

Principle of the Closest Profession for Creating Vector Vocational Guidance Tests



Vladimir Savelyev

Abstract This article presents the principle of the closest profession, which is a practical application of the principle of vector representation of professions, presented by the authors in the previous works. The latter postulates that any profession can be represented as a vector in the n -dimensional space of activity elements. This representation allows defining measures of similarity (for example, Euclidean distance) between different professions. The principle of the closest profession postulates that if with the help of some test, the vector of a person in the vector space of professions is determined, the Euclidean distances can be calculated from a person to all vectors of professions and the n -closest to him or her can be deduced. Such a test, which is called by the authors a vector test, in contrast to traditional vocational guidance methods, represents not just recommendations for a certain class of professions (for example, socioeconomic or technonomic), but an ordered list of professions recommended for the respondent to master. The paper presents a description of the algorithm for the development of vector vocational guidance tests, which consists of three stages—vectorization of professions, development of test materials, and assessment of the validity and reliability of the method. This algorithm was tested, which resulted in a screening vector questionnaire of professional preferences. An assessment of the retrospective validity of the questionnaire is presented. This questionnaire can serve as a confirmation of the efficiency of the principle of the closest profession for creating vocational guidance tests. However, additional research is required to substantiate its practical use.

Keywords Vocational guidance · Professional self-determination · Vocational guidance tests · Vector space of professions · Vector test

V. Savelyev (✉)
Ural Federal University, Yekaterinburg, Russia
e-mail: vladimir.savelev@urfu.ru

1 Introduction

As a result of the development of high technologies, the rate of appearance, disappearance, and change in the content of various professions has increased by many times. The increase in digital opportunities and professional mobility has led to the fact that people are faced with the choice of a profession not only before entering a university but throughout their lives. In these conditions, the development of new vocational guidance methods that meet the requirements of the rapidly changing world becomes especially urgent.

Vocational guidance tests are one of the methods of supporting professional self-determination. Traditionally, they are professional preference questionnaires or aptitude tests. Due to diagnostics, a person receives information about which classes of professions are closest to him or her. Examples of such tests are the Differential Diagnostic Questionnaire by Klimov [2] or the Test by Holland to determine the professional type of personality [5].

Back in the late 1980s, Gavrilov identified a fundamental problem that limits the possibilities of vocational guidance tests [1]. It concerns the problem of psychological classifications of professions on which the available tests are based. It lies in the fact that an attempt to describe all the diversity of the world of professions with the help of a small number of classes leads to a significant loss of information about their features, which are important from the point of view of vocational guidance.

Gavrilov proposed a modular approach to solving this problem. Within the framework of this approach, each profession is divided into separate elements of activity (modules), which are assigned a set of psychological traits. Gavrilov also described an algorithm for identifying such modules.

Gavrilov's ideas were developed in the works by Savelyev, who proposed to consider professions as vectors in the n -dimensional space of activity elements [4]. The values for each coordinate of the vector determine the degree to which these elements are inherent in a particular profession. This representation allows performing a number of mathematical operations on professions, including determining the degree of similarity between professions. According to the author, the vectorization of professions is a fairly good alternative to their classification.

Savelyev also proposed and tested two algorithms that can be used to perform vectorization. The first of them, described in the original study, involves expert evaluation of job descriptions, classification of individual labor functions according to a pre-selected criterion, and counting the number of labor functions of different classes.

The second one, described in an earlier work, involves the use of topic modeling methods for analyzing job descriptions [3]. Topic modeling is a set of methods for computer analysis of a large set of texts that determine the topics of these texts. A topic is a set of words that, with a certain frequency, occur together in a number of texts. For example, the words "ball, goalkeeper, referee" form the topic "football", and "jury, case, judge"—something related to jurisprudence. For each text, the belonging to a particular topic is determined. Formally, topics define the coordinate axes of some linear space, while each text is a vector in this space. This allows interpreting the

results of the topic modeling of job descriptions as a set of vectors of professions defined in an n -dimensional vector space.

However, it remains unclear how exactly the principle of vector representation of professions can help in the development of vocational guidance tests. This work closes this gap. Another principle is proposed in it—the principle of closest professions, which will allow the development of such tests. The authors will also present the results of testing one screening vector questionnaire of professional preferences, which will illustrate the efficiency of the proposed principle.

2 Methods

To be able to use the principle of vector representation of professions for the development of vocational guidance tests, the following proposition is introduced: if the vector of a person is determined in the n -dimensional space of professions, then it is possible to calculate the Euclidean distances from a person to each of the professions and determine the k -closest to him or her. These professions will be recommended for further mastery. By analogy with the k -closest neighbor method used in machine learning, this provision is called the “principle of closest professions”.

Thus, the test that determines that the vector of the respondent in n -dimensional space does not just indicate the priority class of professions—it displays an ordered list of professions that are more likely to suit this respondent. It is proposed to call such tests vector tests.

The development of such tests involves several stages.

At the first stage, it is required to vectorize professions, having received a vector space of professions. Of the two proposed vectorization algorithms, the priority is the one that uses topic modeling. It is the least labor-intensive and less error-prone than the expert one. However, it requires some training in machine learning from the researcher. The result of this stage will be a table with vectors of professions, as well as sets of words describing the coordinate of the vector (element of activity).

At the second stage, a question or a test task is developed for each element of the activity. This task should meet the criterion of apparent validity and quantify the degree to which the respondent’s abilities/preferences correspond to this element of activity. For all questions/tasks, it is mandatory to establish a single dimension for scales with a quantitative assessment.

The third step is to determine the validity and reliability of the resulting test. Reliability can be assessed both by retesting and by determining the degree of internal consistency. However, the assessment of validity requires additional clarification.

In the psychological literature, it is common practice to assess convergent validity as an indicator of the overall validity of a test. The results of the new method are compared with the indicators of the old one, which measures the same psychological construct. However, in the authors’ case, it is difficult due to the lack of similar tests.

Instead of assessing convergent validity, it is proposed to assess retrospective validity; namely, to see how accurately a given test can predict the profession in

which the respondent is already working. An example of such an assessment is shown in the current study.

Due to the peculiarities and scope of this type of questionnaire, it makes no sense to perform standardization, since the main goal is to determine the pool of professions most suitable for the respondent. Standardization is necessary in order to obtain standards for comparing respondents in terms of the manifestation of one or another psychological trait.

The study presents the first results of testing this algorithm for the development of a vector questionnaire and is, in fact, a pilot one. The developed questionnaire is called the screening questionnaire of professional preferences. This study is limited only to assessing its retrospective validity. The general goal of the study is to determine the possibility and feasibility of developing a vocational guidance test according to the proposed algorithm.

The vector space for the development of this questionnaire was obtained by the authors by computer analysis of the descriptions of 431 basic groups of professions presented in the International Standard Classification of Occupations (MSKZ-08 or OKZ) using topic modeling. A specific method is Latent Dirichlet allocation (LDA).

Based on the results, the authors developed a screening questionnaire consisting of 17 questions.

To test this questionnaire, a sample of 222 people working in various specialties was taken. This sample is characterized by a bias toward highly qualified specialists and managers.

They were requested to complete the received questionnaire. The wording of the questionnaire was slightly changed—instead of asking them to indicate the desired profession, they were asked to assess how much their current profession includes this or that element. In addition, they had to indicate this profession in the questionnaire, which was subsequently assigned to one of the basic groups of professions in the OKZ.

The respondents' answers determined the vector of their current profession. Further, for each respondent, the Euclidean distances from the vector of their current profession to the vectors of professions in the OKZ were calculated. After that, the occupations of the OKZ were ordered in order of increasing distance from the smallest to the largest. Finally, in this series, the place (rank) occupied by the base group to which the respondent's profession belongs was determined. The lower the obtained rank was, the more accurately the respondent's current profession was determined by the test.

The ranks of the respondents' current occupation were used for further processing. A graph of accumulated frequencies was built, reflecting the number of subjects whose current profession rank is less than or equal to the given one. In addition, the first, second, and third quartiles were calculated by ranks, as well as the maximum rank in the sample.

The reliability was not assessed, since the purpose of the study was to test the very possibility of creating a vector questionnaire.

Microsoft Excel and the RStudio software development environment were used for data processing.

3 Results

Table 1 provides a description of the coordinate axes of the 17-dimensional vector space of professions, reflecting typical elements of activity. It is reminded that the authors obtained them as a result of LDA topic modeling, performed on job descriptions from the OKZ. Accordingly, the description is a set of words that define the topic.

It also provides wording for items on the screening questionnaire of professional preferences that correspond to each typical activity element.

General instruction for respondents:

- Assess how your real profession corresponds to the following statements (to assess the validity)
- Rate the extent to which what you would like to do corresponds to the following statements (to use the questionnaire).

The assessment is conducted on a six-point scale, where 1 is absolutely incorrect, and 6 is absolutely correct.

Table 2 provides examples of vectors of the three basic groups of specialties. In total, there are 431 vectors.

Please note that managers have the highest value along the T9 axis (organization, coordination, and control), doctors—along the T10 axis (treatment of diseases, receiving patients), and lawyers are distributed along three axes:

- T3—registration, recording, and paperwork; making reports; organization of information storage
- T4—research and experimentation; study of various objects and phenomena; expert judgment in any field
- T17—work in sales; consulting clients; acceptance of payment; organization of the sale and delivery of goods to the final recipient

Figure 1 shows a graph of accumulated frequencies. The abscissa shows the rank of the respondents' current profession. The ordinate is the number of subjects whose current profession rank is less than or equal to the given one. Thus, 17% of the subjects have the rank of their current profession less than or equal to 5. This means that for 17% of the subjects the test placed their own profession in the list of 5 professions closest to them.

The gray lines represent the 95% confidence interval.

Table 3 presents the quartiles according to the ranks of the base groups of the respondents' current professions.

Thus, for 25% of respondents, their own profession was in the top 9 of the closest, for 50%—in the top 25, for 75%—in the top 58. For all respondents, the rank of their basic professions was less than or equal to 290.

Table 1 Description of the coordinate axes of the vector space of professions obtained on the basis of the OKZ using LDA and the corresponding items of the questionnaire

No	Descriptions	Questionnaire items
T1 ^a	Product, manufacture, use, form, fabric, clothing, machine, conformity, detail, glass	Making various products by hand; work with various materials: fabric, glass, wood, clay, etc
T2	Security, assistance, performance, rule, provision, compliance, implementation, definition, relevance, identification	Work in the field of security; ensuring that other people comply with various rules and regulations
T3	Information, maintenance, document, verification, accounting, provision, compilation, registration, receipt, storage	Registration, recording and paperwork; making reports; organization of information storage
T4	Conduct, analysis, development, research, method, definition, consultation, study, evaluation, report	Research and experimentation; study of various objects and phenomena; expert judgment in any field
T5	Repair, installation, service, tool, test, equipment, check, specification, component, device	Assembly, repair, and maintenance of technical systems; installation, testing, adjustment
T6	Products, animal, collection, species, production, operation, service, work, stock, planning	Working in the field of agriculture; growing plants and animals, taking care of trees, picking mushrooms and berries
T7	System, provision, purpose, development, data, problem, recommendation, transmission, network, communication	Work with computer systems; software development; work with data; maintenance of computer networks
T8	Equipment, operation, process, control, work, water, treatment, cleaning, sample, log	Work in the heavy and mining industries; control of machine tools and machines; production process control
T9	Organization, activity, leadership, control, personnel, planning, coordination, provision, implementation, use	Organization, coordination, and control of the activities of other people; drawing up plans, developing projects
T10	Patient, help, treatment, conduct, care, health, condition, disease, procedure, specialist	Treatment of diseases, reception of patients, provision of patient care, assistance in health matters
T11	Material, use, structure, building, placement, construction, site, measurement, stone, quality	Building; wall construction; roofing; finishing work
T12	Means, goal, place, movement, vessel, passenger, case, driving, way, loading	Work on sea, rail, or road transport
T13	Preparation, program, evaluation, conduct, student, maintenance, development, learning, event, child	Teaching children and adults; preparation of training programs; conducting classes; issuance and verification of educational assignments

(continued)

Table 1 (continued)

No	Descriptions	Questionnaire items
T14	Work, management, machine, control, processing, leather, coating, service, purpose, application	Work in the light industry; control of machines that produce paper products, fabrics, leather, and plastic
T15	Product, food, cooking, cleaning, washing, drink, take, mix, food, zone	Work in the food industry; preparing or serving food; cleaning and washing dishes
T16	Preparation, quality, creation, material, choice, image, object, performance, form, paper	Work in the field of media; writing articles and books; image production; creation and performance of musical works; work in a theater
T17	Customer, product, sale, service, receipt, condition, demand, decision, delivery, provision	Work in sales; consulting clients; acceptance of payment; organization of the sale and delivery of goods to the final recipient

^aT—topic. One of the entities resulting from topic modeling, corresponding to the coordinate axis

Table 2 An example of vectors of three basic groups of professions

Coordinate	1120 Heads of institutions, organizations, and enterprises	2211 General practitioners	2611 Lawyers
T1	0.02	0.02	0.02
T2	0.03	0.04	0.07
T3	0.05	0.07	0.13
T4	0.05	0.06	0.15
T5	0.02	0.02	0.02
T6	0.02	0.06	0.04
T7	0.04	0.04	0.03
T8	0.02	0.03	0.04
T9	0.47	0.03	0.06
T10	0.02	0.46	0.03
T11	0.02	0.02	0.02
T12	0.02	0.03	0.07
T13	0.10	0.03	0.06
T14	0.02	0.02	0.02
T15	0.02	0.03	0.03
T16	0.02	0.02	0.09
T17	0.04	0.02	0.13

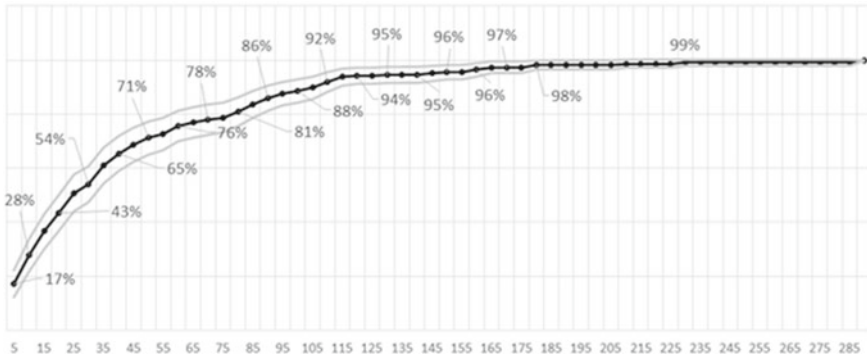


Fig. 1 A graph of accumulated frequencies

Table 3 Quartiles according to the ranks of the subjects' current professions

Indicator	Value
First quartile (25th percentile)	9.25
Second quartile (50th percentile)	25
Third quartile (75th percentile)	58
Maximum rank (100th percentile)	290

4 Discussion

It is reminded that the main goal of this study was to determine the possibility of developing a vocational guidance test based on the principle of vector representation of professions and the principle of the closest profession.

Based on the results of the first stage, a 17-dimensional vector space of professions was obtained. Moreover, each coordinate axis is fairly well interpreted from the point of view of the content of professional activity. Thus, the T9 axis clearly indicates managerial functions, while the T3 axis is connected with working with documents.

The optimal dimension of the space was chosen here by selecting and assessing the degree of interpretability of the resulting topics (coordinate axes). Thus, in the process of research, the authors have examined them in the range from 10-dimensional to 20-dimensional. 17-dimensional space appeared to be the most easily interpreted.

Topic modeling also enabled to vectorize 431 basic occupational groups from the OKZ, each of which received 17 coordinates. The examples were presented in Table 2.

The result of the second stage was a questionnaire consisting of 17 items. For each item, respondents needed to rate on a 6-point scale how much their current/desired profession includes the relevant activity.

At the third stage, the retrospective validity of the questionnaire was assessed by determining how much the questionnaire was able to predict the respondent's current occupation. The questionnaire showed acceptable accuracy—for 50% of the

respondents, the base group of their current profession was included in the list of the 25 closest ones. Taking into account the examination of 431 base groups, this is a rather decent indicator. However, in the absence of accuracy standards for vector tests, it cannot be stated whether it is acceptable for practical use in vocational guidance work.

It is also noted that the sample of respondents was biased toward specialists with higher qualifications, which does not allow speaking about the applicability of this methodology for other groups of specialties.

Finally, the factors are noted that negatively affect the test accuracy indicators:

1. Insufficiently clear wording of questions
2. The expert's mistakes when correlating the current profession declared by the respondents to one of the basic groups of professions
3. Inconsistency of the descriptions of the basic groups of professions presented in the OKZ with the real professional activity of the respondents.

5 Conclusions

In this article, the principle of the closest professions was formulated. It is based on the principle of the vector representation of professions, and is as follows: if with the use of any test, one determines the vector of a person in the vector space of professions, the Euclidean distances can be calculated from a person to all vectors of professions and the n -closest to him or her can be deduced. Guided by this principle, the authors have developed an algorithm for creating vector questionnaires. Such questionnaires make it possible not only to recommend a particular class of profession for the respondents but also to display an ordered list of professions recommended for mastering.

The algorithm was tested by creating a screening test of professional preferences based on the 17-dimensional vector space of basic groups of professions presented in the International Standard Classification of Occupations (MSKZ-08 or OKZ).

It is noted that this is the first experience in creating such questionnaires nowadays. The methodology for their development needs further discussion and clarification. Especially the part that concerns the validity check—it is obvious that for this type of questionnaire the practice of assessing convergent validity using a similar methodology is not suitable. Therefore, in this article, the authors proposed a method for assessing retrospective validity, trying to determine how accurately the test “guesses” a person's current profession.

The screening questionnaire of professional preferences, created by the authors, cannot currently be used as a vocational guidance test due to sample bias. To get the possibility of its practical use, an extensive study on a representative sample is required. However, as a demonstration of the applicability of the principle of closest occupations for creating vocational guidance tests, it is applicable.

References

1. Gavrilov, V. E. (1987). Using a modular approach for the psychological classification of professions for vocational guidance. *Questions of Psychology, 1*, 111–117.
2. Klimov, E. A. (1995). Methodology “Differential-diagnostic questionnaire”. In N. V. Kirshaeva & N. V. Ryabchikov (Eds.), *Psychology of personality: tests, questionnaires, methods*.
3. Savelyev, V. V. (2017). Application of topic modeling methods to professional standards to implement a modular approach to the classification of professions. *Pedagogical education in Russia, 3*, 22–28.
4. Savelyev, V. V. (2018). Profession as a vector in the n-dimensional space of activity elements. *Psychological Journal, 39*(5), 37–45.
5. Holland, J. L., Fritsche B. A., & Powell A. B. (1994). *SDS (self-directed search): Technical manual*. Odessa, FL: Psychological Assessment Resources.