Question Answering Challenge

Mateusz, Biesiadowski
`mb406097@students.mimuw.edu.pl`

Mateusz, Błajda
`mb406098@students.mimuw.edu.pl`

Kajetan, Husiatyński
`kh406160@students.mimuw.edu.pl`

Jakub, Walendowski
`jw406397@students.mimuw.edu.pl`

Supervisor: mgr Sebastian Jaszczur

May 31, 2022

1 Introduction

Our project would consist of building an open-domain question answering system based on Task 4 from PoleEval 2021. The goal of the task is to develop a solution capable of providing answers to general knowledge questions typical for popular TV quiz shows, such as 1 z 10. The evaluation will be carried out on the test-B dataset. Checking if the two answers match will depend on the question type: For non-numerical questions, assess textual similarity. A Levenshtein distance should be computed between the two (lowercased) strings and if it is less than $\frac{1}{2}$ of the length of the gold standard answer, the candidate answer is accepted. For numerical questions (e.g. In which year. . .), assess numerical similarity. Specifically, use a regular expression to extract a sequence of characters that could be interpreted as a number. If such sequences can be found in both answerss and represent the same number, the prediction is accepted.

1.1 Datasets

1. Dataset provided by the competition organizer

2. "Czy wiesz" dataset

1.2   Baseline

Baseline wiki search.

1. Break the question into tokens with spaCy (model pl_core_news_sm) and ignore single character tokens.

2. Send the tokens separated by spaces as a query to Polish Wikipedia API.

3. For each item returned:

   3.1 Split its title into tokens using spaCy,

   3.2 IIf none of the tokens of the title has at least 50 percent overlap (measured as in Evaluation) with any of the tokens of the question:

      3.2.1 Remove the part of the title starting from '(', if found

      3.2.2 Return the title as an answer,

   3.3 Otherwise, continue to the next result.

4. If no answer has been found, delete the first one question token and go back to step 2.

1.3   What experiments will you carry out?

1.3.1   Open questions

We will test how well our model answers questions based on popular Polish TV trivia show "1 z 10". We will check the accuracy over all questions in test set.
If the results would be satisfactory we might test it on diffrent trivia shows like "Kocham cie polsko" or "Awantura o kase".

1.3.2   Closed questions

We will also check how our model performs if possible answers are given like in the TV show "Milionerzy" where with question contestant receives also four possible answers. Our baseline here would be 25% which is equal to answering every answer randomly.

1.3.3   Questions augumentation

We will try paraphrasing questions to extend our training dataset. With extra data we will be able to check if it changed model accuracy on test dataset.

1.3.4   Question classifier

We will test whether by first dividing the questions into 3 categories numeric, closed, open and then processing them with the appropriate specialized models we will achieve higher scores than using one general model.

2   Related Work

Marcińczuk M., Radziszewski A., Piasecki M., Piasecki D. and Ptak M. (2013). Evaluation of a Baseline Information Retrieval for a Polish Open-domain Question Answering System. In Proceedings of the International Conference Recent Advances in Natural Language Processing (RANLP 2013), pp. 428–435. Association for Computational Linguistics. Przybyła P. (2016). Boosting Question Answering by Deep Entity Recognition. arXiv:1605.08675. Vetulani Z. (1988). PROLOG Implementation of an Access in Polish to a Data Base. In Studia z automatyki XII, pp. 5–23. PWN. Walas M. and Jassem K. (2010). Named Entity Recognition in a Polish Question Answering System. In Kłopotek M. A., Marciniak M., Mykowiecka A., Penczek W. and Wierzchoń S. T. (eds.), Intelligent Information Systems, pp. 181–191. Publishing House of University of Podlasie. [Guu+20] Kelvin Guu et al. REALM: Retrieval-Augmented Language Model Pre-Training. 2020. arXiv: 2002.08909 [cs.CL]. [Jos+17] Mandar Joshi et al. TriviaQA: A Large Scale Distantly Supervised Challenge Dataset for Reading Comprehension. 2017. arXiv: 1705.03551 [cs.CL]. [Kar+20] Vladimir Karpukhin et al. Dense Passage Retrieval for Open-Domain Question Answering. 2020. arXiv: 2004.04906 [cs.CL].