

Accepted Manuscript

Adaptive Hash Retrieval with Kernel Based Similarity

Xiao Bai, Cheng Yan, Haichuan Yang, Lu Bai, Jun Zhou,
Edwin Robert Hancock

PII: S0031-3203(17)30131-0
DOI: [10.1016/j.patcog.2017.03.020](https://doi.org/10.1016/j.patcog.2017.03.020)
Reference: PR 6093

To appear in: *Pattern Recognition*

Received date: 11 September 2016
Revised date: 17 March 2017
Accepted date: 20 March 2017

Please cite this article as: Xiao Bai, Cheng Yan, Haichuan Yang, Lu Bai, Jun Zhou, Edwin Robert Hancock, Adaptive Hash Retrieval with Kernel Based Similarity, *Pattern Recognition* (2017), doi: [10.1016/j.patcog.2017.03.020](https://doi.org/10.1016/j.patcog.2017.03.020)



This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Adaptive Hash Retrieval with Kernel Based Similarity

Xiao Bai^a, Cheng Yan^a, Haichuan Yang^a, Lu Bai^b, Jun Zhou^c, Edwin Robert Hancock^d

^a*School of Computer Science and Engineering, Beihang University, Beijing, China*

^b*School of Information, Central University of Finance and Economics, Beijing, China*

^c*School of Information and Communication Technology, Griffith University, Nathan, QLD, Australia*

^d*Department of Computer Science, University of York, York, UK*

Abstract

Indexing methods have been widely used for fast data retrieval on large scale datasets. When the data are represented by high dimensional vectors, hashing is often used as an efficient solution for approximate similarity search. When a retrieval task does not involve supervised training data, most hashing methods aim at preserving data similarity defined by a distance metric on the feature vectors. Hash codes generated by these approaches normally maintain the Hamming distance of the data in accordance with the similarity function, but ignore the local details of the distribution of data. This objective is not suitable for k-nearest neighbor search since the similarity to the nearest neighbors can vary significantly for different data samples. In this paper, we present a novel adaptive similarity measure which is consistent with k-nearest neighbor search, and prove that it leads to a valid kernel if the original similarity function is a kernel function. Next we propose a method which calculates hash codes using the kernel function. With a low-rank approximation, our hashing framework is more effective than existing methods that preserve similarity over an arbitrary kernel. The proposed similarity function, hashing framework, and their combination demonstrate significant improvement when compared with several alternative state-of-the-art methods.

Keywords:

Hashing, K-NN, Kernel, Binary Indexing, Normalized Euclidean Distance

Download English Version:

<https://daneshyari.com/en/article/6939647>

Download Persian Version:

<https://daneshyari.com/article/6939647>

[Daneshyari.com](https://daneshyari.com)