A New SSOPMV Learning for Matrix Data Sets

Changming Zhu, Chengjiu Mei, Rigui Zhou, Lai Wei, Xiafen Zhang and Min Yao

College of Information Engineering, Shanghai Maritime University, No. 1550, LinGang Ave., Shanghai, China.
Email: cmzhu@shmtu.edu.cn

Abstract. In real-world applications, most multi-view data sets are semi-supervised and large-scale. In order to process these data sets, scholars have developed semi-supervise done-pass multi-view learning (SSOPMV). While SSOPMV cannot process matrix data sets. Thus this manuscript extends the model of SSOPMV to matrix version and the new learning machine is named matrix-instance-based SSOPMV, i.e. (MSSOPMV). Related experiments validate that MSSOPMV can process multi-view, semi-supervised, large-scale, and matrix data sets well.

1. Introduction

1.1. Background-Classical Data Sets

In real-world applications, there are many different kinds of data sets, for example, multi-view data sets, semi-supervised data sets, large-scale data sets, and matrix data sets [1].

Multi-view data sets consist of instances with multiple views. For example, each video of a video data set appears in multiple different forms, i.e., visual, audio, and text. Each form is regarded as a view. Moreover, each view has many features. Take the text view for example, information of text view can be represented by text size, text color, and text shape, then size, color, and shape are three features of text view and these features compose a feature set. Then we name this data set as a multi-view data set.

Semi-supervised data sets consists of labeled training instances and unlabeled training instances. As we know, most of real-world multi-view data sets consist of label-known instances (i.e., labeled instances) and label-unknown instances (i.e., unlabeled instances). For example, there is a multi-view video data set in the database of YouTube and hundreds of hours of videos are uploaded to YouTube every minute, then due to the lack of staff, only a small part of these videos are labeled as different classes including art video, entertainment video, sport video and so on. We call them labeled videos. Most videos are not labeled and we call them unlabeled videos. For such a data set with labeled and unlabeled instances, if we only adopt labeled ones for updating and optimizing the model of a learning machine, we define it as supervised data set. If we adopt labeled and some unlabeled instances for training, we define it as semi-supervised data set.

Due to the coming of big-data age, more and more data sets are updated frequently. For example, hundreds of hours of videos are uploaded to YouTube every minute and we can regard YouTube as a large-scale video data set. Limited by the computation and storage ability, it is impossible for the traditional learning machines to store and process the whole data sets simultaneous.

Instances represented by a d-dimensional vector are convenient in mathematics and we name them as vector instances. While in real-world applications, more and more instances are represented in matrix form, i.e., a matrix instance $A \in R^{m \times n}$ and its dimension is $m \times n$. Data sets which consist of those matrix instances are matrix data sets and images and videos are classical matrix ones. Moreover, processing vector instances brings three potential problems. First one is the loss of some implicit structural or local contextual information of matrix instances. Second one is the requirement of a large memory. Third one is the high risk of overtraining. On the contrary, according to [1] said, if one uses matrix instances for learning and designs a matrix learning machine, we can reduce the computational complexity and improve the classification performance. So, studying matrix data sets is important.

1.2. Background-Corresponding Learning Machines

In order to process those kinds of data sets, corresponding learning machines are developed in recent years.

First, in order to process multi-view data sets, some scholars develop corresponding learning machines and according to reference [2], multi-view leaning machines can be classified into four groups: (1) multi-view subspace learning methods [3-4]; (2) pre-fusion methods [5]; (3) late-fusion methods [6-8]; (4) disagreement-based methods [9-11].

Second, locality sensitive discriminant feature (LSDF) [12], constraint scores for semi-supervised feature selection (CSSSFS) [13], semi-supervised online weighted multiple instance learning (SSOWMIL) [14], multi-view semi-supervised learning proposed by Zhu (MvSs-Zhu) [15], multiple-view multiple-learner (MVML) [16], adaptive multi-view selection (AMVS) [17] are widely used semi-supervised learning machines to process the related semi-supervised data sets.

Third, in order to process large-scale data sets, one-pass strategy is developed. For example, one-pass multi-view learning (OPMV) [18] and semi-supervised one-pass multi-view learning (SSOPMV) [1]. OPMV is a classical one-pass learning machine to process supervised multi-view data sets while SSOPMV aims to process semi-supervised multi-view data sets.

Fourth, in order to process matrix data sets, many learning machines including matrix-pattern-oriented Ho-Kashyap (HK) learning machine with regularization learning (MatMHKS) [19], new least squares support vector classification based on matrix patterns (MatLSSVC) [20], and one-class support vector machines based on matrix patterns (OCSVM) [21] have been developed.

What's more, in order to process more complicated data sets, some scholars combine these learning machines into together and develop more feasible learning machines. For example, double-fold localized multiple matrix learning machine with Universum (UDLMMLM) [22] is a combination of matrix learning and semi-supervised learning and UDLMMLM can process both matrix data sets and semi-supervised data sets. Furthermore, the mentioned SSOPMV is the combination of multi-view learning, semi-supervised learning, and one-pass learning and it can process semi-supervised data sets, large-scale data sets, and multi-view data sets simultaneous. Those above mentioned learning machines have been validated the effectiveness for different kinds of data sets.

1.3. Problem, Solution, Contributions

While to the best of our knowledge, there is no learning machine can process semi-supervised data sets, large-scale data sets, multi-view data sets, and matrix data sets simultaneous. Thus, in this manuscript, we adopt the SSOPMV as the basic and develop a matrix-instance-based SSOPMV (MSSOPMV). The contributions of the proposed MSSOPMV are (1) compared with traditional matrix learning, semi-supervised learning, multi-view learning, and one-pass learning, MSSOPMV is the combination of them and has an ability to process semi-supervised, large-scale, multi-view, and matrix data sets simultaneously; (2) compared with SSOPMV, MSSOPMV is feasible for both vector and matrix data sets.

1.4. Framework

Section 2 reviews OPMV and SSOPMV. Section 3 describes MSSOPMV. Section 4 shows the experimental results. Conclusions are given in section 5.

2. Review of OPMV and SSOPMV

2.1. OPMV

OPMV is used for supervised large-scale multi-view data sets and according to [18], suppose there is a supervised two-view data set S_n with n instances. Each labeled instance is $x_i = \{x_i^1, x_i^2; y_i\}$ denotes that this instance consists of information from the two views where $i \in \{1,2,\dots,n\} = [n]$. x_i^1 represents information from the first view and x_i^2 represents information from the second view. y_i is the class label. If $y_i = 1$, this instance is classified into class +1 while if $y_i = -1$ this instance is classified into class -1. Suppose the learning machine for the first view is ω_1 and the one for the second view is ω_2, and since OPMV processes instances one by one, then when the *i*-th instance arrives, the objective function of OPMV is given below where α_i is dual variable, ρ and λ are the penalty parameters, $\phi_i(\omega_1) = \ell(\langle\omega_1, x_i^1\rangle, y_i) + \lambda\Omega(\omega_1)$ and $\psi_i(\omega_2) = \ell(\langle\omega_2, x_i^2\rangle, y_i) + \lambda\Omega(\omega_2)$. Here, $\ell(\langle\omega_1, x_i^1\rangle, y_i)$ represents the classification result of x_i^1under the first view. For $\ell(\langle\omega_2, x_i^2\rangle, y_i)$, the meaning is same. If we have a corrected classification, the result is 0, otherwise, the result is 1.$\Omega(\omega_1) = \|\omega_1\|_2^2$ and $\Omega(\omega_2) = \|\omega_2\|_2^2$.

$$L_i(\omega_1, \omega_2, \alpha_i) = \phi_i(\omega_1) + \psi_i(\omega_2) - \alpha_i^2 + \frac{\rho}{2}\left(\langle\omega_1, x_i^1\rangle - \langle\omega_2, x_i^2\rangle + \alpha_i\right)^2 \tag{1}$$

OPMV minimizes the Eq. (1) and gets the optimal ω_1 and ω_2.Concretely speaking, OPMV updates them with each instance arrives, i.e,

$$\omega_1^{i+1} = \eta v_1^i - \beta_1^i \omega_1^i \tag{2}$$

where

$$v_1^i = -\nabla\phi_i(\omega_1^i) + \frac{1}{\eta}\omega_1^i + \rho(\langle\omega_2^i, x_i^2\rangle - \alpha_i)x_i^1 \tag{3}$$

$$\beta_1^i = \frac{\rho\eta^2\langle x_i^i, v_1^i\rangle}{1+\rho\eta\langle x_i^i, x_i^i\rangle} \tag{4}$$

$$\omega_2^{i+1} = \eta v_2^i - \beta_2^i \omega_2^i \tag{5}$$

where

$$v_2^i = -\nabla\psi_i(\omega_2^i) + \frac{1}{\eta}\omega_2^i + \rho(\langle\omega_1^{i+1}, x_i^1\rangle + \alpha_i)x_i^2 \tag{6}$$

$$\beta_2^i = \frac{\rho\eta^2\langle x_i^2, v_2^i\rangle}{1+\rho\eta\langle x_i^2, x_i^2\rangle} \tag{7}$$

Among these equations, ω_1^i and ω_2^i represent the ω_1 and ω_2 when *i*-th instance arrive. By these equations, after *n* instances arrive, we can get the optimal ω_1 and ω_2 and denote the optimal results are $\omega_1^o$and ω_2^o.Then for an unlabeled test instance $x_{new} = (x_{new}^1, x_{new}^2)$, one can predict its label as $\mathrm{y} = \mathrm{sign}(\langle\omega_1^o, x_{new}^1\rangle + \langle\omega_2^o, x_{new}^2\rangle)$.

2.2. SSOPMV

SSOPMV is the semi-supervised version of OPMV and it aims to process large-scale semi-supervised multi-view data set classification problems. According to [1] said, the procedure of SSOPMV consists of two main steps. First, once an unlabeled instance arrives, SSOPMV labels it with multiple view-based learning machines. If different view-based learning machine including ω_1 and $\omega_2$get the same labels, we add this unlabeled instance to the labeled set and update the learning machines including ω_1 and ω_2 with OPMV. If the labels are not same, we don't update the learning machines and wait the next instance arriving. Once a labeled instance arrives, we update the learning machines directly. Second, once we come through the present instances, for the left unlabeled ones, we adopt view-based learning machines to label them and update the learning machines according to the labelling results. If all unlabeled instances have been labeled or the learning machines cannot be updated again, we can end the procedure of SSOPMV and get the optimal learning machines.

3. Matrix-Instance-Based Semi-Supervised One-Pass Multi-View Learning (MSSOPMV)

With the limitation of manuscript length, we describe our MSSOPMV in simple. Suppose there is semi-supervised large-scale matrix multi-view data sets (in our manuscript, we use a two-view data set for example) $X = \{A_1, \ldots . A_i, \ldots, A_n, B_1, \ldots . B_j, \ldots, B_{n'}\}$ where i-th labelled instance is $A_i = \{A_i^1, A_i^2; y_i\}$ and j-th unlabelled instance is $B_j = \{B_j^1, B_j^2\}$. A_i^1, A_i^2 and B_j^1, B_j^2 have the same meaning of x_i^1, x_i^2. n and n' are the numbers of labelled instances and unlabeled instances respectively. Then if the initial learning machines are W_1^0 and W_2^0, the procedure is given below.

(1) When there is an instance arrives, if it is a labeled instance A_i, we update the learning machines W_1^i and W_2^i with the Eqs. (2)-(7). Notice, x_i^1, x_i^2 are replaced by A_i^1, A_i^2.

(2) When there is an instance arrives, if it is an unlabeled instance B_j and $\langle W_1^{j-1}, B_j^1 \rangle = \langle W_2^{j-1}, B_j^2 \rangle$, we update the learning machines W_1^{j-1} and W_2^{j-1} with the Eqs. (2)-(7). Notice, x_i^1, x_i^2 are replaced by B_j^1, B_j^2. Then we add B_j into the labeled set.

(3) When there is an instance arrives, if it is an unlabeled instance B_j and $\langle W_1^{j-1}, B_j^1 \rangle \neq \langle W_2^{j-1}, B_j^2 \rangle$, we don't update the learning machines and wait the next instance arrive.

(4) When all $n+n'$ instances have been processed once, we suppose the present learning machines are W_1^q and W_2^q. In terms of the left unlabeled instances, we use W_1^q and W_2^q to label them one by one and if $\langle W_1^q, B_j^1 \rangle = \langle W_2^q, B_j^2 \rangle$, we put B_j into the labelled set and update W_1^q and W_2^q, otherwise, we wait the next time to label them. Similar with what SSOPMV has done, if all unlabeled instances have been labeled or the learning machines cannot be updated again, we can end the procedure of MSSOPMV and get the optimal learning machines, i.e., W_1^o and W_2^o.

(5) Once we get W_1^o and W_2^o, for the unlabelled test instance C_{new}, its label is decided by $y = sign(\langle W_1^o, C_{new}^1 \rangle + \langle W_2^o, C_{new}^2 \rangle)$.

4. Experiments

4.1. Used Data Sets

In our experiments, the used data sets are given below. Details can be found in related references. For each data set, we divide it into 20% training instances, 30% validation ones, and 50% test ones.

Multi-view data sets: Mfeat, Reuters, Corel [23];

Semi-supervised data sets: IMDB, News Group [1];

Large-scale data sets: Video, News [1];

Matrix data sets: Coil-20, Letter-Image, ORL [24].

4.2. Compared Learning Machines

For the comparison, besides our MSSOPMV, the compared learning machines are given below. The details and parameter settings of these learning machines can be found in their respective references.

Multi-view learning machines: MvSs-Zhu [15], MVML [16], co-graph [25];

Semi-supervised learning machines: LSDF [12], CSSSFS [13],SSOWMIL [14], AMVS [17], UDLMMLM[22], SSOPMV [1];

One-pass learning machines: OPMV [18], SSOPMV [1];

Matrix learning machines: MatMHKS[19], MatLSSVC[20], OCSVM [21], DLMMLM (double-fold localized multiple matrixized learning machine) [24], UDLMMLM[22].

4.3. Classification Performance Comparisons

For different kinds of data sets, we adopt corresponding learning machines for experiments. The experiments are derived from 10-fold cross validation and we repeat each experiment for ten times. In order to get the optimal parameters, we adopt the grid-search strategy. Then we can get the average classification accuracy (%) and its standard deviation. Tables 1, 2, 3, and 4 show the related experimental results respectively. From those tables, it is found that our proposed MSSOPMV has a best performance on different data sets in average. Indeed, our MSSOPMV can process multi-view, semi-supervised, large-scale, and matrix data sets simultaneously. Furthermore, according to the result of standard deviation, we find the one of MSSOPMV is smallest which indicates the performance of MSSOPMV is most stable.

Table 1. Average classification accuracy (%) and the standard deviation comparisons with different multi-view learning machines on corresponding data sets.

	MvSs-Zhu	MVML	co-graph	MSSOPMV
Mfeat	82.12	82.24	83.56	85.28
	±5.84	±1.16	±5.42	±0.23
Reuters	73.12	74.86	76.10	78.04
	±5.54	±4.29	±6.59	±0.13
Corel	83.09	83.96	84.01	84.38
	±5.44	±8.20	±1.43	±0.66

Table 2. Average classification accuracy (%) and the standard deviation comparisons with different semi-supervised learning machines on corresponding data sets.

	LSDF	CSSSFS	SSOWMIL	AMVS	UDLMMLM	SSOPMV	MSSOPMV
IMDB	73.87	76.29	76.82	78.90	80.14	81.30	81.53
	±0.55	±4.12	±2.50	±6.50	±4.99	±6.38	±0.43
News Group	82.01	84.39	86.08	87.37	89.21	91.41	93.98
	±4.56	±7.52	±6.99	±2.66	±5.09	±3.32	±0.23

Table 3. Average classification accuracy (%) and the standard deviation comparisons with different one-pass learning machines on corresponding data sets.

	OPMV	SSOPMV	MSSOPMV
Video	86.12	87.27	87.78
	±6.55	±1.28	±1.22
News	84.92	87.50	89.60
	±5.36	±8.64	±0.25

Table 4. Average classification accuracy (%) and the standard deviation comparisons with different matrix learning machines on corresponding data sets.

	MatMHKS	MatLSSVC	OCSVM	DLMMLM	UDLMMLM	MSSOPMV
Coil-20	65.12	67.80	70.43	71.39	73.83	77.10
	±2.27	±6.30	±4.85	±2.09	±1.32	±0.39
Letter-Image	79.31	79.66	80.31	83.52	84.32	85.66
	±6.36	±0.64	±0.92	±6.39	±7.28	±0.15
ORL	73.26	75.72	77.44	80.47	82.88	84.33
	±5.38	±4.09	±6.32	±3.92	±3.30	±1.70

5. Conclusions

In this manuscript, we develop matrix-instance-based semi-supervised one-pass multi-view learning (MSSOPMV) which can process semi-supervised data sets, large-scale data sets, multi-view data sets, and matrix data sets simultaneously. Compared with traditional semi-supervised learning machines, one-pass learning machines, multi-view learning machines, and matrix learning machines, it is found that the developed MSSOPMV has a better and stable classification performance. What's more, with experiments on different kinds of data sets, our MSSOPMV has a wider application fields.

6. 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.

7. References

[1] Zhu CM, Wang Z, Zhou RG, Wei L, Zhang XF and Ding Y, 2018 *Semi-supervised one-pass multi-view learning* Neural Computing and Applications **online** 1-18, DOI: https://doi.org/10.1007/s00521-018-3654-3

[2] Ye HJ, Zhan DC, Miao Y, Jiang Y and Zhou ZH, 2015 *Rank Consistency based Multi-View Learning: A Privacy-Preserving Approach* ACM International on Conference on Information and Knowledge Management pp 991–1000

[3] Hardoon DR, Szedmak S, and Shawe-Taylor J, 2004 *Canonical correlation analysis: an overview with application to learning methods* Neural Computation 16(12): 2639–2664

[4] Sharma A, Kumar A, Daume H, and Jacobs DW, 2012 *Generalized Multiview Analysis: A discriminative latent space* IEEE Conference on Computer Vision and Pattern Recognition 157: 2160–2167

[5] Gonen M and Alpaydin E, 2011 *Multiple Kernel Learning Algorithms* Journal of Machine Learning Research 12: 2211–2268

[6] Ye G, Liu D, Jhuo IH, and Huan J, 2012 *Robust late fusion with rank minimization* Computer Vision and Pattern Recognition pp 3021–3028

[7] Fang YX, Zhang HJ, Ye YM, and Li XT, 2014 *Detecting hot topics from Twitter: A multiview approach* Journal of Information Science 40(5): 578–593

[8] Zhang HJ, Liu G, Chow TWS, and Liu WY, 2011 *Textual and Visual Content-Based Anti-Phishing: A Bayesian Approach* IEEE Transactions on Neural Networks 22(10): 1532–1546

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

[10] Wang W and Zhou ZH, 2010 *Multi-view active learning in the non-realizable case* In Advances in Neural Information Processing System 23 pp 2388–2396

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

[12] Zhao J, Lu K, and He X, 2008 *Locality sensitive semi-supervised feature selection* Neuro computing 71:1842–1849

[13] Kalakech M, Biela P, Macaire L, and Hamad D, 2011 *Constraint scores for semi-supervised feature selection: a comparative study* Pattern Recognition Letters 32: 656–665

[14] Wang ZH, Yoon S, Xie SJ, Lu Y, and Park DS, 2016 *Visual tracking with semi-supervised online weighted multiple instance learning* The Visual Computer 32(3): 307–320

[15] Zhu SH, Sun X, and Jin DL, 2016 *Multi-view semi-supervised learning for image classification* Neuro computing 208: 136–142

[16] Sun SL and Zhang QJ, 2011 *Multiple-view Multiple-learner Semi-supervised Learning* Neural Processing Letters 34: 229–240

[17] Yang ZK, Liu Z, Liu SY, Min L, and Meng WT, 2014 *Adaptive multi-view selection for semi-supervised emotion recognition of posts in online student community* Neuro computing 144: 138–150

[18] Zhu Y, Gao W, and Zhou ZH, 2015 *One-Pass Multi-View Learning* Journal of Machine Learning and Research 30: 1–16

[19] Chen SC, Wang Z, and Tian YJ, 2007 *Matrix-pattern-oriented ho-kashyap classifier with regularization learning* Pattern Recognition 40(5): 1533–1543
[20] Wang Z and Chen SC, 2007 *New least squares support vector machines based on matrix patterns* Neural Processing Letters 26(1): 41–56
[21] Yan Y, Wang Q, Ni G, Pan Z, and Kong R, 2012 *One-class support vector machines based on matrix patterns* International Conference on Informatics, Cybernetics, and Computer Engineering pp 223–231
[22] Zhu CM, 2017 *Double-fold localized multiple matrix learning machine with Universum* Pattern Analysis and Applications 20(4): 1091–1118
[23] Xu YM, Wang CD, and Lai JH, 2016 *Weighted Multi-view Clustering with Feature Selection* Pattern Recognition 53:25–35
[24] Zhu CM, Wang Z, Gao DQ, and Feng X, 2015 *Double-fold localized multiple matrixized learning machine* Information Sciences 295: 196–220
[25] Du YT, Li Q, Cai ZM, and Guan XH, 2013 *Multi-view semi-supervised web image classification via co-graph* Neuro computing 122: 430-440