

# Precedent-based approach for the identification of deviant behavior in social media

Anna V. Kalyuzhnaya, Nikolay O. Nikitin, Nikolay Butakov, Denis Nasonov  
ITMO University, 49 Kronverksky Pr. St. Petersburg, 197101, Russian Federation  
kalyuzhnaya.ann@gmail.com

**Abstract.** The current paper is devoted to a problem of deviant users' identification in social media. For this purpose, each user of social media source should be described through a profile that aggregates open information about him/her within the special structure. Aggregated user profiles are formally described in terms of multivariate random process. The special emphasis in the paper is made on methods for identifying of users with certain on a base of few precedents and control the quality of search results. Experimental study shows the implementation of described methods for the case of commercial usage of the personal account in social media.

**Keywords:** deviant user, social media, behavior pattern, precedent-based search

## 1 Introduction

Nowadays increasing the popularity of social media can be observed. As a result, more and more people have their digital reflection on the Internet. This situation gives new opportunities for analysis of the real-life behavior of person using his digital trace, on the one hand. On the other hand, it creates new behavioral patterns that exist only on the Internet and differs from real-life patterns. Investigation of behavioral patterns of social media users put before us the first task to divide normal (or typical) and deviant behavior, and more than this to identify different types of deviations.

In a frame of this research, we define deviant behavior as behavior that differs from some categorical or quantitative norm. And, therefore, we can distinguish deviations in a broad sense and semantic-normative deviations (in a narrow sense). Deviation in a broad sense (or statistical deviation) can be defined on the basis of behavioral factors that represent a statistical minority of the population. The semantically-normative deviation is defined on the basis of the postulated norm, which explicitly or implicitly divides behavior into conformal (corresponding to expectations) and deviant (not satisfying expectations). Unlike statistical deviation, this form of deviation operates categorical variables. The definition of semantically-normative deviation is based both on the observed behavioral patterns and on the tasks of a specific subject area for which an analysis of the behavior of social media users is conducted. Other words, we can identify the user with semantically-normative deviation if we know how looks like such user and what is the target population for this, or we can find precedents (examples of deviant profile). The main research question of this paper is how to identify users with

certain semantically-normative deviation on a base of few precedents. This question is highly relevant because normally identification of similar objects requires large labeled samples for search algorithms training. Although, labeling of social media data is an enormously time / human resources consuming procedure. And whether we can find a solution for identification of profiles similar to few labeled ones, the secondary question is how we can control the quality of search results.

## 2 Related work

The area of user behavior profiling studied in several works in different domains. The most widespread approaches are either profiling based on analysis of user posts and comments texts or based on statistical aggregating different types of activities. Text analysis approaches [1][2][3] represented mostly by generative stochastic models based on LDA and by recurrent neural networks while the second class of approaches [4][5][6] use vectors built on aggregated users' features to find outliers using supervised and unsupervised methods. Aggregated features may include topological, temporal, and other user's behavior characteristics, but all of them represented in profile as aggregated value or set of values. The advantage of this class is a generalized representation of user's features and activities. It should be noticed that mentioned works use user's profiles for specific goals in certain domains. E.g., works [1][2][5] use users profiles for cyber-bullying detection and employ only those features that may help in this task. That reduces the ability of such approaches to be used in other areas.

Discussed works also use supervised and unsupervised methods for outlier detection. Supervised methods need labeled data, unsupervised, in their own turn, needs results interpretation after the moment when separated clusters are found and outliers detected, which may be uninterpretable. In contrary, the suggested approach is intended to be used as semi-supervised, which means that for training we use both: labeled samples and information about structural differences in the users' profiles, represented as features vectors. Despite the variety of existing approaches like [8], there are no methods of unification for approaches and algorithms that can be used to different forms of deviant behavior and different tasks of detection.

## 3 The identification of deviant users in social media

### 3.1 Aggregated social media profile model and its components

From formal point of view, the task of designing a model of behavioral aggregated user profiles can be designed in terms of a multidimensional random function.

$\mathbf{X} = \{\mathbf{X1}, \dots, \mathbf{Xn}\}$  –  $n$ -dimensional random process, defined in the  $n$ -dimensional space of attributes (features) of the aggregated user profile, for which the mathematical model is formalized. For a random process  $\mathbf{X}$  there is a family of  $n$ -dimensional distribution functions

$$\{F_t(X, t)\} = \{F_{t1}[(X1, \dots, Xn)_{t1}], \dots, F_{tM}[(X1, \dots, Xn)_{tM}]\} \quad (1)$$

that are defined on quasi-stationary intervals  $t = t1 \dots tM$ . Within the quasi-stationary interval, the realization of the n-dimensional random process  $X$  can be considered as an n-dimensional random variable (regardless of time). Then, the time aggregated profile can also be represented as a sequence of profiles on quasi-stationary intervals. On the basis of the general probabilistic model can be constructed aggregated n-dimensional profile.

For correct building an aggregated behavior profile it is necessary to look on the problem from two different sides: how user's state drives the user to leave traces in social media and how these traces can be used to restore user's state. For clarity let introduce the following definitions. User behavior profile is an aggregation of events in user's trace in a way that can be used to (a) characterize user's main aspects of behavior; (b) make users comparable and distinguishable. An event is an elementary action performed by the user in social network or media. Trace is a set of events generated by a certain user for a defined interval of time. The behavior of users in social networks is a reflection of activities and processes taking place in the real world. The user interacts with social network by creating posts on important for his topics, commenting existing posts and discussing different things with other users – e.g., the user generates events. These events are combined into an explicit digital trace of implicit internal state of the user which stays behind each event.

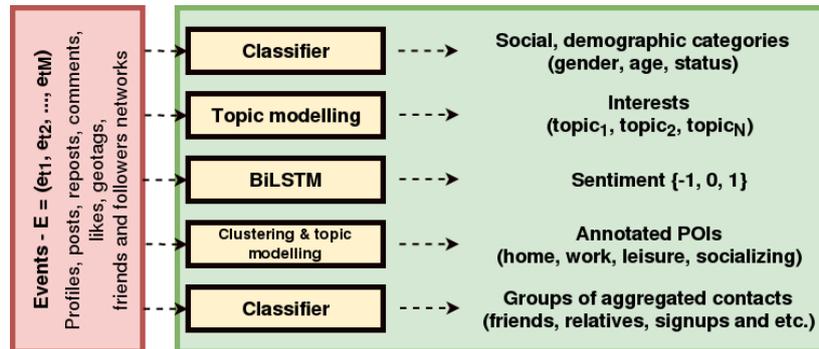


Fig 1. Aggregation of main aspects of user behavior profile

User behavior is conditioned by two main groups of factors internal (or individual, which is specific for a concrete user) and external (or social, which depend on how user's relation with external for him real world). External factors can be represented as an environment where the user lives, including cultural, social, political and other contexts. This environment influences user's activities as in the real world as in social networks. These factors can be seen as latent variables having specific values for individual users. To estimate values of the hidden factors mentioned above, available digital traces have to be processed and aggregated into components. These components represent user behavior too but as a result of available observations. Components can be organized in four main aspects of the way they are being aggregated (static information, sentiment-semantic, topological and geo-temporal, see Fig 1), starting

with data collecting from social networks. User behavior can change over time and may be addressed by aggregating user's events only for time intervals of a certain length with overlapping to catch his evolution and development trends. But this topic is out of the scope for the current work.

### 3.2 Precedent-based algorithm for deviant users identification

To identify deviant users, two methods have been developed, applied depending on the setting of a specific task. The first method was developed to identify deviants at the population level (unsupervised). This method is applicable to the task of searching for deviant users and subpopulations, considering the unknown form of their deviation in a given population. The first method directly follows from the descriptive model of ordinary users.

More interesting and complicated is the second method that is designed for the identification of profiles with specific deviation, described as a range of aggregated profile features. Since the concrete features values, associated with deviation, are not always known, the approach based on initial set of expertly-confirmed deviant profile examples can be more suitable.

The small number of confirmed deviants is a common problem for this task. A manual search of deviant profiles in a social network is quite difficult, so identification process can be started from 2-3 confirmed profiles. For this reason, a stage of preliminary semiautomatic profile set extension can be added. The main idea is to find the subset of profiles feature space, where the known deviants are similar to each other and differ from non-deviant profiles.

The search for an optimal subset of profile features is a complex task. When we have 48 features, and the problem is to found a subset with unknown size with the best quality of deviations identification, the number of combinations to check will be  $2^{48}$ , that is hard to compute. Therefore, the evolutionary approach based on kofnGA algorithm[7] was applied to reduce the time of optimization, with the size of the population set to 30 and mutation probability is 0.3. Since the algorithm deals with the fixed size of the subset, it's performing for every variant of size is needed. Then, comparison of obtained quality metrics was performed. The additional penalty was added to subspace when deviants or normal profiles are indistinguishable (to avoid the trivial cases with space with insignificant dimensions)

To identify the additional deviants, that are similar to already known, we develop the iterative algorithm, based on the several k-nearest neighbors classifiers. It's presented in Fig 2.

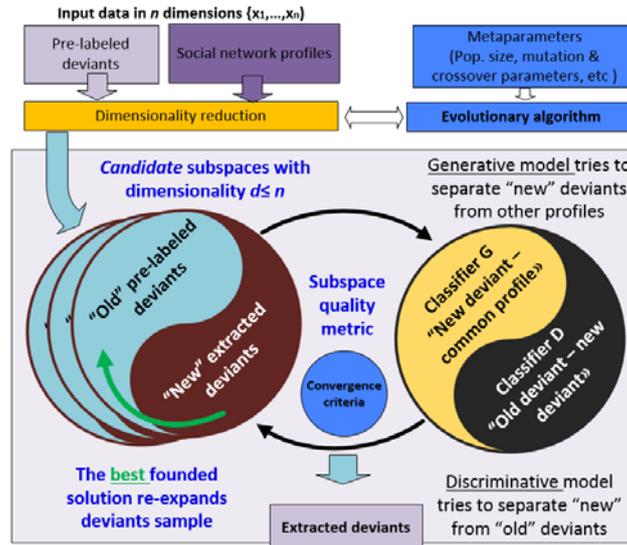


Fig. 2. Algorithm for deviant's identification based on expertly confirmed profiles

The generative classifier G, which trained on the manually pre-labeled deviant profiles subsample (“old” deviants), that mixed with random sample the unlabeled profiles (to solve the problem of only positive markup [9]), separates the whole set of profiles as deviant and non-deviant, in order to found the group of “new” deviants. For this model, the evolutionary algorithm tries to maximize the quality of classification.

Otherwise, a discriminative model D tries to separate expertly-confirmed deviant profiles and profiles that are recognized as deviant by the model G. Evolutionary algorithm tries to minimize the quality of classification, that means that the automatically extracted “new” deviant profiles must be indistinguishable from pre-approved “old” deviant profiles. The quality of every classifier is measured as AUC (the area under the receiver operating characteristic curve – ROC, that describes the effectiveness of a binary classifier). While the approach is iterative, the obtained deviants for every cycle can be used to extend the base deviant set and then repeat the classification if needed, that allow controlling the expected quality and size of result sample.

## 4 Experimental study

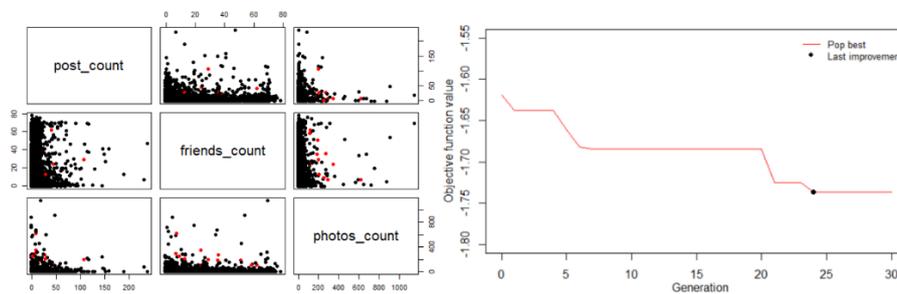
To provide the experiments with profiling and user class matching, we parse the data from public user pages of VK social network. The obtained dataset contains behavioral features for 48K user accounts aggregated for an entire lifetime.

The case study of profiling is devoted to the analysis of social network profiles, that are intended for commercial purposes – for example, as an online clothing store. First of all, we manually labeled several examples of such pages. Then previously-described iterative algorithm for profile set extension was applied to this task. The initial set of

commercial accounts contains 17 entities. The example of commercial accounts localization in some 3-dimensional space is presented in fig. 3a.

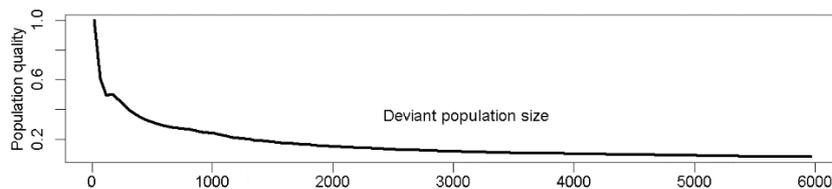
To verify the results of deviants' identification, the dictionary of domain-specific key phrases was created. Then, it was used to check the correctness of automated deviants markup and measure the population quality – the ratio of correctly recognized deviants to total deviant sample size (that can be changed from 0 to 1).

The optimization problem is to determine the optimal subset of behavioral features, that allows finding the accounts with similar deviation. The experiment was conducted with the dimensionality of space is varied from 5 to 20. The convergence of evolutionary optimization algorithm was achieved in 24 epochs (Fig. 3b).



**Fig. 3.** a) The subspace of accounts sample (red dots in commercial accounts) b) The convergence of evolutionary algorithm of optimal dimensions' search

The founded optimal space contains 10 dimensions (statistics of posts, friends, followers, photos, likes, comments, emotions, etc.). Then, since the quality of extracted accounts depends on the expected size of the sample, the sensitivity analysis of deviant population was provided. When the initial expertly-confirmed sample is 100% correct, after the first iteration only 60% of accounts founded are really online shops, and the other are fans of sales, photographers and artists, active travelers. The further expansion of the sample tends to lower quality. The dependency of the quality metric from deviant sample size is presented in fig. 4.



**Fig. 4.** Quality of extended population

In this case, the quality decreases rapidly, that can be explained by the significant similarity of commercial accounts with some other.

## 5 Conclusion and future works

Identification of deviant users in social media is highly relevant task nowadays. Relevance can be explained by increasing necessity for behavioral analysis and control in cyberspace, from one point of view, and the possibility to explore additional open information about the real person to analyze his/her integrity, reliability or compliance with some standards. This paper discusses an approach for precedent-based identification of deviant users in a targeted population. Such approach allows searching for profiles of users similar to few examples which can be easily found manually. Also, it was shown how the quality of searched look-alike population is related to the population size.

Current results make possible next research step – prediction of future deviant behavior. This ambitious task can be subdivided into two research directions: prediction of the next (the nearest) deviant or even delinquent action of the deviant user (on the level of population), and the probabilistic forecast with different horizons of deviant changes in user behavior.

## 6 Acknowledgments

This research financially supported by Ministry of Education and Science of the Russian Federation, Agreement #14.578.21.0196 (03.10.2016). Unique Identification RFMEFI57816X0196.

## References

1. Raisi E., Huang B. Cyberbullying identification using participant-vocabulary consistency. 2016. P. 46–50.
2. Sax S. Flame Wars: Automatic Insult Detection.
3. Wang Y. et al. Topic-Level Influencers Identification in the Microblog Sphere. 2016. Vol. 0. P. 4–5.
4. Angeletou S., Rowe M., Alani H. Modelling and analysis of user behaviour in online communities // Semant. Web–ISWC 2011. 2011. P. 35–50.
5. Dadvar, M., Ordelman, R., de Jong, F. & Trieschnigg D. Towards User Modelling in the Combat against Cyberbullying // Lect. Notes Comput. Sci. 2012. Vol. 7337/2012. P. 277–283.
6. Galal A., Elkorany A. Dynamic Modeling of Twitter Users Dynamic Modeling of Twitter Users // Proc. 17th Int. Conf. Enterp. Inf. Syst. 2015. Vol. 2. P. 585–593.
7. Wolters, M. A. (2015). A genetic algorithm for selection of fixed-size subsets with application to design problems. *J Stat Softw*, 68(1), 1-18.
8. Ma Q. et al. A sub-linear, massive-scale look-alike audience extension system //Proceedings of the 5th International Workshop on Big Data, Streams and Heterogeneous Source Mining: Algorithms, Systems, Programming Models and Applications. – 2016.
9. Elkan C., Noto K. Learning classifiers from only positive and unlabeled data //Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining. – ACM, 2008. – C. 213-220.