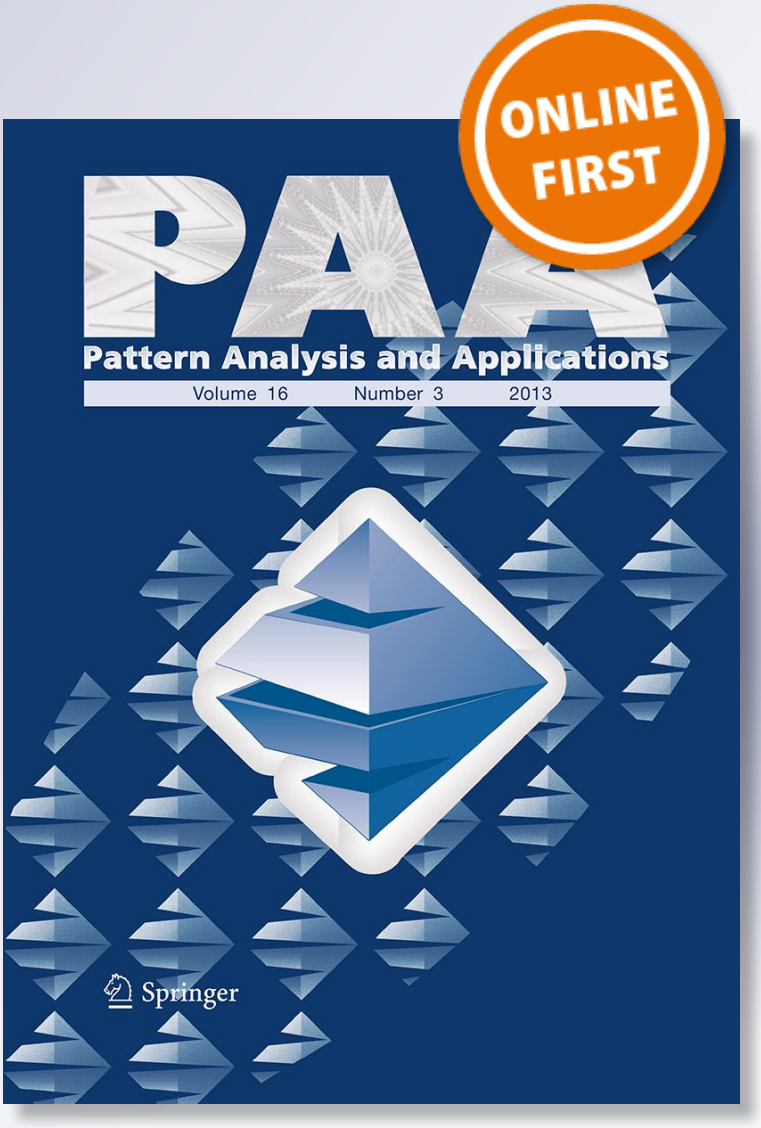
Entropy-based multi-view matrix completion for clustering with side information

Changming Zhu & Duoqian Miao

Pattern Analysis and Applications

ISSN 1433-7541

Pattern Anal Applic
DOI 10.1007/s10044-019-00797-0



Springer

Pattern Analysis and Applications
https://doi.org/10.1007/s10044-019-00797-0

THEORETICAL ADVANCES

Entropy-based multi-view matrix completion for clustering with side information

Changming Zhu[1,2] · **Duoqian Miao**[1]

Received: 11 August 2017 / Accepted: 12 February 2019

Abstract
Multi-view clustering aims to group multi-view samples into different clusters based on the similarity. Since side information can describe the relation between samples, for example, must-links and cannot-links, thus multi-view clustering with the consideration about side information along with samples can get more feasible clustering results. As a recent developed multi-view clustering approach, multi-view matrix completion (MVMC) constructs similarity matrix for each view and casts clustering into a matrix completion problem. Different from traditional multi-view clustering approaches, MVMC enforces the consistency of clustering results on different views as constraints for alternative optimization and the global optimal solution can be obtained. Although related experiments show that MVMC exhibits impressive performance, it still neglects the possibility of a sample belonging to a cluster. In this paper, we consider the possibility on the base of entropy and develop an entropy-based multi-view matrix completion for clustering with side information (EMVMC). Experiments on multi-view datasets Course, Citeseer, Cora, WebKB, NewsGroup, and Reuters validate the effectiveness of EMVMC.

Keywords Multi-view clustering · Fuzzy membership · Matrix completion

1 Introduction

1.1 Background

Multi-view datasets consist of samples with multiple views, and each view corresponds to a feature group. Suppose, we collect multiple videos from YouTube and each video appears in multiple varied forms, e.g., visual, audio, and text. Then, each form also has multiple features. For example, text possesses color, size, shape, and some other features. We regard these forms as views, and then a multi-view video dataset is composed of these collected videos.

In order to process multi-view datasets, some multi-view learning machines or multi-view approaches [1–8] are developed and they are applied to different applications including multi-view classification, multi-view representation, multi-view clustering [9–15], etc. Among these applications, multi-view clustering has been paid more attention by recent researchers. Multi-view clustering aims to group multi-view samples into different clusters based on the similarity, and it has plenty of real applications, such as data summarization [16], text mining [17, 18], bioinformatics [19]. According to these applications, some multi-view clustering approaches are developed including the following four cases. First, Rabbouch et.al developed a clustering approach to count and recognize the vehicles [16]. Second, Azadani et. al. [17] proposed a well-established data mining technique called frequent itemset mining to employ a minimum spanning tree based clustering algorithm to discover various subthemes of the document and then the most informative sentences in a document can be selected. Third, Zheng et. al. [18] developed a corpus-based enrichment approach for short text clustering. Fourth, Dougherty et. al. [19] designed a clustering approach to divide the gene-expression microarrays. Related experiments have validated the effectiveness of these developed clustering approaches in corresponding fields.

✉ Duoqian Miao
dqmiao@tongji.edu.cn

Changming Zhu
cmzhu@shmtu.edu.cn

[1] Department of Computer Science and Technology, Tongji University, Shanghai 201804, People's Republic of China

[2] College of Information Engineering, Shanghai Maritime University, Shanghai 201306, People's Republic of China

Published online: 19 February 2019

1.2 Problem and recent solutions

But it is found that these approaches exist the following problems. First, in order to process multi-view datasets, some clustering approaches concatenate all features from multiple views into a single one. Although this operation brings a convenience in mathematic, it has an vital drawback, namely the dimension of concatenation feature matrices is usually high which may trigger the curse of dimensionality and result in a high computational cost. Second, the approach of concatenation treats different views equally which is not appropriate since the difference between views is ignored. Third, these approaches ignore the side information of multi-view samples. Indeed, side information can describe the relation between samples, for example, must-links and cannot-links and this kind of information can guide the multi-view clustering from a new point.

In order to process these problems, researchers have developed many solutions including multi-view clustering via robust nonnegative matrix factorization (MVCRNMF) [20] and multi-view matrix completion (MVMC) approach [21]. MVCRNMF is mainly used for community detection and its key idea is to build a multi-view robust nonnegative matrix factorization (NMF) model with the co-regularized constraint on community indicator matrices of link view and content view. This can make link and content information complement each other during the factorization process of NMF.

Different from MVCRNMF, MVMC has more widely application fields and it can solve above problems. As [21] said, side information describes the relation between samples and there are two kinds of side information, label-level and sample-level. Side information with label-level is hard to be used in real-world applications [21] while the one with sample-level is convenient to be collected since this kind of information can be reflected through pairwise constraints. In terms of pairwise constraints, they are always consisted of two parts, must-link (M) and cannot-link (C). A must-link (cannot-link) represents that the pair of samples should (not) be assigned into the same cluster. Due to it is convenient to gather pairwise constraints, thus, some scholars pay attention to clustering with sample-level side information and develop some multi-view clustering approaches. Among these approaches, MVMC is a new developed (which is proposed in 2017) and effective one for clustering with side information. Related experiments have shown that MVMC constructs a pairwise similarity matrix S_v for the vth view independently and casts clustering task into a matrix completion problem based on given pairwise constraints and feature information from multiple views. Then, the final pairwise similarity matrix S is learned by controlling S and S_v in different views to approach each other. The global optimal solution is obtained by projective alternative optimization since the objective function is jointly convex.

1.3 Proposal

Although related experiments [21] have shown that MVMC can efficiently utilize side information and outperform some previous multi-view clustering approaches, it has a key defect. In MVMC, one adopts Eq. (1) to construct a pairwise similarity matrix S_v in vth view where R is the number of clusters and $\mathbf{u}_r^v \in \{0,1\}^n$ is the membership vector of the rth cluster in the vth view, where $u_{i,j}^v = 1$ if the jth sample x_j is assigned to the rth cluster and zero, otherwise. Here, n is the total number of samples. According to S_v, it is found the belonging of sample x_j has only two cases, belongs or not belongs. This indicates that the relationship between cluster and sample is clear. While as we know, in real-world applications, each sample belongs to a cluster with a possibility. For example, a sample belongs to cluster A with a 30% possibility while belongs to cluster B with a 70% possibility.

$$S_v = \sum_{r=1}^{R} \mathbf{u}_r^v (\mathbf{u}_r^v)^{\mathrm{T}} \tag{1}$$

During the procedure of MVMC, the possibility is not taken into consideration, thus we will overcome it. Due to entropy is used to evaluate the class certainty of each sample and fuzzy membership between samples [22], thus we combine entropy with MVMC and propose an entropy-based multi-view matrix completion for clustering with side information (EMVMC).

1.4 Motivation, novelty, contribution

According to what we have said before, the motivation of EMVMC is introducing entropy into the MVMC so as to consider the possibility of a sample belonging to a cluster.

Since MVMC is developed in 2017 and to the best of our knowledge, there is no recent multi-view clustering approach is developed to group multi-view samples into different clusters based on the class certainty of each sample, thus our developed EMVMC has a novelty.

What's more, since compared with MVMC, our EMVMC considers the possibility of a sample belonging to a cluster so as to boost the performance of a multi-view clustering with side information, thus contributions of EMVMC are (1) developing a more feasible multi-view clustering approach for real-world applications; (2) reflecting the degree of membership of a sample to a cluster.

1.5 Framework

The rest of this paper is organized as below. Review about MVMC can be found in Sect. 2. Description of EMVMC is given in Sect. 3. Experiments are given in Sect. 4. The conclusions are given in Sect. 5.

2 Review about MVMC

MVMC includes two main parts, one is clustering with side information and the other is matrix completion (MC) [21]. Thus here, we first to review these two main parts, and then review the framework of MVMC.

2.1 Clustering with side information

Clustering with side information has been widely used and developed well in recent years. We know that side information describes relation between samples and the relation can be reflected by label-level and sample-level. In terms of sample-level, must-link and cannot-link are two ways to collect pairwise constraints which reflect side information. There are plenty of algorithms about clustering with side information are proposed based on distance metric learning. For example, the information theoretic metric learning algorithm (ITML) proposed in [23] learns a metric matrix with side information based on information theory. Matrix completion based constraint clustering (MCCC) proposed in [24] converts clustering to a matrix completion problem. Internet traffic clustering with side information (ITCSI) proposed in [25] adopts a constrained expectation maximization (EM) algorithm for clustering.

2.2 Matrix completion

Matrix completion problem was original proposed by [26] for collaborative filtering in 1992. As [26] said, suppose that there is a low-rank matrix should be recovered and MC will find a matrix X that minimizes the difference with the given observation. However, it is still challenging because rank minimization problem is NP-hard and this challenging confuses the scholars almost 20 years until 2012. In 2012, [27] found that minimizing $rank(X)$ can be realized by the minimization of $||X||_\star$ which is the nuclear norm of X under broad conditions. Moreover, [28] proposed an approach to speed up the process of MC by utilizing side information. After that, MC problem gets a fast development and has been applied into clustering and some approaches including graph-based clustering proposed by [29], crowd-sourced clustering proposed by [30], side information-related clustering approach proposed by [24], subspace-learning-based clustering proposed by [31] have been developed.

2.3 Framework of MVMC

Suppose there is a dataset D with n samples and V views (see Eq. 2) and features in vth view is denoted as X_v (see Eq. 3) where $x_i^v \in \mathbb{R}^{1\times d_v}$ is the feature of x_i in the vth view, and d_v is dimension of the vth view. Then, D can be rewritten with Eq. (4).

$$D = \{x_1, x_2, \ldots, x_{n-1}, x_n\} \tag{2}$$

$$X_v = (x_1^v; x_2^v; \ldots; x_n^v) \in \mathbb{R}^{n\times d_v} \tag{3}$$

$$D = \{X_1, X_2, \ldots, X_V\} \in \mathbb{R}^{n\times \sum_{v=1}^{V} d_v} \tag{4}$$

After that, let M (C) denote the set of must-link (cannot-link) constraints, $(i,j) \in M((i,j) \in C)$ implies x_i and x_j should (not) be assigned into the same cluster. $\Omega = M \bigcup C$ is used to represent all pairwise constraints. Then, the framework of MVMC is given below.

According to Sect. 1.3, in the rth cluster in the vth view, $\mathbf{u}_r^v \in \{0,1\}^n$ is the membership vector where $u_{r,j}^v = 1$ if the jth sample x_j is assigned to the rth cluster and zero, otherwise. Then, in the vth view, the pairwise similarity matrix S_v is given by Eq. (1) and S_v is a $n \times n$ matrix. Each element of S_v is 0 or 1. If the ith row and jth column element of S_v, i.e., $[S_v]_{i,j}$ is 1(0), then x_i and x_j are (not) assigned to the same cluster in vth view.

Once we get $S_1, S_2, \ldots, S_V$, we can adopt Eq. (5) to get a finally similarity matrix S.

$$S = \frac{1}{V}\sum_{v=1}^{V} S_v \tag{5}$$

What's more, according to [21] said, S_v can be expanded by Eqs. (6) and (7) where $M_v \in \mathbb{R}^{k\times k}$ is a positive semidefinite matrix and $\mathbf{z}_j^v$s ($j = 1, \ldots, k$) are the first k left singular vector of features in X_v. Although k is able to vary over different views, this does not make difference to the essence of the problem.

$$S_v = Z_v M_v Z_v^{\mathrm{T}} \tag{6}$$

$$Z_v = [\mathbf{z}_1^v, \mathbf{z}_2^v, \ldots, \mathbf{z}_k^v] \tag{7}$$

Moreover, due to M_v is constrained as a positive semidefinite matrix, thus we have the expression given in Eq. (8) where σ_i and eig_i are the ith singular value and eigenvalue of M, respectively. Then, since Z_v is the orthogonal matrix, so we have the expression given in Eq. (9).

$$||M_v||_\star = \sum_{i=1}^{k} |\sigma_i| = \sum_{i=1}^{k} |\mathrm{eig}_i| = tr(M_v) \tag{8}$$

$$||S_v||_\star = \left|\left|Z_v M_v Z_v^{\mathrm{T}}\right|\right|_\star = ||M_v||_\star \tag{9}$$

Thus according to the above information, the optimization problem of MVMC can be given with Eqs. (10), (11), and (12). Then, Eqs. (13) and (14) show the constraints of this optimization problem. Among these equations, S_{ob} is the partial observations matrix, $R_{\Omega}(\cdot) : \mathbb{R}^{n\times n} \mapsto \mathbb{R}^{n\times n}$ is a linear operator which preserves the entry of S in Ω and 0 outside, $C_1 > 0$ and $C_2 > 0$ are two regularization parameters, $||\cdot||_{\text{F}}$ is Frobenius norm.

$$\min_{S,\{M_v\}_{v=1}^m} \sum_{v=1}^{V} (tr(M_v) + C_1 A + C_2 B) \tag{10}$$

$$A = \left|\left|R_{\Omega}(Z_v M_v Z_v^{\text{T}} - S_{\text{ob}})\right|\right|_{\text{F}}^2 \tag{11}$$

$$B = \left|\left|(Z_v M_v Z_v^{\text{T}} - S)\right|\right|_{\text{F}}^2 \tag{12}$$

$$0 \le S_{i,j} \le 1, \forall i,j \in \{1,2,\ldots,n\} \tag{13}$$

$$M_v \in S_+^k, v = 1,2,\ldots,m \tag{14}$$

In order to solve this problem, MVMC first to initial $\{M_v\}_{v=1}^V$ and S by S_{ob}, then it repeats the step (a), minimizing $\{M_v\}_{v=1}^V$ over S and step (b), minimizing S over $\{M_v\}_{v=1}^V$, until convergence. Details can be referred to [21]. After optimization of the problem, we can get the pairwise similarity matrix S and the clustering results.

3 The framework of entropy-based multi-view matrix completion for clustering with side information (EMVMC)

3.1 Entropy-based fuzzy membership of samples

Suppose there are n multi-view samples with two clusters, l samples belong to cluster + 1 (positive cluster) and other $n - l$ samples belong to cluster − 1 (negative cluster). Then, we assume the fuzzy membership of a sample x_i from positive cluster is m_i^+ while the one of a sample x_j from negative cluster is m_j^-. As [22] said, fuzzy membership can be measured by entropy, i.e., a higher cluster certainty denotes a lower entropy and a lower entropy means a higher fuzzy membership. Thus, we adopt entropy to decide the fuzzy membership of a sample.

First, for a sample x_i, we select its k nearest neighbors $\{x_{i1}, x_{i2}, \ldots, x_{ik}\}$. Second, among these k nearest neighbors, we count the number of both positive samples and negative samples and denote the numbers as num_{+i} and num_{-i}, respectively. Third, we suppose that x_i belongs to positive cluster with a p_{+i} possibility while belongs to negative cluster with a p_{-i} possibility and calculate p_{+i} and p_{-i} by Eqs. (15) and (16). Fourth, the entropy of x_i is defined by Eq. (17) where ln represents the natural logarithm operator. Fifth, we design a function $f(x_i)$ whose expression is given in Eq. (18). Sixth, if x_i is a positive sample, its fuzzy membership is m_i^+ (see Eq. 19), otherwise, if x_i is a negative sample, its fuzzy membership is m_i^- (see Eq. 20) where r is the minority-to-majority cluster ratio.

$$p_{+i} = \frac{\text{num}_{+i}}{k} \tag{15}$$

$$p_{-i} = \frac{\text{num}_{-i}}{k} \tag{16}$$

$$H_i = -p_{+i}\ln(p_{+i}) - p_{-i}\ln(p_{-i}) \tag{17}$$

$$f(x_i) = 1 - H_i \tag{18}$$

$$m_i^+ = f(x_i) \tag{19}$$

$$m_i^- = f(x_i) \times r \tag{20}$$

3.2 The solution of EMVMC

Once we get the fuzzy membership m_i of each sample x_i, we combine them with MVMC and get the solution of EMVMC. Indeed, the model of EMVMC is very similar with the one of MVMC and the only difference is that EMVMC introduces entropy-based fuzzy membership of samples. Although it looks that this introduction is not worth mentioning, the later experiments will validate that the introduction brings an impressive performance.

Indeed, the introduction is reflected by $\mathbf{u}_r^v$. In EMVMC, we let $\mathbf{u}_r^v \in \{u_{r,j}^v\}^n$ be the membership vector of the rth cluster in the vth view and different from the one in MVMC, that in EMVMC, $u_{r,j}^v \in [0,1]$, i.e., jth sample belongs to rth cluster with a $u_{r,j}^v$ possibility. Then, we still get S_v and S. Indeed, in S_v, when $[S_v]_{i,j}$ is larger, the x_i and x_j have a larger probability to be assigned to a same cluster. After that, we optimize Eqs. (10), (11), and (12) with the constraints given by Eqs. (13) and (14) to get the clustering results.

4 Experiments

Our experiments include four parts. First part is experimental setting, second one is clustering performance comparison, third one is the relation between the number of nearest neighbors and clustering performance, and the fourth one is significance analysis.

4.1 Experimental setting

4.1.1 Datasets

Since EMVMC is a multi-view clustering approach, so we conduct the experiments on multi-view datasets Course, Citeseer, Cora, WebKB, NewsGroup, and Reuters. These datasets are also used in rank consistency based multi-view learning (RANC) [32] and the configurations and details can be referred to [32]. Here, we summary their information in Table 1 and describe them in simple.

Course consists of course web pages and non-course ones and each page has two views [5].

Both Citeseer and Cora consist of samples with two views, i.e., content and cites [33].

WebKB consists of web pages collected from four universities: Cornell, Texas, Wisconsin, and Washington which have five categories, i.e., student, project, course, stuff, and faculty. Each web page is described with two views: content and citation and we treat WebKB in four separate subdatasets grouped by universities [32].

NewsGroup consists of samples from six groups, i.e., M2, M5, M10, NG1, NG2, NG3 and we treat NewsGroup in six separate subdatasets corresponding to each group. For each sub-one, there are three views, partitioning around methods, supervised mutual information, and unsupervised mutual information [34].

Reuters consists of machine translated documents which are written in five different languages, i.e., English, French, German, Italian, and Spanish. Each language is treated as a view and each document can be translated from one language to another language. Moreover, documents are also categorized into six different topics (classes), i.e., C15, CCAT, E21, ECAT, GCAT, and M11 [35, 36].

Table 1 Brief dataset description

Dataset	R	n	V	$d_v\ (v = 1, 2, \ldots, V)$
Course	2	1051	2	66,5
Citeseer	6	3264	2	3703, 3264
Cora	7	2708	2	1433, 2708
Cornell	5	195	2	1703, 195
Texas	5	185	2	1703, 185
Washington	5	217	2	1703, 217
Wisconsin	5	262	2	1703, 262
News-M2	2	1200	3	2000, 2000, 2000
News-M5	5	500	3	2000, 2000, 2000
News-M10	10	500	3	2000, 2000, 2000
News-NG1	2	500	3	2000, 2000, 2000
News-NG2	5	400	3	2000, 2000, 2000
News-NG3	8	1000	3	2000, 2000, 2000
Reuters	6	1600	5	2000, 2000, 2000, 2000, 2000

R is the number of clusters, n is the number of samples, V is the number of views, and d_v is the dimension of each view

4.1.2 Compared approach and parameter setting

Since [21] has validated that MVMC outperforms some classical multi-view clustering algorithms including co-regularized spectral clustering (Co-Reg) [37], multi-view kernel k-means algorithm (MKKM) [38], robust multi-view spectral clustering based on Markov chain (RMSC) [39], ITML [23], MCCC [24], etc., so we only adopt MVMC as the compared approach.

Moreover, since MVMC is the basic model of EMVMC and compared with MVMC, EMVMC only has one more particular parameter, i.e., the number of nearest neighbors, k. So the setting of most parameters can refer to [21] and k is selected from the set $\{1, 2, \ldots, 20\}$.

4.1.3 How to get and measure clustering performance

Here, we take a dataset for instance and describe that how to get and measure the clustering performance.

First, for a multi-view dataset with C classes and n samples, we use a multi-view clustering approach to group the multi-view samples into R clusters. Sometimes samples of one class are fallen into multiple clusters and a cluster maybe includes samples from multiple classes. Here, n_r^c denotes the number of samples in the rth cluster from cth class, n_r denotes the total number of samples in the rth cluster, n^c represents the number of samples in cth class.

Second, in terms of this dataset, we select one class as the positive class and other classes form the negative one. Then, we can get a binary class subdataset. After the repetition of this operation for C times, we can form C different binary class subdatasets.

Third, for each binary class subdataset, we define the label of each cluster. Simply speaking, in each cluster, if the number of positive samples is larger than the number of negative samples, we regard this cluster is positive. Otherwise, this cluster is negative. Then, observed class labels of samples in a positive (negative) cluster are positive (negative) no matter their true class labels are positive or negative.

Fourth, for each binary class subdataset, we count the numbers of true positive (TP), false positive (FP), false negative (FN), and true negative (TN), respectively [40]. TP means the samples whose true and observed labels are both positive, FP indicates the samples whose true labels are negative while the observed labels are positive, FN denotes the samples whose true labels are positive while the observed labels are negative, TN represents the samples whose true and observed labels are both negative.

Fifth, we repeat the third step and fourth step for C times and then we can get C TPs, C FPs, C FNs, and C TNs. For

convenience, we sum C TPs and still use TP to denote the sum. For others, we carry out the same operations.

Sixth, in order to measure the clustering performances, we adopt the following widely used criteria: accuracy, true positive rate (acc^+), true negative rate (acc^-), positive predictive value (PPV), F-Measure, G-Mean, normalized mutual information (NMI), adjusted rand index (Adj-RI), and average entropy (AE). The computation expressions of accuracy, acc^+, acc^-, PPV, F-Measure, and G-Mean are given in Eqs. (21)–(26), and we can see the definitions of these equations from [22]. The computation expression of NMI is given in Eq. (27) where $H(\pi)$ represents the entropy of the clusters (see Eq. 28) and $H(\zeta)$ represents the entropy of the classes (see Eq. 29). We can refer to [41] to see the definition of NMI. For Adj-RI, its computation expression can be found in Eq. (30) and [42] where RI is the rand index (see Eq. 31) and $E(\star)$ represents the expectation of $\star$. In Eq. (31), a is the number of samples whose true labels and observed labels are same while b is the number of samples whose true labels and observed labels are different. For AE, its computation expression is given in Eq. (32) and for each sample, its entropy is given by Eq. (17). What's more, clustering time (CT) (in seconds) is also a widely used criterion. Its computation expression is given in Eq. (33) where t_s represents the time when clustering is start and t_e represents the time when clustering results are given.

$$\text{accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{21}$$

$$\text{acc}^+ = \frac{\text{TP}}{\text{TP} + \text{FN}} = \text{recall}^+ = \text{sensitivity} \tag{22}$$

$$\text{acc}^- = \frac{\text{TN}}{\text{TN} + \text{FP}} = \text{recall}^- = \text{specificity} \tag{23}$$

$$\text{PPV} = \frac{\text{TP}}{\text{TP} + \text{FP}} = \text{precision} \tag{24}$$

$$\text{F-Measure} = \frac{2 \times \text{precision} \times \text{recall}^+}{\text{precision} + \text{recall}^+} \tag{25}$$

$$\text{G-Mean} = \sqrt{\text{acc}^+ \times \text{acc}^-} \tag{26}$$

$$\text{NMI} = \frac{2\sum_{r=1}^{R}\sum_{c=1}^{C}\frac{n_r^c}{n}\ln\left(\frac{n_r^c n}{\sum_{p=1}^{R} n_p^c \sum_{q=1}^{C} n_r^q}\right)}{H(\pi) + H(\zeta)} \tag{27}$$

$$H(\pi) = -\sum_{r=1}^{R}\frac{n_r}{n}\ln\left(\frac{n_r}{n}\right) \tag{28}$$

$$H(\zeta) = -\sum_{c=1}^{C}\frac{n^c}{n}\ln\left(\frac{n^c}{n}\right) \tag{29}$$

$$\text{Adj-RI} = \frac{\text{RI} - E[\text{RI}]}{\max \text{RI} - E[\text{RI}]} \tag{30}$$

$$\text{RI} = \frac{a + b}{Cn(Cn-1)/2} \tag{31}$$

$$\text{AE} = \sum_{i=1}^{n} H_i \tag{32}$$

$$\text{CT} = t_e - t_s \tag{33}$$

Seventh, in order to show the generalization performance of these clustering approaches, for each dataset, 10 test runs are conducted and the average performance as well as standard deviation are presented. Concretely speaking, according to Table 1, we know for each dataset, they consist of multiple samples. But indeed, these samples are collected from a larger data warehouse. So each time, we select some samples from this warehouse and carry out clustering. After 10 test runs, we can get 10 groups of clustering performances and the corresponding average performance and standard deviation are also gotten.

4.2 Comparison about clustering performance

According to the definitions of the clustering criteria, we know that a smaller CT or AE indicates a better clustering performance while for other criteria, a higher value denotes a better clustering performances. Then, we adopt Table 2 and Table 3 to show the related experimental results. According to these two tables, we find that (1) on all datasets, EMVMC has a better result in terms of accuracy, acc^-, PPV, F-Measure, G-Mean; (2) in terms of acc^+, EMVMC outperforms MVMC except for Washington; (3) for NMI, Adj-RI, and AE, EMVMC performs better on half of the datasets; (4) In terms of CT, although on some datasets, EMVMC has to cost more time due to EMVMC should compute the entropy firstly, but the extra time is less than 10% which is acceptable for us; (5) compared with MVMC, EMVMC has a smaller average standard deviation on each criterion which means the performance of EMVMC is more stable. Generally speaking, EMVMC has a better clustering performance than MVMC in average.

4.3 Relation between the number of nearest neighbors and clustering performance

According to the procedure of EMVMC, we know that the number of nearest neighbors k influences the fuzzy

Table 2 Comparisons of clustering performance on the used datasets

	Accuracy↑	acc^+ ↑	acc^- ↑	PPV↑	F-Measure↑	G-Mean↑	NMI↑	Adj-RI↑	AE↓	CT↓
EMVMC										
Course	0.87	0.89	0.85	0.86	0.86	0.87	0.63	0.63	0.01	**103.00**
Citeseer	0.62	0.70	0.60	0.26	0.37	0.65	0.80	0.85	0.10	299.00
Cora	0.71	0.79	0.70	0.30	0.43	0.74	0.85	0.85	**0.53**	**248.16**
Cornell	0.74	0.73	0.74	0.42	0.53	0.74	0.88	0.83	**0.62**	**17.89**
Texas	0.67	0.74	0.65	0.35	0.46	0.69	0.98	0.91	0.16	16.69
Washington	0.72	**0.78**	0.70	0.40	0.51	0.74	**0.56**	**0.60**	**0.55**	**21.28**
Wisconsin	0.64	0.65	0.64	0.31	0.42	0.64	0.79	0.81	**0.18**	22.55
News-M2	0.89	0.97	0.80	0.83	0.86	0.88	0.88	0.84	0.02	158.52
News-M5	0.91	0.92	0.90	0.71	0.80	0.91	**0.70**	0.75	0.09	65.66
News-M10	0.78	0.79	0.78	0.28	0.42	0.78	**0.78**	**0.78**	0.13	**68.97**
News-NG1	0.88	0.94	0.82	0.84	0.86	0.88	**0.39**	**0.37**	0.03	66.89
News-NG2	0.89	0.93	0.87	0.65	0.75	0.90	0.78	0.76	0.55	**58.10**
News-NG3	0.83	0.93	0.81	0.41	0.55	0.87	**0.60**	**0.62**	0.37	133.42
Reuters	0.71	0.70	0.72	0.33	0.45	0.71	**0.79**	**0.79**	0.33	353.33
MVMC										
Course	0.84	0.84	0.84	0.84	0.84	0.84	0.52	0.57	0.02	94.88
Citeseer	0.59	0.63	0.58	0.23	0.33	0.61	0.72	0.73	0.40	299.00
Cora	0.70	0.76	0.69	0.29	0.41	0.72	0.71	0.70	0.25	243.48
Cornell	0.73	0.71	0.73	0.40	0.52	0.72	0.63	0.64	0.32	16.99
Texas	0.65	0.72	0.63	0.33	0.43	0.67	0.70	0.66	0.37	17.20
Washington	0.70	0.79	0.68	0.38	0.49	0.73	0.71	0.74	0.40	19.23
Wisconsin	0.61	0.63	0.60	0.28	0.39	0.61	0.58	0.58	0.05	24.27
News-M2	0.87	0.95	0.79	0.82	0.84	0.87	0.62	0.65	0.02	162.53
News-M5	0.89	0.91	0.89	0.67	0.76	0.90	0.72	0.67	0.21	68.32
News-M10	0.74	0.73	0.74	0.24	0.36	0.74	0.84	0.84	0.21	68.95
News-NG1	0.86	0.93	0.78	0.81	0.83	0.85	0.84	0.86	0.04	71.94
News-NG2	0.87	0.92	0.86	0.62	0.72	0.89	0.52	0.51	0.55	51.88
News-NG3	0.81	0.88	0.81	0.39	0.53	0.84	0.66	0.65	0.42	146.68
Reuters	0.66	0.69	0.65	0.28	0.40	0.67	0.92	0.92	0.38	368.64

For the first eight criteria, the higher, the better. For EMVMC, result in bold represents that the performance of EMVMC is worse

membership of a sample and as a result, the clustering performance will also be affected. Thus here, we show the relation between k and clustering performance. For convenience, we only show the relation in terms of accuracy with Fig. 1 due to for other clustering criteria, we can draw a similar conclusion. According to this figure, it is found that the accuracy of EMVMC has an approximate monotonically increasing property when $k \in [1, 10]$ while an approximate monotonically decreasing property when $k \in [11, 20]$. This indicates that if we want to get a better accuracy, maybe setting k be 9, 10, 11 is much better. Besides that, we also find that for other criteria, we can get the similar results.

4.4 Significance analysis

In order to show the effectiveness of the proposed EMVMC, we adopt significance analysis including the paired t test [43] and Friedman–Nemenyi statistical test [44]. In terms of the paired t test [43], it is used to analyze if the differences between two compared approaches on one dataset are significant or not. In terms of Friedman–Nemenyi statistical test [44], Friedman test is used to analyze if the differences between all compared approaches on multiple datasets are significant or not while Nemenyi test is used to analyze if the differences between two compared approaches on multiple datasets are significant or not.

(A) In our experiments, the paired t test [43] is used to analyze that for a clustering criterion, if the difference between EMVMC and MVMC on one dataset is significant or not. Concretely speaking, since for each dataset, we run the test for 10 times, so with EMVMC and MVMC used, for each clustering criterion A, each dataset B has 10 results and we record them as $A - B$(EMVMC) and $A - B$(MVMC), respectively. Then, we carry out paired t

Table 3 Comparisons of the standard deviation of clustering performance on the used datasets

	Accuracy	acc^+	acc^-	PPV	F-Measure	G-Mean	NMI	Adj-RI	AE	CT
EMVMC										
Course	0.05	0.07	0.08	0.03	0.05	0.02	0.05	0.03	0.00	6.84
Citeseer	0.02	0.05	0.05	0.02	0.03	0.06	0.00	0.05	0.01	17.91
Cora	0.03	0.02	0.05	0.03	0.04	0.03	0.01	0.05	0.05	23.49
Cornell	0.01	0.01	0.06	0.02	0.03	0.02	0.02	0.00	0.03	0.34
Texas	0.05	0.03	0.01	0.01	0.03	0.03	0.06	0.08	0.00	1.61
Washington	0.05	0.07	0.03	0.02	0.02	0.00	0.03	0.02	0.00	0.42
Wisconsin	0.01	0.02	0.02	0.03	0.01	0.02	0.07	0.01	0.02	1.60
News-M2	0.02	0.01	0.05	0.05	0.04	0.06	0.06	0.04	0.00	12.80
News-M5	0.05	0.05	0.09	0.01	0.00	0.06	0.01	0.05	0.00	5.29
News-M10	0.04	0.03	0.04	0.00	0.04	0.06	0.02	0.01	0.00	5.46
News-NG1	0.07	0.06	0.01	0.06	0.02	0.03	0.00	0.02	0.00	5.35
News-NG2	0.07	0.01	0.08	0.05	0.03	0.08	0.02	0.07	0.03	3.95
News-NG3	0.06	0.00	0.00	0.01	0.03	0.08	0.03	0.06	0.02	2.61
Reuters	0.07	0.05	0.03	0.00	0.04	0.03	0.05	0.01	0.01	31.38
MVMC										
Course	0.14	0.05	0.08	0.05	0.16	0.11	0.08	0.11	0.00	5.14
Citeseer	0.02	0.08	0.01	0.01	0.04	0.09	0.09	0.00	0.00	10.93
Cora	0.11	0.04	0.08	0.01	0.02	0.14	0.04	0.09	0.05	19.79
Cornell	0.02	0.04	0.11	0.08	0.09	0.06	0.03	0.00	0.04	1.65
Texas	0.06	0.06	0.05	0.05	0.04	0.13	0.02	0.05	0.04	1.44
Washington	0.05	0.13	0.14	0.06	0.06	0.10	0.13	0.12	0.03	3.32
Wisconsin	0.08	0.03	0.08	0.02	0.04	0.12	0.03	0.07	0.00	2.99
News-M2	0.07	0.12	0.15	0.16	0.15	0.17	0.04	0.13	0.00	20.76
News-M5	0.01	0.07	0.07	0.08	0.14	0.13	0.03	0.13	0.04	8.14
News-M10	0.04	0.09	0.06	0.01	0.00	0.02	0.12	0.16	0.02	1.47
News-NG1	0.14	0.03	0.07	0.14	0.15	0.02	0.07	0.11	0.00	1.24
News-NG2	0.06	0.08	0.03	0.02	0.03	0.14	0.08	0.07	0.10	1.33
News-NG3	0.12	0.11	0.12	0.03	0.04	0.08	0.09	0.01	0.04	26.61
Reuters	0.10	0.11	0.09	0.01	0.02	0.11	0.13	0.17	0.01	60.87

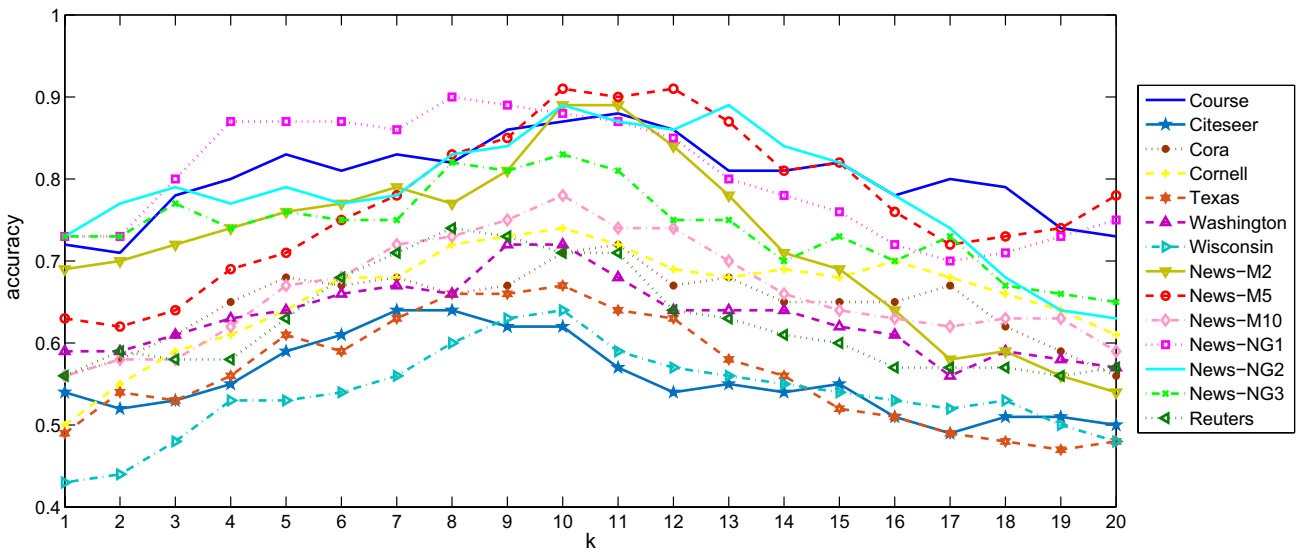


Fig. 1 Relation between the number of nearest neighbors and clustering performance in terms of accuracy

test on $A-B$(EMVMC) and $A-B$(MVMC) and get a sig-value. If sig-value is less than 0.05, it denotes that EMVMC and MVMC have a significant difference in this clustering criterion A on this dataset B. Otherwise, if the sig-value is more than 0.5, it denotes that EMVMC and MVMC have not a significant difference in A on B. What's more, if the sig-value is smaller, the difference between EMVMC and MVMC is more significant.

(A-1) According to the definition of paired t test and sig-value, we use Table 4 to show the sig-value of each dataset for each clustering criterion. From this table, it is found that only on NMI, Adj-RI, AE, and CT, the differences between EMVMC and MVMC on some datasets are not significant while on other criteria, their differences are significant on each dataset. What's more, according to Table 2 and Table 4, we find that for those cases which differences are not significant except for CT on Reuters, the performance of our EMVMC is not significant worse than the one of MVMC. In other words, for each clustering criterion, EMVMC outperforms MVMC on each dataset in statistical except for CT on Reuters.

(B) In our work, we adopt Friedman–Nemenyi statistical test [44] to analyze the significance of difference between these multi-view clustering approaches on multiple datasets. According to the definitions of Friedman test and Nemenyi test, since the number of compared approaches is two, thus Friedman test and Nemenyi test should draw a same conclusion. Now we carry out the Friedman test and Nemenyi test on each clustering criterion as below.

(B-1) For each clustering criterion, Friedman test ranks the approaches for each dataset separately, the best performing approach getting the rank of 1, the second best rank 2,⋯, as shown in Table 5. If the performances are same, the ranks should be averaged. Further, the average ranks on each clustering criterion for the approaches are also assigned. We define that r_i^j be the rank of jth of k approaches on the ith of N datasets. Then, for jth approach, the Friedman test compares the average ranks, namely R_j (see Eq. 34). The Friedman statistic χ_{F}^2 (see Eq. 35) is distributed according to χ_{F}^2 with $k-1$ degrees of freedom. On the base of χ_{F}^2, another statistic F_{F} is given in Eq. (36). In Eq. (36), F_{F} is distributed according to the F-distribution with $k-1$ and $(k-1)(N-1)$ degrees of freedom. According to [44] said, by carrying out the Friedman test, if χ_{F}^2 or F_{F} is larger than the critical value $F_\alpha(k-1,(k-1)(N-1))$ where α is a confidence level, we should reject the null-hypothesis, namely the differences between all compared approaches on the multiple datasets are significant. Moreover, if F_{F} is N/A, i.e., the denominator of Eq. (36) is 0, and χ_{F}^2 is still larger than $F_\alpha(k-1,(k-1)(N-1))$, we can say that the differences between all compared approaches are absolute significant. The table of critical values can be found in any statistical book and in generally, α is set to be 0.05 or 0.10.

$$R_j = \frac{1}{N}\sum_{i=1}^{N} r_i^j \tag{34}$$

$$\chi_{\mathrm{F}}^2 = \frac{12N}{k(k+1)}\left[\sum_{j=1}^{k} R_j^2 - \frac{k(k+1)^2}{4}\right] \tag{35}$$

$$F_{\mathrm{F}} = \frac{(N-1)\chi_{\mathrm{F}}^2}{N(k-1)-\chi_{\mathrm{F}}^2} \tag{36}$$

(B-2) After that, we carry out Nemenyi test and in our experiments, for each clustering criterion, we compute

Table 4 The paired t test comparisons of EMVMC and MVMC on different clustering criteria for the used datasets

Sig-value	Accuracy	acc$^+$	acc$^-$	PPV	F-Measure	G-Mean	NMI	Adj-RI	AE	CT
Course	0.034	0.050	0.011	0.018	0.026	0.034	0.047	0.040	0.044	0.033
Citeseer	0.044	0.026	0.031	0.045	0.012	0.003	0.037	0.008	0.036	0.000
Cora	0.021	0.032	0.019	0.041	0.043	0.026	0.030	0.035	**0.058**	**0.069**
Cornell	0.017	0.028	0.014	0.039	0.031	0.021	0.020	0.042	0.039	0.012
Texas	0.032	0.029	0.033	0.015	0.049	0.031	0.042	0.028	0.001	0.031
Washington	0.029	0.013	0.038	0.039	0.044	0.019	0.047	0.013	0.035	0.036
Wisconsin	0.007	0.030	0.026	0.042	0.024	0.043	0.024	0.002	0.003	0.039
News-M2	0.017	0.018	0.016	0.014	0.015	0.017	0.028	0.032	0.003	0.025
News-M5	0.019	0.013	0.021	0.006	0.042	0.017	**0.056**	0.020	0.009	0.040
News-M10	0.048	0.029	0.045	0.010	0.020	0.017	**0.057**	0.045	0.015	0.000
News-NG1	0.025	0.013	0.039	0.030	0.028	0.026	0.022	**0.053**	0.043	0.047
News-NG2	0.016	0.011	0.018	0.046	0.034	0.014	0.049	0.027	0.005	0.036
News-NG3	0.013	0.019	0.007	0.048	0.037	0.030	**0.061**	0.028	0.001	0.021
Reuters	0.010	0.007	0.005	0.011	0.040	0.040	0.008	0.020	0.040	**0.063**

The sig-values which are not smaller than 0.5 are given in bold

the CD_α with Eq. (37) where critical value q_α is given in Table 6 [44]. If the difference of the ranks for two compared approaches on a clustering criterion is larger than CD_α, we say that in terms of this criterion, the two compared approaches have a significant difference on all used datasets when confidence level is α.

$$CD_\alpha = q_\alpha \sqrt{\frac{k(k+1)}{6N}} \tag{37}$$

(B-3) Now according to Table 5 and the procedures of Friedman–Nemenyi statistical test, we can get the test results for each clustering criterion. The results are shown in Table 7. According to the results given in Tables 5 and 7, it is found that (1) for clustering criteria accuracy, acc^-, PPV, F-Measure, and G-Mean, since χ_F^2 is larger than $F_{0.05}(1, 13)$ and $F_{0.10}(1, 13)$ and then F_F is N/A, so EMVMC and MVMC have an absolute significant difference on multiple datasets in terms of these criteria. Moreover, since the average rank difference of EMVMC and MVMC is $2 - 1 = 1$, and 1 is larger than $CD_{0.05}$ and $CD_{0.10}$, thus we can also draw a same conclusion; (2) for acc^+, since both χ_F^2 and F_F are larger than $F_{0.05}(1, 13)$ and $F_{0.10}(1, 13)$ and the average rank difference of EMVMC and MVMC is $1.93 - 1.07 = 0.86$ which is larger than $CD_{0.05}$ and $CD_{0.10}$, so we can draw a conclusion that EMVMC and MVMC have a significant difference on multiple datasets in terms of acc^+; (3) for other four criteria, since both χ_F^2 and F_F are smaller than $F_{0.05}(1, 13)$ and $F_{0.10}(1, 13)$ and the average rank differences of EMVMC and

Table 5 Rank comparisons of multi-view clustering approaches on the used datasets for each clustering criterion

EMVMC/MVMC	Accuracy	acc^+	acc^-	PPV	F-Measure	G-Mean	NMI	Adj-RI	AE	CT
Course	1/2	1/2	1/2	1/2	1/2	1/2	1/2	1/2	1/2	2/1
Citeseer	1/2	1/2	1/2	1/2	1/2	1/2	1/2	1/2	1/2	1.5/1.5
Cora	1/2	1/2	1/2	1/2	1/2	1/2	1/2	1/2	2/1	2/1
Cornell	1/2	1/2	1/2	1/2	1/2	1/2	1/2	1/2	2/1	2/1
Texas	1/2	1/2	1/2	1/2	1/2	1/2	1/2	1/2	1/2	1/2
Washington	1/2	2/1	1/2	1/2	1/2	1/2	2/1	2/1	2/1	2/1
Wisconsin	1/2	1/2	1/2	1/2	1/2	1/2	1/2	1/2	2/1	1/2
News-M2	1/2	1/2	1/2	1/2	1/2	1/2	1/2	1/2	1/2	1/2
News-M5	1/2	1/2	1/2	1/2	1/2	1/2	2/1	1/2	1/2	1/2
News-M10	1/2	1/2	1/2	1/2	1/2	1/2	2/1	2/1	1/2	2/1
News-NG1	1/2	1/2	1/2	1/2	1/2	1/2	2/1	2/1	1/2	1/2
News-NG2	1/2	1/2	1/2	1/2	1/2	1/2	1/2	1/2	1/2	2/1
News-NG3	1/2	1/2	1/2	1/2	1/2	1/2	2/1	2/1	1/2	1/2
Reuters	1/2	1/2	1/2	1/2	1/2	1/2	2/1	2/1	1/2	1/2
Average	1/2	1.07/1.93	1/2	1/2	1/2	1/2	1.43/1.57	1.36/1.64	1.29/1.71	1.46/1.54

Table 6 Critical values for the two-tailed Nemenyi test

No. approaches	2	3	4	5	6	7	8	9	10
$q_{0.05}$	1.960	2.343	2.569	2.728	2.850	2.949	3.031	3.102	3.164
$q_{0.10}$	1.645	2.052	2.291	2.459	2.589	2.693	2.780	2.855	2.920

Table 7 Friedman–Nemenyi statistical test comparisons of multi-view clustering approaches on the used datasets for each clustering criterion

	accuracy	acc^+	acc^-	PPV	F-Measure	G-Mean	NMI	Adj-RI	AE	CT
N	14.00	14.00	14.00	14.00	14.00	14.00	14.00	14.00	14.00	14.00
k	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00
χ_F^2	14.00	10.29	14.00	14.00	14.00	14.00	0.29	1.14	2.57	0.07
F_F	N/A	36.00	N/A	N/A	N/A	N/A	0.27	1.16	2.93	0.07
$F_{0.05}(1, 13)$	4.6672	4.6672	4.6672	4.6672	4.6672	4.6672	4.6672	4.6672	4.6672	4.6672
$F_{0.10}(1, 13)$	3.1362	3.1362	3.1362	3.1362	3.1362	3.1362	3.1362	3.1362	3.1362	3.1362
$CD_{0.05}$	0.52	0.52	0.52	0.52	0.52	0.52	0.52	0.52	0.52	0.52
$CD_{0.10}$	0.44	0.44	0.44	0.44	0.44	0.44	0.44	0.44	0.44	0.44

MVMC are both smaller than $CD_{0.05}$ and $CD_{0.10}$, so we can say that in terms of these four criteria, EMVMC and MVMC have not a significant difference on multiple datasets. Indeed, according to Table 4, we know that on some datasets including Cora, News-M5, News-M10, News-NG1, News-NG3, and Reuters, the differences between our EMVMC and the compared MVMC are not significant and that's why on multiple datasets, they cannot have a significant difference. But according to what we have said in paired *t* test, namely on most of those datasets and corresponding criteria, the performance of EMVMC is not significant worse than the one of MVMC, so we can also draw a conclusion here that for each clustering criterion, EMVMC outperforms MVMC on multiple datasets in statistical.

5 Conclusions

Multi-view matrix completion (MVMC) is a multi-view clustering approach with side information including must-link and cannot-link and in MVMC, the relationship between cluster and sample has only two cases, sample belongs to a cluster and sample does not belong to a cluster. But in real-world applications, each sample belongs to a cluster with a possibility. Thus in our work, we introduce entropy to evaluate the class certainty of each sample and develop an entropy-based multi-view matrix completion for clustering with side information (EMVMC). Different from MVMC, EMVMC reflects the fuzzy membership between samples.

In order to validate the effectiveness of our EMVMC, multi-view datasets Course, Citeseer, Cora, WebKB, News-Group, and Reuters are adopted for experiments. Then, we show the clustering performance comparison, relation between the number of nearest neighbors and clustering performance, and significance analysis.

According to the experimental results, it is found that (1) EMVMC outperforms MVMC on each dataset in terms of accuracy, acc^{-}, PPV, F-Measure, and G-Mean; (2) EMVMC has a better performance on most of the datasets in terms of acc^{+}, NMI, Adj-RI, AE, and CT; (3) the performance of EMVMC is more stable; (4) when the number of nearest neighbors k is set to be 9, 10, 11, the performance of EMVMC is much better; (5) in terms of each clustering criterion, according to paired *t* test, EMVMC has a better performance than MVMC on each dataset in statistical except for CT on Reuters and according to Friedman–Nemenyi statistical test, EMVMC is significant better than MVMC on multiple datasets in statistical. In generally, our proposed EMVMC combines the entropy with MVMC and boosts the clustering performance in average.

Acknowledgements This work is supported by (1) National Key R&D Program of China (Grant No. 213), (2) National Natural Science Foundation of China (Grant Nos. 61673301, 61602296, 51575336), (3) Natural Science Foundation of Shanghai (Grant No. 16ZR1414500) and the authors would like to thank their supports.

References

1. Tang JJ, Li DW, Tian YJ, Liu DL (2018) Multi-view learning based on nonparallel support vector machine. Knowl Based Syst 158:94–108
2. Tang JJ, Tian YJ, Liu XH, Li DW, Lv J, Kou G (2018) Improved multi-view privileged support vector machine. Neural Netw 106:96–109
3. Zhao Y, You XG, Yu SJ, Xu C, Yuan W, Jing XY, Zhang TP, Tao DC (2018) Multi-view manifold learning with locality alignment. Pattern Recognit 78:154–166
4. Li JH, Wang CD, Li PZ, Lai JH (2018) Discriminative metric learning for multi-view graph partitioning. Pattern Recognit 75:199–213
5. Blum A, Mitchell T (1998) Combining labeled and unlabeled data with co-training. In: Eleventh conference on computational learning theory, pp 92–100
6. Xue Z, Li GR, Huang QM (2018) Joint multi-view representation and image annotation via optimal predictive subspace learning. Inf Sci 451–452:180–194
7. Li JX, Zhang B, Lu GM, Zhang D (2019) Generative multi-view and multi-feature learning for classification. Inf Fusion 45:215–226
8. Zhu Y, Gao W, Zhou ZH (2015) One-pass multi-view learning. J Mach Learn Res 30:1–16
9. Huang SD, Kang Z, Xu ZL (2018) Self-weighted multi-view clustering with soft capped norm. Knowl Based Syst 158:1–8
10. Brbić M, Kopriva I (2018) Multi-view low-rank sparse subspace clustering. Pattern Recognit 73:247–258
11. Huang L, Chao HY, Wang CD (2019) Multi-view intact space clustering. Pattern Recognit 86:344–353
12. Chao GQ, Sun SL (2019) Semi-supervised multi-view maximum entropy discrimination with expectation Laplacian regularization. Inf Fusion 45:296–306
13. Huang FR, Zhang XM, Zhao ZH, Li ZJ, He YY (2018) Deep multi-view representation learning for social images. Appl Soft Comput 73:106–118
14. Sun SL, Taylor JS, Mao L (2017) PAC-Bayes analysis of multi-view learning. Inf Fusion 35:117–131
15. Houthuys L, Langone R, Suykens JAK (2018) Multi-view least squares support vector machines classification. Neurocomputing 282:78–88
16. Rabbouch H, Saâdaoui F, Mraihi R (2017) Unsupervised video summarization using cluster analysis for automatic vehicles counting and recognizing. Neurocomputing 260:157–173
17. Azadani MN, Ghadiri N, Davoodijam E (2018) Graph-based biomedical text summarization: an itemset mining and sentence clustering approach. J Biomed Inform 84:42–58
18. Zheng CT, Liu C, Wong HS (2018) Corpus-based topic diffusion for short text clustering. Neurocomputing 275:2444–2458
19. Dougherty ER, Barrera J, Brun M, Kim S, Cesar RM, Chen Y, Bittner M, Trent JM (2002) Inference from clustering with application to gene-expression microarrays. Comput Biol 9(1):105–126
20. He CB, Tang Y, Liu H, Fei X, Li HC, Liu SY (2019) A robust multi-view clustering method for community detection combining link and content information. Phys A Stat Mech Appl 514:396–411
21. Zhao P, Jiang Y, Zhou ZH (2017) Multi-view matrix completion for clustering with side information. In: Proceedings of the 21st

Pacific-Asia conference on knowledge discovery and data mining, pp 403–415
22. Zhu CM, Wang Z (2017) Entropy-based matrix learning machine for imbalanced data sets. Pattern Recognit Lett 88:72–80
23. Davis JV, Kulis B, Jain P, Sra S, Dhillon IS (2007) Information-theoretic metric learning. In: International conference on machine learning, pp 209-216
24. Yi J, Zhang L, Jin R, Qian Q, Jain AK (2013) Semi-supervised clustering by input pattern assisted pairwise similarity matrix completion. In: International conference on machine learning, pp 1400–1408
25. Wang Y, Xiang Y, Zhang J, Zhou WL, Xie BL (2014) Internet traffic clustering with side information. J Comput Syst Sci 80(5):1021–1036
26. Goldberg D, Nichols D, Oki BM, Terry D (1992) Using collaborative filtering to weave an information tapestry. Commun ACM 35(12):61–70
27. Candès EJ, Recht B (2012) Exact matrix completion via convex optimization. Commun ACM 55(6):111–119
28. Xu M, Jin R, Zhou ZH (2013) Speedup matrix completion with side information: application to multi-label learning. In: Conference on neural information processing systems, vol 27, pp 2301–2309
29. Jalali A, Chen Y, Sanghavi S, Xu H (2011) Clustering partially observed graphs via convex optimization. In: International conference on machine learning, pp 1001–1008
30. Yi J, Jin R, Jain AK, Jain S, Yang T (2012) Semi-crowdsourced clustering: generalizing crowd labeling by robust distance metric learning. In: Conference on neural information processing systems, vol 25, pp 1772–1780
31. Liu Z, Hu ZX, Nie FP (2018) Matrix completion and vector completion via robust subspace learning. Neurocomputing 306:171–181
32. Ye HJ, Zhan DC, Miao Y, Jiang Y, Zhou ZH (2015) Rank consistency based multi-view learning: a privacy-preserving approach. In: ACM international on conference on information and knowledge management, pp 991–1000
33. Sen P, Namata GM, Bilgic M, Getoor L, Gallagher B, Eliassi-Rad T (2008) Collective classification in network data. AI Mag 29(3):93–106
34. Bisson G, Grimal C (2012) Co-clustering of multi-view datasets: a parallelizable approach. In: Proceedings of the IEEE 12th international conference on data mining, pp 828–833
35. Amini MR, Usunier N, Goutte C (2009) Learning from multiple partially observed views—an application to multilingual text categorization. In: Neural information processing systems (NIPS), pp 28-36
36. http://multilingreuters.iit.nrc.ca/ReutersMultiLingualMultiView.htm
37. Kumar A, Rai P, Daume H (2011) Co-regularized multi-view spectral clustering. In: Conference on neural information processing systems, vol 24, pp 1413–1421
38. Tzortzis G, Likas A (2012) Kernel-based weighted multi-view clustering. In: IEEE international conference on data mining, pp 675–684
39. Xia R, Pan Y, Du L, Yin J (2014) Robust multi-view spectral clustering via low-rank and sparse decomposition. In: AAAI conference on artificial intelligence, pp 2149–2155
40. Gu Q, Zhu L, Cai ZH (2009) Evaluation measures of the classification performance of imbalanced data sets. Comput Intell Intell Syst 51:461–471
41. Tzortzis GF, Likas AC (2009) The global kernel k-means algorithm for clustering in feature space. IEEE Trans Neural Netw 20(7):1181–1194
42. Hubert L, Arabie P (1985) Comparing partitions. J Classif 2(1):193–218
43. Vapnik V (1998) Statistical learning theory. Wiley, New York
44. Demsar J (2006) Statistical comparisons of classifiers over multiple data sets. J Mach Learn Res 7:1–30