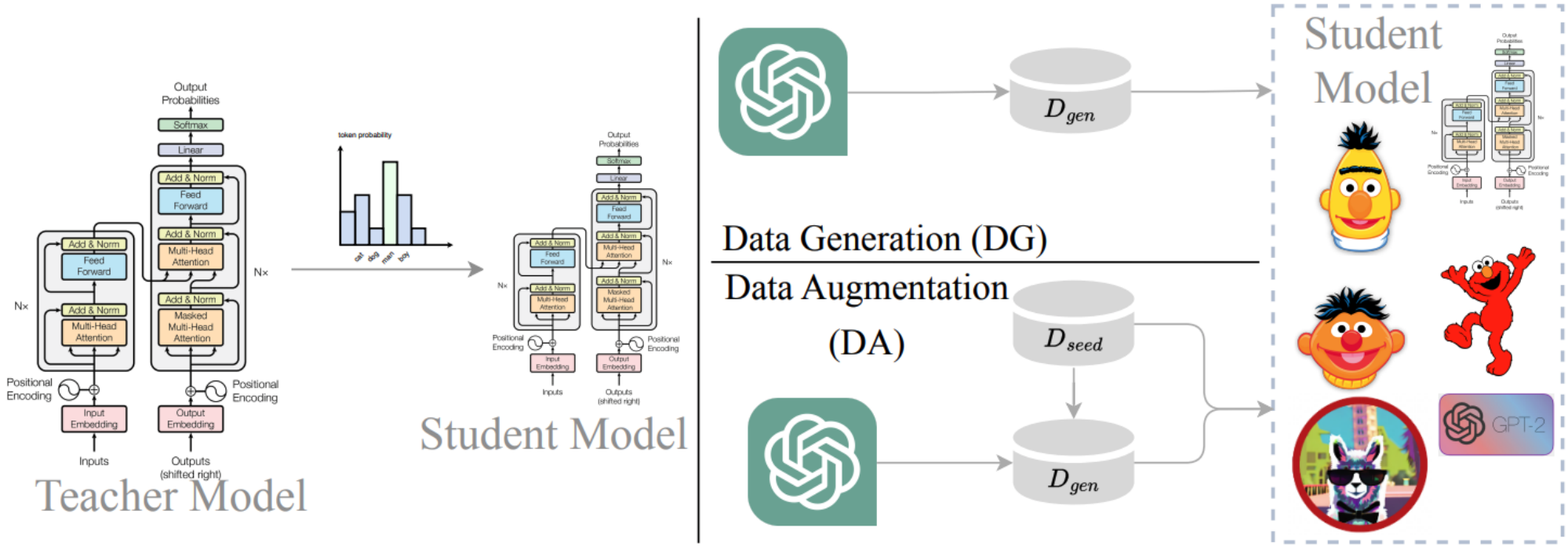
Evolving Knowledge Distillation with Large Language Models and Active Learning

Chengyuan Liu[1,2,†] , **Yangyang Kang**[2] , **Fubang Zhao**[2] , **Kun Kuang**[1,*] ,
Zhuoren Jiang[3,*] , **Changlong Sun**[2] , **Fei Wu**[1,4]
liucy1@zju.edu.cn, {yangyang.kangyy, fubang.zfb}@alibaba-inc.com, kunkuang@zju.edu.cn
jiangzhuoren@zju.edu.cn, changlong.scl@taobao.com, wufei@zju.edu.cn
[1]College of Computer Science and Technology, Zhejiang University
[2]Institute for Intelligent Computing, Alibaba Group
[3]School of Public Affairs, Zhejiang University
[4]Shanghai Institute for Advanced Study of Zhejiang University

Speaker: Chengyuan Liu

Introduce to Knowledge Distillation



Motivation

Under-utilization

- previous studies have regarded LLMs as mere text generators and sentence rewriters, solely relying on their capabilities for text generation and labeling. However, they have neglected the knowledge embedded in the downstream task and the powerful comprehensive ability of LLMs, which may resulting in a hindrance to the quality of the generated text.

Inflexibility

- prior KD studies have primarily been conducted in an offline and static manner. They construct the entire training data in one go, without considering any dynamic changes that may arise in the status and weaknesses of the student model. Consequently, the generated data often lacks specificity and diversity, limiting its effectiveness in improving the performance of the student model.

Contribution

Evolving Knowledge Distillation

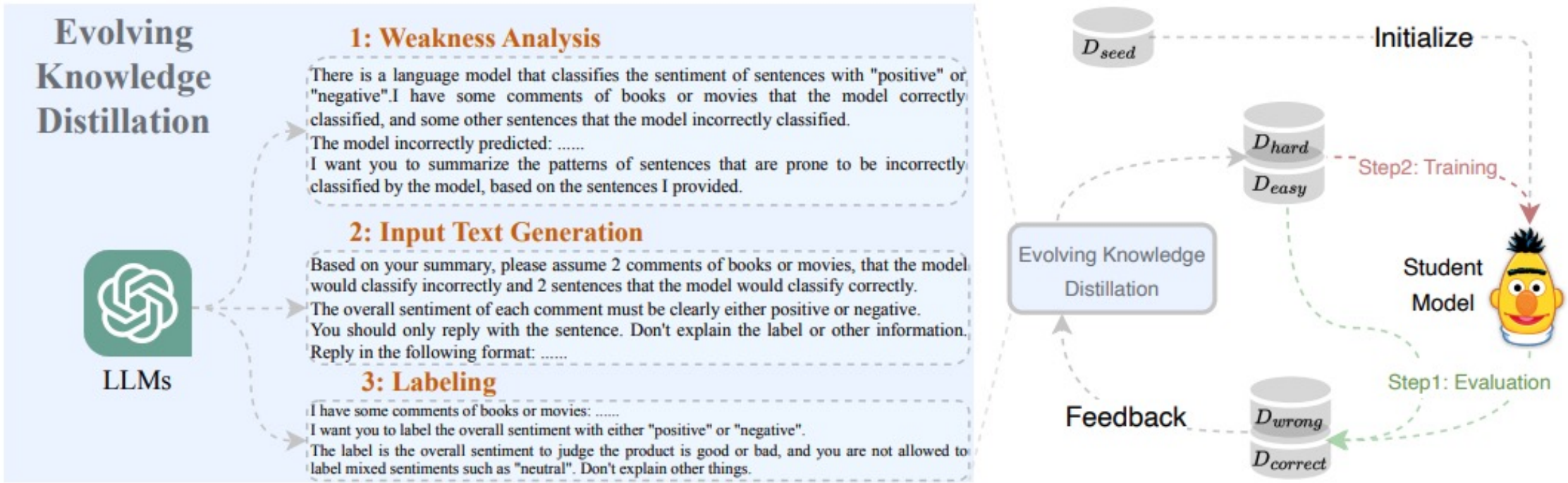
- We introduce the concept of Evolving Knowledge Distillation, which uses dynamically teaching strategy to distill the knowledge about learning the task, understanding the input texts, labeling and evaluating the predictions of student model.

EvoKD

- A novel approach called EvoKD is proposed in this paper incorporating the evolving KD and Active Learning, which leverages LLM's potential to comprehend the target task and acquire valuable knowledge.

Experiments Verification

- Experiments on text classification and NER tasks are conducted under few-shot settings comparing EvoKD with other baselines. EvoKD significantly outperformed all baseline approaches. Notably, EvoKD achieved up to 90% of the full-shot text classification performance with only 1-shot

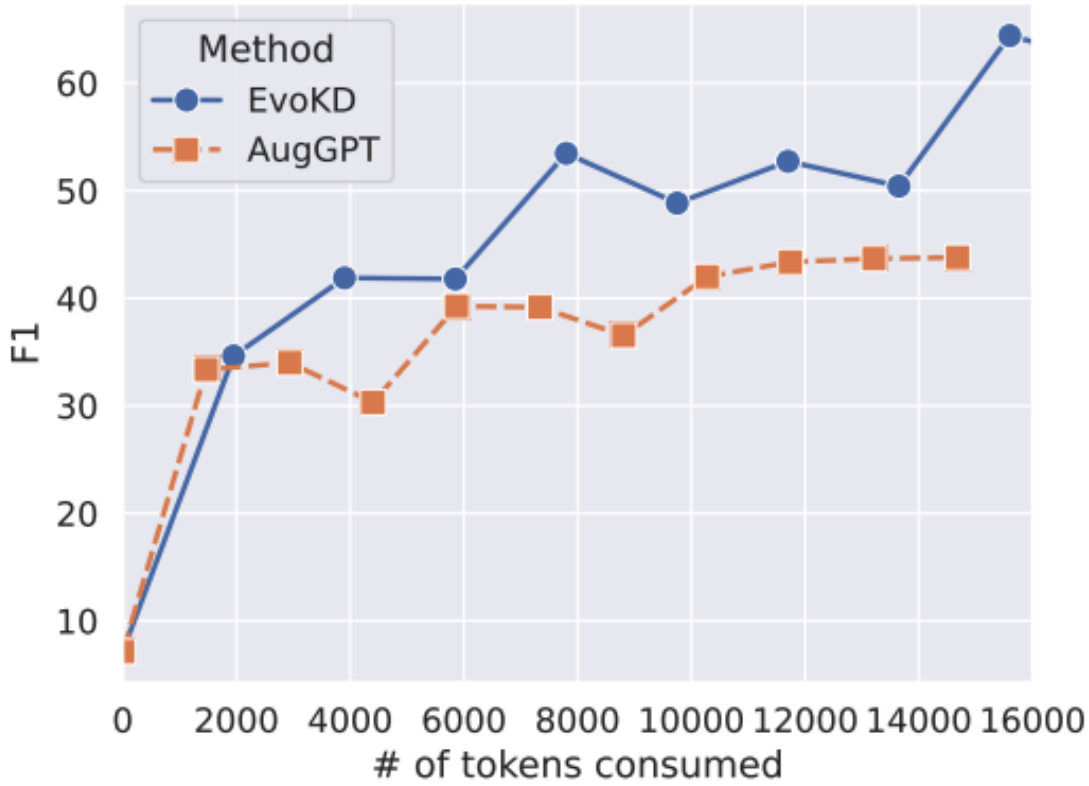

Evolving Knowledge Distillation
1: Weakness Analysis
There is a language model that classifies the sentiment of sentences with "positive" or "negative".I have some comments of books or movies that the model correctly classified, and some other sentences that the model incorrectly classified.
The model incorrectly predicted: .....
I want you to summarize the patterns of sentences that are prone to be incorrectly classified by the model, based on the sentences I provided.
2: Input Text Generation
Based on your summary, please assume 2 comments of books or movies, that the model would classify incorrectly and 2 sentences that the model would classify correctly.
The overall sentiment of each comment must be clearly either positive or negative.
You should only reply with the sentence. Don't explain the label or other information.
Reply in the following format: .....
3: Labeling
I have some comments of books or movies: .....
I want you to label the overall sentiment with either "positive" or "negative".
The label is the overall sentiment to judge the product is good or bad, and you are not allowed to label mixed sentiments such as "neutral". Don't explain other things.
LLMs
D_{seed}
Initialize
D_{hard}
D_{easy}
Step2: Training
Evolving Knowledge Distillation
Student Model
Step1: Evaluation
Feedback
D_{wrong}
$D_{correct}$

Method	English			Chinese		AVG
	Amazon	IMDB	Inshorts	TouTiao	CAIL2019	
Full Shot	0.9480	0.9495	0.9705	0.8495	0.9683	0.9372
No Augment	0.6030 ± 0.0880	0.5833 ± 0.0853	0.6408 ± 0.1528	0.3638 ± 0.0737	0.4422 ± 0.0604	0.5266
EDA	0.6314 ± 0.0838	0.6189 ± 0.0599	0.6776 ± 0.1669	0.3848 ± 0.0843	0.6120 ± 0.0319	0.5849
ZeroGen	0.7054 ± 0.1134	0.5087 ± 0.2204	0.8334 ± 0.0429	0.6442 ± 0.0369	0.7620 ± 0.1118	0.6907
SunGen	0.6257 ± 0.1288	0.5769 ± 0.0521	0.8103 ± 0.0456	0.2533 ± 0.0827	0.8305 ± 0.1020	0.6193
Gradual	0.5826 ± 0.0771	0.6857 ± 0.0109	0.7608 ± 0.0144	-	-	-
AugGPT	0.6234 ± 0.1712	0.6903 ± 0.0788	0.7902± 0.0759	0.6514 ± 0.0459	0.7122 ± 0.1117	0.6935
EvoKD	**0.8425 ± 0.0317**	0.7982 ± 0.0565	0.8516 ± 0.0257	0.6874 ± 0.0199	**0.9148 ± 0.0411**	0.8189
+Init	0.8403 ± 0.0240	**0.8359 ± 0.0272**	**0.8688 ± 0.0167**	**0.7112 ± 0.0237**	0.9137 ± 0.0355	**0.8340**

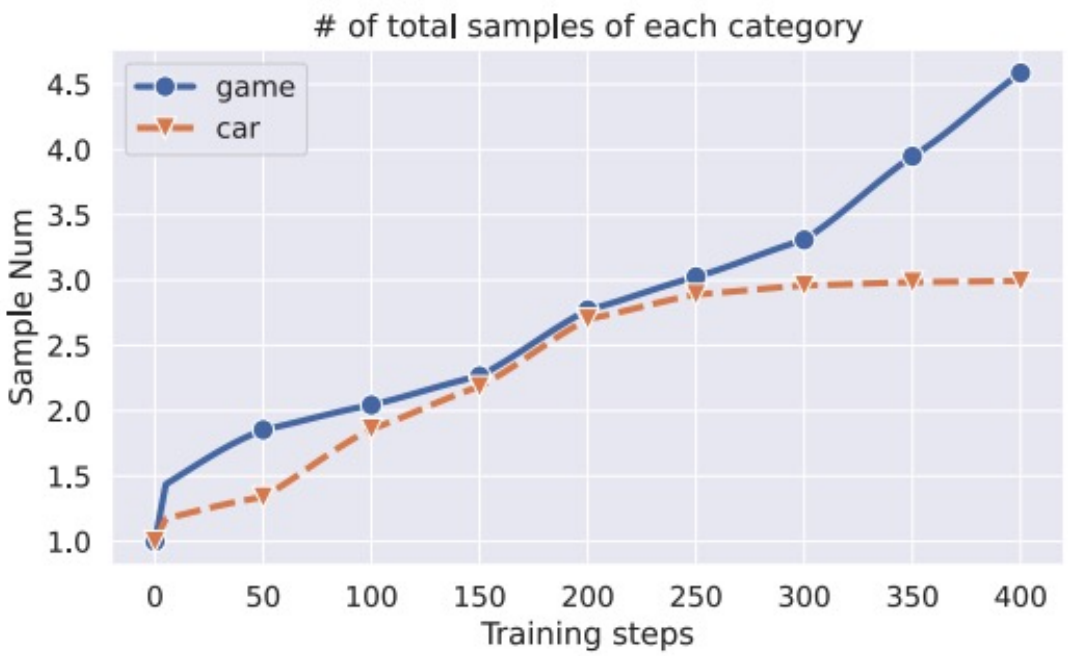
Experiment results under 1-shot text classification.

Method	CoNLL03	CoNLL04	AVG
Full Shot	0.9322	0.8766	0.9044
No Augment	0.3143	0.4929	0.4036
EDA	0.3062	0.5058	0.4060
AugGPT	0.6315	0.6683	0.6499
EvoKD	0.6538	**0.6848**	**0.6693**
+Init	**0.6629**	0.6628	0.6629

Experiment results under 1-shot NER.



F1 versus the number of tokens used during training.



EvoKD concentrates on the samples with lower performance.

THANKS FOR WATCHING!

Speaker: Chengyuan Liu