# CLARIFICATION QUESTION GENERATION

Lokesh JK
*AI21MTECH14001*

Karthikeyan Mohanraj
*AI21MTECH14007*

Tejas Arya
*AI21MTECH14004*

Kaushiki Dwivedi
*AI21MTECH14003*

Sarvani Mathigetta
*SM21MTECH12004*

## Abstract

*The problem statement is to create a model that asks questions to fill information gaps of the posted question, typically through generating clarification questions.* [1]

## 1. Introduction

Identifying missing information in the given context which is currently missing from text is an under explored aspect of text understanding. Recently proposed task of clarification question generation can aid machine learning models reduce the ambiguity in a given context. Rao and Daumé III (2018, 2019) proposed models for this task which is successful at generating fluent and relevant questions but still falls short in terms of usefulness and identifying missing information.Even with advent of large-scale pretrained generative models (Radford et al., 2019; Lewis et al., 2019; Raffel et al., 2019) were not successful in going beyond fluency and relevance.To do so, we must first recognise what is lacking, which, if included, would be beneficial to the consumer of the information. This could be achieved by taking humans as inspiration , who in general identify missing information by using global knowledge,i.e recollecting previous experiences and comparing them to the current one to ascertain what information is missing and if added would be the most useful. This project is about inculcating the above mentioned idea into the model so that it generates clarification questions on missing information.

### 1.1. Motivation

The main motivation is with the advent of high quality speech recognition and text generation systems, we are increasingly using dialog as a mode to interact with devices (Clark et al., 2019). However, these dialog systems still struggle when faced with ambiguity and could greatly benefit from having the ability to ask clarification questions.

[1]Our code is available at https://github.com/LokeshJatangi/Diverse_specific_clarification_questions

## 2. Literature Review

In paper[1] the author describes a novel approach to the problem of clarification question generation and proposed a Generative Adversarial Network (GAN) based model where the generator is a sequence-to-sequence model that generates questions, and the discriminator is a utility function that models the value of updating the context with the answer to the clarification question. They developed three model variants and compared their performance with the baseline model. The models were trained with Amazon product description and stack Exchange posts datasets, and they evaluated the model on automated evaluation metrics and crowd sourced human judgments. Finally, they concluded that the model with an adversarial training approach produces more useful and specific questions compared to both a model trained using maximum likelihood objective and a model trained using utility reward-based reinforcement learning.

In paper[2], the author formulated the task to tackle the problem of missing information in the product description on e-commerce websites by proposing the task of Diverse CQGen to request various unstated aspects in writing with a group of semantically different questions. To deal with the specificity challenge, determined by the size of its applicable range, proposed a novel model named Keyword Prediction and Conditioning Network(KPCNet) Keywords in CQs. Its keywords can capture the main semantic of a question. The clustering method also explored a novel use of producing coherent keyword groups for keyword selection to generate the correct, specific, and diverse questions. The model is trained on the Home and Kitchen category of the Amazon dataset. The model is evaluated on both automatic metrics and human judgments. They concluded that the model covers various information needs and improves the robustness to problematic generations.

In paper [3], authors described a model that identifies useful missing information in each context (schema) i.e., it

generates clarification questions to reduce ambiguity. They stated that the model fills the missing information from global knowledge or from previous experience just like how humans do. In the first stage, they found what is missing by taking a difference between the global knowledge's schema and schema of the local context and then fed that missing schema to a fine-tuned BART-model to generate a question which is further made more useful using PPLM. They tested this model on two scenarios community-QA (product-description from amazon.com) and dialog history from the Ubuntu Chat forum and evaluated on Automatic and human judgment metrics. Finally, they concluded that the framework works across domains, shows robustness towards information availability, and responds to the dynamic change in global knowledge.

In paper[4], authors investigated the problem of generating informative questions in information asymmetric conversations. In this paper they worked on the scenario where the questionnaire is not given the context from which answers are drawn. The core challenges they worked upon are defining the informativeness of potential questions and exploring the prohibitively large space of potential questions to find good candidates. This paper is the first attempt at studying question generation to seek information in open-domain communication. Authors found out that optimizing metrics(quantify how much new information question reveal) to quantify informativeness of questions via reinforcement learning leads to a better system that behaves pragmatically and has improved communication efficiency.

In paper[5], the author formulated the task of asking clarifying questions in open-domain information-seeking conversational systems. Proposed an offline evaluation methodology for the task and collected a dataset, called Qulac (Questions for lack of clarity), through crowd sourcing. Dataset is built on top of the TREC Web Track 2009-2012 data collections and consists of over 10K question-answer pairs for 198 TREC topics with 762 facets. At the second stage, the proposed model, called NeuQS, aims to select the best question to be posed to the user based on the query and the conversation context. The Question Retrieval Model is described as the BERT- Language Representation based, called BERT-LeaQuR. The aim is to maximize the recall of the retrieved questions, retrieving all relevant clarifying questions to a given query in the top k questions. Hence, illustrated the workflow of a conversational search system, focusing on asking clarifying questions, addressing all the facets in the collection.

In paper [6], introduce the task of clarification question generation about missing information in a given context.They developed a model which is inspired by Expected value of perfect information ( EVPI) in which joint neural network composed of LSTM's, models the probability distribution of answer given a (post , question) tuple by generating answer representation and also models a Utility function that calculates utility of updated post as a binary classification problem.The authors have defined task as ranking the most useful questions , and they evaluated against expert human annotations using a StackExchange dataset made up of (posts, questions, and responses) triples.Finally, The authors conclude EVPI model with answer candidates is a promising formalism for the clarification question generation task as they outperform neural baseline models.

## 3. Implementation of Reference paper[2]

### 3.1. Dataset preparation

The collected dataset contained many noise and unhelpful questions. The following steps are used to remove those questions. The techniques are Fixing Unescaped HTML characters(removal of unescaped HTML special characters), Removing non-question parts(declarative sentences which are not relevant to the context are removed), and removing noise questions(Questions that are universal are removed). The data preprocessing was done only on the training and validation dataset. But, the test dataset is preserved.

### 3.2. Hyperparameters and other settings

The max length of context and the question for the KPCNet-beam is 100 and 20, respectively, and the number of GRU layers set for both the encoder and decoder is set to be 2 with 100 hidden units. The learning rate of 0.0003 to train at most 60 epochs. The KPCNet is implemented in PyTorch and trained on Ubuntu 18.04.4 LTS server with one NVIDIA GeForce RTX 2080 Ti.

For the Home  Kitchen dataset, the model is operated on 200D word embeddings borrowed from Rao and Daumé III [1], which are pre-trained from in-domain data with Glove [7] and are frozen during training. For the Office dataset, all models are operated on 200D word embeddings that we pre-trained from in-domain data with Word2vec[8] in gensim5.

### 3.3. Inference results

In the case of Figure 1, Even though the description emphasizes the design, the luxury of the product, the questions generated are focused on keywords like dimensions, height, and a number of pieces of the product.

In the case of Figure 2, Generated questions are on the thickness of the foam and the density of foam which are missing information in Context. These questions are informative and diverse, which provides clarity to the buyer of the product.

In the case of Figure 3, generated questions provide pertinent information to the buyer of the product as the answer to these questions provide clarity on the product's utility.

In the case of Figure 4, the description concentrates on the features and material of the cover. The questions gen-

**Description:** Manhattan dark cherry finish glass top dining 7-piece table set . an exquisite masterpiece , inspired by today 's most modern designs , the Manhattan dark cherry finish glass top dining table set is the perfect addition for those looking to add a casual and luxurious tone to their dining room . plush leatherette seats paired with the grand 72 " glass table top gives this set its contemporary and trend-setting appeal while its unique pedestal base provides maximum style and sturdiness .

**Questions:**

- what are the dimensions of the seat ?'
- 'what is the height of the chair ?',
- 'what are the dimensions of this table'?

Figure 1. Clarification question generated on dining table

**Description:** Sleep innovations 4-inch dual layer mattress topper . 10-year limited warranty . made in the USA . queen size

**Questions:**

- 'what are the dimensions of this topper ?',
- 'what is the warranty on this topper ?',
- 'how long is the mattress ?'
- 'what is the thickness of the memory foam ?

Figure 2. Clarification question generated on Bed

**Description:**

Whistling tea kettle in purple . kitchen-works 2.5 qt . whistling tea kettle ( purple ).

**Questions:**

- 'how long does it take to boil water ?'
- 'how long does this kettle last ?'
- 'does this tea kettle whistle ?'

Figure 3. Clarification question generated on Tea kettle

erated requests information on the suitability of the mixer cover to the user.

In the case of Figure 5, the description provides complete information on the technical specifications of the product. The keywords generated could not capture missing information like the warranty of the product, so the questions generated are redundant.

## 3.4. Evaluation Metrics

### 3.4.1 BLEU

Bilingual Evaluation Understudy is a metric for machine translation. The BLEU score is a number between zero and one that measures the similarity of the machine-translated text to a set of high-quality reference translations. BLEU

**Description:**

Kitchen-aid khaki cover for tilt-head stand mixer ksmct1kb . keep your kitchen-aid tilt-head stand mixer looking like new with this khaki mixer cover . designed to neatly fit over any kitchen-aid tilt-head style stand mixer , this convenient quilted cover is made from 100 % cotton and features a zipper closure and front pocket to stow favourite recipes . to wash your cover , turn inside out , machine wash cold , gentle cycle . only wash with like colors . use mild detergent and non-chlorine bleach only . tumble dry low . remove promptly from the dryer .

**Questions:**

- 'will this fit the 720 ?'
- 'will this fit a kitchen-aid <unk> ?'
- ,'will this cover fit a 4.5 quart mixer ?'
- 'does this cover have a lid ?'
- 'will this cover fit the kitchen-aid stand mixers ?'

Figure 4. Clarification question generated on Mixer cover

**Description:**

gourmet grade gr1100n 14-cup food processor , ultra quiet powerful heavy duty motor . ultra-quiet , 1400w `` dynamic drive '' motor delivers up to 1400 watts of power for faster , better & more consistent results massive 14 cup bowl with easy-lock lid & ultra-wide fast feed tube . includes precision milled stainless steel blades disks for shredding & slicing+ juicer attachment also includes the ultimate kugel blade- a whole new kind of blade for a whole new caliber of kugel includes fine shredder thin slicer disk , French fry disk , citrus juicer , s-shaped slicer , the ultimate kugel blade .

**Questions:**

- 'does it come with the food processor ?'
- 'does this machine come with a blade ?',
- 'where can I get replacement parts for this model ?'

Figure 5. Clarification question generated on Mixer

scores are used as a benchmark for text generation, used Pairwise –BLEU and Avg BLEU as an evaluation metric for diverse machine translation.

### 3.4.2 Distinct-3

Distinct-3 is an algorithm for evaluating the textual diversity of the generated text by calculating the number of distinct n-grams.

### 3.4.3 METEOR

METEOR (Metric for Evaluation of Translation with Explicit Ordering) is a metric for evaluating the machine-translated output. It is used for individual-level automatic evaluation.

## References

[1] Rao, S. and Daumé III, H., 2019. ” Answer-based adversarial training for generating clarification questions.” arXiv preprint arXiv:1904.02281.

[2] Zhang, Z. and Zhu, K., 2021, April. " Diverse and Specific Clarification Question Generation with Keywords ". In Proceedings of the Web Conference 2021 (pp. 3501-3511).

[3] Majumder, B.P., Rao, S., Galley, M. and McAuley, J., 2021. " Ask what's missing and what's useful: Improving Clarification Question Generation using Global Knowledge. " arXiv preprint arXiv:2104.06828.

[4] Qi, P., Zhang, Y. and Manning, C.D., 2020. " Stay hungry, stay focused: Generating informative and specific questions in information-seeking conversations. " arXiv preprint arXiv:2004.14530.

[5] Aliannejadi, M., Zamani, H., Crestani, F. and Croft, W.B., 2019, July. " Asking clarifying questions in open-domain information-seeking conversations." In Proceedings of the 42nd international acm sigir conference on research and development in information retrieval (pp. 475-484).

[6] Rao, S. and Daumé III, H., 2018. " Learning to ask good questions: Ranking clarification questions using neural expected value of perfect information." arXiv preprint arXiv:1805.04655.

[7] Jeffrey Pennington, Richard Socher, and Christopher Manning. 2014. GloVe: Global Vectors for Word Representation. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). 1532–1543.

[8] Tomás Mikolov, Ilya Sutskever, Kai Chen, Gregory S. Corrado, and Jeffrey Dean. 2013. Distributed Representations of Words and Phrases and their Compositionality. In Advances in Neural Information Processing Systems 26: 27th Annual Conference on Neural Information Processing Systems 2013. Proceedings of a meeting held December 5-8, 2013, Lake Tahoe, Nevada, United States, Christopher J. C. Burges, Léon Bottou, Zoubin Ghahramani, and Kilian Q. Weinberger (Eds.). 3111– 3119.