Improving word mover's distance by leveraging self-attention matrix

Hiroaki Yamagiwa[1], Sho Yokoi[2,3], Hidetoshi Shimodaira[1,3]
[1]Kyoto University [2]Tohoku University [3]RIKEN AIP

EMNLP 2023

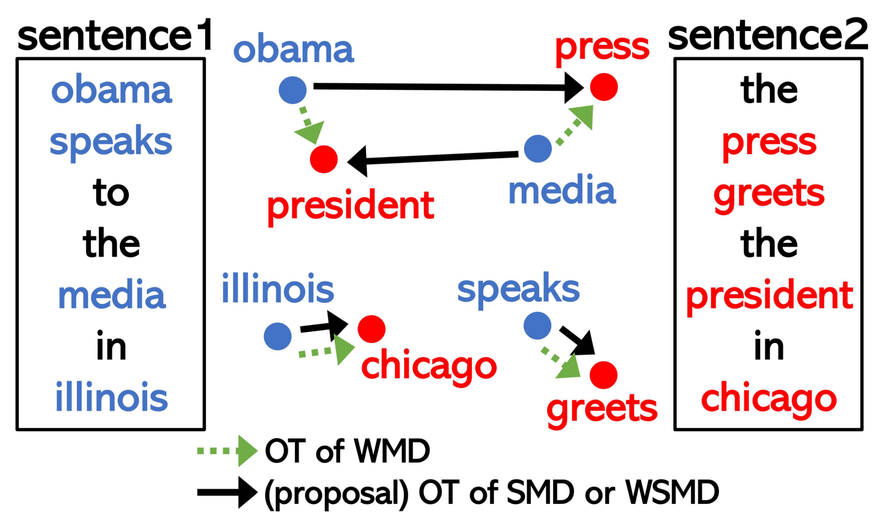
arXiv

Background

- **Word Mover's Distance** (**WMD**) [1] uses the **Wasserstein distance** to measure semantic textual similarity.
- WMD cannot address the order of words within a sentence.

Approach

- Use the **Self-Attention Matrix** (**SAM**) from BERT-based models as **structure information**.
- Propose a novel method that combines **WMD** and **SAM** using the **Fused Gromov-Wasserstein** distance [2].

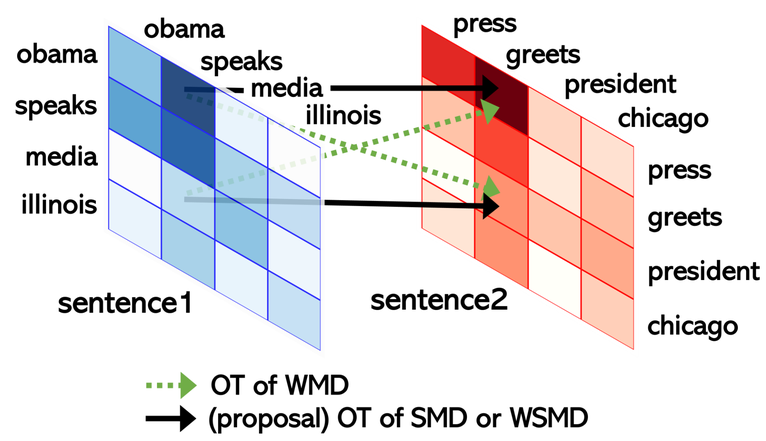


Results

- The method improved the performance of WMD-like methods in a paraphrase identification task [3].

Methods

Self-Attention Matrix (SAM)



Both pairs, (obama, speaks) and (press, greets), show a high attention weight.

Proposed Method

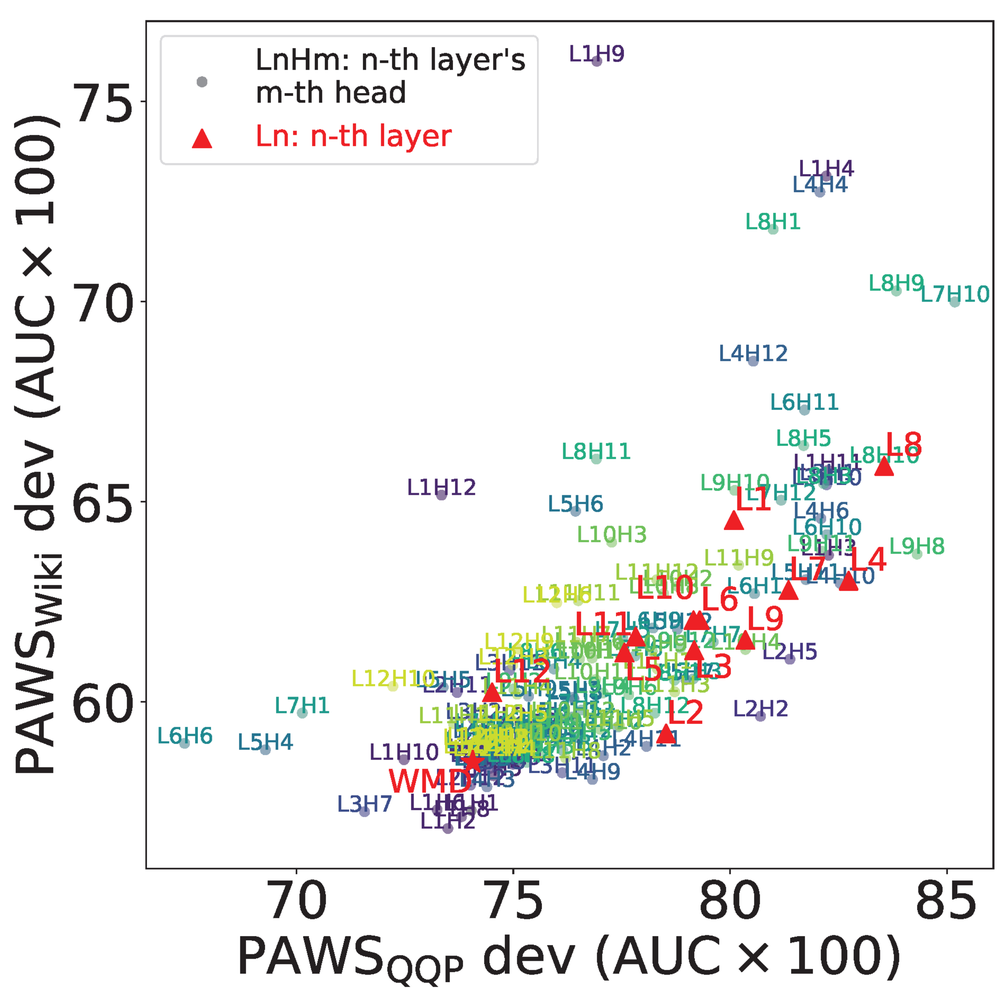
Let A and A' be the **SAM**s for sentences s and s'. Define the **Word and sentence Structure Mover's Distance** (**WSMD**) as follows:

$$\mathrm{WSMD}\big((s,A),(s',A')\big) = \min_{\boldsymbol{P}\in\boldsymbol{\Pi}(\mathrm{u},\mathrm{u}')} \sum_{i,j,i',j'} \Big\{(1-\lambda)C_{ij} + \lambda\, k\big|A_{ii'} - A'_{jj'}\big|^2\Big\} P_{ij}P_{i'j'}$$

Normalization parameter

Wasserstein — Gromov-Wasserstein — Fused Gromov-Wasserstein

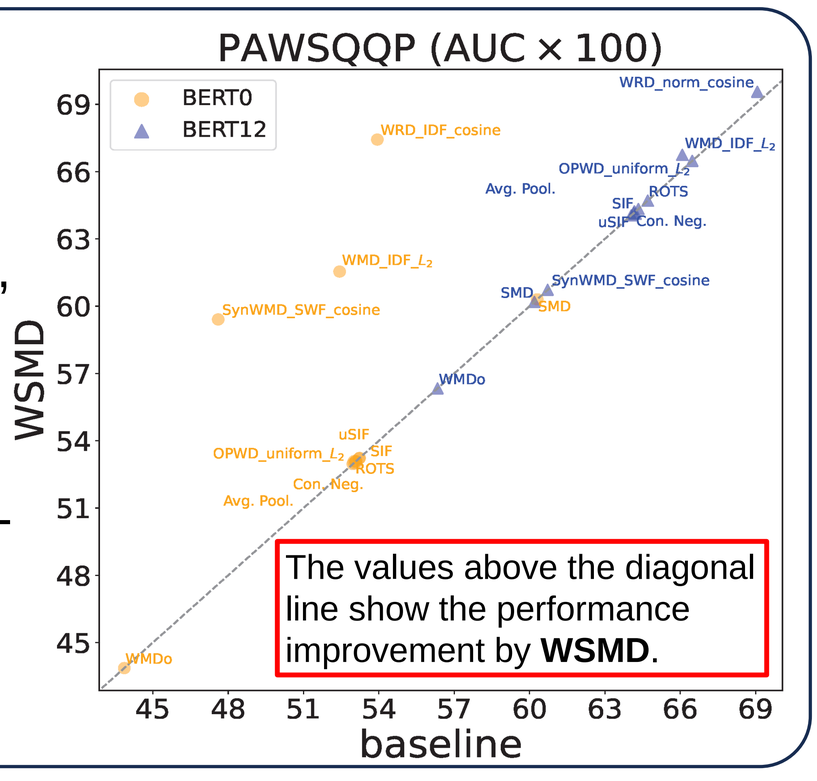
Head Selection



- The performance varies significantly among different attention heads.
- By selecting layers and using the heads within them, we average the **WSMD** to evaluate the performance.

Experiments

- For paraphrase identification, we used the PAWS [3] dataset, which contains sentence pairs with high word overlap.
- **WSMD** was effective for WMD-like methods such as WMD, WRD, and SynWMD.



References

1. Matt J. Kusner, Yu Sun, Nicholas I. Kolkin, and Kilian Q. Weinberger. 2015. *From word embeddings to document distances*. ICML.
2. Titouan Vayer, Nicolas Courty, Romain Tavenard, Laetitia Chapel, and Rémi Flamary. 2019. *Optimal transport for structured data with application on graphs*. PMLR.
3. Yuan Zhang, Jason Baldridge, and Luheng He. 2019. *PAWS: Paraphrase adversaries from word scrambling*. NAACL.