Clustering Nominal and Numerical Data: A New Distance Concept for a Hybrid Genetic Algorithm

Laetitia Jourdan, Clarisse Dhaenens, El-Ghazali Talbi

Clustering nominal and numerical data: a new distance concept for an hybrid genetic algorithm

L. Vermeulen-Jourdan and C. Dhaenens and E-G. Talbi

LIFL-Université de Lille1, Bât M3-Cité Scientifique,
59655 Villeneuve d'Ascq Cedex FRANCE jourdan@lifl.fr,
WWW home page: http://www.lifl.fr/jourdan

Abstract. As intrinsic structures, like the number of clusters, is, for real data, a major issue of the clustering problem, we propose, in this paper, CHyGA (Clustering Hybrid Genetic Algorithm) an hybrid genetic algorithm for clustering. CHyGA treats the clustering problem as an optimization problem and searches for an optimal number of clusters characterized by an optimal distribution of instances into the clusters. CHyGA introduces a new representation of solutions and uses dedicated operators, such as one iteration of K-means as a mutation operator. In order to deal with nominal data, we propose a new definition of the cluster center concept and demonstrate its properties. Experimental results on classical benchmarks are given.

1 Introduction

Clustering is used to identify classes of objects sharing common characteristics and its methods can be applied to many human activities and particularly to the automatic decision making problem. The data clustering or unsupervised classification, can isolate similarities and differences in a database and make groups of similar data which are called classes, groups or clusters. It can reveal some intrinsic structures (e.g. the number of clusters). There exists many clustering methods like graph-based ones, model-based ones, genetic algorithm-oriented ones, distance based approaches or their hybridizations. Most of these methods require as an input the number of clusters to determine. This requirement is a major problem for real-life problems, where the number of clusters is not known in advance.
Genetic algorithms have been successfully applied to partitioning problems [12] and in particular to clustering problems. In [14], the authors couple the fuzzy K-means algorithm with a GA, where one iteration of the Fuzzy K-means is used to compute the fitness of the classification. However, all those algorithms require as input the number of clusters.
In this paper, we present an hybrid genetic algorithm using a specific encoding with dedicated operators. The hybridization consists in using a move of K-means as one of these operators. We recall that K-means is devoted to deal with numerical data, and conversely we aim at dealing with nominal data. Therefore,

in order to realize the proposed hybridization, we need to redefine the step of the K-means algorithm. In addition, we propose a new definition for the cluster center concept. An associated distance is also presented.
Section 2 presents the K-means algorithm, the definition of the center concept for nominal data and the proposed associated distance. Section 3 presents CHyGA, its encoding and the dedicated operators. Section 4 provides experimental results for several classical numerical and nominal datasets.

2 Clustering and center concept

Clustering aims to group similar objects into clusters which can be described by theirs centers. Each object is described by a set of attributes. Each attribute A_i has a domain definition Ω and takes a value in this domain.
The K-means algorithm is one of the most famous algorithm for clustering [6]. We firstly present the classical algorithm dedicated to numerical data, then we introduce a definition of the center concept more adapted to nominal data. We also introduce the associated distance.

2.1 The K-means algorithm

The K-means algorithm is an iterative procedure where an iteration is given below:

Input: Partition P of k clusters: $C_1, ..., C_k$;
Compute $Center(C_1), ..., Center(C_k)$;
Remove all objects from all cluster;
for each object O_i **do**
  Let C_j, $(j \in [1, k])$ be the cluster whose center is the closest to O_i;
  Assign O_i to C_j;
end for
Compute the resulting new partition $P = C_1, ..., C_l (l \leq k)$;
Remove all empty clusters.

The major drawback of K-means algorithm is that it often terminates on a local optimum and works only on numerical values because it minimizes a cost function calculating the means of clusters. Moreover, it needs to compute centers. The center of a cluster is easy to define on numerical values because the mean makes sense, but for nominal data it is not so simple.

2.2 A new definition for the cluster center concept

In some works, authors have proposed some definitions for the center of categorical or nominal data. For example, Huang proposes to compute the center of a cluster by using the mode of a set [15].

Definition 1. *Let X be a set of n nominal objects described by attributes $A_1, ..., A_m$, Ω the set of all possible combinations of values of the attributes.*
A mode of X is a vector $Q \in \Omega$, $Q = [q_1 q_2 ... q_m]$ that minimizes $D(Q, X) =$

$\sum_{j=1}^{n} d(X_j, Q)$, *where* $d(X, Y)$ *is a simple matching [18]. Q is not necessarily an element of* X.

This definition has two drawbacks: first, Q is not always unique; second, if we consider an attribute having the following values $Y, ?, Y, N, Y, N, ?$ the mode will choose the value Y which has a frequency of 3, but is it really significant ?
We decide not to use such a center election. We propose here to consider a center election based on a majority vote (frequency $\leq 1/2$). When there is no satisfiable candidate none is chosen and we use a partially defined center. We will use the notation * to denote values of attributes for which there is no satisfiable candidate.

Definition 2. *) Let* X *be a set of* n *nominal objects described by attributes* $A_1, ..., A_m$.
The center of X *is the vector* $Q = [q_1 q_2 ... q_m]$ *with* $q_i \in \Omega \cup \{*\}$ *which minimizes* $D(Q, X) = \sum_{j=1}^{n} d_v(X_j, Q)$ *where a possible* d_v *is a distance measure defined in paragraph 2.3.* Q *is not necessarily an element of* X *and is unique.*

The defined center will be called a partial center and represents an hyperplane whereas for commonly used definition, a center is reduced to a single point of the space $\mathbb{R}^m$ (like for mode). The dimension of the hyperplane in $\mathbb{R}^m$ is the *number of attributes - number of determined attributes of the center.*

2.3 A proposed associated distance measure

In order to realize a clustering with K-means, we have to define a distance between objects and the defined center. The proposed distance is based on the Hamming distance, but has to be adapted in order to deal with the partial center concept.
Let $d_v(O, C)$ be the distance between an object O and a center C. We define $d_v = \sum_{i=1}^{m} d_\sigma(O_i, C_i)$ as the following:

$$\forall x \text{ and } y, \quad d_\sigma(x, y) = \begin{cases} 0 \text{ if x=y} \\ 1/2 \text{ if x or y equals to * (x}\neq\text{y)} \\ 1 \text{ if x} \neq \text{ y} \end{cases} \tag{1}$$

2.4 Properties of the center

In this section, we must verify that the center Q defined in Section 2.2 respects the fundamental properties of a center regarding the distance described in Section 2.3.

Theorem 1. *The function* $D(X, Q) = \sum_j d_v(X_j, Q)$ *is minimized by the center* Q *defined as in definition 2.*

Proof. Let n be the number of objects in the cluster X, let m be the number of attributes of an object. Let d_v be the previous defined distance ($d_v = \sum_i d_\sigma(O_i, C_i)$). Let Q be a center of the cluster: $\sum_{j=1}^{n} d_v(X_j, Q) = \sum_{j=1}^{n} \sum_{i=1}^{m} d_\sigma(X_{j_i}, q_i)$.

As $\forall i \in [1,m]\ d_\sigma(X_{j_i}, q_i) \geq 0$, if each element of the sum is minimal $D(X,Q)$ is minimal.
Let n_{q_i} be the frequency of the value q_i of the attribute i chosen to be the representative center. Recall that the distance between * and an attribute is always 1/2. As n_{q_i} is the appearance frequency of q_i : $n - n_{q_i} \geq 0$.
For each $d_\sigma(X_{i_j}, q_j)$ two cases are possible :
1. Let the i^{th} attribute of the center be determined. By definition of the center, the attribute has a frequency greater than the half of the voices ($n_{q_i} \geq \frac{n}{2}$). As $d_\sigma(X_{j_i}, q_i) = n - n_{q_i}$ then $n - n_{q_i} \leq \frac{n}{2}$. Thus $d_\sigma(X_{j_i}, q_i)$ obtained by the majority vote is minimal.
2. Let the i^{th} attribute of the center be undetermined. There exists no value representing the majority for this attribute. So $\forall q_i,\ n - n_{q_i} > n/2$ then $d(X_{j_i}, q_i)$ obtained with the * is minimal.
Hence, the proposed definition of center minimizes the presented intra-cluster distance.

3 CHyGA

The clustering problem is NP-hard [19], and may be treated as an optimization problem. The literature shows that GAs are well adapted to explore the very large search space of this problem [16]. Figure 1, shows the different stages of CHyGA, the genetic algorithm we propose. We present here the main characteristics of this algorithm.

3.1 Encoding

The encoding is used to describe potential solutions and must be carefully chosen. For clustering or grouping problems, there exist different representations [3, 12, 17, 20]. For example the group number representation which indicates for each object the group it belongs to [17]. All those representations have advantages and drawbacks.
In CHyGA, we choose to use a double hybrid representation (see Figure 2) which merges the group number representation and a description of clusters. This representation allows to find very quickly information on object's affectations and cluster's composition. In the group number structure, for each object is indicated the cluster it belongs to. In the complex structure, for each cluster a double linked list gives its composition. Moreover, an array allows to point directly on each object (to move it from one cluster to another in constant time).

3.2 Fitness Function

The fitness function is used to evaluate the suitability of a solution to the problem. This function is used for the selection phase.
In the literature, several cluster validity indicators have been used [1, 4, 9, 10].

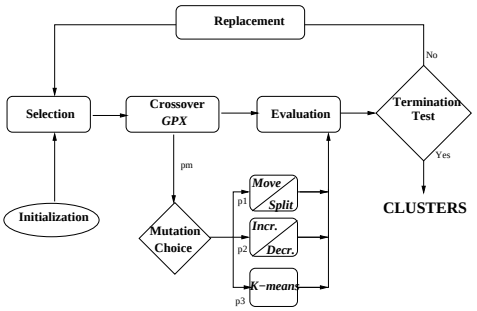

Fig. 1. Stages of CHyGA.

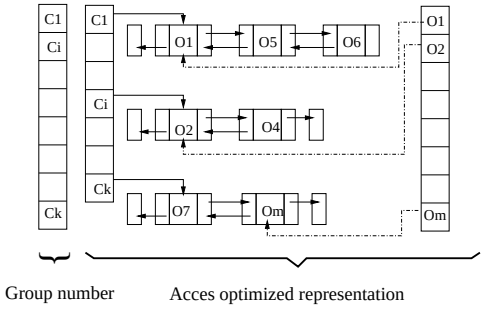

Fig. 2. The encoding.

To measure the quality of the proposed clustering, we use as fitness function the criterion of Calinski and Harabasz (CH) that measures the good distribution of the objects [7].
Given k clusters, n features, let χ_j be the j^{st} cluster with j=1...k.
Let m be the mean vector $m = \begin{pmatrix} m_1 & m_2 & \cdots & m_n \end{pmatrix}$ and m_j the vector of means for the j^{st} cluster $m_j = \begin{pmatrix} m_{j_1} & m_{j_2} & \cdots & m_{j_n} \end{pmatrix}$. X_i is an element of the cluster χ_j and $(X_i - m_j)^t$ is the transposed vector. Then the Calinski and Harabasz criterion is:

$$CH = \frac{(n-k) \times trace(\sum_{j=1}^{k} n_j (m_j - m)(m_j - m)^t)}{(k-1) \times trace(\sum_{j=1}^{k} \sum_{X_{i \in \chi_j}} (X_i - m_j)(X_i - m_j)^t)} \quad (2)$$

3.3 Operators

Crossover
In CHyGA, we use the GPX crossover (Greedy Partition Crossover) [13] which is a partition crossover. Our representation is well adapted to this operator, as elements of a same cluster are grouped.

Mutations
Our GA uses 5 different mutation operators. For clustering problems mutation operators must able to introduce new groups and to remove existing groups. In [11], Falkenauer shows that the mutation of a single affectation (object/group) is not enough because it doesn't increase significantly the fitness and the new candidate solution will be quickly lost from the population. Our GA uses two standard mutations for grouping problems : split and move [8], two specifically designed mutation operators adapted to clustering and a single iteration of K-means.
The split operator selects some objects in a particular cluster and moves these objects to a new cluster. We also define two other operators that are more specific for the clustering. The first operator, called "increase", is able to increase the number of clusters. This mutation operator chooses the cluster that has the

maximum internal variance and splits it into two clusters. The objects are reallocated between the two new clusters. Our second specifically designed mutation operator, "decrease", aims at merging two clusters into one. This operator detects the two closest centers and groups the corresponding clusters into one.
We also use one iteration of the K-means [16] algorithm as a mutation operator to make a local search and improve the quality of a solution. It takes as input the distribution given by the chromosome, calculates centers of clusters and reassigns objects to clusters whose center are the closest.
At the end of the genetic search, we try to improve the quality of all the solutions found with a local search. The local search consists in applying an iteration of the K-means algorithm.

4 Experimental results

We have experimented CHyGA on some common datasets available at the UCI repository [5], or in the articles of Bandyopadhyay [2] and Ruspini [21]. The specificities of each dataset used to evaluate the performances of the proposed algorithm are summed up in Table 1. It indicates the name of the dataset its number of instances (n), the number of attributes (m), the original number of classes (k), the percentage of missing data (l). Therefore a dataset is summed up by **Name (n, m, k, l)**. Table 1 also indicates the CH value obtained with the initial repartition of the instances in their classes as all the datasets are basically intended to the classification data mining task.

Table 1 also indicates for each dataset some descriptive statistics of the results obtained by CHyGA. We indicate for the Calinski and Harabasz criteria (CH) the mean of the best results obtained over ten runs, the maximum, the standard deviation (σ) and the median. Information is given about the number of discovered clusters.
We can observe that CHyGA is able to discover a number of clusters equal to the original number of classes. Only on one dataset, Iris, it sometimes obtains, for the best solution of an execution, a number of clusters of four instead of three. However, in this case, the best solution ever obtained for CH has been found for a number of clusters equal to three (559.258(3)).
Concerning the robustness of the method, we can observe that the method is very robust on some of the numerical datasets where the standard deviation of the best results obtained by CHyGA over the runs is equal to zero.
In addition, we compare our method to some classical methods: K-means, Kmediods, Single link and Complete link. To make a fair comparison, we give as input, for all the classical methods, the number of clusters found by CHyGA. For non deterministic methods (K-means, Kmediods), we ran ten times the algorithms. Table 2 indicates for each method, the best solution found, the standard deviation (σ) of the solution given by the algorithms for CH. As often in clustering, interesting objectives are to minimize the intra cluster distance and to maximize the inter cluster distance, Table 2 also indicates their value.
For numerical datasets, we can observe that our method finds for all the datasets

Table 1. Descriptive statistics of the results obtained by CHyGA on different numerical and nominal datasets.

Numerical Datasets		
AD_5_2		
(250, 2, 5, 0)		
CH = 387.75		
	Criteria	# cluster
Mean	386.20	5
Max.	387.75	5
σ	0.18	0
Med.	386.44	5
AD_4_3		
(400, 3, 4, 0)		
CH = 3207.41		
	Criteria	# cluster
Mean	3207.41	4
Max.	3207.41	4
σ	0	0
Med.	3207.41	4
Ruspini		
(75, 2, 4, 0)		
CH = 425.32		
	Criteria	# cluster
Mean	425.32	4
Max.	425.32	4
σ	0	0
Med.	425.32	4

Numerical Datasets		
Iris		
(150, 4, 3, 0)		
CH = 486.32		
	Criteria	# cluster
Mean	526.42	3
Max.	559.25 (3)	4
σ	11.78	0.82
Med.	527.83	3
Cancer		
(683, 9, 2, 0)		
CH = 912.20		
	Criteria	# cluster
Mean	1025.81	2
Max.	1025.81	2
σ	0	0
Med.	1025.81	2
Diabetes		
(768, 8, 3, 0)		
CH = 24.29		
	Criteria	# cluster
Mean	1136.18	3
Max.	1142.49	3
σ	6.98	0
Med.	1139.2	3

Nominal Datasets		
Lung		
(32, 57, 2, 4.13)		
CH =9.609		
	Criteria	# cluster
Mean	23.38	2
Max.	23.68	2
σ	0.14	0
Med.	23.34	2
Vote		
(435, 16, 2, 0.22)		
CH =455.865		
	Criteria	# cluster
Mean	508.79	2
Max.	522.34	2
σ	9.76	0
Med.	507.52	2
Breast Cancer		
(286,9, 2, 0.34)		
CH =212.251		
	Criteria	# cluster
Mean	243.63	2
Max.	247.49	2
σ	2.99	0
Med.	243.91	2

Table 2. Best results obtained with different method on classical dataset from UCI.

Data	Criteria	CHyGA	K-means	Kmed.	Single Link	Compl. Link
AD_5_2	Intra	**5.06**	5.17	6.08	93.90	5.62
	Inter	218.85	218.80	218.02	**218.89**	218.88
	Calinski	**388.12**	386.58	163.92	11.74	6.25
	σ	0	28.84	24.68	-	-
AD_4_3	Intra	**6.54**	6.63	8.75	6.55	19.67
	Inter	410.88	410.88	411.10	**410.83**	410.99
	Calinski	**3207.41**	3206.68	1720.86	3190.08	0.57
	σ	0	385.02	313.31	-	-
Ruspini	Intra	**38.75**	**38.75**	54.18	40.57	162.15
	Inter	19782.42	19782.42	**19787.28**	19783.07	19782.52
	Calinski	**425.32**	**425.32**	263.92	422.40	2.63
	σ	0	161.56	74.2	-	-
Iris	Intra	**2.36**	**2.36**	2.86	2.69	2.50
	Inter	46.6264	46.6264	**46.93**	46.62	46.62
	Calinski	**559.26**	**559.26**	322.92	343.85	42.35
	σ	11.78	1.98	108.55	-	-
Cancer	Intra	14.14	17.08	13.42	28.16	20.03
	Inter	235.31	235.31	237.68	235.17	235.34
	Calinski	**1026.06**	1008.74	1201.66	298.09	12.06
	σ	0	0	385.02	-	-
Diabetes	Intra	**246.56**	247.44	276.17	454.01	264.00
	Inter	105763.16	105763.25	105736.98	**105763.90**	105760.35
	Calinski	**1135.07**	133.97	276.17	20.19	0.95
	σ	6.98	17.56	47.51	-	-

the best solution of CH also found sometimes by the K-means algorithm. For intra cluster distance which measures the compactness of the cluster, we can observe that CHyGA finds for each dataset the smaller value which is for two datasets, Ruspini and Iris, also obtained by K-means. For inter cluster criteria which measures the separation of the obtained clusters, we remark that even if CHyGA doesn't find the best value of the criteria, the value obtained is really closed to the best found. Indeed, when we look at the error made by CHyGA in comparison with the best found, this error is less than 0.02% for Ruspini dataset, 0.66% for Iris dataset and 0.0006% for Diabetes dataset.
For the three stochastic methods, we indicate the standard deviation of the solutions obtained over the ten executions of the algorithms (σ). We can observe that CHyGA has the smaller standard deviation except for the Iris dataset.
Concerning the computational time, it is obvious that our method is longer than a simple K-means algorithm because we use a single iteration of it as a local search in our hybridization. Methods such as Single or Complete link are faster than our method on small datasets but longer on larger datasets (Diabetes for example).
Hence, those results show that CHyGA has best or at least comparable performances than the best clustering methods. Nevertheless, it is important to recall that CHyGA does not require as input the number of clusters to look for and can deal with nominal or numerical data.

5 Conclusion

This paper has proposed a specific hybrid Genetic Algorithm for the clustering problem: CHyGA. This algorithm is very interesting because it makes clustering with no indication on the number of clusters by using the Calinski and Harabasz criteria. We have used a specific encoding and an one iteration K-means to hybridize the algorithm. To deal with nominal data we proposed a new center conception and an associated distance. Results obtained are very encouraging. We obtained on both numerical and nominal data the same number of clusters than the number of given classes.
This algorithm works for an unknown number of clusters and the final population is compound of different solutions with different number of clusters. Here, only the best solution has been used for the evaluation, but it could be interesting to look at several good solutions obtained thanks to CHyGA.
A multi-criteria approach would be interesting to look for the best compromise between different quality criteria that may be used.

Appendix

We demonstrate the properties of d_v and show that it is a distance.

Proof. Positivity
By construction, d_σ is always positive or equal to zero then
$d_v(O, C) = \sum_i d_\sigma(O_i, C_i) \geq 0$.

Proof. Symmetry
$d_v(O,C) = \sum_i d_\sigma(O_i, C_i)$ then if d_σ is symmetrical, d_v is symmetrical.
Let us consider the different cases:
If $d_\sigma(O_i, C_i) = 0 \Longrightarrow O_i = C_i \Longleftrightarrow C_i = O_i$ then $d_\sigma(C_i, O_i) = 0$.
If $d_\sigma(O_i, C_i) = \frac{1}{2} \Longrightarrow C_i = *$ (O_i always defined) then $d_\sigma(C_i, O_i) = \frac{1}{2}$.
If $d_\sigma(O_i, C_i) = 1 \Longrightarrow O_i \neq C_i$ then $C_i \neq O_i$ then $d_\sigma(C_i, O_i) = 1$.
d_σ is symmetrical then d_v is symmetrical.

Proof. Triangular inequality
Let O_1, O_2, O_3 be three objects that can be centers.
We want to show that $d_v(O_1, O_2) \leq d_v(O_1, O_3) + d_v(O_3, O_2)$.
We must show that $\sum_i d_\sigma(O_{1_i}, O_{2_i}) \leq \sum_i d_\sigma(O_{1_i}, O_{3_i}) + \sum_i d_\sigma(O_{3_i}, O_{2_i})$.
Let $A = \sum_i d_\sigma(O_{1_i}, O_{2_i})$ et $A_i = d_\sigma(O_{1_i}, O_{2_i})$.
Let $B = \sum_i d_\sigma(O_{1_i}, O_{3_i}) + \sum_i d_\sigma(O_{3_i}, O_{2_i})$ and $B_i = d_\sigma(O_{1_i}, O_{3_i}) + d_\sigma(O_{3_i}, O_{2_i})$.
We compute attribute by attribute. For attribute i, three cases may appear:
1. If $O_{1_i} = O_{2_i}$ then $A_i = 0$:
   - If $O_{1_i} = O_{3_i}$ then $O_{2_i} = O_{3_i}$ then $B_i = 0$
   - If $O_{1_i} \neq O_{3_i}$ then $B_i += \frac{1}{2}$ or 1 and then $O_{3_i} \neq O_{2_i}$ then $B_i += \frac{1}{2}$ or 1
$\Longrightarrow B_i = 0, 1$ or 2. In this case $A_i \leq B_i$.
2. If $O_{1_i} \neq O_{2_i}$ and $\neq *$ then $A_i = 1$:
   - If $O_{1_i} = O_{3_i}$ then $B_i += 0$ but $O_{3_i} \neq O_{2_i}$ then $B_i += 1$.
   - If $O_{1_i} \neq O_{3_i}$ then if $O_{3_i} = *$ then $B_i = \frac{1}{2} + \frac{1}{2}$ else $B_i = 1 + (0 \; or \; 1)$
$\Longrightarrow B_i = 1$ or 2. In this case, $A_i \leq B_i$.
3. If $O_{1_i} \neq O_{2_i}$ and one of them equals to *. As the distance is symmetrical, let $O_{1_i} = *$ then $A_i = \frac{1}{2}$:
   - If $O_{1_i} = O_{3_i}$ then $B_i += 0$ but $O_{3_i} \neq O_{2_i}$ then $B_i += \frac{1}{2}$.
   - If $O_{1_i} \neq O_{3_i}$ then $B_i = \frac{1}{2} + (0$ or $1)$.
   - If $O_{3_i} = *$ then $B_i = \frac{1}{2}$.
$\Longrightarrow B_i \geq \frac{1}{2}$. In this case $A_i \leq B_i$.

To conclude, for all objects O_1, O_2, O_3:
$\forall i \; A_i \leq B_i \Longrightarrow \sum_i A_i \leq \sum_i B_i \Longrightarrow A \leq B$
$\sum_i d_\sigma(O_{1_i}, O_{2_i}) \leq \sum_i d_\sigma(O_{1_i}, O_{3_i}) + \sum_i d_\sigma(O_{3_i}, O_{2_i})$
$d_v(O_1, O_2) \leq d_v(O_1, O_3) + d_v(O_3, O_2)$

Thus d_v is positive, symmetrical and transitive so, we can conclude that d_v is a distance.

References

1. S. Bandyopadhyay and U. Maulik. Nonparametric genetic clustering: Comparison of validity indices. *IEEE Trans. Syst., Man, Cybern-Part C: Applications and Reviews*, 2001.
2. S. Bandyopadhyay and U. Maulik. Genetic clustering for automatic evolution of clusters and application to image classification. *Pattern Recognition*, 35:1197–1208, 2002.

3. J.C. Bezdeck, S. Boggavaparu, L.O. Hall, and A. Bensaid. Genetic algorithm guided clustering. In *Proc. of the First IEEE Conference on Evolutionary Computation*, pages 34–38, 1994.
4. J.C. Bezdek. Some new indexes of cluster validity. *IEEE Trans. on Systems, Man, and Cybernetics*, 28(3):301–315, 1998.
5. C.L. Blake and C.J. Merz. Uci repository of machine learning databases. http://www.ics.uci.edu/~mlearn/MLRepository.html, 1998. University of California, Irvine, Dept. of Information and Computer Sciences.
6. L. Bottou and Y. Bengio. Convergence properties of the K-means algorithms. In G. Tesauro, D. Touretzky, and T. Leen, editors, *Advances in Neural Information Processing Systems*, volume 7, pages 585–592. The MIT Press, 1995.
7. T. Calinski and J. Harabasz. A dendrite method for cluster analysis. *Communications in statistics*, 3(1):1–27, 1974.
8. R.M. Cole. Clustering with genetic algorithms. Master's thesis, University of Western Australia, Australia, 1998. http://citeseer.nj.nec.com/cole98clustering.html.
9. D.L. Davies and D.W. Bouldin. A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1, 1979.
10. J.C. Dunn. A fuzzy relative of the isodata process and its use in detecting compact, well-seperated clusters. *Journal of Cybernetics*, 3(3):32–57, 1973.
11. E. Falkenauer. A new representation and operators for genetic algorithms applied to grouping problems. *Evolutionary Computation*, 2(2):123–144, 1994.
12. E. Falkenauer. *Genetic Algorithms and Grouping Problems.* John Wiley, 1998.
13. P. Galinier and JK. Hao. Hybrid evolutionary algorithms for graph coloring. *Journal of Combinatorial Optimization*, 3:379–397, 1999.
14. L. O. Hall, I. B. Oezyurt, and J. C. Bezdek. Clustering with a genetically optimized approach. *IEEE Transactions on evolutionary Computation*, 3(2):103–112, July 1999.
15. Z. Huang. Extensions to the k-means algorithm for clustering large data sets with categorical values. *Data Mining and Knowledge Discovery*, 1998.
16. A.K. Jain, M.N. Murty, and P.J. Flynn. Data clustering: A review. *ACM Computing Surveys*, 31(3):264–323, September 1999.
17. D.R. Jones and M.A. Beltramo. Solving partitioning problems with genetic algorithms. In Eds. R. Belew, L. B. Booker, editor, *Proc. of the Fourth International Conference on Genetic Algorithms*, pages 442–449. Morgan Kaufman Publishers, 1991.
18. L. Kaufman and P. Rousseuw. *Finding Groups in Data- An Introduction to Cluster Analysis.* Wiley Series in Probability and Mathematical Sciences, 1990.
19. G. L. Liu. *Introduction to combinatorial Mathematics.* McGraw Hill, 1968.
20. Z. Michalewizc. *Genetic Algorithms + Data Structures = Evolution Programs.* Springer-Verlag, third, revised and extend edition, 1996.
21. E.H. Ruspini. Numerical methods for fuzzy clustering. *Inform. Sci.*, 2:319–350, 1970.