Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 169 (2020) 158–163

Procedia Computer Science

www.elsevier.com/locate/procedia

Postproceedings of the 10th Annual International Conference on Biologically Inspired Cognitive Architectures, BICA 2019 (Tenth Annual Meeting of the BICA Society)

Career guidance based on machine learning: social networks in professional identity construction

Pavel Kiselev[a]*, Boris Kiselev[b], Valeriya Matsuta[c], Artem Feshchenko[c], Irina Bogdanovskaya[d], Alexandra Kosheleva[d]

[a]*Psychological Institute, Russian Academy of Education, 125009, 9-4 Mokhovaya str., Moscow, Russian Federation*
[b]*National Research Nuclear University MEPhI, 115409, 31 Kashirskoe shosse, Moscow, Russian Federation*
[c]*National Research Tomsk State University, 634050, 36 Lenin Ave., Tomsk, Russian Federation*
[d]*Herzen State Pedagogical University of Russia, 48 Moika emb., 191186, St. Petersburg, Russian Federation*

Abstract

Earlier research has shown that personality traits can be predicted by mining social networks data. This paper describes the social constructivism grounds of machine learning methods in career guidance and broadens understanding role of social networks in psychological researches. The theoretical grounds are empirically confirmed by AUC-ROC measure calculation in career guidance modeling. Implications for career guidance practice will also be presented.

Keywords: machine learning, career guidance, social networks, social constructivism.

1. Introduction

Career guidance in the 21st century is responding to the challenges of rapid transformation of society in a post-industrial and informational way. The enormous amount of data held about modern economics, labor markets and

* Corresponding author. Tel.: +7-926-852-50-39.
E-mail address: forestfield@yandex.ru

10.1016/j.procs.2020.02.128

cultural diversity cannot exceed human cognitive abilities for decision making. This is best illustrated by the example of searching the Internet which is practically impossible without search systems based on machine learning. Improving career guidance can be achieved by embedding machine learning into the career consulting ecosystem. Our study emphasizes the importance of the machine learning for the development of computer-assisted career guidance systems [1].

2. Background

The Big-Five theories offer career guidance professionals worldwide a set of principles and concepts that they can use to communicate about practice and research. These five theories are Theory of Work-Adjustment, Holland's Theory of Vocational Personalities in Work Environment, the Self-concept Theory of Career Development formulated by Super and more recently by Savickas, Gottfredson's Theory of Circumscription and Compromise, and Social Cognitive Career Theory. Every of these theories accept importance of values in career guidance. For example, Holland postulated [3] that people search for an environment that would allow them to express their values.

In addition, gender, social class, family background and cultural characteristics impacts on career development. These values assist to better career construction based on 'calling' not on simply surviving. Career choice and development as a process of personal and career construction settled by values. Unlike constructs that are more peripheral to an individual (e.g., attitudes, opinions), values are relatively permanent, although capable of being changed under certain conditions [4]. Values directly influence behavior in that they encourage individuals to act in accordance with their values [5].

According to Hogan and Blake [6] needs, values, and interests are closely related concepts for career guidance. Needs, values, and interests differ primarily in their breadth and level of abstraction [7]. Hence, the domains of vocational interests and work values are conceptually similar [8].

Career success is often associated with successful professional identity construction - especially in contemporary careers that are characterized by shifting boundaries in occupational, organizational, national, and global work arrangements [9].

One of the leading career researchers Savickas referred to using social constructivism to unite segmented career theories [10]. Social constructivism provides a theoretical framework and the key concepts to understand professional identity formation in social networks. Social constructivism holds that reality is constructed through the interaction with others and is influenced by biological factors, personality, social, cultural and historical experience. For young adults the primary mode of interaction is participation in digital practices within social networks. This participation results in identity formation.

1. Self-disclosure, which involves an iterative process of sharing personally relevant information and receiving feedback, is central to identity formation. Social networks allow them to express themselves in broader ways and to receive feedback from others. This could potentially lead to subtle changes and more rapid shifts in their identity that go beyond alterations that would already take place through face-to-face feedback.
2. Joining groups that reflect different aspects of their identity which they wish to explore or deepen. Social networks may simultaneously amplify dimensions of self-identify and extend group identities.

Second way can be understood that a young adult has repertoire of identities based on different groups used for constructing his narrative [11].

Values are not only footprints in social networks data, but also some external factor. Collaborating users in social networks is determined by values and permanently constructs values at the same time. Values are socially constructed within relationships and contexts. Values connect the personal and the social.

There are various research works in mining social networks datasets. The most popular direction deals with the study of reflecting in user's Facebook activities personality traits as Openness to experience, Neuroticism, Extraversion, Agreeableness and Conscientiousness [12, 13]. There is research in other social networks as Sina Weibo [14] and another theoretical constructs as Dark Triad [15]. All these researchers in turn are based on early blogs study by Oberlander, Gill and Nowson [16, 17].

This approach, however, has a major drawback. Because social networks are only indicated personality traits it's hard to understand their role in human development including career development. Differently from those research works we assume on theoretical basis of social constructivism and career theories that social networks are more than passive environment for footprint of user's personality. Collaborating in social networks is the key part of young adult's identity formation and the special case of professional identity formation. This process is moderated by values. For effective use of machine learning in career guidance it should be embedded in this process.

3. Method

3.1. Participants

The analysis for this study were conducted on the data collected from the website www.digitalhuman.ru. Participants were social networks users and were recruited from adware of computerized career guidance on the Internet. All of the participants approved the agreement with privacy policy and approved the accessing of their account through official API of VKontakte social network. VKontakte is the main Russian social networks for young adults, more than 95% of this demographic have accounts with this social network. Participants were provided with career guidance recommendations when the questionnaire was completed.

Before processing the data, the following categories of users who agreed to participate were removed:

1. Users with less than 10 group subscriptions Users tend to follow many groups in social networks, on average, participants follow 136 groups. Having a small number of subscriptions means that participant hides their main account and instead used a fictitious account for online questionnaires. Subsequently, 393 subjects were removed from dataset.
2. Users with several first-rank vocational interests. Subsequently, another 310 users were removed from the dataset.

The final sample included 1252 participants, 65% female and 35% male. The gender distribution is quite usual for online psychological services. All of participants confirmed age from 18 to 22 before test was began.

3.2. Questionnaire

All participants completed questionnaire for measuring of vocational interests according the model as described below. The questionnaire consists of 30 items which make up ten scales. Each of these items came from the Digital Human battery for career guidance.

The model implies predictions by vocational interests type appropriate to work values. Every type corresponds to professional sphere supports by Russian higher education and has more than several million workplaces in Russian economy. This typology extents classical Holland's vocational interest's typology for post-industrial society.

1. Sphere of administration and management: This sphere concerns accounting and human relation in corporation or government. Resource administration is important here to keep processes and systems working efficiently and productively, it is important to excel at bringing people together and maintaining harmony within a group. This sphere corresponds to Enterprising type in Holland types (RIASEC).
2. Health sphere: It is important to listen to others' problems, do not let emotions interfere with performance and respond constructively to the situation. In Russia, this is the education under patronage of Ministry of Health Care. This sphere corresponds to Social type in RIASEC.
3. Education sphere: The main work values are developing others, awareness of individual's progress, belief in human's potential. There are dozens pedagogical colleges in Russia for training future teachers. This sphere corresponds to Social type in RIASEC.
4. Creative sphere: The main work value is creating new art objects. Vacancies for this sphere are primarily for graphic, textile and interior designers. Creation of fiction or nonfiction literature by writers, journalists and editors corresponds to this sphere also. This sphere is supported by undergraduate and graduate programs in arts and corresponds to Artistic type in RIASEC.
5. Media sphere. The main work value for this sphere is performing on stage for positive feedback from auditory. The typical vacancies are actors, musician and dancers. YouTube bloggers correspond to this sphere too. In Russia, these are special academies, such as ballet academy, conservatory, etc. Sphere corresponds to Artistic type in RIASEC.

6. Sphere of manufacture: For engineers and manual workers in this sphere values are based on principles of being hands-on and useful. Besides mining, constructions and transport this sphere includes agriculture because modern agricultural manufacture looks like some factory. The Sphere of manufacture is the heart of industrial society. This sphere supported by undergraduate and graduate programs in engineering. Sphere corresponds to Realistic type in RIASEC.
7. Security and emergency sphere: The main work values are to serve the country, protect law-abiding citizens, provide judgement. In Russia, this is the education controlled by Ministry of Defense, police and secret service. This Sphere corresponds to Realistic type in RIASEC.
8. Sphere of sales and marketing: The main focus for this sphere is the importance to discover and meet customer's underlying needs, become a trusted advisor to customers. It may be achieved by analysis in marketing or maintaining networks of contacts and work-related friendships in sales. This Sphere corresponds to Enterprising type in RIASEC.
9. IT sphere: This sphere is concerned primarily with the creation of software with user-friendliness for the internal client as its primary objective. This sphere is supported by undergraduate and graduate programs in computer science and corresponds to Investigative type in RIASEC.
10. Research and development sphere: Curiosity is the primary motive for those involved in this sphere. This sphere is supported by undergraduate and graduate programs in fundamental research in natural sciences, math and humanities. This Sphere corresponds to Investigative type in RIASEC.

Alpha reliabilities of the subscales ranged from 0.66 to 0.89 for the different samples. The evaluation items were presented in a graphic rating scale format, each having nine possible response categories ranging from “completely not attracts” to “maximally attracts”. Rating vocational interests independently also permits the use of more sophisticated statistical analyses [18]. A rank ordering extracted from normative ratings for the purposes of analysis and for comparing ratings with rankings. Item parcels, formed to serve as observed indicators of each latent construct, all produced substantial loadings on their respective latent dimensions (range from .51 to .90), and confirmatory factor analyses indicated support for the hypothesized latent structure of the ten constructs. Subscales account for 72.76 percent of the total variance.

Cause dataset preprocessing rules a participant has one first-rank vocational interest. The distribution of participants across vocational interests as follows:

Table 1. Sample distribution of participants across vocational interests.

Vocational interest	Number of participants	Percent of participants
Sphere of administration and management	272	22
Health sphere	154	12
Education sphere	43	3
Creative sphere	179	14
Media sphere	165	13
Sphere of manufacture	28	2
Security and emergency sphere	116	9
Sphere of sales and marketing	87	7
IT sphere	115	9
Research and development sphere	93	8

3.3. Machine learning approach

Each vocational interest has been modeled in isolation by binary classification. Positive class corresponds to respondents first ranked the vocational interest get by questionnaire for this social network user. Negative class corresponds everyone else.

According to Table 1 classes are significantly unbalanced for all spheres. In such conditions, accuracy may be misleading because a majority class default classifier would obtain high accuracy, whereas the minority class is mainly ignored. To improve the measure predictive ability of models we used the area under the receiver-operating characteristic curve (AUC-ROC).

To avoid the overfitting was used K-fold (K = 5) cross-validation method. Dataset was randomly partitioned into five subsets of equal size, in each subset positive and negative classes were balanced. Of the five subsets, one subset is retained as the validation data for testing, and the remaining four subsets were used for training.

Each social network profile was coded in terms of gender. The set of features have been derived from information from followed groups. All features are categorical, so, we used Python realization of open-source gradient boosting on decision trees library CatBoost optimized for categorical features.

4. Results and Discussion

Table 2 presents the prediction of vocational interests on the social networks profile. Mean and standard deviation of AUC-ROC measure was calculated on 5-folds cross validation run of the train set using CatBoost model.

Table 2. Classification results based on AUC-ROC measure

Vocational interest	AUC-ROC	
	Mean	SD
Sphere of administration and management	0.70	0.03
Health sphere	0.73	0.04
Education sphere	0.83	0.08
Creative sphere	0.69	0.02
Media sphere	0.68	0.05
Sphere of manufacture	0.85	0.08
Security and emergency sphere	0.79	0.03
Sphere of sales and marketing	0.72	0.07
IT sphere	0.83	0.08
Research and development sphere	0.74	0.05

The AUC-ROC range from 0.68 (media sphere) to 0.85 (sphere of manufacture). Values AUC-ROC around 0.7-0.8 were considered fair [19] and indicates reasonably good performance on predicting vocational interests. Creative sphere and media sphere affiliation is more difficult to predict.

Feature extraction from joining social network groups for vocational interest prediction is a practical realization of social constructivism grounds of professional identity construction regulated by values. Theoretical grounds of modern career theories empirically confirmed by AUC-ROC measure calculation results.

The current findings have important implications for practice. Information about young adult's vocational interests can be obtained by themselves and career counselors without questionnaire completing to save time and get possibility of regular and longitudial monitoring. Questionnaires can be expensive to administer and time and resource intensive. Biases, such as the social desirability and the observer effect are less significant for social network data analysis [20].

The study results also suggest a number of other promising lines of research on the use of machine learning and social networks data in career counseling. For example, the practical task is the selection of a college major. The decision which college major to pursue is one of the farthest-reaching decisions for young adults. This study finds that using machine learning and data of social network profile will guide students to a major can be useful and enjoyable for students. Thus, advising becomes custom-fitted to these students.

Conclusion

The study has used theoretical background of social constructivism and machine learning to draw two main conclusions. One is that social networks datasets are more than digital footprint of values unlike early research of correspondence between personality and social networks. Collaborating users in social networks is essential of values, it's a permanent process of constructing values and identity formation.

The other is that machine learning models can be part of professional identity formation. Artificial intelligence doesn't bind career construction, but mediates this process through helping awareness about vocational interest appropriate to work values. Machine learning supports individuals and counsellors in constructing the personal development in a world of unprecedented and ongoing rapid changes occurring within the workplace and in individual careers. A novel approach to career recommendation without questionnaire completing appears to be a promising example of career guidance based on machine learning.

References

[1] Gati, I., Saka, N., & Krausz, M. (2001). Should I use a computer-assisted career guidance system? It depends on where your career decision-making difficulties lie. British Journal of Guidance & Counselling, 29(3), 301-321.

[2] Athanasou, J. A., & Van Esbroeck, R. (Eds.). (2008). International handbook of career guidance (Vol. 21). Springer Science & Business Media.

[3] Holland, J. L. (1997). Making vocational choices: A theory of vocational personalities and work environments. Psychological Assessment Resources.

[4] Meglino, B. M., & Ravlin, E. C. (1998). Individual values in organizations: Concepts, controversies, and research. Journal of management, 24(3), 351-389.

[5] Williams, R. M., Jr. (1979). Change and stability in values and value systems: A sociological perspective. In M. Rokeach (Ed.), Understanding human values (pp. 15-46). New York: Free Press.

[6] Hogan, R., & Blake, R. J. (1996). Vocational interests: Matching self-concept with the work environment. Individual differences and behavior in organizations, 89-144.

[7] Hogan, J., & Hogan, R. (1996). Motives, values, preferences inventory manual. Tulsa, OK: Hogan Assessment Systems.

[8] Ones, D. S., Anderson, N., Viswesvaran, C., & Sinangil, H. K. (Eds.). (2015). The SAGE Handbook of Industrial, Work & Organizational Psychology: V1: Personnel Psychology and Employee Performance. Sage.

[9] Slay, H. S., & Smith, D. A. (2011). Professional identity construction: Using narrative to understand the negotiation of professional and stigmatized cultural identities. Human relations, 64(1), 85-107.

[10] Savickas, M. L. (2005). The theory and practice of career construction. In S. D. Brown & R. W. Lent (Eds.), Career development and counseling: Putting theory and research to work (pp. 42-70). Hoboken, NJ: John Wiley.

[11] De Fina, A. (2006). Group identity, narrative and self-representations. Studies in interactional sociolinguistics, 23, 351.

[12] Markovikj, D., Gievska, S., Kosinski, M., & Stillwell, D. J. (2013, June). Mining Facebook data for predictive personality modeling. In Seventh International AAAI Conference on Weblogs and Social Media.

[13] Peng, K. H., Liou, L. H., Chang, C. S., & Lee, D. S. (2015, October). Predicting personality traits of Chinese users based on Facebook wall posts. In 2015 24th Wireless and Optical Communication Conference (WOCC) (pp. 9-14). IEEE.

[14] Xue, D., Hong, Z., Guo, S., Gao, L., Wu, L., Zheng, J., & Zhao, N. (2017). Personality recognition on social media with label distribution learning. IEEE Access, 5, 13478-13488.

[15] Bogolyubova, O., Panicheva, P., Tikhonov, R., Ivanov, V., & Ledovaya, Y. (2018). Dark personalities on Facebook: Harmful online behaviors and language. Computers in Human Behavior, 78, 151-159.

[16] Nowson, S. (2006). The Language of Weblogs: A study of genre and individual differences. Ph.D. Dissertation, University of Edinburgh.

[17] Oberlander, J., & Gill, A. J. (2006). Language with character: A stratified corpus comparison of individual differences in e-mail communication. Discourse Processes, 42(3), 239-270.

[18] Hicks, L. E. (1970). Some properties of ipsative, normative, and forced-choice normative measures. Psychological Bulletin, 74: 167-184.

[19] El Khouli, R. H., Macura, K. J., Barker, P. B., Habba, M. R., Jacobs, M. A., & Bluemke, D. A. (2009). Relationship of temporal resolution to diagnostic performance for dynamic contrast enhanced MRI of the breast. Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine, 30(5), 999-1004.

[20] Ziemer, K. S., & Korkmaz, G. (2017). Using text to predict psychological and physical health: A comparison of human raters and computerized text analysis. Computers in Human Behavior, 76, 122-127.