Pattern Recognition 75 (2018) 1–3

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/patcog

Editorial

Distance metric learning for pattern recognition

1. Introduction

Distance metric learning techniques have played a central role in pattern recognition, and a variety of distance metric learning methods have been developed for various pattern recognition applications over the past decade. These methods aim to learn an appropriate distance function given some constrains between samples. To better discover the geometric property of high-dimensional feature spaces and exploit the complementary information of different feature spaces, manifold learning and multi-view learning strategies have also been integrated into distance metric learning to further improve the performance of various distance metric learning methods. While these methods are helpful to learn the similarity of data such as images, videos, texts, radars, and voices, how to develop task-specific distance metric learning algorithms for different pattern recognition tasks still remains unsolved, especially for big data which are captured in the wild. Moreover, how to develop transferable and nonlinear distance metric learning methods for large-scale pattern recognition systems still requires many efforts. This special issue focuses on state-of-the-art methods, algorithms and applications of distance metric learning for various pattern recognition applications.

2. Overview of accepted papers

Accepted papers in this special issue cover a broad range of pattern recognition applications with distance metric learning techniques including face analysis, person re-identification, image retrieval, object detection, image classification, multi-modal matching and so on. In total, we received 63 submissions from all over the world and 31 of them have been finally accepted based on a strict and comprehensive review process. Each submission has been assigned to two peer reviewers and gone through at least two rounds of revision at least.

2.1. Face analysis

The paper "Learning Structured Ordinal Measures for Video based Face Recognition" proposes a structural ordinal measure method by utilizing output structures, which learns ordinal filters and structured ordinal features simultaneously by a non-convex integer program. Experimental results show that their method achieves the state-of-the-art performance with fewer features and samples in video face recognition.

Yan and Hu investigate the problem of video-based kinship verification via human face analysis in "Video-Based Kinship Verification Using Distance Metric Learning". They present a video face dataset for kinship verification and evaluate several state-of-the-art metric learning based kinship verification methods.

In the paper "Supervised Deep Hashing for Scalable Face Image Retrieval", Yan et al. utilize both quantization errors and classification errors to supervise the training procedure of the deep hashing network. Experiments demonstrate its effectiveness compared with several state-of-the-art hashing methods in face recognition.

Tang et al. combine the multiple metric distances induced by complementary features in the article "Collaborative Discriminative Multi-Metric Learning for Facial Expression Recognition in Video". It learns the multimetric to push the negative video away, and also maximizes the correlation of different feature channels. Experimental results on two datasets demonstrate the effectiveness of their method.

The paper "Multiple Face Tracking and Recognition with Identity-specific Localized Metric Learning" simultaneously addresses the tasks of multiple face recognition and multiple face tracking, which introduces the constraints of inter-frame temporal smoothness and within-frame identity. Experimental results on the several multiple face recognition sequences demonstrate that the method improves the accuracy in various challenging scenarios.

Li et al. design an end-to-end CNN to integrate the feature extraction and metric learning in the paper "Distance Metric Optimization Driven Convolutional Neural Network for Age Invariant Face Recognition". The experiments contain three parts, which include age-invariant face identification on the MORPH database, age-invariant face retrieval on the CACD database and age-invariant face verification on the CACD-VS database.

2.2. Person Re-identification

Dai et al. propose a cross-view semantic projection learning method to extract latent semantics form hand-crafted features through matrix factorization in the paper "Cross-view Semantic Projection Learning for Person Re-identification". Experimental results on several challenging datasets show the advantage of their algorithm.

The paper "Multi-Type Attributes Driven Multi-Camera Person Re-identification" presents a weakly supervised multi-type attribute learning framework to obtain mid-level human attributes, where the attribute labels are predicted for the distance metric. Experiments on the four datasets show the effectiveness of the method.

https://doi.org/10.1016/j.patcog.2017.10.032

In the article "Person Re-identification via Integrating Patch-based Metric Learning and Local Salience Learning", Zhao et al. learn the metric on the patch-level and uses local salience learning algorithm based on K-means to learn the weights of patches. Experimental results indicate that their method achieves comparable performance with most CNN-based methods.

Ren et al. introduce a mean distance of multi-metric subspace to address the over-fitting problem of metric learning in the paper "Feedback Mechanism based Iterative Metric Learning for Person Re-identification". Experiments on three datasets show that their method achieves the state-of-the-art performance.

The paper "Gestalt Laws Based Tracklets Analysis for Human Crowd Understanding" aims to detect crowded human groups and learn semantic regions with a unified hierarchical clustering framework. The method defines a unified similarity metric for both original tracklets and representative tracklets with the Gestalt laws of grouping. The performance on three datasets demonstrates the effectiveness of the approach.

2.3. Image retrieval

In the paper "Robust Discrete Code Modeling for Supervised Hashing", the authors propose a robust discrete code modeling method to learn high-quality discrete binary codes and hash functions by suppressing the influence of unreliable binary codes and potentially noisily-labeled samples. Experimental results on various real-life datasets show the superiority of the approach compared with several state-of-the-art hashing methods.

In the article "Adaptive Hash Retrieval with Kernel Based Similarity", Bai et al. present an adaptive similarity metric which is consistent with k-nearest neighbor search, and present a kernel-based hashing code calculation. Performance evaluation on the four datasets demonstrates significant improvement of the method.

Chen and Zhou propose a method in the paper "Collaborative Multi-view Hashing", which aims to incorporate multi-view representations into the binary code learning for scalable visual retrieval by exploiting both interview and representation-label collaborations. Experimental results on two datasets validate the superiority of the approach in comparison with several state-of-the-art hashing methods.

Pedronette et al. propose a manifold learning approach that exploits the intrinsic dataset geometry for improving the effectiveness of image retrieval tasks in the paper "Unsupervised Manifold Learning through Reciprocal kNN Graph and Connected Components for Image Retrieval Tasks". Experiments on six public datasets show the effectiveness of their method.

Authors in the paper "Quantization-based Hashing: A General Framework for Scalable Image and Video Retrieval" present a quantization-based hashing framework to combine the quantization error reduction methods and property preserving hashing methods. Experimental results demonstrate that their method outperforms existing unsupervised and supervised hashing methods.

2.4. Multi-modal matching

In the paper "Efficient Multi-modal Geometric Mean Metric Learning", Liang et al. develop an effective and efficient metric learning algorithm in a joint optimization problem by pulling similar pairs close whereas pushing dissimilar pairs away and assign a weight to each modality. Experimental results on five image datasets show that their method outperforms the state-of-the-art metric learning methods.

Li et al. develop a multi-view discriminative metric learning method to obtain better graph partitioning results in the article "Discriminative Metric Learning for Multi-view Graph Partitioning". Experimental results on both synthetic and real-world graphs show the effectiveness of their method.

In the paper "Cross Modal Similarity Learning with Active Queries", Gao et al. propose a method for active cross modal similarity learning. They actively query the most important supervised information with the intra-modal and inter-modal similarities, and select the instances which are closer to the decision boundary. Experiments on multiple benchmark datasets validate the effectiveness of the approach.

The paper "Topic Driven Multi-modal Similarity Learning with Multi-view Voted Convolutional Features" analyzes the similarities between objects by operating over multiple modalities, and fine-tunes the model variables with a ranking loss. Experimental results demonstrate the improvements over previous methods of their method.

Pereira and Torres establish pairwise constraints between the source and labeled target data, and exploits the global structure of the unlabeled data to build a domain invariant subspace in the paper "Semi-Supervised Transfer Subspace for Domain Adaptation". Results on 49 cross-domain problems show the effectiveness of their method on domain shift reduction and classification accuracy.

The paper "Parametric Local Multi-view Hamming Distance Metric Learning" proposes a multi-view local metric learning method to locally adapt to the data structure of multi-modal data for text-image matching. The weak-supervisory information provided by pairwise and triplet constraints are incorporated in a coherent way to achieve semantically effective hash codes. Experimental evaluations on cross-media retrieval tasks demonstrate the effectiveness of the approach.

2.5. Object detection

In the paper "3D Neuron Tip Detection in Volumetric Microscopy Images Using an Adaptive Ray-Shooting Model", Liu et al. propose an adaptive ray-shooting model by changing the length of the shooting rays and the number of adjacent slices according to the local diameter of the neuron. Experimental results show that their method improves the detection accuracy rate by about 10% in challenging datasets.

The paper "Locality Constraint Distance Metric Learning for Traffic Congestion Detection" defines the congestion in an accurate and unified manner, and regards the congestion level analysis as a regression problem. The method ensures the local smoothness and preserves the correlations between samples based on the low-level texture feature and kernel regression. The experiments on the constructed dataset confirm the effectiveness of the method.

2.6. Image classification

Hu and Tan employ a feed-forward neural network to seek hierarchical feature projection matrices and dictionary simultaneously to exploit nonlinear structure of samples in the paper "Nonlinear Dictionary Learning with Application to Image Classification". They also extend original unsupervised method to a supervised extension by learning the class-specific dictionary. Experimental results on four image datasets demonstrate the effectiveness of the methods.

The paper "A Parasitic Metric Learning Net for Breast Mass Classification Based on Mammography" trains a deep CNN to obtain more discriminative description of breast tissues, and proposes a parasitic metric learning net to improve the performance. Experimental results show that their method achieves the state-of-the-art performance.

2.7. Other applications

The paper "A One-Pass Closed-Form Solution for Online Metric Learning" adopts a one-pass triplet construction strategy and employs a closed-form solution to update the metric for new coming samples. This paper also presents an extension to enhance the robustness when the first several samples come from the same class. Experiments are on the various typical tasks show that the methods obtain the promising performance with low computational cost.

The authors in the paper "Learning Local Metrics from Pairwise Similarity Data" learn a set of local metrics, which are aggregated for the final distance metric. Experiments show that their method achieves state-of-the-art results on several datasets.

In the paper "Learn to Model Blurry Motion via Directional Similarity and Filtering", Li et al. propose a hybrid framework by conducting a learnable directional filtering layer to encode the angle and distance similarity matrix between blur and camera motion. Their method yields competitive performance with the state-of-the-art approaches on a synthetic dataset.

Acknowledgments

We would like to thank Professor Edwin R. Hancock, Editor-in-Chief of the *Pattern Recognition* journal for giving us the opportunity to organize this special issue. We also thank all the authors for their contributions to this special issue, and reviewers for their timely and insightful comments and suggestions.

Jiwen Lu
Department of Automation, Tsinghua University, Beijing 100084, China

Ruiping Wang
Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China

Ajmal Mian
School of Computer Science and Software Engineering, University of Western Australia, Crawley WA6009, Australia

Ajay Kumar
Department of Computing, Hong Kong Polytechnic University, Kowloon, Hong Kong

Sudeep Sarkar
Department of Computer Science and Engineering, University of South Florida, Tampa, FL 3364, USA

E-mail address: lujiwen@tsinghua.edu.cn (J. Lu)