UNIPD@SimpleText2024:

A Semi-Manual Approach on Prompting ChatGPT for Extracting Terms andWrite Terminological Definitions

Giorgio Maria Di Nunzio, Elena Gallina, Federica Vezzani
University of Padova

9 September 2024, SimpleText @ CLEF 2024, Grenoble, France

Objective

- Identifying and explaining difficult content using Large Language Models (LLMs) to enhance text simplification.
- Research question:
    - Can a dedicated prompt set a reasonable baseline for automative text extraction by means of a LLM?
    - Can a (limited) human intervention of the results improve the output?

Methodology

- We implement a methodology involves iterative experimentation with various prompting strategies to optimize the performance of the model in this task.
  - Initially analyze a diverse set of complex texts
  - Design and test a series of prompts to guide the LLM
  - Refining prompts based on feedback and evaluation metrics

Experimental setting

Training the prompt

Dense Re-Ranking with Weak Supervision for RDF Dataset Search

Qiaosheng Chen, Zixian Huang, Zhiyang Zhang, Weiqing Luo, Tengteng Lin, Qing Shi, and Gong Cheng(✉)

State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing, China
{qschen,zixianhuang,zhiyangzhang,wqluo,tengtenglin, qingshi}@smail.nju.edu.cn, gcheng@nju.edu.cn

Abstract. Dataset search aims to find datasets that are relevant to a keyword query. Existing dataset search engines rely on conventional sparse retrieval models (e.g., BM25). Dense models (e.g., BERT-based) remain under-investigated for two reasons: the limited availability of labeled data for fine-tuning such a deep neural model, and its limited input capacity relative to the large size of a dataset. To fill the gap, in this paper, we study dense re-ranking for RDF dataset search. Our re-ranking model encodes the metadata of RDF datasets and also their actual RDF data—by extracting a small yet representative subset of data to accommodate large datasets. To address the insufficiency of training data, we adopt a coarse-to-fine tuning strategy where we warm up the model with weak supervision from a large set of automatically generated queries and relevance labels. Experiments on the ACORDAR test collection demonstrate the effectiveness of our approach, which considerably improves the retrieval accuracy of existing sparse models.

Experimental setting

Training the prompt - first round

EL **You**

You are a terminologist working within the framework of the SimpleText track, which is a part of the CLEF initiative. CLEF (Conference and Labs of the Evaluation Forum) is an organization promoting research in multilingual information access, and the SimpleText track uses a corpus of scientific literature abstracts and popular science requests. Based on the abstract provided below, you are required to perform three tasks:
Task 1: Identify all terms
Task 2: Determine the difficulty level of the terms extracted in Task 1 (Easy, Medium, Difficult)
Task 3: Provide an intensional definition for the terms identified as difficult in Task 2

Abstract (Dense Re-Ranking with Weak Supervision for RDF Dataset Search, by Qiaosheng Chen et al.): Dataset search aims to find datasets that are relevant to
a keyword query. Existing dataset search engines rely on conventional sparse retrieval models (e.g., BM25). Dense models (e.g., BERT-based) remain under-investigated for two reasons: the limited availability of labeled data for fine-tuning such a deep neural model, and its limited input capacity relative to the large size of a dataset. To fill the gap, in this paper, we study dense re-ranking for RDF dataset search. Our re-ranking model encodes the metadata of RDF datasets and also their actual RDF data—by extracting a small yet representative subset of data to accommodate large datasets. To address the insufficiency of training data, we adopt a coarse-to-fine tuning strategy where we warm up the model with weak supervision from a large set of automatically generated queries and relevance labels. Experiments on the ACORDAR test collection demonstrate the effectiveness of our approach, which considerably improves the retrieval accuracy of existing sparse models.

Experimental setting

Training the prompt - first round

Task 2: Determine the difficulty level of the terms

- Easy: Dataset search, Keyword query, Sparse retrieval models, BM25, Metadata, RDF data, Training data, Automatically generated queries, Relevance labels, Retrieval accuracy
- Medium: Dense models, BERT-based, Dense re-ranking, Coarse-to-fine tuning strategy, Weak supervision, ACORDAR test collection
- Difficult: RDF dataset search

Task 3: Provide an intensional definition for the terms identified as difficult in Task 2

- RDF dataset search: RDF (Resource Description Framework) is a standard model for data interchange on the web. RDF dataset search involves finding datasets structured using RDF that are relevant to a given query. It entails searching for datasets where information is organized according to the RDF model, which represents data as triples of subject-predicate-object statements. This task typically involves searching for datasets that contain RDF metadata and possibly RDF data, and it may require specialized techniques for indexing, querying, and retrieving RDF-encoded information efficiently.

Experimental setting

Training the prompt - second/third round

Task 2: Determine the difficulty level of the terms extracted in Task 1 (Easy, Medium, Difficult) for the general public

- Easy: Dataset search, Keyword query, Sparse retrieval models, BM25, Metadata, RDF data, Fine-tuning, Deep neural model, Relevance labels, Training data, Retrieval accuracy
- Medium: Dense models, BERT-based, Dense re-ranking, RDF dataset search, Input capacity, Automatically generated queries, ACORDAR test collection
- Difficult: Coarse-to-fine tuning strategy, Weak supervision

Task 3: Provide an intensional definition for the terms identified as difficult in Task 2

- Coarse-to-fine tuning strategy: A strategy used in machine learning and optimization where the model or algorithm is initially trained or adjusted at a high level of granularity (coarse) and then progressively refined (fine-tuned) at increasingly finer levels of detail. This approach allows the model to first capture the general trends or features of the data before focusing on more specific and nuanced patterns. Coarse-to-fine tuning strategies are often employed to improve the efficiency and effectiveness of learning algorithms, particularly when dealing with large and complex datasets.
- Weak supervision: A method in machine learning where the training data is labeled with noisy, imprecise, or incomplete annotations instead of precise labels. Unlike traditional supervised learning where each training example is associated with a fully accurate label, weak supervision

Experimental setting

Runs

- **unipd_t21t22_chatgpt** - contains the ChatGPT output without any modification;

- **unipd_t21t22_chatgpt_mod1** - remove manually non terms

- **unipd_t21t22_chatgpt_mod2** - additional manual corrections
  - remove partial/not meaningful multi-word terms
  - split “body mass (BM)” into “body mass” and “BM”
  - incomplete terms are completed
  - reassigned terms to the correct sentence (wrong id)

Experimental setting

Removal of non terms

G01.1_1000902583_1	mobile devices	e
G01.1_1000902583_1	applications	e
G01.1_1000902583_1	wildfire confrontation	d
G01.1_1000902583_1	end-users	e
G01.1_1000902583_1	data and information	e
G01.1_1000902583_1	sharing of intelligence	m
G01.1_1000902583_1	coordination	m
G01.1_1000902583_1	personnel	e

Experimental setting

Additional modifications

G01.1_1000902583_5	transferring of information	m
G01.1_1000902583_5	knowledge	e
G01.1_1000902583_5	operation centers	m
G01.1_1000902583_5	field	e
M1_11_1	healthy physically active male subjects	m
M1_11_1	age	e
M1_11_1	body mass (BM)	m
M1_11_1	peak oxygen uptake (VO2peak)	d

Experimental setting

Additional modifications

G01.1_1000902583_5	transferring of information	m
G01.1_1000902583_5	knowledge	e
G01.1_1000902583_5	operation centers	m
G01.1_1000902583_5	field	e
M1_11_1	healthy physically active male subjects	m
M1_11_1	age	e
M1_11_1	body mass (BM)	m
M1_11_1	peak oxygen uptake (VO2peak)	d

G01.1_1000902583_5	wildfire confrontation operation centers	m
G01.1_1000902583_5	field	e
M1_11_1	healthy physically active male subjects	m
M1_11_1	age	e
M1_11_1	body mass	m
M1_11_1	BM	m
M1_11_1	peak oxygen uptake	d
M1_11_1	VO2peak	d

Results

runid	recall overall	precision overall	f1 overall
unipd_t21t22_chatgpt	0.116	0.562	0.192
unipd_t21t22_chatgpt_mod1	0.227	0.398	0.289
unipd_t21t22_chatgpt_mod2	0.331	0.338	0.334
unipt_t21t22_manual	0.545	0.469	0.504
median score for all the runs in the task	0.109	0.561	0.186

Results

runid	recall overall (difficult)	precision overall (difficult)	f1 overall (difficult)
unipd_t21t22_chatgpt	0.077	0.612	0.137
unipd_t21t22_chatgpt_mod1	0.226	0.591	0.327
unipd_t21t22_chatgpt_mod2	0.385	0.682	0.492
unipt_t21t22_manual	0.364	0.904	0.519
median score for all the runs in the task	0.091	0.563	0.157

Results

runid	bleu n1 average	bleu n2 average	bleu n3 average	bleu n4 average
unipd_t21t22_chatgpt	0.309	0.185	0.089	0.049
unipd_t21t22_chatgpt_mod1	0.311	0.181	0.082	0.045
unipd_t21t22_chatgpt_mod2	0.294	0.184	0.091	0.052
unipt_t21t22_manual	0.299	0.189	0.095	0.054
median score for all the runs in the task	0.258	0.143	0.045	0.021

Conclusion

- Lots of questions...

UNIPD@SimpleText2024:

A Semi-Manual Approach on Prompting ChatGPT for Extracting Terms andWrite Terminological Definitions

Giorgio Maria Di Nunzio, Elena Gallina, Federica Vezzani
University of Padova

9 September 2024, SimpleText @ CLEF 2024, Grenoble, France