Generating Parallel Mathematic Items Using Automatic Item Generation

Teguh Bharata Adji, Feddy Setio Pribadi, Hendro Eko Prabowo, R. Rosnawati, and Ariyadi Wijaya
Department of Electrical Engineering and Information Engineering, Engineering Faculty, Universitas Gadjah Mada
adjji@ugm.ac.id

Abstract. Computer assisted assessment (CAA) objective question has been proven to be able to overcome weaknesses in the evaluation process using paper. Problems that can be overcome include reducing the time needed for correction and the burden of the process of handling thousands of participants simultaneously. However, the preparation for objective questions still constrained by the provision of parallel questions. Parallel questions are made so that each examinee gets a different question but has the same level of difficulty. Manually creating parallel questions at this time is still experiencing problems, that is, in addition to costly, it also has problems when making parallel questions with the same level of difficulty. Therefore several researchers develop the Automatic Item Generation system to overcome these problems. Automatic Item Generation developed in this paper is used to make parallel questions for mathematical problems. The parallel questions can be generated as the AIG system has 3 components, namely stem, variables and options that can be dynamically changed using variables mapping, combination and permutation. The result is that the system is able to generate 70% of the variant questions from dataset questions of junior high school.

Keywords: Computer assisted assessment, Automatic item generation, Mathematics, Parallel questions.
INTRODUCTION

Assessment is an important process in learning (Hunt, Hughes, & Rowe, 2002; Mohler & Mihalcea, 2009). Through assessment, the teacher can find out how many students have understood the material and whether the student has fulfilled the planned learning objectives (Chang, Chuang, & Lin, 2010).

The rapid development of information and communication technology has led to a system that is able to assist the assessment process. The system is called Computer Assisted Assessment (CAA). Perez (Pérez & Alfonseca, 2005) stated that the CAA is a study of how a computer is automatically able to give a score (score) to the test participants' answers.

The CAA for objective-type questions has been proven to provide accurate assessment (Burrows, Gurevych, & Stein, 2015; Tian, 2009). However, the CAA for this type of question still has weaknesses in the process of preparing the item. Chang and Mark (Chang, Chuang, & Lin, 2010; Mark J. Gierl, 2015) stated that the process of preparing item items that are carried out through human roles requires a long time, hard effort and expensive costs. One way that can be done to overcome the problem of preparing question items is to use the Automatic Item Generation (AIG) system (Mark J. Gierl, 2015; Gierl, Lai, & Turner, 2012). In addition to cope with the problems in preparing assessment tools, AIG is also intended to make variations of questions so that they have the same level of difficulty (Chang, Chuang, & Lin, 2010; Mark J. Gierl, 2015).

This paper presents the results of the development of the AIG system, especially for mathematics subjects. Some researchers who have developed AIG for mathematics include (Chang, Chuang, & Lin, 2010; Mark J. Gierl, 2015). Chang tried to expand AIG that has the ability to produce math words. Meanwhile Mark developed an AIG system that could be used for formative assessment. However, all previous systems were yet to have mathematical editor.

Unlike the system that had been developed previously, the aim of this research is to develop AIG system that able to generate mathematical questions for the National Examination (UN / “Ujian Nasional”). Some features developed include text editors to enter the stem, equation editor to enter mathematical equations, and export file facilities.

DATASET

This study using the National Exam (UN) questions dataset on mathematics subjects for junior high school level. The dataset consists of 40 questions having topics as presented in Table 1.

As shown in Table 1, there are 9 topics tested. The topic about Numbers has 8 questions, Comparison there are 2 questions, geometry 11 questions, algebra 8 questions, Set 2 questions, trigonometry 2 questions, statistics and probabilistic 1 question, statistical 4 questions, and the last probabilistic 2 questions.

| No. | Question | Topic               |
|-----|----------|---------------------|
| 1,3,5,6,7,8,9,10 | Numbers |
| 2,11 | Comparison |
| 12,15,17,18,19,20,21,22 | Algebra |
| 13,14 | Set |
| 23,24 | Trigonometry |
| 16 | Statistical and Probability |
| 35,36,37,38 | Statistical |
| 39,40 | Probability |

METHOD

AIG can be described as a process for multiplying question items using a model and computer technology (Chang, Chuang, & Lin, 2010; Mark J. Gierl, 2015). The description implies that AIG has two main components, namely the question model and computer technology. The model was developed by an item designer in the field related to the subject. The model contains three elements, namely the statement (stem), variable and option (choice of answers) (Mark J. Gierl, 2015). Computer technology in that definition means a programming algorithm that is used to manipulate variables. Figure 1 is the AIG scheme proposed in this paper.
stem and the mathematical formula to generate the answer choices automatically. All input data is then processed using Algorithm to Generate Parallel Items, which results in the same complexity parallel questions.

As an illustration of how the system works, Figure 2 shows the process of entering data so that the system is able to generate parallel questions.

Figure 2. Illustration of the input process

Figure 2 shows an example of a question text used as input of the AIG system. The text is used as a basis for creating a stem that consists of two parts namely statements and variables. The variables are written in square brackets. The statement will be a question statement and the variable will contain the value that will be involved in the calculation process.

The next process is to give values to variables. The developed AIG system can accept variables in integer, float, and range forms. In addition, the system is able to provide variables that are functions of other variables. For example, variable \((n5)\) is a function of variable \((r1)\). The last process is to enter a mathematical formula that will be used to generate options automatically based on the value of the variables.

Figure 3. Illustration of entering external formula

The developed AIG system also can solve the option generation using external variables as illustrated in Figure 3. Figure 3 shows that variable \((p)\) is not known (external) in the stem, but this variable is a variable needed in the calculation process to produce an option. This process can be done using programming language editor. In other words, the developed AIG system is capable of behaving like a programming language editor that is very flexible for users.

The developed AIG system is a web-based system that is developed using python and html programming languages. Python is one programming language that has a lot of libraries such as nltk, numpy, and sympy that is powerful in manipulating texts and numbers. The three libraries are very helpful in developing AIG for mathematics subject.

RESULT AND DISCUSSION

Figure 4 and Figure 5 are the core parts of the developed AIG system in this study. In Figure 4, the web page for defining stems and variables is shown in the **Stem Soal** dialog box. The picture shows a simple text editor to enter Stem. Stem can be in the form of text, images or mathematical equations.

Figure 4. Form to enter stem

Figure 5 shows the following process after the user enters the data on the page shown in Figure 4. Figure 5 shows the process of initializing the variable values and defining the mathematical formula used for calculating the value of the variables that will be displayed on the options on Figure 6.

Figure 5. Form to enter values of variables and option formulas
One of the strengths of the developed AIG system is the mathematics editor feature that can capture arithmetic structures. The feature will ease the user to type flexible mathematical questions without the help of the system programmer. Figure 5 shows the illustration of developing questions by involving complex mathematical formulas. Figure 6 shows the results of the generation process as in Figure 5.

Table 2 shows the results of the system tests on 40 math UN questions. The table shows that not all forms of mathematical questions can be accommodated by the developed AIG system.

Table 2. System test result

| No. | Question | Topic      | Auto Generate |
|-----|----------|------------|---------------|
| 1,3,5,6,7,8,9,10,13 | Numbers | 1,5,7,8,10 | Y              |
| 2,11 | Comparison | 2,11       | N              |
| 4,25,26,27,28,29,30,31,32,33,34 | Geometry | 26,27,28,30,31,32,33,34 | Y |
| 12,15,17,18,19,20,21,22 | Algebra  | 12,22      | Y              |
| 13,14 | Set      | 14         | Y              |
| 23,24 | Trigonometry | 23,24     | Y              |
| 35,36,37,38 | Statistics and Probability | 35,36 | N |
| 39,40 | Probability | 39,40 | N              |

Table 2 also shows that among 40 tested items, 22 items can be used as basic stems to generate parallel questions, while 18 items cannot.

Some types of questions that cannot be used as basic stems in generating parallel questions are: 1) question naturally cannot be paralleled such as questions no. 4, 37, 29, 25, and 19, 2) questions requiring image manipulation such as questions no. 24, 23, 13, and 9, 3) questions involving symbolic calculation such as questions no. 15, 17, 20, and 21, 4) questions involving inequality such as questions no. 3, 18.

**CONCLUSION**

About 55% of the mathematical questions can be processed by the developed AIG system to generate parallel questions. This capability is supported by mathematics editor so that users can type several equations flexibly.

Some future research to improve the AIG system is by developing libraries that able to calculate equations involving symbolic calculation and inequality, and by developing graphical editors that able to create graphs dynamically.

*****

**REFERENCES**

Burrows, S., Gurevych, I., & Stein, B. (2015). The eras and trends of automatic short answer grading. International Journal of Artificial Intelligence in Education. http://doi.org/10.1007/s40593-014-0026-8

Chang, K., Chang, W., & Lin, Y. (2010). Online Math Word Problem Test System Based on Dynamic Item Generation Mechanism. IEEE: 2010 International Computer Symposium (ICS2010), 354–357.

Hunt, N., Hughes, J., & Rowe, G. (2002). Formative Automated Computer Testing (FACT). British Journal of Educational Technology, 33, 525–535. http://doi.org/10.1111/1467-8535.00289

Mark J. Gierl, H. L. (2015). Using Automated Processes to Generate Test Items And Their Associated Solutions and Rationales to Support Formative Feedback. Interaction Design and Architecture, 25, 9–20.

Mohler, M., & Mihalcea, R. (2009). Text-to-text semantic similarity for automatic short answer grading. In Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics on - EACL ’09 (pp. 567–575). http://doi.org/10.3115/1609067.1609130

Pérez, D., & Alfonseca, E. (2005). Application of the Bleu algorithm for recognising textual entailments. Workshop Recognising Textual Entailment, 1–4.

Tian, Y. (2009). Online automated assessment of student english-Chinese translation. In 2009 WRI World Congress on Computer Science and Information Engineering, CSIE 2009 (Vol. 5, pp. 339–342). http://doi.org/10.1109/CSIE.2009.785
Appendix

Types of Question in UN

- **Questions naturally cannot be paralleled**
  Roni diperbolehkan ibunya untuk mengambil satu permen dari sebuah kantong. Dia tidak dapat melihat warna permen tersebut. Banyaknya permen dengan masing-masing warna dalam kantong tersebut ditunjukkan dalam grafik berikut.

![Pie Chart](image)

Berapakah peluang Roni mengambil sebuah permen warna merah?

- **Questions involving mathematical logic**
  “Toko Pakaian”
  Ada empat toko menjual jenis barang yang sama. Daftar harga barang dan diskon seperti pada tabel.

| Barang | Harga       | Diskon |
|--------|-------------|--------|
| Baju   | Rp80.000,00 | 25%    |
| Celana | Rp100.000,00| 10%    |

| Barang | Harga       | Toko Rame | Toko Damai | Toko Seneng | Toko Indah |
|--------|-------------|-----------|------------|-------------|------------|
| Baju   | Rp80.000,00 | 25%       | 20%        | 15%         | 10%        |
| Celana | Rp100.000,00| 10%       | 15%        | 20%         | 25%        |

Ali akan membeli sebuah baju dan celana di toko yang sama. Di toko manakah Ali berbelanja agar diperoleh harga yang paling murah?

- **Questions that require manipulation of images dynamically**
  Perhatikan gambar di samping!
  Besar $\angle ABC$ adalah ....

![Diagram](image)

- **Questions that include the calculation process by involving symbolic mathematic**
  Diketahui fungsi $f(x) = 5x - 15$. Nilai $f(a + 2)$ adalah ....