Monitoring Governmental Topics on Social Media Using Dynamic Topic Modeling

Alena Mamaeva[a], Ivan Mamaev [a, b]

[a] *Saint Petersburg State University,11 Universitetskaya Emb., Saint Petersburg, 199034, Russia*
[b] *Baltic State Technical University "Voenmeh" named after D.F. Ustinov, 1 Krasnoarmeyskaya St. Saint Petersburg, 190005, Russia*

Abstract

The paper discusses the experiments on dynamic topic modeling of the corpus of Russian governmental posts from VKontakte social network. The study is aimed at detecting hidden topical relations and tracking the evolvement of main topics within the text collection. The experiments were conducted on ministerial posts from 15 communities, we give explanations on the resultant dynamic topic models, and establish links with the issues that were important at a specific period in the Russian government. The results justify the use of dynamic topic modeling as a means of social media analysis that can be applied to Russian corpora of Internet texts.

Keywords

Social Network, Governmental Post, Corpus Linguistics, Russian, Dynamic Topic Modeling

1. Introduction

The past decade of the XXI century was marked by the rapid transition of life to the virtual space, as a result, it allowed a great number of scientists to get access to a large amount of textual information, which helps them track the development of the language levels on the Internet. There have been papers dedicating to automatic analysis of Internet texts [2, 3, 4]. Recently it has become popular to use methods of semantic compression for dealing with corpora. One of these methods is topic modeling.

The algorithms of topic modeling can be defined as the compressed representation of documents in order to highlight the main topics. It may be valuable for different spheres of life: political campaigns, business etc. A number of topic modeling methods are widely used in practical research nowadays, some of them are probabilistic models such as pLSA (probabilistic Latent Semantic Analysis) or LDA (Latent Dirichlet Allocation). Using them, one can detect main constant topics in a set of documents. Later other extensions of probabilistic models, that focus on tracking topics over time, appeared in computational linguistics – dynamic topic modeling (DTM). We decided to analyze the evolution of the topical structure of the Russian ministerial corpus proposed in [13] as its texts represent the current situation both in the country and in the world. We give explanations on the final dynamic topic models and establish links with the issues that were important at a specific period in the Russian government.

2. Related works

The main application of dynamic topic modeling is analyzing evolution of topics in large texts collections in different areas of science. For instance, in [15] linguists revealed some niche topics in Russian prose of the first third of the XX century that characterize the main events in the history of Imperial Russia and Soviet Russia: philosopher's ships, revolutions etc. In [7] economists studied how the evolutions of topics on cryptocurrency on forums were interconnected with big events in the cryptocurrency area. They concluded that if any cryptocurrency related service (currency exchanges or mining hardware manufactures) was hacked, users would instantly express their opinions on forums. As a result, the resultant dynamic topic models would change.

IMS 2021 - International Conference "Internet and Modern Society", June 24-26, 2021, St. Petersburg, Russia
EMAIL: az998@mail.ru (A. 1); mamaev_96@mail.ru (A. 2); mamaev_id@voenmeh.ru (A. 2)
ORCID: 0000-0003-2041-4238 (A. 1); 0000-0003-3362-9131 (A. 2)

It is also important to note that the procedures of dynamic topic modeling are widely used for examining governmental texts. In 2020 the pandemic of the coronavirus became one of the issues being discussed both in real life and on the Internet. In [10] the authors analyzed tweets posted by U.S. Governors and Presidential cabinet members to track the decisions made by federal or state authorities. They used a Hawkes binomial topic model [9]. The final evolving models were dedicated to businesses issues, research in creating a vaccine, and calls for social distancing and staying at home. In [14] the authors also discussed the problems of the pandemic from the social network corpus, but they used another approach for obtaining topics – Dynamic LDA. The models partly overlap with the ones described in [10] as both corpora were based on the same social network.

In [5] the evolution of political agenda of the European Parliament plenary was analyzed with the help of dynamic topic modeling based on Non-negative Matrix Factorization (NMF). The authors created a corpus of speeches from 1994 to 2014. The results show that the political agenda of the EP reacts to exogenous events such as the Euro-crisis of 2008.

The paper [4] is dedicated to analyzing politically oriented posts on the US 2016 elections and detecting trolls. They proposed a graph-based algorithm called Dynamic Exploratory Graph Analysis (DynEGA). It helped to reveal the following topics: the right-wing trolls posted messages on supporting Donald Trump's presidential campaign, antiterrorism content, as well as attacking the Democrats; the left-wing discussed supporting the Black Lives Matter movement and activities against black culture and music.

It is also worth mentioning that during the past years Russian scholars started paying special attention to describing automatic analysis of Russian governmental messages from social media, especially in terms of dynamic topic modeling. Papers [11, 12] are dedicated to the analysis of politically oriented texts of RBK Group and governmental websites. The authors ran a number of experiments, they including three different topic modeling algorithms: LSI, LDA and DTM with NMF. As a result, DTM with NMF algorithm proved to be less time-consuming, and its results can be as precise as the results of LSI and LDA algorithms are.

Our experiment is going to continue the contemporary research of Russian corpora with the help of dynamic topic models, we try to focus on the texts of governmental communities on social networks.

3. Experiment

3.1. The corpus of ministerial posts

The material for collecting the dataset was based on the corpus described in [13], but it was enlarged, as the previous corpus contained posts of 2019 and the beginning of 2020. We added posts from other periods of 2020. The corpus con-sists of posts of 15 ministerial communities from VKontakte social network. We divided all the posts into eight periods: 1) winter 2019, 2) spring 2019, 3) sum-mer 2019, 4) autumn 2019, 5) winter 2019-2020 (December 2019, January 2020, and February 2020), 6) spring 2020, 7) summer 2020, 8) autumn and winter 2020. It allows tracking the change of topics during the periods and create the final picture of the governmental development.

3.2. Corpus preprocessing

To implement further procedures of dynamic topic modeling, the corpus needs to be processed using standard NLP approaches.

1. The first step is extracting tokens from the posts.
2. All the tokens are normalized with the help of the pymorphy2 library[2] [6].
3. Then we created a stop-list that is based on a Frequency Dictionary of Contemporary Russian by O.N. Lyashevskaya and S.A. Sharov[3]. This list contains about 1400 words: they are high-frequency conjunctions, prepositions, particles, and common words that can reduce the quality of the resultant models (*прочий (other), накануне (on the eve), etc.*)
4. As any text consists of unigrams and n-grams, we need to enrich the bags-of-words with lexical constructions. We use the gensim library[4] for this purpose. As a result, we obtain lexical

[2] https://pymorphy2.readthedocs.io/en/stable/
[3] http://dict.ruslang.ru/freq.php
[4] https://radimrehurek.com/gensim/

constructions that are typical for ministerial posts: *оказывать_помощь (accord_assistance), первый_медицинский_помощь (first_aid), тушение_пожар (put_out_fire), московский_область (moscow_region), эпидемиологический_обстановка (epidemic_situation), министерство_внутренний_дело (ministry_of_internal_affairs),* etc.

After preprocessing the size of the final corpus turned out to be 61 591 063 words.

3.3. Dynamic topic modeling with non-negative matrix factorization

There are a lot of ways to implement dynamic topic modeling: using the gensim library, FastDTM [1] etc. In [5] the authors used dynamic topic modeling with non-negative matrix factorization for analyzing the speeches of the EP. Also, papers [11, 12, 15] proved the consistency of DTM with NMF. We chose the DTM procedure provided by derekgreene GitHub user[5]. We consider this approach to be effective for the Russian corpus of ministerial posts and try to adapt it for Russian governmental texts on social networks. Below we present the steps to implement dynamic topic modeling.

1. First there is a need to build a skip-gram word2vec model of the entire corpus. The following parameters are used: minimum number of documents for a term to appear – 10, minimum document length – 50 characters, the dimensionality of word vectors – 500, window – 5.
2. We specify a comma-separated range of topics (5, 15) in each time window in order to calculate topic coherence based on the pre-built word2vec model. The top recommended number of topics for each time window was saved in a csv-file.
3. Finally, we automatically search for the optimal number of dynamic topics, specifying the range of topics and basing on the word2vec model.

After applying all the steps, we figured out that six main topics evolve during two years.

Table 1
Main dynamic topics within the ministerial corpus

Dynamic topic	Topic
1	*проект, россия, российский, новый, образование, минпросвещения, работа, школа, производство, школьник (project, russia, russian, new, education, ministry of education, work, school, production, pupil)*
2	*полиция, россия, мчс, мвд, пожарный, полицейский, сотрудник, спасатель, область, служба (police, russia, ministry of emergency situations, ministry of internal affairs, firefighter, policeman, employee, rescuer, region, service)*
3	*военный, учение, россия, флот, боевой, полигон, условный, оборона, военнослужащий, стрельба (military, exercise, russia, navy, combat, firing field, conditional, defense, serviceman, shooting)*
4	*россия, российский, страна, дело, министр, иностранный, международный, вопрос, федерация, оон (russia, russian, country, affair, minister, foreign, international, issue, federation, united nations)*
5	*театр, культура, музей, россия, область, спектакль, фильм, российский, портал, выставка (theater, culture, museum, russia, region, performance, film, russian, site, exhibition)*
6	*россия, спорт, российский, олимпийский, чемпион, день, мир, поздравлять, чемпионат, чемпионка (russia, sport, russian, olympic, champion, day, world, congratulate, championship, champion)*

In the following sections we will comment on each topic and overall situation.

[5] https://github.com/derekgreene/dynamic-nmf

4. Interpretation and Results

4.1. The first dynamic topic

The first set of topical words describes the sphere of education in Russia in 2019-2020. Below we present the evolvement of the topic in five time windows.

Table 2

The evolvement of the first topic

Time window	Set of topical words	Situation
winter 2019	*российский, россия, производство, проект, мантуров, новый, завод, промышленность, просвещение, автомобиль (russian, russia, production, project, manturov, new, plant, industry, education, car)*	-
summer 2019	*россия, российский, проект, развитие, новый, производство, работа, предприятие, министр, промышленность (russia, russian, project, development, new, production, work, enterprise, minister, industry)*	-
spring 2020	*россия, проект, работа, онлайн, студент, российский, время, новый, образование, университет (russia, project, work, online, student, russian, time, new, education, university)*	The beginning of the coronavirus pandemic, everyone starts the remote study.
summer 2020	*россия, проект, университет, новый, российский, спорт, программа, работа, наука, студент (russia, project, university, new, russian, sport, program, work, science, student)*	The enrollment of students in universities, the great number of online and real events on sports and science are held in Russia.
autumn and winter 2020	*минпросвещения, просвещение, учитель, школьник, педагог, школа, всероссийский, образование, страна, новый (ministry of education, education, teacher, student, teacher, school, all-Russian, education, country, new)*	The ministry of education starts publishing information about upcoming exams (the Russian state exam) and competitions for teachers.

Basing on the table above, we can conclude that the education topic was acute during the pandemic of the coronavirus. In 2020 almost all the topics can be compared with the situation in 2019 in which we have only two periods when the topic on education evolved: in spring and summer. Unfortunately, they are hard to connect with real-based events. It may be linked to the focus of the government on the development of education in technical spheres such as engineering, manufacturing etc.

4.2. The second dynamic topic

Unlike the first topic, we can track the evolvement of the second one during all the periods of two years.

Table 3
The evolvement of the second topic

Time window	Set of topical words	Situation
winter 2019	*полиция, мвд, полицейский, россия, сотрудник, мужчина, водитель, помощь, мчс, служба (police, ministry of internal affairs, policeman, russia, employee, man, driver, help, ministry of emergency situations, service)*	-
spring 2019	*полиция, россия, мвд, сотрудник, мчс, полицейский, пожарный, опасность, внутренний, весна (police, russia, ministry of internal affairs, employee, ministry of emergency situations, policeman, firefighter, danger, internal, spring)*	-
summer 2019	*мчс, россия, пожарный, полиция, спасатель, человек, область, пожар, вода, спасательный (ministry of emergency situations, russia, firefighter, police, rescuer, man, region, hot, water, rescue)*	The beginning of forest fires in the Far East of Russia.
autumn 2019	*россия, мчс, пожарный, полиция, мвд, спасатель, сотрудник, полицейский, человек, работа (russia, ministry of emergency situations, firefighter, police, ministry of internal affairs, rescuer, employee, policeman, man, work)*	-
winter 2019-2020	*полиция, мчс, россия, мвд, пожарный, полицейский, сотрудник, спасатель, служба, область (police, ministry of emergency situations, russia, ministry of internal affairs, fire, policeman, employee, rescuer, service, region)*	-
spring 2020	*полиция, россия, пожарный, мвд, мчс, коронавирус, сотрудник, спасатель, дезинфекция, помощь (police, russia, firefighter, ministry of internal affairs, ministry of emergency situations, coronavirus, employee, rescuer, disinfection, help)*	The ministry of emergency situations disinfects a great number of facilities in the Russian Federation.
summer 2020	*полиция, россия, мчс, мвд, пожарный, сотрудник, полицейский, область, ребёнок, служба (police, russia, ministry of emergency situations, ministry of internal affairs, fire, employee, policeman, region, child, service)*	-
autumn and winter 2020	*полиция, мвд, россия, мчс, сайт, сотрудник, спасатель, мошенник, распродажа (police, ministry of internal affairs, russia, ministry of emergency situations, site, employee, rescuer, fraud, sale)*	Posts try to pay attention to frauds on the Internet, especially during the Black Friday.

After analyzing the table, it is clear that the topic, dedicated to the police and rescue operations, has almost the same distribution in all the time windows. We can state that all the posts of these communities are written on the only topics: work of policemen and rescuers. There are only three well-interpreted topics. For instance, the third time window (summer 2019) is notable as its topical words like *пожарный, спасатель, пожар, вода (firefighter, rescuer, fire, water)* indicate the topic of forest fires in Russia that are typical for this season. At the same time, it should be noted that the topic is not fully covered in the seventh time window although it is also summer. This fact can be explained that in 2020 there were less fires than in 2019.

4.3. The third dynamic topic

The third niche topic describes the military and navy service in the Russian Federation that also can be seen in all eight time windows.

Table 4

The evolvement of the third topic

Time window	Set of topical words	Situation
winter 2019	*военный, флот, боевой, корабль, стрельба, условный, россия, учение, оборона, самолёт (military, navy, combat, ship, shooting, conditional, russia, exercises, defense, aircraft)*	-
spring 2019	*военный, сирийский, рукбан, лагерь, флот, россия, боевой, оборона, стрельба, этап (military, syrian, rukban, camp, navy, russia, combat, defense, shooting, stage)*	Russia-Syrian diplomatic and military delegation tried to resolve the conflict in the Rukban camp between the USA and the Syrian refugees.
summer 2019	*военный, россия, армия, учение, конкурс, оборона, условный, боевой, международный, войско (military, russia, army, exercises, competition, defense, conditional, combat, international, army)*	The increasement of number of international competitions and exhibitions (Armygames 2019 and others).
autumn 2019	*военный, учение, россия, оборона, войско, условный, боевой, флот, корабль, противник (military, exercises, russia, defense, army, conditional, combat, navy, ship, enemy)*	-
winter 2019-2020	*военный, россия, боевой, оборона, учение, армия, полигон, флот, российский, шойгу (military, russia, combat, defense, exercises, army, firing field, navy, russian, shoygu)*	The number of official visits to military facilities were made by Sergey Shoygu, the minister of Defence.
spring 2020	*военный, россия, сербия, военнослужащий, стрельба, полигон, российский, минобороны, коронавирус, специалист (military, russia, serbia, serviceman, shooting, firing field, russian, ministry of defense, coronavirus, specialist)*	Russian servicemen deliver medical equipment to Serbia to help the nation.
summer 2020	*военный, конкурс, россия, полигон, учения, армия, российский, военнослужащий, команда, экипаж (military, competition, russia, firing field, exercises, army, russian, serviceman, team, crew)*	Returning to calling to military service after the stabilization of the situation with the pandemic of the coronavirus.
autumn and winter 2020	*военный, россия, российский, учение, полигон, минобороны, условный, военнослужащий, пациент, сила (military, russia, russian, exercises, firing field, ministry of defense, conditional, military officer, patient, force)*	Servicemen build hospitals for patients diagnosed with a new coronavirus.

In the area of military and navy service there are more topics to be interpreted. For instance, in spring 2019 despite the measures taken by Syrian and Russin authorities, the Rukban camp of internally displaced people still existed up to the present moment, and its residents are still unable to return home due to tough opposition from the side of the USA, so it was one of the acute topics that time. In winter 2019-2020 the minister of defense had a series of official visits to the military facilities and held some meetings with ministers of defense of other countries. Most of these events were held because of the upcoming Victory Day to commemorate the 75th Diamond Jubilee of the capitulation of Nazi Germany. Later, in autumn and winter 2020 main topics on social networks were dedicated to building a number of permanent and temporary hospitals for patients diagnosed with a coronavirus. Although servicemen started building in spring 2020, the problem became pivotal only at the end of 2020 when the number of coronavirus cases had increased greatly compared to spring 2020.

4.4. The fourth dynamic topic

The fourth set of topics describes the sphere of external affairs in all the time windows.

Table 5
The evolvement of the fourth topic

Time window	Set of topical words	Situation
winter 2019	*россия, российский, дело, министр, страна, иностранный, международный, лавров, февраль (russia, russian, affair, minister, country, foreign, international, lavrov, february)*	-
spring 2019	*россия, российский, министр, лавров, встреча, страна, федерация, международный, вопрос, обсе (russia, russian, minister, lavrov, meeting, country, federation, international, issue, osce)*	Sergey Lavrov conducted some official meetings in Moscow including the meeting with OSCE general secretary.
summer 2019	*россия, российский, дело, министр, иностранный, международный, страна, вопрос, лавров (russia, russian, affair, minister, foreign, international, country, question, lavrov)*	-
autumn 2019	*россия, дипломат, страна, молодой, международный, министр, вопрос, иностранный, оон, федерация (russia, diplomat, country, young, international, minister, issue, foreign, united nations, federation)*	Russia hosts the third international meeting of young diplomats from Russia, India and China.
winter 2019-2020	*россия, российский, страна, дело, вопрос, международный, министр, иностранный, федерация, сотрудничество (russia, russian, country, affair, issue, international, minister, foreign, federation, cooperation)*	-
spring 2020	*россия, российский, страна, сша, вопрос, посольство, дело, международный, рейс, федерация (russia, russian, country, usa, issue, embassy, affair, international, flight, federation)*	Russia launches export flights for the Russians who are abroad. The discussion of riots in the USA.
summer 2020	*россия, российский, страна, независимость, дело, международный, индия, отношение, сингапур, иностранный (russia, russian, country, independence, affair, international, india, relationship, singapore, foreign)*	The community often posts congratulations on the anniversaries of the Independence Day (India, Singapore, etc.).
autumn and winter 2020	*россия, беларусь, страна, союз, международный, лукашенко, министр, лавров, вопрос, государство (russia, belorussia, country, union, international, lukashenko, minister, lavrov, issue, country)*	Lavrov has several meetings with Belorussian official representatives both in Russia and Belorussia and discussing the development of the Union State.

The obtained topics are rather stable as the lemmata don't change a lot within all the topics. If we have a close look at the posts of the ministry of external affairs, we will see that the posts are usually describe the main events in which Sergey Lavrov took part, special days in the lives of other countries and some official meetings. Only 2020 has certain burning issues like the organization of flights for the Russians that are not in the country because of closing the borders or the discussion of US riots by reason of the presidential race or the Black Lives Matter movement.

4.5. The fifth dynamic topic

The fifth topic on the cultural events was represented in five periods, the spring and summer of 2019 and the spring of 2020 weren't mentioned.

Table 6

The evolvement of the fifth topic

Time window	Set of topical words	Situation
winter 2019	*театр, хороший, сцена, спектакль, россия, театральный, артист, музеи, фильм, картина (theatre, good, stage, performance, russia, theatrical, artist, museums, film, picture)*	-
autumn 2019	*россия, культура, российский, театр, фильм, проект, музей, ночь, искусство, выставка (russia, culture, russian, theatre, film, project, museum, night, art, exhibition)*	The beginning of "The Artnight" festival, which is timed to the Day of National Unity.
winter 2019-2020	*культура, театр, церемония, россия, музей, фильм, ольга, спектакль, официальный, новый (culture, theatre, ceremony, russia, museum, film, olga, performance, official, new)*	Olga Lyubimova became a new Minister of Culture, she being presented to the public. She also made some official visits to Saint Petersburg and Svetlogorsk.
summer 2020	*культура, россия, область, музей, новый, театр, библиотека, возвращаться, российский, работа (culture, russia, region, museum, new, theatre, library, return, russian, work)*	Russian regional museums and exhibitions prepare to welcome visitors after the lockdown.
autumn and winter 2020	*театр, культура, музей, россия, фестиваль, ссылка, библиотека, спектакль, концерт, искусство (theatre, culture, museum, russia, festival, link, library, performance, concert, art)*	-

According to the table, the cultural sphere on the social networks is well-reflected: the topics describe upcoming festivals, real and online performances, the visit of the new Minister of culture to theatres and libraries, etc. Only two periods cannot be interpreted. As we consider, these periods were rather stable in this sphere.

4.6. The sixth dynamic topic

The sport topic is shown in six periods excluding the spring and summer of 2020 when Russia couldn't hold any sports events.

Table 7
The evolvement of the sixth topic

Time window	Set of topical words	Situation
winter 2019	*россия, праздновать, российский, поздравлять, чемпион, завоевать, чемпионка, мир, сегодня, спорт (russia, celebrate, russian, congratulate, champion, win, championess, world, today, sport)*	-
spring 2019	*россия, российский, проект, кратный, день, министр, образование, первый, развитие, антон (russia, russian, project, multiple, day, minister, education, first, development, anton)*	-
summer 2019	*россия, олимпийский, спорт, день, российский, сборная, чемпион, чемпионат, мир, чемпионка (russia, olympic, sport, day, russian, national team, champion, championship, world, championess)*	-
autumn 2019	*россия, праздник, российский, министр, развитие, образование, день, конкурс, страна, первый (russia, feast, russian, minister, development, education, day, competition, country, first)*	-
winter 2019-2020	*россия, российский, проект, хоккейный, развитие, золото, страна, сборная, работа (russia, russian, project, hockey, development, gold, country, combined team, work)*	The Youth hockey team won the third Winter Youth Olympic Games in the USA.
autumn and winter 2020	*россия, спорт, российский, минобрнауки, проект, фальков, развитие, новый, наука, спортивный (russia, sport, russian, ministry of education, project, falkov, development, new, science, sports)*	-

Unfortunately, the resultant topics don't allow us to highlight pivotal events in the sports sphere. The only well-described topic is dedicated to winning in the third Winter Youth Olympic Games. At the same time, we see that educational sphere somehow interact with the sports one as different topical lemmata can be in one set (*sport – ministry of education* etc.). It can be explained by the fact that the ministry of sports tries to promote sports activities in Russian school and make PE lessons one of the most important one for students.

5. Discussions

Below we present a summary table denoting the statistics of the resultant topics.

Table 8.
Statistics on topics

Dynamic Topic	Number of time windows	Number of interpreted topics
Education	5	3
Emergency situations	8	3
Military cases	8	6
External affairs	8	5
Culture	5	3
Sport	6	1
Mean	6,7 (6-7 time windows)	3,5 (3-4 interpreted topics)

While applying the algorithm of dynamic topic modeling to the corpus of ministerial posts, we can describe main advantages and disadvantages. First of all, according to Table 8, more than a half of the corpus turned out to be well-interpreted. Despite the similar sets of lemmata within each period, there can be some special words that help us to understand a described situation (for instance, the *rukban* lemma denotes the place of a possible military conflict). Moreover, there are few verbs in all the topics, it makes the interpretation of topical sets easier. If there had been more verbs (*declare, say, state, claim* etc.), it would have been harder to name the topics. Also, we can distinguish some relations between obtained lemmata like in the LDA topic models [8]: *россия – страна (russia – country), военный – оборона (military – defense), учитель – школьник (teacher – student)* and others.

As for disadvantages, the final dynamic topic models don't include the collocations that we used for the enrichment of the corpus. In this case, further development of models can be connected with the using another application for detecting lexical constructions. For instance, we can use NLTK that provides the detection based on different measures (t-score, log-likelihood, etc.). The combination of the measures may improve the chance of their appearing in the models. Of course, the models based on the word2vec corpus need more training in the future. Changing the parameters may allow us to obtain more precise corpus, it leading to appearing time periods that weren't covered in the present paper.

Unfortunately, DTM with NMF failed to highlight topics that are on everyone's lips: for instance, there is a topic dedicated to health that was acute in 2020. In [13] it is explained that the coronavirus topic is scattered across all the ministerial communities, and it can be absorbed by other topical sets. For instance, when speaking about dynamic topics on external affairs, we can distinguish a set that is indirectly connected with the coronavirus topic: *россия, российский, страна, сша, вопрос, посольство, дело, международный, рейс, федерация (russia, russian, country, usa, issue, embassy, affair, international, flight, federation)*. As it was previously mentioned, this one is dedicated to Russian export flights. There are a lot of specific topics that might be unknown for an average inhabitant: the celebration of the Independence Day of certain countries, the conflict in the Rukban camp, etc. Further tuning of the algorithm and corpus enlargement may help to improve the quality of topics.

6. Conclusion

In the present paper, we have analyzed the development of ministerial post on VKontakte social network for two years. We prove that if some issues discussed in the posts of social networks are pivotal, they will be reflected in a certain time window of the dynamic topic models. At the same time, it will be hard to detect any changes if the topics are evenly distributed within all the time periods. From the point of view of linguistics, we can highlight different syntagmatic and paradigmatic relations in each topic.

Further experiments can be aimed at:

- comparing the results of dynamic topic models and the "openness" of the state in online communities;
- using other algorithms of dynamic topic modeling to distinguish their common and different features;
- involving other Russian social networks to compare the activity of ministries in them and see if there are any difference compared to the posts on VKontakte;
- enriching the existing corpus with collocations as it may help to interpret certain periods.

7. References

[1] Bhadury, A., Chen, J., Zhu, J., Liu, S.: Scaling up dynamic topic models. In: International World Wide Web Conference, 381–390 (2016).

[2] Bodrunova, S., Blekanov, I., Kukarkin, M.: Topic modeling for Twitter discussions: Model selection and quality assessment. In: Proceedings of the 6th SGEM International Multidisciplinary Scientific Conferences on SOCIAL SCIENCES and ARTS SGEM2018, Science and Humanities, 207–214. STEF92 Technology Ltd., Sofia, Bulgaria (2019).

[3] Bodrunova, S., Blekanov, I., Kukarkin, M.: Topics in the Russian Twitter and relations between their interpretability and sentiment. In: Sixth International Conference on Social Networks Analysis, Management and Security, 549–554 (2019).

[4] Golino, H., Christensen, A. P., Moulder, R. G., Kim, S., Boker, S. M.: Modeling latent topics in social media using Dynamic Exploratory Graph Analysis: The case of the right-wing and left-wing trolls in the 2016 US elections. In: PsyArXiv, 1–19 (2020).

[5] Greene, D., Cross, J.P.: Exploring the political agenda of the European parliament using a dynamic topic modelling approach. In: Political Analysis, volume 25, issue 1, 77–94 (2017).

[6] Korobov, M.: Morphological Analyzer and Generator for Russian and Ukrainian Languages. In: Analysis of Images, Social Networks and Texts, 320–332 (2015).
[7] Linton, M., Teo, E.G.S., Bommes, E., Chen, C.Y., Härdle, W.K.: Dynamic Topic Modelling for Cryptocurrency Community Forums. In: Härdle W., Chen CH., Overbeck L. (eds) Applied Quantitative Finance. Statistics and Computing, 355–372. Springer, Berlin, Heidelberg (2017).
[8] Mitrofanova, O.: Probabilistic Topic Modeling of the Russian Text Corpus on Musicology. In: LMAC 2015, CCIS 561, 69–76. Springer Nature (2015).
[9] Mohler, G., McGrath, E., Buntain, C., LaFree, G.: Hawkes binomial topic model with applications to coupled conflict-twitter data. In: Annals of Applied Statistics, volume 14, number 2, 1984–2002 (2020).
[10] Sha, H., Hasan, M.A., Mohler, G., Brantingham, P.J.: Dynamic Topic Modeling of the COVID-19 Twitter Narrative Among US Governors and Cabinet Executives. In: Workshop Proceedings of the 14th International AAAI Conference on Web and Social Media, 1–6 (2020).
[11] Skitalinskaya, G., Aleksandrov, M., Danilova, V., Stefanovsky, D.: Website materials of governmental and regional administrations of Russia in terms of dynamic topic modeling (Materialy saytov pravitel'stva i regional'nykh administraciy Rossii v zerkale dinamicheskogo tematicheskogo modelirovaniya). In: Mathematical modeling of social processes (Matematicheskoe modelirovanie social'nykh processov), volume 20, 166–174 (2018).
[12] Skitalinskaya, G.: Analysis of news dynamics using two-step dynamic topic modeling algorithms (Analiz dinamiki novostey s pomoshch'yu dvukh-shagovykh algoritmov dinamicheskogo tematicheskogo modelirovaniya). In: Mathematical modeling of social processes (Matematicheskoe modelirovanie social'nykh processov), volume 19, 97–104 (2017).
[13] Zaitseva, A., Mamaev, I.: Automatic detection of the topical structure of the ministerial posts on social networks. In: Proceedings of the Computational Models in Language and Speech Workshop (CMLS 2020) co-located with 16th International Conference on Computational and Cognitive Linguistics (TEL 2020), CEUR Workshop Proceedings, volume 2780, 32–42 (2020).
[14] Zamani, M., Schwartz, H.A., Eichstaedt, J., Guntuku, S.C., Ganesan, A.V., Clouston, S., Giorgi, S.: Understanding weekly COVID-19 concerns through dynamic content-specific LDA topic modeling. In: Proceedings of the 4th Workshop on Natural Language Processing and Computational Social Science (NLP+CSS), 193–198 (2020).
[15] Zamiraylova, E., Mitrofanova, O.: Dynamic topic modeling of Russian prose of the first third of the XXth century by means of non-negative matrix factorization. In: Proceedings of the III International Conference on Language Engineering and Applied Linguistics (PRLEAL-2019), CEUR Workshop Proceedings, volume 2552, 321–339 (2020).