

大模型SFT数据筛选

起源：《LIMA: Less Is More for Alignment》 Meta AI

Meta发布了LIMA模型，在LLaMA-65B的基础上，只用1000个精心准备的样本数据进行微调，无需RLHF，就达到了和GPT-4相媲美的程度。据此发布了论文LIMA: Less Is More for Alignment。如标题所示，核心思想就是对于一个强大的LLM，在它预训练的时候就已经学到了绝大部分知识，后续只需要一些精品数据微调就足以让模型产生高质量的内容。不需要搞RLHF哪些复杂的东西。

提出了表面对齐假设：模型的知识 and 能力几乎完全是在预训练期间学习的，而对齐则教会它在与用户互动时应该使用哪种子分布的格式。如果这个假设是正确的，并且对齐主要是关于学习风格，那么表面对齐假设的一个推论是，可以用相当小的一组示例对预训练的语言模型进行充分调整。

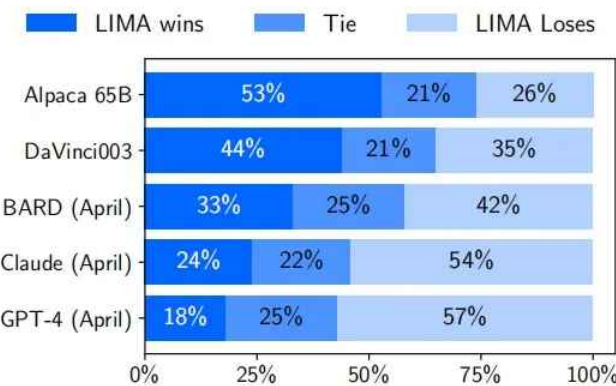


Figure 1: Human preference evaluation, comparing LIMA to 5 different baselines across 300 test prompts.

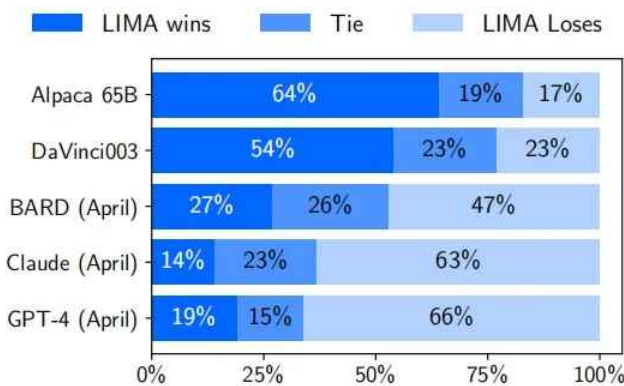


Figure 2: Preference evaluation using GPT-4 as the annotator, given the same instructions provided to humans.

知乎 @北方的郎

所以重要的问题在于，SFT阶段模型需不需要学到新知识？如果假设模型的知识 and 能力几乎完全是在预训练期间学习的，相当于一个已经知道答案的机器，SFT的作用则只是教它如何把答案按人类偏好说出来，而这种任务只需要一小部分数据就能学会。不需要大量的数据提供额外的知识来学习。

这里来具体说明一下：
其实就是base模型到chat模型，到底需要多少数据：

base模型：预训练任务是无监督，只会续写，没有指令遵循能力。

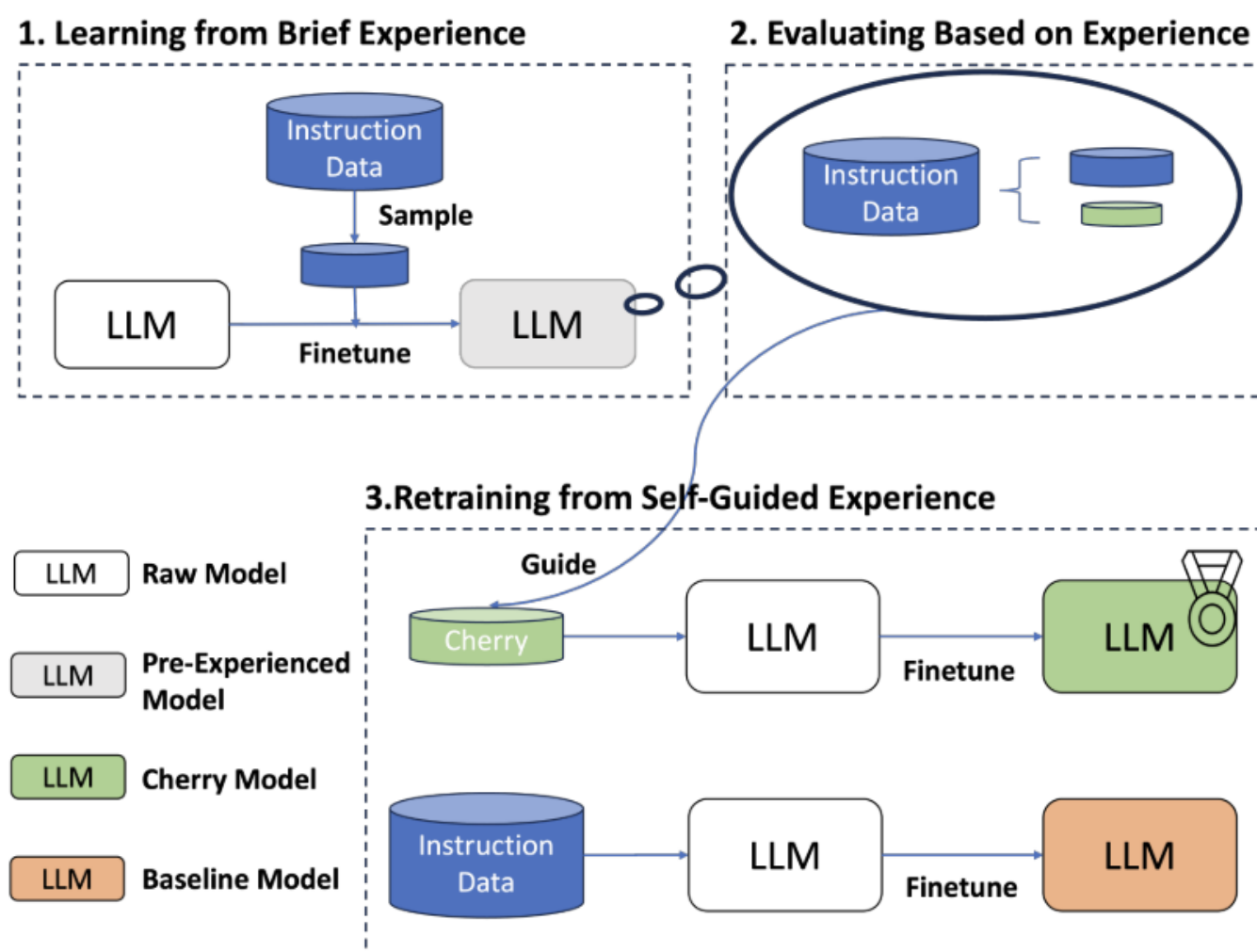
chat模型：能和人类对话的模型，训练任务是SFT，有监督微调。

一.从大量可用数据集中自动识别高质量数据

《From Quantity to Quality: Boosting LLM Performance with Self-Guided Data Selection for Instruction Tuning》

利用IFD指标自动筛选樱桃数据，再利用樱桃数据进行模型指令微调，获取更好地微调模型，主要涉及三个步骤：

- Learning from Brief Experience：利用少量数据进行模型初学；
- Evaluating Based on Experience：利用初学模型计算原始数据中所有IFD指标；
- Retraining from Self-Guided Experience：利用樱桃数据进行模型重训练。



第一步：Learning from Brief Experience

利用少量数据进行模型初学习的原因如下：

- 一些模型为Base模型，只能进行续写，并没有指令遵循的能力；
- LIMA已经证明高质量数据可以让模型具有指令遵循能力；
- 如果采用大量数据进行学习，时间成本和资源成本较高。

在少量数据的选择上，数量选择1k条样本，为了保证数据的多样性，采用K-Means方法对指令进行聚类，共聚出100个簇，每个簇里选择10个样本。并且仅在初始模型上训练1个epoch获取简要预经验模型（Brief Pre-Experience Model）。

第二步：Evaluating Based on Experience

利用简要预经验模型可以对数据集中所有样本进行预测，通过指令内容预测答案内容，并可以获取预测答案与真实答案之间的差异值（利用交叉熵），即条件回答分数（Conditioned Answer Score，CAS），如下：

$$L_{\theta}(A|Q) = \frac{1}{N} \sum_{i=1}^N \log P(w_i^A|Q, w_1^A, w_2^A, \dots, w_{i-1}^A; \theta)$$

根据CAS的高低，可以判断出模型对指令Q生成答案A的难易，但也可能收到模型生成答案A的难易程度的影响。我们利用模型直接对答案进行续写，再根据答案真实内容获取之间的差异值，即直接答案分数（Direct Answer Score，DAS），如下：

$$s_{\theta}(A) = -\frac{1}{N} \sum_{i=1}^N \log P(w_i^A|w_1^A, \dots, w_{i-1}^A; \theta).$$

DAS得分越高，可能表明该答案对模型生成来说本身就更具挑战性或更复杂。为了获取更好的指令数据，也就是哪些指令对模型的影响更高，需要刨除答案本身的影响，因此提出了指令跟随难度（Instruction-Following Difficulty，IFD）分数，如下：

$$s_{\theta}(A|\dot{Q}) = L_{\theta}(A|Q).$$

$$\text{IFD}_{\theta}(Q, A) = \frac{s_{\theta}(A|Q)}{s_{\theta}(A)}$$

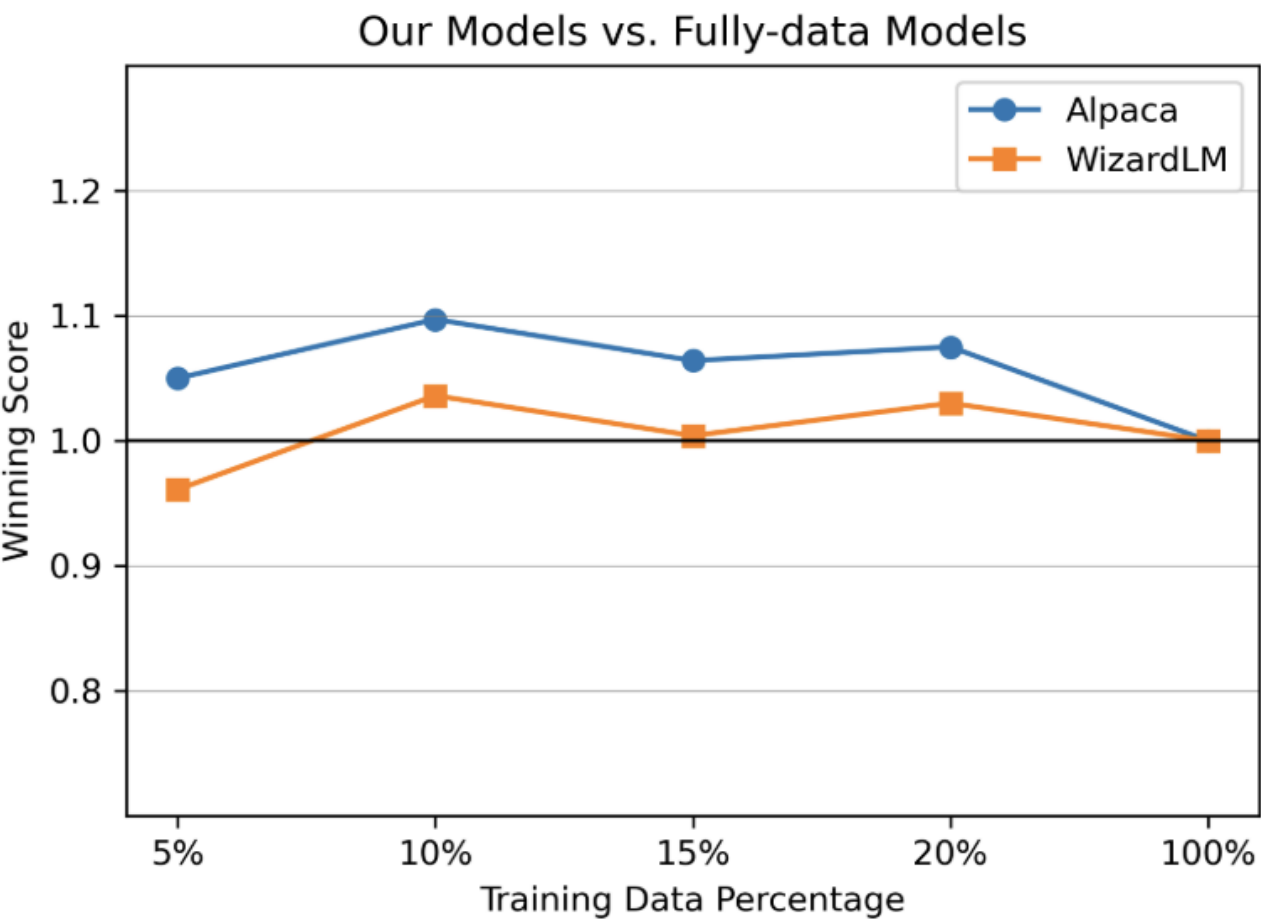
IFD大于1，说明指令让答案更难生成，数据质量低，直接剔除。

第三步：Retraining from Self-Guided Experience

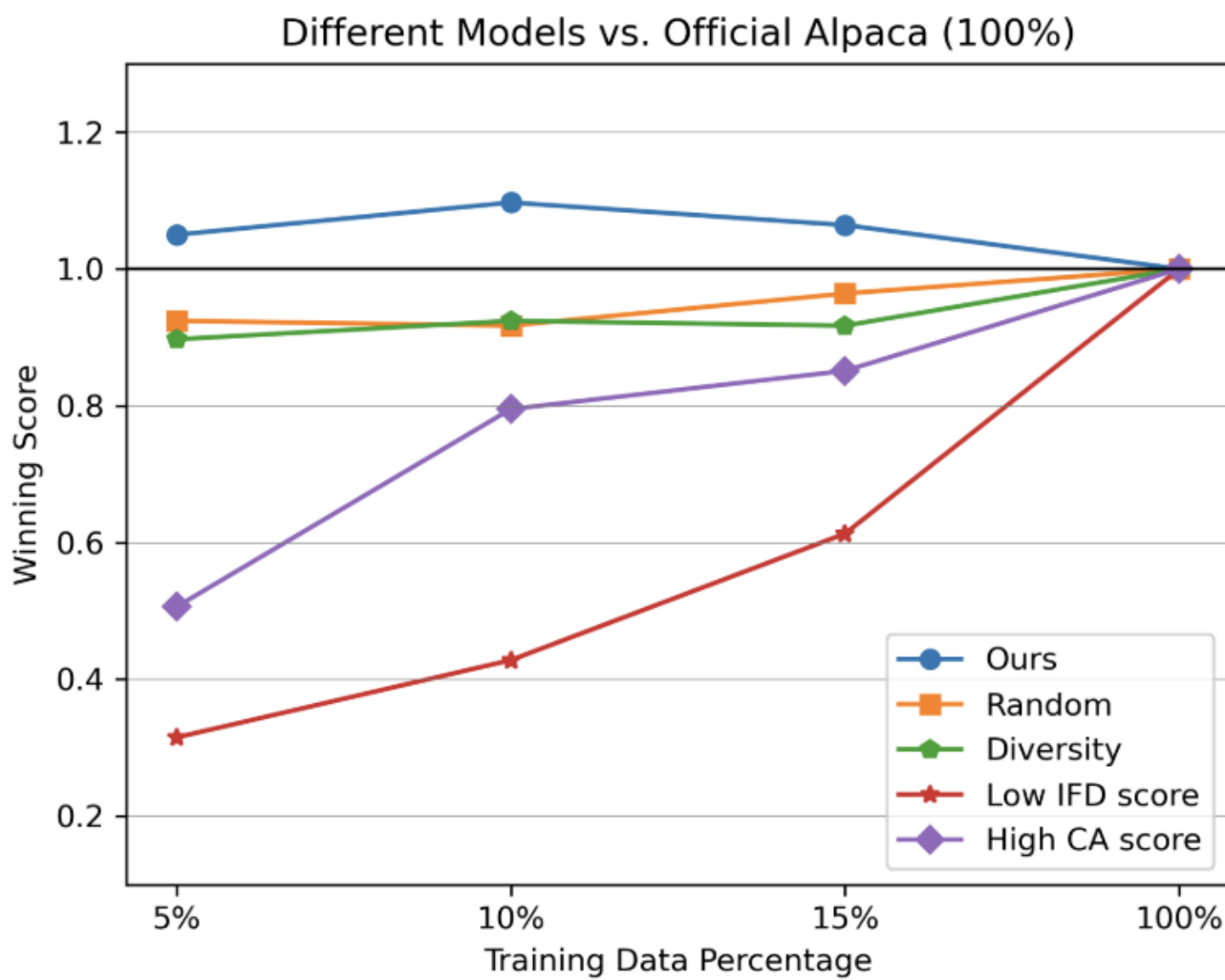
利用IFD指标，对原数据集进行排序，选择分数靠前的数据作为cherry数据，对原始模型进行指令微调，获取cherry模型。

实验结果：

在Alpaca和WizardLM两个数据集上利用Llama-7B进行实验，发现在5%的Alpaca樱桃数据上进行训练就超过了全量数据训练结果。



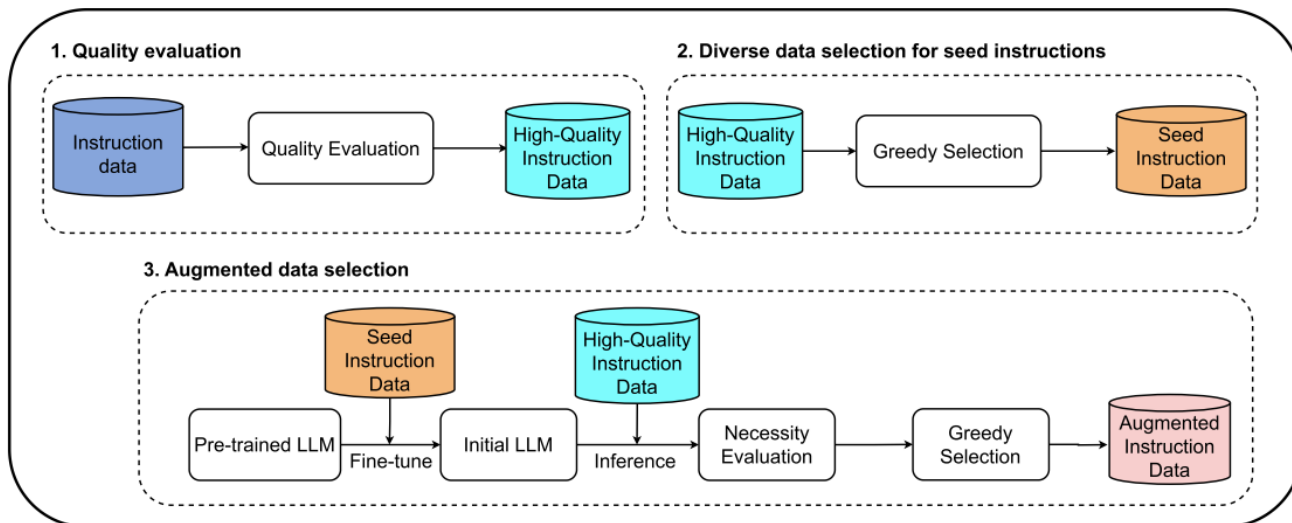
如何判断IFD指标是有效的？对比随机采样、IFD采样、IFD低分采样、CAS采样四种方法对模型指令微调的影响，发现IFD采样在不同数据比例下，均高于全量数据微调效果，但其他采样方法均低于全量数据微调方法。



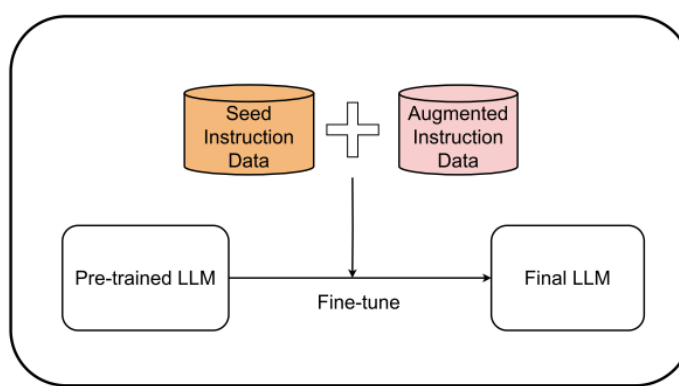
《MoDS: Model-oriented Data Selection for Instruction Tuning》

MoDS方法主要通过质量、覆盖范围、必要性三个指标来进行数据的筛选，其中数据质量是为了保证所选的指令数据的问题和答案都足够好；数据覆盖范围是为了让所选择的数据中指令足够多样、涉及知识范围更广；数据必要性是选择对于大模型较复杂、较难或不擅长的数据以填补大模型能力的空白。整体流程如下图所示，

Instruction Data Selection



Fine-tuning with Selected Data



1.质量筛选:

对于数据进行质量过滤时，采用OpenAssistant的reward-model-debertav3-large-v2模型（一个基于DeBERTa架构设计的奖励模型）对数据进行质量打分。将原始数据的Instruction、Input、Output的三个部分进行拼接，送入到奖励模型中，得到一个评分，当评分超过 α 时，则认为数据质量达标，构建一份高质量数据集-Data1。

Index	instruction	input	output	quality score
1	Classify the following data with three labels.	fjsklfjdsklfjsklfjsklfjs	Label1: Character data\nLabel2: Alphabetical data\nLabel3: Non-alphabetical data	-2.71
2	Change the following sentence from the singular form to the plural form.	The cat is drinking.	The cats are drinking.	2.62
3	Write a simple definition of the word \"economics\".	-	Economics is the study of how society uses its limited resources to produce and distribute goods and services.	6.44
4	Assign this task to the design team.	Design a communication tool for internal team members.	The design team has been assigned the task of designing a communication tool for internal team members.	-0.006
5	Compute the following expression.	$2^3 + 3^2$	11	-3.09

2.多样性筛选

为了避免所选质量数据高度相似，通过K-Center-Greedy算法进行数据筛选，在最大化多样性的情况下，使指令数据集最小。获取种子指令数据集（Seed Instruction Data）-SID。

K中心问题的目标是从给定的数据点集中选择K个中心的子集，以最小化任何数据点与其最近中心之间的最大距离。

Algorithm 1 K-Center-Greedy (Sener and Savarese, 2017)

Input: data x_i , existing pool s^0 and a budget b
Initialize $s = s^0$
repeat
 $u = \operatorname{argmax}_{i \in [n] \setminus s} \min_{j \in s} \Delta(x_i, x_j)$
 $s = s \cup u$
Until $|s| = b + |s^0|$
return $s \setminus s^0$

3.必要性筛选

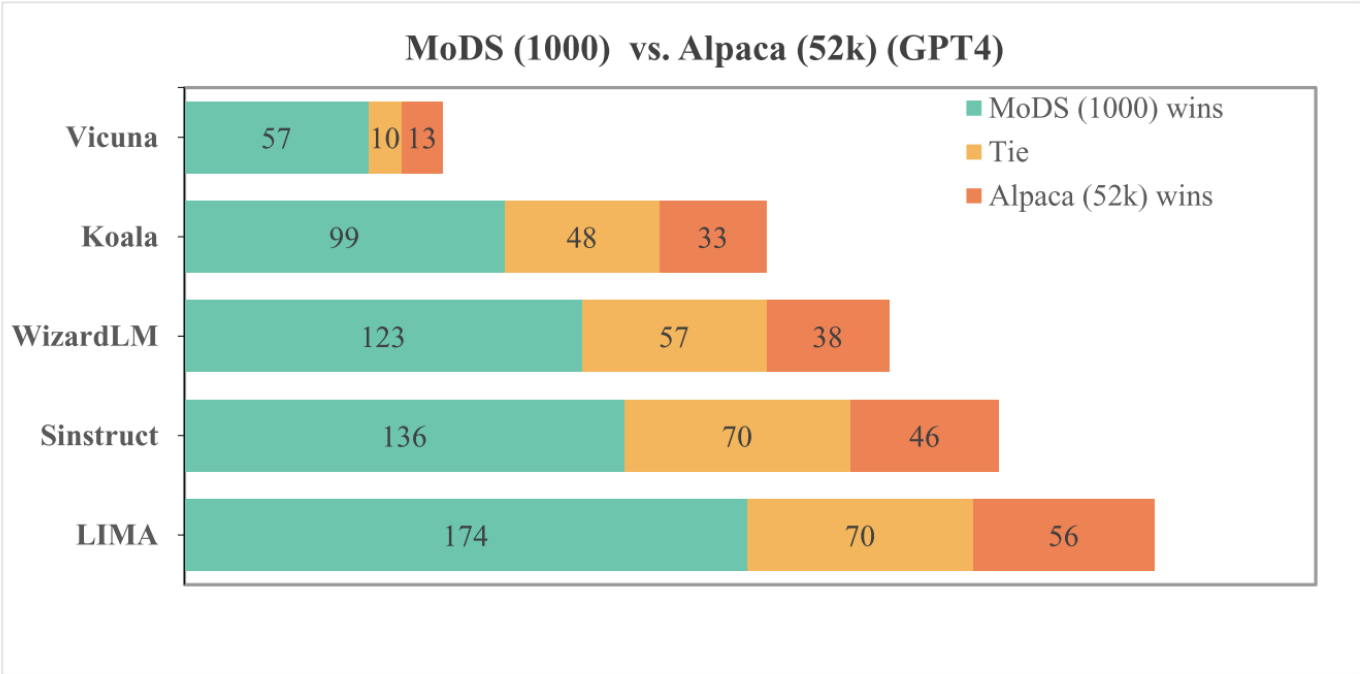
不同的大型语言模型在预训练过程中所学到的知识和具有的能力不同，因此在对不同的大型语言模型进行指令微调时，所需的指令数据也需要不同。

对于一条指令，如果给定的大型语言模型本身能够生成较好的回答，则说明给定的大型语言模型具有处理该指令或者这类指令的能力，反之亦然，并且哪些不能处理的指令对于模型微调来说更为重要。

- 使用SID数据集对模型进行一个初始训练
- 用训练好的初始模型对整个高质数据集-Data1中的指令进行结果预测
- 利用奖励模型对结果进行评分，当分值小于 β 时，说明初始模型不能对这些指令生成优质的回复，不具有处理这些类型指令的能力，获取必要性数据集-Data2
- 对Data2进行多样性筛选，获取增强指令数据集（Augmented Instruction Data）-AID。

最终利用种子指令数据集和增强指令数据集一起对模型进行指令微调，获得最终模型。

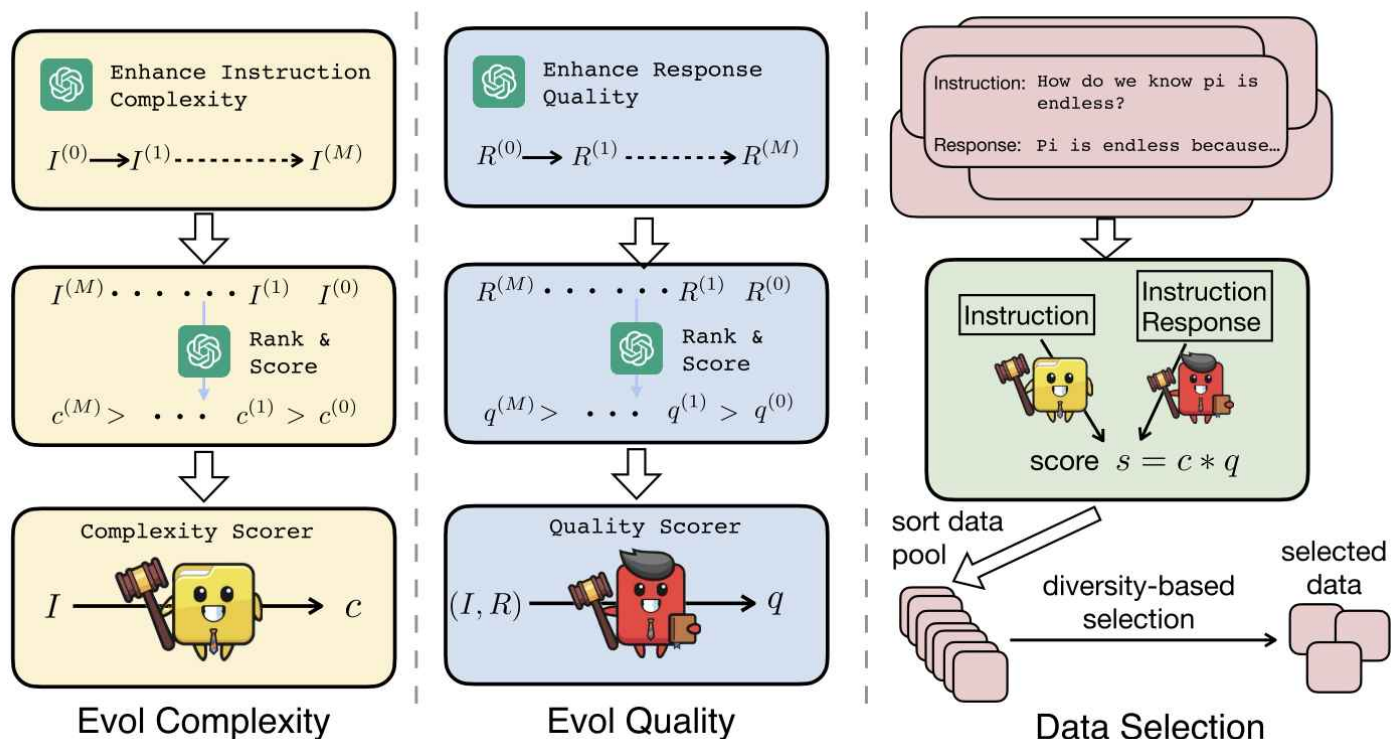
实验结果



	Test sets														
	Vicuna			Koala			WizardLM			Sinstruct			LIMA		
	win	tie	lose	win	tie	lose	win	tie	lose	win	tie	lose	win	tie	lose
Self-guided	60	7	13	112	44	24	127	57	34	122	82	48	197	48	55
MoDS(1000)	57	10	13	99	48	33	123	57	38	136	70	46	174	70	56
MoDS(2000)	67	3	10	116	40	24	132	57	29	134	77	41	195	58	47

《WHAT MAKES GOOD DATA FOR ALIGNMENT? A COMPREHENSIVE STUDY OF AUTOMATIC DATA SELECTION IN INSTRUCTION TUNING》

DEITA，核心观点是Score-First, Diversity-Aware Data Selection，即先对数据进行复杂性和质量评分，再通过多样性进行数据筛选，如下图所示



底层指令数据收集

为了对比大模型数据池选择，对大模型下游任务的影响，构建了2个具有不同属性的数据池。

- X_{sota} : 收集大模型的SOTA方法中的数据集，假设这些数据使用了相对复杂、多样和高质量的指令数据，涉及WizardLM (Alpaca)、WizardLM (ShareGPT)、UltraChat和ShareGPT (Vicuna)，共计300k数据。
- X_{base} : 收集Alpaca、Dolly、OAssit和FLAN 2022数据作为基础指令数据集，共计100k数据。

Data Pool	Dataset Source	Sample Size
X_{sota}	ShareGPT	58 K
	UltraChat	105 K
	WizardLM	143 K
X_{base}	Alpaca	52 K
	Dolly	15 K
	OAssit	10 K
	FLAN 2022	23 K

复杂性评分

当前对于复杂性评估的方法有：

- Random Selection：随机选择样本。
- Instruction Length：按照指令的长度计算复杂性。
- Perplexity：通过预训练模型计算回复的困惑度作为复杂性指标，困惑值越大意味着数据样本越难。
- Direct Scoring：利用ChaGPT给指令的复杂性打分。
- Instruction Node：利用ChatGPT将指令转换成语义树，通过树的节点数作为复杂性指标。
- Instag Complexity：利用ChatGPT对部分数据进行打标签，再训练一个Llama模型，再利用训练后的Llama模型对全量数据预测，标签越多说明数据约复杂。
- IFD：指令跟随难度作为复杂性指标。

DEITA评估复杂性的方法，主要先对一个小规模种子数据集（2k）进行数据复杂性扩展，再利用ChatGPT对扩展数据进行打分，并训练一个Llama-7B的模型，最后利用训练后的模型对数据的打分作为复杂性评估指标。

Case A		
Instruciton	Direct Scroing	Ranking & Scoring
Rate the given restaurant based on your experience. Restaurant: Pizzeria Garden	4	2
Using your experience as a guide, provide a comprehensive evaluation of Pizzeria Garden, taking into account factors such as the quality of the food, the level of service provided, the ambiance of the restaurant, and any other relevant aspects that may contribute to your overall impression. Please provide a detailed analysis that requires multiple steps of reasoning to arrive at a final rating.	8	3
Based on your extensive knowledge and expertise, please conduct a thorough and intricate assessment of Pizzeria Garden, encompassing various elements such as the excellence of the cuisine, the degree of attentiveness exhibited by the staff, the atmosphere and mood of the establishment, and any other pertinent factors that could influence your overall perception. Your evaluation should entail a meticulous analysis that necessitates multiple layers of reasoning to arrive at a conclusive rating.	8	4
In order to provide a thorough and nuanced evaluation of Pizzeria Garden, we require your extensive knowledge and specialized expertise. Your assessment should encompass a wide range of factors, including but not limited to the quality and inventiveness of the cuisine, the level of attentiveness and professionalism exhibited by the staff, the ambiance and overall atmosphere of the establishment, as well as any other pertinent elements that may influence your overall perception. Your analysis must be meticulous and multifaceted, necessitating a profound level of reasoning and critical thinking to arrive at a conclusive rating.	8	5
Case B		
Instruction	Direct Scroing	Ranking & Scoring
Create a birthday wish for someone who loves animals.	8	1
Craft a heartfelt birthday message for an individual who has a deep affection for a specific type of animal, such as dogs, cats, birds, or reptiles. Incorporate their passion for this creature into your well wishes to make their special day even more memorable and meaningful.	8	2
Compose a sincere and thoughtful birthday greeting for an individual who holds a profound fondness for a particular species of fauna, be it canines, felines, avians, or cold-blooded creatures. Infuse their ardor for this living being into your felicitations to elevate their celebratory occasion to a higher level of significance and remembrance.	7	4
Compose a sincere and introspective birthday greeting for an individual who harbors a profound and enduring fondness for a particular species of fauna, be it domestic canines, felines, avians, or reptiles. Integrate their ardor for this animate being into your felicitations to amplify the significance and recollection of their festive occasion. Furthermore, contemplate incorporating individual reminiscences or minutiae about the selected genus to further individualize the message and showcase your comprehension and admiration of their passions.	8	6

数据复杂性扩展，采用WizardLM模型的深度优化方法，通过添加约束、深化、具体化和增加推理步骤等技术来增强原始指令的复杂性，再让ChatGPT对原始指令和扩展后的指令一起评分，可以让ChatGPT发现不同复杂程度的指令之间的细微差别，指令之间具有可比性，不同复杂度指令之间的分数更加可靠，最终提升模型训练效果。

	Prompt Template
Direct Scoring	<p>We would like you to evaluate and rate the difficulty and complexity of the following question. You should give an overall score on a scale of 1 to 10, where a higher score indicates higher difficulty and complexity. You must just give a score without any other reasons.</p> <p>Question: <Instruction></p> <p>Score:</p>
Instruction Node	<p>You need to rewrite the following "instruction" to a TREE through Semantic Parsin in the natural language processing field and only count the total node number of the TREE. You must just give node number without any other reasons.</p> <p>Instruction: <Instruction></p> <p>Node number:</p>

	Prompt Templates
adding constraints	<p>I want you act as a Prompt Rewriter.</p> <p>Your objective is to rewrite a given prompt into a more complex version to make those famous AI systems (e.g., ChatGPT and GPT4) a bit harder to handle. But the rewritten prompt must be reasonable and must be understood and responded by humans.</p> <p>Your rewriting cannot omit the non-text parts such as the table and code in #Given Prompt#. Also, please do not omit the input in #Given Prompt#.</p> <p>You SHOULD complicate the given prompt using the following method:</p> <p>Please add one more constraints/requirements into #Given Prompt#</p> <p>You should try your best not to make the #Rewritten Prompt# become verbose, #Rewritten Prompt# can only add 10 to 20 words into #Given Prompt#.</p> <p>'#Given Prompt#', '#Rewritten Prompt#', 'given prompt' and 'rewritten prompt' are not allowed to appear in #Rewritten Prompt#</p> <p>#Given Prompt#:</p> <p><Here is instruction></p> <p>#Rewritten Prompt#:</p>
deepening	<p>I want you act as a Prompt Rewriter.</p> <p>Your objective is to rewrite a given prompt into a more complex version to make those famous AI systems (e.g., ChatGPT and GPT4) a bit harder to handle. But the rewritten prompt must be reasonable and must be understood and responded by humans.</p> <p>Your rewriting cannot omit the non-text parts such as the table and code in #Given Prompt#. Also, please do not omit the input in #Given Prompt#.</p> <p>You SHOULD complicate the given prompt using the following method:</p> <p>If #Given Prompt# contains inquiries about certain issues, the depth and breadth of the inquiry can be increased. or</p> <p>You should try your best not to make the #Rewritten Prompt# become verbose, #Rewritten Prompt# can only add 10 to 20 words into #Given Prompt#.</p> <p>'#Given Prompt#', '#Rewritten Prompt#', 'given prompt' and 'rewritten prompt' are not allowed to appear in #Rewritten Prompt#</p> <p>#Given Prompt#:</p> <p><Here is instruction></p> <p>#Rewritten Prompt#:</p>

质量评分

当前对于质量评估的方法有：

- Random Selection：随机选择样本。
- Response Length：采用输出长度作为质量评估指标。
- Direct Scoring：利用ChatGPT直接评估对特定指令输出结果的准确性。
- MoDs：用RM算质量

DEITA评估质量的方法，与评估复杂性方法一致。先对一个小规模种子数据集（2k，与复杂性数据一致）进行数据质量扩展，再利用ChatGPT对扩展数据进行打分并训练一个Llama1-7B的模型，最后利用训练后的模型对数据的打分作为质量评估指标。

Instruction A

Find the title of a book about a political movement in the 1940s.

Response	Direct Scoring	Ranking & Scoring
The title of a book about a political movement in the 1940s is "A Matter Of Resistance: The Struggle for Social Justice and Human Rights in India's 1940s".	4.5	2
One book that delves into the political movement of the 1940s is "A Matter Of Resistance: The Struggle for Social Justice and Human Rights in India's 1940s". This book provides a detailed account of the struggle for social justice and human rights during that time period. It sheds light on the challenges faced by the people and the efforts made to overcome them. The book is a valuable resource for anyone interested in understanding the political climate of the 1940s.	4.5	3
One book that delves into the political movement of the 1940s is "The Struggle for Social Justice and Human Rights in India's 1940s". This book offers a comprehensive account of the challenges faced by the people during that time period and the efforts made to overcome them. It provides valuable insights into the political climate of the 1940s, making it an excellent resource for anyone interested in gaining a deeper understanding of this era. The book is highly recommended for those who want to explore the political movement of the 1940s in greater depth.	4.5	4
The Struggle for Social Justice and Human Rights in India's 1940s" is a book that provides a detailed account of the political movement during that time period. It offers a comprehensive analysis of the challenges faced by the people and the efforts made to overcome them. The book provides valuable insights into the political climate of the 1940s, making it an excellent resource for anyone interested in gaining a deeper understanding of this era. It covers various aspects of the political movement, including the key players, their ideologies, and the impact of their actions. Overall, "The Struggle for Social Justice and Human Rights in India's 1940s" is a must-read for those who want to explore the political movement of the 1940s in greater depth.	4.5	5

Instruction B

Generate three new lines for an incomplete poem.

Response	Direct Scoring	Ranking & Scoring
That filled the morning air Spreading joy and cheer As far as one could see.	4.5	2
Here are three new lines to complete the poem: The flowers bloomed in the meadow, Their colors vibrant and bright, A scene of beauty, a true delight.	4.5	3
Here are three new lines to complete the poem: The birds chirped their melody Welcoming the day with glee Nature's beauty on full display.	4.5	4
The sun rose in the sky, Its rays spreading warmth and light, A new day dawning, full of promise and delight. The birds joined in with their sweet melody, A symphony of nature's harmony, Welcoming the day with joy and glee. And as the world awakens, we see, Nature's beauty on full display, A breathtaking sight, for all to see and say.	4.5	5

数据质量扩展，通过特殊的提示词利用ChatGPT对数据的回复部分进行改写，主要是增强回复的有用性、相关性、丰富深度、创造力和提供额外的细节描述。

	Prompt Templates
enhancing helpfulness	<p>I want you to act as a Response Rewriter</p> <p>Your goal is to enhance the quality of the response given by an AI assistant to the #Given Prompt# through rewriting.</p> <p>But the rewritten response must be reasonable and must be understood by humans.</p> <p>Your rewriting cannot omit the non-text parts such as the table and code in #Given Prompt# and #Given Response#. Also, please do not omit the input in #Given Prompt#.</p> <p>You Should enhance the quality of the response using the following method:</p> <p>Please make the Response more helpful to the user.</p> <p>You should try your best not to make the #Rewritten Response# become verbose, #Rewritten Response# can only add 10 to 20 words into #Given Response#.</p> <p>'#Given Response#', '#Rewritten Response#', 'given response' and 'rewritten response' are not allowed to appear in #Rewritten Response#</p> <p>#Given Prompt#:</p> <p>Give three tips for staying healthy.</p> <p>#Given Response#:</p> <p><Response></p> <p>#Rewritten Response#:</p>
augmenting relevance	<p>I want you to act as a Response Rewriter</p> <p>Your goal is to enhance the quality of the response given by an AI assistant to the #Given Prompt# through rewriting.</p> <p>But the rewritten response must be reasonable and must be understood by humans.</p> <p>Your rewriting cannot omit the non-text parts such as the table and code in #Given Prompt# and #Given Response#. Also, please do not omit the input in #Given Prompt#.</p> <p>You Should enhance the quality of the response using the following method:</p> <p>Please make the Response more relevant to #Given Prompt#.</p> <p>You should try your best not to make the #Rewritten Response# become verbose, #Rewritten Response# can only add 10 to 20 words into #Given Response#.</p> <p>'#Given Response#', '#Rewritten Response#', 'given response' and 'rewritten response' are not allowed to appear in #Rewritten Response#</p> <p>#Given Prompt#:</p> <p>Give three tips for staying healthy.</p> <p>#Given Response#:</p> <p><Response></p> <p>#Rewritten Response#:</p>
enriching depth	<p>I want you to act as a Response Rewriter</p> <p>Your goal is to enhance the quality of the response given by an AI assistant to the #Given Prompt# through rewriting.</p> <p>But the rewritten response must be reasonable and must be understood by humans.</p> <p>Your rewriting cannot omit the non-text parts such as the table and code in #Given Prompt# and #Given Response#. Also, please do not omit the input in #Given Prompt#.</p> <p>You Should enhance the quality of the response using the following method:</p> <p>Please make the Response more in-depth</p> <p>You should try your best not to make the #Rewritten Response# become verbose, #Rewritten Response# can only add 10 to 20 words into #Given Response#.</p> <p>'#Given Response#', '#Rewritten Response#', 'given response' and 'rewritten response' are not allowed to appear in #Rewritten Response#</p> <p>#Given Prompt#:</p> <p>Give three tips for staying healthy.</p> <p>#Given Response#:</p> <p><Response></p> <p>#Rewritten Response#:</p>

知乎 @刘聪NLP

多样性筛选

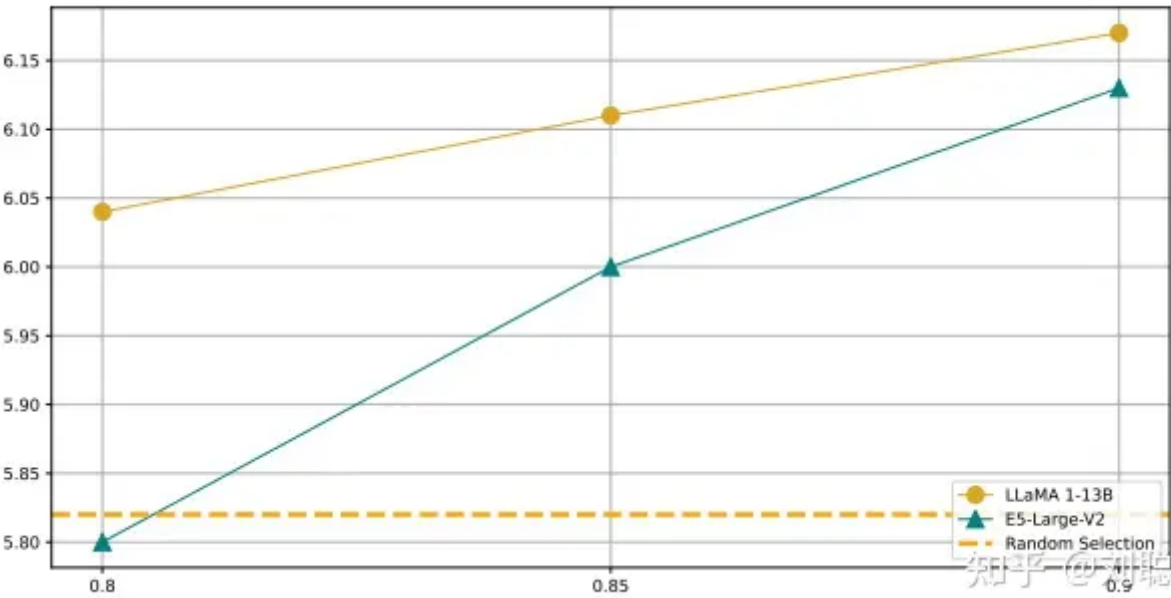
多样性筛选方法，首先将数据池中的数据按照复杂性和质量的综合得分（复杂性分数*质量分数）进行降序排序，然后按顺序逐个取出样本数据 x_i ，计算 x_i 与筛选池中相邻最近的样本之间距离值，其中，数据利用Llama-13B模型进行向量表征，距离计算采用余弦相似度。如果距离值小于 t 时，认为该样本与筛选池中数据相似程度不高，可以纳入筛选池；否则不纳入筛选池。当筛选池中样本数达到规定样本个数，完成多样性筛选。

Algorithm 1 Score-First, Diversity-Aware Data Selection

```
1: Input: The data pool  $X$ , data budget  $m$ 
2: Output: The selected subset  $S_{\pi_{\text{DEITA}}}^{(m)}$ 
3: Initialize Empty Dataset  $S_{\pi_{\text{DEITA}}}^{(m)}$ 
4: Sorting  $X$  with the combined complexity score and quality score  $s = q * c$ ;
5: Getting the sorted Pool  $X^*$ ;
6: for Each Sample  $x \in X^*$  do
7:   //  $d(x, S)$  denotes the distance between  $x$  and its nearest neighbor in  $S$ 
8:   if  $d(x, S_{\pi_{\text{DEITA}}}^{(m)}) < \tau$  then
9:      $S_{\pi_{\text{DEITA}}}^{(m)} \leftarrow S_{\pi_{\text{DEITA}}}^{(m)} \cup \{x\}$ 
10:  else
11:    Continue
12:  end if
13:   $X \leftarrow X \setminus \{x\}$ 
14:  if  $|S_{\pi_{\text{DEITA}}}^{(m)}|$  equals to  $m$  then
15:    Break
16:  end if
17: end for
```

知乎 @刘聪NLP

当 $t=0.9$ 时效果比较优异，如下图所示。



知乎 @刘聪NLP

实验结果

比较DEITA与其他数据筛选方法的效果，如下表所示。

Model	Data Size	MT-Bench	AlpacaEval(%)
Random	6K	5.84	73.91
Alpagasus (Pool=50K)	6K	5.61	71.21
LIMA	1K	4.29	41.98
TAGLM [†]	6K	6.09	72.80
DEITA-LLaMA1-13B _{6K}	6K	6.46	77.08

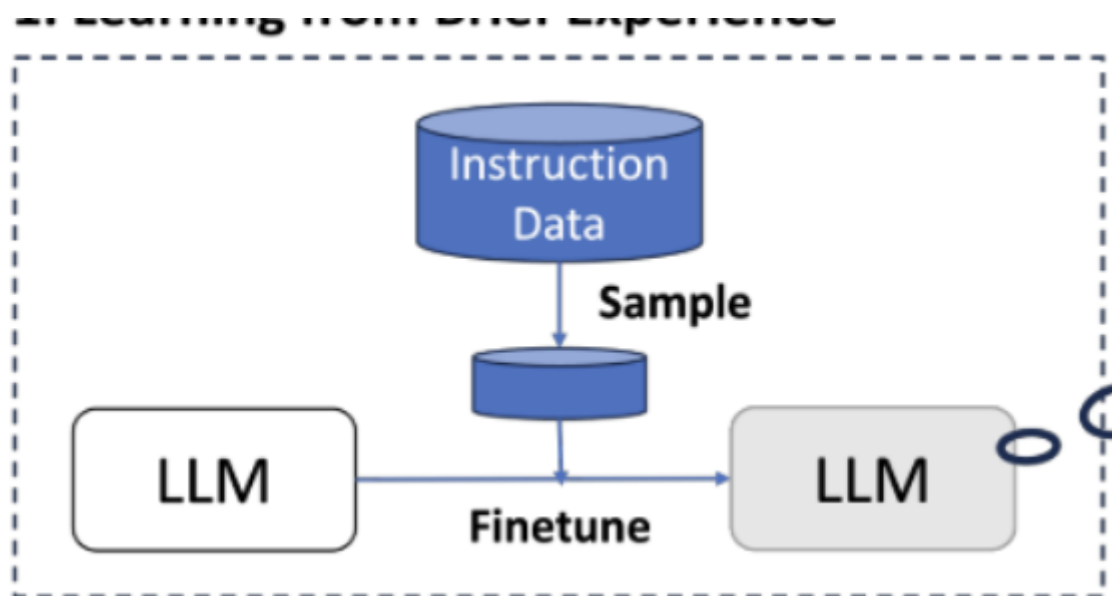
同时使用DEITA方法筛选6K和10K数据分别训练LLaMA1-13B、LLaMA2-13B和Mistral-7B模型，与随机数据筛选、其他SOTA的SFT模型以及对齐模型对比，取得了较为优异的成绩。

Model	Data Size / Alignment	MT-Bench	AlpacaEval(%)
Proprietary Models			
GPT-4	—	8.99	95.28
Claude-v2	—	8.06	91.36
gpt-3.5-turbo	—	7.90	89.37
Open-sourced Models based on LLaMA-1-13B			
Alpaca-13B	52K / SFT	4.53	—
WizardLM-13B	70K / SFT	6.35	75.31
Vicuna-13B-v1.3	125K / SFT	6.39	82.11
TAGLM-13B [†]	6K / SFT	6.09	72.80
Random-Select	10K / SFT	6.03	71.52
DEITA-LLaMA1-13B _{6K}	6K / SFT	6.46	77.08
DEITA-LLaMA1-13B _{10K}	10K / SFT	6.60	78.01
Open-sourced Models based on LLaMA-2-13B			
LLaMA2-13B-Chat	>100K / SFT + >1M / RLHF	6.65	81.09
Vicuna-13B-v1.5	125K / SFT	6.57	78.80
TÜLÜ 2 13B	326K / SFT	6.70	78.90
TÜLÜ 2 + DPO 13B	326K / SFT + 60K / DPO	7.00	89.50
Random-Select	10K / SFT	5.78	65.19
DEITA-LLaMA2-13B _{6K}	6K / SFT	6.65	80.75
DEITA-LLaMA2-13B _{10K}	10K / SFT	6.79	81.09
Open-sourced Models based on Mistral-7B			
Mistral-7B-Instruct-v0.1	—	6.84	69.65
Mistral-7B-Instruct-v0.2	—	7.60	93.65
zephyr-beta-sft*	200K / SFT	5.32	75.12
zephyr-beta	200K / SFT + 60K / DPO	7.34	90.60
Random-Select	10K / SFT	5.89	56.90
DEITA-Mistral-7B _{6K}	6K / SFT	7.22	80.78
DEITA-Mistral-7B _{10K}	10K / SFT	7.32	81.67
DEITA-Mistral-7B _{6K} + DPO	6K / SFT + 10K / DPO	7.55	90.05

Code: 《From Quantity to Quality: Boosting LLM Performance with Self-Guided Data Selection for Instruction Tuning》

github: https://github.com/tianyi-lab/Cherry_LLM

1. Learning from Brief Experience:



第一步sample: 用kmeans, 100个中心, 每个中心10条

要用kmeans, 就要得到所有指令的数字表示(sentence_embedding)
(output.hidden_states[-1]) :

代码: data_analysis.py

```
1 def get_perplexity_and_embedding_whole_text(tokenizer, model, text,
2       max_length):
3     input_ids = tokenizer.encode(text, return_tensors="pt", truncation=True,
4       max_length=max_length).to(device)
5     with torch.no_grad():
6         outputs = model(input_ids, labels=input_ids.contiguous())
7         loss = outputs.loss
8         perplexity = torch.exp(loss)
9
10    hidden_states = outputs.hidden_states
11    embeddings = hidden_states[-1]
12    sentence_embedding = embeddings.mean(dim=1)
13
14    return perplexity.to('cpu'), sentence_embedding.to('cpu')
```

第二步：聚类，kmeans，

代码：data_by_cluster.py

```
1 def do_clustering(args, high_dim_vectors):
2
3     clustering_algorithm = args.cluster_method
4     if clustering_algorithm == 'kmeans':
5         clustering = KMeans(n_clusters=args.kmeans_num_clusters,
6                             random_state=0).fit(high_dim_vectors)
7     return clustering
```

第三步：微调一轮

代码：

Cherry_LLM-main/training/stanford_alpaca/train.py

```
1
2     data_module = make_supervised_data_module(tokenizer=tokenizer,
3         data_args=data_args)
4     trainer = Trainer(model=model, tokenizer=tokenizer, args=training_args,
5         **data_module)
6     trainer.train()
7     trainer.save_state()
8     trainer.save_model(output_dir=training_args.output_dir)
```

2: Evaluating Based on Experience

第一步：分别算CAS和DAS：

代码：data_analysis.py

```
1     elif args.mod == 'cherry':
2         instruct_i_input_ids = tokenizer.encode(instruct_i,
3             return_tensors="pt", truncation=True, max_length=args.max_length).to(device)
4         instruct_i_len = instruct_i_input_ids.shape[1]
5         ppl_out_alone, _, loss_list_alone =
6             get_perplexity_and_embedding_part_text(tokenizer, model, direct_answer_text,
7                 output_i, args.max_length-instruct_i_len+4)
8         ppl_out_condition, _, loss_list_condition =
9             get_perplexity_and_embedding_part_text(tokenizer, model, whole_text, output_i,
```

```

    args.max_length)
7
8         temp_data_i['ppl'] = [0,ppl_out_alone,ppl_out_condition]
9         temp_data_i['token_loss'] =
    [[],loss_list_alone,loss_list_condition]

```

第二步：算IFD，筛选数据

代码：data_by_IFD.py，注意大于1的数据直接丢弃

```

1         mean_1 = loss_list_1.mean()
2         mean_2 = loss_list_2.mean()
3         mean_rate = mean_2/mean_1
4         if mean_rate > 1:
5             continue
6
7         mean_rate_list.append((mean_rate,i))
8         mean_list_1.append((mean_1,i))
9         mean_list_2.append((mean_2,i))

```

3: Retraining from Self-Guided Experience

应用筛选数据微调：

Cherry_LLM-main/training/stanford_alpaca/train.py

```

1
2     data_module = make_supervised_data_module(tokenizer=tokenizer,
    data_args=data_args)
3     trainer = Trainer(model=model, tokenizer=tokenizer, args=training_args,
    **data_module)
4     trainer.train()
5     trainer.save_state()
6     trainer.save_model(output_dir=training_args.output_dir)

```