Jointly Multiple Hash Learning

Xingbo Liu,1 Xiushan Nie,2,* Yingxin Wang,1 Yilong Yin3,*
1School of Computer Science and Technology, Shandong University, Jinan, P.R. China
2School of Computer Science and Technology, Shandong Jianzhu University, Jinan, P.R. China
3School of Software, Shandong University, Jinan, P.R. China
sclxb@mail.sdu.edu.cn, wyxwalker@163.com, niexiushan@163.com, ylyin@sdu.edu.cn

Motivation
Hashing can compress heterogeneous high-dimensional data into compact binary codes while preserving the similarity to facilitate efficient retrieval and storage, and thus hashing has recently received much attention from information retrieval researchers. Most of the existing hashing methods first define a fixed length (e.g., 32, 64, or 128 bit) for the hash codes, before learning them with this fixed length. When the length of the hash code changes, these methods need to be retrained and learn the hash codes again. However, we know that hash codes are compact representations of the original samples, and one sample can be represented by various hash codes. In addition, hash codes with different lengths that represent the same sample would reflect different relationships and associations should exist among different types of specific information from the original sample. If they are treated as different views of the original sample, some relationships and associations should exist among different views that could improve the performance of hash learning according to multi-viewing theory (Xu, Tao, and Xu 2013). Therefore, harnessing these relationships will boost the performance of hashing methods.

Inspired by this possibility, in this study, we propose a new model called jointly multiple hash learning (JMH), which can learn hash codes with multiple lengths simultaneously. In the proposed JMH method, three types of information are used for hash learning, which come from hash codes with different lengths and the original features of the samples and labels. They are represented by the projections $P_k$, $U_k$, and $W_k$. To learn these three projections and multi-length hash codes, we propose three loss terms comprising $Q$-Loss, $L$-Loss, and $M$-Loss in the objective function of the JMH, where $Q$-Loss represents the loss between the $k_{th}$ hash code $H_k$ and the other hash codes, $L$-Loss represents the loss between the hash codes and original features and $M$-Loss describes the classification error, respectively.

Formulation
It is assumed that we have a training set $X$ comprising $n$ samples (i.e., $X = \{x_i\}_{i=1}^n$), where $x_i \in \mathbb{R}^d$ is the feature vector. A semantic label matrix $Y = \{y_{ij}\}_{i,j=1}^n$ is also available, where $y_i = \{y_{ij}\} \in \mathbb{R}^c$ is the label vector of the $i_{th}$ instance and $c$ is the number of categories in the training set. If the $i_{th}$ instance belongs to the $j_{th}$ semantic category, $y_{ij} = 1$; otherwise, $-1$. In addition, it is assumed that we want to obtain $K$ different hash codes (vectors) with each length $L_k$. The hash matrix of all $n$ samples for the $k_{th}$ ($1 \leq k \leq K$) hash codes is defined as $H_k = \{h_{ki}\}_{i=1}^n$, where $h_{ki}$ is the $k_{th}$ hash code of the $i_{th}$ sample with length $L_k$. $P_k$, $U_k$, and $W_k$ are the matrices that describe the linear projections.

In the proposed JMH, three types of information are used, which come from hash codes with different lengths and the original features of the samples and labels. They are represented by the projections $P_k$, $U_k$, and $W_k$. To learn these three projections and multi-length hash codes, we propose three loss terms comprising $Q$-Loss, $L$-Loss, and $M$-Loss in the objective function of the JMH, where $Q$-Loss represents the loss between the $k_{th}$ hash code $H_k$ and the other hash codes, $L$-Loss represents the loss between the hash codes and original features and $M$-Loss describes the classification error, respectively.

\[
Q = \sum_{k=1}^{K-1} \gamma_k ||H_k - P_k^T H_{k+1}||^2 + \sum_{k=1}^{K-1} \lambda_k ||P_k||^2, \tag{1}
\]

\[
L = \sum_{k=1}^{K} \alpha_k ||H_k - U_k^T X||^2, \tag{2}
\]

\[
M = \sum_{k=1}^{K} \mu_k ||Y - W_k^T H_k||^2, \tag{3}
\]

where $\text{Loss}(\cdot)$ is a loss function, and $\lambda_k$ and $\gamma_k$ are the parameters.

By combining Equations (1), (2), and (3), the final formulation for the JMH can be defined as:

\[
\begin{align*}
&\min_{H_k, W_k, P_k, U_k} \sum_{k=1}^{K} \mu_k ||Y - W_k^T H_k||^2 \\
&+ \sum_{k=1}^{K-1} \gamma_k ||H_k - P_k^T H_{k+1}||^2 + \sum_{k=1}^{K} \alpha_k ||H_k - U_k^T X||^2 \\
&+ \sum_{k=1}^{K} \lambda_k ||P_k||^2
\end{align*}
\]

s.t. $H_k \in \{-1, 1\}^{L_k \times n}, \tag{4}$

where $\mu_k$, $\alpha_k$, $\lambda_k$ and $\gamma_k$ are the parameters.
Optimizing Equation (4) directly is a challenge because it is both non-convex and non-smooth. However, obtaining a solution becomes easier when we take one variable while another is fixed. In the JMH, Equation (4) can be solved by an iterative framework and cyclic coordinate descent until convergence, and we can directly obtain a discrete solution for the hash codes without relaxation.

**Experiments**

We used three datasets comprising CIFAR-10, MS-COCO, and NUS-WIDE in the experiments, which have been employed widely in image retrieval studies. To evaluate the proposed method, we used the evaluation metric comprising the mean average precision (MAP), which is used widely in image retrieval.

We compared the proposed JMH with the following methods: Spectral Hashing (SH) (Weiss, Torralba, and Fergus 2009), PCA-Iterative Quantization (PCA-ITQ) (Gong and Lazebnik 2011), PCA-Random Rotation (Gong and Lazebnik 2011), Collective Matrix Factorization Hashing (MFH) (Ding, Guo, and Zhou 2014), Supervised Discrete Hashing (SDH) (Shen et al. 2015), Column Sampling based Discrete Supervised Hashing (COSDISH) (Kang, Li, and Zhou 2016), Fast Supervised Discrete Hashing (FSDH) (Gui et al. 2018), and another method.

During training, four lengths of hash codes (i.e., $K=4$) ranging from 12 to 48 bits were employed in comparisons with several state-of-the-art methods. In addition, five types of hyperparameters were used with our proposed JMH method. We empirically set $\{\mu\}_{k=1}^{K} = \{1, 0.8, 0.7, 0.6\}$, $\{\alpha\}_{k=1}^{K-1}$, $\{\gamma\}_{k=2}^{K}$ to $1e-6$, $\{\beta\}_{k=1}^{K}$ and $\{\lambda\}_{k=0}^{K}$ to 0.1.

The MAP results obtained for all of the methods based on CIFAR-10, MS-COCO and NUS-WIDE using hash codes with different lengths are listed in Table 1. Table 1 shows that the proposed JMH method performed substantially better than the other hashing methods, where the MAP results showed that our method improved the performance by 3–10% in most cases. The results demonstrated that the relationships among multi-length hashes could boost the retrieval performance.

**Conclusion**

In this study, we propose a novel supervised discrete hashing method for learning multi-length hash codes in a united framework. In the proposed model, every hash code is treated as a view of the corresponding sample. After training using the cyclic coordinate descent method, discrete multi-length hashes are obtained based on the associations among different hash codes. Experiments conducted based on three benchmark datasets confirmed the superior performance of our method in various retrieval scenarios. One challenge with the proposed method is that the number of parameters increases with the number of hash codes in the framework. We will try to address this challenge in future research.

**Acknowledgement**

This work is supported by the National Natural Science Foundation of China (61876098, 61671274, 61573219), and the Fostering Project of Dominant Discipline and Talent Team of Shandong Province Higher Education Institutions.

**References**


---

Table 1: Performance in terms of MAP score.

<table>
<thead>
<tr>
<th>Method</th>
<th>CIFAR-10</th>
<th></th>
<th>MS-COCO</th>
<th></th>
<th>NUS-WIDE</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>12 bits</td>
<td>24 bits</td>
<td>36 bits</td>
<td>48 bits</td>
<td>12 bits</td>
<td>24 bits</td>
</tr>
<tr>
<td>SH</td>
<td>0.2704</td>
<td>0.2908</td>
<td>0.2898</td>
<td>0.2961</td>
<td>0.6490</td>
<td>0.6501</td>
</tr>
<tr>
<td>PCA-ITQ</td>
<td>0.2767</td>
<td>0.3554</td>
<td>0.3398</td>
<td>0.3562</td>
<td>0.5328</td>
<td>0.6281</td>
</tr>
<tr>
<td>PCA-RR</td>
<td>0.2903</td>
<td>0.3054</td>
<td>0.2938</td>
<td>0.3172</td>
<td>0.5590</td>
<td>0.5677</td>
</tr>
<tr>
<td>MFH</td>
<td>0.2991</td>
<td>0.3345</td>
<td>0.3473</td>
<td>0.3623</td>
<td>0.6171</td>
<td>0.6330</td>
</tr>
<tr>
<td>SDH</td>
<td>0.3473</td>
<td>0.3830</td>
<td>0.4111</td>
<td>0.4984</td>
<td>0.5482</td>
<td>0.6037</td>
</tr>
<tr>
<td>COSDISH</td>
<td>0.3820</td>
<td>0.4366</td>
<td>0.4854</td>
<td>0.5295</td>
<td>0.5230</td>
<td>0.5369</td>
</tr>
<tr>
<td>FSDH</td>
<td>0.5370</td>
<td>0.6218</td>
<td>0.6526</td>
<td>0.6632</td>
<td>0.7287</td>
<td>0.7780</td>
</tr>
<tr>
<td>JMH</td>
<td>0.4576</td>
<td>0.6517</td>
<td>0.6624</td>
<td>0.6695</td>
<td>0.7741</td>
<td>0.8096</td>
</tr>
</tbody>
</table>

Acknowledgements

This work is supported by the National Natural Science Foundation of China (61876098, 61671274, 61573219), and the Fostering Project of Dominant Discipline and Talent Team of Shandong Province Higher Education Institutions.