Curriculum and Teaching Methodology (2023)
Clausius Scientific Press, Canada
DOI: 10.23977/curtm.2023.062017
ISSN 2616-2261 Vol. 6 Num. 20

A Study on Similarity and Difficulty Evaluation of Elementary School Mathematics Application Problems Based on Cosine Similarity and AHP

Xiaohan Zhu*,#, Huanpeng Liu#, Xin Peng#

School of Computer Science, National University of Defense Technology, Changsha, Hunan, 410000, China

**Corresponding author: 13212268601@163.com*

#These authors contributed equally.

Keywords: Elementary School Math Problems, Cosine Similarity, Hierarchical Analysis

Abstract: With the rapid development of online education, how to design a model for evaluating the similarity and difficulty of elementary math application problems has become an important research direction in the field of education. In order to measure the similarity and difficulty of elementary school math application problems, in this paper, after transforming the problems into feature vectors, the similarity of the two problems can be calculated by using measures such as cosine similarity and Euclidean distance. Then, six indicators, namely, the difficulty of the knowledge point of the problem, the difficulty of solving the problem, the logical difficulty of the problem, the linguistic difficulty of the problem, the difficulty of the picture of the problem, and the practicality of the problem, are selected, and the weights of each indicator are assigned by using the method of hierarchical analysis to determine the degree of importance of each indicator, which is used to determine the difficulty of the mathematical application problems at last.

1. Introduction

For a MOOC online education platform to enable personalized instruction and self-directed learning, it must address the following key questions:

How to measure the similarity between topics?

In order to be able to automatically recommend other topics with similar question types for extended practice based on students' learning and question answering, it is necessary to analyze the difficulty and similarity of the topics.

A measure of similarity between topics is required. There are several ways to measure the similarity between topics, such as text similarity algorithm (TSC), which compares keywords and syntax in topics.

There are many ways to measure the similarity between topics, such as text similarity algorithm to compare the keywords and grammatical structures in the topics, or graphical neural network to compare the structure and relationship between topics.

The most appropriate method must be chosen according to the actual situation.

How to evaluate the difficulty of the topics?

In order to be able to recommend the right level of difficulty of the practice questions according to the students' learning level and answer situation, the difficulty of the questions needs to be evaluated.

Evaluate the difficulty of the questions. There are many ways to evaluate the difficulty of the questions, such as using the AHP model to assign weights, or using the BP neural network to predict the difficulty.

There are various methods to evaluate the difficulty of questions, such as using AHP model to assign weights or using BP neural network to predict the difficulty. The most appropriate method must be chosen according to the actual situation.

After a survey of the results of previous studies, it was found that there is an approach to analyze mathematical topics using the Beltran pre-trained Iransformer model and dataset expander, the advantage of this approach is that using the Beltran pre-trained Iransformer model, its multiple attention mechanism can be utilized to capture long-distance dependency and denotative relationships, which is very important for solving elementary school level mathematical application problems is very important. This model has made great strides in natural language processing, so one can expect better results in the analysis of math application problems as well.

Extensive expansion of the MAWPS dataset using the dataset expander can increase the diversity of the training data and improve the generalization ability of the model. This is very helpful in solving different types of math application problems and can improve the accuracy and robustness of the model in various situations. However, there are also the following problems, the method relies on Beltogen to pre-train the Iransformer model, which means that a lot of computational resources and time are needed for training and pre-training the model. This may be a challenge for some resource-limited environments.

The effectiveness of the dataset expander depends on the quality of the expansion algorithm and the diversity of the dataset. If the expansion algorithm is not accurate enough or the diversity of the dataset is not wide enough, it may lead to poor quality of the expanded dataset and affect the performance of the model.[1]

It is also possible to use the method of BP neural network for mathematical problem assessment, [2] but the model has artificially set assessment criteria. Since the assessment criteria are set by human beings, they may be influenced by subjective factors, leading to subjectivity and inconsistency in the assessment results. This means that different people may give different assessment results based on their subjective judgment, which affects the reliability and objectivity of the assessment.[3]

In terms of the treatment of similarity, the method of measuring the similarity of two elementary school math application problems needs to take into account factors such as the content, the solution method, and the difficulty of the questions. Designing an appropriate measurement method is the key.

Among the assumptions of similarity

a. it is believed that the similarity of two topics can be measured by comparing the content of the text, the solution method and the degree of difficulty of the two topics

Degree of similarity.

b. It is assumed that after transforming the questions into feature vectors, using measures such as cosine similarity and Euclidean distance, the degree of similarity between the two questions can be calculated.

Similarity between the two questions.

In the difficulty analysis of application problems, the diversity and difficulty level of the problems are considered. The model should be able to handle different types of topics and differentiate them according to the difficulty.

Among the assumptions of difficulty

a. it is assumed that topic difficulty can be measured by factors such as topic length, keywords, and solution steps.

b. it is assumed that mathematical models, such as regression models or neural networks, can be constructed to predict the difficulty of a question.[4]

2. Main content

2.1 Modeling and solving for topic similarity

Cosine similarity is a common similarity measure that can be used to describe the similarity between two elementary school math application problems. The steps are as follows:

Create feature vectors: the extracted mathematical features are formed into feature vectors. The feature vector can be represented by a series of numbers, and each number represents the weight of a mathematical feature.

Calculate cosine similarity: Calculate the cosine similarity between the feature vectors of two topics. The cosine similarity is calculated as follows:

Cosine similarity = (A-B)/(||A||||B||)

where A and B are the feature vectors of the two problems, ||A|| and ||B|| denote the modulus of A and B, and - denotes the dot product of the vectors.

Describe the similarity metric: describes the degree of similarity between two topics based on the calculation of the cosine similarity. The value of cosine similarity ranges from [-1, 1]. If the cosine similarity is 1, it means that these two issues are exactly the same; if the cosine similarity is 0, it means that there is no similarity between the two subjects; if the cosine similarity is -1, it means that these two issues are exactly opposite.[5-7]

It should be noted that when constructing the feature vectors, the features need to be appropriately normalized to avoid the effect of too large or too small feature weights on the cosine similarity calculation results.

Cosine similarity is a measure of similarity between two vectors. The basic idea is to represent the vectors as vectors in a multidimensional space and then calculate the cosine of the angle between them. The closer the cosine is to 1, the more similar the two vectors are.

For two elementary school math application problems, we can consider them as vectors, extract the key information of each problem as the dimensions of the vectors, and then use the cosine similarity to calculate the similarity between them.

The following is the exact process of creating and solving the model:

Topic vector representation

The key information for each question is extracted and weights are assigned according to its importance in the question. For example, for a question involving area calculation, we can take the area calculation formula as the most important information and assign it the highest weight. Other important information can be weighted according to the actual situation. Finally, we can represent each problem as a vector, the dimension of the vector is the amount of key information, and the value of each dimension represents the weight of the corresponding information in that dimension.

Calculation of cosine similarity

Calculate the cosine similarity between problem vectors as follows:

Normalize the two problem vectors by dividing each vector by its length to make them unit vectors.

Compute the dot product of the two vectors.

Compute the length product of the two vectors.

Divide the dot product by the length product to get the cosine similarity of the two vectors.

Model Solution

Compute the cosine similarity in pairs with all problem vectors to get the similarity matrix. You can represent the matrix as an undirected graph, where each vertex represents a problem and the weights of the edges indicate the similarity between the corresponding problems. Such undirected graphs can be processed analytically with graph-theoretic algorithms such as clustering algorithms to discover underlying problem organization and patterns of similarity.

In conclusion, cosine similarity is a simple and effective metric that can be used to describe the similarity between elementary math application problems. By representing the topics as vectors and calculating the cosine similarity between the topics, the similarity matrix between the topics as well as the similarity patterns, organizational structure and other useful information can be obtained.

2.2 Modeling and solving for topic difficulty

Evaluating the level of difficulty of elementary school mathematics application problems can be done by building a mathematical model and using the hierarchical analysis method (AHP) to assign weights to each indicator, thus determining the level of importance of each indicator. The following model and six indicators:

Difficulty of the topic's knowledge points: this indicator measures the difficulty of the knowledge points involved in the topic. For example, a problem involving the multiplication of decimals may be more difficult than a problem involving the addition of whole numbers. This indicator can be measured by the difficulty of the knowledge points involved in the topic.

Problem Solving Difficulty: this indicator measures the difficulty of solving the problem, i.e., how much time and effort the student needs to spend to solve the problem. For example, problems that require multi-stage calculations or require the use of specific problem-solving skills may be more difficult. This indicator can be measured in terms of the computational difficulty of the topic, the complexity of the problem-solving skill, and the difficulty of the reasoning.

Logical difficulty of the topic: this indicator measures the logical difficulty of the topic, i.e., how much reasoning and analysis is required for students to solve the problem. For example, questions that require multiple levels of reasoning or consideration of multiple conditions may be more difficult. This indicator can be measured by the logical complexity of the question, the complexity of the conditions, and the difficulty of the reasoning.

Linguistic Difficulty of the Question: This indicator measures the linguistic difficulty of the question, that is, whether the language used in the question is easy to understand. For example, questions with difficult vocabulary or grammatical structures may be more difficult. This indicator can be measured by the linguistic complexity, vocabulary, and grammatical difficulty of the topic.

Picture Difficulty of the Topic: this indicator measures the picture difficulty of the topic. For example, it may be difficult to understand complex diagrams or solve problems that require geometric reasoning. This metric can be measured by the image complexity of the topic, geometric reasoning difficulty, and data interpretation difficulty.

Utility of the topic: This indicator measures the utility of the topic, i.e., whether the question is relevant to real life. For example, problems that deal with everyday life or social issues may be more practical. This indicator can be measured in terms of the practical application, relevance and social concern of the topic.

The establishment of hierarchical analytical modeling, or AHP for short, refers to a decision-making method that decomposes the elements that are always relevant to decision-making into objectives, guidelines, options, etc. And on this basis, qualitative and quantitative analysis is carried out to provide reference for decision makers to choose the optimal resource isolation program.

The steps of hierarchical analysis method are as follows:

(1) Establish a hierarchical structure, goals, factors, decision-making objects can be divided into the highest level, middle level and bottom level according to their relationship, and draw a

hierarchical diagram. After a detailed analysis of the problem, identify the factors in it and determine the relationship and affiliation between the factors.

(2) Construction of judgment matrix: when determining the weights of various factors at different levels, it is often not easy to be accepted by others if the results are only qualitative. Therefore, a unified matrix method is proposed, which does not compare all factors together, but compares them in pairs. At this point, a relative scale is used, which minimizes the difficulty of comparing the various factors and improves accuracy. The general form of the relevant Arab judgment matrices is shown in the table 1.

Table 1: General form of the Arab judgment matrix

A	B_1	B_2	……	B_n
B_1	b_{11}	b_{12}	……	b_{1n}
B_2	b_{21}	b_{22}	……	b_{2n}
……	……	……	……	……
B_n	b_{n1}	b_{n2}	……	b_{nn}

In the judgment matrix. Usually a nine-point scale is used, which is determined after reviewing a large amount of literature, soliciting expert opinions, and repeated research. The results of the comparison of two factors are known. The comparison is made by quantifying the importance of each factor. When you compare two things, you can best compare their strengths and weaknesses. The use of relative scales avoids the difficulty of comparing factors of different natures to each other and improves accuracy. The so-called judgment matrix is a comparison of the relative importance of all the factors in this layer to the relative importance of a factor in the previous layer. The elements of the judgment matrix are given by Santy's 1-9 scale, the meaning of which is shown in the table 2.

Table 2: Arab nine-point standard system and its definitions

resizing	**significance**
1	suggests that both factors are equally important in the comparison.
3	Indicates that one factor is slightly heavier than the other when compared to two factors.
5	Indicates that one factor is significantly more important than the other when compared to the two factors.
7	Indicates that one factor is more important than the other when comparing two factors.
9	Indicates that one factor is more important than the other when compared to two factors
2,4,6,8	The median of the two neighboring judgments above
Count down	The judgment matrix for comparing factors i and j is a_{ij} , Then the judgment matrix for comparing factors j and i is a_{ji} $a_{ji} = 1/a_{ij}$

(3) Calculate the weight of each element: by calculating the judgment matrix (consistency matrix), the weight of all the elements of the layer on the relevant elements of the previous layer is calculated, and then the weight of the elements of the previous layer is further synthesized by using the calculation results of the weight of the single layer. The order of the elements is obtained through the calculation results.

Determination of the scope of factors to be evaluated P

An evaluation indicator, $u = \{u_1, u_2, \cdots\cdots, u_p\}$.

Setting the annotation hierarchy scope

$v = \{v_1, v_2, \cdots\cdots, v_p\}$. That is, a collection of layers. Each level can correspond to a fuzzy subset.

Construction of fuzzy relationship matrix R

After constructing the hierarchical fuzzy subsets, we need to evaluate the evaluated things one by one from each factor. $u_i\left(i=1,2,\cdots\cdots,p\right)$ quantize. That is, the affiliation of the thing being evaluated in the hierarchical fuzzy subset is determined from a single-factor perspective. $\left(R\,|\,u_i\right)$ Then we get the fuzzy relationship matrix:

$$R=\begin{bmatrix} R| & u_1 \\ R| & u_2 \\ & \cdots \\ R| & u_p \end{bmatrix}=\begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1m} \\ r_{21} & r_{22} & \cdots & r_{2m} \\ \cdots & \cdots & \cdots & \cdots \\ r_{p1} & r_{p2} & \cdots & r_{pm} \end{bmatrix}_{p.m} \tag{1}$$

The matrix R intermediate stage i first row. j column element indicates that the thing being evaluated depends on a factor r_{ij}. The v_j degree of affiliation of a subset of hierarchical fuzzy. A thing being evaluated lies on a particular factor. The aspect is represented by a fuzzy vector u_i. $\left(R\,|\,u_i\right)=\left(r_{i1},r_{i2},\cdots\cdots,r_{im}\right)$ In other evaluation methods, it is mainly characterized by the actual value of a particular indicator, so from this point of view, fuzzy comprehensive evaluation needs more information

Determine the vector of evaluation factor weights

In fuzzy comprehensive evaluation, determine the weight vector of the evaluation factors: $A=\left(a_1,a_2,\cdots\cdots,a_p\right)$. The element a_i in the weight vector A It is essentially a factor. a_i is the degree of affiliation. In this paper, the hierarchical analysis is used to determine the relative importance ranking between the evaluation indicators. Therefore, the weight factors are determined and normalized before synthesis.

$$\sum_{i=1}^{p} a_i=1 \;,\; a_i\geq 0 \;,\; i=1,2,\cdots\cdots,n \tag{2}$$

Fuzzy composite judgment result vector

The fuzzy comprehensive evaluation result vector B for each evaluated object is obtained by synthesizing using the appropriate operator A as well as the thing being evaluated R. This means that:

$$A\circ R=\left(a_1,a_2,\cdots\cdots,a_p\right)\begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1m} \\ r_{21} & r_{22} & \cdots & r_{2m} \\ \cdots & \cdots & \cdots & \cdots \\ r_{p1} & r_{p2} & \cdots & r_{pm} \end{bmatrix}=\left(b_1,b_2,\cdots\cdots,b_m\right)=B \tag{3}$$

In between them. b_1 is the first j a column operation of A and R indicating that the thing being evaluated is generally correct. v_j denotes the degree of affiliation of the hierarchical fuzzy subset.

In practice, the most commonly used method is the principle of maximum affiliation, but in some cases a large amount of information can be lost very unwillingly or even unreasonable evaluation results can be obtained. A method of calculating affiliation using weighted averages is proposed, which makes it possible to rank multiple evaluated objects in terms of their hierarchical position.

This means that the difficulty of the knowledge points of the topic is the most important indicator for judging the difficulty of elementary school math application problems, followed by the difficulty of solving the problem and the logical difficulty. The linguistic difficulty, figurative difficulty and practicality of the topic are of relatively low importance.[8-10]

3. Conclusion

In elementary school math application problems, the level of difficulty and similarity of the topics is often a major concern for students. In order to characterize the similarity between two elementary math application problems, the cosine similarity can be used to measure the similarity between the topics. Cosine similarity is a similarity measure based on a vector space model that can be used to calculate the cosine value of the angle between two vectors. The larger the value, the more similar the two vectors are.

Specifically, suppose the keyword vector of topic A is denoted as a=(a1, a2, ... , an) and the keyword vector of topic B is denoted as b = (b1, b2, ... , bn), then the cosine similarity between them can be defined as: cos(a, b) = (a-b) / (||a|| ||b||) where - denotes the dot product operation of vectors, and ||a|| and ||b|| denote the modes of vectors, respectively. In order to evaluate the level of difficulty of elementary school mathematics, a mathematical model can be developed using the hierarchical analysis method (AHP). Hierarchical analysis is a multilevel decision analysis method that can be used to solve multifactor and multilevel decision problems.

In particular, assume that the difficulty of question A is a and the difficulty of question B is b. The difficulty relationship between them can be defined as: a/b=w1/w2, where w1 and w2 are the weights of question A, respectively. Question B is presented in the difficulty assessment, which can be calculated by the AHP model.

Here's a breakdown of the model's strengths, weaknesses, and growth prospects.

The AHP method is mainly based on the evaluator's understanding of the nature and elements of the evaluation problem, and is more concerned with qualitative analysis and judgment than the general quantitative method.

The disadvantage of AHP is that the establishment of weights is too subjective.

The innovation of this paper is to grasp the connectivity of the problem, to consider the difference with the original model every time a new model is established, and to compare the new model with the actual problem after it is given, and to carry out the error assessment and correction in time, so that the established model is more generalizable.

References

[1] Mingzhe Li, Analysis of Elementary Math Word Problems Based on AI Deep Learning[J]. Academic Journal of Mathematical Sciences, 2023, 4(3)

[2] WU Man-yi, XU Liang-ji, ZHANG Kun. Research on the projected parameter seeking of probabilistic integration method based on SSA-BP neural network[J/OL]. Metal Mining: 1-10 [2022-05-03].

[3] ZHAO Lixing, HOU Xingming, XU Zhaowen, and LIN Zi. Equipment maintenance capability assessment based on GA-wavelet-BP neural network[J]. Modern Defense Technology, 2022, 50(02):84-95

[4] Jiang Qiyuan et al. Mathematical modeling [M]. Beijing: Higher Education Press, 2003.

[5] WANG Xinmin, LI Weiying, ZHOU Yu, JING Xinyu, CHEN Sheng. Evaluation of surface water quality based on random forest and fuzzy comprehensive evaluation. [J]. Water Supply and Drainage, 2022, 58(02):128-132.

[6] YANG Kun-Yu, PAN Jian, LUO Xian-Ting. Evaluation of multifunctional marching stool design scheme based on fuzzy comprehensive evaluation method[J]. Industrial Design, 2022(01):158-160.

[7] Gao Feng, Wu Dan. Study on the application of cost control evaluation based on fuzzy comprehensive evaluation method--Taking the teaching building of A middle school in Tianjin as an example [J]. A Middle School Teaching Building in Tianjin [J]. Real Estate World, 2022(01):12-14.

[8] WANG Guinan, SU Xianhe, SHANG Guojing, FAN Jiwei, BIAN Xian. State evaluation of generator excitation system based on fuzzy comprehensive evaluation Modeling and application[J]. Jilin Electric Power, 2021, 49(06):18-21.

[9] JIN Yang, YU Minjian, YU Lu. Research on route generation assistance decision-making based on fuzzy comprehensive evaluation [J]. Science and Technology Innovation Science and Technology Innovation, 2021(34):73-75.

[10] Zhong Hongxing. Fuzzy comprehensive evaluation of construction quality of DaDawang gate[J]. Heilongjiang

Water Conservancy Science and Technology, 2021, 49(11):221- 225.