Possibilistic One Mean Clustering Using Particle Swarm Optimization

Thomas A. Runkler
Siemens AG
Munich, Germany
thomas.runkler@siemens.com

***Abstract*—Two popular methods for soft clustering are fuzzy clustering and possibilistic clustering. Possibilistic c–means (PCM) clustering is less sensitive to noise and outliers in data than fuzzy c–means clustering (FCM), but PCM may find non–distinct (duplicate) clusters. Both PCM and FCM can also be used to find single clusters, but FCM yields trivial clusters (the complete set with all memberships equal to one), whereas PCM can find meaningful single clusters, which is called possibilistic one–mean (P1M). P1M is also used for multiple clusters by sequentially finding one cluster after the other, the so–called sequential possibilistic one–means (SP1M), which can avoid non–distinct clusters, as an alternative to conventional PCM. P1M clustering can be done by alternating optimization (AO) of the P1M objective function, but P1M AO may converge to local minima. To the best of our knowledge this is the first paper that examines local minima of the P1M objective function. Moreover, we propose a particle swarm optimization (PSO) algorithm to avoid such local minima. Experiments with synthetic and real–world data indicate that PSO finds the global minimum of the P1M objective function considerably more often than AO, and also on average leads to lower values of the P1M objective function, which yields a higher probability of obtaining better clustering results, at the cost of longer run time, but still feasible even for larger data sets.**

I. Introduction

Clustering is used to partition a set into subsets called clusters, where elements in the same cluster are similar in some well–defined sense. For numerical data sets, points in each cluster should have low pairwise distances, e.g. Euclidean distances. Soft clustering allows to partially assign objects to possibly multiple clusters. Fuzzy clustering such as *fuzzy c–means* (FCM) [1] uses membership values between zero and one to quantify to which degree each object is assigned to each cluster, and requires the sum of these membership values in all clusters to be equal to one for each object. This constraint leads to a high sensitivity to noise and outliers, because these are often assigned large memberships in one or several clusters. This constraint is abandoned in possibilistic clustering, for example in *possibilistic c–means* (PCM) [2], which is therefore often more robust against noise and outliers, but PCM sometimes finds non–distinct clusters, i.e. some pairs of clusters may (almost) be the same. Both fuzzy and possibilistic clustering can be used to find a single cluster in a given set, but fuzzy clustering will yield trivial single clusters equal to the complete set (because the memberships have to sum up to one), whereas possibilistic clustering can find non–trivial single clusters which are proper subsets of the given set. An example for such a clustering method is *possibilistic one–mean* (P1M) [3]. Often P1M is used to construct one cluster after the other, until the desired number of clusters in found. This is called *sequential possibilistic one–means* (SP1M) [3], which is an alternative to conventional PCM that can avoid non–distinct clusters. All of the mentioned clustering methods (FCM, PCM, P1M, SP1M) find cluster centers and partition matrices by minimizing their specific cluster objective functions. A widely used approach is to optimize the clustering objective functions by *alternating optimization* (AO), i.e. the partition matrices and cluster centers are alternatingly updated using the necessary conditions for extrema of the objective function. Depending on the considered data set, the cluster objective functions may have local minima, so AO may yield sub–optimal solutions when converging to such local minima. Many approaches have been suggested to avoid local minima, in particular stochastic optimization methods such as *genetic algorithms* (GA) [4], *ant colony optimization* (ACO) [5], or *particle swarm optimization* (PSO) [6]. For example, the optimization of the FCM objective function by GA has been proposed in [7], by ACO in [8], and by PSO in [9].

To the best of our knowledge this is the first paper that investigates local extrema of the PCM or more specifically P1M objective functions, and the first paper that proposes a stochastic optimization approach to optimize these. More specifically, we propose the new P1M PSO algorithm that optimizes the P1M objective function by particle swarm optimization. P1M PSO can be executed sequentially to find more than one cluster. We call this *sequential possibilistic one–means with particle swarm optimization* (SP1M PSO). We compare our new P1M PSO algorithm with the traditional P1M AO algorithm in experiments with three data sets: two Gaussians, the BIRCH data set, and the Iris data set. Our paper is structured as follows: In Section II we provide an overview of fuzzy and possibilistic clustering models for multiple and also for single clusters. In Section III we briefly review PSO and propose its application to P1M clustering. In section IV we present our experiments and results when we apply P1M AO and P1M PSO clustering to the three considered data sets. In section V we summarize our conclusions and outline ideas for future research.

II. Fuzzy and Possibilistic Clustering

In this section we provide an overview of fuzzy and possibilistic clustering models for multiple and also for single clusters. A popular fuzzy clustering method is *fuzzy c–means* (FCM) [1]. Given a data set $X = \{x_1, \ldots, x_n\} \subset \mathbb{R}^p$, FCM finds a set $V = \{v_1, \ldots, v_c\} \subset \mathbb{R}^p$ of cluster centers and a $c \times n$ matrix U of membership values $u_{ik} \in [0,1]$, $i = 1, \ldots, c$, $k = 1, \ldots, n$, that quantify to which degree each point k belongs to each cluster i, where the sum of memberships of all points in each cluster is one.

$$\sum_{k=1}^{n} u_{ik} = 1 \tag{1}$$

for all $i = 1, \ldots, c$. The cluster centers and the membership matrix can be found by minimizing the FCM objective function

$$J_{\text{FCM}}(U, V; X) = \sum_{i=1}^{c} \sum_{k=1}^{n} u_{ik}^m d_{ik}^2 \tag{2}$$

with the parameter $m > 1$. In this paper we will always use the standard value $m = 2$ and the Euclidean distance

$$d_{ik}^2 = \|x_k - v_i\|^2 \tag{3}$$

The minimization of the objective function can be done by *alternating optimization* (AO), where U and V are updated alternatingly using the necessary conditions for extrema of the objective function, which for FCM yield

$$u_{ik} = \frac{1}{\sum\limits_{j=1}^{c} \left(\frac{d_{ik}}{d_{jk}}\right)^{\frac{2}{m-1}}}, \quad v_i = \frac{\sum\limits_{k=1}^{n} u_{ik}^m x_k}{\sum\limits_{k=1}^{n} u_{ik}^m} \tag{4}$$

If we choose $c = 1$, then FCM will yield one trivial cluster where all $u_{1k} = 1$ and v_1 is the mean of the data set X. Many successful applications of FCM are reported in the literature, for example in [10]–[14], but FCM is known to be sensitive to noise and outliers, because the constraint (1) often leads to implausibly high membership values for noise or outlier points [15]. This constraint is abandoned in *possibilistic c–means* (PCM) [2], and trivial zero memberships are avoided by adding a penalty term to the objective function

$$J_{\text{PCM}}(U, V; X) = \sum_{i=1}^{c} \sum_{k=1}^{n} u_{ik}^m d_{ik}^2 + \sum_{i=1}^{c} \eta_i \sum_{k=1}^{n} (1 - u_{ik})^m \tag{5}$$

with parameters $\eta_1, \ldots, \eta_c > 0$, which yields the PCM AO update equations

$$u_{ik} = \frac{1}{1 + \left(\frac{d_{ik}^2}{\eta_i}\right)^{\frac{1}{m-1}}}, \quad v_i = \frac{\sum\limits_{k=1}^{n} u_{ik}^m x_k}{\sum\limits_{k=1}^{n} u_{ik}^m} \tag{6}$$

Each parameter η_i, $i = 1, \ldots, c$, specifies the quadratic radius of cluster i. Also for PCM many successful applications are reported in the literature, for example in [16]–[20]. Notice that the FCM AO update equations (4) are coupled, i.e. u_{ik} depends on all d_{jk} and therefore also on all v_j, $j = 1, \ldots, c$, whereas the PCM AO update equations (6) are independent of each other, i.e. u_{ik} depends on d_{ik} and v_i but not from any other v_j, $j \neq i$. As an undesirable effect of this, PCM may yield non–distinct clusters, i.e. it may find (almost) the same cluster several times. However, as a desirable effect of this, PCM can also be used to find one non–trivial single cluster [21], which is then called *possibilistic one–mean* (P1M) [3], [22], [23]. This is in contrast to FCM which can also find one single cluster, but this will be trivial with all memberships equal to one and the cluster center equal to the mean of the data set, as pointed out above. P1M finds one single cluster center $v \in \mathbb{R}^p$ and a vector of membership values $(u_1, \ldots, u_n)$ by minimizing the objective function

$$J_{\text{P1M}}(U, V; X) = \sum_{k=1}^{n} u_k^m d_k^2 + \eta \sum_{k=1}^{n} (1 - u_k)^m \tag{7}$$

with the parameter $\eta > 0$ specifying the quadratic radius of the cluster. The P1M AO update equations are

$$u_k = \frac{1}{1 + \left(\frac{d_k^2}{\eta}\right)^{\frac{1}{m-1}}}, \quad v = \frac{\sum\limits_{k=1}^{n} u_k^m x_k}{\sum\limits_{k=1}^{n} u_k^m} \tag{8}$$

Often P1M is used in a sequential manner called *sequential possibilistic one–means* (SP1M), where one cluster is found at a time using P1M until a total of c clusters is found. When SP1M initializes cluster centers, it can take the previously found clusters into account and therefore avoid finding duplicate clusters, as an alternative to conventional PCM. However, also using P1M for finding one *single* cluster is an important task, for example for finding the most relevant cluster in a data set with many clusters, noise, and outliers, or when processing streaming data. Successful applications of (S)P1M are reported in [24]–[26], for example. In this paper we will see that P1M AO (just as PCM AO and SP1M AO) may yield sub–optimal solutions because of converging to local minima of the objective function. This can be avoided by using alternative approaches to optimize the objective function, for example stochastic optimization methods. In this paper we propose to use *particle swarm optimization* (PSO) [6], [27] as a very promising example of such stochastic optimization methods to optimize the P1M objective function.

III. Particle Swarm Optimization

In this section we briefly review *particle swarm optimization* (PSO) [6], [27] and propose its application to P1M clustering. PSO is a heuristic stochastic optimization method that can be used to minimize an objective function over a continuous space without using any gradient information (gradient–free method). Each solution candidate (called *particle*) is represented as a point in the p–dimensional space. A swarm of q such particles moves through the solution space and explores the objective function. Each particle is influenced by the (personal and global) best found solutions, which leads to an interaction between the particles and to a balance between exploration and exploitation. More specifically, each particle has a position v_i, a velocity $\dot{v}_i$, and a personal best solution $pBest_i$, and the whole swarm has a global best solution $gBest$. In each step the velocity of each particle is updated as

$$\dot{v}_i := w\dot{v}_i + c_1 r_1 (pBest_i - v_i) + c_2 r_2 (gBest - v_i) \qquad (9)$$

with r_1 and r_2 randomly chosen from $(0, 1)$, and with the parameters w, c_1, and c_2. The parameter w is called the *inertia weight*, and it controls how much the particle's previous velocity affects its current velocity. A higher inertia weight leads to more exploration and a lower one to more exploitation. c_1 is called the *cognitive coefficient* (or *personal best acceleration coefficient*), and it controls how much a particle is attracted to its personal best solution. c_2 is called the *social coefficient* (or *global best acceleration coefficient*), and it controls how much a particle is attracted to the global best solution. In this paper we will not perform any parameter tuning but use standard settings from the literature: number of particles $q = 30$, inertia weight $w = 0.7$, cognitive coefficient $c_1 = 2$, and social coefficient $c_2 = 2$. After the update of the velocities, the position of each particle is updated as

$$v_i := v_i + \dot{v}_i \qquad (10)$$

Finally, the objective function is evaluated at the positions of each particle, and the personal and global best solutions are updated. This algorithm can be briefly summarized by the following pseudo code:

Algorithm: Particle Swarm Optimization

Initialize $v_i, \dot{v}_i$ for $i = 1, \ldots, q$
$pBest_i := v_i$ for $i = 1, \ldots, q$
$gBest := pBest_k$ where $k = \arg\min J(pBest_j)$
For $t = 1, \ldots, t_{\text{max}}$
    For $i = 1, \ldots, q$
        $r_1, r_2 \in (0, 1)$
        $\dot{v}_i := w\dot{v}_i + c_1 r_1 (pBest_i - v_i) + c_2 r_2 (gBest - v_i)$
        $v_i := v_i + \dot{v}_i$
        If $J(v_i) < J(pBest_i)$, then $pBest_i := v_i$
        If $J(v_i) < J(gBest)$, then $gBest := v_i$

We apply this PSO algorithm to P1M clustering by using the P1M objective function (7) for J and letting each particle represent a potential location of the cluster center v. We initialize all particles to the same initial position, and initialize the velocities of each particle by randomly drawing from a p–dimensional Gaussian normal distribution. As the resulting cluster center v we use the global best solution $gBest$ after t_{max} steps.

IV. Experiments

In this section we present our experiments and results when we apply P1M AO and P1M PSO clustering to three different data sets: a simple data set with two Gaussians, the BIRCH data set, and the well-known Iris data set. For all three data sets we compare the performance of AO and PSO for the

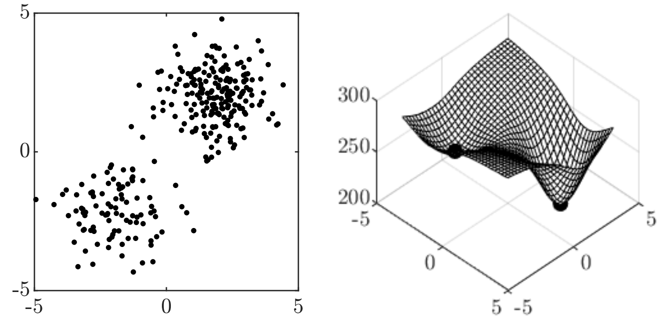

Fig. 1. Two Gaussians: data set (left) and P1M objective function (right).

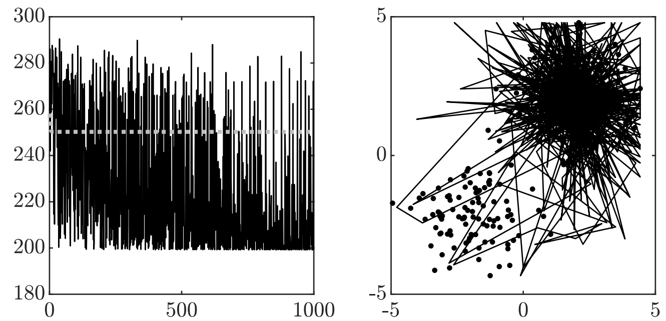

Fig. 2. Two Gaussians data set: values of the P1M objective function for one run of AO (grey dotted line) and PSO (black solid line) (left); positions of the cluster center for this P1M run (right).

optimization of the P1M objective function. We visualize the first two data sets and the corresponding P1M objective functions, and for all three data sets we present results from selected single optimization runs and show histograms of the P1M AO and P1M PSO results for 100 runs with random initialization.

A. Two Gaussians

In a first set of experiments we consider a two–dimensional data set constructed by randomly drawing 100 points from a Gaussian normal distribution centered at $(-2, -2)$, and another 200 points from a Gaussian normal distribution centered at $(+2, +2)$, so we have a total of 300 data points. Fig. 1 shows this data set (left) and its P1M objective function (right). For this data set, the P1M objective function has a global minimum at $v_1 \approx (2.0076, 1.9536)$ close to the center of the top right Gaussian with $J_1 \approx 199.2122$ and a local minimum at $v_2 \approx (-1.9078, -1.9131)$ close to the center of the bottom left Gaussian with $J_2 \approx 250.2279$, both are marked with black circles. Obviously, J_1 is much lower than J_2. For the initialization $V = (-1, -1)$, Fig. 2 (left) shows the values of J for the first 1000 iterations of AO (grey dotted line) and PSO (black solid line). AO quickly converges to the suboptimal local minimum v_2 with $J_2 \approx 250.2279$. PSO explores the search space and yields values of J in the range from about 200 to 290, with a moving average starting at

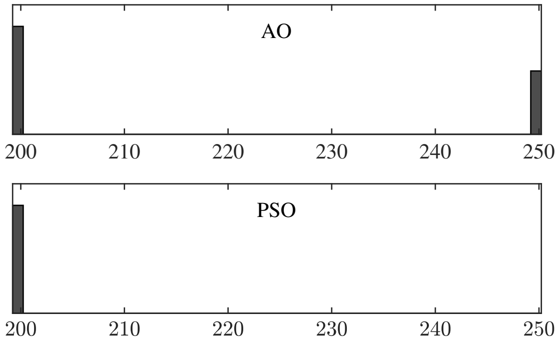

Fig. 3. Two Gaussians data set: histograms of the P1M objective function for AO and PSO.

around 270 and declining towards 200. After 35 iterations it finds the solution $J \approx 200.3326$, and the minimum of this run of PSO is $J \approx 199.2132$, which is very close to the global minimum $J_1 \approx 199.2122$. For this run of PSO, Fig. 2 (right) shows the pathway of v. It explores many positions in the minimal bounding box of X, also the region of the bottom left Gaussian around $(-2, -2)$ which corresponds to the local minimum J_2, but concentrates on the region of the top right Gaussian around $(+2, +2)$ corresponding to the global minimum J_1. We repeat this experiment with 100 random initializations of V in the minimal bounding box of X. Fig. 3 shows the histograms of the minimal values of J found by AO (top) and PSO (bottom). AO finds the global minimum J_1 (left histogram bar in the top view) in about 63% of the runs, and converges to the local minimum J_2 (right histogram bar in the top view) in 37% of the runs, whereas PSO finds the global minimum J_1 (histogram bar on the left in the bottom view) in *all* of the runs.

B. BIRCH

In a second set of experiments we consider an instance of the well–known two–dimensional BIRCH data set [28], which is widely used for comparing clustering algorithms, for example in [3], [29], [30]. Our instance of the two–dimensional BIRCH data set has an array of 4×4 clusters located at the grid points $(-6, -2, 2, 6) \times (-6, -2, 2, 6)$, each drawn from a Gaussian normal distribution with the point counts

$$n = \begin{pmatrix} 100 & 200 & 150 & 50 \\ 200 & 150 & 50 & 100 \\ 150 & 50 & 100 & 200 \\ 50 & 100 & 200 & 150 \end{pmatrix} \tag{11}$$

so the total number of data points is 2000. Fig. 4 shows this data set (left) and its P1M objective function (right). For this data set, the P1M objective function has one global minimum at $v_1 \approx (5.8167, 1.9528)$ with $J_1 \approx 1856.9$, which corresponds to the Gaussian at $(6, 2)$ with 200 points, and it has 15 local minima, each close to one of the Gaussian centers at $(-6, -2, 2, 6) \times (-6, -2, 2, 6)$, with higher values of J for lower numbers of data points, and lower values of J for higher numbers of data points. For the initialization $V = (0, 0)$, Fig. 5 shows the values of J for the first 1000 iterations of AO (grey dotted line) and PSO (black solid line). AO quickly converges to the suboptimal local minimum $v_2 \approx (1.9178, 2.0922)$ with $J_2 \approx 1868.9$, which corresponds to the Gaussian at $(2, 2)$ with only 100 points. PSO yields values of J in the range from about 1860 to 1980, with a moving average starting at around 1920 and declining towards 1860. The minimum of this run of PSO is $J \approx 1856.9$ after 834 iterations, which is almost equal to the global minimum J_1. We repeat this experiment with 100 random initializations of V in the minimal bounding box of X. Fig. 6 shows the histograms of the minimal values of J found by AO (top) and PSO (bottom). AO finds the global minimum J_1 (leftmost histogram bar in the top view) in only about 6% of the runs, and solutions with values of J up to about 1912.9 in all other cases, with an average of $J \approx 1876.8$, whereas PSO finds the global minimum J_1 (leftmost histogram bar in the bottom view) in about 47% of the runs, and solutions with values of J up to about 1864.3 in all other cases, with an average of only $J \approx 1857.9$.

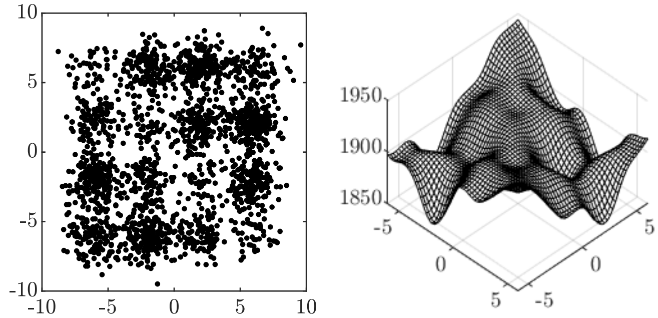

Fig. 4. BIRCH: data set (left) and P1M objective function (right).

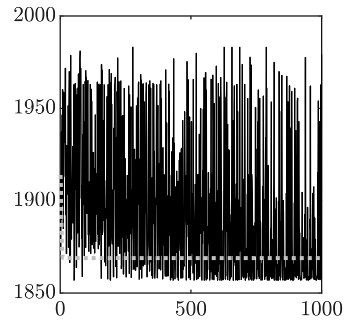

Fig. 5. BIRCH data set: values of the P1M objective function for one run of AO (grey dotted line) and PSO (black solid line) (left).

C. Iris

In a third set of experiments we consider the well–known Iris data set [31] which contains 150 4–dimensional data points [32]. Each data point corresponds to an individual Iris flower,

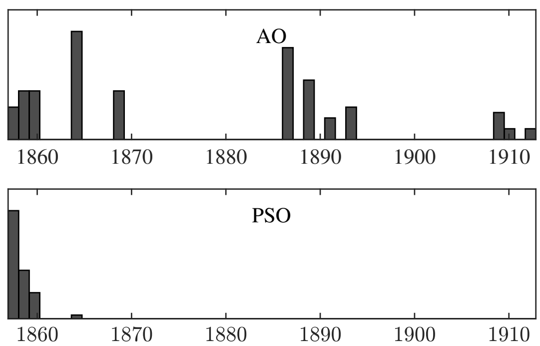

Fig. 6. BIRCH data set: histograms of the P1M objective function for AO and PSO.

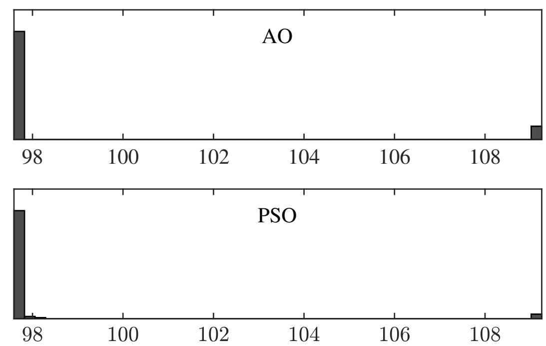

Fig. 8. Iris data set: histograms of the P1M objective function for AO and PSO.

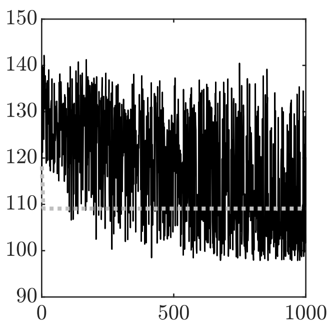

Fig. 7. Iris data set: values of the P1M objective function for one run of AO (grey dotted line) and PSO (black solid line).

which belongs to one of the three distinct classes (species): Iris setosa, Iris versicolor, or Iris virginica. Each feature represents a specific measurement of that flower: sepal length, sepal width, petal length, and petal width (in centimeters). We normalize each feature of the data set to mean zero and variance one. Since the data set is 4–dimensional, we can not simply plot the data set nor its objective function. For the initialization $V = (-0.599, 0.061, -1.036, -0.295)$, Fig. 7 shows the values of J for the first 1000 iterations of AO (grey dotted line) and PSO (black solid line). AO quickly converges to $v_2 \approx (-0.9885, 0.7022, -1.227, -1.1868)$ with $J_2 \approx 109.08$, which is a local minimum of the P1M objective function. The nearest neighbor of this local minimum is of the class Iris setosa. PSO yields values of J in the range from about 98 to 142, with a moving average starting at around 128 and declining towards 110. The minimum of this run of PSO is $J \approx 97.879$ after 969 iterations. This is close to the global minimum of the P1M objective function at $v_1 \approx (0.3696, -0.3823, 0.5444, 0.4937)$ with $J_1 \approx 97.59$. The nearest neighbor of this global minimum is of the class Iris versicolor, but the distances of this global minimum to the means of the classes Iris versicolor and Iris virginica are almost the same. This corresponds to the finding often reported in the literature that the classes Iris versicolor and Iris virginica are highly overlapping and can be considered as one single cluster with 100 data points, as opposed to the cluster Iris setosa with only 50 data points. This explains why the joint cluster Iris versicolor and Iris virginica with 100 points yields a better (lower) value of the P1M objective function (global minimum) than the cluster Iris setosa with only 50 data points (local minimum). We repeat this experiment with 100 random initializations of V in the minimal bounding box of X. Fig. 8 shows the histograms of the minimal values of J found by AO (top) and PSO (bottom). AO finds the global minimum J_1 (left histogram bar in the top view) in about 89% of the runs, and the local minimum J_2 (right histogram bar in the top view) in about 11% of the runs, with an average of $J \approx 98.8548$. PSO finds solutions close to the global minimum J_1 (left histogram bar and the tiny bar next to it in the bottom view) in about 96% of the runs, and the local minimum J_2 (right histogram bar in the bottom view) in about 4% of the runs, with an average of only $J \approx 98.1202$.

To summarize the histograms for the three considered data sets at Figs. 3, 6, and 8, PSO on average finds considerably better solutions than AO, and has higher probabilities to find solutions close to the global minimum of the P1M objective function for all three data sets.

V. Conclusions

In this paper we have examined the objective functions of *possibilistic one–mean* (P1M) clustering for three data sets: two Gaussians, BIRCH, and Iris data set. Our experiments have shown that traditional *alternating optimization* (AO) often converges to local minima of the P1M objective function and therefore yields sub–optimal clustering results. To overcome this, we proposed an alternative approach to minimize the P1M objective function based on *particle swarm optimization* (PSO). In our experiments, AO converges very quickly, but PSO needs about 100 times more steps, which translates to run times in tens of milliseconds for AO and seconds for PSO on a conventional office PC. However, PSO finds the global minimum of the P1M objective function in considerably more cases than AO. For the two Gaussians data set, PSO even finds the global minimum of the P1M objective function in *all* out of 100 runs, whereas AO finds it in only 63%. This indicates that PSO is an attractive alternative to AO for finding better possibilistic clusters, at the cost of longer run time. For simplicity our experiments were focused on finding single clusters, but our findings can be easily extended to finding multiple clusters using *sequential possibilistic one–means* (SP1M).

This work is a first study of local minima in possibilistic clustering and how to avoid these. We are planning to extend this work in several different directions: We plan an extensive parameter study to speed up the PSO algorithm for clustering tasks. We will consider also other stochastic optimization methods such as *genetic algorithms* (GA) or *ant colony optimization* (ACO). We want to find non–spherical clusters by replacing the Euclidean with the Mahalanobis distance. And we are planning to extend our experiments to other large–scale real–world data sets, in particular in data streaming settings.

REFERENCES

[1] J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*. New York: Plenum Press, 1981.

[2] R. Krishnapuram and J. M. Keller, "A possibilistic approach to clustering," *IEEE Transactions on Fuzzy Systems*, vol. 1, no. 2, pp. 98–110, May 1993.

[3] T. A. Runkler and J. M. Keller, "Sequential possibilistic one–means clustering," in *IEEE International Conference on Fuzzy Systems*, Naples, Italy, Jul. 2017.

[4] J. H. Holland, *Adaptation in Natural and Artificial Systems*. The University of Michigan Press, 1975.

[5] M. Dorigo and V. Maniezzo, "Ant system: Optimization by a colony of cooperating agents," *IEEE Transactions on Systems, Man, and Cybernetics B*, vol. 26, no. 1, pp. 29–41, 1996.

[6] J. Kennedy and R. C. Eberhart, "Particle swarm optimization," in *IEEE International Conference on Neural Networks*, Perth, Australia, Nov. 1995, pp. 1942–1948.

[7] F. Klawonn and A. Keller, "Fuzzy clustering with evolutionary algorithms," *International Journal of Intelligent Systems*, vol. 13, no. 10–11, pp. 975–991, 1998.

[8] T. A. Runkler, "Ant colony optimization of clustering models," *International Journal of Intelligent Systems*, vol. 20, no. 12, pp. 1233–1261, Dec. 2005.

[9] T. A. Runkler and C. Katz, "Fuzzy clustering by particle swarm optimization," in *IEEE International Conference on Fuzzy Systems*, Vancouver, Jul. 2006, pp. 3065–3072.

[10] Y. W. Lim and S. U. Lee, "On the color image segmentation algorithm based on the thresholding and the fuzzy c–means techniques," *Pattern Recognition*, vol. 23, no. 9, 1990.

[11] S. Araki, H. Nomura, and N. Wakami, "Segmentation of thermal images using the fuzzy c–means algorithm," in *IEEE International Conference on Fuzzy Systems*, San Francisco, Mar. 1993, pp. 719–724.

[12] T. Kaczmarzyk and W. Pedrycz, "Content-based image retrieval: an application of MPEG-7 standard and fuzzy c–means," in *Annual Meeting of the North American Fuzzy Information Processing Society*, 2006, pp. 172–177.

[13] S. Zhang, R. S. Wang, and X. S. Zhang, "Identification of overlapping community structure in complex networks using fuzzy c–means clustering," *Physica A: Statistical Mechanics and its Applications*, vol. 374, no. 1, pp. 483–490, 2007.

[14] D. Czerwinski, A. Kiersztyn, R. Łopucki, P. Karczmarek, I. Kitowski, and A. Zbyryt, "An application of fuzzy c–means, fuzzy cognitive maps, and fuzzy rules to forecasting first arrival date of avian spring migrants," in *IEEE International Conference on Fuzzy Systems*, 2020.

[15] T. A. Runkler and J. C. Bezdek, "Alternating cluster estimation: A new tool for clustering and function approximation," *IEEE Transactions on Fuzzy Systems*, vol. 7, no. 4, pp. 377–393, Aug. 1999.

[16] O. Sjahputera, J. M. Keller, J. W. Davis, K. H. Taylor, F. Rahmatpanah, H. Shi, D. T. Anderson, S. N. Blisard, R. H. Luke, M. Popescu, G. Author, and C. Caldwell, "Relational analysis of CpG islands methylation and gene expression in human lymphomas using possibilistic c–means clustering and modified cluster fuzzy density," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 4, no. 2, pp. 176–189, 2007.

[17] P. Ganesan and V. Rajini, "A method to segment color images based on modified fuzzy–possibilistic–c–means clustering algorithm," in *Recent Advances in Space Technology Services and Climate Change*, 2010, pp. 157–163.

[18] S. Bahrampour, B. Moshiri, and K. Salahshoor, "Weighted and constrained possibilistic c–means clustering for online fault detection and isolation," *Applied Intelligence*, vol. 35, no. 2, pp. 269–284, 2011.

[19] Z.-X. Ji, Q.-S. Sun, and D.-S. Xia, "A modified possibilistic fuzzy c–means clustering algorithm for bias field estimation and segmentation of brain MR image," *Computerized Medical Imaging and Graphics*, vol. 35, no. 5, pp. 383–397, 2011.

[20] M. M. B. dos Santos and M. P. Segundo, "Continuous biometric authentication using possibilistic c–means," in *IEEE International Conference on Fuzzy Systems*, Rio de Janeiro, Brazil, 2018.

[21] J. M. Keller and I. J. Sledge, "A cluster by any other name," in *Annual Meeting of the North American Fuzzy Information Processing Society*, San Diego, 2007, pp. 427–432.

[22] W. Wu, J. M. Keller, and T. A. Runkler, "Sequential possibilistic one–means clustering with dynamic eta," in *IEEE International Conference on Fuzzy Systems*, Rio de Janeiro, Brazil, Jul. 2018.

[23] T. A. Runkler and J. M. Keller, "Sequential possibilistic one–means clustering with variable eta," in *GMA/GI Workshop Computational Intelligence, Dortmund*. KIT Scientific Publishing, Dec. 2017, pp. 103–116.

[24] W. Wu and J. M. Keller, "Sequential possibilistic local information one-means clustering for image segmentation," in *IEEE International Conference on Fuzzy Systems*, 2020.

[25] W. Wu, J. M. Keller, M. Skubic, M. Popescu, and K. R. Lane, "Early detection of health changes in the elderly using in-home multi-sensor data streams," *ACM Transactions on Computing for Healthcare*, vol. 2, no. 3, pp. 1–23, 2021.

[26] W. Wu, J. M. Keller, J. Dale, and J. C. Bezdek, "StreamSoNG: A soft streaming classification approach," *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 6, no. 3, pp. 700–709, 2021.

[27] Y. Shi and R. Eberhart, "A modified particle swarm optimizer," in *IEEE International Conference on Evolutionary Computation*, Anchorage, Alaska, May 1998, pp. 69–73.

[28] T. Zhang, R. Ramakrishnan, and M. Livny, "BIRCH: An efficient data clustering method for very large databases," in *ACM SIGMOD International Conference on Management of Data*, 1996, pp. 103–114.

[29] B. Zhang, M. Hsu, and U. Dayal, "K–harmonic means — a data clustering algorithm," HP Laboratories Palo Alto, Tech. Rep. HPL–1999-124, 1999.

[30] O. Nasraoui and C. Rojas, "Robust clustering for tracking noisy evolving data streams," in *SIAM International Conference on Data Mining*, 2006, pp. 619–623.

[31] E. Anderson, "The Irises of the Gaspe Peninsula," *Bull. of the American Iris Society*, vol. 59, pp. 2–5, 1935.

[32] J. C. Bezdek, J. M. Keller, R. Krishnapuram, L. I. Kuncheva, and N. R. Pal, "Will the real Iris data please stand up?" *IEEE Transactions on Fuzzy Systems*, vol. 7, no. 3, pp. 368–369, 1999.