DEFINE: Decision-Making with Analogical Reasoning over Factor Profiles

Yebowen Hu,[1] Xiaoyang Wang,[3] Wenlin Yao,[3] Yiming Lu,[4] Daoan Zhang[23]
Hassan Foroosh,[1] Dong Yu,[3] Fei Liu[4]

[1] University of Central Florida, [2] University of Rochester, [3] Tencent AI Lab, [4] Emory University

Introduction

- DEFINE is a new framework for constructing probabilistic **factor profiles** from speech transcripts that describe complex scenarios, and leveraging these profiles with **analogical reasoning** to enhance decision-making

- We demonstrates how to effectively extract decision factors from lengthy transcripts and use them to forecast post-earning stock movements. This approach has the potential to generalize to other domains, such as consulting, and political debates

DEFINE

Step 1: Construct Factor Profile

Simona Jankowski (Vice President, Investor Relations): ... During this call, we will discuss non-GAAP financial measures. You can find a reconciliation of these non-GAAP financial measures to GAAP financial measures in our CFO commentary, ... With that, let me turn the call over to Colette.

Colette Kress (Executive Vice President, Chief Financial Officer): Thanks, Simona. Q1 was another record quarter. Revenue of $26 billion was up 18% sequentially and up 262% year on year and well above our outlook of $24 billion. Starting with Data Center. Data Center revenue of $22.6 billion was a record, up 23% sequentially and up 427% year on year, driven by continued strong demand for the NVIDIA Hopper GPU computing platform...

```
"instance_id": "NVDA_2024-05-22",
"factor profile": {
  "economic_health": {
    "summary": "NVIDIA reported a record quarter with significant revenue growth, particularly in the Data Center segment, driven by strong demand for AI infrastructure. The company expects continued growth across all market platforms in the next quarter, indicating a positive economic environment for NVIDIA's operations.",
    "outcomes": {
      "positive-outlook": "very likely",
      "unknown-or-uncertain": "very unlikely"
    }
  },
  "natural_disasters_and_other_black_swan_events": {
    "summary": "The transcript does not mention any natural disasters or black swan events affecting NVIDIA's operations or market conditions.",
    "outcomes": {
      "major_impact": "very unlikely",
      "unknown-or-uncertain": "likely"
    }
  } ...
```

Decision: "Strong Buy"

Step 3: Apply Analogical Reasoning

Analogous Example A   Analogous Example B   Analogous Example C   Analogous Example D

Step 2: Identify Comparable Profiles

We construct a factor profile for each transcript, defining a set of factors: $F = \{F_1, ..., F_i\}$ where each factor F has potential outcomes $P(O_{ij}|X)$, These probabilities are inferred using a methodology that optimally integrates textual reasoning with quantitative analysis

Then we calculate the Kullback-Leibler (KL) divergence to measure the similarity between profiles.

$$D_{KL}(P||Q) = \sum_{i=1}^{n} \sum_{i=1}^{m} P(O_{ij}|X) \log \frac{P(O_{ij}|X)}{Q(O_{ij}|X_c)}$$

Finally, we identify Top-K profiles that show the lease divergence from a test instance's profile and present these as **analogical examples** for the LLM to consider when reasoning about **stock movements**

Strong Buy | **Buy** | **Hold** | **Sell** | **Strong Sell**

Dataset Statistics

Num. of Transcripts	11,950
Num. of Companies	869
Avg. #Tokens per Transcript	10,187
Avg. #QA Pairs per Transcript	10
Avg. #Trans per Company	14
Avg. #Speakers per Transcript	12
Year Range	2017–2024

Decision Factors

We construct a factor profile for each transcripts based on 15 key factors

Macroeconomic Influences
- Economic Health
- Market Sentiment and Investor Psychology
- Political Events and Government Policies
- Natural Disasters and Black Swan Events
- Geopolitical Issues

Company Specific Dynamics
- Mergers and Major Acquisitions
- Regulatory Changes and Legal Issues
- Financial Health
- Company Growth
- Company Product Launches
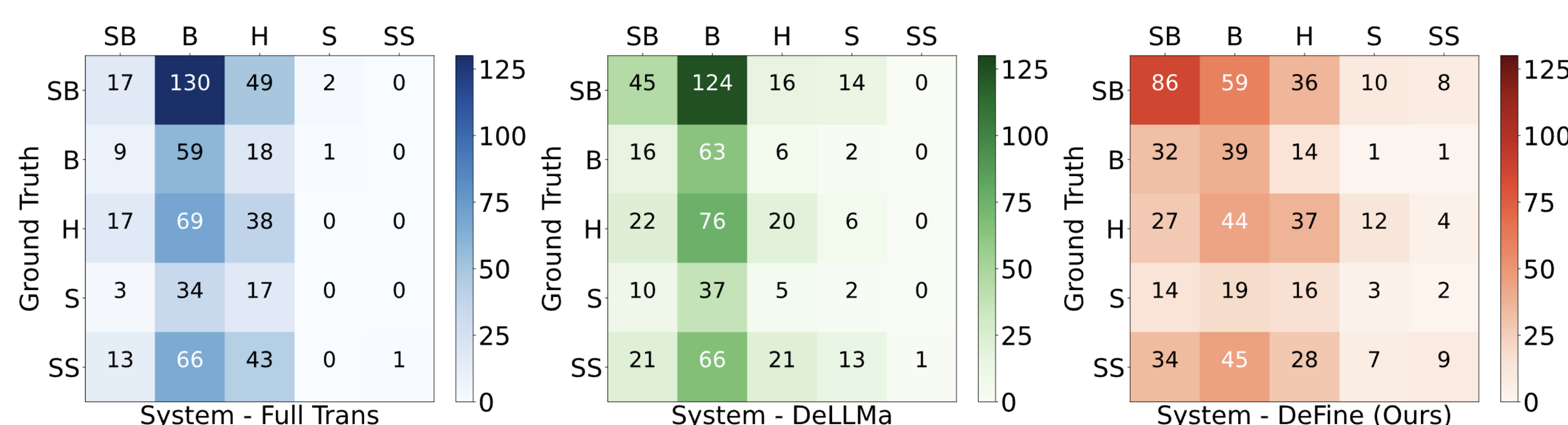- Supply Chain
- Technological Innovation

Historical Financial Metrics
- Historical Earnings Per Share (EPS)
- Historical Revenue
- Historical Stock Prices

Experiments

System	Recall	Prec.	F₁	Accu.
LLM+CoT+Trans	21.56	33.66	13.52	19.59
LLM+CoT+Summ	22.77	16.17	14.12	20.61
LLM+CoT+Factors	24.38	28.58	17.26	22.32
DeLLMa	38.30	23.14	16.68	22.35
DEFINE (Ours)	26.15	27.67	**23.73**	**29.64**

Label	Recall	Prec.	F₁
Strong Sell	7.32	37.50	12.24
Sell	5.56	9.09	6.90
Hold	29.84	28.24	29.02
Buy	44.83	18.93	26.62
Strong Buy	43.22	44.56	**43.88**

We present the accuracy and macro-averaged F-scores for various systems, all using *GPT-4o-2024-08-06*. Our new system, DEFINE, which combines factor profiles with analogical reasoning, achieves the best performance.

We analyze the **key decision factors** using the Bradley-Terry model, ranking to factors which is more likely to lead to favorable stock movement by comparing pairs of **factor profiles**

Factor (Outcome)	Salience
- Regulatory changes and legal issues happened (positive outlook)	0.0364
- Natural disasters and other black swan events (major impact)	0.0360
- Political events and government policies (major upheaval)	0.0349
- Geopolitical issues (escalation to conflict)	0.0345
- Supply chain (positive outlook)	0.0322
- Tech innovation (positive outlook)	0.0317
- Historical stock price change (bullish)	0.0316
- Historical EPS (bullish)	0.0315
- Financial health (positive outlook)	0.0311

Factor (Outcome)	Salience
- Economic health (unknown or uncertain)	0.0362
- Market sentiment and investor psychology (unknown or uncertain)	0.0350
- Company growth (unknown or uncertain)	0.0338
- Supply chain (unknown or uncertain)	0.0326
- Geopolitical issues (escalation to conflict)	0.0322
- Historical revenue (decline)	0.0319
- Historical stock price change (bullish)	0.0318
- Tech innovation (unknown or uncertain)	0.0315
- Natural disasters and other black swan events (major impact)	0.0315
- Political events and government policies (major upheaval)	0.0313

Consumer Defensive Sector   *Technology Sector*
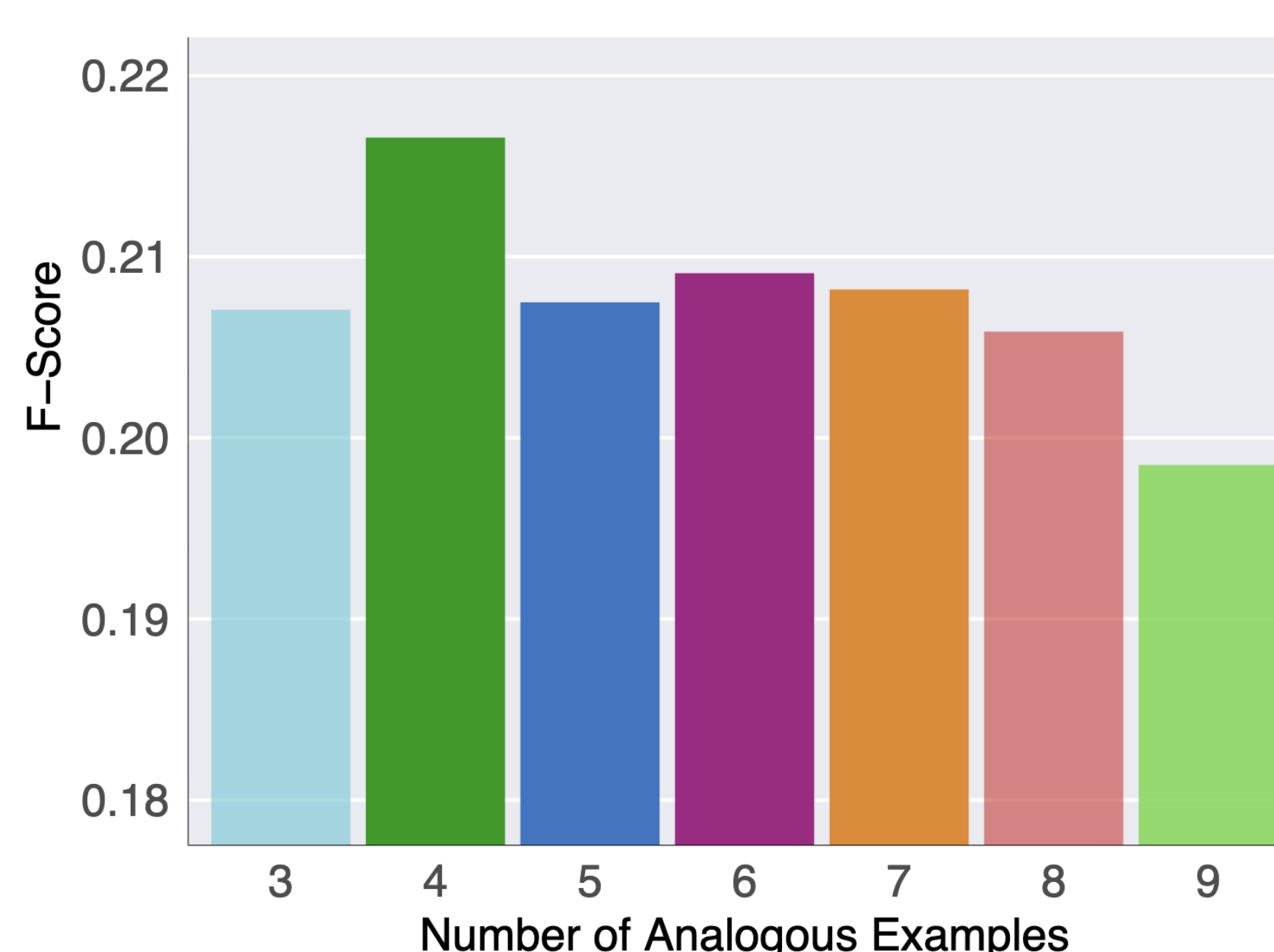
Majority voting from K examples

A comparison of confusion matrices from the LLM + Trans, DeLLMa, and DEFINE methods. While LLM + Trans and DeLLMa lean towards 'Buy (B),' DEFINE *offers more balanced outcomes across all decision categories*, showing notable improvement in 'Strong Buy (SB),' 'Buy (B),' 'Hold (H),' and 'Sell (S)' decisions.

SCAN ME

Code, Data