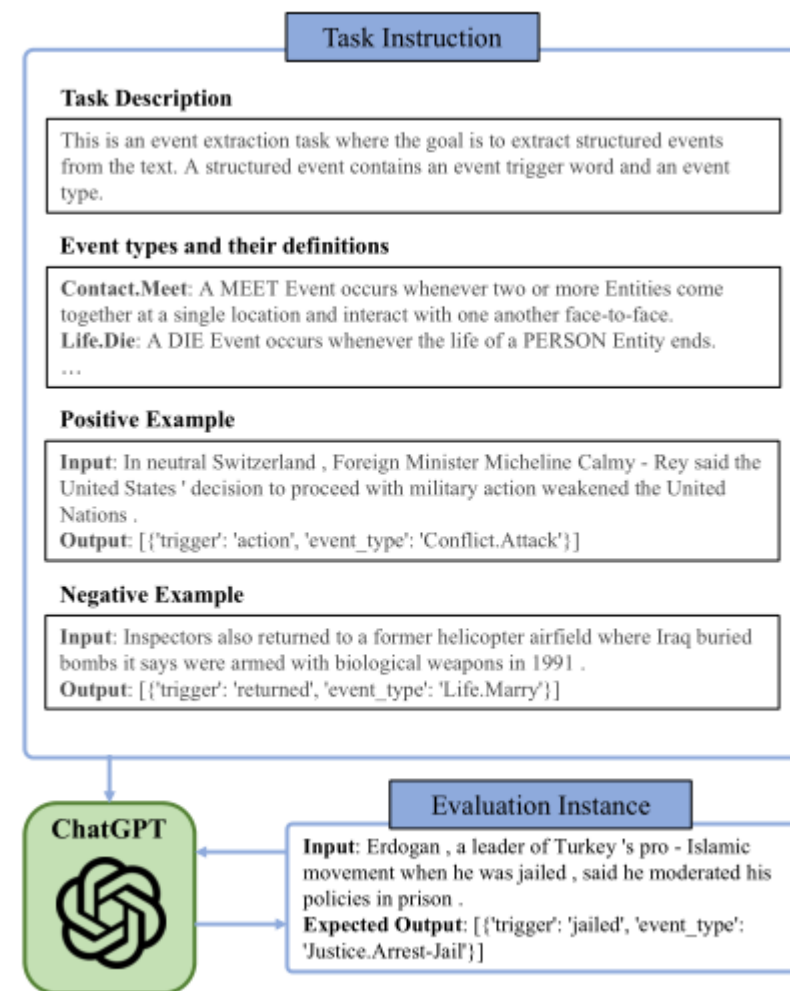# chatgpt阅读汇报

2023.4.19

《Exploring the Feasibility of ChatGPT for Event Extraction》

- 解决问题：zero-shot event extraction --ChatGPT

- 主要方法：ED转为QA(multi-turn)

- 输入：任务指令以及演示示例

- 输出：结构化形式（json）

Task Instruction

Task Description

This is an event extraction task where the goal is to extract structured events from the text. A structured event contains an event trigger word and an event type.

Event types and their definitions

Contact.Meet: A MEET Event occurs whenever two or more Entities come together at a single location and interact with one another face-to-face.
Life.Die: A DIE Event occurs whenever the life of a PERSON Entity ends.
...

Positive Example

Input: In neutral Switzerland , Foreign Minister Micheline Calmy - Rey said the United States ' decision to proceed with military action weakened the United Nations .
Output: [{'trigger': 'action', 'event_type': 'Conflict.Attack'}]

Negative Example

Input: Inspectors also returned to a former helicopter airfield where Iraq buried bombs it says were armed with biological weapons in 1991 .
Output: [{'trigger': 'returned', 'event_type': 'Life.Marry'}]

ChatGPT

Evaluation Instance

Input: Erdogan , a leader of Turkey 's pro - Islamic movement when he was jailed , said he moderated his policies in prison .
Expected Output: [{'trigger': 'jailed', 'event_type': 'Justice.Arrest-Jail'}]

《Exploring the Feasibility of ChatGPT for Event Extraction》

- 五个方面性能比较（数据集:ace2005，随机选择20个测试样例测试，baselines:Text2Event[1]、EEQA[2]）：

- 1、与各微调模型比较（见表1）

	Event Detection		
	P	R	F1
ChatGPT	57.14	72.73	64.00
Text2Event (T5-base)	75.76	75.76	75.76
Text2Event (T5-large)	**82.76**	72.73	77.42
EEQA (BERT-base)	81.82	**81.82**	**81.82**

Table 1: Comparison with task-specific models.

- 1、Lu Y, Lin H, Xu J, et al. Text2event: Controllable sequence-to-structure generation for end-to-end event extraction[J]. arXiv preprint arXiv:2106.09232, 2021.

- 2、Du X, Cardie C. Event Extraction by Answering (Almost) Natural Questions[C]//Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP). 2020: 671-683.

《Exploring the Feasibility of ChatGPT for Event Extraction》

- 2、与不同prompt比较（探究各个输入因素对chatgpt性能影响）：

	Event Detection		
	P	**R**	**F1**
Full Prompt	57.14	72.73	64.00
- Positive Example	48.00	72.73	57.83
- Negative Example	62.50	75.76	68.49
- Event Type Definition	47.50	57.58	52.05

Table 2: Instructing with different elements.

- 3、是否持续输出机构化事件?
  - --测试结果显示：19/20可输出结构化事件，当遇到无法识别的事件触发词或者类型，输出[]

《Exploring the Feasibility of ChatGPT for Event Extraction》

- 4、分析长尾和复杂情况（一句话包含多个事件）的表现：

- low-frequency(<10);high-frequency(>=10)

- Simple sample(=1);complex samples(>=2)

	High Frequency			Low Frequency			Simple Examples			Complex Examples		
	P	**R**	**F1**	**P**	**R**	**F1**	**P**	**R**	**F1**	**P**	**R**	**F1**
ChatGPT	23.08	45.00	30.51	23.40	55.00	32.84	49.00	50.00	28.99	42.31	52.38	46.81
Text2Event (t5-base)	55.56	50.00	52.63	**61.90**	**65.00**	**63.41**	**79.17**	**95.00**	**86.36**	81.82	64.29	72.00
Text2Event (t5-large)	54.55	60.00	57.14	60.00	60.00	60.00	75.00	90.00	81.82	**88.57**	**73.81**	**80.52**
EEQA	**59.09**	**65.00**	**61.90**	61.11	55.00	57.89	78.26	90.00	83.72	82.35	66.67	73.68

Table 3: Performance of ChatGPT and task-specific models in long-tail and complex scenarios.

《Exploring the Feasibility of ChatGPT for Event Extraction》

- 5、评估在最小次数下得到想要的输出

- 参数如下：4个专业标注人员评估结果；ace05test中随机选择10个样例；每人尝试5次；

- 结论：

- 1、不够健壮，不存在持续向上的性能

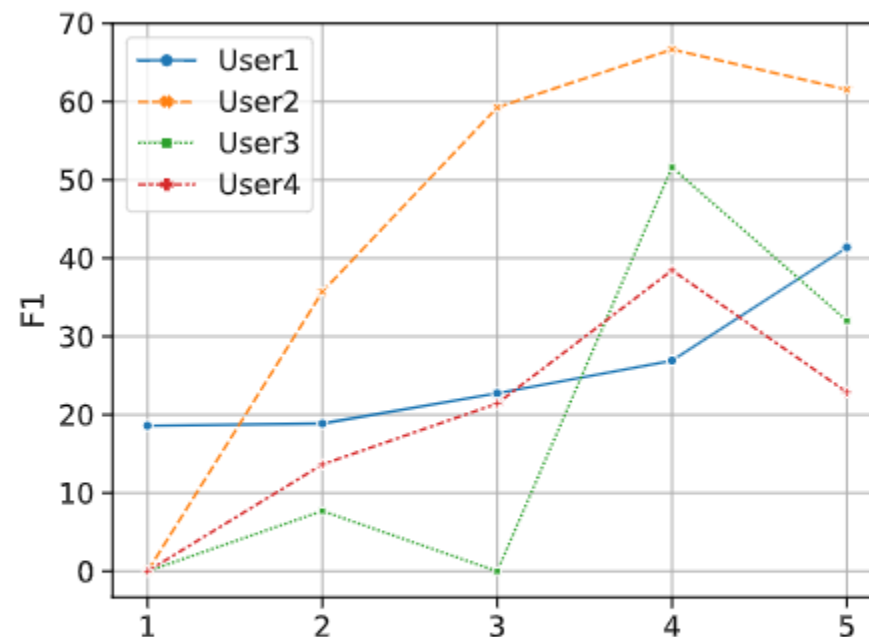- 2、对不同prompt敏感，第4次尝试时最高66.67，
最低26.92

Figure 2: Performance of ChatGPT with different number of attempts by four annotators.

《WavCaps: A ChatGPT-Assisted Weakly-Labelled Audio Captioning Dataset for Audio-Language Multimodal Research》

- 1、解决问题：在线获取的原始数据噪音大（不包含音频内容、只有单词或短语等），利用大模型--chatgpt自动过滤和转换原始数据，提出针对audio-language (AL) multimodal learning任务的数据集WavCaps

- 主要方法：以往的方法（iamage-based filter、text-based filter）导致高丢弃率,因此，提出三阶段的pipline模式过滤噪声并生成高质量的数据，首先基于文本出现的频率过滤掉无关数据，然后基于chatgpt将原始数据生成类似caption的句子，最后我们在后处理阶段细化第二阶段不理想的输出，最终得到第一个大规模，弱标签的音频字幕数据集WavCaps

《WavCaps: A ChatGPT-Assisted Weakly-Labelled Audio Captioning Dataset for Audio-Language Multimodal Research》

- 介绍三步piplines

- 收集原始数据：https://sound-effects.bbcrewind.co.uk/、https://soundbible.com/、BBC、AudioSet

- 1、pre-filtering:删除音频持续时长小于1秒的数据；对freesound数据采用高频文本过滤（音频片段中共享的高频描述通常是由同一个用户在同时上传多个音频片段时上传的，并且与音频通常无关），该阶段过滤掉265000条无关数据。

- 2、ChatGPT-based Transformation：希望的输出以下特征：1）使用简洁的语法对音频对容进行准确描述；2）避免使用无法单独从音频信号中推断出的命名实体，如：人名，位置等；3）排除任何与声音无关的主观信息，如个人感受等，prompt如下：

《WavCaps: A ChatGPT-Assisted Weakly-Labelled Audio Captioning Dataset for Audio-Language Multimodal Research》

TABLE I
EXAMPLE PROMPTS TO CHATGPT FOR FREESOUND AND AUDIOSET STRONGLY-LABELLED SUBSET. TRANSFORMATION EXAMPLES ARE ARE IGNORED.
'SL' REFERS TO 'STRONGLY-LABELLED'.

Data sources	Prompts
FreeSound BBC Sound Effects SoundBible	I will give you a list of descriptions of sounds. Process each individually. Extract the type of the sound and generate an audio caption describing the sound events. The audio caption should be less than 20 words. Delete the author of the sound. Delete locations, city names, country names. Delete the time. Delete device names. Delete the proper noun modifiers, number modifiers, and unit modifiers. Summarize each output into one sentence. Replace all named entities with their hypernyms. Replace people names with "someone". Do not write introductions or explanations. Only describe the sound events and do not use "heard", "recorded". Start each output sentence with its index. Make sure you are using grammatical subject-verb-object sentences. Output "Failure." if the description is not related to sound.
AudioSet SL	I will give you a number of lists containing sound events occurred sequentially in time. Process each individually. Write an one-sentence audio caption to describe these sounds. Make sure you are using grammatical subject-verb-object sentences. Directly describe the sounds and avoid using the word "heard". The caption should be less than 20 words.

《WavCaps: A ChatGPT-Assisted Weakly-Labelled Audio Captioning Dataset for Audio-Language Multimodal Research》

● 接2、ChatGPT-based Transformation：输入输出

TABLE II
EXAMPLES OF RAW DESCRIPTIONS AND CHATGPT'S OUTPUTS.

Data sources	raw descriptions	ChatGPT's output
FreeSound	This sound is of a book falling down the staircase in the Stanford University library west stacks. After 2h and 15min. Excerpt.	A book is falling down a staircase. Failure.
BBC Sound Effects	Timber & Wood - Rip saw, carpenters' workshop. Motor Car: Ford Ecsort 1300 (Automatic) - Ford Escort 1300, exterior, passing with horn. (1300cc engine, automatic transmission.)	Someone is using a rip saw in a carpenter's workshop. A car is passing with its horn.
SoundBible	Tasmanian Devil growling screaming hissing. Warning sounds from a Tasmanian Devil in Zoo. Large Tibetan Bells ringing in a temple. Could also use for Monastery or Monks.	An animal is growling, screaming, and hissing. Bells are ringing.
AudioSet SL	['Accelerating, revving, vroom', 'Race car, auto racing'] ['Female speech, woman speaking', 'Whoosh, swoosh, swish']	A race car is accelerating and revving. A woman is speaking while something whooshes.

《WavCaps: A ChatGPT-Assisted Weakly-Labelled Audio Captioning Dataset for Audio-Language Multimodal Research》

- 3、Post-Processing：由于经chatgpt处理后的数据可能存在如很难删除原文中的数量、人名、地名等信息，此外还有小部分与原音频无关，后处理的策略为：首先使用ner模型识别包含数量、人命、地名的不合理输出；然后再次使用chatgpt,一般可删除命名实体信息，如果还不能删除，则丢弃该数据。

- WavCaps数据分析：处理前后对比

TABLE III
STATISTICS OF HARVESTED RAW DATA AND WAVCAPS DATASET.

Data Sources	Before Processing			After Processing		
	num. of audio	avg. audio duration (s)	avg. text length	num. of audio	avg. audio duration (s)	avg. text length
FreeSound	567078	56.87	17.74	262300	85.98	6.77
BBC Sound Effects	33064	115.75	15.91	31201	115.04	9.67
SoundBible	1576	11.20	17.90	1232	13.12	5.87
AudioSet SL subset	108317	10.00	-	108317	10.00	9.79
WavCaps	-	-	-	403050	67.59	7.80

《WavCaps: A ChatGPT-Assisted Weakly-Labelled Audio Captioning Dataset for Audio-Language Multimodal Research》

● 接WavCaps数据分析：

top100个单词云分布

Fig. 2. Word clouds of top 100 words in WavCaps dataset (top) and the entire harvested raw descriptions (bottom), stop words are ignored.