Discrete Semantic Matrix Factorization Hashing for Cross-Modal Retrieval

Jianyang Qin, Lunke Fei, Shaohua Teng, Wei Zhang, Genping Zhao, Haoliang Yuan

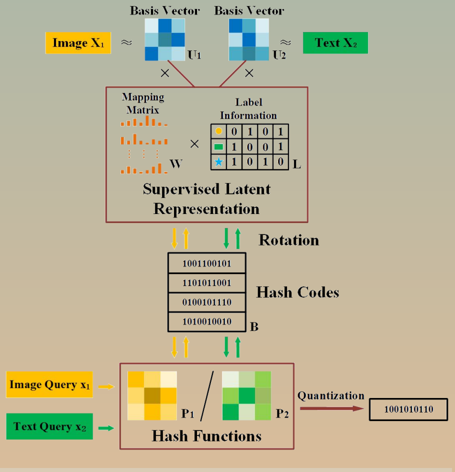
School of Computer Science and Technology, Guangdong University of Technology, Guangzhou, China

ICPR 2020
25th INTERNATIONAL CONFERENCE ON PATTERN RECOGNITION
Milan, Italy 10 | 15 January 2021

INTRODUCTION

Cross-modal retrieval is becoming a trend because of the tremendous explosion of multimedia information. For example, Twitter needs to retrieve both the relevant images and text from a query word or image. Due to the efficiency and effectiveness, hashing has been a promising technique for cross-modal retrieval in massive data analysis. In this paper, we propose a new supervised hashing method, namely, Discrete Semantic Matrix Factorization Hashing (DSMFH), for cross-modal retrieval.

- ***First, we conduct the collective matrix factorization via directly using the available label information to obtain a latent representation***
- ***Then, we simultaneously learn the hash codes and corresponding hash functions by deriving the collective matrix factorization into a discrete optimization.***
- ***Finally, we employ an alternatively iterative optimizing procedure to efficiently solve the collective matrix factorization and discrete learning.***



MOTIVATION

While most existing supervised cross-modal hashing methods have achieved encouraging performance in many scenarios, there usually exists two limitations.

- ***How to efficiently and effectively utilize the available label information for more accurate and reliable cross-modality retrieval needs to be further investigated.***
- ***How to reduce accumulated errors during hashing learning remains a central and challenging problem.***

MODEL

Let $X^{(t)}$ be the sets of n training samples in the t-th modalities, class labels L is available for training samples. We aim to learn the discrete hash codes B and hash functions P for cross-modal retrieval

$$\underbrace{\sum_{t=1}^{M}\lambda_t\left\|X^{(t)}-U^{(t)}WL\right\|_F^2}_{\text{Learning Supervised Latent Representation}}+\underbrace{\left\|Q^TB-WL\right\|_F^2+\left\|Q\right\|_F^2}_{\text{Learning Hash Codes}}+\underbrace{\sum_{t=1}^{M}\left\|B-P^{(t)}X^{(t)}\right\|_F^2}_{\text{Learning Hash Functions}},\quad s.t.\;\; B\in\{-1,1\}^{r\times n}$$

By embedding the labels into the collective matrix factorization, we can learn a discriminative latent representation which preserve the semantic consistency between different modalities and corresponding label information

We directly learn the discrete hash codes via the quantization of rotation technique, and the binary codes can be optimized bit by bit by the Discrete Cyclic Coordinate Descent method.

Having obtaind different hash functions, the binary code of a new query data $x'^{(t)}$ can be easily encoded by $b'^{(t)} = sign(P^{(t)}x'^{(t)})$, which is used to retrieve similar data.

We jointly learn the supervised latent subspace, hash functions and hash codes. Thus, we can not only maintain the discrete constraints to generate better hash codes, but also minimize the projection errors so that the gaps between training data and unseen data are minimized.

EXPERIMENTS

We conduct experiments on three widely used databases, including Wiki, MIRflickr25k and NUSWIDE, to evaluate the proposed method compared with the state-of-the-art cross-modal hashing methods. The experimental results on three benchmark databases demonstrate that our DSMFH method outperforms the state-of-the-art cross-modal hashing methods.

TABLE I. The mAP results of seven methods with various code lengths on the Wiki, MIRFlickr and NUS-WIDE databases

Task	Method	Wiki				Mirflick				NUS-WIDE			
		24bits	48bits	96bits	128bits	24bits	48bits	96bits	128bits	24bits	48bits	96bits	128bits
Image query Tag	CMFH	0.2093	0.2320	0.2357	0.2407	0.5580	0.5869	0.5848	0.5841	0.3909	0.3922	0.3931	0.3939
	SCM	0.1496	0.1511	0.1501	0.1509	0.5993	0.6019	0.6060	0.6073	0.4706	0.4732	0.4747	0.4713
	FSH	0.1539	0.1586	0.1636	01645	0.6114	0.6111	0.6093	0.6094	0.4803	0.4861	0.4874	0.4876
	SDMFH	0.3019	0.3307	0.3491	0.3495	0.6476	0.6628	0.6692	0.6687	0.5821	0.5812	0.6421	0.6242
	GSPH	0.2499	0.2809	0.2940	0.2823	0.6919	0.7022	0.7105	0.7118	0.5902	0.5967	0.5986	0.6067
	LCMFH	0.3512	0.3603	0.3752	0.3752	0.6681	0.6728	0.6776	0.6785	0.6225	0.6379	0.6395	0.6458
	DSMFH	**0.3553**	**0.3688**	**0.3921**	**0.3841**	**0.6923**	**0.7200**	**0.7261**	**0.7310**	**0.6236**	**0.6596**	**0.6695**	**0.6680**
Tag query Image	CMFH	0.4936	0.5246	0.5317	0.5416	0.5977	0.5959	0.5949	0.5928	0.4049	0.4109	0.4123	0.4169
	SCM	0.5354	0.5443	0.5423	0.5427	0.6775	0.6904	0.7084	0.7142	0.6126	0.6334	0.6495	0.6390
	FSH	0.4869	0.5129	0.5245	0.5256	0.6479	0.6491	0.6478	0.6472	0.5546	0.5676	0.5611	0.5667
	SDMFH	0.6605	0.7061	0.7115	0.7075	0.7448	0.7744	0.7765	0.8020	0.7100	0.7676	0.7665	0.7824
	GSPH	0.6041	0.6027	0.6300	0.6331	0.7459	0.7563	0.7695	0.7705	0.6827	0.6911	0.6898	0.6962
	LCMFH	0.7073	0.7369	0.7400	0.7315	0.7301	0.7468	0.7557	0.7615	0.7279	0.7393	0.7437	0.7470
	DSMFH	**0.7257**	**0.7526**	**0.7554**	**0.7564**	**0.7544**	**0.7919**	**0.8029**	**0.8100**	**0.7576**	**0.7871**	**0.7994**	**0.7956**

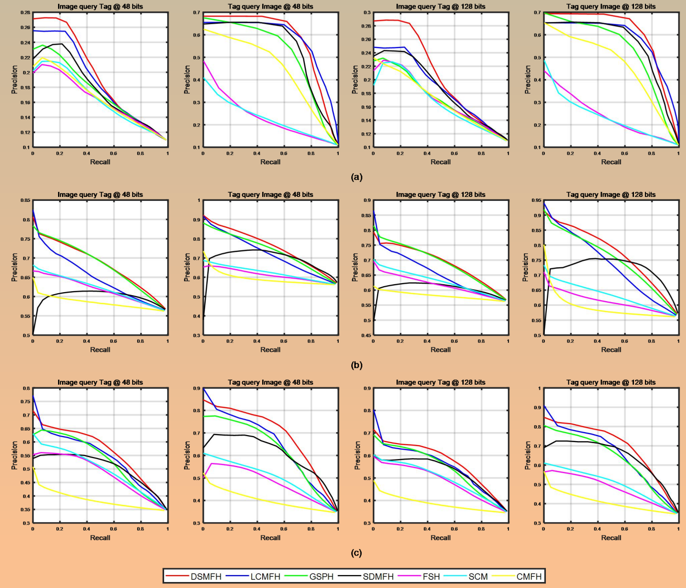


Fig. 1. The PR curves of different methods on the (a) Wiki, (b) MIRFlickr25k, (c) NUS-WIDE databases, respectively.