



# 论文汇报

1. Summlt: Iterative Text Summarization via ChatGPT
2. CLUSTERLLM: Large Language Models as a Guide for Text Clustering

汇报人：陈鑫宇

时间：2023年5月31日





養天地正氣 法古今完人

# SummIt: Iterative Text Summarization via ChatGPT





1. 参考不完美的摘要优化
2. 语法错误，内容不正确
3. 可控性缺乏





## Refined Summary

You are a summarizer that follows the output pattern. You revise the summary based on the given instructions. You follow all the instructions without commenting on them.

**Refine:** [Revise Suggestions] Revise the summary. Follow all the suggestions and you are not make more comments. [Format Instructions]

## Summarizer

Knowledge  
Extractor

Topic  
Extractor

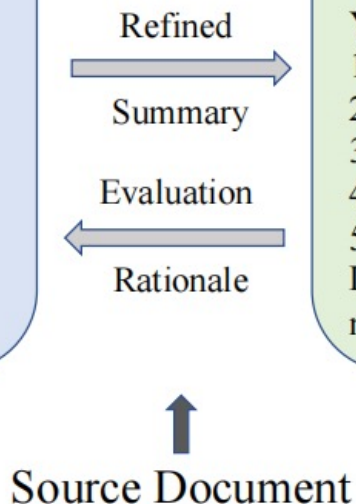
You are a summary evaluator that gives scores for the summaries with revise suggestions.

Your suggestions can be:

1. Add the information of <insert>
2. Remove the information of <insert>
3. Rephrase the information of <insert>
4. Shorten the summary
5. Keep the summary unchanged

If you think there's no further revision is needed, you must add "<STOP>" at the end.

## Evaluator







User

**Summarize the Document:**

Chelsea's on loan midfielder Oriol Romeu goes up against Sportsmail's Martin Keown as the pair predict the scores for the weekend's Premier League and FA Cup matches.....



ChatGPT  
Summarizer

**Summary:** Chelsea's on loan midfielder Oriol Romeu and Sportsmail's Martin Keown predict the scores for the weekend's Premier League and FA Cup matches, including the standout fixture between Chelsea and Manchester United.....



ChatGPT  
Evaluator

**Revision Comments:** The summary is too long and contains unnecessary details. It should be shortened to only include the main point, which is Oriol Romeu and Martin Keown predicting the scores for the weekend's Premier League and FA Cup matches.

**Evaluation Score:** [0.1, 0.6, 0.2, 0.1, 0]



**Summary:** Chelsea's on loan midfielder Oriol Romeu and Sportsmail's Martin Keown predict the scores for the weekend's Premier League and FA Cup matches.



**Revision Comments:** The summary is concise and includes the main point of the document. There's no need to further revise.<STOP>

**Evaluation Score:** [0.9, 0.1, 0, 0, 0]



Chelsea's on loan midfielder Oriol Romeu and Sportsmail's Martin Keown predict the scores for the weekend's Premier League and FA Cup matches.

Iterative Summarization  
(Run in Background)  
Iteration 1  
Iteration n





Model	CNN/DM				XSum			
	R1	R2	RL	GPT-Eval	R1	R2	RL	GPT-Eval
Zero-shot setting								
PEGASUS <sub>ZS</sub>	32.90	13.28	29.38	3.23	19.27	3.00	12.72	3.52
BART <sub>ZS</sub>	32.83	13.30	29.64	3.42	19.26	3.30	14.67	3.49
T5 <sub>ZS</sub>	<b>39.68</b>	<b>17.24</b>	26.28	3.47	19.66	2.91	15.31	3.55
ChatGPT	39.44	16.14	<b>29.83</b>	3.46	21.61	<b>5.98</b>	17.60	3.47
SummIt (ours)	36.50	13.49	26.76	<b>4.33</b>	<b>21.92</b>	5.93	<b>17.62</b>	<b>4.24</b>
Few-shot setting								
ChatGPT	<b>40.00</b>	<b>16.39</b>	<b>30.02</b>	3.57	<b>23.96</b>	<b>7.36</b>	<b>19.36</b>	3.57
SummIt (ours)	37.29	13.60	26.87	<b>4.35</b>	22.04	6.20	17.46	<b>4.32</b>





養天地正氣 法古今完人

	PEGASUS	BART	T5	ChatGPT	SummIt
CNN/DM	0.00	0.04	0.10	0.34	<b>0.52</b>
XSum	0.00	0.30	0.08	0.24	<b>0.38</b>





	XSum				
	R1	R2	RL	GPT-Eval	FactCC
ChatGPT-Init	21.61	5.98	17.60	3.47	28.00
ChatGPT-Final	21.92	5.93	17.62	4.24	36.00
OpenIE-Init	22.01	5.11	17.06	3.85	51.68
OpenIE-Final	19.72	3.85	15.36	4.95	47.24





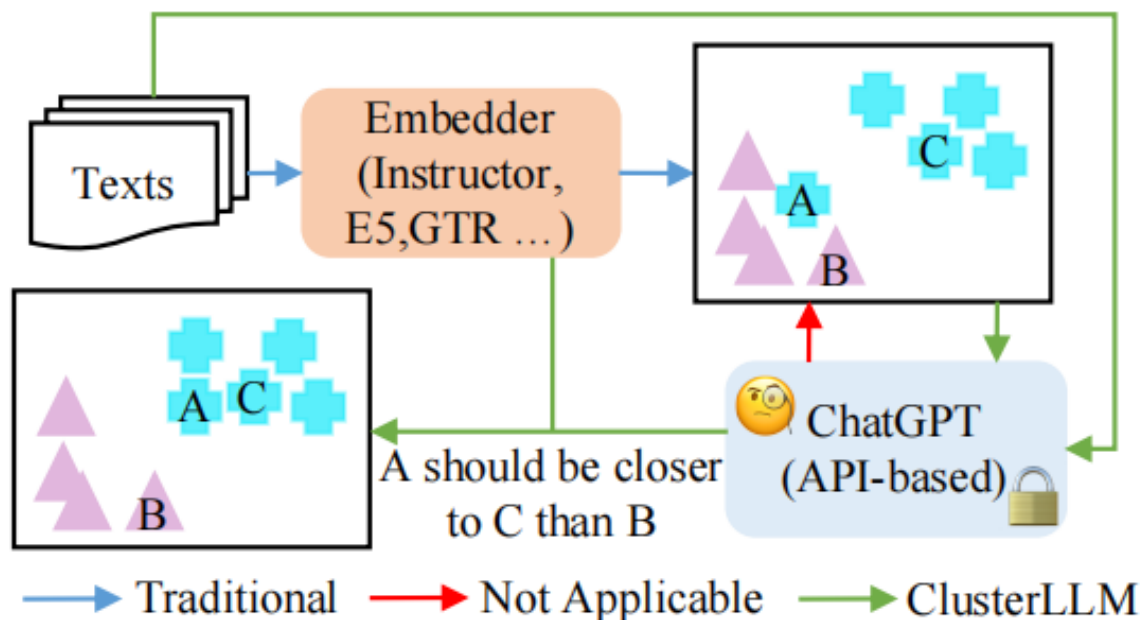




養天地正氣 法古今完人





---

# CLUSTERLLM: Large Language Models as a Guide for Text Clustering

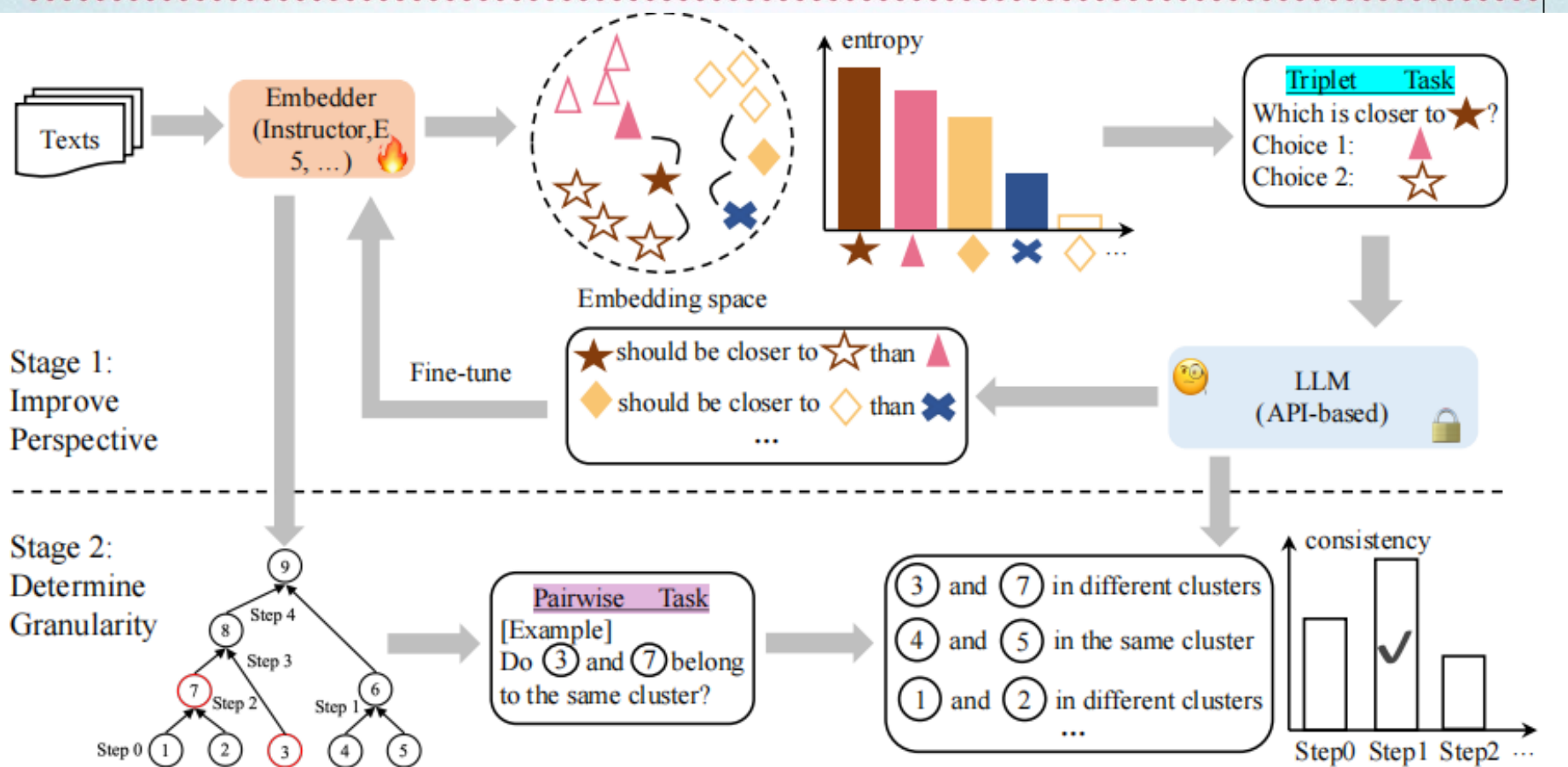




“Is  the same kind of shape as ?”,

is unknown. However, when asking “Which shape looks more like  ?  or ?”, it is much easier to pick up  . We query LLMs with a collection







lab.	Method	Intent Discovery		Information Extraction			Topic Mining		Emotion	Avg	All
		Bank77	CLINC(I)	FewRel	FewNerd	FewEvent	StackEx	ArxivS2S	GoEmo		
✓	Contrast (Vaze et al., 2022) <sup>†</sup>	31.56 <sub>(2.80)</sub>	34.68 <sub>(2.10)</sub>	39.10 <sub>(1.32)</sub>	30.42 <sub>(0.80)</sub>	47.61 <sub>(2.48)</sub>	16.79 <sub>(1.10)</sub>	14.99 <sub>(0.32)</sub>	16.34 <sub>(1.25)</sub>	28.94	-
	DAC (Zhang et al., 2021b)	37.31 <sub>(4.72)</sub>	61.84 <sub>(1.39)</sub>	43.24 <sub>(3.11)</sub>	40.84 <sub>(3.75)</sub>	33.22 <sub>(3.56)</sub>	19.53 <sub>(1.51)</sub>	16.33 <sub>(0.65)</sub>	18.88 <sub>(1.81)</sub>	33.90	-
	DPN (An et al., 2022)	31.02 <sub>(4.59)</sub>	45.56 <sub>(9.57)</sub>	33.86 <sub>(6.71)</sub>	38.13 <sub>(5.79)</sub>	43.43 <sub>(3.02)</sub>	13.31 <sub>(1.58)</sub>	12.66 <sub>(1.64)</sub>	14.86 <sub>(1.90)</sub>	29.10	-
	CLNN (Zhang et al., 2022)	47.84 <sub>(3.84)</sub>	75.64 <sub>(.89)</sub>	48.39 <sub>(3.95)</sub>	40.59 <sub>(3.75)</sub>	28.05 <sub>(1.93)</sub>	28.36 <sub>(2.03)</sub>	18.66 <sub>(1.70)</sub>	20.01 <sub>(1.35)</sub>	38.44	-
✗	E5 (Wang et al., 2022)	59.60 <sub>(0.91)</sub>	75.83 <sub>(0.79)</sub>	39.62 <sub>(1.22)</sub>	25.49 <sub>(1.22)</sub>	37.30 <sub>(1.97)</sub>	37.31 <sub>(0.95)</sub>	30.85 <sub>(0.37)</sub>	22.13 <sub>(1.04)</sub>	41.02	47.70
	CLUSTERLLM-E	69.06 <sub>(1.99)</sub>	79.51 <sub>(1.10)</sub>	47.53 <sub>(1.00)</sub>	28.52 <sub>(0.63)</sub>	42.17 <sub>(1.24)</sub>	43.01 <sub>(1.58)</sub>	<b>34.93</b> <sub>(0.36)</sub>	22.69 <sub>(0.41)</sub>	45.93	50.92
	Instructor (Su et al., 2022)	64.49 <sub>(1.52)</sub>	79.29 <sub>(1.03)</sub>	41.23 <sub>(0.60)</sub>	30.02 <sub>(1.24)</sub>	41.99 <sub>(2.04)</sub>	44.81 <sub>(0.94)</sub>	24.31 <sub>(0.77)</sub>	25.19 <sub>(0.98)</sub>	43.92	49.90
	SCCL-I (Zhang et al., 2021a)	65.48 <sub>(1.36)</sub>	80.85 <sub>(0.74)</sub>	41.15 <sub>(1.51)</sub>	31.09 <sub>(0.87)</sub>	39.97 <sub>(0.52)</sub>	45.11 <sub>(0.93)</sub>	25.63 <sub>(0.53)</sub>	<b>34.33</b> <sub>(0.86)</sub>	45.45	50.73
	CLUSTERLLM-I	70.77 <sub>(0.49)</sub>	82.77 <sub>(1.20)</sub>	47.94 <sub>(1.37)</sub>	34.75 <sub>(1.58)</sub>	46.17 <sub>(2.07)</sub>	47.21 <sub>(1.07)</sub>	26.61 <sub>(0.48)</sub>	27.49 <sub>(1.25)</sub>	47.96	53.09
	CLUSTERLLM-I-iter	<b>71.20</b> <sub>(1.59)</sub>	<b>83.80</b> <sub>(0.41)</sub>	<b>51.22</b> <sub>(1.43)</sub>	<b>40.60</b> <sub>(0.77)</sub>	<b>50.60</b> <sub>(0.79)</sub>	<b>47.75</b> <sub>(1.24)</sub>	26.34 <sub>(0.38)</sub>	26.75 <sub>(1.76)</sub>	<b>49.78</b>	<b>53.99</b>







Type	Model	Bank77	CLINC(I)	FewRel	FewNerd	FewEvent	StackEx	ArxivS2S	GoEmo
Random	#GT Triplets	23	6	41	156	105	14	22	117
	Instructor	100	100	80.49	71.15	98.10	85.71	95.45	68.38
	GPT3.5	100	100	85.37	82.05	94.29	71.43	81.82	68.38
	$\Delta$	(+0)	(+0)	(+4.88)	(+10.90)	(-3.81)	(-14.28)	(-13.63)	(+0)
Entropy-based	#GT Triplets	510	462	266	347	259	271	145	206
	Instructor	64.12	76.19	62.41	59.65	70.66	68.27	59.31	64.08
	GPT3.5 †	76.67	79.44	76.69	68.88	83.78	71.22	73.79	64.56
	$\Delta$	(+12.55)	(+3.25)	(+14.28)	(+9.23)	(+13.12)	(+2.95)	(+14.48)	(+0.48)
	GPT4	79.41	80.74	87.22	82.13	85.71	79.70	77.93	61.65
	$\Delta$	(+15.29)	(+4.55)	(+24.81)	(+22.48)	(+15.05)	(+11.43)	(+18.62)	(-2.43)





Method	Intent Discovery		Information Extraction			Topic Mining		Emotion	Avg	All
	Bank77	CLINC(I)	FewRel	FewNerd	FewEvent	StackEx	ArxivS2S	GoEmo		
Instructor	64.49 <sub>(1.52)</sub>	79.29 <sub>(1.03)</sub>	41.23 <sub>(0.60)</sub>	30.02 <sub>(1.24)</sub>	41.99 <sub>(2.04)</sub>	44.81 <sub>(0.94)</sub>	24.31 <sub>(0.77)</sub>	25.19 <sub>(0.98)</sub>	43.92	49.90
+self-supervise	68.18 <sub>(.73)</sub>	80.82 <sub>(0.75)</sub>	41.72 <sub>(0.47)</sub>	31.39 <sub>(0.74)</sub>	43.94 <sub>(1.15)</sub>	46.15 <sub>(1.17)</sub>	25.65 <sub>(0.37)</sub>	24.11 <sub>(2.02)</sub>	45.25	51.39
+CLUSTERLLM-random	59.88 <sub>(2.56)</sub>	74.40 <sub>(0.91)</sub>	40.65 <sub>(0.89)</sub>	27.15 <sub>(0.53)</sub>	44.23 <sub>(1.72)</sub>	40.81 <sub>(0.94)</sub>	22.03 <sub>(0.28)</sub>	28.62 <sub>(1.95)</sub>	42.22	48.27
+CLUSTERLLM-GPT3.5	70.77 <sub>(0.49)</sub>	82.77 <sub>(1.20)</sub>	47.94 <sub>(1.37)</sub>	34.75 <sub>(1.58)</sub>	46.17 <sub>(2.07)</sub>	47.21 <sub>(1.07)</sub>	26.61 <sub>(0.48)</sub>	27.49 <sub>(1.25)</sub>	47.96	53.09
+CLUSTERLLM-GPT4	69.71 <sub>(1.13)</sub>	81.91 <sub>(1.20)</sub>	<b>48.96</b> <sub>(1.14)</sub>	<b>37.54</b> <sub>(0.54)</sub>	47.98 <sub>(1.45)</sub>	46.82 <sub>(0.78)</sub>	26.16 <sub>(0.22)</sub>	27.41 <sub>(1.13)</sub>	48.31	53.22
+CLUSTERLLM-GT&GPT3.5	<b>71.35</b> <sub>(1.97)</sub>	<b>84.00</b> <sub>(1.04)</sub>	48.91 <sub>(1.20)</sub>	37.27 <sub>(0.61)</sub>	<b>48.12</b> <sub>(1.52)</sub>	<b>47.55</b> <sub>(1.17)</sub>	26.14 <sub>(0.57)</sub>	<b>30.91</b> <sub>(1.16)</sub>	<b>49.28</b>	<b>53.96</b>

