

A Semi-supervised Three-Way Clustering Framework for Multi-view Data

Hong Yu^(✉), Xincheng Wang, and Guoyin Wang

Chongqing Key Laboratory of Computational Intelligence,
Chongqing University of Posts and Telecommunications,
Chongqing 400065, People's Republic of China
yuhong@cqupt.edu.cn

Abstract. A new semi-supervised clustering framework for uncertain multi-view data is proposed inspired by the theory of three-way decisions, which is an alternative formulation different from the ones used in the existing studies. A cluster is represented by three regions such as the core region, fringe region and trivial region. The three-way representation intuitively shows which objects are fringe to the cluster. The proposed method is an iterative processing which includes two parts: (1) the three-way spectral clustering algorithm which is devised to obtain the three-way representation result; and (2) the active learning strategy which is designed to obtain the prior supervision information from the fringe regions, and the pairwise constraints information is used to adjust the similarity matrix between objects. Experimental results show that the proposed method can cluster multi-view data effectively and is better in performances than the compared single-view clusterings and other semi-supervised clustering approaches.

Keywords: Multi-view data · Three-way decisions · Semi-supervised clustering · Spectral clustering · Active learning

1 Introduction

In some applications such as computer video, social computing, and multimedia area, objects are usually represented in several different ways. This kind of data is termed as the multi-view data. Multi-view clustering, which is also one kind of multi-view learning, has attracted more and more attentions [1–3, 12, 19]. In the existing methods, spectral clustering [4, 13] is a popular one for multi-view data because it represents multi-view data via graph structure and makes it possible to handle complex data such as high-dimensional and heterogeneous as well as it can easily use the pairwise constraint information provided by users. Therefore, some scholars research on spectral clustering for multi-view data [5, 8, 17].

Generally speaking, there are two types of typical prior supervised information, namely, class labels and pairwise constraints [5, 6, 16]. In practice, it is difficult to obtain the independent class labels, yet it could be relatively easy to ensure correlated or uncorrelated information among data objects. Therefore,

pairwise constraints describe two objects whether they should be assigned to the same class or the different classes. However, choosing the supervised information is random in most of existing methods, and it does not produce positive effect on improving the clustering result when the algorithm itself can find the prior information or there are amounts of noises in the prior information. Therefore, the active learning method is introduced to optimize the selection of the constraints for semi-supervised clustering [15, 18, 28].

Most of the existing researches on the topic has focused on selecting an initial set of pairwise constraints before performing semi-supervised clustering. This is not suitable if we wish to iteratively improve the clustering model by actively querying users. In fact, many clustering approaches are based on iterative framework. Obviously, it is much better in each iteration that we determine objects with the most important information toward improving the current clustering result and form queries accordingly than just choosing the information randomly. The responses to the queries (i.e., constraints) are then used to update the clustering. This process repeats until we reach the stop conditions. Such an iterative framework is widely used in active learning for semi-supervised clustering.

In this paper, we focus on how to improve the quality of clustering for multi-view data with the aid of pairwise constraints. Therefore, we propose a semi-supervised clustering framework based on active learning by using three-way decisions. In order to further choosing the supervision information during the iterative processing, we introduce the idea of three-way decisions into this work, inspired by the three-way decisions theory as suggested by Yao [21, 22]. Three-way decisions extend binary-decisions in order to overcome some drawbacks of binary-decisions. The basic ideas of three-way decisions have been widely used in real-world decision-making problems, such as decision making [23], email spam filtering [29], three-way investment decisions [9] and many others [25]. Interval sets provide an ideal mechanism to represent soft clustering. Lingras and Yan [10] introduced interval sets to represent clusters. Lingras and West [11] proposed an interval set clustering method with rough k-means for mining clusters of web visitors. Yao et al. [20] represented each cluster by an interval set instead of a single set as the representation of a cluster. Inspired by these results, we have introduced a framework of three-way cluster analysis [26, 27].

In our work, a three-way representation for a cluster is presented, where a cluster is represented by three regions, i.e., the core region, fringe region and trivial region, instead of two regions as the other existing methods. Objects in the core region are typical elements of the cluster, objects in the fringe region are fringe elements of the cluster, and objects in the trivial region do not belong to the cluster definitely. A cluster is therefore more realistically characterized by a set of core objects and a set of boundary objects. The three-way representation intuitively shows which objects are fringe to the cluster. Thus, we can reduce the search space to the fringe regions when selecting the pairwise constraints. The basic idea of the work is to propose an iterative processing, in which a three-way clustering algorithm is devised to obtain the three-way clustering result and an active learning strategy is designed to obtain the prior supervision information from the fringe regions.

The remainder of this paper is organized as follows. Section 2 introduces some basic concepts. Section 3 describes the proposed framework, the three-way spectral clustering algorithm and the active learning strategy. Section 4 reports the results of comparative experiments and conclusions are provided in Sect. 5.

2 Preliminaries

In this section, some basic concepts of multi-view and semi-supervised clustering are introduced.

2.1 Multi-view Data

In the multi-view setting, an object (data point) \mathbf{x} is described with several different disjoint sets of features. Let $X = \{\mathbf{x}_1, \dots, \mathbf{x}_i, \dots, \mathbf{x}_N\}$ be a universe with N objects. There are H numbers of views to describe the objects, and $X^{(1)}, X^{(2)}, \dots, X^{(h)}, \dots, X^{(H)}$ be the data matrix of each view respectively.

For h -th view, $X^{(h)} \in \mathbb{R}^{N \times d^{(h)}}$, and $d^{(h)}$ is the feature dimension of the h -th view. $X^{(h)} = \{\mathbf{x}_1^{(h)}, \mathbf{x}_2^{(h)}, \dots, \mathbf{x}_i^{(h)}, \dots, \mathbf{x}_N^{(h)}\}$, where $\mathbf{x}_i^{(h)} = (x_{i,h}^1, x_{i,h}^2, \dots, x_{i,h}^j, \dots, x_{i,h}^{d^{(h)}})$ is its i -th object, and $x_{i,h}^j$ is the j -th feature of i -th object in the h -th view.

2.2 Pairwise Constraints

Pairwise constraints is one kind of typical prior information for semi-supervised clustering. Wagstaff and Cardie [14] introduce must-link (positive association) and cannot-link (negative association) to reflect the constraint relations between the data points.

For the universe $X = \{\mathbf{x}_1, \dots, \mathbf{x}_i, \dots, \mathbf{x}_N\}$, let $Y = \{y_1, \dots, y_i, \dots, y_N\}$ be the class labels of objects respectively. Must-link constraint requires that the two points must belong to the same cluster, and this relation is denoted by $ML = \{(\mathbf{x}_i, \mathbf{x}_j) \mid y_i = y_j, \text{ for } i \neq j, \mathbf{x}_i, \mathbf{x}_j \in X, y_i, y_j \in Y\}$. Cannot-link constraint requires that the two points must belong to different clusters, and this relation is denoted by $CL = \{(\mathbf{x}_p, \mathbf{x}_q) \mid y_p \neq y_q, \text{ for } p \neq q, \mathbf{x}_p, \mathbf{x}_q \in X, y_p, y_q \in Y\}$. Klein et al. [7] found that must-link constraint has the transitivity properties on objects, namely, for $\mathbf{x}_i, \mathbf{x}_j, \mathbf{x}_k \in X$,

$$\begin{aligned} (\mathbf{x}_i, \mathbf{x}_j) \in ML \ \& \ (\mathbf{x}_j, \mathbf{x}_k) \in ML \Rightarrow (\mathbf{x}_i, \mathbf{x}_k) \in ML, \\ (\mathbf{x}_i, \mathbf{x}_j) \in ML \ \& \ (\mathbf{x}_j, \mathbf{x}_k) \in CL \Rightarrow (\mathbf{x}_i, \mathbf{x}_k) \in CL. \end{aligned} \quad (1)$$

In fact, simply using the constraint information in the algorithm may cause a deflection problem of the singular points during the clustering process. The so-called deflection of singular points is that the points belonging to ML are assigned to CL or the points belonging to CL are assigned to ML . Therefore, it is not always true that there are more pairwise constraints the better the clustering result is. We hope to obtain the best possible result with fewer constraints which is just the purpose of active learning.

2.3 Representation of Three-Way Clustering

The purpose of clustering is to divide the N objects of a universe X into some clusters. If there are K clusters, the family of clusters, \mathbf{C} , is represented as $\mathbf{C} = \{C_1, \dots, C_k, \dots, C_K\}$. A cluster is usually represented by a single set in the existing works, namely, $C_k = \{\mathbf{x}_1, \dots, \mathbf{x}_i, \dots, \mathbf{x}_{|C_k|}\}$, and it is abbreviated as C by removing the subscript when there is no ambiguity. From the view of making decisions, the representation of a single set means that, the objects in the set belong to this cluster definitely, the objects not in the set do not belong to this cluster definitely. This is a typical result of two-way decisions. For hard clustering, one object just belong to one cluster; for soft clustering, one object might belong to more than one cluster. However, this representation cannot show which objects might belong to this cluster, and it cannot show the degree of the object influence on the form of the cluster intuitively. Thus, the use of three regions to represent a cluster is more appropriate than the use of a crisp set, which also directly leads to three-way decisions based interpretation of clustering.

In contrast to the general crisp representation of a cluster, we represent a three-way cluster C as a pair of sets:

$$C = (Co(C), Fr(C)). \tag{2}$$

Here, $Co(C) \subseteq X$ and $Fr(C) \subseteq X$. Let $Tr(C) = X - Co(C) - Fr(C)$. Then, $Co(C)$, $Fr(C)$ and $Tr(C)$ naturally form the three regions of a cluster as Core Region, Fringe Region and Trivial Region respectively. That is:

$$\begin{aligned} CoreRegion(C) &= Co(C), \\ FringeRegion(C) &= Fr(C), \\ TrivialRegion(C) &= X - Co(C) - Fr(C). \end{aligned} \tag{3}$$

If $\mathbf{x} \in CoreRegion(C)$, the object \mathbf{x} belongs to the cluster C definitely; if $\mathbf{x} \in FringeRegion(C)$, the object \mathbf{x} might belong to C ; if $\mathbf{x} \in TrivialRegion(C)$, the object \mathbf{x} does not belong to C definitely.

These subsets have the following properties.

$$\begin{aligned} X &= Co(C) \cup Fr(C) \cup Tr(C), \\ Co(C) \cap Fr(C) &= \emptyset, \\ Fr(C) \cap Tr(C) &= \emptyset, \\ Tr(C) \cap Co(C) &= \emptyset. \end{aligned} \tag{4}$$

If $Fr(C) = \emptyset$, the representation of C in Eq. (2) turns into $C = Co(C)$; it is a single set and $Tr(C) = X - Co(C)$. This is a representation of two-way decisions. In other words, the representation of a single set is a special case of the representation of three-way cluster.

Furthermore, according to Eq. (4), we know that it is enough to represent a cluster expediently by the core region and the fringe region.

In another way, we can define a cluster by the following properties:

$$\begin{aligned} (i) \quad &Co(C_k) \neq \emptyset, 1 \leq k \leq K; \\ (ii) \quad &\bigcup Co(C_k) \cup Fr(C_k) = X, 1 \leq k \leq K. \end{aligned} \tag{5}$$

Property (i) implies that a cluster cannot be empty. This makes sure that a cluster is physically meaningful. Property (ii) states that any object of X must definitely belong to or might belong to a cluster, which ensures that every object is properly clustered.

With respect to the family of clusters, \mathbf{C} , we have the following family of clusters formulated by three-way decisions as:

$$\mathbf{C} = \{(Co(C_1), Fr(C_1)), \dots, (Co(C_k), Fr(C_k)), \dots, (Co(C_K), Fr(C_K))\}. \quad (6)$$

Obviously, we have the following family of clusters formulated by two-way decisions as:

$$\mathbf{C} = \{Co(C_1), \dots, Co(C_k), \dots, Co(C_K)\}. \quad (7)$$

3 The Proposed Semi-supervised Clustering Method

In this section, a semi-supervised three-way clustering framework for multi-view data is proposed. The three-way spectral clustering algorithm and the active learning strategy are described.

3.1 The Framework

The proposed semi-supervised three-way clustering framework for multi-view data (or SS-TWC, for short) is shown in Fig. 1, which is an iterative processing. In short, the framework consists of two parts, i.e., the three-way clustering and the active learning. The main goal of Part 1 is to produce the clustering result in three-way representation. In other words, the other clustering algorithm also works as long as we alter it to adopt to the three-way representation. The task of Part 2 is to choose some objects (points) to query experts. The responses to the queries (i.e., constraints) are then used to update the clustering in Part 1.

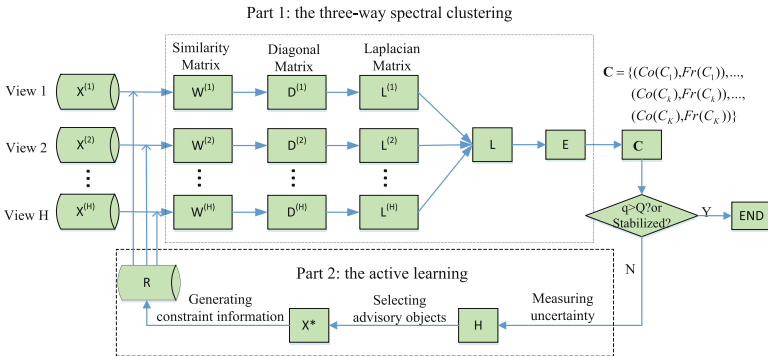


Fig. 1. SS-TWC: semi-supervised three-way clustering framework for multi-view data

In this paper, the spectral clustering approach is used to produce the three-way clustering in Part 1. The framework of three-way spectral clustering algorithm is described in Algorithm 1 in Subsect. 3.2. The algorithm computes on the multiple views $X^{(1)}, X^{(2)}, \dots, X^{(H)}$, to find a low-dimensional feature space E of original data points by calculating eigenvectors of fused Laplacian matrix L . In order to obtain more accurate partitions, uncertain objects are assigned to the fringe region of corresponding cluster. For these uncertain objects, they can get further decision when information is sufficient.

In Part 2, the SS-TWC uses the active learning method to learn dynamically objects with most important information toward improving the current clustering result. The framework of active learning algorithm is described in Algorithm 2 in Subsect. 3.3. In each iteration, the active learning measures uncertain objects in fringe regions with a certain strategy. The produced pairwise constraints information is applied to adjust the similarity matrix between data points in Algorithm 1, which makes objects being more compact in one cluster and more discrete in different clusters.

We need to note that the final result of clustering can be expressed by two-way or three-way representation according to the user demands. In the framework, the result of the first iteration is in three-way representation, and the fringe regions reduce after processing iterations. In each iteration, we query experts to acquire the prior information by choosing the objects from fringe regions. The algorithm obtains the two-way clustering result finally when the iteration is going on until the fringe regions are empty, which is the results in our experiments. The algorithm also can obtain the three-way clustering result finally, if it stops when the clustering result is stable or the iterative times q reaches the maximum number Q .

3.2 The Three-Way Spectral Clustering

First, we need to map the data set $X = \{\mathbf{x}_1, \dots, \mathbf{x}_i, \dots, \mathbf{x}_N\}$ to a similarity matrix W . We refer to the concept of k -nearest neighbors in consideration of the nearer neighbors contribute more to the classified information than the more distant ones.

In spectral clustering, the Gaussian kernel function is widely used as the similarity measure. However, it is difficult to determine the optimal value of the kernel parameter, which reflects the neighborhood of the data points. In addition, with a fixed kernel parameter, the similarity between two objects is only determined by their Euclidean distance. Inspired by the idea of local scaling parameter can be determined by shared nearest neighbors [24], we proposed an adaptive parameter based on shared neighbors instead of traditional kernel parameter.

Let $N(\mathbf{x}) = \{\mathbf{x}_s | \text{dist}(\mathbf{x}, \mathbf{x}_s) \leq r, \mathbf{x}_s \in kNN(\mathbf{x})\}$ contains objects that are members of neighborhood of \mathbf{x} with the neighbor radius r , where $\text{dist}(\mathbf{x}, \mathbf{x}_s)$ describes the Euclidean distance between \mathbf{x} and \mathbf{x}_s , $kNN(\mathbf{x})$ denotes k neighbor points of \mathbf{x} . The neighbor radius r of object \mathbf{x} is defined as $r = \frac{1}{k} \sum_{s=1}^k$

$dist(\mathbf{x}, \mathbf{x}_s)$, such that $\mathbf{x}_s \in kNN(\mathbf{x})$. Neighbor radius of each object can be confirmed by its k neighbor points. In addition, the number of points in the join neighborhood of two objects indicates their closeness. Therefore, a similarity function that considers global distribution and local consistency is given by:

$$w_{ij} = \begin{cases} \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{r_i^2 r_j^2 |N(\mathbf{x}_i) \cap N(\mathbf{x}_j)| + 1}\right), & \mathbf{x}_i \in kNN(\mathbf{x}_j) \text{ or } \mathbf{x}_j \in kNN(\mathbf{x}_i), \\ 0, & \text{others.} \end{cases} \quad (8)$$

where r_i and r_j are the neighbor radius of \mathbf{x}_i and \mathbf{x}_j respectively, $|N(\mathbf{x}_i) \cap N(\mathbf{x}_j)|$ is the number of objects in the join neighborhood of \mathbf{x}_i and \mathbf{x}_j .

We adopt graphs $G^{(1)}, \dots, G^{(H)}$ to describe the multiple views $X^{(1)}, \dots, X^{(H)}$ respectively. $G^{(h)} = (V^{(h)}, E^{(h)}, W^{(h)})$, where $W^{(h)}$ represents the similarity relationship among data points of h -th view. $L^{(h)}$ denotes the normalized graph Laplacian matrix of $G^{(h)}$ and is defined as:

$$L^{(h)} = I - (D^{(h)})^{-1/2} W^{(h)} (D^{(h)})^{-1/2}, \quad (9)$$

where $D^{(h)} \in \mathbb{R}^{N \times N}$ denotes the degree matrix of graph $G^{(h)}$ whose i -th diagonal element is $d_i^{(h)} = \sum_{j=1}^N w_{ij}^{(h)}$.

The objective function of the normalized spectral clustering is defined as:

$$\min_{G \in \mathbb{R}^{N \times K}} \sum_{h=1}^H \text{tr}(G^T L^{(h)} G), \quad \text{s.t. } G^T G = I \quad (10)$$

Due to G is the identical matrix of all views, Eq. 10 can be converted to:

$$\min_{G \in \mathbb{R}^{N \times K}} \text{tr}\left(G^T \left(\sum_{h=1}^H L^{(h)}\right) G\right), \quad \text{s.t. } G^T G = I \quad (11)$$

In fact, the three-way spectral clustering algorithm (see Algorithm 1) implements the initial clustering and the iterative clustering. In the initial processing, i.e. the iteration times be 1, the constraint set $R = \emptyset$. That means there is no prior information and the spectral clustering is a unsupervised learning for multi-view data. In the iterative clustering processing, i.e. the iteration times more than 1, pairwise constraints information produced by the active learning algorithm (see Algorithm 2) are added to the constraint set R . The similarity matrixes of spectral clustering are adjusted by the following formula:

$$\begin{aligned} \text{if } (\mathbf{x}_i, \mathbf{x}_j) \in ML, \text{ then } w_{ij}^{(h)} &= w_{ji}^{(h)} = 1; \\ \text{if } (\mathbf{x}_i, \mathbf{x}_j) \in CL, \text{ then } w_{ij}^{(h)} &= w_{ji}^{(h)} = 0. \end{aligned} \quad (12)$$

Based on the idea of three-way decisions, the proposed framework assigns the current uncertain objects to the corresponding fringe region. First, for the objects need to be divided, the proposed algorithm finds the neighborhood $N(\mathbf{x}_i)$ with neighbor radius r_i . Then, it calculates the proportion of the objects in $N(\mathbf{x}_i)$

Algorithm 1. The three-way spectral clustering algorithm

Input: the multi-view data $\{X^{(1)}, X^{(2)}, \dots, X^{(H)}\}$, the number of clusters K , the constraint set R , the threshold values α and β .

Output: $\mathbf{C} = \{(Co(C_1), Fr(C_1)), \dots, (Co(C_K), Fr(C_K))\}$.

for each view $X^{(h)}$ do

- if $R \neq \emptyset$ then**
 - └ to adjust similarity matrix $W^{(h)}$ according to Eq. 12;
- else**
 - └ to compute similarity w_{ij} according to Eq. 8;
- └ to construct the normalized graph Laplacian matrix $L^{(h)}$ according to Eq. 9;
- to compute the fused normalized Laplacian matrix L according to Eq. 11;
- to compute the K smallest eigenvectors of L and construct eigenvector matrix E ;
- to normalize the rows of E to have unit norm; to cluster E by using k-means algorithms, to assign objects to core regions, fringe regions and trivial regions, by using the three-way rules Eq. 14.

belong to each cluster; and it assigns \mathbf{x}_i to the corresponding core region or fringe region. The proportion that objects in $N(\mathbf{x}_i)$ belong to C_k is given by:

$$P(N(\mathbf{x}_i)|C_k) = \frac{|N(\mathbf{x}_i) \cap C_k|}{|N(\mathbf{x}_i)|} \quad (13)$$

Naturally, we have the three-way decision rules as follows.

$$\begin{aligned} & \text{if } P(N(\mathbf{x}_i)|C_k) \geq \alpha, \text{ then } \mathbf{x}_i \text{ is assigned to } Co(C_k); \\ & \text{if } \beta < P(N(\mathbf{x}_i)|C_k) < \alpha, \text{ then } \mathbf{x}_i \text{ is assigned to } Fr(C_k); \\ & \text{if } P(N(\mathbf{x}_i)|C_k) \leq \beta, \text{ then } \mathbf{x}_i \text{ is assigned to } Tr(C_k); \end{aligned} \quad (14)$$

where α and β are the three-way decision thresholds.

3.3 The Active Learning to Acquire Pairwise Constraints

In this work, we consider active learning of constraints in the iterative framework. The search space is reduced to the fringe regions in the proposed method. In the current iteration, we need to decide which objects have the most important information toward improving the current clustering result and form queries accordingly. The responses to the queries (i.e., constraints) are then used to update the similarity matrix by using Eq. 12.

Specifically, we define the uncertainty in terms of the concept of entropy, which is a classic measure of uncertainty. In the h -th view, $w_{ij}^{(h)}$ denotes similarity between points $\mathbf{x}_i^{(h)}$ and $\mathbf{x}_j^{(h)}$, the probability of $\mathbf{x}^{(h)}$ belongs to different core regions $Co(C_k)$ ($1 \leq k \leq K$) is defined as:

$$p^{(h)}(\mathbf{x} | Co(C_k)) = \frac{\frac{1}{|Co(C_k)|} \sum_{\mathbf{x}_j \in Co(C_k)} w_{ij}^{(h)}}{\sum_{p=1}^K \left(\frac{1}{|Co(C_p)|} \sum_{\mathbf{x}_j \in Co(C_p)} w_{ij}^{(h)} \right)}, \quad (15)$$

where $|Co(C_k)|$ is the cardinality of $Co(C_k)$.

Then, the maximum entropy of \mathbf{x} among H views is calculated by the following formula.

$$H_{Max}(\mathbf{x}) = \underset{h=1}{\operatorname{argmax}}^H \left(-\frac{1}{K} \sum_{k=1}^K (p^{(h)}(\mathbf{x} | Co(C_k)) \log_2 p^{(h)}(\mathbf{x} | Co(C_k))) \right). \quad (16)$$

An object with the bigger entropy will have more classification information to help decision-making. Thus, the object with most important information is selected by the following formula.

$$\mathbf{x}^* = \underset{\mathbf{x} \in U}{\operatorname{argmax}} H_{Max}(\mathbf{x}), \quad (17)$$

where U denotes the set of unlabeled data.

In order to reduce the cost of queries, we first sort the K probabilities, $p^{(h)}(\mathbf{x}^* | Co(C_k))$ for $1 \leq k \leq K$, in descending order. Then, we begin query the core of the cluster from the higher one until a must-link constraint satisfied.

Algorithm 2. The strategy of active learning

Input: the clustering result \mathbf{C}

Output: the constraint set R

for each view $X^{(h)}$ **do**

for $k = 1$ to K **do**

for $\mathbf{x} \in Fr(C_k)$, to compute $p^{(h)}(\mathbf{x} | Co(C_k))$ according to Eq. 15;

 to compute $H_{Max}(\mathbf{x})$ according to Eq. 16;

 to select the most information object \mathbf{x}^* according to Eq. 17;

 to sort probabilities $p(\mathbf{x}^* | Co(C_k))$ in descending order;

for $k = 1$ to K **do**

 to query whether \mathbf{x}^* belongs to $Co(C_k)$;

if the response is True **then**

 to select a point \mathbf{x} from $Co(C_k)$ randomly, set $(\mathbf{x}^*, \mathbf{x}) \in ML$, and

$Co(C_k) = Co(C_k) \cup \{\mathbf{x}^*\}$;

 to update the constraint set R according to Eq. 12; break;

else

 to construct pairwise constraint information $(\mathbf{x}^*, \mathbf{x}_i) \in CL$;

4 Experimental Results

In this section, we validate the proposed method on some real-world datasets. Table 1 gives the summary information about the datasets. SensIT¹ uses two sensors to classify three types of vehicle. We randomly sample 100 data for each class, and then conduct experiments on 2 views and 3 classes. Reuters²

¹ <https://www.csie.ntu.edu.tw/~cjlin/libsvmtools/datasets/multiclass.html>.

² <https://archive.ics.uci.edu/ml/datasets.html>.

contains feature characteristics of documents originally written in five different languages, and their translations over a common set of 6 categories. We use documents originally in English as the first view and their French and German translations as the second and the third view respectively. We randomly sample 1200 documents from this text collection, with each of the 6 clusters having 200 documents. Cora³ and CiteSeer⁴ collect kinds of scientific publications, with the first view being the textual content of documents and the second view being citation links between documents.

Table 1. Information about the datasets

Datasets	Size	#View	#Cluster
SensIT	300	2	3
Reuters	1200	3	6
Cora	2708	2	7
CiteSeer	3312	2	6

We compare the proposed SS-TWC method with some representative multi-view clustering strategies.

- Best Single View(BSV): running the proposed semi-supervised spectral clustering on each input view, and then reporting the results of the view that achieves the best performance.
- Feature Concatenation(FeatCon): concatenating the features of all views to form a single representation, and then applying the proposed semi-supervised spectral clustering on the concatenated view.
- AMVNMF: the adaptive multi-view semi-supervised nonnegative matrix factorization, it is an iterative multi-view semi-supervised clustering algorithm from the reference [16].
- SS-TWC(R): the method is similar with the SS-TWC except it obtains equivalent constraint information by using the random strategy instead of the active learning strategy in Algorithm 2.

The quality of the final clustering is evaluated by the traditional indices such as the accuracy (AC) and normalized mutual information (NMI). To ensure the objectivity of the experimental results, the results of AMVNMF are from the reference [16] and the other methods are programmed in C++. Each test runs 10 times, the average values of AC and the NMI are recorded in Table 2. The k , the number of neighbors, is set to be the 5% of the universe in the tests.

Obviously, the SS-TWC outperforms the other compared methods on the four datasets. Unlike the BSV and FeatCon, the SS-TWC profits from the correlative and complementary information among multiple views. Compared to the

³ <http://linqs.umiacs.umd.edu/projects//projects/lbc/index.html>.

⁴ <http://linqs.umiacs.umd.edu/projects//projects/lbc/index.html>.

Table 2. Comparison of experimental results

Indices	Datasets	BSV	FeatCon	AMVNMF	SS-TWC(R)	SS-TWC
AC	SensIT	68.73	76.39	71.33	67.71	77.67
	Reuters	54.67	58.43	59.88	57.20	66.22
	Cora	42.70	46.27	48.71	40.67	51.88
	Citeseer	46.16	53.36	53.14	45.39	56.04
NMI	SensIT	34.58	38.02	31.73	24.87	41.16
	Reuters	44.35	45.62	42.75	32.13	47.88
	Cora	30.71	34.41	34.59	20.19	36.72
	Citeseer	21.49	23.38	26.13	18.27	30.26

AMVNMF, the proposed SS-TWC has the benefit of processing the uncertainty in multi-view data by using the three-way decisions. In addition, the compared results between the SS-TWC and the SS-TWC(R) show that the proposed strategy of selecting pairwise constraints dynamically is much effective. In short, the proposed method work well in dealing with multi-view data.

5 Conclusions

In many scenarios, more than one view can be provided to describe the data due to the fact that data may be collected from different sources or be represented by different kind of feature sets for different tasks. Clustering multi-view data is an important problem. In this paper, we proposed a semi-supervised three-way clustering framework for multi-view data. The framework is an iterative processing, which consists of two parts, i.e., the three-way clustering and the active learning. The main goal of three-way clustering is to produce the clustering result in three-way representation, in which the objects in fringe regions intuitively gives the clue to query. Thus, the task of the active learning is to choose some objects (points) from fringe regions to query experts, and the responses to the queries (i.e., constraints) are then used to update the clustering result in iterations. The spectral clustering and the active learning strategies are used to implement the framework. The experimental results show that the proposed method achieves better performance in both the accuracy and the NMI than the compared methods. However, we need further work to improve the time complexity though the cost for computing has reduced little by constructing kNN graph. To consider the different contribution of different view is another direction for future research.

Acknowledgments. This work was supported in part by the National Natural Science Foundation of China under Grant Nos. 61379114 & 61533020.

References

1. Blum, A., Mitchell, T.: Combining labeled and unlabeled data with co-training. In: *Proceeding of the Eleventh Annual Conference on Computational Learning Theory*, pp. 92–100. ACM (1998)
2. Bickel, S., Scheffer, T.: Multi-view clustering. In: *ICDM*, vol. 4, pp. 19–26 (2004)
3. Chaudhuri, K., Kakade, S.M., Livescu, K., Sridharan, K.: Multi-view clustering via canonical correlation analysis. In: *Proceedings of the 26th Annual International Conference on Machine Learning*, pp. 129–136. ACM (2009)
4. Chen, W., Feng, G.: Spectral clustering: a semi-supervised approach. *Neurocomputing* **77**(1), 229–242 (2012)
5. Ding, S., Jia, H., Zhang, L., Jin, F.: Research of semi-supervised spectral clustering algorithm based on pairwise constraints. *Neural Comput. Appl.* **24**(1), 211–219 (2014)
6. Grira, N., Crucianu, M., Boujemaa, N.: Unsupervised and semi-supervised clustering: a brief survey (2005)
7. Klein, D., Kamvar, S.D., Manning, C.D.: From instance-level constraints to space-level constraints: making the most of prior knowledge in data clustering. Stanford (2002)
8. Li, Y., Nie, F., Huang, H., Huang, J.: Large-scale multi-view spectral clustering via bipartite graph. In: *AAAI*, pp. 2750–2756 (2015)
9. Liang, D., Liu, D.: Systematic studies on three-way decisions with interval-valued decision-theoretic rough sets. *Inf. Sci.* **276**, 186–203 (2014)
10. Lingras, P., Yan, R.: Interval clustering using fuzzy and rough set theory. In: *Proceedings 2004 IEEE Annual Meeting, Fuzzy Information, Banff, Alberta*, pp. 780–784 (2004)
11. Lingras, P., West, C.: Interval set clustering of web users with rough k-means. *J. Intell. Inf. Syst.* **23**(1), 5–16 (2004)
12. Liu, J., Wang, C., Guo, J., Han, J.: Multi-view clustering via joint nonnegative matrix factorization. In: *Proceeding of the 2013 SIAM International Conference on Data Mining*, pp. 252–260. Society for Industrial and Applied Mathematics (2013)
13. Luxburg, U.: A tutorial on spectral clustering. *Stat. Comput.* **17**(4), 395–416 (2007)
14. Wagstaff, K., Cardie, C.: Clustering with instance-level constraints. In: *AAAI/IAAI*, vol. 1097 (2000)
15. Wang, W., Zhou, Z.H.: On multi-view active learning and the combination with semi-supervised learning. In: *Proceedings of the 25th International Conference on Machine Learning*, pp. 1152–1159. ACM (2008)
16. Wang, J., Wang, X., Tian, F., Liu, C.H., Yu, H., Liu, Y.: Adaptive multi-view semi-supervised nonnegative matrix factorization. In: Hirose, A., Ozawa, S., Doya, K., Ikeda, K., Lee, M., Liu, D. (eds.) *ICONIP 2016. LNCS*, vol. 9948, pp. 435–444. Springer, Cham (2016). doi:[10.1007/978-3-319-46672-9_49](https://doi.org/10.1007/978-3-319-46672-9_49)
17. Xia, T., Tao, D., Mei, T., Zhang, Y.: Multiview spectral embedding. *IEEE Trans. Syst. Man Cybern. Part B (Cybern.)* **40**(6), 1438–1446 (2010)
18. Xiong, S., Azimi, J., Fern, X.Z.: Active learning of constraints for semi-supervised clustering. *IEEE Trans. Knowl. Data Eng.* **26**(1), 43–54 (2014)
19. Xu, C., Tao, D., Xu, C.: A survey on multi-view learning. arXiv preprint [arXiv:1304.5634](https://arxiv.org/abs/1304.5634) (2013)

20. Yao, Y., Lingras, P., Wang, R., Miao, D.: Interval set cluster analysis: a reformulation. In: Sakai, H., Chakraborty, M.K., Hassanien, A.E., Ślęzak, D., Zhu, W. (eds.) RSFDGrC 2009. LNCS, vol. 5908, pp. 398–405. Springer, Heidelberg (2009). doi:[10.1007/978-3-642-10646-0_48](https://doi.org/10.1007/978-3-642-10646-0_48)
21. Yao, Y.: An outline of a theory of three-way decisions. In: Yao, J.T., Yang, Y., Słowiński, R., Greco, S., Li, H., Mitra, S., Polkowski, L. (eds.) RSCTC 2012. LNCS, vol. 7413, pp. 1–17. Springer, Heidelberg (2012). doi:[10.1007/978-3-642-32115-3_1](https://doi.org/10.1007/978-3-642-32115-3_1)
22. Yao, Y.: Three-way decisions and cognitive computing. *Cogn. Comput.* **8**(4), 543–554 (2016)
23. Yao, J., Azam, N.: Web-based medical decision support systems for three-way medical decision making with game-theoretic rough sets. *IEEE Trans. Fuzzy Syst.* **23**(1), 3–15 (2015)
24. Ye, X., Sakurai, T.: Robust similarity measure for spectral clustering based on shared neighbors. *ETRI J.* **38**(3), 540–550 (2016)
25. Yu, H., Wang, G.Y., Li, T.R., Liang, J.Y., Miao, D.Q., Yao, Y.Y.: *Three-Way Decisions: Methods and Practices for Complex Problem Solving*. Science Press, Beijing (2015). (in Chinese)
26. Yu, H., Zhang, C., Wang, G.: A tree-based incremental overlapping clustering method using the three-way decision theory. *Knowl.-Based Syst.* **91**, 189–203 (2016)
27. Yu, H., et al.: Methods and practices of three-way decisions for complex problem solving. In: Ciucci, D., Wang, G., Mitra, S., Wu, W.-Z. (eds.) RSKT 2015. LNCS, vol. 9436, pp. 255–265. Springer, Cham (2015). doi:[10.1007/978-3-319-25754-9_23](https://doi.org/10.1007/978-3-319-25754-9_23)
28. Zhang, X., Zhao, D.Y., Wei, S., Xiao, W.X.: Active semi-supervised clustering based on multi-view learning. In: WRI Global Congress on Intelligent Systems, GCIS 2009, vol. 3, pp. 495–499. IEEE (2009)
29. Zhou, B., Yao, Y., Luo, J.: Cost-sensitive three-way email spam filtering. *J. Intell. Inf. Syst.* **42**(1), 19–45 (2014)