JID: KNOSYS [m5G;February 6, 2018;20:39]

Knowledge-Based Systems 000 (2018) 1–17

Contents lists available at ScienceDirect

Knowledge-Based Systems

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

An overview of incremental feature extraction methods based on linear subspaces

Katerine Diaz-Chito [a,*], Francesc J. Ferri [b], Aura Hernández-Sabaté [a]

[a] Centre de Visió per Computador, Universitat Autònoma de Barcelona, Spain
[b] Department d'Informàtica, Universitat de València, Spain

ARTICLE INFO

Article history:
Received 2 May 2017
Revised 19 January 2018
Accepted 20 January 2018
Available online xxx

Keywords:
Incremental learning
Incremental feature extraction
Dimensionality reduction
Linear subspace methods
Covariance-based
Covariance-free

ABSTRACT

With the massive explosion of machine learning in our day-to-day life, incremental and adaptive learning has become a major topic, crucial to keep up-to-date and improve classification models and their corresponding feature extraction processes. This paper presents a categorized overview of incremental feature extraction based on linear subspace methods which aim at incorporating new information to the already acquired knowledge without accessing previous data. Specifically, this paper focuses on those linear dimensionality reduction methods with orthogonal matrix constraints based on global loss function, due to the extensive use of their batch approaches versus other linear alternatives. Thus, we cover the approaches derived from Principal Components Analysis, Linear Discriminative Analysis and Discriminative Common Vector methods. For each basic method, its incremental approaches are differentiated according to the subspace model and matrix decomposition involved in the updating process. Besides this categorization, several updating strategies are distinguished according to the amount of data used to update and to the fact of considering a static or dynamic number of classes. Moreover, the specific role of the size/dimension ratio in each method is considered. Finally, computational complexity, experimental setup and the accuracy rates according to published results are compiled and analyzed, and an empirical evaluation is done to compare the best approach of each kind.

1. Introduction

Processing large amounts of data is nowadays a challenging task in the field of pattern recognition, which aims to extract meaningful information embedded in the data. As a first general step, appropriate structured descriptors or features must be selected or extracted from raw data through a learning process using any prior information available. This leads to a more discriminative data representation with lower dimensionality, facilitating the following steps on machine learning and data mining pipelines. The traditional way to extract these features is usually based on batch learning. However, this requires that all the data must be available from the beginning and used as whole, which is not convenient or even feasible in most online, interactive or stream-based processing applications. Several application domains such as autonomous navigation systems [1], human-robot interaction [2], object tracking [3], image classification [4], stream processing [5], face recognition [6] or recommendation systems [7–10] have been shown as examples where a complete set of training samples is usually not known in advance but generally provided little by little. Moreover, in some cases the properties of data may change as new data is considered. For instance, in face recognition tasks, human faces may show large variations depending on expressions, lighting conditions, make-up, hairstyles, aging and so forth. When a human is registered in a person identification system, it is quite difficult to consider all this facial variability in advance [11] but instead it is more convenient to discover it during the operation of the system.

As an effective alternative, the paradigm of incremental or adaptive learning has been considered and deeply studied as its own pattern recognition and machine learning subfield. By using incremental learning, feature extraction processes should be capable of incorporating the new information available while retaining the previously acquired knowledge, without accessing the previously processed training data. This fact is very challenging specially in the era of big data, where new chunks of data is continuously appearing and new classification objectives arise.

Among the huge amount of incremental learning schemes, this paper focuses on linear subspace-based incremental feature extraction methods with orthogonal matrix constraints based on global loss function, due to the extensive use of their batch approaches

* Corresponding author.
E-mail addresses: kdiaz@cvc.uab.es (K. Diaz-Chito), francesc.ferri@uv.es (F.J. Ferri), aura@cvc.uab.cat (A. Hernández-Sabaté).

https://doi.org/10.1016/j.knosys.2018.01.020
0950-7051/© 2018 Elsevier B.V. All rights reserved.

Table 1
Summary of the methods presented in the paper.

		System updating way			
		covariance-based	Using SVD updates	Adaptive	Covariance free
Approach based on	PCA	Murakami and Kumar [40]	Chandrasekaran et al. [41]		Weng et al. [42]
		Hall et al. [43–45]	Levy and Lindenbaum [12]		Skočaj and Leonardis [46]
		Ozawa et al. [11,47,48]	Kwok and Zhao [49]		Qu and Yao [50]
		Li [51]	Zhao et al. [52]		Yan and Tang [53]
		Huang et al. [54]	Li et al. [55]		Zeng and Li [5]
		Duan and Chen [56]			
		Arora et al. [57]			
		Jin et al. [58]			
	LDA	Pang et al. [59,60]		Zhao and Yuen [61]	
		Ye et al. [62]		Liu et al. [63]	
		Kim et al. [64,65]		Lu et al. [66,67]	
		Uray et al. [68]		Yeh and Wang [69]	
		Song et al. [70]		Chu et al. [71]	
		Zheng and Tang [72]		Zhang et al. [73]	
		Lamba et al. [74]			
		Peng et al. [6]			
		Lu et al. [75]			
		Dhamecha et al. [76]			
	DCV	Diaz et al. [77,78]			Ferri et al. [79,80]
		[4,81]			Diaz et al. [77]
					Lu et al. [82]
					Zhu et al. [83]

versus other linear alternatives with unconstrained objectives, such as probabilistic PCA [3,12], or matrix factorization methods [7–10,13–16], mostly popular for building collaborative filtering on recommender systems. Note that not all linear feature extraction methods need to produce orthogonal projections, or indeed projections at all. While subspace-based methods can be based on linear and non-linear subspaces, linear methods are the most extensively used, even in highly non-linear problems where the non-linearity is modeled in the subsequent feature extraction and classification stages instead. An example of this is the use of linear dimensionality reduction methods in modern deep learning architectures as preprocessing step to reduce the number of parameters to be learned and the number training samples [17–19]. Moreover, these techniques have been used in the last years in many successful problems as object tracking [20–22] or in other application fields, such as pharmaceutics [23], medical image [24,25], agriculture [26], industrial applications [27], chemometrics [28,29] pattern recognition [30] or bioinformatics [31,32].

Therefore, this paper presents a categorized overview of the research done over the past decades on linear subspace-based incremental feature extraction and dimensionality reduction for matrices and general applications. Special emphasis is put on those methods with orthogonal matrix constraints based on global loss function, such as Principal Components Analysis (PCA), Linear Discriminative Analysis (LDA), and Discriminative Common Vector (DCV) methods, over methods with unconstrained objectives, such as probabilistic PCA [3,12] or matrix factorizations [7–10,14–16]. Similarly, we consider that those incremental methods which are more related to subspace-to-subspace matching [33,34], and tensor factorization [35–38] are out of the scope. By restricting ourselves to these methods, we can both keep our survey to a manageable size and also concentrate at the basic ideas behind the different incremental approaches that are usually shared across a wider range of works. For the same reason, we have obviated incremental non-linear extensions of the above methods [39].

In the present work we will differentiate methods according to the *subspace model* used. From this viewpoint, two main categories of incremental subspace-based methods are usually considered depending on whether or not the above matrices are explicitly considered and computed (using different forms of decompositions) or not. Some of these variants are referred to in the literature as

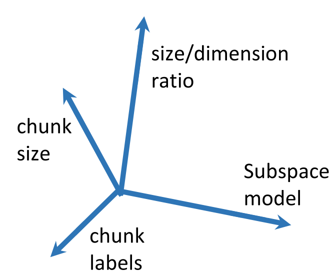

Fig. 1. Proposed taxonomy for subspace-based incremental feature extraction methods.

covariance-based or covariance-free methods. Table 1 summarizes all the papers considered in the present work according to the subspace model used and the computation (or not) of the above matrices.

To complete this multidimensional taxonomy, which is graphically illustrated in Fig. 1, different ways of feeding incremental algorithms are considered. The first one is in terms of the data size required for each update, which may range from one single sample at a time to moderate chunks of data. The second one is in terms of data labels, i.e. whether or not the set of labels in the corresponding classification problem is fixed beforehand or may grow arbitrarily along the incremental process. We will refer to these two aspects as *chunk size* and *chunk label* structure, respectively. Finally, we will also consider the *size/dimension* ratio, where we explicitly distinguish between the case in which the input space dimension is much greater than the expected data stream size and this constitutes a requirement or strongly conditions a particular method. Facing very small values of this ratio is usually referred as the small sample size (SSS) case.

The paper is organized around the above taxonomy paired with the discussion of the advantages and disadvantages inherent to each approach. The remainder of the paper is structured as follows. Section 2 describes the problem setting. Sections 3–5 contain an organized overview of incremental feature extraction based on PCA, LDA and DCV approaches, respectively. Section 6 shows a performance analysis of the incremental methods regarding their experimental setup and accuracy rates available in published results,

Please cite this article as: K. Diaz-Chito et al., An overview of incremental feature extraction methods based on linear subspaces, Knowledge-Based Systems (2018), https://doi.org/10.1016/j.knosys.2018.01.020