Application of Automated Tools in Researching Internet Discourses: Experience of Using the Recurrent Neural Networks for Studying Discussions on Pension Reform

Petr Begen[1[0000-0002-0613-3133]], Yuri Misnikov[1[0000-0002-1948-619X]]
and Olga Filatova[2[0000-0001-9568-1002]]

[1] ITMO University, Kronverksky pr., 49, 197101, St. Petersburg, Russia
[2] St. Petersburg State University, Universitetskaya nab., 7–9, 199034, St. Petersburg, Russia
peetabegen@yandex.ru, yuri.misnikov@gmail.com,
o.filatova@spbu.ru

Abstract. The paper presents the results of an experiment that applied the Recurrent Neural Network (RNN) and long short-term memory (LSTM) networks to assess how accurately they can determine the attitude of 998 participants towards the pension reform policy in Russia who posted 10,592 comments on 16 online forums in 11 cities. The training set was assembled and coded according to a proposed conceptual model of a moral discourse based on Jurgen Habermas's discourse ethics theory. The main conclusion of this experiment is that the discourse-based approach — based on the identification of basic validity claims — can be instrumental in building training datasets for deep machine learning on a socially salient topic. The experiment also shows benefits and limitations of using artificial neural networks for a deeper understanding of the results of public discussions in an online environment. The main benefit was that the built neural networks have proven to be sufficiently accurate in predicting positions of discourse participants towards the pension reform policy, with almost 90% in the case of binary classification (two "For" and "Against" positions). However, the accuracy level drops with the inclusion of a third "Neutral" category (to 78%), which was a major limitation of the research; that is, the variation in the prediction accuracy is due to the uneven distribution of data among categories and an increase of new data. Yet this indicator is still acceptable when working with Internet discourse data.

Keywords: recurrent neural networks, machine learning, Internet discourse, e-participation, validity claims, deliberation.

1 Introduction

An interest in creating and testing tools for electronic participation, started over two decades ago, still continues to address the growing needs of involving citizens in government policy- and decision making as new technologies emerge and expand. While there are many digital platforms that have been created to facilitate government-citizen interaction over the Internet both in Russia and other countries, there are few of them that help citizens debate public policy issues among themselves in a more collaborative and dialogical manner by, for example, understanding how their positions compare with those of other participants, and how their contributions to the debate shape its intersubjectively aggregated outcomes. Awareness of this individual-collective linkage would help discussants anticipate their role in and an impact on a debate course and thus become more responsible in choosing the way they write their posts; that is, to be attentive not only to the contributions of other participants, but also to "predict" the moral and ethical side of their own inputs.

In the context of broader public communication experience, citizen-to-citizen interaction is considered in this paper an essential discursive practice of "will formation", in Jurgen Habermas' terminology. The presented in the paper study examines real-life socio-political Internet-based public discussions as a form of e-participation in collective will formation among interacting discourse participants. The main emphasis is placed on using artificial intelligence research methods to assess predictability of connection between individual discursive practices and respective worldview positions on a publicly salient issue.

Typically, the goal of analyzing discussions in the Internet space concerning important for society political decisions is to examine the efficacy of tools of citizens' participation in politics and define their role in government-citizen interaction [1]. This paper focuses instead on citizen-to-citizen discursive interaction. It instrumentalizes a discourse ethics theory of Jurgen Habermas as a conceptual premise of his model of deliberative democracy [2]. Habermas's theory is based on the concept of basic validity claims that discourse participants apply to validate the normative (also moral, ethical) rightness of their utterances. These claims are the requests, statements addressed to other participants with the purpose to articulate certain "truths" and seek (and predict) their response as an act of validation of such claims.

As a rule, the contents of such validation can be revealed through either consent or disagreement, often supported by arguments revealing reasons for that. Such discourses are viewed as an ethically justified form of political organization of society allowing to overcome political differences within the free citizenry under the model of deliberative democracy. This study further advances the Habermasian concept of basic validity claims applied to online political debate among lay citizens from the will formation perspective [1, 3].

It is also a convenient methodological tool for studying online discursive practices, in contrast to a semantically shallow tonal (sentiment) analysis, which is methodologically less "attached" to explanatory social theories and displays little interest in interactional aspects existing between discourse participants. The tonal analysis tools are unable to take account of such deliberative aspects of Internet discourses as dialogicality and interactivity. To address the latter, the posted texts need to be presented as interconnected sequences of claims to validity examined through agreement-disagreement as a core feature of discursive interactivity.

The rise of information and communication technologies coupled with a significant increase in the amount of data and the growth of computing power have made it possible to use deep machine learning algorithms for text analysis primarily in computer linguistics studies [4, 5]. In this context, an experiment was conducted to train a neural network to distinguish discursive situations based on agreement and disagreement among citizens when they discuss a publicly significant socio-political topic. The Russian government's policy of the retirement pension reform was used as a topic of such experimentation. In addition, an automated tool was developed to help construct training datasets comprised of the labeled posts as an input to the neural network.

2 Research Design and Methodology

2.1 Conceptualizing Internet discourse

From the analytical perspective, an Internet discussion becomes an ethically oriented discourse when imagined as a particular combination and pattern of validity claims to normative rightness revealed via agreement-disagreement on a certain issue. That is, it occurs when negative and positive sentiments of individual texts disclose intersubjective debating communities whose participants express either similar or diverging positions by respectively agreeing or disagreeing with one another, with support of arguments (a claim validation act).

The coding of the agreement-disagreement dichotomy is subject to identifying the availability of argumentation behind the validated in such a way positions. The presence or absence of such positions is an evidence of whether the discourse is morally oriented or remains at the level of a rationally pragmatic exchange of opinions and views. The translation of texts into the claims to normative rightness and moral worldviews can conceptually be imagined in the form of a discourse pyramid, as schematically illustrated below in Fig. 1.

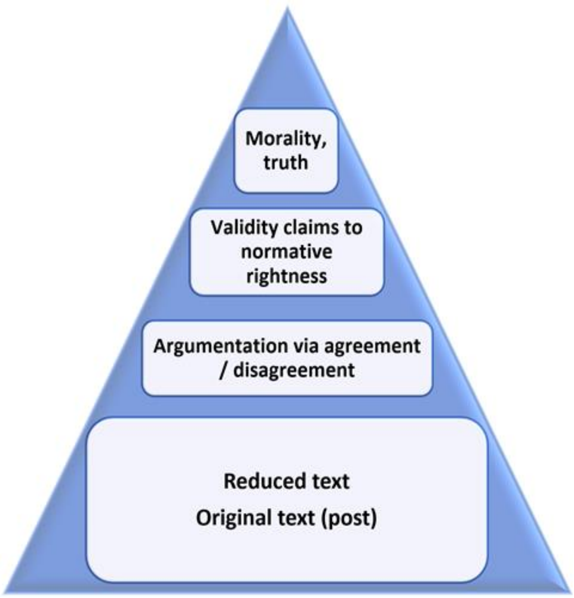

Fig. 1. Aggregation model of moral discourse

At the lower level, a human coder trained in content analysis analyzes the original author's text with the help of the coding tools to reduce it to a minimum set of words without the loss in meaning by removing those parts of the post that are redundant from the point of view of understanding the main meaning of the text (prepositions, conjunctions, other); however, the text is not altered or edited by the coder so as its intended meaning is intact.

The middle-lower level aggregates the original text to conclude whether it contains an act of agreement or disagreement in relation to the preceding posts, and whether any arguments are presented in support of the articulated agreement-disagreement (there could be no opinion expressed too). Two types of the actualization of agreement-disagreement are distinguished: one type looks whether there is a directly manifested agreement-disagreement by using such built-in functions as "answer", "quote", naming the author(s) of preceding posts; the second type is an indirect form of expressing agreement-disagreement without addressing particular posts. Such posts (usually one post) can be logically identified thanks to their intended meaning and with the help of other cues. The identification of the second type can cover the contents of ten preceding posts.

The identification of agreement-disagreement sequences helps identify reasons (arguments) behind that and move to the 3rd middle-upper level of discourse. The coder at this level formulates the validated through agreement-disagreement claims. The upper level aggregates these claims further by formulating broader worldview positions (as normatively, morally right) belonging to intersubjective solidarities formed among discourse participants. This is usually achievable when analyzing disputes on socially and politically salient topics in the format, for example, "The government is right in implementing pension reform" or "It is wrong to accept the pension reform is such a form".

Filling these levels with appropriate content means labeling the source text as an input for training the neural network, which should learn how the original text is linked to specific validity claims and agreement-disagreement sequences. It can be done by forming pairs of posts connected with one another through a semantically similar content along the bottom-up axis of aggregation process. This has proven to be a reliable way of step-by-step coding; for example, the analysis of online discussions regarding the attitude towards the policy of destroying the embargoed western food products allowed to reveal "For" and "Against" intersubjective solidarities around several important for them issues [1, 6, 7, 8].

Such studies, based on manual coding, require a lot of time, attention on the part of coders and can be implemented only on a relatively small data sample. Moreover, for each new study, it is necessary to conduct a new coding. At the same time, the once trained neural networks may be repeatedly used on a different (although topically similar) material due to continuous learning. The experience of applying machine learning principles for discourse analysis is presented below.

2.2 Data sample and coding procedure

The purpose of the experiment was to identify possibilities of using the learning potential of artificial neural networks for a deeper understanding of the results of public discussions. This experiment aimed at training the neural network so that it learns to determine the “For”, “Against” and “Neutral” attitude discourse participants towards the pension reform policy. The first step was to form a dataset for feeding into the neural network. The Internet discussions on the websites of eleven cities in Russia were selected for analysis according to the typology of cities set by the Ministry of Economic Development — the largest, large, big, medium and small [9]. Two cities were selected from each group, their most popular Internet sites were identified and online discussions on pension topics were analyzed. The cities included: St. Petersburg and Volgograd (largest), Kaliningrad and Sevastopol (large), Bratsk and Nalchik (big), Belorechensk and Khanty-Mansiysk (medium), Uryupinsk and Snezhinsk (small); special attention in the study was given to Moscow.

Three maximally different forums representing various social groups were selected as additional data sources for machine learning: the All-Russian female portal Woman.ru, the otzovik.com review site and the website of the electronic newspaper KM.ru.

A total of 16 forums were analyzed, containing 10,592 comments were posted by 998 people. There were 304 “For” posts (3%); 2,510 “Against” posts (24%); 7,778 “Neutral” (73%). Data for machine learning were fed into in Excel tables and then exported in .csv format for convenient work with them in a software environment. All collected posts were labeled by the coders in three possible ways: 0 — category “Against”, 1 — “For” and 2 — “Neutral”.

To implement automated tools for researching Internet discourse and writing software, we used the Python programming language and third-party libraries for working with data and machine learning methods. The .csv data were loaded into the program using the pandas library. The following two columns of data were used: (1) “message text” and (2) “For-1 / Against-0 / Neutral-2”, loaded as the semantically connected pairs X and Y, respectively, where X is a dataset for training at the input end (post text), and Y — the output end (participants’ position “For”, “Against”, “Neutral”).

2.3 Applying machine learning algorithms for discourse analysis

To make the data compatible with the requirements of the machine learning algorithms, the text was preprocessed using the built-in Tokenizer class of Keras library, which allows for deleting redundant characters, reduce the words to lower case, calculate the frequency of occurrence of words, removing punctuation marks and invisible characters, numbers, etc. In accordance with international practice of text processing [10, 11], the most common and rare words were not taken into account leaving a set of 3,000 most relevant words. Further, the labeled data were divided into training and test samples in the 80/20 ratio, i.e., 20% of the total data set was used for the final testing of the trained model. The test data were not included into the training of the model; accordingly, the model “saw” these data for the first time during the final testing.

The preprocessed text was presented in the form of vector sequences of a maximum of 200 characters. All words in the training set were represented as vectors with a given dimension. Initially, such a vector was filled with random numbers (most often these were zeros). During the learning process, these values were changed so that the words used in the same semantic context are as close as possible in the vector space. A recurrent neural network (RNN) with LSTM blocks was used as a machine learning algorithm. This choice is due to the fact that recurrent networks can use their internal memory to process sequences of arbitrary length, whilst the LSTM blocks (blocks with long short-term memory) as a recurrent network can cope very well with classification and forecasting problems [12, 13, 14].

Such an approach allowed the model to memorize during the training the previous values in the vector sequences for further decision-making and adjustment of weights on hidden layers of the neural network. The initial number of iterations for which the model had to learn was 100 cycles. To prevent overtraining of the network, the EarlyStopping function was used. It checks the error between the loss functions and network errors. It was noticed that this "early stop" function worked after the 10th training cycle. The used recurrent neural model consisted of the following layers and blocks:

- input layer (i.e., a sequential data set) — Input;
- a connection layer that translates the entire vocabulary of words into a custom dimension (200 characters) for further training — Embedding;
- LSTM-block that stores memorized information from previous sequences — LSTM;
- a fully connected layer with a linear rectification unit (ReLU activation function), which is responsible for determining the values of neurons and their settings — Dense;
- a layer that does not allow overtraining of the network on data (by eliminating neurons that return 0 for any values and parameters) — Dropout;
- an output layer with a customizable number of outputs (in our case 2 or 3, with sigmoid or softmax activation functions) — Output.

Two approaches were used to build the output layer:

1. The use of binary classification: only the categories "For" or "Against", while the category "Neutral" was not used in this case; respectively, 2,814 statements were used (1/4 of all the posts).
2. The use of classification in three categories: "For", "Against" and "Neutral". All the coded 10,592 posts were processed.

3 Research Results

Applying a binary classification approach on the output layer (two categories "For" and "Against"), the model was able to determine such categories with the accuracy of approximately 89%. This is acceptable for determining tonality of the posts; however, with an increase in the data volume and the appearance of new words that were not involved in compiling the vocabulary by the model, the level of accuracy rapidly drops with the value of the loss function reaching 4%. In the second case, when applying all three classification categories, the accuracy indicators for determining the category were around 78%. The result is less accurate compared with the binary classification approach due to the uneven distribution of data among categories and an increase of new data. Yet this indicator is also acceptable for further use when working with Internet discourse. An open test was also conducted, i.e., random data were taken from the test sample and fed into the trained model. The accuracy of the determined categories was compared with the previous "true" results.

Such testing showed that the utterances falling under the category "Neutral" towards pension reform were determined with the higher level of probability (higher accuracy) than the posts related to the categories "Against" and "For" due to the uneven distribution of data among the categories (three-fourth of the total dataset volume falls into the category "Neutral"). Accordingly, for further interpretation of the obtained results, new experiments are needed either with a large amount of more evenly distributed data across the categories or using other machine learning methods.

4 Conclusions

The main conclusion of this experiment is that the discourse-based approach — based on the identification of basic validity claims — applied for building a training set for deep machine learning has appeared to be a workable solution to predict positions of discourse participants in relation to a certain topic. In this case it was a morally loaded topic of the fairness of the government pension reform policy in the people's eyes.

Another important conclusion is that built Recurrent Neural Network (RNN) has proven to be sufficiently accurate in predicting the intended meaning of the posts. In this experiment, for the binary classification, the accuracy rate was visibly higher at the level of 89% than for the classification of three positions (categories "For", "Against", "Neutral"), which was about 78%. The variation in the prediction accuracy demonstrates that its level decreases if the data are unevenly distributed across position categories, i.e., there is not enough data for the neural network to learn dominant patterns and eliminate errors.

These are promising results considering a rather small dataset of about 10,000 posts, which was dominated by the "Neutral" position category 2, with 74% of all discourse posts and just 3% galling under "For" category. This is an important research outcome requiring more care when building a more balanced training dataset, which, in turn, requires choosing adequate social media sources to improve classification accuracy.

The future work will continue experimenting with the conceptual model of discourse and its modification to obtain better accuracy results in combination with different approaches to machine learning. An automated toolkit was developed for the study of Internet discourses based on recurrent neural networks with an LSTM block. Possible new approaches that can be used are the following: (a) increasing or adjusting the array of training datasets balanced across chosen categories; (b) using pre-trained dictionaries of terms (distribution thesaurus) [15, 16]; (c) using the "attention" layer [17, 18] to prevent the effect of "oblivion" in a recurrent neural network; (d) using the algorithms

Word2Vec [19], Doc2Vec [20], GloVe [21] to take a better account of the utterance context and thus determine the semantically related words with higher accuracy.

This work was supported by the Russian Science Foundation, project No. 18-18-00360 "E-participation as Politics and Public Policy Dynamic Factor".

References

1. Misnikov, Y.G., Filatova, O.G., Chugunov, A.V.: Elektronnoe vzaimodejstvie vlasti i obshchestva: napravleniya i metody issledovanij. Nauchno-tekhnicheskie vedomosti SPbGPU. Gumanitarnye i obshchestvennye nauki (1), 52–60 (2016).
2. Habermas, J.: Moral consciousness and communicative action. Polity Press, Cambridge (1992).
3. Misnikov, Y.: Public Activism Online in Russia: Citizens' Participation in Webbased Interactive Political Debate in the Context of Civil Society. Development and Transition to Democracy. University of Leeds (2011).
4. Wang, P., Xu, J., Xu, B., Liu, C., Wang, H.Z.F., Hao, H.: Semantic Clustering and Convolutional Neural Network for Short Text Categorization. Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing, vol.2. pp. 352–357. Association for Computational Linguistics, Beijing (2015). http://dx.doi.org/10.3115/v1/P15-2058, last accessed 2019/12/06.
5. Socher, R., Perelygin, A., Wu, J.Y., Chuang, J., Manning, C.D., Ng, A.Y., Potts, C.: Recursive deep models for semantic compositionality over a sentiment treebank. EMNLP, 1631–1642 (2013).
6. Misnikov, Y., Chugunov, A., Filatova, O.: Converting the outcomes of citizens' discourses in cyberspace into policy inputs for more democratic and effective government. In: Paulin, A.A., Anthopoulos, L.G., Reddick, C.G. (eds). Beyond Bureaucracy: Towards Sustainable Governance Informatisation, Public Administration and Information Technology, vol. 25, pp. 259–287. Springer (2017), https://doi.org/10.1007/978-3-319-54142-6_15, last accessed 2019/12/06.
7. Misnikov, Y., Chugunov, A., Filatova, O.: Online Discourse as a Microdemocracy Tool: Towards New Discursive Epistemics for Policy Deliberation. In: 9th International Conference on Theory and Practice of Electronic Governance, ICEGOV 2016, pp. 40–49. ACM, Montevideo (2016).
8. Misnikov, Y., Chugunov, A., Filatova, O. Citizens' deliberation online as will-formation: The impact of media identity on policy discourse outcomes in Russia. In: Lecture Notes in Computer Science, vol. 9821, pp. 67–82. Springer (2016).
9. SP 42.13330.2011 Gradostroitel'stvo. Planirovka i zastrojka gorodskih i sel'skih poselenij. Aktualizirovannaya redakciya SNiP 2.07.01-89 (s popravkoj). Moscow (2011), http://docs.cntd.ru/document/1200084712, last accessed 2019/12/05.
10. Ingersoll, G.S., Morton, T.S., Ferris, E.L.: Obrabotka nestrukturirovannyh tekstov. Poisk, organizaciya i manipulirovanie. Per. s angl. Slinkin, A.A. DMK Press, Moscow (2015).
11. Barsegyan, A.A., Kupriyanov, M.S., Holod I.I., Tess M.D., Elizarov S.I.: Analiz dannyh i processov: ucheb. Posobie. 3d izd., pererab. i dop. BHV-Peterburg, St. Petersburg (2009).
12. Aggarwal, C.C.: Data Classification: Algorithms and Applications. 1st edn. Chapman & Hall/CRC (2014).
13. Zhou, C., Sun, C., Liu, Z., Lau, F.C.M.: A C-LSTM Neural Network for Text Classification. (2015), https://arxiv.org/abs/1511.08630, last accessed 2019/12/06.

14. Prasanna P.L., Rao, D.R.: Text classification using artificial neural networks. International Journal of Engineering & Technology 7 (1.1), 603–606, (2018).
15. RusVectōrēs, https://rusvectores.org/ru/about/, last accessed 2019/12/06.
16. Nacional'nyj korpus russkogo yazyka, http://ruscorpora.ru, last accessed 2019/12/06.
17. Colin, R., Ellis, D.P.W.: Feed-Forward Networks with Attention Can Solve Some Long-Term Memory Problems. (2015), https://arxiv.org/abs/1512.08756, last accessed 2019/12/05.
18. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.: Attention Is All You Need. (2017), https://arxiv.org/abs/1706.03762, last accessed 2019/12/06.
19. Mikolov, T., Chen, K., Corrado, G., Dean, J.: Efficient Estimation of Word Representations in Vector Space. (2013), https://arxiv.org/abs/1301.3781, last accessed 2019/12/06.
20. Mikolov, T., Le, Q.: Distributed Representations of Sentences and Documents. https://cs.stanford.edu/~quocle/paragraph_vector.pdf, last accessed 2019/12/06.
21. Pennington, J., Socher, R., Manning, C.D.: GloVe: Global Vectors for Word Representation. https://nlp.stanford.edu/pubs/glove.pdf, last accessed 2019/12/06.