Canonical Correlation Analysis with Bhattacharya Similarity Distance for Multiview Data Representation

Surendra Gupta[1], Urjita Thakar[1], and Sanjiv Tokekar[2]

[1] Department of Computer Engineering,
Shri Govindram Seksaria Institute of Technology and Science, Indore
{sgupta,uthakar}@sgsits.ac.in
[2] Department of Electronics and Telecommunication Engineering,
Institute of Engineering & Technology (IET) DAVV, Indore
stokekar@ietdavv.edu.in

Abstract. Many problems can be correctly tackled only when the related information is available from a various viewpoints. Thus when a problem is to be solved by machine, the data from multiple views needs to be well presented to a machine learning algorithm that suitably processes and learns. For a machine learning algorithm to be able to process such data, finding complementary information between view pairs is important. In this paper, a similarity distance based canonical correlation analysis(SDCCA) has been proposed to determine the complementary information by finding the similarity between view pairs for multiview data representation. The proposed approach uses the Bhattacharya similarity distance. It is evident from experimental results based on real-world data sets that the SDCCA approach performs better compared to existing canonical correlation analysis based approaches. Thus is a more effective and promising approach for solving real-life problems where consideration of complementary information in multiple views is essential.

Keywords: Canonical correlation analysis, Bhattacharya similarity distance, Feature extraction, Cross correlation, Multiview data representation.

1 Introduction

Many important real-life problems such as video surveillance, medical diagnosis, fashion industry[1], sentiment analysis[2], etc. can not be handled well through a single view. This is because a single view contains only partial information related to the problem. Thus consideration of data from multiple views of the same problem is essential. Such multiview data is heterogeneous in nature. Also feature set obtained from a single view cannot represent the essential latent space for machine learning.

Multiview learning (MVL) is to exploit the complementary and related information across multiple views [3]. Few researchers have [4–6] classified the exiting

MVL approaches into three major categories: co-training[7], multi-kernel learning[8] and subspace learning representation [9]. One of the classical approaches in MVL is canonical correlation analysis (CCA) first proposed by Hotelling[10].It is a subspace learning representation approach.
CCA finds pairs of projections from different views, such that the correlation between the transformed views in the common subspace is maximized. CCA looks for latent low dimensional information from two views of common subspace. To overcome this problem, Rupnik J. et al. [11] proposed multi-view CCA (MCCA) that generalizes two-views CCA. Also, PCA is used in MCCA to handle data set from multiple views jointly. The generalized CCA (GCCA)[12] explores the pairwise correlation of multiple views by searching the projection of each view by maximizing the sum of all pairwise correlation and maximizing the variance between views in the common subspace. GCCA and MCCA methods only preserve the correlation between different views but fail to explore the semantics information between views. All these variants of CCA find the linear transformation of views. To overcome this problem, Galen Andrew [13] proposed a Deep CCA (DCCA) that uses a complex nonlinear transformation of two views. DCCA also does not require an inner product that is used in CCA.
Multiple individual views may be high dimensional and redundant. To overcome this issue sparse based CCA (SCCA) [14, 15] and orthogonal canonical correlation analysis (OCCA)[16, 17] exist. Sparse based CCA enforces sparsity by using $l_1 - norms$ either as a regularization or by putting constraints on the projection vectors. Orthogonal based CCA is enforced by the construction of the orthogonal relationship between distinctive views.
To learn a compact and consistent representation by aggregating the variables from each view by exploiting inherent correlation across multiple views. CCA is equipped with different regularizations [18, 19]. CCA ignores the manifold structure of the single view itself. Manifold structure issues have been addressed by locality preserving CCA (LPCCA) [20]. For global structure preservation issues, multiset canonical correlations analysis based on low-rank representation (LRM-CCA) [21] has been proposed by researchers.
In real-world applications, data from multiple views often contain similar or complementary information. In general, most of the proposed CCA based approaches do not exploit the similarity and complementary information. Therefore existing approaches are not able to build comprehensive latent space for multiview learning. As a result, existing multiview learning algorithms do not perform well and fail to produce correct results. To address this issue, in this paper, a canonical correlation analysis with Bhattacharya similarity distance for multiview data representation (SDCCA) is proposed. Main contributions of the work are:

- Multiview learning is formulated by constructing latent space that exploits the similarly and complementary information between views.
- Within the latent space, LLP is also applied to explore the local manifold structure of the single view itself.

The rest of the paper is organized as follows: In section 2, the related work in the field of CCA is described. Section 3 presents the proposed SDCCA approach.

The theoretical analysis of the proposed methodology is also presented in this section. The experimental setup and result in a real data set are given in section 4, followed by a conclusion in section 5.

2 Related work

In this section, the work carried out by different researchers in the field of CCA, and its variants are presented. The concept of Bhattacharya Similarity Distance is also discussed.

2.1 Canonical Correlation Analysis (CCA)

The CCA finds a set of two projections so that the correlation between paired data sets is maximized in the common feature subspace. Let $X = \mathbb{R}^{n\times d_1}$ and $Y = \mathbb{R}^{n\times d_2}$ be the two input vectors. Where d_1 and d_2 are the dimensionality of X and Y views, respectively, and n is the total number of samples. CCA aims to find two projection vectors w_x and w_y that can maximize the correlation between them. The following optimization problem represents CCA objective function:

$$\max_{w_x,w_y} \rho_{x,y} = \frac{Con(w_x^T x, w_y^T y)}{\sqrt{\rho(w_x^T x)\rho(w_y^T y)}} \tag{1}$$

$$= \frac{w_x^T X\ Y^T w_y}{\sqrt{(w_x^T X X^T w_x)(w_y^T Y Y^T w_y)}} \tag{2}$$

The eq. (2) cab be reformulated as

$$\max_{w_x,w_y} w_x^T X Y^T w_y \tag{3}$$

$$s.t. \quad w_x^T X X^T w_x = 1, w_y^T Y Y^T w_y = 1$$

Equivalently it is represented in f-norm by

$$\max_{w_x,w_y} \|w_x X - Y w_y\|_F^2 \tag{4}$$

$$s.t. \quad w_x^T X X^T w_x = 1, w_y^T Y Y^T w_y = 1$$

CCA is viewed as learning a common subspace such that correlation between two data views is maximized. The common subspaces between two views can be formulated as

$$C = \frac{X w_x + Y w_y}{2} \tag{5}$$

Eq. (5) can be written in general form as:

$$U = \frac{1}{m}\sum_{v=1}^{m}(X_v w_v) \tag{6}$$

Where, m is total number of views, and U is universe subspace of all views.

2.2 Locality Preserving CCA (LP-CCA)

The LP-CCA finds a projection such that two input samples which are close to each other preserve the local structures of the input data into feature space. LP-CCA objective function can be written as

$$\min_{w_x,w_y} \sum_{i=1}^{n} \|(W_x^T(x_i - \bar{X}) - W_y^T(y_i - \bar{Y})\|^2 \tag{7}$$

$$s.t. \quad \sum_{i=1}^{n} \|W_x^T(x - \bar{x})\|^2 = 1 \ , \sum_{i=1}^{n} \|W_y^T(y - \bar{y})\|^2 = 1$$

2.3 A new Locality Preserving CCA (ALPCCA)

ALPCCA[22] exploits the local manifold structure of sample data by embedding the neighborhood information into CCA by adding the cross-view correlation between neighbors, besides the original correlation between paired views of themselves. ALPCCA use 'adding strategy' that is good for high dimensional data. ALPCCA is represented by the following optimization problem:

$$\max_{w_x,w_y} \frac{W_x^T c_{xy} W_y}{\sqrt{W_x^T c_{xx} W_x}\sqrt{W_y^T c_{yy} W_y}} \tag{8}$$

$$s.t. \quad W_x^T c_{xx} W_x = 1, W_y^T c_{yy} W_y = 1$$

where

$$c_{xy} = \sum_{i=1}^{n} x_i y_i^T + \sum_{i=1}^{n}\sum_{j=1}^{n} s_{ij}^x x_i y_j^T + \sum_{i=1}^{n}\sum_{j=1}^{n} s_{ij}^y x_i y_j^T$$

Here, s_{ij}^x and s_{ij}^y are local manifold information between samples, which are define as below

$$s_{ij}^x = \begin{cases} exp(-\|x_i - x_j\|^2/t_x) & if \ x_j \epsilon NE(x_i) \ or \ x_i \epsilon NE(x_j) \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

$$s_{ij}^y = \begin{cases} exp(-\|y_i - y_j\|^2/t_y) & if \ y_j \epsilon NE(y_i) \ or \ y_i \epsilon NE(y_j) \\ 0 & \text{otherwise} \end{cases} \tag{10}$$

2.4 Bhattacharya Distance

Bhattacharya distance(BD) [23] is used to measure the similarity between two views that realize that similar view pairs share less complementary information between the views. Let X and Y be two views over the same domain D, then Bhattacharya distance is defined as:

$$D_B(X\|Y) = -ln\,(BC(X\|Y)) \tag{11}$$

where, BC is Bhattacharya coefficient and it is defined as:

$$BD(X\|Y) = \sum_{s\epsilon D} \sqrt{X(s).Y(s)}$$

If both views are represented by normal vectors then Bhattacharya distance between two views is given by

$$D_B(X\|Y) = \frac{1}{4} ln\left(\frac{1}{4}\left(\frac{\sigma_X^2}{\sigma_Y^2} + \frac{\sigma_Y^2}{\sigma_X^2} + 2\right)\right) + \frac{1}{4}\left(\frac{(\mu_X - \mu_Y)^2}{\sigma_X^2 + \sigma_Y^2}\right) \tag{12}$$

where σ_X^2 is the variance of the X-th view, μ_X is the mean of the X-th view and same with Y view, where X,Y are the two different views.

3 Proposed Method

In this section, the proposed similarity based canonical correlation analysis is presented. Through this method, latent space for multiview learning is determined by incorporating similarities and complementary information from multiple views with cross correlation analysis. It also explores the local manifold structures within the view. SDCCA objective function is directly developed from original objective function of CCA.

Let a set of pair-wise sample data (X,Y) for multiview learning where X and Y are represented by $X = [x_1, x_2, x_3, ...x_n] \epsilon \mathbb{R}^{d_x}$ and $Y = [y_1, y_2, y_3, ...y_n] \epsilon \mathbb{R}^{d_y}$ with n instances and d_x and d_y dimension respectively. To target the goal, the objective function of SDCCA is given in the following form:

$$\max_{w_x, w_y} \frac{W_x^T \tilde{c}_{xy} W_y}{\sqrt{W_x^T c_{xx} W_x} \sqrt{W_y^T c_{yy} W_y}} \tag{13}$$

$$s.t. \quad W_x^T c_{xx} W_x = 1, W_y^T c_{yy} W_y = 1$$

Where

$$\begin{aligned} \tilde{c}_{xy} &= \sum_{i=1}^{n} x_i y_i^T + \sum_{i=1}^{n}\sum_{j=1}^{n} s_{ij}^x x_i y_j^T + \sum_{i=1}^{n}\sum_{j=1}^{n} s_{ij}^y x_i y_j^T + \sum_{i=1}^{n}\sum_{j=1}^{n} s_{ij}^{xy} x_i y_j^T \\ &= XY^T + Xs^xY^T + Xs^yY^T + Xs^{xy}Y^T \\ &= X(I + s^x + s^y + s^{xy})Y^T \end{aligned}$$

Here c_{xx} and c_{yy} are the XX^T and YY^T respectively. s^x, s^y are the neighborhood information between samples define in the eq. (9) and eq. (10).

s^{xy} is a dissimilarity between sample and it is defined as $1 - \tilde{s}^{xy}$, where $\tilde{s}^{xy}$ is similarity between views x and y. It is calculated using Bhattacharya similarity coefficient defined as:

$$\tilde{s}^{xy} = cos(\theta) = \sum_{i\epsilon N} \sqrt{x_i . y_i} \tag{14}$$

Thus, if the two samples are same then:

$$\tilde{s}^{xy} = cos(\theta) = \sum_{i \epsilon N} \sqrt{x_i . y_i} = 1 \tag{15}$$

It can be easily seen from eq. (15) that $s^{xy} = 0$, if and only if view x is equal to view y, and $s^{xy} = 1$, if and only if view x is orthogonal to view y. The dimensionality of data sets of different views may be different. Dimensionality of the data set with higher dimension is reduced by applying principal component analysis(PCA) to match it with that of the data set with smaller dimensionality. Thus resulting both the data sets become equal in dimension.
It can be easily seen that SDCCA will be equivalent to CCA if all s^x, s^y and s^{xy} contain all zero elements. i.e. local manifold information within the view and similarity information between views is absent. To solve SDCCA, the objective optimization function of eq. (13) can be reformulated as following equation:

$$\max_{w_x, w_y} W_x^T \tilde{c}_{xy} W_y \tag{16}$$

$$s.t. \quad W_x^T c_{xx} W_x = 1, W_y^T c_{yy} W_y = 1$$

The optimization problem of SDCCA given by eq. (16) can be rewritten as the Lagrangian optimization as follows:

$$\begin{bmatrix} & \tilde{c}_{xy} \\ \tilde{c}_{xy} & \end{bmatrix} \begin{bmatrix} w_x \\ w_y \end{bmatrix} = \lambda \begin{bmatrix} \tilde{c}_{xy} & \\ & \tilde{c}_{xy} \end{bmatrix} \begin{bmatrix} w_x \\ w_y \end{bmatrix} \tag{17}$$

Eq. (17) can be efficiently computed via the well-know singular value decomposition(SVD) to obtain the projective vector w_x and w_y. The complete algorithm of SDCCA is given in Algorithms 1.

Details of experiments carried out and discussion on results is given in next section.

4 Experiments and Results

To evaluate the performance and effectiveness of the proposed approach, experiments were conducted on two real-world data sets namely as Coil-20 data set[24] and UCI multiple features digit data set[25]. The Coil-20 data set consists of 20 categories of different objects, with 72 images of each object with total of 1440 color images. UCI digit data set consists of 10 classes of handwritten numbers '0'-'9'. There are 200 patterns per class with total of 2,000 patterns. PCA is applying on each view to find the PCA features set of views to get same dimensionality for all views. Using the Bhattacharya distance, the similarity matrix between the views is calculated. Next, SDCCA applied to extract features from multiview data. The well-known nearest neighbor classifier is used to evaluate the performance of the proposed SDCCA. Some samples images of coil-20 data set and UCI digit data set are shown in Fig. 1. (a) and (b) respectively.

Algorithms 1: SDCCA algorithm

Input : multiview training datasets $X\epsilon\mathbb{R}^{nxd_1}$,$Y\epsilon\mathbb{R}^{nxd_2}$
where d_1 and d_2 are the dimensionality of X and Y views.
Output: w_x and w_y

1. Construct $S^x, S^y and S^{xy}$:
    1.1 if x_i and y_j are in neighbors within the t_x kernel windows
        $S^x = exp(-\|x_i - x_j\|^2/t_x)$; otherwise 0.
    1.2 if y_i and x_j are in neighbors within the t_y kernel windows
        $S^y = exp(-\|y_i - y_j\|^2/t_y)$; otherwise 0.
    1.3 $s^{xy} = 1$ - $\sum_{i\epsilon N} \sqrt{x_i.y_i}$.
2. Define S $= I + S^x + S^y + S^{xy}$
3. Co-variance matrix:
    $\tilde{C}_{xy} = XSY^T$
5. Compute C_{xx} *and* C_{yy}
6. Compute matrix H $=C_{xx}^{-\frac{1}{2}}\tilde{C}_{xy}C_{yy}^{-\frac{1}{2}}$
7. Perform SVD decomposition on H: H$=UDV^T$.
8. Obtain $w_x = C_{xx}^{-\frac{1}{2}}U$ *and* $w_y = C_{yy}^{-\frac{1}{2}}V$

(a)
(b)

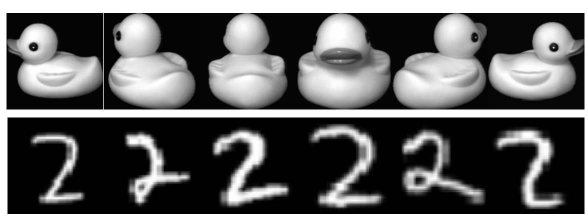

Fig. 1: (a) Sample of Coil-20 images (b) Sample of UCI digit images

All the experiments are implemented using the Python programming language on PC with 16GB RAM.
The proposed SDCCA algorithm was compared with CCA and ALPCCA to demonstrate the effectiveness of the proposed approach. Accuracy rate criteria are adopted for the evaluation of the proposed approach.

4.1 Performance Evaluation

Coil-20 data set: The Coil-20 data set is created by Columbia University, which consists of 20 categories of different objects, with 72 images of each object with a total of 1440 color images. These images were captured at a pose interval of 5^0. For experimental purpose (i) all images are resized into 64x64 pixels. (ii) for each view, three features gray scale intensity (GIS), Local binary pattern (LBS), and histogram of oriented gradient(HOG) have been extracted. For training of the SDCCA with Coil-20 data set 6/8/10/12 samples out of 72 views have been taken for four different training sessions and remaining samples used for testing. The latent space obtained after training is then subjected to nearest neighbor classification to verify appropriate inclusion of features and views.
Average results of 10 independent runs for each possible combination of views are shown in Table 1 for three different algorithms CCA, ALPCCA and SDCCA. It is observed that for all combinations of views SDCCA performed better. Hence using this method wider latent space gets created for multiview data representation that contains more complementary information between views.

Table 1: Comparison of accuracy of three algorithms on Coil-20 data set

Train	View1	View2	CCA	ALPCCA	SDCCA
TR-6	gis	lbs	71.23	71.89	81.24
	gis	hog	72.23	74.43	84.94
	lbs	hog	72.65	75.13	81.23
TR-8	gis	lbs	74.02	76.85	89.65
	gis	hog	79.93	79.82	91.12
	lbs	hog	77.34	79.76	89.32
TR-10	gis	lbs	82.12	84.23	91.23
	gis	hog	79.23	81.23	93.97
	lbs	hog	77.35	79.23	89.86
TR-12	gis	lbs	82.12	84.54	93.63
	gis	hog	83.87	88.75	94.23
	lbs	hog	78.43	84.12	95.13

UCI digit multiview data set: UCI-handwritten digit data set is a multiview data set that consists of features of handwritten numerals (‘0’–‘9’) extracted from a collection of Dutch utility maps. There are 200 patterns per class of 10 classes with total of 2,000 patterns. Data set has been digitized in binary images. These digits are represented in terms of six feature sets that include 6 morphological features(mor); 47 Zernike moments(zer); 64 Karhunen-Love coefficients(kar); 78 Fourier coefficients of the character shapes(fou); 216 profile correlations(fac) and 240 pixel averages in 2 x 3 windows(pix). Each time two sets from these 6 features set are chosen, so there are total 15($\binom{6}{2}$) different data combinations. For each combination 4-fold method is applied and two fold(50 percent) are applied for training and remaining fold for testing. Thus, total 6 ($\binom{4}{2}$)combination are used for the single iteration. Average results of independent runs for each possible combination are shown in Table 2 for three different algorithms CCA, ALPCCA and SDCCA. It is observed that for all combinations of views except one case SDCCA performed better. Hence using this method wider latent space get created for multiview data representation that consist of more complementary information between views.

Table 2: Comparison of accuracy of three algorithms on UCI digit multiview data set

S.No	View1	View2	CCA	ALPCCA	SDCCA
1	fac	fou	81.63	84.83	89.23
2	fac	kar	94.23	93.23	95.34
3	fac	mor	75.89	77.74	77.76
4	fac	pix	84.23	83.04	89.29
5	fac	zer	83.92	86.04	88.61
6	fou	kar	89.01	90.13	92.53
7	fou	mor	76.19	76.21	81.07
8	fou	pix	77.17	76.53	81.18
9	fou	zer	80.14	80.97	79.42
10	kar	mor	79.61	82.45	88.23
11	kar	pix	83.45	87.06	92.43
12	kar	zer	91.12	92.54	95.45
13	mor	pix	73.92	74.34	76.23
14	mor	zer	71.87	74.79	75.24
15	pix	zer	80.72	83.24	93.04

5 Conclusion

In this paper, an algorithm for multiview learning has been presented. The proposed similarity-based CCA method based on Bhattacharya Similarity Distance builds a wider latent space for multiview data representation. Similarity and cross-correlation between pair views have been embedded in the objective function. This enables complement views to contribute more in the latent space. It performed significantly better than CCA and ALPCCA. The performance has been validated over the two real-world data sets, namely Coil-20 and UCI digit multiview data set.

The current work is validated on the classification task with two views. It can be used for further tasks such as data visualization, clustering, etc., with more than two views.

References

1. Giovanni Battista Gardino, Rosa Meo, and Giuseppe Craparotta.: Multi-view latent learning applied to fashion industry, Information Systems Frontiers, pp 1-17, April 2020. doi:10.1007/s10796-020-10005-8
2. Siamak Faridani: Using canonical correlation analysis for generalized sentiment analysis, product recommendation and search, In Proceedings of the fifth ACM conference on Recommender systems, pp. 355–358, Oct. 2011. doi:10.1145/2043932.2044005
3. Shiliang Sun.:A survey of multi-view machine learning, Neural Computing and Applications, vol. 23, pp. 2031-2038, Feb. 2013. doi:10.1007/s00521-013-1362-6.
4. Guo Chenfeng and Wu Dongrui.: Canonical correlation analysis (cca) based multi-view learning: An overview, arXiv preprint arXiv:1907.01693, July 2019.doi:abs/1907.01693.
5. Tadas Baltrušaitis, Chaitanya Ahuja, and Louis-Philippe Morency.: Multimodal machine learning: A survey and taxonomy, IEEE transactions on pattern analysis and machine intelligence, vol. 41(2), pp. 423-443, Jan. 2018. doi:10.1109/TPAMI.2018.2798607
6. Jing Zhao, Xijiong Xie, Xin Xu, and Shiliang Sun.: Multi-view learning overview: Recent progress and new challenges, Information Fusion, vol. 38, pp. 43-54, Nov. 2017. doi:10.1016/j.inffus.2017.02.007
7. Avrim Blum and Tom Mitchell.: Combining labeled and unlabeled data with co-training, In Proceedings of the Eleventh Annual Conference on Computational Learning Theory, COLT-98, pp. 92-100, New York, USA, 1998. doi:10.1145/279943.279962
8. Mehmet Gönen and Ethem Alpaydin.: Multiple kernel learning algorithms. Journal of Machine Learning Research, vol. 12(64), pp. 2211-2268, 2011. URL http://jmlr.org/papers/v12/gonen11a.html.
9. C. Xu, D. Tao, and C. Xu.: Multi-view intact space learning, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 37(12), pp. 2531-2544, 2015. doi:10.1109/TPAMI.2015.2417578.
10. Hotelling and Harold.: Relations between two sets of variates, Biometrika, vol. 28(3-4), pp. 321-377, Dec. 1936. doi:10.1093/biomet/28.3-4.321.
11. Jan Rupnik and John Shawe-Taylor.: Multi-view canonical correlation analysis,In Conference on Data Mining and Data Warehouses(SiKDD 2010), pp. 1-4, Oct. 2010.
12. Horst Paul.: Generalized canonical correlations and their application to experimental data, Journal of clinical psychology, vol. 17,Oct. 1961. doi:10.1002/1097-4679(196110)17:4⟨331::AID-JCLP2270170402⟩3.0.CO;2-D
13. Galen Andrew, Raman Arora, Jeff Bilmes, and Karen Livescu.: Deep canonical correlation analysis, In International conference on machine learning, pp. 1247-1255, June 2013. doi:10.5555/3042817.3043076
14. Daniela M Witten, Robert Tibshirani, and Trevor Hastie.: A penalized matrix decomposition, with applications to sparse principal components and canonical correlation analysis, Biostatistics, vol. 10(3), pp. 515–534, April 2009. doi:10.1093/biostatistics/kxp008.

15. David R Hardoon and John Shawe-Taylor.: Sparse canonical correlation analysis, Machine Learning, vol. 83(3), pp. 331-353, 2011. doi:10.1007/s10994-010-5222-7.
16. Li Wang, Lei-hong Zhang, Zhaojun Bai, and Ren-Cang Li.: Orthogonal canonical correlation analysis and applications, Optimization Methods and Software, vol. 35(4), pp. 787-807, Jan. 2020. doi:10.1080/10556788.2019.1700257.
17. Hai Shu, Xiao Wang, and Hongtu Zhu.: D-cca: A decomposition-based canonical correlation analysis for high-dimensional datasets. In Journal of the American Statistical Association, vol. 115(529), pp. 292-306, 2020. doi:10.1080/01621459.2018.1543599.
18. Arthur Tenenhaus and Michel Tenenhaus.: Regularized generalized canonical correlation analysis, Psychometrika, vol. 76(2), pp. 257-284, April 2011. doi:10.1007/S11336-011-9206-8
19. M. Xu, Z. Zhu, X. Zhang, Y. Zhao, and X. Li.: Canonical correlation analysis with l2,1-norm for multiview data representation, IEEE Transactions on Cybernetics, vol. 50(11), pp. 4772-4782, April 2019. doi:10.1109/TCYB.2019.2904753.
20. Xiaofei He and Partha Niyogi.: Locality preserving projections, In Advances in neural information processing systems, pp. 53-160, Dec. 2003. doi:10.5555/2981345.2981365.
21. H. Zhang, J. Zhang, Y. Liu, and L. Jing.: Multiset canonical correlations analysis with global structure preservation, IEEE Access, vol. 8, pp. 53595-53603, March 2020. doi:10.1109/ACCESS.2020.2980964.
22. Zhang D. Wang F.: A new locality-preserving canonical correlation analysis algorithm for multi-view dimensionality reduction, Neural Process Lett., vol. 37, pp. 135-146, July 2012. doi:10.1007/s11063-012-9238-9
23. A. Bhattacharyya.: On a measure of divergence between two multinomial populations, SankhyÄ: The Indian Journal of Statistics, vol. 7(4), pp. 401-406, 1946. URL http://www.jstor.org/stable/25047882.
24. S Nayar, S Nene, and Hiroshi Murase.: Columbia object image library (coil 20), department of computer Science, Columbia University, Tech. Rep. CUCS-005-96, v2, Feb. 1996. https://www.cs.columbia.edu/CAVE/software/softlib/coil-20.php
25. Dheeru Dua and Casey Graff.: UCI machine learning repository - handwritten digits data set, 2017, URL https://archive.ics.uci.edu/ml/datasets/Multiple+Features.