Neurocomputing 136 (2014) 152–161

Contents lists available at ScienceDirect

Neurocomputing

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

Comparative study among three strategies of incorporating spatial structures to ordinal image regression

Qing Tian, Songcan Chen*, Xiaoyang Tan

College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, PR China

A R T I C L E I N F O

Article history:
Received 26 July 2013
Received in revised form
19 November 2013
Accepted 3 January 2014
Communicated by Qingshan Liu
Available online 28 January 2014

Keywords:
Ordinal regression
Vector pattern
Matrix pattern
Spatial structure
Euclidean distance
Bilinear

A B S T R A C T

Images usually have specific spatial structures, and related researches have shown that these structures can contribute to the establishment of more effective classification algorithms for images. So far though there have been many solutions of making use of such spatial structures separately proposed, little attention has been paid to their systematic summary, let their comparative study alone. On the other hand, we find that the existing image-oriented ordinal regression (OR) methods do not utilize such structure information, which motivates us to compensate a comparative study through embedding such spatial structure into ORs. Towards the end, in this paper, we (**1**) through a summary, find three typical strategies of using image prior spatial information, i.e., structure-embedded Euclidean distance strategy, structure-regularized modeling strategy for classifier learning, and direct manipulation strategy on images without vectorization for image; *more importantly*, (**2**) apply these strategies to establish corresponding ORs for classifying data with ordinal characteristic, conduct comprehensive comparisons and give analysis on them under three evaluation criteria. Experimental results on typical ordinal image datasets JAFFE, UMIST and FG-NET show that the latter two strategies can, on the whole, achieve distinct gain in OR performance and while the first one cannot necessarily as expected, which is due to whether the spatial information is directly embedded into the objective function involved or not.

1. Introduction

1.1. Background

Images have two-dimensional inherent spatial structures, in which explicit and implicit discriminative information beneficial to image classification is involved. For example, in human faces, the eyes, nose and mouth are distributed in different regions, and specific geometric relations exist between them. However, most current developed pattern recognition and machine learning algorithms are based on vector patterns, in which the process of matrix-to-vector conversion is performed, consequently, useful spatial structure information to classification is lost seriously, thus leaving the room of performance promotion.

Over past years, though strategies of taking advantage of spatial structure information have been separately developed for improving performance of image classification, a systematic summary and comparative study among them is still lacked. For this purpose, in this paper, we will ***first*** make a summary from those scattered related literature and group them into three categories; then for making a comparison among them, we choose one of currently popular topics in image classification, i.e., image-oriented OR, as the comparative platform. OR is a special machine learning paradigm and possesses the duality of classification and regression, thus often is applied in such scenarios in which the predicted labels are discrete but ordered [26,31], e.g., human facial age estimation and movie scoring and so on. Besides the duality, further reasons of choosing image-oriented OR as the comparative study paradigm are (***1***) *these specially-designed ORs for ordinal image classification have so far hardly exploited such spatial information, and (**2**) the multi-index-based synthetic evaluation originated from their duality of classification and regression can more be reflected from multi-facets for such information utilization than single-index evaluation for classification or regression.* And ***next***, we develop three image-oriented OR variants by the compensation of spatial information using the aforementioned three strategies and then make an extensive comparison from a joint view of regression and classification under three evaluation criteria of MAE, Acc and OCI.

1.2. Categorization of the strategies of utilizing spatial structure information

In this subsection we analyze the existing scattered schemes designed to utilize the spatial structure information and summarize them into three main families as follows:

1.2.1. Structure-embedded Euclidean distance strategy

It is known that Euclidean distance (ED) is one of the most often-used metric in pattern recognition. However, when it is used

* Corresponding author.
E-mail addresses: tianqing@nuaa.edu.cn (Q. Tian), s.chen@nuaa.edu.cn (S. Chen), x.tan@nuaa.edu.cn (X. Tan).

0925-2312/$ - see front matter
http://dx.doi.org/10.1016/j.neucom.2014.01.017

to similarity/distance measure between two images, the spatial structure information involved in them is not sufficiently reflected such that classification performance for the images is unfavorably affected. In order to compensate such loss, many attempts [1–8] have been done, among which Ref. [1] can be viewed as their representative. In Ref. [1], the authors developed an IMage Euclidean Distance (IMED) by means of embedding spatial structure of images to ED and applied it to handwritten digit and human face recognition with better performance than ED. Due to its insensitiveness to small distortion of images and generality able to be embedded into such classifiers as SVM, IMED can successively be extended. For example, Li et al. [4] extended the IMED to multi-view gender classification and achieved higher classification accuracy; Liu et al. [5] further proposed multi-linear locality-preserved maximum information embedding for face recognition with more stable performance. Moreover, Li and Lu [8] developed an adaptive IMED (AIMED) by further fusing gray level knowledge of image to IMED besides the prior spatial information to achieve more satisfactory identification performance for human face and handwritten digit. In summary, these methods originating from IMED are either modified to different applications or embedded into other learning tasks such as SVM for performance gain. Thus in the following comparative study, we just adopt IMED as basic embedding, but any of its effective variants can straightforwardly be utilized in a similar way.

1.2.2. Structure-regularized modeling strategy

In this family, the strategy of exploiting spatial structure usually adopts the regularization technique to penalize a related objective function such that the resulted solution (by optimizing the objective) is spatially smooth as much as possible [9–13]. The spatial smooth subspace learning (SSSL) proposed in Ref. [9] can be regarded as the representative, in which a Laplacian penalty is imposed to constrain the projection coefficients to be spatially smooth. Zuo et al. [12] went further by weighting the Laplacian penalty function with Gaussian weights to realize multi-scale image smoothing. Chen et al. in Ref. [13] developed a regularized metric learning framework by again imposing the Laplacian penalty and achieved competitive face recognition performance on several benchmark datasets. From these related researches it can be easily found that the structure-regularized modeling indeed can also compensate the spatial information loss induced by tensor- or matrix-to-vector conversion. Therefore, we also try to adopt such a spatially-regularized strategy for image-oriented OR. Considering that adapting those successive strategies from the spatial regularization [9] to our problem is trivial, thus without loss of generality, we here take the spatial smooth constraint in Ref. [9] as the basic regularization strategy to conduct the following comparative study.

1.2.3. Direct manipulation strategy on images

The strategies in former two families are all vector-pattern-oriented. Though the spatial structure information of images can get utilized and thus related learning performance is boosted, these strategies usually suffer from (1) high computational complexity; and (2) the so-called "*small sample problem*", i.e., the dimensionality of feature vector is higher than the training set size, leading to over-fitting. Hence, a natural way to mitigate or address these problems is operating directly image (or reshaped image) patterns. Along this line, many studies have been developed, for example Refs. [14–25], in which the works of Chen et al. [14–18] and Tao et al. [20–25] can be regarded as their representatives. More specifically, Chen et al. developed a series of classifiers, such as MatMHKS [14] and MatFE+MatCD [18], by bilinear projection on image (or reshaped image) patterns and achieved competitive performance in such classification tasks as human face and handwritten digit identification, against the vector-pattern-oriented counterparts; while Tao et al. developed their dimensionality reduction or classification modeling directly manipulated on (higher order) tensor patterns and applied them respectively for human gait recognition [20] and visual tracking [25]. It is the direct manipulation on matrix or tensor as operating unit such that the schemes like the bilinear projection on image (second-order tensor) can make more sufficient use of the inherent spatial structure information involved in the data than their vectorized counterparts. Out of the similar consideration, in the following comparative study, we take the bi-lateral manipulation on image as a direct learning scheme to make a comparison on ordinal learning performance with the other methods.

Finally, we tabulate a brief comparative summarization for the aforementioned three strategies in Table 1.

1.3. Review of OR

Following the categorization and summary for spatial structure information utilization strategies, our next step is in position to taking the OR as a research platform, on which we will make extensive empirical comparison on three image sets among we afore-summarized three groups of categories. Before that let us briefly give a review for OR, OR is actually a special learning strategy used to design classifiers for ordinal classes, e.g., human age estimation. Due to its duality of regression and classification and powerful ability, OR has so far been widely applied in domains such as the recommender system [26], web page ranking [27], image retrieval [28], medical image diagnosis [29–30] and age estimation [31–32]. In implementing them, various approaches have been put forward [33–44], including KDLOR [44], one of distinguished ORs. Though most of these ORs have achieved performance to different extents, however, when manipulated on images, almost all these methods neglect the compensation of spatial structure information for vectorized images, thus choosing the image-oriented OR as the research platform to give a comparison among the summarized three categories of using spatial structure is reasonable. *Though such a work of incorporating the spatial information to existing OR seems trivial, to the best of our knowledge, there has indeed no related research done yet.*

Now for the sake of clarity but without loss of generality, we will just take the linear version of KDLOR, a typical OR model proposed in Ref. [44], as the basic OR approach (herein denoted as LDLOR), and select IMED [1], SSSL [9] and bilinear modeling [14] as the comparative representatives of the three families of spatial structure information utilization to re-model LDLOR, thus yielding three modified LDLOR versions, respectively named as IMED–LDLOR, SSSL–LDLOR and Bil–LDLOR, and for which we conduct a series of experiments on several image benchmark datasets and report comparison results in terms of the OR-specific evaluation criteria.

The remainder of the paper is organized as follows. In Section 2, we briefly review a representative OR, i.e., LDLOR, which is taken as the base model (baseline). In Section 3, three re-modeled LDLOR counterparts derived from three spatial structure information utilizing

Table 1
A comparative summarization of the three strategies.

Strategy type	Input pattern type	Embedding fashion to the objective
Metric embedding	Vector	Indirect
Structure regularization	Vector	Direct
Direct manipulation on images	Matrix	Direct