

Literature Review on Automatic MCQs generation from text.

1st Vyshak Puthusseri

Dept of Computer Applications
College of Engineering Trivandrum
Kerala, India
vyshakputhusseri@cet.ac.in

2nd Baby Sylal

Dept of Computer Applications
College of Engineering Trivandrum
Kerala, India
syla@cet.ac.in

Abstract—Multiple Choice Questions (MCQs) has a major role in the assessment test conducted for major exams in the education field. The process of generating the MCQs manually was time-consuming and requires a huge human expert knowledge. Usually, the MCQs contain one right answer and 3 distractors. A distractor can be a wrong answer for that particular question which is having similarity with the correct answer. In the real world, the MCQ questions are created by domain-specific experts. They have to analyze the text in depth to make questions that can examine the knowledge level of the candidates. Our study focuses on generating distractors which are having the same semantics but which are not suitable as a correct answer. The first step is the sentence selection which has to be informative. From the selected sentence MCQs has to be generated along with distractor.

Index Terms—NLP, distractor generation, MCQ generation

I. INTRODUCTION

The knowledge level of a person is assessed through the examinations. The current education system uses various methods for the assessments and multiple-choice questions are the most prominent ones in the case of the competitive exams. Even though the multiple-choice questions are having advantages like easy evaluation(also digital evaluation), reliable scoring mechanism, etc, the process of generating the MCQs is difficult. It requires a domain expert for the cognitive creation of the questions along with the options. The quality of the generated questions can be evaluated through various methods. If a human was evaluating the quality of the MCQs, then the rubrics will be the amount of distraction that the wrong answer can make in the user.

The research in this field started in the late 1990s itself. Among them, a popular one is by David Coniam in 1997 which describes three ways of generating the MCQs [1]. The main aim of this review is to summarise the recent research in the field of generating the MCQs. The distractor generation for the MCQs is also another strong area of research related to this field. Another research area is on the evaluation of the generated questions. It's observed that there is no standardized way for evaluating the quality of the generated questions and the distractors. As the distractor has to distract the candidate who is taking the test.

II. MULTIPLE CHOICE QUESTIONS

Most of the competitive exams use the MCQs as the standard form of questions for their assessments. Objective

type questions can be considered as the superclass of the MCQs [2]. Some of the researchers also called the MCQs as factual questions by the assumption that the questions are generated from the facts from the text ([3],[4],[5]). An MCQ contains three elements: a stem, a key, and a set of distractors. A stem is a sentence from which the questions are generated. The target word or the key is the correct answer for that particular question. The set of distractors which seems to be the right answer but are not. They are used to distract the candidate from choosing the right answer. For example:

The planet which is farthest from the sun is

- a). Saturn
- b). Uranus
- c). Neptune
- d). Pluto

In the above example, the stem is "The planet which is farthest from the sun is". Neptune is the correct answer among the four given options. And the other three wrong options are called distractors.

III. RESEARCH MOTIVATIONS AND OBJECTIVES

The main aim of MCQs is to evaluate the knowledge level of a huge set of candidates. But the process of generating the MCQs is complex, which motivates the researchers in the natural language processing community to explore the possibilities of automatic multiple choice question generation. Another use of the automated MCQ generator is to help the candidate to revise the content by evaluating himself from the generated questions. Many researchers have undergone various domains for the MCQs generation which includes language learning to the assessment of knowledge in a particular domain([6],[7],[8]).

The main objective of this study is to learn the workflow of the automatic multiple-choice questions generation and critical analysis of the techniques used to generate the questions from the paragraph and the ways which are used to evaluate the quality of the generated questions.

IV. PREVIOUS WORKS

The generation of the automatic multiple-choice questions follows a similar workflow, even though the architecture may slightly vary in different studies. Graesser and Wisher prove in

their work that the number of distractors is three for a standard MCQ[9].

Different methods in NLP can conceivably contribute to the automatic generation of MCQs. Mitkov and Ha (2003)[10] used the WordNet (Fellbaum 1998)[11] as a lexical dictionary resource for performing the MCQ generation process. Susanti et al. (2015)[12] also worked on WordNet for generating the MCQs. As the problem becomes harder among the community, few studies emerged focusing on a particular section of the MCQ generation process.

A. Sentence Selection

Sentence selection is an important step in the MCQ generation. All the sentences in the paragraph may not be appropriate for generating valid questions. Only the sentence with the informative content can act as the candidate sentence. Several approaches had been used in the literature for sentence selection. Based on each factor those approaches can be summarised as follows. One approach is to choose the word count for the candidate sentence selection. A short sentence cannot be used to produce the MCQ as it lacks in the reproduction of the information. In the case of a long sentence also the difficulty in the generation process is high, as they contains complex relations and the dependencies are hard to parse. [13],[14],[15]

The second approach for the candidate sentence selection is based on the information of the parts-of-speech. Mitkov et al. used the verb forms and the sentence length for the selection process[16]. Lin, Sung, and Chen used the adjective-noun pairs for the selection of the sentence[17].

Mukta Majumder et al. used the topic word and parse structure similarity for the sentence selection process[18]. Hybrid approaches are also used in the literature. Agarwal and Mannem, 2011 [13] selected the sentence based on factors like the position of the sentence, frequency of the words, abbreviations, starting word of the sentence, proper nouns. The Keys are chosen by analyzing the parts of speech of that particular sentence, frequency of the word in the entire document, and the similarity with the title of the document.

B. Distractor generation

The quality of the MCQ depends on the generated distractors. The distractors should be able to distract the candidate from choosing the option. If the distractors are not enough to confuse the candidate from choosing the option he can easily identify the right answer. Several studies based on the parts-of-speech information had occurred for the generation of the distractors[19]. Another approach was to use the frequency of the word in the given paragraph[8]. Goodrich proposed a method to determine the distractor's efficiency [20]. Two concepts are used for measuring efficiency: potency and discrimination. Potency tells the percentage of candidates who choose a particular option. If more students are choosing a distractor, it indicates the distracting efficiency of the distractor. Discrimination has to do with the ability to differentiate students of different levels of proficiency.

CONCLUSION

MCQs are one of the best methods for the assessment tests. This study analyses the works that are done in the field of automatic MCQ generation and its workflow. From the study, we observed that the process of generating the MCQs will differ as to the domain of input can vary. It will be better to choose a hybrid approach for sentence selection. We had identified that the use of deep learning techniques can outperform the present results of the distractors generation. The problem is still having a huge scope for research in the future.

REFERENCES

- [1] David Coniam, "A Preliminary Inquiry Into Using Corpus Word Frequency Data in the Automatic Generation of English Language Cloze Tests", Proceedings of the HLT-NAACL 03 Workshop on Building Educational Applications Using Natural Language Processing - Volume 2, Association for Computational Linguistics, Stroudsburg, PA, USA
- [2] R. D. Nielsen, J. Buckingham, G. Knoll, B. Marsh, and L. Palen, "A taxonomy of questions for question generation," in Proceedings of the Workshop on the Question Generation Shared Task and Evaluation Challenge, 2008.
- [3] R. Mitkov, L. An Ha, and N. Karamanis, "A computer-aided environment for generating multiple-choice test items," *Nat. Lang. Eng.*, vol. 12, no. 2, pp. 177–194, Jun. 2006.
- [4] M. Liu, V. Rus, and L. Liu, "Automatic chinese factual question generation," *IEEE Transactions on Learning Technologies*, vol. 10, no. 2, pp. 194–204, 2017.
- [5] D. Metzler and W. B. Croft, "Analysis of statistical question classification for fact-based questions," *Inf. Retr.*, vol. 8, no. 3, pp. 481–504, May 2005.
- [6] J. Mostow and W. Chen, "Generating instruction automatically for the reading strategy of self-questioning," in *AIED 2009: 14th International Conference on Artificial Intelligence in Education*, 2009, pp. 465–472.
- [7] J. C. Brown, G. A. Frishkoff, and M. Eskenazi, "Automatic question generation for vocabulary assessment," in Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing, ser. HLT '05. Stroudsburg, PA, USA: Association for Computational Linguistics, 2005, pp. 819–826.
- [8] N. Karamanis, L. A. Ha, and R. Mitkov, "Generating multiple choice test items from medical text: A pilot study," in Proceedings of the Fourth International Natural Language Generation Conference, ser. INLG '06. Stroudsburg, PA, USA
- [9] Graesser, Arthur C. and Wisher, R. A., "Question Generation as a Learning Multiplier in Distributed Learning Environments", U.S. Army Research Institute for the Behavioral and Social Sciences (Alexandria, Va), A654993, 2001
- [10] Mitkov, R., Ha, L.A. (2003). Computer-aided generation of multiple-choice tests. In Proceedings of the HLT-NAACL 03 Workshop on Building Educational Applications Using Natural Language Processing - Volume 2, Association for Computational Linguistics, Stroudsburg, PA, USA, HLT-NAACL-EDUC '03. <http://dx.doi.org/10.3115/1118894.1118897>, (pp. 17–22).
- [11] Fellbaum, C. (1998). WordNet: An Electronic Lexical Database. Cambridge: MIT Press.
- [12] Susanti, Y., Iida, R., Tokunaga, T. (2015). Automatic generation of english vocabulary tests. In Proceedings of the 7th International Conference on Computer Supported Education. INSTICC, Setubal, (pp. 77–87).
- [13] Agarwal, M. and Mannem, P. (2011) 'Automatic gap-fill question generation from text books', Proceedings of the 6th Workshop on Innovative use of NLP for Building Educational Applications, Hyderabad, AP, India, pp.56–64.
- [14] E. Sumita, F. Sugaya, and S. Yamamoto, "Measuring nonnative speakers' proficiency of english by using a test with automatically-generated fill-in-the-blank questions," in Proceedings of the Second Workshop on Building Educational Applications Using NLP, ser. EdAppsNLP 05. Stroudsburg, PA, USA: Association for Computational Linguistics, 2005, pp. 61–68.

- [15] "A system for generating multiple choice questions: With a novel approach for sentence selection," ACL-IJCNLP Workshop on NLPTEA 2015, pp. 64–72, 2015.
- [16] I. Aldabe, M. Maritxalar, and R. Mitkov, "A study on the automatic selection of candidate sentences distractors." in AIED 2009: 14th International Conference on Artificial Intelligence in Education Workshops Proceedings, 2009, pp. 656–658.
- [17] Y.-C. Lin, L.-C. Sung, and M. C. Chen, "An automatic multiple choice question generation scheme for english adjective understanding," in Workshop on Modeling, Management and Generation of Problems/Questions in eLearning, the 15th International Conference on Computers in Education (ICCE 2007), 2007, pp. 137–142.
- [18] Mukta Majumder, Sujana Kumar Saha "A System for Generating Multiple Choice Questions: With a Novel Approach for Sentence Selection", Proceedings of the 6th Workshop on Innovative use of NLP for Building Educational Applications, Hyderabad, AP, India, pp.56–64.
- [19] R. Correia, J. Baptista, N. Mamede, I. Trancoso, and M. Eskenazi, "Automatic generation of cloze question distractors," in Proceedings of the Interspeech 2010 Satellite Workshop on Second Language Studies: Acquisition, Learning, Education and Technology, Waseda University, Tokyo, Japan, 2010.
- [20] Goodrich, H. C., "Distractor efficiency in foreign language testing", TESOL Quarterly, vol. 11, no. 1, pp. 69–78, 1977.