Automatic Question Generation from Text

Article · April 2015

Mohd Shahid Husain
Integral University

ISSN 2347 - 7911
International Journal for Innovations in Engineering, Science and Management
Available online at: www.ijiesm.com Volume 3, Issue 4, April 2015

Automatic Question Generation from Text

Himanshu Jethwani
Student M.Tech CSE
Integral University(India)
hhjj86@gmail.com

Mohd Shahid Husain
Assistant Professor Dept. Of CSE
Integral university(India)
shahid@iul.ac.in

Mohd Akbar
Assistant Professor Dept. Of CSE
Integral University(India)
akbar@iul.ac.in

Abstract:

Humans are curious by nature. They ask questions to satisfy their never-ending quest for knowledge. For instance, students ask questions to learn more from their teachers, teachers ask questions to help themselves evaluate performance of the students, and our day-to-day lives involve asking questions in conversations. Questions are the significant constituent of countless learning interactions from one-to-one tutoring sessions to extensive assessments in addition to real life discussions

One noticeable fact is that humans are not often very skilled in asking good questions because of their inconsistent mind in certain situations. It has been found that most people have trouble to identify their own knowledge deficiency This becomes our main motivation to automate the generation of questions with the hope that the potential benefits from an automated QG system could assist humans in meeting their useful inquiry needs.

Question Generation (QG) and Question Answering (QA) became the two major challenges for natural language understanding communities, recently QG has turned into an essential element of learning environments, systems, information seeking systems, etc Considerable interest from the Natural Language Processing (NLP), Natural Language Generation , Intelligent Tutoring System, and Information Retrieval (IR) communities have currently identified the Text-to-Question generation task as a promising candidate for the shared task In the Text-to-Question generation task, a QG system is given a text (such as a word, a set of words, a single sentence, a text, a set of texts, a stretch of conversational discourse, an inadequate question, and so on), and its goal would be to make a set of questions for which the text contains answers.

I. INTRODUCTION

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. A good question*[3]* is relatively short, clear, and unambiguous.

A few qualities[5] of a good question are as follows:

1. Evokes the truth.

2. Asks for an answer on only one dimension. A question that asks for a response on more than one dimension will not provide the information you are seeking. For example, a researcher investigating a new food snack asks *Do you like the texture and flavor of the snack?* If a respondent answers *no*, then the researcher will not know if the respondent dislikes the texture or the flavor, or both.

3. Can accommodate all possible answers. Multiple choice items are the most popular type of survey questions because they are generally the easiest for a respondent to answer and the easiest to analyze.

4. Has mutually exclusive options. A good question leaves no ambiguity in the mind of the respondent. There should be only one correct or appropriate choice for the respondent to make.

5. Does not presuppose a certain state of affairs.

6. Does not use emotionally loaded or vaguely defined words. Quantifying adjectives (e.g., most, least, majority) are frequently used in questions. It is important to understand that these adjectives mean different things to different people.

Importance of questions in Learning

One of the most important uses of questions is reflection, improving our understanding of things we have found out. People often spend hours by themselves contemplating ideas and working through issues raised by what they have read. These ideas and issues are often articulated in the form of questions. Questions are used from the most elementary stage of learning to original research. In the scientific method, a question often forms the basis of the investigation and can be considered a transition between the observation and hypothesis stages. Students of all ages use questions in their learning of topics, and the skill of having learners creating “investigatable” questions is a central part of inquiry education.

The method of questioning student responses may be used by a teacher to lead the student towards the truth without direct instruction, and also helps students to form logical conclusions. Questions have also been used to develop students' interest in a topic. Another use of questions is to give students a map for self-recognition[6] of reaching milestones of understanding as they study a unit, by asking, when a unit is begun, some difficult things that require understanding and insight into the material to be able to answer. A widespread[3] and accepted use of questions in an educational context is the assessment of students' knowledge through exams.

Question Generation

Question Generation (QG) is that the task of mechanically generating queries from numerous inputs like raw text, database, or linguistics illustration. though automatic QG may be approached with numerous techniques, QG is largely thought to be a discourse task involving the subsequent*[2]* four steps:

(1) When to raise the question,
(2) What the question is concerning, i.e. content choice,
(3) Question kind identification, and
(4) Question construction.

Question Generation and its applications

Question generation – [1]the purposeful asking and response of questions about what's browse – serves the goal of reading comprehension instruction not solely of its own accord, however conjointly in conjunction with multiple reading comprehension methods. QG methods, additionally to being a natural precursor to Question respondent (QA), is a superb compliment to a different and proved strategy - comprehension observance.

Often observed as self-regulation, comprehension observance interprets into meta-cognitive awareness and students' skills to self-select and use questioning methods on a situational basis. Students learn to severally and actively choose and use methods that facilitate them comprehend text material. Notably, a number of the foremost favorable gains in students skills to critique and improve the standard of their own queries and people of different students are found to occur in conjunction with comprehension observance instruction. QG has done an outsized role in reciprocal teaching. The National Reading Panel made self-questioning as the single handiest reading comprehension strategy to learn - that's, teaching kids to raise themselves concerning text as they browse it, as critics asking queries, except as examples to demonstrate the self-questioning strategy .

Of the four principal methods*[1]* (e.g. summarization, question generation, clarification, and prediction) utilized in combination throughout reciprocal teaching, question generation is that the strategy most often incorporated. QG and QA area unit key challenges facing systems that move with natural languages. The potential advantages of victimization automatic systems to get queries helps cut back the dependency on humans to get queries and different desires related to systems interacting with natural languages. Ultimately, QG permits humans, and in several cases AI systems, to know their surroundings and every different. analysis on QG contains a long history in AI, psychology, education, and language process.

One thread[2] of analysis has been theoretical, trying to know and specify the triggers (e.g., data discrepancies) and mechanisms (e.g., association between sort of data discrepancy and question type) underlying QG. The opposite thread of analysis has targeted on automatic QG. Applications of automatic QG facilities are endless A couple of examples are given below:

1. Instructed smart queries that learners would possibly raise whereas reading documents and different media.
2. Queries that human and pc tutors would possibly raise to market and assess deeper learning.
3. Instructed queries for patients and caretakers in drugs.
4. Instructed queries that may be asked in legal contexts by litigants or in security contexts by interrogators.
5. Queries generated from data repositories as candidates for commonly asked Question (FAQ) facilities.

Classification of queries

It is necessary to categorize queries as different*[4]* class of queries need different methods for automatic generation of Question-type

• **Yes/No Questions**: In linguistics, a question, formally called a polar question, could be a question whose expected answer is either affirmative or no. Formally, they gift associate degree exclusive disjunction, a combine of alternatives of that just one is appropriate. In English, such queries may be shaped in each positive and negative forms (For example, can you be here tomorrow? and Won't you be here tomorrow?).

• **Wh-Questions**: Wh-questions use interrogative words1, like why, when, who, where,which, etc., to request data. They can't be answered with a affirmative or no. Non-polar wh- queries area unit in distinction with polar yes-no queries, that don't essentially gift a variety of other answers, or essentially prohibit that vary to 2 alternatives. (For example, What time did you get across last night?)

• **Fill-in-the-blank queries OR Objective queries** Fill-in-the-blank question, conjointly called diagnostic test question, could be a sentence with one or a lot of blanks in it with four alternatives to fill those blanks.

Example:

…………carried the burden of our nation for twenty one years, before Kohli.

(a) Sachin (b) Sehwag (c) Raina (d) Dhoni

ISSN 2347 - 7911
International Journal for Innovations in Engineering, Science and Management
Available online at: www.ijiesm.com Volume 3, Issue 4, April 2015

II. PROBLEM STATEMENT

The main task is to automatically generate questions from given text ie sentences or group of sentences.The basic input is an affirrmative or declarative sentence or a group of sentences.The output should be different kind of questions depending on the type of sentences .Basically we can say

Input: text

Output: questions for reading assessment

For eg we have “Ram is a good boy”

The output should be “Who is a good boy?”

Another Example is “Rahul is always fighting”

The output should be “Describe Rahul.”

The basic target of this system is to provide a solution to the problem of first breaking down the paragraph and then convert it to questions.

OUR SOLUTION

The basic approach to solve the problem is integration and conversion

First the sentence/s should be broken into parts and classified and then it should be converted into questions.

The basic approach for extracting simple statements to generate question is as follows

Here we have

Input: complex sentence/s

Output: set of simple declarative sentences

Our method:

1. Uses rules to extract and simplify sentences
2. Is motivated by linguistic knowledge
3. Outperformed a sentence compression baseline

Our idea here is too extract and simplify multiple statements from complex sentences including operations for various syntactic constructions encoded with pattern matching rules for trees.

PSUEDOCODE FOR SENTENCE EXTRACTION

Take as input a tree t.

Extract a set of declarative sentence trees $T_{extracted}$ from constructions in t.

For each t’ in $T_{extracted}$:

Simplify t’ by removing modifiers.

Extract trees $T_{conjuncts}$ from conjunctions in t’.

For each $t_{conjunct}$ in $T_{conjuncts}$:

$T_{result} = T_{result}$ $\{t_{conjunct}\}$

Return T_{result}

PROPOSED ALGORITHM FOR QG

```
BtnGenerateQuestions()
{// initialize the variables
  StringBuilder output = new StringBuilder();
  string strNounName = "";
  string strMainNoun = "";
  string[] sentences =
SplitSentences(txtBoxEnteredText.Text);
  string strQuestion;
  int iFlag = 0;
  int iMainNounFlag = 0;
foreach (string sentence in sentences)
 {  int i = 0;
  string[] strText =
ParseSentence(sentence).Show().ToString().Split(' ');
  strText = removeBrackets(strText);
  string strAdjective ="";
foreach (string s in strText)
    {
     if ((s == "NNP") || (s == "PRP"))
    {
        strNounName = strText[i + 1];
        if (iMainNounFlag == 0)
     {          strMainNoun = strNounName;
          iMainNounFlag = 1;         }
          iFlag = 1;
        }
     if ((s == "NN") && (iFlag != 1))
      {
        strNounName = strText[i + 1];
        if (iMainNounFlag == 0)
         {          strMainNoun = strNounName;
 iMainNounFlag = 1;
         }
         iFlag = 2;
         }
```

ISSN 2347 - 7911
International Journal for Innovations in Engineering, Science and Management
Available online at: www.ijiesm.com Volume 3, Issue 4, April 2015

```
        if ((s=="NN") || (s=="JJ"))
         {
          if (iMainNounFlag == 0)
          {
            strMainNoun = strNounName;
            iMainNounFlag = 1;
          }
            iFlag = 3;
            if (s == "JJ")
          {
               strAdjective = strText[i + 1];
          }
            }
             i = i + 1;           }
strQuestion = removeFullStop(sentence);
 if (iFlag == 1)
  {
    strQuestion = replaceString(strQuestion, strNounName,
"Who");   }
 if (iFlag == 2)
  {    strQuestion = replaceString(strQuestion,
strNounName, "Who");   }
 if (iFlag == 3)
  {    strQuestion = replaceString(strQuestion,
strAdjective, "________________"); }
 if (iMainNounFlag == 1)
  {    strQuestion = "Describ " + strMainNoun + "?";   }
 iFlag = 0;
 strQuestion = "";}
}
```

The functions used by the algorithm to generate the questions are

// Replaces the Noun and Adjective's in questions

```
string replaceString(string sentence, string toReplace,
string replacedWith)
 {   return (sentence.Replace(toReplace, replacedWith) +
"?");  }
string[] removeBrackets(string[] strText)
 {   int i = 0;
   foreach (string s in strText)
  {
     strText[i] = strText[i].Trim(new Char[] { '(', ')', '.', ' '});
     i = i + 1;    }
   return strText; }
```

// Used as preprocessing to remove the parser added text

```
string removeFullStop(string sentence)
 {   return (sentence.Trim(new Char[] { '.' })) ;
 }
```

// Used to split the sentences

```
string[] SplitSentences(string paragraph)
 {   if (mSentenceDetector == null)
   {      mSentenceDetector = new
OpenNLP.Tools.SentenceDetect.EnglishMaximumEntropy
SentenceDetector(mModelPath + "EnglishSD.nbin");
   }   return
mSentenceDetector.SentenceDetect(paragraph); }
```

// Generates the tokens from the sentences given

```
string[] TokenizeSentence(string sentence)
 {   if (mTokenizer == null)
   {      mTokenizer = new
OpenNLP.Tools.Tokenize.EnglishMaximumEntropyToke
nizer(mModelPath + "EnglishTok.nbin");
    }   return mTokenizer.Tokenize(sentence); }
```

// This tags the token generated from sentences

```
string[] PosTagTokens(string[] tokens)
 {   if (mPosTagger == null)
   {      mPosTagger = new
OpenNLP.Tools.PosTagger.EnglishMaximumEntropyPos
Tagger(mModelPath + "EnglishPOS.nbin", mModelPath +
@"Parser\tagdict");     }
   return mPosTagger.Tag(tokens);}
```

// Used to parse the sentence to generate a single parse tree

```
Parser.Parse ParseSentence(string sentence){
   if (mParser == null)
   {      mParser = new
OpenNLP.Tools.Parser.EnglishTreebankParser(mModelPa
th, true, false);     }
   return mParser.DoParse(sentence); }
```

// Used to find "Names" like name of location, money, organization, percentage, person and time from the postags

```
string FindNames(string sentence)
 {   if (mNameFinder == null)
   {      mNameFinder = new
OpenNLP.Tools.NameFind.EnglishNameFinder(mModelP
ath + "NameFind\\");     }
   string[] models = new string[] { "location", "money",
"organization", "percentage", "person", "time" };
   return mNameFinder.GetNames(models, sentence); }
```

III. RESULT AND ANALYSIS

The sentences were tested on the system created and compared to questions generated by a normal human proficient in English language and asking questions. The Human subject as assumed as ideal for comparing the results.

ISSN 2347 - 7911
International Journal for Innovations in Engineering, Science and Management
Available online at: www.ijiesm.com Volume 3, Issue 4, April 2015

The results are showed in the table below followed by inferences

S. NO	NO OF SENTENCES	NO OF CORRECT QUESTIONS	NO OF INCORRECT QUESTIONS	NO OF QUESTIONS BY HUMAN
1	1	2	0	2
2	1	2	0	2
3	2	3	1	2
4	1	2	0	2
5	1	2	0	2
6	1	2	0	2
7	2	1	1	3
8	1	2	0	2
9	1	2	0	3
10	1	3	0	3

Total no of sentences = 12
Total no of correct questions = 21
Total no of incorrect questions = 2
Total no of questions by the human = 23
The above result shows that the system is working fairly correct with an accuracy of over 90% which can be further improved.

IV. CONCLUSION AND FUTURE WORK

The present system made by me works wonderfully on good amount of text i.e. sentences .Fairly accurate questions are developed with a minor here and there.The system not only generates factual questions but also some descriptive questions. That is the added flavor of the system.The system also fails in some cases subject to very complex sentences. In this work I have buried the roots of a tree that will grow high in future.
This work is a true reflection of what NLP can do.
The future seems bright. After successful implementation of this work ,in future it will lead to Question Generation from entire document.In that case it would be required to first find out those sentences from a document on whom questions can be formed.After that the current system algorithm with minor improvement can be deployed.The ultimate system will be able to independently handle any pdf or word or any other type of text file ,analyze it and find important sentences for QG.Further from those sentences a variety of questions could be formed with minimum inaccuracy.

V. REFERENCES

[1] “ Automatic Question Genaration Asyntactical Approach” Husam Deeb Abdullah Deeb Ali, 2012

[2].” Automatic Factual Question Generation from Text” Michael Heilman CMU-LTI-11-004

[3] “Automatic Question Generation using Discourse Cues and Distractor Selection for Cloze Questions” Rakshit Shah Language Technology and Research Center (LTRC)

[4] “Towards Automatic Topical Question Generation” Y llias ,Chali Sadid , A.Hasan University of Lethbridge, Lethbridge, AB, Canada

[5] https://www.statpac.com/surveys/question-qualities.htm

[6] “Using Questions to Teach Better”Rick Garlikov - http://www.garlikov.com/teaching/usingquestions.html