

The Application of Genetic Algorithm and an Evaluation Algorithm in Online Examination System

Yang Yue, Mingbo Zhao*
School of Computer Science and Technology
Shandong University of Technology
Zibo, Shandong Province, China

Abstract:- Online examination system plays a significant role in education. However, there are varieties of disadvantages in non-optimized systems, such as randomly selecting questions that make the exam paper has an imbalance difficulty, the unanticipated weight of knowledge points, and so on. A genetic algorithm is an efficient and achievable way to improve the ability to generate exam paper.

Besides, a massive amount of data are generated when the system is running. Nevertheless, some of the systems only store the data, in another word, they do not make full use of the generated data. An evaluation algorithm is put forward in this essay to give objective and scientific evaluations on students' learning and teachers' teaching via using the data that are generated in examinations, which is based on the degree of difficulty. To make this algorithm working well, the degree of difficulty of questions stored in the database is supposed to be updated dynamically when the samples of questions' answers become large enough.

Keywords:- Online Examination System; Genetic Algorithm; Difficulty-Based Evaluation Algorithm; Education Feedback.

I. INTRODUCTION

Although there are a great number of online examination systems, lots of them are nonoptimized, which means they haven't had a superb algorithm scheme to generate examination paper effectively as well as an effective way to fully utilize the data generated after examinations[1].

It is confirmed that using a genetic algorithm is a significant method to optimize the function of automatic paper generation[2]. Furthermore, it is easy to achieve that updates the degrees of questions' difficulty with examinations' results, which makes the evaluation algorithm based on the degree of difficulty could be implemented.

This essay shows how to use a genetic algorithm and difficulty-based evaluation algorithm in the online examination system, what the difficulty-based evaluation algorithm is as well.

II. GENETIC ALGORITHM

A genetic algorithm is based on the law of biological evolution in nature, which searches the optimal solution by using a computer to simulate the process of biological evolution[3]. The main idea of the genetic algorithm is to encode the interfering factors of the question to biological chromosomes[3]. Then it converts the procedure of solving the question to the behavior of chromosomes, such as crossover and mutation, and implements population evolution via natural selections[4]. Finally, the optimal solution can be found at the end of evolution.

Comparing with other methods, a genetic algorithm is usually more efficient to find a better solution, especially for the complex combinatorial optimization problems[5].

A. Encoding

First and foremost, it's essential to encode the restrictive conditions of the question to chromosomes, otherwise, using a genetic algorithm to find the optimal solution is impossible.

In the online examination system, a chromosome is represented for an examination paper and each gene on the chromosome is stand for a question that is included in the paper. So, lots of chromosomes are constituted a set which means the solution domain of the optimization problem, paper generation problem. The optimal solution, an examination paper that we need is a specific chromosome that has already evolved through the following steps.

Generally, the chromosome is represented by a bit-string, which consists of symbols 0 and 1[4]. In this traditional encoding type, questions in the database are ordered by ID number[6], and there is a one-to-one correspondence between each question's position and each index of bit-string. If the paper contains the question, the bit is 1 in the corresponding position of the bit-string, otherwise, the bit is 0. The example can be seen as the following picture.



Fig 1:- Example of Traditional Encoding

However, this encoding way is really a poor design for the online examination system, for the reason that there are massive questions in the database of an online examination system, which means the bit-string might be very long, as a result, it is not beneficial for follow-up decoding and processing.

Therefore, an improved encoding method that makes the chromosome reserve some positions for genes, which presented by ID numbers of questions is more efficient for the online examination system. At the beginning of paper generation, each chromosome reserves many positions whose number is equivalent to the number of questions in the examination paper. The ID number of a question is filled to an empty position on the chromosome when the question is selected by the paper generation program. Other positions are filled in the same way till all positions are filled with nonduplicate ID numbers. The example of this encoding method can be seen as the following picture.

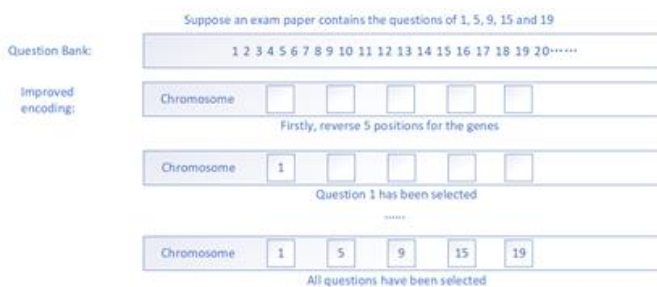


Fig 2:- Example of Improved Encoding

Obviously, the size of the chromosome is not too big even though the database has a large number of questions, which makes the chromosome easy to decode as well as easy to process.

B. Population Initialization

Before the process of evolution, an original population is needed. Assume a population has N individuals (in the experiment, N is set to 20), each individual has 3 chromosomes that respectively stand for selection questions, judgment questions and essay questions, which depends on the questions' types stored in the database. For each chromosome, select questions from the database of this type randomly and uniquely till every individual's chromosomes has filled with genes, no empty positions. It is not necessary to make every individual unique, but in the same chromosome, there must be no duplicate genes.

C. Fitness

After the original population generation, the process of evolution could start, but wait for a moment. Before evolution starts, the problem we should pay attention to is how to evaluate whether the population is better or worse, in another word, what are the evidences why we say a population is better than another one.

It is fitness that quantitative measures how good a population is. Each individual has a fitness that in the set of $[0, 1]$. The bigger the fitness is, the better the individual is. It

is stipulated that the biggest fitness of an individual in the population is the population's fitness. So different populations can be compared by their fitness[4].

A specific fitness function must be given, which is a programming implementation of a mathematical formula that is pre-defined. In the experiment, the interfering factor of fitness is the degree of difficulty, the weight of knowledge points' chapter and the frequency of questions have been selected, which is aim to make the difficulty of examination paper and the percentages of each chapter are in line with expectations, moreover, to make the question that has a low frequency in past examinations to be selected with a higher probability as possible.

D. Select Parents to Crossover

The primary method to realize population evolution is crossover, which is resemble with biological hybridization.

Firstly, select two different individuals from the parent population as parents randomly, moreover, this step can be optimized by selecting the best one in a random subset of parent population, if the second parent is duplicate, repeat the step till the two parents are different.

Secondly, generate a random positive integer S for each chromosome, which is smaller than the number of genes on the chromosome. The genes at the left of index S of the first parent are copied to the child's chromosome, while the right part genes of the second parent are copied to the child's chromosome starts at the index of S . If there are any duplicate genes, select another nonduplicate one from the database randomly. This step is called crossover. The example can be seen as follows.

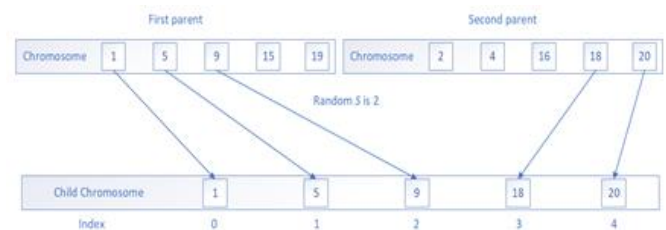


Fig 3:- Example of Crossover

E. Mutation

Although crossover is the main method to achieve evolution, it usually has no ability to produce new genes, which is similar to that in biology, in other words, a particular fitness is pre-generated before the evolution beginning and limits the evolution of the population. To solve this problem, the mutation has been introduced, which randomly changes a gene in a random position to a random and nonduplicate one in the database with a certain and quite low probability[3].

When the step of crossover is ended, there is a child population generated, which contains N individuals as same as the parent population. The system of mutation is applied to the child population. Therefore, maybe the parent population only contains a few genes, the child population

can acquire more novel genes that are not included in the parent population. However, the probability of mutation cannot be too high to avoid frequently destroy the positive result of evolution, which actually makes the genetic algorithm to be a completely random algorithm.

F. Elitism

It is similar to biology that evolution is the process that directional changes the frequency of the population's genes, but variation is not directional. Therefore, the fitness of the child population is lower than that of the parent is really possible, which means after the step of evolution, the population actually becomes vestigial.

Elitism is of virtual importance to avoid the problem of deterioration, which persists the best individual of parent population. It looks like clone in the biology, which directly copies all genes of the individual to the child whose genotype is exactly the same as a parent(only one parent).

As a result of keeping the best one of the parent population, the fitness of the child population must not become lower, at least remain the same as the parent.

However, one elite individual is sufficient. Do not hold more individuals even they are excellent to refrain from converging earlier to a local optimum solution.

G. Termination Condition

The process of evolution is running as the above steps repeatedly until the termination condition has met. The process can be terminated when the fitness of the population is higher enough(e.g. more than 0.98) or the process has run for a long time(e.g. population has evolved for 100 times)[2].

Finally, a good enough examination paper is generated. Maybe it is not the truly best one, but it might be better than most of the papers that are generated by a simple random algorithm.

III. DIFFICULTY BASED EVALUATION ALGORITHM

Online examination systems can do more things than a paper generation. People usually evaluate students' learning and teachers' teaching by the result of examinations, which can be done by online examination systems automatically. Generally, examination systems just simply statistic the results of examinations and show to users.

A. What the Evaluation Algorithm is

In this essay, an improved method which is based on the degree of difficulty is come up, which could remind students of the general situation of learning and each knowledge point's learning situation, as well as reminds the teacher of the situation of teaching. The sample of results for students is shown follows.



Fig 4:- Example of Evaluation's Result

Each length of bars stands for how good the student learns. The longer the bar is, the better the student learns.

The top bar is the overall evaluation that represents the result of this course:

- The red bar means this student had poor learning of this course, the below-average level.
- The blue bar means this student learned this course just OK, the average level.
- The green bar means this student had a better knowledge of this course, the above-average level.
- Bars following the overall bar are each chapters' evaluation that represents the result of each chapter:
- The red bar means this chapter is a weakness of this student.
- The blue bar means this chapter is neither weakness nor strength.
- The green bar means this chapter is the strength of this student.

Carefully, the result of chapters' evaluation, the color of chapter bars is always compared with the overall evaluation of the student himself, which means maybe a chapter bar is long, but the color of the bar is red that means this chapter is a weakness of this student, if the student is extremely excellent, the overall bar is much longer, therefore, this chapter is a relative weakness of the excellence and vice versa.

B. Prepare for Evaluation

In order to make the result of evaluation valuable, scientific as well as objective, the online examination system is supposed to have the ability of self-evolution, which automatically updates the degree of difficulty of questions whose answer samples' size are large enough[7]. In this way, the degree of difficulty of questions becomes more scientific and objective, which is the basis of the evaluation algorithm.

The main idea of the evaluation algorithm is that the degree of difficulty represents the average ratio of the score in an examination or a question within the case of large samples[7]. Therefore, the result of difficulty degree of an examination multiply by the total mark of the examination is a board sense average mark, which is affected by everyone who has answered any question that is included in the

examination although he has not attended this examination, as a result, the sample participated in the calculation is greatly expanded by the degree of difficulty.

The board sense average mark can be considered as if numerous students, more than that participate in, took part in the examination, the average of their marks might be equivalent to this value. If a student who has already participated in the examination got a mark bellows this value, it meant the student had poor learning of this course even though the mark that he got is not low, maybe greater than the actual average mark of the examination, which meant lots of the students that took part in this examination did not during well in the examination.

A new mark in the set of [0, 100] that remaps the board sense average mark to 50, marks below it is remapped to the set of [0, 50) and marks above it is remapped to the set of (50, 100]. The steps to calculate the evaluation mark can be seen as the next section.

C. Calculation of Evaluation Mark of Examination

There is an example of calculating the evaluation mark of a student for an examination.

- Calculate the degree of difficulty of the examination using (1). *De* stands for the degree of difficulty of the examination, *Dq* is the degree of difficulty of the question and *Mq* is the mark of this question.

$$De = \frac{\sum(Dq \times Mq)}{Exam\ Full\ Mark} \tag{1}$$

- Calculate the ratio of the score using (2). *Re* stands for the ratio of the score for the student in the examination.

$$Re = \frac{Student's\ Mark\ of\ The\ Exam}{Exam\ Full\ Mark} \tag{2}$$

- Calculate the evaluation mark of the examination using (3) if the result of (2) is bigger than that of (1), otherwise using (4). *Ee* stands for the evaluation mark of the exam.

$$Ee = \frac{Re - De}{1 - De} \times \frac{100}{2} + 50 \tag{3}$$

$$Ee = \frac{Re}{De} \times \frac{100}{2} \tag{4}$$

As is shown above, 50 is a line of demarcation that means the student is on average position. If the evaluation mark is smaller than 50, the student is below average, as multiplying the value by 2 means how many percent of the basic marks, below-average marks, the student gets. If the evaluation mark is greater than 50, the student is above average, as subtracting 50 from the value then multiplying the result by 2 means how many percent of the extra marks, above-average marks, the student gets. Therefore, each level's students can be quantitatively compared in the same region, which is distributed according to the difficulty gradient.

D. Calculation of Evaluation Mark of Chapter

The step of calculating the evaluation mark of a chapter is quite similar to that of examination, which just adds a constraint of the chapter. There is an example of calculating the evaluation mark of a student for a chapter.

- Calculate the degree of difficulty of the examination using (1). *Dc* stands for the degree of difficulty of the chapter, *Dqc* is the degree of difficulty of the questions that are included in the examination, *Mqc* is the mark of this question and *Mfc* is the total mark of this chapter's questions in the examination.

$$Dc = \frac{\sum(Dqc \times Mqc)}{Mfc} \tag{5}$$

- Calculate the ratio of score using (2). *Rc* stands for the ratio of the score for the student of the chapter in the examination.

$$Rc = \frac{Student's\ Mark\ of\ The\ Chapter}{Mfc} \tag{6}$$

- Calculate the evaluation mark of the chapter using (5) if the result of (6) is bigger than that of (5), otherwise using (8). *Ec* stands for the evaluation mark of the chapter in the examination.

$$Ec = \frac{Rc - Dc}{1 - Dc} \times \frac{100}{2} + 50 \tag{7}$$

$$Ec = \frac{Rc}{Dc} \times \frac{100}{2} \tag{8}$$

The meaning of the chapter's evaluation mark is similar to that of examination, which just changes the target from examination to chapter in the examination.

However, comparing the chapter's evaluation mark to that of examination could conclude whether the student has poor learning of this chapter if the chapter's evaluation mark is smaller or the student has a better knowledge of this chapter if the chapter's evaluation mark is larger.

E. Extend for Evaluation Mark

Above is the application of the evaluation mark in evaluating students' learning. It also can be used to evaluating teacher's teaching. The step is much as above, just sum the evaluation mark of all students that the teacher taught and calculate the average of the sum. There's no need to repeat the details of the steps for it.

IV. RESULT

The system improved by the two above algorithms is completely developed and is been running for a while, which has achieved the desired effect.

A. Result of Testing Genetic Algorithm

In the testing, the aim of the degree of difficulty of the examination paper is set to 0.7, 8 chapters whose weights are equivalent are selected.

One of the outputs of debugging is as table 1. From the left to right are the times of evolution that deletes some steps whose fitness are unchanged, the fitness of the population, the degree of difficulty of the examination paper, the PSI value which means how large the gap between expected chapters' weights and actual weights, the coefficient of questions' selection frequency.

No	Fitness	Difficulty	PSI	Selection
init	0.830652	0.766000	0.129478	0.454000
1	0.860387	0.750667	0.114302	0.381000
3	0.862074	0.756667	0.095097	0.366333
4	0.862737	0.740667	0.088301	0.408000
5	0.862822	0.748667	0.056277	0.399667
6	0.867615	0.750667	0.106605	0.355167
7	0.877655	0.716667	0.049698	0.426167
9	0.884776	0.710667	0.117070	0.386333
11	0.886873	0.708667	0.106605	0.387333
12	0.889987	0.702667	0.094467	0.395333
15	0.890988	0.700667	0.061618	0.409667
16	0.895405	0.700667	0.061618	0.392000
21	0.898314	0.700000	0.115615	0.360500
30	0.899514	0.702000	0.115615	0.350500
35	0.901394	0.704000	0.075893	0.353667
40	0.901427	0.706000	0.075893	0.348333
45	0.902909	0.702000	0.080822	0.350833
50	0.904323	0.702000	0.100439	0.337333
52	0.906248	0.700000	0.100439	0.334833
56	0.908524	0.702000	0.095510	0.322500
63	0.910430	0.708000	0.061618	0.312833
67	0.913303	0.702000	0.064386	0.315833
68	0.913432	0.702000	0.088931	0.305500
81	0.914352	0.702000	0.069727	0.309500
result	0.914352	0.702000	0.069727	0.309500

Table 1:- Record of Genetic Algorithm

The fitness is expected to become larger, at least keeping unchanged because of the elitism. The difficulty is supposed to converge to 0.7. The overall trends of PSI and selection are down.

As is expected, the fitness becomes larger, at least keeping unchanged, the difficulty converges to 0.7 with narrower fluctuation, and the overall trends of PSI and selection are down. On the whole, all trends of targets are in line with expectations. As we expect, the genetic algorithm makes an efficient improvement for a paper generation.

In order to view the result more intuitively, table 2 is the statistic of questions. From the left to right are the ID numbers of questions, the degree of difficulty of the question, the chapter that the question belongs to, the times that the question has been selected.

QID	Difficulty	Chapter	Selection
226	0.7	18	6
194	0.8	17	6
130	0.7	15	4
263	0.7	19	7
268	0.8	19	5
186	0.8	16	6
222	0.8	18	5
158	0.8	15	1
18	0.9	13	11
281	0.8	19	3
195	0.9	17	7
150	0.7	15	2
104	0.7	21	14
128	0.7	15	3
155	0.7	15	6
207	0.7	17	7
230	0.8	18	5
233	0.5	18	8
282	0.8	19	5
308	0.5	21	8
24	0.8	13	19
105	0.7	21	15
181	0.5	15	5
191	0.7	16	14
210	0.7	17	10
236	0.8	18	3
291	0.7	19	1
309	0.9	21	15
321	0.5	20	11
324	0.9	20	10
39	0.8	15	11
42	0.7	15	12
44	0.5	15	11
53	0.8	16	10
56	0.8	16	10
63	0.9	17	10
67	0.7	17	10
73	0.7	18	12
74	0.7	18	10
75	0.7	18	11
76	0.7	18	11
95	0.3	20	13
97	0.8	20	11
98	0.5	20	12
109	0.8	21	11
110	0.5	21	11
113	0.3	21	12
244	0.5	18	10
303	0.7	19	11
304	0.7	19	11

Table 2:- Statistic of Question

The distribution of difficulty is shown in the figure 5.

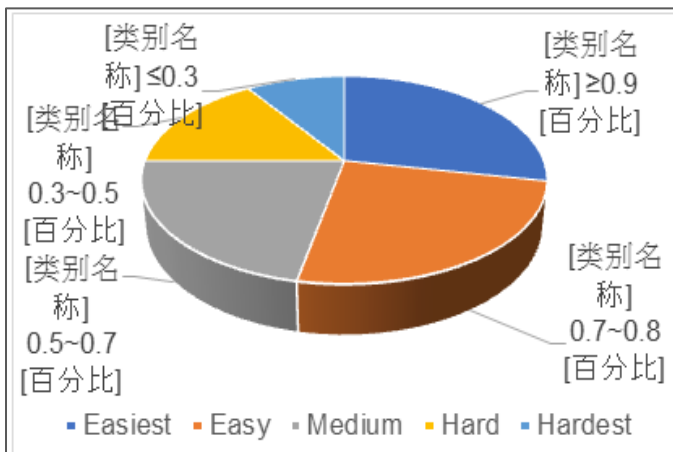


Fig 5:- Distribution of Difficulty

As is shown in figure 5, as a result of the difficulty of the examination paper is set to 0.7, most of the questions are in the set of easiest, easy and medium whose degree of difficulty is bigger than or equals 0.5. Only a little questions are the hard and hardest whose degree of difficulty is smaller than 0.5.

The distribution of chapters that questions belong to is shown in figure 6.

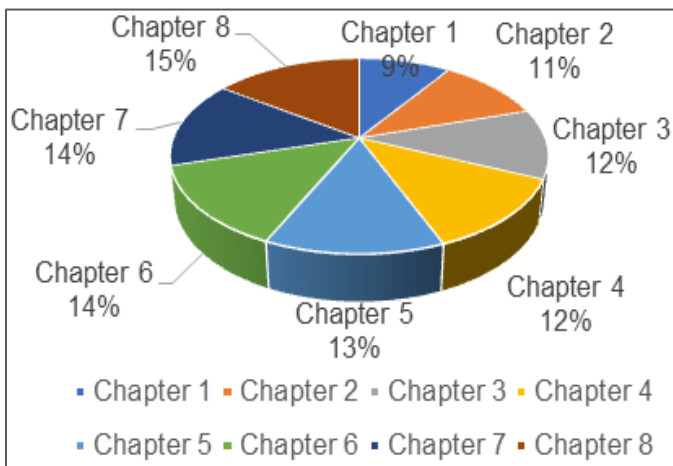


Fig 6:- Distribution of Chapters

As is shown in figure 6, most of the chapters' percentage are close, which are in line with expectations that the weights of chapters are equivalent, only chapter 1 and chapter 8 have a few gaps, for the reason that the degree of difficulty is more significant than PSI in this experiment.

The distribution of the times of questions has been selected is shown in figure 7.

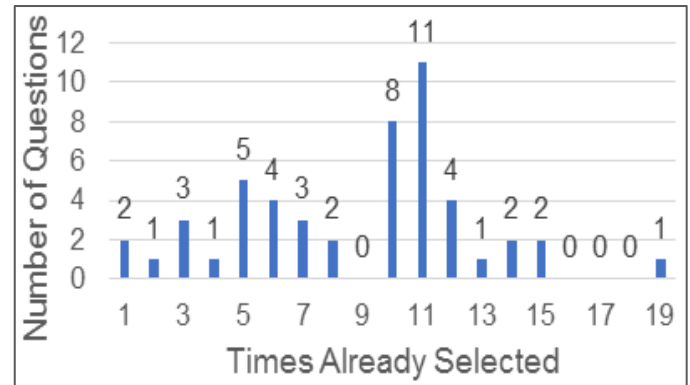


Fig 7:- Distribution of Selection Times

As is shown in figure 7, most of the times of questions have been selected are 10 and 11, and the number of questions whose selection times bellow average is greater than that of questions whose selection times above average. Therefore, it has achieved a good result in making a balance of the frequency of questions been selected.

B. Result of Evaluation Algorithm

The result of an examination for a student is shown in figure 8. The degree of difficulty of this examination paper is about 0.83, which means the difficulty of the examination is simple, and the total mark of the paper is 100, as a result, the average mark is expected to about 83.



Fig 8:- Result of Testing Evaluation Algorithm

As is shown in figure 8, although 71.5 marks whose total mark is 100 usually mean the student is at medium level, this examination's average mark is expected to 83, so the system gives the student a low evaluation, which reminds the student by a red overall bar. Each chapter's evaluation is compared to the overall evaluation, chapters 1,2,3,4, and 7 are the advantages parts for the student, and others are the disadvantage parts.

The numeric value of the evaluations can be seen in table 3.

Chapter	Evaluation Mark
Chapter1 Introduction	46.57
Chapter2 Basis of C Language	53.12
Chapter3 Elementary Programming	50.00
Chapter4 Selections	47.29
Chapter5 Loops	28.23
Chapter6 Arrays	42.45
Chapter7 Functions	73.68
Chapter8 Pointers	35.43
Overall	43.07

Table 3:- Value of Evaluation

As is shown in table 3, the overall evaluation's value is about 43, which means the student only get 86 percent(43 multiply by 2) of the basic marks(lower than board sense average mark), therefore, the student has poor learning of the course. Although the evaluation value of chapter 1 is below 50 that means the average level of massive students, it greater than that of the overall evaluation, so chapter 1 is an advantage part of the student for this course.

V. CONCLUSION

Using a genetic algorithm and the difficulty-based evaluation algorithm to improve the online examination system has achieved a positive result, which makes the function of a paper generation more effective and useful. Besides, benefiting by data generated in the online examination system, the difficulty-based evaluation algorithm could give scientific and objective evaluations to students' learning or teachers' teaching.

REFERENCES

- [1]. Jianjun Hu, Yuehong Sun, Qingzhen Xu, "The Genetic Algorithm in the Test Paper Generation," *Web Information Systems and Mining*, vol. 6987, pp.109-113, 2011.
- [2]. Liu Chao, Hu Yan, "Research On Intelligent Test Paper Generating Strategy Based On Genetic Algorithm And BP Neural Network," *PROCEEDINGS OF THE 2017 2ND JOINT INTERNATIONAL INFORMATION TECHNOLOGY, MECHANICAL AND ELECTRONIC ENGINEERING CONFERENCE (JIMEC 2017)*, no. 62, pp.157-160, 2017.
- [3]. Han LL (Han, LingLing), Li XD (Li XiaoDong), "The Analysis of Exam Paper Component based on genetic algorithm," *2014 FOURTH INTERNATIONAL CONFERENCE ON COMMUNICATION SYSTEMS AND NETWORK TECHNOLOGIES (CSNT)*, vol. 1, pp.561-564, 2014.

- [4]. Yongming Li, Dong Li, "The Design and Implementation of Genetic Algorithm with Directed Mutation in Intelligent Generating Test Paper System," in *2011 3rd International Conference on Computer and Network Technology(ICCNT2011)*, 2011, pp.159-161.
- [5]. Harik G R, Lobo F G, "A parameter-less genetic algorithm," *IEEE Transactions on Evolutionary Computation*, pp.523—528, 2015.
- [6]. Jeffrey D. Ullman, Jennifer Widom, *A First Course in Database Systems 3rd Edition*, Prentice Hall, 2007, pp.15-366.
- [7]. David Forsyth, *Probability and Statistics for Computer Science*, Springer International Publishing, 2018.