

Calls to Action on Social Media: Potential for Censorship and Social Impact

Anna Rogers, Olga Kovaleva, Anna Rumshisky

Department of Computer Science

University of Massachusetts Lowell

Lowell, MA 01854, USA

{arogers, okovaleva, arum}@cs.uml.edu

Abstract

Calls to action on social media are known to be effective means of mobilization in social movements, and a frequent target of censorship. We investigate the possibility of their automatic detection and their potential for predicting real-world protest events, on historical data of Bolotnaya protests in Russia (2011-2013). We find that political calls to action can be annotated and detected with relatively high accuracy, and that in our sample their volume has a moderate positive correlation with rally attendance.

1 Introduction

Calls to action (CTAs) are known to be effective means of mobilization in social networks (P.D. Guidry et al., 2014; Savage et al., 2016), and they are also known to be a target for censorship by authoritarian states (King et al., 2013, 2014). However, to the best of our knowledge, they have not been systematically evaluated for their potential for automatic detection and predicting offline protest events.

We contribute a case study on political CTAs in historical data on Bolotnaya protests in Russia (2011-2013). We identify 14 core and borderline types of political CTAs, and we show that they are relatively easy both to annotate (with IAA 0.78) and to classify (F1 of 0.77, even with a small amount of annotated data). All of that puts them at high risk for censorship, but also opens the possibilities to track such censorship. We also find that in Bolotnaya data, the volume of CTAs on social media has a moderate positive correlation with actual rally attendance.

2 Related Work

2.1 Prediction of social unrest with social media data

Social movements differ in their goals (reforms or preservation of status quo), size of the group they are targeting, methods, and other factors (Snow et al., 2004), but their success always ultimately depends on successful mobilization of new participants. The role of social media in that has been clear since the Arab Spring (Dewey et al., 2012). Social media fundamentally changed the social movements, enabling new formats of protest, a new model of power, and greater activity outside of formal social organizations (Earl et al., 2015).

Expert judgement is famously unreliable for predicting political events (Tetlock, 2017). So, if social media play such an important role in social movements, can they also be used to track and perhaps predict the real-world events? By now hundreds of studies explored various kinds of forecasting based on social media (Phillips et al., 2017; Agarwal and Sureka, 2015), from economic factors to civil unrest. Most of them show that their techniques do have predictive merits, although some skepticism is warranted (Gayo-Avello et al., 2013).

Most of the civil unrest prediction work is done on Twitter and news, sometimes in combination with other sources such as blogs and various economic indicators (Ramakrishnan et al., 2014; Manrique et al., 2013). The basic instrument of analysis in most of these studies is time series of social media activity on a given topic (Hua et al., 2013). Data filtering is typically performed via protest-related keywords, hashtags, geolocation or known activist accounts. Many studies also rely on some combination of spatiotemporal features (e.g. Ertugrul et al. (2019); Zhao et al. (2015)). The texts of posts could be mined for extracting struc-

tured event-related information, or dense meaning representations could be used without identifying specific features, such as doc2vec representations of news articles and social media streams (Ning et al., 2016). Additionally, social network structure (Renaud et al., 2019) and activity cascades (Cadena et al., 2015) were also found useful, as well as mining and incorporating demographic information (Compton et al., 2014).

The typically-used features that are extracted from social media text include time, date, and place mentions, sentiment polarity of the post, and presence of violent words (Benkhelifa et al., 2014; Bahrami et al., 2018). Another popular approach relies on manually created lexicons of protest-related vocabulary (such as “police”, “molotov”, “corruption”, etc.) combined with event-specific names of politicians, activists etc. (Spangler and Smith, 2019; Mishler et al., 2017). Korolov et al. (2016) identifies possible stages of mobilization in a social movement (sympathy, awareness, motivation, ability to take part). To the best of our knowledge, CTAs have not been systematically investigated for their predictive potential.

2.2 Censorship in social media

Similarly to the systems used to predict offline events, many current censorship systems seem to rely on keywords (MacKinnon, 2009; Verkamp and Gupta, 2012; Chaabane et al., 2014; Zhu et al., 2013). However, it is highly likely that states engaging in suppression of collective action are researching more sophisticated options, and it is therefore imperative that censorship monitors also have better tools to monitor what gets deleted.

Much of research on Internet censorship focuses on China, where there does not seem to be a single policy enforced everywhere: local organizations and companies show significant variation in their implementations (Knockel et al., 2017; Miller, 2018; Knockel, 2018). This depends on only on the goals of the platform, its ties or dependence on the government, but also the market forces: a competing platform that would find a way to censor less would be more attractive for the users (Ling, 2010). The actual process also varies based on the available resources: it is likely that larger companies have significant censor staff (Li and Rajagopalan, 2013), while others might rely only on simple keyword filtering. Finally, even at government level not all criticism is disal-

lowed: a significant degree of freedom seems to be allowed with respect to local social movements that are unlikely to become a threat to the regime (Qin et al., 2017).

Calls to action seem to be an obvious candidate for types of verbal messages strongly associated with social movements, and they are known to be effective means of mobilization in social networks (P.D. Guidry et al., 2014; Savage et al., 2016). In particular, King et al. (2013, 2014) report that sometimes the censors let through materials that are simply critical of government, but flag materials with collective action potential (such as calls to attend a rally or support opposition). The effort to shut down the collective action is clear, for example, from the fact that Instagram was simply shut down for 4 days while photos of Hong Kong protests were trending (Ma, 2016).

To the best of our knowledge, the censorship potential of CTAs has also not been specifically addressed in the context of political protests.

3 Case study: Bolotnaya protests, Russia

Our case study is the 2011–2013 Russian protests, of which the best known is the “March of the Millions” on May 6, 2012 in the Bolotnaya square in Moscow. The movement was widespread, with protests in many smaller Russian cities and towns. The protesters were opposing fraudulent elections and government corruption. This was the largest protest movement in Russia since 1990s.

The experiments discussed below rely on the “Bolotnaya” dataset that contains posts, likes and groups of users from VKontakte, the largest Russian social network. The main statistics for the dataset are shown in Table 1. It was created by the New Media Center (Moscow, Russia) in 2014 on the basis of a list of 476 protest groups, which was compiled by Yandex (the largest Russian search engine). The data is used by an agreement with New Media Center.

Enikolopov et al. (2018) report that the number of users of VK social network in different locations was in itself associated with higher protest activity, and locations where the user base was fractured between VKontakte and Facebook had fewer protests, which overall suggests that the main role of social media was the ease of coordination (rather than actual spreading of information critical of the government). This is consistent with the reported role of Facebook in Egypt’s

Protest groups	476
Posts in the protest groups	196017
Protest groups' members	221813
Posts posted by the group	77604
Posts shared from other groups	81403
Time frame covered	08.2010 - 10.2014
Users liking the posts by the protest groups	57754

Table 1: Basic statistics on the Bolotnaya dataset

Tahrir Square protests (Tufekci and Wilson, 2012). If these conclusions are correct, then higher volume of CTA should in itself also be a factor in higher protest attendance.

4 Defining Calls to Action

Prototypical CTAs are imperatives prompting the addressee to perform some action, such as “*Don’t let the government tell you what to think!*”. This seems like a straightforward category to annotate, but in reality CTAs may be expressed in various ways, including both direct and indirect speech acts. There are many borderline cases that would in the absence of clear guidelines decrease inter-annotator agreement (IAA). There is relevant work on the task of identification of requests in emails (Lampert et al., 2010) and intention classification for dialogue agents (Quinn and Zaiane, 2014), but, to the best of our knowledge, this work is the first to create a detailed schema for CTA annotation in the context of a political protest.

The current work on censorship is concerned not so much with CTAs in particular, but with a broader category of “material with collective action potential”. King et al. (2013) defines such materials as those that ‘(a) involve protest or organized crowd formation outside the Internet; (b) related to individuals who have organized or incited collective action on the ground in the past; or (c) relate to nationalism or nationalist sentiment that have incited protest or collective action in the past.’ In other words, this definition only concerns offline events, and does not include various forms of “crowd protesting” such as calls to share information critical of the government.

Based on extensive manual analysis of samples from Bolotnaya dataset, we identified 5 core and 9 borderline cases for political CTAs, shown in Figure 2. Since we were interested in CTAs for social movements, we excluded any other CTAs that would formally fit the criteria, such as invitations, marketing CTAs etc. We also excluded

any other protest-related posts, such as reports of protest events. Of the core and borderline CTA cases, we chose to consider 8 as CTAs.

This choice does not have a firm theoretical underpinning and would vary depending on the researcher’s perspective and the case study. For example, in our Bolotnaya data we opted to not include broad rhetorical questions like “*For how much longer shall we put up with this?*”, but in a different context (especially in a different culture) they could be key. Inter-annotator agreement depends on how the guidelines’ describing the chosen policy explicitly.

5 Annotation study

Pilot data analysis made it clear that the CTA and non-CTA classes were not balanced. Since CTAs overall constitute a small portion of all posts, we pre-selected the data for annotation using a manually created seed list of 155 protest-related keywords and phrases, such as “participate”, “share”, “join”, “fair elections”, etc.

We used our schema to develop detailed annotation guidelines for an annotation study on 1000 VKontakte posts from Russian Bolotnaya data. The annotation was performed on the level of full post, not individual sentences. We considered a post as CTA if it included even one instance of a political CTA as defined above. Ambiguous cases were treated as political CTA, as long as they *could* function as such: for example, *Join us tomorrow!* could refer to both a protest or a birthday party.

Each post was annotated by 3 native Russian speakers, using the classification interface of Prodigy¹ annotation tool. The inter-annotator agreement as estimated by Krippendorff’s alpha was .78. In the end, we obtained 871 posts on which at least 2 annotators agreed. 300 of them were identified as CTAs, and 571 - as non-CTAs. This was used as the training dataset for the work to be described in subsequent sections.

6 Classification

In our experiments, we randomly split the collected CTA dataset into the train and test parts in the 80/20% ratio. We selected Logistic Regression (LR) and Support Vector Machine classifier with a linear kernel (SVC) as our baseline models. Both models were used as implemented in the

¹<https://prodi.gy/>

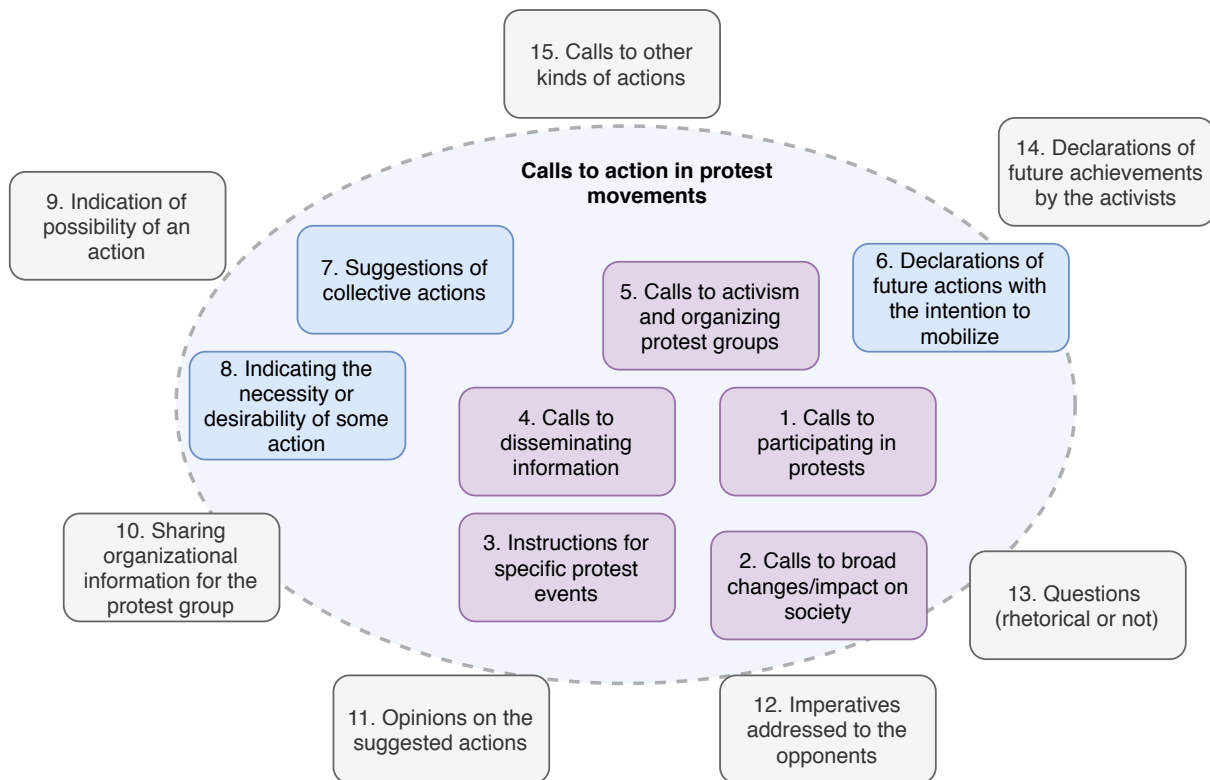


Figure 1: The core and peripheral cases of political CTAs. This study focuses on types 1-8. Examples for each type:

1. Everybody, join us tomorrow in Sakharov square!
2. If you love Russia, if you love your home city of Smolensk, start the fight with the crooks and thieves!
3. Do not form a line or arrange to meet in a specific place.
4. Invite foreign press and TV – let them see what is going on in our capital!
5. Observers in Kaluga, please respond!
6. That's ok, we will tell them what we think of them even in the square in front of the Central market!
7. I suggest we put on white stripes on our arms as a symbol of honest elections. That's easy to do!
8. On the 10th of March we should come in large numbers!
9. You can download the leaflet with the invitation here.
10. This is the beginning! We will start activities when we will have 50 members. We repeat, participation in this group can only be active.
11. I do like the idea of the government's resignation, but I think your slogans are too emotional. Furthermore, I'm against calling an early election.
12. Out with you, McFaul! And take Putin and Medvedev with you, together with Nemtsov and Chirikova!
13. Is THAT really our choice? (rhetorical)
Today at 10 pm Vlad and I are going to post the leaflets around the city. Who wants to help us? (factual)
14. Together we will get rid of Putin's lies and dictatorship!
15. Everybody, come to my birthday party on Saturday!

Figure 2: Types of calls to actions in Russian social networks, with examples

scikit-learn² library. In both cases we used TF-IDF representations of both original posts and posts lemmatized with pymorphy³ library (Korobov, 2015). We picked the best regularization hyperparameters for each model through cross-validation based on the average F1 score over 5 folds.

The current state-of-the-art deep learning approaches rely on large Transformer-based models pre-trained on large text corpora and then fine-tuned for a given task. In particular, we tried two versions of BERT (Devlin et al., 2019): the multilingual model released in the PyTorch repository of BERT⁴, and the Russian version (RuBERT) released by DeepPavlov⁵. The latter model is initialized as multilingual BERT and further fine-tuned on Russian Wikipedia and news corpora (Kuratov and Arkhipov, 2019). Both models have 12 layers and 180M parameters. We trained both models for 40 epochs with the batch size of 32 and the learning rate of $5e^{-5}$.

Additionally to BERT representations, we experimented with the contextual embedder of the ELMo model (Peters et al., 2018) pre-trained for Russian and released by DeepPavlov⁶. The posts were split into sentences using the NLTK library⁷ and each sentence token was encoded by the ELMo embedder into a 1024-dimensional vector. The classification was performed by a standard LSTM network (Hochreiter and Schmidhuber, 1997) with a hidden size of 256 units followed by a linear layer. We trained the network for 25 epochs and with the learning rate of 0.001.

The results of all the classification experiments are shown in Table 2. The best performance was achieved by RuBERT, with LSTM on ELMo close second. The effect of lemmatization with linear classifiers is inconsistent. It is interesting that simple logistic regression with lemmatized TF-IDF representation of the posts is only 4 F1 points below ELMo, which suggests that the overall classification task is not very difficult.

²<https://scikit-learn.org>

³<https://github.com/kmike/pymorphy2/>

⁴<https://github.com/huggingface/pytorch-transformers>

⁵<http://docs.deep-pavlov.ai/en/master/components/bert.html>

⁶<https://github.com/deepmipt/DeepPavlov/tree/master/deep-pavlov>

⁷<https://www.nltk.org/>

Classifier	Acc.	F1
LR (no lemmatization)	0.78	0.67
LR (lemmatization)	0.82	0.71
SVC (no lemmatization)	0.80	0.68
SVC (lemmatization)	0.78	0.65
BERT multilingual	0.8	0.73
RuBERT	0.86	0.78
LSTM on ELMo	0.83	0.75

Table 2: CTA classification results

7 CTAs for Predicting social unrest

To estimate the potential usefulness of CTAs as indicators of offline protest events, we ran the trained RuBERT CTA classifier over 91K posts falling in the date range between Dec 2011 through Jul 2013 from the Bolotnaya dataset.

Figure 3 shows the volume of posts identified as CTAs, plotted against the Wikipedia data about attendance of individual rallies⁸. When no attendance data is available, we assume that there were 0 protest events. The two green lines correspond to upper and lower attendance estimates. The blue line shows the detected CTAs.

Despite the noisiness and incompleteness of the available protest data (see subsection 9.1), the Pearson’s correlation between attendance estimates and the number of detected CTAs is about 0.4, which is considered to be “moderate”. This could make CTAs a useful additional factor to systems based on spatiotemporal, demographic, and/or network activity features.

⁸Russian Wikipedia, Protest movement in Russia (2011-2013): <https://tinyurl.com/y46qyb9w>.

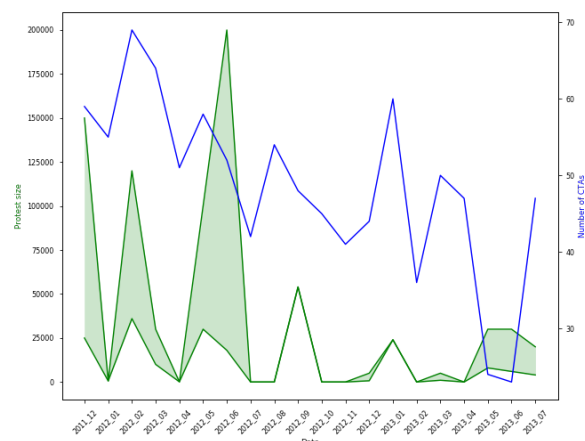


Figure 3: The correlation between the detected CTAs (blue) and the rally attendance (green) per month. The two green lines reflect the upper and lower attendance estimates depending on the source of data used.

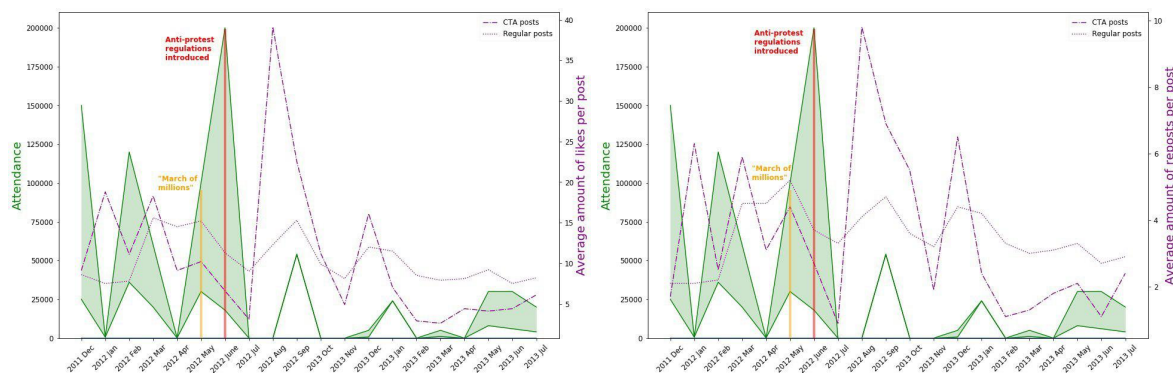


Figure 4: Average number of likes and reposts on CTA and non-CTA posts vs rally attendance

We also conducted experiments to estimate the real-world effect of likes and reposts of CTA posts. Intuitively, one would expect that a higher number of likes and reposts of CTA posts should result in higher attendance for protest rallies. To see whether that was the case for Bolotnaya data we calculated the number of shares and likes on posts detected as CTAs by our classifier, and all other posts in the sample. Figure 4 shows these numbers plotted against the attendance of the protest events.

The pattern we actually observed in Bolotnaya data is different: before the March of the Millions the average number of both reposts and likes is spiking before a protest event, and going down after it. This corresponds to preparation and the aftermath of a major event. Interestingly, after the March of the Millions there was much like/repost activity which did *not* result in any larger events. This can be attributed to the introduction of the anti-protest laws that effectively stifled the movement: the link between social media and real-world activity clearly becomes weaker.

8 Discussion

8.1 Censoring CTAs

Our annotated dataset is quite small (only 871 posts), and this is on purpose: our point is that even with such a small (and unbalanced) dataset it is already possible to obtain a reasonably good classifier (and its performance would likely improve with more data). This is an additional factor in censorship potential: if a system for detecting CTAs could be built quickly and cheaply, it is highly likely that such systems are already being developed by the well-sponsored research teams employed by the authoritarian states. Our study should at least level the playing field for censor-

ship monitors, as will be discussed below. The guidelines we developed will be made available on request by researchers.

The data specific to our Bolotnaya case study would not be openly released because, 8 years after the events, the issues that were driving them continue to be the key factors in the activities of the Russian opposition movements. In particular, Russia has just experienced a new wave of protests estimated to be the largest since 2012 (Wilpert, 2019), also driven by the issues of corruption and fair elections, and resulting in hundreds of arrests (BBC, 2019a,b). Many of the key political figures on both sides are also the same. All this makes our Bolotnaya data potentially useful for censoring new protests.

The situation actually became worse for the protesters because since 2011 a range of new laws went into action to restrict activity on social media. The social network users and popular bloggers are personally identifiable (via their phone numbers), VPNs are illegal, and social network operators are obliged to store activity data for 6 months and decrypt them for authorities (House, 2018; Wesolowsky, 2019). Activists can be imprisoned for sharing “inauthentic and illegal information of social importance”, a broad formulation that is interpreted freely by the authorities (Schreck, 2017).

9 Web monitoring potential

As discussed above, materials with collective action potential are already undergoing active censorship in authoritarian states, and it is highly likely that classifiers similar to ours are actually already in place. We hope that our study would somewhat level the playing field for those who combat the censorship.

In particular, if authoritarian states are able to detect CTAs for censorship, it is equally possible to use CTA classifiers in monitoring systems that would scan the web for content that is removed, and report on the ongoing censorship. At present monitoring efforts rely on manual and keyword analysis (MacKinnon, 2009; Verkamp and Gupta, 2012; Chaabane et al., 2014; Zhu et al., 2013). Note that if data on what is being censored were continually collected, the censors would actually “help” the monitoring efforts: by flagging and removing content they would essentially be providing free annotation.

One more use of CTA classifiers would be to help the protesters to find new creative ways of expressing their views that would pass the automatic filters, such as the Chinese egao phenomenon (Horsburgh, 2014; Yates and Hasmath, 2017). Providing an independent web-service with which the activists could check how easy their message is to flag would arguably boost such creativity and provide the activists with their own weapons in the linguistic race against the authoritarian states.

9.1 Limitations

The present study is limited in several ways. First, the small size of annotated data only provides a lower bound on the performance on CTA classifiers, which would likely increase with more annotated data. However, our point was not in achieving the best possible performance, but in showing that automatic detection of CTAs is possible even with relatively little data.

The second limitation comes from the lack of reliable attendance data for Bolotnaya protests - a situation pervasive in authoritarian states with tight control over media and civic organizations. For example, the official police report for the Kaluga square event stated 8,000 people, while opposition politicians reported 100-120,000. According to bloggers, there were 30,000 people, and a Russian parliamentarian estimated 50-60,000⁹. However, this limitation would impact any prediction method, and it arises precisely in the situations in which the most important protest activity is happening.

Last but not the least, the whole field of forecasting in with social media data is suffering from

the lack of common best practices, which is aggravated by the impossibility to replicate most of the results due to data sharing concerns (Phillips et al., 2017). This study is not an exception: Bolotnaya data in this study was used by an agreement with New Media Center, and we cannot release it publicly. Without major changes in accessibility of social network data for researchers, the only way forward in the field seems to be partial validation by similar patterns uncovered in other case studies.

10 Conclusion

Calls to action are a vital part of mobilization effort in social movements, but, to the best of our knowledge, their potential for censorship and predicting offline protest events has not yet been evaluated.

We examine political calls to action in a case study on historical data on Bolotnaya protests in Russia (2011-2013). We identify 14 core and borderline types of political CTAs, and we show that they are relatively easy to annotate (with IAA 0.78) and detect automatically (F1 of 0.77, even with a small amount of annotated data), which puts them at high risk for censorship in authoritarian states. We also find that in Bolotnaya data, the volume of CTAs on social media has a moderate positive correlation with actual rally attendance.

11 Acknowledgements

We thank Viktoria Khaitchuk, Gregory Smelkov (University of Massachusetts Lowell), and Mikhail Gronas (Dartmouth College) for their help in developing the annotation schema and the training set. We also thank the anonymous reviewers for their valuable feedback. This work was supported in part by the U.S. Army Research Office under Grant No. W911NF-16-1-0174.

References

- Swati Agarwal and Ashish Sureka. 2015. *Applying Social Media Intelligence for Predicting and Identifying On-line Radicalization and Civil Unrest Oriented Threats*. *arXiv:1511.06858 [cs]*.
- Mohsen Bahrami, Yasin Findik, Burcin Bozkaya, and Selim Balcisoy. 2018. *Twitter Reveals: Using Twitter Analytics to Predict Public Protests*. *arXiv:1805.00358 [cs]*.
- BBC. 2019a. *More than 600 detained in banned Russian protest*. *BBC News*.

⁹Russian Wikipedia, Protest movement in Russia (2011-2013): <https://tinyurl.com/y46qyb9w>.

- BBC. 2019b. [Thousand arrests at Moscow election protest](#). *BBC News*.
- E. Benkhelifa, E. Rowe, R. Kinmond, O. A. Adedugbe, and T. Welsh. 2014. [Exploiting Social Networks for the Prediction of Social and Civil Unrest: A Cloud Based Framework](#). In *2014 International Conference on Future Internet of Things and Cloud*, pages 565–572.
- Jose Cadena, Gizem Korkmaz, Chris J. Kuhlman, Achla Marathe, Naren Ramakrishnan, and Anil Vullikanti. 2015. Forecasting social unrest using activity cascades. *PLoS one*, 10(6):e0128879.
- Abdelberi Chaabane, Terence Chen, Mathieu Cunche, Emiliano De Cristofaro, Arik Friedman, and Mohamed Ali Kaafar. 2014. [Censorship in the wild: Analyzing Internet filtering in Syria](#). In *Proceedings of the 2014 Conference on Internet Measurement*, pages 285–298.
- Ryan Compton, Craig Lee, Jiejun Xu, Luis Artieda-Moncada, Tsai-Ching Lu, Lalindra De Silva, and Michael Macy. 2014. [Using publicly visible social media to build detailed forecasts of civil unrest](#). *Security Informatics*, 3(1):4.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. [BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding](#). In *Proceedings of the 2019 NAACL-HLT*, pages 4171–4186.
- Taylor Dewey, Juliane Kaden, Miriam Marks, Shun Matsushima, and Beijing Zhu. 2012. [The impact of social media on social unrest in the Arab Spring](#). Technical report, Defense Intelligence Agency, Stanford University.
- Jennifer Earl, Jayson Hunt, R. Kelly Garrett, and Aysenur Dal. 2015. [New technologies and social movements](#). *The Oxford handbook of social movements*, pages 355–366.
- Ruben Enikolopov, Alexey Makarin, and Maria Petrova. 2018. [Social Media and Protest Participation: Evidence from Russia](#). SSRN Scholarly Paper ID 2696236, Social Science Research Network, Rochester, NY.
- Ali Mert Ertugrul, Yu-Ru Lin, Wen-Ting Chung, Muheng Yan, and Ang Li. 2019. [Activism via attention: Interpretable spatiotemporal learning to forecast protest activities](#). *EPJ Data Science*, 8(1):5.
- Daniel Gayo-Avello, Panagiotis Takis Metaxas, Eni Mustafaraj, Markus Strohmaier, Harald Schoen, Peter Gloor, Evangelos Kalampokis, Efthimios Tambouris, and Konstantinos Tarabanis. 2013. [Understanding the predictive power of social media](#). *Internet Research*.
- Sepp Hochreiter and Jürgen Schmidhuber. 1997. [Long Short-Term Memory](#). *Neural Computation*, 9(8):1735–1780.
- Nicola Horsburgh. 2014. *Chinese Politics and International Relations: Innovation and Invention*, 1 edition.
- Freedom House. 2018. [Freedom on the Net 2018: Russia](#).
- T. Hua, C. Lu, N. Ramakrishnan, F. Chen, J. Arredondo, D. Mares, and K. Summers. 2013. [Analyzing Civil Unrest through Social Media](#). *Computer*, 46(12):80–84.
- Gary King, Jennifer Pan, and Margaret E. Roberts. 2013. [How Censorship in China Allows Government Criticism but Silences Collective Expression](#). *American Political Science Review*, 107(2):326–343.
- Gary King, Jennifer Pan, and Margaret E. Roberts. 2014. [Reverse-engineering censorship in China: Randomized experimentation and participant observation](#). 345(6199):1251722.
- Jeffrey Knockel. 2018. *Measuring Decentralization of Chinese Censorship in Three Industry Segments*. Ph.D., The University of New Mexico.
- Jeffrey Knockel, Lotus Ruan, and Masashi Crete-Nishihata. 2017. [Measuring Decentralization of Chinese Keyword Censorship via Mobile Games](#). In *7th USENIX Workshop on Free and Open Communications on the Internet (FOCI 17)*.
- Mikhail Korobov. 2015. [Morphological Analyzer and Generator for Russian and Ukrainian Languages](#). In *Analysis of Images, Social Networks and Texts, Communications in Computer and Information Science*, pages 320–332. Springer.
- R. Korolov, D. Lu, J. Wang, G. Zhou, C. Bonial, C. Voss, L. Kaplan, W. Wallace, J. Han, and H. Ji. 2016. [On predicting social unrest using social media](#). In *ASONAM 2016*, pages 89–95.
- Yuri Kuratov and Mikhail Arkhipov. 2019. [Adaptation of deep bidirectional multilingual transformers for russian language](#). *arXiv preprint arXiv:1905.07213*.
- Andrew Lampert, Robert Dale, and Cecile Paris. 2010. [Detecting Emails Containing Requests for Action](#). In *Proceedings of NAACL-HLT*, pages 984–992, Los Angeles, California. ACL.
- Hui Li and Megha Rajagopalan. 2013. [At Sina Weibo’s censorship hub, China’s Little Brothers cleanse online chatter](#).
- Yutian Ling. 2010. [Upholding Free Speech and Privacy Online: A Legal-Based and Market-Based Approach for Internet Companies in China](#). *Santa Clara Computer & High Technology Law Journal*, (1):175–218.
- Veronica Ma. 2016. [Propaganda and Censorship: Adapting to the Modern Age](#). *Harvard International Review; Cambridge*, 37(2):46–50.

- Rebecca MacKinnon. 2009. China’s censorship 2.0: How companies censor bloggers. *First Monday*, 14(2).
- P. Manrique, H. Qi, A. Morgenstern, N. Velásquez, T. Lu, and N. Johnson. 2013. Context matters: Improving the uses of big data for forecasting civil unrest: Emerging phenomena and big data. In *2013 IEEE International Conference on Intelligence and Security Informatics*, pages 169–172.
- Blake Andrew Phillip Miller. 2018. [The limits of commercialized censorship in China](#).
- A. Mishler, K. Wonus, W. Chambers, and M. Bloodgood. 2017. [Filtering Tweets for Social Unrest](#). In *2017 IEEE 11th International Conference on Semantic Computing (ICSC)*, pages 17–23.
- Yue Ning, Sathappan Muthiah, Huzefa Rangwala, and Naren Ramakrishnan. 2016. Modeling precursors for event forecasting via nested multi-instance learning. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1095–1104. ACM.
- Jeanine P.D. Guidry, Richard D. Waters, and Gregory D. Saxton. 2014. [Moving social marketing beyond personal change to social change: Strategically using Twitter to mobilize supporters into vocal advocates](#). *Journal of Social Marketing*, 4(3):240–260.
- Matthew Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. 2018. [Deep Contextualized Word Representations](#). In *Proceedings of NAACL-HLT*, pages 2227–2237.
- Lawrence Phillips, Chase Dowling, Kyle Shaffer, Nathan Hodas, and Svitlana Volkova. 2017. [Using Social Media to Predict the Future: A Systematic Literature Review](#). *arXiv:1706.06134 [cs]*.
- Bei Qin, David Strömberg, and Yanhui Wu. 2017. [Why Does China Allow Freer Social Media? Protests versus Surveillance and Propaganda](#). *Journal of Economic Perspectives*, 31(1):117–140.
- Kevin Quinn and Osmar Zaiane. 2014. Identifying questions & requests in conversation. In *Proceedings of the 2014 International C* Conference on Computer Science & Software Engineering*, page 10. ACM.
- Naren Ramakrishnan, Patrick Butler, Sathappan Muthiah, Nathan Self, Rupinder Khandpur, Parang Saraf, Wei Wang, Jose Cadena, Anil Vullikanti, Gizem Korkmaz, Chris Kuhlman, Achla Marathe, Liang Zhao, Ting Hua, Feng Chen, Chang Tien Lu, Bert Huang, Aravind Srinivasan, Khoa Trinh, Lise Getoor, Graham Katz, Andy Doyle, Chris Ackermann, Ilya Zavorin, Jim Ford, Kristen Summers, Youssef Fayed, Jaime Arredondo, Dipak Gupta, and David Mares. 2014. [‘Beating the News’ with EMBERS: Forecasting Civil Unrest Using Open Source Indicators](#). In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD ’14, pages 1799–1808, New York, NY, USA. ACM.
- Molly Renaud, Rostyslav Korolov, David Mendonça, and William Wallace. 2019. Social Network Structure as a Predictor of Social Behavior: The Case of Protest in the 2016 US Presidential Election. In *Recent Developments in Data Science and Intelligent Analysis of Information*, Advances in Intelligent Systems and Computing, pages 267–278. Springer International Publishing.
- Saiph Savage, Andrés Monroy-Hernández, and Tobias Hollerer. 2016. [Botivist: Calling Volunteers to Action using Online Bots](#). In *19th ACM Conference on Computer-Supported Cooperative Work and Social Computing*. ACM.
- Carl Schreck. 2017. [Russian Activist Convicted For Repost...About Repost He Was Convicted For](#).
- David A. Snow, Sarah A. Soule, and Hanspeter Kriesi. 2004. Mapping the Terrain. In *The Blackwell Companion to Social Movements*, pages 3–16. Blackwell Pub. Ltd.
- Ethan Spangler and Ben Smith. 2019. [Let Them Tweet Cake: Estimating Public Dissent using Twitter](#). SSRN Scholarly Paper ID 3418050, Social Science Research Network, Rochester, NY.
- Philip E. Tetlock. 2017. *Expert Political Judgment: How Good Is It? How Can We Know?*, new edition. Princeton University Press, Princeton Oxford.
- Zeynep Tufekci and Christopher Wilson. 2012. [Social Media and the Decision to Participate in Political Protest: Observations From Tahrir Square](#). *Journal of Communication*, 62(2):363–379.
- John-Paul Verkamp and Minaxi Gupta. 2012. [Inferring Mechanics of Web Censorship Around the World](#).
- Tony Wesolowsky. 2019. [Six Steps Russia Is Taking Toward Restricting Its Internet](#).
- Greg Wilpert. 2019. [Russia’s Anti-Gov’t Protests Draw Large Crowds, but Putin Remains Secure](#).
- Mathew Yates and Reza Hasmath. 2017. [When a joke is more than a joke: Humor as a form of networked practice in the Chinese cyber public sphere](#). *The Journal of Chinese Sociology*, 4(1):17.
- Liang Zhao, Feng Chen, Chang-Tien Lu, and Naren Ramakrishnan. 2015. Spatiotemporal event forecasting in social media. In *Proceedings of the 2015 SIAM International Conference on Data Mining*, pages 963–971. SIAM.
- Tao Zhu, David Phipps, Adam Pridgen, Jedidiah R. Crandall, and Dan S. Wallach. 2013. The velocity of censorship: High-fidelity detection of microblog post deletions. In *Presented as Part of the 22nd USENIX Security Symposium (USENIX Security 13)*, pages 227–240.