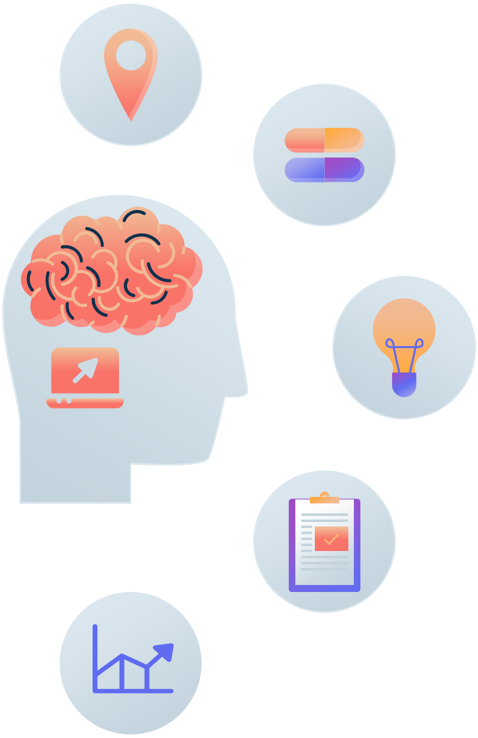
Paper

GitHub Project

NEURAL INFORMATION PROCESSING SYSTEMS



From News to Forecast: Integrating Event Analysis in LLM-Based Time Series Forecasting with Reflection

Xinlei Wang[1], Maike Feng[2], Jing Qiu[1], Jinjin Gu[1], Junhua Zhao[2,3]

[1]School of Electrical and Computer Engineering, The University of Sydney

[2]School of Science and Engineering, The Chinese University of Hong Kong, Shenzhen

[3]Shenzhen Institute of Artificial Intelligence and Robotics for Society

Introduction

Time Series Forecasting (TSF) is crucial for decision-making across economic, infrastructural, and social domains, decoding the evolving relationships within complex real-world systems.

Traditional forecasting methods work well with stable time series, but they

- struggle with **sudden disruptions or anomalies** caused by **external random events**
- fail to systematically link **complex social events** to data fluctuations

Our Purpose:

- Integrating insights from **real-world events** and **their effects** on social and economic behavior
- **Reasoning between textual input and numerical time series**
- Improving the **reliability and accuracy** of TSF

News Article and Other Textual Context

- News articles offer insights into **unexpected events, policy changes, and sentiment shifts** not captured by numerical data.
- Integrating news enriches forecasting with **real-time context and qualitative information for non-linear influences**.

We propose a unified approach that **embeds news and supplementary information into time series data using textual prompts:**

- Next-token prediction tasks
- **Fine-tuning pretrained large language models (LLMs)**

Why Large Language Models?

News Analysis and Selection with LLM Agents

Effective news filtering is a key issue for enhancing time series forecasting as input diversity increases. This requires a deep understanding of how news interact with forecast variables.

Human-like Reasoning Ability and Multi-modal Modeling

The inductive reasoning capabilities of pre-trained LLMs, along with their ability to model multi-modal distributions, **enable few-shot predictions in time series**.

Potential in Forecasting needs to be Further Explored

The potential of language models in TSF through effective tokenization has been proven, but **existing studies mainly utilize the mapping capabilities of LLMs for numerical regression**.

Methods Overview

News Reasoning Agent

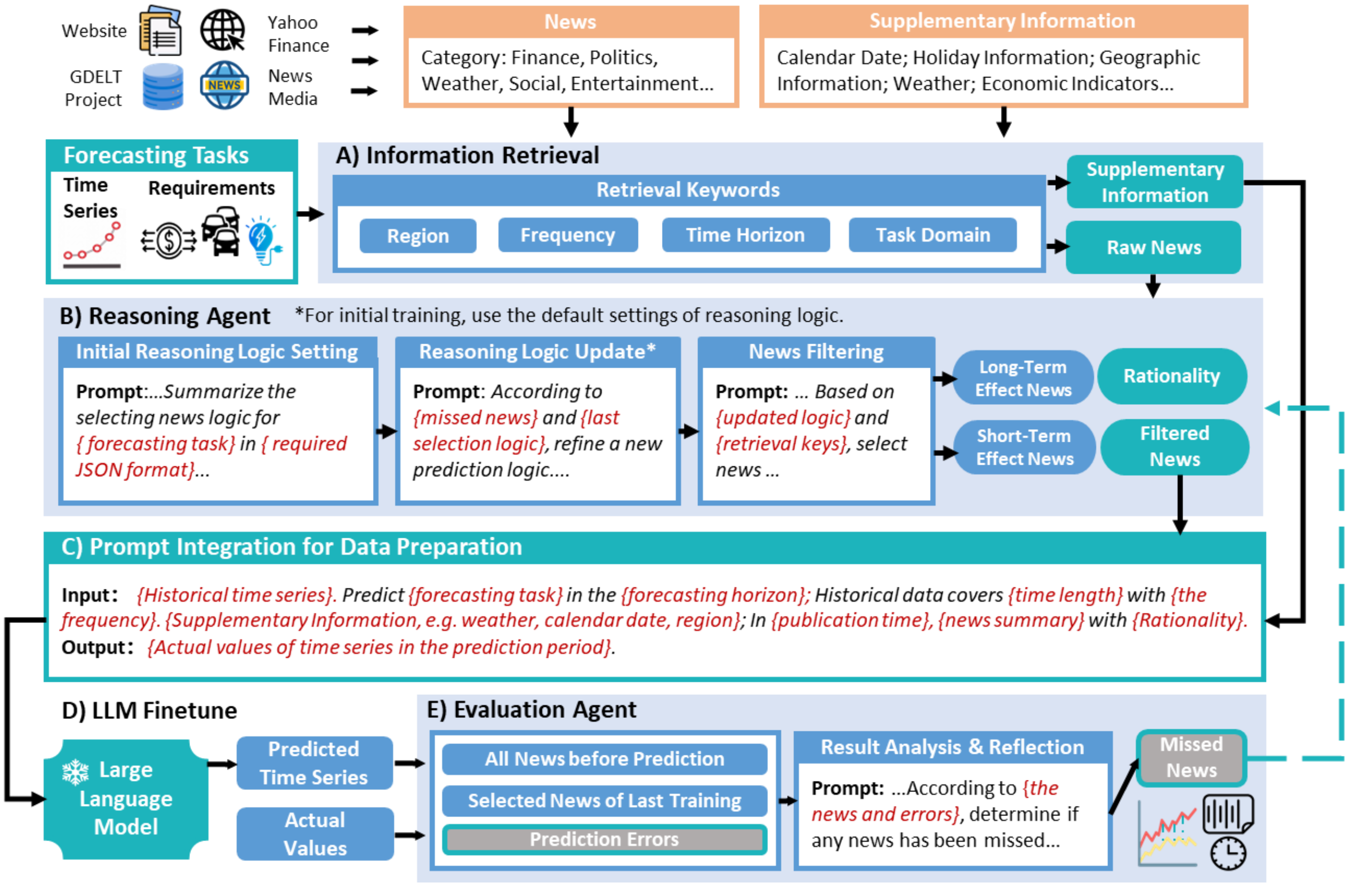
- News filtering

Fine-Tuning LLMs

- Unified input-output data

Evaluation Agent

- Comparing discrepancies between predicted and actual series
- Identify missed news
- Optimize news selection and filtering logic



Rethinking Time Series Forecasting Problem and Elements

Assume a series of event $e_{0:u}$ and a time series $x_{0:t}$: $P(x_{t+1} \mid x_{0:t}) \Rightarrow P(x_{t+1} \mid x_{0:t}; e_{0:u})$

Short-Term Effect News

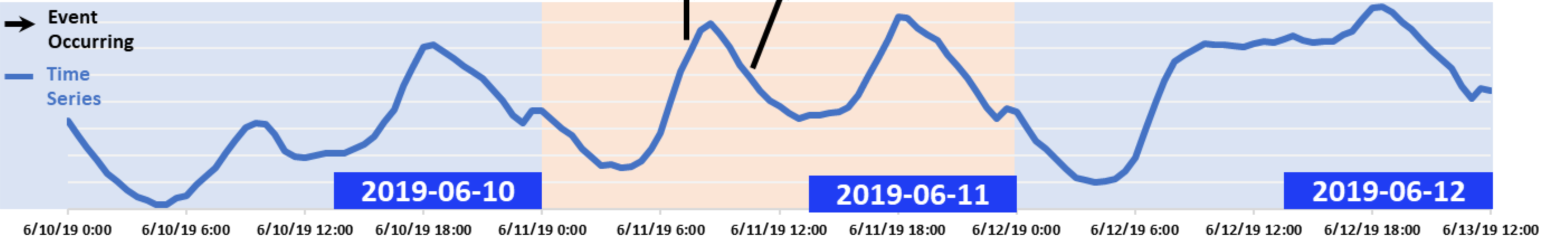
Example:

- **News:** Peak hour delays for Melbourne with a major freeway closed after a serious crash; VIC,2019-06-11 7:18:00 AM.
- **Rationality:** A traffic event can immediately influence electricity consumption, with increased use of public transportation systems, altered business opening hours, and traffic management responses.

Short-Term Effect News

Example:

- **News:** A deluge approaches southern states; Various southern states,2019-06-11 11:00:00 AM.
- **Rationality:** Weather warnings and severe weather can lead to adjustments in electricity consumption due to increased use of heaters, preparing for potential outages, and emergency responses.



Long-Term Effect News

Example:

- **News:** Optus has activated early parts of its 5G network, with more areas to follow; General Australia, 2019-01-30 14:05:00;
- **Rationality:** 5G network expansion requires extensive infrastructure, increasing long-term electricity demands due to additional facilities and greater reliance on 5G technologies.

Short-Term Effect News

Example:

- **News:** Train services were suspended in Melbourne after a train hit a car at a level crossing near Parkdale. VIC, 2019-06-12 3:41:00 PM.
- **Rationality:** The suspension of train services may directly affect electricity load as alternative transportation and emergency services require increased power consumption following the incident.

Time Series and News/Events

The news filtered by the agent mainly includes:

Social Sentiment Changes

Political Events
Elections or legislation affecting exchange rates and regional electricity demand.

Economic Events
Changes in fiscal policy affecting exchange rates

Selected News

Technological Developments
Breakthroughs in AI technology potentially influencing Bitcoin prices

Natural Disasters

Public Health Crises
Public health crises (like the COVID-19 pandemic) impacting traffic flow and electricity load

Evaluation Agent helps identify hidden, unexpected, or counterintuitive events

- Saudi Arabia's "carbon neutrality" goal indirectly impacts Australia's economy and exchange rates by affecting oil prices.

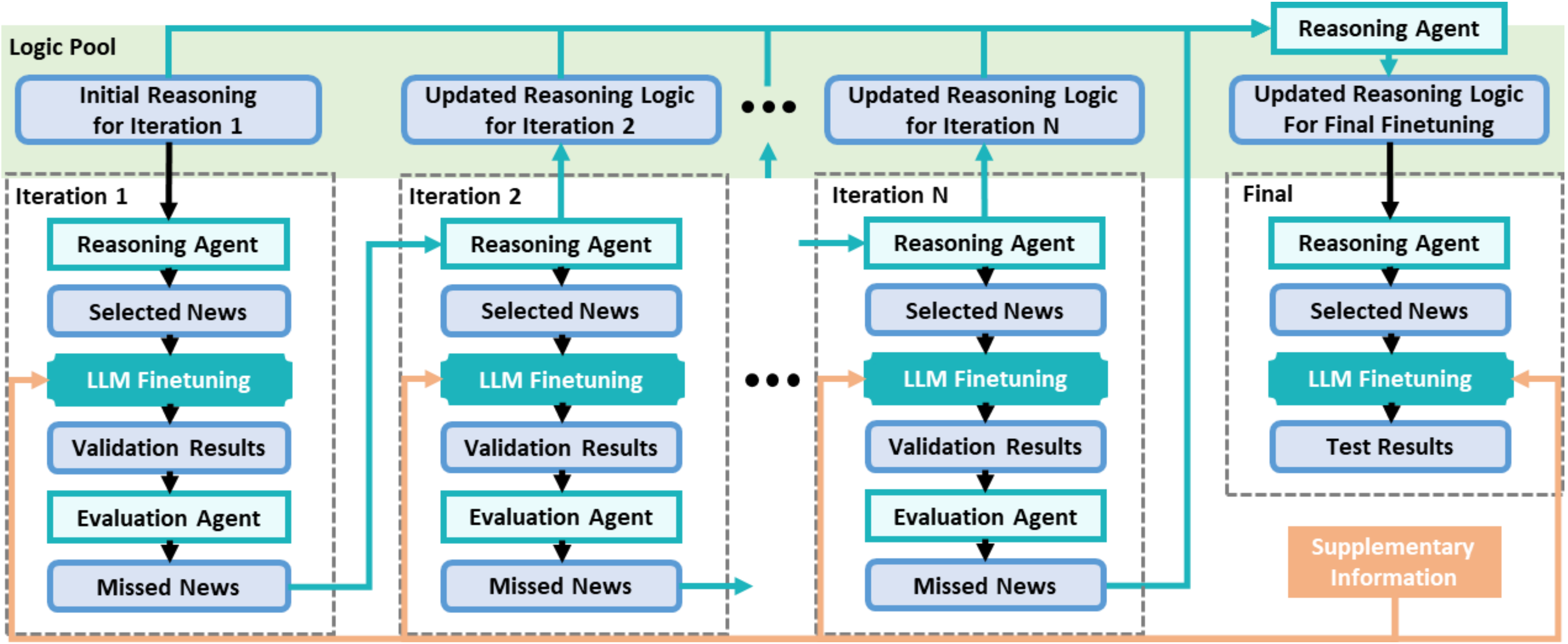
Pairing Time Series Data and Text

Examples of Background Information

1. **Instruction:** "The historical load data is: ...7015.7,6875.1,6634.6,6334.6,6134.7,6007.9,..."
2. **Input:** "Based on the historical load data, please predict the load consumption in the next day. The region for prediction is NSW. The start date of historical data was on 2019-11-9 that is Weekend, and it is not a public holiday. The data frequency is 30 minutes per point. Historical data covers 1 day. The date of prediction is on 2019-11-10 that is Weekend, and it is not a public holiday. Weather of the start date: the minimum temperature is 286.5; the maximum temperature is 297.96; the humidity is 34.0; the pressure is 1012.0. Weather forecast of the prediction date: the minimum temperature is 284.92; the maximum temperature is 301.04; the humidity is 46.0; the pressure is 1016.0. On 2019-11-09 08:51:00, the news can change the time series fluctuation that The ongoing fires lead to an immediate and direct effect on today's load consumption mostly due to loss of infrastructure, increased demand from firefighting efforts, and the need for emergency communications. On 2019-11-09 20:20:00, the news can change the time series fluctuation that The devastating bushfires in NSW lead to increased short-term electricity consumption due to emergency services' operations, resident evacuations, and heightened communication needs. "
3. **Output:** "...6592.6,6467.0,6312.3,6066.8,5902.9,5795.0..."

Selected News

Overall Pipeline



Experiments

Time Series Data

Electricity
Australian State-level Electricity Demand (Half-hourly)

Exchange
Australian Exchange Rate (Daily)

Bitcoin
Bitcoin Price (Daily)

Traffic
California Street-level Traffic Volume (Hourly)

News Data

420,000 news articles are collected

Yahoo

News AU

GDELT Dataset

Supplementary Information

Calendar Date; Holiday Information; Geographic Information; Weather; Economic Indicators...

Effectiveness of news integration.

we conducted ablation experiments comparing four scenarios:

Table 1: Performance comparison of different prompt designs. Red font indicates the best.

	Electricity				Exchange			
	RMSE	MSE$_{\times 10^{-3}}$	MAE	MAPE	RMSE$_{\times 10^{3}}$	MSE$_{\times 10^{5}}$	MAE$_{\times 10^{3}}$	MAPE
Only Numeric Prompt	337.10	113.64	204.89	5.27%	7.80	6.10	5.74	0.77%
Textual Prompt without News	336.41	113.17	206.08	5.29%	7.41	5.49	5.44	0.73%
Textual Prompt with Non-Filtered News	407.86	166.35	250.75	6.84%	8.28	6.86	6.37	0.85%
Textual Prompt with Filtered News	**280.39**	**78.62**	**180.96**	**5.15%**	**6.46**	**4.17**	**4.83**	**0.65%**

	Traffic			Bitcoin			
	RMSE$_{\times 10^{2}}$	MSE$_{\times 10^{3}}$	MAE$_{\times 10^{2}}$	RMSE$_{\times 10^{-3}}$	MSE$_{\times 10^{-6}}$	MAE $_{\times 10^{-3}}$	MAPE
Only Numeric Prompt	4.55	2.07	1.66	4.46	19.94	3.07	5.72%
Textual Prompt without News	4.44	1.97	1.54	3.87	14.97	2.76	5.08%
Textual Prompt with Non-Filtered News	4.89	2.39	1.89	4.02	16.13	2.88	5.35%
Textual Prompt with Filtered News	**4.22**	**1.78**	**1.43**	**3.67**	**13.41**	**2.68**	**4.95%**

Inclusion of agent-filtered news significantly enhances forecasting performance.

Effectiveness of the evaluation agent.

Table 2: Comparison of Iterative Analysis. The baseline case is the initial selection. The arrow means the comparison of each case with baseline cases. A red downward arrow indicates an improvement, a blue upward arrow indicates performance degradation.

	Electricity				Exchange			
	RMSE	$MSE_{\times 10^{-3}}$	MAE	MAPE	$RMSE_{\times 10^{3}}$	$MSE_{\times 10^{5}}$	$MAE_{\times 10^{3}}$	MAPE
1. Initial selection	313.89	98.53	190.79	5.36%	6.61	4.37	4.83	0.65%
2. The second selection	↓287.35	↓82.57	↓180.49	↓4.93%	↓6.46	↓4.17	↓4.83	↓0.65%
3. The third selection	↓303.03	↓91.83	↑192.30	↑5.38%	↑7.69	↑5.92	↑5.63	↑0.75%
4. The fourth selection	↓280.39	↓78.62	↓180.96	↓5.15%	↓6.60	↓4.36	↓4.82	↓0.65%

	Traffic			Bitcoin			
	$RMSE_{\times 10^{2}}$	$MSE_{\times 10^{3}}$	$MAE_{\times 10^{2}}$	$RMSE_{\times 10^{-3}}$	$MSE_{\times 10^{-6}}$	$MAE_{\times 10^{-3}}$	MAPE
1. Initial selection	4.36	1.90	1.45	4.12	16.98	2.97	5.50%
2. The second selection	4.36	1.90	↑1.52	↓3.67	↓13.41	↓2.68	↓4.95%
3. The third selection	↓4.22	↓1.78	↓1.43	↓3.75	↓14.08	↓2.83	↓5.18%

Introducing an evaluation agent that iteratively refines news selection achieve progressively improved time series forecasting results .

Compare to other forecasting methods.

Table 3: Comparison of baselines for time series forecasting on different metrics. A lower value indicates better performance. Red: the best. Blue: the second best.

Domains	Metrics	Ours	Auto. [59]	In. [67]	Dlin. [64]	iTrans. [38]	FiLM [68]	Times. [58]	Pyra. [36]	PatchTST[40]	FED. [69]	GPT4TS [70]
Electricity	MAE	180.96	349.43	282.56	255.7	233.58	254.05	237.49	220.32	234.46	238.77	236.91
	$MSE_{\times 10^{-3}}$	78.62	251.79	166.07	161.59	135.27	153.90	134.42	97.61	133.53	133.96	142.60
	RMSE	280.39	501.78	407.52	401.98	367.79	392.3	366.64	312.42	365.41	366	377.62
	MAPE	5.15%	10.63%	8.94%	7.29%	6.86%	7.36%	6.81%	6.87%	6.56%	6.75%	6.61%
Exchange	$MAE_{\times 10^{3}}$	4.83	9.27	1.75	6.96	5.12	6.44	5.24	14.6	6.73	8.98	15.05
	$MSE_{\times 10^{4}}$	0.42	1.36	4.76	0.91	0.45	0.77	0.45	3.55	0.77	1.28	4.01
	$RMSE_{\times 10^{2}}$	0.65	1.17	2.18	9.52	0.671	0.875	0.673	1.88	0.875	1.13	2.00
	MAPE	0.65%	1.23%	2.32%	0.92%	0.68%	0.85%	0.70%	1.94%	0.90%	1.21%	1.34%
Traffic	$MAE_{\times 10^{2}}$	1.43	2.49	4.44	1.70	1.56	1.65	1.61	1.51	1.84	1.74	1.64
	$MSE_{\times 10^{3}}$	1.78	2.19	5.27	1.67	1.54	1.71	1.49	0.98	1.54	1.43	1.45
	$RMSE_{\times 10^{2}}$	4.22	4.68	7.26	4.09	3.93	4.14	3.86	3.13	3.92	3.79	3.81
Bitcoin	$MAE_{\times 10^{-3}}$	2.68	4.28	12.27	5.74	3.20	3.28	3.17	9.22	2.85	3.96	2.84
	$MSE_{\times 10^{-6}}$	13.41	27.64	162.47	50.90	16.21	17.65	16.38	123.71	13.52	24.60	13.66
	$RMSE_{\times 10^{-3}}$	3.67	5.26	12.75	7.13	4.03	4.20	4.05	11.12	3.68	4.96	3.70
	MAPE	4.95%	7.61%	21.28%	10.39%	5.70%	5.84%	5.64%	16.16%	5.13%	6.97%	5.08%

Conclusions

Enhanced Forecasting Performance

Integrating news into time series forecasting with LLM-based methods improves prediction accuracy and model intelligence.

Potential Paradigm Shift in TSF

Encourages forecasting aligned with real-world dynamics through the effective utilization of unstructured text data.

Expand Applicability

Equip agents with sophisticated analytical tools for better data processing. Enhance precision and relevance, broadening use in predictive analytics.

Thanks!

Do you have any questions?

Xinlei.wang@sydney.edu.au

Paper

GitHub Project