Distilled ChatGPT Topic & Sentiment Modeling with Applications in Finance

Olivier Gandouet, Mouloud Belbahri, Armelle Jezequel, Yuriy Bodjov

TD Asset Management, Layer 6 AI

Edge Intelligence Workshop at AAAI 2024, February 26, 2024 | Vancouver, Canada

Abstract

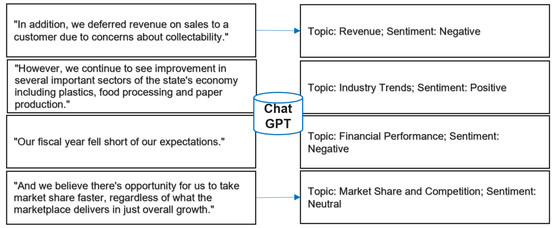
In this study, ChatGPT is utilized to create streamlined models that generate easily interpretable features. These features are then used to evaluate financial outcomes from earnings calls. We detail a training approach that merges knowledge distillation and transfer learning, resulting in lightweight topic and sentiment classification models without significant loss in accuracy. These models are assessed through a dataset annotated by experts. The paper also delves into two practical case studies, highlighting how the generated features can be effectively utilized in quantitative investing scenarios.

Introduction

The objective of this work is to distil ChatGPT to construct topic/sentiment classification models based off earning calls transcripts. The three steps are:

1. Identify a comprehensive list of topics that adequately represent a significant portion of the subject matter in the field.
2. Create a labeled dataset of sentences from the corpus based on the teacher topic/sentiment model.
3. Train a "small" topic/sentiment model using a supervised approach.

Below are examples of sentences labeled by ChatGPT.



Benchmark Datasets: We collaborated with three financial experts who tagged the topic and sentiment of 1,000 sentences. Each sentence was carefully reviewed to ensure the accuracy of the assigned sentiment, and a secondary round of tagging was conducted for instances where there were discrepancies in the initial tagging. For the topic models specifically, we also set aside a separate set of data tagged by ChatGPT, which accounted for 20% of our scored sample. This reserved dataset allowed us to assess the effectiveness of our supervised topic modeling approach.

Distillation Pipeline

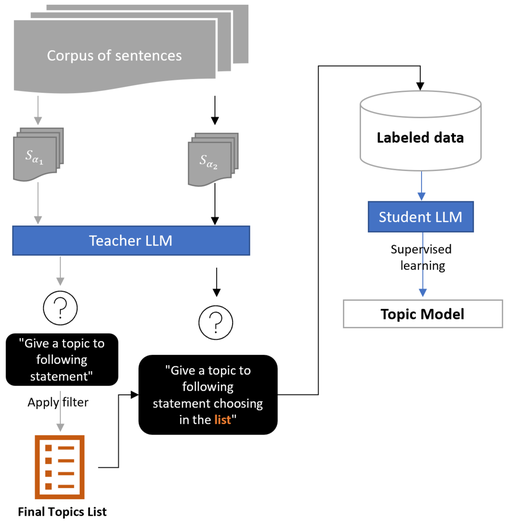


Figure 1:Earning Calls Topic Classification Pipeline.

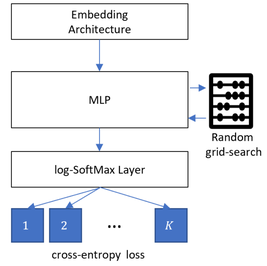


Figure 2:Topic Classification Student Model Architecture.

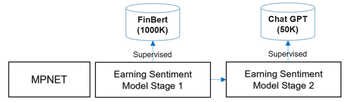


Figure 3:Sentiment Classification Model Pipeline.

Results

Category	Detected Topic	Proportion	Average Sentiment
Financial Performance Metrics	Revenue	8.9%	0.49
	Earnings Per Share (EPS)	2.4%	0.59
	Gross Margin	1.1%	-0.03
	Operating Margin	1.0%	0.68
	Cost of Goods Sold	0.6%	-2.07
	Operating Expenses	1.1%	-1.06
	Guidance	1.8%	-0.45
	Financial Performance	13.3%	0.06
Balance Sheet and Cash Flow Metrics	Cash Flow from Operations	1.7%	0.60
	Capital Expenditures	2.5%	0.40
	Debt Levels and Financing	0.8%	0.09
	Balance Sheet Metrics	2.7%	0.37
	Dividends and Buybacks	1.8%	1.42
Operational Efficiency and Management	Cost Management and Efficiency	8.8%	0.29
	Mergers and Acquisitions (M&A)	3.0%	0.88
	Strategic Initiatives	3.2%	1.29
	Management Changes	3.3%	-0.58
	Workforce	1.0%	-1.31
	Product/Service Updates	9.8%	1.25
	Innovation and R&D	2.6%	1.46
Market and Industry Analysis	Industry Trends	0.7%	-0.19
	Market Share and Competition	3.2%	0.55
	Customer Acquisitions	2.1%	1.43
Geographical and Economic Considerations	Economic Factors	1.3%	-1.48
	Foreign Exchange	1.4%	-0.93
	Geographic Performance	2.6%	0.34
Regulation and Risk	Regulatory Changes	2.5%	-1.00
	Risk Factors	1.5%	-1.85
	Environmental, Social and Governance	1.8%	0.21
Others	Others	11.5%	-1.20

Figure 4:Identified topics distribution and average sentiment per topic on the labeled sentences dataset.

Model	#Tokens	Size	F_1 vs. Teacher	F_1 vs. Human
Paraphrase Albert	256	43MB	46.8%	61.9%
MiniLM-L6	256	120MB	55.1%	60.3%
MPNET	384	420MB	63.1%	72.8%
DistilBERT	512	420MB	61.3%	74.4%
FinBERT	512	438MB	48.8%	54.5%

Table 1:Topic Classification Models Performance.

Model	MPNET	FinBERT	ChatGPT 3.5
F_1 vs. Human	77.8%	65.3%	83.1%

Table 2:Sentiment Classification Models Performance.

Applications

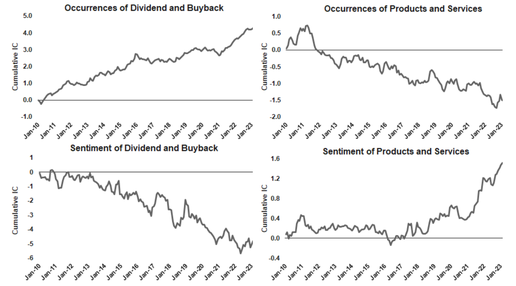


Figure 5:Cumulative IC Trends with respect to propensity and sentiment for Dividend & Buyback (left panels) and Products & Services (right panels).

	Earnings		Revenue	
Filter	Outlook	Trailling	Outlook	Trailling
Earnings	High	High	Medium	Medium
Revenue	Medium	Medium	High	High
Guidance	High	Low	High	Low
Others	Low	Low	Low	Low

Table 3:Filter intensity for earnings and revenue sentiments trends.

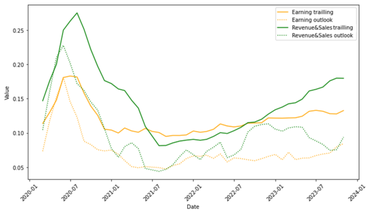


Figure 6:Trends in negativity value: earning, sales with or without outlook.

Conclusion

Our research introduces a new methodology for analyzing earnings calls using a knowledge distillation framework which is better adapted for use on resource-constrained devices. This method has shown potential to identify signals related to stock movements and to provide deeper insights into the content of earnings calls.