Technological University Dublin

ARROW@TU Dublin

Conference Papers

Social Media Research Group (SMRG)

2019

Classification of Schoolchildren on Professional Trajectories using Experience of Successful Specialists

John Cardiff
Technological University Dublin, john.cardiff@tudublin.ie

Elena Shushkevich
Technological University Dublin, shushkevich@yandex.ru

Svetoslav Zverev
Technological University Dublin

Mikhail Alexandrov
Autonomous University of Barcelona

Angels Catena
Autonomous University of Barcelona

Follow this and additional works at: https://arrow.tudublin.ie/smrgcon

Part of the Communication Technology and New Media Commons, Educational Sociology Commons, and the Social Media Commons

See next page for additional authors

Authors

John Cardiff, Elena Shushkevich, Svetoslav Zverev, Mikhail Alexandrov, Angels Catena, and Dmitry Srefanovskiy

Classification of Schoolchildren on Professional Trajectories using Experience of Successful Specialists

Svetoslav Zverev

Technological University of
Dublin, Dublin, Ireland
RANEPA, Moscow, Russia
svetoslav.zverev@gmail.com

Mikhail Alexandrov

Autonomous University of
Barcelona
Barcelona, Spain
RANEPA, Moscow, Russia
malexandrov@mail.ru

Angels Catena

Autonomous University of
Barcelona
Barcelona, Spain
angells.catena@uab.cat

John Cardiff

Technological University of
Dublin, Dublin, Ireland
john.cardiff@it-tallaght.ie

Elena Shushkevich

Technological University of
Dublin, Dublin, Ireland
e.shushkevich@yandex.ru

Dmitry Srefanovskiy

RANEPA
Moscow, Russia
dstefanovskiy@gmail.com

Abstract – **In the paper, we propose a new approach to vocational guidance of schoolchildren based on classification of pupil wishes between given professional trajectories, which are presented by profiles of successful professionals. Both wishes and profiles are replies in free text form on a questionnaire proposed by skilled psychologists. Such an approach avoids the well-known deficiencies of traditional methods including binary questioning, talks about concrete professions, and interviews with school psychologists. We use the simple terms selection for preprocessing and the traditional method of voting for classification. The mentioned procedures are discussed and the proposed approach is preliminary checked on invited specialists. This joint Russian-Irish research has been carried out with Moscow schoolchildren (2 schools) and Moscow specialists (2 trajectories). The results of presented pilot study look very promising. It is the basis for current applied research in Moscow and the future activities in Dublin.**

Keywords – education; vocational guidance; text processing; classification

I. Introduction[1]

A. Vocational guidance, state-of-the-art

It is well-known that vocational guidance of schoolchildren is a multi-dimensional problem, whose solution is usually considered on different levels. It is:

- governmental level, where one deals with needs of national economy and current planes of a given government;

- social level, where one should take into account mentality and traditions of people of a given region;

- personal level, when it is necessary to coordinate psychological characteristics of a given schoolchild with the specificity of professions, which he/she thinks about.

In this paper we consider only the personal aspect of vocational guidance and here the traditional approaches meet the following well-known difficulties [1, 2]:

- There are a lot of professional areas in modern industry and children lose themselves among professions. Really, it is impossible to understand the profession's requirements by means of a single lecture or single excursion to corresponding company;

- It is difficult to coordinate internal motivation of schoolchild and specificity of particular professions. Really, the former and the latter are fuzzy and hidden in various manifestations.

Existing methodologies based on numerous questionnaires and talks with psychologists do not cover these difficulties. As a result we often have incorrect choice of a future profession made by schoolchildren. This circumstance has the following two important consequences: (a) avoidable governmental expenses when pupils change their career; and (b) non-effective work of these pupils when they keep their career but complete their functions without any wish and interest.

The problem of vocational guidance proves to be in a depressed state for Russian economy during the last decade. A good example is presented by one of the biggest job search site career.ru [3]. It shows that only half of students from colleges and universities answer on expert's questions that

---

they are interested in their professional area and they are going to find such employment, as shown in Fig.1. However, the statistic of employment of graduates shows a worse picture (see Fig. 2).

Similar statistics have emerged from the Russian Presidential Academy of National Economy and Public Administration (RANEPA). Namely, only 37% of graduates work in the preliminary selected specialty [4].
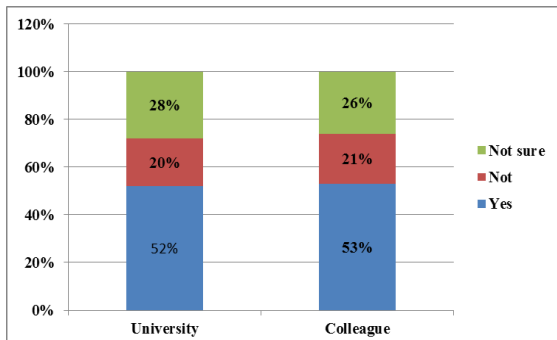

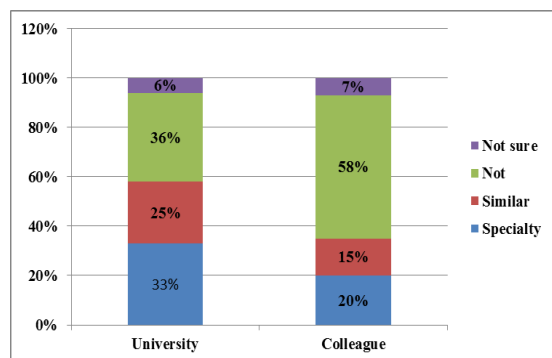Fig.1. Interest of students to their future profession, 2018.

Fig. 2. Employment of graduates, 2018.

The presented statistic is more than an ordinary figure. It is a huge financial loss for government and economy in general. For this reason last year RANEPA supported the project related to vocational guidance, which we are involved in.

B. Related work

The contemporary approach to determination of vocational orientation is based on the well-known three-factor model of American sociologist T. Parsons that includes the following elements [5]:

- informational (e.g. reflection of various aspects of professions in media);

- experienced (e.g. excursions to enterprises, internships, open days);

- psycho-diagnostic (e.g. talks with psychologists, analysis of personal parameters via various tests).

According this model a lot of vocational guidance tests are developed, which take into account all mentioned elements or some of them. However most of these questionnaires have the same deficits:

1. answer involves choosing one option from a list that limits possibility of schoolchildren to manifest their hidden motivation;

2. methods of vocational guidance are usually oriented on the interest to some aspects of profession but not to the profession in total;

3. analysis of results is very subjective because it is carried out by experts.

An interesting approach to avoid these limitations is described in the monograph [6]. Here, invited experts in a given area (in this case, military doctors) formulated requirements to interns of the Saint Petersburg Military Medical Academy. These requirements were presented in the form of a model of multi-criteria selection to tune a professional decision-support system. Its scales of parameters, set of criteria, and automated ranking of results partially covered the problems (1, 2, 3) mentioned above.

It should be noted that information concerning different professions from social media is often contradictive that reflects an information confrontation [7]. It has effects on pupil preferences but this task is out of the scope of this research.

C. Problem setting

In our research we prefer to talk about a professional trajectory but not about a concrete specialty. We consider the professional trajectory as a designation of a set of specialties having a similar character of work. The reasons are:

- unlike students, schoolchildren have no clear presentation of their future profession;

- we have no representative samples of successful specialists related to different specialties.

Our goal is to demonstrate possibilities to reveal hidden motivations of schoolchildren, which can be associated with different professional trajectories. Unlike the work [6] our approach is essentially less sophisticated and more universal. It includes:

- questionnaire with free answers without any mentions about concrete specialties;

- simple procedure of terms selection for transformation of replies on questionnaire;

- classifier based on the method of voting.

The form of questionnaire is developed by RANEPA's psychologists. Term selection uses its distribution between specialists of each group. The method of voting is similar to the method of potential functions, where potentials reflect closeness between marked objects from given classes and an object to be classified [8].

The number of trajectories is fixed and each trajectory is represented by its successful specialists. In the paper we consider 2 trajectories:

- the social-economical one including sociology, economy, finance, business administration, etc;

- the computer one including system and applied programming, computer security, computer vision, etc.

The pupils to be classified belong to 2 schools, where each school includes classes related to the mentioned trajectories.

The principal hypothesis consists in that the closeness of hidden motivations of pupils and specialists can help to detect the preferential trajectory for each schoolchild.

We use replies of specialists in order to know the distinguishability of trajectories, the best measure of distance/closeness, and the potential possibilities of an approach.

The text is organized as follows. Section 2 describes the experimental data and preprocessing. Section 3 gives the short description of classifier and measures used in calculations. In section 4 we study trajectories. Section 5 includes the results of classification of specialists and pupils. Section 6 concludes the paper

II. DATA

A. General information

RANEPA psychologists proposed the questionnaire including 8 questions. The questionnaire has the same questions both for ordinary schoolchildren and for successful professionals taking part in the project:

- schoolchildren should prepare their replies reflecting various aspects of their ideal work (the work of dream) in a free text form;

- specialists should prepare their replies reflecting their current work (independently positive or negative) in a free text form.

TABLE I. QUESTIONNAIRE WITH EXAMPLES OF WORDS FROM TEXTS

Questions	Schoolchildren	Specialists
Chiefs	interest professionalism	supervision responsibility
Colleagues	sincerity adequacy	friendship experience
Users	politeness understanding	exactingness punctuality
Self – realization	society competition	success hard work
Salary	investments teaching	responsibility success
Reputation	professionalism adequacy	efforts charity
Health	medicine sport	life weakness
Time	result overloading	money old age

Table I contains the list of mentioned 8 questions together with words related to these questions we frequently met in texts.

We consider such a questionnaire as the indirect one because it doesn't contain any direct questions concerning concrete professions. We think that these indirect questions and free responses can help to reveal hidden motivation of schoolchildren.

Answers of specialists are considered as their profiles, which belong to a concrete trajectory from a given list of trajectories. On the current stage of our research we consider the following 2 trajectories mentioned above:

- social-economic;
- computational.

In the classification problem presented below these trajectories are considered as classes.

In the experiments we used replies of 48 specialists and 27 schoolchildren. Table II contains general information about all participants of the experiments. The numbers in parenthesis are the number of schoolchildren belonging to social-economic trajectory and computer trajectory respectively.

TABLE II. NUMBER OF SPECIALISTS AND SCHOOLCHILDREN

Trajectories and schools	No. of specialists and schoolchildren
social-economical	21
computer	27
totally of specialists	48
school 2005	15 (5+10)
school 'kvartal'	12 (6+6)
totally of schoolchildren	27

B. Examples of short replies

The examples presented below are shortened replies on the questionnaire translated from Russian language.

Example 1. Text prepared by a schoolchild.

First of all, I believe that we should work for life but not to live for work. With this point, I'd like to feel a pleasure of communication with my colleagues and chiefs. I appreciate an interest of chiefs for their young colleagues and also I appreciate openness and sincerity between colleagues.

My self-realization will depend on my satisfaction by work and with communication with friends and relatives. To have my mind active I have to be a pupil till old ages and to have my health good and for this I need to be on friendly terms with sports. Money is not my priority; I need enough money only for a normal level of life and no more.

Example 2. Text prepared by a specialist.

I deal with risk projects and it is easier for me to talk about the negative aspects of a career that should be avoided than about the positive ones. Speaking of bosses and colleagues, I had the most of my problems from short-sighted and

narcissistic persons. Their experience was not the principal thing, since I had my own head on my shoulders.

Self-realization depends on achievements estimated by the society but not me myself. Naturally success in career has no any connection with dishonesty. Every success should be paid by time and often health.

C. Preprocessing and vocabularies

Preprocessing consists in transformation of all documents (replies on questionnaires) to their numerical form. The principal step here consists in term selection, which forms terms space for evaluation of relation between documents. In our model we use the following 2 rules:

- terms space is formed by profiles of specialists but not schoolchildren

- terms space is formed separately for each trajectory and then they are merged

The 1-st rule is the consequence of that the specialists are considered as ideal representatives having successful experience in their trajectories.

The 2-nd rule is the consequence of application of the method of voting, where specialists act together being inside the same trajectory and separately being between different trajectories.

The initial steps of preprocessing are well-known in text mining [9]: filtering stop words and stemming. Then we select rare terms. Speaking 'rare terms' we mean terms that belong not more than 10%-20% of profiles. These thresholds indirectly reflect our observations that just approximately 10%-20% of successful specialists have very specific experience reflected in their vocabularies.

It proved that the threshold 20% gives very poor vocabularies when images of pupils' documents are empty. From the other hand the threshold 5% forms redundant vocabularies, which worsen results of experiments. The threshold 10% was taken as a compromise. The sizes of vocabularies are presented in Table III.

TABLE III. SIZES OF VOCABULARIES (DIMENSION OF TERMS SPACES)

Trajectories	Non-filtered	Filtered
social-economical	408	62
computer	446	151
common	637	145

We would like to give a certain comment concerning preprocessing: we do not use the popular tf×idf procedure [9] because it works badly with short texts; we do not use very effective criterion of specificity [10] because our vocabularies includes general lexis; we also do not use any statistical and information criteria [9] because of limited numbers of specialists in a given trajectories.

Preprocessing of schoolchildren's replies reduces to filtering stop words, stemming and vectorization in the space of terms of specialists.

III. CLASSIFICATION

A. Method of voting

It is very old and simple method of classification having the title 'method of potentials' in the Russian literature [8]. It can be shortly described by the following way. Let we have an object presented in any R^n-space: $X=\{x_i\}$, $i=1,…n$. Let we have m-classes presented by their representatives in the same space: $Y_j=\{y_{i,j}\}$, $i=1,…,n$, $j=1,…,m_k$, where j is number of objects in k-class, m_k is quantity of objects in the k-th class. Let there is a measure that defines the closeness between objects X and Y: $S(X, Y_j)$. This measure can be named as a potential of object Y_j on object X.

The average potential of class k on the object X is equal:

$$U_k(X) = \Sigma_j S(X, Y_j) / m_k, k = 1, 2, ..., m.$$

The object belongs to class having the largest potential. This method was a predecessor of well-known method of k-nearest neighbors or k-NN.

B. Measures

In the paper, we test 3 well-known measures for evaluation of relation between schoolchildren and trajectories:

- Jaccard measure of similarity

- Cos-measure of closeness

- Euclidean measure of distance

Jaccard measure is popular for work with weak-filled matrixes. We deal with such matrixes. Cos-measure is popular in text mining. It doesn't depend on absolute frequencies of terms and depends only on their proportion. Euclidean measure doesn't require any comments. The Euclidean distance D is recalculated to closeness C according to the simple formula $C=1/(1+D)$ in order to be used in the method of voting.

C. Accuracy of classification

The assessment of accuracy of the results proved to be not an easy problem. Indeed, the natural way to do it would be to ask the schoolchildren participating in the experiments about their preferable profession or even trajectory. But with this question they were losing themselves because their choice was too big. So, we only asked the schoolchildren about the trajectory recommended by the program: whether they agree or not agree.

IV. TRAJECTORIES

A. Intersection of trajectories

We assess intersection of trajectories vocabularies in order to evaluate potential possibilities to have significant results of classification. Naturally if these vocabularies have the empty intersection then we can hope on high quality of results. From the other hand, if we have completed or almost completed intersection than our results cannot be high.

Intersection of vocabularies on the vocabulary of social-economical trajectory is equal to 47% and intersection of

vocabularies on the computer vocabulary is equal to 36%. These values are typical for many applications in text mining, where authors could get good results.

B. *Variety of profiles in trajectories*

Variety of profiles gives us information about representativeness of each trajectory by specialists taking part in the experiments. Obviously such representativeness should not be small.

We evaluate the variety of profiles using coefficient of variation. For social-economic trajectory it is equal 22%, and for computer trajectory it is equal 26%. Such variations can be accepted as the significant ones for the representativeness of given trajectories in the experiments.

V. EXPERIMENTS

A. *Results of experiments with specialists*

We classified specialists in order to evaluate the potential possibilities of the proposed method. Table IV presents the results of experiments with filtered vocabulary.

TABLE IV. ACCURACY OF CLASSIFICATION ON FILTERED VOCABULARY (SPECIALISTS)

Trajectories	Jaccard	Cos	Euclidean
social-economical	1,00	0,90	1,00
computer	0,80	1,00	1,00
general	0,92	0,94	1,00

One can see that the method of voting shows its efficiency with all measures.

B. *Results of experiments with schoolchildren*

Experiments with schoolchildren were the principal ones in our research. Tables V and VI present the results of experiments with filtered and non-filtered vocabularies.

TABLE V. ACCURACY OF CLASSIFICATION ON FILTERED VOCABULARY (SCHOOLCHILDREN)

Trajectories	Jaccard	Cos	Euclidean
social-economical	1,00	0,91	0,91
computer	0,29	0,71	0,86
general	0,72	0,83	0,89

These experiments show that the procedure of filtering of rare terms described in p. 2 is completely justified and the Euclidean measure is the best one.

TABLE VI. ACCURACY OF CLASSIFICATION ON NON-FILTERED VOCABULARY (SCHOOLCHILDREN)

Trajectories	Jaccard	Cos	Euclidean
social-economical	0,06	0,45	0,82
computer	1,00	0,86	0,71
general	0,39	0,61	0,78

VI. CONCLUSIONS

In the paper we proposed the new approach for assessment of professional preferences of schoolchildren. In this approach we use experience of successful specialists in given professional trajectories.

We tested the proposed methodology on real data and results proved to be very promising. It will be the basis for our future research and its practical applications in Dublin and Moscow.

We suppose: to increase essentially the quantity of schoolchildren, specialists, and trajectories. We also intend to consider the other methods of classifications including those based on GMDH technics [11].

REFERENCES

[1] B. Hopson, J. Hayes, The theory and practice of vocational guidance, Pergamon, 1968.

[2] A.Yu. Kobzov, A.V. Kobzova, "The importance of vocational guidance testing of young people in the choice of their future profession," Ed. Board 'Young scientist', 2013, no. 11, pp. 354-357, [rus] https://moluch.ru/archive/58/8096/

[3] Russian job resource // https://hh.ru/article/21826

[4] RANEPA resource // https://www.ranepa.ru/images/News/2018-05/28-05-2018-monitoring-vypuskniki.pdf

[5] T. Pasons, Wikipedia // https://en.wikipedia.org/wiki/Talcott_Parsons

[6] S. Mikoni, Multicriteria selection on a finite set of alternatives, Saint Petersburg, 'Lany', 2009. [rus]

[7] A. Petrov, O. Proncheva, "Modeling Propaganda Battle: Decision-Making, Homophily, and Echo Chambers," In: Springer, series CCIS, 2018, vol. 930, pp. 197-209. doi: 10.1007/978-3-030-01204-5_19

[8] A. Arkadiev, E. Braverman, Learning machine to classification of objects, Moscow: Nauka, 1971. [rus]

[9] R. Baeza-Yates, B. Ribero-Neto, Modern Information Retrieval, Addison Wesley, 2010.

[10] R. Lopez, "LexisTerm – the program for term selection by the criterion of specificity," Artificial Intell. Application to Business and Engineering Domain, ITHEA Publ., Rzeszov-Sofia, 2011, vol. 24, pp. 8-15.

[11] V. Stepashko, "Developments and Prospects of GMDH-Based Inductive Modeling," Advances in Intelligent Systems and Computing II, CSIT 2017, AISC series, vol. 689, Cham: Springer, 2018, pp. 474-491.