Instruction Embedding: Latent Representations of Instructions Towards Task Identification

Yiwei Li[1*], Jiayi Shi[1*], Shaoxiong Feng[2], Peiwen Yuan[1], Xinglin Wang[1], Boyuan Pan[2], Heda Wang[2], Yao Hu[2], Kan Li[1]

[1] School of Computer Science, Beijing Institute of Technology
[2] Xiaohongshu Inc

Motivation

The latent representation of instructions is essential for tasks like data selection for instruction tuning and prompt retrieval for in-context learning. While previous studies obtain text embeddings by capturing their overall semantic information, the embeddings of instructions should focus on **identifying their task categories**.

We propose a new concept called instruction embedding, a specialized subset of text embedding that **prioritizes task identification** for instructions over the extraction of sentence-level semantic information.

Sample1 - different tasks

- Tell me the main idea of this article.
- Tell me the gender of the author of this blog post.

Similarity with text embedding: 0.9943
Similarity with instruction embedding: -0.0254

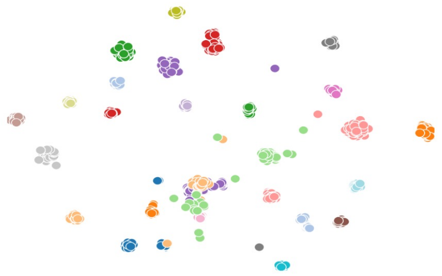
Sample2 – similar tasks

- Create a poem with at least 5 lines, rhyming pattern aabb.
- Write a limerick based on the following noun.

Similarity with text embedding: 0.3239
Similarity with instruction embedding: 0.8287

(a)

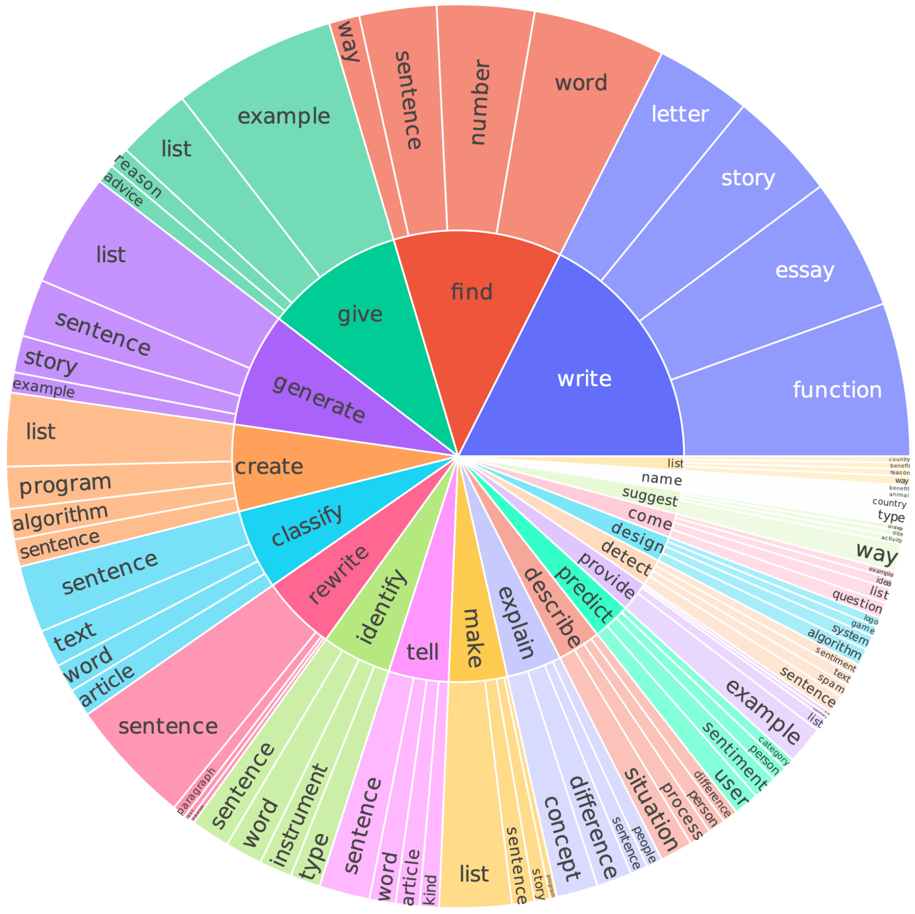
(b)



(c)

Instruction Embedding Benchmark (IEB)

We construct IEB, a new benchmark for instruction embedding training and evaluation which is **labeled by task categories of instructions**. We define the task as a category of activities or work that we expect the LLM to perform, which can be represented by a key phrase (mostly verb-noun phrases). We parse the instructions to extract the phrase through syntactic analysis.



Instruction Embedding Benchmark (IEB)

Parsing Tag	Task Annotation	Examples
VP	verb + noun	Write an sessay about my favourite season. Compose a song about the importance of computer science.
SBARQ	wh- + knowledge	What is the difference between machine learning and deep learning? Why are matrices important in linear algebra? How is a liquid chromatography differs from gas chromatography? Who wrote the song House of Love? When was the "No, They Can't" book released? Where was 52nd International Film Festival of India held?
	what + math	What is the result when 8 is added to 3? What is the value of (x - y)(x + y) if x = 10 and y = 15?
SQ	yes/no + knowledge	Was Furze Hill an established community in the 19th century? Did Sir Winston Churchill win the Nobel Peace Prize?
	yes/no + task	Are the following two sentences grammatically correct? Should this comma be included or omitted?
Others	verb + knowledge	Summarize the Challenger Sales Methodology for me. Describe the Three Gorges Dam of China.
	verb	Translate "Bonjour" into English. You need to translate ‘‘I have been to Europe twice" into Spanish.
	verb + math	Multiply 12 and 11. Simplify 2w+4w+6w+8w+10w+12.
	noun + knowledge	Short Summary about 2011 Cricket World Cup. iPhone 14 pro vs Samsung s22 ultra.

Task categories with examples of IEB

Prompt-based Instruction Embedding

PIE-Prompt: guiding the model in extracting the tasks embedded within given instructions.

The essence of an instruction is its task intention. With this in mind, given the instruction below:

{Instruction}

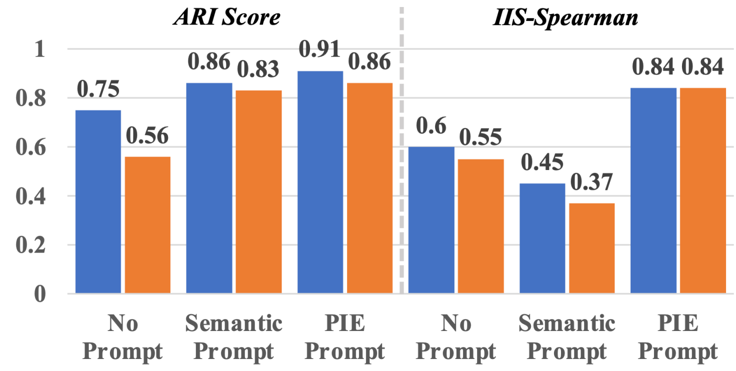
after thinking step by step, the task of the given instruction is:

Fine-tune PIE-model following the contrastive learning (CL) framework in SimCSE.
We replace the dropout-based positive sample pairs construction method with a method based on instruction task labels from training set.

$$\ell_i = -log \frac{e^{sim(h_{ij}, h_{ik})} / \tau}{\sum_{m=1}^{N} e^{sim(h_{ij}, h_{mk'})/\tau}}$$

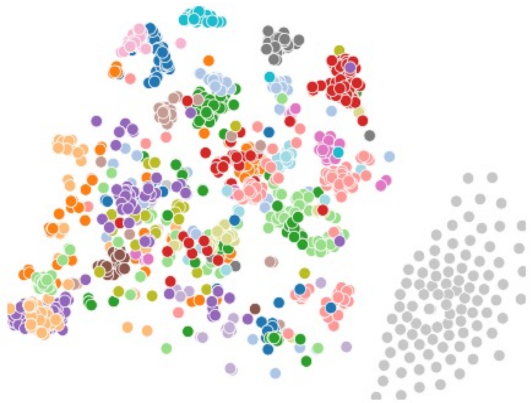
Experiments

Method			ARI	CP	Homo	Silh	IIS-Spearman
None-Fine-tuned							
BERT			0.3113	0.4853	0.6777	0.0792	0.5522
BERT (semantic-prompt)			0.2840	0.4524	0.6570	0.0936	0.5335
BERT (PIE-prompt)			0.2474	0.4038	0.6210	0.0706	0.4724
Llama			0.1813	0.3151	0.5439	0.0995	0.1565
Llama2 (semantic-prompt)			0.4238	0.5947	0.7549	0.1298	0.5893
Llama2 (PIE-prompt)			0.4814	0.6305	0.8014	0.1611	0.7189
Vicuna			0.1198	0.2859	0.4828	0.0934	0.1211
Vicuna (semantic-prompt)			0.1871	0.3145	0.5133	0.1081	0.6934
Vicuna (PIE-prompt)			0.5305	0.6633	0.8242	0.1732	0.7534
Unsupervised Fine-tuned							
Wiki	w/o prompt	Llama2	0.3306	0.4877	0.6891	0.2185	0.1714
		BERT	0.4741	0.6187	0.7741	0.1225	0.7460
	semantic-prompt	Llama2	0.1776	0.3087	0.5412	0.0818	0.1476
		BERT	0.3371	0.5084	0.6974	0.1161	0.6804
Supervised Fine-tuned with hard negative sampling							
EFT-train	w/o prompt	Llama2	0.7541	0.8469	0.9143	0.3608	0.6038
		BERT	0.8837	0.9392	0.9695	0.4574	0.8436
	semantic-prompt	Llama2	0.8651	0.9204	0.9619	0.4542	0.8433
		BERT	0.8876	0.9377	0.9683	0.4946	**0.8450**
	PIE-prompt	Llama2	**0.9125**	0.9432	0.9697	0.4803	**0.8450**
		BERT	0.8974	**0.9453**	**0.9721**	**0.5180**	0.8446

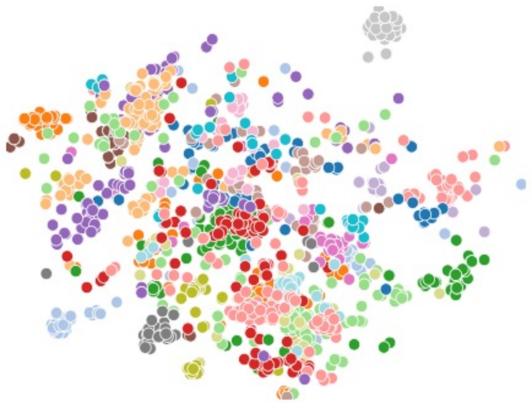


- The PIE-Prompt plays a crucial role to guide LLMs to extract task categories in both learning-free and SFT scenarios.
- Both Llama2 and BERT succeed to learn to identify instructions task categories through fine-tuning
- The hard negative sampling strategy helps model to distinguish positives and negatives through instruction tasks instead of the shortcut of word overlap.

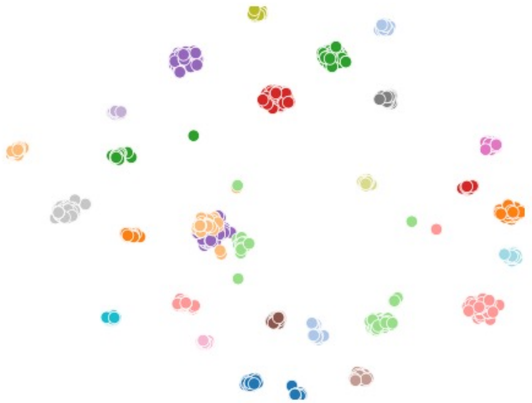
Visualization Analysis



BERT

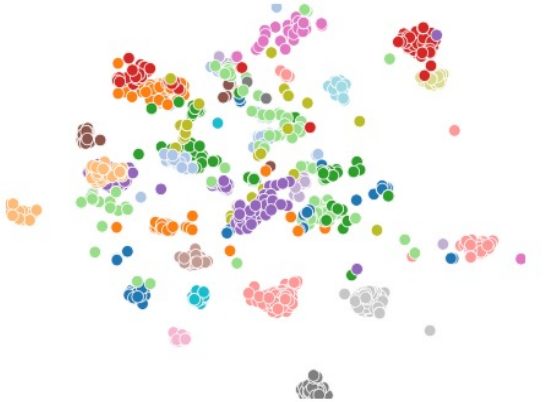


BERT (PIE-Prompt)

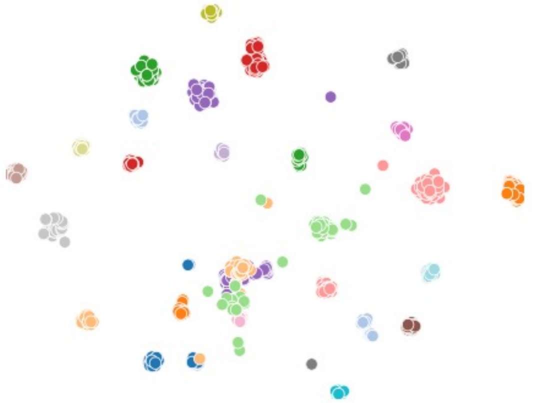


Supervised BERT (PIE-Prompt)

Llama2

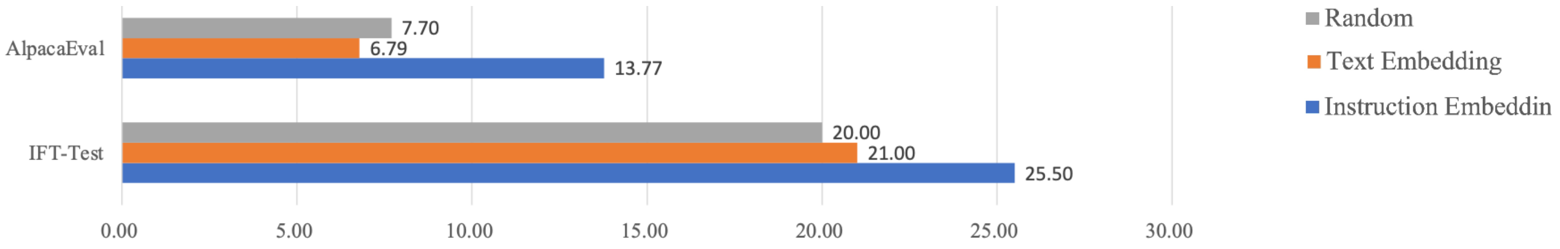


Llama2 (PIE-Prompt)

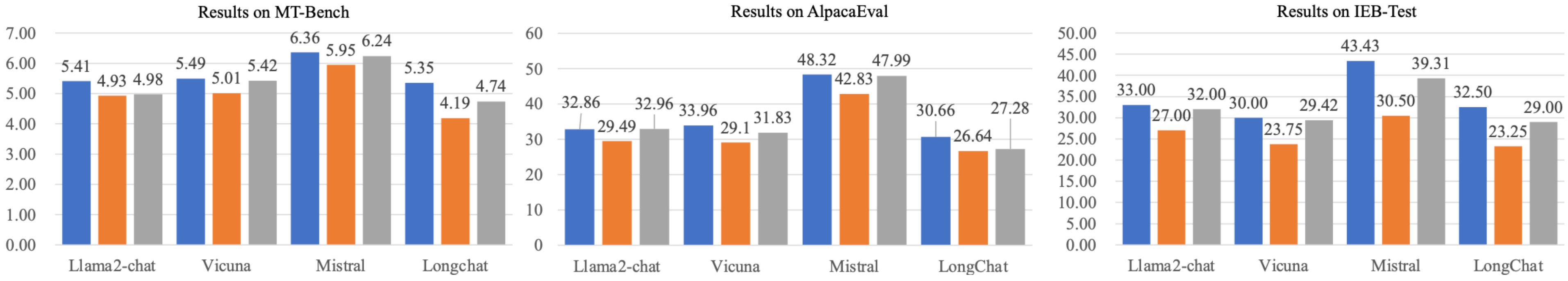


Supervised Llama2 (PIE-Prompt)

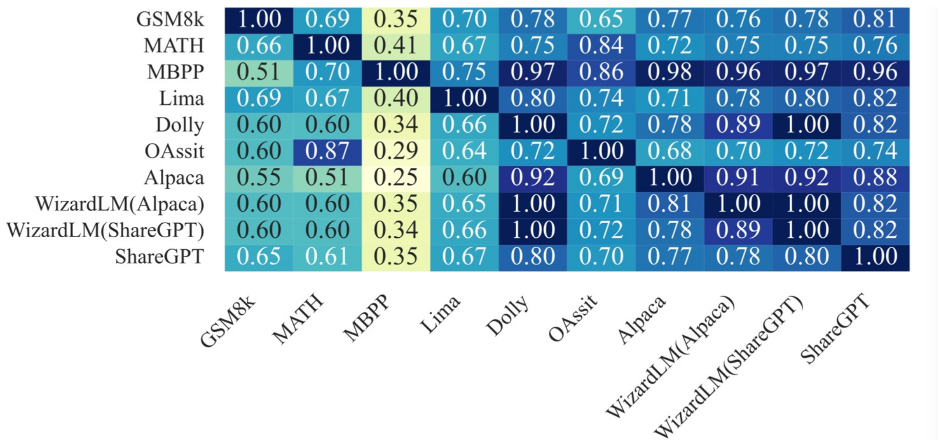
Evaluation on Downstream Tasks



Data selection for instruction tuning



Demonstration retrieval for in-context learning



Dataset task correlation analysis

Model	Instruction Embedding			Text Embedding			Random		
	10	50	100	10	50	100	10	50	100
Llama2-chat	18.40	6.89	**3.17**†	33.50	**5.35**	3.34	**13.46**	5.68	3.97
Vicuna	**8.92**†	**3.76**‡	**3.43**‡	13.22	8.56	3.61	11.53	5.88	4.61
Mistral	7.92‡	**4.27**‡	**2.14**‡	**2.98**	5.05	3.29	10.94	5.67	3.35
Longchat	**7.76**‡	**4.69**‡	3.70	28.82	4.74	**3.47**	12.07	6.11	4.22

Tiny benchmark

Summary

- We introduce the concept of instruction embedding, which prioritizes task identification over traditional sentence-level semantic analysis.

- We propose a prompt-based approach for generating instruction embeddings, applicable in both learning-free and supervised fine-tuning scenarios.

- We demonstrate the superiority of instruction embedding on two basic evaluation tasks and four downstream tasks

Thanks

- liyiwei@bit.edu.cn
- shijiayi@bit.edu.cn

北京理工大学
BEIJING INSTITUTE OF TECHNOLOGY

小红书