Automatic Question Paper Pattern Generation using GA Approach

DIMPLE VALAYIL PAUL
Department of Computer Science, Dnyanprassarak Mandal's College and Research Centre,
Assagao, Bardez, Mapusa-403507, Goa, India
Corresponding Author: dimplevp@rediffmail.com

Abstract - This paper focuses on question paper template generation and its use in dynamic generation of examination question paper. Question paper template generation is a constrained based optimization problem. Choosing an efficient, scientific and rational algorithm to generate a template is the key to dynamic examination question paper generation system. By using Genetic Algorithm (GA) and educational taxonomies, this paper analyses the initial population generation, does chromosome encoding, applies genetic manipulations and experimentally proves that the generated question paper templates are best suited for the dynamic examination paper generation system. This new approach outperforms traditional algorithms that randomly generate examination papers in terms of their topic coverage, learning domains and marks distribution.

Keywords – GA approach

(Received June 1st, 2020/ Accepted June 11st, 2020)

1. Introduction

Every educational system consists of an examination system through which the qualities and abilities of the students are assessed by giving them grades and positions [1]. Preparing an examination question paper is challenging, tedious and time consuming for the instructors. A high-quality question paper (test paper) must necessarily address the following issues.

- Are contents of question paper appropriate in terms of coverage of topics?
- Can questions (test items) examine student ability at different cognitive levels of Taxonomy?
- Is it possible to avoid similarity among questions?
- Can we generate question papers of different difficulty levels?
- Does the question paper satisfy both time and marks constraints?

The efficiency and quality of dynamic examination paper generation system depends entirely on its algorithm design. Traditional question paper generation algorithms have the disadvantages of slow convergence, low success rate and poor quality. This motivated us to use genetic algorithms in question paper generation [2-4]. In this paper, we present the way in which educational taxonomy and improved genetic algorithm can be applied in question paper template

generation. This new approach can successfully generate examination question papers dynamically by considering the changing requirement of each instructor.

.

2. Research Background & Related Work

Three approaches commonly used till date for question paper generation are as under –

a. the question database is viewed as the questions set for selection by experienced instructors;

b. previously generated question paper set is used for selection;

c. intelligent algorithm is used for automatic generation of question paper.

The first and second approaches are able to guarantee quality. But the first one works similar to manual operation, while the second limits the number of papers. The third type is widely used in educational field at present and this approach is already implemented by many researchers [5-6]. But the main limitation of many of these existing implementations is that they use a random generation strategy and create thousands of question papers with repetitive questions lacking in coverage of topics and learning objectives. In our approach, we generate a dynamic template and use it for generation of question paper that has proper weightage allotted to subject content, learning domain, type of question, etc. and can be used to generate several question papers almost without repetition depending on the instructor's choice. The number of unique question papers (without any overlap) that can be prepared using a generated template depends on the quality and size of the Question Bank created by the Instructor for that subject.

The Approach presented in this paper monitors the quality of question paper based on a wide range of instructor requirements such as the general topics, the average degree of difficulty, kinds of questions, selection of modules, selection of cognitive levels, etc. In order to incorporate the above requirements, we have applied the well-established concept of Educational Taxonomies. GA offers a robust non-linear search technique that is particularly suited to problems involving large numbers of variables. The strength of Genetic algorithms is derived from their ability to exploit, in a highly efficient manner, information about a large number of individuals. This search method is modeled on natural selection by Holland in [7] whose motivation was to design and implement a robust adaptive system. GA [13] is being used to solve a variety of problems and is becoming an important tool in machine learning and function optimization.

GA applies genetically inspired operators to populations of potential solutions in an iterative fashion, creating new populations while searching for an optimum solution [8-9]. The evolution of new generation is based on two primary operators: mutation and crossover. The power of genetic algorithms is the technique of applying these recombination operators (crossover and mutation) to a population of individuals. Despite their randomized nature, GA is not a simple random search. GA takes advantage of the old knowledge held in a parent population to generate new solutions with improved knowledge.

3. Methodology

3.1 Select Units or Modules of a Subject

Examination is conducted for a subject of a course having pre-defined university specified syllabus file with unit-wise contents. It is necessary to mention whether the question paper template is designed for all units or modules of a subject or selected unit of a subject.

3.2 Decide Cognitive Processing Levels

Instructors always attempt to include questions that measure higher levels of cognitive processing. This is not a good approach to evaluate performance of students at different levels of learning such as Excellent, Good, Average, etc. It should be the goal of the instructor to ensure that their questions have cognitive characteristics testing understanding, problem-solving, critical thinking, analysis, synthesis, evaluation and interpretation rather than just declarative knowledge. There are many theories that provide frameworks on levels of thinking which have serious impact on framing good questions. Bloom's taxonomy [10-12] is often cited as a tool to use different cognitive levels in choosing the questions. Cognitive processing levels for a question paper template are decided on the basis of the taxonomy that is selected by the instructor for each examination. Difficulty level of an examination paper is defined on the basis of the percentage of weights allotted to different cognitive levels of taxonomy. In our system, we have considered three major difficulty levels such as low, medium and high which are differing based on the weights assigned to different cognitive levels.

3.3 Design Question paper Template

The question paper template (Figure 1.) can be thought of as an engineering design plan. It lays out exactly how the question paper will be created and essentially depicts how the questions are measured. A question paper template is the framework for the structure of the question paper. The template specification gives a plan for constructing a quality examination question paper. The template defines the scope of the paper with respect to syllabus contents and content of skills being measured by the examination, as well as the hierarchical importance of each content area along with its cognitive processing levels.

3.4 Frame Questions for Question Paper

Good question papers always have a well-defined plan for measuring the important skills and knowledge of credential candidates. Question paper should always be developed on a sound scientific basis with a clear rationale for item writing and test score interpretation.

An examination needs to have a specification that depicts a clear relationship between performance of a job function, mastery of skills and knowledge area. A question paper should have a planned distribution of questions among different level of difficulty to classify students as Excellent, Good and Average.

3.5 Manage a Question Bank

For each subject, questions are arranged in the question bank in a unit-wise manner. Questions are of objective and descriptive types. In order to generate a question paper, questions are selected from the question bank based on the marks allotted to the level of the unit specified in the generated question paper template. Updates to the Question Bank are done manually.

4. Problem Formulation

4.1 Steps for applying GA

- **a. Generate Population:** Population is a set of chromosomes used to find the optimal feature subsets. In this work, a matrix represents a chromosome, and a population consists of templates that are either generated at initialization (first generation) or generated by applying genetic algorithm operators.
- **b. Initialization:** In this step, we generate the initial population, which is the initialization of random marks to each cell of the matrix.
- **c. Calculate Fitness:** In this step, the fitness score of each chromosome (template) is calculated. The details of Fitness Function are explained in section 4.2.1. Then, the generated population is searched to find the chromosome (template) with best fitness value.
- **d.** If for any generation, a template is found with fitness value closer to 1, then we stop.
- **e.** Else apply GA operators and produce a new generation.

Repeat the above steps (c) to (e) after each generation until optimum solution is arrived or the specified numbers of generations are completed.

4.2 Problem statement

4.2.1 Fitness Value for a test paper template (F)

Let N be the number of marks allotted for the question paper. Let m be the number of units selected by instructor and n be the number of taxonomy levels. Let $U=<u_1, u_2,..,u_m>$ be the vector of weights where u_i is the weight assigned to the i^{th} unit, and, $L=<l_1,l_2,\ldots,l_n>$ be the vector of weights where l_i is the weight assigned to the i^{th} level.

Let $X=<x_{11},x_{12},x_{ij},\ldots,x_{mn}>$ be the set of weights where x_{ij} is the weight assigned to the j^{th} level of i^{th} unit.

For a unit i, $\sum_{j=1}^{n} x_{ij}=u_i$, and
for a level j $\sum_{i=1}^{m} x_{ij}=l_i$

4.2.1.1 The weakness of a unit (WU_i)

For a unit i, $X_i=<x_{i1},x_{i2},\ldots,x_{in}>$ and x_i cX. Before calculating W_i, normalize X_i to obtain $X_i=<x_{i1},., x_{in}>$ such that $x_{ij}= x_{ij} *N/u_i$.

$$WU_i = (\sum_{j=1}^{n} |x_{ij} - l_j) /N$$

4.2.1.2 The weakness of a level (WL_i)

For a level j, $X_j=<x_{1j},x_{2j},\ldots,x_{mj}>$ and x_jcX. Before calculating L_j, normalize X_j to obtain $X_j=<x_{1j},., x_{1m}>$ such that $x_{ij}= x_{ij} *N/l_j$.

$$WL_j= (\sum_{i=1}^{m} |x_{ij} - u_i) /N$$

4.2.1.3 The fitness of a paper

Unit wise fitness, $Funit = \sum_{i=1}^{m} (1-WU_i)/m$

Level wise fitness, $F_{level} = \sum_{j=1}^{n} (1-WL_j)/m$

Fitness of paper, $F= (F_{unit} + F_{level})/2$

4.2.1.4 Simplex Representation

$$\text{Max } F= ((\sum_{i=1}^{m} (1-(\sum_{j=1}^{n} |(x_{ij}*N)/u_i - l_j|/N))/m) +$$

$$(\sum_{j=1}^{n} (1-(\sum_{i=1}^{m} |(x_{ij}*N)/l_j - ui|/N))/n)) /2$$

Subject to the constraint

$$\sum_{j=1}^{n} x_{ij}=u_i$$

$$\sum_{i=1}^{m} x_{ij}=l_i$$

$x_{ij}>0$

5. Implementation Details

5.1 Hardware and Software Platform Used:

Implementation is done using Microsoft Visual Basic.NET as Front-End Tool and SQL Server as Back End Tool on a 2 GHz processor with 1GB RAM.

5.2 Datasets used:

Study was conducted for Goa University Examination System. Software Engineering (SE) and Information Technology (IT) are two subjects offered in the third year of the three year bachelor's degree course of computer science (B.Sc Computer Science) at Goa University. Blooms taxonomy with six levels such as knowledge, understanding, Application, Analysis, Synthesis and Evaluation were considered as cognitive processing levels.

5.3 Sample Input screen:

Instructor is given the flexibility to choose some/all units of a subject and also some learning objectives of the specified subject. Question paper template for different examinations such as in-semester (20marks), end-semester (80marks), practical (50marks), etc. can be generated. Provision is also made to prepare question paper template on different difficulty levels such as Low, Medium and High. Figure 2 below shows the sample input for question paper template generation.

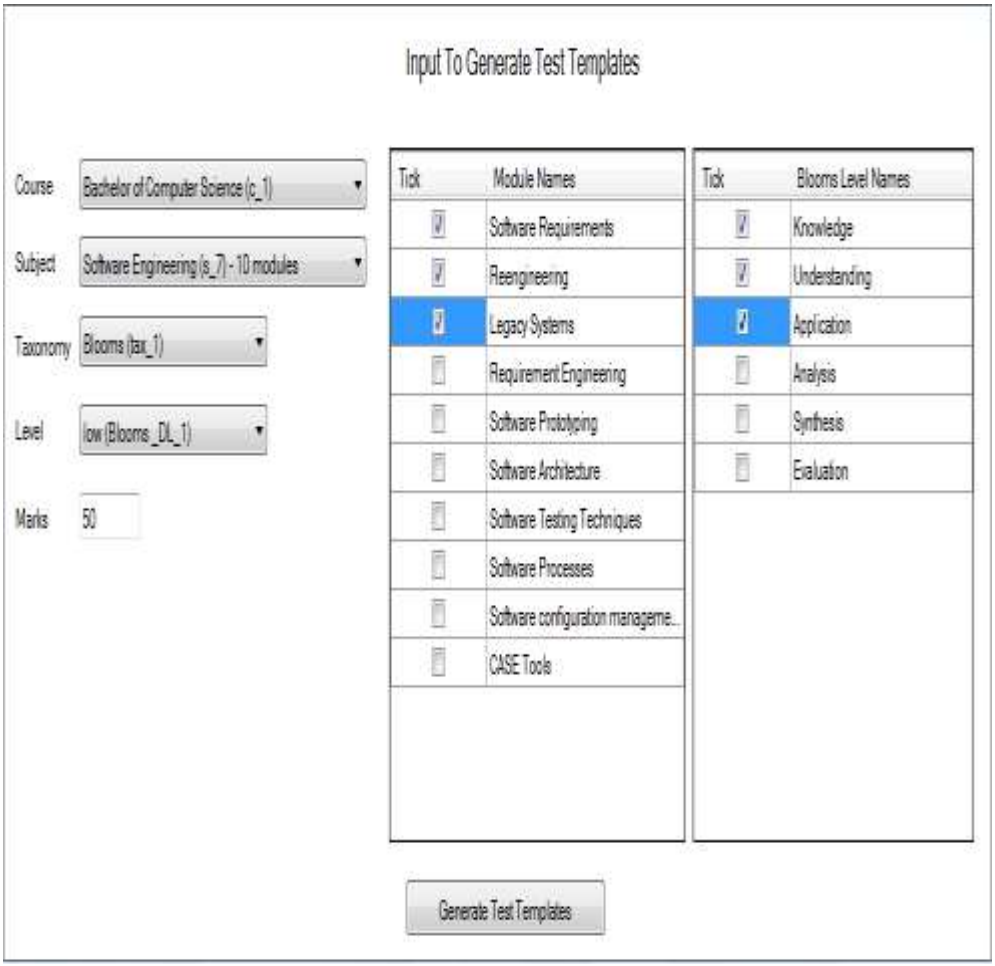
Table 1. Terminology used

Term	Meaning
Course	Refers to a Degree/Diploma program offered at the university. Example: Bachelor of Science (Computer Science), Bachelor of Computer Application
Subject	Subject is a paper which is offered at different semesters of a course. Example: Software Engineering (SE) is offered at the third year of B.Sc Computer Science.
Modules or Units	For each subject, there is a university prescribed syllabus which consists of different modules and each of these modules is allotted particular weightage. Example: Software Requirement in SE is given a weightage of 20 marks.
Instructor	An instructor is a faculty who teaches a subject in a particular course.
Taxonomy	Taxonomy is a classification system of educational objectives based on the level of student understanding necessary for achievement or mastery. Example: Blooms, Solo etc.
Taxonomy Levels	Each taxonomy has its cognitive stages in learning and is called taxonomy levels. Example of Blooms Taxonomy Levels: Knowledge, Comprehension, Application, Analysis, Synthesis and Evaluation
Difficulty level with respect to taxonomy	Difficulty level with respect to taxonomy refers to the value used to assign weightage to each taxonomy level. Examination for different subjects of a course is conducted with varying difficulty levels (Low/Medium/High).
Question paper template	A matrix with rows representing units, columns representing cognitive levels of taxonomy and cells representing weightage assigned to a level of a unit. (Figure 1.)

Figure 1. Question Paper Template Format

Units	**level 1**	**level 2**	**level 3**	**level 4**	**…**	**level n**	**Units weight**
unit 1	x_{11}	x_{12}	x_{13}	x_{14}	⋯	x_{1n}	u_1
unit 2	x_{21}	x_{22}	x_{23}	x_{24}	⋯	x_{2n}	u_2
unit 3	x_{31}	x_{32}	x_{33}	x_{33}	⋯	x_{3n}	u_4
unit 4	x_{41}	x_{42}	x_{43}	x_{43}	⋯	x_{4n}	u_4
…	…	…	…	…	…	…	…
unit m	xm_1	x_{m2}	x_{m3}	x_{m3}	⋯	x_{mn}	u_m
Level weight	l_1	l_2	l_3	l_3	…	l_n	**N**

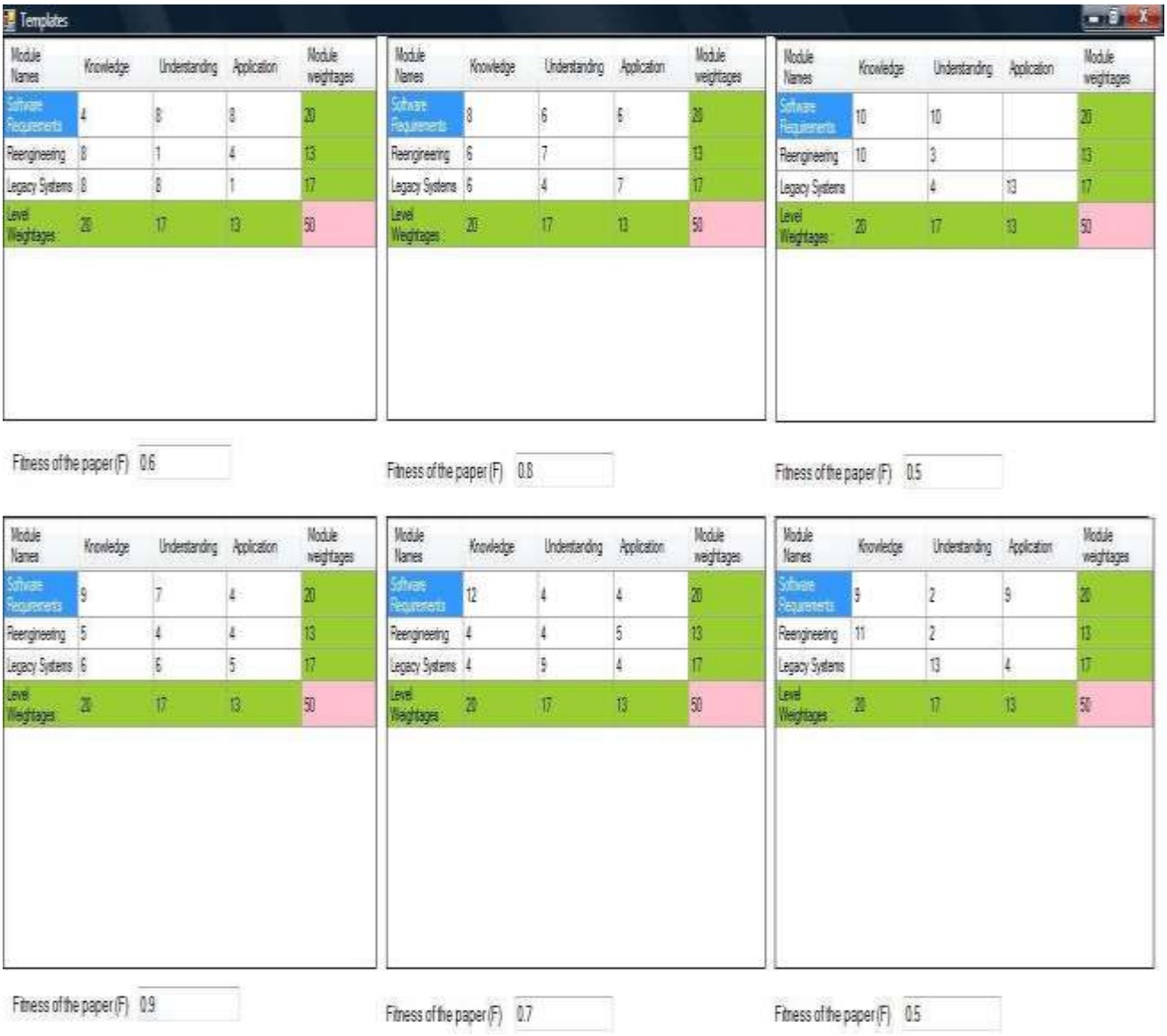
Figure 2. Question Paper Template Format



5.4 Results Obtained:

Figure 3 below shows the question paper templates generated for SE by selecting only three units of the subject and choosing only three levels of blooms taxonomy namely Knowledge, Understanding and Application.

Figure 3:Generated question paper templates with calculated fitness values



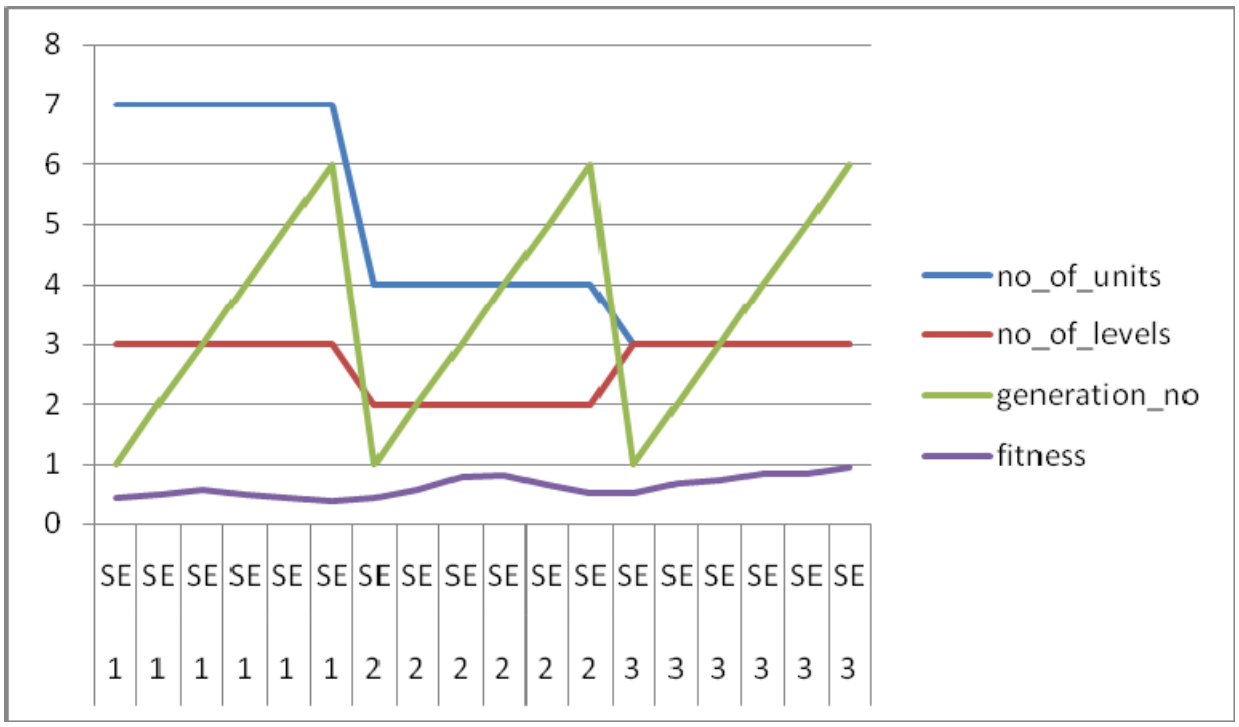
5.5 Accuracy of the Result

Table 2 below shows the results obtained after implementation. Figure 4 below shows that SE template for matrix of size 3 by 3 is successfully getting generated with a fitness value of 0.9456 which is the optimum solution.

Table 2. Calculation of fitness value for three question paper templates of SE

Template_No	Subject	No_of_Units	No_of_Levels	Generation_No	Fitness_Value
1	SE	7	3	1	0.4698
1	SE	7	3	2	0.5178
1	SE	7	3	3	0.5834
1	SE	7	3	4	0.5122
1	SE	7	3	5	0.4532
1	SE	7	3	6	0.4123
2	SE	4	2	1	0.4678
2	SE	4	2	2	0.5785
2	SE	4	2	3	0.7976
2	SE	4	2	4	0.8245
2	SE	4	2	5	0.6745
2	SE	4	2	6	0.5323
3	SE	3	3	1	0.5296
3	SE	3	3	2	0.6795
3	SE	3	3	3	0.7435
3	SE	3	3	4	0.8833
3	SE	3	3	5	0.8434
3	SE	3	3	6	0.9456

Figure 4: Comparison of fitness values based on the above table


8
7
6
5
4
3
2
1
0
SE 1
SE 1
SE 1
SE 1
SE 1
SE 1
SE 2
SE 2
SE 2
SE 2
SE 2
SE 2
SE 3
SE 3
SE 3
SE 3
SE 3
SE 3
no_of_units
no_of_levels
generation_no
fitness

6. Conclusion

This paper focused on a new approach for dynamic question paper generation by using question paper templates that are obtained using genetically inspired algorithm. The primary objective of this study was to generate question paper template, using genetic algorithm that can be used in dynamic generation of examination question paper. The main advantage of this new approach is the application of strengths of GAs for use in dynamic question paper generation. Complexity of this template generation algorithm is proportional to the number of units and the number of levels.

7. Future Work

In our work, we have focused on generation of the question paper template by using genetic algorithm. This template can be used to generate question paper, either by manually selecting the questions from the question bank or using intelligent algorithms for the selection of questions from the question bank. Our future work will focus on developing algorithms to automate question paper generation which will require the minimum number of inputs from instructors.

Reference

[1] Ahmad,Z. "Existing System of Examination and the Need of Reform" ,Journal of Elementary Education vol .1,3,56,1993.

[2] D. E. Goldberg, "Genetic Algorithms in Search, Optimization and Machine Learning", Addison-Wesley Longman Publishing Co., Inc. Boston, MA, USA, 1989.

[3] Huang, M., Tang, L., Liang, X., "Research on IntelligentTest Paper Generation Based on Improved Genetic Algorithm," Chinese Control and Decision Conference (CCDC 2009), China, pp.2967-2970, 2009.

[4] Ming-Hsiung Ying, Shao-Hsuan Huang, Luen-Ruei Wu,"An Item Selection Strategy Based on Association Rules and Genetic Algorithms", Journal of Software, Vol. 5, No. 12, December 2010.

[5] Adeli, H., & Cheng, N.,Concurrent genetic algorithms for optimization of large structures. ASCE Journal of Aerospace Engineering, 7(3), 276–296, 1994.

Ting Long, Houjun Wang, Shulin Tian," Test Generation Algorithm for Linear Systems Based on Genetic Algorithm", Journal of Electronic Testing: Theory and Applications, Volume 26 Issue 4, August 2010.

[6] Anbo Meng, Luqing Ye, Daniel Roy Pierre Padilla,"Genetic algorithm based multi-agent system applied to test generation", Computers & Education 49, 1205–1223, 2007.

[7] K.G.Srinivasa,K.R,Venugopal,L.M.Patnaik,"A Self-Adaptive Migration Model Genetic Algorithms for Data Mining Applications", In Proceedings of Inf. Sci.. 4295-4313,2007.

[8] Bloom (1956). Taxonomy of educational objectives: The classification of educational goals. In B. Bloom (Ed.), Handbook 1, cognitive domain. New York: Longman.

[9] Krathwohl, D. R., "A revision of Bloom's taxonomy: An overview, Theory into Practice", Vol.41 (4), pp 212-219,2002.

[10] Yeh L. C. and LIN S. P., "The research of Bloom's taxonomy for educational objects," Journal of Education Research, Vol.105, pp. 94-106, 2003.

[11] Khairuddin, N. N., and Hashim, K.," Application of Bloom's Taxonomy in Software Engineering Assessments", Proceedings of the 8th WSEAS International Conference on Applied Computer Scince, 2008.

[12] Gandomi, A. Yang, H., Talatahari, X.S S. & Alavi, A.H. (2013). Meta-heuristics in Modeling and Optimization, Metaheuristic Applications in Structures and Infrastructures, First edition, Elsevier, 2013, 1- 24.