MODELLING OF AN EDUCATIONAL PROFILE OF A STUDENT BY ANALYZING PUBLIC USER DATA FROM SOCIAL NETWORKS

Conference Paper · March 2018

DOI: 10.21125/inted.2018.1115

5 authors, including:

Artem Feshchenko
Tomsk State University
27 PUBLICATIONS **28** CITATIONS
SEE PROFILE

Vyacheslav Goiko
Tomsk State University
9 PUBLICATIONS **5** CITATIONS
SEE PROFILE

Valeriia Matsuta
Tomsk State University
5 PUBLICATIONS **0** CITATIONS
SEE PROFILE

Andrey Stepanenko
Tomsk State University
4 PUBLICATIONS **2** CITATIONS
SEE PROFILE

MODELLING OF AN EDUCATIONAL PROFILE OF A STUDENT BY ANALYZING PUBLIC USER DATA FROM SOCIAL NETWORKS

A. Feshchenko[1], V. Goiko[1], V. Matsuta[1], A. Stepanenko[1], P. Kiselev[2]

[1] National Research Tomsk State University (RUSSIAN FEDERATION)
[2] Psychological Institute of the Russian Academy of Education (RUSSIAN FEDERATION)

Abstract

The present paper discusses the prospects of using the VK (VKontakte) social network for identifying psychological traits, interests and professional hobbies, that are important for spotting gifted senior high school students.

Methods: psychological testing and questionnaire survey («#Careerguidance» method), academic record analysis, structural and content analysis of a social network profile, content analysis, percentile normalization, machine learning.

Using a sample of 1692 senior high school students giftedness was defined as a combination of intelligence (analogies, convergent and divergent thinking styles), creativity (fluency, flexibility, originality, elaboration, independence), conative and personal features (conation for knowledge, self-actuating, proactivity, determination, resoluteness, social intelligence).

The paper further presents «psychological profiles» of gifted senior high school students based on the analysis of their VK profile data. Female users mostly join communities with educational (informative), social and commercial content, male users mostly with entertaining and educational (informative) content. Female users mostly post and share in order to inform and entertain friends and subscribers, to encourage action. Male users do the same in order to inform and entertain friends and subscribers.

Machine learning proved to be useful in researches with a large number of participants and attributes. To solve the problem a binary classification was used. So, gifted and non-gifted students were compared. Support vector machine proved to be the most efficient model to solve this problem. It allows to identify participants with highly developed psychological features, calculate correlations between participants with different degrees of giftedness features and VK communities. Predictive modeling is performed based on regional and federal communities as markers. A part of marker communities is targeted on the participants' gender.

The novelty of the research lies in the fact that social network user data is examined specifically in order to identify giftedness.

Practical application: recruitment of prospective students with significant cognitive and personal potential; express diagnostics of cognitive, personal features and soft skills of students in order to develop individual educational paths.

Keywords: giftedness, social networks, senior high school students, psychological testing and questionnaire survey, profile analysis, big data analysis, machine learning.

1 INTRODUCTION

Social networks generate massive information flows, which show high levels of dynamics and scale. Channel capacities of social networks are steadily growing, along with that grows data transfer speed, therefore grows the speed of data processing.

Information collected from social networks using state-of-the-art big data analysis technologies might look chaotic at first sight, but can be classified based on a big number of criteria, including those targeted on specific user groups, as well as personalized ones.

These criteria are formulated in accordance with social and economic objectives of both national and regional levels. Thus, in an effort to improve quality of education and expand education market, regions are aiming at providing a high level of competency development of prospective students and reaching better admission statistics of regional universities, regions are interested in gifted senior high school students.

Aim of the research is to evaluate how social networks can be used to determine which psychological traits and interests of senior high school students are important for identifying giftedness. Evaluation was made based on the analysis of the psychological testing data and the data taken from user profiles in VK social network («digital footprints» of the users).

2 METHODOLOGY

Methodological approaches used are based on modern scientific understandings of giftedness that are enunciated in works of both Russian and foreign researchers: the conceptions of giftedness by Babaeva, Bogoyavlenskaya [9], Leites [8], Matyushkin [7], Melik-Pashaev[15], Ushakov [12], Kholodnaya [14], Shadrikov [9], Shumakova [17], Shcheblanova [18], Yurkevich [20] et al., the multidimensional giftedness models by Renzulli [10, 38], Heller et al [13, 29].

According to these understandings, giftedness is defined as an interaction between intelligence, creativity and motivational and personal characteristics. A person is considered gifted if these qualities are highly developed or if a person is able to develop and utilize them in any potentially valuable field of work [18].

Total sample consisted of 1692 people, 969 female and 723 male senior high school students from Tomsk. Research methods. Data collection: psychological testing, academic record analysis, «manual» analysis of the structure and contents of a social network profile. Data processing and analysis methods: percentile (non-linear) normalization, machine learning (binary classification based on: support vector machine, random forest, gradient boosting), content analysis.

3 RESULTS

At the first stage data was collected utilizing the psychological testing method traditionally used for giftedness diagnostics. Computerized «#Careerguidance» method was also used, which is often applicable for students of various types of senior high schools [1].

Apart from the main purpose, to identify professional orientation in general, this method can be utilized to identify the intellectual potential of a participant, to study their personality (emotional, volitional, communicative powers), professional interests and values, career orientation.

«Analogies» and «Cognitive styles» subtests were used to study intellectual aspect of giftedness. «Creativity self-esteem» and «Creative behavior questionnaire» subtests were used to study creative aspect of giftedness. «Volitional powers», «Social intelligence», «Career choice motives» subtests were used to study motivational and personal aspect of giftedness.

To study academic performance of the participants, their school record for the last semester was analyzed. Russian language, literature, history, social studies, foreign language for humanitarian, algebra, geometry, physics, computer science for physico-mathematical and biology, chemistry, geography for natural sciences subjects.

At the second stage of the research collected data was processed utilizing percentile (non-linear) normalization, which made possible to cluster participants into groups according to discovered levels of intelligence, creativity and motivational and personal traits development.

At the third stage of the research data analysis and its interpretation were done for a group of participants, whose results were in the upper quartile (75-100%).

At the fourth stage of the research gifted students were identified on VK. The following groups of data were analyzed:

1 Profile: surname, name and pseudonym, contacts, education and career information, life philosophy, number of friends.

2 User posts and shares.

3 Followed groups and pages.

Based on the results of the «manual» analysis of the structure and contents of participants social network profiles, it is possible to mark some psychological features typical for gifted senior high school students. However, such an analysis has limitations which make it difficult to determine features indicating giftedness in a «digital footprint» of a user:

- specific nature of entertainment content dominating VK, while «useful» activity is almost not present;
- specific nature of graphic content dominating VK - in this case, psycholinguistic methods are not applicable, as images replace user text;
- the number of attributes for «manual» analysis is too large - some users may have around 200 groups and pages subscriptions, thus adding texts, connections, reactions, etc., to the analysis can accumulate up the number of attributes to several hundreds.

For this reason, at the fifth stage of the research, machine learning method was used in order to identify gifted senior high school students, allowing to conduct a large-scale research regardless of a number of participants or attributes.

Similarly to [30, 37], the main task was a binary classification of the participants. Positive class includes participants who had above average results for a corresponding psychological trait (dependent variable equals «1»), negative class includes participants who had below average results for a corresponding psychological trait (dependent variable equals «0»). Therefore, predictions were based on matching the participant with one of the two classes: above or below average results.

In order to solve a binary classification task a standard set of metrics was used: precision, F-measure and area under the ROC curve.

Attributes included participant's gender, VK groups, tests results. The hypothesis was that members of a certain group would have similar psychological features, that would make it possible to consider group membership for predicting psychological features of a group's members. Previous research has shown [21] that user membership on Facebook or Twitter communities can be used to predict psychological features of a community's members. Group membership, same as text data, is an example of «digital footprint».

Main methods included support vector machine, random forest and gradient boosting. These were implemented through Python libraries.

It is worth noting that in the present research prediction model was built for participants with characteristics known beforehand:

1 Senior high school students (n=1240, including 698 females and 542 males). This limitation allowed to simplify modelling process, avoid taking into consideration age specifics. Moreover, VK is known to be the most popular social network among senior high school students.

2 Students from the Tomsk region. This limitation allowed to improve modelling quality due to taking regional specifics into consideration.

Depending on a certain psychological feature, this strategy allowed to make predictions for 50-76% of all participants. For the other 24-50% the amount of data was not enough to make predictions based on psychological features.

Comparing the models by calculating the area under the ROC curve showed that support vector machine was the most effective for the binary classification task, as the area under the ROC curve value was the highest (table 1).

Cross-validation was used, while calculating the area under the ROC curve. The sample was divided into five equal parts, four parts were used for training, one part was used for testing. Data sample was divided five times. Results shown in table 1 are averaged after five iterations of training and testing of the model.

Table 1. *Machine learning models comparison by the area under the ROC curve*

Subtest scale	Support vector machine	Random forest	Gradient boosting
Analogies	0,78	0,72	0,76
Divergent style	0,64	0,6	0,62
Convergent style	0,70	0,63	0,67
Fluency	0,67	0,61	0,65
Semantic flexibility	0,69	0,64	0,69
Originality	0,71	0,64	0,69
Creative behavior	0,71	0,62	0,68
Proactivity	0,69	0,65	0,7
Resoluteness	0,71	0,67	0,7
Determination	0,68	0,6	0,65
Self-actuating motive	0,76	0,68	0,72
Social intelligence	0,74	0,69	0,73

Additionally precision and F-measure were calculated for a model based on support vector machine (table 2).

Table 2. *Metrics for the model based on support vector machine*

Subtest scale	Support vector machine	Random forest	Gradient boosting
Analogies	0,75	0,53	0,78
Divergent style	0,64	0,41	0,64
Convergent style	0,66	0,59	0,7
Fluency	0,62	0,37	0,67
Semantic flexibility	0,67	0,57	0,69
Originality	0,67	0,57	0,71
Creative behavior	0,73	0,39	0,71
Proactivity	0,67	0,59	0,69
Resoluteness	0,67	0,64	0,71
Determination	0,63	0,65	0,68
Self-actuating motive	0,65	0,76	0,76
Social intelligence	0,65	0,76	0,74

Therefore, average precision of the prediction model was 0,67. This allowed to confirm that it is indeed the exact level of precision machine learning techniques provide for identifying participants with high level of psychological qualities, and discovering positive or negative correlations between the level of development of participants and their group membership in social networks.

Therefore, it is possible to say that machine learning techniques allow to identify users, who have highly developed psychological features which are important for identifying giftedness with satisfactory accuracy.

Using social networks data does not require a prior psychological testing, predictions can be made for users who were not surveyed to identify their giftedness. In other words, prediction model developed in the present research can be applied for identifying giftedness of any senior high school student in

4 CONCLUSIONS

Social networks can be fairly effective for identifying psychological features – intellectual, creative and motivational and personal aspects of giftedness. Utilizing machine learning methods, that allow to analyze large numbers of participants and attributes, provides an opportunity to identify social network users who have highly developed psychological features with satisfactory accuracy and to discover correlations between the development level and their group membership in social networks.

Novelty of the research lies in the fact that social network user data is examined specifically in order to identify giftedness: works of both Russian and foreign researchers show the results of social network data analysis for predicting gender, age, nationality, psychological type, political views, but not giftedness.

Research perspectives are connected with the opportunity to apply machine learning methods to perform large-scale, quick, longitude and less time consuming testing of senior high school students, compared to standard paper testing, in order to identify qualities and skills important for a successful professional fulfillment and to monitor student development.

As a direction for future research, we plan to increase the number of examined attributes: apart from communities membership, user reaction (likes and shares count), social graphs and user posts can be utilized. It would allow to sophisticate prediction model and increase its precision.

ACKNOWLEDGEMENTS

The reported study was funded by RFBR according to the research project №17-16-70004 “Research of social networks’ potential for identification, attraction and retention of talented youth at regional higher educational institutions based on big data analytics”.

REFERENCES

[1] Bogdanovskaya I. M. Komp'yuternaya psikhodiagnostika v proforientatsionnoi rabote so starsheklassnikami / I. M. Bogdanovskaya, P. B. Kiselev, A. N. Kosheleva, V. A. Ruban // Psikhologicheskie problemy obrazovaniya i vospitaniya v sovremennoi Rossii: materialy IV konferentsii psikhologov obrazovaniya Sibiri. – Irkutsk, 2016. – S. 296-302.

[2] Leont'ev D. A. Lichnostnoe v lichnosti: lichnostnyi potentsial kak osnova samodeterminatsii // Uchenye zapiski kafedry obshchei psikhologii MGU im. M. V. Lomonosova. – Vyp. 1. – M. : Smysl, 2002. – S. 56-65.

[3] Leont'ev D. A. Lichnost' kak opredelyayushchii faktor // Voobrazhenie i tvorchestvo v obrazovanii i professional'noi deyatel'nosti: materialy Chetvertoi mezhdunarodnoi konferentsii pamyati L. S. Vygotskogo. – M. : RGGU, 2004. – S. 214-223.

[4] Litvinova T. A. K probleme stabil'nosti kharakteristik idiostilya // Izvestiya Yuzhnogo federal'nogo universiteta. Filologicheskie nauki. – 2015. – №. 3. – S. 98-106.

[5] Mal'tsev V. Ne vo frendakh schast'e [Elektronnyi resurs] // Expert Online. 2012. URL: http://expert.ru/2012/09/12/ne-vo-frendah-schaste (data obrashcheniya: 04.12.2017).

[6] Maslou A. Motivatsiya i lichnost' / A. Maslou. – SPb. : Piter, 2011. – 352 s.

[7] Matyushkin A. M. Myshlenie, obuchenie, tvorchestvo / A. M. Matyushkin. – M. : Izd-vo MPSI; Voronezh : NPO «MODEK», 2003. – 720 s.

[8] Psikhologiya odarennosti detei i podrostkov / pod red. N. S. Leitesa. – M : Izdatel'skii tsentr «Akademiya», 1996. – 416 s.

[9] Rabochaya kontseptsiya odarennosti / D. B. Bogoyavlenskaya [i dr.]. – M. : Min-vo obrazovaniya RF, 2003. – 95 s.

[10] Renzulli Dzh. Model' obogashchayushchego shkol'nogo obucheniya / Dzh. Renzulli, S. M. Ris // Osnovnye sovremennye kontseptsii tvorchestva i odarennosti. – M., 1997. – S. 214-242.

[11] Sternberg R. D. Triarkhicheskaya teoriya intellekta // Inostrannaya psikhologiya. – 1996. – № 6. – S. 54-61.

[12] Ushakov D. V. Struktura i dinamika intellektual'nykh sposobnostei : avtoref. dis. ... d-ra psikhol. nauk / D. V. Ushakov. – M., 2004. – 48 s.

[13] Kheller K. A. Diagnostika i razvitie odarennykh detei i podrostkov // Osnovnye sovremennye kontseptsii tvorchestva i odarennosti. – M., 1997. – S. 243-264.

[14] Kholodnaya M. A. Psikhologiya intellekta: paradoksy issledovaniya / M. A. Kholodnaya. – SPb. : Piter, 2002. – 272 s.

[15] Khudozhestvennaya odarennost' detei, ee vyyavlenie i razvitie / A. A. Melik-Pashaev [i dr.]. – Dubna : Feniks+, 2006. – 112 s.

[16] Shmelev A. G. Psikhodiagnostika lichnostnykh chert / A. G. Shmelev. – SPb. : Rech', 2002. – 480 s.

[17] Shumakova N. B. Obuchenie i razvitie odarennykh detei / N. B. Shumakova. – M. : Izd-vo MPSI; Voronezh : NPO «MODEK», 2004 a. – 336 s.

[18] Shcheblanova E. I. Odarennost' kak psikhologicheskaya sistema: struktura i dinamika v shkol'nom vozraste : dis. ... d-ra psikhol. nauk / E. I. Shcheblanova. – M., 2006. – 311 s.

[19] Yung K. G. Vospominaniya, razmyshleniya, snovideniya / K. G. Yung. – Minsk : OOO «Kharvest», 2003. – 496 s.

[20] Yurkevich V. S. Odarennyi rebenok: illyuzii i real'nost' / V. S. Yurkevich. – M. : Prosveshchenie; Uchebnaya literatura, 1996. – 136 s.

[21] A Societal Sentiment Analysis: Predicting the Values and Ethics of Individuals by Analysing Social Media Content / T. Maheshwari et al. // Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics. – 2017. – Vol. 1. – P. 731-741.

[22] Clark B. Growing up gifted / B. Clark. – New York : Macmillan, 1992. – 674 p.

[23] Cropley A. J. Recognizing creative potential: an evaluation of the usefulness of creativity tests // High ability studies. The journal of the European Council for High Ability. – 1996. – Vol. 7. – № 2. – P. 203-219.

[24] Dunbar R. I. M. How many friends does one person need?: Dunbar's number and other evolutionary quirks. – London : Faber and Faber, 2010. – 312 p.

[25] Feldhusen J. F. A conception of giftedness. // Conceptions of giftedness. – Cambridge : Cambridge University Press, 1986. – P. 112-127.

[26] Freeman J. Cultural influences on gifted gender achievement // High ability studies. The journal of the European Council for High Ability. – 2004. – № 1. – P. 7-23.

[27] Freeman J. Gifted children growing up / J. Freeman. – London : Cassell, 1991.

[28] Freeman J. Recent development for the high able in Britain // Education of the gifted in Europe: theoretical and research issues. – Amsterdam, 1992. – P. 58-70.

[29] Heller K. A. International trends and issues of research into giftedness // Proceedings of the Second Asian Conference on giftedness: growing up gifted and talented. – 1992. – P. 93-110.

[30] Large scale personality classification of bloggers / F. Iacobelli et al. // Affective computing and intelligent interaction. – Springer, Berlin, Heidelberg, 2011. – P. 568-577.

[31] Mining facebook data for predictive personality modeling / D. Markovikj et al. // Proceedings of the 7th international AAAI conference on Weblogs and Social Media (ICWSM 2013). – Boston, MA, USA. – 2013. – P. 23-26.

[32] Monks F. J. Development of gifted children: The issue of identification and programming // Talent for the friture. – Assen, 1992. – P. 191-202.

[33] myPersonality Project [Elektronnyi resurs] // URL: http://mypersonality.org/wiki/doku.php#mypersonality_in_numbers (data obrashcheniya: 04.12.2017).

[34] New machine learning methods demonstrate the existence of a human stylome / H. Van Halteren et al. // Journal of Quantitative Linguistics. – 2005. – Vol. 12. – № 1. – P. 65-77.

[35] Nowson S. Weblogs, genres and individual differences / S. Nowson, J. Oberlander, A. J. Gill // Proceedings of the 27th Annual Conference of the Cognitive Science Society. – 2005. – Vol. 1666. – P. 1671.

[36] Pennebaker J. W. Linguistic styles: language use as an individual difference / J. W. Pennebaker, L. A. King // Journal of personality and social psychology. – 1999. – Vol. 77. – № 6. – P. 1296.

[37] Personality, gender, and age in the language of social media: The open-vocabulary approach / H. A. Schwartz et al. // PloS one. – 2013. – Vol. 8. – № 9. – P. e73791.