Chapter 3

Self-Supervised Learning Based on Discriminative Nonlinear Features and Its Applications for Pattern Classification

Qi Tian, University of Texas at San Antonio, USA

Ying Wu, Northwestern University, USA

Jie Yu, University of Texas at San Antonio, USA

Thomas S. Huang, University of Illinois, USA

ABSTRACT

For learning-based tasks such as image classification and object recognition, the feature dimension is usually very high. The learning is afflicted by the curse of dimensionality as the search space grows exponentially with the dimension. Discriminant expectation maximization (DEM) proposed a framework by applying self-supervised learning in a discriminating subspace. This paper extends the linear DEM to a nonlinear kernel algorithm, Kernel DEM (KDEM), and evaluates KDEM extensively on benchmark image databases and synthetic data. Various comparisons with other state-of-the-art learning techniques are investigated for several tasks of image classification, hand posture recognition and fingertip tracking. Extensive results show the effectiveness of our approach.

INTRODUCTION

Invariant object recognition is a fundamental but challenging computer vision task, since finding effective object representations is generally a difficult problem. Three dimensional (3D) object reconstruction suggests a way to invariantly characterize objects. Alternatively, objects could also be represented by their visual appearance without explicit reconstruction. However, representing objects in the image space is formidable, since the dimensionality of the image space is intractable. Dimension reduction could be achieved by identifying invariant image features. In some cases, domain knowledge could be exploited to extract image features from visual inputs, such as in content-based image retrieval (CBIR). CBIR is a technique which uses visual content to search images from large-scale image databases according to user's interests, and has been an active and fast advancing research area since the 1990s (Smeulders, 2000). However, in many cases machines need to *learn* such features from a set of examples when image features are difficult to define. Successful examples of learning approaches in the areas of content-based image retrieval, face and gesture recognition can be found in the literature (Tieu et al., 2000; Cox et al., 2000; Tong & Wang, 2001; Tian et al., 2000; Bellhumeur, 1996).

Generally, characterizing objects from examples requires huge training datasets, because input dimensionality is large and the variations that object classes undergo are significant. Labeled or supervised information of training samples are needed for recognition tasks. The generalization abilities of many current methods largely depend on training datasets. In general, good generalization requires large and representative labeled training datasets. Unfortunately, collecting labeled data can be a tedious, if not impossible, process. Although unsupervised or clustering schemes have been proposed (e.g., Basri et al., 1998; Weber et al., 2000), it is difficult for pure unsupervised approaches to achieve accurate classification without supervision.

This problem can be alleviated by *semisupervised* or *self-supervised* learning techniques which take hybrid training datasets. In content-based image retrieval (e.g., Smeulders et al., 2000; Tieu et al., 2000; Cox et al., 2000; Tong & Wang, 2001; Tian et al., 2000), there are a limited number of labeled training samples given by user query and relevance feedback (Rui et al., 1998). Pure supervised learning on such a small training dataset will have poor generalization performance. If the learning classifier is over-trained on the small training dataset, *over-fitting* will probably occur. However, there are a large number of unlabeled images or unlabeled data in general in the given database. Unlabeled data contain information about the joint distribution over features which can be used to help supervised learning. These algorithms assume that only a fraction of the data is labeled with ground truth, but still take advantage of the entire data set to generate good classifiers; they make the assumption that nearby data are likely to be generated by the same class. This learning paradigm could be seen as an integration of pure supervised and unsupervised learning.

Discriminant-EM (DEM) (Wu et al., 2000) is a self-supervised learning algorithm for such purposes that use a small set of labeled data with a large set of unlabeled data. The basic idea is to learn discriminating features and the classifier simultaneously by inserting a multiclass linear discriminant step in the standard expectation-maximization (EM) (Duda et al., 2001) iteration loop. DEM makes the assumption that the probabilistic

Related Content

The Factors that Influence E-Instructors' Performance in Taiwan: A Perspective of New
Human Performance Model
Chun-Yi Shenand Chiung-Sui Chang (2010). *International Journal of Multimedia Data
Engineering and Management (pp. 50-59).*
www.irma-international.org/article/factors-influence-instructors-performance-taiwan/49149

Unsupervised Video Object Foreground Segmentation and Co-Localization by
Combining Motion Boundaries and Actual Frame Edges
Chao Zhangand Guoping Qiu (2018). *International Journal of Multimedia Data Engineering and
Management (pp. 21-39).*
www.irma-international.org/article/unsupervised-video-object-foreground-segmentation-and-co-localization-
by-combining-motion-boundaries-and-actual-frame-edges/226227

Multimodal Data Integration and User Interaction for Avatar Simulation in Augmented
Reality
Anchen Sun, Yudong Tao, Mei-Ling Shyu, Angela Blizzard, William Andrew Rothenberg,
Dainelys Garciaand Jason F. Jent (2022). *International Journal of Multimedia Data Engineering
and Management (pp. 1-19).*
www.irma-international.org/article/multimodal-data-integration-and-user-interaction-for-avatar-simulation-in-
augmented-reality/304391

A Multimedia Database Supports Internet-Based English Learning
Ying-Hong Wang (2002). *Distributed Multimedia Databases: Techniques and Applications (pp.
274-292).*
www.irma-international.org/chapter/multimedia-database-supports-internet-based/8627

Trends in Managing Multimedia Semantics
Roberto Poli, Achilles D. Kameasand Lambrini Seremeti (2018). *Digital Multimedia: Concepts,
Methodologies, Tools, and Applications (pp. 1690-1702).*
www.irma-international.org/chapter/trends-in-managing-multimedia-semantics/189549