object of event analysis is not the events themselves, but messages about events, mainly from the media. These events are systematized, analyzed, classified and processed using software and mathematical methods [1].

Currently, event analysis is used in conflictology, sociology, political science and natural sciences. Its wide scope of application is explained by the possibility of comparing different events, analyzing them by the number of participants, duration and scale of interaction. This allows not only to compare events, but also to build multi-variant scenarios, which increases the accuracy of not only tactical but also strategic forecasting. Thus, event analysis provides a more detailed idea of changes in the political and social situation compared to traditional research methods.

The event analysis methodology is aimed at monitoring the course of events and their intensity in order to identify the main trends in the evolution of the situation both at the national and international levels. Initially, the analysis process most often included two approaches: the first, based on the analysis of data «from below», and the second, where the researcher formulated normative models for subsequent filling with facts - the «top-down» approach. The first approach means that the researcher does not predetermine the important aspects of the process being studied, with the exception of the main object of observation. In the second case, the study is based on a structured collection of information, where certain elements of the process are identified in advance as the most significant. Usually, both approaches are used together in research, enriching the view of the situations being analyzed [1].

The practical implementation of event analysis can be divided into two main phases. The first phase involves the formalized presentation of event messages using a specific coding scheme, which creates «event data». The second phase involves using the data to formulate meaningful hypotheses and conclusions regarding the political processes being studied, as well as to build and test models. In modern political science, this stage uses a wide range of statistical methods and mathematical approaches, such as factor, discriminant, correlation, cluster analysis, etc. At the final stage, the results are validated, followed by the preparation of forecasts and expert assessments [1]. Modifications in the form of the introduction of graph algorithms, segmentation and clustering of time series, and the selection of keywords were introduced at the stage of analytical research, and all other stages were performed in accordance with the established methodology. A description of each stage of event analysis implemented in this research is given below.

3.1. Compilation of an information array

The first stage of working with event analysis is the creation of an information array or data bank. Data sources can be different: official documents, reports, news articles, incident statistics, etc. In this research, an information and news resource was chosen as a source of data collection, since news reflects almost all aspects of social interactions and events. Data collection was carried out automatically using web scraping in the Python programming language with the BeautifulSoup and requests libraries. For this purpose, only one Russian information and news resource was used to avoid problems of duplication and aggregation of data from different sources. The news is collected for a full four years between 2020 and 2023, along with metadata such as date, category, original news title, and headlines of news mentioned in context [9]. In this research, we assume that each news item in the selected information and news resource can be classified into one of the social categories, although among them there may be news items that are difficult to assess in terms of their contribution to the social sphere – for example, some news items on scientific topics.

3.2. Construction of a system of classifiers

The second stage is the development of a system for classifying news reports on social events to formalize events and phenomena and analyze their interactions. The results of observation can be recorded using coding. A more complex system can be used to record events - a coding form that includes various details of the phenomenon being studied, such as data on the initiators of events, the social context of what is happening, the type of event, the objects to which the actors' actions are directed, etc. Today, there are many event analysis databases that are constantly being updated with new studies. All databases can be divided into two main groups. The first group is subject-oriented databases focused on participants in international political processes, including information on the interaction of a certain set of participants over a certain period of time. The second group is problem-oriented databases focused on specific historical events, such as major conflicts [1].

A new classification system based on previously identified social spheres that may require the development of social forecasts is created in this research, see Table 1. If necessary, some of them can be expanded or combined, which was done using the Word2Vec model. The Word2Vec mathematical model created by Google is a neural network that processes text data and includes two learning models: Continuous Bag of Words (CBOW) and Skip-gram. CBOW is a «continuous bag of words» architecture that predicts the current word based on the context surrounding it. Skip-gram architecture uses the current word to predict the words surrounding it. The Word2Vec training model is fed a text data array as input, and word vectors are generated at the output. Then, the cosine distances between all words from the input sample are calculated. This means that for each word from the submitted text, a list of the closest words to it can be found, that is, those that are most often mentioned in the same context, based on the similarity of their vectors [10].

Thus, the Word2Vec model was trained on the entire array of news data, and the result was a table of correlations between each sphere from Table 1 and the list of the words that are most contextually close to it. After a little manual filtering, a final list of all spheres and the category words that are included in it was compiled. It is important to take into account that each new category represents a social process or phenomenon, so that not one, but a number of messages about specific social events can be attributed to it.

In scientific papers, when preparing classification systems, it was assumed that such categories could be given a specific emotional coloring - either positive or negative: for example, in the field of conflict studies. This feature was also adopted in the research, and an emotional component was added to each of the newly formed categories. Some of the categories initially imply one or another emotional coloring - for example, poverty and crime are negative phenomena, and cooperation and import substitution are positive. Some categories are presented in a more generalized way by adding the prefixes «development», «achievements», «problems» - for example, «educational development». The final list of categories is presented in the results.

The next important step is to classify the news array into the selected categories. For this purpose, the topic modeling approach using Zero-Shot classification was chosen. Topic modeling is one of the modern applications of machine learning to text analysis, which help to determine which topics each document belongs to and which words form each of them [11]. At the same time, one of the main difficulties is the formation of training data for each category, which is very problematic in the context of the study, since it is not possible to make a high-quality markup of data for each category due to the imbalance of the entire sample. In this regard, the Zero-Shot classification approach was chosen for topic modeling, which help to bypass all these limitations.

Zero-Shot text classification is a classification task in which models can classify text without being trained on a dataset created for this classification task [12]. The model is able to predict which of the proposed classes the text most likely belongs to based on the analysis of keywords and context. To classify news into new categories, a ready-made multilingual topic modeling model was chosen using the Zero-Shot classification approach [13]. This stage was performed in the Python programming language using the transformers library.

3.3. Calculating the results of filling the matrix classifier

The third stage in the classical event analysis method is the calculation of the results of filling the matrix classifier. Quantitative data for analytical comparison of qualitative characteristics of the situation are expressed through the definition of their relative values, as well as through the construction of indices. Determining relative values is advisable if statistical processing of data is required, especially when using event information. The construction of an index is used to combine various quantitative data into a single complex indicator for the purpose of subsequent monitoring of the situation [1].

Since the main stage of event analysis consists of a set of analysis methods, each of them requires its own approach to calculating the final results. In most cases, it is assumed that the absolute number of news reports on each topic for a certain period of time is calculated.

3.4. Conducting an analytical study

The most labor-intensive stage of the study is conducting analytical comparisons of the obtained values of indicators describing the types of events or their aspects at different time stages. The entire analyzed period is divided into intervals, and the events observed in each of them are compared according to various criteria within these periods.

The analysis methods were selected in such a way that after completing this stage it would be possible to evaluate the parameters proposed by C. McClelland, in accordance with which the data is processed, see Table 2 [8].

Table 2. Event analysis parameters

Parameter	**The question that the parameter answers**
Plot evaluation	What's happening?
Evaluation of the initiating subject	Who is behind this?
Property evaluation	In relation to whom?
Event Time evaluation	When?

This stage is divided into three main parts: graph analysis, time series analysis, and keyword extraction. The choice of these approaches is justified by the parameters outlined above. Graph analysis is used to evaluate the object («in relation to whom/what?») and partially the initiating subjects («who/what is behind this?»), time series analysis is used to evaluate the time of the event («when?»), and keyword extraction is used to evaluate the plot («what is happening?»).

Graph analysis is very rarely described within the framework of event analysis, but it is the best way to see the strength of the relationships between different social categories [14]. The graph is a pair G = (V, E), where V is a set whose elements are called vertices, and E is a set of unordered pairs of vertices whose elements are called edges [15]. Relationship graphs are constructed based on the appearance of categories in the context of one news item. Nodes are the categories themselves, and edges show the presence of a connection between them within one news text. Edges also have weights, and the more news items link two categories, the greater the values of these weights. To track the dynamics of relationships between social events and phenomena, the researcher has the opportunity to construct graphs for an annual interval and compare values across four years.

For the event analysis algorithm, calculating the centrality is also an important step, since it allows identifying the most inter-industry and connecting categories. Centrality is one of the most important indicators used to show the relevance or structural importance of a node in the network. For each category in terms of years, it is possible to calculate the centrality indicator by degree: according to this approach, nodes with a large number of connections receive a higher centrality value, and the indicator itself is calculated as the ratio of the number of nodes with which the node in question has connections to the total number of nodes [15]. The graph analysis was performed in the Python programming language using the networkx and plotly libraries to visualize the results.

To estimate the event time, the time series analysis method was chosen. Time series are built separately for each category, where the x-axis is the date, and the y-axis is the number of news items for this category for this date. In scientific papers on event analysis, time series are usually used to build a frequency distribution of classifiers for specified periods [6]. However, this approach may not be very informative, and here the main goal is to identify key events that resulted in extreme points or periods that stand out from the general series, and manual processing to solve this issue can be labor-intensive and ineffective. Because of this, it was decided to add a time series segmentation stage, as a result of which such events will be determined automatically and show changes more clearly.

The Pruned Exact Linear Time (PELT) algorithm is used for this purpose. This algorithm searches for a set of «inflection points» for a given time series such that their number and location minimize a given «cost» of segmentation. The basic steps of the algorithm are to define a «cost» function for a segment, then iterate over all possible starting and ending points of the segment and check whether splitting into new segments reduces the value of the cost function compared to the unsplit segment. One commonly used approach to identifying multiple change points is to minimize the sum, presented in the formula below [16].

$$\sum_{i=1}^{m+1}\left[C_{\left(y_{(\tau_{i-1}+1):\tau_i}\right)}+\beta f(m)\right]$$

Here is C – the segment «cost» function, τ is the «inflection» point, m – is the total number of «inflection» points, $\beta f(m)$ – is a regularizer to prevent overfitting.

It was found that some points and periods of reduced news activity in the time series coincide with holidays and weekends. Therefore, data on these days were removed from the sample so that the time series would be smoother and no false correlations would be found. This stage was implemented in the Python programming language using the ruptures library for segmentation and the plotly library for visualizing the results.

Another method borrowed from scientific papers is clustering of time series, which allows to detect correlations between time series representing different categories. To perform clustering, time series are built not by days, but by months, as this allows to smooth time series and get rid of noise. For this task, the k-means method is taken, according to which the Euclidean distance between vectors of unshifted time series is calculated, centroids are found for them and finally clusters are determined as a result of moving the centroids by the number of iterations [17]. The k-means algorithm itself clusters time series built by months for the year of interest, as this helps to determine correlations between them better than by dates.

Since the optimal number of clusters is not known in advance, it is necessary to determine this value using the elbow method and the silhouette metric [17]. The elbow method shows the optimal number of clusters according to the following principle: if after the visual elbow on the graph there is a sharp decrease in the total error, then this number is considered optimal, but if there are many clusters, the error will be minimized, but there will be no point in clustering in principle. The sum of the squares of the distances from the objects to the cluster center is calculated - in other words, errors. According to the silhouette method, the optimal number of clusters is the peak value on the graph, after which there is a sharp decline. The metric calculates for each object the average distance between it and the objects inside the cluster (a) and between it and the objects in the nearest cluster (b). The larger the normalized b-a, the better. To implement this task, the k-means algorithm was used, implemented in the Python programming language using the sklearn libraries for normalizing time series, tslearn for clustering and plotly for visualizing the results.

To assess the essence of events, the subjects involved in them, and to determine a more meaningful correlation between categories, another method has been added – identifying key terms from news headlines for extreme values in time series. Such values are understood to mean dates that coincide with the maximum frequency activity for specific categories, as well as «inflection points» that indicate the emergence of a new segment in the time series: some turning points, as a result of which news activity in certain periods began to differ sharply.

To implement this step, an approach was used to identify key terms in the text, based on both linguistic tools (part-of-speech identification, tokenization) and relative frequency [18]. For each extreme date within one category and time period, a list of the most frequent key terms is derived, which allows for more informative conclusions about key events, their actors and the emotional component that influenced the change in news activity. For this task, a model from the TermExtractor library was taken, the code is implemented in the Python programming language.

3.5. Verification of results and formulation of conclusions

Since this version of the methodology does not involve forecasting using software and mathematical methods, this stage is completely manual, and the main result of the implemented methodology is the analyzed data, presented in the form of interactive graphs on the final analytical panel, implemented in the Python programming language using the dash library [19]. By selecting various filters, such as the time period and category of interest, it is possible to compare the results of the analysis by year, determine the change in the behavior of the analyzed objects in dynamics and, based on this, build social forecasts.

The final stage of the event analysis methodology – validation of results and social forecasting – is carried out on the basis of the final analytical panel, which includes:

- a set of interactive graphs constructed using the methods listed above: graph analysis, segmentation and clustering of time series, distribution of key terms;
- select filters – «Category» and «Year»;
- statistics on the absolute and relative number of news items by «Category» and «Year», as well as the central categories for the given period.

4. Experimental results

A fragment of the collected data array from the Russian information and news resource is presented in the figure (Fig. 1).

date	rubric	title	news mentioned in the context
2022/06/28	Россия	Свердловский губернатор вмешался в конфликт коммунистов с Ельцин Центром	['Никита Михалков потребовал признать Ельцин Центр иноагентом', 'Российский депутат предложил Шойгу закрыть Ельцин Центр']
2020/10/04	Экономика	Названы самые популярные валюты для сбережений у россиян	['Россияне решили массово забрать валюту из банков']
2023/10/10	Мир	ХАМАС начало обстрел аэропорта Тель-Авива	['В Израиле опровергли проникновение боевиков на север страны']
2023/08/18	Моя страна	В Северной Осетии появились пляжные спортивные площадки	['Во Владикавказе появятся скульптуры двух героев романа «12 стульев»']
2021/06/15	Бывший СССР	Зеленский оценил позицию НАТО насчет членства Украины	['В НАТО подтвердили вступление в альянс Украины и Грузии']

Fig. 1. Fragment of a news array

The results of the second and third stages of the methodology are combined in one figure – a list of formed social categories is shown, as well as their absolute number in thousands over four years (Fig. 2).

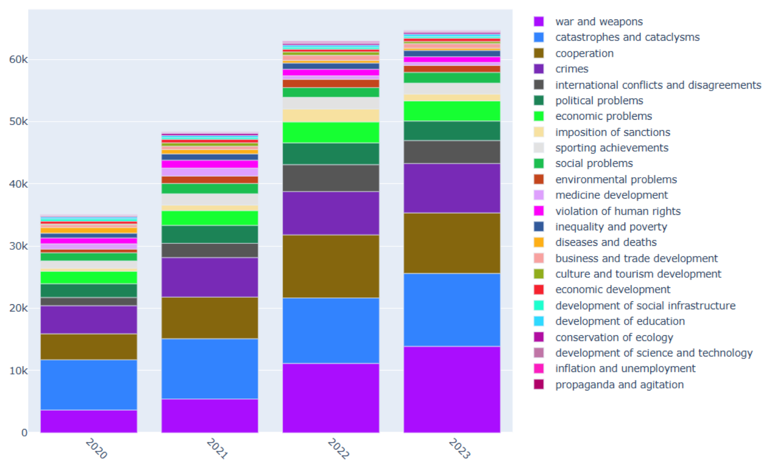

Fig. 2. Distribution of the number of news items by categories and years

The figure shows a graph constructed for 2023: when hovering over nodes in the analytical panel, the name of the corresponding category is seen (Fig. 3). The size of the category nodes is proportional to the number of news items for a given category for the period taken. The thickness of the edges is proportional to the number of news items that mentioned pairs of news items belonging to the corresponding categories.

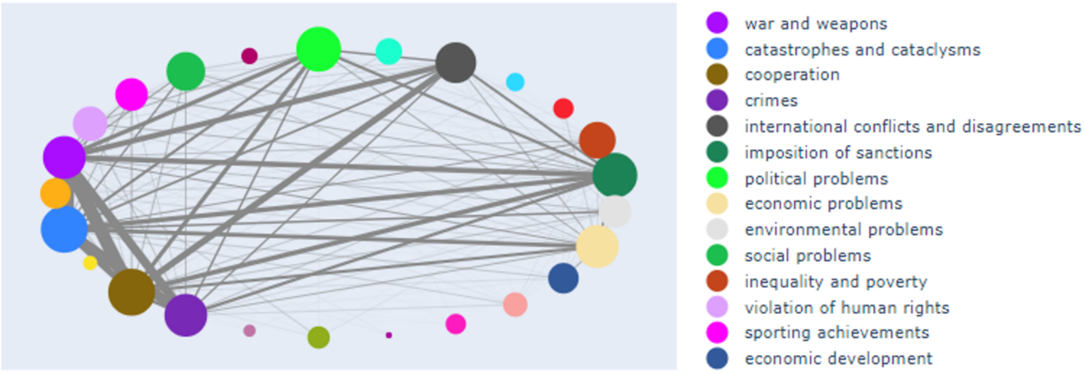

Fig. 3. Category graph for 2023

The result of determining the top 5 central category nodes for 2023 is presented in the table, see Table 3.

Table 3. Top 5 most central category nodes of 2023

Category	**Centrality**
Catastrophes and cataclysms	0.958
Cooperation	0.958
Imposition of sanctions	0.917
Political problems	0.917
Crimes	0.875

To demonstrate the work of the methodology, the category «development of science and technology» was taken as an example, for which a social forecast will be compiled in the future. The figure shows an example of a graph for segmenting a time series by category «development of science and technology» for 2023 (Fig. 4), where the x-axis indicates the date, and the y-axis indicates the number of news items. In this case, the choice was to combine into segments no smaller than a week. The regularization parameter is selected experimentally to prevent overtraining of the algorithm. One of the popular approaches used in the research is to take regularization as two logarithms of the length of the original series. The smaller the value of the regularization parameter, that is, the smaller the «penalty», the more segments are allocated.

Each segment represents a separate period of special news activity: there are segments with a small amplitude of frequency activity spread, and there are with a high one. It can be assumed that in the first case there were no strong news hooks. The vertical red stripe marks the inflection points, i.e. the dates that are the segment boundaries and indicate a change in the behavior of the time series over a certain period. In addition, the top 5 points with maximum values of news activity for the entire period are highlighted in red. Since the graph is interactive, there is no need to display the dates that coincide with the maximum points, and they can be seen by hovering the cursor over the corresponding point. This graph can be used to determine the dates on which the most interesting events from the point of view of analysis occurred, which led to the emergence of a news hook. A more informative description of these events on the specified dates is given at the stage of keyword analysis.

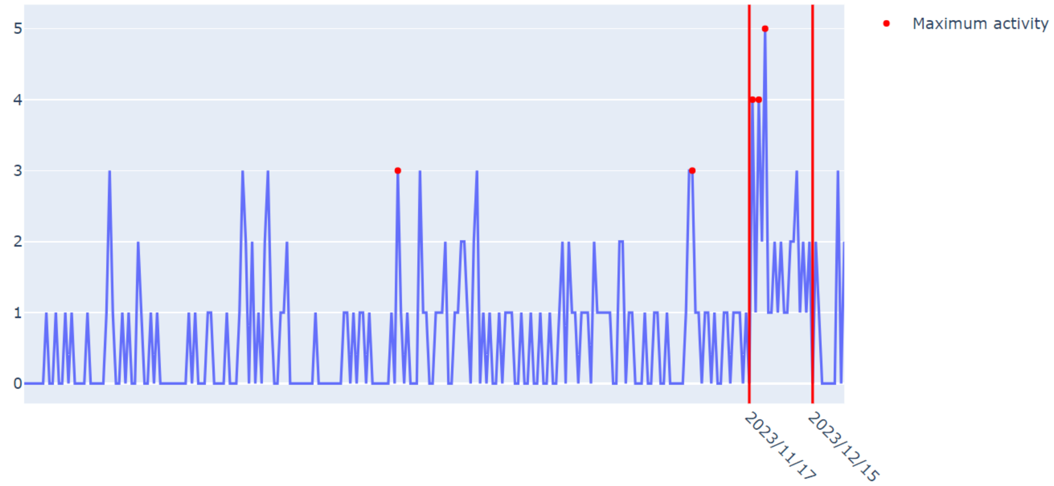

Fig. 4. Graph with the results of time series segmentation for the category «development of science and technology» for 2023

As a result of the clustering algorithm, the graph displays the cluster of time series that contains the category selected in the analytical panel filter. The figure shows an example of a cluster for the category «development of science and technology» for 2023), where the x-axis indicates the date, and the y-axis indicates the number of news items (Fig. 5). By constructing this graph, it is possible to trace how the correlation of time series of different categories changes over the years, to determine the reasons why the behavior of some categories in a certain period is similar to the selected one, and in another period is strikingly different.

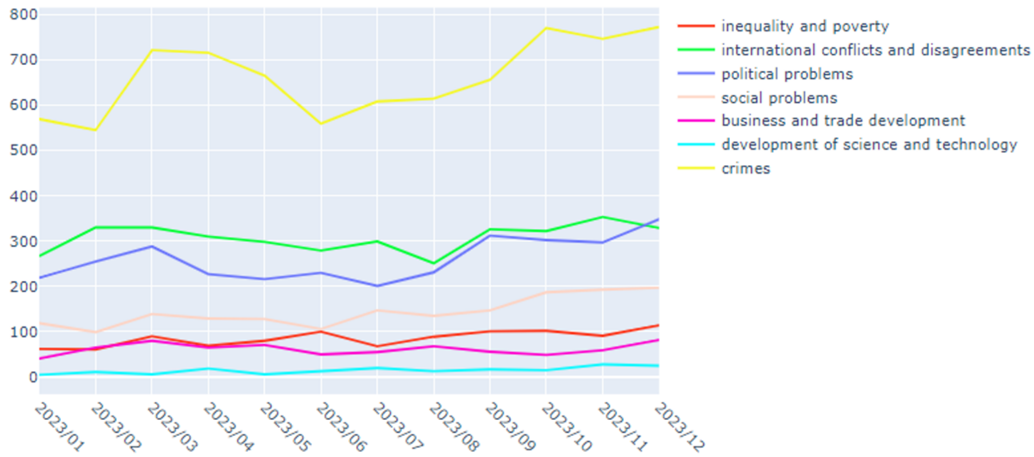

Fig. 5. Graph with the results of time series clustering for the category «development of science and technology» for 2023

A method used for a more informative description of events that occurred on dates that coincide with the points of extremum or inflection in a time series is the selection of key terms. The figure shows the result of the algorithm for the category «development of science and technology» for 2023, where the x-axis shows the date, and the y-axis shows the total number of key terms (Fig. 6). Note that since the source text is in Russian, the key terms are also in Russian.

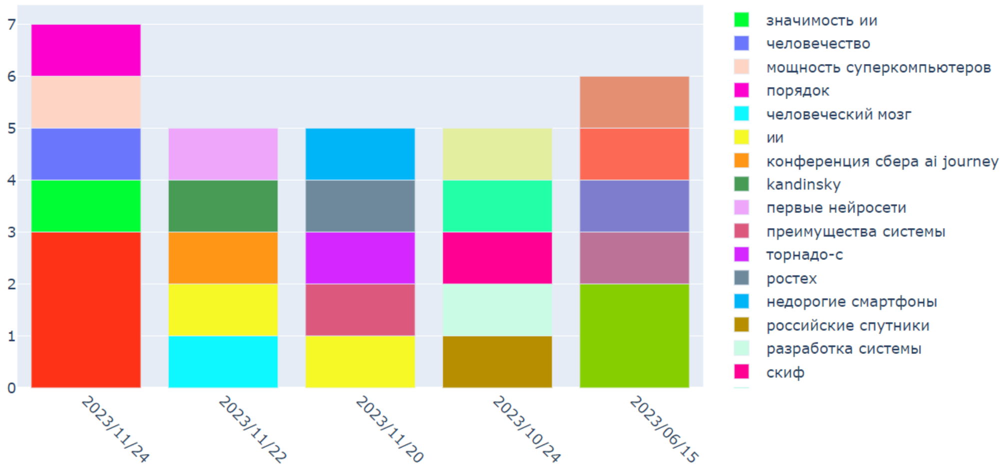

Fig. 6. Graph with the results of defining key terms for the category «development of science and technology» for 2023

The analytical panel is implemented in the Python programming language using the dash library, the results are displayed on the local server. An example of an analytical panel for the category «development of science and technology» for 2023 is shown in the figure (Fig. 7).

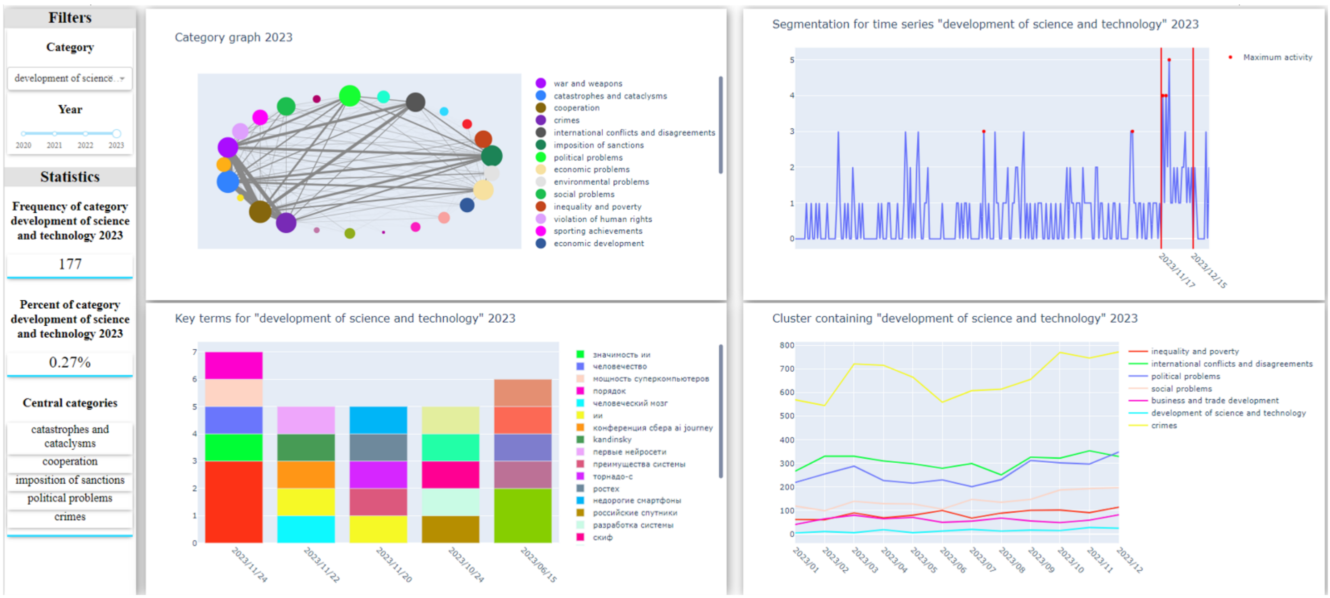

Fig. 7. The result of constructing the final analytical panel with filters «Category» - «development of science and technology», «Year» - 2023.

The method for constructing a social forecast is to write an analytical note on the development trends in the sphere of «development of science and technology» in Russia for the coming years based on the analysis of retrospective data from the analytical panel. For convenience, the analytical note is presented in the form of a table below with an analysis of trends, putting forward the most probable scenarios and recommendations for both the state and business, see Table 4.

Table 4. Social forecast for the category «Development of science and technology» in Russia

Trends by year			Forecasts and recommendations		
	Trends	Sources	Forecast	State	Business
2020	Medicine: development of vaccination technologies. Ecology: technologies for environmentally friendly construction, innovations in materials science, development of the Arctic zone.	Key terms and time series (TS) segmentation: vaccines, ionospheric sounding, frost-resistant concrete. Graphs and clustering of TS: ecological development, medical development, cooperation.	Development of environmentally friendly technologies, as well as the healthcare sector.	Ensuring sufficient funding for scientific research in the fields of ecology and medicine.	Investing in research to improve healthcare infrastructure and approaches. Active participation in environmental initiatives and sustainable development projects.
2021	Education: digitalization of educational processes, development of technology parks and innovation clusters. Medicine: Expanding the use of AI for diagnostics and treatment, continuing work to improve and distribute the Russian vaccine.	Key terms and segmentation of TS: vaccines, technology parks, artificial intelligence. Graphs and clustering of TS: development of social infrastructure, development of medicine, development of education.	Accelerated development of digital technologies, including artificial intelligence, in response to the needs of the digital transformation of the economy and society.	Formulation of national digitalization strategies, including the development of legislation and infrastructure to support digital technologies. Support for research in the field of artificial intelligence and its applications.	Investment in the development of digital platforms, development of innovative products based on artificial intelligence. Active use of digital technologies to improve the efficiency of business processes.
2022	Industry: development of high-tech industries, including the military-industrial complex, development of new technologies to increase productivity. Innovation and collaboration: strengthening technological sovereignty through the development of domestic technologies.	Key terms and segmentation of TS: Russian drones and radar satellites, Russia's technological sovereignty, artificial intelligence. Graphs and clustering of TS: economic development, business and trade development, innovation and import substitution.	Active development of international scientific and technological initiatives, especially with the participation of East Asian countries.	Support for international scientific projects and exchange programs, stimulation of technology transfer and joint research projects.	Active participation in international scientific and technological partnerships, creation of joint research and innovation laboratories with foreign partners.

2023	Security: development of technologies in the field of security and defense, including cyber security and radar technologies. IT: investments in the development of artificial intelligence, especially in the areas of cybersecurity and process automation.	Key terms and segmentation of TS: artificial intelligence, supercomputers, Russian radar satellites. Graphs and clustering of TS: political problems, the introduction of sanctions, international conflicts and disagreements.	Active development of domestic technologies aimed at strengthening national security and sovereignty.	Increasing government funding for security technologies, cyber defense, and high-tech product development. Introducing support measures for national manufacturers and research centers.	Investment in development and innovation aimed at strengthening the country's technological base. Cooperation with government customers and scientific institutions.

5. Discussion of results

Event analysis is an adaptable technology for use as a social forecasting tool, since the choice of data classification system and methods at the stage of analytical research depends on the tasks set. The main result of the research was the author's modification of the event analysis algorithm, in particular, for its use as a social forecasting tool. To put forward specific forecasts, an expert assessment is necessary, but it should be based on data analyzed from various sides, which confirms the relevance and significance of the study.

Compared to the classical event analysis method, where the main method at the stage of analytical comparison, as a rule, is the calculation of the frequency distribution of categories in certain periods, additional approaches have been added that help to consider the relationships and trends in more detail. The modified stages made it possible to more reasonably answer the main questions posed in the classical method: about the assessment of the plot, the initiating subjects, objects, and the time of the event.

The conclusions drawn from the data in the interactive graphs helped to determine the behavior trends in the sphere of «science and technology development» for the corresponding years. The reliability of the obtained results was assessed by manual verification of the identified trends by comparing them with the conclusions previously drawn up by experts. This proves that the adapted event analysis methodology has a sufficient degree of reliability and completeness, which allows using its capabilities as one of the stages in conducting large-scale social research and forecasts. The most probable scenarios for the development of the object of social forecasting and recommendations developed on its basis will allow government agencies to take timely measures to support relevant research by private organizations, and businesses to look for new and profitable niches for development. This will lead to synchronization of the actions of the state and business, their mutually beneficial cooperation and accelerated development of various spheres of life.

The main advantage of the developed method is that it can be adapted to some social forecasting tasks that require classification of data into different categories and their comprehensive analysis both in static and dynamic states. As further work, it is planned to increase the number of data collection sources, improve the quality of data classification by categories and improve the methods used at the stage of analytical comparison.

6. Conclusions and Summary

In the course of the study, an adapted event analysis method was developed and tested, which demonstrated its effectiveness in the context of social forecasting. The main conclusions made as a result of the research can be summarized as follows. The adapted event analysis method demonstrated a high degree of universality, allowing it to be used in various areas of social research. It can be adapted to specific tasks, which makes it suitable for the analysis of both short-term and long-term social phenomena.

The introduction of new analytical tools, such as graph and time series analysis, significantly increased the level of reliability of the data obtained. Manual verification of the identified trends confirmed that the analysis results correspond to real events and their context.

The creation of interactive graphs with the ability to filter data improved the perception of information and provided the ability to dynamically compare different scenarios. This allows not only to visualize the results, but also to deepen the analysis, revealing hidden relationships and patterns.

The results of the study can be used by both government agencies and businesses. Scenarios for the development of events formulated on the basis of the data obtained can serve as a basis for making strategic decisions. This helps to synchronize the actions of various entities and ensure mutually beneficial cooperation.

In the future, we plan to expand data sources, including new information flows and social media. This will improve the classification of events and improve analytical methods. It is also worth considering the possibility of integrating machine learning to automate forecasts and improve their accuracy.

Thus, the adapted event analysis methodology is a powerful tool for social forecasting that combines traditional approaches with modern analytical technologies. Its use can significantly improve the quality of forecasts and promote more effective interaction between government agencies and businesses in the field of scientific research and technology.

References

1. Maltseva, A. V. Using the event analysis technique to study processes in the labor market / A. V. Maltseva et al. // Bulletin of Eurasian Science. - 2012. - No. 3 (12). - P. 12.
2. Novoselov, M. V. Horizons of sociological application of EVENT analysis / M. V. Novoselov // Social and humanitarian sciences: theory and practice. - 2018. - No. 1 (2). - P. 497-506.
3. Azarov, A. A. Predictor Mining: application of data mining methods in social computing tasks / A. A. Azarov // Computer Science and Automation. - 2013. - No. 26. - P. 136-161.
4. Kudryavtseva, V. I. Features of social forecasting in international relations / V. I. Kudryavtseva // Bulletin of BSU. - 2004. - No. 2. - P. 67-71.
5. Ortega, J. L. «How do media mention research papers? Structural analysis of blogs and news networks using citation coupling» / JL Ortega // Journal of Informetrics. – 2021. – No. 15 (3). – P. 101175. – doi.org/10.1016/j.joi.2021.101175.
6. Blokh, I. «News clustering based on similarity analysis» / I. Blokh, V. Alexandrov // Procedia Computer Science. – 2017. – No. 122. – P. 715-719. – doi.org/10.1016/j.procs.2017.11.428.
7. Nekhamkin, A. N. Social forecasting: achievements, shortcomings, ways of improvement / Nekhamkin A. N., Nekhamkin V. A. // Bulletin of the Moscow State Regional University. Series: Philosophical Sciences. - 2020. - No. 2. - P. 57-68.
8. McClelland, C. A. Let the user beware / C. A. McClelland // International Studies Quarterly. – 1983. – No. 27(2). – P. 169-177. – doi: 10.2307/2600544.
9. Lenta.ru: website / founder OOO MINS. – Moscow. – Updated within 24 hours. – URL: https://lenta.ru/ (date accessed: 20.08.2024).
10. Johnson, S. J. A detailed review on word embedding techniques with emphasis on word2vec / S. J. Johnson, M. R. Murty, I. Navakanth // Multimedia Tools and Applications. – 2024. – No. 13 (83). – P. 37979-38007.
11. A, Daud. Knowledge discovery through directed probabilistic topic models: a survey / Daud A., Li J., Zhou L., Muhammad F. // Frontiers of Computer Science in China. - 2010. - No. 2 (4). - P. 280-301. – doi:10.1007/s11704-009-0062-y.

12. Z, Ji. Zero-shot classification with unseen prototype learning / Ji Z., Cui B., Yu Y., Pang Y., Zhang Z. // Neural computing and applications. – 2023. – P. 1-11. – doi: 10.1007/s00521-021-05746-9

13. Hugging Face: website / MoritzLaurer/mDeBERTa-v3-base-mnli-xnli. – Updated within 24 hours. – URL: https://huggingface.co/MoritzLaurer/mDeBERTa-v3-base-mnli-xnli (date of access: 20.08.2024).

14. Ulizko, M. S. Visual analytics of twitter and social media data flows: a case study of covid-19 rumors / M. S. Ulizko, E. V. Antonov, M. A. Grigorieva, E. S. Tretyakov, R. R. Tukumbetova, A. A. Artamonov // Scientific Visualization. – 2021. – No. 4 (13). – P. 144-163.

15. Camacho, D. The four dimensions of social network analysis: An overview of research methods, applications, and software tools / D. Camacho, Á. Panizo-LLedot, G. Bello-Orgaz, A. Gonzalez-Pardo, E. Cambria // Information Fusion. – 2020. – No. 63. – P. 88-120. – doi: 10.1016/j.inffus.2020.05.009.

16. Killick, R. Optimal detection of changepoints with a linear computational cost / R. Killick, P. Fearnhead, I. A. Eckley // Journal of the American Statistical Association. – 2012. – No. 107(500). – P. 1590-1598. – doi: 10.1080/01621459.2012.737745.

17. Ulizko, M. S. Clustering Thematic Information in Social Media / M. S. Ulizko, A. A. Artamonov, J. E. Fomina, E. V. Antonov, R. R. Tukumbetova // Proceedings of the International Conference on Computer Graphics and Vision «Graphicon». – 2022. – No. 32. – P. 403-413.

18. Motovskikh, L. V. Keyword extraction for text classification / L. V. Motovskikh // Bulletin of the Moscow State Linguistic University. Humanities. - 2020. - No. 9 (838). - P. 235-242.

19. Dabbas, E. Interactive Dashboards and Data Apps with Plotly and Dash: Harness the power of a fully fledged frontend web framework in Python—no JavaScript required. / E. Dabbas. – Packt Publishing Ltd, 2021. – 336 p. – ISBN 978-1-80056-891-4.