Centre for
Data Analytics

Insight

HTML: Hierarchical Transformer-based Multi-task Learning for Volatility Prediction

23/04/2020

Presenter: Linyi Yang
Supervisors: Ruihai Dong; Barry Smyth
Afflation: Insight Centre, University College Dublin (UCD)

DCU NUI Galway OÉ Gaillimh UCC University College Cork, Ireland Coláiste na hOllscoile Corcaigh UCD DUBLIN | sfi Science Foundation Ireland

Contents

Financial Forecasting

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

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

State-of-the-art Model

Volatility Forecasting using the Earnings Conference Call data:

- Interesting idea: What you say and how you say it matters
- Dataset: S&P 500 Companies' Earnings Call Transcripts and Audio Recordings in 2017

- Model: Multimodal Deep Regression Model (MDRM)

What You Say and How You Say It Matters: Predicting Financial Risk Using Verbal and Vocal Cues

Yu Qin
School of Information
Renmin University of China
qinyu.gemini@gmail.com

Yi Yang *
HKUST Business School
Hong Kong University of Science and Technology
imyiyang@ust.hk

[Qin and Yang, ACL-19]

Earnings Conference Call Data

Insight

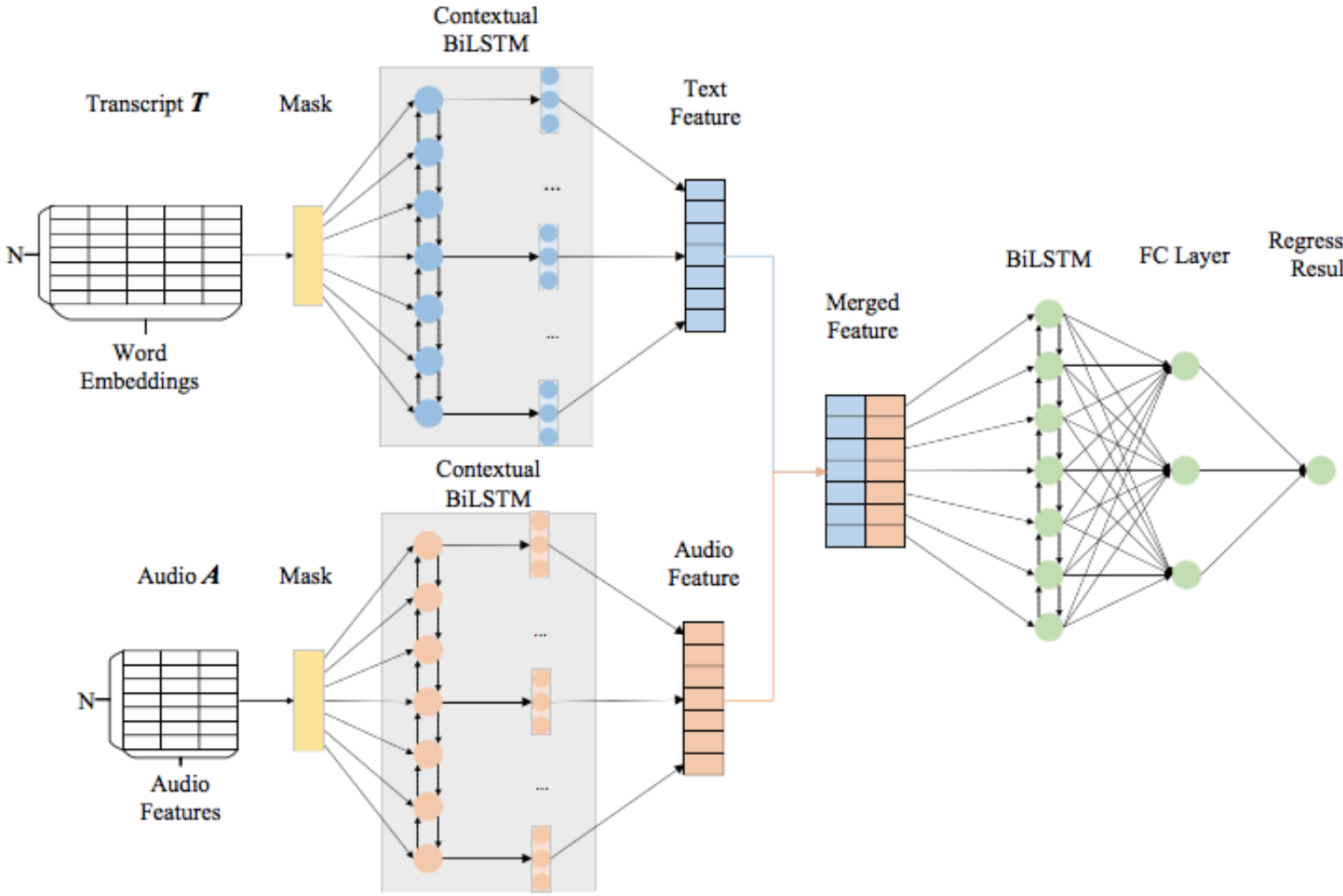
Brian Nowak, Analyst: Thanks for taking my questions. One on YouTube, I guess. Could you just talk to some of the qualitative drivers that are really bringing more advertising dollars on to YouTube? And then I think last quarter you had mentioned the top 100 advertiser spending was up 60% year-on-year 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.

[Katherine and Amanda, ACL-19]

State-of-the-art Model



The State-of-the-art Multimodal Deep Regression Model (MDRM)

Motivation

A straightforward thinking is that transformer could help with feature extractions, but the use of the transformer meets three main issues:

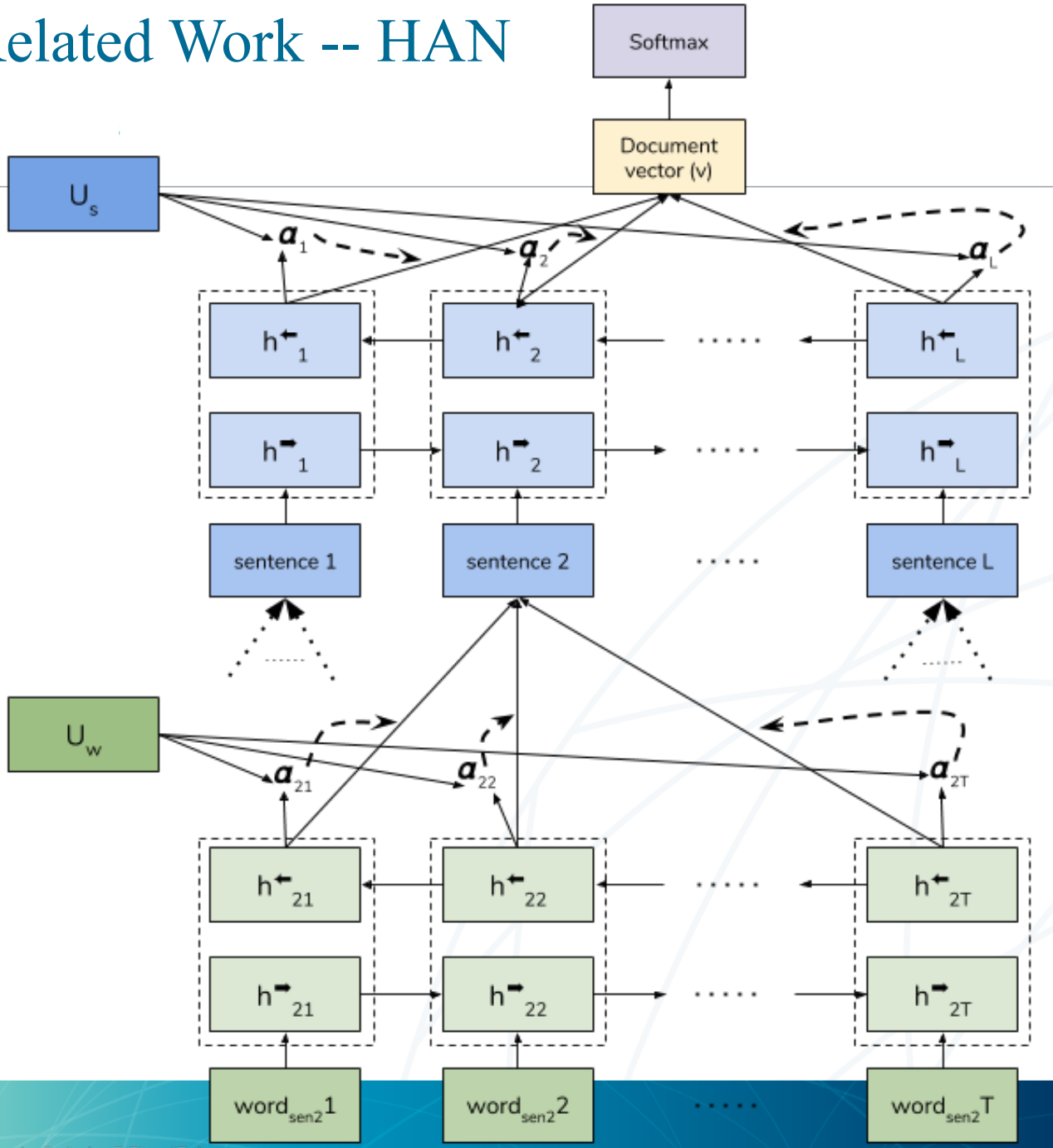
1. It is challenging for transformers to process such a long document avg_len=2000+ tokens -- limited by the computation cost --

2. The large-scale pre-trained models rely on the original training goal too much -- predict-masked-token –

3. Audio features and textual features cannot be concatenated at the sentence-level by the original transformers

→ The idea of using hierarchical transformer came out: HTML

Related Work -- HAN



Sentence attention

U_s: sentence context vector

Sentence encoder

Word attention

α : importance weight
U_w : word context vector (*what is the informative word?*)

Word encoder

$\overrightarrow{h}$: forward hidden state
$\overleftarrow{h}$: backward hidden state
$word_{sen2}1$: The first word in sentence 2

Motivation

Insight

Using hierarchical transformer seems like a good idea, however, the overfitting problem is easy to occur for these reasons:

1. HAN vs. HTML: Hierarchical attention network (HAN) does not rely on any pre-trained model, while HTML relies on transformers at dual levels

2. The parameters of the sentence-level transformer are randomly initialized

3. The financial index is hard to strictly follow the experience of the past

→ Multi-task Learning may help model enhance generalization: **HTML**

Pick a good auxiliary task

Insight

Assumption:

1) Multi-task learning is a natural fit in finance or economics forecasting where we may want to predict the value of more than one possibly related indicators.

2) The auxiliary task should be related to the main task in some way and that it should be helpful for predicting the main task.

Prediction Goals

Main Task: Log volatility as our basic measure of the average n-day volatility

$$v_{[0,n]} = \ln\left(\sqrt{\frac{\sum_{i=1}^{n}(r_i - \bar{r})^2}{n}}\right) \quad (1)$$

In Equation 1, r_i is the stock return on day i and $\bar{r}$ is the average stock return in a window of n days. The return is defined as $r_i = (P_i - P_{i-1})/P_{i-1}$, where P_i is the adjusted closing price of a stock on day i.

Auxiliary Task: The single day log volatility

$$v_n = \ln\left(\left|\frac{P_n - P_{n-1}}{P_{n-1}}\right|\right)$$

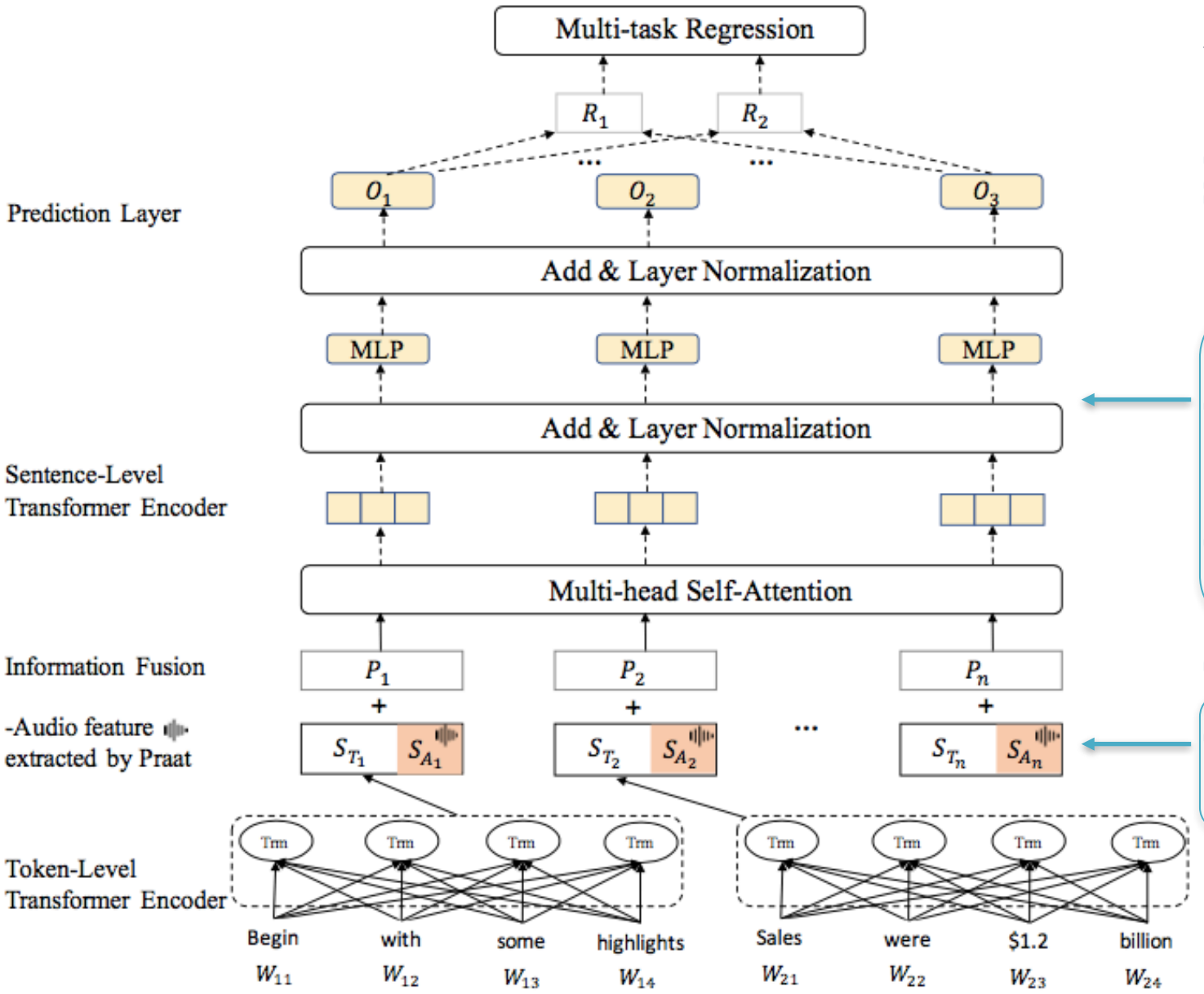
Motivation

Insight

Our motivations can be summarized:

- Better process the long document
- Using co-evolutionary methods to build multimodal
- Alleviate the overfitting problem by using the multi-task learning
- Reduce the training time largely by the lightweight calculation

Model



Two options:
1) Joint Training
2) Agile Training

Information Fusion

Research Questions

1) Is the vocal cue really important for the market?

If it is, to what extent?

2) Does HTML work well for the volatility forecasting?

If it is, what is the result of the ablation study?

Experimental Results

Earnings Conference Call Dataset (Transcripts + Audios) the same as [Qin, ACL-19]:

- Collected from Seeking Alpha and EarningsCast
- In total, there are 576 instances with 88,829 sentences including 280 listed companies during 2017

Experimental Results

Price-based Methods		**n=3**	**n=7**	**n=15**	**n=30**
Linear Regression		1.710	0.526	0.330	0.284
LSTM		1.970	0.459	0.320	0.235
LSTM+ATT		1.852	0.470	0.308	0.231
MTLSTM+ATT		1.983	0.435	0.304	0.233
Text-based Methods		**n=3**	**n=7**	**n=15**	**n=30**
SVR+RBF(TF-IDF)		1.695	0.498	0.342	0.249
SVR+RBF(Glove)		1.667	0.549	0.345	0.275
HAN(Glove)		1.426	0.461	0.308	*0.198*
Multimodal Methods		**n=3**	**n=7**	**n=15**	**n=30**
SVR(Glove+Audio)	Text+Audio	1.722	0.501	0.307	0.233
bc-LSTM(Glove+Audio)[45]	Text+Audio	1.418	0.436	0.304	0.219
MDRM [46]	Text Only	1.431	0.439	0.309	0.219
	Audio Only	1.412	0.440	0.315	0.224
	Text+Audio	*1.371*	*0.420*	*0.300*	0.217
HTML (Ours)	Text Only	1.175	0.372	**0.153**	**0.133**
	Text+Audio	**0.845**	**0.349**	0.251	0.158

Improvements:
3-days (+38.4%)
7-days (+16.9%)
15-days(+49.0%)
30-days(+38.7%)

Ablation Studies

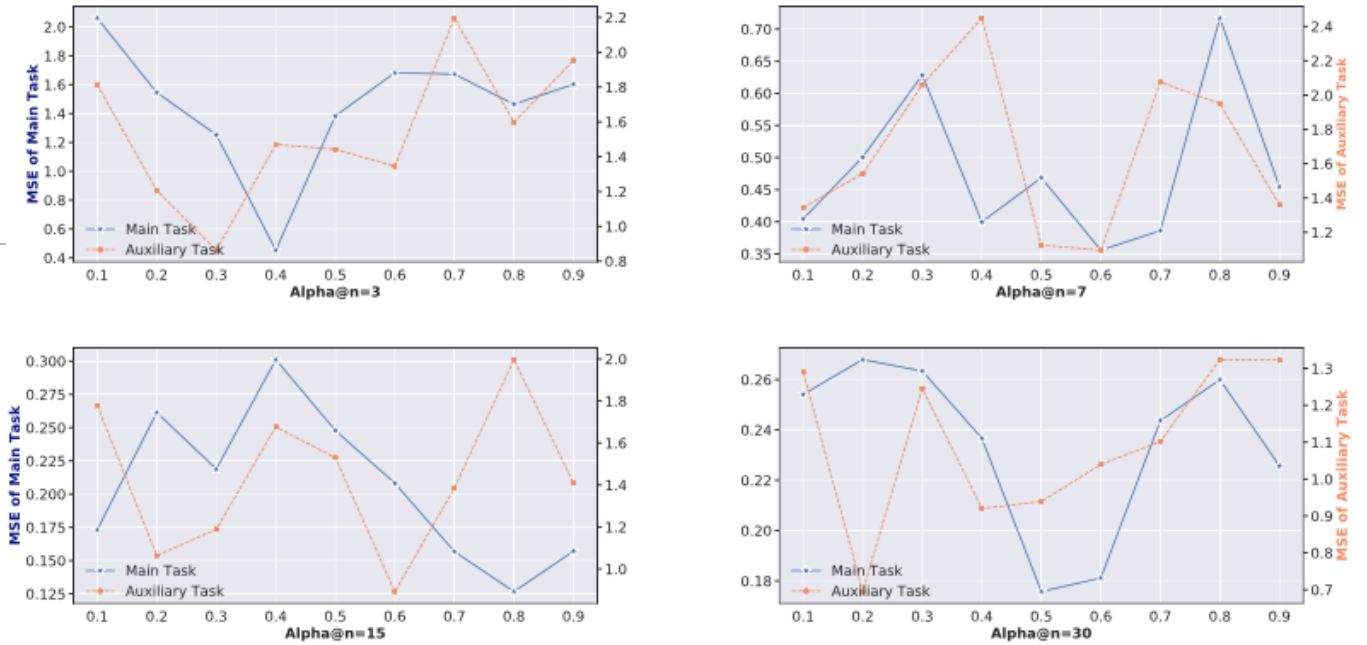
Table 3: Ablation studies on the multi-task learning and embeddings. HTSL and HTML are short for Hierarchical Transformer-based Single-task Learning and Hierarchical Transformer-based Multi-Task Learning respectively

Model	Embeddings	n=3	n=7	n=15	n=30
HTSL	Glove	1.558	0.469	0.291	0.181
	Glove+Audio	1.313	0.389	0.330	0.238
	WWM-BERT	1.344	0.363	0.271	0.162
	WWM-BERT+Audio	1.087	0.432	0.308	0.181
HTML	Glove	1.574	0.474	0.276	0.164
	Glove+Audio	1.278	0.370	0.282	0.201
	WWM-BERT	1.175	0.372	**0.153**	**0.133**
	WWM-BERT+Audio	**0.845**	**0.349**	0.251	0.158

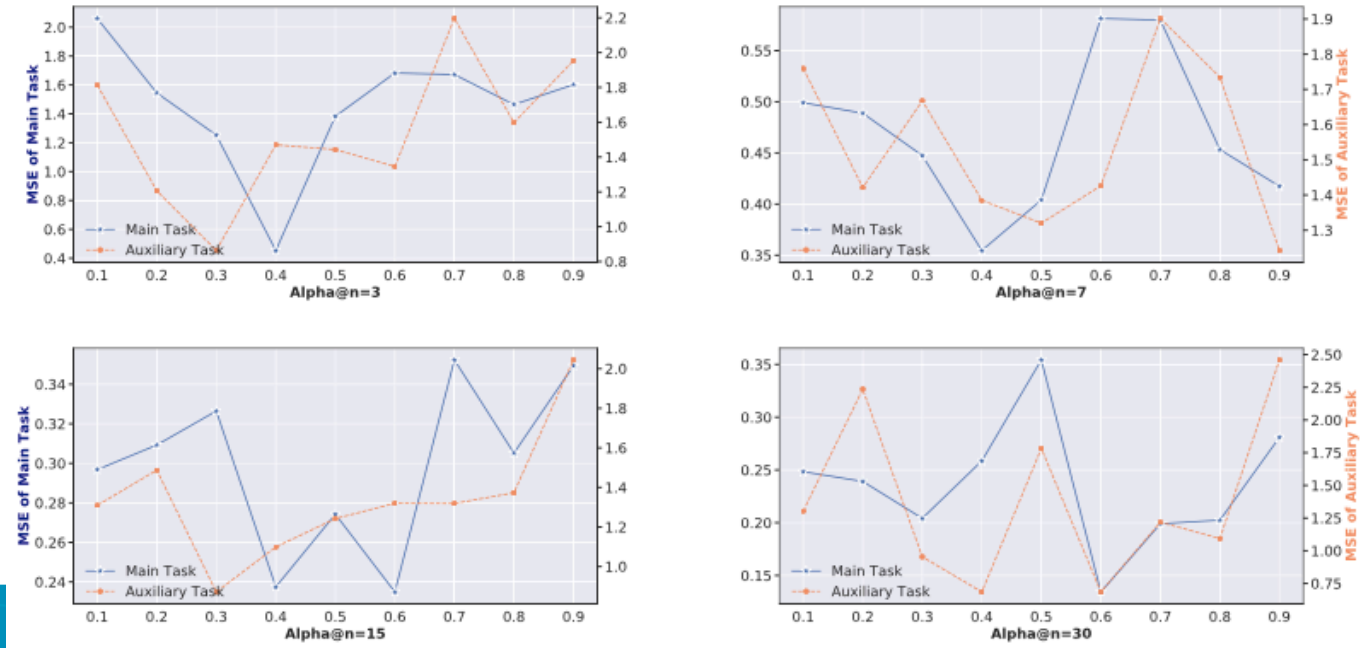
Interesting Findings:
1) Vocal features guarantee more benefits for short-term prediction
2) Multi-task learning guarantee more benefits for long-term prediction

Hyper-parameters selection

Optimal hyper-parameters are generally consistent between two tasks



(a) Text-only data as input



(b) Multimedia data as input

Conclusion

- We have proposed a novel hierarchical, multi-task, transformer learning model for volatility prediction, based on the text and/or audio of earning calls
- Our HTML model builds on very recent work and delivers substantial performance improvements, providing a new performance benchmark for this task
- The utility of audio data exists a significant opportunity to explore the use of audio features in a range of related tasks (e.g. fraud detection, asset pricing, stock recommendation etc.)

Future Directions

- Extend the size of the dataset -> 576-3900+ instances
- Explore the use of the Numerical Embedding in Finance Domain

"Do NLP Models Know Numbers? Probing Numeracy in Embeddings" [AllenNLP, EMNLP-19]

- Jointly training with the event extraction and event representation

"Open Domain Event Extraction Using Neural Latent Variable Models" [Yue Zhang, ACL-19]

Thanks for listening!

Q&A

Linyi Yang