Universum-based multi-view matrix learning machine

1st Changming Zhu
College of Information Engineering
Shanghai Maritime University
Shanghai, P.R. China
cmzhu@shmtu.edu.cn

2nd Chengjiu Mei
College of Information Engineering
Shanghai Maritime University
Shanghai, P.R. China
544306495@qq.com

3rd Duoqian Miao
College of Electronics and Information
Tongji University
Shanghai, P.R. China
779561898@qq.com

4th Rigui Zhou
College of Information Engineering
Shanghai Maritime University
Shanghai, P.R. China
rgzhou@shmtu.edu.cn

5th Lai Wei
College of Information Engineering
Shanghai Maritime University
Shanghai, P.R. China
weilai@shmtu.edu.cn

6th Xiafen Zhang
College of Information Engineering
Shanghai Maritime University
Shanghai, P.R. China
xfzhang@shmtu.edu.cn

7th Min Yao
College of Information Engineering
Shanghai Maritime University
Shanghai, P.R. China
minyao@shmtu.edu.cn

***Abstract*—In real-world applications, most data sets consist of few labeled instances and many unlabeled instances. For those data sets, labeled ones can provide useful discriminant information for training a learning machines while unlabeled ones only provide less information. Thus most traditional learning machines have no ability to process this kind of data sets. In order to process the small-size-labeled problems, previous scholars develop Universum learning which can generate additional unlabeled instances with some discriminant information so as to enhance useful information for training a feasible learning machine. Moreover, many learning machines have been developed to process matrix instances including images, video, etc. These matrix learning machines have better classification performances compared with the original vector-instance-based learning machines including support vector machine when they process matrix instances. After that, some scholars develop learning machines to process both matrix instances and small-size-labeled data sets, for example, double-fold localized multiple matrix learning machine with Universum (UDLMMLM). While those learning machines have no ability to process multi-view data sets whose instances consist of multiple views and each view represents information for instances from a field. Thus, in this manuscript, we adopt UDLMMLM as a basic learning machine and extend its model to multi-view case. The new learning machine is named as Universum-based multi-view matrix learning machine (UMMLM). Experiments on some different kinds of data sets validate the effectiveness of UMMLM.**

***Index Terms*—Matrix learning, Universum learning, Multi-view learning**

I. Introduction

A. Background

Pattern recognition is a branch of machine learning that focuses on the recognition of instances and regularities in data [1]. For the recognition of instances, there are many widely used approaches, i.e., matrix learning, Universum learning, multi-view learning, etc.

1) Matrix learning: In real-world applications, to choose an appropriate representation for instances is necessary. In traditional pattern recognition applications, an instance is always represented by a point in a d-dimensional space [1]. Such a representation is treated as vector representation and can bring a convenience in mathematics. Instances with vector representation are called vector instances and the corresponding learning machine is named as vector-instance-based learning machine or vector learning machine. While vector learning machine cannot process matrix instances which are represented by a matrix representation including images or videos. Moreover, vectorizing a matrix instance to be a vector one so as to fit to the model of vector learning machine brings three potential problems [2]–[4]. One is the loss of some implicit structural or local contextual information, another is the requirement of a large memory, the third is the high risk of overtraining. To solve these problems, matrix-instance-based learning machine, i.e., matrix learning machine which can process matrix instances directly has been developed. Some classical matrix learning machines are matrix-instance-based Ho-Kashyap (HK) learning machine with regularization learning (MatMHKS) [5], new least squares support vector classification based on matrix instances (MatLSSVC) [6], and one-class support vector machines based on matrix instances (OCSVM) [7]. Related experiments have validated that matrix learning machines can reduce the computational complexity and improve the classification performance [3], [4], [8], [9].

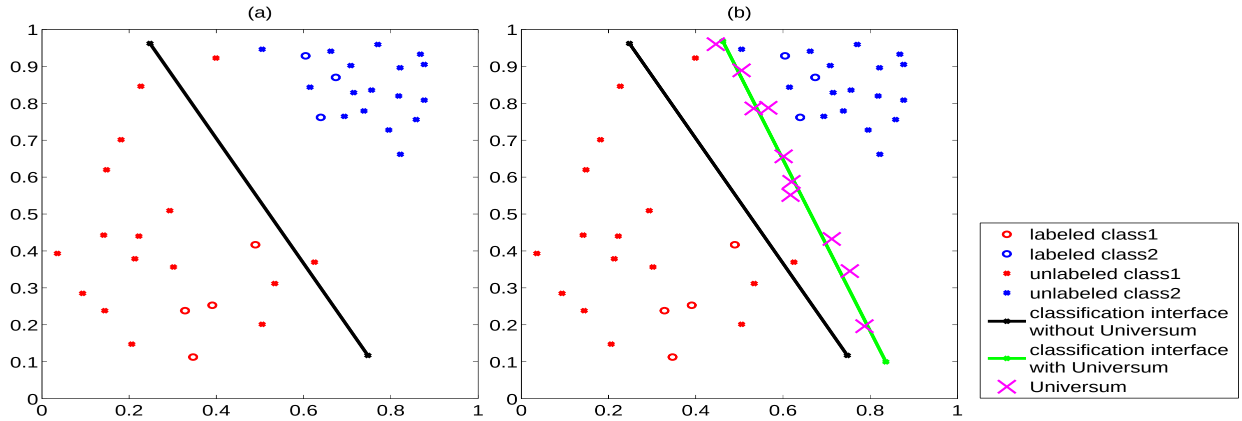

Fig. 1. (a) Example of semi-supervised learning with two classes. (b) Example of semi-supervised learning with two classes and Universum.

2) Universum learning: Moreover, in real-world applications, most data sets consist of labeled instances and unlabeled instances. In terms of these labeled ones, since the labels of instances are known beforehand, thus these instances can provide some useful discriminant information for training a feasible learning machine. On the contrary, those unlabeled ones can not provide enough information due to the labels are not known beforehand. For a data set, if the training set consists of all labeled instances, we name the data set as supervised data set. If the training set consists of labeled instances and some unlabeled ones, we name the data set as semi-supervised data set. Since most real-world data sets consist of few labeled training instances and more unlabeled training instances, thus in order to process the small-size-labeled problems, there are two kinds of solutions. First solution is adopting semi-supervised learning machines which are developed to process semi-supervised data sets directly. For these machines, both labeled and unlabeled original instances are used to train a learning machine. For example, multi-view semi-supervised classification via adaptive regression (MVAR) [10], co-labeling [11], sparse Markov chain-based semi-supervised multi-instance multi-label method (Sparse-Markov) [12], and semi-supervised multi-view hash model (SSMVH) [13] are present popular used semi-supervised learning machines. While due to the size of unlabeled instances is still large, thus these machines have to cost a longer training time. Furthermore, although current most semi-supervised learning machines achieve considerable success in the domain of machine learning [14], availability of only a few labeled instances may affect classification performance [15]. Please see Fig. 1-(a) which is also given in [15], the classifier or learning machine fails to learn a robust hyperplane with insufficient labeled instances and enormous unlabeled instances. Thus, in order to process this issue, some scholars develop the second solution, i.e., Universum learning which was developed by Vapnik et al. [16] initially. In general, Universum learning adopts labeled instances as the basic and collects or generates some instances which do not belong to any class of data, but do belong to the same domain as the problem and these collections or generations are named Universum instances which reflect some discriminant information. By Universum, we can obtain a robust decision hyperplane, please see Fig. 1-(b). Some references [15], [17]–[19] have validated that a learning machine with Universum has a better classification performance in some fields including body pose recognition [20], boosting strategy [21], dimensionality reduction technique [22], and multi-view learning [23], [24].

3) Multi-view learning: Multi-view learning aims to process multi-view data set which consists of instances with multiple views and each view is made up a feature group. For example, a video data set consists of videos from YouTube and each video appears in multiple varied forms, e.g., visual, audio, and text. Each form can be treated as a view of this data set. Furthermore, take text view as example, this view has several features including text color, size, content and these features form a feature group. Learning machines or approaches developed on the base of multi-view instances are named multi-view learning machines or multi-view approaches [25]–[32] and multi-view learning machines has been widely used in multi-view clustering [33], handwritten digit recognition [34], human gait recognition [35], image recognition [36], [37] and so on [38].

B. Problem

According to the above contents, each approach has many related learning machines and these learning machines are feasible for corresponding data sets. Moreover, many scholars combine these approaches into together for processing more complicated data sets. For example, double-fold localized multiple matrix learning machine with Universum (UDLMMLM) [39] is a combination of matrix learning and Universum learning and UDLMMLM can process both matrix instances and small-size-labeled data sets. While it is found that UDLMMLM has no ability to process multi-view data sets and furthermore, to the best of our knowledge, there is no learning machine is developed to process matrix instances, small-size-labeled data sets, and multi-view data sets simultaneously.

C. Proposal, contributions, and paper organization

Thus, according to what we said before, in order to process matrix data sets, small-size-labeled data sets, and multi-view data sets simultaneously, this manuscript adopts UDLMMLM as a basic learning machine and extends its model to multi-view case. The new learning machine is named as Universum-based multi-view matrix learning machine (UMMLM).

The contributions of UMMLM are (1) compared with traditional matrix learning, Universum learning, multi-view learning, UMMLM is the combination of them and has an ability to process matrix data sets, small-size-labeled data sets, and multi-view data sets simultaneously; (2) compared with the original traditional multi-view learning machines, since Universum learning is adopted, thus, our developed UMMLM has a better classification performance; (3) compared with UDLMMLM, UMMLM is feasible for both single-view and multi-view data sets.

The rest organization of this paper is given below. Section II reviews UDLMMLM. Section III gives the description of the proposed learning machine UMMLM. In section IV, the experimental results show the feasibility and effectiveness of UMMLM. Finally, conclusions are given in section V.

II. Review of UDLMMLM

UDLMMLM can process small-size-labeled data sets and matrix data sets simultaneously and its framework consists of two main steps. The first step is generating Universum instances and the second step is training the learning machine with labeled and Universum instances.

In terms of the first step, UDLMMLM adopts creating in-between Universum patterns (CIBU) method. Simply speaking, suppose there is a binary-class data set with N labeled matrix instances (A_i, φ_i), $i = 1, 2, ..., N$ and N' unlabeled matrix instances $A_{ui'}$, $i' = 1, 2, ..., N'$. $A_i \in R^{m\times n}$ or $(A_{ui'} \in R^{m\times n})$ and $\varphi_i \in \{+1, -1\}$ is the class label of A_i. Then, scholars only adopt the N labeled training instances and calculate the neighbor-distance matrix G. The i-th row and j-th column datum of G, i.e., G_{ij} is given below.

$$G_{ij} = \| A_i - A_j \|_2^2 \ , \ if \ A_i \in N_k(A_j) \ or \ A_j \in N_k(A_i) \quad (1)$$

If the condition $A_i \in N_k(A_j)$ or $A_j \in N_k(A_i)$ is not satisfied, $G_{ij} = +\infty$. $N_k(A_j)$ and $N_k(A_i)$ are the sets of the k-nearest neighbors of A_i and A_j, respectively, and A_j and A_i are the instances from different classes, respectively. In practice, k is set to be 3. When the value of G_{ij} is finite, CIBU creates the Universum instances from the middle of the shortest pathes between different training instances as below.

$$A_k^* = \frac{A_i + A_j}{2} \quad (2)$$

With Eq. (2), one can generate L Universum instances and $L \leq kN$. According to [39] said, these Universum instances can provide useful discriminant information and the size of them is always smaller than N'. Thus, when scholars process small-size-labeled matrix data sets, UDLMMLM can train a feasible learning machine and save the training time.

In terms of the second step, suppose there are N labeled matrix instances (A_i, φ_i), $i = 1, 2, ..., N$ and L Universum instances A_j^*, $j = 1, 2, ..., L$. Then similar with DLMMLM (i.e., double-fold localized multiple matrix learning machine [40]), since a matrix instance has multiple diverse matrix representations and a vector instance also can be matrixzed into different matrix representations [5], then each matrix representation has respective function to the training a learning machine, so scholars let *p*-th matrix representation of A_i be $A_i^p \in R^{m_p\times n_p}$ and *p*-th matrix representation of A_j^* be $A_j^{*p} \in R^{m_p\times n_p}$ where $m_p \times n_p = m \times n$. Then the optimization problem of UDLMMLM is given below.

$$\begin{aligned}
\min L = & \sum_{p=1}^{M}\left(\sum_{i=1}^{N}(\varphi_i g^p(A_i^p) - 1 - b_i^p)^2 + F\right) \quad (3)\\
& +\gamma \sum_{i=1}^{N}\sum_{p=1}^{M}\left(g^p(A_i^p) - \sum_{q=1}^{M}\eta_q(A_i^q) g^q(A_i^q)\right)^2 + \\
& E\sum_{r=1}^{M}\left[\sum_{j=1}^{L}\left(g^r(A_j^{*r}) - 1 - b_j^{*r}\right)^2\right] + \\
& D\sum_{j=1}^{L}\sum_{r=1}^{M}\left(g^r(A_j^{*r}) - \sum_{h=1}^{M}\eta_h(A_j^{*h}) g^h(A_j^{*h})\right)^2
\end{aligned}$$

where $F = C(u^{p^T} S_1^p u^p + \tilde{v}^{p^T} S_2^p \tilde{v}^p)$, $g^p(A_i^p) = u^{pT} A_i^p \tilde{v}^p + v_0^p$. M is the number of matrix representations. u^p, $\tilde{v}^p$, and v_0^p represent the left weight, right weight, and bias of a learning machine under the *p*-th matrix representation. $[b_1^p, ..., b_i^p, ..., b_N^p]^T$ composes the vector b^p and $[b_1^{*p}, ..., b_j^{*p}, ..., b_L^{*p}]^T$ composes the vector b^{*p}. b_i^p and b_j^{*p} represent the loose variable of a learning machine for A_i^p and A_j^{*p} respectively under the *p*-th matrix representation. C and E are the regularization parameters while γ and D are coupling parameters. $S_1^p = m_p I_{m_p\times m_p}$, $S_2^p = n_p I_{n_p\times n_p}$ are two regularization matrices corresponding to the u^p and $\tilde{v}^p$ respectively. The lengths of u^p and $\tilde{v}^p$ are respectively m_p and n_p.

In order to minimize Eq. (3), UDLMMLM adopts the two-step alternating optimization algorithm and details can be found in [39]. After the algorithm, one can get the optimal results of parameters u^q, $\tilde{v}^q$, and v_0^q, i.e., u_n^q, $\tilde{v}_n^q$, and ${v_0}_n^q$ under the *q*-th matrix representation. Then the final discriminant function of UDLMMLM which is used to label *i*-th test instance B_i is defined below where B_i^q is its *q*-th matrix representation and η_q is the weight of *q*-th matrix representation. If $g(B_i) > 0$, the instance B_i belongs to class $+1$ while if $g(B_i) < 0$, this instance belongs to class -1.

$$g(B_i) = \sum_{q=1}^{M}\eta_q({u_n^q}^T B_i^q \tilde{v}_n^q + {v_0}_n^q) \quad (4)$$

III. Framework of UMMLM

UMMLM is developed on the base of UDLMMLM and it can further process matrix data sets, small-size-labeled data sets, and multi-view data sets simultaneously. Suppose there is a matrix small-size-labeled multi-view data set $X = \{X^1, X^2, ..., X^v, ..., X^V\} = \{(A_1, \varphi_1), ..., (A_i, \varphi_i), ..., (A_N, \varphi_N), A_{u1}, ..., A_{ui'}, ..., A_{uN'}\}$ with V views, N labeled training instances, and $N^{'}$ unlabeled training instances. Here, $v = 1, 2, ..., V$, $i = 1, 2, ..., N$, and $i^{'} = 1, 2, ..., N^{'}$. X^v represents the v-th view and $X^v = \{A_1^v, ..., A_i^v, ..., A_N^v, A_{u1}^v, ..., A_{ui'}^v, ..., A_{uN'}^v\}$ where A_i^v and $A_{ui'}^v$ represent the features of A_i and $A_{ui'}$ under the v-th view. With such a definition, $A_i = \{A_i^1, ..., A_i^v, ..., A_i^V\}$ and $A_{ui'} = \{A_{ui'}^1, ..., A_{ui'}^v, ..., A_{ui'}^V\}$.

Then we should first to collect some Universum instances under each view. The collection method is same as CIBU given in [39] and under v-th view, there are L^v Universum instances are collected where the j-th Universum instance is A_j^{*v}, $j = 1, 2, ..., L^v$.

Then if we suppose x^v represents the term x under v-th view, the optimization problem of UMMLM is given below.

$$\min L^{'} = \sum_{v=1}^{V} L^v \tag{5}$$

where L^v represents the updated version of Eq. (3) under the v-view and each term x in Eq. (3) is replaced by x^v.

The optimization of Eq. (5) is similar with the one of UDLMMLM. Simply speaking, under v-th view and q-th matrix representation, the optimization of left weight, right weight, and the bias of the learning machine, i.e., u^{q^v}, $\tilde{v}^{q^v}$, and $v_0^{q^v}$ is same as the one of u^q, $\tilde{v}^q$, and v_0^q. If we denote the optimal results are $u_n^{q^v}$, $\tilde{v}_n^{q^v}$, and ${v_0}_n^{q^v}$, then we can use Eq. (6) to label the *i*-th test instance $B_i = \{B_i^1, ..., B_i^v, ..., B_i^V\}$ where $B_i^{q^v}$ is the *q*-th matrix representation of B_i^v, B_i^v is the features of B_i under v-th view, and $sign$ represents sign function. If $g(B_i) > 0$, B_i belongs to class $+1$ while if $g(B_i) < 0$, this instance belongs to class -1.

$$g(B_i) = \sum_{v=1}^{V} sign(\sum_{q=1}^{M} \eta_q ({u_n^{q^v}}^T B_i^{q^v} \tilde{v}_n^{q^v} + {v_0}_n^{q^v})) \tag{6}$$

IV. Experiments

A. Experimental setting

1) Setting of data sets: In order to validate the feasibility and practicality of UMMLM, we adopt some matrix data sets, small-size-labeled data sets, and multi-view data sets for experiments.

The matrix data sets used here are three image data sets which are also used in [39]. They are Coil-20, Letter-Image, and ORL. Information of them are given in Table I and Fig. 2. For Coil-20, it consists of instances from 20 categories and we select 72 instances for each category. Each instance is a 32×32 image. For Letter-Image, it consists of hand written digits from 0-9 and each digit is treated as a class which has 50 instances whose dimensions are 24×18. ORL is a face data sets with 40 persons and each person provides 10 images about faces with variable expressions. Each face has a dimensionality 32×20. For these three data sets, we select 20% instances for training, 30% instances for validation, and the rest 50% instances for testing.

The small-size-labeled data sets are Shaking and Woman. Information of them are given in Table I and Fig. 3. Different from the matrix data sets, these two data sets are collected from two moving objects in continuously changing environments. Both of these data sets are real-world data sets. For the reason that each frame of these two data sets is an image with RGB color, so each image has three dimensions. For Shaking, our objective is catching the head of the person who is playing the guitar. For Woman, we should catch the woman who is walking in the street. Moreover, for Shaking, we collect 365 frames and for Woman, we collect 550 frames. Then for each of them, we label 50% instances and the rest 50% instances are used for testing. Among the 50% instances, 40% (i.e., 20% of the whole data sets) instances are used for training and 60% (i.e., 30% of the whole data sets) instances are used for validation. For the validation part, although we know the labels of them, for the semi-supervised learning machines, we still regard them as unlabeled and adopt them along with the original labeled instances to train a leaning machine.

The multi-view data sets are Mfeat, Reuters, and Corel [41]. In terms of these three data sets, (1) Mfeat consists of hand written digits (0-9) [42] and each instance consists of six views, i.e., Fourier coefficients of the character shapes (fou), profile correlations (fac), Karhunen-Love coefficients (kar), pixel averages in 2×3 windows (pix), Zernike moments(zer), and morphological features (mor). Details of Mfeat can be found in Table II. (2) Reuters consists of machine translated documents which are written in five different languages which are treated as five views [43], [44]. These five languages are English (EN), French (FR), German (GR), Italian (IT), and Spanish (SP) and each document can be translated from one language to another language. Moreover, the documents are also categorized into six different topics, i.e., classes. Details of Reuters can be shown in Table III. (3) Corel is extracted from a Corel image collection [42] and it consists of 68040 photos from various categories. In our experiments, we randomly select 1000 photos from 10 categories and each category has 100 photos. The 10 categories are C0-Africa, C1-Beach, C2-Building, C3-Buses, C4-Dinosaurs, C5-Elephants, C6-Flowers, C7-Horses, C8-Mountains and C9-Food. For this data set, four views are adopted. They are color histogram (abbr. Col-h), color histogram layout (abbr. Col-hl), color moments (abbr. Col-m), and co-occurrence texture (abbr. Coo-t). Each view represents a feature set. Information of this data set is given in Table IV. For each multi-view data set, we also select 20% instances for training, 30% instances for validation, and 50% instances for testing.

2) Setting of learning machines: Then we compare our UMMLM with some matrix learning machines, learning machines to process small-size-labeled data sets, and multi-view learning machines as below.

TABLE I
DESCRIPTION OF THE USED MATRIX AND SMALL-SIZE-LABELED DATA SETS.

Order	Data set	No.dimension	No.class	No. instances
Matrix				
1	Coil-20	32×32	20	1440
2	Letter-Image	24×18	10	500
3	ORL	32×20	40	400
Small-size-labeled				
4	Shaking	$352 \times 624 \times 3$	1	365
5	Woman	$288 \times 352 \times 3$	1	550

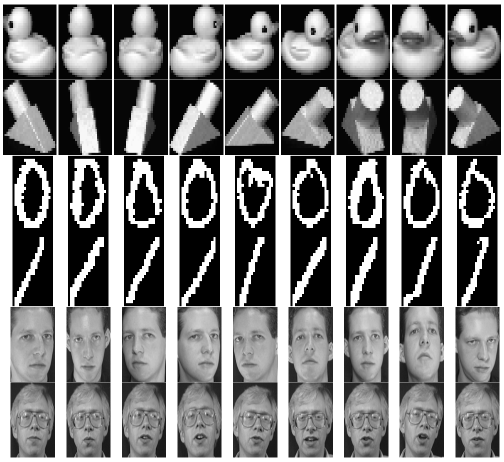

Fig. 2. Image data sets: the first and second rows show images from Coil-20, the third and fourth rows show ones from Letter-Image, and the fifth and sixth rows show ones from ORL.

TABLE II
DETAILED INFORMATION OF MFEAT DATA SET.

View	No. instances	No. features	No. digits
fac	2000	216	10
fou	2000	76	10
kar	2000	64	10
pix	2000	240	10
zer	2000	47	10
mor	2000	6	10

TABLE III
DETAILED INFORMATION OF REUTERS DATA SET.

View	No. documents	Vocabulary size
EN	18,758	21,513
FR	26,648	24,839
GR	29,953	34,279
SP	12,342	11,547
IT	24,039	15,506
Topic	No. documents	Per(%)
C15	18,816	16.84
CCAT	21,426	19.17
E21	13,701	12.26
ECAT	19,198	17.18
GCAT	19,178	17.16
M11	19,421	17.39

TABLE IV
DETAILED INFORMATION OF COREL DATA SET.

View	No. instances	No. features	No. categories
Col-h	1000	32	10
Col-hl	1000	32	10
Col-m	1000	9	10
Coo-t	1000	16	10

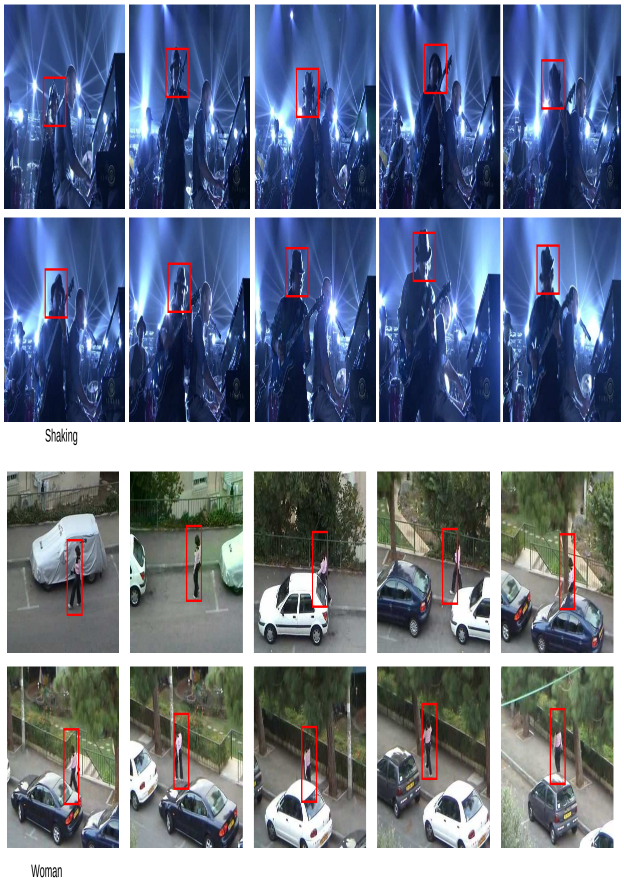

Fig. 3. Small-size-labeled data sets: the first and second rows show frames from Shaking, the third and fourth rows show frames from Woman.

The compared matrix learning machines are MatMHKS [5], MatLSSVC [6], OCSVM [7], DLMMLM [40], and UDLMMLM [39].

The compared learning machines to process small-size-labeled data sets include MVAR [10], co-labeling [11], Sparse-Markov [12], SSMVH [13], regularized matrix-pattern-oriented classification machine with Universum (RMMU) [19], and UDLMMLM [39].

The compared multi-view learning machines are multiple-view multiple-learner (MVML) [34], multi-view low-rank DL (MLDL) [36], multi-view linear discriminant analysis (MV-LDA) [45], multi-view canonical correlation analysis (MV-CCA) [46], multi-view locality preserving projections (MV-LPP) [47].

For all compared learning machines, the parameter settings can be found in respective references. For our UMMLM, its parameter setting is same as the one of UDLMMLM. Specially, the weight of each matrix representation is $\frac{1}{M}$ since according to [39] said, the influence of matrix representation is not too large.

3) Setting of experimental environment: We carry out the experiments with the below experimental environment: Intel dual-core processors, 2.66GHz strobe frequency, 4G RAM DDR, Win 7 operating system, and MATLAB R2014a.

4) Way to obtain the optimal experimental results: In order to get the optimal experimental results, we adopt the grid-search method. Namely, for one combination of the parameters, for each data set with the corresponding learning machines, with the restriction of the ratios of training set, validation set, and test set, we select the training, validation, and test instances in random. Then we repeat the experiments for ten times and get the average results. The parameters whose average experimental results are best are regarded as the optimal experimental parameters and the experimental results are regarded as the best ones.

B. Performance comparison

For different data sets, we adopt corresponding learning machines for experiments. Tables V, VI, and VII show the classification accuracy (%) and the standard deviation comparisons of the corresponding learning machines on different data sets respectively. According to these three tables, it is found that our proposed UMMLM has a best performance on different data sets in average. Indeed, our UMMLM can process matrix data sets, small-size-labeled data sets, and multi-view data sets simultaneously. Moreover, with Universum learning used, during the procedure of training, more instances with useful discriminant information can be generated and this operation enhances the performance of a learning machine. Furthermore, according to the result of standard deviation, we find the one of UMMLM is smallest which indicates the performance of UMMLM is most stable.

TABLE V
AVERAGE CLASSIFICATION ACCURACY (%) AND THE STANDARD DEVIATION COMPARISONS WITH DIFFERENT MATRIX LEARNING MACHINES ON CORRESPONDING DATA SETS.

	MatMHKS	MatLSSVC	OCSVM	DLMMLM	UDLMMLM	UMMLM
Coil-20	82.31	83.29	85.60	88.37	90.49	91.52
	± 0.46	± 0.38	± 0.55	± 0.81	± 0.70	± 0.21
Letter-Image	92.01	93.87	95.74	96.66	96.68	97.76
	± 0.79	± 1.11	± 0.38	± 0.65	± 1.04	± 0.20
ORL	83.41	85.12	87.22	89.00	89.73	90.29
	± 0.82	± 0.73	± 0.95	± 1.07	± 1.19	± 0.63

TABLE VI
AVERAGE CLASSIFICATION ACCURACY (%) AND THE STANDARD DEVIATION COMPARISONS WITH DIFFERENT LEARNING MACHINES TO PROCESS SMALL-SIZE-LABELED DATA SETS ON CORRESPONDING DATA SETS.

	MVAR	co-labeling	Sparse-Markov	SSMVH	RMMU	UDLMMLM	UMMLM
Shaking	73.12	75.20	76.54	79.08	81.50	83.74	85.09
	± 0.47	± 0.58	± 0.53	± 0.69	± 0.79	± 0.22	± 0.21
Woman	72.01	72.10	72.68	74.14	76.13	77.01	79.51
	± 0.48	± 0.07	± 0.44	± 0.45	± 0.36	± 0.62	± 0.02

TABLE VII
AVERAGE CLASSIFICATION ACCURACY (%) AND THE STANDARD DEVIATION COMPARISONS WITH DIFFERENT MULTI-VIEW LEARNING MACHINES ON CORRESPONDING DATA SETS.

	MVML	MLDL	MV-LDA	MV-CCA	MV-LPP	UMMLM
Mfeat	81.93	82.41	83.51	85.41	87.15	89.41
	± 1.10	± 0.55	± 0.86	± 0.94	± 1.06	± 0.48
Reuters	78.33	80.53	80.60	80.72	82.49	84.31
	± 0.30	± 0.80	± 0.47	± 0.49	± 0.49	± 0.20
Corel	84.76	86.24	87.67	88.20	90.78	91.46
	± 1.28	± 1.43	± 1.77	± 1.40	± 1.54	± 0.71

V. CONCLUSION

Matrix learning aims to design feasible learning machines to process matrix instances directly and Universum learning is feasible for processing small-size-labeled data sets. As a matrix learning machine with Universum, double-fold localized multiple matrix learning machine with Universum (UDLMMLM) has an ability to process both matrix instances and small-size-labeled data sets. While UDLMMLM cannot process multi-view data sets. Thus, in this manuscript, we adopt UDLMMLM as the basic and extend its model to multi-view version, the new developed learning machine is named as Universum-based multi-view matrix learning machine (UMMLM) and related experiments on some matrix data sets, small-size-labeled data sets, and multi-view data sets have validated that our proposed UMMLM outperforms the corresponding learning machines.

ACKNOWLEDGMENT

This work is supported by (1) Natural Science Foundation of Shanghai under grant numbers 16ZR1414500 and 16ZR1414400 (2) National Natural Science Foundation of China under grant numbers 61602296, 51575336, 61603245, and 41701523 (3) PuJiang talent plan under grant number 16PJ1403700 and the authors would like to thank their supports.

REFERENCES

[1] R.O. Duda, P.E. Hart, and D.G. Stork, Pattern Classification, Wiley-Interscience Press, New York, 2001.
[2] H. Wang and N. Ahuja, Rank-r approximation of tensors using image-as-matrix representation, IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2:346-353, 2005.
[3] J. Yang, D. Zhang, A.F. Frangi, and J. Yang, Two-dimensional PCA: a new approach to appearance-based face representation and recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence, 26(1):131-137, 2004.
[4] M. Li and B. Yuan, 2D-LDA: A novel statistical linear discriminant analysis for image matrix, Pattern Recognition Letters, 26(5):527-532, 2005.
[5] S.C. Chen, Z. Wang, and Y.J. Tian, Matrix-pattern-oriented ho-kashyap classifier with regularization learning, Pattern Recognition, 40(5):1533-1543, 2007.
[6] Z. Wang and S.C. Chen, New least squares support vector machines based on matrix patterns, Neural Processing Letters, 26(1):41-56, 2007.
[7] Y. Yan, Q. Wang, G. Ni, Z. Pan, and R. Kong, One-class support vector machines based on matrix patterns, International Conference on Informatics, Cybernetics, and Computer Engineering, pp. 223-231, 2012.
[8] J.P. Ye, Generalized low rank approximations of matrices, Machine Learning, 61(1):167-191, 2005.
[9] S.C. Chen, Y. Zhu, D. Zhang, and J.Y. Yang, Feature extraction approaches based on matrix pattern: Mat-PCA and Mat-FLDA, Pattern Recognition Letters, 26(8):1157-1167, 2005.
[10] H. Tao, C.P. Hou, F.P. Nie, J.B. Zhu, and D.Y. Yi, Scalable Multi-View Semi-Supervised Classification via Adaptive Regression, IEEE Transactions on Image Processing, 26(9):4283-4296, 2017.
[11] X.X. Xu, W. Li, D. Xu, and I.W. Tsang, Co-Labeling for Multi-View Weakly Labeled Learning, IEEE Transactions on Pattern Analysis and Machine Intelligence, 38(6):1113-1125, 2016.
[12] C. Han, J. Chen, Q.Y. Wu, S. Mu, and H.Q. Min, Sparse Markov chain-based semi-supervised multi-instance multi-label method for protein function prediction, Journal of Bioinformatics and Computational Biology, 13(5):1543001, 2015.

[13] C.H. Zhang and W.S. Zheng, Semi-supervised Multi-view Discrete Hashing for Fast Image Search, IEEE Transactions on Image Processing, 26(6):2604-2617, 2017.

[14] R. Sheikhpour, M.A. Sarram, S. Gharaghani, and M.A.Z. Chahooki, A Survey on semi-supervised feature selection methods, Pattern Recognition, 64:141-158, 2017.

[15] C.L. Liu, W.H. Hsaio, C.H. Lee, T.H Chang, and T.H. Kuo, Semi-Supervised Text Classification with Universum Learning, IEEE TRANSACTIONS ON CYBERNETICS, 46(2):462-473, 2016.

[16] V. Vapnik and S. Kotz, Estimation of dependences based on empirical data, Springer, 1982.

[17] J. Weston, R. Collobert, F. Sinz, L. Bottou, and V. Vapnik, Inference with the universum, The 23rd International Conference on Machine Learning, pp. 1009-1016, 2006.

[18] D. Zhang, J.D. Wang, F. Wang, and C.S. Zhang, Semi-Supervised Classification with Universum, Siam International Conference on Data Mining, 2(4):323-333, 2008.

[19] D.D. Li, Y.J. Zhu, Z. Wang, C.Y. Chong, and D.Q. Gao, Regularized Matrix-Pattern-Oriented Classification Machine with Universum, Neural Processing Letters, 45(3):1077-1098, 2017.

[20] B. Peng, G. Qian, and Y.Q. Ma, View-invariant pose recognition using multilinear analysis and the universum, Advances in Visual Computing, 5359:581-591, 2008.

[21] C. Shen, P. Wang, F. Shen, and H. Wang, *Uboost: boosting with the universum*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 34(4):825-832, 2012.

[22] X.H. Chen, S.C. Chen, and H. Xue, Universum linear discriminant analysis, Electronics Letters, 48(22):1407-1409, 2012.

[23] Z. Wang, Y.J. Zhu, W.W. Liu, Z.H. Chen , and D.Q. Gao, Multi-view learning with universum, Knowledge-Based Systems, 70:376-391, 2014.

[24] D.L. Liu, Y.J. Tian, R.F. Bie, and Y. Shi, Self-Universum support vector machine, Personal and Ubiquitous Computing, 18:1813-1819, 2014.

[25] D.R. Hardoon, S. Szedmak, and J Shawe-Taylor, Canonical correlation analysis: an overview with application to learning methods, Neural Computation, 16(12): 2639-2664, 2004.

[26] A. Sharma, A. Kumar, H. Daume, and D. W. Jacobs, Generalized Multiview Analysis: A discriminative latent space, IEEE Conference on Computer Vision and Pattern Recognition, 157: 2160-2167, 2012.

[27] M. Gönen and E. Alpaydin, Multiple Kernel Learning Algorithms, Journal of Machine Learning Research, 12: 2211-2268, 2011.

[28] G. Ye, D. Liu, I.H. Jhuo, and J. Huan, Robust late fusion with rank minimization, Computer Vision and Pattern Recognition, pp. 3021-3028, 2012.

[29] A. Blum and T. Mitchell, Combining labeled and unlabeled data with co-training, Eleventh Conference on Computational Learning Theory, pp. 92-100, 1998.

[30] W. Wang and Z.H. Zhou, Multi-view active learning in the non-realizable case, In Advances in Neural Information Processing System 23, pp. 2388-2396, 2010.

[31] Z.H. Zhou and M. Li, Semi-Supervised learning with very few labeled training examples, In Proceeding of the 22nd AAAI conference on Artificial Intelligence, pp. 675-680, 2007.

[32] Y. Zhu, W. Gao, and Z.H. Zhou, One-Pass Multi-View Learning, Journal of Machine Learning and Research, 30: 1-16, 2015.

[33] G. Tzortzis and A. Likas, Kernel-based Weighted Multi-view Clustering, 2012 IEEE 12th International Conference on Data Mining, pp. 675-684, 2012.

[34] S.L. Sun and Q.J. Zhang, Multiple-view Multiple-learner Semi-supervised Learning, Neural Processing Letters, 34:229-240, 2011.

[35] M.Q. Deng, C. Wang, and Q.F. Chen, Human Gait Recognition based on Deterministic Learning through Multiple Views Fusion, Pattern Recognition Letters, 78(C):56-63, 2016.

[36] F. Wu, X.Y. Jing, X.G. You, D. Yue, R.M. Hu, and J.Y Yang, Multi-view low-rank dictionary learning for image classification, Pattern Recognition, 50:143-154, 2016.

[37] S.H. Zhu, X. Sun, and D.L. Jin, Multi-view semi-supervised learning for image classification, Neurocomputing, 208:136-142, 2016.

[38] H.Y. Wang, X. Wang, J. Zheng, J.R. Deller, H.Y. Peng, L.Q. Zhu, W.G Chen, X.L. Li, R.J. Liu, and H.J. Bao, Video Object Matching across Multiple Non-overlapping Camera Views based on Multi-feature Fusion and Incremental Learning, Pattern Recognition, 47(12):3841-3851, 2014.

[39] Zhu C.M, Double-fold localized multiple matrix learning machine with Universum, Pattern Analysis and Applications 20(4):1091-1118, 2017.

[40] C.M. Zhu, Z. Wang, D.Q. Gao, and X. Feng, Double-fold localized multiple matrixized learning machine, Information Sciences, 295:196-220, 2015.

[41] Y.M. Xu, C.D. Wang, and J.H. Lai, Weighted Multi-view Clustering with Feature Selection, Pattern Recognition, 53:25-35, 2016.

[42] A. Asuncion and D. Newman, UCI machine learning repository, http://archive.ics.uci.edu/ml/, 2007.

[43] M.R. Amini, N. Usunier, and C.Goutte, Learning from multiple partially observed viewsłan application to multilingual text categorization, Neural Information Processing Systems (NIPS), pp. 28-36, 2009.

[44] http://multilingreuters.iit.nrc.ca/ReutersMultiLingualMultiView.htm

[45] A. Iosifidis, A. Tefas A, N. Nikolaidis, and I. Pitas, Multi-view human movement recognition based on fuzzy distances and linear discriminant analysis, Computer Vision and Image Understanding, 116(3): 347-360, 2012.

[46] J. Rupnik and J. Shawe-Taylor, Multi-View Canonical Correlation Analysis, In proceeding of Slovenian KDD Conference on Data Mining Data Warehouses, pp. 1-4, 2010.

[47] X. Yin, Q. Huang, and X. Chen, Multiple View Locality Preserving Projections with Pairwise Constraints, Communication Systems and Information Technology, 100: 859-866, 2011.