XRR: Explainable Risk Ranking for Financial Reports

The European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases

Ting-Wei Lin,[1] Ruei-Yao Sun,[1] Hsuan-Ling Chang,[2] Chuan-Ju Wang,[3] Ming-Feng Tsai[1]

[1]Department of Computer Science, National Chengchi University, Taiwan
[2]Department of Finance, National Taiwan University, Taiwan
[3]Research Center for Information Technology Innovation, Academia Sinica, Taiwan

Agenda

- ❖ Introduction

- ❖ Methodology: XRR

- ❖ Experiments

- ❖ Financial Analysis

- ❖ Discussion
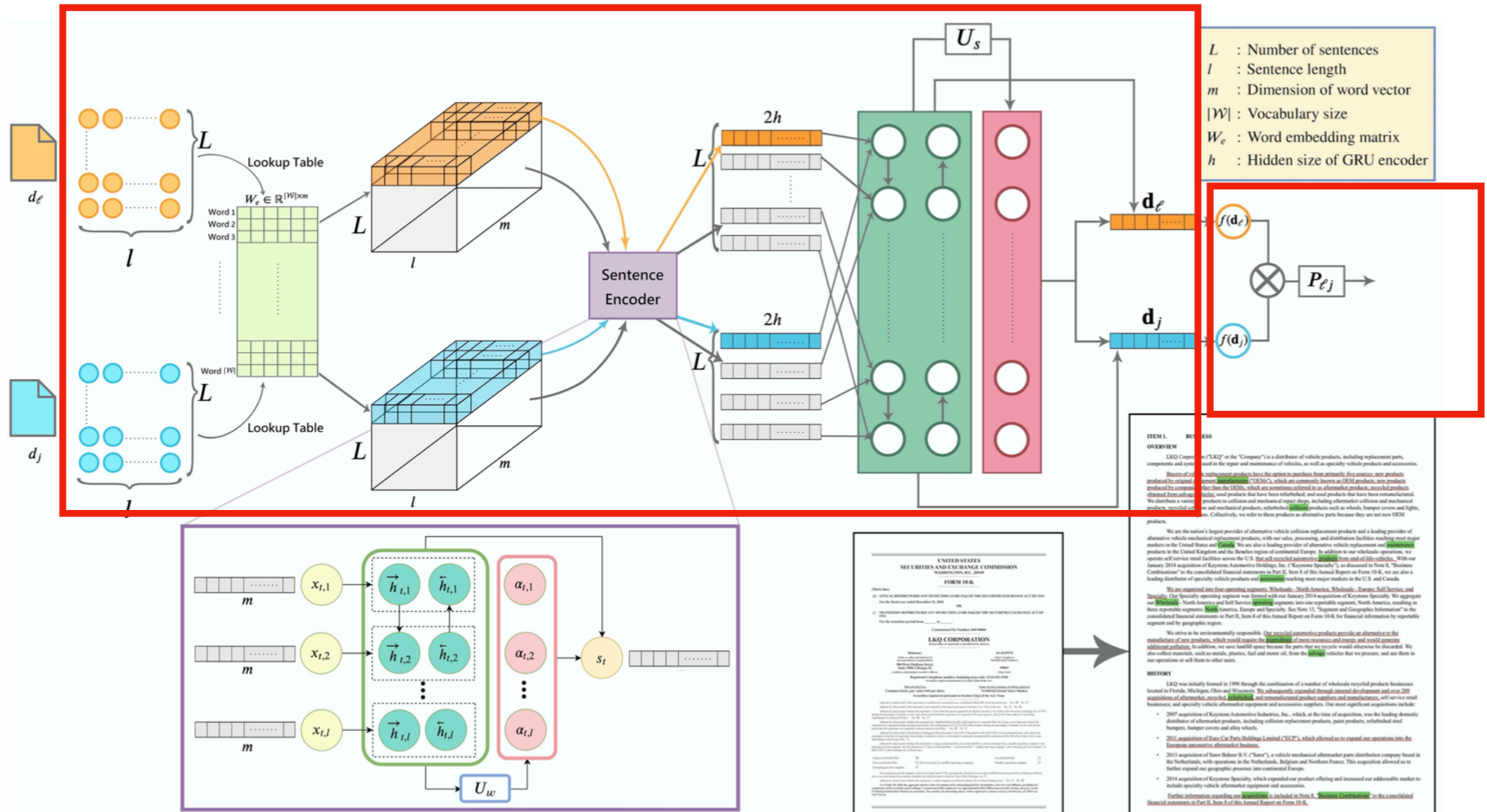
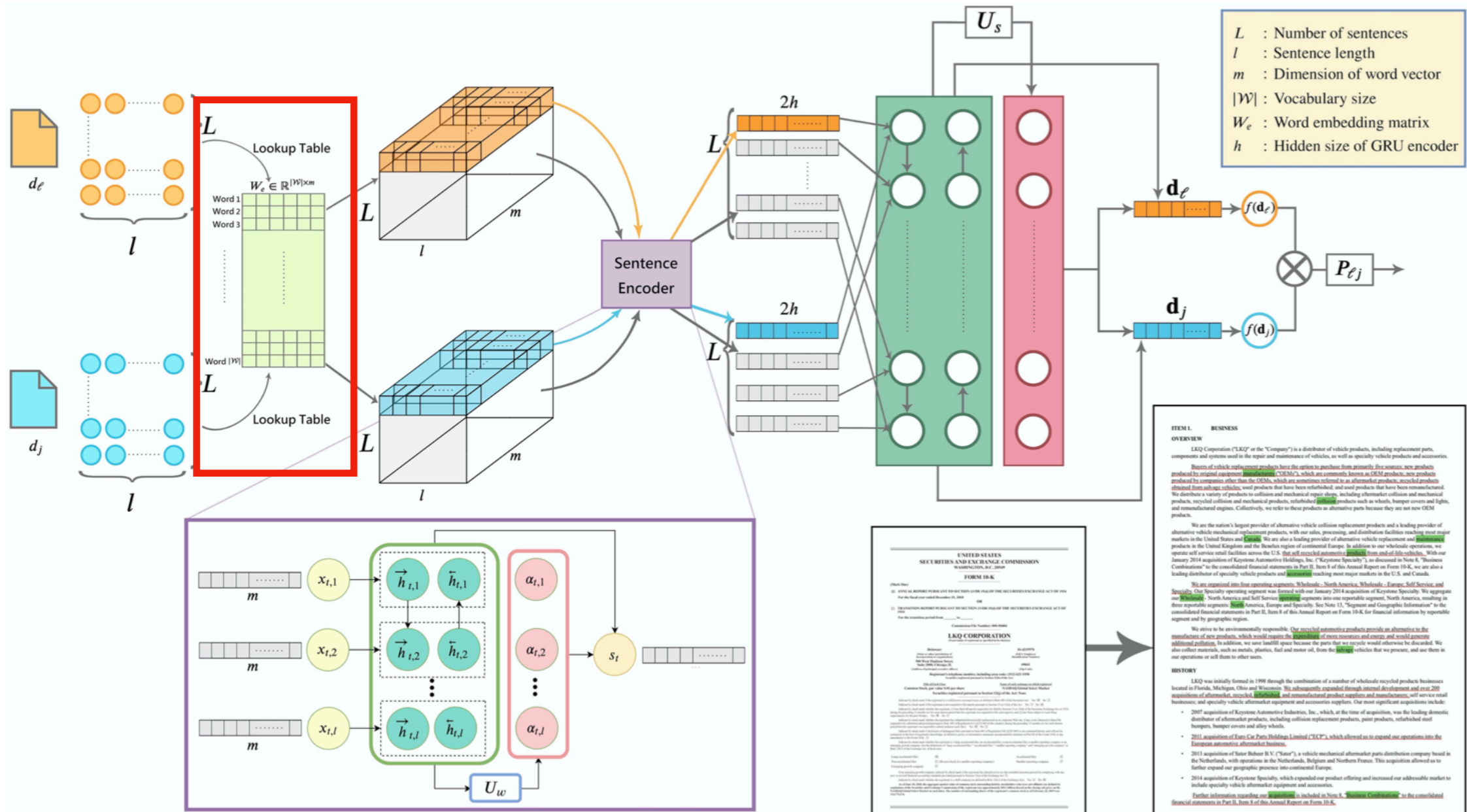- ❖ Conclusion

Introduction
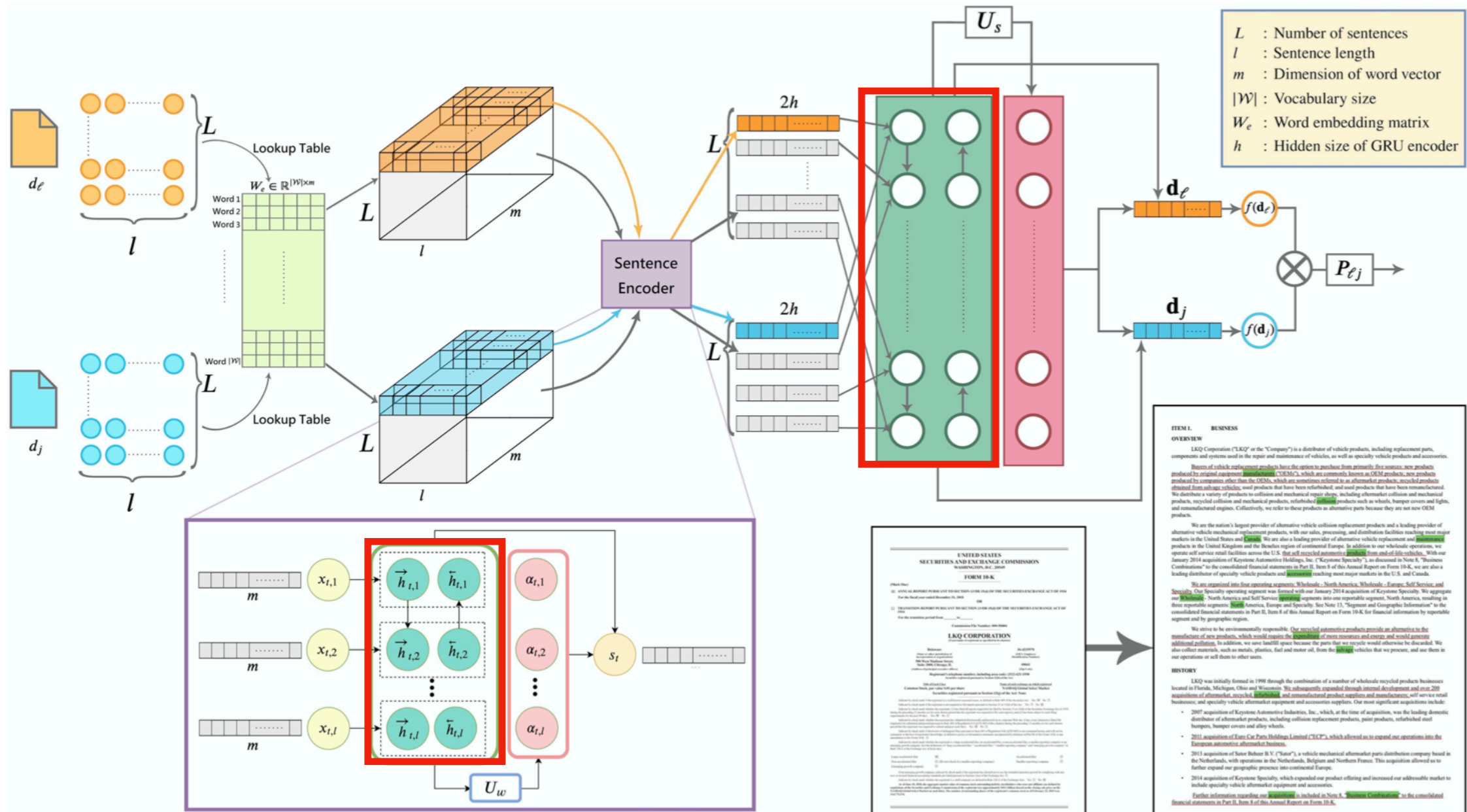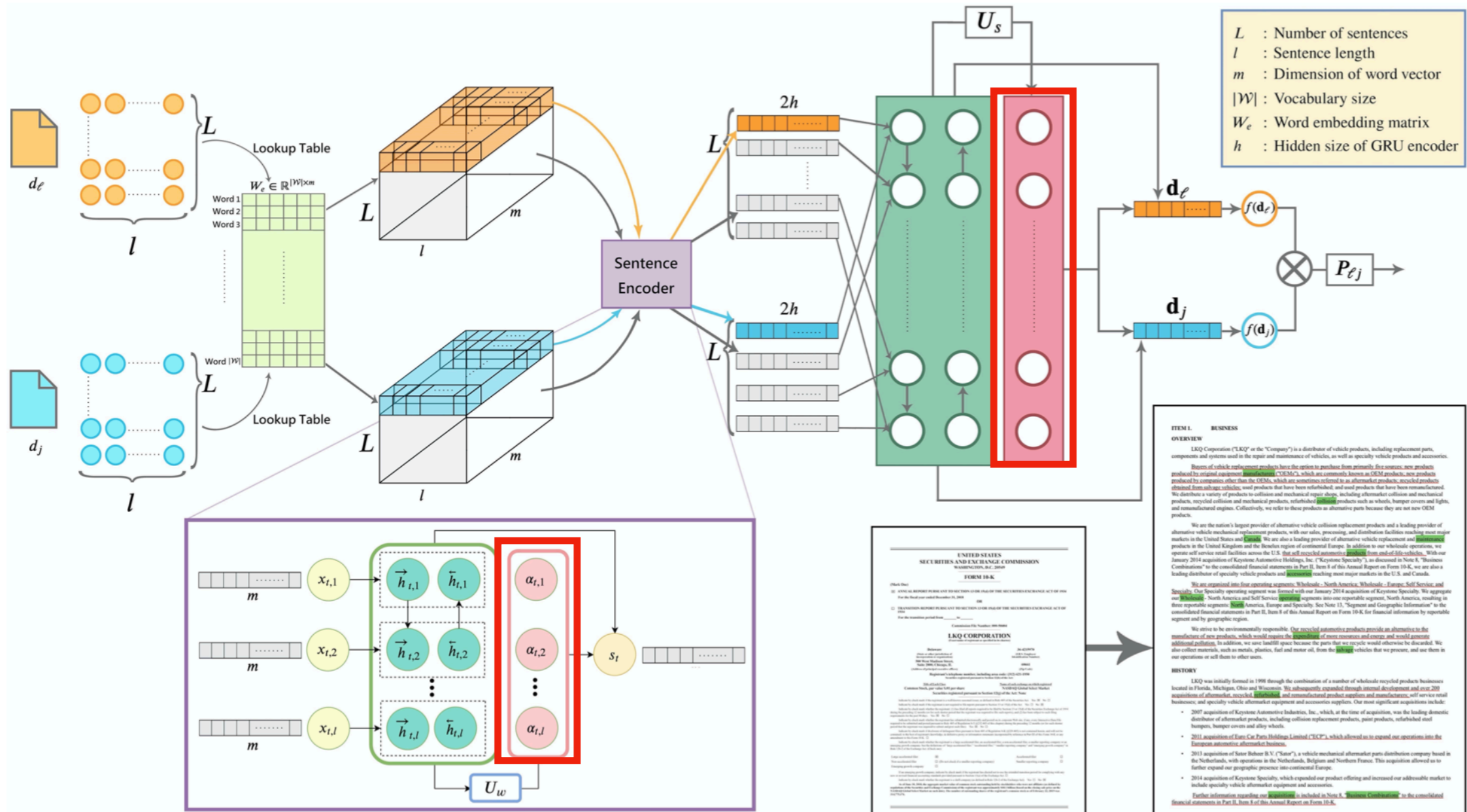
Introduction

Introduction

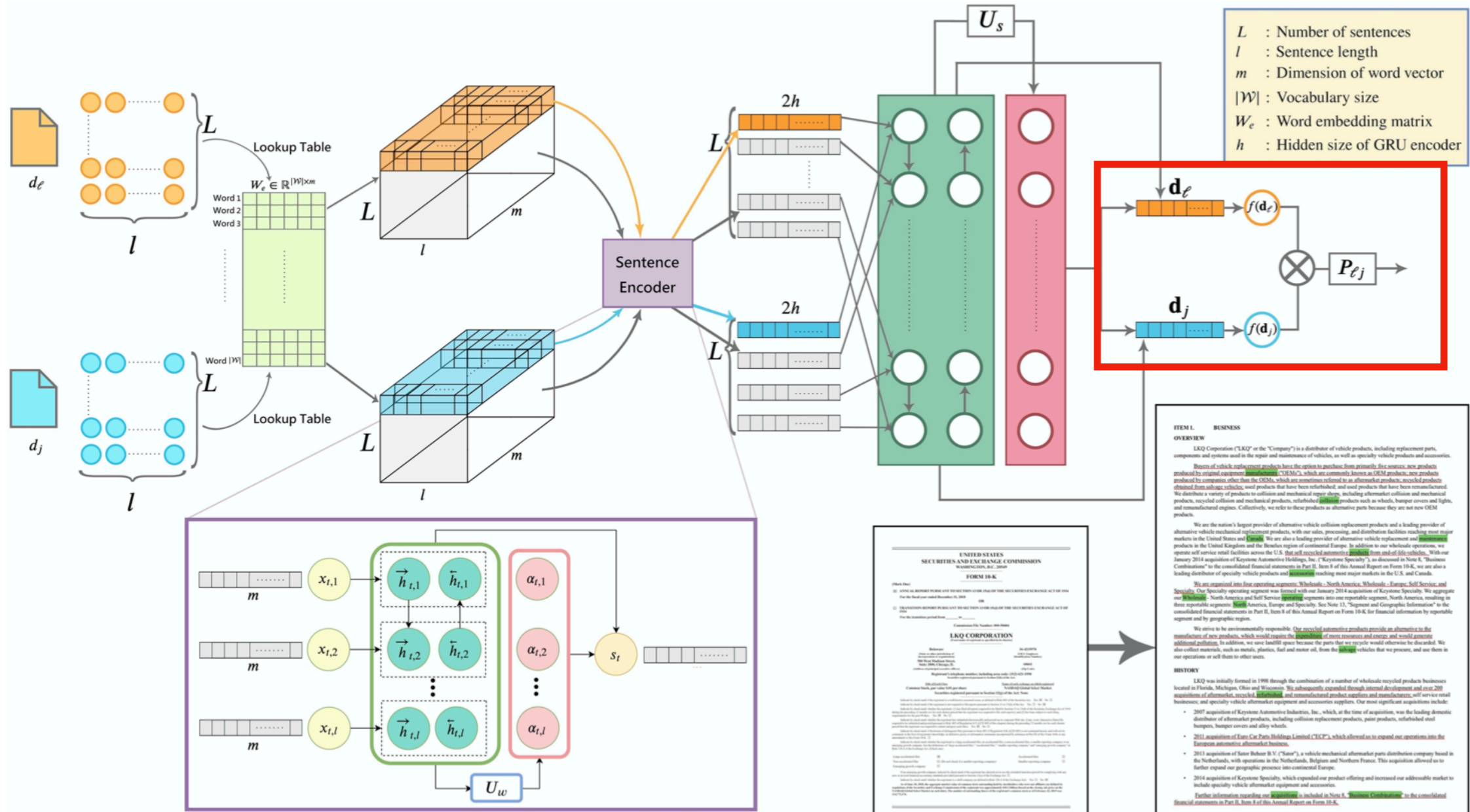# Methodology

# Methodology

Methodology

Methodology

Methodology

# Experiments

❖ Datasets: 10-K financial reports from 1996 to 2013

❖ Method: train in 5 year and test 6 year

❖ Baselines:

   ❖ Fasttext

   ❖ Hierarchical Attention Network (HAN)

   ❖ RankSVM

❖ Evaluation:

   ❖ Correlation Matrix: Kendall's Tau, Spearman's Rho

   ❖ Information Retrieval: Recall, Precision

❖ Pre-train Word Embedding:

   ❖ GloVe (G)

   ❖ BERT (B)

   ❖ Fin-Word2Vec (F)

Experiments

Metric	Method	Model	2001	2002	2003	...	2010	2011	2012	2013	Average
τ	Classification	Fasttext	0.475	0.388	0.401	...	0.449	0.460	0.452	0.463	0.426
		HAN	0.527	0.474	0.582	...	0.557	0.569	0.590	0.593	0.535
	Ranking	RankSVM	0.549	0.521	0.525	...	0.589	0.592	0.593	0.591	0.547
		XRR (G)	0.536	0.501	0.502	...	0.580	0.607	0.623	0.607	0.547
		XRR (B)	0.541	0.525	0.518	...	0.591	0.616	0.632	0.625	0.559
		XRR (F)	**0.570**	**0.541**	**0.553**	...	**0.605**	**0.616**	**0.637**	**0.629**	**0.573***
ρ	Classification	Fasttext	0.589	0.493	0.506	...	0.573	0.583	0.568	0.585	0.540
		HAN	0.648	0.587	0.599	...	0.690	0.702	0.720	0.727	0.661
	Ranking	RankSVM	0.685	0.657	0.661	...	0.733	0.733	0.731	0.732	0.686
		XRR (G)	0.671	0.632	0.636	...	0.720	0.750	0.762	0.748	0.684
		XRR (B)	0.675	0.659	0.657	...	0.732	0.756	0.772	0.766	0.697
		XRR (F)	**0.702**	**0.675**	**0.691**	...	**0.749**	**0.760**	**0.773**	**0.768**	**0.711***

Notation $*$ denotes significance compared to the best baseline under a permutation test with $p < 0.05$.
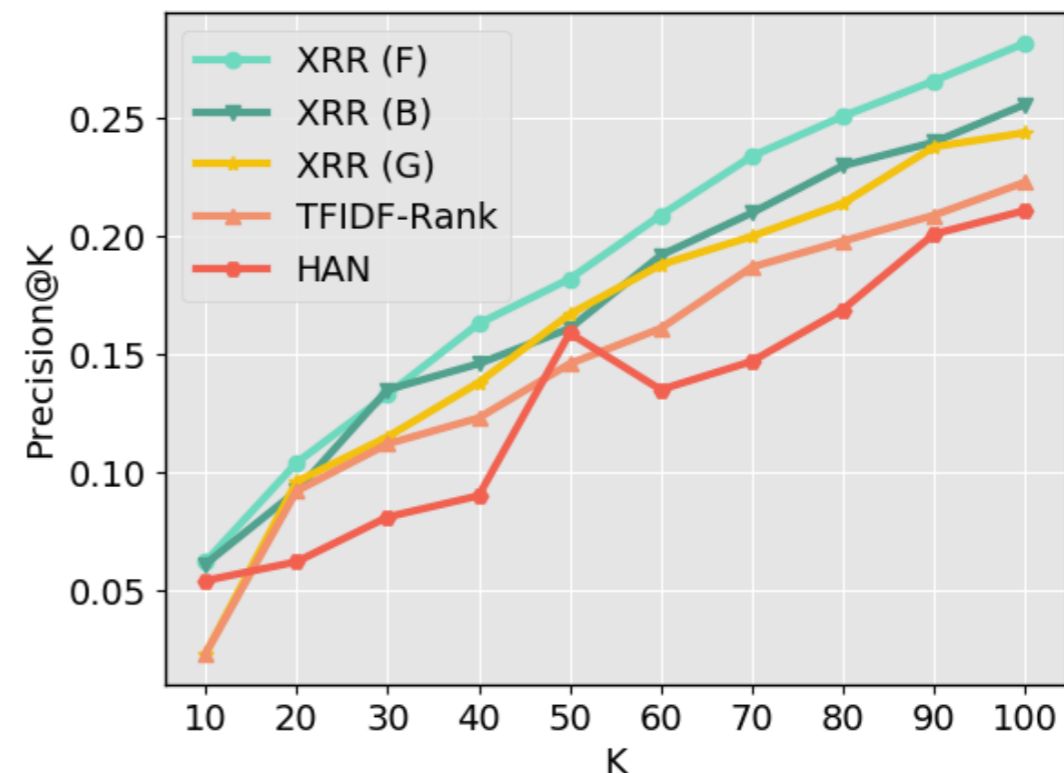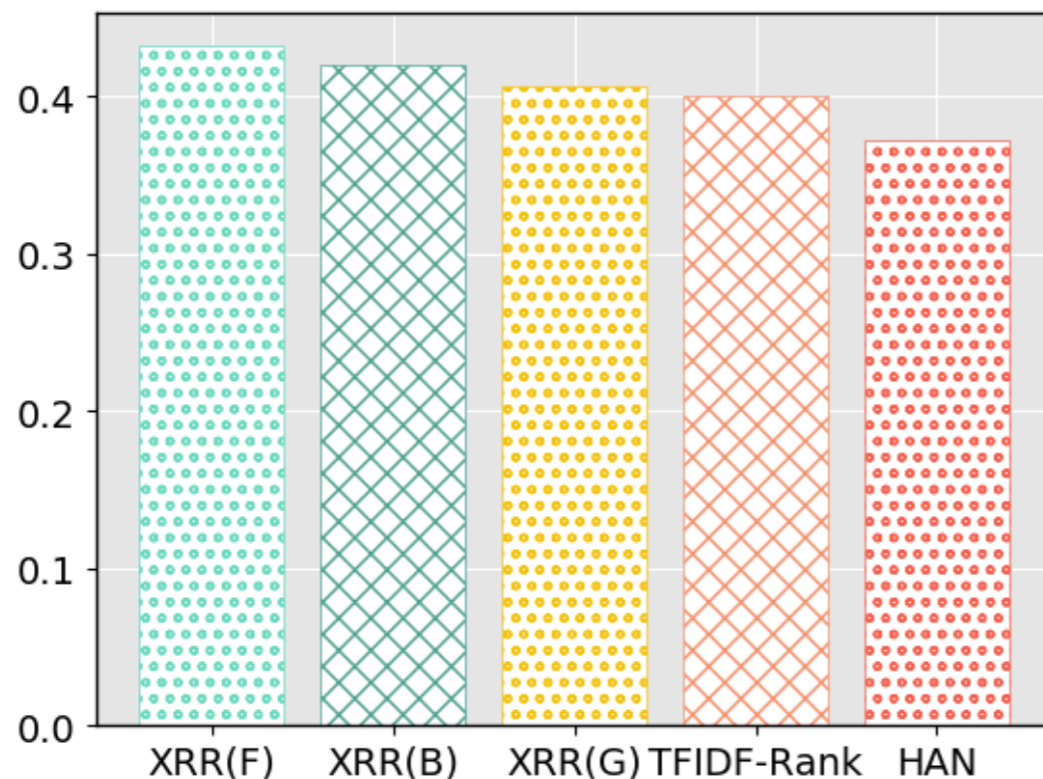
Financial Analysis

Variables	Rank 1	Rank 2	Rank 3	Rank 4	Rank 5
Firm Size	8.5052	7.8410	6.9821	6.1892	5.7281

Financial Analysis

(a) τ

(b) ρ
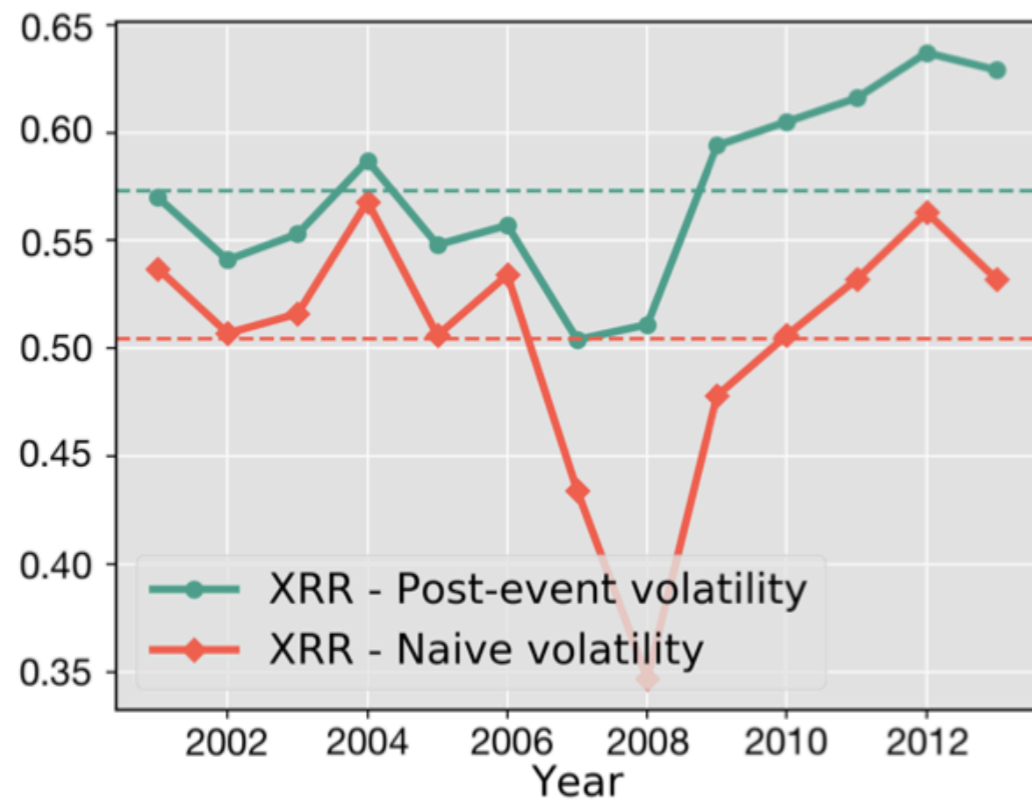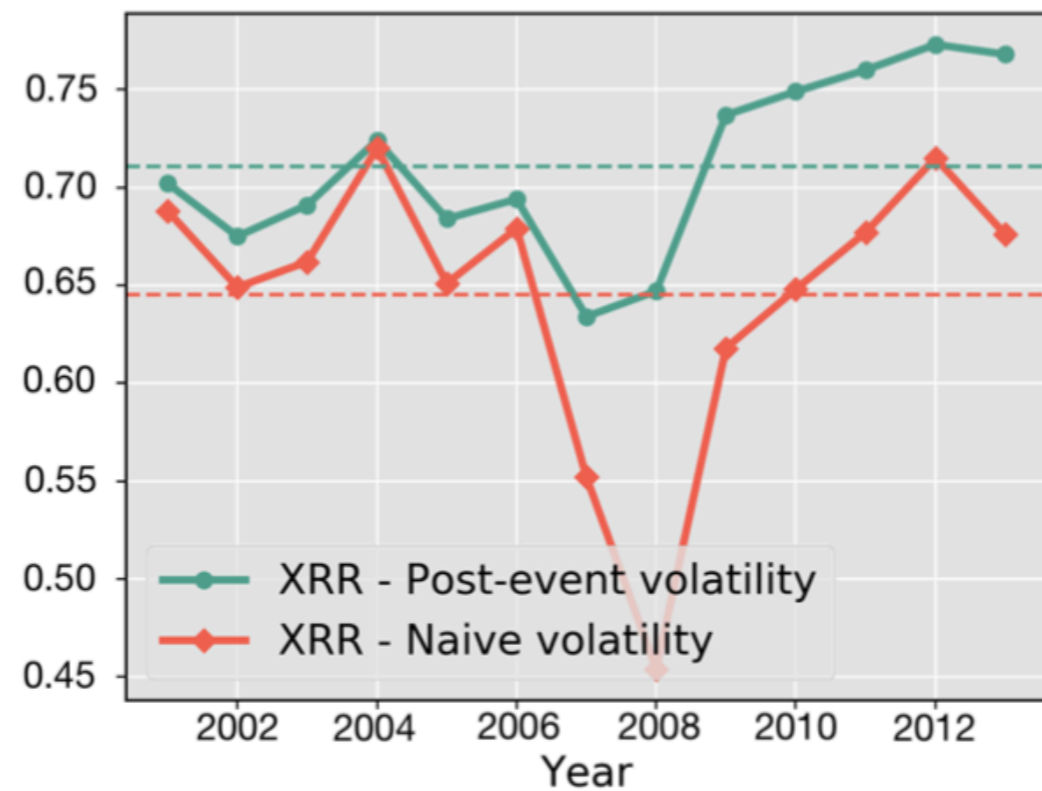
Discussion
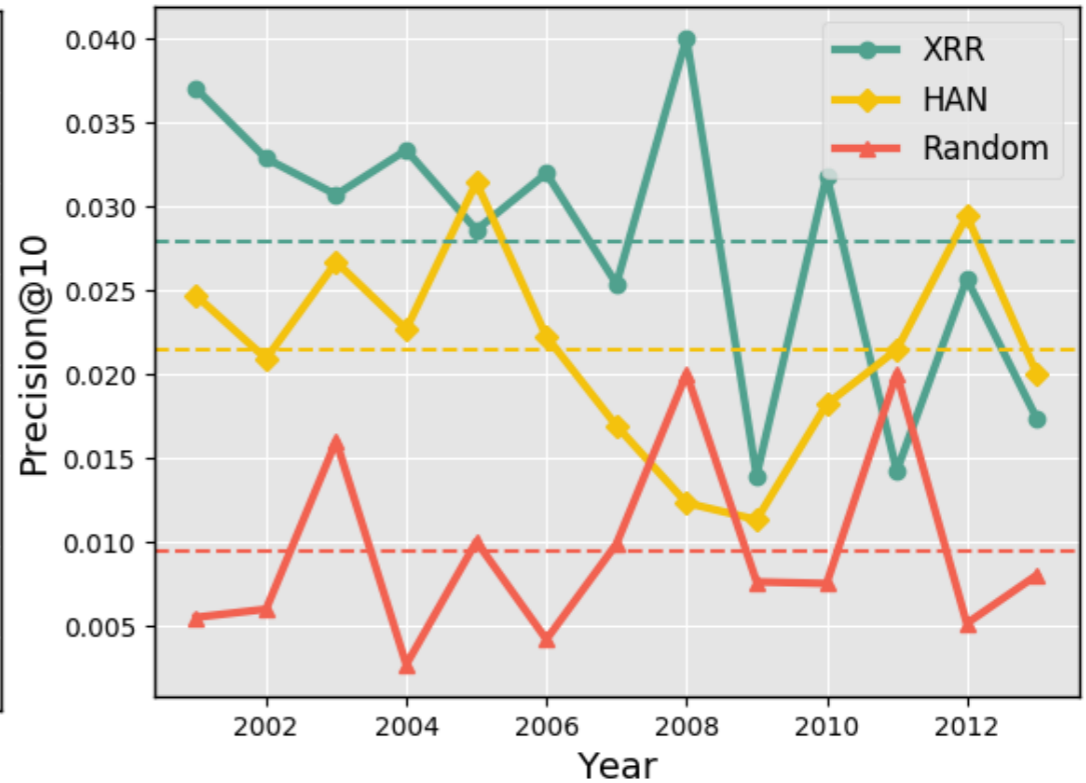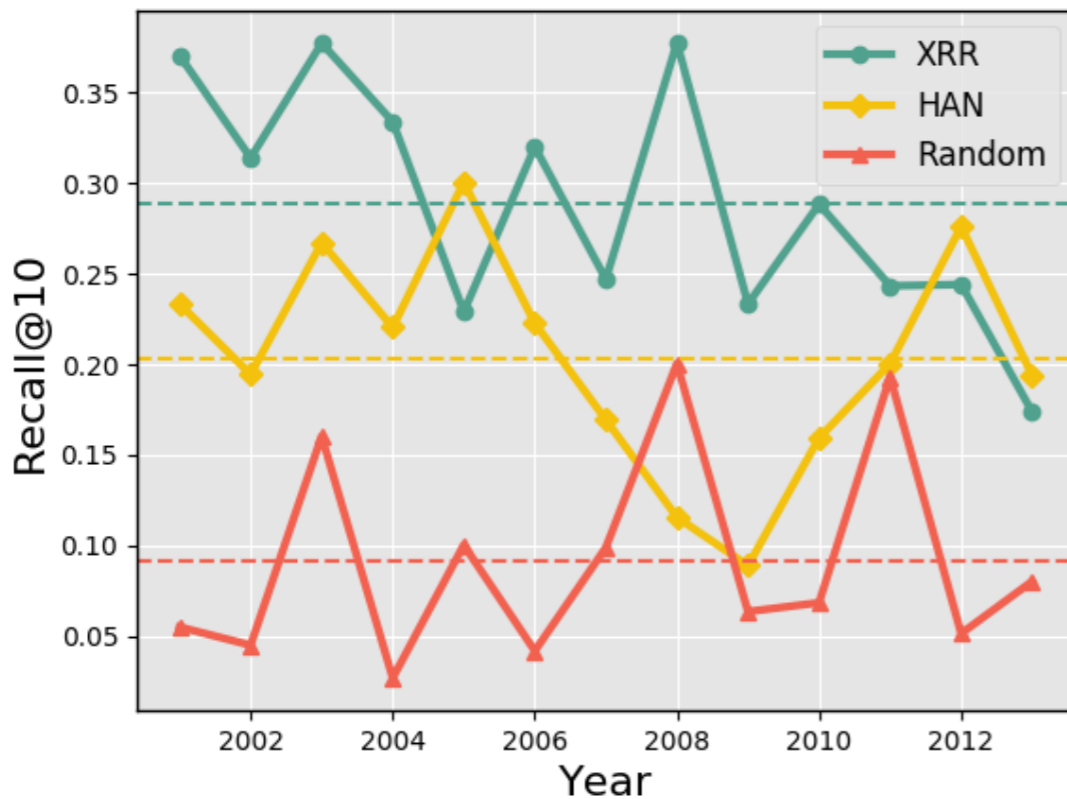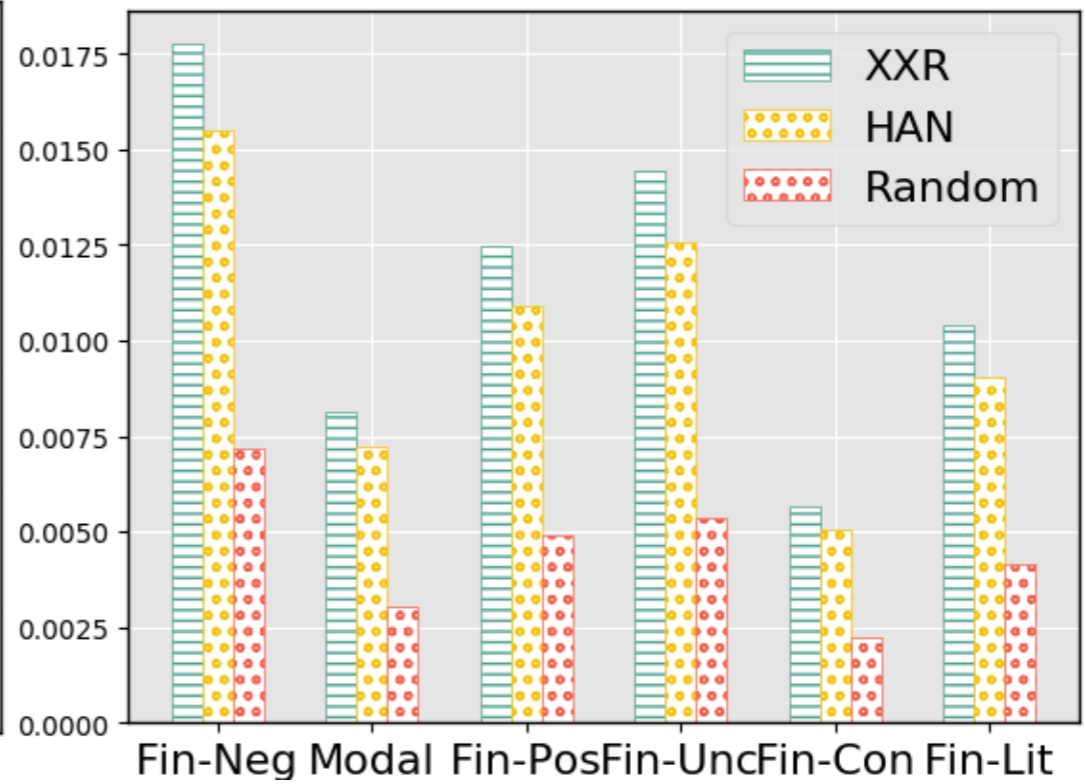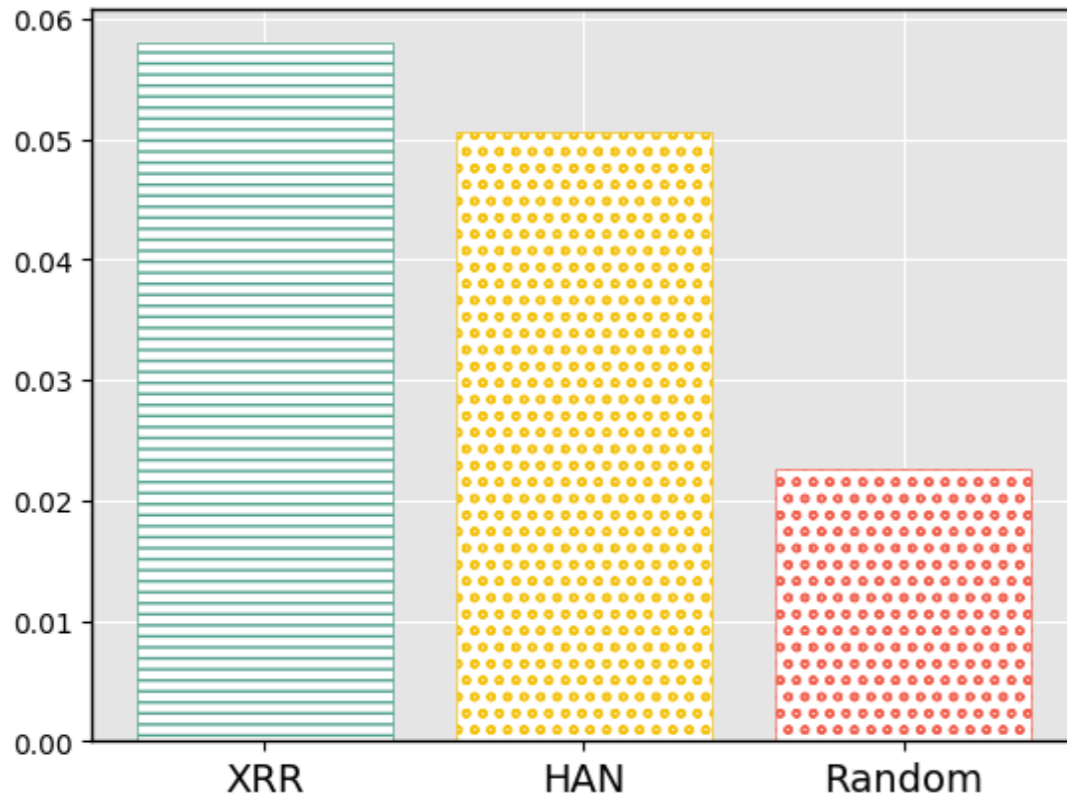
Conclusion

1. We propose XRR for risk ranking with financial reports, allowing for modeling financial texts with more complex structures and highlighting important information at both the word and sentence levels.

2. We propose using the post-event return volatility as a risk proxy for such text analytic tasks, and our experiments also attest the appropriateness of the proxy for the tasks.

3. We conduct extensive experiments and analyses on a large collection of financial reports, the results of which attest the effectiveness of the proposed method in terms of both ranking performance and explainability.

Thanks for Listening

Q&A