Questionator - Automated Question Generation using Deep Learning

Animesh Srivastava
Computer Engineering
Fr.C Rodrigues Institute Of Technology
Navi Mumbai, India
animeshsrivastava698@gmail.com

Shantanu Shinde
Computer Engineering
Fr.C Rodrigues Institute Of Technology
Navi Mumbai, India
shantanushinde.shinde@gmail.com

Naeem Patel
Computer Engineering
Fr.C Rodrigues Institute Of Technology
Navi Mumbai, India
naeempatel010@gmail.com

Siddhesh Despande
Computer Engineering
Fr.C Rodrigues Institute Of Technology
Navi Mumbai, India
siddhesh991998@gmail.com

Anuj Dalvi
Computer Engineering
Fr.C Rodrigues Institute Of Technology
Navi Mumbai, India
anuj.hp07@gmail.com

Shweta Tripathi
Computer Engineering
Fr.C Rodrigues Institute Of Technology
Navi Mumbai, India
shwetripathi09@gmail.com

2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE) 978-1-7281-4142-8/20/$31.00 ©2020 IEEE 10.1109/ic-ETITE47903.2020.212

***Abstract*—Due to a boom in the amount of data generated every day, there is a need for automation in the education domain where it is humanly impossible for a single individual to make sense out of the data even for a simple task such as generating questions for a quiz or a test. Automatic question generation for textual inputs is valuable in academics where answering questions helps students to learn and improve their understanding of their field of study. Automatic question generation finds application in dialog systems or virtual assistants where asking questions is an important part of interactions between humans and machines. In this paper, we propose a state-of-the-art solution using a pipeline that utilizes natural language processing and image captioning techniques capable of generating questions not only for textual but also for visual inputs. Along with the question, distractors for the generated questions and their answers are also created.**

***Index Terms*—Natural Language Processing, Automatic Question Generation, Option Generation, Image Captioning**

I. Introduction

Asking and answering questions is an effective method for testing ones understanding of any topic. Answering questions helps to improve the process of learning and is an integral part of academics. With the huge abundance of data, it contains a lot of valuable knowledge that has the potential to be used in the field of education. For example, consider blogs, news articles, research papers, educational videos etc. which are published every day. The data may be of various formats like text, images, videos etc. It is not possible to manually go through and make sense of such huge amounts of data.

Technology has been evolving everyday and there is more and more automation everyday. In the field of education especially where newer methods have been adopted to enhance the teaching and learning experience, Objective questions aims at improving the learning experience of students as well as provide better metrics for assessment. Also with the abundant data, we have a opportunity to generate huge number of these objective questions primarily Multiple Choice Question(MCQ). However, developing effective objective tests has some disadvantages, it is time-consuming and requires expertise.

Apart from the education domain, Quiz Question Generation for the entertainment industry is also gaining importance. There are millions of people playing quiz games as a fun activity and at times serious money making option. This requires automated generation of millions of questions each day manually.

Our Automated Question Generation system aims at automating this process of question generation by providing a end-to-end system which will take the data which may be in the form of image/sentence as input and provide a question as output along with the Distractors. These Distractors consist of answers which are closely related to the actual answer. More technically, we have used a Convolutional Neural Network(CNN) [1] as an encoder which extracts features from the image, and a Long Short Term Memory(LSTM) [2] network as a decoder which turns the extracted features into natural language. This encoder-decoder system of the CNN and LSTM are part of Image Captioning [3]. Then using natural language processing we extract the different components of the sentence like subject, object and the predicate, and form the question. We generate the distractors using GloVe [4].

This paper is organised as follows. Section II consist of the literature survey in the domain of Question Generation. Section III describes the pipeline of our proposed system. Section IV provides the evaluation of the work. Section V outlines our conclusion.

II. Related Works

A. Automatic Question Generation from Childrens Stories for Companion Chatbot

Che-Hao Lee et al. [5] proposed generating questions for the Chinese language based on a novel method for generating and ranking questions. For generating questions they replaced target answer phrases with the interrogatives. Their method proved to be effective as the Chinese language, unlike English, does not require sentence decomposition in generating questions. The generated questions were then passed through a logistic regression-based Ranking model which was trained using supervised learning based on data created by the question generation model.

B. Automatic Question Generation System

P. Pabitha et al. [6] proposed a similar approach of Automatic Question Generation System generates questions by using a Nave Bayes supervised learning-based Key Phrase extraction model and a Noun filter. The extraction model extracts key phrases and the noun filter is used to find nouns in the input both of which are used to generate questions.

C. Design of Question Answering System with Automated Question Generation

Min-Kyoung K. et al. [7] suggests question generation as a means for question answering. The question generation is performed using sentence splitting and Named Entity Recognition. The question generation generates two types of questions based on different formats. The generated question is passed through a question filter to analyze question quality.

D. Dual Learning for Visual Question Generation

Xing X. et al [8] performs visual question generation and visual question answering based on a dual learning approach which makes use of two agents based on pre-trained models that form a closed loop and improve each other by reinforcement learning.

E. Thematic Question Generation over Knowledge Bases

Tanguy R et al. [9] introduces and tackles the issue of automatically generating typical multiple-choice questions over knowledge graphs. A template-based approach for generating questions and their distractors to make the questions multiple choice is proposed. This approach identifies topic boundaries and generates questions based on templates used for question answering. They were able to achieve a success rate of 0.69 in the topic assignment and a very high success rate of over 0.9 for the quality of the generated questions and distractors as well as substantial inter-rater agreements.

III. Proposed System

Our proposed system mainly comprises of three sections namely Image Captioning, Question Generation and the Option Generation. The entire pipeline has been outlined in figure 1.

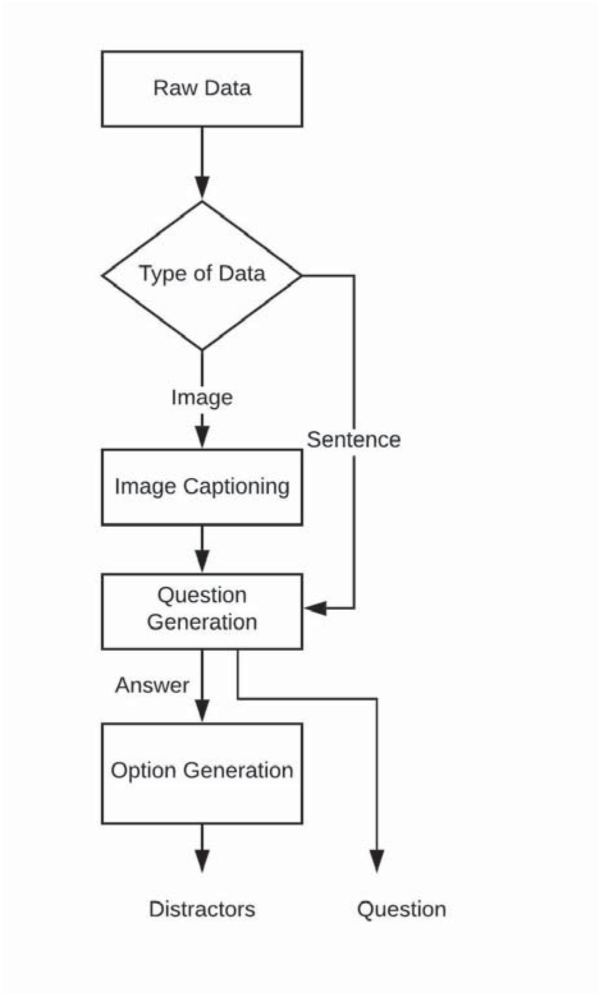

Fig. 1. Pipeline for the proposed automated Question Generation System

A. Image Captioning

Step one in this pipeline is Image Captioning. The image captioning module converts a given input image into a natural language description. The natural language provides a good solution for describing the semantic information of images. Typically, a encoder-decoder framework is used for this task. The architecture of Image Captioning has been illustrated in figure 2. The Encoder used is a convolutional neural network (CNN). We have used ResNet-152 model [10] as a global image feature extractor which is fed to a Long short Term network(LSTM) to generate sequence of words. The CNN based Encoder compresses the information in the original image into a smaller representation. We process this feature vector and use it as an initial input to the following LSTM. The LSTM then decodes the process feature vector and turns it into a natural language. The ResNet-152 model was pre-trained on the ILSVRC-2012-CLS Image classification dataset [11]. The ILSVRC-2012-CLS dataset contains over 1000 classes. Our proposed model requires syntactically correct simple sentences hence we used transfer learning on our factual dataset.

Our dataset consisted of 1128 simple factual sentences. These sentences were generated by crowdsourcing, that is, we provided 30 students with 400 different images from different domains and asked them to come up with simple sentences describing each image and selected the best ones out of the 1200 sentences.

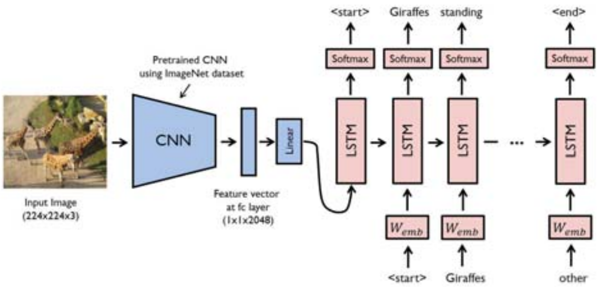

Fig. 2. Architecture of Image Captioning model [13]

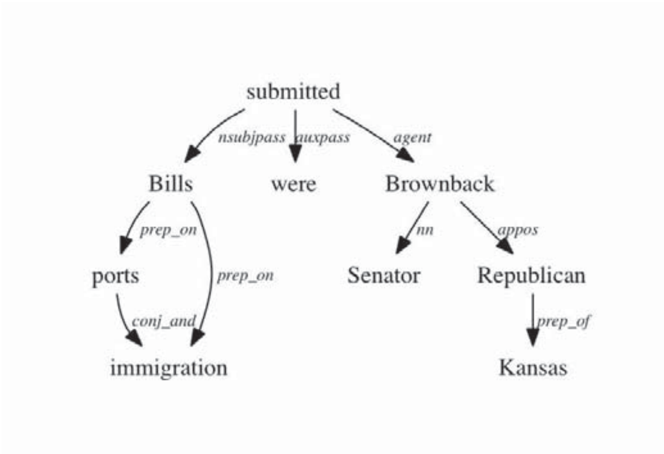

Fig. 3. dependency parse of a short sentence [16]

All the training was performed on Google Colaboratory (a.k.a. Colab) [12], which is a cloud service based on Jupyter Notebooks for disseminating machine learning education and research. It provides a run-time fully configured for training deep learning models and free-of-charge access to a robust GPU. Google Colab provides Nvidia Tesla K80 GPU with 12GB of DDR5 memory. We trained the model with the pre-trained weights as the starting point for 5 epochs with a batch size of 128 and a learning rate of 0.001.

B. Question Generation

The caption generated by the image captioning is now fed through the question generation. The question is generated by extracting the subject, object and predicate which are the three different components when breaking down a sentence. The subject is the word or set of words which the sentence talks about, the predicate is the verb, and the object is the noun that is acted upon by the subject. This task of extracting the subject, object and the predicate is obtaining through the use of Dependency Parsing [14].

A dependency parser examines the grammatical composition of a sentence, building up connections between ”head” words and words which alter those heads. It performs the task of understanding the syntactical structure of the sentence. The most utilized syntactic structure is parse tree [15]. But a given sentence can have many parse trees called ambiguities because of the ambiguous part of speech tags. These ambiguities can be resolved by using an appropriate parsing algorithm.

Figure 3 below shows a dependency parse of the sentence - *”Bills on ports and immigration were submitted by Senator Brownback, Republican of Kansas ”.*

Dependency parsing captures the relationship among verb, subject and object. Once we detect the subject and object we construct the ”Wh” question types. We take the subject and pass it through a Part of Speech(POS) tagger [17] which helps us to replace the subject in the sentence with a proper Wh question.(who, what). Also, If an input sentence has a location(LOC) and Time(TMP), we can construct the where and when questions. Next, we replace the subject with the correct ”Wh” question. Then, verb tense and subject position is altered while preserving the remaining words in the sentence. Finally, we produce the question for our caption and also the answer to the question. For example, consider the sentence given below.

Messi was born in 1987.

We can construct the following questions after we extract the subject ”Messi” and replace it with a ”Who”, and then generate a question. The answer(”Messi”) is given to the Option Generation.

Who was born in 1987 ?

Another way a question can be generated is a ”When” question. As the sentence - *Messi was born in 1987.* contains a Time(TMP), we switch the verb tense and the subject position. So another question with the answer(”1987”) that can be generated is:-

When was Messi born?

To perform the dependency parsing of the sentences, we have used the Stanford CoreNLP module [18] which is based on the super-fast transition-based parser powered by a neural network which accepts word embedding inputs [19]

C. Option Generation

The question generation module generates a question as well as determines the expected output for the generated question. The expected answer is given as input to the option generation module. This module generates options so as to make the end result a set of multiple choice questions.

This module makes use of GloVe [4] to make a Vector representation for words. Glove stands for Global Vectors for word representation. It was developed by Stanford and is an unsupervised learning algorithm for generating word embeddings. Training for GloVe is done using a word-word co-occurrence matrix determined using a corpus.

GloVe makes use of a word-word co-occurrence matrix. Only the non-zero entries in the matrix are considered while training. The co-occurrence matrix tabulates how frequently words occur with each other. This co-occurrence is subject to the corpora being used for training. The matrix is populated in a single pass through the entire training corpora. This single pass algorithm can sometimes prove to be computationally expensive especially for large training Corpora.

GloVe is essentially a log-bilinear model with a weighted least-squares objective. The main idea that underlines this model can be explained with the help of an example. Consider the following table.

Probability and Ratio	$k = solid$	$k = gas$	$k = water$	$k = fashion$
$P(k\|ice)$	1.9×10^{-4}	6.6×10^{-5}	3.0×10^{-3}	1.7×10^{-5}
$P(k\|steam)$	2.2×10^{-5}	7.8×10^{-4}	2.2×10^{-3}	1.8×10^{-5}
$P(k\|ice)/P(k\|steam)$	8.9	8.5×10^{-2}	1.36	0.96

Fig. 4. Co-occurrence probabilities of for the words ”ice” and ”steam” [4]

Here, we have the co-occurrence probabilities of for the words ”ice” and ”steam” determined using a 6 billion word corpus. It is natural for ”ice” to co-occur more with the word ”solid” than with ”gas” and for ”steam” to co-occur more with ”gas” than with ”solid”, which is seen in the table as well. Also the probabilities for co-occurrence of both ”ice” and ”steam” will be high with ”water” as it is a shared property and will be very low with ”fashion” as both of them are completely unrelated. One interesting observation that can be made here is that in the case ratio of probabilities the noise from non-discriminative words like ”water” and ”fashion” cancel out, such that large values i.e. very much greater than 1 will correlate well with ”ice” and small values i.e. very much lesser than 1 will correlate well with ”steam”. Thus, it can be said that the ratio of probabilities encode some form of meaning.

Thus, the main intuition underlying the GloVe model is the simple observation that ratios of word-word co-occurrence probabilities have the potential for encoding some form of meaning.

GloVe represents words as vectors such that their dot products gives the logarithm of the words probability of co-occurrence. Thus, vector difference in essence represents the logarithm if ratios of probabilities of co-occurrence. This is because, the logarithms of ratios is the same as the difference of logarithms.

For our aim in this module we have made use of pre-trained word vectors. This model was trained using Wikipedia 2014 and English Gigaword fifth edition dataset [20]. This dataset contains 6 billion tokens and 4,00,000 vocabulary. The dataset was used to train separate models for 50, 100, 200 and 300 dimensions.

IV. Results

In order to evaluate our proposed system, we have split the evaluation into two sections. The quality of the captions, questions and options generated from our system was evaluated manually. This is because utilizing standard metrics such as BLEU score [21] are not a true indication of the generated questions and captions. BLEU cannot effectively reflect the naturalness of the generated questions/caption i.e. it cannot check if the questions are fluent and grammatically accurate.

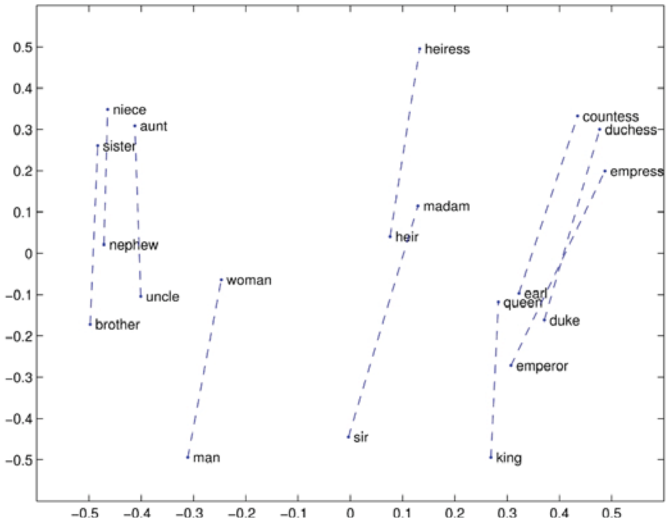

Fig. 5. Visual Representation of word Vectors [4]

A. Images

Figure 6 shows the result of our system on images. The images are first passed through the Image Captioning and then through the question generation.

Input Image	Caption Generated	Question
	A man is playing football on the field.	Who is playing football on the field?
	dog is playing with a frisbee on the beach.	What is playing with a frisbee on the beach?
	A man is riding a wave on a surfboard in the ocean.	Where is a man riding a wave on a surfboard?

Fig. 6. Result of the question generation for images

B. Sentence

Table 1 shows the result of independently running the Question Generation on a randomly selected set of sentences.

TABLE I: Results of Question Generation of sentence input

Input Sentence	Generated Question	Answer
The pen is mightier than the sword.	What is mightier that the sword?	Pen
Jack went up a hill.	Who went up a hill?	Jack
The staff performed well.	Who performed well?	The staff
Joe went to the store.	Who went to the store?	Joe
Eiffel Tower was built in 1887.	When was Eiffel Tower built?	1887
I live in Washington D.C.	Where do you live?	Washington D.C.
Statue of is Liberty located in New York.	Where is Statue of Liberty located?	New York

C. Option Generation

Table II shows the results of using GloVe for option generation.

TABLE II: Results of Option Generation

Answer	Distractors
Messi	Ronaldo, Pogba, Ronadinho
Apple	Orange, Banana, Mango
London	Mumbai, New York, Tokyo
Gun	Sword, Knife, Pistol
Lincoln	Roosevelt, Washington, Nixon

V. Acknowledgment

We would like to thank Myraa Technologies and Ms. Shilpa Karkera, CEO of Myraa Technologies who provided with insight and expertise to this research. Their contributions to our efforts are greatly appreciated. We would like to thank our families and friends who supported us in the course of writing this paper.

VI. Conclusion

The proposed question generation pipeline has demonstrated that automated generation of questions can find use in the education domain for generating Multiple Choice Questions(MCQ). This is a robust system which not only works for sentences but also for images. This system has achieved outstanding results as shown above. However, there were outliers as well. Further additions to the captioning dataset , for instance adding complex sentences and then improving the question generation model to generate questions on these complex sentences could yield better results.

References

[1] Yann LeCun, Patrick Haffner, Léon Bottou, and Yoshua Bengio. *Object Recognition with Gradient-Based Learning*, pages 319–345. Springer Berlin Heidelberg, Berlin, Heidelberg, 1999.

[2] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural Comput.*, 9(8):1735–1780, November 1997.

[3] Zhongliang Yang, Yu-Jin Zhang, Sadaqat Rehman, and Yongfeng Huang. *Image Captioning with Object Detection and Localization*, pages 109–118. Springer International Publishing, 12 2017.

[4] Jeffrey Pennington, Richard Socher, and Christopher Manning. Glove: Global vectors for word representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1532–1543, Doha, Qatar, October 2014. Association for Computational Linguistics.

[5] C. Lee, T. Chen, L. Chen, P. Yang, and R. T. Tsai. Automatic question generation from childrens stories for companion chatbot. In *2018 IEEE International Conference on Information Reuse and Integration (IRI)*, pages 491–494, July 2018.

[6] P. Pabitha, M. Mohana, S. Suganthi, and B. Sivanandhini. Automatic question generation system. In *2014 International Conference on Recent Trends in Information Technology*, pages 1–5, April 2014.

[7] M. Kim and H. Kim. Design of question answering system with automated question generation. In *2008 Fourth International Conference on Networked Computing and Advanced Information Management*, volume 2, pages 365–368, Sep. 2008.

[8] X. Xu, J. Song, H. Lu, L. He, Y. Yang, and F. Shen. Dual learning for visual question generation. In *2018 IEEE International Conference on Multimedia and Expo (ICME)*, pages 1–6, July 2018.

[9] T. Raynaud, J. Subercaze, and F. Laforest. Thematic question generation over knowledge bases. In *2018 IEEE/WIC/ACM International Conference on Web Intelligence (WI)*, pages 1–8, Dec 2018.

[10] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, June 2016.

[11] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, and Li Fei-Fei. ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision (IJCV)*, 115(3):211–252, 2015.

[12] Tiago Pessoa, Raul Medeiros, Thiago Nepomuceno, Gui-Bin Bian, V.H.C. Albuquerque, and Pedro Pedrosa Filho. Performance analysis of google colaboratory as a tool for accelerating deep learning applications. *IEEE Access*, PP:1–1, 10 2018.

[13] Yunjey Choi. Architecture of image captioning, 2018. [Online; accessed May 10, 2018].

[14] Peng Qi, Timothy Dozat, Yuhao Zhang, and Christoper Manning. Universal dependency parsing from scratch, 01 2019.

[15] Karen Jensen and George E. Heidorn. The fitted parse: 100english. In *Proceedings of the First Conference on Applied Natural Language Processing*, ANLC '83, pages 93–98, Stroudsburg, PA, USA, 1983. Association for Computational Linguistics.

[16] The Stanford Natural Language Processing Group. Dependency parsing of a sample sentence, 2010. [Online; accessed May 10, 2018].

[17] Eric Brill. A simple rule-based part of speech tagger. *Proceedings of the Third Conference on Applied Natural Language Processing, Trento, Italy*, 05 2002.

[18] Christopher D. Manning, Mihai Surdeanu, John Bauer, Jenny Finkel, Steven J. Bethard, and David McClosky. The Stanford CoreNLP natural language processing toolkit. In *Association for Computational Linguistics (ACL) System Demonstrations*, pages 55–60, 2014.

[19] Danqi Chen and Christopher Manning. A fast and accurate dependency parser using neural networks. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 740–750, Doha, Qatar, October 2014. Association for Computational Linguistics.

[20] Philadelphia: Linguistic Data Consortium. English Gigaword Fifth Edition LDC2011T07, 2011. [Web Download].

[21] Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, pages 311–318, Philadelphia, Pennsylvania, USA, July 2002. Association for Computational Linguistics.