Factual Question Generation System Using Pro Semantic Ranker

J Dafni Rose\textsuperscript{a,1}, U Sakthi\textsuperscript{b} and S S Ajithkumar\textsuperscript{c}

\textsuperscript{a,b}Dept of CSE, St. Joseph's Institute of Technology, Chennai, India
\textsuperscript{c}Member Technical Staff Zoho Corporation, Chennai

Abstract. In today's internet world the amount of materials available to learn and gain knowledge is immense in numbers, which has given access to a lot of people to gain knowledge easily. Is it possible to find out if someone has read this sentence? To arrive at a conclusion of whether or not someone has read something, we can ask them to summarize its contents or question them about it. We can start with basic questions that deal with the title or the abstract, and progress to more challenging questions. A good example of this process would be a primary school teacher questioning his or her students on the basics of what they are reading to make sure that they are learning. It is a time-consuming process to create the basic questions and reading assessments that are commonly used, and can be quite taxing on the educator. This paper focuses on automating that time-intensive process. To be precise, this paper deals with the problem of generation of the factual questions in an automated manner from stand-alone texts.

Keywords. Question Generation, Pro Sematic Ranker, Natural Language Processing, Ranking

I. Introduction

Asking questions from learners is said to facilitate interest and learning, to recognize problem learning areas to assess vocabulary and reading comprehension, to provide writing support, to support inquiry needs etc. The process of creating exercises and tests for practice is time and labour consuming, but has been an integral part of teaching for the longest time. With the rapidly expanding amount of learning material on the internet, there is an urgency to make this task scalable and less extensive. Subsequently, there is also a rising demand for intelligent tutoring systems that utilizes computer-generated or computer-assisted instructional material to evaluate the learner’s comprehension. Inevitably, the task of the Automatic Question Generation (QG) caught the attention of NLP researchers from across the globe. Automatic QG has been defined as the task of automatically generating questions from various inputs like raw text, database or semantic representation\textsuperscript{1}. Apart from its direct application in the educational domain, in general, the core NLP areas like Question Answering, Dialogue Generation, Information Retrieval, Summarization, etc. also benefit from large scale

\textsuperscript{1}DafniRose.J, Dept of CSE, St. Joseph’s Institute of Technology, Chennai; E-mail: jdafnirose@yahoo.co.in
automatic Question Generation. The architecture diagram describes the main question. The paper is organized as follows. Section II includes the survey of how the questions are scored and ranked according to features of the source sentences, input sentences, the question, and the transformations used in the generation. Section III specifies the system functionality of pro semantic ranker. Section IV specifies the three stages of pro semantic ranker. Section V includes the performance analysis of proposed fog computing based smart knowledge based system. In the last section, we conclude some considerations and discuss future work of this research work.

2. Related Works

For the purpose of generating meaningful questions from the given natural language sentences, a neural encoder-decoder model is utilized [1]. The input text is read along with the answer position to generate an answer input representation. Question generation from supplied text using neural networks that automatically crawled and processed large-scale QA pairs from a community QA website was proposed in [2,1] and used as training data. Application of rules and templates to the given input’s syntactic representation is how sentence-to-question transformation is carried out [3,7]. High levels of abstraction cannot be obtained in question generation using these syntactic representations. The semantic roles of words must be taken into consideration for there to be an elevated level of comprehension in the generated questions [4,5]. Removal of stop words is performed and Natural Language Processing is carried out. Key phrase extraction is done by using TF – IDF algorithm and checking the wiki presence of terms. TheWordNet tool is used to for the creation of triplets for the generation of question papers as well as for the conduction of input clarity [6]. Automatic Question Generation (AQG) takes a passage and an answer as input. The aim of the AQG is to generate a respective question based on the input provided to it [5,3]. A process of generating questions that attempts to mimic a human approach of

![Figure 1. Architecture diagram of pro semantic ranker](image-url)
creating an initial draft of questions and subjecting that to refinement was proposed in [8] as a solution to this shortcoming [9].

3. Structure of Pro Sematic Ranker

The structure, views of a system and its behaviour is defined by the system architecture, which is conceptual. An architecture description represents the system formally, and is designed to support its structural and behavioural reasoning. Figure 1 depicts the architecture diagram of a pro semantic ranker. The first step is to extract relevant questions from the text file provided as input, containing the actual answers needed for the questions. After removing redundant text found in the file, relevant sentences containing answers can be extracted. Stage 2 involves the system taking a declarative sentence as input and processing it to generate a set of possible questions as output. The generated questions may be vague and users of this system only take into consideration a short set of candidate questions, making it very important to identify and present the questions that are most likely to be accepted first. Stage 3 of the system utilizes a statistical model to determine the question acceptability of each candidate question generated in stage 2 and ranks it accordingly. A data structure stores the generated questions and displays them in sets to the user when he or she is finished with studying. These questions can be taken as a form of revision or as a test.

4. Stages of Pro Sematic Ranker

The architecture diagram of the system defines the architecture of the main question generator, as well the process flows and dependencies required generating questions from the given input file. This system progresses over three stages namely, NLP transformation, Question transducer, and Ranking along with User Interface.

Stage 1- Transforming Source Sentence: Two forms of transformations are included in this operation. One of them deals with the extraction of simplified factual statements in sets from complex input sentences that have embedded constructions. The other form of transformation deals with replacing pronouns with their respective antecedents which takes place after the extraction of simplified statements.

Stage 2: Question Transducer: The system generates a set of possible questions derived from the declarative sentence it receives as input. Answer phrases that may potentially be targets for WH-movement are identified and converted into question phrases. Questions are not generated from all of the phrases of these types, or from verb phrases.

Stage 3: Question Ranker and Statistical Model: Users of this system only take into consideration a short list of candidate questions, making it very important to identify and present the questions that are most likely to be accepted first. Stage 3 takes the candidate questions obtained from Stage 2 and ranks them on the basis of question acceptability using a statistical model. The statistical model, along with its features is described in this section. The acceptability y of a question x is modelled is

\[ y = w^T f(x) \]
Here, $f$ is called as feature function. It produces a real-valued numbers vector that considers the various aspects of the question after processing the question provided as its input. The $w$ used here is a vector of real-valued weight parameters for each feature in the feature vector returned by $f$. A weight vector $w^*$ is achieved that decreases the total adjusted errors on the training data values, matter to a consequence on the type of the weight vector, to evade over fitting.

\[
\hat{w} = \arg\min_w \sum_{i=1}^{N} (y_i - w^T f(x_i))^2 + \lambda |w|^2
\]

In the equation describes above, $x_i$ is a particular occurrence in the training data, possessing an tolerability range of $y_i$, and $N$ represents the count of trainings value of data. Such a format of linear regression is termed as edge regression.

5. Performance Analysis

The performance efficiency of factual question generation system using pro semantic ranker is analysed with various algorithms as shown in Table 1. The score obtained by various question generator system and the score is based on how much the question is semantically correct and whether it makes any sense is represented in Figure 2. From this table we can observe that the Pro-Semantic rankerSystem generates better and more semantically correct questions. Useful questions may often come across as semantic transformations of declarative sentences, or even as lexical and syntactic sometimes. This is carried out by transforming lexical items, semantics and syntactic structure.

| Algorithms                  | Average Score on 10 |
|-----------------------------|---------------------|
| Syntax Based                | 3.3                 |
| Semantic Based              | 5.8                 |
| Pro-Semantic Ranker System  | 8.2                 |
6. Conclusion

The proposed system of factual question generation helps in generating better and semantically correct questions when compared to the other traditional framework. It can also be widely adopted in online examinations for reading comprehensions by which malpractice can be avoided and can also be used to generate small unit tests in schools which enables the students to learn the concepts then and there. Automatically generating factual questions is a problem that is only slightly acknowledged in this proposed process. The system utilizes high-level semantic types with a super sense tagger to annotate source sentences and thereby generate question phrases.

References

[1] Qingyu Zhou, Nan Yang, Furu Wei, Chuanqi Tan, Hangbo Bao, Ming Zhou. Neural Question Generation from Text: A Preliminary Study. Apr 06, 2017.
[2] Nan Duan, Duyu Tang, Peng Chen, Ming Zhou. Question Generation for Question Answering. Sep 2017.
[3] Onur KEKLICH, Automatic Question Generation using Natural Language Processing Techniques, July 2018 IZMIR.
[4] K.Surya, R.Aarthipriya, S.Iswarya. Automatic Question and Answer Pair Generation. Special Issue Published in International Journal of Trend in Research and Development (IJTRD), ISSN: 2394-9333, www.ijtrd.com, Sep 15, 2016.
[5] Xingdi Yuan, Tong Wang, Caglar Gulcehre, Alessandro Sordoni, Philip Bachman, Sandeep Subramanian, Saizheng Zhang, Adam Trischler, Machine Comprehension by Text-to-Text Neural Question Generation, Aug 3, 2017.
[6] Aleena Susan Mathew, Vidya. N, Implementation of Automatic Question Paper Generator System, International Research Journal of Engineering and Technology (IRJET) Volume: 06 Issue: 02, Feb 2019.
[7] Han Xiao, Feng Wang, Jianfeng Yan, Jingyao Zheng. Dual Ask-Answer Network for Machine Reading Comprehension, Sep 10, 2018.
[8] Xingdi Yuan, Tong Wang, Caglar Gulcehre, Alessandro Sordoni, Philip Bachman, Sandeep Subramanian, Saizheng Zhang, Adam Trischler, Machine Comprehension by Text-to-Text Neural Question Generation, May 15, 2017.
[9] Ambeth Kumar.V.D. Precautionary measures for accidents due to mobile phone using IOT. Clinical eHealth, Volume 1, Issue 1, March 2018, Pages 30-35.