Leaf: Multiple-Choice Question Generation

Kristiyan Vachev\textsuperscript{1}  Momchil Hardalov\textsuperscript{1}  Georgi Karadzhov\textsuperscript{2}  Georgi Georgiev\textsuperscript{3}  Ivan Koychev\textsuperscript{1}  Preslav Nakov\textsuperscript{4}  \\
\textsuperscript{1}Sofia University  \textsuperscript{2}University of Cambridge  \textsuperscript{3}Releva.ai  \\
\textsuperscript{4}Qatar Computing Research Institute, HBKU

Summary

- Testing with quiz questions helps learning and retention, but creating such questions is a tedious and time-consuming task.
- We built a system that automatically generates multiple-choice questions.
- We fine-tuned T5 models to generate
  - answers
  - answer distractors
  - answer-aware questions
- We trained on open-source datasets.
- We open-sourced our code and training scripts.

System

**Leaf**

Multiple-Choice Question Generation

Generating multiple-choice questions suitable for knowledge assessment in classroom or industry settings.

Question 1

What group is Oxygen a member of?

- Chalcogen group  \( \checkmark \)
- Oxygen group  \( \times \)
- Halogen group  \( \times \)

Add answer...

Question 2

What is the chemical element with the symbol O and atomic number 8?

- Oxygen  \( \checkmark \)
- Deuterium  \( \times \)
- Carbon dioxide  \( \times \)
- Nitrogen  \( \times \)

Add answer...

Number of questions...  Generate

In schools:

- Students can self-assess.
- Instructors can quickly and efficiently detect knowledge gaps.

In industry:

- New employees can be onboarded faster.
- Managers can make better strategic decisions.

On other platforms:

- Content for learning management systems, which increases user retention.
- Data for training conversational agents or question answering systems.

Question Generation

- Based on a fine-tuned T5 Transformer model.
- Using the open-source dataset SQuAD1.1 containing 100,000+ question-answer pairs.
- Multitask model for answer-aware questions generation and question and answer generation from context.

Distractor Generation

- Fine-tuned a version of the T5 Transformer model to generate three distractors.
- Trained on RACE, an open-source dataset containing 100,000 multiple-choice questions.
- Used sense2vec word embeddings for greater variety of distractors.

Architecture

- The instructor can input their educational content and desired number of questions and can then review the generated questions in our web platform.
- After pre-processing the text, the Multiple Choice Question Generator Module generates and returns the required number of question-answer pairs and distractors.
- Separating the ML Models achieves better abstraction and higher flexibility.