International Journal of Future Generation Communication and Networking
Vol. 13, No.2s, (2020), pp. 1586–1590

Question Generation from Text

Yash Oswal[#1], Isha Pisal[*2], Mehul Shah[#3], Rushikesh Rote[#4]

[#]*Computer Department, VIIT, Pune University*

[1]yash.oswal@viit.ac.in

[2]isha.pisal@viit.ac.in

[3]mehul.shah@viit.ac.in

[3]rushikesh.rote@viit.ac.in

Abstract

A goal to generate questions based on text which is given as input to the system. Our proposed system aims at generating the questions by analyzing vivid texts. Soft-copy of the text is given to the system, and wh-question is the expected output. Question Generation is a major challenge for natural language understanding communities, recently QG is turned now into an information seeking systems. Interest of the Natural Language Processing , Information Retrieval and Natural Language Generation communities now have identified the Text-to-Question generation as a prominent task for a system, where, given a text which can be anything such as: a single sentence, a word or set of words, a text or set of texts, an inadequate question, and so on; and its goal would be to make a set of questions. Also, semantics of the words extracted are been checked for appropriate sentence formation.

Keywords— *Question Generation, Rule Based Technique, POS Tagging, Semantics.*

I. INTRODUCTION

To generate question manually, professors/teachers require a lot of time and efforts to find right questions and its marks respectively. Hence, an approach to do this task in less time and efficiently is through Data Science, and Natural Language Processing.

II. LITERATURE SURVEY

Li Fei-Fei, Andrej Karpathy. DeepVisual-Semantic Alignments for Generating Image Descriptions in CVPR 2015 for Research paper on Image to Text Extraction stating various techniques for the same.

Sanja Filder, Raquel Urtasun. A Sentence Is Worth a Thousand Pixels. Conference Paper in Proceedings / IEEE Computer Society Conference on Computer Vision and Pattern Recognition, IEEE Computer Society Conference on Computer Vision and Pattern Recognition · June 2013. Research on to form sentences and find relationships between the objects. They find out words for the images and form sentences from various words defining the relationship between the objects in the images. They search for presence of class, objects cardinality and then their relationships. POS tagging is used for the purpose of text extraction where relationships are defined using the prepositions from POS tagging.

Hao Fang, Forrest Iandola, Saurabh Gupta, Rupesh K Srivastava. From Captions to Visual Concepts and Back. CVPR 2015 PAPER.Resarch on Divide the image into several different parts containing various objects in the same image. Identify suitable words for the divided parts of the image using weakly supervised learning. Find the correct match for various objects in bounding box and discard other words using probabilistic based search. They use Microsoft training dataset to make the machine learn on weakly supervised learning. They extract nouns, verbs, adjectives from the images in form of text. Use of language model to generate sentences from the texts. They use deep multimodal similarity model to re rank the sentences generated.

ISSN: 2233-7857 IJFGCN

Christian Jauvin, Yoshua Bengio, Pascal Vincent, Réjean Ducharme. A Neural Probabilistic Language Model. Journal of Machine Learning Research 3 (2003) 1137–1155. Research on topic of **Curse of dimensionality** in which a word sequence of which model will be tested, will be likely to be different from all the word sequences seen earlier during training. Every word is associated deterministically or probabilistically with discrete class, and there is similarity in words of same class. This paper also concerns the challenge of training such large neural networks for very large data sets.

Courses notes on NLP by Michael Collins, Columbia University, research on Tagging problems and Hidden Markov Model.

III. METHODOLOGY

1. **Text to Question Generation:** This is the second module of our proposed system, where the text generated from the input images is used to form questions. These questions are generated using Rule-Based techniques where various question formation rules are defined. We define context free grammar (CFG) rules for formation of questions using texts or sentences.
2. **Semantics Checker for generated Questions:** The generated questions may be in WH format but not necessary be in grammatical correct format. Hence our third module comes into picture, where we check the grammar of the generated questions. For this we propose to use Parts-of-Speech Tagging (POS) integrating with Markov Model.

Question Generation is the task of mechanically generating queries from various inputs like raw text, database, though automatic QG may be approached with numerous techniques, QG is largely thought to be a discourse task involving the subsequent 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 – 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. Students learn to severally and actively choose and use methods that facilitate them comprehend text material. Notably, a number of the foremost favourable gains in student's 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. 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, except as examples to demonstrate the self-questioning strategy . The four principal methods: summarization, question generation, clarification, and prediction are utilized in combination throughout reciprocal teaching, question generation is that, the strategy most often incorporated. QG and QA are domains in which the key challenges are facing systems that move with natural languages.

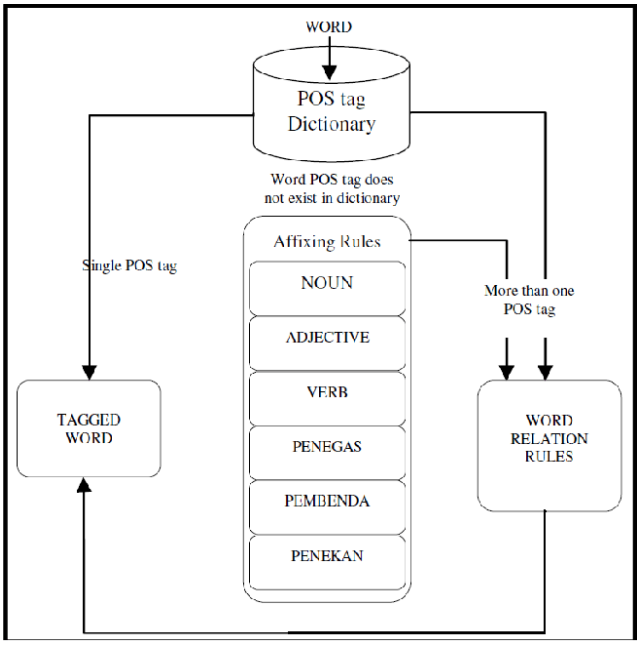
Text to Question Generation: This is the second module of our proposed system, where the text generated from the input images is used to form questions. These questions are generated using Rule-Based techniques where various question formation rules are defined. We define context free grammar (CFG) rules for formation of questions using texts or sentences.

ISSN: 2233-7857 IJFGCN

Part of Speech	Tag
Noun	n
Verb	v
Adjective	a
Adverb	r

POS table

Rule-based Technique: - This is a very common and known technique where the inference engine, goes through a simple recognize-assert cycle i.e. the control scheme in which called backward chaining for goal-driven reasoning and forward chaining for data-driven reasoning. The basic concept for forward chaining lies that the premises of a rule i.e., if portion are satisfied by the data, the system asserts the conclusions of the rule i.e. then portion as true.
We define Context Free Grammar (CFG) rules to the system and frame the questions according to the rules defined[9 10]. One of the famous technique of tagging is rule-based POS tagging. Rule-based taggers use lexicon i.e. knowledge or dictionary for getting possible tags for tagging each word present in the text. If the word has more than one tag, then rule-based taggers use hand-written predefined rules to appropriately identify the tag. Disambiguation is also a feature performed in rule-based tagging which is done by analysing the linguistic features of a word along with its following and preceding words present in the text. A common example, suppose if the preceding word, of a word is article then word must be a noun.



POS Tagging

All such kind of data in rule-based POS tagging is coded to form rules, these rules may be −

- Context pattern rules,
- Regular expression compiled in finite-state automata, present with lexically ambiguous sentence representation.

ISSN: 2233-7857 IJFGCN

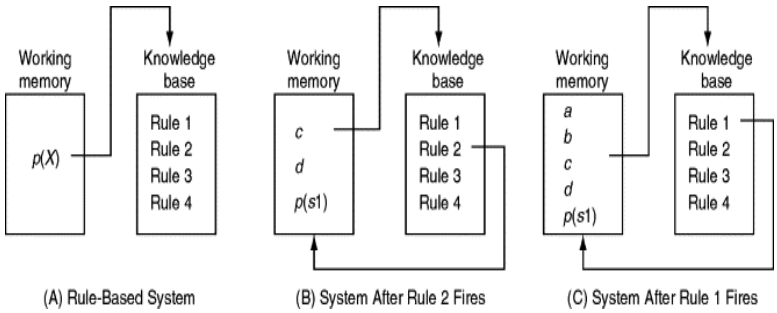
Rule-based POS tagging has two-stage architecture where −

- **First stage** − list of appropriate parts-of-speech is assigned by using a dictionary knowledge.
- **Second stage** − single part-of-speech for each word is defined using lists of hand-written disambiguation rules.

Properties of Rule-Based POS Tagging

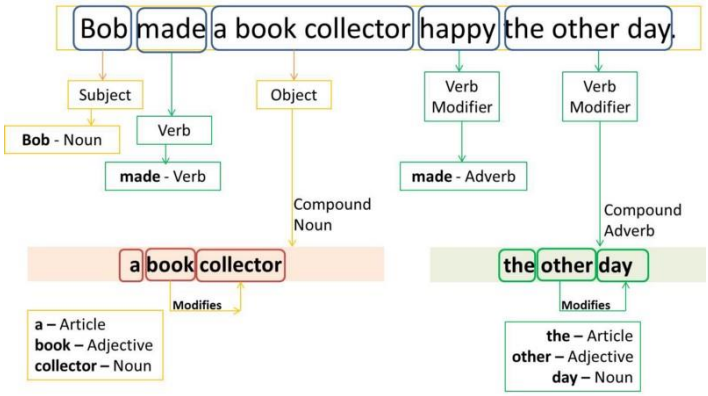
Rule-based POS taggers possess the following properties −

- These taggers are knowledge-driven taggers.
- The rules in Rule-based POS tagging are built manually by a human.
- Rules are been coded.
- Limited number of rules approximately.
- Language modeling and smoothing is defined explicitly in rule based tagger.



Rule – based Technique

Semantics Checker for generated Questions: The generated questions may be in WH format but not necessary be in grammatical correct format. Hence our third module comes into picture, where we check the grammar of the generated questions. For this we propose to use Parts-of-Speech Tagging (POS) integrating with Markov Model.



Semantics Checker

ISSN: 2233-7857 IJFGCN

IV. CONCLUSION & FUTURE WORK

This paper is all about the proposed system for performing text extraction, whose usage is day by day increasing exponentially, making it significantly important to understand them faster for storage and its usage. The proposed approach explores various techniques and features to predict characters from the text using neural networks. Experimental results show that our approach paves way for some positive exploration. Also the results show some meaningful work for text extraction.

V. ACKNOWLEDGEMENT

We would like to thank Prof. Pranali Chavhan for giving us all the help and guidance we needed. We are grateful to her, for her kind support and suggestions.

We are also grateful to Dr. S. R. Sakhare, HOD Computer department, Vishwakarma Institute of Information Technology for his support.

We are also grateful to our project expert Rrof. Snehal Rathi for the indispensable support, supervision and suggestions.

VI. REFERENCES

[1] Andrej Karpathy, Li Fei-Fei. DeepVisual-Semantic Alignments for Generating Image Descriptions.
[2] Sanja Filder, Raquel Urtasun. A Sentence Is Worth a Thousand Pixels.
[3] Hao Fang, Saurabh Gupta, Forrest Iandola, Rupesh K Srivastava. From Captions to Visual Concepts and Back.
[4] Yoshua Bengio, Réjean Ducharme, Pascal Vincent, Christian Jauvin. A Neural Probabilistic Language Model.
[5] Courses notes on NLP by Michael Collins, Columbia University.
[6] A. Farhadi, M. Hejrati, M. A. Sadeghi, P. Young, C. Rashtchian, J. Hockenmaier, and D. Forsyth. Every picture tells a story: Generating sentences from images. In ECCV. 2010.
[7] Kai Yu, Zijian Zhao, Xueyang Wu, Hongtao Lin and Xuan Liu. Rich Short Text Conversation Using Semantic Key Controlled Sequence Generation. At IEEE Transaction on ASL.
[8] Itzair Aldabe and Montse Maritxalar. Semantic Similarity Measures for the Generation of Science Tests in Basque. At IEEE Transactions on Learning Technologies
[9] Dr. C. Nalini, Shwetambari Kharabe, Sangeetha S," Efficient Notes Generation through Information Extraction", International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249 – 8958, Volume-8 Issue-6S2, August 2019.
[10] Shwetambari Kharabe, C. Nalini , R. Velvizhi," Application for 3D Interface using Augmented Reality", International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249 – 8958, Volume-8, Issue-6S2,August 2019.

ISSN: 2233-7857 IJFGCN