cases	doc_1		doc_2		decision	id
	authors	• Lingfei Wu • I. E. Yen • Kun Xu • Fangli Xu • Avinash Balakrishnan • Pin-Yu Chen • Pradeep Ravikumar • M. Witbrock	**authors**	• Lingfei Wu • Ian E. H. Yen • Kun Xu • Fangli Xu • Avinash Balakrishnan • Pin-Yu Chen • Pradeep Ravikumar • Michael J. Witbrock	DUPLICATES	366
	title	Word Moverâ€™s Embedding: From Word2Vec to Document Embedding	**title**	Word Mover's Embedding: From Word2Vec to Document Embedding		
	publication_date	2018-10-26 00:00:00	**publication_date**	2018-10-30 19:43:17+00:00		
	source	SupportedSources.SEMANTIC_SCHOLAR	**source**	SupportedSources.ARXIV		
	journal	ArXiv	**journal**	None		
	volume	abs/1811.01713	**volume**			
	doi	10.18653/v1/D18-1482	**doi**			
	urls	• https://www.semanticscholar.org/paper/ef56afa4cefb90e263b1434dd3ca650904f54c6e	**urls**	• http://arxiv.org/pdf/1811.01713v1 • http://arxiv.org/abs/1811.01713v1 • http://arxiv.org/pdf/1811.01713v1		
	id	id-8103817222637774579	**id**	id8616581425248519 66		
	abstract	While the celebrated Word2Vec technique yields semantically rich representations for individual words, there has been relatively less success in extending to generate unsupervised sentences or documents embeddings. Recent work has demonstrated that a distance measure between documents called Word Moverâ€™s Distance (WMD) that aligns semantically similar words, yields unprecedented KNN classification accuracy. However, WMD is expensive to compute, and it is hard to extend its use beyond a KNN classifier. In this paper, we propose the Word Moverâ€™s Embedding (WME), a novel approach to building an unsupervised document (sentence) embedding from pre-trained word embeddings. In our experiments on 9 benchmark text classification datasets and 22 textual similarity tasks, the proposed technique consistently matches or outperforms state-of-the-art techniques, with significantly higher accuracy on problems of short length.	**abstract**	While the celebrated Word2Vec technique yields semantically rich representations for individual words, there has been relatively less success in extending to generate unsupervised sentences or documents embeddings. Recent work has demonstrated that a distance measure between documents called \emph{Word Mover's Distance} (WMD) that aligns semantically similar words, yields unprecedented KNN classification accuracy. However, WMD is expensive to compute, and it is hard to extend its use beyond a KNN classifier. In this paper, we propose the \emph{Word Mover's Embedding } (WME), a novel approach to building an unsupervised document (sentence) embedding from pre-trained word embeddings. In our experiments on 9 benchmark text classification datasets and 22 textual similarity tasks, the proposed technique consistently matches or outperforms state-of-the-art techniques, with significantly higher accuracy on problems of short length.		
	versions		**versions**			