Using Linguistic Activity in Social Networks to Predict and Interpret Dark Psychological Traits

Arseny Moskvichev[1(✉)], Marina Dubova[1], Sergey Menshov[2], and Andrey Filchenkov[2]

[1] Saint Petersburg State University, Saint Petersburg, Russia
arseny.moskvichev@gmail.com
[2] ITMO university, Saint Petersburg, Russia

Abstract. Studying the relationships between one's psychological characteristics and linguistic behaviour is a problem of a profound importance in many fields ranging from psychology to marketing, but there are very few works of this kind on Russian-speaking samples. We use Latent Dirichlet Allocation on the Facebook status updates to extract interpretable features that we then use to identify Facebook users with certain negative psychological traits (the so-called Dark Triad: narcissism, psychopathy, and Machiavellianism) and to find the themes that are most important to such individuals.

1 Introduction

The problem of linking individual characteristics and the digital records of one's behaviour has been given much attention in recent literature. Often, the primary goal is to predict individual characteristics based on the user's activity in social networks. This idea was applied to a broad range of target variables, and it was repeatedly demonstrated that it is possible to predict demographic (age, gender, sexual orientation, ethnicity) [7,13,31] and psychological characteristics (agreeableness, neuroticism, happiness) [24–26], as well as political preferences [1, 19]. Another dimension along which one can compare the works in this field is the choice of features. The most common options include user likes, geotags, and wall-posts, but sometimes more original sources of information are used, as in [10], where authors analyzed mobile device logs in order to predict user's personality.

The best predictive performance is usually achieved by combining different sources of information, as it was done, for example, in [14], where the authors improved venue recommendations by combining information from several social networks, or in [11], where the authors described an efficient substance use detection system. A similar approach was applied with considerable success in [21] and in [5] for the problem of predicting psychological variables measured using the Big Five personality model. In these works, a broad set of features was used,

A. Filchenkov et al. (Eds.): AINL 2017, CCIS 789, pp. 16–26, 2018.
https://doi.org/10.1007/978-3-319-71746-3_2

ranging from a number of photos uploaded by user to word forms extracted through linguistic analysis.

The downside of this attitude, however, is that interpretability is often sacrificed for the sake of achieving higher accuracy. Since the primary purpose of our article is to explore the relationship between certain psychological traits and language, we restrict our further analysis to the works that mostly rely on text-based features.

Among the works that utilize texts as the primary source of information, the results are most impressive for the predictions of demographic variables such as gender or age, with the achieved accuracy and R-squared metrics reaching numbers as high as 0.9 and 0.8 for gender and age respectively [27]. There are also works of this kind that focus on Russian-speaking samples, for example, predicting age based on users' wallposts [3].

At the same time, the achieved accuracy values are relatively low, when it comes to predicting psychological characteristics. For example, in one twitter-based study [29], the authors hosted an open competition on Kaggle, with the winning model achieving an AUC of 0.641 for Psychopathy (the results for other psychological traits they used were even worse). Other psychological variables could be even harder to predict, with standard methods giving accuracy values in the sub-0.6 range [2]. This might be due to the fact that the psychological variables themselves are difficult to define and measure, so there is a large amount of noise in the target variable [20].

On the other side of the research spectrum, in the fields of psychology, psychiatry, and sociology, there is a lasting effort to understand how the specific personality traits manifest themselves through behaviour and language. Such studies usually focus on the correlations between psychological traits and specific words or word categories (usually predefined), paying less attention to the predictive performance. The most commonly used predefined word categories include dictionaries like ANEW (Affective Norms fro English Words) that maps words to their emotional values and LIWC (Linguistic Inquiry and Word Count) that provides a number of "psychologically meaningful" word interpretations [22,30]. The problem with this approach is that it lacks flexibility. Not only relevant categories can emerge or disappear from the public discourse with time, it is also difficult to adapt these dictionaries to other languages, since the translations require thorough validation. Therefore, the data-driven approaches to category extraction are becoming more and more popular, and, as shown in [27], they could also lead to superior predictive performance.

In our work, we focus on the following two questions:

1. Are there specific semantic preferences related to the Dark Triad of psychological traits?
2. Can we predict individual's psychological characteristics based on the high-level semantic content of the texts they write?

For English-speaking samples, the answer is "yes", as it can be seen from [16, 27,29]. However, it is unclear, whether the same results can be achieved on the Russian segment of Facebook users. It is especially true for the second question,

since while there were studies that study the linguistic correlates of the Dark Triad of psychological traits [23] in Russian samples, the predictive performance was not investigated in that article.

2 Method

2.1 Psychometrics

In order to measure individual psychological traits constituting the psychological Dark Triad, we used the Russian version [12] of the Short Dark Triad questionnaire [18]. We chose the short version to maximize the chances of survey completion.

We also introduced three questions from the classical social desirability scale questionnaire [9] to detect cases when a participant provides dishonest answers in order to seem a "better" person according to social standards.

In addition, one "trap question" was used. It is a simple instruction of the form "please, choose the third option" that is used to check whether the participant is actually paying attention and reading the questions rather than choosing random answers.

2.2 Topic Models

In order to extract high-level topics relevant to the Russian-speaking segment of the Facebook audience, we used the Latent Dirichlet Allocation, which is one of the standard techniques for this task [4].

LDA is based on several assumptions. Each document is assumed to contain text related to several topics and relatedness to a topic is precisely described by containing words related to this topic. More formally, each document is considered to be generated in the following way: given a distribution of its topics and a distribution of words for each topic, a new word in the document is generated by choosing its topic and then choosing the word of that topic. All the choices are independent. Distributions of words and topics are assumed to be Multinomial, while distribution of their parameters is Dirichlet.

2.3 Predictive Models

We used standard classification algorithms, such as Support Vector Machine with a radial basis function kernel, Random Forest ensemble classifier and a Multinomial Naive Bayes classifier [15].

In order to obtain the binary labels from the ordinal measurements of personality, we used the median split on all available data, as it was done in [29]. It should be noted that since there are multiple posts associated with each user, there are different ways to approach this classification problem. One possibility is to train classifiers on single posts entities and to average the predictions on the test phase. In this case, the cross-validation scheme should be chosen appropriately, so as to preclude the event when the posts from one participant are

present in both training and test sets. Another option is to average the features for each participant before training the classifier.

Both options were explored and gave almost identical results. Because we use the median split, care should be taken when using the first strategy, in order to account for the slightly changing class imbalances (occurring due to the fact that different participants could have significantly different numbers of posts). Overall, the pre-averaging approach is slightly more natural in this scenario, so we only report the classification results obtained using it.

2.4 Statistical Analysis

Although the methods of statistical inference used in this article are limited to the calculation of Pearson's correlations criterion, it is important to note that in order to account for multiple hypothesis testing, we applied the Benjamini-Hochberg correction (FDR) [6]. By doing this we can ensure that the correlations that we found do indeed reflect the presence of a statistical relationship between two variables rather than being a consequence of excessive hypothesis testing.

3 Experiment

3.1 Data Collection

The data were obtained through a Facebook application that was created for this study. The participants were presented with an option to take part in the study by filling-in the psychological questionnaire and by giving access to their Facebook profile demographic information. No monetary incentives were used to attract participants, with their primary motivation being to receive the feedback on their psychological traits. In order to inform more participants about our study we ran an advertising campaign through Facebook Advertising Services.

For each participant, we collected the following data:

1. The measurements of the participant's individual psychological traits, including the measurements of the so-called psychological "Dark Triad" (Psychopathy, Narcissism, and Machiavellianism), on which we focus in this article.
2. User-generated texts, obtained from the Facebook status updates (wall-posts).
3. Demographic and other information from the user's Facebook profile. This portion of data includes age, gender, location, and likes.

3.2 Data Preprocessing

Initially, this procedure resulted in a sample of 8367 participants, with 56% of the sample being women, 41 person (0.5%) of unidentified gender, and the rest being males. The average age was 46 years, with a standard deviation of 13.46 years, 4% of participants did not provide their age.

During the initial filtering stage, we kept the participants who satisfied the following criteria:

1. They completed the questionnaire
2. They answered correctly to the "trap" question
3. The social desirability scale total is less than 13 points (15 being the maximum)
4. The number of "fast" responses (less than 5 s) is fewer than 36.

This resulted in a sample of 3341 participants. After we additionally filtered out participants with no posts containing the non-empty "message" field, we obtained the final sample with the size of 2852.

3.3 Implementation Details

In order to obtain the user-generated texts, we used the "message" field of the Facebook API post object, as it was done in other studies. Unfortunately, the manual inspection revealed a presence of posts that were automatically generated by Facebook applications and the posts containing copied materials from various sources. Since there is no simple and reliable way of sorting such posts out, and since these posts, while not being written by the user, still do reflect his or her interests and attitudes, we decided to leave them in the dataset.

We used the **word_tokenizer** function from the nltk library to separate message strings into words; we also removed the punctuation symbols and English and Russian stop words (also obtained through the nltk library) in order to make the topics more interpretable. In addition to that, we excluded all words with document frequency less than 10^{-4}.

The next step was to build the bag of words document representation. The Russian language exhibits a rich morphological structure, and in order to reduce this complexity and avoid introducing excessive amounts of variables into the document-word matrix, we extracted the normal form of each word using the pymorphy2 package before building the bag of words representation.

In order to extract topics, we used an LDA implementation from the LDA library for Python[1]. For other machine learning methods, we used the scikit-learn Python library.

Lastly, the statistical analysis was performed using the R programming language.

4 Results

4.1 Prediction

To evaluate the predictive performance of different classifiers, we used a 10-fold cross-validation scheme. Results in Table 1 summarize the algorithm predictive performances for the cases when extracted topics were used as features. It is important to note that the Random Forest classifier repeatedly outperformed all other models in all cases, therefore we only report scores obtained by this model.

[1] https://pypi.python.org/pypi/lda.

Table 1. Classification results for topic-based predictions

	Psych.	Mac.	Nar.	Gender
Baseline accuracy	0.52	0.507	0.552	0.531
Random Forest Accuracy	0.558	0.516	0.562	0.691
Random Forest AUC	0.571	0.526	0.558	0.748
Baseline accuracy H/L	0.507	0.531	0.534	-
Random Forest Accuracy H/L	0.572	0.581	0.587	-
Random Forest AUC H/L	0.591	0.576	0.612	-

To make our model comparable to a broader set of works, we also calculated the accuracy for the truncated sample. This truncated sample is obtained by throwing out the cases falling in the interval of ± one standard deviation from the mean.

It is important to note that by using the raw bag-of-words matrix (instead of 25 topics extracted using LDA), we get the accuracies that do not significantly differ from those listed in the Table 1. Moreover, other methods of dimensionality reduction (such as, for example, PCA or feature selection from the elastic net regression) result in worse prediction performance.

4.2 Statistical Analysis

We calculated the Pearson's correlation between self-reported Dark Triad scores and the estimated presence of each LDA-selected topic (averaged across all posts for each user). In order to account for multiple hypothesis testing, we applied the Benjamini-Hochberg false discovery rate correction (FDR) [6].

Machiavellianism. As it can be seen from the Table 2, we found the following patterns in topics for participants with high Machiavellianism scores:

1. Writing less about God, faith and soul. It is consistent with the idea that Machiavellianism is characterized by cynical disregard for morality [17].
2. Writing more about business and work. It is also consistent with the belief that Machiavellianism is described by concentration on self-interest [17].
3. Writing more posts with patriotic feeling: about Homeland and political situation in Russia. Appeal to patriotic feeling could be an effective method of manipulation of others (the key characteristic of Machiavellianism [17]).

Narcissism. These patterns of Facebook activity turned out to be the indicators of Narcissism:

1. Large diversity of topics among the posts.

2. Writing more posts describing friendship and social relationships. It could a way to brag about happy relationships that is largely consistent with Narcissism [8].
3. Writing more about health, body condition and illnesses. It is consistent with the most well-known characteristic of Narcissism: the concentration on oneself [8].

Psychopathy. Psychopathy is characterized by the following topics activity:

1. Writing more posts on Homeland and political situation: about Russia, Ukraine, USA, Putin, Crimea etc. It could be a form of consistent antisocial behavior (Internet terrorism) related to Psychopathy [28,32].
2. Writing more about daily activity. Small stories describing trivial mundane situations could be related to the selfishness characterizing Psychopathy [28].
3. Writing posts describing parties and celebrations.
4. Writing less about weather, season and time of day.
5. Writing more about working activity, projects, earnings and economical situation. It could also be consistent with selfishness characteristic of Psychopathy [28].

5 Discussion

Fist of all, we did not focus on optimizing the achieved accuracies at all costs (for example we avoided engineering new features and performed only a bare minimum of manual hyperparameter optimization (none for the best performing model)). The reasons to avoid extensive optimizations of this kind were as follows: the primary purpose of this article was to provide the proof of concept, and we deemed it reasonable to start with a simple baseline solution that works "from the box". The other reason is that our dataset is very small, therefore we limited the model evaluation to the cross-validation technique and we did not want to introduce the possibility of our conclusions being contaminated by the cross-validation set overfitting.

Having said that, we should first note that the obtained accuracies are lower than the state of the art predictive models applied to English-speaking segments of social networks [27,29]. At the same time, it is important to mention that the accuracies are generally low for the predictions of psychological variables, and the gap is not very big. Indeed, some studies focusing on predicting the Big Five personality traits report that their standard methods give very similar results, despite using a much larger dataset [2]. Moreover, there are very few works focusing specifically on the Dark Triad prediction, which are particularly difficult to predict, judging by the results of Kaggle competition, described in [29]. Lastly, our study replicates the pattern of differing predictive difficulty found in other articles, with Psychopathy being the most predictable among the Dark Triad psychological traits [16].

Table 2. Semantic correlates of the Dark Personality Traits, $*p < 0.05$, $**p < 0.01$, No signs: $p < 0.06$, FDR-corrected

Machiavellianism		Narcissism		Psychopathy	
Topic	Cor.	Topic	Cor.	Topic	Cor.
Faith** (holy, word, God, church, soul, Christ, faith, pray, sin)	−0.068	Diversity of topics in posts**	0.075	Patriotism** (Russia,nation, Putin, Ukraine, federation, politics, Crimea, citizen, west, USA)	0.068
Business* (money, Russia, work, rouble, company, price, business, project)	0.052	Friendship* (best, good, friend, love, attitude, true)	0.059	Daily Routine* (talk, car, go, think, money, road, phone, decide, do, see, stand, buy)	0.064
Patriotism (Russia, nation, Putin, Ukraine, federation, politics, Crimea, citizen, west, USA)	0.049	Health (water, help, body, doctor, organism, health, energy, illness, treatment)	0.051	Celebration* (celebration, congratulate, Birthday, love, health, greeting)	0.056
				Environment* (morning, summer, good, evening, Moscow, night, weather, autumn, rain)	−0.055
				Business (money, Russia, work, rouble, company, price, business, project)	0.050

There are a few potential explanations for the fact that the achieved performance metrics are not very high. The first and the most obvious is that the amounts of data that we have are smaller by an order of magnitude than the amounts data used in most cases, which may very well be a decisive factor [27]. Another possibility is that the texts that we collected contain too many copied or irrelevant material and are thus more noisy and less reliable. Lastly, there is a chance that the psychometric methods adapted to Russian are less precise in identifying psychological traits.

In order to partially answer to this question, we measured the accuracy of gender prediction (assuming that the self-reported gender is measured with equal precision in Russian and English-speaking samples). The achieved accuracy of (0.69) is very similar that achieved in another study (0.72) [33], where a relatively small dataset and similar prediction techniques were used. At the same time, the studies on larger datasets [27] usually achieve accuracies around 0.9. This observation corroborates the view that the size of the dataset might have been the primary limiting factor.

On the psychological side, we can see that by using topic modeling, we can indeed identify interpretable topics that give insightful information on the ways in which the psychological traits manifest themselves through the linguistic behaviour in social networks.

6 Conclusion

In this paper, we analyzed relationship between Russian-speaking Facebook users' texts and their psychological characteristics. We used topic modeling approach to represent user-generated texts as the mixtures of automatically generated high-level semantic categories. This model was used for two purposes corresponding to the two research questions of this paper.

Firstly, we identified specific semantic preferences related to the Dark Triad of psychological traits, including the following observations:

- Machiavellianists have a tendency to write about business-related and patriotic topics more often, while religious discourse is rare in their texts.
- Narcissistic users have a tendency to write about personal and social aspects of well-being, writing more often about wellness and social acceptance, as well as showing increased diversity in their choice of topics.
- Users with high Psychopathy scores show semantic preferences to business and patriotism topics. They are also more prone to describing the details of their daily routine and actions, while giving less attention to the properties of their surroundings like weather or the time of year.

Secondly, we have shown that it is possible to use these extracted features to predict the psychological characteristics of social network users. Although the accuracies were low in general sense, they were significantly above the chance level, which is a good result, considering the intrinsic noisiness of psychological measurements. Moreover, while not being applicable on practice for individual user profiling, these results could be applied to detect groups of people exhibiting certain negative psychological traits.

We see the main impact of this article in that we have shown that the flexible data-driven methodology previously only applied to English-speaking samples can be successfully adapted to the Russian segment of social networks in order to predict and better understand personal traits based on user-generated texts.

Acknowledgements. The authors acknowledge Saint Petersburg State University for a research grant 8.38.351.2015.

References

1. Agarwal, S., Sureka, A.: Applying social media intelligence for predicting and identifying on-line radicalization and civil unrest oriented threats. arXiv preprint arXiv:1511.06858 (2015)
2. Alam, F., Stepanov, E.A., Riccardi, G.: Personality traits recognition on social network-facebook. In: WCPR (ICWSM-13), Cambridge, MA, USA (2013)
3. Alekseev, A., Nikolenko, S.I.: Predicting the age of social network users from user-generated texts with word embeddings. In: Artificial Intelligence and Natural Language Conference (AINL), IEEE, pp. 1–11. IEEE (2016)
4. Alghamdi, R., Alfalqi, K.: A survey of topic modeling in text mining. Int. J. Adv. Comput. Sci. Appl. (IJACSA) **6**(1) (2015)
5. Bachrach, Y., Kosinski, M., Graepel, T., Kohli, P., Stillwell, D.: Personality and patterns of facebook usage. In: Proceedings of the 4th Annual ACM Web Science Conference, pp. 24–32. ACM (2012)
6. Benjamini, Y., Yekutieli, D.: The control of the false discovery rate in multiple testing under dependency. Ann. Stat. **29**, 1165–1188 (2001)
7. Buraya, K., Farseev, A., Filchenkov, A., Chua, T.-S.: Towards user personality profiling from multiple social networks. In: AAAI, pp. 4909–4910 (2017)
8. Campbell, W.K., Miller, J.D.: The handbook of narcissism and narcissistic personality disorder: theoretical approaches, empirical findings, and treatments. Wiley, Hoboken (2011)
9. Crowne, D.P., Marlowe, D.: A new scale of social desirability independent of psychopathology. J. Consult. Psychol. **24**(4), 349 (1960)
10. de Montjoye, Y.-A., Quoidbach, J., Robic, F., Pentland, A.S.: Predicting personality using novel mobile phone-based metrics. In: Greenberg, A.M., Kennedy, W.G., Bos, N.D. (eds.) SBP 2013. LNCS, vol. 7812, pp. 48–55. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-37210-0_6
11. Ding, T., Bickel, W.K., Pan, S.: Social media-based substance use prediction. arXiv preprint arXiv:1705.05633 (2017)
12. Egorova, M., Sitnikova, M.: Parshikova ov adaptatsiia korotkogo oprosnika temnoi triady [adaptation of the short dark triad]. Psikhologicheskie issledovaniia **8**(43), 1 (2015)
13. Farseev, A., Nie, L., Akbari, M., Chua, T.-S.: Harvesting multiple sources for user profile learning: a big data study. In: Proceedings of the 5th ACM on International Conference on Multimedia Retrieval, pp. 235–242. ACM (2015)
14. Farseev, A., Samborskii, I., Chua, T.-S.: bbridge: A big data platform for social multimedia analytics. In: Proceedings of the 2016 ACM on Multimedia Conference, pp. 759–761. ACM (2016)
15. Friedman, J., Hastie, T., Tibshirani, R.: The elements of statistical learning. Springer series in statistics, vol. 1. Springer, Berlin (2001)
16. Garcia, D., Sikström, S.: The dark side of facebook: Semantic representations of status updates predict the dark triad of personality. Pers. Individ. Differ. **67**, 92–96 (2014)
17. Jakobwitz, S., Egan, V.: The dark triad and normal personality traits. Pers. Individ. Differ. **40**(2), 331–339 (2006)
18. Jones, D.N., Paulhus, D.L.: Introducing the short dark triad (sd3) a brief measure of dark personality traits. Assessment **21**(1), 28–41 (2014)
19. Kosinski, M., Stillwell, D., Graepel, T.: Private traits and attributes are predictable from digital records of human behavior. Proc. Natl. Acad. Sci. **110**(15), 5802–5805 (2013)

20. Lambiotte, R., Kosinski, M.: Tracking the digital footprints of personality. Proc. IEEE **102**(12), 1934–1939 (2014)
21. Markovikj, D., Gievska, S., Kosinski, M., Stillwell, D.J.: Mining facebook data for predictive personality modeling. In: Seventh International AAAI Conference on Weblogs and Social Media (2013)
22. Nielsen, F.Å.: A new anew: evaluation of a word list for sentiment analysis in microblogs. arXiv preprint arXiv:1103.2903 (2011)
23. Panicheva, P., Ledovaya, Y., Bogolyubova, O.: Lexical, morphological and semantic correlates of the dark triad personality traits in russian facebook texts. In: Artificial Intelligence and Natural Language Conference (AINL), IEEE, pp. 1–8. IEEE (2016)
24. Peng, Z., Hu, Q., Dang, J.: Multi-kernel svm based depression recognition using social media data. Int. J. Mach. Learn. Cybern. 1–15 (2017)
25. Preotiuc-Pietro, D., Carpenter, J., Giorgi, S., Ungar, L.: Studying the dark triad of personality through twitter behavior. In: Proceedings of the 25th ACM International on Conference on Information and Knowledge Management, pp. 761–770. ACM (2016)
26. Preoţiuc-Pietro, D., Carpenter, J., Giorgi, S., Ungar, L.: Studying the dark triad of personality using twitter behavior (2016)
27. Schwartz, H.A., Eichstaedt, J.C., Kern, M.L., Dziurzynski, L., Ramones, S.M., Agrawal, M., Shah, A., Kosinski, M., Stillwell, D., Seligman, M.E., et al.: Personality, gender, and age in the language of social media: the open-vocabulary approach. PloS One **8**(9), e73791 (2013)
28. Skeem, J.L., Polaschek, D.L., Patrick, C.J., Lilienfeld, S.O.: Psychopathic personality: bridging the gap between scientific evidence and public policy. Psychol. Sci. Public Interest **12**(3), 95–162 (2011)
29. Sumner, C., Byers, A., Boochever, R., Park, G.J.: Predicting dark triad personality traits from twitter usage and a linguistic analysis of tweets. In: 11th International Conference on Machine Learning and Applications (ICMLA), 2012, vol. 2, pp. 386–393. IEEE (2012)
30. Tausczik, Y.R., Pennebaker, J.W.: The psychological meaning of words: Liwc and computerized text analysis methods. J. Lang. Soc. Psychol. **29**(1), 24–54 (2010)
31. Wang, P., Guo, J., Lan, Y., Xu, J., Cheng, X.: Multi-task representation learning for demographic prediction. In: Ferro, N., Crestani, F., Moens, M.-F., Mothe, J., Silvestri, F., Di Nunzio, G.M., Hauff, C., Silvello, G. (eds.) ECIR 2016. LNCS, vol. 9626, pp. 88–99. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-30671-1_7
32. Williams, K., McAndrew, A., Learn, T., Harms, P., Paulhus, D.L.: The dark triad returns: entertainment preferences and antisocial behavior among narcissists, machiavellians, and psychopaths. In: Poster presented at the 109th Annual Convention of the American Psychological Association, San Francisco, CA (2001)
33. Zhang, C., Zhang, P.: Predicting gender from blog posts. University of Massachusetts Amherst, USA (2010)

http://www.springer.com/978-3-319-71745-6

Artificial Intelligence and Natural Language
6th Conference, AINL 2017, St. Petersburg, Russia,
September 20–23, 2017, Revised Selected Papers
Filchenkov, A.; Pivovarova, L.; Žižka, J. (Eds.)
2018, XI, 305 p. 39 illus., Softcover
ISBN: 978-3-319-71745-6