Quant-GPT

Authors: Yuzhe Yang, Kangqi Yu, Junquan Peng

Student IDs: 121090684, 121090735, 120090556

Email: {yuzheyang, kangqiyu, junquanpeng}@link.cuhk.edu.cn

Introduction

In this paper, we present Quant-GPT, a Large Language Model (LLM) for A-share market investment.

The core recipe of Quant-GPT is to leverage both distilled sentiment analysis data from **ChatGPT** and real-world announcements from the **A shares market** in the supervised fine-tuning stage. This is not only because purely using **ChatGPT**-distilled data might cause "model collapse" and the weak causality between sentiment and expected return, but also because real-world data from **the A shares market** reflects the common expectation of all the investors.

To synergize the strengths of finance news, we introduce RAG (Retrieval-Augmented Generation) where a searching tool is designed to retrieve related news of company announcements, assisting Quant-GPT make more accurate judgments on the expected return of the announcement.

Experimental results (backtest metrics including annualized return, max drawdown, sharpe ratio) demonstrate that Quant-GPT achieves state-of-the-art results in investment decisions among open-source LLMs. It is worth noting that by using additional real-world data and RAG, the distilled language model (i.e., Quant-GPT) outperforms its teacher model (i.e., ChatGPT) in most cases. See our demo by scanning the upper QR code.

Motivation

2023-03-31 20:09:46
三一重工股份有限公司
2022年社会责任报告
……

主要内容
1. 公司介绍
2. 可持续发展管理
3. 诚信经营
4. 科技创新
5. 绿色运营
6. 员工关怀
7. 社会责任
8. 环境绩效
9. 利益相关方沟通

……

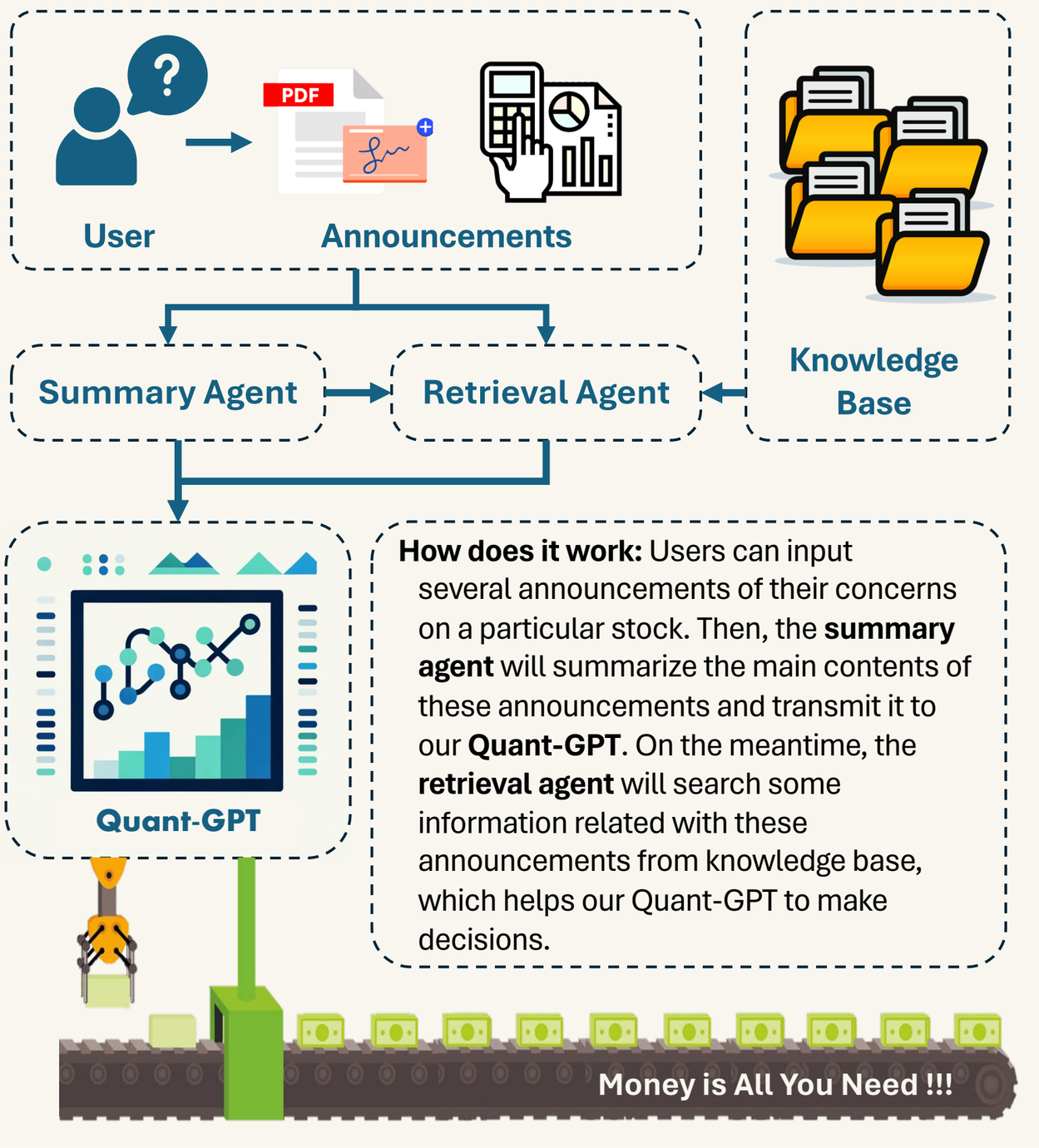
- 2023-03-09 04:16:39 新闻:【引活水、增活力、促转型，创业板助力实体经济高质量发展】
- 2023-03-14 05:22:00 新闻:【69家上市公司公布2022年度分红预案，拟合计派现290.37亿元】
- 2023-01-13 20:35:01 新闻:【深交所：加强退市风险公司2022年年度报告信息披露工作】
- 2023-01-12 05:45:26 新闻:【A股公司ESG风险频发 环境信息披露水平亟待提升】
- 2023-01-12 02:41:20 新闻:【A股公司ESG风险频发，环境信息披露水平亟待提升】

The RAG system is an agent that summarizes company announcements using a summary LLM trained on augment data from GPT. It is activated when a user inputs an announcement, focusing on key aspects such as the announcement time, company introduction, and relevant quantitative values.

The system has embedded over 90,000 news articles using dense and lexical embeddings. **Dense embeddings** capture the overall semantics of the text by converting it into vectors, while **lexical embeddings** enhance the extraction of key information by assessing the importance of keywords in the text.

The top five news articles within a user-defined time window are retrieved through a hybrid similarity measure. Users can tailor the time window duration and news types in the UI, enabling the model to perform precise sentiment analysis of announcements.

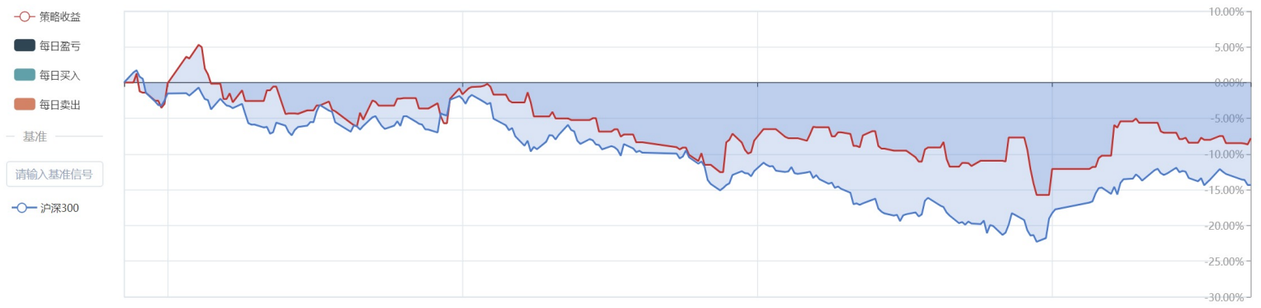
Methodology



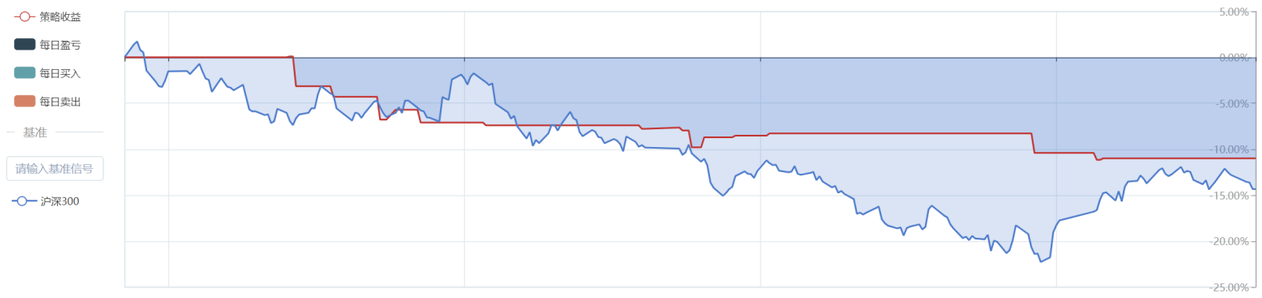
How does it work: Users can input several announcements of their concerns on a particular stock. Then, the **summary agent** will summarize the main contents of these announcements and transmit it to our **Quant-GPT**. On the meantime, the **retrieval agent** will search some information related with these announcements from knowledge base, which helps our Quant-GPT to make decisions.

Results

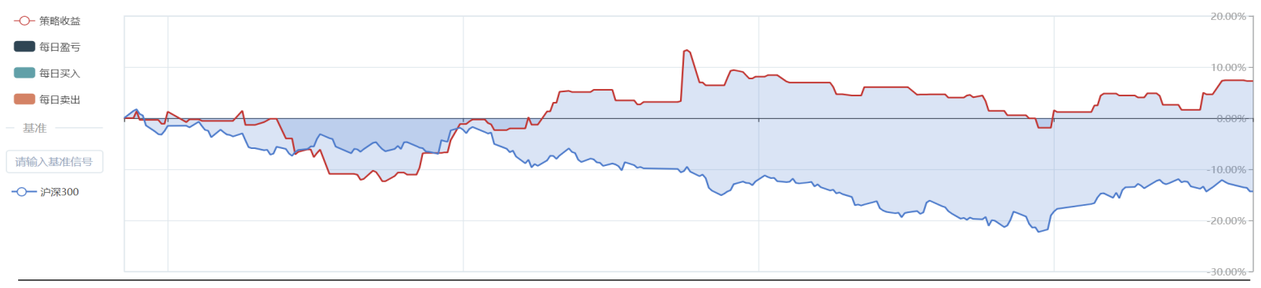
Trade with Zero-Shot Instructions (gpt3.5-turbo)



Trade with Zero-Shot Instructions (Qwen-7B)



Trade with Quant-GPT Instructions



Metrics	gpt-3.5-turbo	Qwen-7B	Quant-GPT
Annualized Return	-7.81%	-11.02%	7.26%
Max Drawdown	19.96%	11.26%	13.61%
Sharpe Ratio	-0.49	-2.02	0.40