Research on the Evaluation Effect of English Network Based on CADATS

Chunling Liu[1]

[1]School of Foreign Languages, Wuhan Polytechnic University, Wuhan 430072 China, liuchunlingedu@yeah.net

Corresponding author: Chunling Liu, liuchunlingedu@yeah.net

Abstract. In order to solve various problems in the implementation of computer adaptive English test, this study used CADATS to conduct adaptive test of students' English ability and tested the performance of CADATS in exams and adaptive tests. At the same time, the information gain in the decision tree is introduced as the attribute weight calculation method. In addition, with the help of Dreamweaver web editor, in the ASP scripting environment, the design and development of the system main interface is completed by combining Visual Basic scripting language (VBScript) and HTML code. Finally, performance analysis is performed through experiments. The research results show that the proposed algorithm has certain performance and can provide theoretical reference for subsequent related research.

Keywords: CADATS; English; network assessment; system testing; network teaching.
DOI: https://doi.org/10.14733/cadaps.2020.S1.23-33

1 INTRODUCTION

Educational testing can not only detect the effects of previous learning (how much knowledge information is kept in long-term memory), it can also reverse the knowledge information in long-term memory. Moreover, compared with simple repetitive learning and over-learning, one or several tests in the knowledge learning process (even if the test does not give any feedback) can significantly promote the long-term retention of learning content and improve learning efficiency. Researchers refer to this phenomenon as the Testing Effect, also known as the Retrieval Effect [1]. The existing computer-based test system is only a simple platform transfer for traditional paper testing. Because computers have the ability to process data information at high speed, computer-based test systems have the following advantages: improve the marking rate, reduce the labor intensity of teachers, reuse the test questions library and test paper library, shorten the statistical period of performance, and facilitate the management of relevant data information. Computer-based test systems still need to be further strengthened for the study of system versatility and ease of use, and the advantages of the network are not fully utilized [2].

The development of foreign computer-based test systems is relatively mature. Moreover, due to the strong portability, portability and time-space of the network-based test system, many international large-scale test systems have already completed the support of the network test system. The more famous ones are the TOEFL test and the GRE test. [3]. The TOEFL test is an English proficiency test conducted by the American Educational Testing Service (ETS), which is called the "English proficiency test for non-native English speakers." TOEFLiBT (TOFEL Internet. based Test) was applied. The TOEFL iBT was first used in the United States, Canada, France, Germany and Italy.[4]. The new GRE test (Revised GRE), after five years of research by ETS, officially began the reform of the GRE regular test under the guidance of the education sector represented by the US Graduate School [5].

Since the 1990s in the country, the authoritative examination department has gradually implemented a computer-based test system, but it was difficult to be accepted and popularized at that time. Many national computer-based examinations in China have begun to use computers to organize tests [6].The test system related to the computer test took the lead in replacing the traditional paper test in China, which triggered the reform of the test content, test methods and test forms in the domestic education industry and had a profound impact on the domestic informatization of teaching test. Subsequently, some university English-related test systems gradually emerged in the domestic market. However, most of them are test systems designed for common tests in the undergraduate English teaching process, such as the oral test system promoted by the author's school (China University of Science and Technology) and the college English four or six computer test. The author's school took the lead in reforming the English exam, began using the oral test system, conducted oral tests on all undergraduates, and achieved such a large-scale “paperless” test for the first time [7]. sThe College English Test (CET) is a national teaching test conducted by the Ministry of Education. It is divided into four levels (CET-4) and six (CET-6), which are held once a year in June and December. The country began to try the college English four or six computer test, which is the symbol of the transformation of China's large-scale English standardized test to computer language test system [8]. More than 50 colleges and universities set up by the Ministry of Education piloted the fourth- and sixth-level college English test. The form of the test was the same as that of the TOEFL iBT and the GRE online test. In the next few years, the scale of the four- or six-level computer test is gradually expanding. Although the computer test has not yet been fully popularized, the promotion of the four- or six-level computer test has become the development trend of the network-based test system [9].

In China, China's teaching and testing informatization has developed rapidly, and there are many computer testing systems developed and applied in China, and these systems are gradually shifting from local applications to Web applications [10]. However, from the perspective of the functional improvement of the test system and the convenience and versatility of the system, China's network test system is still in the stage of exploration and improvement, and there are still many shortcomings that need further research. The popularity of test systems based on computers and networks has become more widespread. The existing test system is relatively complete in the implementation of relevant core functions and has initially completed the platform transfer from traditional paper testing to computer-based paperless testing [11]. Existing test systems have a good reference role in the implementation of related functions, system data communication and test security. However, research and implementation of existing systems in terms of versatility and test aids that improve system convenience are still scarce.

2 RESEARCH METHODS

2.1 Improved KNN algorithm

In the KNN algorithm, each attribute has the same weight and takes a weight of 1. However, in reality, this is often not the case. The size of each attribute's impact on the result is different. Taking this article as an example, in the process of scoring English compositions, different attributes occupy different weights, and the length and depth of sentences and the intrinsic link of composition may have a relatively large proportion. Moreover, some attributes may have a lower weight. Therefore, if the classification is performed according to the K nearest neighbor algorithm, the effect may be less than ideal. In this paper, the attribute selection method in the decision tree—information gain is introduced as the attribute weight calculation method [12].

In information gain, the metric is to see how much information a feature can bring to a classification system. The more information it brings, the more important it is. For a feature, the amount of information will change when the system has it and without it, and the difference between the amount of information before and after is the amount of information that this feature brings to the system. The so-called amount of information is entropy [13].

The decision tree algorithm ID3 uses the information gain as an attribute selection metric. Node N represents a tuple stored in partition D. The attribute with the highest information gain is selected as the split attribute of node N. This attribute minimizes the amount of information needed to classify tuples in the result partition and reflects the minimum randomness or "impurity" in those partitions. This approach minimizes the number of expected tests required to classify an object and ensures that a simple (but not necessarily the simplest) tree is found.

The expected information required for tuple classification in D is given by equation (1) [14].

$$Info(D) = -\sum_{i=1}^{m} p_i \log_2(p_i) \tag{1}$$

Among them, p_i is the probability that any tuple in D belongs to class c_i ,The base 2 logarithmic function is used because the information is encoded in binary. $Info(D)$ is the average amount of information needed to identify the category number of the tuple in D. It should be noted that the information we have here is only the percentage of tuples in each class. $Info(D)$ is also called entropy of D.

Now, let's assume that we want to divide the tuples in D by attribute A. Among them, the attribute A has different values $\{a_1, a_2, \cdots, a_v\}$ of V numbers according to the observation of the training data. If A is a discrete value, then these values directly correspond to the output of the v number tested on A. Attribute A can be used to divide D into a subset $\{D_1, D_2, \cdots, D_v\}$ of V numbers. Among them, D_j contains the tuples in D, which have a value of h on a_j. These divisions will correspond to the branches growing from node N. Ideally, we want this partition to produce an accurate classification of the tuples, that is, we want each partition to be simple. However, most of these divisions are impure (that is, divisions may contain tuples from different classes rather than from the tuples of their respective types.). After re-division, how much information is needed to get an accurate classification is measured by equation (2) [15].

$$Info_A(D) = -\sum_{j=1}^{m} \frac{|D_j|}{|D|} \times Info(D_j) \tag{2}$$

$|D_j|/|D|$ is the weight of the j-th division. The smaller the desired information is also needed, the higher the purity of the partition. The information gain is defined as the difference between the original information requirement (ie, based on the class ratio only) and the new requirement (ie, obtained after dividing A), that is $Gain(A) = Info(D) - Info_A(D)$. In other words, $Gain(A)$ tells us how much we got through the division of A. It knows the value of A, so the expectation of information needs is reduced. The attribute A with the highest information gain $Gain(A)$ is selected as the split attribute of the node N.

The following example illustrates the process of information gain calculation. Taking the training tuple data marked by the All Electronics customer database class as an example, a training set D of a tuple that has been class-labeled is given and randomly selected from the All Electronics customer database. The category number attribute buys_computer has two different values (that is $\{yes, no\}$), so there are two different classes (that is $m=2$). Class C_1 corresponds to yes and class C_2 corresponds to no. Class yes has 9 tuples, and class no has 5 tuples. The (root) node N is created by the tuple in D. In order to find the splitting criteria for these tuples, the information gain of each attribute must be planned. First, equation (4) is used to calculate the expected information needed to classify tuples in D [16].

$$Info(D) = -\frac{9}{14}\log_2 - \frac{5}{14}\log_2\left(\frac{5}{14}\right) = 0.940 \tag{3}$$

Next, the expected information needs for each attribute need to be calculated. Starting from the attribute age, the distribution of the yes and no tuples for each class of age needs to be observed. For the youth class of age, there are 3 yes tuples and two no tuples. If the tuple is divided according to age, the expected information needed to classify the tuples in D is:

$$\begin{aligned} Info_{age}(D) = &-\frac{5}{14}\times\left(-\frac{2}{5}\log_2\frac{2}{5}-\frac{3}{5}\log_2\frac{3}{5}\right)+\frac{4}{14}\times\left(-\frac{4}{4}\log_2\frac{4}{4}-\frac{0}{4}\log_2\frac{0}{4}\right) \\ &+\frac{4}{15}\times\left(-\frac{3}{5}\log_2\frac{2}{5}\right)=0.694 \end{aligned} \tag{4}$$

Therefore, the information gain of this division is:

$$Gain(age) = Info(D) - Info_{age}(D) = 0.940 - 0.694 = 0.246 \tag{5}$$

$$\begin{aligned} &Gain(income) = 0.029, Gain(student) = 0.151, \\ &Gain(credit\_rating) = 0.048 \end{aligned} \tag{6}$$

Since age has the highest information gain in the attribute, it is selected as the split attribute. Node N is marked with age, and each attribute value grows a branch, and then the tuple is divided accordingly.

2.2 Calculate the information gain value of each attribute in article

The weights of hitting four-level words and modal verbs are smaller because they are more evenly distributed between classes. The weight of the LSA value is relatively large, so it can be seen that the LSA algorithm is very effective for measuring the intrinsic information in the composition, while the other attribute weights are relatively average. This result is not only consistent with statistical cognition, but also consistent with the results of common-sense inference. It can be inferred that if the assignment of feature weights is appropriate, the classification effect should be improved. According to the attribute values calculated in the above table, smaller and larger weights are assigned to those attributes with smaller and larger information gain values.

The distance calculation formula of Euclidean after the weight vector is introduced is shown in formula (7). $v_{i,j}$ is the weight of the j-th attribute of sample point i:

$$Sim(x_1, x_2, v_i) = \sqrt{\sum_{r=1}^{n}\left(v_{i,j}a_r(x_i) - a_r(x_j)\right)^2} \tag{7}$$

After introducing the distance calculation formula of the weight function, the classification result can indeed improve. Figure 1 shows the KNN algorithm processing flow using the weighting function.

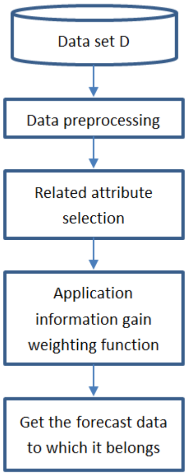

Figure 1: KNN algorithm processing flow using weighting function.

2.3 Weight the distance

In the process of using KNN algorithm to perform classification prediction, the closer the test sample point is, the more influence the accuracy of classification prediction is, especially when the data set is not balanced. However, points farther from the test sample point are likely to produce errors in the results of the classification prediction. Therefore, a significant improvement to the KNN algorithm is to weight the contributions of the k-number neighbors and assign larger weights to the nearest neighbors based on their distance from the query point x_q. Larger weights are assigned to closer neighbors. The distance is calculated by calculating the following formula (8):

$$f'(x_q) \leftarrow \sum_{i=1}^{k}\left(w_i f(x_i) \Big/ \sum_{i=1}^{k} w_i \right), w_i = 1 \big/ d(x_q, x_i)^2 \tag{8}$$

In order to deal with the case where the query point x_q exactly matches a training example x_i, which causes $d(x_q, x_i)^2$ to be 0, we set $f'(x_q) = f(x_i)$. At this time, the sample points that are very far away have little effect on $f'(x_q)$, and the sample points that are closer to each other have a great influence on $f'(x_q)$.

After weighting the attributes using the information gain method and weighting the distance from the sample points to the test points, a total weighting formula is obtained. Equation (9) gives the total weighting formula for the improved KNN. $v_{i,j}$ is the weight of the j-th attribute of sample point i:

$$f'(x_q) \leftarrow \sum_{i=1}^{k}\left(w_i \sqrt{\sum_{r=1}^{n}\left(v_{i,j} a_r(x_i) - a_r(x_j)\right)^2} \Big/ \sum_{i=1}^{k} w_i \right), w_i = 1 \big/ d(x_q, x_i)^2 \tag{9}$$

3 SYSTEM CONSTRUCTION

The development of this test system mainly relies on the Dreamweaver web page editor, and in the ASP scripting environment, the design and development of the system main interface is completed by combining Visual Basic scripting language (VBScript) and HTML code. Since the operation of the system needs to be associated with the back-end database, after weighing the commonly used database management system, Microsoft Office Access is finally selected as the database management platform. At present, there are many resources for webpage production software, but Dreamweaver is undoubtedly occupying an important position with its unique advantages. It not only supports the editing of static web pages, but also supports the production of dynamic web pages. Its "what you see is what you get" visualization feature greatly facilitates the programmer's control of the interface's effects, and cross-platform, cross-browser-restricted web page production results show its strong portability. Its powerful site management capabilities and the ability to support real-time updates of site resources ease the burden on administrators, so the development tools for this system use Dreamweaver.

Based on the previous theoretical analysis and demand analysis, each functional module of the universal English test system is designed and implemented using the B/S mode. The field test module is intended to use the B/S mode for pre-development. After the development is completed, it is properly packaged and improved through Qt, a cross-platform C++ graphical user interface application framework. Because the browser has certain test security and cheating problems during the field test, the database server uses the MySQL relational database management system. The background part is developed based on the PHP Symfony2 framework. The front-end part is developed using a combination of three web development technologies: JavaScript, CSS, and HTML.

The general English test system is divided into four parts: the management end, the teacher invigilator, the field test end, and the student end. The overall architecture of the system is shown in Figure 2.

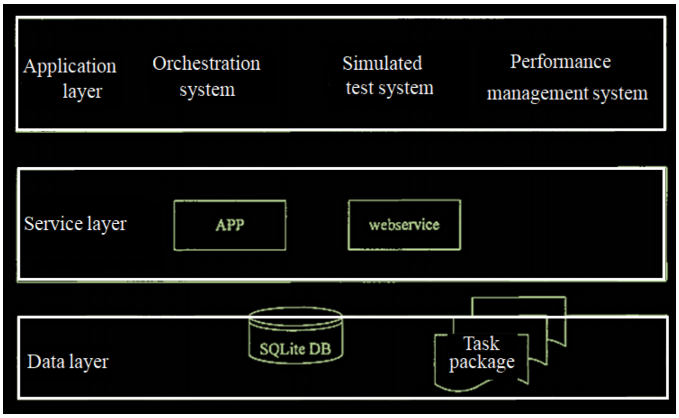

Figure 2: Overall architecture of the system.

The layout system architecture is shown in Figure 3. With the overall system architecture, a three-tier architecture model is adopted. The client UI uses WPF representation. The data layer includes an APP interface and a SQLite database. The file library includes test paper resources and task packages. In principle, one task package of the system corresponds to one test paper. The common library is Log and Common, and Log is used as a debugger. It is essential to monitor the running of the program.

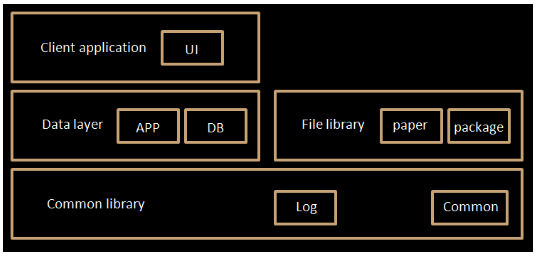

Figure 3: Architecture diagram of grade management.

Figure 4 is a physical topology diagram of the simulated test system, which is divided into examination management, invigilator (examination machine), and commentary paper.

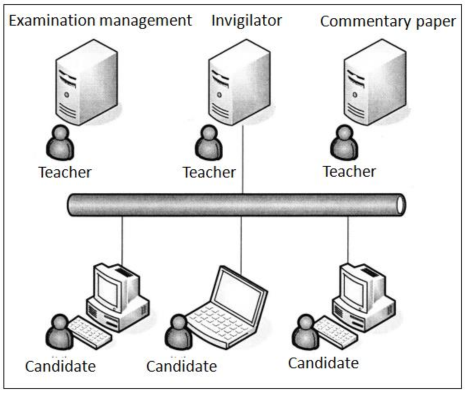

Figure 4: Physical topology of the simulated test system architecture.

System function is implemented as shown below:

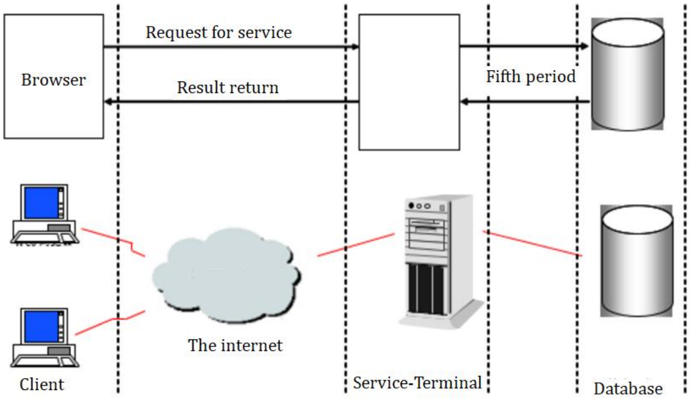

Figure 5: system function structure

Exam Paper, as shown in Table 1, contains the ID, number, and name fields of the test paper. Among them, PaperId is the primary key.

Key	Name	DataType	Notnull	Description
primarykey	PaperId	Integer		Text paper ID
	PaperNo	NVarChar[100]		Text paper number
	PaperName	NVarChar[100]	Notnull	Text paper name
	Reserved			Reserved field 1

Table 1: Exam Paper.

The paper test record table PaperRecord is shown in Table 2, including the ID of the test paper record, the ID of the test paper, and the ID of the task. The PaperRecordId is the primary key.

Key	Name	DataType	Notnull	Description
primarykey	PaperRecordId	Integer		ID of the test paper record
	PaperNo	NVarChar[100]		ID of the test paper
	taskId	NVarChar[100]		ID of the task

Table 2: Examination task record table PaperRecord.

4 ENGLISH NETWORK EVALUATION EFFECT BASED ON CADATS

CADATS: For the adaptive test, the Rasch model is used for capability estimation and topic selection during the test process. Once there is a standard question bank (ie, the difficulty factor for each topic is known), a set of tests containing several topics based on the project response theory can be designed. Its purpose is to be able to measure the level of special ability of the subject. A subset of the standard test questions can also be selected for computer-adaptive test questions. Tests based on classical response theory and project reflection theory will be a fixed number of test questions, all questions will be displayed, and the subject can return to the answers and questions. However, in the adaptive test, although the subject knows how many questions have been completed, he cannot return to the previous question.

Generally speaking, the increase or decrease of the difficulty of the topic has a great relationship with the answer of the subject. The next topic is affected by the answer to the previous question. After a certain amount of test, the ability value of the subject can be estimated. The execution diagram of the subject "huangrong" can be compared with the difficulty table of the title.

As can be seen from Figure 6, the "huangrong" title execution map is the same as the adaptive test flow chart. If the title is correct, the next one will give a more difficult problem, and if it is wrong, the next one will give a less difficult problem, and the difficulty level of the problem will gradually decrease. From the above results, it can be seen that CADATS is effective for the implementation of computer adaptive tests. In the test editing, the first question of the test will be the main test as the question with the smallest difficulty value. The CADATS software used in this study is based on the IRT model for parameter estimation. The title in the question bank is to extract the English test questions from a simulated volume in a high school in a city to do the problem analysis, and to eliminate the inappropriate questions, so the amount of questions is not large.

Figure 6: Schematic diagram of the topic execution.

The question bank also needs to be constantly updated, that is, the amount of questions is increased, and some problems of insufficient quality are constantly eliminated, which increases the construction and maintenance cost of the question bank. The test subjects were college students. Most college students are more skilled at using computers. After giving certain guidance and practice, they can conduct computer-adaptive tests well, but some students may also improperly use computers. The adaptive test of CADATS is intended to better test the students' ability, so the target is more for primary and secondary school students. For the primary and middle school students in big cities, the use of computers may be more skilled, but for some remote areas where education resources are relatively backward, the implementation of computer adaptive tests is still difficult.

The test transmission component is mainly used to implement a computer test, and the computer test is implemented by calling an executable application file (as shown in Fig. 7). This file can be copied to the network for multi-user testing and can be copied to a single machine for testing.

Figure 7: test transmission component.

Before each design of the test questions, the test questions need to be incorporated into the CADATS system. Although it is feasible to input only one or two sets of questions during a certain test, the compilation of a large number of test questions in various subjects is also time consuming and laborious. However, once the completion of the national and even the province's national test questions, it will bring great convenience to the students' actual examinations. Once all the test questions have been compiled and the adaptive test is performed, and the appropriate samples are selected, the test questions can be collected for the schools in a certain area according to the appropriate sample. However, this is also the need for professional people to make constant repairs and choices, so the difficulty is also conceivable. The update and improvement of the question bank needs to be changed according to the actual situation of education, and the collection of samples also needs to be more advanced. At present, the application of CADATS is still implemented in the examination of a certain school in a certain area. If the system is fully applied to the national examinations, there is still a long way to go.

After the successful development of the system, a small-scale test was conducted. The test results show that the system has a high level of reliability and validity. A series of empirical studies have shown that proper CAT training prior to testing can significantly improve the satisfaction of subjects with adaptive test systems. Individuals with high computer self-efficacy showed significant satisfaction with the adaptive test system. Test anxiety does not have a significant impact on individual system satisfaction, but its interaction with computer self-efficacy can have a significant impact on system satisfaction. Therefore, in the future, the promotion process of CAT can be broken from the perspective of training and self-efficacy.

5 CONCLUSION

The English online examination system can, to a large extent, demonstrate the practical application of electronic information technology in school teaching. At the same time, the basic functions of the system can meet the test management requirements of the school to a certain extent. More importantly, the system is able to maintain the safety, convenience and efficiency of the operation, thus providing a great help for the school's test management. This paper discusses in detail the overall requirements of the online English test system and carries out system design and functional module design. The CADATS software used in this study is based on the IRT model for parameter estimation. At present, the application of CADATS is still carried out in the examination of a certain school in a certain area. If the system is fully applied to the national examination, further research is needed. The purpose of this study is to use the software to conduct an adaptive test of students' English proficiency in order to test whether the software can be used in China. Another goal is to better promote the development of computer-adaptive tests and to test the performance of CADATS in exams and adaptive tests.

6 ORCID

Chunling Liu, https://orcid.org/0000-0001-5241-7623

REFERENCES

[1] Jingyu, C.; Ying, L.: Research on English pronunciation training based on intelligent speech recognition, International Journal of Speech Technology, 2018, https://doi.org/10.1007/s10772-018-9523-8

[2] Ratnam, D. J.; et al: Sense disambiguation of English simple prepositions in the context of English–Hindi machine translation system, 2018, https://doi.org/10.1007/978-981-10-6680-1_13

[3] Zhang, Y.: Research on English machine translation system based on the internet, International Journal of Speech Technology, 20(4), 2017, 1017-1022, https://doi.org/10.1007/s10772-017-9465-6

[4] Baboota, R.; Kaur, H.: Predictive analysis and modelling football results using machine learning approach for English Premier League, International Journal of Forecasting, 2018, S0169207018300116., https://doi.org/10.1016/j.ijforecast.2018.01.003

[5] Elaraby, M.; et al: Gender aware spoken language translation applied to English-Arabic, 2018

2nd International Conference on Natural Language and Speech Processing (ICNLSP), IEEE, 2018, https://doi.org/10.1109/ICNLSP.2018.8374387
[6] Chen, H. B.; et al: A simplification–translation–restoration framework for domain adaptation in statistical machine translation: A case study in medical record translation, Computer Speech & Language, 42, 2017, 59-80, https://doi.org/10.1016/j.csl.2016.08.003
[7] Bansal, S.; et al: Towards speech-to-text translation without speech recognition, 2017, https://doi.org/10.18653/v1/E17-2076
[8] Wang, L.; et al. A novel and robust approach for pro-drop language translation, Machine Translation, 31(1-2), 2017, 65-87, https://doi.org/10.1007/s10590-016-9184-9
[9] Ranade, P.; et al: Using deep neural networks to translate multi-lingual threat intelligence, 2018, https://doi.org/10.1109/ISI.2018.8587374
[10] Stig-Arne, G.; et al: The MeMAD submission to the WMT18 multimodal translation task, Proceedings of the Third Conference on Machine Translation: Shared Task Papers. 2018, https://doi.org/10.18653/v1/W18-6439
[11] Lu, Z.; Sun, J.: Presenting English polysemous phrasal verbs with two metaphor-based cognitive methods to Chinese EFL learners, System, 2017, S0346251X17306589, https://doi.org/10.1016/j.system.2017.07.016
[12] Wong, M. H. I.; Zhao, H.; Macwhinney, B.: A cognitive linguistics application for second language pedagogy: The English Preposition Tutor, Language Learning, 68(2), 2018, https://doi.org/10.1111/lang.12278
[13] Wang, X.; et al: Translating phrases in neural machine translation, 2017, https://doi.org/10.18653/v1/D17-1149
[14] Zhang, B.; et al: A Context-aware recurrent encoder for neural machine translation, IEEE/ACM Transactions on Audio, Speech, and Language Processing, 25(12), 2017, 2424-2432, https://doi.org/10.1109/TASLP.2017.2751420
[15] Mohamadi, Z.: Comparative effect of online summative and formative assessment on EFL student writing ability, Studies in Educational Evaluation, 59, 2018, 29-40, https://doi.org/10.1016/j.stueduc.2018.02.003
[16] Feng, Y.; et al: Memory-augmented neural machine translation, 2017, https://doi.org/10.18653/v1/D17-1146