MANTECH Publications

Journal of Artificial Intelligence, Machine Learning and Soft Computing
Volume 4, Issue 1, January-June 2019

Automatic Question Generation From Given Paragraph

Deokate H. G,[1] Jogdand P. P[2], Satpute P. S[3], Shaikh S. B.[4]

SVPM's, SPPU, INDIA

Email id: *prasadjogdand2552@gmail.com*

DOI: *http://doi.org/10.5281/zenodo.2652778*

Abstract

In this project we have presented an approach to generate questions from a paragraph and the size of the paragraph is defined by its scope. A mix of syntax and semantic based approach to natural language processing is used to generate the questions from the paragraph. Important sentences from the paragraph are selected based upon the certain features and the questions are generated for these selected sentences. Our system implements generation of question from paragraph and also generating simple and complex types of questions. And the research till date works on either implementing question generation from single sentences or implementing generation of simple questions from paragraph or implementing question generation of complex questions from paragraph.

Keywords: *Question Generation, Tokenization, POS, NLP, Data mining, etc*

INTRODUCTION

Researchers have proved that the humans are not skilled in generating the questions. The Automatic Question Generation (AQG) is necessary one to generate good questions. Questions are indeed to have in-depth knowledge in any domain. Question generation can help a person to generate questions from the given text automatically. It is a process in which given an input text to the system it will create reasonable questions from the input as output. The potential benefits of using automated systems to generate questions helps reduce the dependency on humans to generate questions and other needs associated with systems interacting with natural languages. Question generation can

be applied in many fields like intelligent tutoring systems, MCQ generation, FAQ generation etc. New e-learning methodologies require assessment procedures that automatically measure the student's achievements during the teaching and learning process. These procedures must be compatible with other solutions that provide personalized feedback to students for understanding and improving the quality of their learning experience. Many e-learning proposals use Multiple-Choice Questions (MCQ) as an assessment tool. Basically, a MCQ consists of a question text and a few (e.g. four) choices, from which one is the correct answer and the others are incorrect alternatives (called distractors). MCQs are labor intensive, time consuming and difficult to construct. For this reason, recent efforts have focused on the automatic generation of well-constructed MCQs, mainly for vocabulary assessment or grammar exercises. The automated creation of tests involves generating distractors based on certain knowledge and, subsequently, using these distractors to create the assessment test.

Motivation

Our motivation is to demonstrate that there is still value in single-sentence QG, and given the right approach, we can escape the realm of the factoid and pose questions that are not just syntactically and semantically sound but also more useful pedagogically.

We want to demonstrate that by leveraging the semantic content of learning materials, we can generate questions that are more pedagogically-useful than factoids.

The system will take a paragraph as input and generate important questions from the important sentences extracted from the paragraph. It will generate different wh-type of questions from those selected sentences.

The questions will be generated from both simple and complex sentences.

Automatic Question Generation (AQG) tool that serves as a guide for students by providing questions which are meaningful and grammatically correct. A system to assess the vocabulary of the human knowledge is given in. Questions are to be generated automatically to test the vocabulary.

Problem Definition

The existing question generation system in paves a way to generate questions based on extracting key phrases by using unsupervised method and then mapping it with Wikipedia article in order to get

additional information about the key phrases. The obtained information may tend to be implicit. Nevertheless, an analysis of the existing systems for generating MCQ tests has allowed to identify some relevant drawbacks.

Firstly, before starting the execution of the system, the teachers must define the input knowledge database (anontology, a corpus, or a conceptual map, for instance) that will be used to extract distractors. These knowledge models are usually created for a concrete learning domain (a course or a subject).

Secondly, the quality of distractors depends on the expressiveness and richness of this input knowledge.

Thirdly, the automatic solutions restrict the type of questions to be included into a test. These solutions are based on the use of simple test templates. And, finally, the developed systems are not network-accessible, not even they were designed to be integrated into learning platforms. The system will take a paragraph as input and generate important questions from the important sentences extracted from the paragraph. It will generate different wh-type of questions from those selected sentences. The questions will be generated from both the complex and the simple sentences. It will consider only those complex sentences which consists the following discourse connective: because, and, since, when, as a result, for example, for instance.

LITERATURE SURVEY

Automatic Question Generation System

The process of automating the question generation consists of many tasks. Selecting the target content (what to ask), question type (who, why, how) and actual question generation are the major issue of Automatic Question Generation. Certain definitions retrieved is available in Wikipedia either directly or is the outcome of executing set of sub queries for each key phrase categories The problem in the existing system is that some of the definition sentences which are taken out from Wikipedia were implicit. The proposed system overcomes the problems by using Supervised Learning Approach, Naïve Bayes method. It also extends its work to use Summarization, Noun Filtering and Question Generation in the aim of generating semantically correct questions.

Semantics and service technologies for the automatic generation of online MCQ tests

Active learning requires that students receive a continuous feedback about their understanding. Multiple-Choice Questions (MCQ) tests have been frequently used to provide students the required feedback and to measure the effectiveness of this learning model. To construct a test is a challenging task, which is time consuming and requires experience. For these reasons, research efforts have been focused on the automatic generation of well-constructed tests.

The semantic technologies have played a relevant role in the implementation of these test generation systems. Nevertheless, the existing proposals present a set of drawbacks that restrict their applicability to different learning domains and the type of test to be composed. In this paper, we propose a service-oriented and semantic-based system that solves these drawbacks. The system consists of a dynamic strategy of generating candidate distractors (alternatives to the correct answer), a set of heuristics for scoring the distractors' suitability, and a selection of distractors that considers the difficulty level of tests. Besides, the final version of tests is created using the Google Form service, a de-facto standard for elaborating online questionnaires.

Automatic Question Generation from Text for Self-Directed Learning

Question generation from text is a Natural Language Generation task of vital importance for self-directed learning. Learners have access to learning materials from a wide variety of sources, and these materials are not often accompanied by questions to help guide learning.

Prior question generation techniques have focused primarily on generating factoid questions, which are often not the most pedagogically important questions for a learner. Furthermore, prior techniques have not fully leveraged the semantic content of learning materials and have not often been evaluated in a pedagogically-inspired framework. This thesis introduces a novel template-based approach to question generation that combines semantic roles with a method of generating both general and domain-specific questions. We evaluate our approach in a way that is mindful of the context in which the generated questions are to be used. This evaluation shows our approach to be effective in generating pedagogically-useful questions.

Automatic Question Generation using Discourse Cues and Distractor Selection for Cloze Questions

A question may be either a linguistic expression used to make a request for information, or else the request itself made by such an expression. This information may be provided with an answer. Asking questions is a fundamental cognitive process that underlies higher-level cognitive abilities such as comprehension and reasoning. The ability to ask questions is the central cognitive element that distinguishes human and animal cognitive abilities. Questions are used from the most elementary stage of learning to original research. Question Generation (QG) is the task of automatically generating questions from various inputs such as raw text, database, or semantic representation.

Ultimately, QG allows humans, and in many cases artificial intelligence systems, to understand their environment and each other. Research on QG has a long history in artificial intelligence, psychology, education, and natural language processing.

Automatic Factual Question Generation from Text

Texts with potential educational value are becoming available through the Internet (e.g., Wikipedia, news services). However, using these new texts in classrooms introduces many challenges, one of which is that they usually lack practice exercises and assessments. Here, we address part of this challenge by automating the creation of a specific type of assessment item. Specifically, we focus on automatically generating factual WH questions. Our goal is to create an automated system that can take as input a text and produce as output questions for assessing a reader's knowledge of the information in the text. The questions could then be presented to a teacher, who could select and revise the ones that he or she judges to be useful. After introducing the problem, we describe some of the computational and linguistic challenges presented by factual question generation.

We then present an implemented system that leverages existing natural language processing techniques to address some of these challenges. The system uses a combination of manually encoded transformation rules and a statistical question ranker trained on a tailored dataset of labeled system output. We present experiments that evaluate individual components of the system as well as the system as a whole. We found, among other things, that the question

ranker roughly doubled the acceptability rate of top-ranked questions.

ARCHITECTURAL DIAGRAM

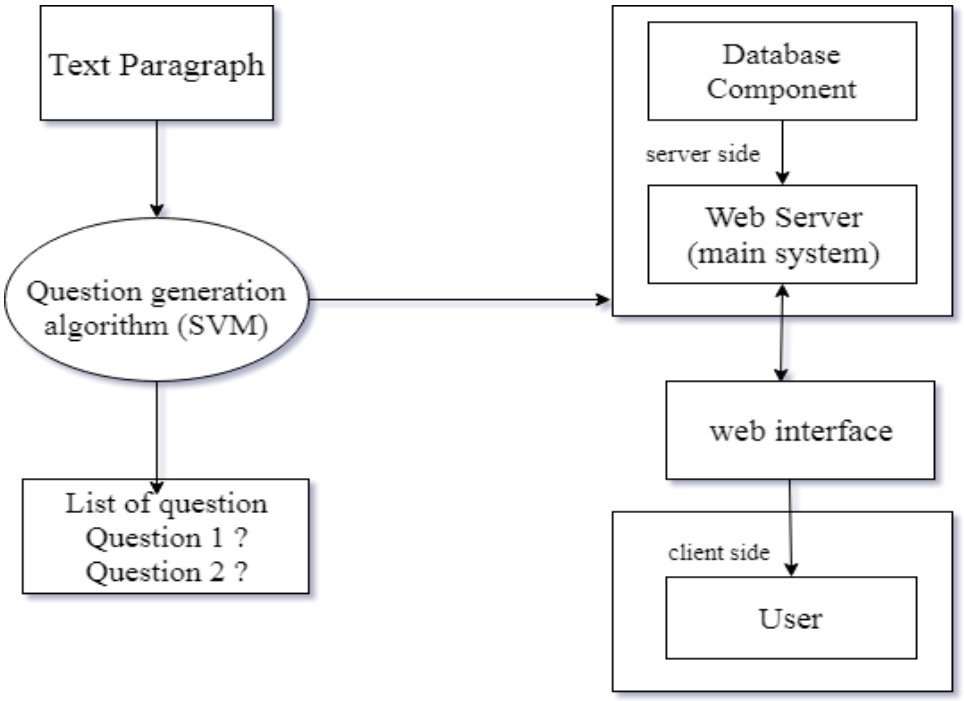

Figure: Architecture Diagram

ALGORITHM

Support Vector Machine

Classifying data is a common task in machine learning. In machine learning, support vector machines (SVMs, also support vector networks [1]) are supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis. Given a set of training examples, each marked as belonging to one or the other of two categories, an SVM training algorithm builds a model that assigns new examples to one category or the other, making it a non-probabilistic binary linear classifier (although methods such as Platt scaling exist to use SVM in a probabilistic classification setting).

In addition to performing linear classification, SVMs can efficiently perform a non-linear classification using what is called the kernel trick, implicitly mapping their inputs into high dimensional feature spaces.

Methods used in Question Generation System

Part of Speech tagger

Part-of-Speech Tagger identifies part of speech of each word, such as noun, verb, adjective, and so on. In our system, we preprocessed the sentences by adding tags, which were also features of the word, to each word through Part-of-Speech Tagger.

Dependency Parser

A Dependency Parser analyzes the grammatical structure of the sentence and that provides a view of relationships between words and words. The result helps us to realize the roles that the word playing is in a sentence.

Logistic Regression

Logistic regression is a method of statistically analyzing data which is also used to understand the relationship between variables, thereby presenting suitable for the problem of general binary classification. The calculation is precise and specific at the time of classification, the calculation speed must be fast. In addition, little storage resources are needed. The classification can be modeled directly, without the need to presume data distribution in advance, thus avoiding the problem of inaccurate distribution of hypotheses. Not only categories, but the approximate probability can be predicted, as well as the task of ranking.

SYSTEM DESCRIPTION

Relevant Mathematics Associated with the Project

Sr. No	Description	Observation
1	Problem Description and System	
	Let S be Closed system defined as S = { Ip, Op, Ss, Su, Fi,A} To select the training documents and give the path of the folder and perform various actions from the set of actions A, so that Su state can be attained.	System
	S={Ip,Op,Ss,Su,Fi,A} Where, IP1={Username,Password}	State of available document for training and testing
	Set of actions=A={F1,F2,F3,F4} Where,	

	F1=Tokenization F2=Removing Stop words F3= Keyword Extraction F4=Comparison Ss={rest state, login state, upload a document, Preprocessing, Keyword Extraction, Question Generation} Su- success state if question generated Successfully Fi- failure state if failed to generate question.	

System S can be defined as collection of following set:

S={Ip,Op,Ss,Su,A}

Mapping Functionsf(x)	**X**	**Y**
F2(Ip1) →Op1	Ip1	Op1
F3(Ip2) →Op2	Ip2	Op2
F4(Op2) →Op3	Op2	Op3
F6(Ip2) →Su	Op2	Su

OBJECTS

1. Input1:Ip1={Username, Password}
2. Input2:Ip2={Paragraph}
3. Input3:Ip3={Training Database}
4. Output1:Op1={Extracted Keyword}
5. Output2:Op2={Generated Questions}

Advantages

1. Wh-type questions and MCQ generation makes easy.
2. Saving time of question paper generation
3. Question generation from simple as well as complex Sentences is possible.
4. User Friendly GUI
5. Remotely used because its web application.
6. Avoid repetition of questions.
7. Create variants or versions of a particular question or the whole question paper.
8. Easy to adapt and use, intuitive interface.
9. Create partially or fully different versions of the same question paper.

10. Save time, money, and headache so that you can focus on other critical tasks in your business operations.

CONCLUSION

In this project, we proposed an approach to automatically generate questions given a paragraph. We have used human effort to evaluate the system. We extract simple and complex sentences from the paragraph and generate question based on subject verb object and prepositions present in the sentence by mapping it to certain predefined rules. Our system does not support anaphora resolution i.e. pronoun resolution. Also our system has a human evaluation so steps can be taken for semantically providing proper sentences. Also many different types of questions like the yes/no question, summary question can be generated.

REFERENCES

1. 2010 Question generation shared task and evaluation challenge, http://questiongeneration.org/QG2010

2. Manish Agarwal and Prashanth Mannem, Automatic Gap-fill Question Generation from Text Books, Language Technologies Research Center International Institute of Information Technology Hyderabad, AP, India – 500032

3. Husam Ali, Yllias Chali, and Sadid A. Hassn, Automation Of Question Generation From Sentences, In proceedings of third workshop of Question Generation, June 18, 2010.

4. Manish Agarwal, Rakshit Shah and Prashanth Mannem. Automatic Question Generation using Discourse Cues. Language Technologies Research Center International Institute of Information Technology Hyderabad, AP, India – 500032.

5. Ming Liu,Rafael A. Calvo, Anindito Aditomo, and Luiz Augusto Pizzato,"Using Wikipedia and Conceptual Graph Structures to Generate

Cite this Article

Prof. S. S. More, Smita, Amruta, Harshada, Dhanshri, Akshay, (2019). **Internet of Things for Smart City** " Journal of Internet of Things and Information Technology", 2(2), 76-80

http://doi.org/10.5281/zenodo.265277

AUTHORS' PROFILE

[1] Deokate H. G

Department: Computer Engineering

College: Svpm's, Sppu, India

[2] Jogdand P. P

Department: Computer Engineering

College: Svpm's, Sppu, India

[3] Satpute P. S

Department: Computer Engineering

College: Svpm's, Sppu, India

[4] Shaikh S. B.

Department: Computer Engineering

College: Svpm's, Sppu, India