Expert Systems With Applications 192 (2022) 116276

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Multi granularity based label propagation with active learning for semi-supervised classification

Shengdan Hu [a,b], Duoqian Miao [a,b,*], Witold Pedrycz [a,c,d]

[a] Department of Computer Science and Technology, Tongji University, Shanghai, 201804, China
[b] Key Laboratory of Embedded System and Service Computing, Ministry of Education, Tongji University, Shanghai, 201804, China
[c] Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB T6G 2G7, Canada
[d] System Research Institute, Polish Academy of Sciences, Warsaw, PL-01447, Poland

ARTICLE INFO

Keywords:
Semi-supervised learning
Granular computing
Multi granularity
Label propagation
Active learning
Three-way decision

ABSTRACT

Semi-supervised learning (SSL) methods, which exploit both the labeled and unlabeled data, have attracted a lot of attention. One of the major categories of SSL methods, graph-based semi-supervised learning (GBSSL) learns labels of unlabeled data on an adjacency graph, where neighborhood sparse graph is often used to reduce computational complexity. However, the neighborhood size is difficult to set. Instead of assigning a concrete value of neighborhood size, we propose a new label propagation algorithm called multi granularity based label propagation (MGLP) and developed from the view of granular computing. In MGLP, labels of unlabeled data are learned by two classic label propagation processes with diverse neighborhood size k, where granular computing delivers a guiding strategy to leverage multiple level neighborhood information granules, and three-way decision acts as an active learning strategy to select the unlabeled data for further annotating. Through the iterative procedures of label propagating, data annotating and data subset updating, the ultimate pseudo label accuracy of unlabeled data may be higher. Theoretically, the accuracy of pseudo labels is enhanced in some scenarios. Experimentally, the results of simulation studies on ten benchmark datasets, show that the proposed method MGLP can rise pseudo labels accuracy by 8.6% than LP (label propagation), 6.5% than LNP (linear neighborhood propagation), 6.4% than LPSN (label propagation through sparse neighborhood), 4.5% than Adaptive-NP (adaptive neighborhood propagation) and 4.6% than CRLP (consensus rate-based label propagation). It also provides a novel way to annotate data.

1. Introduction

Nowadays, procurement of a deluge of data from various sources is easy. How to make full use of these data and reveal potential information from them plays a key role in many aspects of society. Machine learning, which can learn from and make predictions on data, received a lot of attention in industry and academia. Typically, depending on whether the labels of data are available, machine learning tasks are classified into three categories: supervised learning, unsupervised learning and semi-supervised learning.

Raw data are easily obtained, while annotating all of them is time-consuming and laborious. Semi-supervised learning (SSL) tackles this problem by trying to effectively combine a large amount of unlabeled data with labeled data to learn and predict (Hu et al., 2018; Zhu, 2008). Up to present, many SSL methods were proposed, thereinto the graph-based semi-supervised learning (GBSSL) methods (Gan et al., 2018; Wang & Zhang, 2008; Yu & Kim, 2018; Zhou et al., 2004; Zhu, 2002; Zhu et al., 2003; Zoidi et al., 2018) form a major category, and are applied in several fields (Dornaika et al., 2020; Francisquini et al., 2017; Giasemidis et al., 2020; Hong et al., 2019; Huang et al., 2020; Sun et al., 2018). In GBSSL techniques, it is often assumed that two close nodes in the instance space should have the same label, which refers to smoothness assumption, and two points which are connected by a path going through dense regions should have the same label, which refers to cluster assumption (Zhou, 2018). On account of these assumptions, labels can be learned through the graph.

Generally speaking, the crux of GBSSL methods come with two procedures: constructing graph and then learning labels. As for constructing graph, vertices refer to data (labeled and unlabeled) and edges represent the similarities of data. In term of learning labels, the pseudo labels of unlabeled data can be learned according to the optimization of the objective function (Zhou et al., 2004; Zhu et al., 2003). However, uncertainties arise in these two processes. For example, what is the suitable value of neighborhood size parameter when constructing

* Corresponding author at: Department of Computer Science and Technology, Tongji University, Shanghai, 201804, China.
E-mail addresses: hushengdan@163.com (S. Hu), dqmiao@tongji.edu.cn (D. Miao), wpedrycz@ualberta.ca (W. Pedrycz).

https://doi.org/10.1016/j.eswa.2021.116276
Received 12 June 2020; Received in revised form 11 November 2020; Accepted 21 November 2021
Available online 18 December 2021

graph? How to effectively measure the similarities between data? What is the precise form of objective function to fit data? Recently, some sparse reconstruction based methods are proposed (Cheng et al., 2010; Jia et al., 2016; Zang & Zhang, 2012; Zhang et al., 2018, 2020, 2015) to avoid determining parameters and assigning weights. In l_1-graph (Cheng et al., 2010) and sparse neighborhood graph (Zang & Zhang, 2012), the graph structure and edge weight are obtained simultaneously. Adaptive-NP (Jia et al., 2016), AELP-WL (Zhang et al., 2018) and ALP-TMR (Zhang et al., 2020) are unified frameworks integrating weight learning and label propagation. In this paper, we will only show solicitude for uncertainties caused by constructing graph from the view of granular computing.

Granular computing (Pedrycz, 2013), which has appeared in many fields, and manifested as fuzzy sets (Zadeh, 1997), rough sets (Yao, 1999, 2020a), shadowed sets (Pedrycz & Vukovich, 2002), formal concept analysis (Wu et al., 2009; Yao, 2020a) and alike, has been an important paradigm to tackle uncertainties. Granular computing embraces theories, methodologies, techniques, tools for structured thinking, structured problem solving and structured information processing. As it exactly coincides with the procedure of problem abstracting and resolving of we human beings explicitly or implicitly, granular computing has attracted the attention of many scholars (Ouyang et al., 2019; Qian et al., 2020; Zhang & Miao, 2016; Zhou et al., 2020). Selecting a suitable level of information granularity and constructing corresponding information granules is one of the fundamental issues in granular computing. From the standpoint of data-driven, Pedrycz and Homenda (2013) proposed the notion of justifiable granularity, and pointed out that information granules should be justified in light of experimental evidence and have specific enough meaning semantics. That is to say that an information granule has the characteristics of coverage and specificity (Xu et al., 2018), where coverage refers to consisting as many data as possible, and specificity involves a compact data field.

It is fairly straightforward to associate information granule with the local neighborhood size in graph-based label propagation methods, and the granularity of a neighborhood information granule is intimately connected with the value of neighborhood size parameter k. The value of k influence the results of pseudo labels of unlabeled data: if it is too large to cover the field outside the manifold, the information granule will lose its specificity, and if it is too small to cover limited area, the information granule will lose its coverage. However, there are few of studies about the value of k, that empirically suggest that k should be small (Fan et al., 2018; Zhu, 2008), but determining its exact value is still an open problem.

For the sake of reducing the uncertainty caused by the value of k and improving the pseudo label accuracy, a novel label propagation algorithm called multi granularity based label propagation (MGLP) is proposed by employing multiple level granularity. The workflow of the proposed algorithm MGLP is: first, we learn the pseudo labels of unlabeled data through two classic label propagation algorithms with diverse k, next, applying three-way decision (Yao, 2020a, 2020b) in active learning (Settles, 2010) to choose and annotate the unlabeled data with different pseudo labels learned by the first step, then add these data into the labeled dataset, finally iteratively execute the whole steps mentioned above till convergency. In short, the basic steps in MGLP include label propagating, data annotating and data subset updating.

Main innovations of MGLP can be identified as follows:

(1) Granular computing offers some guidelines for sound structured thinking, to leverage multiple level neighborhood information granules in graph-based semi-supervised learning.

(2) Learn labels of unlabeled data by two classic label propagation processes with diverse k, to strike a balance between the coverage and specificity of neighborhood information granules.

(3) Iteratively adopt three-way decision as an active learning strategy to select the unlabeled data with distinct pseudo labels for further annotating.

Table 1
Key notations and corresponding descriptions.

Notation	Description
C	Number of classes
$\mathcal{X}$	Training data matrix of size $n \times d$
$\mathcal{X}_L$	Labeled data matrix of size $l \times d$
$\mathcal{X}_U$	Unlabeled data matrix of size $u \times d$
W	Weight matrix of size $n \times n$
$\mathcal{Y}$	Label matrix of size $n \times C$
$\mathcal{Y}_L$	Given label matrix of size $l \times C$ for $\mathcal{X}_L$
$\mathcal{Y}_U$	Pseudo label matrix of size $u \times C$ for $\mathcal{X}_U$
D	Diagonal degree matrix of size $n \times n$
P	Probabilistic transition matrix of size $n \times n$

The paper is organized as follows: To make the study self-contained, in Section 2, we outline some basic notations and review related works. In Section 3, we present the proposed multi granularity based label propagation algorithm with three-way decision, and some discussions are included. Extensive simulation experiments on several datasets are conducted and results analyzed in Section 4. Moreover, final conclusions are drawn in Section 5.

2. Related works

In this section, we briefly review some preliminaries which are closely related to this paper. For convenience, key notations and corresponding descriptions are shown in Table 1.

Assuming a C-classification task, let the training data matrix be $\mathcal{X} = [\mathcal{X}_L; \mathcal{X}_U] \in R^{n\times d}$, where $\mathcal{X}_L = [x_1, x_2, \dots, x_l]^T$ represents the labeled data subset, $\mathcal{X}_U = [x_{l+1}, x_{l+2}, \dots, x_{l+u}]^T (l + u = n)$ stands for the unlabeled data subset, and the labels of the labeled data are $y_i \in \{1, 2, \dots, C\}, i \in \{1, 2, \dots, l\}$. The goal of GBSSL is to learn the labels of unlabeled data $\{y_{l+1}, y_{l+2}, \dots, y_{l+u}\}$ through labeled data $\{\langle x_1, y_1\rangle, \langle x_2, y_2\rangle, \dots, \langle x_l, y_l\rangle\}$ over the graph, which are consistent with the labeled data and it should be smooth on the whole graph $g = (\mathcal{V}, \mathcal{E})$. In the graph g, the vertices $\mathcal{V} = \{x_1, x_2, \dots, x_n\}$ represent all the instance data, and the edges $\mathcal{E}$ correspond to similarities (distances) between instances. The similarities are often defined in the form of an weight matrix $W \in R^{n\times n}$, where the entry $w_{ij} \geq 0$ characterizes the similarity between the vertices x_i and x_j $(1 \leq i, j \leq n)$, and $w_{ij} = 0$ if there is no edge $e \in \mathcal{E}$ connects x_i and x_j. w_{ij} can be defined by a Gaussian kernel (Zhu et al., 2003):

$$w_{ij} = \begin{cases} exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right) & x_i \in N(x_j) \text{ and } x_j \in N(x_i) \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

where σ is a bandwidth hyper-parameter, $\|\bullet\|$ refers to the Euclidean norm, and $N(\bullet)$ denotes the neighbors of $x_i (i = 1, 2, \dots, n)$. The neighbors could be the k nearest neighbors (k-NN) or ϵ nearest neighbors (ϵ-NN). Another method to construct W is applying zero–one weighting:

$$w_{ij} = \begin{cases} 1 & x_i \in N(x_j) \text{ and } x_j \in N(x_i) \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

Denote a label matrix $\mathcal{Y} = [\mathcal{Y}_L; \mathcal{Y}_U] = (y_{ij})_{n\times C}$, where $\mathcal{Y}_L$, $\mathcal{Y}_U$ corresponding to labeled and unlabeled data respectively. For labeled data, $y_{ij} = 1$ if x_i is labeled as j and $y_{ij} = 0$ otherwise, and for unlabeled data, the initial value of y_{ij} could be arbitrary, but should satisfy $\sum_{j=1}^{C} y_{ij} = 1, i \in \{l + 1, l + 2, \dots, n\}$. Then, the objective function with respect to the constructed graph can be obtained by minimizing the following general cost (Belkin et al., 2006; Chapelle et al., 2006):

$$\hat{\mathcal{Y}}^* = \arg\min_{\hat{\mathcal{Y}}} \left(\|\hat{\mathcal{Y}}_L - \mathcal{Y}_L\|^2 + \lambda_1 tr(\hat{\mathcal{Y}}^T L \hat{\mathcal{Y}}) + \lambda_2 \|\hat{\mathcal{Y}}\|^2\right), \tag{3}$$

where λ_1 and λ_2 $(\lambda_1, \lambda_2 \geq 0)$ are regularization parameters. The small regularization term $\|\hat{\mathcal{Y}}\|^2$ in (3) is to avoid overfitting, and the first term

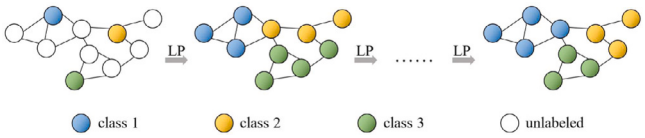


Fig. 1. An illustration of label propagation: pseudo labels of unlabeled data are iteratively updated until convergence occurs.

measures the consistency with the labels of labeled data, which is given as:

$$\|\hat{\mathcal{Y}}_L - \mathcal{Y}_L\|^2 = \sum_{i=1}^{l} \|\hat{y}_i - y_i\|^2. \tag{4}$$

The second term of (3) regularizes the smoothness, where $L = D - W$ is called graph Laplacian, D is the diagonal degree matrix with each elements $D_{ii} = \sum_{j=1}^{n} W_{ij}$, which can be represented as:

$$\begin{aligned} tr(\hat{\mathcal{Y}}^T L \hat{\mathcal{Y}}) &= tr(\hat{\mathcal{Y}}^T (D - W)\hat{\mathcal{Y}}) \\ &= \frac{1}{2}\left(2\sum_{i=1}^{n}\hat{y}_i^2 \sum_{j=1}^{n} W_{ij} - 2\sum_{i,j=1}^{n} W_{ij}\hat{y}_i\hat{y}_j\right) \\ &= \frac{1}{2}\sum_{i,j=1}^{n} W_{ij}\left(\hat{y}_i - \hat{y}_j\right)^2, \end{aligned} \tag{5}$$

In a nutshell, the whole process of GBSSL can be decomposed into two subproblems: graph construction and label propagation. When constructing Laplacian weight graphs, there are usually two main strategies: k-NN and ϵ-NN, where k is the number of instances in the neighborhood, ϵ refers to the radius of neighborhood, and their values are often not easy to be determined. As for label propagation, various methods may be reduced to the framework as shown in (3), involving the variants of the terms or different trade-off between the terms.

Label propagation algorithm (Zhu, 2002) propagates labels based on the known labels and a weighted graph through dense data regions iteratively until convergence has occurred (see Fig. 1), and the output are pseudo labels of unlabeled data. It is noted that there should be no noise in the available labels, namely the known labels will not change during the propagation process.

While the labels of labeled data are fixed in LP, the fitness term $\sum_{i=1}^{l} \|\hat{y}_i - y_i\|^2$ of (3) is zero, then the objective function mainly depends on the smoothness term $\frac{1}{2}\sum_{i,j=1}^{n} W_{ij}\left(\hat{y}_i - \hat{y}_j\right)^2$, and the expression $\hat{y}_i \approx \hat{y}_j$ should hold for those node pairs with large W_{ij}. That is to say, larger weights enable labels to travel through the graph more easily, and the label of an instance can be obtained by its nearest neighbors, which can be written as:

$$y_i = \sum_{x_j \in N(x_i)} w_{ij} y_j. \tag{6}$$

Define the probabilistic transition matrix as $P = (P_{ij})_{n\times n}$, and

$$P_{ij} = P(i \leftarrow j) = \frac{W_{ij}}{\sum_{k=1}^{n} W_{kj}}. \tag{7}$$

In fact, P is a row-normalized affinity matrix and $P = D^{-1}W$. The diffusion process of labels to all nodes on the graph can be formulated as:

$$\mathcal{Y} \leftarrow P\mathcal{Y}, \tag{8}$$

and the process is iteratively executed until the label probability distribution of instances has converged. Then one can obtain the pseudo labels of unlabeled data using the following formula:

$$y_i^* = \arg\max_j \left(y_{ij}\right), \tag{9}$$

where $i \in \{l+1, l+2, \ldots, n\}$ and $j \in \{1, 2, \ldots, C\}$. As the labels of labeled data may be differ from the initial ones after propagation, they need to be clamped in each iteration to avoid fading away, which act as sources being pushed out to unlabeled data.

3. Problem statement and proposed method

3.1. Uncertainties of label propagation algorithms

Many graph-based label propagation algorithms have been proposed, however there are still several open problems with these methods. The main challenges can be categorized as: ① how to construct appropriate graph to reflect the topology structure of instances, ② how to calculate the weight matrix to show the geometry relationship of data, and ③ how to set sound strategy to propagate labels over the weighted graph.

As the uncertainty principle proposed by German physicist Werner Heisenberg indicates, it is impossible to achieve precise values, and this is the intrinsic property of nature. There are also many uncertainties in graph-based label propagation methods, and they will affect the methods' effectiveness.

For graph construction, the uncertainty mainly lies in the edges and nodes of a graph. Since over a full-connected graph, it is often time-consuming to compute in propagation process, and some semantically-unrelated information between instances beyond local region may be conveyed, k-NN or ϵ-NN sparse graph (Zhu, 2008) is often used, where k and ϵ are free local nearest neighbor parameters. However, the graph-based label propagation methods suffer from choosing the optimal local neighborhood size. If the value of local neighborhood size is too large, the edges span outside the manifold may be included in the graph, and if the value of local neighborhood size is too small, the local topology may be not preserved (Wang & Zhang, 2008). Furthermore, it is harder or even vainfruitlessly to set neighborhood size when there exist bridge points or overlapping region among classes (Wang & Zhang, 2008; Zhou et al., 2018), because through such points labels from one class could be wrongly propagated to the nodes from another class.

As to weight matrix calculation, the uncertainty will reside with the similarity measure between nodes. As depicted in Section 2, Gaussian kernel based similarity measure $exp\left(-\|x_i - x_j\|^2/(2\sigma^2)\right)$, which relies on Euclidean distance, is often used. But the variable σ is a bandwidth hyper-parameter, and its value is not easy to set. Some other similarity measures are also used, such as geodesic distance, Kullback–Leibler divergence (Fan et al., 2018), Jensen–Shannon divergence (Chen et al., 2006), and cosine distance (Yu et al., 2019; Zhang et al., 2019). Based on the assumption that the original linearly inseparable instances can be linearly separated in a higher-dimensional space, the pairwise similarity between x_i and x_j can be measured as $\langle\phi(x_i), \phi(x_j)\rangle$ in the kernel space K, where $\phi : x \in \mathcal{X} \rightarrow \phi(x) \in K$ is the nonlinear mapping (Zhang et al., 2019). On the contrary, in feature low-dimensional subspace (Li et al., 2019), similarity measurement is implemented based on the feature subsets of data. In adaptive graph, the similarities can be calculated and updated during the model selection processing (Zhao et al., 2019). All in all, diverse methods are applied to decrease the uncertainty caused by similarity measure.

The main uncertainty in label propagation processing stems from the model definition (Belkin et al., 2006), including the terms of a model and corresponding trade-off parameters of each term. The optimal model selection approach or solution process of a model (Chapelle et al., 2006) may also greatly affect the efficiency and effectiveness of a graph-based label propagation method.

From the above analysis, uncertainty can be seen as an essential characteristic in graph-based semi-supervised learning.

3.2. Illustrative examples

In what follows, a series of experiments are conducted on an artificial dataset to illustrate the uncertainties of the classic label propagation algorithm based on pseudo labels, and since the graph is the crux of the graph-based methods, we concentrate only on the diverse local neighborhood size k when constructing graph.

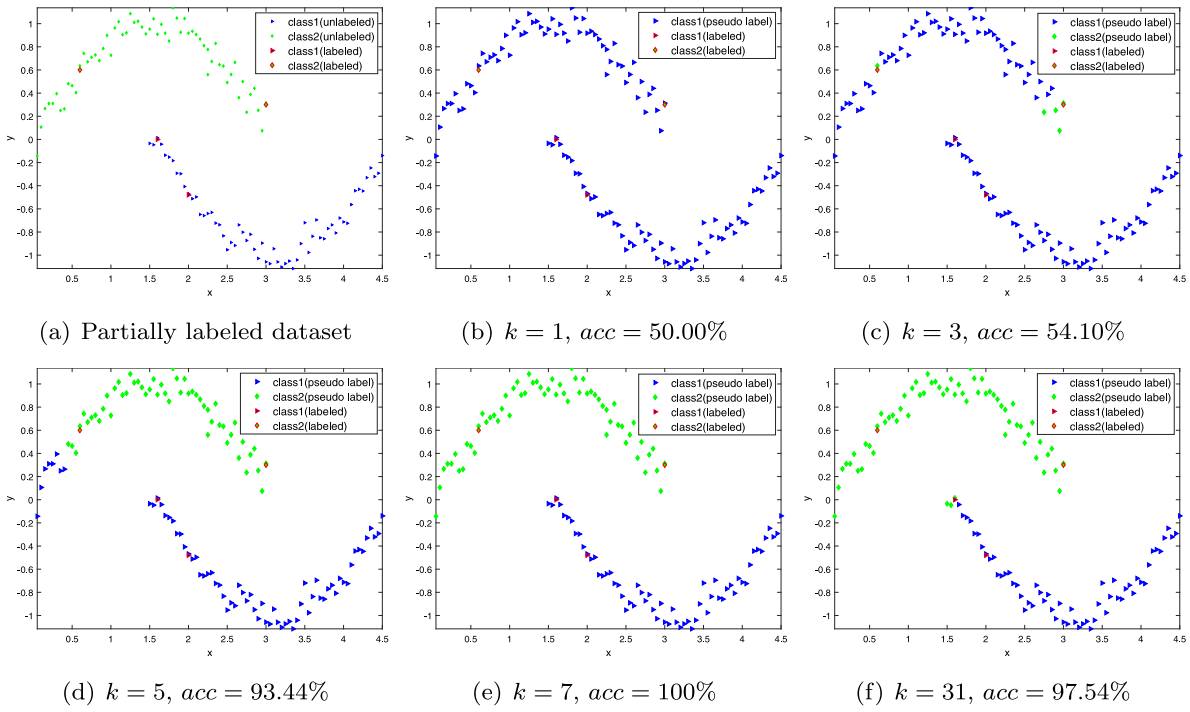


Fig. 2. Various propagation results according to diverse nearest neighbor parameter k on the toy dataset. (a) Toy data with two labeled data for each class (large triangle, diamond with red edge) and 122 unlabeled data (small triangle, diamond). (b)–(f) Classification results after label propagation, and the pseudo label accuracy varies according to different value of k.

The two-class dataset in Fig. 2(a) consists of 126 data, among which four are labeled (two for each class) and the other 122 are unlabeled. The labels of unlabeled data are predicted by adopting classic label propagation procedure (viz. Algorithm 1), where Euclidean distance-based similarities are used and the weight matrix W is defined by zero–one weighting (see (2)). Classification results after label propagation are shown in Fig. 2(b)–(f), and we can figure out the relativity between pseudo label accuracy and nearest neighbor parameter k in this toy dataset. The pseudo label accuracy of unlabeled data is comparatively low when $k \leq 3$, and with the rise of k's value, the accuracy increases rapidly, meanwhile, the accuracy is 100% when $k \in \{6, 7, \ldots, 30\}$, see Fig. 4(a), then it oscillates and declines as $k \geq 31$.

As for real world datasets, there may exist outliers or data located in overlapping region of classes, and they are considered as bridge nodes in graph. If the bridge nodes are originally labeled in dataset, their labels may be more easily propagated to more than one class, consequently, the pseudo label accuracy will be affected. Fig. 3 visualizes the same dataset as Fig. 2(a) all but a bridge node shown in red solid circle, and it pertains to a labeled datum. After propagation, the pseudo label accuracy of unlabeled data with respect to neighborhood size k is shown in Fig. 4(b). Comparing to the original dataset with no bridge node (see Fig. 4(a)), the pseudo label accuracies are impacted more by k, especially when $k \in \{15, 16, \ldots, 37\}$, furthermore, the curve varies rapidly as $k = 17$ and 28.

In a nutshell, the value of local neighborhood size k is pivotal to the classification results of unlabeled data by giving a thorough insight into the results. Empirically, it is suggested that the value of k should be set small to obtain better performance with lower computational complexity (Fan et al., 2018; Zhu, 2008). But determining the exact value remains to be an open problem.

3.3. Multi granularity label propagation

In light of the analysis made above, it is obvious that uncertainty is entrenched in GBSSL methods and their performance needs to be further improved. Zhou (2018) expounds that in order to achieve better

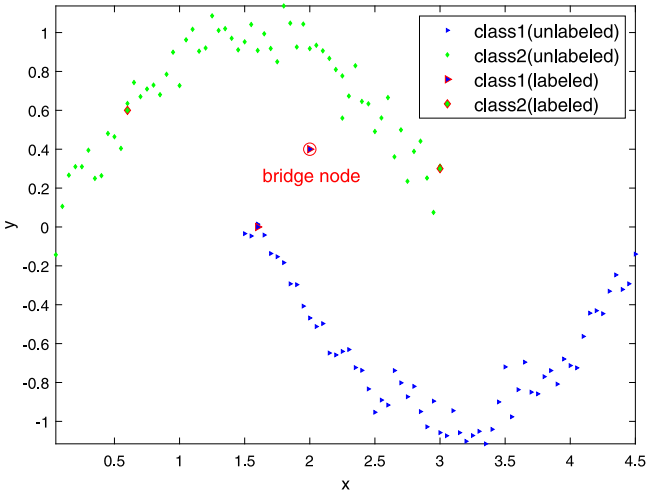
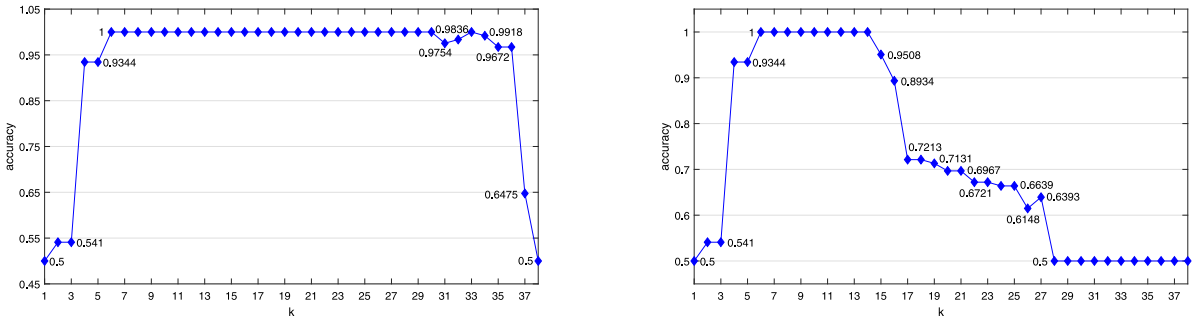


Fig. 3. Toy data as in Fig. 2(a) except for a bridge node.

performance, ensemble mechanisms are often incorporated in semi-supervised learning method. In this section, we will adopt the ensemble strategy to improve the classic label propagation algorithm from the multiple level granularity point of view.

When constructing a neighborhood information granule with respect to a datum according to the local neighborhood size k, the granule will scale as the value of k varies: a smaller k presents more specific neighborhood information granules with finer granularity, however, a greater k presents more expansive neighborhood information granules with coarser granularity. In label propagation algorithm, if the neighborhood information granule of a datum is too coarse to cover the data points outside the manifold, the pseudo label of this datum propagated by its neighbors is likely to inaccurate, similarly, if the neighborhood information granule is too specific to contain only a



(a) Original dataset (b) Dataset with a bridge node

Fig. 4. Pseudo label accuracy of unlabeled data with respect to nearest neighbor parameter k after label propagation. (a) Results of the original dataset as shown in Fig. 2(a). (b) Results of the dataset with a bridge node as shown in Fig. 3.

few data points, the pseudo label of this datum will be inaccurate as lack of information (see Fig. 4). In order to give consideration to both the coverage and specificity of information granules, we propose a novel label propagation algorithm called multi granularity based label propagation (MGLP) to enhance pseudo label accuracy by employing multiple level granularity.

In a formal manner, let LP_{k1} and LP_{k2} be two label propagation algorithms with different neighborhood granularity, after propagation by applying these algorithms, we separate the unlabeled dataset $\mathcal{X}_U$ to three parts:

$$\begin{cases} \mathcal{X}_{U1} = \left\{x \in \mathcal{X}_U | y^{LP_{k1}} = y^{LP_{k2}} = y\right\}, \\ \mathcal{X}_{U2} = \left\{x \in \mathcal{X}_U | y^{LP_{k1}} = y^{LP_{k2}} \neq y\right\}, \\ \mathcal{X}_{U3} = \left\{x \in \mathcal{X}_U | y^{LP_{k1}} \neq y^{LP_{k2}}\right\}. \end{cases} \tag{10}$$

where $\mathcal{X}_{U1}$, $\mathcal{X}_{U2}$ and $\mathcal{X}_{U3}$ are pairwise disjoint, and $\mathcal{X}_{U1} \bigcup \mathcal{X}_{U2} \bigcup \mathcal{X}_{U3} = \mathcal{X}_U$. $y^{LP_{k1}}$ and $y^{LP_{k2}}$ are pseudo labels of x learned by the two label propagation algorithms respectively, besides y is the authentic label of x. In detail, $\mathcal{X}_{U1}$ denotes the dataset in which the two pseudo labels of every datum are just the same as the authentic label, and $\mathcal{X}_{U2}$ denotes the dataset in which the two pseudo labels of every datum are the same but not equal to the authentic label, nevertheless, $\mathcal{X}_{U3}$ refers to the dataset in which the two pseudo labels of every datum are different. Since the real labels of unlabeled data are unknown, we cannot distinguish $\mathcal{X}_{U1}$ and $\mathcal{X}_{U2}$ directly, and we take $\mathcal{X}_{U3}$ for further consideration. For this most uncertain subset, the diverse pseudo labels of data indicate ambiguous results after label propagation carried out with diverse value of k, so we can annotate them and add them into the labeled dataset to decline the uncertainty, thus through iteratively label propagation, data annotating and dataset updating, the pseudo label accuracy will be enhanced comparing to the regular label propagation.

The procedure of the multi granularity based label propagation (see Algorithm 1) algorithm is described as follows: the pseudo labels of unlabeled data are obtained by two classic label propagation algorithms with diverse k, then according to the pseudo labels, the unlabeled dataset are divided into three disjoint subsets, among which we apply three-way decision to choose the unlabeled data that have different pseudo labels and annotate them. Add these annotated data into the initial labeled dataset and update labeled dataset and unlabeled dataset, then iterate the procedures mentioned, namely propagation and annotation, till all the pseudo labels learned by two label propagation algorithms are the same. The iteration processes leverage multiple neighborhood information granules to learn the pseudo labels of unlabeled data.

Theorem 1. *The MGLP algorithm is convergent.*

Proof. First, assume that the probabilistic transition matrix P is partitioned into four block matrices:

$$P = \begin{bmatrix} P_{LL} & P_{LU} \\ P_{UL} & P_{UU} \end{bmatrix}, \tag{11}$$

Algorithm 1 Multi granularity based label propagation (MGLP)

Input: Training data $\mathcal{X} = [\mathcal{X}_L; \mathcal{X}_U] \in R^{n \times d}$, given labels $y_l, l \in \{1, 2, \cdots, l\}$, k-NN numbers $k1$ and $k2$, parameter σ.

Output: Pseudo labels of all the unlabeled data $y_u, u \in \left\{l' + 1, l' + 2, \cdots, n\right\} (l' \geq l)$.

```
Call LP_k1, LP_k2 to obtain Y_U^{LP_k1}, Y_U^{LP_k2} respectively;
while Y_U^{LP_k1} ≠ Y_U^{LP_k2} do
    For every x ∈ X_U
    if y^{LP_k1} ≠ y^{LP_k2} then
        Annotate x;
        X_L = X_L ⋃ {x};
        X_U = X_U − {x};
    end if
    Call LP_k1, LP_k2 to obtain Y_U^{LP_k1}, Y_U^{LP_k2} respectively;
end while
return y_u;
```

the LP_k algorithm is convergent, and the pseudo labels matrix of unlabeled data $\mathcal{Y}_U$ converges to $(\boldsymbol{I} - P_{UU})^{-1} P_{UL} \mathcal{Y}_L$ after propagation (Zhu, 2002).

After that, we prove the convergence of MGLP algorithm, in which, one can achieve two pseudo labels matrices $\mathcal{Y}_U^{LP_{k1}}$ and $\mathcal{Y}_U^{LP_{k2}}$, then manually label the unlabeled data with diverse pseudo labels. Take the worst case into account, if $\forall x \in \mathcal{X}_U$, $y^{LP_{k1}} \neq y^{LP_{k2}}$, all the unlabeled data will be annotated by oracle, and the number of iterations could be the greatest as $|\mathcal{X}_U|$. Obviously, MGLP algorithm is convergent in this scenario, let alone in the situation of $\mathcal{X}_{U1} \bigcup \mathcal{X}_{U2} \neq \emptyset$. □

3.4. Some discussions about MGLP

As the computational complexity of LP_k algorithm is $O(kn^2)$, in the worst case, the computational complexity of MGLP algorithm is $O((k1 + k2)un^2)$, namely, $O(kun^2)$.

Let $|\mathcal{X}_{U1}| = u1$, $|\mathcal{X}_{U2}| = u2$, and $|\mathcal{X}_{U3}| = u3$, the accuracy of pseudo labels acc satisfies the following inequality:

$$\frac{u1}{u1 + u2 + u3} \leq acc \leq \frac{u1 + u3}{u1 + u2 + u3}, \tag{12}$$

then, after the data in subset $|\mathcal{X}_{U3}|$ are annotated and added into labeled dataset, the accuracy of updated pseudo labels will satisfy:

$$\frac{u1'}{u1 + u2} \leq acc' \leq \frac{u1' + u3'}{u1 + u2}. \tag{13}$$

where $u1 + u2 = u1' + u2' + u3'$.

Since

$$\frac{u1'}{u1 + u2} \geq \frac{u1'}{u1 + u2 + u3}, \tag{14}$$

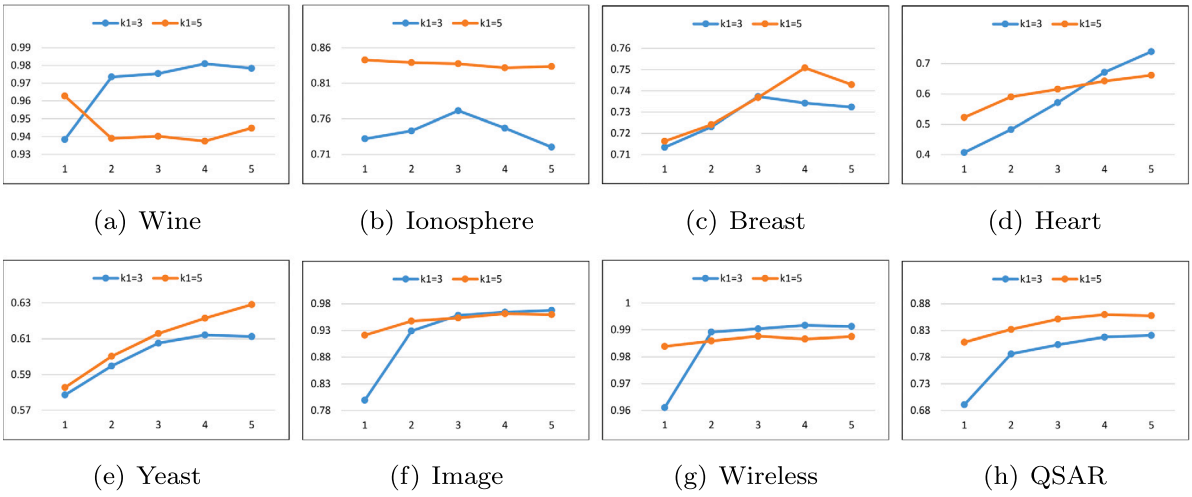


Fig. 5. The accuracy results of diverse groups of $k1$ and $k2$.

Table 2
Summary of datasets.

Name	#Instances	#Attributes	#Classes
Wine	178	13	3
Ionosphere	351	34	2
Breast	286	9	2
Heart	303	14	5
Yeast	1484	8	10
Image	2310	19	7
Wireless	2000	7	4
QSAR	1055	41	2

if $u1' \geq u1 + u3$, it can be deduced that $acc' \geq acc$ holds, namely the pseudo label accuracy of MGLP is definitely higher than that of regular label propagation. However in other cases, clear relationship between acc' and acc cannot be given theoretically.

It is rather intuitive that as the labeled ratio increases, the accuracy of pseudo labels improves, but in fact, this is not always the case, which has been reported experimentally in the literature. Therefore, the data initially chosen to be annotated also play an important role in semi-supervised learning paradigm, and MGLP algorithm provides another way to annotate data.

4. Experiments

In this section, we set up several simulation experiments completed for eight real datasets from the UCI repository[1] to validate the effectiveness of the proposed method MGLP. Specifically, experimental datasets are prepared at first, and then some comparative experiments are conducted.

4.1. Datasets and experiment settings

The eight classification datasets are: Wine, Ionosphere, Breast Cancer (Breast for short), Heart Disease (Heart for short), Yeast, Image Segmentation (Image for short), Wireless Indoor Localization (Wireless for short) and QSAR biodegradation (QSAR for short). They are shown in Table 2, and preprocessed as follows: ① encode the nominal values with dummy variables (one-hot coding), ② normalize every value of attributes with Z-score normalization. All the experiments are implemented in MATLAB on a PC with CPU 2.6 GHz and 8 GB memory.

[1] http://archive.ics.uci.edu/ml

For convenience, in each of the following experiments, the convergence thresholds $\max_t$ and $\min_\varepsilon$ are set to be 100000 and 1e-10 respectively, and Euclidean distance-based similarities are adopted and the weight matrix W is defined by zero–one weighting (see (2)). Moreover, to maintain the distribution information of the original data, we assume the instances of different class in the labeled data subset are of the same proportion as the instances of different class in the original dataset.

4.2. Impact of information granularity

As the neighborhood size parameter k presents the granularity of an information granule, in this section, we explore the impact of information granularity on pseudo label accuracy of unlabeled data by setting different groups of $k1$ and $k2$. The process of experiment is as below: firstly, specify the labeled ratio of dataset to be 0.1, and select the initial labeled data randomly. Then, MGLP algorithm is applied to the dataset to get the pseudo labels according to $k1$, $k2$ and other parameters provided in Section 4.1. Finally, calculate the pseudo label accuracy to measure the performance of various level granularity. We repeat each experiment ten times and obtain the average values of pseudo label accuracies, as shown in Tables 3 and 4, where the bold ones are the greatest of the dataset.

In Table 3, the value of $k1$ is fixed as 3, and the value of $k2$ varies ($k2 = 5, 7, 9, 11, 13$). The results show that the pseudo label accuracies of dataset Heart, Image and QSAR change greatly, especially in Heart, from 0.4067 as the lowest to 0.7393 as the highest. However, in the other five datasets, the changes are mild comparatively. For Table 4, the value of $k1$ is fixed to 5, and as the values of $k2$ altering ($k2 = 7, 9, 11, 13, 15$), the pseudo label accuracies of dataset Heart change a lot, and values in other seven datasets change mildly. In sum, the pseudo labels of unlabeled data in dataset Heart are influenced enormously by information granularity, and the impact of information granularity to dataset Wireless is limited.

We can also see the impact of the ten groups of $k1$ and $k2$ on pseudo label accuracy for each dataset in Fig. 5, where 1, 2, 3, 4, 5 refer to the value of $k2$, namely, 5, 7, 9, 11, 13 when $k1 = 3$ and 7, 9, 11, 13, 15 when $k1 = 5$ respectively. The polygonal lines fluctuate slightly when $k1 = 5$ in dataset Ionosphere, Image and Wireless, while in other circumstances, the tendency of lines may be affected significantly by the values of k.

So, it can be concluded that the impact of information granularity is different for diverse datasets, and the values of $k1$ and $k2$ need to be selected based on specific datasets in real applications.

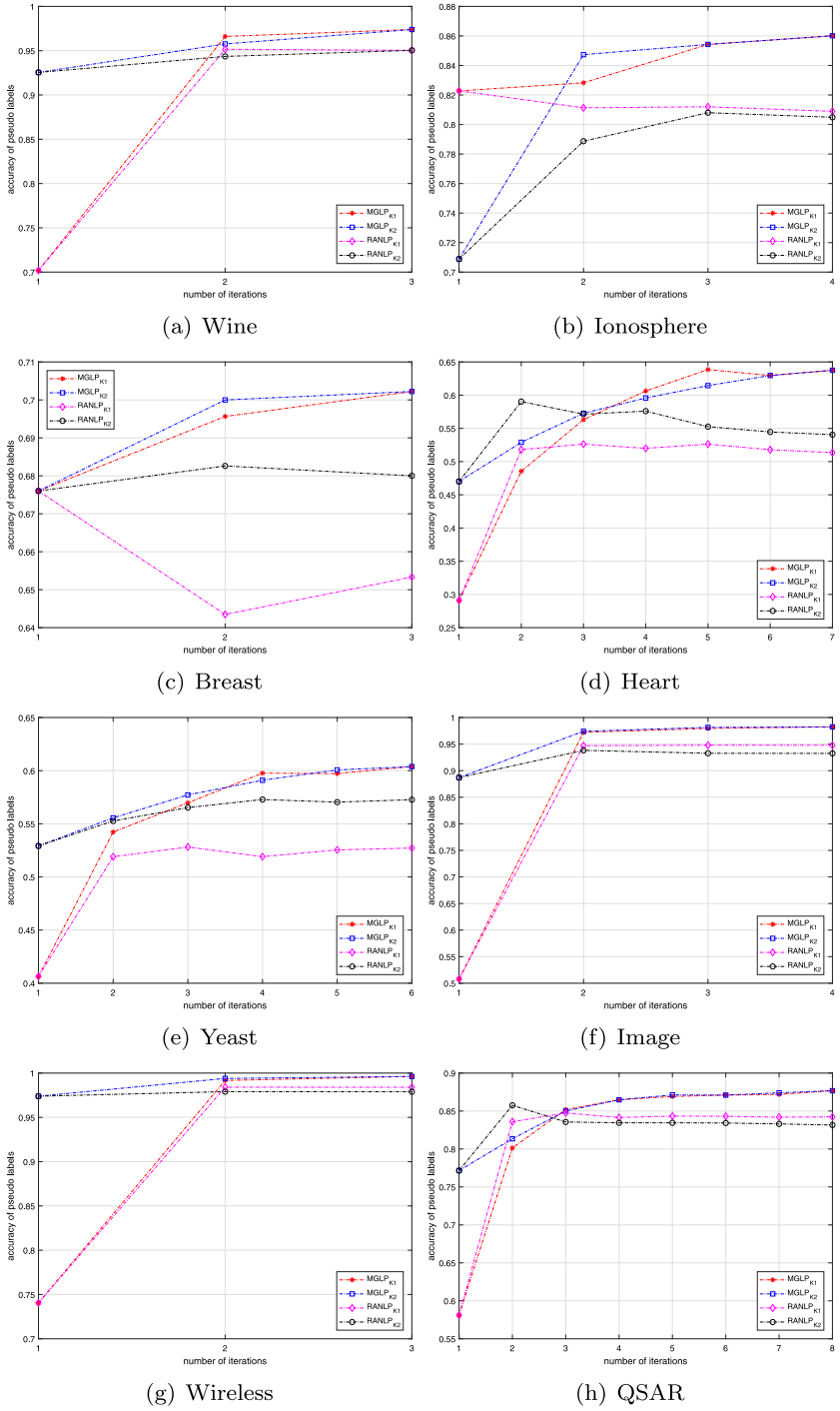


Fig. 6. Comparison of MGLP and classic label propagation.

Table 3
The accuracy results of pseudo labels on the eight datasets ($k1 = 3$).

Dataset	Accuracy				
	$k2 = 5$	$k2 = 7$	$k2 = 9$	$k2 = 11$	$k2 = 13$
Wine	0.9384 ± 0.0822	0.9735 ± 0.0096	0.9754 ± 0.0099	$\mathbf{0.981 \pm 0.0066}$	0.9784 ± 0.0079
Ionosphere	0.732 ± 0.0704	0.7431 ± 0.0957	$\mathbf{0.7713 \pm 0.1019}$	0.747 ± 0.0974	0.7203 ± 0.0905
Breast	0.7134 ± 0.0292	0.7231 ± 0.0363	$\mathbf{0.7373 \pm 0.0416}$	0.7342 ± 0.0397	0.7323 ± 0.032
Heart	0.4067 ± 0.0589	0.4825 ± 0.089	0.5713 ± 0.0854	0.6714 ± 0.0697	$\mathbf{0.7393 \pm 0.0468}$
Yeast	0.5786 ± 0.0181	0.5948 ± 0.0149	0.6076 ± 0.0173	$\mathbf{0.612 \pm 0.0234}$	0.6112 ± 0.0169
Image	0.7993 ± 0.0273	0.9288 ± 0.0315	0.9583 ± 0.0331	0.9642 ± 0.0232	$\mathbf{0.9676 \pm 0.0185}$
Wireless	0.961 ± 0.0139	0.9892 ± 0.0045	0.9904 ± 0.0029	$\mathbf{0.9917 \pm 0.0028}$	0.9913 ± 0.0029
QSAR	0.691 ± 0.031	0.7862 ± 0.0411	0.8035 ± 0.0286	0.8176 ± 0.0248	$\mathbf{0.8205 \pm 0.032}$

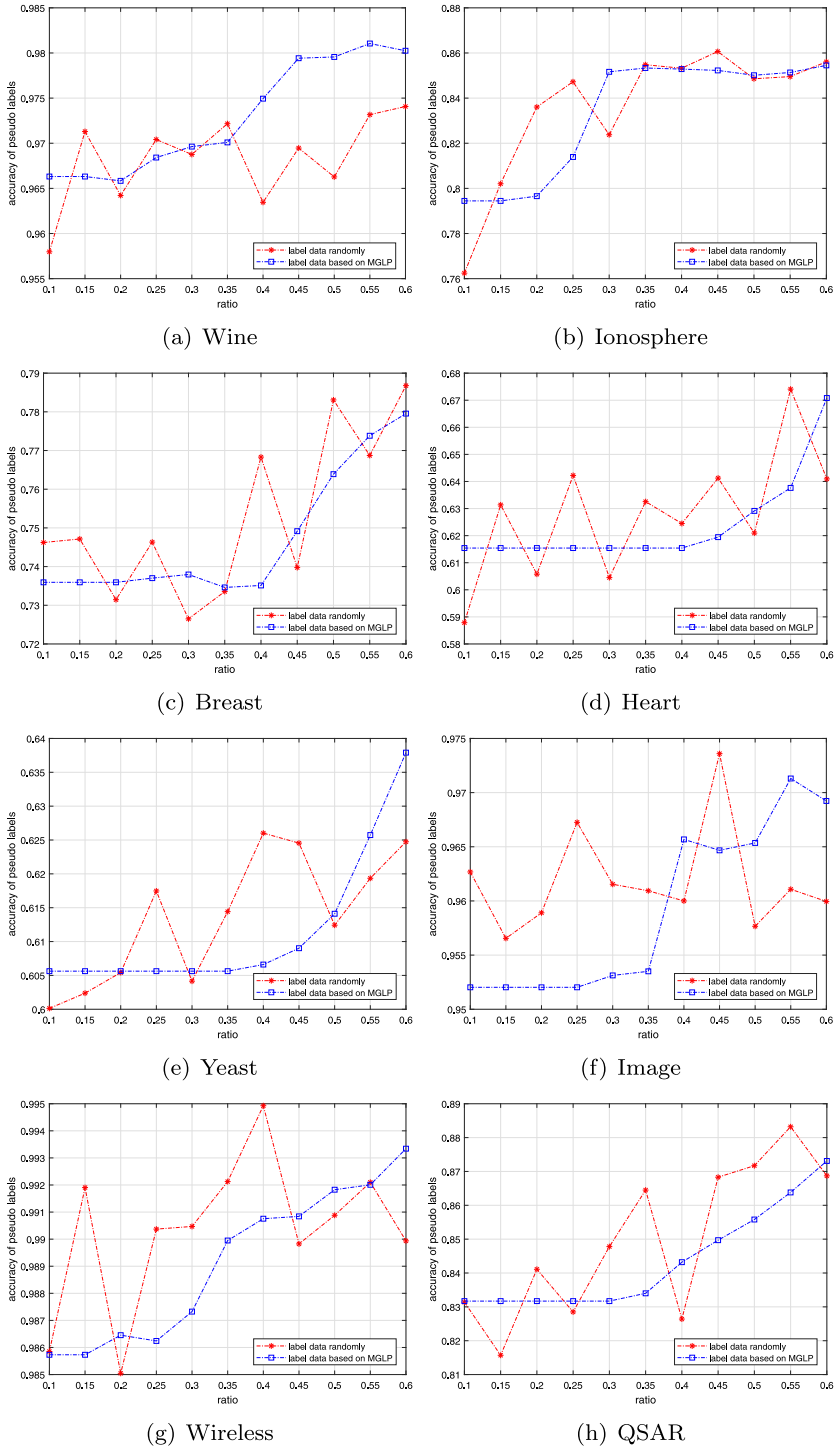


Fig. 7. Impact of labeled ratio on pseudo label accuracy.

4.3. Impact of data chosen strategy

Here, several experiments are conducted to illustrate the superiority of the proposed method MGLP compared to the classic label propagation algorithm. In particular, we scrutinize the impact of strategy to choose, annotate and add unlabeled data into labeled subset. In MGLP algorithm, after propagating, the initially unlabeled data in $\mathcal{X}_{U3}$, will be annotated and then added into the labeled dataset, actually, the labeled ratio increases in this procedure. So, we annotate the same number of unlabeled data and add them into the labeled dataset in each iteration of classic label propagation algorithm, but these data are selected randomly from the initially unlabeled dataset contrasting to finding out $\mathcal{X}_{U3}$ based on the three-way decision strategy in MGLP.

In the experiment of this section, the initial labeled ratio is also 0.1, while let the neighborhood parameters be $k1 = 3$ and $k2 = 11$. Each experiment is executed ten times and the average pseudo label accuracies of eight datasets are shown in Fig. 6, where MGLP_{K1} and MGLP_{K2} refer to annotating unlabeled data according to MGLP algorithm, however, RANLP_{K1} and RANLP_{K2} refer to annotating the same number of unlabeled data randomly.

Table 4
The accuracy results of pseudo labels on the eight datasets ($k1 = 5$).

Dataset	Accuracy				
	$k2 = 7$	$k2 = 9$	$k2 = 11$	$k2 = 13$	$k2 = 15$
Wine	$\mathbf{0.9628 \pm 0.0107}$	0.939 ± 0.0903	0.9402 ± 0.091	0.9375 ± 0.089	0.9448 ± 0.0926
Ionosphere	$\mathbf{0.8426 \pm 0.06}$	0.8392 ± 0.0656	0.8374 ± 0.0655	0.8318 ± 0.0638	0.8337 ± 0.0637
Breast	0.7163 ± 0.0278	0.7241 ± 0.0274	0.7368 ± 0.0384	$\mathbf{0.7508 \pm 0.0299}$	0.7429 ± 0.0328
Heart	0.5224 ± 0.0754	0.5903 ± 0.0883	0.6154 ± 0.0963	0.6423 ± 0.0717	$\mathbf{0.6613 \pm 0.0774}$
Yeast	0.5828 ± 0.013	0.6002 ± 0.0145	0.6129 ± 0.0192	0.6215 ± 0.0278	$\mathbf{0.6291 \pm 0.0224}$
Image	0.9208 ± 0.0081	0.9473 ± 0.0071	0.9535 ± 0.0094	$\mathbf{0.9614 \pm 0.0032}$	0.9596 ± 0.0042
Wireless	0.9839 ± 0.0036	0.9859 ± 0.0021	$\mathbf{0.9877 \pm 0.0025}$	0.9866 ± 0.0029	0.9875 ± 0.0018
QSAR	0.8079 ± 0.0223	0.8318 ± 0.0221	0.8513 ± 0.0245	$\mathbf{0.8596 \pm 0.0226}$	0.8574 ± 0.015

From Fig. 6, we can see the following phenomenons: ① All the pseudo label accuracies of MGLP_{K1} and MGLP_{K2} increase monotonically as the number of iterations rises, except for the sixth iteration of MGLP_{K1} in dataset Heart. ② The results of RANLP_{K1} and RANLP_{K2} decrease in many situations, such as the second iteration of RANLP_{K1} in dataset Ionosphere, the second iteration of RANLP_{K1} in dataset Breast, and the third iteration of RANLP_{K2} in dataset Heart etc. ③ When MGLP converges, the pseudo label accuracies are all higher than the results of regular label propagation (the values of the first iteration), and they are also higher than the results of regular label propagation with randomly annotating unlabeled data (RANLP_{K1} and RANLP_{K2}). ④ Evidently, there exist some monotonical decrease cases of annotating unlabeled data randomly, for example, from the second to the fourth iteration of RANLP_{K1} in dataset Ionosphere, from the second to the seventh iteration of RANLP_{K2} in dataset Heart. ⑤ Also, some vibration phenomena appear in annotating unlabeled data randomly, such as from the second to the seventh iteration of RANLP_{K1} in dataset Heart, and from the second to the sixth iteration of RANLP_{K1} in dataset Yeast.

What is shown in Fig. 6 illustrates that the proposed is superior than the classic label propagation algorithm from the accuracy perspective. The pseudo label accuracies of MGLP increase ordinarily in the procedure of iteration, however they may vibrate, or even decrease when annotating unlabeled data randomly. Maybe bridge nodes or outliers should be responsible. If the labels of these data are propagated, the labels of their neighbors may be wrong, and as the iterations go, the wrong labels will propagate cumulatively to affect the accuracies of all the dataset. It also proves that as the number of labeled data increases, the accuracy of pseudo labels does not necessarily increase. So, the three-way decision strategy in MGLP is important to select data to be labeled compared with random selection.

4.4. Impact of labeled ratio

To verify the relationship of the pseudo label accuracies of MGLP and labeled ratio, some experiments are conducted here. The labeled ratio increases from 0.1 to 0.6, and the compartment is 0.05. We compare two kinds of results: the first is based on the data labeled randomly according to the given ratio, and the second is based on the data annotated by MGLP. In detail, the annotating process is as follows: firstly, annotate initial data with the ratio 0.1 randomly and execute MGLP, next, compare the number of unlabeled data needed to be annotated (and will be added into the labeled dataset) in MGLP and the target number of labeled data. If the former is bigger, annotate all the selected data by MGLP and quit the iteration procedure. If the latter is bigger, annotate all the selected data by MGLP and the differential number of data randomly.

The algorithms are evaluated 100 times, and the average values of pseudo label accuracies are shown in Fig. 7. On the whole, the curves in the figure show that pseudo label accuracies of MGLP are increased as the labeled ratio rises. Specifically, they are strict increased in dataset Heart, Yeast and QSAR. Furthermore, they decrease occasionally in the other five datasets, for instance, when ratio=0.2 in dataset Wine, ratio=0.5 in dataset Ionosphere, ratio=0.35 in dataset Breast, ratio=0.45 and ratio=0.6 in dataset Image, ratio=0.25 in dataset Wireless, however, the changes are small. Contrast to the relatively smooth curves of MGLP, the curves of random data labeling according to the given ratio jitter greatly, and there are no overt change tendencies. Some definite changes are: ratio=0.35 to ratio=0.4 in dataset Wine, ratio=0.4 to ratio=0.45 in dataset Breast, ratio=0.45 to ratio=0.5 in dataset Image, etc.

From the results described above, we can see that if we label data randomly according to the labeled ratio, the accuracy of pseudo labels may be likely to decrease or fluctuate as the labeled ratio increases. But the accuracy of pseudo labels in MGLP is more stable, and increases normally. The conclusion exactly coincides with the one in Section 4.3.

As a result, based on the guidelines of multiple level granularity and active learning strategy of three-way decision, the propose algorithm MGLP is not only an effective means to obtain pseudo labels with higher accuracy of unlabeled data, but it also can be viewed as a novel way to annotate data through the iteration procedure and we can apply it in real applications.

Table 5
Parameters for different methods.

Method	Parameters
MGLP	$k1 = 3,\ k2 = 11$
LP	$k = 3$
LNP	$k = 3,\ \alpha = 0.8$
LPSN	$\epsilon = 1e-4$
Adaptive-NP	$\alpha = 1e-5,\ \beta = 1e-5$, iteration=10
CRLP	$B = 100,\ \alpha = 0.8$

4.5. Comparisons with some typical methods

In addition, we also conduct some experiments to verify the performance of MGLP comparing to some typical methods, including LP (classic label propagation) (Zhu, 2002), LNP (linear neighborhood propagation) (Wang & Zhang, 2008), LPSN (label propagation through sparse neighborhood) (Zang & Zhang, 2012), Adaptive-NP (adaptive neighborhood propagation) (Jia et al., 2016) and CRLP (consensus rate-based label propagation) (Yu & Kim, 2018).

Besides the eight datasets mentioned in Table 2, two real image datasets are added here. The COIL20 database has 1440 gray object images for 20 different subjects, and the size of each image is 32×32 pixels. There are 9298 images with 16×16 grayscale pixels in handwritten database USPS.

The initial labeled ratio is 0.1, and the other parameters of the six methods are shown in Table 5. In MGLP, the values of $k1$ and $k2$ are the same as in Section 4.3. In CRLP, B is the number of random subspaces, α is the coefficient to control the tradeoff between spreading and retaining, and they are set referring to Yu and Kim (2018). In LNP, α is the same as in CRLP. The values of k in LP and LNP are set to 3 as $k1$ in MGLP. In LPSN, the radius of sparse neighborhood ϵ is fixed at 1e-4 as Zang and Zhang (2012). In Adaptive-NP, the trade-off parameters α and β are set to be 1e-5, and the number of iterations is 10 according to Jia et al. (2016).

Table 6
Mean pseudo label accuracies of different methods.

Dataset	MGLP	LP	LNP	LPSN	Adaptive-NP	CRLP
Wine	**0.9754 ± 0.0113**	0.8918 ± 0.0136	0.9001 ± 0.0153	0.8955 ± 0.0141	0.9199 ± 0.0126	0.9079 ± 0.0090
Ionosphere	0.7749 ± 0.0194	0.7733 ± 0.0195	0.7736 ± 0.0225	0.7684 ± 0.0326	**0.7765 ± 0.0184**	0.7693 ± 0.0085
Breast	**0.7411 ± 0.0119**	0.6796 ± 0.0106	0.6848 ± 0.0128	0.6765 ± 0.0057	0.7095 ± 0.0162	0.6884 ± 0.0114
Heart	**0.6281 ± 0.0180**	0.5118 ± 0.0139	0.5641 ± 0.0122	0.5137 ± 0.0364	0.5646 ± 0.0241	0.5267 ± 0.0265
Yeast	**0.6132 ± 0.0063**	0.5135 ± 0.0063	0.5331 ± 0.0056	0.5496 ± 0.0057	0.5529 ± 0.0039	0.5116 ± 0.0053
Image	**0.9681 ± 0.0039**	0.9025 ± 0.0042	0.9038 ± 0.0040	0.9247 ± 0.0048	0.9403 ± 0.0041	0.9389 ± 0.0037
Wireless	**0.9822 ± 0.0006**	0.9214 ± 0.0043	0.9238 ± 0.0039	0.9318 ± 0.0012	0.9348 ± 0.0042	0.9325 ± 0.0044
QSAR	0.8196 ± 0.0110	0.7726 ± 0.0068	0.7857 ± 0.0067	0.7843 ± 0.0081	**0.8243 ± 0.0153**	0.7709 ± 0.0307
COIL20	**0.9103 ± 0.0432**	0.8423 ± 0.0328	0.8649 ± 0.0421	0.8917 ± 0.0427	0.8786 ± 0.0532	0.8332 ± 0.0395
USPS	**0.8932 ± 0.0576**	0.8359 ± 0.0412	0.8632 ± 0.0645	0.8664 ± 0.0448	0.8421 ± 0.0643	0.8217 ± 0.0611
average	**0.8306 ± 0.0183**	0.7645 ± 0.0153	0.7797 ± 0.0190	0.7803 ± 0.0196	0.7944 ± 0.0216	0.7701 ± 0.0200

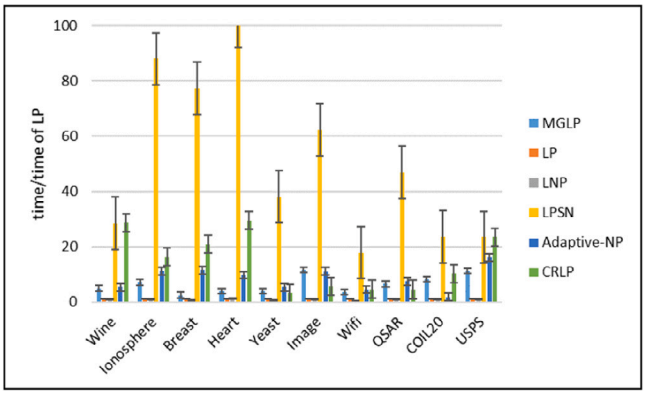


Fig. 8. Execution time ratio of different methods to LP.

Since the labeled ratio will increase in MGLP, we augment the labeled subset in other five methods by selecting unlabeled data randomly to the same labeled ratio as MGLP. Every method is evaluated 20 times, and the mean pseudo label accuracies are shown in Table 6.

It can be observed that the pseudo label accuracies of MGLP are definitely higher than the other five methods on all the ten datasets except for Ionosphere and QSAR. Especially in dataset Heart and Yeast, the increased accuracies are more than 10 percent. On average, the accuracy of MGLP rise by 8.6% than LP, 6.5% than LNP, 6.4% than LPSN, 4.5% than Adaptive-NP and 7.8% than CRLP.

Fig. 8 summarizes the ratio of the methods' execution time to the time of LP. As it can be seen, the computational cost of MGLP is higher than that of LP and LNP in all the ten datasets, lower than that of CRLP, Adaptive-NP and LPSN in most of the ten datasets.

The experimental results demonstrate that the proposed algorithm MGLP is effective, and outperforms other methods on most of the ten benchmark datasets with higher computational complexity. It can be supposed that the proposed method may reveal intrinsic data structure more effectively by employing multi neighborhood granules and active learning with three-way decision.

5. Conclusions

In this paper, we propose a novel label propagation algorithm called multi granularity based label propagation (MGLP) to determine the pseudo labels of unlabeled data. In this method, the concept of multiple level granularity is adopted to set neighborhood size, and three-way decision works as a active learning strategy to choose unlabeled data. Specific procedure and some discussions of the proposed method are given. Experimental results show that through the iterative process of propagating labels, annotating particular unlabeled data and adding them into labeled dataset, the pseudo label accuracy will be improved in comparison with the random data labeling and some other typical methods. It is also demonstrated that we can apply the proposed method to annotate data. Therefore, MGLP is effective.

Since the graphs used in MGLP are only k-NN graphs in this work, we will investigate the feasibility of applying different graphs in this model to make MGLP as a more general framework. In addition, the multi granularity based label propagation evaluated in this work is transductive, for future work, we can focus on extending the model to out-of-sample scenario. There often exist big data, real-time streaming data and class-imbalance data in many practical applications, so in the future, it will also be interesting to investigate the effectiveness and modifications of MGLP to these kinds of data.

CRediT authorship contribution statement

Shengdan Hu: Writing – original draft, Writing – review & editing, Conceptualization, Formal analysis, Methodology, Data curation, Validation. **Duoqian Miao:** Conceptualization, Formal analysis, Supervision, Writing – review, Funding acquisition. **Witold Pedrycz:** Conceptualization, Formal analysis, Supervision, Writing – review.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors thank both the editors and the anonymous referees for their valuable suggestions, which substantially improved this paper. This work is supported by the National Natural Science Foundation of China (Grant No. 61976158, 61673301, 62076182).

All authors reviewed the manuscript critically for scientific content, and all authors gave final approval of the manuscript for publication.

References

Belkin, M., Niyogi, P., & Sindhwani, V. (2006). Manifold regularization: a geometric framework for learning from labeled and unlabeled examples. *Journal of Machine Learning Research*, *7*, 2399–2434.

Chapelle, O., Schölkopf, B., & Zien, A. (2006). Label propagation and quadratic criterion. In *Semi-supervised learning* (pp. 193–216). MIT Press.

Chen, J., Ji, D., Tan, C., & Niu, Z. (2006). Relation extraction using label propagation based semi-supervised learning. In *Proceeding of the 21st international conference on computational linguistics and the 44th annual meeting of the association for computational linguistics* (pp. 129–136).

Cheng, B., Yang, J., Yan, S., Fu, Y., & Huang, T. (2010). Learning with l1-graph for image analysis. *IEEE Transactions on Image Processing*, *19*, 858–866.

Dornaika, F., Wang, K., Arganda-Carreras, I., Elorza, A., & Moujahid, A. (2020). Toward graph-based semi-supervised face beauty prediction. *Expert Systems with Applications*, *142*, Article 112990.

Fan, M., Zhang, X., Du, L., Chen, L., & Tao, D. (2018). Semi-supervised learning through label propagation on geodesics. *IEEE Transactions on Cybernetics*, *48*, 1486–1499.

Francisquini, R., Rosset, V., & Nascimento, M. C. (2017). GA-LP: A genetic algorithm based on label propagation to detect communities in directed networks. *Expert Systems with Applications*, *74*, 127–138.

Gan, H., Li, Z., Wu, W., Luo, Z., & Huang, R. (2018). Safety-aware graph-based semi-supervised learning. *Expert Systems with Applications*, *107*, 243–254.

Giasemidis, G., Kaplis, N., Agrafiotis, I., & Nurse, J. (2020). A semi-supervised approach to message stance classification. *IEEE Transactions on Knowledge and Data Engineering*, *32*, 1–11.

Hong, D., Yokoya, N., Chanussot, J., Xu, J., & Zhu, X. (2019). Learning to propagate labels on graphs: an iterative multitask regression framework for semi-supervised hyperspectral dimensionality reduction. *ISPRS Journal of Photogrammetry and Remote Sensing*, *158*, 35–49.

Hu, S., Miao, D., Zhang, Z., Luo, S., & Zhang, Y. (2018). A test cost sensitive heuristic attribute reduction algorithm for partially labeled data. In *International joint conference on rough sets* (pp. 257–269).

Huang, R., Feng, W., Wang, Z., Xing, Y., & Zou, Y. (2020). Exemplar-based image saliency and co-saliency detection. *Neurocomputing*, *371*, 147–157.

Jia, L., Zhang, Z., Wang, L., Jiang, W., & M.B., Z. (2016). Adaptive neighborhood propagation by joint l2, 1-norm regularized sparse coding for representation and classification. In *IEEE international conference on data mining* (pp. 201–210).

Li, J., Jing, M., Lu, K., Zhu, L., & Shen, H. (2019). Locality preserving joint transfer for domain adaptation. *IEEE Transactions on Image Processing*, *28*, 6103–6115.

Ouyang, T., Pedrycz, W., & Pizzi, N. J. (2019). Record linkage based on a three-way decision with the use of granular descriptors. *Expert Systems with Applications*, *122*, 16–26.

Pedrycz, W. (2013). *Granular computing: analysis and design of intelligent systems*. CRC Press.

Pedrycz, W., & Homenda, W. (2013). Building the fundamentals of granular computing: a principle of justifiable granularity. *Applied Soft Computing*, *13*, 4209–4218.

Pedrycz, W., & Vukovich, G. (2002). Granular computing with shadowed sets. *International Journal of Intelligent Systems*, *17*, 173–197.

Qian, J., Liu, C., Miao, D., & Yue, X. (2020). Sequential three-way decisions via multi-granularity. *Information Sciences*, *507*, 606–629.

Settles, B. (2010). Active learning literature survey. *Technical Report*, University of Wisconsin.

Sun, M., Hao, S., & Liu, G. (2018). Semi-supervised vehicle classification via fusing affinity matrices. *Signal Processing*, *149*, 118–123.

Wang, F., & Zhang, C. (2008). Label propagation through linear neighborhoods. *IEEE Transactions on Knowledge and Data Engineering*, *20*, 55–67.

Wu, W., Leung, Y., & Mi, J. (2009). Granular computing and knowledge reduction in formal contexts. *IEEE Transactions on Knowledge and Data Engineering*, *21*, 1461–1474.

Xu, J., Wang, G., Li, T., & Pedrycz, W. (2018). Local-density-based optimal granulation and manifold information granule description. *IEEE Transactions on Cybernetics*, *48*, 2795–2808.

Yao, Y. (1999). Rough sets, neighborhood systems, and granular computing. In *Proceeding of the 1999 IEEE canadian conference on electrical and computer engineering* (pp. 1553–1558).

Yao, Y. (2020). Three-way granular computing, rough sets, and formal concept analysis. *International Journal of Approximate Reasoning*, *116*, 106–125.

Yao, Y. (2020). Tri-level thinking: models of three-way decision. *International Journal of Machine Learning and Cybernetics*, *11*, 947–959.

Yu, J., & Kim, S. B. (2018). Consensus rate-based label propagation for semi-supervised classification. *Information Sciences*, *465*, 265–284.

Yu, E., Sun, J., Li, J., Chang, X., Han, X., & Hauptmann, A. (2019). Adaptive semi-supervised feature selection for cross-modal retrieval. *IEEE Transactions on Multimedia*, *21*, 1276–1288.

Zadeh, L. (1997). Towards a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic. *Fuzzy Sets and Systems*, *90*, 111–117.

Zang, F., & Zhang, J. (2012). Label propagation through sparse neighborhood and its applications. *Neurocomputing*, *97*, 267–277.

Zhang, Z., Jia, L., Zhao, M., Liu, G., Wang, M., & Yan, S. (2019). Kernel-induced label propagation by mapping for semi-supervised classification. *IEEE Transactions on Big Data*, *5*, 148–165.

Zhang, Z., Li, F., Jia, L., Qin, J., Zhang, L., & Yan, S. (2018). Robust adaptive embedded label propagation with weight learning for inductive classification. *IEEE Transactions on Neural Networks and Learning Systems*, *29*, 3388–3403.

Zhang, X., & Miao, D. (2016). Quantitative/qualitative region-change uncertainty/certainty in attribute reduction: comparative region-change analyses based on granular computing. *Information Sciences*, *334–335*, 174–204.

Zhang, H., Zhang, Z., Zhao, M., Ye, Q., Zhang, M., & Wang, M. (2020). Robust triple-matrix-recovery-based auto-weighted label propagation for classification. *IEEE Transactions on Neural Networks and Learning Systems*, *31*, 4538–4552.

Zhang, Z., Zhao, M., & Chow, T. (2015). Graph based constrained semi-supervised learning framework via label propagation over adaptive neighborhood. *IEEE Transactions on Knowledge and Data Engineering*, *27*, 2362–2376.

Zhao, M., Zhang, Y., Zhang, Z., Liu, J., & Kong, W. (2019). Alg: adaptive low-rank graph regularization for scalable semi-supervised and unsupervised learning. *Neurocomputing*, *370*, 16–27.

Zhou, Z. (2018). A brief introduction to weakly supervised learning. *National Science Review*, *1*, 48–57.

Zhou, D., Bousquet, O., Lal, T., Weston, J., & Scholkopf, B. (2004). Learning with local and global consistency. *Neural Information Processing Systems*, *16*, 321–328.

Zhou, J., Lai, Z., Miao, D., Gao, C., & Yue, X. (2020). Multigranulation rough-fuzzy clustering based on shadowed sets. *Information Sciences*, *507*, 553–573.

Zhou, K., Martin, A., Pan, Q., & Liu, Z. (2018). SELP: Semi-supervised evidential label propagation algorithm for graph data clustering. *International Journal of Approximate Reasoning*, *92*, 139–154.

Zhu, X. (2002). Learning from labeled and unlabeled data with label propagation. *Technical Report*, Carnegie Mellon University.

Zhu, X. (2008). Semi-supervised learning literature survey. *Technical Report*, Department of Computer Sciences, University of Wisconsin at Madison.

Zhu, X., Ghahramani, Z., & Lafferty, J. D. (2003). Semi-supervised learning using Gaussian fields and harmonic functions. In *Proceedings of the twentieth international conference (ICML 2003)* (pp. 21–24).

Zoidi, O., Tefas, A., Nikolaidis, N., & Pitas, I. (2018). Positive and negative label propagation. *IEEE Transactions on Circuits and Systems for Video Technology*, *28*, 342–355.