

Combined use of correlation measures for selecting semantically close concepts of the ontology

A Yu Timofeeva¹, T V Avdeenko¹, E S Makarova¹ and M Sh Murtazina¹

¹Novosibirsk State Technical University, K. Marks ave. 20, Novosibirsk, Russia, 630073

Abstract. The paper suggests a new approach to the selection of correlated concepts for the ontology. It is based on the principal component analysis, but, unlike the standard approach, not Pearson correlation coefficients, but other correlation measures are used. This is due to the fact that the selection of concepts is based on data on the semantic association between concepts and cases, which are represented in the form of weight coefficients that take discrete values and a significant number of zero values. For such cases, the most appropriate is the polychoric correlation coefficient. It allows one to detect a monotonous dependence on the contingency table. However, for a certain table structure, the coefficient erroneously indicates a close relationship. This problem has been analysed in detail, and it has been suggested to use the correlation ratio in problem cases. Using the example of the problem of selecting concepts for the ontology in the IT consulting practice, the advantages of the proposed approach are shown. The first one is the increase in the percentage of variance of concepts explained by the principal components. The second one is that more concepts are selected based on unsupervised feature selection using weighted principal components.

1. Introduction

One of the key trends in the development of artificial intelligence is associated with the transition from the storage and processing of data to the accumulation and processing of knowledge. In this process, ontology, as a form of representation of knowledge, plays an important role. The main components of ontology are concepts of the subject domain. It is important to select concepts in such a way as to avoid their redundancy. So the semantically close concepts should be selected. This problem can be considered as one of the tasks of machine learning - the feature selection or the feature extraction.

There are several approaches to solving the problem of feature selecting [1]: filter methods, wrapping techniques, embedded methods.

Filter methods [2] are the simplest. They evaluate each variable according to individual criteria (information gain, chi-square statistics, etc.). An example is the selection algorithm Relief [2, 3]. The disadvantage of filter methods is that the correlation between the features is not taken into account, therefore, redundant attributes can be selected.

Embedded methods perform the feature selection as part of the model construction process. An example is the LASSO regression [4], constraining the weights of some features and shrinking of others to zero. Thereby a sparse solution is achieved. It includes only relevant features. Estimates of such regression, however, have no analytical expression. It requires the use of numerical optimization algorithms. In addition, the solution is very sensitive to the regularization parameter, which affects the degree of sparseness of the solution.

This drawback is attempted to be eliminated by using "wrappers", search procedures that include learning and evaluating the model using a potential subset of features. However, such procedures require, ideally, a search from all possible subsets of the feature set. So that algorithms are characterized by exponential complexity in terms of the number of features. This, as a rule, is unacceptable, and one must resort to "greedy" search algorithms, which never revise the earlier choice. For example, forward selection and backward elimination are used. However, they can give a locally optimal solution.

Typically, the described approaches are used in supervised learning. This requires a response variable. Based on the quality of its prediction, the attributes are selected. For example, the selection of ontology concepts could be done in order to improve the quality of classification of cases. However, the cases do not always have a class label. In this situation, the unsupervised feature selection is performed. This is a more difficult problem [5]. The approaches used can be categorized into cluster recognition and redundancy minimization.

Methods, that involve clustering, [6] select attributes to group data points (in our example, cases) in the best way. Other approaches are not restricted to clustering problems. Their goal is to select the smallest subset of attributes while preserving the most relevant information about the data [7]. The simplest criterion for selecting such subset can be the data variance. The explained variance can be a criterion for both selection and extraction of variables. The most popular approach here is the principal component analysis [8], which uses the decomposition of the covariance (correlation) matrix. Its results are used in the feature selection on the basis of weighted principal components [9]. These approaches are described in section 2.

However, when using correlation coefficients, it is necessary to take into account that the data, as a rule, are not continuous. Typically, the Pearson correlation coefficient is used, which can give biased results in the case of discrete data. For example, it was shown in [10] that when the validity of constructs is analyzed from ordinal values measured in the Likert scale, the results of factor analysis better reflect the theoretical model when factorization is performed using polychoric correlations rather than Pearson correlation coefficients. Nevertheless, the polychoric correlation coefficient has a number of drawbacks; in particular, with a certain structure of the contingency table, it erroneously reveals the presence of a strong relationship. This is a particular problem for sparse tables with a large number of zero values. Further, in Section 3, situations with poor behavior of the polychoric correlation are analyzed and other appropriate correlation measures are considered. Section 4 compares the various correlation measures and suggests ways to combine them to select concepts. Section 5 presents the results of applying the proposed approach for the selection of ontology concepts in IT consulting practice. Finally, Section 6 gives an interpretation of the results obtained and discusses the directions for their further application.

2. Dimensionality reduction techniques

Traditionally, dimensional reduction techniques [11] were developed for the analysis of either quantitative (principal component analysis) [12] or categorical data (correspondence analysis). Lately a lot of attention has been paid to approaches to the analysis of discrete data. It is suggested in [10] to use the polychoric correlation coefficients to reduce the dimensionality of such data. In addition, exploratory analysis methods for mixed data are being actively investigated. For example, the French school *Analyse des données*, founded by Jean-Paul Benzécri, develops factor analysis of mixed data [13]. These approaches differ in the way the correlation matrix is calculated. In general, the procedure of the principal component analysis remains standard, it is described below.

2.1. Principal component analysis

Let M be the correlation matrix of k features. On its basis one can obtain weights, which are the association between the variables and the components, so called loadings. Loadings vector for j -th principal component is calculated as

$$a_j = v_j \sqrt{\lambda_j} \quad (1)$$

where v_j is eigenvector of matrix M corresponding to the eigenvalue λ_j . The values of the eigenvector are normalized to the sum of the squares of the values. The matrix A of loadings for q principal components contains the vectors a_1, \dots, a_q , $q \leq k$, k is the number of features. The matrix of values of principal components (factor scores) can therefore be given as

$$F = XA(A'A)^{-1},$$

where X is initial data matrix $n \times k$, n is the number of cases. Note that the columns of matrix X are normalized, i.e. the sample mean of each variable has been shifted to zero, and the sample variance has been shifted to unit. The choice of number q is usually based on a scree plot, which shows the proportion of variance explained by each component.

2.2. Unsupervised feature selection

For the feature selection the results of the principal component analysis described in the previous subsection are used. The approach is based on the calculation of the weighted sum of the loadings for the i -th feature [9]:

$$\omega_i = \sum_{j=1}^q |a_{ij}| s_j \quad (2)$$

where a_{ij} is i -th element of the vector a_j , i.e. the loading for j -th principal component by i -th feature, s_j is the fraction of the explained variance calculated as

$$s_j = \frac{\lambda_j}{\sum_{l=1}^k \lambda_l}.$$

The ordering of the features in order of decreasing weights ω_i allows us to separate the essential concepts from the irrelevant ones. It is proposed in [9] to determine the threshold of weights on the basis of the ideas of a moving average control chart that has been widely used in quality control [14]. The difference is that the weights are not ordered. Therefore, it is proposed to use their random permutations and calculate the indicator

$$\overline{MR}_i = \frac{|\omega_{i_1} - \omega_{i_2}| + |\omega_{i_2} - \omega_{i_3}| + \dots + |\omega_{i_k} - \omega_{i_{k-1}}|}{k}$$

where $\omega_{i_1}, \omega_{i_2}, \dots, \omega_{i_k}$ is i -th random permutation. The number of permutations I should be taken sufficiently large to obtain stable results, for example, 1000. Further, the results are averaged:

$$\overline{MR}^* = \frac{1}{I} \sum_{i=1}^I \overline{MR}_i.$$

Finally, the threshold is calculated as follows:

$$\gamma = \bar{\omega} + \Phi^{-1}(1 - \alpha) \frac{\sqrt{\pi}}{2} \overline{MR}^*$$

where $\bar{\omega} = \frac{1}{k} \sum_{j=1}^k \omega_j$, $\Phi^{-1}(1 - \alpha)$ is a quantile of the standard normal distribution of order $(1 - \alpha)$, α is a given significance level, usually, 0.05. Based on the threshold the indicator of relevance of i -th feature can be constructed:

$$P(i) = \begin{cases} 1, & \omega_i \geq \gamma, \\ 0, & \omega_i < \gamma. \end{cases} \quad (3)$$

All the features for which $P = 1$ are recognized as relevant and selected.

In the article [9], which offers the described approach, it is not specified how the number q of extracted principal components is chosen. For a different number of components, different weights

will be obtained, which will affect the ordering of the features and the threshold γ . Further, this problem is investigated using the example of concept selecting.

3. Correlation measures

To analyze the association between attributes which are difficult to give objective quantification and whose values are ordered categories, the polychoric correlation coefficient is intended. It can be used also when counting data is analyzed, that is, discrete, taking a limited number of numerical values. It can also be rounded data, as well as data are subjectively and inaccurately, for example, expert ratings.

3.1. Polychoric correlation

Polychoric correlation ρ indicates an association between two theorized normally distributed continuous latent variables, from two observed ordinal variables. Its estimation is usually based on maximum likelihood method [15]. Polychoric correlation has the following properties:

- $-1 \leq \rho \leq 1$.
- It is symmetrical.
- $\rho = 0$ in the case of independence.
- If $\rho = 1$ then there is a strong monotonic relation.

The latter property is an advantage over the Pearson correlation coefficient, which reveals only a linear relationship. At the same time, the advantage of the polychoric correlation coefficient may turn out to be a significant drawback. So let us consider a number of examples of tables of relative frequencies:

$$D_1 = \begin{pmatrix} 0.5 & 0.25 \\ 0.25 & 0 \end{pmatrix}, D_2 = \begin{pmatrix} 0.74 & 0.01 \\ 0.25 & 0 \end{pmatrix}, D_3 = \begin{pmatrix} 0.74 & 0.25 \\ 0.01 & 0 \end{pmatrix}.$$

In all three cases, the value of the polychoric correlation is -1. Thus, the result does not depend on the frequency of non-zero values, as long as $d_{22} = 0$, and the rest were non-zero frequency. But if in the first case it is still possible to presume the presence of some nonlinear dependence, in other cases the small relative frequency of 0.01 can simply be a consequence of the presence of outliers.

The problem also remains for tables of higher dimension that satisfy conditions:

$$d_{li} \neq 0 \forall i, d_{jl} \neq 0 \forall j, d_{kl} = 0, \forall k, l \neq 1. \quad (4)$$

If the matrix is close to such a structure, the coefficient will be close to -1 and erroneously indicate an association. A similar problem is characteristic for the Yule coefficient [16], which reveals the relationship between binary variables. It is noted that it is unstable to small frequencies. However, the scientific literature does not offer approaches to solving this problem, which could be directly applied to the problem of feature selection.

Obviously, if the contingency table has a structure described by relations (4), then the use of the polychoric correlation coefficient leads to incorrect results. For this reason, it is necessary to involve other correlation measures that make it possible to identify nonlinear relationships and be appropriate for analysis discrete variables. In this case, they should be more sensitive to non-zero values of the observed frequencies $d_{li} \neq 0 \forall i, d_{jl} \neq 0 \forall j$.

The simplest approach would be to replace the polychoric correlation coefficient by the Pearson correlation coefficient in those cases where the first one falsely indicates a strong relationship. Such a trivial approach will also be analyzed, but it is better to choose a measure that is more suitable for analyzing relationships on discrete data.

3.2. Polyserial correlation

One option is the polyserial correlation coefficient ρ_{XY} . It reveals a latent correlation between a continuous variable X and a ordered categorical variable Y . It has the following properties:

- $-1 \leq \rho_{XY} \leq 1$.
- It is not symmetric, that is, $\rho_{XY} \neq \rho_{YX}$.

- $\rho_{XY} = 0$ in the case of independence.
- If $\rho_{XY} = 1$ then there is a strong association between X and Y .

Like polychoric correlation, an estimate of polyserial correlation is the result of maximizing the likelihood function [17]. Since, according to the properties, the coefficient ρ_{XY} is not symmetrical, it is therefore important here which of the variables is assumed to be continuous and which is discrete.

3.3. Correlation ratio

Presumably, the same drawbacks as the polychoric correlation may be inherent in the polyserial correlation coefficient. Therefore, in addition, we consider the correlation ratio of a random variable Y at random variable X defined as

$$\eta_{Y|X}^2 = 1 - \frac{\overline{D}_{Y|X}}{D_Y}, \quad (2)$$

where $\overline{D}_{Y|X}$ is the mean value of the conditional variance of the random variable Y under the condition X , D_Y is unconditional variance of a random variable Y . It is obvious from relation (2) that the correlation ratio is always nonnegative. The correlation ratio is asymmetric, that is, $\eta_{Y|X}^2 \neq \eta_{X|Y}^2$. A zero value indicates that there is no association. For comparison with the correlation coefficients, it is better to consider the value $\eta_{Y|X}$ or $\eta_{X|Y}$.

To analyze the possibilities of combined use of correlation measures, the polychoric, polyserial correlation coefficients and the correlation ratio have been calculated using the free software for statistical analysis R. For these purposes, a number of user-functions have been implemented [18].

4. Combined use of correlation measures

The selection of the ontology concepts is based on their semantic relations with the cases. The closeness of the semantic relation is determined by some weights that take values from 0 to 1. As a rule, weights are assigned expertly, so take discrete values (for example, rounded). The values of weight coefficients can be calculated on the basis of associative relationships between the case and the ontology concepts [19]. In this case they take a limited number of rational values as a result of multiplication of simple fractions. Thus, the weights are discrete.

The empirical study used data on the semantic association of cases and ontology concepts in the practice of IT consulting [20]. The data contains 120 cases and 20 concepts. First, a matrix of polychoric correlations between all the concepts was constructed. In total, the matrix (lower triangle) contains 190 correlation coefficients. As a result, it was found that 99 coefficients (about half) are close to -1. It should be noted that in the problem cases the numerical optimization of maximum likelihood function does not always give the estimate values exactly equal to -1, since values, close to -1, give approximately the same value of the objective function.

For comparison, the Pearson correlation coefficients r_{xy} are calculated. Figure 1 shows the results in the form of a scatter plot. Here and below (figures 2-4), the line represents the equality of the correlation coefficients, that is, for figure 1, it is a graph of equation $r_{xy} = \rho$.

It can be clearly seen from figure 1 that in most cases the polychoric correlation coefficient indicates a closer association between the concepts than the Pearson correlation coefficient. However, there are also clearly visible problem points - close to -1 values of the polychoric coefficient. In these cases, the values of the Pearson correlation coefficient range from 0 to -0.2, which indicates a rather weak relationship. Nevertheless, the Pearson correlation coefficient does not reveal a non-linear relationship, and therefore may underestimate the closeness of relation.

Both the correlation ratio and the polyserial coefficient are asymmetric. So, next, the correlation ratio η was calculated as the average value between $\eta_{Y|X}$ and $\eta_{X|Y}$. In the same way the polyserial correlation ρ_s was calculated as the average between ρ_{XY} and ρ_{YX} .

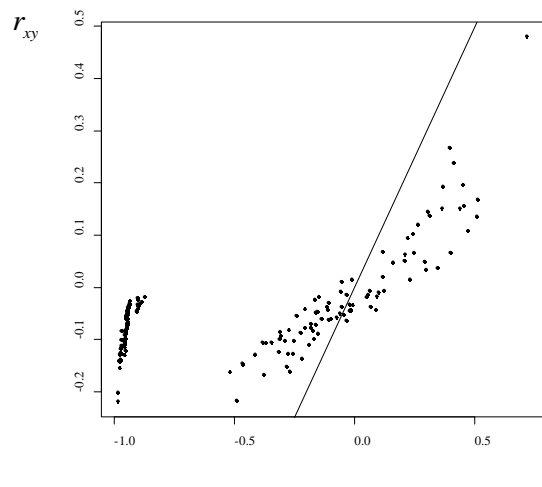


Figure 1. Scatterplot of polychoric and Pearson correlation coefficients.

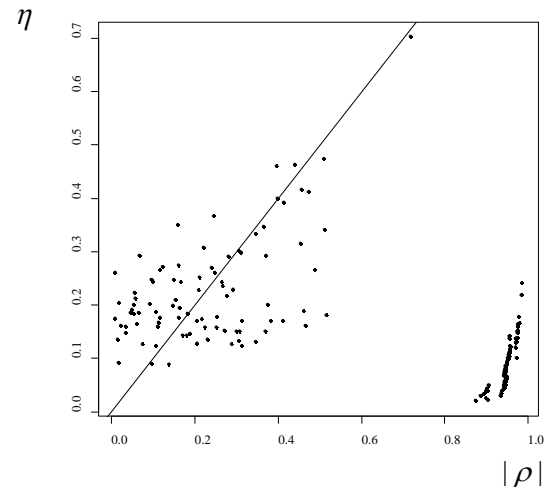


Figure 2. Scatterplot of polychoric coefficient and correlation ratio.

As noted above, the correlation ratio does not indicate the direction of the relationship, since it takes only non-negative values. For this reason, it is more correct to compare it with the absolute values of correlation coefficients. So figure 2 compares its values with the absolute values of the polychoric correlation coefficient. It can be seen that in a number of cases the correlation ratio shows a closer relationship, and in others the polychoric coefficient. There are also problem situations, in these cases the correlation ratio takes values close to the absolute values of the Pearson correlation coefficient, and indicates a weak relationship. But the values of the correlation ratio η , according to its properties, are always greater or equal to $|r_{xy}|$. So it is better to use the correlation ratio. In order to take into account the direction of the relationship, one must take the sign of the polychoric correlation coefficient.

If we compare the polychoric and polyserial correlation coefficients (figure 3), in most cases (77 coefficients out of 91, not related to the problem ones), the polychoric coefficients indicate a closer relationship than the polyserial ones. Thus, polyserial coefficients systematically underestimate the closeness of the relation. What can not be said about the correlation ratio: out of 91 non-problematic coefficients, only 44 polychoric correlation coefficients have the absolute value greater than the correlation ratio.

At the same time, in the problem cases, the polyserial coefficient shows a closer relation between the concepts, since it takes values from -0.6 to -0.2. However, this indicates, rather, that this coefficient also negatively reacts to a certain structure of the contingency tables. This is clearly seen in figure 4, which compares the absolute values of the polyserial coefficient and the correlation ratio. Situations, in which the values of the polychoric coefficient are close to -1, are highlighted in gray. They obviously stand out from the rest of the points on the graph.

Thus, we propose an approach to ontology concepts selection consisting of the following steps.

Step 1. Calculation of polychoric correlations ρ .

Step 2. Identification of problem situations by frequency tables satisfying (4), as well as by the values of the polychoric correlations close to -1.

Step 3. Replacement of the polychoric correlations in the problem situations, revealed at the step 2, by the values of the correlation ratios η , calculated as the mean between $\eta_{Y|X}$ and $\eta_{X|Y}$, multiplied by $\text{sign}(\rho)$.

Step 4. Based on the resulting correlation matrix M , consisting of polychoric correlations and correlation ratios, calculation of loadings vector for j -th principal component by the formula (1). Interpretation of results allows us to identify blocks of interrelated concepts.

Step 5. For the concept selection, the calculation of the weighted sum of the loadings by the formula (2). The weights $\omega_1, \dots, \omega_k$ allow us to order concepts by relevance. The calculation of indicator P by the formula (3), and the selection of most informative concepts.

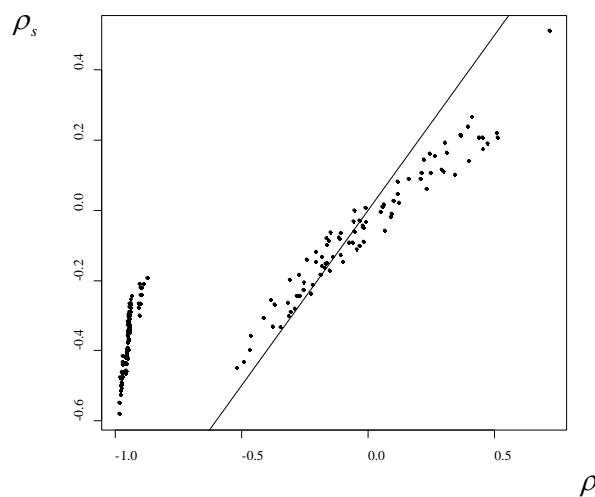


Figure 3. Scatterplot of polychoric and polyserial correlation coefficients.

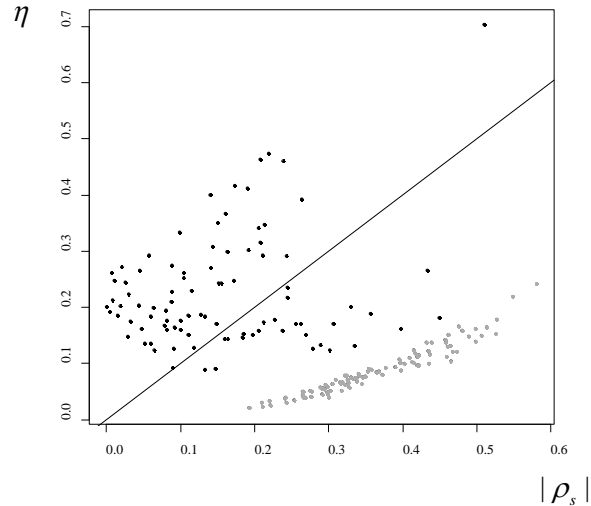


Figure 4. Scatterplot of polyserial coefficient and correlation ratio.

The advantages of the proposed approach as compared to the standard one (calculation of the Pearson correlation) should consist in increasing the percentage of variance of concepts explained by the extracted components. As a result, this allows us to partition the concepts into a smaller number of groups, the interrelation within which are closer.

5. Application in the practice of IT- consulting

Despite the fact that the concepts are carefully organized into the ontology by a domain specialist, the IT problem of the user being solved is often at the junction of various concepts. Therefore, the cases often refer to different hierarchical branches. The application of methods of grouping concepts could identify the most informative groups of concepts, as well as the most frequent combination of concepts describing the user's problems. The latter can be used in the decision support system for intellectual help for the user what additional concepts (in addition to the one already selected) to choose for the link with the current case (user problem).

In the above example of ontology in the practice of IT consulting we selected the most relevant concept using the proposed approach. For comparison the standard method based on Pearson correlation coefficients was used. The number q of extracted principal components was varied from 1 to 15. Table 1 shows the values of indicator function P defined given by (3). The results are presented for a limited number of principal components in order to demonstrate the differences between approaches. It can be seen that number and composition of selected concepts vary depending on q . Generally, the number of selected concepts is very small. It is difficult to detect any pattern of changes in a subset of selected concepts with an increase in the number of extracted components. Using the proposed approach, it is possible to increase the number of selected concepts. The maximum number of them is achieved when five principal components are extracted. Based on this, the optimal number q is chosen equal to five.

The loadings matrices for five principal components are presented in table 2. It allows to present concepts of the ontology of IT consulting in a space of small dimension. For clarity, only significant values of the loadings are given in the table. Their absolute values indicate the closeness of the relationship between concepts and the principal components. The weak relation of concepts (for example, "Vacation" in the standard approach) with all five components indicates that such concepts could not be included in the identified groups. The results of the feature selection (table 1) can be

compared with the results of the principal component analysis (table 2). For example, if one uses only the first component, only the attributes that closely correlate with this component remain in the set.

Table 1. The results of unsupervised feature selection.

Concept	The standard approach					The proposed approach				
	q=1	q=3	q=5	q=7	q=9	q=1	q=3	q=5	q=7	q=9
Vacation	0	0	0	0	0	1	1	1	0	0
Sick leave	0	1	1	0	0	1	1	1	0	1
Time-keeping	1	1	1	1	1	1	1	1	1	1
Calculation	0	0	0	0	0	0	0	1	0	0
Calculationof deductions	0	0	1	1	1	0	0	1	1	1
2-NDFL	1	1	1	0	0	1	1	1	1	1
6-NDFL	1	1	1	0	0	1	1	1	1	1

Table 2. Loadings on principal components and cumulative percentage of explained variance.

Concept	The standard approach					The proposed approach				
	1	2	3	4	5	1	2	3	4	5
Order on admission		0.445					0.664			
Order of dismissal		0.409			0.375		0.573			
Vacation						-0.629				
Sick leave		-0.469				-0.514				
Time-keeping	0.514	-0.440				-0.549				
Reporting		0.673					0.806			
Calculation prepayment				-0.457						0.573
Calculation					-0.399				0.748	
Payment at the average wage					0.396				-0.543	
Calculation of deductions			0.446	-0.422						0.476
Salary				-0.569						0.576
Recalculation			0.645					-0.607		
2-NDFL	-0.744					0.838				
6-NDFL	-0.681		-0.459			0.727				
Insurance payment			0.425	0.404	-0.337			-0.677		
Other taxes								-0.491		
Wirings								-0.461	0.415	
Cumulative explained variance, %	9.7	18.1	25.9	32.7	38.9	14.7	27.0	37.7	46.6	55.1

As can be seen, from the results of table 2, due to the proposed approach, the percentage of the explained variance significantly increases. So, with the standard approach, the five extracted components sum up only 38.9% of the initial variation of the concepts, whereas the proposed approach allows to explain 55.1% of the variance. In addition, five identified groups were able to include more concepts, additionally included "Vacation", "Other taxes" and "Wirings". Thus, the desired effect is achieved.

6. Results and discussion

The obtained results can be interpreted from the point of view of IT consulting practice.

The concepts, combined the first principal component, reflect the most common user errors in the calculations. If there is an incorrect calculation, then as a rule the error arises either in the incorrect

formulation of vacation or sick leave, and the problem with the time-keeping. At the same time, problems with vacation and sick leave can lead to the errors in reporting on taxes (2-NDFL and / or 6-NDFL). Reports on personal income tax are also interrelated, if there is an error or a question on one report, then the second one most likely will also have an error.

The second group of concepts deals with problems in personnel reporting. If there is a question on the admission / dismissal orders, there will be a problem with personnel reporting, and vice versa, if there is an error in the report, then it is worth checking the personnel orders (admission, dismissal). The concept "Recalculation" is connected with the third principal component. When recalculating, as a rule, users forget to remake taxes, so there are errors in taxes, insurance payment and wirings as a consequence.

Wirings also fell into the fourth group. The problem with wirings also arises when the calculation is incorrectly. These are interdisciplinary issues. The calculation and the payment at the average wage are mutually exclusive types, that is, at the same time a sick leave (payment at the average wage) and calculation (salary payment) can not meet together, this is a mistake. So, the user needs to make changes.

The fifth principal component associates with calculation prepayment, calculation of deductions and salary. In the payment documents, it is always necessary to check the calculation of deductions, so that everything is reflected correctly in the 6-NDFL statements. Also through salary payment documents a prepayment is formed. The prepayment is usually a fixed amount, sometimes as half of the salary, then in the payment document deductions are reflected. But such questions are rare.

Thus, concepts are combined into the groups by how often the errors occur when working with the software products. The first group of concepts is the most frequently encountered problematic situation, since the calculation errors are usually more frequent. The second most popular are the problems with personnel documents (the errors of the second group). The problems with taxes and the average wage are not very frequent operations and this part is fairly well implemented in the programs. Therefore, there are fewer questions on this part. The prepayment, deductions and salary are, as a rule, the most recent operations in the general list of all operations, and if everything was done correctly in the previous steps, then there are very few errors in this part.

As a result, concepts of different hierarchical branches were combined, i.e. errors often arise at the junction of the concepts that fall into groups. As a recommendation to improve the decision support system, this can be used, for example, if the consultant chooses one concept for linking the case, the system may recommend him choosing other concepts from the group identified on the basis of the principal components.

Thus, during the ontology concept selection based on their semantic relationships with cases, the use of the principal component analysis requires the choice of appropriate correlation measures. Due to the characteristics of the data representing the weight coefficients and taking discrete values, it is suggested to use the polychoric correlation. However, as it turned out, it gives incorrect results for a certain structure of the contingency tables. At the same time, in the conducted empirical study of the ontology of the IT consulting domain, this structure occurs quite often (in half the cases). Therefore, it is suggested in the problem situations to replace the polychoric correlation coefficient by the correlation ratio, which reveals nonlinear relationships and is appropriate for discrete data. As a result, such a combined use of correlation measures makes it possible to increase the percentage of the explained variance of principal components. So it allows to increase the number of selected concepts based on unsupervised feature selection using weighted principal components.

7. References

- [1] Flach P 2012 *Machine learning: the art and science of algorithms that make sense of data* (Cambridge University Press)
- [2] Kira K and Rendell L A 1992 *Proceedings of the ninth international workshop on Machine learning* 249-256
- [3] Robnik-Sikonja M and Kononenko I 2003 *Machine learning* **53** 23-69
- [4] Tibshirani R 1996 *Journal of the Royal Statistical Society. Series B (Methodological)* **58** 267-288

- [5] Dy J G and Brodley C E 2004 *Journal of machine learning research* **5** 845-889
- [6] He X, Cai D, and Niyogi P 2006 *Advances in neural information processing systems* **18** 507-514
- [7] Golay J and Kanevski M 2017 *Knowledge-Based Systems* **135** 125-134
- [8] Parveen A N, Nisthana H, Inbarani H H and Kumar E N S 2012 *Proc. Int. Conf. on Computing, Communication and Applications (ICCCA)* 1-7
- [9] Kim S B and Rattakorn P 2011 *Expert systems with applications* **38** 5704-5710
- [10] Holgado-Tello F P, Chacón-Moscoso S, Barbero-García I and Vila-Abad E 2010 *Quality & Quantity* **44** 153-166
- [11] Myasnikov E V 2017 *Computer Optics* **41(4)** 564-572 DOI: 10.18287/2412-6179-2017-41-4-564-572
- [12] Spitsyn V G, Bolotova Yu A, Phan N H and Bui T T T 2016 *Computer Optics* **40(2)** 249-257 DOI: 10.18287/2412-6179-2016-40-2-249-257
- [13] Pagès J 2014 *Multiple Factor Analysis by Example Using R* (London, Chapman & Hall/CRC The R Series)
- [14] Vermaat M B, Ion R A, Does R J M M and Klaassen C A J 2003 *Quality and Reliability Engineering International* **19** 337-353
- [15] Olsson U 1979 *Psychometrika* **44** 443-460
- [16] Kendall M and Stuart A 1961 *The Advanced Theory of Statistics: Inference and relationship* (London: Charles Griffin and Co., Ltd.) p 676
- [17] Drasgow F 1988 *Encyclopedia of statistical sciences* (John Wiley & Sons)
- [18] Timofeeva A Y 2017 *CEUR Workshop Proceedings* **1837** 188-194
- [19] Avdeenko T V and Makarova E S 2017 *CEUR Workshop Proceedings* **2005** 11-20
- [20] Avdeenko T V and Makarova E S 2017 *Journal of Physics: Conference Series* **803** 012008

Acknowledgments

The reported study was funded by Russian Ministry of Education and Science, according to the research project No. 2.2327.2017/4.6.