



HAL
open science

Evidential clustering based on transfer learning

Kuang Zhou, Mei Guo, Arnaud Martin

► **To cite this version:**

Kuang Zhou, Mei Guo, Arnaud Martin. Evidential clustering based on transfer learning. International Conference on Belief Functions, Oct 2021, Shanghai, China. hal-03405161

HAL Id: hal-03405161

<https://hal.science/hal-03405161v1>

Submitted on 27 Oct 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Evidential clustering based on transfer learning

Kuang Zhou¹, Mei Guo¹, and Arnaud Martin²

¹ School of Mathematics and Statistics, Northwestern Polytechnical University,
Xi'an, Shaanxi 710072, PR China

² Univ Rennes, CNRS, IRISA, DRUID, Rue E. Branly, 22300 Lannion, France
kzhoumath@nwpu.edu.cn, guomei237@163.com, Arnaud.Martin@univ-rennes1.fr

Abstract. Clustering is an essential part of data mining, which can be used to organize data into sensible groups. Among the various clustering algorithms, the prototype-based methods have been most popularly applied due to the easy implementation, simplicity and efficiency. However, most of them such as the c -means clustering are no longer effective when the data is insufficient and uncertain. While the data for the current clustering task may be sparse, there is usually some useful knowledge available in the related scenes. Transfer learning can be adopted to address such cross domain learning problems by using information from data in a related domain and transferring that data/knowledge to the target task. The inconsistency between different domains can increase the uncertainty in the data. To handle the insufficiency and uncertainty problems in the clustering task simultaneously, a prototype-based evidential transfer clustering algorithm, named transfer evidential c -means (TECM), is introduced in the framework of belief functions. The proposed algorithm employs the cluster prototypes of the source data as references to guide the clustering process of the target data. The experimental studies are presented to demonstrate the advantages of TECM in both synthetic and real-world data sets.

Keywords: Belief functions · Clustering · Transfer learning · Uncertainty · Source domain.

1 Introduction

Clustering is an unsupervised technique aiming to classify patterns into groups [10, 6]. It has been widely used in many fields such as image segmentation, market research and data analysis. Traditional clustering methods, such as c -means, usually work well when the data are sufficient. However, in real world, uncertain and noisy data are omnipresent. Moreover, sometimes we can not get enough data to train a fine clustering model. To address the problems of a lack of information and data impurity, several advanced cluster models have been developed, such as semi-supervised learning [1], multi-view clustering [7], transfer learning [2, 4] and so on.

Transfer learning can learn an effective model for the target domain by effectively leveraging useful information from the source domain [2]. Fig.1 illustrates

a situation where transfer learning is useful. As we can see, it is difficult to obtain an ideal partition for the target data (the left figure) as they are too sparse. However, if information from the source domain (the right figure) is considered, more promising clustering results can be expected. In general, two kinds of information can be transferred from the source to the target domain: raw data or knowledge [9]. Due to the necessity of privacy protection in some applications, such as users' personal information, the original raw data in the source domain are not always accessible. Thus to employ some advanced knowledge from the source domain instead of raw data is more practical. For example, in the clustering task, the cluster prototypes of the source data (red triangles in the right figure) can be regarded as good references for the target domain.

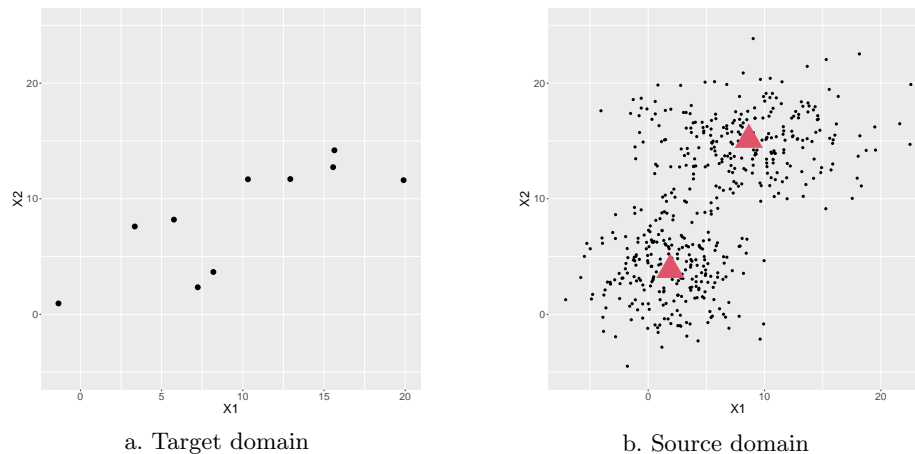


Figure 1. An example where transfer learning is required for the clustering task.

The available knowledge in the source domain can help us improve the cluster model, but the inconsistency between information from the two domains may increase the uncertainty in the data. The theory of belief functions is an efficient mathematical tool for uncertain information representation and fusion. The concept of credal partitions defined in the framework of belief functions is first proposed by Dencœux *et.al.* [3] to deal with the uncertain cluster structure, and following many evidential clustering methods have been designed and widely applied [8, 5]. In this paper, we combine the idea of evidential clustering and transfer learning to develop a new clustering method, named transfer evidential *c*-means (TECM), for insufficient and uncertain data. It first identifies cluster prototypes based on the source domain, which are then transferred into the target domain to guide the clustering procedure. The experimental results on generated and UCI data show the effectiveness of the proposed method.

The remainder of this paper is organized as follows. The proposed TECM algorithm is presented in detail in Section 2. Numerical experiments are conducted in Section 3. Conclusions are drawn in the final section.

2 Transfer evidential c -means

Inspired by the idea of evidential clustering and transfer learning, in this section we will introduce the transfer evidential c -means (TECM) clustering algorithm.

Denote the n data samples in the target domain by $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$ and assume that there are c clusters. The frame of discernment is $\Omega = \{\omega_1, \dots, \omega_c\}$. The available supervised knowledge in a related domain is represented by prototypes $\mathbf{V}^{(s)} = \{\mathbf{v}_1^{(s)}, \mathbf{v}_2^{(s)}, \dots, \mathbf{v}_c^{(s)}\}$. The superscript (s) indicates that the prototypes are from the source domain. The objective function of TECM and the optimization approach will be introduced in the following.

2.1 The objective function of TECM

As an evidential clustering method in the framework of belief functions, TECM aims to look for the optimal credal partition $\mathbf{M} = (\mathbf{m}_1, \dots, \mathbf{m}_n) \in \mathbb{R}^{n \times 2^c}$ and the matrix $\mathbf{V} = (\mathbf{v}_1, \dots, \mathbf{v}_c)$ of size $(c \times p)$ of cluster centers in the target data by minimizing the following objective function:

$$J_{\text{TECM}}(\mathbf{M}, \mathbf{V}) = \sum_{i=1}^n \sum_{\substack{A_j \subseteq \Omega \\ A_j \neq \emptyset}} c_j^\alpha m_{ij}^\beta d_{ij}^2 + \sum_{i=1}^n \delta^2 m_{i\emptyset}^\beta + \beta_1 \left[\sum_{i=1}^n \sum_{\substack{A_j \subseteq \Omega \\ A_j \neq \emptyset}} c_j^\alpha m_{ij}^\beta d_{ij}^{2(s)} + \sum_{i=1}^n \delta^2 m_{i\emptyset}^\beta \right] + \beta_2 \sum_{k=1}^c \|\mathbf{v}_k^{(s)} - \mathbf{v}_k\|^2, \quad (1)$$

subject to:

$$\sum_{A_j \subseteq \Omega, A_j \neq \emptyset} m_{ij} + m_{i\emptyset} = 1, \quad (2)$$

where m_{ij} denotes $m_i(A_j)$ and $m_{i\emptyset}$ denotes $m_i(\emptyset)$. $c_j = |A_j|$ denotes the cardinal of A_j . d_{ik} denotes the distance between \mathbf{x}_i and the barycenter (prototype, denoted by $\bar{\mathbf{v}}_k$) associated with A_k :

$$d_{ik}^2 = \|\mathbf{x}_i - \bar{\mathbf{v}}_k\|^2, \quad (3)$$

where prototype $\bar{\mathbf{v}}_k$ can be defined mathematically by:

$$\bar{\mathbf{v}}_k = \frac{1}{c_k} \sum_{h=1}^c s_{hk} \mathbf{v}_h, \quad \text{with } s_{hk} = \begin{cases} 1 & \text{if } \omega_h \in A_k \\ 0 & \text{else} \end{cases}. \quad (4)$$

Notation \mathbf{v}_h denotes the center of samples in cluster ω_h . Parameters α , β and δ control the degree of penalization for imprecise classes with high cardinality,

the fuzziness of the partition, and the amount of data considered as outliers respectively. These parameters have the same meaning as those in ECM [8].

The objective functions in Eq.(1) has four terms. The first two terms are directly inherited from ECM, which are mainly used to learn from the target data. The third and fourth terms enable the model to learn with the knowledge from the source domain, where the knowledge in the form of cluster prototypes is available for the clustering task. β_1 and β_2 are nonnegative parameters which can balance the influence of data in the target domain and knowledge in the source domain. In the experiments, we suggest the default values for these parameter $\alpha = 1, \beta = 2, \beta_1 = \beta_2 = 1, \delta = 10$.

2.2 Optimization

To minimize the objective function J_{TECM} , the Lagrange multipliers method is adopted. First, consider that the prototype sets in the target domain, \mathbf{V} , is fixed. To solve the constrained minimization problem with respect to the membership matrix \mathbf{M} , n Lagrange multipliers $\lambda_i (i = 1, \dots, n)$ are introduced and the Lagrangian can be written as:

$$L(\mathbf{M}; \lambda_1, \dots, \lambda_n) = J_{\text{TECM}} - \sum_{i=1}^n \lambda_i \left(\sum_{\substack{A_j \subseteq \Omega \\ A_j \neq \emptyset}} m_{ij} + m_{i\emptyset} - 1 \right). \quad (5)$$

Differentiating the Lagrangian with respect to m_{ij} , $m_{i\emptyset}$, and λ_i and setting the derivatives to zero, the necessary condition of optimality for \mathbf{M} can be got as:

$$\frac{\partial L}{\partial m_{ij}} = c_j^\alpha \beta m_{ij}^{\beta-1} \left(d_{ij}^2 + \beta_1 d_{ij}^{2(s)} \right) - \lambda_i = 0. \quad (6)$$

$$\frac{\partial L}{\partial m_{i\emptyset}} = \beta m_{i\emptyset}^{\beta-1} (\delta^2 + \beta_1 \delta^{2(s)}) - \lambda_i = 0. \quad (7)$$

$$\frac{\partial L}{\partial \lambda_i} = \sum_{\substack{A_j \subseteq \Omega \\ A_j \neq \emptyset}} m_{ij} + m_{i\emptyset} - 1 = 0. \quad (8)$$

From Eqs.(6) and (7), it is easy to obtain

$$m_{ij} = \left(\frac{\lambda_i}{c_j^\alpha \beta \left(d_{ij}^2 + \beta_1 d_{ij}^{2(s)} \right)} \right)^{1/(\beta-1)}. \quad (9)$$

$$m_{i\emptyset} = \left(\frac{\lambda_i}{\beta (\delta^2 + \beta_1 \delta^{2(s)})} \right)^{1/(\beta-1)}. \quad (10)$$

Substituting Eqs.(9) and (10) into Eq.(8), we can get

$$\left(\frac{\lambda_i}{\beta}\right)^{1/(\beta-1)} = \frac{1}{\sum_{\substack{A_j \subseteq \Omega \\ A_j \neq \emptyset}} \Delta_{ij} + \left(\frac{1}{\delta^2 + \beta_1 \delta^{2(s)}}\right)^{\frac{1}{\beta-1}}}, \quad (11)$$

where

$$\Delta_{ij} = \left(\frac{1}{c_j^\alpha (d_{ij}^2 + \beta_1 d_{ij}^{2(s)})}\right)^{\frac{1}{\beta-1}}. \quad (12)$$

Return in Eqs.(9) and (10),

$$m_{ij} = \frac{\left(1 / \left(c_j^\alpha \left(d_{ij}^2 + \beta_1 d_{ij}^{2(s)}\right)\right)\right)^{\frac{1}{\beta-1}}}{\sum_{\substack{A_k \subseteq \Omega \\ A_k \neq \emptyset}} \left(1 / \left(c_k^\alpha \left(d_{ik}^2 + \beta_1 d_{ik}^{2(s)}\right)\right)\right)^{\frac{1}{\beta-1}} + \left(\frac{1}{\delta^2 + \beta_1 \delta^2}\right)^{\frac{1}{\beta-1}}}, \quad (13)$$

and

$$m_{i\emptyset} = \frac{\left(\frac{1}{\delta^2 + \beta_1 \delta^2}\right)^{\frac{1}{\beta-1}}}{\sum_{\substack{A_k \subseteq \Omega \\ A_k \neq \emptyset}} \left(1 / \left(c_k^\alpha \left(d_{ik}^2 + \beta_1 d_{ik}^{2(s)}\right)\right)\right)^{\frac{1}{\beta-1}} + \left(\frac{1}{\delta^2 + \beta_1 \delta^2}\right)^{\frac{1}{\beta-1}}}. \quad (14)$$

Next we consider that the credal membership matrix \mathbf{M} is fixed. It is easy to see the minimization of J_{TECM} with respect to \mathbf{V} is an unconstrained optimization problem. The partial derivatives of J_{TECM} with respect to the prototypes of the specific classes in the target domain can be given by:

$$\frac{\partial J_{\text{TECM}}}{\partial \mathbf{v}_l} = \sum_{i=1}^n \sum_{\substack{A_j \subseteq \Omega \\ A_j \neq \emptyset}} c_j^\alpha m_{ij}^\beta \frac{\partial d_{ij}^2}{\partial \mathbf{v}_l} - 2\beta_2 (\mathbf{v}_l^{(s)} - \mathbf{v}_l), \quad (15)$$

$$\frac{\partial d_{ij}^2}{\partial \mathbf{v}_l} = 2(\mathbf{x}_i - \bar{\mathbf{v}}_j) \left(-s_{lj} \frac{1}{c_j}\right), \quad (16)$$

where $\bar{\mathbf{v}}_j$ is defined by Eq.(4). Thus we have:

$$\begin{aligned} \frac{\partial J_{\text{TECM}}}{\partial \mathbf{v}_l} &= -2 \sum_{i=1}^n \sum_{\substack{A_j \subseteq \Omega \\ A_j \neq \emptyset}} c_j^{\alpha-1} m_{ij}^\beta s_{lj} \left(\mathbf{x}_i - \frac{\sum_{k=1}^c s_{kj} \mathbf{v}_k}{c_j}\right) \\ &\quad - 2\beta_2 (\mathbf{v}_l^{(s)} - \mathbf{v}_l). \end{aligned} \quad (17)$$

Setting these derivatives to zero, we can get l linear equations of \mathbf{v}_k :

$$\sum_i \mathbf{x}_i \sum_{\substack{A_j \subseteq \Omega \\ A_j \neq \emptyset}} c_j^{\alpha-1} m_{ij}^\beta s_{lj} = \sum_{k=1}^c \mathbf{v}_k \sum_{i=1}^n \sum_{\substack{A_j \subseteq \Omega \\ A_j \neq \emptyset}} c_j^{\alpha-2} m_{ij}^\beta s_{kj} s_{lj} - \beta_2 \left(\mathbf{v}_l^{(s)} - \mathbf{v}_l \right). \quad (18)$$

Let \mathbf{B} be a matrix of size $(c \times p)$, and it can be defined by:

$$\mathbf{B}_{lq} = \sum_{i=1}^n x_{iq} \sum_{\substack{A_j \subseteq \Omega \\ A_j \neq \emptyset}} c_j^{\alpha-1} m_{ij}^\beta s_{lj} = \sum_{i=1}^n x_{iq} \sum_{A_j \ni \omega_l} c_j^{\alpha-1} m_{ij}^\beta, \quad (19)$$

and \mathbf{H} be a matrix of size $(c \times c)$ given by:

$$\mathbf{H}_{lk} = \sum_{i=1}^n \sum_{\substack{A_j \subseteq \Omega \\ A_j \neq \emptyset}} c_j^{\alpha-2} m_{ij}^\beta s_{lj} s_{kj} = \sum_i \sum_{A_j \supseteq \{\omega_k, \omega_l\}} c_j^{\alpha-2} m_{ij}^\beta. \quad (20)$$

Let \mathbf{I} be the $(c \times c)$ identity matrix. The prototype matrix \mathbf{v} can be got by solving the following linear system:

$$(\mathbf{H} + \beta_2 \mathbf{I}) \mathbf{v} = \mathbf{B} + \beta_2 \mathbf{v}^{(s)}. \quad (21)$$

3 Experiments

Some experiments are provided in this section. Generated Gaussian data and some UCI data sets are considered to show the performance of the proposed evidential transfer clustering method. In all experiments, the credal partitions provided by ECM and TECM are transformed into hard partitions by using maximum the corresponding Pignistic probability [11]. The parameters in ECM and TECM are all set as default ($\alpha = 1, \beta = 2, \beta_1 = \beta_2 = 1, \delta = 10$). Then, the Adjusted Rand Index (ARI) and Normalized Mutual Information (NMI) to measure closeness of a hard partition to the ground truth are adopted as performance index.

3.1 Gaussian data sets

As mentioned, TECM has advantages in the situation when the data in the target domain are insufficient and uncertain to train a good model. This experiment is to illustrate the application scope of TECM. Assume that source data and target data both follow two-dimensional Gaussian distribution. The mean values and covariance matrices of the source data and target data are listed in Tab.1.

There are three clusters in both the target data and the source data. Denote the number of data samples in each cluster of the target and source domain by

Table 1. Distributions of source data and target data.

Mean	Covariance	Mean	Covariance
$\mu_1^{(s)} = [2, 4]$	$\begin{bmatrix} 10 & 0 \\ 0 & 10 \end{bmatrix}$	$\mu_1 = [3, 4]$	$\begin{bmatrix} 10 & 0 \\ 0 & 11 \end{bmatrix}$
$\mu_2^{(s)} = [9, 15]$	$\begin{bmatrix} 25 & 0 \\ 0 & 7 \end{bmatrix}$	$\mu_2 = [10.5, 12.5]$	$\begin{bmatrix} 25 & 0 \\ 0 & 7 \end{bmatrix}$
$\mu_3^{(s)} = [8, 30]$	$\begin{bmatrix} 30 & 0 \\ 0 & 20 \end{bmatrix}$	$\mu_3 = [9, 29]$	$\begin{bmatrix} 30 & 0 \\ 0 & 19.5 \end{bmatrix}$

n_t and n_s respectively. As mentioned, when n_t is small, it is difficult to cluster the samples in the target domain correctly.

The experiment is designed by increasing n_t gradually (from 10 to 500) and applying both ECM and TECM algorithms. For each n_t , Gaussian data are generated 100 times under the fixed parameters in Tab.1. ECM and TECM algorithms are evoked each time. Noted that here in TECM the prototypes of clusters in the source domain are got by evoking c -means clustering method on the source data. The average values of ARI and NMI are reported and the results are shown in Fig.2. As can be seen, the clustering results obtained by TECM is significantly better than those by ECM in terms of both ARI and NMI.

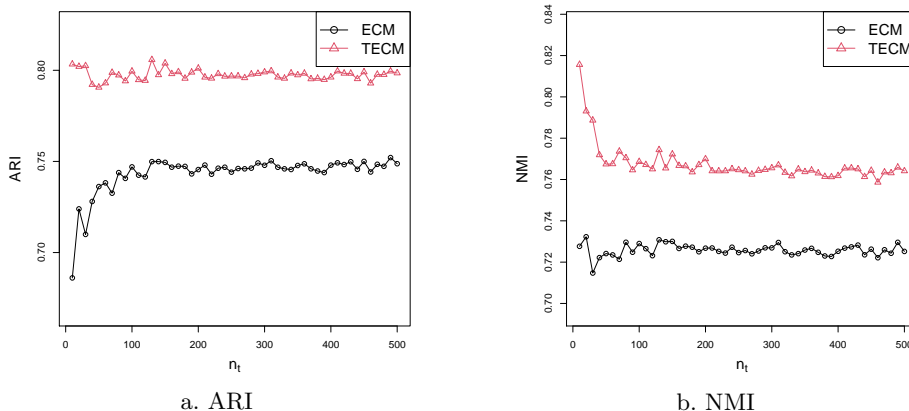


Figure 2. The ARI and NMI value of the clustering results on Gaussian data.

3.2 Iris data

This experiment is to show the effects of the prototypes available in the source domain on the clustering performance for the target data. We consider the Iris data set consisting of 50 samples from each of three species of Iris. Four features are measured from each sample: Sepal.Length (SL), Sepal.Width (SW), Petal.Length (PL) and Petal.Width (PW). The four features are divided into two parts $FT1$ and $FT2$. The six cases are listed in Tab.2. The samples with features in $FT1$ are regarded as the target data to be clustered.

Table 2. The feature division for Iris data.

Case	<i>FT1</i>	<i>FT2</i>	Case	<i>FT1</i>	<i>FT2</i>
Case 1	SL, SW	PL, PW	Case 4	SW, PL	SL, PW
Case 2	SL, PL	SW, PW	Case 5	SW, PW	PL, PL
Case 3	SL, PW	SW, PL	Case 6	PL, PW	SL, SW

In order to generate the prototypes in the source domain which are required before evoking TECM, for each case we first apply c -means clustering on the samples with features in *FT2* and get the best hard partition for the 150 samples. Then the following two schemes are designed to get the prototypes $v_k^{(s)}$:

Scheme A: By the feature mean of samples in each group with feature set *FT2*;

Scheme B: By the feature mean of samples in each group with feature set *FT1*;

The methods with two schemes are termed by TECM-A and TECM-B respectively. We can see that in Scheme B the prototypes are from the target data (with *FT1*) based on a clustering rule learned with the source domain (with *FT2*), while in Scheme A the prototypes are from the source data (with *FT2*) based on a clustering rule learned with the same domain (with *FT2*). The ARI and NMI for the results by ECM, TECM-A and TECM-B are displayed in Fig.3. From the figure we can see:

- For TECM-B, it performs better than ECM in all the cases except Cases 1 and 4, where the behavior of the two methods (TECM-B and ECM) is similar.
- For TECM-A, it performs worse than TECM-B in all the cases. It is not better than ECM in Cases 1, 3, 4, and 5.

In TECM-B, the transferred knowledge of prototypes have the same feature set as the target samples (this corresponds to the illustrative example in the introduction). The results show that the clustering performance is indeed improved by the use of information from the source. On the contrary, in TECM-A, the feature sets in the source and target domain are different. The knowledge from the source has a negative influence on the performance of transfer clustering in this situation. The imperfect matching between information provided by the two domains can degrade the clustering performance. We will study how to avoid such kind of negative transfer in the future.

3.3 UCI data sets

Three UCI data sets are used in this experiment: Seeds, Wine and Karate Club network. The first two data sets are object data while the last is a graph data. The number of samples of Seeds data is 210 while 178 for Wine data. The Karate Club network is a graph with 34 nodes and 78 edges.

There are 7 features in Seeds data and 13 for Wine data. In the experiment, nf features are randomly selected from the original data to form the target data set. For Karate graph, the vector embedding is first calculated by spectral

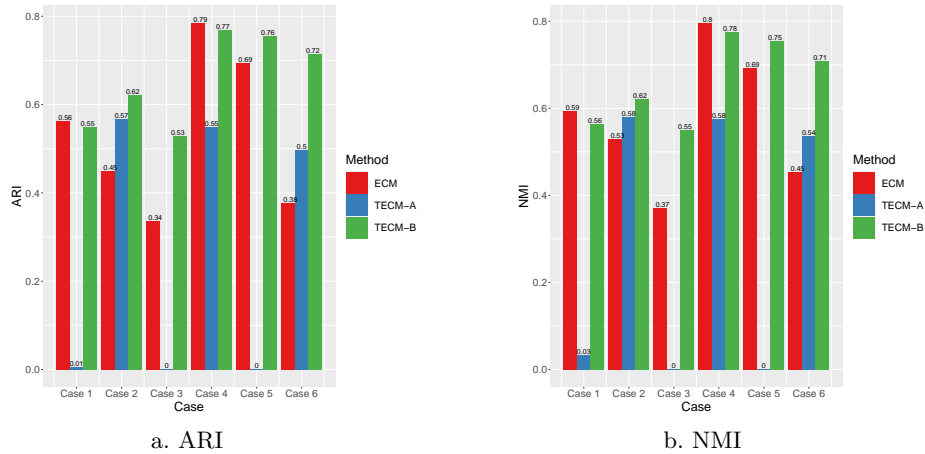


Figure 3. The ARI and NMI for the clustering results on Iris set by TECM and ECM.

Table 3. The ARI and NMI of clustering results on UCI data sets.

Dataset	$nf = 2$				$nf = 3$			
	ARI		NMI		ARI		NMI	
	ECM	TECM	ECM	TECM	ECM	TECM	ECM	TECM
Seeds	0.4748	0.5074	0.4685	0.4999	0.5046	0.5276	0.4907	0.5177
Wine	0.3497	0.3938	0.357	0.3838	0.3233	0.3785	0.3419	0.3694
Karate	0.2636	1	0.3173	1	0.7717	1	0.7329	1

decomposition of its adjacency matrices [12]. The embedding dimension is set to nf . Then ECM and TECM algorithms are used. We note here as the benchmarks for these data sets are known, in TECM the average values of the samples in the target data are directly used to simulate the prototypes in the source domain.

The ARI and NMI values of the clustering results provided by ECM and TECM are listed in Tab.3. In all the experiments, the results by TECM are better than those by ECM as they generally have higher ARI and NMI values. This confirms the advantages of the evidential transfer clustering approach when there is some available positive transferred knowledge in the source domain.

4 Conclusion

In this study, the concept of knowledge transfer has been used to develop an evidential transfer clustering method named TECM for the application of clustering task when the target data are uncertain or insufficient. The proposed TECM algorithm can learn from not only the data of the target domain but also from the knowledge of the source domain in the form of prototypes as well. The experimental results on generated data and UCI data have demonstrated the effectiveness of TECM algorithm compared with ECM which is without the transfer learning ability.

In TECM, the number of clusters in the source domain and in the target domain is assumed to be identical, which may be difficult to satisfy in real applications. How to deal with the case when the number of classes in the two domains is different will be studied in the future.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (No.61701409), the Natural Science Basic Research Plan in Shaanxi Province of China (No.2018JQ6005), the Aero Science Foundation of China (No.20182053023), the Science Research Plan of China (Xi'an) Institute for Silk Road Research (2019ZD02), and the Fundamental Research Funds for the Central Universities of China (No.310201911cx041).

References

1. Bai, L., Liang, J., Cao, F.: Semi-supervised clustering with constraints of different types from multiple information sources. *IEEE transactions on pattern analysis and machine intelligence* (2020)
2. Deng, Z., Jiang, Y., Chung, F.L., Ishibuchi, H., Choi, K.S., Wang, S.: Transfer prototype-based fuzzy clustering. *IEEE Transactions on Fuzzy Systems* **24**(5), 1210–1232 (2015)
3. Denceux, T., Masson, M.H.: EVCLUS: evidential clustering of proximity data. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* **34**(1), 95–109 (2004)
4. Gargees, R., Keller, J.M., Popescu, M.: TLPCM: Transfer learning possibilistic *c*-means. *IEEE Transactions on Fuzzy Systems* (2020)
5. Gong, C., Su, Z.g., Wang, P.h., Wang, Q.: An evidential clustering algorithm by finding belief-peaks and disjoint neighborhoods. *Pattern Recognition* **113**, 107751 (2021)
6. Jain, A.K.: Data clustering: 50 years beyond k-means. *Pattern recognition letters* **31**(8), 651–666 (2010)
7. Liu, X., Li, M., Tang, C., Xia, J., Xiong, J., Liu, L., Kloft, M., Zhu, E.: Efficient and effective regularized incomplete multi-view clustering. *IEEE transactions on pattern analysis and machine intelligence* (2020)
8. Masson, M.H., Denceux, T.: ECM: An evidential version of the fuzzy *c*-means algorithm. *Pattern Recognition* **41**(4), 1384–1397 (2008)
9. Qian, P., Jiang, Y., Deng, Z., Hu, L., Sun, S., Wang, S., Muzic, R.F.: Cluster prototypes and fuzzy memberships jointly leveraged cross-domain maximum entropy clustering. *IEEE transactions on cybernetics* **46**(1), 181–193 (2015)
10. Saxena, A., Prasad, M., Gupta, A., Bharill, N., Patel, O.P., Tiwari, A., Er, M.J., Ding, W., Lin, C.T.: A review of clustering techniques and developments. *Neuro-computing* **267**, 664–681 (2017)
11. Smets, P.: Decision making in the TBM: the necessity of the pignistic transformation. *International Journal of Approximate Reasoning* **38**(2), 133–147 (2005)
12. Sussman, D.L., Tang, M., Fishkind, D.E., Priebe, C.E.: A consistent adjacency spectral embedding for stochastic blockmodel graphs. *Journal of the American Statistical Association* **107**(499), 1119–1128 (2012)