

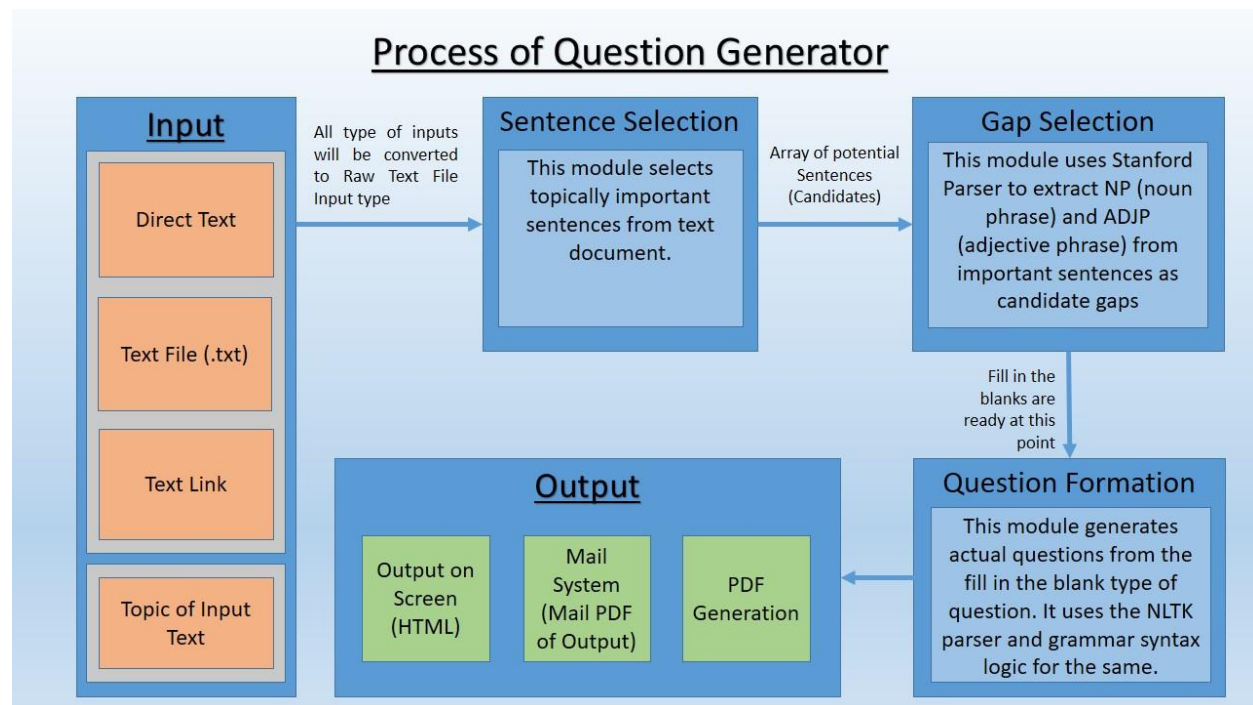
Using Natural Language Processing for Smart Question Generation

This paper represents the study of automatic question generation through NLP with statistical pattern and novel neural network. This paper also discusses a review of research work carried out so far in this area. Question Generation (QG) aims to create natural questions from a given a sentence or paragraph or paragraphs. NLP (Natural Language Processing) is an area of research and application that explore how computers can be used to understand and manipulate natural language text or speech to do useful things. Our aim is to create the technique which can generate various logical questions from the given text input. Right now, only humans are capable of accomplishing this with higher accuracy.

The goal of question generation is to generate questions according to some given information (e.g., a sentence or a paragraph). It has been applied in many scenarios, e.g., generating questions for reading comprehension [(2)Du et al., 2017] and generating data for large-scale question-answering training [(2)Duan et al., 2017]. Since questioning is an important communication skill, question generation plays an important role in both general-purpose Chatbot systems and goal-oriented dialogue systems. In the context of dialogue, many researchers have studied the problem [(3)Mostafazadeh et al., 2016; (4)Bordes & Weston, 2017]. The generated questions are mainly used to start a conversation or to obtain some specific information.

Much of the existing work focuses on generating questions based only on the provided information such as sentences, paragraph or paragraphs. In this paper, we argue that "Topic-based question generation" is also very important. That is, in addition to the given sentence or paragraph, it is also useful to specify a relevant topic contained in the text. The main reason is that a sentence or paragraph often involves multiple topics or concepts that questions can be generated, only arbitrarily choose one or mixing them may be of limited use because we found that in practical applications, questions need to be targeted toward some topics related to the current conversation. One almost never asks a random question in a conversation. Generating a question without knowing what it is about is not very useful. To solve the proposed problem, we propose a novel neural network that can generate topic-based questions. One major advantage of our model is that it can be trained directly using a question-answering corpus without requiring any additional annotations like annotating the topics in the questions or answers.

One key application of question generation is in the area of education - to generate questions for reading comprehension materials [(1) Heilman and Smith, 2010]. The researcher has studied the examination process and researched if questions can be generated automatically with the help of the computerized application. So it will reduce the task of an educator. Work in this sector can also help the students and learners on another side, to become more efficient by self-evaluation with the help of this method while reading and learning new things.



((5)Using Natural Language Processing for Smart Question Generation – Aditya Sarvaiya)

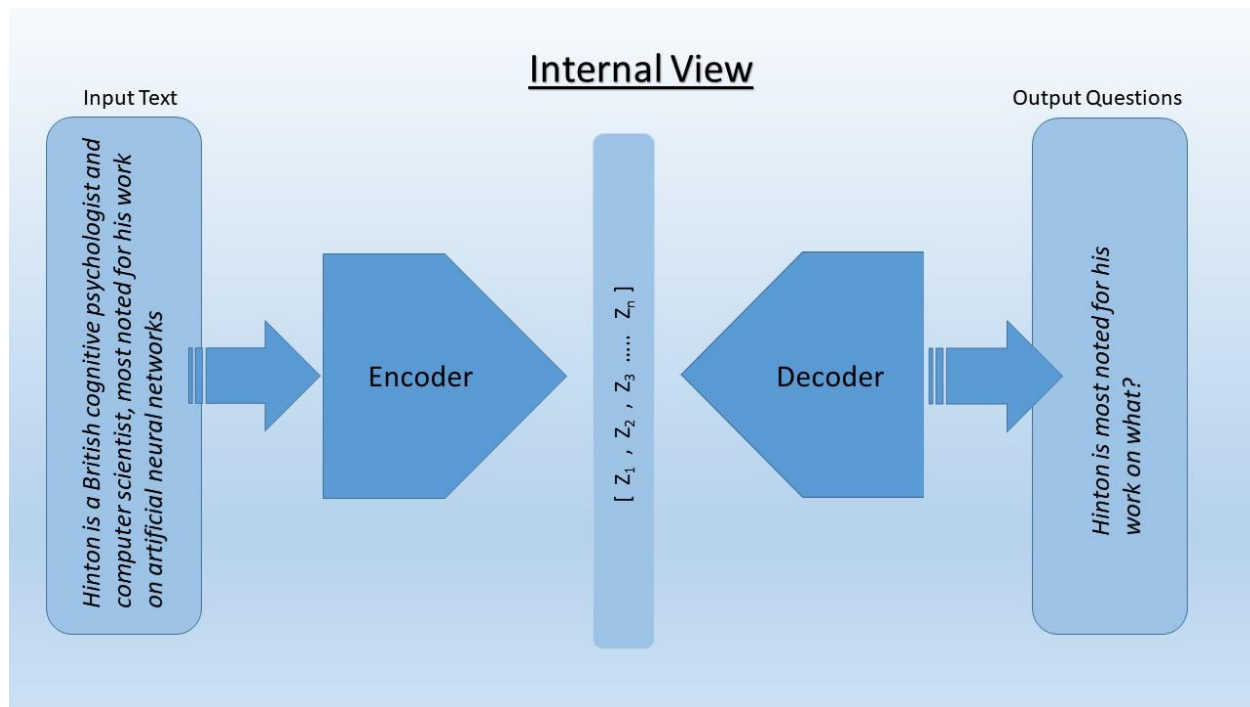
We aim to implement it using encoder-decoder nets, which will increase the quality drastically[(5)Aditya Sarvaiya, 2018]. Encoder-decoder nets have been used by Google for its neural machine translation (language translation) and recurrent neural networks. By keeping encoder-decoder at the core, we also take help from Stanford Parser and NLTK for grammar analysis and more basic natural language analysis.

Encoder

The encoder takes a preprocessed sentence from the input text and converts it according to the weights of the hidden layer. This hidden layer creates an intermediate representation of the input text and passes it to the decoder.

Decoder

The decoder converts the hidden-layer information into question form. Machine translation uses the same concept. Here, we treat questions, essentially, as another language.



((5)[Using Natural Language Processing for Smart Question Generation – Aditya Sarvaiya](#))

This image shows how the encoder-decoder network works internally.

In summary, this paper makes the following contributions:

- It proposes the new problem of question generation based on a given sentence and a topic (or concept) in the sentence. To our knowledge, this topic-based question generation has not been studied before. The model can also take a question type because for the same topic, different types of questions can be asked.
- It proposes a novel neural network model to solve the problem. A pre-decode mechanism is also explored to improve the model performance.
- The proposed model can be directly trained using a normal question-answering corpus without requiring additional labelling of topics in each input sentence.
- The proposed model is evaluated using the Amazon community question-answering corpus. Experimental results show that our model is effective.

References

- 1) Michael Heilman and Noah A. Smith. 2010. Good question! Statistical ranking for question generation. In Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics. Association for Computational Linguistics, Los Angeles, California, pages 609–617. (<http://www.aclweb.org/anthology/N10-1086>)
- 2) Nan Duan, Duyu Tang, Peng Chen, and Ming Zhou. Question generation for question answering. In EMNLP, 2017. ()
- 3) Nasrin Mostafazadeh, Ishan Misra, Jacob Devlin, Margaret Mitchell, Xiaodong He, and Lucy Vanderwende. Generating natural questions about an image. In ACL, 2016.
- 4) Antoine Bordes and Jason Weston. Learning end-to-end goal-oriented dialog. In ICLR, 2017.
- 5) Aditya Sarvaiya. Using Natural Language Processing for Smart Question Generation. In Intel Ai Academy 2018. (<https://software.intel.com/en-us/articles/using-natural-language-processing-for-smart-question-generation>)