

PCA-guided k -Means Clustering With Incomplete Data

Katsuhiro Honda, Ryoichi Nonoguchi, Akira Notsu, Hidetomo Ichihashi

Department of Computer Science and Intelligent Systems

Osaka Prefecture University

1-1 Gakuen-cho, Nakaku, Sakai, Osaka 599-8531, Japan

Email: {honda, notsu, ichi}@cs.osakafu-u.ac.jp

Abstract—This paper considers k -Means clustering of incomplete data sets including missing values. Although the main purpose of k -Means clustering is to partition samples into several homogeneous clusters by minimizing within-cluster errors, it has been shown that a relaxed solution of k -Means can be recovered in a PCA-guided manner. In this paper, the PCA-guided k -Means procedure is extended to a situation in which some observations are missing. Principal component scores, which can be identified with a rotated solution of cluster indicators of k -Means clustering, are estimated in an iterative process without imputation. Besides solving the eigenvalue problem of covariance matrices, k -Means-like partitions are derived through lower rank approximation of the data matrix ignoring missing elements. Several experimental results demonstrate that the PCA-guided process is more robust to initialization problems even though it is based on iterative optimization, just as the k -Means procedure is.

Keywords— k -means clustering, missing value, principal component analysis.

I. INTRODUCTION

k -Means clustering [1] is a fundamental technique for revealing intrinsic cluster structures of multivariate data sets. Although the k -Means process has been demonstrated to be very useful for extracting spherical clusters based on iterative optimization, it often suffers from initialization problems such that bad initializations lead to convergence to the wrong local minima. Recently, Ding and He [2] discussed a close relation between k -Means solutions and principal component scores, and proposed the principal component analysis (PCA)-guided k -Means procedure based on a deterministic method for finding the optimal (unique) k -Means partition. In the clustering model, a relaxed k -Means cluster indicator is identified with a rotated principal component score matrix, although the rotation matrix cannot be explicitly estimated and an approach such as visualization by ordering of samples in connectivity matrices [3] has to be used in order to visually access the cluster structures.

In [4], a noise rejection mechanism was introduced on the basis of noise fuzzy clustering [5]. In the fuzzy PCA-guided robust k -Means (FPR k -Means) procedure, both the responsibility weight of each sample with respect to the k -Means process and a rotated-cluster indicator are estimated by

This work was funded in part by the Ministry of Internal Affairs and Communications, Japan, under the Strategic Information and Communications R&D Promotion Programme (SCOPE) and in part by the Ministry of Education, Culture, Sports, Science and Technology, Japan, through a Grant-in-Aid for Scientific Research (20700214).

an iterative procedure. Removing the noise samples, a robust cluster indicator is estimated in a deterministic manner.

In addition to noisy data, incomplete data are common in real applications where some observations are missing. In order to apply k -Means-like clustering algorithms to incomplete data sets without a priori imputation, Hathaway and Bezdek [6] introduced four strategies in the Fuzzy c -Means (FCM) clustering algorithm [7]. Although the simple whole data strategy (WDS) eliminating all incomplete samples can be adopted if there are only a few missing elements, it is useless in many real applications. In the partial distance strategy (PDS), clustering criteria of distances between samples and prototypes are calculated using observed coordinates only. In [6], it was also shown that some iterative completion strategies such as the optimal completion strategy (OCS) and nearest prototype strategy (NPS) also work well, although they require more computational effort.

In the application of PCA to incomplete data sets, several methods that extract principal components without any pre-processing of data have been proposed. Ruhe [8], Wiberg [9], and Shibayama [10] proposed PCA-like methods for capturing the structure of incomplete multivariate data without any imputations or statistical assumptions. Especially, Shibayama's approach made it possible to calculate principal components by a computationally efficient algorithm. Shum *et al.* [11] applied such techniques to object modeling. These methods are based on the lower rank approximation of the data matrix, which accomplishes the minimization of the least square criterion, and the methods extract principal components of the covariance matrix of the sample data set. Honda *et al.* [12] extended the PCA model to a linear fuzzy clustering algorithm by considering the close connection between PCA and linear fuzzy clustering algorithms such as Fuzzy c -Varieties (FCV) and Fuzzy c -Elliptotypes (FCE) [13].

In this paper, Shibayama's PCA-like method is applied to the k -Means problem with incomplete data in order to estimate a relaxed solution. Several experimental results demonstrate that the PCA-guided process is more robust to initialization problems, even though it is also based on iterative optimization, the same as the k -Means procedure.

The remainder of this paper is organized as follows: Section II briefly reviews k -Means and PCA-guided k -Means. Section III introduces Shibayama's PCA-like method for incom-

plete data, which is then applied to k -Means tasks. Section IV presents several experimental results to demonstrate the characteristic features of the proposed approach. Section V summarizes the conclusions of this paper.

II. BRIEF REVIEW OF k -MEANS AND PCA-GUIDED k -MEANS

A. k -Means Clustering and Partial Distance Strategy

Assume that we have n samples with m -dimensional observation $\mathbf{x}_i$, $i = 1, \dots, n$, and the goal is to partition the samples into several clusters where samples belonging to the same cluster are similar while samples belonging to different clusters are dissimilar.

k -Means [1] is a non-hierarchical prototype-based clustering in which prototypes (centroids) are used for representing clusters and the objective function is defined as the sum of within-cluster errors:

$$\begin{aligned} L_{km} &= \sum_{k=1}^K \sum_{i \in G_k} d_{ki} \\ &= \sum_{k=1}^K \sum_{i \in G_k} \|\mathbf{x}_i - \mathbf{b}_k\|^2, \end{aligned} \quad (1)$$

where K is the pre-defined number of clusters and $\mathbf{b}_k$ is the representative prototype (centroid) of cluster G_k .

The k -Means process is composed of two phases: prototype estimation and sample assignment. The two phases are iterated until the solution is trapped at a local minimum. Because of the greedy nature of the updating algorithm, the k -Means procedure often suffers from the initialization problem that the process converges to a different local optimum for a different initialization.

When some observations are missing, we need a modified clustering criterion d_{ki} that can be reconstructed using only the remaining observations. PDS is a simple but useful approach in which squared distances are calculated considering the observed coordinates only.

$$d_{ki} = \sum_{j=1}^m z_{ij} (x_{ij} - b_{kj})^2, \quad (2)$$

where z_{ij} is a bivariate variable as follows:

$$z_{ij} = \begin{cases} 1 & ; \text{ if } x_{ij} \text{ is observed.} \\ 0 & ; \text{ if } x_{ij} \text{ is missing.} \end{cases} \quad (3)$$

B. PCA-guided k -Means

Ding and He [2] revealed a close relation between PCA and k -Means clustering. The k -Means objective function of Eq.(1) can be re-defined by a centroid-less formulation as follows:

$$L_{km} = \sum_{i=1}^n \|\mathbf{x}_i\|^2 - \sum_{k=1}^K \frac{1}{n_k} \sum_{i,j \in G_k} \mathbf{x}_i^\top \mathbf{x}_j, \quad (4)$$

where n_k is the number of samples belonging to cluster G_k . Because the first term is a constant, the optimal k -Means partition is derived by minimizing the second term, which is the

sum of within-cluster (inner product) similarities. In this context, the (hard) k -Means cluster partition is represented by K non-negative bivariate indicator vectors $H_K = (\mathbf{h}_1, \dots, \mathbf{h}_K)$,

$$h_{ki} = \begin{cases} \frac{1}{n_k^{1/2}} & ; \text{ if sample } i \text{ belongs to cluster } G_k. \\ 0 & ; \text{ otherwise.} \end{cases}$$

In order to relax the k -Means problem, Ding and He introduced a rotated solution Q_K by considering $K \times K$ orthogonal transformation $T = \{t_{ij}\}$,

$$Q_K = (\mathbf{q}_1, \dots, \mathbf{q}_K) = H_K T, \quad (5)$$

where, avoiding redundancies, the last column of T is set as:

$$\mathbf{t}_n = (\sqrt{n_1/n}, \dots, \sqrt{n_K/n})^\top, \quad (6)$$

which brings $\mathbf{q}_K = (\sqrt{1/n}, \dots, \sqrt{1/n})^\top$. Considering the mutual orthogonality of $\mathbf{h}_k$, we have the following conditions:

$$Q_{K-1}^\top Q_{K-1} = I_{K-1}, \quad (7)$$

$$\sum_{i=1}^n q_{ki} = 0, \quad k = 1, \dots, K-1, \quad (8)$$

where $Q_{K-1} = (\mathbf{q}_1, \dots, \mathbf{q}_{K-1})$ and $\mathbf{q}_k = (q_{k1}, \dots, q_{kn})^\top$. Then, the k -Means objective function can be written as

$$L_{km} = \sum_{i=1}^n \|\mathbf{x}_i\|^2 - \frac{1}{n} \mathbf{e}^\top X^\top X \mathbf{e} - \text{Tr}(Q_{K-1}^\top X^\top X Q_{K-1}), \quad (9)$$

where $X = (\mathbf{x}_1, \dots, \mathbf{x}_n)$, and $\mathbf{e}$ is the n -dimensional vector whose elements are all 1.

Here, we can see that the k -Means problem is free from the influences of parallel transformation and that the original data $\mathbf{x}_i$ can be replaced with the centered data $\mathbf{y}_i$ without loss of generality. Then, the objective function can be written as:

$$L_{km} = \sum_{i=1}^n \|\mathbf{y}_i\|^2 - \text{Tr}(Q_{K-1}^\top Y^\top Y Q_{K-1}), \quad (10)$$

where $Y = (\mathbf{y}_1, \dots, \mathbf{y}_n)$ and $Y\mathbf{e} = \mathbf{0}$.

The optimal solutions for the transformed discrete membership indicator vectors Q_{K-1} are derived by maximizing $\text{Tr}(Q_{K-1}^\top Y^\top Y Q_{K-1})$, and continuous (relaxed) solutions are the eigenvectors corresponding to the $K-1$ largest eigenvalues of $Y^\top Y$. Therefore, the optimal Q_{K-1} can be identified with the principal component scores in PCA, and a continuous solution for k -Means clustering is derived in a PCA-guided manner.

If we want to know the cluster structure from the (continuous) solution for Q_{K-1} , we need to compute the optimal transformation T , although this is not easy in many case. In [2], cluster connectivity analysis is performed by calculating the following connectivity matrix C :

$$C = \frac{1}{n} \mathbf{e} \mathbf{e}^\top + \sum_{k=1}^{K-1} \mathbf{q}_k \mathbf{q}_k^\top, \quad (11)$$

and a probability for the connectivity between samples i and j is given as:

$$p_{ij} = \frac{c_{ij}}{\sqrt{c_{ii}c_{jj}}}. \quad (12)$$

Cluster structure is shown in the diagonal block structure using the spectral ordering [3].

III. PRINCIPAL COMPONENTS ESTIMATION FROM INCOMPLETE DATA AND ITS APPLICATION TO k -MEANS CLUSTER INDICATOR ESTIMATION

A. PCA-like Method for Incomplete Data

Assume that a data set is represented by the $m \times n$ data matrix $X = (\mathbf{x}_1, \dots, \mathbf{x}_n)$ and the goal is to extract a p ($p < m$)-dimensional principal component score for each sample. The PCA process is achieved by estimating the lower rank approximation of X as follows:

$$X \approx AF + \mathbf{v}\mathbf{e}^\top, \quad (13)$$

where A is the $m \times p$ principal component matrix $A = (\mathbf{a}_1, \dots, \mathbf{a}_p)$, and F is the $p \times n$ principal component score matrix $F = (\mathbf{f}_1, \dots, \mathbf{f}_n)$. $\mathbf{v}$ is the centering vector.

In order to estimate principal components from incomplete data, Shibayama [10] considered minimization of the following least square criterion:

$$\begin{aligned} L &= \text{Tr}(Z \odot (X - AF + \mathbf{v}\mathbf{e}^\top))^\top (Z \odot (X - AF + \mathbf{v}\mathbf{e}^\top)) \\ &= \sum_{i=1}^n \sum_{j=1}^m z_{ij} \left(x_{ij} - \sum_{k=1}^p f_{ik} a_{kj} - v_j \right)^2, \end{aligned} \quad (14)$$

where $Z = \{z_{ij}\}$ is given by Eq.(3) and $\odot$ denotes the Hadamard product.

This model is based on a concept similar to PDS in k -Means clustering, namely that the partial errors calculated only by the observed elements are minimized.

If the data matrix X does not include any missing elements, the PCA model is equivalent to the standard formulation in which the centering vector $\mathbf{v}$ is given by the mean vector and the principal component vectors $\mathbf{a}_k$, $k = 1, \dots, p$ are derived as the eigenvectors corresponding to the principal eigenvalues of the scatter matrix. However, if the data matrix includes missing elements, the PCA process cannot be reduced to an eigen problem.

In the Shibayama's PCA-like method [10], principal components are estimated on the basis of an iterative optimization scheme. Under the constraints $F\mathbf{e} = \mathbf{0}$ and $FF^\top = I$, the updating rule for each step is derived as follows:

$$\tilde{\mathbf{a}}_j = (F\tilde{Z}_jF^\top)^{-1}F\tilde{Z}_j(\tilde{\mathbf{x}}_j - v_j\mathbf{e}), \quad (15)$$

$$v_j = (\mathbf{e}^\top \tilde{Z}_j \mathbf{e})^{-1} \mathbf{e}^\top \tilde{Z}_j (\tilde{\mathbf{x}}_j - F^\top \tilde{\mathbf{a}}_j), \quad (16)$$

$$\mathbf{f}_i = (A^\top Z_i A)^{-1} A^\top Z_i (\mathbf{x}_i - \mathbf{v}), \quad (17)$$

where $A = (\tilde{\mathbf{a}}_1, \dots, \tilde{\mathbf{a}}_m)^\top$, $X = (\tilde{\mathbf{x}}_1, \dots, \tilde{\mathbf{x}}_m)^\top$ and $Z_i = \text{diag}(z_{i1}, \dots, z_{im})$, $\tilde{Z}_j = \text{diag}(z_{1j}, \dots, z_{nj})$.

Beside random initialization, initial values for A , F and $\mathbf{v}$ can be estimated by WDS in which the standard PCA model is applied after removing all samples including missing elements. The iterative procedure of updating A , F , and $\mathbf{v}$ is continued until convergence while normalizing F in each iteration.

B. k -Means Cluster Indicator for Incomplete Data

Although the iterative optimization process of the PCA-like method with bad initialization may become trapped at local minima, we can expect that a plausible estimator of principal component scores can be obtained using the multi-starting strategy even when data matrices include missing elements. Therefore, the derived principal component scores F are promising candidates for the relaxed cluster indicator in the PCA-guided k -Means procedure.

The procedure for deriving the relaxed cluster indicator matrix Q_K in the K cluster case can be summarized as follows:

Step.1 Set the dimension of principal components as $p = K - 1$ and initialize A , F , and $\mathbf{v}$.

Step.2 Repeat the PCA-like process until convergence.

Step.3 Substitute F into Q_{K-1} .

Step.4 Set $\mathbf{q}_K$ as $\mathbf{q}_K = (\sqrt{1/n}, \dots, \sqrt{1/n})^\top$.

Once the relaxed cluster indicator Q_K is estimated, cluster structures can be captured by a technique such as cluster connectivity analysis.

IV. NUMERICAL EXPERIMENTS

In this section, several experimental results are shown in order to demonstrate the characteristic features of the proposed approach.

A. Iris Data Set

The first experiment was performed with the Iris data set that is available from the UCI Machine Learning Repository [14]. In this experiment, 150 instances ($n = 150$) with four-dimensional observation ($m = 4$) drawn from three classes were partitioned into three clusters without using the class label. The data set does not include any missing elements. It is well known that the k -Means model can capture the three dense masses of the three labeled classes in many trials using the multi-starting strategy, although whereas one class is well separated from others, the remaining two classes somewhat overlap.

In this experiment, the frequencies of good results in the proposed approach using the multi-starting strategy are compared with that in the k -Means using PDS. Here, a "good" result does not necessarily mean that all instances are correctly classified into the three classes without using class labels, but rather it means that the three cluster centers were located near the center of each class.

Table I compares the clustering performances of the proposed approach and the k -Means procedure using PDS with incomplete data. Various incomplete data sets were generated by withholding one observation (attribute) from randomly selected instances. The table shows the average frequencies of good results in 100 different data sets with 100 trials

TABLE I
AVERAGE FREQUENCIES (%) OF “GOOD” RESULTS IN 10000 TRIALS WITH INCOMPLETE DATA SET (IRIS DATA SET)

# missing elements	PCA-guided	<i>k</i> -Means (PDS)
0	100.00	79.01
50	100.00	85.01
60	95.84	82.46
70	92.23	79.04
80	90.37	78.58
90	84.19	75.49
100	82.46	73.20
110	82.34	69.57
120	79.34	67.47
130	74.84	65.15
140	71.91	63.42
150	70.14	60.21

with random initialization. The result of the no missing value situation for the PCA-guided approach is the same as that given by the conventional PCA-guided *k*-Means procedure and is free from random initialization based on the deterministic procedure. Note that the *k*-Means procedure may become trapped at a local minimum even when the data set includes no missing values.

The table demonstrates that the proposed PCA-guided approach is still robust to initialization when the number of missing values is not too large, e.g., 50 missing elements, although the PCA process is based on an iterative optimization scheme. As the number of missing elements increase, the frequency of good results decrease. However, the proposed PCA-guided approach is always more robust than the *k*-Means procedure using PDS. This result implies that the PCA-guided approach is more reliable than the conventional *k*-Means procedure even when data sets includes missing elements.

B. Document Clustering

Next, the characteristic feature of the proposed approach is demonstrated in a document clustering task using the Japanese novel *Kokoro* by Soseki Natsume, which can be downloaded from Aozora Bunko (<http://www.aozora.gr.jp>). The novel is composed of three chapters (“Sensei and I”, “My Parents and I”, “Sensei and His Testament”) and the chapters include 36, 18, and 56 sections, respectively. In this experiment, the sections are regarded as individual text documents (number of samples is $n = 110$) and the documents are partitioned into three clusters ($K = 3$) without using the chapter information. This data set forms roughly three clusters in the 83-dimensional data space [16], as shown in Fig. 1-(a). Thus, the goal of the experiment is to reconstruct the intrinsic chapter information from a word analysis. The text documents were preprocessed using a morphological analysis system software “Chasen” (<http://chasen.naist.jp/hiki/ChaSen/>), which segments Japanese text strings into morphemes and tags these morphemes with their parts of speech and pronunciations. The 83 most frequently used substantives and verbs (they were used more than 50 times in the novel) were used as attributes to be analyzed with their tf-idf weights [15].

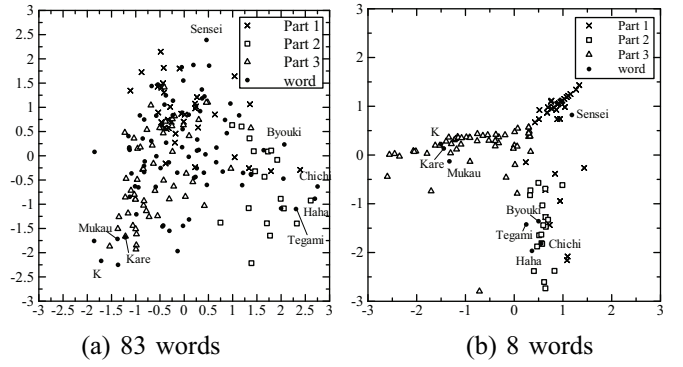


Fig. 1. 2-D biplots of Japanese novel *Kokoro* [16]: (a) 83 words, (b) 8 words

TABLE II
AVERAGE FREQUENCIES (%) OF “GOOD” RESULTS USING AN INCOMPLETE DATA SET (JAPANESE NOVEL *Kokoro* WITH 83 WORDS (ATTRIBUTES))

# missing elements	PCA-guided	<i>k</i> -Means (PDS)
0	100.00	100.00
10	100.00	88.12
20	99.30	76.24
30	99.40	72.28
40	98.73	59.45
50	98.68	48.56
60	98.69	43.61
70	98.50	38.64
80	95.51	26.81
90	96.20	18.87
100	94.91	7.98
110	93.39	1.10

A comparison of the frequencies of good results is shown in Table II. Various incomplete data sets were generated in a similar manner as in the previous experiment. The proposed PCA-guided approach is much more robust than the *k*-Means procedure using PDS when data sets include many incomplete instances. This is because very ambiguous cluster boundaries easily trap the *k*-Means procedure using PDS at local minima whereas the PCA-guided approach is seldom influenced by a slight loss of information (in only 1 of 83 attributes).

Second, the same experiment was repeated, but using fewer words (attributes). In [16], it was shown that the three-cluster structure is clarified by using eight significant words. Therefore, only eight words (eight attributes) were used for this experiment. Table III shows a comparison of the results. The performance of the proposed PCA-guided approach is inferior to that shown in Table II because information loss became relatively large (in 1 of 83 attributes). On the other hand, the performance of the *k*-Means procedure using PDS is almost equivalent to that shown in Table II, although in this case as well, it is inferior to the PCA-guided approach.

These experimental results imply that the PCA-guided procedure for *k*-Means is also useful for incomplete data sets.

TABLE III
AVERAGE FREQUENCIES (%) OF "GOOD" RESULTS USING AN INCOMPLETE
DATA SET (JAPANESE NOVEL *Kokoro* WITH 8 WORDS (ATTRIBUTES))

# missing elements	PCA-guided	k -Means (PDS)
0	100.00	100.00
10	98.62	88.12
20	96.81	76.25
30	93.32	72.29
40	91.86	59.47
50	90.05	48.57
60	88.92	43.61
70	77.59	38.61
80	67.05	26.80
90	65.84	18.93
100	54.35	8.00
110	54.64	1.07

V. CONCLUSION

In this paper, a PCA-guided procedure for k -Means clustering of incomplete data sets is proposed. Although the PCA-like method is not a deterministic procedure but rather is based on an iterative optimization scheme, the PCA-guided approach is still more robust to initialization problems than the k -Means procedure using PDS.

Comparison with other techniques for handling missing values, such as OCS, remains for future work. Other potential future work involves component-wise robust PCA [17], [18], in which component-wise noise is ignored by means of an M-estimation scheme [19].

REFERENCES

- [1] J. B. MacQueen, "Some methods of classification and analysis of multivariate observations," *Proc. of 5th Berkeley Symposium on Math. Stat. and Prob.*, pp. 281-297, 1967.
- [2] C. Ding and X. He, " K -means clustering via principal component analysis," *Proc. of Int'l Conf. Machine Learning (ICML 2004)*, pp. 225-232, 2004.
- [3] C. Ding and X. He, "Linearized cluster assignment via spectral ordering," *Proc. of Int'l Conf. Machine Learning (ICML 2004)*, pp. 233-240, 2004.
- [4] K. Honda, A. Notsu and H. Ichihashi, "Fuzzy PCA-guided robust k -means clustering," *IEEE Transactions on Fuzzy Systems*, vol. 18, no. 1, pp. 67-79, 2010.
- [5] R. N. Davé, "Characterization and detection of noise in clustering," *Pattern Recognition Letters*, vol. 12, no. 11, pp. 657-664, 1991.
- [6] R. J. Hathaway, and J. C. Bezdek, "Fuzzy c -means clustering of incomplete data," *IEEE Trans. Syst., Man, Cybern.*, vol.B31, pp. 735-744, 2001.
- [7] J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*, Plenum Press, 1981.
- [8] A. Ruhe, "Numerical computation of principal components when several observations are missing," Tech. Rep. UMINF-48-74, Dept. Information Processing, Umea Univ., 1974.
- [9] T. Wiberg, "Computation of principal components when data are missing," in *Proc. 2nd Symposium on Computational Statistics*, pp. 229-236, 1976.
- [10] T. Shibayama, "A PCA-like method for multivariate data with missing values," *Japanese Journal of Educational Psychology*, vol.40, pp. 257-265, 1992. (in Japanese)
- [11] H. Shum, K. Ikeuchi, and R. Reddy, "Principal component analysis with missing data and its application to polyhedral object modeling," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.17, pp. 854-867, 1995.
- [12] K. Honda and H. Ichihashi, "Linear fuzzy clustering techniques with missing values and their application to local principal component analysis," *IEEE Trans. Fuzzy Systems*, vol. 12, no. 2, pp. 183-193, 2004.
- [13] J. C. Bezdek, C. Coray, R. Gunderson and J. Watson, "Detection and characterization of cluster substructure 2. Fuzzy c -varieties and convex combinations thereof," *SIAM J. Appl. Math.*, vol.40, pp. 358-372, 1981.
- [14] A. Asuncion and D.J. Newman, UCI Machine Learning Repository [<http://www.ics.uci.edu/mllearn/MLRepository.html>], Irvine, CA: University of California, School of Information and Computer Science, (2007).
- [15] G. Salton and C. Buckley: Term-weighting approaches in automatic text retrieval, *Information Processing and Management*, vol.24, issue 5, pp.513-523, 1988.
- [16] K. Honda, H. Wada, H. Ichihashi and A. Notsu: An intuitive text mining method based on variable selection in principal component analysis, *Proc. of 8th International Symposium on Advanced Intelligent Systems*, pp.334-337, 2007.
- [17] F. de la Torre and M. J. Black: A framework for robust subspace learning; *International Journal of Computer Vision*, vol. 54, pp. 117-142 (2003)
- [18] K. Honda and H. Ichihashi, "Component-wise robust linear fuzzy clustering for collaborative filtering," *International Journal of Approximate Reasoning*, vol. 37, no. 2, pp. 127-144, 2004.
- [19] P. W. Holland and R. E. Welsch, "Robust regression using iteratively reweighted least-squares," *Communications in Statistics*, vol. A6, no. 9, pp. 813-827 (1977)