https://doi.org/10.20965/jaciii.2015.p0023

Paper:

On Cluster Extraction from Relational Data Using L_1-Regularized Possibilistic Assignment Prototype Algorithm

Yukihiro Hamasuna* **and Yasunori Endo****

*Department of Informatics, School of Science and Engineering, Kinki University
3-4-1 Kowakae, Higashi-osaka, Osaka 577-8502, Japan
E-mail: yhama@info.kindai.ac.jp
**Faculty of Engineering, Information and Systems, University of Tsukuba
1-1-1 Tennodai, Tsukuba, Ibaraki 305-8573, Japan
E-mail: endo@risk.tsukuba.ac.jp

[Received April 20, 2014; accepted August 25, 2014]

This paper proposes entropy-based L_1-regularized possibilistic clustering and a method of sequential cluster extraction from relational data. *Sequential cluster extraction* means that the algorithm extracts cluster one by one. The assignment prototype algorithm is a typical clustering method for relational data. The membership degree of each object to each cluster is calculated directly from dissimilarities between objects. An entropy-based L_1-regularized possibilistic assignment prototype algorithm is proposed first to induce belongingness for a membership grade. An algorithm of sequential cluster extraction based on the proposed method is constructed and the effectiveness of the proposed methods is shown through numerical examples.

Keywords: assignment prototype algorithm, possibilistic clustering, L_1-regularization, sequential cluster extraction, relational data

1. Introduction

The aim of cluster analysis called *clustering*, is to discover important structures and features from massive complex databases. Clustering, a data analysis method, divides a set of objects into groups called clusters. Objects classified in the same cluster are considered similar, while those classified in a different cluster are considered dissimilar. Hard c-means [1] and fuzzy c-means clustering (FCM) are the most well known clustering methods [2–4], as is possibilistic clustering (PCM), known to be useful from the viewpoint of robustness against noise and outliers [5]. Robustness against noise and outliers is essential if clustering methods are to be useful in real-world applications. Several variations have been proposed and studied based on PCM [6, 7]. The noise clustering concept has been proposed to overcome the negative effect of noise and outliers [8–10]. A method of sequential cluster extraction has even been proposed using this drawback [8]. Significant algorithms that extract "one cluster at a time" have also been proposed [7, 11]. The sequential cluster extraction algorithm need not determine the number of clusters in advance – an advantage important when handling massive and complex data sets for detecting dense clusters.

Conventional clustering methods such as FCM and PCM generally handle numerical objects in the form of p-dimensional vector sets. Relational data, in contrast, are obtained from various social phenomena, e.g., social networks and e-commerce [12]. Relational data consist of measures of similarity or dissimilarity between objects. Many clustering methods have been proposed to handle such relational data [13–18]. The cluster center referred to as representative of a cluster is not used in these methods, and the membership degree of each object to each cluster is calculated directly from dissimilarities between objects. In the fuzzy non-metric model (FNM) of relational clustering methods, in which the membership degree of each object to each cluster is calculated directly from dissimilarities between objects [13, 16]. The assignment-prototype algorithm (AP) [18] and relational fuzzy c-means clustering (RFCM) [14] are also clustering methods used for relational data. FNM and AP handle both Euclidean and non-Euclidean relational data, whereas RFCM handles only Euclidean relational data. Non-Euclidean relational fuzzy c-means has been proposed to handle non-Euclidean relational data from that sense [15].

A probabilistic constraint on membership grade is used in FCM to denote that the sum of membership degrees is equal to 1. Noise and outliers cause negatively affect clustering results by FCM because of such a constraint. A possibilistic constraint is used instead of a probabilistic constraint in PCM [5], so a particular additional term for the membership degree is added in an objective function to obtain nontrivial solutions. In the machine learning field and related research, the L_1-regularization term is well known and useful for inducing sparseness [19, 20]. Sparseness means that a small variables are calculated as zero. In the clustering field, the sparse possibilistic clustering method has been proposed by introducing L_1-regularization [21] to obtain sparse results by calculating small degrees of membership as zero. This means that sparseness is induced from the viewpoint of not being in the cluster. Sparseness for being in the cluster is also considered in the clustering field. In that sense, L_1-regularized possibilistic clustering and its classification function have been proposed and described [22]. The sequential cluster extraction algorithm has also been constructed based on the L_1-regularized possibilistic clustering.

In this paper, we propose an entropy-based L_1-regularized assignment prototype algorithm (eL_1PAP) to handle relational data. We construct the sequential cluster extraction algorithm to extract "one cluster at a time" from relational data. We also show the effectiveness of the proposed method through numerical examples.

This paper is organized as follows: In Section 2, we introduce symbols, fuzzy clustering and possibilistic clustering. In Section 3, we formulate an optimization problem and construct the eL_1PAP algorithm. In Section 4, we construct the sequential cluster extraction algorithm based on eL_1PAP. In Section 5, we review the results for the proposed method with datasets, and in Section 6, we present conclusions.

2. Preparation

The set of objects to be clustered is given and denoted by $X = \{x_1, \ldots, x_n\}$ in which $x_k\,(k = 1, \ldots, n)$ is an object. In most cases, $x_1, \ldots, x_n$ are p-dimensional vectors $\Re^p$, that is, object $x_k \in \Re^p$. A cluster is denoted as $C_i (i = 1, \ldots, c)$. The membership degree of x_k belonging to C_i is denoted as u_{ki} and a partition matrix is denoted as $U = (u_{ki})_{1\le k\le n,\ 1\le i\le c}$.

2.1. Assignment Prototype Algorithm

Several clustering methods for relational data have been proposed. One example is fuzzy non-metric model (FNM) of FCM-type clustering methods for relational data based on optimizing an objective function under the membership grade constraint [13, 16].

FNM depends strongly on initial values [16]. An assignment prototype algorithm (FAP) has been proposed as a clustering method for relational data to achieve robustness for initial values [18]. The objective function of FAP is similar to and yet a bit different from the objective function of FNM. Two variables, i.e., membership grade u_{ki} and prototype weight w_{ti}, are used in the assignment prototype algorithm. Windham considers the following objective function:

$$J_{fap}(U,W) = \sum_{i=1}^{c}\sum_{k=1}^{n}\sum_{t=1}^{n} (u_{ki})^2 (w_{ti})^2 r_{kt}$$

where $W = (w_{ti})_{1\le t\le n,\ 1\le i\le c}$, is the prototype weight and r_{kt} is the distance measure between objects. The probabilistic constraint for u_{ki} is as follows:

$$\mathscr{U}_f = \left\{ (u_{ki}) : u_{ki} \in [0,1],\ \sum_{i=1}^{c} u_{ki} = 1,\ {}^{\forall}k \right\}. \qquad (1)$$

The constraint for w_{ti} is as follows:

$$\mathscr{W}_f = \left\{ (w_{ti}) : w_{ti} \in [0,1],\ \sum_{t=1}^{n} w_{ti} = 1,\ {}^{\forall}i \right\}. \qquad (2)$$

An entropy-based assignment prototype algorithm (eFAP) is considered in the same manner as in [3, 13]. The objective function of eFAP is as follows:

$$J_{efap}(U,W) = \sum_{i=1}^{c}\sum_{k=1}^{n} u_{ki} w_{ti} r_{kt} + \lambda_u \sum_{i=1}^{c}\sum_{k=1}^{n} u_{ki}\log u_{ki} + \lambda_w \sum_{i=1}^{c}\sum_{t=1}^{n} w_{ti}\log w_{ti}, \quad \ldots \qquad (3)$$

where $\lambda_u > 0$ and $\lambda_w > 0$ are fuzzification parameters. Constraints for u_{ki} and w_{ti} are the same as for Eqs. (1) and (2).

The formulation of FAP and eFAP is similar to that of fuzzy c-medoids [12] and related to the FCM-type co-clustering model [23].

2.2. Possibilistic Clustering

FCM is not robust against noise and outliers due to constraint $\mathscr{U}_f$. The main idea of possibilistic clustering (PCM) is to reduce the influence of noise and outliers on a data set [5, 9]. We introduce two objective functions of PCM:

$$J_{ps}(U,V) = \sum_{k=1}^{n}\sum_{i=1}^{c} (u_{ki})^m d_{ki} + \sum_{i=1}^{c} \eta_i \sum_{k=1}^{n} (1-u_{ki})^m,$$

$$J_{pe}(U,V) = \sum_{k=1}^{n}\sum_{i=1}^{c} u_{ki}\left\{ d_{ki} + \lambda u_{ki}(\log u_{ki} - 1)\right\},$$

where $m > 1$, $\eta_i > 0$ and $\lambda_u > 0$ are parameters for PCM.

The constraint $\mathscr{U}_f$ is not used in PCM, but additional terms for u_{ki} are added to derive a nontrivial solution. Condition $\mathscr{U}_p$ is written as follows:

$$\mathscr{U}_p = \left\{ (u_{ki}) \ : \ u_{ki} \in [0,1],\ {}^{\forall}k \right\}, \quad \ldots\ldots \qquad (4)$$

where we have omitted original constraint $0 < \sum_{k=1}^{n} u_{ki} \le n$.

The above objective functions are typical examples of PCM [4]. Sparse possibilistic clustering has also been proposed by introducing L_1-regularization [21]:

$$J_{sp}(U,V) = \sum_{k=1}^{n}\sum_{i=1}^{c} \left\{ (u_{ki})^m d_{ki} + \gamma |u_{ki} - \alpha| \right\},$$

where $m > 1$ and $\gamma > 0$ are parameters. Parameter $\alpha > 0$ is called a baseline constant [20]. This method induces sparseness for calculating the small membership grade as zero.

3. Proposed Method

3.1. Entropy-Based L_1-Regularized Possibilistic Assignment Prototype Algorithm

We propose a new possibilistic clustering method for relational data, called entropy-based L_1-regularized possibilistic assignment prototype algorithm (eL_1PAP). We consider the following objective function for eL_1PAP:

$$\begin{aligned} J_{elpap}(U,W) = & \sum_{i=1}^{c}\sum_{k=1}^{n}\sum_{t=1}^{n} u_{ki}w_{ti}r_{kt} + \lambda_w \sum_{i=1}^{c}\sum_{t=1}^{n} w_{ti}\log w_{ti} \\ & + \lambda_u \sum_{i=1}^{c}\sum_{k=1}^{n} u_{ki}\left(\log u_{ki} - 1\right) \\ & + \gamma \sum_{i=1}^{c}\sum_{k=1}^{n} |1-u_{ki}|. \end{aligned}$$

$\lambda_w > 0$, $\lambda_u > 0$, and $\gamma > 0$ are the parameters of eL_1PNM. The condition for u_{ki} remains the same as for Eq. (4) and that for w_{ti} remains the same as for Eq. (2). In the following discussion of eL_1PAP, d_{ki} and g_{ti} are as follows:

$$d_{ki} = \sum_{t=1}^{n} w_{ti}r_{kt},$$

$$g_{ti} = \sum_{k=1}^{n} u_{ki}r_{kt}.$$

The optimal solution of w_{ti} is derived as follows:

$$w_{ti} = \frac{\exp\left(-\lambda_w^{-1}g_{ti}\right)}{\sum_{q=1}^{n}\exp\left(-\lambda_w^{-1}g_{qi}\right)}. \quad \ldots\ldots\ldots (5)$$

The main problem in constructing the algorithm of eL_1PAP is how to derive the optimal solution of u_{ki}. Each membership u_{ki} could be solved separately in an eL_1PAP procedure because of a condition $\mathscr{U}_p$. We thus first consider the following semi-objective function:

$$J_{elpap}^{ki}(u_{ki}) = u_{ki}d_{ki} + \lambda_u u_{ki}\left(\log u_{ki} - 1\right) + \gamma|1-u_{ki}|.$$

We decompose $1-u_{ki} = \xi^+ - \xi^-$, to obtain partial derivatives for u_{ki} in which all element of ξ^+ and ξ^- are non-negative. The semi-objective function is rewritten by using a decomposition method [20] as follows:

$$J_{elpap}^{ki}(u_{ki}) = u_{ki}d_{ki} + \lambda_u u_{ki}\left(\log u_{ki} - 1\right) + \gamma\left(\xi^+ + \xi^-\right).$$

Constraints are as follows:

$$1-u_{ki} \le \xi^+, \quad 1-u_{ki} \ge -\xi^-, \quad \xi^+, \xi^- \ge 0.$$

Introducing Lagrange multipliers β^+, β^-, ψ^+, and $\psi^- \ge 0$, Lagrangian L^{lp} is as follows:

$$\begin{aligned} L^{elpap} = & u_{ki}d_{ki} + \lambda_u u_{ki}\left(\log u_{ki} - 1\right) + \gamma\left(\xi^+ + \xi^-\right) \\ & + \beta^+\left(1-u_{ki}-\xi^+\right) + \beta^-\left(-1+u_{ki}-\xi^-\right) \\ & - \psi^+\xi^+ - \psi^-\xi^-. \quad \ldots\ldots\ldots (6) \end{aligned}$$

From $\partial L^{elpap}/\partial\xi^+ = 0$ and $\partial L^{elpap}/\partial\xi^- = 0$,

$$\gamma - \beta^+ - \psi^+ = 0, \quad \gamma - \beta^- - \psi^- = 0. \quad \ldots (7)$$

Since ψ^+, $\psi^- \ge 0$, conditions $0 \le \beta^+ \le \gamma$ and $0 \le \beta^- \le \gamma$ are obtained from Eq. (7). Substituting Eq. (7) into Eq. (6), Lagrangian L^{elpap} is simplified as follows:

$$L^{elpap} = u_{ki}d_{ki} + \lambda_u u_{ki}\left(\log u_{ki} - 1\right) + \beta\left(1-u_{ki}\right). \quad (8)$$

Here, $\beta = \beta^+ - \beta^-$ and satisfies $-\gamma \le \beta \le \gamma$.
From $\partial L^{elpap}/\partial u_{ki} = 0$

$$u_{ki} = \exp\left(-\frac{d_{ki}-\beta}{\lambda_u}\right). \quad \ldots\ldots\ldots (9)$$

Substituting the above for L^{elpap}, the Lagrangian dual problem is written as follows:

$$L_d^{elpap} = \beta - \lambda_u \exp\left(-\frac{d_{ki}-\beta}{\lambda_u}\right).$$

From $\partial L_d^{elpap}/\partial\beta = 0$, this dual problem is solved as

$$\beta = d_{ki}. \quad \ldots\ldots\ldots (10)$$

The optimal solution of the primary problem is derived by considering Eq. (9), Eq. (10) and $-\gamma \le \beta \le \gamma$. $\beta < 0$ is not realized because d_{ki} is always positive. For $0 \le \beta \le \gamma$, the optimal solution is $u_{ki} = 1$ because $\beta = d_{ki}$. For $\gamma < \beta$, the optimal solution is $u_{ki} = \exp\left(-\frac{d_{ki}-\gamma}{\lambda_u}\right)$. The optimal solution for u_{ki} of eL_1PAP is derived as follows:

$$u_{ki} = \begin{cases} 1 & 0 \le d_{ki} \le \gamma \\ \exp\left(-\frac{d_{ki}-\gamma}{\lambda_u}\right) & \gamma < d_{ki} \end{cases} \quad (11)$$

3.2. eL_1PAP Algorithm

The eL_1PAP algorithm is described as **Algorithm 1** based on the previous discussion:

Algorithm 1 eL_1PAP algorithm

eL_1PAP 1 Set initial values and parameters.

eL_1PAP 2 Calculate $u_{ki} \in U$ as follows:

$$u_{ki} = \begin{cases} 1 & 0 \le d_{ki} \le \gamma \\ \exp\left(-\frac{d_{ki}-\gamma}{\lambda_u}\right) & \gamma < d_{ki} \end{cases}$$

eL_1PAP 3 Calculate $w_{ti} \in W$ as follows:

$$w_{ti} = \frac{\exp\left(-\lambda_w^{-1}g_{ti}\right)}{\sum_{q=1}^{n}\exp\left(-\lambda_w^{-1}g_{qi}\right)}.$$

eL_1PAP 4 If the convergence criterion is satisfied, stop. Otherwise go back to **eL_1PAP 2**.

4. Sequential Cluster Extraction Algorithm

The objective function of PCM is minimized separately because the probabilistic constraint used in FCM is not used in PCM [9]. This implies the drawback that a cluster centers obtained by PCM would have exactly the same results. The sequential cluster extraction algorithm is constructed by developing this drawback. The basis of this algorithm was previously proposed and discussed [7, 9, 10]. We construct the sequential cluster extraction algorithm by using the proposed method. The object has small d_{ki} classified in a compact cluster. The membership grade of such data can then be calculated as $u_{ki} = 1$ by using Eq. (11). These data should be classified in the same cluster, which satisfies $u_{ki} = 1$. eL_1PAP can thus extract one cluster at a time by minimizing the objective function in $c = 1$. The sequential cluster extraction algorithm by eL_1PAP is described as **Algorithm 2**.

Algorithm 2 Sequential cluster extraction algorithm based on proposed methods

STEP 1 Give X, initial values u_{ki}, w_{ti}, and parameters λ_u, λ_w, and γ.

STEP 2 Repeat the eL_1PAP algorithm with $c = 1$ until the convergence criterion is satisfied.

STEP 3 Extract $\{x_k \mid u_{ki} = 1\}$ from X.

STEP 4 If $X = \emptyset$ or the maximum number of repetitions is satisfied, stop. Otherwise give initial values and go back to **STEP 2**.

Parameter γ plays an important role in extracting clusters one by one. Previous studies mentioned a procedure for determining the sequential cluster extraction parameter [8, 10]. Based on that research, we construct the updating procedure for γ as follows:

$$\gamma = \rho \left[\frac{\sum_{k=1}^{|X|} \sum_{i=1}^{c} d_{ki}}{|X|} \right] \quad \text{. . . . . (12)}$$

$\rho > 0$ is a scale multiplier that must be determined in advance. $|X|$ means the number of objects handled in STEP 2. Parameter γ is updated in **Algorithm 2** by using Eq. (12).

5. Numerical Examples

We show the numerical examples of sequential cluster extraction using the Polaris and Iris datasets. The parameters which are used in eL_1PAP are fixed $\lambda_u = 1.0$ and $\lambda_w = 100.0$. The squared Euclidean norm is used as the distance measure between x_k and x_t, i.e., $r_{kt} = \|x_k - x_t\|^2$.

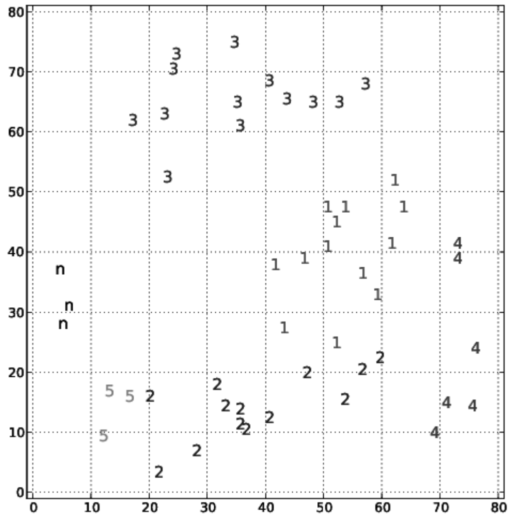

Fig. 1. Sequential cluster extraction by eL_1PAP with $\lambda_u = 1.0$, $\lambda_w = 100.0$, and $\rho = 0.4$. The number of extracted clusters is 5 and the number of noise objects is 3.

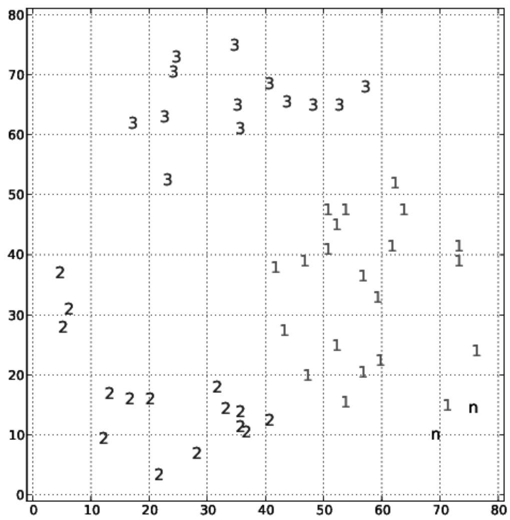

Fig. 2. Sequential cluster extraction by eL_1PAP with $\lambda_u = 1.0$, $\lambda_w = 100.0$, and $\rho = 0.7$. The number of extracted clusters is 3 and the number of noise objects is 2.

5.1. Polaris Dataset

We start with the results of the proposed method with the Polaris dataset, which consists of 51 objects with 2 attributes and should be classified into three clusters.

Figures 1 and **2** are illustrative examples of sequential cluster extraction by eL_1PAP. The value displayed at each data point denotes the sequence for extracting clusters. The letter 'n' denotes the noise objects. We show the effectiveness of the proposed method by evaluating the results of the average and standard derivation of the Rand index (RI) [24], the number of extracted clusters, and the number of noise objects as listed in **Table 1**. The proposed algorithm extracts some *good* clusters and does

Table 1. Results of 100 trials of sequential cluster extraction by eL_1PAP with the Polaris dataset.

	Rand Index		Num. of clusters		Num. of noise	
ρ	Ave	SD	Ave	SD	Ave	SD
0.1	0.703	0.009	9.71	0.516	11.18	2.075
0.2	0.770	0.009	8.57	0.863	1.55	1.445
0.3	0.779	0.016	7.16	0.946	1.04	1.174
0.4	0.804	0.022	5.59	0.722	1.90	1.005
0.5	0.863	0.028	5.18	0.712	1.89	0.646
0.6	0.887	0.030	5.31	0.956	1.58	0.961
0.7	0.923	0.053	3.95	0.792	1.86	0.510
0.8	0.866	0.036	5.64	0.625	0.19	0.560
0.9	0.759	0.018	5.87	0.336	0.00	0.000
1.0	0.693	0.014	5.02	0.346	0.00	0.000

Table 2. Results of 100 trials of sequential cluster extraction by eL_1PAP with the Iris dataset.

	Rand Index		Num. of clusters		Num. of noise	
ρ	Ave	SD	Ave	SD	Ave	SD
0.1	0.329	0.000	0.00	0.000	150.00	0.000
0.2	0.518	0.006	1.00	0.000	113.62	1.420
0.3	0.806	0.024	2.24	0.472	37.37	5.416
0.4	0.795	0.003	3.01	0.099	13.23	1.535
0.5	0.790	0.003	3.49	0.806	4.20	2.631
0.6	0.782	0.001	3.96	0.196	3.17	0.788
0.7	0.780	0.001	3.42	0.494	2.62	0.690
0.8	0.779	0.000	3.37	0.730	3.07	1.645
0.9	0.778	0.000	5.00	0.000	0.00	0.000
1.0	0.777	0.000	5.00	0.000	0.00	0.000

not extract noise objects. Noise objects are the objects not extracted until the convergence criterion is satisfied and consists of a noise cluster . In **Table 1**, the number of clusters means the number of *good* clusters. The number of repetitions is set to 10 as the convergence criterion in this numerical experiment. In these tables, Ave and SD means average and standard variation.

Figures 1 and **2** show that some rough clusters are extracted and some noise objects are not extracted by the proposed algorithm. It also shows that parameter ρ used in Eq. (12) affects the results of sequential cluster extraction. The case with $\rho = 0.7$ takes better results and seems to be suitable for the Polaris dataset, the larger the ρ, the smaller the number of extracted clusters. The number of noise objects is from **Table 1**. Large ρ induces large γ that takes a broad area that satisfies $u_{ki} = 1.0$. These tables show that the proposed algorithm is robust for initial values because the value of SD is quite small. Parameter γ further influences the classification results and sequential cluster extraction.

5.2. Iris Dataset

Next, we show the results of proposed methods with the Iris dataset published in the UCI machine learning repository.[1] The Iris dataset consists of 150 objects with 4 attributes and should be classified into three clusters. We also show the results of proposed method with the average and standard derivation RI, the number of extracted clusters, and the number of noise objects in **Table 2**. In the case of $\rho = 0.1$, the algorithm does not extract a *good* cluster and all objects are classified into noise clusters. Many objects are also classified into noise clusters in the case of $\rho = 0.2$ and $\rho = 0.3$. By evaluating the value of RI, the number of extracted clusters, and the number of noise objects comprehensively, parameter ρ is suitable for the Iris dataset between 0.4 and 0.8.

1. http://archive.ics.uci.edu/ml/ [Accessed September 8, 2014]

6. Conclusions

We have proposed an entropy-based L_1-regularized possibilistic assignment prototype algorithm (eL_1PAP). We have also constructed a sequential cluster extraction algorithm based on eL_1PAP. The algorithm handles both p-dimensional objects $x_k \in R^p$ and relational data (r_{kt}). We have shown the effectiveness of the sequential cluster extraction algorithm through numerical examples.

In future work, we will compare the performance of the proposed sequential cluster extraction algorithm with other methods and benchmark datasets. We will also consider how to determine a better or more suitable parameter ρ, which plays an important role in the proposed method and algorithm.

Acknowledgements

This study was supported in part by Telecommunications Advancement Foundation, JSPS KAKENHI Grant Numbers 26330270, and 26330271.

References:

[1] A. K. Jain, "Data clustering: 50 years beyond K-means," Pattern Recognition Letters, Vol.31, No.8, pp. 651-666, 2010.
[2] J. C. Bezdek, "Pattern Recognition with Fuzzy Objective Function Algorithms," Plenum Press, New York, 1981.
[3] S. Miyamoto and M. Mukaidono, "Fuzzy c-means as a regularization and maximum entropy approach," Proc. of the 7th Int. Fuzzy Systems Association World Congress (IFSA'97), Vol.2, pp. 86-92, 1997.
[4] S. Miyamoto, H. Ichihashi, and K. Honda, "Algorithms for Fuzzy Clustering," Springer, Heidelberg, 2008.
[5] R. Krishnapuram, J. M. Keller, "A possibilistic approach to clustering," IEEE Trans. on Fuzzy Systems, Vol.1, No.2, pp. 98-110, 1993.
[6] Y. Hamasuna, Y. Endo, and S. Miyamoto, "On Tolerant Fuzzy c-Means Clustering and Tolerant Possibilistic Clustering," Soft Computing, Vol.14, No.5, pp. 487-494, 2010.3.
[7] S. Miyamoto, Y. Kuroda, K. Arai, "Algorithms for Sequential Extraction of Clusters by Possibilistic Method and Comparison with Mountain Clustering," J. of Advanced Computational Intelligence and Intelligent Informatics (JACIII), Vol.12, No.5, pp. 448-453, 2008.
[8] R. N. Davé, "Characterization and detection of noise in clustering," Pattern Recognition Letters, Vol.12, No.11, pp. 657-664, 1991.
[9] R. N. Davé and R. Krishnapuram, "Robust clustering methods : A unified view," IEEE Trans. on Fuzzy Systems, Vol.5, No.2, pp. 270-293, 1997.

[10] R. N. Davé and S. Sen, "Robust Fuzzy Clustering of Relational Data," IEEE Trans. on Fuzzy Systems, Vol.10, No.6, pp. 713-727, 2002.

[11] R. R. Yager and D. P. Filev, "Approximate clustering via the mountain method," IEEE Trans. on Systems, Man and Cybernetics, Vol.2, No.8, pp. 1279-1284, 1994.

[12] R. Krishnapuram, A. Joshi, O. Nasraoui, and L. Yi, "Low-complexity fuzzy relational clustering algorithms for Web mining," IEEE Trans. on Fuzzy Systems, Vol.9, No.4, pp. 595-607, 2001.

[13] Y. Endo, "On Entropy Based Fuzzy Non Metric Model – Proposal, Kernelization and Pairwise Constraints –," J. of Advanced Computational Intelligence and Intelligent Informatics (JACIII), Vol.16, No.1, pp. 169-173, 2012.

[14] R. J. Hathaway, J. W. Davenport, J. C. Bezdek, "Relational Duals of the *c*-Means Clustering Algorithms," Pattern Recognition, Vol.22, No.2, pp. 205-212, 1989.

[15] R. J. Hathaway and J. C. Bezdek, "Nerf *c*-Means: Non-Euclidean Relational Fuzzy Clustering," Pattern Recognition, Vol.27, No.3, pp. 429-437, 1994.

[16] M. Roubens, "Pattern classification problems and fuzzy sets," Fuzzy Sets and Systems, Vol.1, pp. 239-253, 1978.

[17] E. Ruspini, "Numerical methods for fuzzy clustering," Information Science, Vol.2, No.3, pp. 319-350, 1970.

[18] M. P. Windham, "Numerical classification of proximity data with assignment measures," J. of Classification, Vol.2, pp. 157-172, 1985.

[19] E. J. Candes, M. B. Wakin, and S. Boyd, "Enhancing Sparsity by Reweighted l1 Minimization," J. of Fourier Analysis and Applications, Vol.14, No.5, pp. 877-905, 2008.

[20] K. Tsuda and T. Kudo, "Clustering graphs by weighted substructure mining," Proc. of the 23rd Int. Conf. on Machine Learning (ICML), pp. 953-9960, 2006.

[21] R. Inokuchi and S. Miyamoto, "Sparse Possibilistic Clustering with L1 Regularization," Proc. of The 2007 IEEE Int. Conf on Granular Computing (GrC2007), pp. 442-445, 2007.

[22] Y. Hamasuna and Y. Endo, "Sequential Extraction By Using Two Types of Crisp Possibilistic Clustering," Proc. of the IEEE Int. Conf. on Systems, Man, and Cybernetics (IEEE SMC 2013), pp. 3505-3510, 2013.

[23] C.-H Oh, K. Honda, and H. Ichihashi, "Fuzzy clustering for categorical multivariate data," Proc. of Joint 9th IFSA World Congress and 20th NAFIPS Int. Conf., pp. 2154-2159, 2001.

[24] W. M. Rand., "Objective criteria for the evaluation of clustering methods," J. of the American Statistical Association, Vol.66, No.336, pp. 846-850, 1971.

Name:
Yukihiro Hamasuna

Affiliation:
Lecturer, Department of Informatics, School of Science and Engineering, Kinki University

Address:
3-4-1 Kowakae, Higashi-osaka, Osaka 577-8502, Japan

Brief Biographical History:
2009-2010 Research Fellow of the Japan Society for the Promotion of Science
2011-2014 Assistant Professor, Kinki University
2014- Lecturer, Kinki University

Main Works:
- Y. Hamasuna, Y. Endo, and S. Miyamoto, "On Tolerant Fuzzy *c*-Means Clustering and Tolerant Possibilistic Clustering," Soft Computing, Vol.14, No.5, pp. 487-494, 2010.
- Y. Hamasuna, Y. Endo, and S. Miyamoto, "Fuzzy *c*-Means Clustering for Data with Clusterwise Tolerance Based on L_2- and L_1-Regularization," J. of Advanced Computational Intelligence and Intelligent Informatics (JACIII), Vol.15, No.1, pp. 68-75, 2011.
- Y. Hamasuna and Y. Endo, "On Semi-supervised Fuzzy *c*-Means Clustering for Data with Clusterwise Tolerance by Opposite Criteria," Soft Computing, Vol.17, No.1, pp.71-81, 2013.

Membership in Academic Societies:
- Japan Society for Fuzzy Theory and Systems (SOFT)
- The Japanese Society for Artificial Intelligence (JSAI)
- The Institute of Electrical and Electronics Engineering (IEEE)

Name:
Yasunori Endo

Affiliation:
Professor, Faculty of Engineering, Information and Systems, University of Tsukuba

Address:
1-1-1 Tennodai, Tsukuba, Ibaraki 305-8573, Japan

Brief Biographical History:
1994-1997 Research Assistant, Waseda University
1997-2001 Assistant Professor, Tokai University
2001-2004 Assistant Professor, University of Tsukuba
2004-2013 Associate Professor, University of Tsukuba
2012 Guest Research Scholar, IIASA
2013- Associate Professor, University of Tsukuba

Main Works:
- Y. Endo and N. Kinoshita, "On Objective-Based Rough c-Means Clustering," Proc. 2012 IEEE Int. Conf. on Granular Computing, #SA002, 2012.
- Y. Endo, R. Murata, H. Haruyama, and S. Miyamoto, "Fuzzy *c*-Means for Data with Tolerance," Proc. of Int. Symp. on Nonlinear Theory and Its Applications, pp. 345-348, 2005.
- Y. Endo, H. Haruyama, and T. Okubo, "On Some Hierarchical Clustering Algorithms Using Kernel Functions," Proc. IEEE Int. Conf. on Fuzzy Systems, #1106, 2004.

Membership in Academic Societies:
- Japan Society for Fuzzy Theory and Systems (SOFT)
- The Institute of Electronics, Information and Communication Engineers (IEICE)
- The Institute of Electrical and Electronics Engineering (IEEE)