A Compare-and-contrast Multistage Pipeline for Uncovering Financial Signals in Financial Reports

Jia-Huei Ju[1], Yu-Shiang Huang[1,3], Cheng-Wei Lin[1], Che Lin[2,3], and Chuan-Ju Wang[1,3]

[1]Academia Sinica(AS), [2]National Taiwan University(NTU), [3]Graduate Program of Data Science, NTU & AS

Motivation

(*Form 10-K is an annual report and required by U.S. Securities and Exchange Commission.)

An empirical problem

- Analyzing financial reports (e.g., Form 10-K report*) requires lots of efforts.
- In fact, texts considered as signal is **extremely fewer** than those considered as unimportant.
- Reviewing financial reports requires **finance-specific** knowledge but also the **company-specific** understanding.

Thus, we propose

- A **compare-and-contrast pipeline** to tackle such empirical problems for financial applications.
- An evaluable **financial signal highlighting task** with datasets and evaluation measurements.

(a) Segment pairs in $\mathcal{T}^{\beta}$	
2017 (ref.)	*Our most critical accounting policies relate to revenue recognition, inventory, pension and other post-retirement benefit costs, goodwill, ...*
2018 (target)	*Our most critical accounting policies relate to revenue recognition, inventory, pension and other post-retirement benefit costs, goodwill, ...*
(b) Segment pairs in $\mathcal{T}^{\alpha}$	
2017 (ref.)	*Net sales in the Americas **increased 5%**, or $201.8 million, to $4,302.9 million.*
2018 (target)	*Net sales in the Americas **decreased 1%**, or $58.5 million, to $4,513.8 million.*

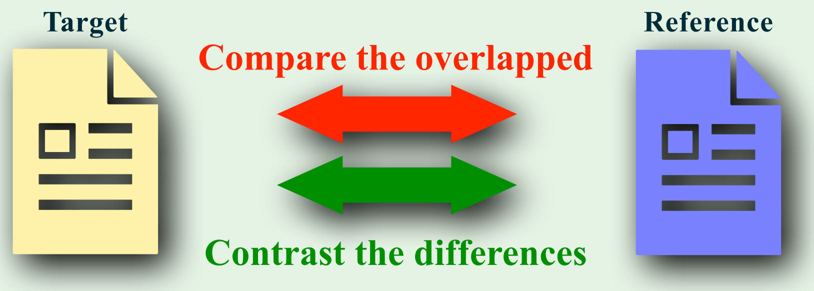
The Reference-to-target structure (of year-to-year financial reports)

- Consider the report of interest as the target, and its last year report as the reference.
- Break down the two reports into **multiple target-to-reference segment pairs**.
- The relationships of each pairs can be classified into:
  - Insignificant relation ($\mathcal{T}^{\beta}$); Revised relations ($\mathcal{T}_1^{\alpha}$) and Mismatched relations ($\mathcal{T}_2^{\alpha}$)

Financial Signal Highlighting

Definition of highlighting tasks

- Predict the **rationales** based on the reference-to-target pairs.
- Rationales indicate the predicted **word importance of a target segment** t **conditioned on reference segment** r as $\mathbf{R} \leftarrow P_f(t\,|\,r), \quad (r,t) \in \{\mathcal{T}_1^{\alpha} \cup \mathcal{T}_2^{\alpha}\}$



Human annotations

- Hired annotators labeled the important words of 200 revised and 200 mismatched pairs.

Automatic evaluations

- R-Prec: measure the precision under the truncations; the truncation is the amount of annotated signals.
- PCC (Pearson's Correlation Coefficient): correlation between predicted word importance and annotations.

The Compare-and-contrast Multistage Pipeline

0. Document Segmentation

- Use cross-segment BERT to break documents into segments

1. Relation Recognition

- Calculate pairwise text similarity based on syntactic and semantics.
- Classify them into **revised and mismatched relations**

2. Signal Highlighting — Out-of-domain Fine-tuning

- Fine-tune on e-SNLI contradicted pairs as the **Zero-shot** highlighter f
- Recast the highlighting task into **binary token classification** task
- Contextualized representation of a reference-to-target pair:

$$h_{(r,t)} = \mathrm{BERT}([\mathrm{CLS}]\, r\, [\mathrm{SEP}]\, t)$$

2+. Signal Highlighting — In-domain Fine-tuning

- Fine-tune on the **hard and soft pseudo-labels** using
  - Hard labels: the **revised words** as labels (CrossEntropy)
  - Soft labels: the **probabilities** of Zero-shot model's prediction (KL Divergence)

$$L_{CE} = \sum_j -\left(Y_t^j \log P_f^j(t\,|\,r)\right) + \left(1 - Y_t^j\right)\log\left(1 - P_f^j(t\,|\,r)\right)$$

$$L_{KL} = -\sum_j KL\left(P_f^j(t\,|\,r) \| P_{f+}^j(t\,|\,r)\right)$$

- Fine-tune with **the warmed-up Zero-shot** highlighter for the final **domain-adaptive** highlighter f^+

Empirical Evaluation

Datasets

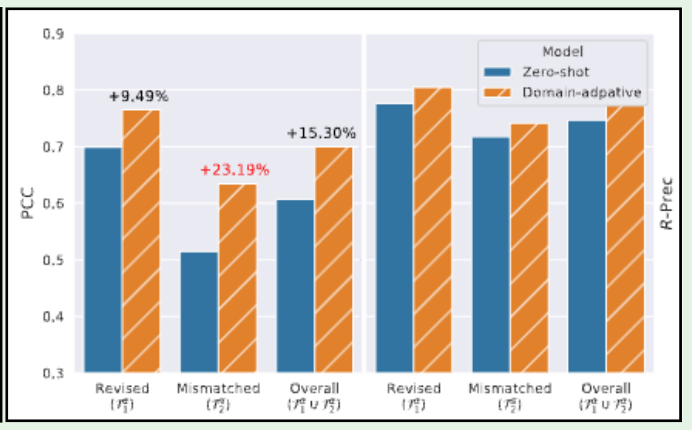
- 400 pairs of our released FINAL Eval set.
- 3,237 pairs of e-SNLI contradicted Test sets.

Results (of our domain-adaptive highlighter)

- Better performance on two types of pairs in FINAL
- Retrain generalization capability in e-SNLI
- Improve even more on unseen relation (the mismatched pairs)

#	W.U.	Labeling		FINAL		e-SNLI$_c$	
		P	S	*R*-Prec	PCC	*R*-Prec	PCC
Zero-Shot							
1	✓	✗	✗	0.7469	0.6067	0.8565	0.7555
Pseudo few-shot							
2	✗	✓	✗	0.6968	0.6368	0.6302	0.5752
Domain-adaptive							
3	✓	✓	✗	0.7160	0.6555	0.8475	0.7305
4	✓	✓	✓	**0.7865***	**0.7290***	**0.8605**	**0.7566**

Table 3: Highlighting performance



The 61st Annual Meeting of the Association for Computational Linguistics (ACL'23)