WESTLAKE UNIVERSITY

NumHTML: Numeric-Oriented Hierarchical Transformer Model for Multi-task Financial Forecasting

PRESENTER: LINYI YANG

Jan 2022

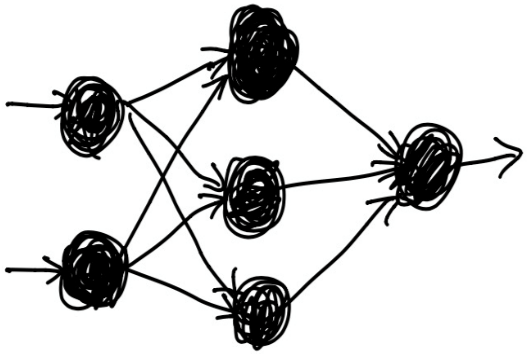
The Booming Growth of FinNLP

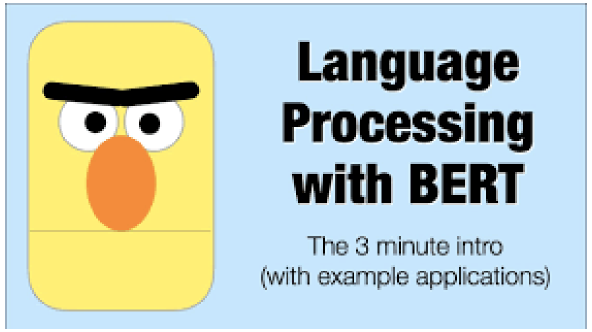
Traditional financial analysis

Great progress

The explosion of unstructured data

Neural structures for NLP





https://gilmartinir.com/earnings-video-conference-calls/

Background: Financial Forecasting

- Stock Price Prediction/Forecasting:
  - ✓ Time-series Prediction: Historical pricing (Quant)
  - ✓ **Textual Data:** Financial news, Social Media (Twitter), and 10K/10Q Report, **Earnings Conference Calls**

- Volatility Forecasting:
  - ✓ Time-series Models: VIX Index, ARCH, GARCH
  - ✓ Textual Data: Financial news, Analyst Reports, Social Media, and **Earnings Conference Calls**

Background: Earnings Conference Calls

Brian Nowak, Analyst: Thanks for taking my questions. One on YouTube, I guess. Could you just talk to some of the qualitative drivers that are really bringing more advertising dollars on to YouTube? And then I think last quarter you had mentioned the top 100 advertiser spending was up 60% year-on-year on YouTube, wondering, if you could update us on that? And the second one on search, it sounds like mobile is accelerating. Where are you now in the mobile versus desktop monetization gap? And, Sundar, how do you think about that long-term? Do you see mobile being higher, reaching equilibrium? How do you see that trending?

Sundar Pichai, CEO: On the YouTube one. Look, I mean, the shift to video is a profound medium shift and especially in the context of mobile, you know and obviously users are following that. You're seeing it in YouTube as well as elsewhere in mobile. And so, advertisers are being increasingly conscious. They're being very, very responsive. So, we're seeing great traction there and we'll continue to see that. They are moving more off their traditional budgets to YouTube and that's where we are getting traction. On mobile search, to me, increasingly we see we already announced that over 50% of our searches are on mobile. Mobile gives us very unique opportunities in terms of better understanding users and over time, as we use things like machine learning, I think we can make great strides. So, my long-term view on this is, it is as-compelling or in fact even better than desktop, but it will take us time to get there. We're going to be focused till we get there.

Figure 1: Earnings calls are extremely complex examples of naturally-occurring discourse. In this example question-answer pair from a Google earnings call on October 27, 2016, the analyst asks six distinct questions in a single turn. Because the interaction originates as speech, there are discourse markers and hedging. The analyst and executive discuss concrete entities and performance statistics and past, present and future performance.

Background: Representing Numbers in Language Models

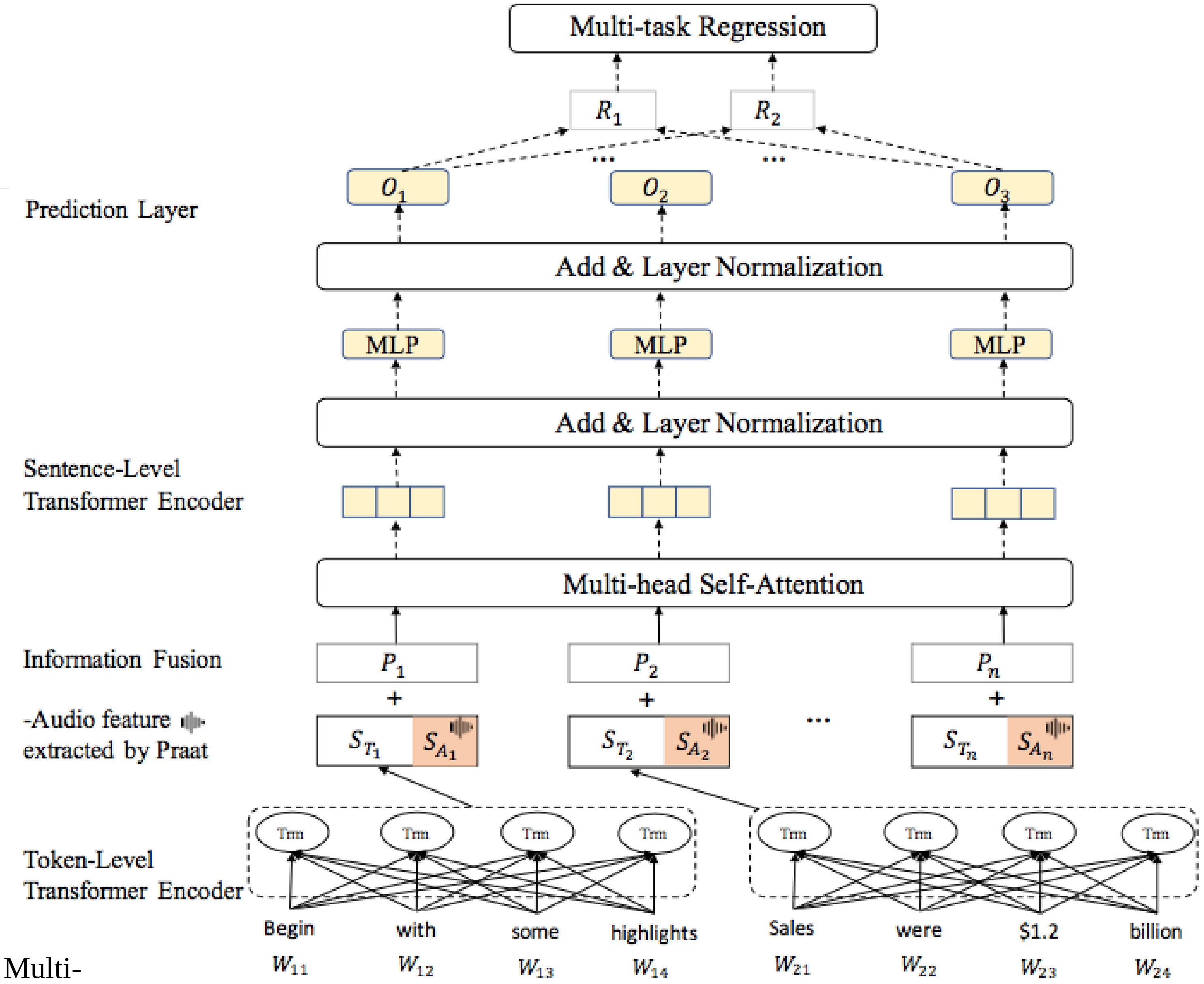
- Numeral Category Classification:
  - ✓ Financial Domain: Monetary, Temporal, Percentage [Chen, 2019, 2021]

- Magnitude Comparison [Wallace, 2019] :
  - ✓ List Maximum from 5-numbers: [$1.2, $2.5, $5, $9.8, **$9.9**]
  - ✓ Numbers Selection: Near values from the same category

[Chen, 2019] Numeracy-600k: learning numeracy for detecting exaggerated information in market comments. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, ACL ’19, 6307–6313.
[Wallace, 2019] Do NLP Models Know Numbers? Probing Numeracy in Embeddings. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP),* 5310–5318.
[Chen, 2021]. Evaluating the Rationales of Amateur Investors. In *Proceedings of the World Wide Web Conference (WWW-21).*

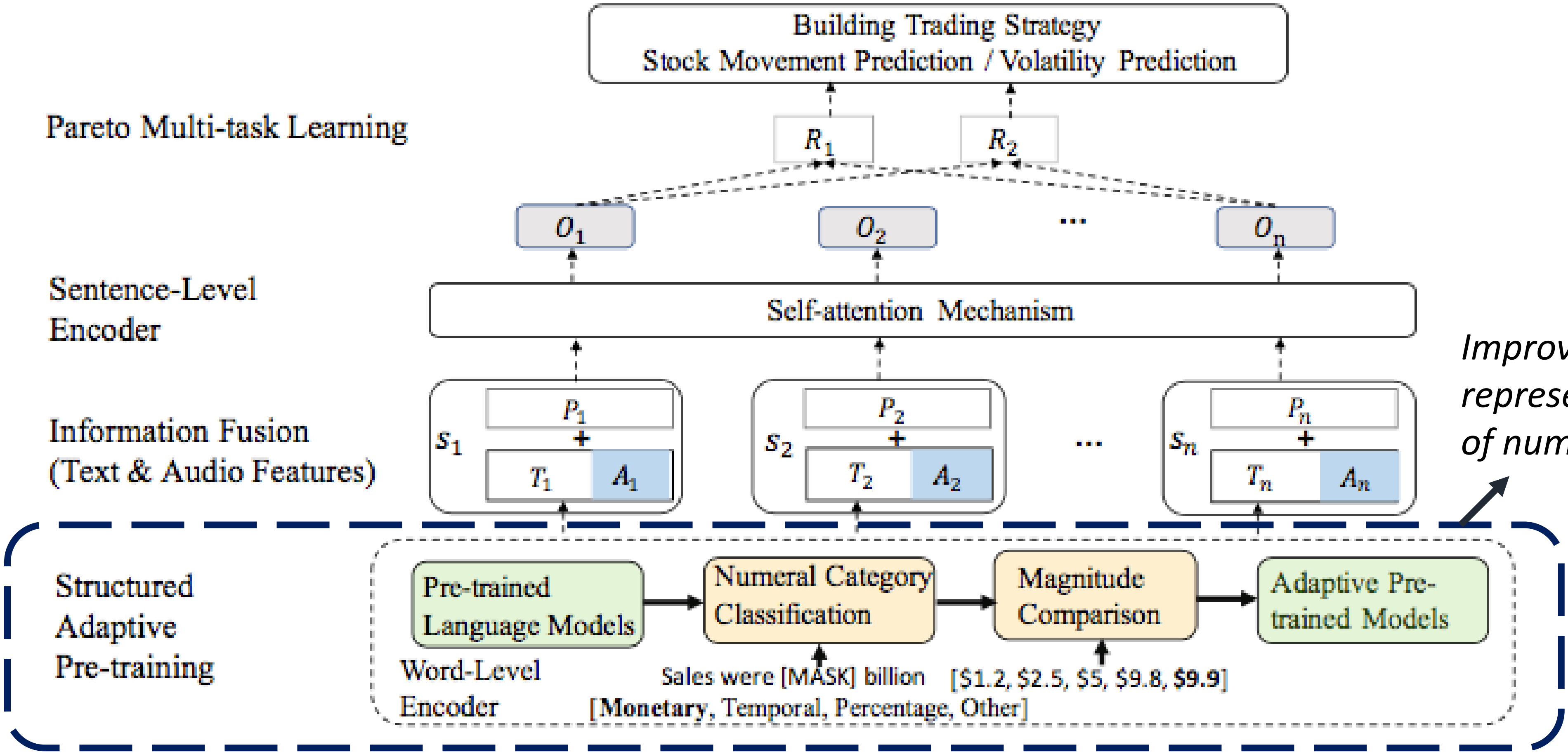
Related Work: HTML

- ➢ Long-form document encoding
- ➢ Multi-modal Information Fusion
- ➢ Multi-task Learning



[Yang, 2020] HTML: Hierarchical Transformer-based Multi-task Learning for Volatility Prediction. In *Proceedings of the WWW-20, The Web Conference 2020*, 441-451

Methodology: NumHTML



Improving the representation quality of numbers

(1) word-level encoder; (2) multimedia information fusion; (3) sentence-level encoder; and (4) Pareto Multi-task Learning.

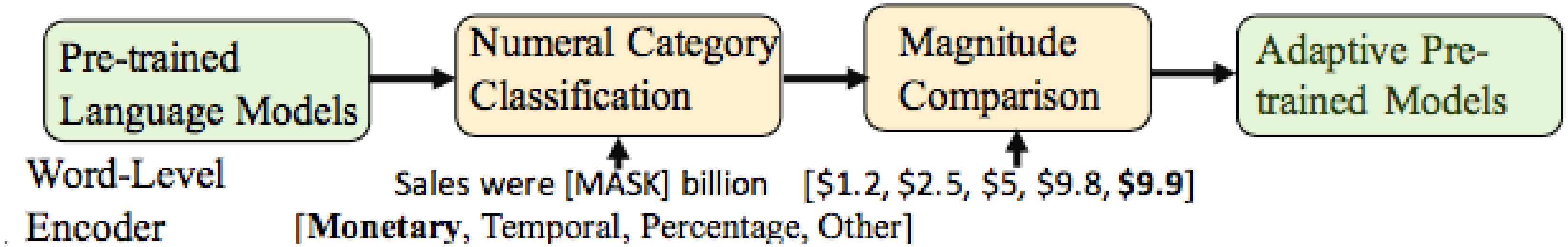
Methodology: Structured Adaptive Pre-training

(1) Category Classification:

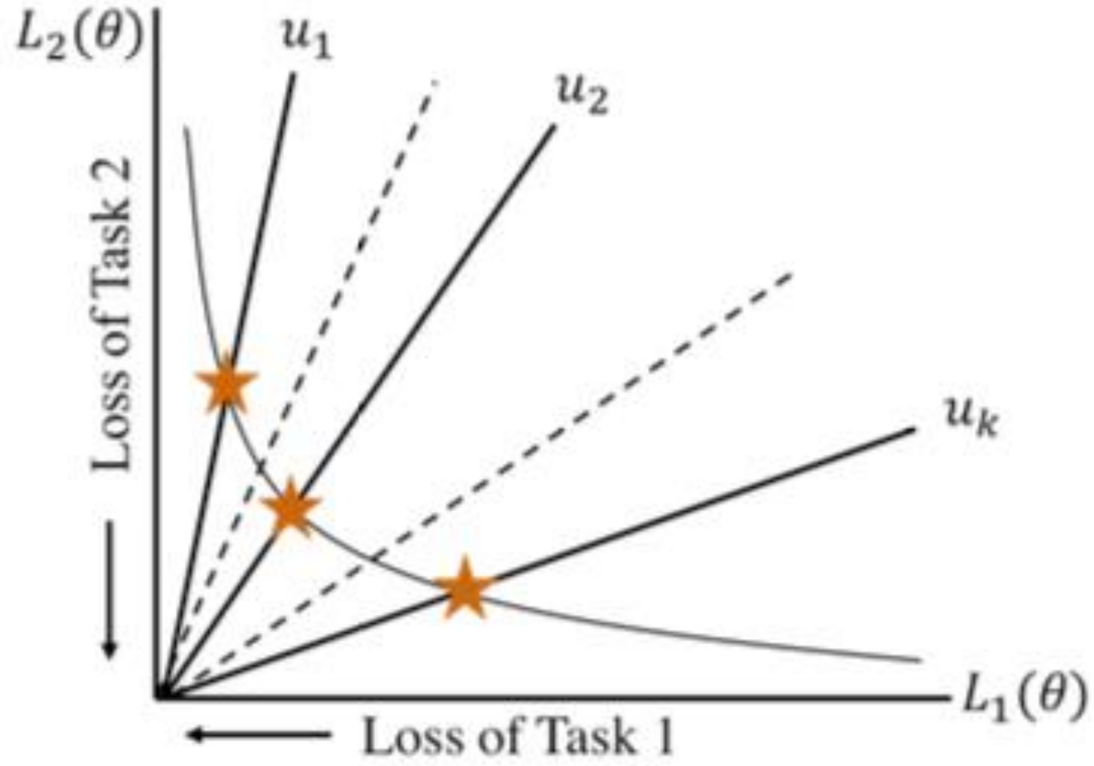
‘During 2020 profits increased by 13% to $205m’ presumably this is tagged as monetary (because of the $205m), temporal (2020) and percentage (%).

(2) Magnitude Comparison:

Input: five monetary numbers [$1.2, $2.5, $5, $9.8, $9.9]
Golden label: [0, 0, 0, 0, 1]



Methodology: Pareto Multi-task Learning



Pareto MTL aims to find a set of Pareto solutions in different restricted preference regions (Lin et al. 2019).

$$\mathcal{F} = \alpha_{patero_1} \sum_i (\hat{y}_i - y_i)^2 + \alpha_{patero_2} \sum_j (\hat{y}_j - y_j)^2$$

Dataset: Multi-modal Earnings Conference Calls

This dataset contains 576 earning calls recordings, correspond to 88,829 text-audio aligned sentences, for S&P 500 companies in U.S. stock exchanges.

We split the dataset into mutually exclusive training/validation/testing sets in the ratio of 7:1:2 (refers to instances) in chronological order, since future data cannot be used for prediction.

Experiments

(1) Stock Prediction Task – Regression (MSE)
(2) Stock Volatility Prediction Task – Regression (MSE)
(3) Trading Simulation Task (Profit & Sharp Ratio)

Pareto Multi-task Learning

Experiments

Table 3: Results for the future n-day stock movement prediction (higher is better). * and ** indicate statistically significant improvements over the state-of-the-art ensemble method with p<0.05, p<0.01 respectively, under Wilcoxon's test.

Model	Price Movement Predictions							
	MCC_3	MCC_7	MCC_{15}	MCC_{30}	$F1_3$	$F1_7$	$F1_{15}$	$F1_{30}$
Price-based LSTM	0.069	0	0.097	0	0.271	0.694	0.200	0.765
Price-based BiLSTM-ATT	0	0	0	0	0.149	0.342	0.200	0.721
SVM	-0.069	0.015	-0.048	-0.003	0.524	0.683	0.645	0.734
HAN (Glove)	0.090	-0.005	0.266	-0.042	0.591	0.621	0.598	0.703
Text-only Methods								
MDRM	0.117	-0.107	0.032	-0.085	0.675	0.500	0.571	0.601
HTML	0.195	0.007	0.119	0.022	0.623	0.688	0.648	0.700
Ensemble	0.204	0.008	**0.132**	0.024	0.675	0.690	0.636	0.703
NumHTML	**0.229****	**0.009**	0.122	**0.031**	**0.689***	**0.691**	**0.644****	**0.727***
Multi-modal Methods								
MDRM (Multi-modal)	0.095	0.056	0.159	-0.065	0.628	0.690	0.452	0.590
HTML (Multi-modal)	0.280	0.125	0.196	0.131	0.696	0.695	0.703	0.748
Ensemble (Multi-modal)	0.321	0.128	0.191	0.128	0.702	0.698	0.702	0.761
NumHTML (w/o Pareto MTL)	0.293	**0.129**	0.198	0.133	0.701	**0.700**	0.711	0.759
NumHTML (w/o Adaptive Pre-training)	0.282	0.121	0.199	0.130	0.697	0.668	0.705	0.746
NumHTML (Multi-modal)	**0.325****	0.126	**0.206****	**0.136***	**0.722***	0.697	**0.716***	**0.770****

Experiments

Table 4: Cumulative profit across different trading strategies.

Strategy	**Profit**	**Sharpe Ratio**
Simple Baselines		
Buy-all	\$36.59	0.76
Short-sell-all	-\$36.59	-0.77
Random	-\$25.78	-0.58
Multi-modal Methods		
MRDM	\$38.75	0.81
HTML	\$72.47	1.52
NumHTML (w/o Pareto)	\$73.90	1.53
Ensemble	\$75.73	1.59
NumHTML	**\$77.81**	**1.62**

Ablation Studies

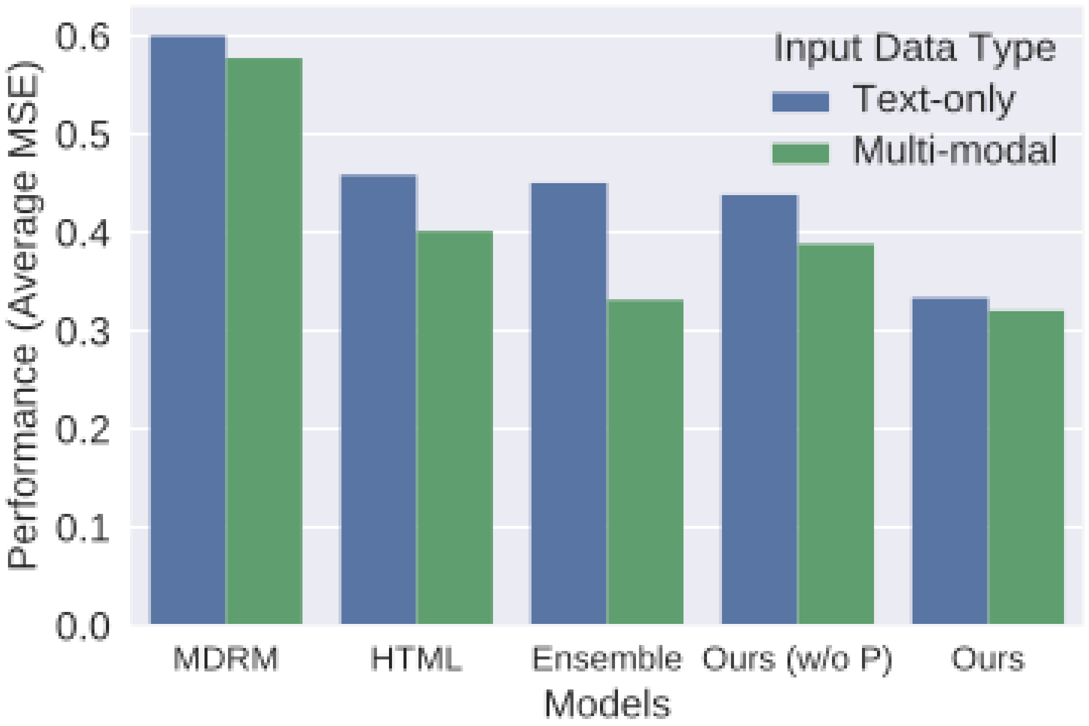


Figure 3: The results of volatility prediction. 'Ours (w/o P)' indicates NumHTML without Pareto MTL. Text-only and multi-modal methods are presented by different colors.

THANKS FOR YOUR ATTENTIONS!!

Q&A

THANKS TO ANONYMOUS REVIEWERS FOR YOUR HELPFUL FEEDBACK.