

cases	doc_1		doc_2		decision	id
					DUPLICATES	374
	authors	<ul style="list-style-type: none">Yunchao WeiYao ZhaoZ. ZhuShikui WeiYanhui XiaoJiashi FengShuicheng Yan	authors	<ul style="list-style-type: none">Yunchao WeiYao ZhaoZhenfeng ZhuShikui WeiYanhui XiaoJiashi FengShuicheng Yan		
	title	Modality-Dependent Cross-Media Retrieval	title	Modality-dependent Cross-media Retrieval		
	publication_date	2015-06-22 00:00:00	publication_date	2015-06-22 14:33:39+00:00		
	source	SupportedSources.SEMANTIC_SCHOLAR	source	SupportedSources.ARXIV		
	journal	ACM Transactions on Intelligent Systems and Technology (TIST)	journal	None		
	volume	7	volume			
	doi	10.1145/2775109	doi			
	urls	<ul style="list-style-type: none">https://www.semanticscholar.org/paper/11b9c878142f97897686aa15eb5f497883c9dd11	urls	<ul style="list-style-type: none">http://arxiv.org/pdf/1506.06628v2http://arxiv.org/abs/1506.06628v2http://arxiv.org/pdf/1506.06628v2		
	id	id45365066602409991	id	id7931685561672651779		
	abstract	In this article, we investigate the cross-media retrieval between images and text, that is, using image to search text (I2T) and using text to search images (T2I). Existing cross-media retrieval methods usually learn one couple of projections, by which the original features of images and text can be projected into a common latent space to measure the content similarity. However, using the same projections for the two different retrieval tasks (I2T and T2I) may lead to a tradeoff between their respective performances, rather than their best performances. Different from previous works, we propose a modality-dependent cross-media retrieval (MDCR) model, where two couples of projections are learned for different cross-media retrieval tasks instead of one couple of projections. Specifically, by jointly optimizing the correlation between images and text and the linear regression from one modal space (image or text) to the semantic space, two couples of mappings are learned to project images and text from their original feature spaces into two common latent subspaces (one for I2T and the other for T2I). Extensive experiments show the superiority of the proposed MDCR compared with other methods. In particular, based on the 4,096-dimensional convolutional neural network (CNN) visual feature and 100-dimensional Latent Dirichlet Allocation (LDA) textual feature, the mAP of the proposed method achieves the mAP score of 41.5%, which is a new state-of-the-art performance on the Wikipedia dataset.	abstract	In this paper, we investigate the cross-media retrieval between images and text, i.e., using image to search text (I2T) and using text to search images (T2I). Existing cross-media retrieval methods usually learn one couple of projections, by which the original features of images and text can be projected into a common latent space to measure the content similarity. However, using the same projections for the two different retrieval tasks (I2T and T2I) may lead to a tradeoff between their respective performances, rather than their best performances. Different from previous works, we propose a modality-dependent cross-media retrieval (MDCR) model, where two couples of projections are learned for different cross-media retrieval tasks instead of one couple of projections. Specifically, by jointly optimizing the correlation between images and text and the linear regression from one modal space (image or text) to the semantic space, two couples of mappings are learned to project images and text from their original feature spaces into two common latent subspaces (one for I2T and the other for T2I). Extensive experiments show the superiority of the proposed MDCR compared with other methods. In particular, based the 4,096 dimensional convolutional neural network (CNN) visual feature and 100 dimensional LDA textual feature, the mAP of the proposed method achieves 41.5%, which is a new state-of-the-art performance on the Wikipedia dataset.		
	versions		versions			