# A Comparative Study of Different Information Entropy Index in Personalized Exercise Recommendation

Qiulei Zheng[1], Huifan Gao[1], Fan Yang[1], Yinghui Pan[2] and Yifeng Zeng[3*]
[1]*Department of Automation, Xiamen University*
[2]*College of Computer Science and Software Engineering, Shenzhen University*
[3]*Department of Computer and Information Sciences, Northumbria University*

***Abstract*—It is meaningful to recommend exercises to students in an online education system with a large amount of learning resource. Many recommendation methods usually rely on strategy in the recommendation system in order to predict an exercise score. In this paper, we compare different information entropy index in the exercise recommendation. These index consider how well exercise matches the student's knowledge ability. In order to compare different methods, we introduce an interactive platform for dynamic exercise recommendation. We conduct a set of exercise recommendation experiments, compare the effect of different index and find the optimal index in the experiment.**

***Index Terms*—exercise recommendation, information entropy index, online education system**

## I. Introduction

In an online education system, it is necessary to recommend learning resource to students to consolidate and improve the knowledge ability [1]. To achieve this goal, recommendation system plays an important role, helping to find suitable exercises from a large number of learning resources to meet different demand. With the development of online educational system, different exercise recommendation methods have been proposed. They provide convenience for students to obtain learning resource from online educational system, which is exactly the purpose of recommendation system [2].

Many exercise recommendation methods usually imitate recommendation system, by analyzing answering record, predicting students' response on exercises, and then recommending exercises to students [3] [4]. However, these methods exist shortcoming in two respects. First of all, one-time exercise recommendation ignores the feedback of students after completing these exercises. Secondly, different from recommendation system, the score of exercises cannot be considered as a recommendation index. Obviously, these methods have shortcomings in educational system.

In order to solve problems in these methods, we introduce a recommendation method based on information entropy of exercise. It does not only rely on predicting scores on these exercises, but also recommend the most suitable exercises of students' knowledge ability. To compare the effect of method, a simulation test system is applied to recommend exercise based on method. Instead of setting the difficulty threshold of exercises, methods try to find exercises that are most suitable for the student's current knowledge level. We find the optimal index in the experiment in exercise recommendation. We hope to reduce the burden of students, and achieve the purpose of the recommendation system.

The information entropy index of exercises can provide a method for the selection of exercises. In this way, we can recommend exercises that are most suitable for student's knowledge ability. In terms of metrics, a dynamic recommendation model satisfies requirement in exercise recommendation scenery. At the same time, exercise recommendation aims to help students find knowledge ability with as less exercises as possible.

The rest of the paper is structured as follows. A brief introduction of background is given in Section 2. In Section 3, we describe methodology in detail, including five different recommendation indices. The experiment and result analysis are provided in Section 4. Finally, we conclude the paper in Section 5.

## II. Background

### A. Exercise Recommendation

Recommendation system has been successfully applied in many fields. For example, there are abundant resource in the e-commerce field. Recommendation system can bring convenience to user of different demand [5]. In the education field, with abundant exercise interactive record in the online education system, scholars have applied many recommendation strategies to recommend exercise, including collaborative filtering, content-based recommendation and hybrid recommendation method. Specifically, the content-based recommendation system provides students with exercises that are not done well in the past [6]. Collaborative filtering methods find similar behavior to make decision for students [7]. The hybrid method combines the advantages of the content-based recommendation system and collaborative filtering method while considering the influence of other factors [8].

### B. Cognitive Diagnosis

An important step in the online education system is cognitive diagnosis. With cognitive diagnosis, analyzing student's knowledge ability can guide student in next study task. Cognitive diagnosis can be divided into two categories, linear models and nonlinear models. The typical nonlinear model, *Item Response Theory (IRT)* model, regards students as a potential characteristic variable and uses a single value as the comprehensive ability of students [9]. In addition, the typical nonlinear model, *Deterministic Inputs, Noisy-And gate (DINA)*, treats students as a binary vector to indicate whether they have mastered each knowledge point [10]. Compared with IRT, DINA considers more exercise information like knowledge point. In this paper, we select DINA as cognitive diagnostic model.

DINA is a kind of nonlinear cognitive diagnosis model. DINA reflects the response probability, where the parameters *s* represents probability of response to exercise incorrectly and and *g* represents probability of response to exercise correctly by guessing. The response probability of correct response to exercise is as (1).

$$P(X_{ij} = 1 \,|s_j, g_j\,) = (1 - s_j)^{\xi_{ij}} {g_j}^{(1-\xi_{ij})} \tag{1}$$

The detail of $\xi_{ij}$ in (1) is as shown in (2), where $\alpha_{ik}$ indicates whether student $i$ has mastered knowledge point *h*(The value is 0 or 1) and $q_{jk}$ indicates whether exercise $j$ involves the knowledge point $h$ (The value is 0 or 1). Then the probability of response to exercise *q* of knowledge vector is as (2).

$$\xi_{ij} = \prod_{k}^{K} \alpha_{ik}^{q_{jk}} \tag{2}$$

The knowledge vector in DINA is discrete values, 0 and 1. Other factors considered by the DINA model include the guessing rate and error rate of response to exercise. The model is designed to calculate the probability of answering exercise correctly. When the probability is calculated to beyond 0.5, we will judge the response to exercise is correct. If the probability is calculated to below 0.5, we will judge the response to exercise is incorrect. In this way, we can simulate the process of students answering the exercise.

## III. Methodology

We calculate information entropy of exercise recommendation index, and recommend the exercise with highest information entropy index to students. The following section shows five different information entropy of exercise.

### A. KL Index

*Kullback-Leibler(KL)*, also known as relative entropy, is a measure of distance between two probability distributions [11].

$$KL[f(x), g(x)] = E_f \left[ \log \frac{f(x)}{g(x)} \right] \tag{3}$$

In (3), *f(x)* represents the true distribution of the data, and *g(x)* represents the approximate distribution for *f(x)*. In other words, the larger KL index, the greater difference between *f(x)* and *g(x)* [12].

With KL index of exercise, we hope to find exercise *h* that can reflect knowledge vector of student *i*. The knowledge vector is one-dimensional vector, and each dimension represents a knowledge point. When students master the knowledge point, the value of the knowledge dimension is 1, otherwise it is 0. The larger KL index, it means that this exercise is better for student *i*. The current knowledge vector of student *i* is calculated based on answering record. $P(X_{ih} = q)$ represents the correct probability of the knowledge vector $\alpha_i$ response to exercise *h*.

In exercise recommendation , $P(X_{ih} = q)$ represents *f(x)*, $P(X_{ch} = q)$ represents *g(x)*. Cognitive diagnosis model DINA is used to calculate the probability $P(X_{ih} = q)$ and $P(X_{ch} = q)$. Therefore, we can find the effect of exercise *h* in distinguishing the knowledge vector $\alpha_i$ and the knowledge vector $\alpha_c$. The formula is as

$$D_h(\alpha_i \,\|\alpha_c\,) = \sum_{q=0}^{1} \log \left( \frac{P(X_{ih} = q)}{P(X_{ch} = q)} \right) \cdot P(X_{ih} = q). \tag{4}$$

After considering all knowledge vectors (There are *K* knowledge points, then there are $2^K$ knowledge vectors), KL index of exercise *h* is

$$KL_h(\alpha_i) = \sum_{c=1}^{2^K} D_h(\alpha_i \,\|\alpha_c\,). \tag{5}$$

Choosing the exercise with the largest KL index is to choose the exercise that best suits the student's current knowledge ability, so exercise is recommended to the students.

### B. PWKL Index

When answering record is obtained, the posterior probability of each knowledge vector is different. Therefore, the posterior probability of each knowledge vector can be calculated through exercises. KL is upgraded to *Posterior Weighted Kullback-Leibler (PWKL)*. The posterior weighted value is calculated for each knowledge vector. The PWKL index of exercise *h* is

$$PWKL_h(\alpha_i) = \sum_{c=1}^{2^K} D_h(\alpha_i \,\|\alpha_c\,) \cdot P(\alpha_c \,|Y_n\,). \tag{6}$$

In (6), $Y_n$ represents answering record, which is the posterior probability calculated based on answering record $Y_n$. After adding the posterior probability, the weight of each knowledge vector is different.

### C. HKL Index

Difference between knowledge vectors should be considered. Some knowledge vectors have a greater degree of difference, while others are smaller. Euclidean distance is used to calculate the difference between knowledge vectors.

$$d(\alpha_i, \alpha_j) = \frac{1}{K} \sum_{k=1}^{K} \sqrt{|\alpha_{ik} - \alpha_{jk}|^2}. \tag{7}$$

In (7), the dimension of each knowledge vector is *K*. Adding the difference between knowledge vector $\alpha_i$ and knowledge vector $\alpha_c$ will have an impact on PWKL index and obtain *Hybird Kullback-Leibler (HKL)* index in (8).

$$HKL_h(\alpha_i) = \sum_{c=1}^{2^K} D_h(\alpha_i \| \alpha_c) \cdot P(\alpha_c | Y_n) \cdot \frac{1}{d(\alpha_c, \alpha_i)} \quad (8)$$

### D. MI Index

Considering two random variables, *X* and *Y*, *Mutual Information (MI)* is defined as the joint distribution *f(X,Y)* and the KL divergence distance of the product of the edge distribution *f(X)* and *f(Y)* [13]. Therefore, the mutual information of random variables *X* and *Y* is as (9).

$$I(X,Y) = \sum_x \sum_y f(x,y) \cdot \log\left[\frac{f(x,y)}{f(x)f(y)}\right] \quad (9)$$

Mutual information is used to measure the degree of dependence between *X* and *Y*. When *X* carries more useful information about *Y*, and the larger *I(X,Y)* is. The MI index is applied to exercise recommendation, *f(x)* is regarded as the posterior probability after the completion of *n-1* exercises, and *f(y)* is regarded as the probability of answering the nth exercise. Then the mutual information here refers to the information gain for the uncertainty of the student's knowledge point vector after adding a new exercise, the larger the better. According to the derivation in the paper, calculating the maximum mutual information of these two random variables is equivalent to calculating KL index of the two random variables in (11).

$$\begin{aligned} &KL(\pi(\alpha | y_{n-1}, y_n) \| \pi(\alpha | y_{n-1})) \\ &= \sum_{c=1}^{2^K} \pi(\alpha_c | y_{n-1}, y_n) \cdot \log \frac{\pi(\alpha_c | y_{n-1}, y_n)}{\pi(\alpha_c | y_{n-1})} \end{aligned} \quad (10)$$

Since the result of exercise is unknown, we turn to calculate the maximum expected mutual information as (11).

$$\begin{aligned} KL(\pi(\alpha | y_{n-1}, y_n) \| \pi(\alpha | y_{n-1})) &= \sum_{y=0}^{1} p(Y_n = y | y_{n-1}) \\ &\cdot \left[\sum_{c=1}^{2^K} \pi(\alpha_c | y_{n-1}, y_n) \log \frac{\pi(\alpha_c | y_{n-1}, y_n)}{\pi(\alpha_c | y_{n-1})}\right] \end{aligned} \quad (11)$$

### E. SHE Index

SHE index stands for Shannon Entropy [14]. Shannon entropy is abbreviated as entropy, which represents the uncertainty of a probability distribution. The Shannon entropy formula of discrete distribution is as (12).

$$SHE = -\sum_i (p_i \cdot \log p_i) \quad (12)$$

The Shannon entropy is applied to the exercise recommendation, and expected Shannon entropy is set as the amount of information of the exercise. The posterior probability of knowledge vector after the student *i* has done n-1 exercises, the prior probability of the knowledge point vector is $\pi_{i,n-1}(\alpha_l)$, and the probability that the knowledge point vector answers question *t* correctly is $P_{lt} = P(Y_{jt} = 1 | \alpha_j = \alpha_l)$, then the knowledge point vector posterior probability formula is:

$$\pi_{i,n-1}(\alpha_l) \propto \lambda_l \prod_{t=1}^{n-1} (P_{lt})^{Y_{it}} \cdot (1 - P_{lt})^{1-Y_{it}}. \quad (13)$$

Combined with the Shannon entropy, the Shannon entropy of the distribution can be written as

$$SHE(\pi_{i,n-1}(\alpha_l)) = -\sum_{l=1}^{2^K} \pi_{i,n-1}(\alpha_l) \cdot \log_b(\pi_{i,n-1}(\alpha_l)). \quad (14)$$

When recommending next exercise, expected Shannon entropy information of exercise *h* is

$$SHE_h(\pi_{i,n}) = \sum_{z=0}^{1} \left[SHE(\pi_{i,n}, Y_{ih} = z) \cdot P(Y_{ih} = z \left| Y_i^{n-1} \right.)\right]. \quad (15)$$

We recommend the exercises that the smallest excepted Shannon entropy to students. After students have done these exercises, they can estimate knowledge ability more accurately with less exercises.

## IV. Experiment And Result

### A. Exercise Recommendation Experiment

Exercise recommendation experiment is as follows. Students complete the first exercise. Then we estimate knowledge ability according to the answering record, and select the next exercise from the exercise bank. When students complete this exercise, repeat the process. When the amount of exercises reaches a fixed value, stop recommending and get the final student knowledge ability. The method of selecting exercises is exercise recommendation index in the paper. The exercises selected by different indices are different. After students have completed all exercises, we estimate their knowledge ability. We compare the estimated knowledge ability with the real knowledge ability to get the accuracy rate. The higher the accuracy rate, the better the discrimination of exercises. Therefore, exercise recommendation index of producing these exercises is better.

Before exercise recommendation experiment, we need to generate students and exercise bank, set knowledge vector evaluation method, length of test, exercise recommendation indices and Knowledge vector evaluation.

- Students generation. Students generation is based on such assumption that each student has a 50% probability of mastering each knowledge point. For example, for a 6-knowledge point scenery, 64 knowledge points vector are equally likely appeared. 100 students are generated in the experiment, and it is equally possible for each student to master each knowledge vector.

- Exercise bank generation. Exercises are generated based on such a assumption that each exercise has a 20% probability of involving each knowledge point.
- Knowledge Vector Evaluation. The knowledge vector is one dimension, and each dimension represents a knowledge point. The value of each knowledge point is a discrete value of 0 or 1. Estimating the student's knowledge vector based on answering record by using the maximum likelihood estimation method.
- Length of test. The experiment is a test of fixed length, and different test lengths are set to check the effect of different exercises recommendation indices. The length of test is set to six, eight, ten, and twelve respectively.
- Exercise recommendation indices. In order to compare different exercise recommendation indices, simulation exercise recommendation is designed. Exercise recommendation indices include KL index, PWKL index, HKL index, SHE index, and MI index.
- Knowledge vector evaluation. In order to compare the effect of different exercise recommendation methods, accuracy is designed. After students finish all exercises, model estimate their knowledge vector based on answering records, and compare knowledge vector predicted with the real knowledge vector.

### B. Metrics

The accuracy of model is divided into attribute accuracy and pattern accuracy. Pattern represents knowledge vectors like [1,0,0,1,1], and attribute in pattern represents each knowledge point (the first 1 in vector represents the first knowledge point).

*Pattern Correct Classification Rate (PCCR)* is based on whether the pattern estimated is the same as the real pattern. For example, if the student's true pattern is [1,0,1,0,0] and pattern estimated is [1,0,1,0,0], this prediction is accurate. The formula of PCCR is shown in (16), where *M* represents the number of students, *I* is an indicator function, $\alpha_i$ represents true pattern and $\widehat{\alpha}_i$ represents pattern estimated.

$$PCCR = \sum_{i=1}^{M} I(\alpha_i = \widehat{\alpha}_i)/M. \tag{16}$$

*Average Attribute Correct Classification Rate (AACCR)* is based on whether the $k$ dimension of pattern estimated is the same as the $k$ dimension of real pattern. The attribute accuracy is a one-dimensional vector, and each dimension represents each attribute. The average attribute accuracy is the average value of accuracy of each attribute. The formula of AACCR is as

$$AACCR = \frac{1}{K}\sum_{k=1}^{K}(\sum_{i=1}^{M} I(\alpha_{ik} = \widehat{\alpha}_{ik})/M). \tag{17}$$

Where *M* represents the number of students, and *K* represents the number of attributes. In order to reveal the difference of exercise recommendation indices, total time of experiment is set as metric. In the experiment, the effect of different index is compared in PCCR, AACCR and Time in each exercise recommendation experiment.

### C. Experiment Result

In the first experiment, the amount of exercises in the test is set to six. The experiment are based on five exercise recommendation indices, and time of test is also set as a metric. It can be seen from Table I that the MI index gets better accuracy in both PCCR and AACCR, which recommend better exercises to find students' knowledge ability. Compared with KL, PWKL and HKL, SHE index has a certain improvement in both PCCR and AACCR.

TABLE I
EXPERIMENT RESULT

| Index | PCCR | AACCR | Time* |
|---|---|---|---|
| KL | 20 % | 80 % | 2.72 |
| HKL | 18 % | 79.7 % | 5.38 |
| PWKL | 27 % | 82.2 % | 6.27 |
| SHE | 39 % | 88.3 % | 23.51 |
| MI | 84 % | 97.3 % | 20.7 |

*The unit of time is minute.

In the second experiment, the amount of exercises in the test is set to eight. The result is shown in Table II. Three indices, including KL, PWKL, HKL, are quite different from SHE index and MI index in PCCR and AACCR. It is obvious that MI index is not only less time-consuming than SHE index, but also high accuracy in PCCR and AACCR than other indices.

TABLE II
EXPERIMENT RESULT

| Index | PCCR | AACCR | Time* |
|---|---|---|---|
| KL | 41 % | 87.7 % | 3.47 |
| HKL | 43 % | 88.5 % | 6.95 |
| PWKL | 42 % | 88.2 % | 7.81 |
| SHE | 92 % | 98.7 % | 29.82 |
| MI | 100 % | 100 % | 26.45 |

*The unit of time is minute.

In the third experiment, the amount of exercises in the test is set to ten. Experiment result is shown in Table III. Among three indices, including KL, PWKL and HKL, the effect of PWKL index is better than KL index and HKL index, but the effect is quite different from SHE index and MI index. When length of test is set to ten, the accuracy of SHE index is the same as MI index. In addition, MI index is better than SHE index in time metric.

TABLE III
EXPERIMENT RESULT

| Index | PCCR | AACCR | Time* |
|---|---|---|---|
| KL | 21 % | 81.2 % | 4.22 |
| HKL | 23 % | 82.3 % | 8.61 |
| PWKL | 28.9 % | 84 % | 9.31 |
| SHE | 100 % | 100 % | 36.4 |
| MI | 100 % | 100 % | 31.92 |

*The unit of time is minute.

In the fourth experiment, twelve questions are set in the test. The result is shown in Table IV. In this test, SHE index and MI index of exercise recommendation indices are ahead of the KL index, achieving a high accuracy rate of 100%. On the one hand, compared with shorter test, more answering record increase accuracy of prediction. On the other hand, MI index and SHE index can both recommend exercises of high discrimination.

TABLE IV
EXPERIMENT RESULT

| Index | PCCR | AACCR | Time* |
|---|---|---|---|
| KL | 47 % | 90.2 % | 5.42 |
| HKL | 45 % | 89.5 % | 11.2 |
| PWKL | 50 % | 90.7 % | 12.06 |
| SHE | 100 % | 100 % | 47.25 |
| MI | 100 % | 100 % | 41.44 |

*The unit of time is minute.

## V. CONCLUSION

In this paper, we compare five different information entropy indices in exercise recommendation. In addition, the effect of information entropy index are verified in an interactive dynamic recommendation model, in order to solve shortcomings in current exercise recommendation methods. In the experiment, we find the optimal exercise recommendation index which can provide better exercises to students. In the evaluation metrics, whether exercise can measure knowledge ability of students exactly is more important.

## ACKNOWLEDGMENTS

This work is supported in part by the National Natural Science Foundation of China (Grants No.61772442, 61836005 and 61806089).

## REFERENCES

[1] Huang, Zhenya and Liu, Qi and Zhai, Chengxiang and Yin, Yu and Chen, Enhong and Gao, Weibo and Hu, Guoping, "Exploring multi-objective exercise recommendations in online education systems," , in Proceedings of the 28th ACM International Conference on Information and Knowledge Management, 2019, pp. 1261-1270.

[2] Vukicevic, Milan and Jovanovic, Milos and Delibasic, Boris and Suknovic, Milija, "Recommender system for selection of the right study program for Higher Education students," RapidMiner: Data Mining Use Cases and Business Analytics Applications, pp. 145, 2013.

[3] Zhu, Tianyu and Huang, Zhenya and Chen, Enhong and Liu, Qi and Wu, Runze and Wu, Le and Hu, Guoping, "Cognitive diagnosis based personalized question recommendation(in chinese)," Chinese Journal of Computers, vol. 40, no. 1, pp. 176-191, 2017.

[4] Jiang, Peichao and Wang, Xiaodong and Sun, Bingxin, "Preference Cognitive Diagnosis for Predicting Examinee Performance," , in 2020 IEEE 2nd International Conference on Computer Science and Educational Informatization (CSEI), 2020, pp. 63-69.

[5] Zhou, Meizi and Ding, Zhuoye and Tang, Jiliang and Yin, Dawei, "Micro behaviors: A new perspective in e-commerce recommender systems," , in Proceedings of the eleventh ACM international conference on web search and data mining, 2018, pp. 727-735.

[6] Thai-Nghe, Nguyen and Drumond, Lucas and Krohn-Grimberghe, Artus and Schmidt-Thieme, Lars, "Recommender system for predicting student performance," Procedia Computer Science, vol. 1, no. 2, pp. 2811-2819, 2010.

[7] Shishehchi, Saman and Banihashem, Seyed Yashar and Zin, Nor Azan Mat and Noah, Shahrul Azman Mohd, "Review of personalized recommendation techniques for learners in e-learning systems," , in 2011 international conference on semantic technology and information retrieval, 2011, pp. 277-281.

[8] Tarus, John K and Niu, Zhendong and Mustafa, Ghulam, "Knowledge-based recommendation: a review of ontology-based recommender systems for e-learning," Artificial intelligence review, vol. 50, no. 1, pp. 21-48, 2018.

[9] Fan, Xitao, "Item response theory and classical test theory: An empirical comparison of their item/person statistics," Educational and psychological measurement, vol. 58, no. 3, pp. 357-381, 1998.

[10] De La Torre, Jimmy, "DINA model and parameter estimation: A didactic," Journal of educational and behavioral statistics, vol. 34, no. 1, pp. 115-130, 2009.

[11] Eguchi, Shinto and Copas, John, "Interpreting kullback-leibler divergence with the neyman-pearson lemma," Journal of Multivariate Analysis, vol. 97, no. 9, pp. 2034-2040, 2006.

[12] Hershey, John R and Olsen, Peder A, "Approximating the Kullback Leibler divergence between Gaussian mixture models," , in 2007 IEEE International Conference on Acoustics, Speech and Signal Processing-ICASSP'07, 2007, pp. IV-317.

[13] Amelio, Alessia and Pizzuti, Clara, "Correction for closeness: Adjusting normalized mutual information measure for clustering comparison," Computational Intelligence, vol. 33, no. 3, pp. 579-601, 2017.

[14] Zhang, Xiaoyu and Wang, Yongbin and Wu, Lin, "Research on cross language text keyword extraction based on information entropy and TextRank," , in 2019 IEEE 3rd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC), 2019, pp. 16-19.