International Journal of Machine Learning and Cybernetics
https://doi.org/10.1007/s13042-021-01450-1

ORIGINAL ARTICLE

Weak-label-based global and local multi-view multi-label learning with three-way clustering

Changming Zhu[1,2] · Dujuan Cao[1] · Shuaiping Guo[1] · Rigui Zhou[1] · Lai Wei[1] · YiLing Dong[1] · Duoqian Miao[2]

Received: 26 November 2020 / Accepted: 8 October 2021

Abstract

This paper develops a weak-label-based global and local multi-view multi-label learning with three-way clustering (WL-GLMVML-ATC) to solve multi-view multi-label data sets and exploit more authentic global and local label correlations of both the whole data set and each view simultaneously. Different from the traditional learning methods, WL-GLMVML-ATC pays more attention to the solutions of weak-label cases and uncertain relationships of clusters with the usage of Universum and active three-way clustering. According to Universum notion, even though the size of labeled instances is much more smaller than the unlabeled ones, the useful sample information can still be enhanced. Through the active three-way clustering strategy, the belongingness of instances to a cluster depend on the probabilities of uncertain instances belonging to core regions. This strategy brings a more authentic local label correlation since many traditional methods suppose that instances and the corresponding clusters always exhibit certain relationships such as belong-to definitely and not belong-to definitely. This hypothesis is not ubiquitous in real-world applications. According to the experiments, we can see WL-GLMVML-ATC (1) achieves a better performance, be superior to the classical multi-view learning methods and multi-label learning methods in statistical, advances the development of these learning methods in final; (2) won't add too much running time; (3) has a good convergence and ability to process multi-view multi-label data sets.

Keywords Three-way clustering · Multi-view multi-label · Label correlation · Weak label

1 Introduction

1.1 Intention and motivation

In real-world applications, multi-view, multi-label, and multi-view multi-label data sets are three ubiquitous data sets [1–5] and there are many related tasks to process these data sets are put forward. Among these tasks, two of them are of a general nature, i.e., clustering and classification and in order to process these two tasks, there are many learning methods are developed [4, 6, 7]. But it is found that although these methods performed better when they were put forward, there are some key problems need to be solved at present as below.

First, for the clustering tasks, previous related methods always assume the instances (a data set is composed of many instances) and the corresponding clusters exhibit certain relationships. Namely, an instance belongs (or not belong) to a cluster definitely. While in real-world applications, instances and the clusters might have gradual relationships. In other words, there are three types of relationships between

✉ Changming Zhu
cmzhu@shmtu.edu.cn

Dujuan Cao
202030310126@stu.shmtu.edu.cn

Shuaiping Guo
202030310102@stu.shmtu.edu.cn

Rigui Zhou
rgzhou@shmtu.edu.cn

Lai Wei
weilai@shmtu.edu.cn

YiLing Dong
yldong@shmtu.edu.cn

Duoqian Miao
dqmiao@tongji.edu.cn

1 College of Information Engineering, Shanghai Maritime University, Shanghai 201306, China

2 Computer Science and Technology, Tongji University, Shanghai 201804, China

Published online: 23 October 2021

an instance and a cluster, namely, belong-to definitely, not belong-to definitely, and uncertain. In most of the existing studies, a cluster is represented by a single set. Any set naturally divides the space into two regions. Instances belong to the cluster if they are elements of the set, otherwise they do not. Here, only two relationships are considered, no matter in hard clustering or in soft clustering. They are typically based on two-way (i.e., binary) decisions. While for the third relationship, which means the instance may or may not belong to the cluster, one cannot make decisions based on the presently obtained knowledge, i.e, two-way decisions, and finally the improvement of clustering performance is limited.

Second, for the classification tasks, previous related methods do not take or exploit the global and local label correlations of both the whole data set and each view simultaneously. Indeed, among labels of instances may exist some correlations, for example, label 'nature' and label 'rural' have a belonging relationship. If the label correlations are shared by all instances, the correlations are global. While if they are shared only by a data subset, they should be local. Even though some latest methods including global and local multi-view multi-label learning (GLMVML) [8] can exploit global and local label correlations of the whole data set and each view simultaneously, while they adopt two-way decision (for example, k-means) to exploit local label correlations which are a little inauthentic. This also limits the improvement of classification performance.

Third, as we know, with the development of the technology, trade, finance, etc., more and more data sets are generated with a rapid speed, for example, news data set or Youtube data set. While with the limitation of manpower, people cannot label all instances. Indeed, only few instances are labeled and such a data set is named weak-label data set. For the previous methods including GLMVML, they cannot process weak-label case well.

There is no doubt, these problems will disturb the processing of real-world applications and degrade the performances of corresponding clustering and classification learning methods. Moreover, since these three kinds of data sets and two kinds of tasks are representative and ubiquitous, thus researching and solving the above mentioned problems are essential and important. This is also the original intention and motivation of our work.

1.2 Tasks and targets

In order to solve the above mentioned problems, we start our research and try to finish the below tasks.

First, in real-world applications, there are three types of relationships between an instance and a cluster, namely, belong-to definitely, not belong-to definitely, and uncertain. While many present clustering methods only make decisions based on the two-way decisions (namely consider the first two relationships) and always ignore the uncertain relationship, thus our work wants to use a method to consider such an uncertain relationship.

Second, local label correlations are useful for improving classification performances. But some latest methods cannot exploit authentic local label correlations since they don't consider the uncertain relationship between instances and clusters. Thus our work aims to research that how to exploit a more authentic local label correlations.

Third, with the limitation of manpower, many data sets consist of few labeled instances and abundant unlabeled instances and such a weak-label case leads to a poor classification or clustering performance. Due to many present methods cannot solve this issue well, thus our work try to propose a method to process weak-label case.

Once we solve the above tasks, we want to realize the corresponding targets, namely, considering an uncertain relationship and processing weak-label case. With the consideration of such an uncertain relationship, we want to exploit a more authentic local label correlations. Meanwhile, with the solution of weak-label case, we want to expand the scope of application of the present learning methods and make them be feasible for the data sets consisting of few labeled instances and abundant unlabeled instances. Finally, we expect that the performance of a multi-view multi-label learning method to process clustering and classification tasks can be enhanced and we can advance the development of multi-view multi-label learning.

1.3 Proposal and solution

1.3.1 Proposal

In order to finish the above mentioned three tasks and realize the targets, we select GLMVML as the basic method and develop a weak-label-based global and local multi-view multi-label learning with three-way clustering (WL-GLMVML-ATC). Here, the reason we select GLMVML is that it encounters the same key problems mentioned as before.

1.3.2 Solution

Then, we know that for the uncertain relationship between instances and clusters, we cannot make decisions based on two-way decisions. But we can make further decisions once more information becomes available. This method is referred to as three-way decisions suggested by Yao [9] and advanced by other scholars [10–27]. Thus, we use an active three-way clustering (ATC) which is one kind of three-way decisions developed by Yu et al. [28] to produce the group partition of the data set and improve clustering performance. With the usage of ATC, we can take the uncertain relationship and each instance belongs to a cluster with a probability rather than certainly. This accords to the real scenes. Moreover, replacing k-means which is used to exploit local label correlations in GLMVML with ATC, we can exploit more authentic local label correlations. In other words, with the usage of ATC which is a three-way decision, we can finish the first and second tasks.

Furthermore, in order to solve the weak-label case and finish the third task, we adopt the notion of Universum learning and develop a method named unlabeled instance generated (UIG).

In summary, with the combination of UIG, ATC, and GLMVML, we realize WL-GLMVML-ATC, solve the above mentioned problems, and enhance the ability to process clustering and classification tasks.

1.4 Originality and contributions

1.4.1 Originality

WL-GLMVML-ATC adopts ATC to divide the data set into several clusters and each instance belongs to a cluster according to a probability. Then, on the base of GLMVML, with the usage of ATC, the belongings of instances to clusters are consistent with the real scenes and we can obtain more authentic local label correlations. Moreover, with the usage of UIG which is developed on the notion of Universum, we can process weak-label case. Although on the surface, the WL-GLMVML-ATC is the combination of the existing work Universum, ATC, and GLMVML, but in the field of multi-view multi-label learning, it is the first attempt for the combination of global and local label correlations, the three-way decisions, and Universum. The proposed method can improve the classification and clustering performances simultaneously.

1.4.2 Contributions

The contributions of WL-GLMVML-ATC are (1) it can reflect the global and local label correlations simultaneously and it can exploit more authentic local label correlations; (2) it has a better ability to process multi-view multi-label data sets, especially for weak-label case; (3) it moves forward research of multi-view multi-label learning.

1.5 Framework

The paper is structured as follows. Section 2 elaborates on the ATC approach. Section 3 describes the UIG approach. Section 4 presents the framework of the developed WL-GLMVML-ATC. Section 5 covers experimental results. Conclusions and future studies are presented in Sect. 6.

2 Active three-way clustering (ATC) [28]

In order to describe and exploit the three types of relationships between instances and clusters, we adopt ATC which is one of three-way decisions in our work. Similar with [28], before we elaborate on the ATC approach, we briefly review the basic knowledge of ATC including the three-way representation of clustering and pairwise constraints as below.

2.1 Three-way representation of clustering

Suppose the group partition of the data set $X = \{x_1, ..., x_i, ..., x_n\}$ is $C = \{C_1, ..., C_m, ..., C_g\}$ and x_i represents the ith instance. Then for each cluster in the group partition, namely, C_m, it can be represented by $C_m = (Co(C_m), Fr(C_m))$. Here, $Co(C_m) = CoreRegion(C_m) \subset X$ and $Fr(C_m) = FringeRegion(C_m) \subset X$. Let $Tr(C_m) = X - Co(C_m) - Fr(C_m) = TrivialRegion(C_m)$. Then $Co(C_m), Fr(C_m), Tr(C_m)$ naturally form the three regions related to C_m as core region, fringe region and trivial region respectively. Specifically, if $x \in CoreRegion(C_m)$, the instance x belongs to cluster C_m definitely; if $x \in FringeRegion(C_m)$, the instance x might belong to cluster C_m; if $x \in TrivialRegion(C_m)$, the instance x does not belong to cluster C_m definitely. Under these definitions, for each cluster, we have $X = Co(C_m) \bigcup Fr(C_m) \bigcup Tr(C_m)$, $Co(C_m) \bigcap Fr(C_m) = \emptyset$, $Co(C_m) \bigcap Tr(C_m) = \emptyset$, $Tr(C_m) \bigcap Fr(C_m) = \emptyset$. Then $C = \{(Co(C_1), Fr(C_1)), ..., (Co(C_g), Fr(C_g))\}$ and it can be treated as the three-way representation of clustering for X.

2.2 Pairwise constraints

According to [28] and [29], pairwise constraints offer typical prior information for semi-supervised clustering. In their work, they introduce must-link (positive association) and cannot-link (negative association) to reflect the constraint relations between the data points, i.e., instances. For the data set $X = \{x_1, ..., x_i, ..., x_n\}$, let $Y = \{y_1, ..., y_i, ..., y_n\}$ indicate the label matrix and y_i denotes the class label of x_i. Then must-link constraint requires that the two instances must belong to the same cluster, and this relation is denoted by $ML = \{(x_i, x_j) | y_i = y_j, \quad for \quad i \neq j, \quad x_i, x_j \in X, \quad y_i, y_j \in Y\}$. Cannot-link constraint requires that the two instances must belong to different clusters, and this relation is denoted by $CL = \{(x_p, x_q) | y_p \neq y_q, \quad for \quad p \neq q, \quad x_p, x_q \in X, \quad y_p, y_q \in Y\}$. For instances $x_i, x_j, x_g \in X$, [30] said that must-link constraint shows the following transitivity properties on instances.

$$
\begin{aligned}
&(x_i, x_j) \in ML \quad \& \quad (x_j, x_g) \in ML \\
&\Rightarrow (x_i, x_g) \in ML \\
&(x_i, x_j) \in ML \quad \& \quad (x_j, x_g) \in CL \\
&\Rightarrow (x_i, x_g) \in CL
\end{aligned} \tag{1}
$$

Then in [28], we set a matrix $R \in \mathbb{R}^{n \times n}$ to store the constraint pairs and the update of R is given below. First, R is initialized as $\varnothing$. Then, when a pair of instances are must-link constraint relation, namely $(x_i, x_j) \in ML$, the corresponding value of element in R is updated to 1; when a pair of instances are cannot-link constraint relation, namely $(x_i, x_j) \in CL$, the corresponding value of element in R is updated to 0. At the end of each iteration, we update R according to the response of expert and the transitivity properties of Eq. (1).

Next, one updates the consensus similarity matrix $W^{\star}$ with Eq. (2). Here, $W^{\star} = (Z^{\star} + (Z^{\star})^T)/2$ where $Z^{\star}$ is a consensus low-rank matrix which is derived from X. The method to get $Z^{\star}$ can refer to [28].

$$
\begin{aligned}
&if \quad (x_i, x_j) \in ML, \quad then \quad w_{ij} = w_{ji} = 1 \\
&if \quad (x_i, x_j) \in CL, \quad then \quad w_{ij} = w_{ji} = 0
\end{aligned} \tag{2}
$$

2.3 Description of algorithm ATC

ATC is an iteration processing and it consists of the below five procedures, namely, the spectral clustering processing, initial core regions construction, to extend core regions and construct fringe regions, to select the most informative instance from fringe regions and to construct pairwise query. According to the below procedures, we know that the three-way clustering results are constructed by two steps. First, we produce a preliminary result. Second, we extend core regions and construct fringe regions.

2.3.1 Spectral clustering processing

Spectral clustering processing is a kind of clustering and it aims to produce a preliminary clustering result and optimize the result by iterations. In our work, we refer to [28] and adopt the same spectral clustering algorithm so as to produce a preliminary clustering result of X. The details of this algorithm can be found in [28]. Here, we should notice that the reason we adopt spectral clustering processing is just to produce a preliminary clustering result. This operation can be replaced by other existing clustering approaches. Due to we will use the below four steps to optimize the clustering result, thus we needn't to pay too much attention to the choice of the clustering algorithm in this step.

2.3.2 Initialize core regions construction

According to the preliminary clustering result of X, we initialize core regions construction. The method to realize the initialization is farthest-first traversal scheme [31] and this scheme aims to select the core instances. Core instances indicate the ones which locate on the fringe of a cluster and contain more information than instances in the center of a cluster. The basic idea of farthest-first traversal of a set of instances is to find K instances such that they are far from each other. Details are given below.

First, we let the original X be the *CandidateSet* and l count the number of constructed core regions. Initial value of l is 1. Second, for each cluster $(Co(C_m), Fr(C_m))$, it is initialized as $\varnothing$, namely, $Co(C_m) = \varnothing$ and $Fr(C_m) = \varnothing$. Third, we select a starting instance x from X at random and put x into the $Co(C_1)$. Fourth, we choose the next instance to be farthest from the untraversed set *CandidateSet* by Eq. (3). Being specific, *AllCo* is the set of all core instances, namely, $AllCo = \bigcup_{m=1}^{g} Co(C_m)$. According to min-max criterion [32], the distance between x and $AllCo$ is $d(x, AllCo) = \min_{y \in AllCo} ||x - y||$. Then, the farthest one is determined as follows.

$$
\begin{aligned}
&x \leftarrow arg \max_{x \in CandidateSet} d(x, AllCo) = \\
&arg \max_{x \in CandidateSet} (\min_{y \in AllCo} ||x - y||)
\end{aligned} \tag{3}
$$

After that, we should decide whether x and an instance $x_i \in Co(C_m)$ $(1 \le m \le g)$ are in the same cluster. We make pair-wise queries through the form as: do instances x and x_i belong to the same cluster? If the ML constraint is satisfied, x is assigned to $Co(C_m)$ and it should be removed from *CandidateSet*. If no one ML constraint is satisfied after traversing all core regions, a new core region $Co(C_{g+1})$ is constructed and assign x to the new core region $Co(C_{g+1})$. With the above procedure, we can divide the *CandidateSet* into g clusters.

2.3.3 Extend core regions and construct fringe regions

Once we execute the spectral clustering processing and initialize the core regions construction as long as it is not in the first iteration, we extend these core regions and construct fringe regions. Concretely speaking, set $N(x)$ be a set of q neighbor instances of x and $N(x)$ can be built by the last iteration result. Then we extend the core regions by observing the relationship between the x (which is an unlabeled instance) and x_i (which is a labeled random instance from $Co(C_m)$). Namely, if x is the neighbor of x_i and x_i is also the neighbor of x, then we can say that x is much similar with x_i and they should both belong to the core region. If x is the neighbor of x_i, but x_i is not the neighbor of x, we say they are not similar and x should belong to the fringe region. Otherwise, if x is not the neighbor of x_i, we think it is trivial to the cluster. So, according to this, we have the following three-way decision rules.

$$\begin{aligned} & if \quad (x \in N(x_i)) \wedge (x_i \in N(x)), \\ & \quad then \quad Co(C_m) = Co(C_m) \cup \{x\} \\ & if \quad (x \in N(x_i)) \wedge (x_i \notin N(x)), \\ & \quad then \quad Fr(C_m) = Fr(C_m) \cup \{x\} \end{aligned} \tag{4}$$

2.3.4 Select the most informative instance $x^\star$ from fringe regions

Then we adopt the active learning strategy to improve the performance of clustering. The objective of this strategy is to select the most informative instance $x^\star$ from fringe regions. Concretely speaking, we measure the uncertainty of instance based on the similarity firstly. Namely, on the base of consensus similarity matrix $W^\star$, let $w_{\cdot j}$ denote the similarity between x and x_j and $U = \bigcup_{m=1}^{g} Fr(C_m)$ denote all uncertain instances currently, then we adopt the following formula to estimate the probability of an uncertain instance x belonging to a core region $Co(C_m)$ where $\left|Co(C_m)\right|$ is the number of instances in the core region $Co(C_m)$.

$$P(x \in Co(C_m)) = \frac{\frac{1}{\left|Co(C_m)\right|} \sum_{x_j \in Co(C_m)} w_{\cdot j}}{\sum_{s=1}^{g} \frac{1}{\left|Co(C_s)\right|} \sum_{x_j \in Co(C_s)} w_{\cdot j}} \tag{5}$$

Second, we use Eq. (6) to measure the uncertainty of an instance by the entropy.

$$H(x) = -\frac{1}{g} \sum_{m=1}^{g} (P(x \in Co(C_m)) log_2 P(x \in Co(C_m))) \tag{6}$$

where $x \in U$.

Then, the most informative instance $x^\star$ is selected by Eq. (7).

$$x^\star = arg \max_{x \in U} H(x) \tag{7}$$

2.3.5 Construct pairwise query

Finally, we construct pairwise query with the following method. First, we sort the clusters by $P(x^\star \in Co(C_m))$ in descending order where $1 \leq m \leq g$. Second, for each cluster, we select one instance x_i from $Co(C_m)$ in random and query the constraint relationship between $x^\star$ and x_i. If $(x^\star, x_i) \in ML$, then $Co(C_m) = Co(C_m) \cup \{x^\star\}$. At last, we adopt Eq. (1) to update the matrix R and Eq. (2) to update the matrix $W^\star$.

With iteration processing about the above five steps, we can divide X into g clusters, namely, $C = \{(Co(C_1), Fr(C_1)), ..., (Co(C_g), Fr(C_g))\}$.

3 Unlabeled instance generated (UIG)

In real-world applications, many data sets are generated with a rapid speed. While with the limitation of manpower, only few instances can be labeled and this leads to a weak-label case. As we know, compared with unlabeled instances, labeled ones can bring more useful sample information. So weak-label case will bring a poor performance for learning methods. In order to overcome such a case, scholars develop a series of Universum learning methods. In terms of Universum, it was proposed by Vapnik and Kotz firstly [33] and have been developed by other scholars [34–37]. Universum aims to encode prior knowledge by given instances. With Universum learning, one collects some instances which do not belong to any class of data, but do belong to the same domain as the problem. These collections are named Universum instances which reflect some prior knowledge. By Universum, we can obtain a robust decision hyperplane.

Due to Universum learning can generate Universum instances which provide some useful sample information, thus on the base of the notion of Universum, we propose a method named unlabeled instance generated (UIG) to add more useful sample information. The processing is given below.

(1) In each clsuter, we set a parameter n_m (n_m should smaller than the number of labeled instances in this cluster). For each labeled instance x_i in this cluster, we get the n_m labeled instances which locate nearest from this labeled instance.
(2) According to those n_m nearest labeled instances, we compute the information entropy of x_i.
(3) Select K ($K \leq n_m$) labeled instances which information entropies are smallest and for each selected labeled instance x_j ($j \in [1, K]$), we get the p unlabeled instances in this cluster which locate nearest or farthest from x_j.
(4) In terms of each unlabeled instance x_{ul} ($l \in [1, p]$) and corresponidng labeled instance x_j, we compute the midpoint of them to construct an Universum instance $x^\star_{jl}$ ($j \in [1, K]$ and $l \in [1, p]$).
(5) Collect all generated Universum instances to create the corresponding Universum set $X^\star$ so as to enhance the useful sample information.

Here, we should notice that the Universum instances are not labeled, but since they are generated by labeled and unlabeled instances and the labeled ones have small information

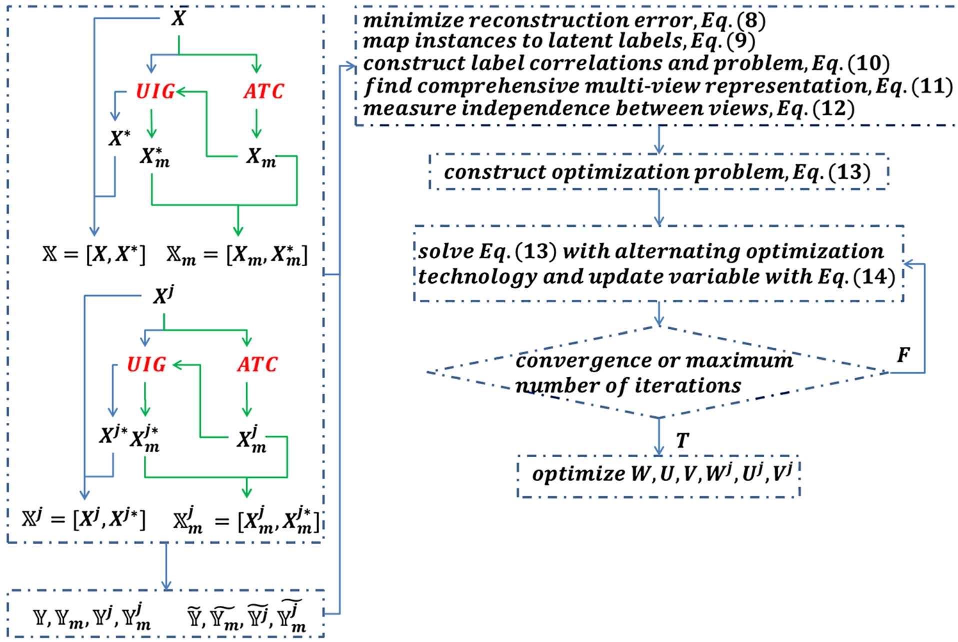

Fig. 1 The main block diagram of WL-GLMVML-ATC

entropies and high certainties, thus those Universum instances can possess more useful sample information.

4 Weak-label-based global and local multi-view multi-label learning with active three-way clustering

On the base of ATC, UIG, and GLMVML, we develop WL-GLMVML-ATC as below and its main block diagram is given in Fig. 1.

4.1 Data preparation

Suppose $X = \{x_1, ..., x_i, ..., x_n\} = \begin{pmatrix} X^1 \\ \vdots \\ X^j \\ \vdots \\ X^v \end{pmatrix} \in \mathbb{R}^{d \times n}$ is a data set with v views and n instances where $i \in [1, n]$ and $j \in [1, v]$. For jth view, its data matrix $X^j \in \mathbb{R}^{d_j \times n}$ consists of information from n instances, namely, $X^j = \{x_1^j, ..., x_i^j, ..., x_n^j\}$ and $x_i^j = \{x_{i1}^j, ..., x_{ip}^j, ..., x_{id_j}^j\}^T \in \mathbb{R}^{d_j \times 1}$ represents the information of jth view for ith instance. pth feature of x_i^j is x_{ip}^j where $p \in [1, d_j]$ and d_j is the feature dimension of jth view. For ith instance, i.e, $x_i \in \mathbb{R}^{d \times 1}$ can be represented by $x_i = \{x_i^{1^T}, ..., x_i^{j^T}, ..., x_i^{v^T}\}^T$ where $d = \sum_{j=1}^{v} d_j$.

Furthermore, X is also a multi-label data set and in different views, an instance always possesses different labels, thus suppose $y_i^j \in \mathbb{R}^{l_j \times 1}$ is a label vector of ith instance in the jth view and each component of y_i^j indicates the label of x_i^j for the corresponding class. l_j represents that at jth view, instances have l_j classes. If the rth component of y_i^j, namely, $y_{ir}^j = 1$, it means x_i^j belongs to rth class definitely. If $y_{ir}^j = -1$, this indicates that x_i^j does not belong to rth class definitely. If $y_{ir}^j = 0$, this means whether x_i^j belongs to rth class or not is not available. Then $y_i = \{y_i^{1^T}, ..., y_i^{j^T}, ..., y_i^{v^T}\}^T$ represents the label of ith instance, $Y^j = \{y_1^j, ..., y_i^j, ..., y_n^j\} \in \mathbb{R}^{l_j \times n}$ represents the label matrix of jth view, and we let $Y = \begin{pmatrix} Y^1 \\ \vdots \\ Y^j \\ \vdots \\ Y^v \end{pmatrix} \in \mathbb{R}^{l \times n}$ indicates the label matrix for X where $l = \sum_{j=1}^{v} l_j$. Furthermore, since in many cases, instances have two kinds of labels, one is predicted labels and the other is real labels. So here, we let Y, Y^j, y_i^j, y_i represent the predicted ones and $\widetilde{Y}, \widetilde{Y^j}, \widetilde{y_i^j}, \widetilde{y_i}$ represent the real ones. Definitions of $\widetilde{\star}$ is similar with the $\star$.

4.2 Processing of WL-GLMVML-ATC

4.2.1 Preprocessing with UIG and ATC

For X, we adopt UIG to generate some Universum instances $X^\star$ which provide useful sample information. Then we carry out ATC on X and produce its group partition, i.e., $C = \{(Co(C_1), Fr(C_1)), ..., (Co(C_g), Fr(C_g))\}$ which represents g clusters. Then we treat instances in each cluster C_m as X_m and still adopt UIG to generate some Universum instances $X_m^\star$ where $m \in [1, g]$.

Similarly, for each view X^j, we also adopt UIG to generate Universum instances $X^{j\star}$. Then, we also adopt ATC to divide X^j into g^j clusters, namely, $C^j = \{(Co(C_1^j), Fr(C_1^j)), ..., (Co(C_{g^j}^j), Fr(C_{g^j}^j))\}$ and for each C_m^j, on the base of its instances X_m^j, we can also generate Universum instances $X_m^{j\star}$ where $m \in [1, g^j]$.

Here, dimensions of $X^\star$, $X^{j\star}$, X_m, $X_m^\star$, X_m^j, $X_m^{j\star}$ are $d \times n^\star$, $d_j \times n^{j\star}$, $d \times n_m$, $d \times n_m^\star$, $d_j \times n_m^j$, $d_j \times n_m^{j\star}$, respectively where $n^\star$, $n^{j\star}$, n_m, $n_m^\star$, n_m^j, $n_m^{j\star}$ are the corresponding numbers of instances in $X^\star$, $X^{j\star}$, X_m, $X_m^\star$, X_m^j, $X_m^{j\star}$.

After that, for convenience, we let $\mathbb{X} = [X, X^\star] \in \mathbb{R}^{d\times(n+n^\star)}$, $\mathbb{X}^j = [X^j, X^{j\star}] \in \mathbb{R}^{d_j\times(n^j+n^{j\star})}$, $\mathbb{X}_m = [X_m, X_m^\star] \in \mathbb{R}^{d\times(n_m+n_m^\star)}$ and $\mathbb{X}_m^j = [X_m^j, X_m^{j\star}] \in \mathbb{R}^{d_j\times(n_m^j+n_m^{j\star})}$.

Finally, according to the processing of X and X^js, Y and Y^j can also be updated as $\mathbb{Y} \in \mathbb{R}^{l\times(n+n^\star)}$, $\mathbb{Y}^j \in \mathbb{R}^{l_j\times(n^j+n^{j\star})}$, $\mathbb{Y}_m \in \mathbb{R}^{l\times(n_m+n_m^\star)}$, $\mathbb{Y}_m^j \in \mathbb{R}^{l_j\times(n_m^j+n_m^{j\star})}$. At the same time, in terms of the $\widetilde{Y}$ and $\widetilde{Y}^j$, their processing are same, and we treat the processing results as $\widetilde{\mathbb{Y}}$, $\widetilde{\mathbb{Y}}^j$, $\widetilde{\mathbb{Y}}_m$, $\widetilde{\mathbb{Y}}_m^j$. The dimensions of them are same as the ones of $\mathbb{Y}$, $\mathbb{Y}^j$, $\mathbb{Y}_m$, $\mathbb{Y}_m^j$, respectively. Here, we should notice that since Universum instances are not labeled, thus their corresponding labels are set to be 0.

4.2.2 Construction of the optimization problem

On the base of the above preprocessing results, we construct the optimization problem of WL-GLMVML-ATC. Since [8] had described the details of GLMVML and the significant difference between WL-GLMVML-ATC and GLMVML is that in order to introduce the local label correlation, WL-GLMVML-ATC divides the data set into several groups with ATC rather than k-means, thus we describe the way to construct the optimization problem directly as below and other information can refer to [8].

1. If $\widetilde{\mathbb{Y}}$ and $\widetilde{\mathbb{Y}}^j$ are low-rank and we get $rank(\widetilde{\mathbb{Y}}) = k$ and $rank(\widetilde{\mathbb{Y}}^j) = k_j$, then we can rewrite them as $\widetilde{\mathbb{Y}} = UV$ and $\widetilde{\mathbb{Y}}^j = U^jV^j$ where $U \in \mathbb{R}^{l\times k}$ or $U^j \in \mathbb{R}^{l_j\times k_j}$ has a function to project the original labels to the latent label space while $V \in \mathbb{R}^{k\times(n+n^\star)}$ or $V^j \in \mathbb{R}^{k_j\times(n^j+n^{j\star})}$ can be treated as the latent labels that are more compact and more semantically abstract than the original labels. Due to in real-world applications, weak-label cases are universal, thus we want to minimize the reconstruction error on the predicted labels, i.e.,
$$\min_{U,V,U^j,V^j} ||\amalg_\Omega(\mathbb{Y} - UV)||_F^2 + \sum_{j=1}^{v} \left|\left|\amalg_{\Omega^j}(\mathbb{Y}^j - U^jV^j)\right|\right|_F^2 \tag{8}$$
where $||\star||_F^2$ represents the square of Frobenius norm for $\star$, Ω (Ω^j) consists of indices of the observed labels in $\mathbb{Y}$ ($\mathbb{Y}^j$). Then if $(i,j) \in \Omega$, we have $[||\amalg_\Omega(A)||]_{ij} = A_{ij}$, otherwise, $[||\amalg_\Omega(A)||]_{ij} = 0$ (similar to Ω^j case).
2. We should adopt a linear mapping $W \in \mathbb{R}^{d\times k}$ ($W^j \in \mathbb{R}^{d_j\times k_j}$) to map instances to the latent labels and we use the following problem to learn the W (W^j).
$$\min_{W,V,W^j,V^j} \left|\left|V - W^T\mathbb{X}\right|\right|_F^2 + \sum_{j=1}^{v} \left|\left|V^j - W^{j^T}\mathbb{X}^j\right|\right|_F^2 \tag{9}$$
3. We should introduce the global and local label correlations and construct the corresponding problem to enhance the performances of methods. Thus, for $\mathbb{X}$, the prediction on instance x_i is $sign(f(x_i))$ where $f(x_i) = UW^Tx_i \in \mathbb{R}^{l\times 1}$, so we let $F_0 = UW^T\mathbb{X} \in \mathbb{R}^{l\times(n+n^\star)}$ represents the output matrix of $\mathbb{X}$. Similar, $F_0^j = U^jW^{j^T}\mathbb{X}^j \in \mathbb{R}^{l_j\times(n^j+n^{j\star})}$, $F_m = UW^T\mathbb{X}_m \in \mathbb{R}^{l\times(n_m+n_m^\star)}$, $F_m^j = U^jW^{j^T}\mathbb{X}_m^j \in \mathbb{R}^{l_j\times(n_m^j+n_m^{j\star})}$ represent the output matrices of $\mathbb{X}^j$, $\mathbb{X}_m$, $\mathbb{X}_m^j$.

   After that, we compute the label correlation matrices on the base of $\mathbb{X}$, $\mathbb{X}^j$, $\mathbb{X}_m$, $\mathbb{X}_m^j$, $\mathbb{Y}$, $\mathbb{Y}^j$, $\mathbb{Y}_m$, $\mathbb{Y}_m^j$. Take $\mathbb{X}$ for example, we let $S_0 = \{[S_0]_{pq}\}$ denote the global label correlation matrix and its Laplacian matrix is L_0. Here, $[S_0]_{pq} = \frac{y_{p,:}y_{q,:}^T}{||y_{p,:}||\,||y_{q,:}||}$ represents the global label correlation of pth label w.r.t. qth label and $y_{p,:}$ ($y_{q,:}$) is the pth (qth) row of $\mathbb{Y}$. Similar, $S_m = \{[S_m]_{pq}\}$, $S_0^j = \{[S_0^j]_{pq}\}$, $S_m^j = \{[S_m^j]_{pq}\}$ are the corresponding local label correlation matrix of $\mathbb{X}_m$, global label correlation matrix of $\mathbb{X}^j$, local label correlation matrix of $\mathbb{X}_m^j$, respectively. Their corresponding Laplacian matrices are L_m, L_0^j, L_m^j, respectively. The dimensions of S_0, S_m, S_0^j, S_m^j, L_0, L_m, L_0^j, L_m^j are $l\times l$, $l\times l$, $l_j\times l_j$, $l_j\times l_j$, $l\times l$, $l\times l$, $l_j\times l_j$, $l_j\times l_j$, respectively.

   According to the above definitions, we construct the corresponding problem as below on the base of these correlation matrices where $tr(A)$ denotes the trace of the matrix A.
$$\min tr(F_0^TL_0F_0) + \sum_{m=1}^{g} tr(F_m^TL_mF_m) + \sum_{j=1}^{v}\left(tr({F_0^j}^TL_0^jF_0^j) + \sum_{m=1}^{g^j} tr({F_m^j}^TL_m^jF_m^j)\right) \tag{10}$$
4. We should introduce a consensus multi-view representation to encode the complementary information from different views. In simple speaking, we use Eq. (11) to find a comprehensive multi-view representation and Eq. (12) to measure the independence between different views where HSIC is a Hilbert-Schmidt independence criterion estimator [38], P represents a consensus multi-view

representation, B^j is the basic matrix corresponding to jth view.

$$\sum_{j=1}^{v} \left|\left|\mathbb{X}^j - B^j P\right|\right|_F^2 \tag{11}$$

$$\sum_{j \neq t} -HSIC(B^j, B^t) \tag{12}$$

According to the above contents, the optimization problem of WL-GLMVML-ATC is given below.

$$\begin{aligned}
&\min_{\triangle} ||\amalg_{\Omega}(\mathbb{Y} - UV)||_F^2 + \lambda_0 \left|\left|V - W^T \mathbb{X}\right|\right|_F^2 \\
&+ \lambda_1 \mathfrak{R}(U, V, W, U^j, V^j, W^j, P, B^j) \\
&+ \sum_{j=1}^{v} (\lambda_2 \left|\left|\amalg_{\Omega^j}(\mathbb{Y}^j - U^j V^j)\right|\right|_F^2 + \lambda_3 \left|\left|V^j - W^{j^T} \mathbb{X}^j\right|\right|_F^2) \\
&+ \lambda_4 tr(F_0^T L_0 F_0) + \lambda_5 \sum_{m=1}^{g} tr(F_m^T L_m F_m) \\
&+ \sum_{j=1}^{v} (\lambda_6^j tr(F_0^{j^T} L_0^j F_0^j) + \lambda_7^j \sum_{m=1}^{g^j} tr(F_m^{j^T} L_m^j F_m^j)) \\
&+ \lambda_8 \sum_{j=1}^{v} \left|\left|\mathbb{X}^j - B^j P\right|\right|_F^2 + \lambda_9 \sum_{j \neq t} -HSIC(B^j, B^t)
\end{aligned} \tag{13}$$

where $\triangle = \{U, V, W, U^j, V^j, W^j, P, B^j\}$, λs (λ^js) are trade-off parameters and $\mathfrak{R}(U, V, W, U^j, V^j, W^j, P, B^j) = ||U||_F^2 + ||V||_F^2 + ||W||_F^2 + ||U^j||_F^2 + ||V^j||_F^2 + ||W^j||_F^2 + ||P||_F^2 + ||B^j||_F^2$.

4.2.3 Solution of the optimization problem

In order to solve the optimization problem, i.e., Eq. (13), we refer to [8] and also adopt alternating optimization technology. According to this technology, in each iteration, we update an variable ϑ in $\triangle$ with gradient descent and leave the others fixed. The update formula is

$$\vartheta := \vartheta - \eta \nabla_{\vartheta} \tag{14}$$

where η is the step size. Through the continuous updating until convergence or maximum number of iterations, we can get the optimal W, U, V and use $W^T UV$ to predict the label matrix of $\mathbb{X}$. Similarly, for $\mathbb{X}^j$, $\mathbb{X}_m$, $\mathbb{X}_m^j$, we can also predict their corresponding label matrices with the optimized results including W^j, U^j, V^j.

4.2.4 Computational complexity

According to [28] and [8], the computational complexity of ATC is $O(ng^2) + O(Frng) + max(O(Fr), O(g))$ where Fr is the number of instances in all fringe regions and the computational complexity of GLMVML is $O(Gn^2)$ where G is a constant. In terms of UIG, its computational complexity is $O(n^2)$. Then in general, the total computational complexity of WL-GLMVML-ATC is still smaller than $O(n^3)$.

Table 1 Detailed information of multi-view data sets

Data set	No. instances	No. label	No. views
Mfeat	2000	10	6
Reuters	111740	6	5
Corel	1000	10	4
VOC	9963	20	2
MIR	23691	38	2
3Source	169	3	3

5 Experiments

5.1 Experimental setting

5.1.1 Data set

We adopt three kinds of data sets for experiments so as to validate the superiority of WL-GLMVML-ATC.

First are 6 multi-view data sets, i.e., Mfeat, [1] Reuters, [2] Corel[3], Pascal VOC 2007 (VOC), [4] MIR-Flickr (MIR), [5] 3Source[6] and their brief descriptions are given in Table 1.

Second are 29 multi-label data sets which are also adopted in [5, 39, 40]. Table 2 shows information of them and *label/instance* represents the average number of labels possessed by each instance.

Third is a classical multi-view multi-label data set, NUS-WIDE [41, 42]. This data set include 810 images (instances). Each instance can be represented with 6 views and 81 labels.

5.1.2 Compared method

In order to process the used data sets, some learning methods are adopted. Concretely speaking, MVMLSS [43], LMSC [44], MLDL [45] are adopted for processing multi-view data sets while LF-LPLC [46], MLCHE [47], GLOCAL [5] are used to process multi-label data sets. In terms of

[1] http://archive.ics.uci.edu/ml/datasets/Multiple+Features

[2] http://archive.ics.uci.edu/ml/datasets/Reuters+RCV1+RCV2+Multilingual%2C+Multiview+Text+Categorization+Test+collection

[3] http://archive.ics.uci.edu/ml/datasets/Corel+Image+Features

[4] http://host.robots.ox.ac.uk/pascal/VOC/

[5] http://press.liacs.nl/mirflickr/

[6] http://mlg.ucd.ie/datasets/3sources.html

Table 2 Detailed information of multi-label data sets

Data set	No. instances	No. features	No. labels	label/instance
Arts	5000	462	26	1.64
Business	5000	438	30	1.59
Computers	5000	681	33	1.51
Education	5000	550	33	1.46
Entertainment	5000	640	21	1.42
Health	5000	612	32	1.66
Recreation	5000	606	22	1.42
Reference	5000	793	33	1.17
Science	5000	743	40	1.45
Social	5000	1047	39	1.28
Society	5000	636	27	1.69
Enron	1702	1001	53	3.37
Corel5K	5000	499	374	3.52
Image	2000	294	5	1.24
Medical	978	1449	45	1.25
Language Log	1459	1004	75	1.18
RCV1V2 (subset1)	6000	944	101	2.88
RCV1V2 (subset2)	6000	944	101	2.63
Bibtex	7395	1836	159	2.4
Delicious	16105	500	983	19.02
Eur-Lex (Sm)	19348	5000	201	2.21
Bookmark	87856	2150	208	2.03
Nuswide	269468	500	81	1.87
TMC2007-5000	28596	500	22	2.16
Stackex-chemistry	6961	540	175	2.11
Stackex-chess	1675	585	227	2.41
Stackex-cooking	10491	577	400	2.23
Stackex-Cs	9270	635	274	2.56
Stackex-philosophy	3971	842	233	2.27

the processing of NUS-WIDE, we select GLMVML [8], MVMLP [42], SSDR-MML [48] for experiments.

5.1.3 Parameter setting and results acquisition

How to set the parameters of the compared methods can be found in the respective references and in terms of WL-GLMVML-ATC, the settings are similar with the ones of GLMVML [8] and ATC [28]. In terms of UIG, the $n_m = 0.3 \times n_{lm}$, $K = 0.3 \times n_{lm}$, $p = 0.7 \times n_{um}$ where n_{lm} and n_{um} represent the number of labeled and unlabeled instances in the mth cluster.

In order to get the optimal results and according to the compared methods' demands, for each data set, we randomly select $\{10\%, 20\%, 30\%, 40\%, 50\%, 60\%\}$ for training and the rest for test. Then for multi-label data sets and NUS-WIDE, in terms of each instance or each view, we randomly remove $\{10\%, 15\%, 20\%, 25\%, 30\%\}$ labels so as to get the observed label matrices. Then we repeat the experiments with each parameter combination for ten times and get the average results and the corresponding standard deviations. The best parameters are the ones whose average precision is the best. Then, the other performance indexes including the AUC (area under the receiver operating characteristic curve), running time, convergence, etc. are given with the optimal parameters. Here, we should notice that for each data set, different methods should process same data.

5.1.4 Experimental environment

All computations are performed on a node of compute cluster with 32 CPUs (Intel Core Due 3.0GHz) running RedHat Linux Enterprise 5 with 48GB main memory. The coding environment is python 3.0.

5.2 Experimental results

5.2.1 AUC and precision

We adopt AUC and precision to show the effectiveness of WL-GLMVML-ATC for the classification tasks. In general, a higher AUC or a higher precision brings a better classification performance. Tables 3 and 4 give the average AUC and precision respectively for the test sets about each data set. In these tables, •/∘ indicates that WL-GLMVML-ATC is significantly better/worse than the corresponding method on a data set (pairwise t-tests at 95% significance level). The best average AUC or precision for each data set is shown in bold. / represents no result since the related method cannot process the corresponding data set. From these tables, it is found that in most cases, WL-GLMVML-ATC has a better performance and according to the win/tie/loss counts, the proposed WL-GLMVML-ATC is clearly superior to other compared learning methods, as it wins for most times and less loses.

5.2.2 Friedman–Nemenyi test

Besides pairwise t-test, Friedman–Nemenyi statistical test is another method to check if the differences between our WL-GLMVML-ATC and other compared methods are significant [49]. Different from pairwise t-test which aims to analyze if the differences between two compared methods on a data set are significant or not, Friedman test is used to analyze if the differences between all compared methods on multiple data sets are significant or not while Nemenyi test is used to analyze if the differences between two compared methods on multiple data sets are significant or not. Thus, we carry out Friedman–Nemenyi test as below and details of the principle for Friedman–Nemenyi test can be found in [49].

For the convenience of the analysis of Friedman–Nemenyi test results, we use Tables 5 and 6 to show the average ranks of WL-GLMVML-ATC and the compared ones on all data sets according to precision and AUC. Due to there are three kinds of compared methods, thus ranks in Tables 5 and 6 are derived from three cases. Case 1 gives the average ranks of WL-GLMVML-ATC, MVMLSS, LMSC, MLDL on 6 multi-view data sets. Case 2 shows the results of WL-GLMVML-ATC, LF-LPLC, MLCHE, GLOCAL on 29 multi-label data sets. Case 3 shows the results of WL-GLMVML-ATC, GLMVML, MVMLP, SSDR-MML on all used data sets. For each case, rank differences between our method and other compared ones are also given. Then according to Tables 5 and 6, we give the related statistical values about Friedman–Nemenyi test with Tables 7 and 8.

Then according to the results of Tables 5, 6, 7, 8 and refer to [49], we carry out Friedman test firstly. Take case 1 in Table 6 for example, since we adopt 6 data sets and 4 methods for experiments in this case, thus $N = 6$ and $k = 4$. Then Friedman statistic $\chi_F^2 = \frac{12\times N}{k(k+1)}[1.17^2 + 3.17^2 + 3.5^2 + 2.17^2 - \frac{k(k+1)^2}{4}] = 12.16$ and $F_F = \frac{(N-1)\chi_F^2}{N(k-1)-\chi_F^2} = 10.40$, further, $F_{0.05}(k-1,(k-1)(N-1)) = F_{0.05}(3, 15) = 3.29$ and $F_{0.10}(k-1,(k-1)(N-1)) = F_{0.10}(3, 15) = 2.49$. Since $F_F > F_{0.05}(3, 15)$ and $F_F > F_{0.10}(3, 15)$, so we can reject the null-hypothesis and say the differences between all compared methods on multiple data sets are significant in this case. For case 2 and case 3 about precision and ones about AUC, we draw a same conclusion.

After that, we use Nemenyi test for pairwise comparisons. According to the principles of Nemenyi test, if rank difference between two compared methods is larger than CD_α ($\alpha = 0.05 \quad or \quad 0.10$), then their differences on multiple data sets are significant. Thus, we still take case 1 in Table 6 for example. Since $N = 6$ and $k = 4$, thus critical value at $q_{0.05}$ is 2.57 and corresponding critical difference (CD) is $CD_{0.05} = q_{0.05}\sqrt{\frac{k\cdot(k+1)}{6\cdot N}} = 1.91$ while the one at $q_{0.10}$ is 2.29 and corresponding CD is $CD_{0.10} = q_{0.10}\sqrt{\frac{k\cdot(k+1)}{6\cdot N}} = 1.71$. Since no matter under the case of $CD_{0.05}$ and under the case of $CD_{0.10}$, rank difference between WL-GLMVML-ATC and MVMLSS (LMSC) is larger than $CD_{0.05}$ and $CD_{0.10}$, so we can say on this case, the performance of WL-GLMVML-ATC is better than the one of MVMLSS (LMSC) and their difference is significant. For case 2 and case 3 in Table 6, WL-GLMVML-ATC is significant better than LF-LPLC, MLCHE, GLOCAL, MVMLP, SSDR-MML, GLMVML since rank differences between our WL-GLMVML-ATC and the compared methods are larger than corresponding $CD_{0.05}$ and $CD_{0.10}$. In the same way, in terms of the statistical values in Table 7, expect for the methods MVMLSS and MLDL under the case of $CD_{0.05}$, WL-GLMVML-ATC is significant better than most compared methods under the cases of $CD_{0.05}$ and $CD_{0.10}$.

In generally, we can validate the effectiveness of WL-GLMVML-ATC from an average view. Specially, for multi-view case, the significance is not very obvious, while for multi-label case and multi-view multi-label case, our method performs significant best in statistics.

5.2.3 Running time

The computational complexity of WL-GLMVML-ATC is always smaller than $O(n^3)$ which is the computational complexity of many traditional methods, it is still larger than some linear learning methods. Thus, we show the running time of these compared methods and observe the difference. Table 9 shows the related experimental results and Avg. (mv) represents the average running time for multi-view data sets while Avg. (ml) represents the average running time for multi-label data sets. From this table, we find that our proposed WL-GLMVML-ATC costs a little more running time which is also accepted by us.

Table 3 Average AUC (mean ± std.) of WL-GLMVML-ATC and compared methods for test instances

Data sets	Ours	LMSC	MVMLSS	MLDL	LF-LPLC	MLCHE	GLOCAL	MVMLP	SSDR-MML	GLMVML
Mfeat	**0.783 ± 0.015**	0.736 ± 0.020 •	0.718 ± 0.018 •	0.668 ± 0.005 •	/	/	/	0.758 ± 0.017 •	0.704 ± 0.017 •	0.731 ± 0.017 •
Reuters	**0.936 ± 0.015**	0.785 ± 0.011 •	0.789 ± 0.020 •	0.915 ± 0.028 •	/	/	/	0.926 ± 0.019 •	0.818 ± 0.020 •	0.900 ± 0.019 •
Corel	**0.829 ± 0.001**	0.772 ± 0.013 •	0.820 ± 0.015 •	0.779 ± 0.014 •	/	/	/	0.823 ± 0.014	0.768 ± 0.014 •	0.801 ± 0.014 •
VOC	**0.729 ± 0.009**	0.687 ± 0.016 •	0.702 ± 0.035 •	0.686 ± 0.035 •	/	/	/	0.701 ± 0.003 •	0.626 ± 0.001 •	0.666 ± 0.049 •
MIR	**0.522 ± 0.017**	0.488 ± 0.008 •	0.516 ± 0.036 •	0.520 ± 0.046	/	/	/	0.512 ± 0.037 •	0.514 ± 0.025 •	0.475 ± 0.002 •
3Source	**0.731 ± 0.001**	0.711 ± 0.037 •	0.693 ± 0.045 •	0.725 ± 0.014	/	/	/	0.647 ± 0.003 •	0.704 ± 0.029 •	0.688 ± 0.023 •
Arts	**0.890 ± 0.007**	/	/	/	0.820 ± 0.005 •	0.782 ± 0.005 •	0.887 ± 0.005	0.788 ± 0.005 •	0.814 ± 0.005 •	0.771 ± 0.005 •
Business	0.958 ± 0.003	/	/	/	0.926 ± 0.003 •	0.961 ± 0.003	0.950 ± 0.003	0.808 ± 0.003 •	0.839 ± 0.003 •	**0.962 ± 0.003**
Computers	**0.904 ± 0.003**	/	/	/	0.813 ± 0.002 •	0.877 ± 0.002 •	0.775 ± 0.002 •	0.774 ± 0.002 •	0.806 ± 0.002 •	0.881 ± 0.002 •
Education	**0.890 ± 0.008**	/	/	/	0.843 ± 0.006 •	0.763 ± 0.006 •	0.803 ± 0.006 •	0.763 ± 0.006 •	0.798 ± 0.006 •	0.800 ± 0.006 •
Entertainment	0.895 ± 0.007	/	/	/	0.846 ± 0.005 •	0.820 ± 0.005 •	0.776 ± 0.005 •	**0.907 ± 0.005** ∘	0.776 ± 0.005 •	0.880 ± 0.005 •
Health	0.968 ± 0.010	/	/	/	0.864 ± 0.007 •	0.839 ± 0.006 •	0.937 ± 0.007 •	0.831 ± 0.007 •	0.846 ± 0.007 •	**0.974 ± 0.008**
Recreation	**0.886 ± 0.000**	/	/	/	0.794 ± 0.000 •	0.845 ± 0.000 •	0.790 ± 0.000 •	0.879 ± 0.000	0.763 ± 0.000 •	0.817 ± 0.000 •
Reference	**0.941 ± 0.004**	/	/	/	0.876 ± 0.004 •	0.825 ± 0.004 •	0.909 ± 0.004 •	0.937 ± 0.004	0.839 ± 0.004 •	0.868 ± 0.004 •
Science	**0.867 ± 0.013**	/	/	/	0.801 ± 0.010 •	0.756 ± 0.009 •	0.783 ± 0.010 •	0.864 ± 0.010	0.814 ± 0.010 •	0.782 ± 0.010 •
Social	**0.967 ± 0.005**	/	/	/	0.921 ± 0.005 •	0.941 ± 0.005 •	0.881 ± 0.005 •	0.895 ± 0.005 •	0.956 ± 0.005 •	0.874 ± 0.005 •
Society	**0.872 ± 0.008**	/	/	/	0.776 ± 0.006 •	0.795 ± 0.006 •	0.752 ± 0.006 •	0.861 ± 0.006 •	0.868 ± 0.006	0.784 ± 0.006 •
Enron	0.923 ± 0.007	/	/	/	0.901 ± 0.005 •	**0.924 ± 0.005**	0.815 ± 0.005 •	0.790 ± 0.005 •	0.877 ± 0.005 •	0.876 ± 0.005 •
Corel5k	**0.840 ± 0.006**	/	/	/	0.787 ± 0.005 •	0.791 ± 0.005 •	0.824 ± 0.005 •	0.832 ± 0.005 •	0.796 ± 0.005 •	0.726 ± 0.010 •
Image	**0.861 ± 0.013**	/	/	/	0.771 ± 0.009 •	0.724 ± 0.009 •	0.785 ± 0.009 •	0.753 ± 0.008 •	0.819 ± 0.008 •	0.858 ± 0.008
Medical	**0.963 ± 0.008**	/	/	/	0.825 ± 0.018 •	0.895 ± 0.008 •	0.903 ± 0.044 •	0.908 ± 0.037 •	0.865 ± 0.042 •	0.961 ± 0.002
Language Log	**0.727 ± 0.001**	/	/	/	0.676 ± 0.024 •	0.646 ± 0.032 •	0.664 ± 0.034 •	0.720 ± 0.044 •	0.619 ± 0.019 •	0.693 ± 0.049 •
RCV1V2(subset1)	**0.923 ± 0.013**	/	/	/	0.800 ± 0.013 •	0.836 ± 0.018 •	0.767 ± 0.040 •	0.889 ± 0.009 •	0.897 ± 0.017 •	0.917 ± 0.010
RCV1V2(subset2)	**0.923 ± 0.002**	/	/	/	0.854 ± 0.050 •	0.896 ± 0.047 •	0.782 ± 0.034 •	0.899 ± 0.030 •	0.837 ± 0.010 •	0.901 ± 0.031 •
Bibtex	0.930 ± 0.006	/	/	/	**0.939 ± 0.044**	0.888 ± 0.017 •	0.887 ± 0.017 •	0.805 ± 0.042 •	0.919 ± 0.021 •	0.804 ± 0.048 •
Delicious	**0.902 ± 0.003**	/	/	/	0.817 ± 0.029 •	0.766 ± 0.038 •	0.837 ± 0.005 •	0.780 ± 0.016 •	0.884 ± 0.043 •	0.758 ± 0.024 •
Eur-Lex(Sm)	0.985 ± 0.010	/	/	/	0.906 ± 0.032 •	**0.987 ± 0.044**	0.978 ± 0.037	0.949 ± 0.015 •	0.875 ± 0.004 •	0.982 ± 0.040
Bookmark	**0.901 ± 0.001**	/	/	/	0.775 ± 0.020 •	0.857 ± 0.009 •	0.878 ± 0.008 •	0.778 ± 0.024 •	0.896 ± 0.040	0.823 ± 0.008 •
Nuswide	**0.688 ± 0.009**	/	/	/	0.657 ± 0.048 •	0.663 ± 0.008 •	0.605 ± 0.048 •	0.610 ± 0.005 •	0.676 ± 0.006 •	0.608 ± 0.047 •
TMC2007-5000	0.949 ± 0.011	/	/	/	0.813 ± 0.004 •	0.853 ± 0.034 •	0.843 ± 0.019 •	**0.957 ± 0.030**	0.925 ± 0.004 •	0.870 ± 0.012 •
Stackex-Chemistry	**0.913 ± 0.009**	/	/	/	0.796 ± 0.046 •	0.912 ± 0.006	0.839 ± 0.031 •	0.865 ± 0.003 •	0.803 ± 0.021 •	0.823 ± 0.013 •
Stackex-Chess	**0.892 ± 0.007**	/	/	/	0.852 ± 0.003 •	0.886 ± 0.028	0.806 ± 0.008 •	0.810 ± 0.024 •	0.749 ± 0.018 •	0.791 ± 0.049 •
Stackex-Cooking	**0.896 ± 0.001**	/	/	/	0.833 ± 0.031 •	0.831 ± 0.039 •	0.767 ± 0.016 •	0.769 ± 0.002 •	0.834 ± 0.036 •	0.863 ± 0.015 •
Stackex-Cs	**0.914 ± 0.011**	/	/	/	0.811 ± 0.020 •	0.853 ± 0.008 •	0.882 ± 0.026 •	0.796 ± 0.008 •	0.780 ± 0.005 •	0.777 ± 0.024 •
Stackex-Philosophy	0.896 ± 0.003	/	/	/	**0.899 ± 0.019**	0.762 ± 0.002 •	0.874 ± 0.004 •	0.876 ± 0.024 •	0.860 ± 0.021 •	0.809 ± 0.009 •
NUS-WIDE	**0.850 ± 0.027**	/	/	/	/	/	/	0.822 ± 0.031 •	0.796 ± 0.056 •	0.801 ± 0.056 •
Win/tie/loss		6 / 0 / 0	6 / 0 / 0	4 / 2 / 0	27 / 2 / 0	24 / 5 / 0	26 / 3 / 0	30 / 5 / 1	34 / 2 / 0	30 / 6 / 0

Table 4 Average precision (mean ± std.) of WL-GLMVML-ATC and compared methods for test instances

Data sets	Ours	LMSC	MVMLSS	MLDL	LF-LPLC	MLCHE	GLOCAL	MVMLP	SSDR-MML	GLMVML
Mfeat	**0.829 ± 0.024**	0.729 ± 0.011 •	0.706 ± 0.017 •	0.798 ± 0.005 •	/	/	/	0.767 ± 0.016 •	0.818 ± 0.016 •	0.814 ± 0.016 •
Reuters	**0.963 ± 0.022**	0.953 ± 0.013 •	0.923 ± 0.020 •	0.812 ± 0.027 •	/	/	/	0.948 ± 0.018 •	0.949 ± 0.018 •	0.953 ± 0.017 •
Corel	0.871 ± 0.014	0.835 ± 0.015 •	0.862 ± 0.014 •	**0.878 ± 0.013**	/	/	/	0.802 ± 0.013 •	0.748 ± 0.013 •	0.763 ± 0.012 •
VOC	**0.701 ± 0.004**	0.617 ± 0.006 •	0.648 ± 0.047 •	0.683 ± 0.028 •	/	/	/	0.594 ± 0.011 •	0.682 ± 0.018 •	0.637 ± 0.034 •
MIR	**0.521 ± 0.002**	0.471 ± 0.042 •	0.440 ± 0.004 •	0.496 ± 0.004 •	/	/	/	0.462 ± 0.002 •	0.464 ± 0.045 •	0.486 ± 0.006 •
3Source	**0.703 ± 0.014**	0.630 ± 0.050 •	0.614 ± 0.034 •	0.655 ± 0.025 •	/	/	/	0.624 ± 0.005 •	0.692 ± 0.035 •	0.680 ± 0.002 •
Arts	**0.655 ± 0.008**	/	/	/	0.634 ± 0.005 •	0.580 ± 0.005 •	0.608 ± 0.005 •	0.603 ± 0.005 •	0.627 ± 0.004 •	0.639 ± 0.004 •
Business	**0.916 ± 0.004**	/	/	/	0.902 ± 0.004 •	0.839 ± 0.004 •	0.803 ± 0.004 •	0.909 ± 0.004	0.905 ± 0.004 •	0.895 ± 0.004 •
Computers	**0.719 ± 0.005**	/	/	/	0.652 ± 0.004 •	0.628 ± 0.004 •	0.658 ± 0.004 •	0.705 ± 0.004 •	0.717 ± 0.004	0.690 ± 0.004 •
Education	**0.646 ± 0.010**	/	/	/	0.561 ± 0.008 •	0.643 ± 0.008	0.617 ± 0.008 •	0.561 ± 0.007 •	0.643 ± 0.008	0.621 ± 0.008 •
Entertainment	**0.691 ± 0.009**	/	/	/	0.683 ± 0.008 •	0.604 ± 0.007 •	0.658 ± 0.008 •	0.702 ± 0.008 ○	0.582 ± 0.008 •	0.684 ± 0.008
Health	**0.771 ± 0.001**	/	/	/	0.680 ± 0.001 •	0.676 ± 0.001 •	0.770 ± 0.001	0.722 ± 0.001 •	0.744 ± 0.001 •	0.682 ± 0.001 •
Recreation	**0.665 ± 0.005**	/	/	/	0.633 ± 0.004 •	0.575 ± 0.004 •	0.638 ± 0.004 •	0.671 ± 0.004	0.655 ± 0.004 •	0.579 ± 0.004 •
Reference	0.721 ± 0.009	/	/	/	**0.733 ± 0.007** ○	0.720 ± 0.007	0.677 ± 0.007 •	0.711 ± 0.007 •	0.642 ± 0.007 •	0.627 ± 0.007 •
Science	**0.600 ± 0.010**	/	/	/	0.596 ± 0.008	0.524 ± 0.008 •	0.524 ± 0.009 •	0.521 ± 0.009 •	0.585 ± 0.008 •	0.594 ± 0.008 •
Social	**0.810 ± 0.010**	/	/	/	0.800 ± 0.008 •	0.686 ± 0.007 •	0.729 ± 0.008 •	0.759 ± 0.008 •	0.708 ± 0.007 •	0.807 ± 0.007
Society	**0.672 ± 0.013**	/	/	/	0.616 ± 0.009 •	0.582 ± 0.008 •	0.665 ± 0.009	0.615 ± 0.009 •	0.652 ± 0.009 •	0.644 ± 0.009 •
Enron	**0.657 ± 0.009**	/	/	/	0.591 ± 0.006 •	0.602 ± 0.006 •	0.572 ± 0.006 •	0.606 ± 0.006 •	0.567 ± 0.006 •	0.602 ± 0.006 •
Corel5k	**0.429 ± 0.005**	/	/	/	0.402 ± 0.004 •	0.387 ± 0.004 •	0.435 ± 0.004 ○	0.430 ± 0.004	0.407 ± 0.004 •	0.416 ± 0.014 •
Image	**0.826 ± 0.010**	/	/	/	0.728 ± 0.007 •	0.800 ± 0.007 •	0.704 ± 0.007 •	0.798 ± 0.007 •	0.832 ± 0.007	0.784 ± 0.006 •
Medical	**0.916 ± 0.016**	/	/	/	0.881 ± 0.026 •	0.886 ± 0.049 •	0.913 ± 0.015	0.823 ± 0.038 •	0.791 ± 0.032 •	0.817 ± 0.008 •
Language Log	**0.395 ± 0.005**	/	/	/	0.361 ± 0.041 •	0.338 ± 0.020 •	0.388 ± 0.015 •	0.371 ± 0.032 •	0.369 ± 0.022 •	0.362 ± 0.029 •
RCV1V2(subset1)	0.628 ± 0.003	/	/	/	**0.637 ± 0.007** ○	0.604 ± 0.040 •	0.612 ± 0.041 •	0.602 ± 0.018 •	0.556 ± 0.021 •	0.615 ± 0.002 •
RCV1V2(subset2)	**0.637 ± 0.008**	/	/	/	0.625 ± 0.047 •	0.555 ± 0.014 •	0.545 ± 0.010 •	0.579 ± 0.018 •	0.561 ± 0.048 •	0.554 ± 0.013 •
Bibtex	**0.615 ± 0.007**	/	/	/	0.611 ± 0.011	0.618 ± 0.043	0.561 ± 0.016 •	0.549 ± 0.042 •	0.540 ± 0.023 •	0.584 ± 0.036 •
Delicious	**0.369 ± 0.012**	/	/	/	0.360 ± 0.014 •	0.346 ± 0.021 •	0.314 ± 0.050 •	0.369 ± 0.032	0.336 ± 0.001 •	0.357 ± 0.006 •
Eur-Lex(Sm)	**0.884 ± 0.006**	/	/	/	0.813 ± 0.050 •	0.833 ± 0.001 •	0.752 ± 0.043 •	0.778 ± 0.002 •	0.762 ± 0.035 •	0.761 ± 0.030 •
Bookmark	**0.508 ± 0.002**	/	/	/	0.469 ± 0.031 •	0.451 ± 0.009 •	0.476 ± 0.040 •	0.441 ± 0.046 •	0.450 ± 0.002 •	0.505 ± 0.002
Nuswide	0.413 ± 0.004	/	/	/	**0.420 ± 0.035** ○	0.407 ± 0.014 •	0.356 ± 0.018 •	0.396 ± 0.045 •	0.397 ± 0.005 •	0.390 ± 0.010 •
TMC2007-5000	**0.868 ± 0.020**	/	/	/	0.834 ± 0.002 •	0.791 ± 0.001 •	0.745 ± 0.022 •	0.847 ± 0.012 •	0.763 ± 0.025 •	0.839 ± 0.043 •
Stackex-Chemistry	**0.462 ± 0.015**	/	/	/	0.424 ± 0.026 •	0.414 ± 0.039 •	0.438 ± 0.041 •	0.445 ± 0.010 •	0.394 ± 0.034 •	0.441 ± 0.014 •
Stackex-Chess	**0.500 ± 0.012**	/	/	/	0.477 ± 0.050 •	0.432 ± 0.013 •	0.422 ± 0.037 •	0.454 ± 0.013 •	0.436 ± 0.045 •	0.451 ± 0.026 •
Stackex-Cooking	**0.511 ± 0.016**	/	/	/	0.458 ± 0.022 •	0.477 ± 0.046 •	0.458 ± 0.003 •	0.479 ± 0.021 •	0.467 ± 0.034 •	0.450 ± 0.008 •
Stackex-Cs	**0.528 ± 0.010**	/	/	/	0.510 ± 0.034 •	0.480 ± 0.005 •	0.505 ± 0.037 •	0.477 ± 0.016 •	0.527 ± 0.047	0.465 ± 0.009 •
Stackex-Philosophy	**0.510 ± 0.009**	/	/	/	0.433 ± 0.019 •	0.464 ± 0.039 •	0.488 ± 0.006 •	0.439 ± 0.020 •	0.437 ± 0.038 •	0.473 ± 0.007 •
NUS-WIDE	**0.871 ± 0.011**	/	/	/	/	/	/	0.782 ± 0.015 •	0.876 ± 0.011	0.853 ± 0.011 •
Win/tie/loss		6 / 0 / 0	6 / 0 / 0	5 / 1 / 0	24 / 2 / 3	26 / 3 / 0	25 / 3 / 1	31 / 4 / 1	31 / 5 / 0	33 / 3 / 0

Table 5 Average ranks of WL-GLMVML-ATC and other compared methods on different cases in terms of AUC

Case 1	Ours	LMSC	MVMLSS	MLDL
Average	1	3.33	2.83	2.83
Rank difference	/	2.33	1.83	1.83
Case 2	Ours	LF-LPLC	MLCHE	GLOCAL
Average	1.17	2.97	2.76	3.1
Rank difference	/	1.8	1.59	1.93
Case 3	Ours	MVMLP	SSDR-MML	GLMVML
Average	1.11	2.78	3.11	3
Rank difference	/	1.67	2	1.89

Table 6 Average ranks of WL-GLMVML-ATC and other compared methods on different cases in terms of precision

Case 1	Ours	LMSC	MVMLSS	MLDL
Average	1.17	3.17	3.5	2.17
Rank difference	/	2	2.33	1
Case 2	Ours	LF-LPLC	MLCHE	GLOCAL
Average	1.17	2.55	3.21	3
Rank difference	/	1.38	2.04	1.83
Case 3	Ours	MVMLP	SSDR-MML	GLMVML
Average	1.14	2.75	3	3.08
Rank difference	/	1.61	1.86	1.94

Table 7 Statistical values on different cases in terms of AUC

Case 1		Case 2		Case 3	
N	6	N	29	N	36
k	4	k	4	k	4
χ_F^2	11.18	χ_F^2	54.06	χ_F^2	67.10
F_F	8.20	F_F	45.94	F_F	57.43
$F_{0.05}(3, 15)$	3.29	$F_{0.05}(3, 84)$	2.71	$F_{0.05}(3, 105)$	2.69
$F_{0.10}(3, 15)$	2.49	$F_{0.10}(3, 84)$	2.15	$F_{0.10}(3, 105)$	2.14
$q_{0.05}$	2.57	$q_{0.05}$	2.57	$q_{0.05}$	2.57
$q_{0.10}$	2.29	$q_{0.10}$	2.29	$q_{0.10}$	2.29
$CD_{0.05}$	1.91	$CD_{0.05}$	0.87	$CD_{0.05}$	0.78
$CD_{0.10}$	1.71	$CD_{0.10}$	0.78	$CD_{0.10}$	0.70

Table 8 Statistical values on different cases in terms of precision

Case 1		Case 2		Case 3	
N	6	N	29	N	36
k	4	k	4	k	4
χ_F^2	12.16	χ_F^2	58.75	χ_F^2	72.94
F_F	10.40	F_F	58.24	F_F	72.81
$F_{0.05}(3, 15)$	3.29	$F_{0.05}(3, 84)$	2.71	$F_{0.05}(3, 105)$	2.69
$F_{0.10}(3, 15)$	2.49	$F_{0.10}(3, 84)$	2.15	$F_{0.10}(3, 105)$	2.14
$q_{0.05}$	2.57	$q_{0.05}$	2.57	$q_{0.05}$	2.57
$q_{0.10}$	2.29	$q_{0.10}$	2.29	$q_{0.10}$	2.29
$CD_{0.05}$	1.91	$CD_{0.05}$	0.87	$CD_{0.05}$	0.78
$CD_{0.10}$	1.71	$CD_{0.10}$	0.78	$CD_{0.10}$	0.70

5.2.4 Convergence

We empirically study the convergence of WL-GLMVML-ATC like GLOCAL [5] does so as to measure the performance of WL-GLMVML-ATC. Fig. 2 shows the objective value with respect to the number of iterations. For convenience and due to the lack of space, we only show the results on multi-view data set Mfeat, multi-label data sets Computers and Language Log, multi-view multi-label data set NUS-WIDE. As can be seen, the objective converges quickly in a few iterations (less than 25). A similar phenomenon can be observed on the other data sets.

What's more, as we know, WL-GLMVML-ATC is developed on the base of ATC, UIG, and GLMVML. Thus, we give the convergence time cost and corresponding ratios to total running time of WL-GLMVML-ATC on all used data sets in Table 10. In this table, 'CT' represents convergence time and 'Ra' indicates the corresponding ratios. According to this table, it is found that on all used data sets, the ratios of convergence time to total running time of

Table 9 Running time (in seconds) of WL-GLMVML-ATC and compared methods

Data sets	Ours	LMSC	MVMLSS	MLDL	LF-LPLC	MLCHE	GLOCAL	MVMLP	SSDR-MML	GLMVML
Mfeat	25.11	24.51	25.57	24.76	/	/	/	25.08	24.77	25.43
Reuters	699.45	683.53	679.05	685.23	/	/	/	656.89	656.54	643.59
Corel	8.37	8.26	7.88	8.28	/	/	/	7.87	8.45	7.90
VOC	80.37	73.03	81.61	78.39	/	/	/	77.72	80.75	80.43
MIR	377.49	376.59	379.69	377.98	/	/	/	360.02	360.96	368.03
3Source	0.23	0.23	0.21	0.23	/	/	/	0.22	0.22	0.22
Avg. (mv)	198.51	194.36	195.67	195.81	/	/	/	187.97	188.62	187.60
Arts	58.31	/	/	/	53.47	58.13	56.16	54.40	57.75	55.69
Business	52.72	/	/	/	49.09	49.72	52.49	53.43	53.76	49.30
Computers	45.58	/	/	/	46.82	45.38	47.05	45.78	45.69	42.43
Education	43.08	/	/	/	39.95	41.78	42.12	40.61	43.01	42.58
Entertainment	41.86	/	/	/	42.10	39.41	41.24	38.93	38.56	39.50
Health	45.36	/	/	/	45.85	41.83	42.34	45.04	43.71	42.75
Recreation	36.56	/	/	/	36.65	33.48	35.20	37.32	35.27	36.59
Reference	27.29	/	/	/	27.53	27.14	27.71	28.60	27.90	27.49
Science	31.60	/	/	/	29.42	31.22	30.67	28.61	32.23	31.67
Social	25.57	/	/	/	24.38	25.69	24.06	25.72	25.51	24.05
Society	32.33	/	/	/	31.49	32.10	32.34	30.75	33.40	30.88
Enron	21.49	/	/	/	22.07	20.24	20.89	21.12	20.73	20.26
Corel5k	64.47	/	/	/	62.07	60.96	63.64	61.99	64.70	64.41
Image	8.97	/	/	/	8.76	8.85	8.78	8.53	8.44	8.35
Medical	8.14	/	/	/	7.85	7.50	7.73	7.85	8.14	8.10
Language Log	10.74	/	/	/	10.71	10.73	10.38	10.80	10.14	10.39
RCV1V2(subset1)	104.02	/	/	/	99.96	102.15	100.85	102.20	103.02	101.33
RCV1V2(subset2)	84.06	/	/	/	79.44	87.21	77.45	82.27	78.89	80.73
Bibtex	99.71	/	/	/	91.00	95.44	91.05	90.98	91.37	96.12
Delicious	1669.68	/	/	/	1662.65	1688.81	1527.99	1656.36	1653.43	1744.80
Eur-Lex(Sm)	209.38	/	/	/	216.85	212.69	209.20	209.40	203.21	198.71
Bookmark	819.19	/	/	/	766.14	752.69	781.22	764.95	825.74	777.30
Nuswide	2165.44	/	/	/	2056.62	2092.68	1998.85	2143.32	2236.02	2130.80
TMC2007-5000	226.82	/	/	/	216.91	219.39	219.10	216.25	210.58	215.68
Stackex-Chemistry	52.49	/	/	/	52.96	53.47	52.54	48.70	51.07	52.52
Stackex-Chess	14.17	/	/	/	14.12	14.07	13.09	13.79	13.48	14.08
Stackex-Cooking	79.42	/	/	/	77.57	73.49	78.75	82.02	75.55	77.58
Stackex-Cs	85.26	/	/	/	78.47	82.44	81.40	83.93	83.15	83.31
Stackex-Philosophy	57.01	/	/	/	58.44	53.50	58.32	58.02	54.71	55.09
Avg. (ml)	214.51	/	/	/	207.22	209.04	201.12	210.06	214.80	212.50
NUS-WIDE	42.09	/	/	/	/	/	/	35.98	37.12	39.81

WL-GLMVML-ATC range from [0.754, 0.982]. Compared with the results of Table 9, refer to the processing of compared methods [5, 8, 42–48], and combine the introduction of 4.2.4, the convergence time is still be accepted by us and their ratios accord to the framework of WL-GLMVML-ATC.

5.2.5 Influence of parameters

According to the experimental setting, there are many adjustable parameters in WL-GLMVML-ATC and two important ones are the numbers of training instances and the numbers of removed labels. Thus we use Fig. 3 to show the influence directly. For convenience, only data set NUS-WIDE and compared methods MVMLP, SSDR-MML, GLMVML are adopted. Indeed, for other data sets and compared methods, the results are similar.

According to this figure, we can see that (1) with the increasing of the numbers of training instances, the AUC, precision, and running time become higher; (2) when more labels are removed, the AUC and precision are worse while

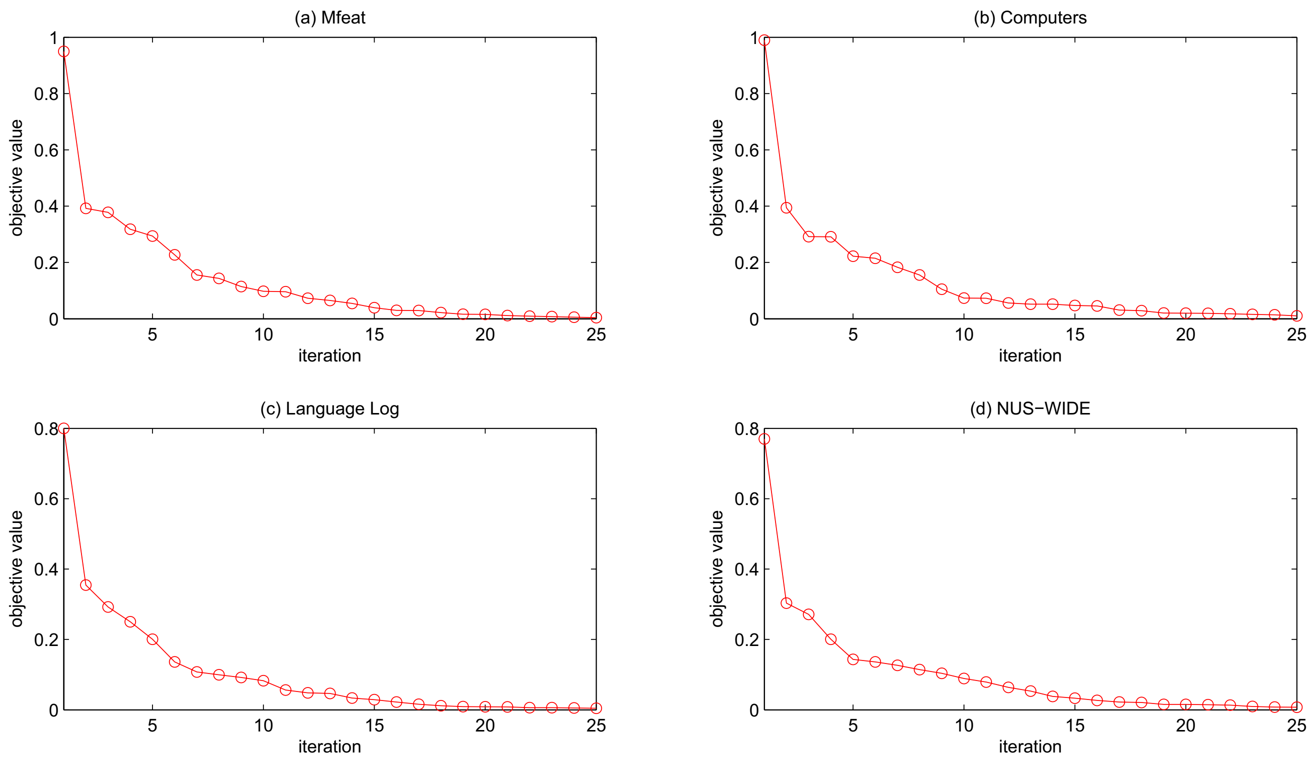

Fig. 2 Convergence of WL-GLMVML-ATC on data sets Mfeat, Computers, Language Log, and NUS-WIDE

Table 10 Convergence time (in seconds) and corresponding ratios of WL-GLMVML-ATC

Data sets	CT	Ra	Data sets	CT	Ra
Mfeat	24.18	0.963	Corel5k	51.32	0.796
Reuters	640.70	0.916	Image	7.42	0.827
Corel	7.55	0.902	Medical	6.75	0.829
VOC	78.92	0.982	Language Log	9.11	0.848
MIR	345.78	0.916	RCV1V2(subset1)	80.20	0.771
3Source	0.22	0.941	RCV1V2(subset2)	66.41	0.790
Arts	46.88	0.804	Bibtex	82.16	0.824
Business	43.86	0.832	Delicious	1258.94	0.754
Computers	34.96	0.767	Eur-Lex(Sm)	161.43	0.771
Education	35.93	0.834	Bookmark	634.87	0.775
Entertainment	31.60	0.755	Nuswide	1699.87	0.785
Health	34.61	0.763	TMC2007-5000	182.14	0.803
Recreation	28.00	0.766	Stackex-Chemistry	42.20	0.804
Reference	21.61	0.792	Stackex-Chess	11.59	0.818
Science	24.65	0.780	Stackex-Cooking	65.84	0.829
Social	19.28	0.754	Stackex-Cs	67.18	0.788
Society	26.48	0.819	Stackex-Philosophy	46.12	0.809
Enron	16.87	0.785	NUS-WIDE	41.25	0.980

the running time is also decreasing; (3) compared with the MVMLP, SSDR-MML, and GLMVML, the AUC and precision of WL-GLMVML-ATC are less affected by the changing of these parameters. In other words, the performances of our WL-GLMVML-ATC are more stable.

5.2.6 Summary of increase proportion of performance

Table 11 shows the detailed comparison between WL-GLMVML-ATC and other compared ones in terms of the increase proportion. The value in this table means that for an index, compared with the other method, how many increase proportion does the WL-GLMVML-ATC get. For example, for LMSC and AUC, 5.86% indicates that compared with LMSC, our WL-GLMVML-ATC brings a better AUC and the increase proportion is 5.86%. Now from this summary table, we can draw a conclusion that with the consideration of ATC, Universum learning, global and local label correlation, WL-GLMVML-ATC can bring a better AUC and precision. Although WL-GLMVML-ATC needs a longer running time, but in practice experiments, a better AUC and precision is always the priority target. What's more, if we improve the experimental environment and adopt GPU or some other better equipments, we can imagine that the increase proportion of running time will be reduced further.

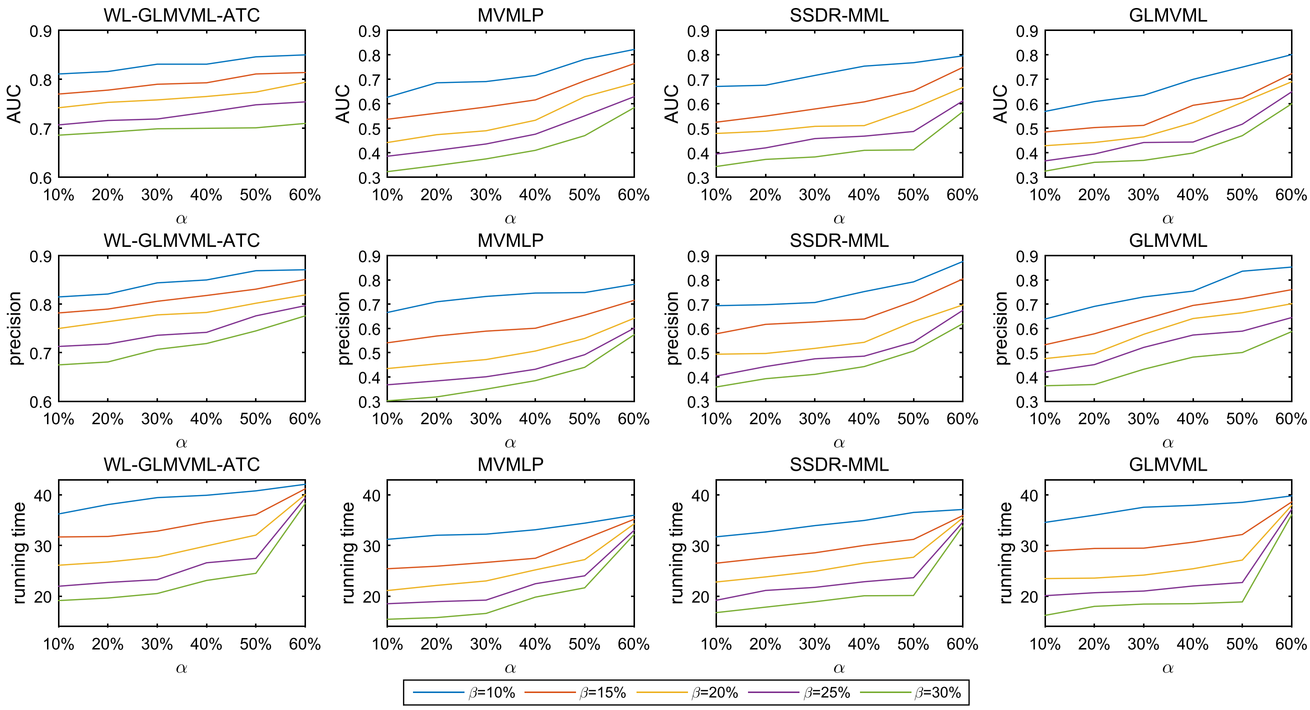

Fig. 3 Influence of two kinds of parameters for WL-GLMVML-ATC, MVMLP, SSDR-MML, GLMVML on data set NUS-WIDE in terms of AUC, precision, and running time. α: ratios of training instances to the total numbers of instances; β: ratios of removed labels to the total numbers of labels

Table 11 Detailed comparison between WL-GLMVML-ATC and other compared ones in terms of the increase proportion

Multi-view			
Index	LMSC	MVMLSS	MLDL
AUC	5.86%	4.88%	3.94%
Precision	5.88%	6.60%	4.44%
Running time	2.13%	1.45%	1.37%
Multi-label			
Index	LF-LPLC	MLCHE	GLOCAL
AUC	7.16%	6.52%	7.04%
Precision	3.43%	5.43%	5.13%
Running time	3.52%	2.61%	6.65%
Multi-view multi-label			
Index	MVMLP	SSDR-MML	GLMVML
AUC	6.01%	6.91%	5.97%
Precision	4.54%	4.71%	4.07%
Running time	2.73%	0.75%	1.72%

6 Conclusions and future studies

In real-world applications, multi-view multi-label data sets are widely used and traditional learning methods always produce worse performances when these data sets exhibit complicate topologies. The reasons for this include that these methods have no ability to exploit global and local label correlations simultaneously, and to reveal the uncertain relationship between instances and the corresponding clusters. In this work, we develop a weak-label-based global and local multi-view multi-label learning with three-way clustering (WL-GLMVML-ATC) to overcome such a problem. In WL-GLMVML-ATC, it makes the belonging of instances to a cluster depend on the probability with active three-way clustering strategy and adopts unlabeled instance generated method to overcome the weak-label case. Then, it makes a learning method can process a complicated data set well with global and local label correlations considered. Although WL-GLMVML-ATC can be treated as the combination of the existing work GLMVML, ATC, and Universum, but in the field of multi-view multi-label learning, it is the first attempt for the combination of global and local label correlations, the three-way decisions, and Universum learning. Experimental results validate that (1) WL-GLMVML-ATC achieves a better average AUC and precision in statistical; (2) the running time of WL-GLMVML-ATC won't add too much; (3) WL-GLMVML-ATC has a good convergence.

While some issues should be solved in the future studies. First, in real-world applications, due to equipment failure, some instances maybe lose partial views. But WL-GLMVML-ATC cannot deal with this. Second, in multi-label learning, there are multiple class labels associated with a single instance simultaneously, and each class label might be determined by some specific features of its own. But WL-GLMVML-ATC does not consider this point. In the future studies, we will pay more attention to these.

Acknowledgements This work is sponsored by 'Chenguang Program' supported by Shanghai Education Development Foundation and Shanghai Municipal Education Commission under grant number 18CG54. This work is also supported by Project funded by China Postdoctoral Science Foundation under grant number 2019M651576, National Natural Science Foundation of China (CN) under grant number 61602296, Natural Science Foundation of Shanghai under grant number 16ZR1414500. The authors would like to thank their supports.

References

1. Liu JL, Teng SH, Fei LK, Zhang W, Fang XZ, Zhang ZX, Wu NQ (2021) A novel consensus learning approach to incomplete multi-view clustering. Pattern Recognit 115:107890
2. Tarekegn A, Giacobini M, Michalak K (2021) A review of methods for imbalanced multi-label classification. Pattern Recognit. https://doi.org/10.1016/j.patcog.2021.107965
3. Hu SZ, Yan XQ, Ye YD (2020) Dynamic auto-weighted multi-view co-clustering. Pattern Recognit 99:107101
4. Yu GX, Xing YY, Wang J, Domeniconi C, Zhang XL (2021) Multiview multi-instance multilabel active learning. IEEE Trans Neural Netw Learn Syst. https://doi.org/10.1109/TNNLS.2021.3056436
5. Zhu Y, Kwok JT, Zhou ZH (2017) Multi-label learning with global and local label correlation. IEEE Trans Knowl Data Eng 99:1–24
6. Tan QY, Yu GX, Wang J, Domeniconi C, Zhang XL (2021) Individuality- and commonality-based multiview multilabel learning. IEEE Trans Cybern 51(3):1716–1727
7. Yin QY, Zhang JG, Wu S, Li HX (2019) Multi-view clustering via joint feature selection and partially constrained cluster label learning. Pattern Recognit 93:380–391
8. Zhu CM, Miao DQ, Wang Z, Zhou RG, Wei L, Zhang XF (2020) Global and local multi-view multi-label learning. Neurocomputing 371:67–77
9. Yao YY (2012) An outline of a theory of three-way decisions. In: Proceedings of the 8th international conference on rough sets and current trends in computing (RSCTC 2012), pp 1–17
10. Yao YY (2016) Three-way decisions and cognitive computing. Cogn Comput 8(4):543–554
11. Yu H (2017) A framework of three-way cluster analysis. In: Proceedings of the international joint conference on rough sets (IJCRS 2017), pp 300–312
12. Yu H, Jiao P, Yao YY, Wang GY (2016) Detecting and refining overlapping regions in complex networks with three-way decisions. Inf Sci 373:21–41
13. Chu XL, Sun BZ, Li X, Han KY, Wu JQ, Zhang Y, Huang QC (2020) Neighborhood rough set-based three-way clustering considering attribute correlations: an approach to classification of potential gout groups. Inf Sci 535:28–41
14. Das P, Das AK, Nayak J, Pelusi D, Ding WP (2019) Group incremental adaptive clustering based on neural network and rough set theory for crime report categorization. Neurocomputing. https://doi.org/10.1016/j.neucom.2019.10.109
15. Zhao B, Ren Y, Gao DK, Xu LZ (2019) Prediction of service life of large centrifugal compressor remanufactured impeller based on clustering rough set and fuzzy Bandelet neural network. Appl Soft Comput 78:132–140
16. Zhou J, Lai ZH, Miao DQ, Gao C, Yue XD (2020) Multigranulation rough-fuzzy clustering based on shadowed sets. Inf Sci 507:553–573
17. Ubukata S, Notsu A, Honda K (2021) Objective function-based rough membership C-means clustering. Inf Sci 548:479–496
18. Zhang PF, Li TR, Wang GQ, Luo C, Chen HM, Zhang JB, Wang DX, Yu Z (2021) Multi-source information fusion based on rough set theory: a review. Inf Fus 68:85–117
19. Zhao J, Liang JM, Dong ZN, Tang DY, Liu Z (2020) Accelerating information entropy-based feature selection using rough set theory with classified nested equivalence classes. Pattern Recognit 107:107517
20. Roy S, Maji P (2020) Rough segmentation of coherent local intensity for bias induced 3-D MR brain images. Pattern Recognit 97:106997
21. Zhao Y, Luo ZW, Quan CQ, Liu DC, Wang G (2020) Cluster-wise learning network for multi-person pose estimation. Pattern Recognit 98:107074
22. Zhang TF, Ma FM, Yue D, Peng C, O'Hare GMP (2020) Interval type-2 fuzzy local enhancement based rough k-means clustering considering imbalanced clusters. IEEE Transactions on Fuzzy Systems 28(9):1925–1939
23. Arnold SD, Radu DG, Horia FP, Costel S (2019) A comparison study of similarity measures in rough sets clustering. In: 2019 IEEE 15th international scientific conference on informatics (ISCI 2019), pp 37–42
24. Feng YF, Chen HM (2019) An improved density peaks clustering based on rough set theory for overlapping community detection. In: 2019 IEEE 14th international conference on intelligent systems and knowledge engineering (ISKE 2019), pp 21–28
25. Roy S, Maji P (2020) Medical image segmentation by partitioning spatially constrained fuzzy approximation spaces. IEEE Trans Fuzzy Syst 28(5):965–977

26. Li DW, Zhang HQ, Li TR, Bouras A, Yu X, Wang T (2021) Hybrid missing value imputation algorithms using fuzzy c-means and vaguely quantified rough set. IEEE Trans Fuzzy Syst. https://doi.org/10.1109/TFUZZ.2021.3058643
27. Gao C, Zhou J, Miao DQ, Wen JJ, Yue XD (2021) Three-way decision with co-training for partially labeled data. Inf Sci 544:500–518
28. Yu H, Wang XC, Wang GY, Zeng XH (2020) An active three-way clustering method via low-rank matrices for multi-view data. Inf Sci 507:823–839
29. Wagstaff K, Cardie C (2000) Clustering with instance-level constraints. In: Proceedings of the 7th international conference on machine learning (ICML 2000), pp 1103–1110
30. Klein D, Kamvar SD, Manning CD (2002) From instance-level constraints to space-level constraints: making the most of prior knowledge in data clustering. Technical Report, Stanford
31. Basu S, Banerjee A, Mooney RJ (2004) Active semi-supervision for pairwise constrained clustering. In: Proceedings of the 4th SIAM international conference on data mining (SDM 2004), pp 333–344
32. Mallapragada PK, Jin R, Jain AK (2008) Active query selection for semi-supervised clustering. In: Proceedings of the 19th international conference on pattern recognition (ICPR 2008), pp 1–4
33. Vapnik V, Kotz S (1982) Estimation of dependences based on empirical data. Springer, Berlin
34. Zhu CM, Miao DQ, Zhou RG, Wei L (2020) Weight-and-Universum-based semi-supervised multi-view learning machine. Soft Comput 24(14):10657–10679
35. Wang Z, Hong SS, Yao LJ, Li DD, Du WL, Zhang J (2020) Multiple Universum empirical kernel learning. Eng Appl Artif Intell 89:103461
36. Richhariya B, Tanveer M (2020) A reduced universum twin support vector machine for class imbalance learning. Pattern Recognit 102:107150
37. Liu CL, Hsaio WH, Lee CH, Chang TH, Kuo TH (2016) Semi-supervised text classification with Universum learning. IEEE Trans Cybern 46(2):462–473
38. Zhang CQ, Yu ZW, Hu QH, Zhu PF, Liu XW, Wang XB (2018) Latent semantic aware multi-view multi-label classification. In: Thirty-second AAAI conference on artificial intelligence, pp 4414–4421
39. Zhang J, Li CD, Cao DL, Lin YJ, Su SZ, Dai L, Li SZ (2018) Multi-label learning with label-specific features by resolving label correlations. Knowl Based Syst 159:148–157
40. Huang J, Qin F, Zheng X, Cheng ZK, Yuan ZX, Zhang WG, Huang QM (2019) Improving multi-label classification with missing labels by learning label-specific features. Inf Sci 492:124–146
41. Chua TS, Tang J, Hong R, Li H, Luo Z, Zheng Y (2009) Nus-wide: a real-world web image database from national university of Singapore. In: Proceedings of the ACM international conference on image and video retrieval, p 48
42. He ZY, Chen C, Bu JJ, Li P, Cai D (2015) Multi-view based multi-label propagation for image annotation. Neurocomputing 168:853–860
43. Sun SL, Zhang QJ (2011) Multiple-view multiple-learner semi-supervised learning. Neural Process Lett 34:229–240
44. Zhang CQ, Fu HZ, Hu QH, Cao XC, Xie Y, Tao DC, Xu D (2018) Generalized latent multi-view subspace clustering. IEEE Trans Pattern Anal Mach Intell. https://doi.org/10.1109/TPAMI.2018.2877660
45. Wu F, Jing XY, You XG, Yue D, Hu RM, Yang JY (2016) Multi-view low-rank dictionary learning for image classification. Pattern Recognit 50:143–154
46. Weng W, Lin YJ, Wu SX, Li YW, Kang Y (2018) Multi-label learning based on label-specific features and local pairwise label correlation. Neurocomputing 273:385–394
47. Kumar V, Pujari AK, Padmanabhan V, Sahu SK, Kagita VR (2018) Multi-label classification using hierarchical embedding. Expert Syst Appl 91:263–269
48. Qian BY, Wang X, Ye JP, Davidson I (2015) A reconstruction error based framework for multi-label and multi-view learning. IEEE Trans Knowl Data Eng 27(3):594–607
49. Demsar J (2006) Statistical comparisons of classifiers over multiple data sets. J Mach Learn Res 7(1):1–30