Comparison of cluster and classical approaches to issuing training information in adaptive learning systems

Konstantin Polyansky[1*], and *Igor* Kovalev[1, 2, 3]

[1]Siberian Federal University, Krasnoyarsk, Russia
[2]Krasnoyarsk State Agrarian University, Krasnoyarsk, Russia
[3]Navoi State University of Mining and Technologies, Navoi, Uzbekistan

Abstract. A comparison of cluster and classical approaches to issuing training information in adaptive training systems is performed. The adaptive training method of L.A. Rastrigin based on the frequency multilingual dictionary on system analysis and information technology is taken as a basis. The quality criterion of training of this method is configured to interact with the correlation matrix of terms obtained during the frequency analysis of the collection of text corpora. The resulting cluster approach shows better results compared to the classical one in the context of issuing terms that are most relevant to the lexical goals of the learner.

1 Introduction

To quickly master vocabulary in narrow subject areas, specialized dictionaries of foreign terms are created. An example is the frequency multilingual dictionary on systems analysis and information technology.

Based on this dictionary, foreign terminology training is conducted in systems analysis and in the IT sphere. The adaptive learning method of L.A. Rastrigin [1] serves as the learning algorithm.

Due to its focus on individual features of memorizing terms, the adaptive learning method allows you to build the learning process as effectively as possible. And in conjunction with a frequency multilingual dictionary, it becomes a means of intensive learning of the most popular foreign terminology.

However, the method has a significant drawback: it does not take into account the personal lexical interests of the student. The student almost always wants to study first of all the terms that are significant for him (in a professional or everyday context), which are related to his type of activity, tasks or hobbies.

Personalization of the adaptive learning method could increase its relevance to the learner's lexical goals. A personal approach to learning makes it more flexible and modern.

The structural diagram of the adaptive learning system (ALS) based on a multilingual frequency dictionary is shown in Figure 1.

* Corresponding author: kostyan785@mail.ru

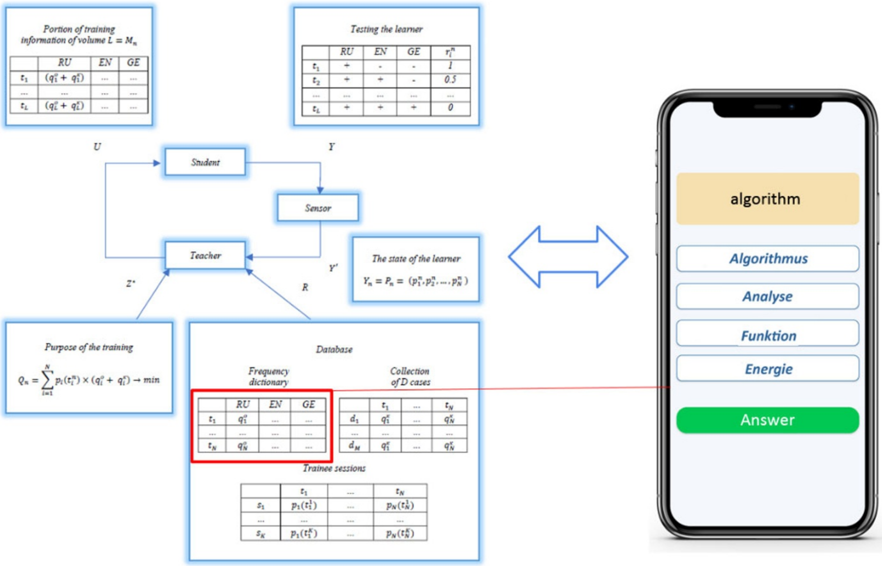


Fig. 1. Structural diagram of the ALS.

As can be seen from Figure 1, the ALS consists of several main blocks: a) database, b) learning goal, c) portion of learning information, d) learning test, e) learning state.

This paper describes the "learning goal" block, which specifies the learning quality criterion. A more detailed description of this block is given below.

2 Statement of the problem

In order to personalize the ALS, we will consider the classical criterion of training quality, identify its shortcomings and propose an alternative cluster criterion to eliminate these shortcomings.

2.1 AOC training quality criterion

2.1.1 Classic version

The adaptive learning algorithm is based on the learning quality criterion. Due to this criterion, in the n-th learning session, the algorithm offers the student exactly those terms that the student remembered the worst in the (n-1)-th session.

$$Q_n = \sum_{i=1}^{N} p_i(t_i^n) \times q_i \to min, \quad (1)$$

where n – training session number,

N – total number of terms of the information-terminological base (ITB) of the frequency dictionary,

$p_i(t_i^n)$ – probability of not knowing the i-th term from the n-th set of training information (TI),

q_i – relative frequency expressing the proportion of a lexical unit (term) in a text that has been statistically processed when compiling a frequency dictionary.

The adaptive learning algorithm aims to minimize the product $p_i(t_i^n) \times q_i$ for each i-th term by the end of the training session. Therefore, it is necessary to submit for training terms for which this product will be maximal:

$$\max(p_i(t_i^n) \times q_i), \tag{2}$$

because in the learning process, ignorance of terms is implied:

$$p_i(t_i^n) \to 0 \tag{3}$$

If we take into account that at the beginning of training the probability of not knowing any i-th term is equal to 1 (since training has not yet begun, and all terms from the set are unfamiliar to the student):

$$p_i(t_i^1) = 1, \tag{4}$$

then minimization of the quality criterion Q_1 at the first step of training is reduced to minimization of q_i:

$$Q_1 = \sum_{i=1}^{N} q_i \to min, \tag{5}$$

and therefore - to include in the portion of the TI terms with the maximum relative frequency. However, often at the beginning of training it is important for the student to study terms that are relevant specifically for him, related to his subject area (forming a single subject cluster). In this case, the classical approach considered cannot ensure personalization of training and becomes less effective.

2.1.2 Cluster version

To achieve personalization in the issuance of training information, it is necessary to modify the criterion of training quality. From the classical version of the operation of this criterion it follows that the issuance of terms depends entirely on the value of the indicator q_i Let us decompose the indicator q_i $into\ components$:

$$q_i = q_i^o + q_i^k, \tag{6}$$

where q_i^o – frequency component of the indicator q_i ,

q_i^k – cluster component of the indicator q_i .

The idea is that terms can be grouped into clusters relative to their co-occurrence in some collection of text corpora. And if one of the clusters is closest to the learner's lexical goals, then the terms of that cluster will have high q_i^k values, thereby displacing terms with high frequency.

The criterion of quality of education, according to the cluster approach, takes the form:

$$Q_1 = \sum_{i=1}^{N} q_i^o + q_i^k \to min \tag{7}$$

With the classical approach, the component $q_i^k = 0$ and terms are given for training only taking into account the frequency component q_i^o. The cluster approach allows reducing the influence of frequency by increasing the weight of terms that are in the same cluster as terms from the student's lexical goals. This makes this approach more personalized in relation to the student.

The problem of finding the components q_i^k is reduced to the analysis of a training collection of text corpora, which can provide information about the mutual relationship of terms and the clusters into which they are combined.

3 Materials and methods

To perform cluster analysis of a collection of text corpora, we will use the popular algorithm of frequency weighting of terms TF-IDF. At the beginning, the frequency component of terms in the collection of corpora is taken.

$$tf(t,d) = \frac{number\ of\ occurrences\ of\ the\ term\ t\ in\ the\ corpus\ d}{total\ number\ of\ terms\ in\ corpus\ d} \tag{8}$$

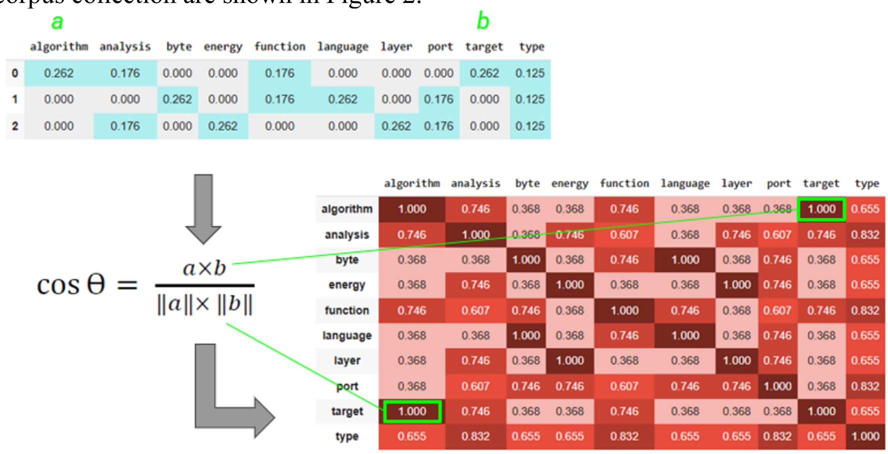
Next, we calculate the "inverse document frequency". This characteristic indicates the significance of the term not within one corpus, but in the collection of corpora under consideration.

$$idf(t,D) = log(\frac{total\ number\ of\ cases\ in\ case\ collection\ D}{the\ number\ of\ corpora\ with\ term\ t\ in\ the\ collection\ of\ corpora\ D}) \quad (9)$$

The product of the values $tf(t,d)$ and $idf(t,D)$ provides a summary characteristic of TF-IDF, reflecting both the significance of the term for the corpus and for the collection of corpora as a whole [2].

$$TF-IDF = tf(t,d) \times idf(t,D) \quad (10)$$

The TF-IDF values calculated (the values are given with the correction for the TfidfVectorizer algorithm [3] of the sklearn module of the python language) for all terms of the corpus collection are shown in Figure 2.



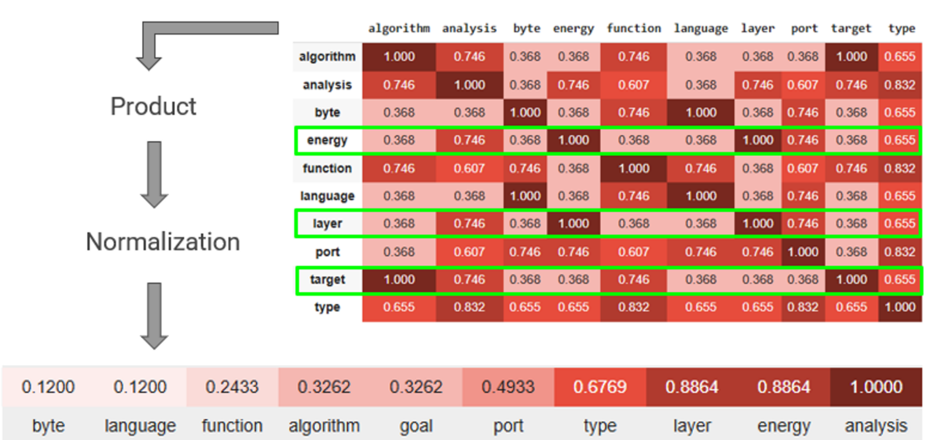
	algorithm	analysis	byte	energy	function	language	layer	port	target	type
0	0.262	0.176	0.000	0.000	0.176	0.000	0.000	0.000	0.262	0.125
1	0.000	0.000	0.262	0.000	0.176	0.262	0.000	0.176	0.000	0.125
2	0.000	0.176	0.000	0.262	0.000	0.000	0.262	0.176	0.000	0.125

	algorithm	analysis	byte	energy	function	language	layer	port	target	type
algorithm	1.000	0.746	0.368	0.368	0.746	0.368	0.368	0.368	1.000	0.655
analysis	0.746	1.000	0.368	0.746	0.607	0.368	0.746	0.607	0.746	0.832
byte	0.368	0.368	1.000	0.368	0.746	1.000	0.368	0.746	0.368	0.655
energy	0.368	0.746	0.368	1.000	0.368	0.368	1.000	0.746	0.368	0.655
function	0.746	0.607	0.746	0.368	1.000	0.746	0.368	0.607	0.746	0.832
language	0.368	0.368	1.000	0.368	0.746	1.000	0.368	0.746	0.368	0.655
layer	0.368	0.746	0.368	1.000	0.368	0.368	1.000	0.746	0.368	0.655
port	0.368	0.607	0.746	0.746	0.607	0.746	0.746	1.000	0.368	0.832
target	1.000	0.746	0.368	0.368	0.746	0.368	0.368	0.368	1.000	0.655
type	0.655	0.832	0.655	0.655	0.832	0.655	0.655	0.832	0.655	1.000

Fig. 2. Formation of a correlation matrix of terms from a “bag of words”.

Here the rows are the text corpora, and the columns are the terms of the frequency dictionary. To perform a vector comparison of a pair of terms, we use their cosine distance [4].

By multiplying the vectors from the table in Figure 2 in pairs, we obtain a correlation matrix [5] of terms. The values of this matrix are the values of the cosines of the angle between the vectors of the terms, written as its rows and columns.



	algorithm	analysis	byte	energy	function	language	layer	port	target	type
algorithm	1.000	0.746	0.368	0.368	0.746	0.368	0.368	0.368	1.000	0.655
analysis	0.746	1.000	0.368	0.746	0.607	0.368	0.746	0.607	0.746	0.832
byte	0.368	0.368	1.000	0.368	0.746	1.000	0.368	0.746	0.368	0.655
energy	0.368	0.746	0.368	1.000	0.368	0.368	1.000	0.746	0.368	0.655
function	0.746	0.607	0.746	0.368	1.000	0.746	0.368	0.607	0.746	0.832
language	0.368	0.368	1.000	0.368	0.746	1.000	0.368	0.746	0.368	0.655
layer	0.368	0.746	0.368	1.000	0.368	0.368	1.000	0.746	0.368	0.655
port	0.368	0.607	0.746	0.746	0.607	0.746	0.746	1.000	0.368	0.832
target	1.000	0.746	0.368	0.368	0.746	0.368	0.368	0.368	1.000	0.655
type	0.655	0.832	0.655	0.655	0.832	0.655	0.655	0.832	0.655	1.000

0.1200	0.1200	0.2433	0.3262	0.3262	0.4933	0.6769	0.8864	0.8864	1.0000
byte	language	function	algorithm	goal	port	type	layer	energy	analysis

Fig. 3. Finding the row vector of parameters q_i^k from the correlation matrix of terms.

 https://doi.org/10.1051/itmconf/20257204010

Using the data of the correlation matrix, it is possible to determine how interconnected (correlated) the terms are within the corpus collection (the highest correlation corresponds to the value 1, the lowest - 0). And the task of finding the joint correlation of several terms is reduced to the row-by-row multiplication of their cosine values. The resulting multiplied values can be normalized using the "minimax" method [6].

Figure 3 schematically shows the process of finding the row vector of parameters q_i^k from the correlation matrix of terms.

4 Results

Based on the conducted cluster analysis of the text corpus collection, a term correlation matrix was obtained. Using the data of this matrix, it is possible to obtain a vector-row of parameters q_i^k . The values of these parameters are substituted into the quality criterion of ALS training.

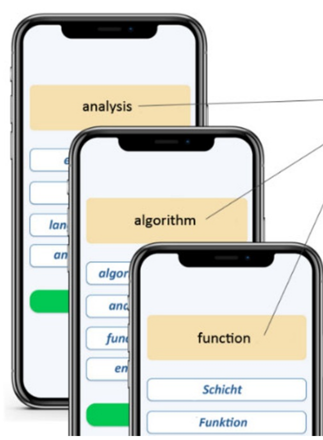
As a result, two fundamentally different algorithms for issuing training information can be obtained: classical and cluster. In the classical algorithm, the parameters q_i^k take the value 0, while in the cluster, the parameters q_i^k are taken from a row vector obtained on the basis of cluster analysis.

The difference is that the classical criterion of the quality of learning ALS issues terms ranked by their frequency. The cluster criterion of the quality of learning issues terms for learning based on the correlation matrix, and therefore, taking into account the lexical goals of the learner.

This difference in the output of terms for the two algorithms can be observed in the example where the learner is primarily interested in learning the terms: *goal*, *energy* and *layer*.

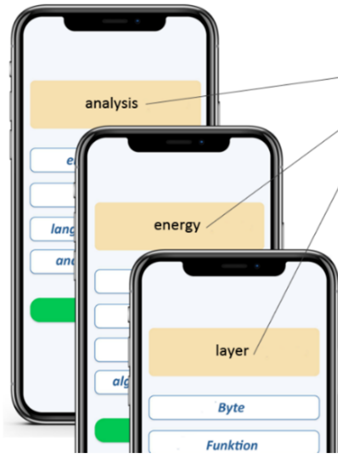
Figure 4 shows that these lexical preferences for terms are not taken into account, and the classical approach outputs terms solely based on their frequency. The terms goal, energy, and layer do not make it to the top of the training information output and are not offered for training first.

On the other hand, the cluster approach shows better results in the context of personalized educational information delivery. The terms goal, energy, and layer are in the top positions of the list, as can be seen in Figure 5.



	RU	q^{o}	$q^{\text{к}}$	$q^{\text{o}}+q^{\text{к}}$
1	analysis, 327	0.327	0.000	0.327
2	algorithm, 126	0.126	0.000	0.126
3	function, 113	0.113	0.000	0.113
4	language, 101	0.101	0.000	0.101
5	target, 94	0.094	0.000	0.094
6	byte, 47	0.047	0.000	0.047
7	energy, 15	0.015	0.000	0.015
8	type, 5	0.005	0.000	0.005
9	port, 4	0.004	0.000	0.004
10	layer, 2	0.002	0.000	0.002

Fig. 4. Issuing terms for training based on their frequency.



cluster window

	RU	q^{o}	$q^{к}$	$q^{o}+q^{к}$
1	analysis, 327	0.327	1.000	1.327
2	energy, 15	0.015	0.886	0.901
3	layer, 2	0.002	0.886	0.888
4	type, 5	0.005	0.677	0.682
5	port, 4	0.004	0.493	0.497
6	algorithm, 126	0.126	0.326	0.452
7	target, 94	0.094	0.326	0.420
8	function, 113	0.113	0.243	0.356
9	language, 101	0.101	0.120	0.221
10	byte, 47	0.047	0.120	0.167

Fig. 5. Issuing terms for training from the cluster.

With the cluster approach, a so-called cluster window is formed, which includes not only target terms (goal, energy and layer), but also terms that form a single semantic cluster with them.

For terms within the cluster window, the value of the parameter q_i^k plays a decisive role and determines their position in the list. While outside the cluster window, the parameter q_i^k takes a fixed value, and here the frequency of use of terms regains its power.

5 Conclusion

In the course of this work, the TF-IDF frequency weighting algorithm was considered, as well as the method of vector comparison of terms based on the cosine distance. As a result, a term correlation matrix was compiled, allowing one to predict the belonging of terms to a particular text corpus (cluster).

The algorithms of adaptive learning and term clustering have been reduced to a single concept. The adaptive learning algorithm of L.A. Rastrigin has been personalized based on a frequency multilingual dictionary using cluster analysis of a text corpus. The learning quality criterion of this algorithm has been modified.

As a result, a comparison was made of the relevance of information output regarding the student's lexical goals using the classical and cluster approaches. The latter shows better results in the context of outputting the terms most closely related to the student's request.

References and links

1. I. V. Kovalev, M. V. Karaseva, E. A. Suzdaleva, *Systemic aspects of the organization and application of multilingual adaptive learning technology* Educational Technology & Society **5,** 207-208 (2002).
2. D. Jurafsky, J. H. Martin, *Question Answering, Information Retrieval, and Retrieval Augmented Generation* Speech and Language Processing **3(14),** 4 (2014).
3. Scikit Learn: website. – 2007. – URL: https://scikit-learn.org/stable/modules/generated/sklearn.feature_extraction.text.TfidfVectorizer.html (date of access: 01.11.2024).

4. Aristoteles, M. U. Syam, Tristiyanto, B. Hermanto, *Implementation of Cosine Similarity Algorithm on Omnibus Law Drafting* (IJACSA) International Journal of Advanced Computer Science and Applications **4(15),** 200-205 (2024).
5. A. R. Hadd, *Correlation Matrices in Cosine Space* (Nashville, Tennessee, 2016).
6. V. V. Starovoytov, Yu. I. Golub, *Data normalization in machine learning* Computer Science **18(3),** 86 (2021).