Journal of Mobile Multimedia, Vol. 12, No.3&4 (2017) 265-276
© Rinton Press

SPARSE CANONICAL CORRELATION ANALYSIS FOR MOBILE MEDIA RECOGNITION ON THE CLOUD

YANJIANG WANG

College of Information and Control Engineering, China University of Petroleum
#66 Changjiang West Road, Huangdao District, Qingdao, Shandong 266580, China
yjwang@upc.edu.cn

BIN ZHOU

Shandong Wide Area Technology Co.,Ltd.
Dongying, Shandong 257081, China
freetzb@163.com

WEIFENG LIU

College of Information and Control Engineering, China University of Petroleum
Qingdao, Shandong 266580, China
liuwf@upc.edu.cn

HUIMIN ZHANG

College of Information and Control Engineering, China University of Petroleum
Qingdao, Shandong 266580, China

With the rapid development of the Internet technology and smartphone, people can easily capture and upload media information including text, audio, photos, and video. And then it becomes one critical demand to effectively and efficiently manage these personal multimedia that are often presented in multiple modalities. Canonical correlation analysis (CCA) has been widely employed for multi-modal data in many applications because of its promising performance in feature extraction and subspace learning for multivariate vectors. However, the traditional CCA may be difficult to interpret especially when the original variables are expected to involve only a few components. In this paper, we develop a mobile media recognition method on the cloud. Particularly, we propose sparse canonical correlation analysis (SCCA) on the cloud. SCCA can find a reasonable trade-off between statistical fidelity and interpretability. Furthermore, we employ a generalised power method to optimise the SCCA algorithm. Finally, we conduct extensive experiments for recognition on several popular databases including UCI datasets and USAA dataset. Experimental results demonstrate that the proposed SCCA algorithm outperforms the traditional CCA algorithm.

Key words: Canonical correlation analysis (CCA), sparse canonical correlation analysis (SCCA), power method, mobile media

1 Introduction

Today, smartphone has become more and more popular, and it can easily obtain many personal multimedia including speeches, images, and video that are often high dimensional data [12, 20, 24, 28]. To effectively manage the huge scale and high dimensional multimedia data, many subspace learning methods have been successfully applied for many applications by mapping the input data to a subspace with lower dimension [11, 21, 23, 25, 26]. The canonical correlation analysis (CCA) [1, 17, 18] is one promising subspace learning method, which exploits the correlation between two multidimensional variables in a linear way and has been widely employed in many applications such as bioinformatics, economics and signal processing. In particular, The CCA measures the correlation between two high-dimensional variables by linear transformation and then learns two corresponding subspaces by maximising the correlation between the pair of variables.

On the other hand, sparsity has been employed into many learning models to leverage the performance [14, 16]. However, similar with principal component analysis (PCA) [9, 13, 20, 21, 22] that gains the principal components (PCs) by linear transformation, the correlations of CCA also cannot be interpreted reasonably as the linear processes of PCA make the principal components difficult to explain. In other words, simple linear combination of original data makes the principal correlations equivocal especially when the original variables are expected to involve only a few components. To tackle this problem, sparse analysis is introduced to reach a balance between statistical fidelity and interpretability [2, 3, 10, 29]. Specifically, the sparse CCA finds a trade-off between the interpretability and the maximal variance projection of the same data, and further obtains better performance than the traditional CCA algorithm.

Secondly, although smartphones can capture image or video effortlessly, it is impossible to conduct image recognition or video classification independently for smartphones given small storage and limited computational resources. Therefore, it should adopt an alternative way to carry out the recognition tasks on mobiles.

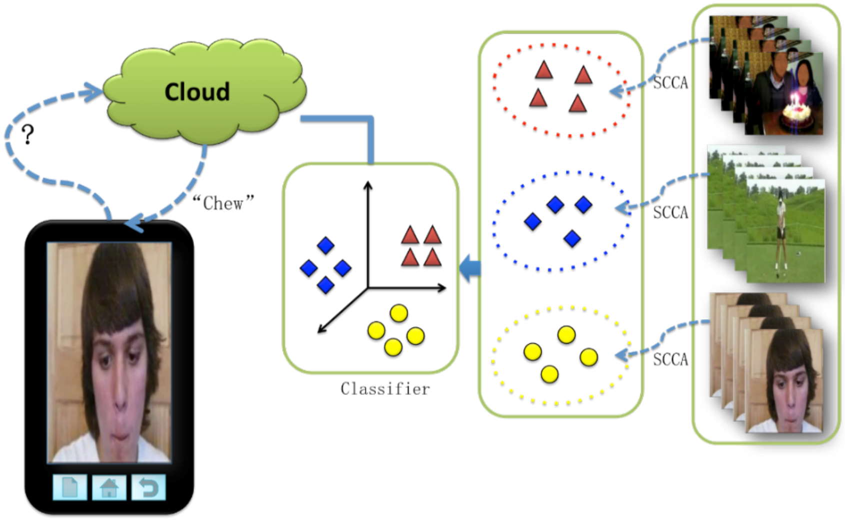

Fig. 1. The architecture of mobile media recognition on the cloud.

On the other hand, prodigious development of Internet technology promote cloud computing on mobile accessible. In this paper, we develop a mobile media recognition method by using cloud computing [19]. The proposed method contains the following steps:

- training the classifier on the cloud,
- capturing the media by mobile or selecting existed one on mobile,
- transferring the media data to the cloud,
- recognising the mobile media and sending the results back to mobile.

Figure 1 illustrates the architecture of the proposed method. In particular, we propose sparse canonical correlation analysis (SCCA) [27] on the cloud and we employ a generalised power method to optimise the SCCA algorithm. To evaluate the effectiveness of the proposed method, we conduct extensive experiments for recognition on several popular databases including UCI datasets and USAA dataset. Experimental results demonstrate that the proposed SCCA algorithm outperforms the traditional CCA algorithm.

The rest of this paper is assigned as follows. Section 2 reviews some related works. Section 3 presents the proposed SCCA algorithm. Section 4 reports experimental results in comparison with the traditional CCA algorithm. Finally, Section 5 gives the conclusion.

2 Related Works

In this section, we briefly review the related works of CCA and SCCA algorithms.

The CCA was first proposed by Hotelling [7] to find pairs of vectors by maximizing the correlation between a set of paired variables. Suppose we are given two variables x and y and their linear combinations $u = a^T x$ and $v = b^T y$. The CCA aims to find the two bases a and b for x and y respectively such that the correlations between the linear projections u and v are mutually maximised. The coefficient can be written as follows.

$$\begin{aligned}\rho &= \frac{E[uv]}{\sqrt{E[u^2]E[v^2]}} \\ &= \frac{x^T C_{uv} y}{\sqrt{x^T C_{uu} x y^T C_{vv} y}} \\ &= \frac{a^T C_{xy} b}{\sqrt{a^T C_{xx} a b^T C_{yy} b}}.\end{aligned}$$

where C_{uu} and C_{vv} are the self-correlation of u and v and $C_{uv} = C_{vu}^T$ is the cross-correlation of u and v. The problem of finding the maximum correlation coefficient equals to the problem of finding the maximum of ρ with respect to a and b is the maximum canonical correlation.

The basis vectors can be acquired by solving the equations following

$$\begin{cases} C_{xx}^{-1}C_{xy}C_{yy}^{-1}C_{yx}a = \gamma^2 a \\ C_{yy}^{-1}C_{yx}C_{xx}^{-1}C_{xy}b = \gamma^2 b \end{cases} \qquad (1)$$

where γ^2 are the squared canonical correlations and a and b are the corresponding normalized basis vectors. a and b can be found by the eigen vectors corresponding to the maximal eigen values. Solve one of the equations and the other equation can be solved correspondingly.

Kernel CCA was proposed by Fyfe and Lai [5], which projects the data into higher dimensional space and then perform the conventional CCA in the new space. Suppose $K_x = X^TX$ and $K_y = Y^TY$ be the linear kernel matrices corresponding to the two variables x and y, then the kernel CCA can be expressed as follows,

$$\max_{\alpha,\beta} \rho = \frac{\alpha^T K_x K_y \beta}{\sqrt{\alpha^T K_x^2 \alpha \beta K_y^T \beta}}.$$

Hardoon and Shawe-Taylor [6] introduced sparsity regularization into CCA to minimize the number of features used in both the primal and dual projections and then proposed sparse CCA as the following problem.

$$\min_{w,e} \|X^T w\text{-}Ke\|^2 + \mu\|w\|_1 + \gamma\|e\|_1, s.t. \|e\|_\infty = 1.$$

And recently, Luo *et al.* [15] proposed tensor CCA for multi-view dimension reduction.

3 Sparse Canonical Correlation Analysis

Figure 2 shows the framework of SCCA for recognition. In this section, we introduce the proposed sparse canonical correlation analysis.

Suppose $C_1 = \left(C_{xx}^{-1}C_{xy}C_{yy}^{-1}C_{yx}\right)\left(C_{xx}^{-1}C_{xy}C_{yy}^{-1}C_{yx}\right)^T$ and $C_2 = \left(C_{yy}^{-1}C_{yx}C_{xx}^{-1}C_{xy}\right)\left(C_{yy}^{-1}C_{yx}C_{xx}^{-1}C_{xy}\right)^T$. The eigen vectors in problem (1) can be easily expressed as

$$\emptyset(r) = \max_{a \in B^n} \sqrt{a^T C_1 a}$$

or

$$\emptyset(r) = \max_{b \in B^n} \sqrt{b^T C_2 b}.$$

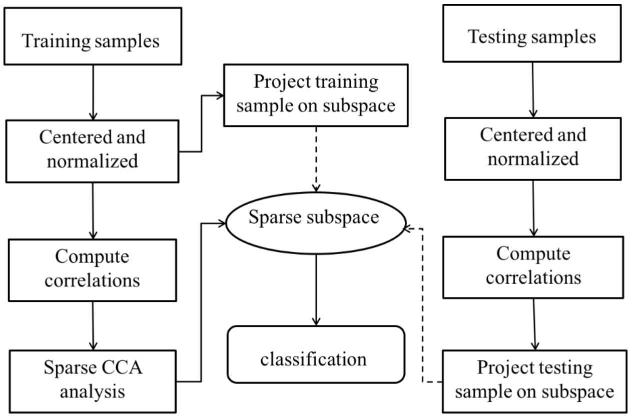

Fig. 2. The framework of SCCA for recognition.

As mentioned above, the SCCA employs sparsity penalty on the eigen vectors in problem (1) (*i.e.* a *or* b) and learns the corresponding subspaces of the two variables. We write the SCCA as the following problems,

$$\emptyset_{l_1}(r) = \max_{a\in B^n} \sqrt{a^T C_1 a}\text{-}r\|a\|_1 \qquad (2)$$

or

$$\emptyset_{l_1}(r) = \max_{b\in B^n} \sqrt{b^T C_2 b}\text{-}r\|b\|_1. \qquad (3)$$

The optimization of problem (3) is similar to that of problem (2). Hence we only consider the problem (2) in the following of this section.

For convenience, denote $A = \left(C_{xx}^{-1}C_{xy}C_{yy}^{-1}C_{yx}\right)^T$ (or $A = \left(C_{yy}^{-1}C_{yx}C_{xx}^{-1}C_{xy}\right)^T$ for problem (3)) and suppose $A \in R^{p\times n}$. Then the problem (2) can be rewritten as

$$\emptyset_{l_1}(r) = \max_{z\in B^n} \sqrt{z^T A^T A_z}\text{-}r\|z\|_1, \qquad (4)$$

where B^n is the unit Euclidean ball in R^n with $B^n = \{y \in R^n | y^T\ y \leq 1\}$ and r ≥ 0 is the parameter to control the sparsity. The above problem (4) provides one way to find the sparse eigen vectors with the largest eigen value.

In this paper, we employ a generalised power method to solve the above problem. Considering that the problem (4) is not convex or concave, Journée *et al.* [10] reformulate the problem as

$$\begin{aligned}\emptyset_{l_1}(r) &= \max_{z\in B^n} \sqrt{z^T A^T A_z}\text{-}r\|z\|_1 \\ &= \max_{z\in B^n} \max_{x\ in\ B^p} x^T A Z\ \text{-}r\|z\|_1 \\ &= \max_{x\ in\ B^p} \max_{z'\in B^n} \textstyle\sum_{i=1}^n \left|z_i'\right|(|a_i^T x|\text{-}r),\end{aligned}$$

where $z_i = sign(a_i^T x) z_i^{'}$. And finally the problem (4) can be translate to the following expression [10].

$$\emptyset^2_{l_1}(r) = \max_{x \in S^p} \sum_{i=1}^{n} [|a_i^T x| \text{-} r]_+^2 \, , \qquad (5)$$

where $S^p = \{y \in R^n | y^T y = 1\}$ is the unit Euclidean sphere.

The optimisation of problem (5) can be divided into two steps:

- finding the locally optimal patterns of zeros and nonzeros for z,
- computing the nonzero elements of z by solving the maximam variance problem.

A gradient method is adopted to find the locally optimal patterns. In particular, firstly it iteratively computes x with $x \leftarrow \sum_i^n [|a_i^T x| \text{-} r]_+ \text{ sign}(a_i^T x)\, a_i$ and $x \leftarrow \frac{x}{||x||}$. Then it constructs the vector $P \in \{0,1\}^n$ such that $p_i = 1$ if $|a_i^T x| > r$ and $p_i = 0$ otherwise.

With the optimal pattern vector P, we construct a matrix A_p that is a submatrix of A containing the columns related to the active entries of P. Conducting the singular value decomposition of the matrix A_p ,we have $A_p = \sigma u v^T$. Then we obtain the solution of problem (5) as $x^* = u, z_p^* = v$ and $z_{p'}^* = 0$, where $P^{'}$ is the complement of P.

After we obtain the sparse eigen vectors of the pair of the variables, we project the original samples onto the sparse subspace and then classify the testing samples to the related categories as illustrated in Figure 2.

4 Experiments

To evaluate the proposed method, we conduct extensive experiments on USAA database [8] and UCI datasets [4] for media recognition.

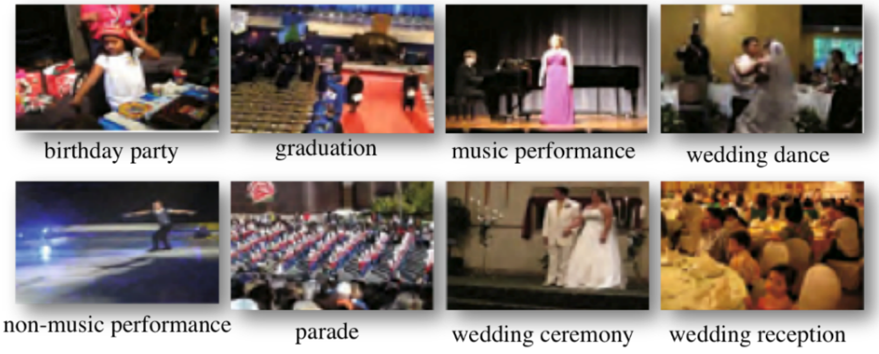

Fig. 3. Examples from the USAA database.

The USAA database is partly separated from CCV database with 8 video categories that are birthday party, graduation, music performance, non-music performance, parade, wedding ceremony, wedding dance, and wedding reception. The videos are with complex video scene information and tagged by 69 multi-modal binary attributes that vary in scene, audio, object, action, and camera movement. Each video sample in the database is represented by a 14000-dimension vector that consists three parts including 5000-dimension SIFT feature, 5000-dimension STIP feature and 4000-dimension MFCC feature. Figure 3 illustrates some examples of the USAA database. For recognition we choose 735 videos as the training set and the rest 731 videos as the testing set.

We choose three datasets from UCI database including Iris dataset, Letter-recognition dataset, and SPECT dataset.

Iris dataset [4] is a dataset about iris plants. This dataset contains 3 categories in 4 attributes. Each category has 50 samples and there are 150 samples in total. Table 1 shows some characteristics of the Iris database. We conduct two recognition experiments on Iris dataset. For the first experiment, we randomly choose 20 samples of each category as the training set, and the rest ones as testing set. For the second experiment, 30 samples are randomly chosen for training, and the other 20 ones for testing. Each experiment is conducted for 5 times.

Table 1. Some characteristics of the Iris database.

	Min	Max	Mean	Standard deviation	Correlation
Sepal length	4.3	7.9	5.84	0.83	0.7826
Sepal width	2.0	4.4	3.05	0.43	-0.4194
Petal length	1.0	6.9	3.76	1.76	0.9490
Petal width	0.1	2.5	1.20	0.76	0.9565

Letter-recognition dataset contains 20000 images of English alphabet from A to Z, and each image has 16 attributes. Figure 4 illustrates the example number of the Letter dataset. For recognition experiment, we conduct tree experiments with the number of training sample are 8000, 10000,12000 respectively.

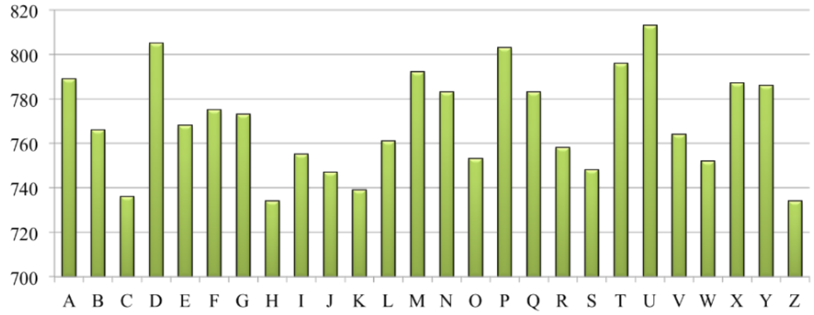

Fig. 4. Example number of the Letter dataset.

SPECT dataset consists of single proton emission computed tomography (SPECT) images of 267 patients with 22 attributes for each image. These patients are classified into two categories: normal and abnormal. For recognition experiment, we choose 80 samples as training set in which half are normal ones and others are abnormal. The rest 15 normal samples and 172 abnormal samples form the testing set.

We split the represent vector of each sample into two parts equally and conduct sparse canonical correlation analysis on the training set to learn the projective function for each subspace. Then we map all samples onto the corresponding subspace. Finally, we employ the nearest neighbour algorithm to classify the testing samples.

Figure 5 shows the performance of USAA database. We can see that the SCCA algorithm outperforms the CCA algorithm in most cases.

Figure 6 demonstrates the mean performance over all classes of Iris database. And Figure 7 and Figure 8 shows the average performance over each single class of different experiments respectively. We observe that the SCCA algorithm performs better than the CCA algorithm for both experiments for Iris recognition.

Figure 9 demonstrates the mean performance over all letters of the Letter-recognition dataset. Figure 10, Figure 11 and Figure 12 shows the performance over each single letter of different experiments respectively. From the results, we can see that the SCCA algorithm outperforms the CCA algorithm for letter recognition experiments.

Table 2 shows the performance of SPECT dataset. From Table 2, it is easy to see that SCCA performs better to recognize the abnormal patients than CCA. In general, SCCA can work more efficiently than CCA on the case of medical diagnosis.

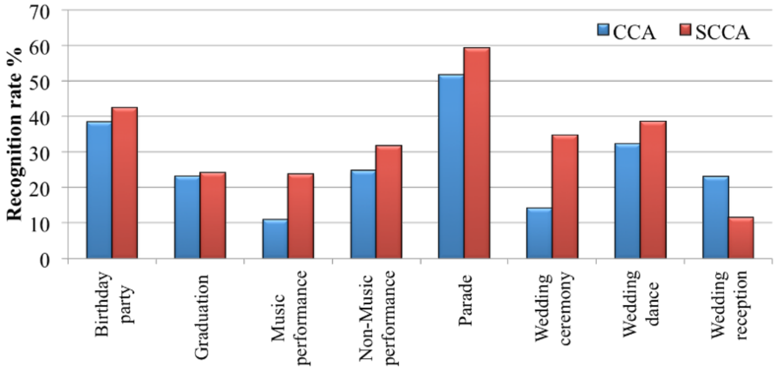

Fig. 5. The performance of USAA database ($r = 0.015$).

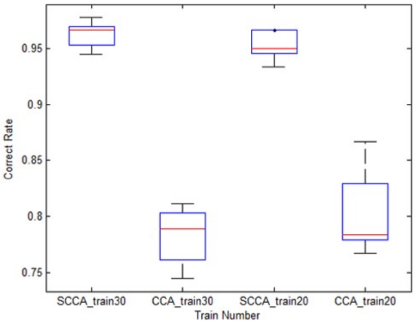

Fig. 6. The mean performance of Iris dataset.

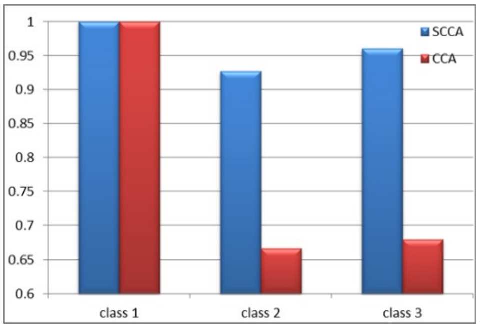

Fig. 7. The average performance of the first experiment on Iris dataset ($r = 0.9$).

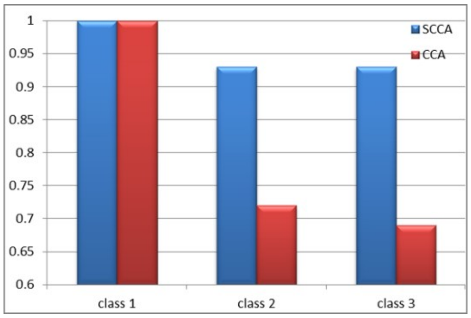

Fig. 8. The average performance of the second experiment on Iris dataset ($r = 0.8$).

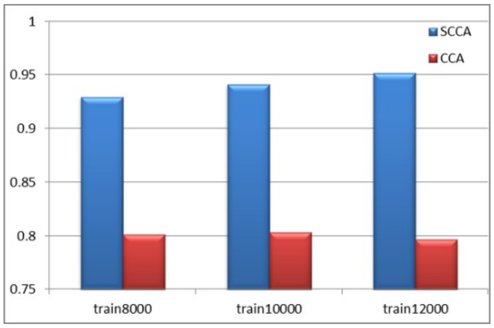

Fig. 9. The mean performance of Letter-recognition dataset.

Table 2. The performance of SPECT dataset ($r = 0.3$).

	CCA (%)	SCCA (%)
total rate	51.87	90.37
normal	20.00	20.00
abnormal	54.65	96.51

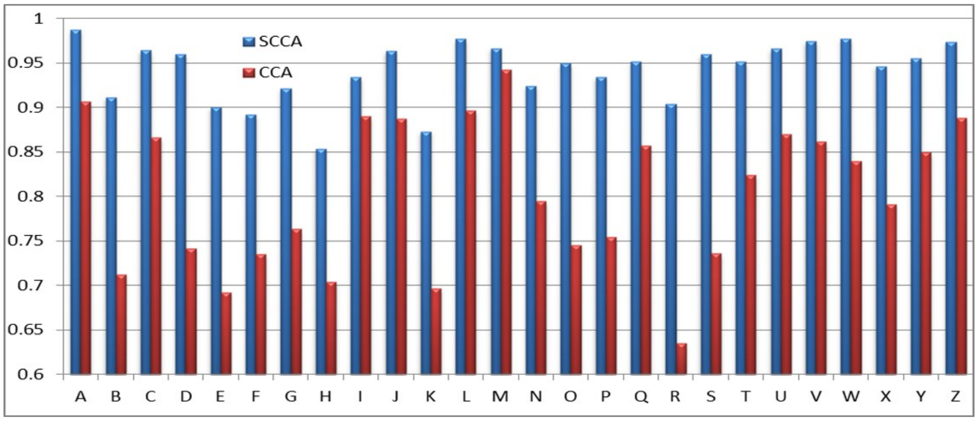

Fig. 10. The performance of the first experiment on Letter-recognition dataset (*r* = 0.07).

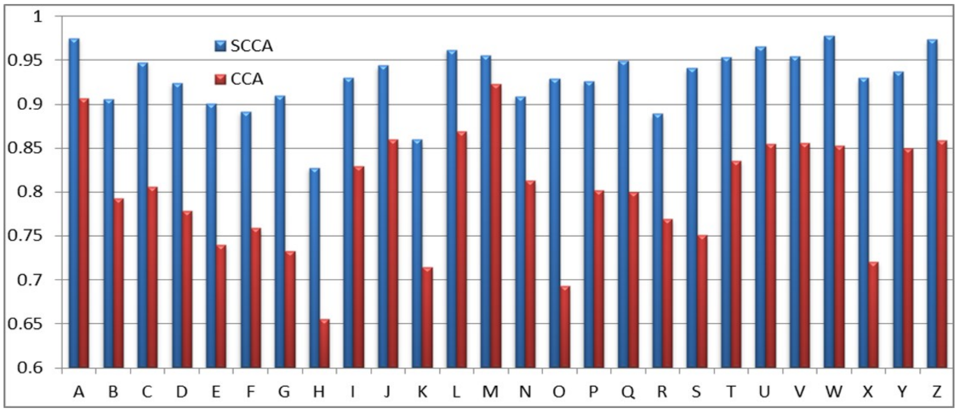

Fig. 11. The performance of the second experiment on Letter-recognition dataset (*r* = 0.07).

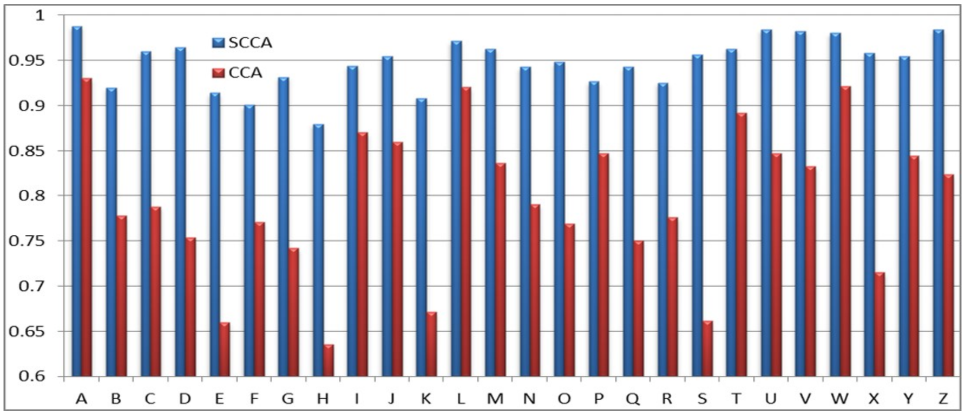

Fig. 12. The performance of the third experiment on Letter-recognition dataset (*r* = 0.08).

5 Conclusion

It is essential to find a proper way for mobile media recognition. And canonical correlation analysis (CCA) has achieved promising performance in feature extraction and subspace learning for multivariate vectors and hence been widely applied for multi-modal data in many applications. In this paper, we develop an mobile media recognition method by using cloud computing. We utilise sparse canonical correlation analysis (SCCA) on the cloud to find a reasonable trade-off between statistical fidelity and interpretability. The extensive experiments on popular databases including UCI datasets and USAA dataset verify the superiority of the proposed method comparing with baseline algorithms.

Acknowledgements

This work is supported by the National Natural Science Foundation of China (Grant No. 61671480, 61 ,61301242, 61271407).

References

1. Akaho, S. (2007), *A kernel method for canonical correlation analysis*, arXiv preprint cs/0609071.
2. d'Aspremont, A., Bach, F., Ghaoui, L.E. (2008), *Optimal solutions for sparse principal component analysis*, Journal of Machine Learning Research, vol. 9, pp. 1269-1294.
3. d'Aspremont, A., El Ghaoui, L., Jordan, M.I., Lanckriet, G.R. (2007), *A direct formulation for sparse pca using semidefinite programming*, SIAM review 49(3), 434-448.
4. Frey, P.W., Slate, D.J. (1991), *Letter recognition using holland-style adaptive classifiers*, Machine learning, 6(2), 161-182.
5. Fyfe, C., Lai, P.L. (2000), *Ica using kernel canonical correlation analysis*, In: In Proc. Int. Workshop on Independent Component Analysis and Blind Signal Separation (ICA2000).
6. Hardoon, D.R., Shawe-Taylor, J. (2011), *Sparse canonical correlation analysis*, Machine Learning, 83(3), 331-353.
7. Hotelling, H. (1936), *Relations between two sets of variates*, Biometrika, vol. 28, pp. 321-377.
8. Jiang, Y.G., Ye, G., Chang, S.F., Ellis, D., Loui, A.C.(2011), *Consumer video understanding: A benchmark database and an evaluation of human and machine performance*, In: Proceedings of the 1st ACM International Conference on Multimedia Retrieval, p. 29. ACM.
9. Jolliffe, I. (2002), *Principal component analysis*, Wiley Online Library.
10. Journée, M., Nesterov, Y., Richtárik, P., Sepulchre, R. (2010), *Generalized power method for sparse principal component analysis*, The Journal of Machine Learning Research, vol. 11, pp. 517-553.
11. Liu, W., Liu, H., Tao, D., Wang, Y., Lu, K. (2015), *Multiview hessian regularized logistic regression for action recognition*, Signal Processing, vol. 110, pp. 101-107.
12. Liu, W., Tao, D., Cheng, J., Tang, Y. (2014), *Multiview hessian discriminative sparse coding for image annotation*, Computer Vision and Image Understanding, vol. 118, pp. 50-60.
13. Liu, W., Zhang, H., Tao, D., Wang, Y., Lu, K. (2016), *Large-scale paralleled sparse principal component analysis*, Multimedia Tools and Applications, 75(3), 1481-1493.
14. Luo, Y., Tao, D., Geng, B., Xu, C., Maybank, S.J. (2013), *Manifold regularized multitask learning for semi-supervised multilabel image classification,* IEEE Transactions on Image Processing, 22(2), 523-536.
15. Luo, Y., Tao, D., Ramamohanarao, K., Xu, C., Wen, Y. (2015), *Tensor canonical correlation analysis for multi-view dimension reduction*, IEEE transactions on Knowledge and Data Engineering, 27(11), 3111-3124.
16. Luo, Y., Wen, Y., Tao, D., Gui, J., Xu, C. (2016), *Large margin multi-modal multi-task feature extraction for image classification*, IEEE Transactions on Image Processing, 25(1), 414-427.
17. Melzer, T., Reiter, M., Bischof, H. (2003), *Appearance models based on kernel canonical correlation analysis*, Pattern recognition, 36(9), 1961-1971.

18. Murtagh, F., Heck, A. (1987), *Multivariate data analysis*, Astrophysics and Space Science Library, vol. 131.

19. Tao, D., Jin, L., Liu, W., Li, X. (2013), *Hessian regularized support vector machines for mobile image annotation on the cloud*, IEEE Transactions on Multimedia, 15(4), 833-844.

20. Tao, D., Li, X., Wu, X., Maybank, S.J. (2007), *General tensor discriminant analysis and gabor features for gait recognition*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 29(10), 1700-1715.

21. Tao, D., Li, X., Wu, X., Maybank, S.J. (2009), *Geometric mean for subspace selection*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 31(2), 260-274.

22. Tao, D., Lin, X., Jin, L., Li, X. (2015), *Principal component 2-d long short-term memory for font recognition on single chinese characters*, IEEE Transactions on Cybernetics, 46(3), 756-765.

23. Tao, D., Tang, X., Li, X., Wu, X. (2006), *Asymmetric bagging and random subspace for support vector machines-based relevance feedback in image retrieval*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 28(7), 1088-1099.

24. Wang, M., Li, W., Liu, D., Ni, B., Shen, J., Yan, S. (2015), *Facilitating image search with a scalable and compact semantic mapping*, IEEE Transactions on Cybernetics, 45(8), 1561-1574.

25. Wang, M., Ni, B., Hua, X.S., Chua, T.S. (2012), *Assistive tagging: A survey of multimedia tagging with human-computer joint exploration*, ACM Computing Surveys (CSUR), 44(4), 25.

26. Yu, J., Wang, M., Tao, D. (2012), *Semisupervised multiview distance metric learning for cartoon synthesis,* IEEE Transactions on Image Processing, 21(11), 4636-4648.

27. Zhang, H., Liu, W., Zha, Z.J.(2015), *Sparse canonical correlation analysis for recognition,* In: Proceedings of the 7th International Conference on Internet Multimedia Computing and Service, p. 17. ACM.

28. Zheng, H., Wang, M., Li, Z. (2010), *Audio-visual speaker identification with multi-view distance metric learning*, In: Image Processing (ICIP), 2010 17th IEEE International Conference on, pp. 4561-4564.

29. Zou, H., Hastie, T., Tibshirani, R. (2006), *Sparse principal component analysis*, Journal of computational and graphical statistics, 15(2), 265-286.