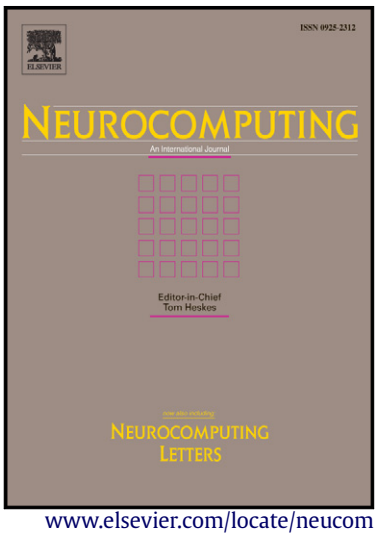

www.elsevier.com/locate/neucom

Local Similarity Preserved Hashing Learning via Markov Graph for Efficient Similarity Search

Hong Liu, Aiwen Jiang, Mingwen Wang, Jianyi Wan

PII: S0925-2312(15)00153-8
DOI: http://dx.doi.org/10.1016/j.neucom.2015.01.073
Reference: NEUCOM15131

To appear in: *Neurocomputing*

Received date: 26 November 2014
Revised date: 23 January 2015
Accepted date: 27 January 2015

Cite this article as: Hong Liu, Aiwen Jiang, Mingwen Wang, Jianyi Wan, Local Similarity Preserved Hashing Learning via Markov Graph for Efficient Similarity Search, *Neurocomputing*, http://dx.doi.org/10.1016/j.neucom.2015.01.073

Local Similarity Preserved Hashing Learning via Markov Graph for Efficient Similarity Search

Hong LIU[a], Aiwen JIANG[a,*], Mingwen WANG[a], Jianyi WAN[a]

[a]*School of Computer and Information Engineer, JiangXi Normal University*

Abstract

Hashing, for its efficiency to nearest neighbor search in high dimensional space, has become an attractive topic in multimedia retrieval area . In this paper, an effective hashing algorithm based on markov graph has been proposed. Through constructing a stable composite affinity graph, it can preserve similarity information well in the embedded subspace. Furthermore, a practical strategy has been supplied to reduce the computational complexity. Comparisons with several state-of-the-art algorithms have been done in three public datasets. The experimental results have demonstrated that the proposed method can achieve competitive performances, and afford large scale similarity search tasks.

Keywords: Hashing, Similarity Search, Markov Graph, Spectral Decomposition

1. Introduction

Efficient similarity search is of great importance for large scale multimedia retrieval tasks. In general, searching the nearest neighbors of a query q requires scanning all n items in database, which will be extremely time-consuming, when dealing with large scale data. The current exact nearest neighbor search methods like kd-tree often suffer from the curse of high dimensionality, while approximately nearest neighbor (ANN) search ones are more practical [1]. Hashing is a popular solution, attracting much attention in recent years. It has been shown that hashing-based methods could achieve fast similarity search by representing high-dimensional data with compact binary codes [2] and have been successfully applied to a variety of pattern recognition problems such as visual object detection[3], and recognition [4], large-scale document analysis [5]and image retrieval[6].

Compared with real-value features, the learned binary codes have many advantages, such as: (1) hashing codes are very compact so that they enable the storage of large scale data in memory; (2) binary codes, performing approximate similarity search through quick bit XOR operations, enable more tasks like information retrieval applicable. Furthermore, binary features will definitely be useful in some emerging practical scenarios like mobile

*Corresponding author

Email addresses: ir_hongliu@foxmail.com (Hong LIU), jiangaiwen@jxnu.edu.cn (Aiwen JIANG), wangmw@jxnu.edu.cn (Mingwen WANG), wanjianyi@jxnu.edu.cn (Jianyi WAN)