Rough Clustering Using an Evolutionary Algorithm

Kevin E. Voges

University of Canterbury, Christchurch, New Zealand

kevin.voges@canterbury.ac.nz

Nigel K. Ll. Pope

Griffith University, Brisbane, Australia

n.pope@griffith.edu.au

Abstract

Cluster analysis is a fundamental technique in traditional data analysis and many clustering methods have been identified, including the commonly used k-means approach, which requires the number of clusters to be specified in advance and is dependent on initial starting points. We present an evolutionary-based rough clustering algorithm, which is designed to overcome these limitations. Rough clusters are defined in a similar manner to Pawlak's rough set concept, with a lower and upper approximation, allowing multiple cluster membership for objects in the data set. The paper describes the template, the data structure used to describe rough clusters. It also provides an overview of the evolutionary algorithm used to develop viable cluster solutions, consisting of an optimal number of templates providing descriptions of the clusters. This algorithm was tested on a small data set and a large data set.

1. Cluster analysis

Cluster analysis is a fundamental technique in both traditional data analysis and in data mining. The technique is defined as grouping 'individuals or objects into clusters so that objects in the same cluster are more similar to one another than they are to objects in other clusters' [8:470]. Many clustering methods have been identified, including partitioning, hierarchical, nonhierarchical, overlapping, and mixture models. One of the most commonly used nonhierarchical methods is the k-means approach [9], [14].

In the k-means approach, objects are randomly selected as initial seeds or centroids, and the remaining objects are assigned to the closest centroid on the basis of the distance between them. The aim is to obtain maximal homogeneity within subgroups or clusters, and maximal heterogeneity between clusters. The data set is partitioned into clusters and an error

term e is calculated, usually based on the Euclidean distance between each object and the cluster centroids. The usual approach is to search for a partition with small e by moving cases from one partition to another [8]. The search through the problem space to find the lowest value of e is considered computationally expensive and local optimization has traditionally been used. In addition, the number of clusters in each partition is decided prior to the analysis, a major limitation of the technique. A k-Modes approach has been developed as an extension of the k-Means algorithm, and has been applied to categorical data clustering by replacing means with modes [4]. However even with this extension, the number of clusters needs to be set in advance.

In the last few decades, as data sets have grown in size and complexity, and the field of data mining has matured, many new techniques based on developments in computational intelligence have started to be more widely used as clustering algorithms. For example, the theory of fuzzy sets developed by Zadeh [31] introduced the concept of partial set membership as a way of handling imprecision in mathematical modeling. This was subsequently applied to cluster analysis [2], [7]

One technique from the field of computational intelligence receiving considerable attention is the theory of rough sets [19], [21], [22]. In most previous applications of rough sets theory, the technique was used for classification problems, where prior group membership is known, and results are usually expressed in terms of rules for group membership [12], [20], [26].

This paper describes a rough clustering technique, based on a simple extension of rough sets theory, applicable where prior group membership is not known. Before describing this technique, a brief introduction to canonical rough sets theory is provided.

IEEE
computer
society

2. Rough sets

The concept of rough or approximation sets, was introduced by Pawlak [19], [21], [22] and is based on the assumption that, with every object of an information system (or data matrix), there is associated a certain amount of information. This information is expressed by means of attributes used as descriptions of the objects. The data is treated from the perspective of set theory and none of the traditional assumptions of multivariate analysis are relevant. For a comprehensive introduction to rough sets (RS) theory, see [16], [21] or [22].

The complete information system expresses all the knowledge available about the objects being studied. More formally, the information system is a pair, $S = (U, A)$, where U is a non-empty finite set of objects called the universe and $A = \{ a_1, \ldots, a_j \}$ is a non-empty finite set of attributes on U. With every attribute $a \in A$ we associate a set Va such that $a : U \to Va$. The set Va is called the domain or value set of a. In statistical terms, this value set equates to the range of values associated with a specific variable. The initial detailed data contained in the information system is used as the basis for the development of subsets of the data that are "coarser" or "rougher" than the original set. As with any data analysis technique, detail is lost, but the removal of detail is controlled to uncover the underlying characteristics of the data. The technique works by 'lowering the degree of precision in data, based on a rigorous mathematical theory. By selecting the right roughness or precision of data, we will find the underlying characteristics' [16:141].

A core concept of RS is that of equivalence between objects (indiscernibility). Objects about which we have the same knowledge form an equivalence relation. Let $S = (U, A)$ be an information system, then with any $B \subseteq A$ there is associated an equivalence relation, $INDA(B)$, the B-indiscernibility relation. It is defined as:

$$INDA(B) = \{ (x, x') \in U^2 \mid \forall a \in B \; a(x) = a(x') \} \; (1)$$

If $(x, x') \in INDA(B)$, then the objects x and x' are indiscernible from each other when considering the subset B of attributes. Equivalence relations lead to the universe being divided into partitions, which can then be used to build new subsets of the universe.

Let $S = (U, A)$ be an information system, and let $B \subseteq A$ and $X \subseteq U$. We can describe the subset X using only the information contained in the attribute values from the subset B by constructing two subsets, referred to as the B-lower and B-upper approximations of X, and denoted as $B_*(X)$ and $B^*(X)$ respectively, where:

$$B_*(X) = \{x \mid [x]B \subseteq X\} \text{ and } B^*(X) = \{x \mid [x]B \cap X \neq \varnothing\} \; (2,3)$$

The lower approximation (LA), defined in (2), contains objects that are definitely in the subset X and the upper approximation (UA), defined in (3), contains objects that may or may not be in X. A third subset is also useful in analysis, the boundary region, which is the difference between the upper and lower approximations. This definition of a rough set in terms of two other crisp sets is the simple but powerful insight contributed by Pawlak, which has led to numerous publications exploring the implications (e.g. [1], [18], [23], [25]).

Rough sets theory has developed an extensive literature well beyond the brief introduction provided here, and the interested reader is referred to [18], [23], [25] and [26] for comprehensive overviews of developments in the field. For example, one major extension of relevance to rough clustering is the development of the concept of *similarity*, a relaxing of the strict requirement of indiscernibility in canonical RS theory, to include objects that are similar rather than identical. A number of ways of defining similarity have been proposed, and [18] provides an introduction to the issues involved. Dissimilarity measures have also been proposed (see for example, [4]).

3. Related work

3.1. Rough clustering

The concept of a rough cluster was introduced in [29], by defining a rough cluster in a similar manner to a rough set - with a lower and upper approximation - allowing multiple cluster membership for objects in the data set. The LA of a rough cluster contains objects that only belong to that cluster, and by definition, the objects belong to the UA as well. The UA of a rough cluster contains objects that may belong to more than one cluster. The clustering algorithm described in [29] used a distance measure to construct a similarity matrix, and each object-object pair in this similarity matrix was assigned to existing or new clusters depending on whether none, one or both objects in the pair were currently assigned. Problems with this approach were the large number of clusters generated and uncertainty as to whether the lower approximations of each cluster provide the most efficient coverage of the data set.

A different approach was followed in [6], who used reducts to develop clusters. Reducts are subsets

of the attribute set A, which provide the same information as the original data set. The reducts are used as initial group centroids, which are then grouped together to form clusters. One problem with this approach is that not all information systems have reducts, and some sets of reducts overlap, which means that the cluster centroids are not necessarily well separated. Other approaches include Herawan [10], who used subsets of the total information set to determine a degree of dependency, or the relationship between the subsets, to determine the best clustering attribute, and Yanto [30], who report a variable precision rough set model.

3.2. Comparing rough and *k*-means clusters

Voges [29] reported a comparison of rough clustering with *k*-means clustering, and found that the two clustering techniques resulted in some clusters that were identified by both techniques, and some clusters that were unique to the particular technique used. The rough clustering solution is necessarily different, because of the possibility of multiple cluster membership of objects. The rough clustering technique also found clusters that were "refined" sub-clusters of those found by *k*-means clustering, and which identified a more specific sub-segment of the data set.

Rough clustering also produces more clusters than *k*-means clustering [29], with the number of clusters required to describe the data dependent on the distance measure. More clusters means an object has a higher chance of being in more than one cluster. A solution with too few clusters does not provide a useful interpretation of the partitioning of the data. On the other hand, too many clusters make interpretation difficult. In addition, the degree of overlap between the clusters needed to be minimized to ensure that each cluster provided information to aid in interpretation.

Rough clustering can be conceptualized as extracting concepts from the data, rather than strictly delineated sub-groupings. Determining a good rough cluster solution requires a trade-off between various factors. As we show below, evolutionary algorithms are a good way of conducting this trade-off.

3.3. Evolutionary algorithms and rough sets

A number of applications of evolutionary algorithms to rough clustering tasks have been reported in the literature.

Mitra [15] proposed an evolutionary rough c-means clustering algorithm to determine the relative importance of upper and lower approximations of rough sets used to model the clusters. The fitness function used in the evolutionary algorithm involved minimizing a specific measure, the Davies–Bouldin clustering validity index. Kumar [11] used an agglomerative hierarchical clustering algorithm for sequential data, where the indiscernibility relation was extended to a tolerance relation with the transitivity property being relaxed.

Bouyer [3] used a Kohonen self-organizing map for pre-processing of data, which was then further divided into clusters using rough sets and genetic algorithms. How the genetic algorithm was applied is not clearly described, but it appears to use a data structure based on inter-neuron distances in the self-organizing map. As this distance measure is based on Euclidean distances, the approach is restricted to continuous attributes. Lingras [13] developed a genome comprising two sections – LA membership and UA membership. The approach required some repair operators, as some randomly generated genes could be invalid. One limitation of this approach was that the number of clusters needed to be specified in advance, and this preliminary knowledge is not always available for larger data sets.

There have also been a number of applications of evolutionary algorithms to classification tasks using rough sets. For example, [5] used a hybrid system to develop linguistic-based technical stock market indicators with rough sets theory used to extract linguistic rules and a genetic algorithm to refine these extracted rules. The effectiveness of the proposed model was verified for both forecasting accuracy and stock returns, and showed that the proposed model was superior to rough sets and genetic algorithms applied independently. Salamó [24] proposed several rough set based measures for estimating attribute relevance for feature dimensionality reduction in Case-Based Reasoning classifiers.

In the following section we present an extension of rough clustering that attempts to overcome the limitations of these previous attempts to apply RS theory to cluster analysis. The approach suggested uses an EA to maximize the coverage of the data set, without pre-specifying the number of clusters required, without relying on structural characteristics of the cluster such as reducts, and using a straightforward fitness function.

4. Rough clustering algorithm

The discussion in the previous section has briefly mentioned some of the shortcomings of previous attempts to apply RS theory to the clustering

problem. To overcome these shortcomings, an EA based approach is proposed that attempts to find the set of lower approximations of the rough clusters, which provides the most comprehensive coverage of the data set with the minimum number of clusters [28]. The number of rough clusters is not specified in advance (except by a broad range between minimum and maximum numbers of clusters, which can be modified within the program's parameters if necessary).

4.1. Data structure

The basic data structure used for describing a rough cluster is the **template** as described in [17]. Let S = (U, A) be an information system. Any clause of the form D = ($a \in Va$) is called a **descriptor**, with the value set Va called the **range** of D. A **template** is a conjunction of unique descriptors defined over attributes from $B \subseteq A$. More formally, any propositional formula $T = \Lambda_{a \in B}$ ($a \in Va$) is called a **template** of S. To create a viable description of a cluster using a template, at least two attributes from B are chosen. This results in compact, but non-trivial, descriptions of the rough cluster.

Template T is *simple* if any descriptor of T has a range of one element. Templates with descriptors having a range of more than one element are called *generalized*. In the example presented below, only simple templates are used. However the technique could be easily extended to include generalized templates, incorporating intervals of attributes (i.e. using a similarity relation rather than an indiscernibility relation).

The data structure acted on by the EA is a cluster solution, C, which is defined as any conjunction of k unique templates,

$$C = T1 \Lambda , \ldots, \Lambda Tk \qquad (4)$$

This data structure was encoded as a simple two-dimensional array with a variable length equal to the number of unique templates in the cluster solution and a fixed width equal to the number of attributes being considered. Possible values in the template were the same as the values in the data set (-2, -1, 1, 2 in Example 1 below, or 3, 2, 1 in Example 2 below), with 0 being used as a "don't care" value. Table 1 shows an example data structure with eight unique templates, using the data presented in Example 1 (Section 5 below).

Table 1.
Data structure example

| T | Variables [1] | | | | |
	Image	Package	Price	Alcohol	Place
1	0	-2	0	0	-1
2	0	-2	0	0	1
3	0	-2	0	0	2
4	0	-1	0	1	0
5	0	-1	0	2	0
6	0	1	2	0	0
7	0	2	2	0	0
8	2	0	-2	0	0

[1] See Section 5 below

A template describes a partition of U and the conjunction of templates contained in a cluster solution results in some templates having both LAs (that is, objects satisfying one template only) and UAs (that is, objects satisfying more than one template). Consequently C is a *rough cluster* solution.

4.2. Pre-processing

The maximum number of templates generated depends on the number of attributes of each object, p, and the range of values for the attribute, q. In a simple case where each attribute has the same range, the maximum number of simple templates generated is p^q. Depending on the data set, many of these templates are of little value in describing the data, either because they cover too small a percentage of the data, or because they are trivial. For example, a cluster solution could be developed using one attribute with four values by simply having four templates - value 1, value 2, value 3 and value 4. This would cover 100% of the data set, but would only provide a simple frequency distribution.

To overcome this problem, objects in the complete set of possible templates were individually checked against the data set, and only those templates containing two or more attributes and matching more than a specified percentage of the data set were considered valid. As will be seen in the following example, this can considerably reduce the number of templates that the EA needs to sample in order to generate useful rough cluster descriptions. This has the effect of reducing the processing time required, as the EA is not dealing with templates that add little to the overall coverage or that are invalid.

For example, all of values shown in the templates presented in Table 1 occur in the original data set, and all have at least two attribute values.

4.3. Fitness measure

A number of objectives need to be considered when developing a fitness measure for rough clustering: (i) maximize the data set coverage c, defined as the fraction of the universe of objects that match the set of templates in the cluster solution, C; (ii) minimize k, the number of templates in the cluster solution, C; (iii) maximize the accuracy a, of each template [21].

More formally, for any $X \subseteq U$, the set of objects $\{ x \in X : \forall_{a \in B} a(x) \in Va \}$ from X satisfying any template T_i is denoted by $[T_i]X$. $[T_i]_*X$ is a lower approximation if x is unique to that set. $[T_i]^*X$ is an upper approximation if x is contained in $[T_i]X$ and at least one other set $[T_j]X$. We therefore define the following values:

$$c = (\Sigma \mid [T_j]_*X \mid) / \mid U \mid, \text{ where } \{1 \le j \le k\} \qquad (5)$$

That is, the **coverage** c, is the sum of the cardinal values of the LAs of each template in the cluster solution, C, divided by the cardinal value of U, the full data set.

$$a = \Sigma (\mid [T_i]_*X \mid / \mid [T_i]^*X \mid), \text{ where } \{1 \le j \le k\} \qquad (6)$$

That is, the **accuracy** a, is the sum of the cardinal value of the LA divided by the cardinal value of the UA for each template in the cluster solution, C.

The *fitness* value, f, of each cluster solution, C, is defined as the coverage multiplied by accuracy divided by the number of templates in C.

$$f = (c \times a) / k \qquad (7)$$

4.4. Recombination and mutation operators

In the current study, a multi-point operator was used to simplify the task of ensuring that only valid rough cluster solutions were generated. The size of the offspring was determined by randomly selecting a value between the sizes of both parents. Templates were then randomly selected from each parent, and then added to the offspring after checking that they were not already present in the cluster solution. In this way, a unique solution, containing material from both parents, was generated. An example of this recombination operator is presented in Table 2.

To provide another source of diversity in the population, two mutation operators were developed. The first operator (MutateAdd) randomly sampled a new template from the list of valid templates and, after checking to ensure that the template was not

already in the cluster solution, added that template to the cluster solution. The second operator (MutateDelete) randomly removed a template from the cluster solution. In this current work, the probability of mutation has been set quite high, at 50%, to ensure that all of the valid templates have been sampled for possible inclusion in a cluster solution.

Repair operators were not required as infeasible solutions were not generated by either the recombination or mutation operators. The only constraint, ensuring each cluster solution contains only unique templates, was easily handled by checking the current set of templates in the cluster solution at the time the offspring were created.

Table 2.
Recombination operator example

Parent 1					
1-1	0	-2	0	0	-1
1-2	0	-2	0	0	1
1-3	0	-2	0	0	2
1-4	0	-1	0	1	0
Parent 2					
2-1	0	-1	0	2	0
2-2	0	1	2	0	0
2-3	0	2	2	0	0
2-4	2	0	-2	0	0
			↕		
Offspring					
1-1	0	-2	0	0	-1
1-2	0	-2	0	0	1
1-3	0	-2	0	0	2
2-1	0	-1	0	2	0

5. Example 1

The evolutionary-based rough clustering algorithm described in the previous section was applied to a study of the beer preferences of "emerging drinkers" (i.e. young adults experiencing alcohol consumption for the first time). As part of a wider study, 174 participants were asked to subjectively rank attributes of beer on a four-point scale (covering Very Important, Important, Unimportant, and Very Unimportant) in terms of which attributes they considered when making a purchasing decision. Five attributes were used: image, packaging, price, alcohol content, and place sold.

The data was used to conduct a rough cluster analysis, partitioning the participants into distinct

clusters depending on which beer attributes were considered important. In this simple example with five attributes, each with a range of values of four plus a "don't care" value, the total number of possible templates was 5^5, or 3125. The pre-processing of the data, as described in Section 4.2, resulted in 112 valid templates. Only templates consisting of two or more attribute-value pairs, matching at least five percent of the data set, were considered. Only simple templates, as defined in [17], were used to develop the cluster descriptions, although the technique could easily be extended to include generalized templates (that is, where an attribute can have a continuous range of values, rather than a single value).

An EA with a varying population size from 100 to 200 was used, evolving over 1000 generations. Initially 100 potential cluster solutions were randomly produced. Both recombination and mutation operators were applied to produce 100 offspring. Parents and offspring were placed in the population pool, sorted by fitness, and the top 100 individuals were available for breeding in the next generation cycle. All individuals in the breeding population had an equal chance of being chosen as parents. The bottom 100 individuals were replaced by the offspring of the breeding population, and the process was repeated.

Table 3.
"Best" cluster solution for rough cluster analysis of beer preference data

T	Variables					$\|[T_i]^*X\|^1$	$\|[T_i]_*X\|^2$
	Image	Package	Price	Alcohol	Place		
1	*	VU	*	*	VU	26	25
2	*	VU	*	*	U	15	14
3	*	VU	*	*	I	12	11
4	*	VU	*	*	VI	10	9
5	*	U	*	U	*	11	9
6	*	U	*	I	*	11	11
7	*	U	*	VI	*	10	8
8	*	I	I	*	*	18	18
9	*	I	VI	*	*	11	11
10	*	VI	I	*	*	13	13
11	*	VI	VI	*	*	15	15
12	I	*	U	*	*	12	6
13	VI	*	VU	*	*	9	7

[1] Size of upper approximation [2] Size of lower approximation
VI-Very important, I-Important, U-Unimportant, VU-Very Unimportant,
*-"Don't care"

The "best" cluster solution obtained is shown in Table 3. This cluster achieved a coverage, c, of 94.8% of the data set. Table 3 shows the thirteen templates that comprise this cluster solution, and the size of the lower and upper approximations for each

template. The accuracy, a, of each template ranged from 0.50 to 1.00. A number of interesting cluster descriptions are apparent in the cluster solution. Template 13 shows an "image conscious" cluster, unconcerned with price, and Template 12 shows a weaker version of this. Templates 8 to 11 show various combination of importance assigned to price and packaging. Templates 6 and 7 show another group whose concern is the level of alcohol content in the beer. Templates 3 and 4 relate to the importance of the purchase location. Templates 1, 2 and 5 are difficult to interpret, as they show the "unimportance" of certain attributes, but don't trade these off against "important" attributes.

6. Example 2

The evolutionary-based rough clustering algorithm described in Section 4 was also applied to a set of city image data. The data used in the analyses was collected as part of a wider study [27], where 6,240 participants were asked to subjectively rank attributes of eleven cities on a three-point scale (High, Moderate, and Low). The eleven city destinations in the Asia Pacific region were Adelaide, Brisbane, Darwin, Melbourne, Perth, Sydney, Auckland, Christchurch, Hong Kong, Singapore, and Tokyo. The cities were ranked on seven factors: Language, Safety, Adventure, Standard of Living, Sun and Sand, Information, and Prices.

An evolutionary algorithm with a varying population size from 40 to 80 was used, evolving over 100 generations. Initially 40 potential cluster solutions were randomly produced. Both recombination and mutation operators were applied to produce 40 offspring. Parents and offspring were placed in the population pool, sorted by fitness, and the top 40 individuals were made available for recombination in the next generation cycle. All individuals in the breeding population had an equal chance of being chosen as parents. The bottom 40 individuals were replaced by the offspring of the breeding population, and the process was repeated.

The "best" cluster solution obtained is shown in Table 4. This cluster achieved coverage of 91.4% of the data set. Table 4 shows the seven templates that comprise this cluster solution, with two of the templates fully enclosed within other templates. Template 1 describes the largest cluster in the solution, comprising 33.6% of the sample. It describes perceptions of cities with a high or medium standard of living and with a high level of interest and adventure. A sub-cluster (Template 1*) completely contained within this cluster (comprising

32.7% of the cluster and 11.0% of the whole sample), describes cities with a high standard of living, a high level of interest and adventure, and for which a high level of tourist information is available. Template 2 describes cities with medium levels of adventure, medium beaches and weather, and medium price. It comprised 28.8% of the sample. Template 3 (10.5% of the sample) describes cities with medium perceptions of (same) language and low adventure levels – these could be considered "safe" cities to visit. A sub-cluster (Template 3*) completely contained within this cluster (comprising 88.1% of the cluster and 9.2% of the whole sample), describes cities with medium perceptions of (same) language and low adventure levels, coupled to medium price levels. Template 4 (9.8% of the sample) comprises cities with medium adventure, low concern with beaches and weather, and medium concern with price. Template 5 (8.8% of the sample) comprises cities with medium adventure, high concern with beaches and weather, and medium concern with price.

Table 4.
"Best" cluster solution for rough cluster analysis of city image data

T	Variables							Size	%
	L	S	A	SL	SS	I	P		
1	*	*	H	H/M	*	*	*	2096	33.6
1*	*	*	H	H	*	H	*	685	11.0
2	*	*	M	*	M	*	M	1799	28.8
3	M	*	L	*	*	*	*	653	10.5
3*	M	*	L	*	*	*	M	575	9.2
4	*	*	M	*	L	*	M	609	9.8
5	*	*	M	*	H	*	M	546	8.8

L-Language, S-Safety, A-Adventure, SL-Standard of Living, SS-Sun and Sand, I-Information, P-Price
H-High, M-Medium, L-Low, *-"Don't care"

7. Conclusions and future research

This paper has presented an extension of previously published work in rough sets theory and rough clustering. Clusters obtained from conventional techniques usually have crisp boundaries, that is, each object belongs to only one cluster, but many real-world data sets do not lend themselves to such a neat solution.

Rough clusters allow an object to belong to multiple clusters. Previous studies using evolutionary algorithms have required that the number of clusters be specified in advance, a major limitation with large or complex data sets. The research presented in this paper uses templates (conjunctions of attribute-value descriptors) to describe the cluster solution. An evolutionary algorithm was used to find a rough cluster solution that covers the largest percentage of the data set with the smallest number of accurate lower approximations. Future extensions will involve extending the template descriptions to include generalized templates, testing the algorithm on further data sets, and comparing this technique with other approaches. Further research can also be conducted on the decision-making processes used by consumers, which would help to validate the reasoning processes suggested by the rough clustering findings.

8. References

[1] Alpigini, J. J., J. F. Peters, A. Skowron, and N. Zhong (eds.), Rough Sets and Current Trends in Computing, Third International Conference, RSCTC 2002, (Lecture Notes in Computer Science, Vol. 2475), Springer-Verlag, Berlin Heidelberg, 2002.
[2] Bezdek, J. C., "Numerical Taxonomy with Fuzzy Sets", Journal of Mathematical Biology, 1, 1974, pp. 57-71.
[3] Bouyer, A., A. Hatamlou, and A. H. Abdullah, "An Optimized Clustering Algorithm Using Genetic Algorithm and Rough Set Theory Based on Kohonen Self Organizing Map", International Journal of Computer Science and Information Security, 8 (4), 2010, pp. 39-44.
[4] Cao F., J. Liang, D. Li, L. Bai, and C. Dang, "A Dissimilarity Measure for the k-Modes Clustering Algorithm", Knowledge-Based Systems, 2011, In Press, doi:10.1016/ j.knosys.2011.07.011.
[5] Cheng, C-H., T-L. Chen, and L-Y. Wei, "A Hybrid Model Based on Rough Sets Theory and Genetic Algorithms for Stock Price Forecasting", Information Sciences, 180, 2010, pp. 1610-1629.
[6] do Prado, H. A., P. M. Engel, and H. C. Filho, "Rough Clustering: An Alternative to Find Meaningful Clusters by Using the Reducts From a Dataset", in Alpigini, J. J., J. F. Peters, A. Skowron, and N. Zhong (eds.), Rough Sets and Current Trends in Computing, Third International Conference, RSCTC 2002. (Lecture Notes in Computer Science, Vol. 2475), Springer-Verlag, Berlin Heidelberg, 2002, pp. 234-238.
[7] Dunn, J. C., "A Fuzzy Relative of the ISODATA Process and its Use in Detecting Compact Well-Separated Clusters", Journal of Cybernetics, 3(3), 1974, pp. 32-57.
[8] Hair, J. E., W. C. Black, B. J. Babin, and R. E. Anderson, Multivariate Data Analysis (7th edn), Prentice-Hall, New Jersey, 2009.
[9] Hartigan, J. A., Clustering Algorithms, Wiley, New York, 1975.
[10] Herawan, T., M. M. Deris, and J. H. Abawajy, "A Rough Set Approach for Selecting Clustering Attribute", Knowledge-Based Systems, 23, 2010, pp. 220–231.
[11] Kumar, P., P. R. Krishna, R. S. Bapi, and S. K. De, "Rough Clustering of Sequential Data", Data & Knowledge Engineering, 63, 2007, pp. 183–199.
[12] Leung, Y., M. M. Fischer, W-Z. Wu, and J-S. Mi, "A Rough Set Approach for the Discovery of Classification

Rules in Interval-Valued Information Systems", International Journal of Approximate Reasoning, 47, 2008, pp. 233-246.

[13] Lingras, P., "Rough Set Clustering for Web Mining", in Proceedings of 2002 IEEE International Conference on Fuzzy Systems, 2002.

[14] MacQueen, J., "Some Methods for Classification and Analysis of Multivariate Observations", in Le Cam, L. M., and J. Neyman (eds.), Proceedings of the Fifth Berkeley Symposium on Mathematics, Statistics and Probability, Volume 1, University of California Press, Berkeley, 1967, pp. 281-298.

[15] Mitra, S., "An Evolutionary Rough Partitive Clustering," Pattern Recognition Letters, 25, 2004, pp. 1439–1449.

[16] Munakata, T., Fundamentals of the New Artificial Intelligence: Neural, Evolutionary, Fuzzy and More (2nd edn.), Springer, New York, 2008.

[17] Nguyen, S. H., "Regularity Analysis and its Applications in Data Mining", in Polkowski, L., S. Tsumoto, and T. Y. Lin (eds), Rough Set Methods and Applications: New Developments in Knowledge Discovery in Information Systems, Physica-Verlag, Heidelberg New York, 2000, pp. 289-378.

[18] Orlowska, E., Incomplete Information: Rough Set Analysis, Physica-Verlag, Heidelberg New York, 1998.

[19] Pawlak, Z., "Rough Sets", International Journal of Information and Computer Sciences, 11(5), 1982, pp. 341-356.

[20] Pawlak, Z., "Rough Classification", International Journal of Man-Machine Studies, 20, 1984, pp. 469-483.

[21] Pawlak, Z. Rough Sets: Theoretical Aspects of Reasoning About Data, Kluwer, Boston, 1991.

[22] Pawlak, Z. and A. Skowron, "Rudiments of Rough Sets", Information Sciences, 177, 2007, pp. 3-27.

[23] Polkowski, L., S. Tsumoto, and T. Y. Lin (eds.), Rough Set Methods and Applications: New Developments in Knowledge Discovery in Information Systems, Physica-Verlag, Heidelberg New York, 2000.

[24] Salamó, M. and M. López-Sánchez, "Rough Set Based Approaches to Feature Selection for Case-Based Reasoning Classifiers", Pattern Recognition Letters, 32, 2011, pp. 280-292.

[25] Szczuka, M., M. Kryszkiewicz, S. Ramanna, R. Jensen, and Q. Hu (eds.), Rough Sets and Current Trends in Computing: 7th International Conference, RSCTC 2010, Warsaw, Poland, June 28-30, 2010 Proceedings, Springer, New York, 2010.

[26] Thangavel, K. and A. Pethalakshmi, "Dimensionality Reduction Based on Rough Set Theory: A Review", Applied Soft Computing, 9, 2009, 1-12.

[27] Voges, K. E., "Rough Clustering of Destination Image Data Using an Evolutionary Algorithm", Journal of Travel and Tourism Marketing, 21(4), 2006, pp. 121-137.

[28] Voges, K. E. and N. K. Ll. Pope "Generating Compact Rough Cluster Descriptions Using an Evolutionary Algorithm", in K. Deb et al. (eds.) GECCO2004: Genetic and Evolutionary Algorithm Conference (LNCS 3103), Springer-Verlag, Berlin, 2004, pp. 1332-1333.

[29] Voges, K. E., N. K. Ll. Pope, and M. R. Brown, "Cluster Analysis of Marketing Data Examining On-line Shopping Orientation: A Comparison of *K*-means and Rough Clustering Approaches", in Abbass, H. A., R. A. Sarker, and C. S. Newton (eds.), Heuristics and Optimization for Knowledge Discovery, Idea Group Publishing, Hershey, PA, 2002, pp. 207-224.

[30] Yanto, I. T. R., P. Vitasari, T. Herawan, and M. Mat Deris, "Applying Variable Precision Rough Set Model for Clustering Student Suffering Study's Anxiety", Expert Systems with Applications, 2011, In Press, doi: 10.1016/j.eswa.2011.07.036.

[31] Zadeh, L. A., "Fuzzy Sets", Information and Control, 8, 1965, pp. 338-353.