cases	doc_1		doc_2		decision	id
	authors	• Hidetoshi Shimodaira • Sho Yokoi • Hiroaki Yamagiwa	**authors**	• Hiroaki Yamagiwa • Sho Yokoi • Hidetoshi Shimodaira	DUPLICATES	143
	title	Improving word mover's distance by leveraging self-attention matrix	**title**	Improving word mover's distance by leveraging self-attention matrix		
	publication_date	2022-11-11 00:00:00	**publication_date**	2022-11-11 00:00:00		
	source	SupportedSources.PAPERS_WITH_CODE	**source**	SupportedSources.SEMANTIC_SCHOLAR		
	journal		**journal**	ArXiv		
	volume		**volume**	abs/2211.06229		
	doi		**doi**	10.48550/arXiv.2211.06229		
	urls	• https://arxiv.org/pdf/2211.06229v1.pdf • https://github.com/ymgw55/WSMD	**urls**	• https://www.semanticscholar.org/paper/6ea52259479f38d59b75e46948518afa83e9c33c		
	id	id-5506090255828225560	**id**	id-2310370535735821246		
	abstract	Measuring the semantic similarity between two sentences is still an important task. The word mover's distance (WMD) computes the similarity via the optimal alignment between the sets of word embeddings. However, WMD does not utilize word order, making it difficult to distinguish sentences with large overlaps of similar words, even if they are semantically very different. Here, we attempt to improve WMD by incorporating the sentence structure represented by BERT's self-attention matrix (SAM). The proposed method is based on the Fused Gromov-Wasserstein distance, which simultaneously considers the similarity of the word embedding and the SAM for calculating the optimal transport between two sentences. Experiments on paraphrase identification and semantic textual similarity show that the proposed method improves WMD and its variants. Our code is available at https://github.com/ymgw55/WSMD.	**abstract**	Measuring the semantic similarity between two sentences is still an important task. The word moverâ€™s distance (WMD) computes the similarity via the optimal alignment between the sets of word embeddings. However, WMD does not utilize word order, making it difficult to distinguish sentences with large overlaps of similar words, even if they are semantically very different. Here, we attempt to improve WMD by incorporating the sentence structure represented by BERTâ€™s self-attention matrix (SAM). The proposed method is based on the Fused Gromov-Wasserstein distance, which simultaneously considers the similarity of the word embedding and the SAM for calculating the optimal transport between two sentences. Experiments on paraphrase identification and semantic textual similarity show that the proposed method improves WMD and its variants. Our code is available at https://github. com/ymgw55/WSMD.		
	versions		**versions**			