Swarm and Evolutionary Computation 17 (2014) 1–13

Contents lists available at ScienceDirect

Swarm and Evolutionary Computation

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

Review

Research on particle swarm optimization based clustering: A systematic review of literature and techniques

Shafiq Alam [a,*], Gillian Dobbie [a], Yun Sing Koh [a], Patricia Riddle [a], Saeed Ur Rehman [b]

[a] Department of Computer Science, The University of Auckland, 303.476, 38 Princes Street, 1010 Auckland, New Zealand
[b] Unitec Institute of Technology, Auckland, New Zealand

ARTICLE INFO

Article history:
Received 11 December 2012
Received in revised form
13 November 2013
Accepted 2 February 2014
Available online 17 February 2014

Keywords:
Swarm intelligence
Particle swarm optimization
Data mining
Data clustering

ABSTRACT

Optimization based pattern discovery has emerged as an important field in knowledge discovery and data mining (KDD), and has been used to enhance the efficiency and accuracy of clustering, classification, association rules and outlier detection. Cluster analysis, which identifies groups of similar data items in large datasets, is one of its recent beneficiaries. The increasing complexity and large amounts of data in the datasets have seen data clustering emerge as a popular focus for the application of optimization based techniques. Different optimization techniques have been applied to investigate the optimal solution for clustering problems. Swarm intelligence (SI) is one such optimization technique whose algorithms have successfully been demonstrated as solutions for different data clustering domains. In this paper we investigate the growth of literature in SI and its algorithms, particularly Particle Swarm Optimization (PSO). This paper makes two major contributions. Firstly, it provides a thorough literature overview focusing on some of the most cited techniques that have been used for PSO-based data clustering. Secondly, we analyze the reported results and highlight the performance of different techniques against contemporary clustering techniques. We also provide an brief overview of our PSO-based hierarchical clustering approach (HPSO-clustering) and compare the results with traditional hierarchical agglomerative clustering (HAC), K-means, and PSO clustering.

Contents

* Corresponding author: Tel.: +64 93737599 82128; fax: +64 93737453.
E-mail addresses: sala038@aucklanduni.ac.nz (S. Alam), gill@cs.auckland.ac.nz (G. Dobbie), ykoh@cs.auckland.ac.nz (Y.S. Koh), pat@cs.auckland.ac.nz (P. Riddle), srehman@unitec.ac.nz (S. Ur Rehman).

http://dx.doi.org/10.1016/j.swevo.2014.02.001
2210-6502

1. Introduction

Regardless of the type, source, or format of the data, when it grows beyond certain limits, it becomes difficult to comprehend. Extracting information from such data becomes difficult. On one hand, the increase in the amount of data adds to the possibility of the data possessing more information but on the other hand, it decreases the speed of pattern extraction. Knowledge Discovery and Data mining (KDD) has made possible the discovery of such useful and hidden patterns in the data. KDD is the process of automatically searching large volumes of data for previously unknown, potentially interesting and informative patterns [1–4]. KDD techniques are mainly influenced by modern information exploration techniques, however, they also rely on traditional computational techniques from statistics, information retrieval, machine learning and pattern recognition [5,4]. Data selection is the first phase of KDD, and specifies the scope of the data to be used for information extraction. The data might come from different sources and need integration before the actual pattern extraction process. Primary and secondary attributes, which will be used in further analysis, are specified in the second stage, the data is analyzed and preprocessed to enhance the reliability of pattern extraction, removing irrelevant data, handling missing values in the data, and removing outlier observations from the data. The preprocessed data is then transformed into a suitable format to be processed by the data mining algorithms. Transformation includes sampling and feature selection. The transformed data is exploited by a data mining method and post-processed to reveal the hidden informative patterns. Finally, the evaluation and interpretation of the resulting patterns take place for decision making (Fig. 1).

Data mining, which is the core of KDD, extracts informative patterns such as clusters of relevant data, classification and association rules, sequential patterns and prediction models from different types of data such as textual data, audio-visual data, and microarray data. This data comes from different sources such as structured databases, semi-structured documents, and data warehouses, where the growing amount of data is challenging the information extraction capabilities of domain experts.

Data clustering, one of the most important techniques in data mining, aims to group unlabeled data into different groups on the basis of similarities and dissimilarities between the data elements [6,5]. A typical clustering process involves feature selection, selection of a similarity measure, grouping of data, and assessment of the output. The process can be performed in a supervised, semi-supervised, or unsupervised manner. Different algorithms have been proposed that take into account the nature of the data, the quantity of the data, and other input parameters in order to cluster the data. Data clustering has received a lot of attention from researchers of various data mining domains. This has resulted in a number of approaches being suggested to address one or other aspects of the data clustering problem. Two of the most commonly used clustering approaches are partition-based clustering and hierarchy-based clustering.

Partition-based clustering divides the data into different groups based on similarities and dissimilarities among the data elements. Similarity measures differ from application to application, but the most common measures are distance-based, pattern-based, and density-based similarity measures. In distance-based similarity measures, a distance function is used to calculate the relative position of a data element inside the cluster by comparing it with the center of the cluster i.e. the centroid. The centroid changes its position during different iterations to improve the quality of the clusters in terms of intra-cluster and inter-cluster distances. The quality of a cluster is relative to an objective function which can be minimizing the intra-cluster distance, maximizing the inter-cluster distance, maximizing similarities, and minimizing dissimilarities among the data elements. K-means clustering technique is one of the foundations of the partitional approaches [6]. It uses a distance measure to divide the dataset into K-clusters. The assignment of data to a particular cluster-centroid is continued until the centroid remains unchanged in successive iterations or a maximum number of iterations is reached. In K-means clustering a data element belongs to only one cluster at a time. K-Harmonic means [7] is another partition-based clustering technique introduced to tackle the problem of sensitivity to the initialization of the K-means method. Partitional approaches are efficient but the quality of the solution depends on domain

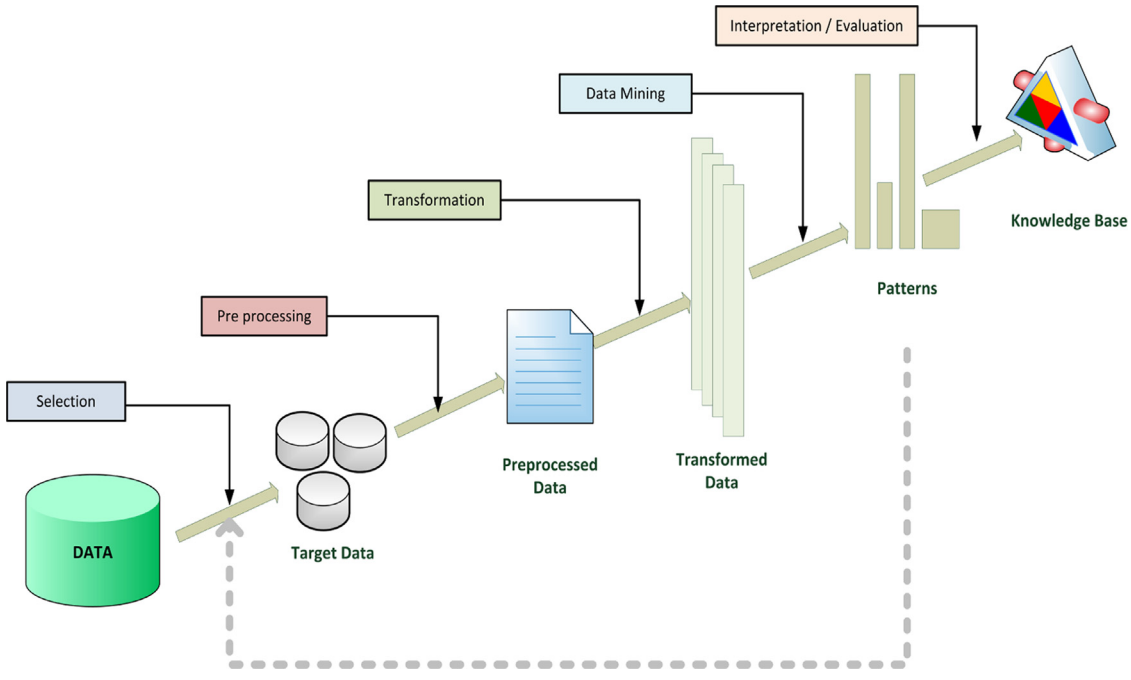

Fig. 1. KDD process [5].

knowledge, initial centroid position, and the number of clusters that are specified in advance.

Hierarchical clustering provides a sequence of nested partitions of the dataset in the form of a hierarchy. It divides the data into a nested tree structure where the levels of the tree show similarity or dissimilarity among the clusters at different levels. A hierarchical approach is either agglomerative or divisive. A divisive approach is based on the splitting of one large cluster into subclusters [8]. In the divisive approach, the clustering process starts with every data element in one large cluster. The cluster is then divided into smaller clusters on the basis of proximity until some criteria related to the number of clusters or number of data elements per cluster have been reached. In the agglomerative approach, the clustering process starts with every data element in an individual cluster. Initially the two most similar objects are merged using a dissimilarity matrix by selecting the smallest value of distance between two data points. In each successive pass, the individual clusters are then merged based on the distance between these clusters which is calculated using any of the linkage methods. Linkage methods calculate the distance between two clusters to find the dissimilarity of the two clusters. There are three commonly used linkage methods, single linkage, average linkage, and complete linkage. In the single linkage method the distance between two clusters is calculated by taking the minimum distance between any two of the members of the clusters. Complete linkage represents inter-cluster distance by the maximum distance between any two of the members of the clusters. In the average linkage method the distance between two clusters is calculated by taking the average distance between all members of the two clusters. The clusters can be cut at some predefined level on the basis of dissimilarity among the data points [6]. The process continues until the last two clusters are merged into a single cluster. The visualization of hierarchical clustering can be shown using a dendrogram.

Apart from traditional hierarchical clustering methods, a number of other hierarchy based clustering techniques have been proposed such as Robust Hierarchical Clustering Algorithm, ROCK [9]. ROCK is a hierarchical clustering approach for categorical data that merges the clusters on the basis of number of common neighbors. Some of the algorithms integrate hierarchical clustering with partitional clustering such as BIRCH (Balanced Iterative Reducing and Clustering using Hierarchies) [10], CURE (Clustering Using Representatives) [11], and Chameleon (Hierarchical Clustering Algorithm Using Dynamic Modelling) [12]. Some hierarchical algorithms such as DBSCAN (Density-Based Spatial Clustering of Applications with Noise) [13] use density instead of distance as a grouping function, and consider the high density areas as a cluster and the low density areas as outliers.

Unlike partitional clustering, hierarchical clustering does not need the number of clusters to be specified in advance. Hierarchical clustering provides a hierarchy of clusters as a solution while partitional clustering provides a single solution. In hierarchical clustering an element assigned to a cluster cannot be reassigned to another cluster in successive passes. While in partitional clustering a data element can go into different clusters in successive passes. Partitional clustering has lower execution time as compared to hierarchical clustering due to its lower algorithmic complexity [6].

Partitional clustering and hierarchical clustering have their own advantages and limitations in terms of generating numbers of clusters, generating different shapes, and overlapping boundaries of clusters. The exact number of natural groups in the data, initial choice of centroids, sensitivity to outliers, and algorithmic complexity are some of the issues which cause bottlenecks in the performance of a particular clustering technique.

Apart from their advantages, both techniques have deficiencies in terms of shape and structure of clusters, exact number of clusters, clustering configuration and degeneracy. To tackle these problems, optimization-based techniques have been investigated for data clustering. When optimization is involved in the process, it either uses an optimization technique as a data clustering algorithm or adds optimization to the existing data clustering approaches. Optimization-based clustering techniques treat data clustering as an optimization problem and try to optimize an objective function either to a minima or maxima. In the context of data clustering, a minimization objective function can be the intra-cluster distance and maximization can correspond to the inter-cluster distance. The results achieved so far from adding optimization to the data mining processes are promising. Optimization has significantly improved accuracy and efficiency while solving some other problems such as global optimization, multi-objective optimization and being trapped in local optima [14–17]. The involvement of intelligent optimization techniques has been found to be effective in enhancing the performance of complex, real time, and costly data mining process. A number of optimization techniques have been proposed to add to the performance of the clustering process. Swarm intelligence (SI) is one such optimization area where techniques based on SI have been used extensively to either perform clustering independently or add to the existing clustering techniques. The next section provides an overview of some of the most important SI-based optimization techniques.

Overall the paper has three main aims. Firstly, it provides the details of evolution of different PSO based clustering techniques and their growth in the literature. We report these results in Section 3 highlighting the number of relevant papers and citations to these papers which were published in the last decade. The second contribution of this paper is to overview the literature of PSO based data clustering. We have highlighted the most cited work in the area of hybrid-PSO clustering and stand-alone PSO clustering. The last objective of this paper is to provide the comparative results of different techniques along with a discussion on pros and cons of each technique.

2. Swarm intelligence and particle swarm optimization (PSO)

Swarm intelligence (SI), inspired by the biological behavior of birds, is an innovative intelligent optimization technique [18,19]. SI techniques are based on the collective behavior of swarms of bees, fish schools, and colonies of insects while searching for food, communicating with each other and socializing in their colonies. The SI models are based on self organization, decentralization, communication, and cooperation between the individuals within the team. The individual interaction is very simple but emerges as a complex global behavior, which is the core of swarm intelligence [20]. Although swarm intelligence based techniques have primarily been used and found very efficient in traditional optimization problems, a huge growth in these techniques has been observed in other areas of research. These application areas vary from optimizing the solution for planning, scheduling, resource management, and network optimization problems. Data mining is one of the contemporary areas of application, where these techniques have been found efficient for clustering, classification, feature selection and outlier detection [21,22]. The use of swarm intelligence has been extended from conventional optimization problems to optimization-based data mining. Section 3 shows the results of a survey that depicts a continuous increase in the number of papers about swarm intelligence and its variants.

A number of SI based techniques with many variants have been proposed in the last decade and the number of new techniques is growing. Among different SI techniques, Ant Colony Optimization (ACO) and Particle Swarm Optimization (PSO) are the two main techniques, which are widely used for solving discrete and continuous

optimization problems. Here we discuss the foundation of PSO followed by its popularity and year-wise growth in KDD literature.

Particle Swarm Optimization (PSO) is a swarm intelligence based metaheuristic algorithm proposed by Kennedy and Eberhart [23] which takes its inspiration from the cooperation and communication of a swarm of birds. The intelligence which emerges from such behavior causes the swarm to mimic complex global patterns. Below we describe general concepts of PSO.

In PSO, each individual in the swarm, called a particle, behaves like an agent of a highly decentralized and intelligent environment. Each particle of the swarm contributes to the environment by following very simple rules, thus cooperating and communicating with other particles of the swarm. A complex global collective behavior emerges in the swarm. This complex global behavior is exploited to solve a complex optimization problem. High decentralization, cooperation amongst the particles and simple implementation make PSO efficiently applicable to optimization problems [19,24,25].

PSO has three main components, particles, social and cognitive components of the particles, and the velocity of the particles. In a problem space where there may be more than one possible solution and the optimal solution of the problem is required, a particle represents an individual solution to the problem. The learning of the particles comes from two sources, one is from a particle's own experience called cognitive learning and the other source of learning is the combined learning of the entire swarm called social learning. Cognitive learning is represented by personal best (*pBest*) and social learning is represented by the global best (*gBest*) value. The *pBest* solution is the best solution the particle has ever achieved in its history. The *gBest* value is the best position the swarm has ever achieved. The swarm guides the particle using parameter *gBest*. Together cognitive and social learning are used to calculate the velocity of particles to their next position.

When applied to optimization problems, a typical PSO algorithm starts with the initialization of a number of parameters. One of the important initializations is selecting the initial swarm. The number of particles in the swarm depends upon the complexity of the problem. An initial choice of solutions is normally made randomly. However an initial guess that spreads the particles uniformly in the solution space can speed up the emergence towards an optimal solution. A typical initial number of particles for PSO in a swarm ranges from 20 to 40 but varies from application to application and problem to problem. The particles start moving from one position to another position in search of a better solution based on the social and cognitive components. The cognitive component *pBest* for minimization problems is calculated as

$$pBest_i(t+1) = \left\{ \begin{array}{ll} pBest_i(t) & \text{if } f(X_i(t+1)) \geq f(pBest_i(t)) \\ X_i(t+1) & \text{if } f(X_i(t+1)) < f(pBest_i(t)) \end{array} \right\} \quad (1)$$

where $X_i(t+1)$ is the particle's new position, $pBest_i(t)$ is the current personal best, and $pBest_i(t+1)$ is the new personal best position of the particle. The value of *gBest* comes from the social learning of the swarm which shows the best fit that any particle of the swarm has ever achieved. Eq. (2) calculates *gBest* for the same minimization problem

$$gBest(t) = \underset{i=1}{\overset{n}{\operatorname{argMin}}} \{f(pBest_i(t))\} \quad (2)$$

where n is the total number of particles. Together *pBest* and *gBest* combine to define the velocity of the particle which guides the particle towards a better solution. The velocity of a particle is thus calculated as

$$V_i(t+1) = \omega \times V_i(t) + q_1 r_1 (pBest_i(t) - X_i(t)) + q_2 r_2 (gBest(t) - X_i(t)) \quad (3)$$

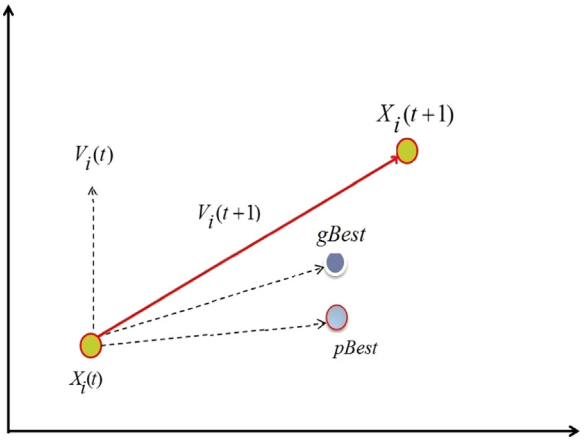

Fig. 2. Movement of a particle influenced by *pBest* and *gBest*.

where $V_i(t)$ represents the current velocity of the particle i, $V_i(t+1)$ represents the new velocity the particle will achieve to move from the current position to the new position. The range of velocities is bounded between V_{Max} and V_{Min} , where V_{Max} is the maximum velocity and V_{Min} is the minimum velocity. The parameters q_1 and q_2 are constants which weight the social and cognitive components, $r_1$and r_2 are two random numbers ranging from 0 to 1, and ω is the inertia of the particle. Velocity added to the current position provides the new position of the particle which is given by

$$X_i(t+1) = X_i(t) + V_i(t+1) \quad (4)$$

The graphical representation of the particle repositioning is shown in Fig. 2. It shows how, *pBest* and *gBest* affect the particles movement from position $X_i(t)$ to $X_i(t+1)$.

Algorithm 1 shows the pseudocode for PSO-based problem solving. Three key steps to implementing PSO for optimization are selection of an objective function(s), representation of the solution space and selection of a stopping criteria. The selection of an objective function depends on the type and nature of the problem. In multi-objective optimization more than one objective function could be used to check the fitness of the swarm. Representation of the solution is done by coding the particle to the solution of the problem space which includes specifying attributes of a particle and a movement mechanism from one place to another place. The stopping criteria of the algorithm can be the maximum number of iterations PSO has performed, accepted error values or some other criteria related to the fitness of the swarm.

Algorithm 1. PSO for finding minima.

Input: function $\int(x)$
Output: *Minima M*
Parameters: Swarm Size *S*, V_{Max}, V_{Min}, ω,q_1,q_2, and number of records *N*
Method:
1: INITIALIZE *S*,V_{Max}, V_{Min}, ω,q_1,q_2, and *N*
2: **for** Each Particle *X* **do**
3: INITIALIZE X_i
4: **end for**
5: **while** (STOPPING CRITERIA(false)) **do**
6: **for** each generation of swarm *S* **do**
7: **for** each iteration
8: CALCULATE *gBest* from Swarm using Eq. (2)
9: CALCULATE Velocity V_i using Eq. (3)

```
10:        if Velocity is greater than V_Max then
11:          SET Velocity = V_Max
12:        end if
13:        if Velocity is less than V_Min max
14:          SET Velocity = V_Min
15:        end if
16:        UPDATE Position X_i(t)
17:      end for
18:      for Each Particle X do
19:      CALCULATE and update pBest from Swarm using Eq. (1)
20:      end for
21:    end for
22: end while
```

3. Popularity and growth of the literature in swarm intelligence

Due to its simplicity and extendibility to different domains, the study and usage of SI techniques have tremendously increased in recent years. An enormous rise in the number of papers published and the number of citations of the SI-based optimization techniques have been recorded. To highlight this trend we conducted a systematic literature review of the area, basing our search on the Scopus database. The database has a variety of options to retrieve data. It supports regular expressions and sophisticated query design and processing using a variety of subjects such as authors, articles, titles, and temporal searches. Various queries were designed to retrieve information from Scopus (http://www.scopus.com) as shown in Appendix A. For the results shown in Fig. 3, a query was designed that retrieves those papers which are related to PSO-based data clustering using keyword based retrieval from the titles of the papers. We limited our subject domains to computing, engineering and mathematics as shown in Query A.1 of Appendix A. The survey of Scopus represented in Fig. 3 shows a dramatic annual increase in the literature published in the area of PSO-based clustering over the last 4 years. The next few paragraphs highlight the type of research in the area over the period of 2002–2011.

The literature reported above does not focus solely on numeric data clustering but also extends to other data clustering domains. Fig. 4 shows a subject based distribution of the literature found for PSO clustering. We categorized the literature into different PSO-based clustering areas, which include data clustering, sensor network clustering, image segmentation and clustering, gene clustering, document and text clustering, and hybrid clustering. We extracted these categories by manually examining the literature, and selected six major categories based on the number of papers retrieved. The number of papers in a category was the main criteria while the nature of techniques used for clustering, and application areas were used as secondary categorization criteria.

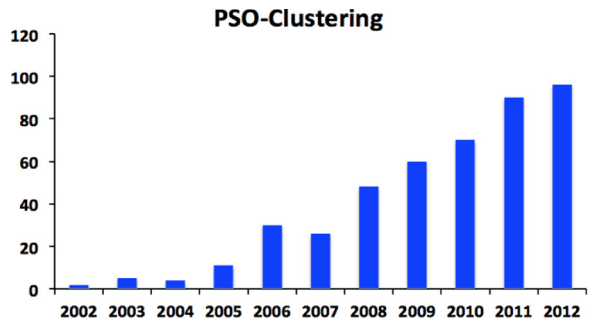

Fig. 3. Clustering using particle swarm optimization.

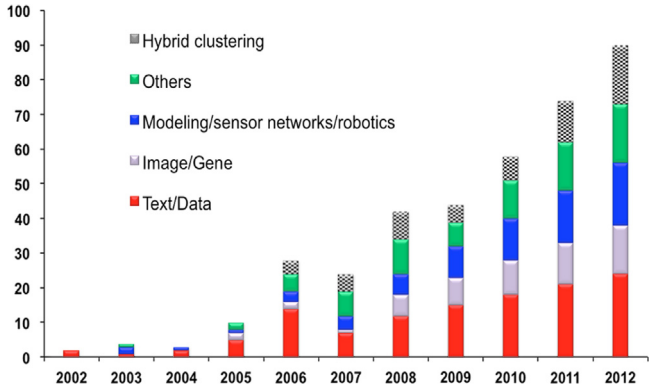

Fig. 4. Year-wise growth of the subject area.

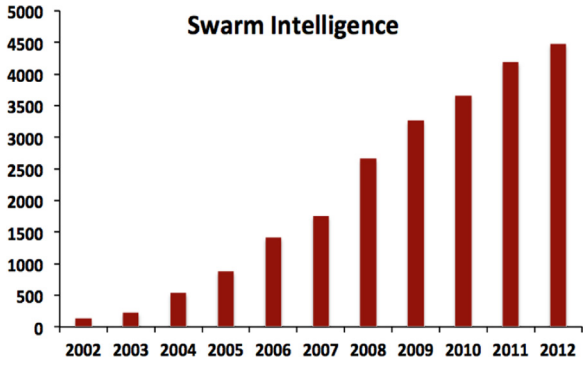

Fig. 5. Literature in swarm intelligence.

Papers which did not fit into any of these categories were termed as others. Some of the areas are overlapping e.g. hybrid clustering methods, which could be applied to data as well as image clustering. We observed that in some cases the hybrid clustering methods are used for modeling purposes and other times for optimization purposes. From Fig. 4 we can see that text, numeric, and categorical data clustering are the highest contributors to the literature of PSO-based clustering over the given time period. Sensor network clustering is an emerging area and a number of studies have been conducted to tackle different aspects of the problem. One reason for using a PSO-based clustering approach is that the nature of the sensor network is similar to the origin of PSO. Each sensor in the network is treated as an agent/particle and different characteristics of that particle are then used to group the sensors into clusters. Some of the papers which do not fall into any of the categories were marked as “others”, including clustering based modeling and fuzzy clustering.

Fig. 4 shows year-wise growth of each subject area text/data clustering, image and gene data clustering, clustering in sensor networks and robotics, and hybrid clustering which includes cross-domain clustering. Text and numeric data clustering leads in the number of papers published annually. Image clustering and gene data clustering have also shown consistent growth highlighting the importance of PSO clustering in these areas. Clustering in sensor networks and other engineering domains is also one of the growing application areas of PSO clustering.

In recent years the implementation of swarm intelligence has been extended from conventional optimization problems to optimization-based data mining. Fig. 5 shows the results of a survey, showing a continuous increase in the number of papers about swarm intelligence. On average there is an increase of more than 90% each year since 2000. This literature includes swarm

intelligence in KDD, modeling, sensor networks optimization and image processing.

The overall statistics show that the contribution of PSO to swarm intelligence literature is more than any other SI based technique, which suggests the importance of PSO and its simplicity and applicability to different application domains. The average increase for the duration from 2002 to 2011 in PSO is above 50% annually.

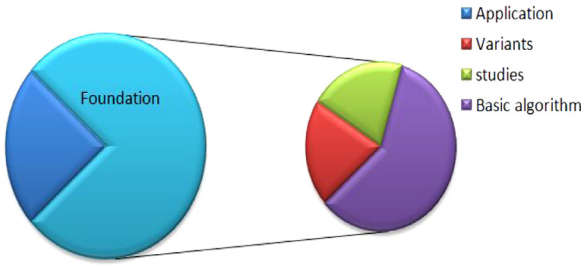

Fig. 6. Citation distribution in PSO.

4. Application areas of PSO

This section highlights different application areas of PSO. Most of the clustering implementations using PSO are in the area of clustering of numeric and text data. An average annual increase of about 50% in the number of papers published is recorded in the last four years, as shown in Fig. 3. Some other KDD areas where PSO has been implemented are PSO-based outlier detection [26], PSO-based classification and association rule mining [27–30], particle swarm based feature selection [31,32], PSO-based text clustering [33], PSO-clustering based recommender systems [34–38], and prediction analysis [39].

One of the important fields where there has been a large increase in recent years is sensor networks. The literature shows a wide implementation of PSO in different areas of sensor networks which includes localization of wireless sensor networks [40,41], optimization and control [42], network coverage [43], and routing and clustering of nodes [44,45]. Swarm robotics is another important area where PSO has successfully been implemented in robot path/motion planning [46], robot collaboration and communication, robot learning and decision making [47,48], and source localization [49]. Image segmentation and image clustering were some of the first application areas of particle swarm optimization [50]. Some of the other PSO-based image processing domains include edge detection, image retrieval and mining, noise removal, image feature selection and image classification. Recently a survey has been conducted [51] that outlines one of the most rapidly growing area of PSO based high dimensional clustering. A number of PSO based cluster techniques have been reviewed which help to enhance the efficiency and accuracy of existing clustering techniques.

Citation counts are one of the factors which can measure the importance and growth of a research field. The next survey takes into account the citation counts in the field of particle swarm optimization. The results of a Scopus query show that the most cited papers are about the foundations of PSO. A classification of the research work into foundation/basics of PSO and application areas of PSO reveals that the foundation literature of PSO has gotten more than 20,000 citations in total, while some individual work [23] has more than 12,000 citations alone. On the other hand, papers about the application of PSO have not been cited as much as papers about the foundations of PSO. However, the most cited papers about the application of PSO are in the field of engineering, which include electromagnetism, voltage control, controller design, multi-objective optimization and task assignment.

Fig. 6 shows the distribution of the citations where 23% of the total citations of PSO belong to applications of PSO while the remaining 77% of the citations are related to the foundation of the PSO algorithms. To further investigate the foundation of PSO, 60% of the literature is related to the origin of basic PSO techniques, 20% of the citations are related to various studies of PSO including surveys and reviews, and the remaining 20% cite the different variants of PSO along with their explanations and implementation.

5. Particle swarm optimization based data clustering

Particle swarm optimization was first used by Van der Merwe and Engelbrecht [17] for data clustering, where randomly created particles were mapped to one data vector. Each particle in that data vector is a representative of one centroid. The particles then moved to a better centroid position during each iteration. The evaluation of the method was based on the cost function that evaluates each candidate solution based on the positions of the proposed clusters' centroids. Xiao et al. [52] hybridized PSO with Self-Organizing Maps (SOM), using SOM for clustering the data and PSO for optimizing the weights of the SOM. In Chen and Ye's approach [14] each particle corresponds to a vector containing the centroids of the clusters. The results were compared with K-means and fuzzy C-means using the objective function which takes into account the intra-cluster distance. A recent similar work [53] reports the results of k-means added with PSO and multiclass merging to perform data clustering. In the image processing domain, Omran et al. [50] suggested a dynamic clustering algorithm based on PSO and K-means for image segmentation. The proposed approach finds the number of clusters in the data automatically by initially partitioning the dataset into a large number of clusters, and the numbers of clusters are optimized using binary PSO. Cohen and De Castro [54] proposed a Particle Swarm Clustering (PSC) algorithm with a fixed number of centroids on a benchmark dataset. Instead of coding the whole solution as a single particle, each particle represents a portion of the solution. The proposed solution is not based on the use of a fitness function, instead it moves the particles to the natural clustering position. The comparison results show that PSC performs better than K-means clustering on the basis of classification error. Another use of PSO clustering is found in the web usage domain where Chen and Zhang [55] cluster web sessions using PSO, combining improved velocity PSO with K-means. The approach is based on enhanced particle search and updating the centroids with the help of K-means. Another similar work [56] recently reported where PSO is used for data clustering independently without being hybridized with any other clustering technique. The authors [57] proposed a K-Harmonic and PSO-based data clustering algorithm to help the K-Harmonic mean algorithm escape from local optima. The authors [58] implemented a K-NM-PSO clustering technique which combines PSO Nelder–Mead (NM) simplex method and the K-means clustering technique. The experiments were run on different benchmark datasets from the UCI machine learning repository [59], demonstrating the robustness of the approach. A bacterial evolutionary algorithm-based strategy [16] was applied to cluster real and synthetic data. The authors [60] proposed Multi-Elitist PSO (MEPSO), a modified version of classical Particle Swarm Optimization (PSO), which employs a kernel-induced similarity measure instead of the conventional distance measure. The proposed technique can find the optimal number of clusters automatically without specifying it in advance. Another similar work has been proposed [61] based on

Differential Evolution (DE). Recently [62–64] PSO has been used to optimize the parameters for subtractive clustering (SC), and PSO hybridised with SVM for feature selection and parameter optimization in the area of predicting bankruptcy and corporate financial distress.

The related work has addressed some of the common aspects of data clustering, such as enhancing the efficiency and improving the accuracy of the techniques, however most of the approaches have ended up either having as good results as the existing techniques or with some minor improvements in one or other aspect. There is still a need for a technique which is efficient and produces accurate results. Fixing the efficiency problem is still an open question. The efficiency and accuracy trade off will be leading the discussion in clustering in the coming years [65]. The next section highlights some of the techniques mentioned above and explains their working principles to tackle the data clustering problem.

We divide PSO-based clustering techniques into two general categories. The first group includes those techniques which use PSO as part of another clustering technique in a hybrid manner. Such techniques use PSO for parameter selection, parameter optimization, and centroid selection or updating. The first part of this section highlights such techniques. The second part of this section explains those techniques where PSO has been used independently for data clustering. The discussion highlights the modeling of the problem for the specified PSO-based technique, selection and initialization of the parameters, number of particles in the initial swarm, and performance measures selected to evaluate the technique. Detailed comparison results of these techniques are discussed in Section 6 with comments on their pros and cons.

5.1. PSO hybridized for data clustering

PSO has been hybridized with a variety of different clustering methods such as K-means, K-Harmonic mean, self organizing maps, and neural networks. It is also used in basic PSO form as well as in discrete PSO form to cluster data. The data which has been used for testing and validation purposes include numeric, alphabetic, microarray and image data. This subsection contains those techniques where PSO has been hybridized with other existing clustering techniques.

PSO and K-means: The credit of starting a research initiative towards PSO-based data clustering goes to Van der Merwe and Engelbrecht [17], who presented the idea of using PSO with K-means clustering for refining the K-means clustering technique. The approach which they presented, uses a fixed number of particles as a swarm. Each particle represents a different clustering solution by assigning centroids of all the clusters to a single particle. Initial particle assignment was done in a random manner. The approach is different from K-means as K-means starts with one solution and tries to optimize that solution in successive iterations, while in this approach the clustering starts from several candidate solutions and the best among them is selected in each iteration. In each successive iteration the particles generate another set of solutions and this continues until it reaches the final best achieved solution. They also hybridized K-means with PSO to initialize the particles and show how performance of the clustering process can be improved by taking the initial seed from K-means and feeding it into PSO clustering. The authors have presented the comparisons of their results with stand alone PSO clustering and PSO hybrid K-means clustering. Evaluation of the method was based on the quantization error that evaluates each candidate solution based on the proposed cluster's centroids. The results will be described in detail in Section 6.

Dynamic PSO (*DCPSO*) *and K-means*: Another hybrid approach was introduced by Omran et al. [50], using a Dynamic Clustering algorithm based on PSO (DCPSO). The proposed approach is a hybridization of PSO and K-means where PSO performs clustering and K-means performs refinement of the clusters. Unlike the previous approach, this approach uses binary PSO. It also finds the number of optimal clusters as compared to the previous approach where the number of clusters needed to be specified in advance. The algorithm initializes swarm parameters such as velocity, number of particles, and initial partition of the dataset. It then calculates the fitness of the particles and updates the velocity and position. The process is performed iteratively until the stopping criteria is met. The approach is capable of automatically detecting clusters in images, where the clusters are well separated, completely overlap or partially overlap. In the proposed approach, large number of clusters were generated and then optimized to a better clustering solution using binary PSO. The clusters are then refined using K-means. The approach was validated using three different validity indices. DCPSO was tested for image segmentation on benchmark natural image datasets and some synthetic images.

Improved velocity PSO with K-means (*RVPSO-K*): Although RVPSO-K also hybridizes K-means with PSO, the technique is different from the two techniques mentioned above. This technique is based on changing the flying trajectory of the particles of the swarm. The authors [55] added random velocity to the particles to enhance the coverage of the swarm in all directions. Centroids of the clusters were updated using the same K-means approach used in the previous work [17]. Stability of the swarm, accuracy of clustering, and convergence speed were the performance measures for the experiments. To the best of our knowledge it is the first work reported, which uses web session data for the evaluation of the experimental results. The authors used web logs and selected user visit attributes for clustering. Although the comparison is done only with K-means, the work defined a new application area, namely the application of PSO-based data clustering for web usage clustering.

PSO, Nelder–Mead (*NM*) *simplex method and K-means*: So far we have studied hybridization of PSO with K-means and have found that the hybridization produces slightly better results in terms of accuracy. Kao et al. [58] proposed an innovative approach which is based on the combination of PSO, Nelder–Mead (NM) simplex method, and the K-means clustering technique. The authors exploited the efficiencies of PSO and NM simplex methods. The NM simplex method is efficient in local search and K-means has low computational cost. The authors overcome the shortcomings of PSO's high computational cost and poor local search behavior by using this hybridization. Because of the insensitivity of PSO to initial clustering centroids and its accuracy, they overcome the initialization sensitivity of the NM simplex method and inaccuracies of K-means. The work reports comparison results for accuracy, intra-cluster distance, and function evaluations of the proposed approach to PSO clustering, K-means clustering, NM-PSO clustering, and K-PSO.

PSO and K-Harmonic means (*PSO-KHM*): In one of the most recent papers in this area [57], the authors proposed a K-Harmonic means (KHM) and PSO-based data clustering algorithm to help the K-Harmonic means algorithm escape from local optima. As discussed earlier, one of the deficiencies of K-means clustering is that it gets trapped in local optima. As PSO is a global stochastic algorithm, it is capable of escaping from local optima. The initial centroid sensitivity is handled by KHM and PSO. KHM also solves the problem of slow convergence of PSO for data clustering. Comparison results for the technique are given in the next section.

PSO and Self Organizing Maps (*SOM*): The work presented by Xiao et al. [52] is another example where PSO is used alongside

another clustering technique. In their proposed SOM/PSO approach, PSO was hybridized with self-organizing maps (SOM) to improve the efficiency of the clustering process. The approach comprises two different phases. Initially SOM are used to cluster the data and then PSO optimizes the weights of the SOM and refines the clusters. The results of the hybrid PSO/SOM were compared with SOM and PSO clustering which shows the efficiency of the technique. The experiment was performed on two well known gene expression datasets, however there are no experiments on the machine learning datasets that are typically used for evaluating data clustering techniques.

5.2. PSO as a data clustering method

Apart from PSO being used as a part of another clustering technique, there is also some work which uses PSO as a stand alone clustering technique. This section discusses different clustering techniques that are solely based on PSO, and do not hybridize PSO with any other clustering technique.

PSC clustering: Cohen and De Castro [54] proposed a novel approach that uses PSO as an independent clustering technique, where the centroids of the clusters were guided by the social and cognitive learning of the particles. As opposed to earlier versions of PSO-based clustering techniques, in this approach each particle represents a portion of the solution instead of representing the whole solution as a single particle. The proposed solution is not based on the use of a fitness function, instead it moves the particles to the centroid of the cluster using the cognitive and social learning of the particles, which cause optimization in the intra-cluster distance. To validate the approach the authors applied the PSC algorithm with a fixed number of centroids on benchmark datasets. PSC performs better than K-means clustering in terms of accuracy of clustering, however the computational cost is not reported.

PSO clustering: Chen and Ye [14] employed a representation in which a particle corresponds to a vector containing the centroids of the clusters. The approach is similar to another proposal [17], where one particle is representative of the clustering solution and compared to the previous work, there is no hybridization with any other algorithm, so the execution time which is not reported, must be higher. The best particle among the swarm is chosen to represent the solution of the problem. The experiments were done on artificial datasets and the results were compared with K-means and fuzzy C-means using an objective function that takes into account the intra-cluster distance.

Evolutionary Particle Swarm (EPSO) for web usage clustering: The authors [66,67] customized PSO and EPSO for web usage clustering. Standard benchmark web usage data was used for the experiments. The web usage data clustering is different from traditional data clustering as the data needs a sophisticated preprocessing stage before it can be grouped into clusters. Another important thing in such data is the selection of appropriate attributes for clustering. The authors have discussed the related work and the significance of the PSO-based clustering approach. The approach they used is based on moving centroids to their natural position based on the related data and the approach is used without any hybridization with any other clustering technique.

Hierarchical PSO clustering (HPSO-Clustering): In this technique the deficiency of the traditional partition based clustering i.e. initialization of particles, trapping in local optima and a lack of the domain knowledge are tackled [68]. On the other hand hierarchical clustering approaches have disadvantages of efficiency and premature clustering of objects into different clusters. The combination of both approaches, partition based techniques that are relatively efficient, and hierarchical clustering techniques that are accurate, could give better results. The authors combined these techniques and added swarm intelligence to the process to give the novel PSO-based hierarchical agglomerative data clustering technique (HPSO-clustering). HPSO-clustering is based on the modeling of each cluster centroid by an individual particle, and so the complete swarm represents the solution of the clustering problem. The number of particles is kept large for the maximum coverage of the problem space. The technique works in an agglomerative manner starting from a relatively large number of particles and combining down to only one final particle. Initially each particle has a small number of associated data vectors while the final particle contains the entire data set. The first generation particles adjust their positions by iterating them for a particular number of iterations. The transition of swarm from one generation to another generation merges two of the selected particles and transforms the swarm into a smaller swarm.

Regardless of how the techniques work, and what their strengths and weaknesses are, the use of PSO in data mining and particularly in data clustering is increasing. A number of applications of these techniques have been reported in the literature which verify the applicability and suitability of PSO for data mining applications. Our own work [15,66,67,69] is one of the promising outcomes of PSO-based data clustering.

6. Clustering comparison of selected techniques

PSO-based clustering has been studied by different researchers for a variety of different application areas. In this section we will overview some of such work and present their results while also discussing the pros and cons of their approaches.

We select different performance comparison measure such as *inter-cluster distance* (separation of clusters), *intra-cluster distance* (compactness of clusters) , *number of function evaluations*, *quantization error* (classification error), and *accuracy* (correctness in clustering configuration). Inter and intra-cluster distances are indicators of how good the clusters are in terms of the position of each data element within its corresponding cluster as well as against the elements of different clusters. Low intra-cluster distance is better than high intra-cluster and vice-versa for inter-cluster distance. Quantization error is a measure that shows the classification accuracy of the clustering technique. Lower quantization error is an indicator of good accuracy. Function evaluations are an alternative to the execution time which shows the efficiency of the clustering techniques. The lower the number of function evaluations the better the efficiency of the algorithm. Overall, these measures describe the validity of the clustering approach as well as the attributes of the datasets. Some of the common datasets which are used for the validation and testing of data clustering techniques are listed in Table 1.

For comparison purposes we chose the widely used UCI machine learning datasets [59] which are freely available to be used for testing and validation. These datasets have many types of

Table 1
Description of commonly used datasets.

Name of dataset	No. of classes	No. of features	No. of observations
Vowel	6	3	871
Iris	3	4	150
Crude oil	3	5	56
CMC	3	9	1473
Breast cancer	2	9	699
Glass	6	9	215
Wine	3	13	179
Ruspini	4	2	75

overlapping behaviors such as no overlap, medium overlap and extreme overlap among the data items of different clusters. Many of the researchers from the KDD and machine learning community have used these datasets as benchmarks for data clustering problems for testing efficiency and accuracy measures. Apart from these benchmark datasets, there are other artificial datasets which are used for testing PSO-based clustering techniques for applications other than numeric data clustering. Such applications include text clustering, image clustering and segmentation, sensor network clustering and gene clustering.

For our experimentation we will comment on the results obtained using independent PSO, hybridized PSO and other relevant techniques. The experimental design consists of four main phases. In the first phase inter and intra cluster distances of the selected techniques will be assessed. In the second phase, we will highlight the function evaluations and execution time of different PSO based clustering techniques, in the third phase accuracy will be assessed while in the last phase cross validation on one of the datasets will be performed for verification of the results.

The scope of the experiments is limited to the measures mentioned above. We have omitted some of the results which we could not verify and presented only those results which could be compared on standard measures as discussed above. Some of the results do not include all of the datasets due to this assumption.

The first PSO-based data clustering by Merwe and Engelbrecht [17] as discussed in Section 5 was compared with the K-means clustering technique. Table 2 shows a summary of the results reported in their work which highlights that PSO-based clustering is better than K-means in terms of inter-cluster (interClus) distance and quantization error (Qnt.Err). The Hybrid PSO with K-mean as shown in the last column performs better on quantization error. Although this approach has a weakness in that it suffers in efficiency compared with K-means, it is still very important due to its novelty, pioneering the research in this direction, and in terms of accuracy. The work has led to some of the very important work that is published later and based on their work has better results than those given here. One of the strengths of this work is to highlight the capability of PSO-based hybrid clustering to converge to lower quantization error based on the fitness function of PSO.

K-NM-PSO [58] is another recent work which hybridizes PSO and K-means based on the Nelder–Mead(NM) simplex method for the local search. Results for the comparison of intra-cluster distance are given in Table 3. The PSO clustering variant they used for comparison is described by Chen and Ye [14] and outlined in Section 5.

The K-NM-PSO results shown in bold show an improvement in the accuracy of clusters using this approach compared with some contemporary clustering algorithms such as K-means and PSO-based data clustering. The number of function evaluations for each approach is given in Table 4. The K-NM-PSO approach needs fewer function evaluations than PSO to reach the optima due to its hybridization with Nelder–Mead (NM) simplex method.

The approach suffers in efficiency in terms of number of function evaluations compared with K-means. K-means needs fewer function evaluations as compared to K-NM-PSO. We have observed this kind of deficiency in almost all PSO-based clustering techniques.

Another recent work in hybridized PSO-based data clustering is PSOKHM [57] where PSO and K-Harmonic means (KHM) are combined to solve some of the problems that PSO aided K-means algorithms suffer from. The results were compared against KHM and PSO-based data clustering. Table 5 outlines the results with respect to $KHM(X,C)$ measure which is

$$KHM(X,C)=\sum_{i=1}^{n}\frac{k}{\sum_{j=1}^{k}\frac{1}{\|x_j-c_i\|^p}}$$

where X is the dataset to be clustered ranging from 1 to n, C is the set of centroids ranging from 1 to k, and p is an input variable for getting good objective functions. $KHM(X,C)$ has been used to evaluate the quality of the clustering solution based on the ratio of number of clusters to the intra-cluster distance. Again the results show that the hybridization has sufficient improvement over KHM and PSO in terms of accuracy and compactness of clusters but the efficiency in terms of execution time is quite low as shown in Table 6. The efficiency in terms of execution time (s) given in Table 6 shows that PSOKHM has sufficiently better execution time as compared to PSO but suffers against KHM.

The techniques which we have looked at so far have one commonality in the results that when PSO is hybridized with any other technique, it performs better in terms of intra-cluster distance, inter-cluster distance, accuracy and quantization error. However, traditional partitional techniques still perform better on efficiency in terms of function evaluations as well as execution time.

Now we will discuss the work where there is no hybridization of PSO with another clustering approach. The approach, PSC [54],

Table 2
Inter-cluster distance and quantization error [17].

Dataset	K-means		PSO		Hybrid	
	interClus	Qnt.Err	interClus	Qnt.Err	interClus	Qnt.Err
Iris	0.88	0.64	0.88	0.77	0.85	**0.63**
Wine	1.01	1.13	2.97	1.49	2.79	**1.07**
Breast cancer	1.82	1.99	3.54	2.53	3.33	**1.89**

Table 3
Total intra-cluster distance for the given datasets [58].

Dataset	K-means	PSO	K-NM-PSO
Vowel	159,242.87	168,477.00	**149,141.40**
Iris	106.05	103.51	**96.67**
Crude oil	287.36	285.51	**277.29**
CMC	5693.60	5734.20	**5532.70**
Cancer	2988.30	3334.60	**2964.70**
Glass	260.40	291.33	**200.50**
Wine	18,061.00	16,311.00	**16,293.00**

Table 4
Function evaluation for KNM-PSO [58].

Dataset	K-means	PSO	KNM-PSO
Vowel	**180**	16,290	9291
Iris	**120**	7260	4556
Crude oil	**150**	11,325	7057
CMC	**270**	36,585	21,597
Cancer	**180**	16,290	10,149
Glass	**630**	198,765	119,825
Wine	**390**	73,245	46,459

Table 5
Comparison of **KHM(X,C)** based on values of $p=3.5$ [57].

Dataset	KHM	PSO	PSOKHM
Iris	113.413	255.763	**110.004**
Glass	1871.812	32,933.349	**1857.152**
Cancer	243,440	240,634	**235,441**
CMC	381,444	423,562	**379,678**
Wine	8,568,319,639	3,637,575,952	**3,546,930,579**

Table 6
Run time of KHM, PSO, and PSOKHM.

Dataset	KHM	PSO	PSOKHM
Iris	0.190(0.007)	3.096(0.010)	**1.826(0.009)**
Glass	**4.042(0.007)**	43.594(0.338)	17.609(0.015)
Cancer	**2.027(0.007)**	16.150(0.144)	9.594(0.023)
CMC	**8.627(0.009)**	148.985(0.933)	39.485(0.056)
Wine	**2.084(0.010)**	35.284(0.531)	6.598(0.008)

Table 7
Classification error for PSC and K-means

Dataset	PSC	K-means
Iris	**7.68**	15.64
Breast cancer	**4.31**	4.14
Glass	**46.26**	48.84

Table 8
Accuracy for HPSO clustering and HAC.

Dataset	HPSO Accuracy	HAC Accuracy
Iris	**92.00**	74.6666
Wine	**70.7865**	61.2359
Breast cancer	96.33	96.33
Glass	**54.6728**	37.3831
Vowel	**47.6463**	45.35017

is based on moving the centroids into better positions in a cluster-based on the cognitive and self organizing of the particles, which comes from the experience of an individual particle rather than the complete swarm. Results on the three benchmark datasets are given in Table 7. Classification error, which is putting the data into the wrong cluster, is used to compare the clustering techniques. The results presented in Table 7 show an improvement over the K-means clustering technique. The results for function evaluation or efficiency on execution times are not mentioned. It would be interesting to look at the result of efficiency as the PSC algorithm may have better execution time compared to other PSO-based clustering approaches.

Recently we proposed using PSO for data clustering without hybridizing it with any other clustering algorithms. The preliminary results of our EPSO-clustering and HPSO-clustering are reported [15,69]. In HPSO-clustering the clustering process initially starts with a large number of clusters and in each subsequent generation only one cluster is consumed to generate a hierarchy of clusters. The merging process continues until it reaches the final single cluster. HPSO-clustering is an extension of EPSO-clustering. HPSO-clustering was tested on seven different datasets against intra-cluster distance, inter-cluster distance, accuracy, efficiency and error rate.

The first experiment compares the accuracy of our proposed technique with hierarchical agglomerative clustering (HAC). The parameters for the experiments were set to, $S=20$–50, $vMax=0.1$–1.0, and max iteration per generation $=100$. Table 8 reports the results of the accuracy of HPSO-clustering and traditional HAC. The results shown in bold verify the improved accuracy of HPSO-clustering against HAC.

The second experiment was carried out to test the accuracy of the approach against partitional and optimization-based clustering approaches. We selected the base algorithms K-means and PSO-clustering. The comparison with K-means was made possible by selecting that generation of HPSO-clustering that matches with the number of clusters of K-means. Table 9 reports the accuracy and standard deviation on accuracy for 20 runs. The accuracy of HPSO-clustering is better than PSO-clustering and K-means clustering on most of the given datasets.

To evaluate the consistency and efficiency of HPSO-clustering in terms of the number of data points, we scaled the CMC data to 250 thousand observations and 2000 attributes. Fig. 7 shows the consistency and efficiency of the HPSO-clustering technique. Execution time is measured against the parameters mentioned and the proposed approach is consistent in execution time with the varying parameters. The reason for scaling the CMC dataset was that it is a benchmark classification data and the ground truth for the accuracy is already known. From the figure we observe the linear growth of time against each of the varied parameter. We did not observe exponential growth in execution time against the number of observations and dimensions.

Table 10 analyzes the Ruspini dataset and cross validates it using different measures such as inter-cluster, intra-cluster and total fitness with mean and standard deviation. The data was manipulated in two different ways, first the centroids were randomly selected and secondly the data was shuffled to make it random. Values reported in the first column represent the best, worst and mean values of fitness when random centroids were selected. The second column reports the values when the dataset was shuffled and the last column shows 10×10 cross validation for fitness when the centroids were randomly initialized. The purpose of this experiment is to show the consistency of HPSO-clustering with different configurations of particles as well as datasets.

The results described in this section have highlighted the trade-offs of efficiency and accuracy of output of the clustering process. Almost all of the PSO-based clustering techniques, whether in stand alone form or in a hybrid form, have an improved accuracy over traditional partitional clustering approaches such as K-means and K-harmonic based clustering. A number of variants to the basic PSO clustering algorithm have been proposed and have resulted in improved efficiency and accuracy. Although, the efficiency of PSO-

Table 9
Comparison of accuracy and StdDev. of accuracy for PSO and K-means clustering.

Dataset	HPSO		K-means		PSO	
	Accuracy	StdDev.	Accuracy	StdDev.	Accuracy	StdDev.
Iris	**90.0333**	1.13374	84.36	8.75	87.47	5.38
Breast cancer	**96.2005**	0.03273	95.86	0.46	94.89	1.32
Wine	70.7584	0.12562	68.88	0.71	**71.29**	0.41
Vowel	**47.6463**	1.29002	45.69	2.15	44.65	2.55
Glass	52.7336	1.66269	51.16	2.41	**54.41**	15.62

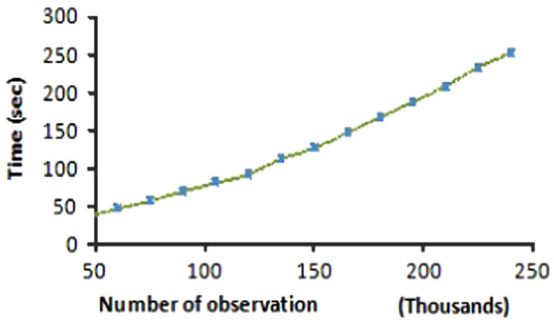

Fig. 7. HPSO-clustering execution time (scaled CMC dataset).

Table 10
Cross validation of Ruspini dataset.

Measures	Random centroids	Random datasets	Cross validation
Intra cluster			
Mean	11.40	11.42	11.34
StdDev.	0.17	0.02	0.26
Best	11.39	11.39	10.80
Worst	11.45	11.45	11.85
Fitness			
Mean	856.85	858.57	761.54
StdDev.	0.16	1.22	18.68
Best	855.00	856.03	721.88
Worst	870.00	860.98	803.33
Inter cluster			
Mean	1079.03	1078.28	1080.77
StdDev.	2.34	4.59	7.27
Best	1092.00	1085.81	1094.24
Worst	1080.23	1070.52	1065.24

based clustering has been considerably improved, it still has poor execution time as compared to partitional clustering techniques.

7. Future work

From the current literature survey in the field of PSO, we notice that there is an increasing trend in the number of papers and citations published in the area. There is evidence that the future of PSO will mostly be dominated by the research in the implementation of PSO in different application domains rather than in the foundations of PSO. Nonetheless there are still some remaining issues in the foundation of existing PSO techniques that still need to be investigated thoroughly. In this regard, researchers are concentrating on two major areas, firstly automation of the techniques, and secondly generalization of PSO based algorithms. These two issues are the core of all problems that optimization based techniques encounter.

The problem of automating different processes and parameters in PSO needs to be addressed such that the techniques can be applied to different application areas with less or no domain knowledge. Current PSO algorithms require us to tune a range of parameters before it is able to find a better solution. Tuning the parameters for different problems and applications leads to the second problem of generalization of the technique. Generalized parameter values and learning components are required so the approach can be used in different problem domains and so better results can be obtained.

In regards to the future work in application areas of PSO, there is still a gap in testing and validation of these techniques, in spite of the large amount of research that has been carried out in PSO based KDD techniques. New areas where such techniques can perform better need to be explored. Thorough testing of these techniques on real data instead of benchmark synthetic data, and validation of the results on the same measures that traditional data mining techniques have used are another future research direction.

8. Summary and conclusion

In this paper we have discussed the evolution of clustering techniques based on Particle Swarm optimization. We systematically surveyed the work, and presented the results of increasing trends in the literature of swarm intelligence, Particle Swarm Optimization and PSO-based data clustering. The literature survey (Sections 2 and 3) and the comparison of results (Section 6) are evidence that there is an enormous increase in the popularity of such techniques. The techniques are novel, collaboration and communication based, and simple to implement. PSO has received prompt attention from optimization-based data mining researchers. PSO-based data clustering and hybrid PSO clustering techniques have outperformed many of the contemporary data clustering techniques. The approach has a tendency to be more accurate and to avoid getting trapped in local optima. PSO-clustering, PSC clustering, EPSO clustering and HPSO-clustering are some of the popular techniques tested on benchmark datasets.

We have also outlined different application areas of PSO relevant to clustering. Each application area has its own requirements and conditions. PSO is simple enough to model and capable of being used for diverse new application domains. Scalability of PSO and its variants allows them to be modified for different application areas. The research also highlights the fact that the past was mostly dedicated to the theory of the foundations of PSO and the study of different variants of PSO, whereas in the future an increase in work is expected in the application areas of PSO.

Appendix A. Scopus queries

A.1. Query to retrieve papers related to PSO clustering

(TITLE(particle swarm optimi*) OR TITLE(pso) AND TITLE (cluster*)) AND PUBYEAR IS 2010 AND (LIMIT-TO(SUBJAREA, “COMP”) OR LIMIT-TO(SUBJAREA, “ENGI”) OR LIMIT-TO(SUBJAREA, “MATH”) OR LIMIT-TO(SUBJAREA, “DECI”) OR LIMIT-TO(SUBJAREA, “MULT”))

A.2. Query to retrieve papers related to swarm intelligence

(TITLE(swarm intelligence) OR TITLE(particle swarm optimi*) OR TITLE(pso) OR TITLE(ant colony optimi*) OR TITLE(aco) OR TITLE(bee colony*)) AND PUBYEAR IS 2009 AND (EXCLUDE(SUBJAREA, “ENVI”) OR EXCLUDE(SUBJAREA, “BIOC”) OR EXCLUDE(SUBJAREA, “ENVI”) OR EXCLUDE(SUBJAREA, “BIOC”) OR EXCLUDE(SUBJAREA, “CENG”) OR EX CLUDE(SUBJAREA, “MEDI”) OR EXCLUDE(SUBJAREA, “CHEM”) OR EXCLUDE(SUBJAREA, “HEAL”) OR EXCLUDE(SUBJAREA, “ECON”) OR EXCLUDE(SUBJAREA, “NEUR”) OR EXCLUDE(SUBJAREA, “PHAR”) OR EXCLUDE(SUBJAREA, “VETE”) OR EXCLUDE(SUBJAREA, “IMMU”))

A.3. Query to retrieve papers related to PSO

(TITLE(particle swarm optimi*) OR TITLE(pso)) AND PUBYEAR IS 2009 AND (EXCLUDE(SUBJAREA, “ENVI”) OR EXCLUDE(SUBJAREA, “BIOC”) OR EXCLUDE(SUBJAREA, “ENVI”) OR EXCLUDE(SUBJAREA, “BIOC”) OR EXCLUDE(SUBJAREA, “CENG”) OR EXCLUDE(SUBJAREA, “MEDI”) OR EXCLUDE(SUBJAREA, “CHEM”) OR EXCLUDE(SUBJAREA, “HEAL”) OR EXCLUDE(SUBJAREA, “ECON”) OR EXCLUDE(SUBJAREA, “NEUR”) OR EXCLUDE(SUBJAREA, “PHAR”) OR EXCLUDE(SUBJAREA, “VETE”) OR EXCLUDE(SUBJAREA, “IMMU”))

A.4. Query to retrieve papers related to ACO

(TITLE(ant colony optimi*) OR TITLE(aco)) AND PUBYEAR IS 2009 AND (EXCLUDE(SUBJAREA, “ENVI”) OR EXCLUDE(SUBJAREA, “BIOC”) OR EXCLUDE(SUBJAREA, “ENVI”) OR EXCLUDE(SUBJAREA, “BIOC”) OR EXCLUDE(SUBJAREA, “CENG”) OR EXCLUDE(SUBJAREA, “MEDI”) OR EXCLUDE(SUBJAREA, “CHEM”) OR EXCLUDE(SUBJAREA, “HEAL”) OR EXCLUDE(SUBJAREA, “ECON”) OR EXCLUDE(SUBJAREA, “NEUR”) OR EXCLUDE(SUBJAREA, “PHAR”) OR EXCLUDE(SUBJAREA, “VETE”) OR EXCLUDE(SUBJAREA, “IMMU”))

References

[1] W.J. Frawley, G. Piatetsky-Shapiro, C.J. Matheus, Knowledge discovery in databases: an overview, AI Mag. 13 (3) (1992) 57.
[2] A.H. Eschenfelder, Data Mining and Knowledge Discovery Handbook, vol. 14, Springer-Verlag New York Incorporated, 1980.
[3] P. Wright, Knowledge discovery in databases: tools and techniques, Crossroads 5 (2) (1998) 23–26.
[4] H.A. Edelstein, Introduction to Data Mining and Knowledge Discovery, Two Crows Corporation 2, USA, 1998.
[5] U.M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, R. Uthurusamy, Advances in Knowledge Discovery and Data Mining, The MIT Press, CA, USA, 1996.
[6] A.K. Jain, M.N. Murty, P.J. Flynn, Data clustering: a review, ACM Comput. Surv. (CSUR) 31 (3) (1999) 264–323.
[7] B. Zhang, M. Hsu, U. Dayal, K-harmonic means – a spatial clustering algorithm with boosting, in: TSDM, 2000, pp. 31–45.
[8] M.R. Anderberg, Cluster Analysis for Applications, Technical Report, DTIC Document, 1973.
[9] S. Guha, R. Rastogi, K. Shim, Rock: a robust clustering algorithm for categorical attributes, Inf. Syst. 25 (5) (2000) 345–366.
[10] T. Zhang, R. Ramakrishnan, M. Livny, Birch: an efficient data clustering method for very large databases, in: ACM SIGMOD Record, vol. 25, ACM, 1996, pp. 103–114.
[11] S. Guha, R. Rastogi, K. Shim, Cure: an efficient clustering algorithm for large databases, in: ACM SIGMOD Record, vol. 27, ACM, 1998, pp. 73–84.
[12] G. Karypis, E.-H. Han, V. Kumar, Chameleon: hierarchical clustering using dynamic modeling, Computer 32 (8) (1999) 68–75.
[13] J. Sander, M. Ester, H.-P. Kriegel, X. Xu, Density-based clustering in spatial databases: the algorithm gdbscan and its applications, Data Min. Knowl. Discov. 2 (2) (1998) 169–194, http://dx.doi.org/10.1023/A:1009745219419.
[14] C.-Y. Chen, F. Ye, Particle swarm optimization algorithm and its application to clustering analysis, in: 2004 IEEE International Conference on Networking, Sensing and Control, vol. 2, IEEE, 2004, pp. 789–794.
[15] S. Alam, G. Dobbie, P. Riddle, An evolutionary particle swarm optimization algorithm for data clustering, in: IEEE Swarm Intelligence Symposium (SIS 2008), 2008, pp. 1–6, http://dx.doi.org/10.1109/SIS.2008.4668294.
[16] S. Das, A. Chowdhury, A. Abraham, A bacterial evolutionary algorithm for automatic data clustering, in: Proceedings of the Eleventh conference on Congress on Evolutionary Computation, CEC'09, IEEE Press, Piscataway, NJ, USA, 2009, pp. 2403–2410.
[17] D. Van der Merwe, A. Engelbrecht, Data clustering using particle swarm optimization, in: The 2003 Congress on Evolutionary Computation, 2003. CEC'03., vol. 1, IEEE, 2003, pp. 215–220.
[18] A. Abraham, H. Guo, H. Liu, Swarm intelligence: foundations, perspectives and applications, in: Swarm Intelligent Systems, Springer, 2006, pp. 3–25.
[19] A.P. Engelbrecht, Fundamentals of Computational Swarm Intelligence, vol. 1, Wiley, Chichester, 2005.
[20] E. Bonabeau, C. Meyer, Swarm intelligence: a whole new way to think about business, Harv. Bus. Rev. 79 (5) (2001) 106–115.
[21] D. Martens, B. Baesens, T. Fawcett, Editorial survey: swarm intelligence for data mining, Mach. Learn. 82 (1) (2011) 1–42.
[22] A. Forestiero, C. Pizzuti, G. Spezzano, A single pass algorithm for clustering evolving data streams based on swarm intelligence, Data Min. Knowl. Discov. 26 (1) (2013) 1–26.
[23] J. Kennedy, R. Eberhart, Particle swarm optimization, in: Proceedings of the IEEE International Conference on Neural Networks, vol. 4, IEEE, 1995, pp. 1942–1948.
[24] J.F. Kennedy, J. Kennedy, R.C. Eberhart, Swarm Intelligence, Morgan Kaufmann, San Francisco, CA. USA, 2001.
[25] R. Poli, J. Kennedy, T. Blackwell, Particle swarm optimization, Swarm Intell. 1 (1) (2007) 33–57.
[26] S. Alam, G. Dobbie, P. Riddle, M.A. Naeem, A swarm intelligence based clustering approach for outlier detection, in: IEEE Congress on Evolutionary Computation (CEC), 2010, pp. 1 –7.
[27] Z. Wang, X. Sun, D. Zhang, Classification rule mining based on particle swarm optimization, in: Rough Sets and Knowledge Technology, Springer, 2006, pp. 436–441.
[28] T. Sousa, A. Silva, A. Neves, Particle swarm based data mining algorithms for classification tasks, Parallel Comput. 30 (5) (2004) 767–783.
[29] B. Alatas, E. Akin, Multi-objective rule mining using a chaotic particle swarm optimization algorithm, Knowl.-Based Syst. 22 (6) (2009) 455–460.
[30] R. Pears, Y.S. Koh, Weighted association rule mining using particle swarm optimization, in: New Frontiers in Applied Data Mining, Springer, 2012, pp. 327–338.
[31] X. Wang, J. Yang, X. Teng, W. Xia, R. Jensen, Feature selection based on rough sets and particle swarm optimization, Pattern Recognit. Lett. 28 (4) (2007) 459–471.
[32] B. Yue, W. Yao, A. Abraham, H. Liu, A new rough set reduct algorithm based on particle swarm optimization, in: Bio-inspired Modeling of Cognitive Tasks, Springer, 2007, pp. 397–406.
[33] X. Cui, T.E. Potok, P. Palathingal, Document clustering using particle swarm optimization, in: Proceedings of the 2005 IEEE Swarm Intelligence Symposium (SIS 2005), IEEE, 2005, pp. 185–191.
[34] S. Alam, G. Dobbie, P. Riddle, Y.S. Koh, Hierarchical PSO clustering based recommender system, in: 2012 IEEE Congress on Evolutionary Computation (CEC), IEEE, 2012, pp. 1–8.
[35] S. Alam, Intelligent web usage clustering based recommender system, in: Proceedings of the fifth ACM conference on Recommender systems, ACM, 2011, pp. 367–370.
[36] S. Alam, G. Dobbie, P. Riddle, Towards recommender system using particle swarm optimization based web usage clustering, in: New Frontiers in Applied Data Mining, Springer, 2012, pp. 316–326.
[37] S. Alam, G. Dobbie, Y.S. Koh, P. Riddle, Clustering heterogeneous web usage data using hierarchical particle swarm optimization, in: IEEE Symposium on Swarm Intelligence (SIS), IEEE, 2013, pp. 147–154.
[38] S. Alam, G. Dobbie, P. Riddle, Y.S. Koh, Analysis of web usage data for clustering based recommender system, in: Trends in Practical Applications of Agents and Multiagent Systems, Springer, 2013, pp. 171–179.
[39] H.-L. Chen, B. Yang, G. Wang, J. Liu, X. Xu, S.-J. Wang, D.-Y. Liu, A novel bankruptcy prediction model based on an adaptive fuzzy k nearest neighbor method, Knowl.-Based Syst. 24 (8) (2011) 1348–1359.
[40] X. Wang, S. Wang, J.-J. Ma, An improved co-evolutionary particle swarm optimization for wireless sensor networks with dynamic deployment, Sensors 7 (3) (2007) 354–370.
[41] W. Jatmiko, K. Sekiyama, T. Fukuda, A pso-based mobile sensor network for odor source localization in dynamic environment: theory, simulation and measurement, in: IEEE Congress on Evolutionary Computation (CEC 2006), IEEE, 2006, pp. 1036–1043.
[42] B. You, G. Chen, W. Guo, Topology control in wireless sensor networks based on discrete particle swarm optimization, in: IEEE International Conference on Intelligent Computing and Intelligent Systems (ICIS 2009), vol. 1, IEEE, 2009, pp. 269–273.
[43] X. Wang, J.-J. Ma, S. Wang, D.-W. Bi, Distributed particle swarm optimization and simulated annealing for energy-efficient coverage in wireless sensor networks, Sensors 7 (5) (2007) 628–648.
[44] J. Hou, X. Fan, W. Wang, J. Jie, Y. Wang, Clustering strategy of wireless sensor networks based on improved discrete particle swarm optimization, in: 2010 Sixth International Conference on Natural Computation (ICNC), vol. 7, IEEE, 2010, pp. 3866–3870.
[45] D. Karaboga, S. Okdem, C. Ozturk, Cluster based wireless sensor network routing using artificial bee colony algorithm, Wirel. Netw. 18 (7) (2012) 847–860.
[46] A. Chatterjee, K. Pulasinghe, K. Watanabe, K. Izumi, A particle-swarm-optimized fuzzy-neural network for voice-controlled robot systems, IEEE Trans. Ind. Electron. 52 (6) (2005) 1478–1489.
[47] J. Pugh, A. Martinoli, Inspiring and modeling multi-robot search with particle swarm optimization, in: IEEE Swarm Intelligence Symposium (SIS 2007), IEEE, 2007, pp. 332–339.
[48] J. Pugh, A. Martinoli, Distributed scalable multi-robot learning using particle swarm optimization, Swarm Intell. 3 (3) (2009) 203–222.
[49] W. Jatmiko, K. Sekiyama, T. Fukuda, A PSO-based mobile robot for odor source localization in dynamic advection-diffusion with obstacles environment: theory, simulation and measurement, IEEE Comput. Intell. Mag. 2 (2) (2007) 37–51.
[50] M.G. Omran, A. Salman, A.P. Engelbrecht, Dynamic clustering using particle swarm optimization with application in image segmentation, Pattern Anal. Appl. 8 (4) (2006) 332–344.
[51] A.A. Esmin, R.A. Coelho, S. Matwin, A review on particle swarm optimization algorithm and its variants to clustering high-dimensional data, Artif. Intell. Rev. (2013) 1–23.
[52] X. Xiao, E.R. Dow, R. Eberhart, Z.B. Miled, R.J. Oppelt, Gene clustering using self-organizing maps and particle swarm optimization, in: Proceedings of the 17th International Symposium on Parallel and Distributed Processing (IPDPS '03), IEEE Computer Society, Washington, DC, USA, 2003, p. 154.2.
[53] Y. Lin, N. Tong, M. Shi, K. Fan, D. Yuan, L. Qu, Q. Fu, K-means optimization clustering algorithm based on particle swarm optimization and multiclass merging, in: Advances in Computer Science and Information Engineering, Springer, 2012, pp. 569–578.
[54] S.C.M. Cohen, L.N. De Castro, Data clustering with particle swarms, in: 2006 IEEE International Conference on Evolutionary Computation, 2006, 1792–1798.
[55] J. Chen, H. Zhang, Research on application of clustering algorithm based on pso for the web usage pattern, in: International Conference on Wireless Communications, Networking and Mobile Computing (WiCom 2007), IEEE, 2007, pp. 3705–3708.
[56] T. Cura, A particle swarm optimization approach to clustering, Expert Syst. Appl. 39 (1) (2012) 1582–1588.
[57] F. Yang, T. Sun, C. Zhang, An efficient hybrid data clustering method based on k-harmonic means and particle swarm optimization, Expert Syst. Appl. 36 (6) (2009) 9847–9852.
[58] Y.-T. Kao, E. Zahara, I.-W. Kao, A hybridized approach to data clustering, Expert Syst. Appl. 34 (2008) 1754–1762.
[59] A. Asuncion, D. Newman, Uci Machine Learning Repository, University of California, School of Information and Computer Science, Irvine, CA, ⟨http://www.ics.uci.edu/~mlearn/mlrepository.html⟩.
[60] S. Das, A. Abraham, A. Konar, Automatic kernel clustering with a multi-elitist particle swarm optimization algorithm, Pattern Recognit. Lett. 29 (5) (2008) 688–699.
[61] S. Das, A. Abraham, A. Konar, Automatic clustering using an improved differential evolution algorithm, IEEE Trans. Syst., Man Cybern., Part A: Syst. Hum. 38 (1) (2008) 218–237.
[62] M.-Y. Chen, A hybrid anfis model for business failure prediction utilizing particle swarm optimization and subtractive clustering, Inf. Sci. 220 (2013) 180–195, http://dx.doi.org/10.1016/j.ins.2011.09.013.

[63] F.S. Shie, M.-Y. Chen, Y.-S. Liu, Prediction of corporate financial distress: an application of the america banking industry, Neural Comput. Appl. 21 (7) (2012) 1687–1696.
[64] M.-Y. Chen, Bankruptcy prediction in firms with statistical and intelligent techniques and a comparison of evolutionary computation approaches, Comput. Math. Appl. 62 (12) (2011) 4514–4524.
[65] J. Han, M. Kamber, Data Mining: Concepts and Techniques, The Morgan Kaufmann Series in Data Management Systems, Morgan Kaufmann, San Francisco, CA. USA, 2000.
[66] S. Alam, G. Dobbie, P. Riddle, Particle swarm optimization based clustering of web usage data, in: Proceedings of the 2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, vol. 3, 2008, pp. 451 –454.
[67] S. Alam, G. Dobbie, P. Riddle, Exploiting swarm behaviour of simple agents for clustering web users' session data, in: L. Cao (Ed.), Data Min. Multi-agent Integr., Springer, US, 2009, pp. 61–75.
[68] S. Alam, G. Dobbie, P. Riddle, M.A. Naeem, Particle swarm optimization based hierarchical agglomerative clustering, in: 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT), vol. 2, IEEE, 2010, pp. 64–68.
[69] S. Alam, G. Dobbie, P. Riddle, M.A. Naeem, Particle swarm optimization based hierarchical agglomerative clustering, in: IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, vol. 2, 2010, pp. 64–68.