# Test paper composition based on fish swarm algorithm

**Wenfa Wang*, Shiyao Wang**

*Department of Computer Science, Yanan University, Shaanxi, 716000, China*

*Received 1 May 2014, www.cmnt.lv*

**Abstract**

Fish swarm algorithm was applied to seek optimization solution of test paper composition. Imitation results showed that fish swarm algorithm had better performance than random drawing algorithm in both composition accuracy and running time. The validity and superiority of the algorithm in this work, therefore, were verified.

*Keywords:* fish swarm algorithm, test paper composition, mathematical model, running time

## 1 Introduction

Development of intelligent item bank is an important aspect of computer application in educational field. By taking advantage of computer and technology and combining modern educational theory, questions are selected from current bank to automatically generate test questions satisfying teaching and examination requirements.

Currently, there are two main algorithms to solve test paper composition:

1) random search in current question bank for questions satisfying conditions with required amount. With great randomness and disconfirmation, this method is poorly intelligent, which cannot satisfy with education demands;

2) questions extracted for the whole test paper by random drawing. This method, which is relatively inflexible, cannot satisfy the different requirements of question types in papers.

## 2 Artificial fish swarm algorithm

Proposed by Li Xiaolei, et al. in 2002, Artificial Fish Swarm Algorithm (AFSA) is a new-type biotic optimization algorithm based on research into animal swarm intelligent behaviour. "Generally, the place with the most existing fish in water area is also the place with the most nutrients". According to this characteristic, this algorithm enables optimization to be realized by imitating foraging behaviour of fish swarm.

As a kind of optimizing method based on autonomous animats behaviour, AFSA is a new-type intelligent biotic algorithm built upon activity characteristics of fish swarm.

Fish has following behaviours by observation: foraging behaviour, swarm behaviour, following behaviour, and random behaviour. Because fish behaviours are tightly related to problem solution in this part, a major problem of algorithm implementation will be the adoption of simple and effective methods for constructing and realizing the behaviours.

### 2.1 BASIC CONCEPT

It is assumed that there is a swarm composed of *N* artificial fish in an n-dimensional target searching space. Status of every artificial fish per day can be expressed as vector quantity $X=(x_1,x_2,\cdots,x_n)$, in which $x_i\ (i=1,2,\cdots,n)$ is the variable for optimization. Food concentration in current location of artificial fish is expressed as $Y=f(x)$, while Y is the target function. Figure 1 shows that $\delta$ is the factor of congestion degree, step is the step length of artificial fish movement, $d=\|x_i-x_j\|$ represents distance between individuals, try number is the maximum try number for foraging by artificial fish, and visual means perception range of artificial fish.

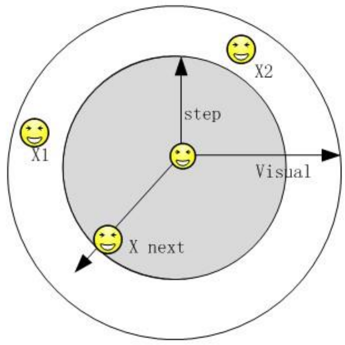


FIGURE 1 Visual field and step length of artificial fish

### 2.2 BEHAVIOUR DESCRIPTION OF ALGORITHM

1) Random behaviour: means that artificial fish move in visual field randomly. Once the food is found, it will lead the fish towards the direction with increasing foods rapidly.

***Corresponding author** e-mail: ydwangwenfa@126.com

2) Foraging behaviour: means that artificial fish move to a place with more foods. An artificial fish $X_i$ will randomly select a status $X_j$ in visual field, and then their target function values should be calculated, respectively. If $Y_j$ is better than $Y_i$ according to comparison, $X_i$ will move one step to the direction of $X_j$. Otherwise, $X_i$ should keep selecting status $X_j$ in its visual field until satisfying the forward condition. After repeated try number times, a random step should be taken to give $X_i$ a new status if the forward condition is not satisfied.

3) Swarm behaviour: means that artificial fish gather into a swarm during swimming to ensure their survival and avoid danger. The following three rules should be observed in a swarm. Division rule: is to avoid excessive congestion with neighbouring fish as much as possible. Alignment rule: is to keep consistent average direction to neighbouring fish as much as possible. And cohesion rule: is to move towards the centre of neighbouring fish as much as possible. An artificial fish $X_i$ will search the fish number $n_f$ and central location $X_c$. When $Y_c / n_f > \delta Y_i$, which represents the optimal and uncrowded status of partner's central location, $X_i$ will move a step towards the central location, or, perform foraging behaviour.

4) Following behaviour: means that artificial fish move to the optimal direction in visual field. An artificial fish $X_i$ will search the optimal partner $X_j$ of function among all partners. When $Y_c / n_f > \delta Y_i$, which represents it is not too crowded around the optimal partner, $X_i$ will move a step towards the central location, or, perform foraging behaviour.

5) Bulletin board: is a space for recording individual status of the optimal artificial fish. The current status of an artificial fish should be compared with recorded status on bulletin board after one time of iteration. When the comparison shows, that the current status is better, the status on bulletin board should be updated with it, or, keep unchanged. The required optimal value is the value on bulletin board outputted after all irritations in the whole algorithm.

### 2.3 ALGORITHM STEPS

Given that above-mentioned behaviours of artificial fish swarm, every artificial fish will select a behaviour for actual performance by exploring the condition of current environment and its partners. Eventually, artificial fish will gather around several local extremums in different numbers. In discussion of maximum problem, however, artificial fish with larger adaptive value will generally gather around larger extremum field, which is beneficial to obtain overall extremum field. Additionally, a greater number of artificial fish will gather around extremum field in larger value, which is beneficial to estimate and obtain overall extremum. Specific steps of artificial fish swarm algorithm are shown as follows:

Step 1. Swarm scale is determined as *N*, and *N* individuals are randomly generated in feasible region of variable. It is assumed that the visual range of artificial fish is Visual, the step length is Step, the factor of congestion degree is $\delta$, and the try number is try number.

Step 2. Adaptive value of every individual in initial fish swarm is calculated, and the status and its value of the optimal artificial fish will be given to bulletin board.

Step 3. Individual will update itself by behaviours, such as foraging, swarm and following to generate a new fish swarm.

Step 4. All the individuals should be evaluated. The bulletin board should be updated with the individual, which is better.

Step 5. Algorithm will be finished when the optimal solution on bulletin board reaches the satisfying error range, or turn to step 3.

## 3 Fish swarm algorithm steps for test paper composition

During test paper composition, logic order must be followed to compose a test paper conforming to composition constraints. Specific steps are as Figure 2.

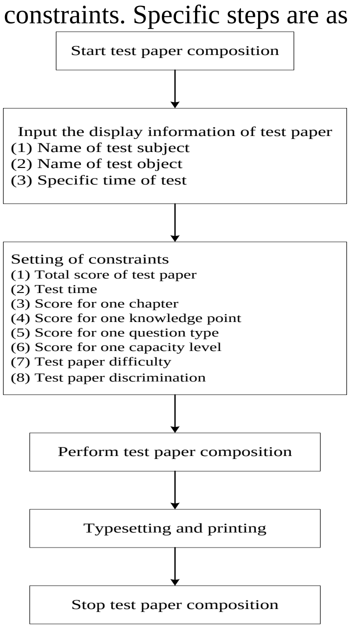


FIGURE 2 Fish swarm algorithm steps for test paper composition

## 4 Constraints for test paper composition

The objective of test paper composition is enabling test system to automatically select a certain amount of test questions from item bank according to constraints. These questions will compose a test paper satisfying examination requirement, which strives to sincerely and effectively reflect teaching level, and test examinees on mastery degree of relevant knowledge. Thus, test quality can be improved, and test objective can be reached.

Consequently, all the constraints affecting test paper problem should be determined according to examination requirement in composition to reach the objective. Generally, there are following constraints on test paper composition according to above analysis of composition steps:

1) Constraint of knowledge point. Knowledge points in teaching are always corresponding to section content in textbook, so this constraint is also the constraint of chapters. In occupational or other tests, however, constraint of knowledge point means the constraint of professional knowledge.

2) Constraint of question type. Constraint of question type means the question types included in the test paper, that is, question types for test paper composition in the examination.

3) Constraint of question amount. Constraint of question amount is the amount of test questions included in the test paper. Specifically, it is the question amount for one question type. This constraint is positively correlated with answer time

4) Constraint of answer time. Constraint of answer time means the longest time for finishing all the questions in the paper. In general, the longer the answer time is, the greater the question amount is.

5) Constraint of difficulty. Constraint of difficulty means the average difficulty of the test paper and the difficulty of test questions. Generally, test difficulty depends on different test targets. For example, in the test selecting minority of outstanding students, test difficulty should properly increase. In qualification test requiring majority of passers, however, the difficulty should properly reduce. Better effectiveness can be reached only if this constraint combines with the constraint of discrimination.

6) Constraint of discrimination. Constraint of discrimination is the discrimination capability of test paper. Total discrimination of test paper is the average discrimination of all test questions.

7) Constraint of exposure time. This means the test times of knowledge. The more test times, the more times the knowledge point will be repeated.

## 5 Mathematical model of test paper composition

The above analysis on constraints shows that the system is always limited by some constraints when selecting a test question from item bank in composition. If every constraint is a local constraint, all the constraints, which should be satisfied in the test paper, will constitute global constraint. Four vectors will be defined here:

Attribute set: $s=\{s_1,s_2,\cdots,s_n\}$, which is attribute variable of test question (question type, affiliated section, knowledge point, difficulty factor, score, etc.) or their quantization value;

Value set: $L=\{l_1,l_2,\cdots,l_n\}$, which corresponds to the value range of every attribute variable of test question;

Constraint set: $R=\{r_1,r_2,\cdots,r_n\}$, which corresponds to the constraint relation between different values for attribute variable of test question;

Question set: $I=\{i_1,i_2,\cdots,i_n\}$, which corresponds to the test question selected from item bank and satisfying with certain constraints.

It can be seen that test paper composition is to randomly select a group I from item bank, while the attribute variable *S* of this group should satisfy the constraint of *R* within value range *L*. Automatic test paper composition, therefore, can be described as a problem of multi-constraint objective satisfaction. In addition, the mathematical model of test paper composition can also be considered as a constraint satisfaction problem. The constraints can be divided into hard constraint and soft constraint in type.

### 5.1 MATHEMATICAL MODEL OF HARD CONSTRAINT

Actually, the test paper composition is to select test questions satisfying all the constraints from item bank. It is assumed that there are m test questions in a test paper, and each of these questions should satisfy n indicators. Therefore, the following *m***n* target matrix is established. (*n*-dimensional vectors include difficulty $a_1$, question type $a_2$, section $a_3$, teaching requirement $a_4$, time $a_5$, score $a_6$, etc. $a_i$ means the indicator No. *i*. *m* refers to question amount in paper.) The matrix is as follows:

$$s=\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}.$$

In this matrix every line represents one question in item bank, while every list is the attribute value of the question. The target matrix should satisfy the following constraints:

1) $\sum a_{i1}*a_{i6}/100=ND$ (difficulty of the whole paper, which is determined by the user of composition), the constraint of difficulty. The determination of question difficulty: scoring rate (d)=1- (average score/full score of the question);

2) $\sum a_{1i}*a_{i2}=m_d$ ($m_d$ is the score of question type *d*), in which $C_{1i}=\begin{cases}1\\0\end{cases}$. If $a_{i2}=d$, take 1 for $C_{1i}$, otherwise, take 0;

3) $\sum c_{2i}*a_{i3}=Z_h$ ($Z_h$ is the score for section *h*), in which $C_{2i}=\begin{cases}1\\0\end{cases}$. If $a_{i3}=h$, take 1 for $C_{2i}$, otherwise, take 0;

4) $\sum c_{3i} * a_{i4} = p_k$ ($p_k$ is the score for teaching requirement $k$), in which $C_{3i}=\begin{cases}1\\0\end{cases}$. If $a_{i2}=k$, take 1 for $C_{3i}$, otherwise, take 0;

5) $\sum a_{i5} = t$ ($t$ is the test time);

6) $\sum a_{i6} = G$ ($G$ is the total score of test paper).

### 5.2 MATHEMATICAL MODEL OF SOFT CONSTRAINT

In test paper composition, “hard constraint” is to completely obtain the target value, while “soft constraint” is to incompletely obtain the target value. In practical process, above target functions (1)–(6) are difficult to be completely satisfied. In order to minimize the deviation between target function and target value, a priority should be established based on the importance of every target according to user requirements. All the targets should be ordered in this priority to achieve more targets as much as possible. The imported positive and negative deviation variables represent the parts more and less than target value, respectively. The process of test paper composition, therefore, becomes the solution process of “soft constraint” target value. The multi-target optimization model based on “soft constraint” is established as follows:

$$\min\ z = p_1 \left| \sum_{i=1}^{p} \left( da_i^- + da_i^+ \right) \right| + \ldots + \mathrm{p}_2 \left| \sum_{j=1}^{q} \left( db_j^- + db_j^+ \right) \right| .$$

Constraints:

$$\sum_{j=1}^{q} \sum_{k}^{n} f_k x_{ijk} + da_i^- + da_i^+ = a_i\ , \quad i = 1, 2K, p$$

$$\sum_{i=1}^{p} \sum_{k}^{n} f_k x_{ijk} + db_j^- + db_j^+ = b_j\ , \quad j = 1, 2K, \mathrm{q}\ .$$

$$\sum_{i=1}^{p} \sum_{j=1}^{q} x_{ijk} = S_k\ , \qquad k = 1, 2K, \mathrm{n}$$

In the constraints, $p_i$, the priority factor, represents relative importance of every target. $p_i$ is always prior to $p_i$+1 to all values of $i$, in which i=1, 2,3… $x_{ijk}$ represents the amount of questions with difficulty $i$, scope $j$ and question type $k$ in the test paper. $da_i^-$ and $da_i^+$ represent the positive and negative deviation variables for total score of questions with difficulty $i$, respectively, which correspond to difficulty requirement $a_i$. $db_j^-$ and $db_j^+$ represent the positive and negative deviation variables for total score of questions with scope $j$, respectively, which correspond to scope requirement $b_j$.

## 6 Algorithm simulation

Table 1 shows the parameter setting in algorithm simulation with MATLAB software according to algorithm steps and mathematical model of test paper composition:

TABLE 1 Parameter setting for fish swarm algorithm

| Parameter | Value | Parameter | Value |
|---|---|---|---|
| Artificial fish swarm | 50 | Perception distance | 1 |
| The maximum iterations | 100 | Factor of congestion degree | 0.618 |
| The maximum try number of foraging | 100 | Step length of movement | 0.1 |

The result of 100 iterations simulation results shows fish better convergence of the algorithm (see Figure 3).

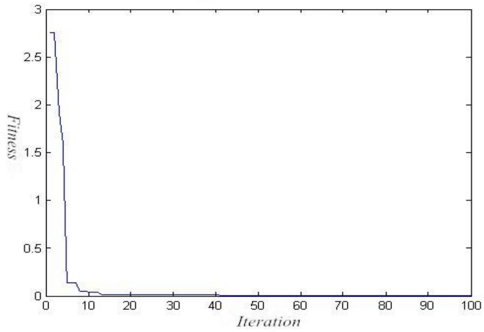


FIGURE 3 Convergence map of fish swarm algorithm

In order to verify the effectiveness and superiority of the algorithm in this work, the algorithm is compared with random drawing in composition accuracy and running time. Figure 4 and Figure 5 show the comparison results:

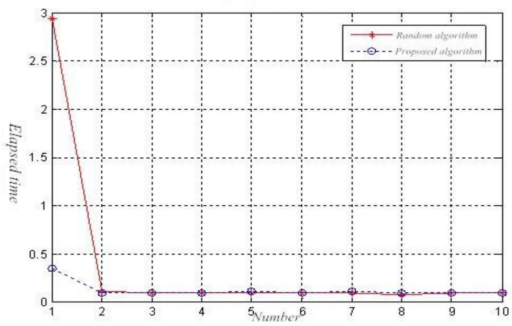


FIGURE 4 Comparison in running time

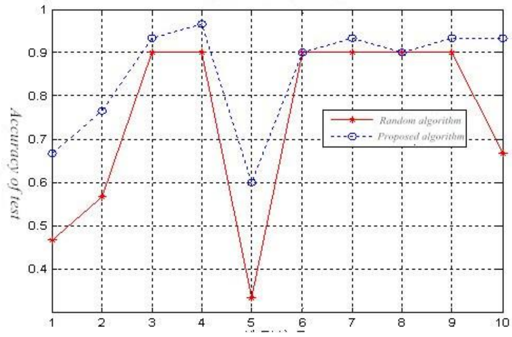


FIGURE 5 Comparison in accuracy

Simulation experiment is conducted by MATLAB software. By comparing with the algorithms in literature, the algorithm in this work is verified in both composition accuracy and running time. It is found in Figure 3 that the accuracy of fish swarm algorithm is higher than that of random drawing. Moreover, it is found in Figure 4 that running time of fish swarm algorithm is also better than that of random drawing.

## 7 Conclusions

Fish swarm algorithm was applied to make optimization solution for test paper composition. Simulation result showed that fish swarm algorithm was better than random drawing in both composition accuracy and running time. The effectiveness and superiority of algorithm in this work, therefore, were verified.

## Acknowledgements

The work was supported by Natural Science Fund of Yan’an (YA2013-12), “Related Models and Algorithms Improving Oil Production of Enterprises”; Natural Science Special Fund of Yan'an University (YDZ2013-02), “Oil Pump Scheduling Policy under Multi-Constrained Condition — A Case Study of Oil Plant in Qili Village”; National Innovation Training Program (201410719001), “Design and Implementation of Universal Questions System in Universities”.

## References

[1] Yuan X 2008 *Research on intelligent examination system based on adaptive genetic algorithm* Central South University Changsha
[2] Yin X, et al. 2009 Analysis and study of volume pattern in network test system *Computer Technology and Development* **19**(2) 206-8
[3] Lei Guoyu, Zhouyong 2006 *Analysis and design of strategy for test paper composition* Southwest University of Science and Technology, Sichuan
[4] Zhang Y, Li H 2007 Maneuver of composing examination papers from questions in view of genetic algorithms *Journal of Inner Mongolia Agricultural University* **27**(3) 103-5 (*in Chinese*)
[5] Zhang Q, Xing C 2008 Application of real-number encoding in intelligent test paper composition in view of genetic algorithms *Science & Technology Information* **1**(3) 44-5
[6] Zhou H 2009 Research on strategy for test paper composition in test system based on web *Northeast Normal University Jilin*
[7] Xia G 2008 *Research on general question bank management system of self-learning examination and its strategy for intelligent test paper composition* Tianjin Normal University, Tianjin
[8] Yun Q, Huang G 1997 *Genetic algorithm and genetic technology* Metallurgical Industry Press Beijing
[9] Wang Y, Hou S, Guo M 2003 Research on algorithm for automatic test paper generation in testing system *Journal of Harbin Institute of Technology* **35**(3) 343-8

## Authors

**Wenfa Wang, born in October, 1968, Tongshan, Shanxi Province, China**

**Current position, grades:** an associate professor of Department of Computer Science, Yanan University, China.
**Scientific interest:** software technology and algorithm design.
**Publications:** 20 papers.

**Shiyao Wang, born in November, 1989, Yanan, Shanxi Province, China**

**Current position, grades:** a master of Department of Computer Science, Yanan University, China.
**Scientific interest:** algorithm design and analysis.
**Publications:** 3 papers.