

# Rough Sets in Machine Learning: A Review

Rafael Bello and Rafael Falcon

**Abstract** This chapter emphasizes on the role played by *rough set theory* (RST) within the broad field of Machine Learning (ML). As a sound data analysis and knowledge discovery paradigm, RST has much to offer to the ML community. We surveyed the existing literature and reported on the most relevant RST theoretical developments and applications in this area. The review starts with RST in the context of *data preprocessing* (discretization, feature selection, instance selection and meta-learning) as well as the generation of both *descriptive and predictive knowledge* via decision rule induction, association rule mining and clustering. Afterward, we examined *several special ML scenarios* in which RST has been recently introduced, such as imbalanced classification, multi-label classification, dynamic/incremental learning, Big Data analysis and cost-sensitive learning.

## 1 Introduction

Information granulation is the process by which a collection of *information granules* are synthesized, with a granule being a collection of values (in the data space) which are drawn towards the center object(s) (in the object space) by an underlying indistinguishability, similarity or functionality mechanism. Note that the data and object spaces can actually coincide [141]. The *Granular Computing* (GrC) paradigm

---

R. Bello (✉)

Department of Computer Science, Universidad Central de Las Villas,  
Carretera Camajuaní km 5.5, Santa Clara, Cuba  
e-mail: rbellop@uclv.edu.cu

R. Falcon

Research & Engineering Division, Larus Technologies Corporation,  
170 Laurier Ave West - Suite 310, Ottawa, ON, Canada  
e-mail: rafael.falcon@larus.com; rfalcon@uottawa.ca

R. Falcon

School of Electrical Engineering and Computer Science, University of Ottawa,  
800 King Edward Ave, Ottawa, ON, Canada

© Springer International Publishing AG 2017

G. Wang et al. (eds.), *Thriving Rough Sets*, Studies in Computational Intelligence 708, DOI 10.1007/978-3-319-54966-8\_5

[7, 183] encompasses several computational models based on fuzzy logic, Computing With Words, interval computing, rough sets, shadowed sets, near sets, etc.

The main purpose behind Granular Computing is to find a novel way to synthesize knowledge in a more human-centric fashion and from vast, unstructured, possibly high-dimensional raw data sources. Not surprisingly, Granular Computing (GrC) is closely related to Machine Learning [83, 95, 257]. The aim of a learning process is to derive a certain rule or system for either the automatic classification of the system objects or the prediction of the values of the system control variables. The key challenge with prediction lies in modeling the relationships among the system variables in such a way that it allows inferring the value of the control (target) variable.

*Rough set theory* (RST) [1] was developed by Zdzisław Pawlak in the early 1980s [179] as a mathematical approach to intelligent data analysis and data mining [180]. This methodology is based on the premise that lowering the degree of precision in the data makes the data pattern more visible, i.e., the rough set approach can be formally considered as a framework for pattern discovery from imperfect data [220]. Several reasons are given in [34] to employ RST in knowledge discovery, including:

- It does not require any preliminary or additional information about the data
- It provides a valuable analysis even in presence of incomplete data
- It allows the interpretation of large amounts of both quantitative and qualitative data
- It can model highly nonlinear or discontinuous functional relations to provide complex characterizations of data
- It can discover important facts hidden in the data and represent them in the form of decision rules, and
- At the same time, the decision rules derived from rough set models are based on facts, because every decision rule is supported by a set of examples.

Mert Bal [3] brought up other RST advantages, such as: (a) it performs a clear interpretation of the results and evaluation of the meaningfulness of data; (b) it can identify and characterize uncertain systems and (c) the patterns discovered using rough sets are concise, strong and sturdy.

Among the main components of the knowledge discovery process we can mention:

- PREPROCESSING
  - Discretization
  - Training set edition (instance selection)
  - Feature selection
  - Characterization of the learning problem (data complexity, metalearning)
- KNOWLEDGE DISCOVERY
  - Symbolic inductive learning methods
  - Symbolic implicit learning methods (a.k.a. lazy learning)

- KNOWLEDGE EVALUATION

- Evaluation of the discovered knowledge

All of the above stages have witnessed the involvement of rough sets in their algorithmic developments. Some of the RST applications are as follows:

- Analysis of the attributes to consider
  - Feature selection
  - Inter-attribute dependency characterization
  - Feature reduction
  - Feature weighting
  - Feature discretization
  - Feature removal
- Formulation of the discovered knowledge
  - Discovery of decision rules
  - Quantification of the uncertainty in the decision rules.

RST's main components are an *information system* and an *indiscernibility relation*. An information system is formally defined as follows. Let  $A = \{A_1, A_2, \dots, A_n\}$  be a set of attributes characterizing each example (object, entity, situation, state, etc.) in non-empty set  $U$  called the universe of discourse. The pair  $(U, A)$  is called an *information system*. If there exists an attribute  $d \notin A$ , called the *decision attribute*, that represents the decision associated with each example in  $U$ , then a *decision system*  $(U, A \cup \{d\})$  is obtained.

The fact that RST relies on the existence of an information system allows establishing a close relationship with data-driven knowledge discovery processes given that these information or decision systems can be employed as training sets for unsupervised or supervised learning models, respectively.

A binary *indiscernibility relation*  $I_B$  is associated with each subset of attributes  $B \subseteq A$ . This relation contains the pairs of objects that are inseparable from each other given the information expressed in the attributes in  $B$ , as shown in Eq. (1).

$$I_B = \{(x, y) \in U \times U : f(x, A_i) = f(y, A_i) \ \forall A_i \in B\}. \quad (1)$$

where  $f(x, A_i)$  returns the value of the  $i$ -th attribute in object  $x \in U$ .

The indiscernibility relation induces a granulation of the information system. The classical RST leaned on a particular type of indiscernibility relations called equivalence relations (i.e., those that are symmetric, reflexive and transitive). An equivalence relation induces a granulation of the universe in the form of a partition. This type of relation works well when there are only nominal attributes and no missing values in the information system.

Information systems having incomplete, continuous, mixed or heterogeneous data are in need of a more flexible type of indiscernibility relation. Subsequent RST formulations relaxed the stringent requirement of having an equivalence relation by

considering either a tolerance or a similarity relation [61, 68, 181, 207, 212, 231, 283, 284, 305, 306]; these relations will induce a *covering* of the system. Another relaxation avenue is based on the probabilistic approach [65, 182, 210, 259, 264, 267, 307]. A third alternative is the hybridization with fuzzy set theory [54, 55, 172, 258, 280]. These different approaches have contributed to positioning RST as an important component within Soft Computing [12].

All of the aforementioned RST formulations retain some basic definitions, such as the lower and upper approximations; however, they defined it in multiple ways. The canonical RST definition for the *lower approximation* of a concept  $X$  is given as  $B_*(X) = \{x \in U : B(x) \subseteq X\}$  whereas its *upper approximation* is calculated as  $B^*(X) = \{x \in U : B(x) \cap X \neq \emptyset\}$ . From these approximations we can compute the *positive region*  $POS(X) = B_*(X)$ , the *boundary region*  $BND(X) = B^*(X) - B_*(X)$  and the *negative region*  $NEG(X) = U - B^*(X)$ . These concepts serve as building blocks for developing many problem-solving approaches, including data-driven learning.

RST and Machine Learning are also related in that both take care of removing irrelevant/redundant attributes. This process is termed *feature selection* and RST approaches it from the standpoint of calculating the system reducts. Given an information system  $S = (U, A)$ , where  $U$  is the universe and  $A$  is the set of attributes, a *reduct* is a minimum set of attributes  $B \subseteq A$  such that  $I_A = I_B$ .

This chapter emphasizes on the role played by RST within the broad field of Machine Learning (ML). As a sound data analysis and knowledge discovery paradigm, RST has much to offer to the ML community. We surveyed the existing literature and reported on the most relevant RST theoretical developments and applications in this area. The review starts with RST in the context of *data preprocessing* (discretization, feature selection, instance selection and meta-learning) as well as the generation of both *descriptive and predictive knowledge* via decision rule induction, association rule mining and clustering. Afterward, we examined *several special ML scenarios* in which RST has been recently introduced, such as imbalanced classification, multi-label classification, dynamic/incremental learning, Big Data analysis and cost-sensitive learning.

The rest of the chapter is structured as follows. Section 2 reviews ML methods and processes from an RST standpoint, with emphasis on data preprocessing and knowledge discovery. Section 3 unveils special ML scenarios that are being gradually permeated by RST-based approaches, including imbalanced classification, multi-label classification, dynamic/incremental learning, Big Data analysis and cost-sensitive learning. Section 5 concludes the chapter.

## 2 Machine Learning Methods and RST

This section briefly goes over reported studies showcasing RST as a tool in data preprocessing and descriptive/predictive knowledge discovery.

## 2.1 Preprocessing

### 2.1.1 Discretization

As mentioned in [195], *discretization* is the process of converting a numerical attribute into a nominal one by applying a set of cuts to the domain of the numerical attribute and treating each interval as a discrete value of the (now nominal) attribute. Discretization is a mandatory step when processing information systems with the canonical RST formulation, as there is no provisioning for handling numerical attributes there. Some RST extensions avoid this issue by, for example, using similarity classes instead of equivalence classes and building a similarity relation that encompasses both nominal and numerical attributes.

It is very important that any discretization method chosen in the context of RST-based data analysis preserves the underlying discernibility among the objects. The level of granularity at which the cuts are performed in the discretization step will have a direct impact on any ensuing prediction, i.e., generic (wider) intervals (cuts) will likely avoid overfitting when predicting the class for an unseen object.

Dougherty et al. [53] categorize discretization methods along three axes:

- *global versus local*: indicates whether an approach simultaneously converts all numerical attributes (global) or is restricted to a single numerical attribute (local). For instance, the authors in [174] suggest both local and global handling of numerical attributes in large data bases.
- *supervised versus unsupervised*: indicates whether an approach considers values of other attributes in the discretization process or not. A simple example of an unsupervised approach is an “equal width” interval method that works by dividing the range of continuous attributes into  $k$  equal intervals, where  $k$  is given. A supervised discretization method, for example, will consider the correlation between the numerical attribute and the label (class) attribute when choosing the location of the cuts.
- *static versus dynamic*: indicates whether an approach requires a parameter to determine the number of cut values or not. Dynamic approaches automatically generate this number along the discretization process whereas static methods require an a priori specification of this parameter.

Lenarcik and Piasta [128] introduced an RST-based discretization method that leans on the concepts of a random information system and of an expected value of classification quality. The method of finding suboptimal discretizations based on these concepts is presented and is illustrated with data from concretes’ frost resistance investigations.

Nguyen [173] considers the problem of searching for a minimal set of cuts that preserves the discernibility between objects with respect to any subset of  $s$  attributes, where  $s$  is a user-defined parameter. It was shown that this problem is NP-hard and its heuristic solution is more complicated than that for the problem of searching for

an optimal, consistent set of cuts. The author proposed a scheme based on Boolean reasoning to solve this problem.

Bazan [5] put forth a method to search for an irreducible sets of cuts of an information system. The method is based on the notion of *dynamic reduct*. These reducts are calculated for the information system and the one with the best stability coefficient is chosen. Next, as an irreducible set of cuts, the author selected cuts belonging to the chosen dynamic reduct.

Bazan et al. [6] proposed a discretization technique named *maximal discernibility* (MD), which is based on rough sets and Boolean reasoning. MD is a greedy heuristic that searches for cuts along the domains of all numerical attributes that discern the largest number of object pairs in the dataset. These object pairs are removed from the information system before the next cut is sought. The set of cuts obtained that way is optimal in terms of object indiscernibility; however this procedure is not feasible since computing one cut requires  $O(|A| \cdot |U|^3)$ . Locally optimal cuts [6] are computed in  $O(|A| \cdot |U|)$  steps using only  $O(|A| \cdot |U|)$  space.

Dai and Li [46] improved Nguyen's discretization techniques by reducing the time and space complexity required to arrive at the set of candidate cuts. They proved that all *bound cuts* can discern the same object pairs as the entire set of initial cuts. A strategy to select candidate cuts was proposed based on that proof. They obtained identical results to Nguyen's with a lower computational overhead.

Chen et al. [26] employ a genetic algorithm (GA) to derive the minimal cut set in a numerical attribute. Each gene in a binary chromosome represents a particular cut value. Enabling this gene means the corresponding cut value has been selected as a member of the minimal cut set. Some optimization strategies such as elitist selection and father-offspring combined selection helped the GA converge faster. The experimental evidence showed that the GA-based scheme is more efficient than Nguyen's basic heuristic based on rough sets and Boolean reasoning.

Xie et al. [249] defined an information entropy value for every candidate cut point in their RST-based discretization algorithm. The final cut points are selected based on this metric and some RST properties. The authors report that their approach outperforms other discretization techniques and scales well with the number of cut points.

Su and Hsu [219] extended the modified Chi2 discretizer by learning the pre-defined misclassification rate (input parameter) from data. The authors additionally considered the effect of variance in the two adjacent intervals. In the modified Chi2, the inconsistency check in the original Chi2 is replaced with the "quality of approximation" measure from RST. The result is a more robust, parameterless discretization method.

Singh and Minz [205] designed a hybrid clustering-RST-based discretizer. The values of each numerical attribute are grouped using density-based clustering algorithms. This produces a set of (possibly overlapping) intervals that naturally reflect the data distribution. Then, the rough membership function in RST is employed to refine these intervals in a way that maximizes class separability. The proposed scheme yielded promising results when compared to seven other discretizers.

Jun and Zhou [116] enhanced existing RST-based discretizers by (i) computing the candidate cuts with an awareness of the decision class information; in this way, the scales of candidate cuts can be remarkably reduced, thus considerably saving time and space and (ii) introducing a notion of cut selection probability that is defined to measure cut significance in a more reasonable manner. Theoretical analyses and simulation experiments show that the proposed approaches can solve the problem of data discretization more efficiently and effectively.

### 2.1.2 Feature Selection

The purpose behind feature selection is to discard irrelevant features that are generally detrimental to the classifier's performance, generate noise, increase the amount of information to be stored and the computational cost of the classification process [222, 302]. Feature selection is a computationally expensive problem that requires searching for a subset of the  $n$  original features in a space of  $2^n - 1$  candidate subsets according to a predefined evaluation criterion. The main components of a feature selection algorithm are: (1) an *evaluation function* (EF), used to calculate the fitness of a feature subset and (2) a *generation procedure* that is responsible for generating different subsets of candidate features.

Different feature selection schemes that integrate RST into the feature subset evaluation function have been developed. The *quality of the classification*  $\gamma$  is the most frequently used RST metric to judge the suitability of a candidate feature subset, as shown in [9–11, 64] etc. Other indicators are *conditional independence* [208] and *approximate entropy* [209].

The concept of *reduct* is the basis for these results. Essentially, a reduct is a minimal subset of features that generates the same granulation of the universe as that induced by all features. Among these works we can list [37, 38, 85, 89, 111, 136, 168, 196, 221, 223, 239, 247, 248, 255, 270, 302]. One of the pioneer methods is the *QuickReduct* algorithm, which is typical of those algorithms that resort to a greedy search strategy to find a relative reduct [136, 202, 247]. Generally speaking, feature selection algorithms are based on heuristic search [97, 164, 302]. Other RST-based methods for reduct calculation are [98, 209].

More advanced methods employ metaheuristic algorithms (such as Genetic Algorithms, Ant Colony Optimization or Particle Swarm Optimization) as the underlying feature subset generation engine [8–11, 15, 64, 102, 119, 241, 242, 245, 246, 268, 274, 297]. Feature selection methods based on the hybridization between fuzzy and rough sets have been proposed in [13, 28, 42–44, 51, 75, 87, 90, 92, 101, 103–105, 125, 193, 197, 203, 225, 299]. Some studies aim at calculating all possible reducts of a decision system [27, 28, 206, 225, 299].

Feature selection is arguably the Machine Learning (ML) area that has witnessed the most influx of rough-set-based methods. Other RST contributions to ML are concerned with providing metrics to calculate the inter-attribute dependence and the importance (weight) of any attribute [120, 222].

### 2.1.3 Instance Selection

Another important data preprocessing task is the editing of the training sets, also referred to as *instance selection*. The aim is to reduce the number of examples in order to bring down the size of the training set while maintaining the system efficiency. By doing that, a new training set is obtained that will bring forth a higher efficiency usually also produces a reduction of the data.

Some training set edition approaches using rough sets have been published in [16, 19]. The simplest idea is to remove all examples in the training set that are not contained in the lower approximation of any of the decision classes. A more thorough investigation also considers those examples that lie in the boundary region of any of the decision classes. Fuzzy rough sets have been also applied to the instance selection problem in [99, 232, 233].

### 2.1.4 Meta-Learning

An important area within knowledge discovery is that of *meta-learning*, whose objective is to learn about the underlying learning processes in order to make them more efficient or effective [234]. These methods may consider measures related to the complexity of the data [79]. The study in [18] explores the use of RST-based metrics to estimate the quality of a data set. The relationship between the “quality of approximation” measure and the performance of some classifiers is investigated in [17]. This measure describe the inexactness of the rough-set-based classification and denotes the percentage of examples that were correctly classified employing the attributes included in the indiscernibility relationship [224]. The authors in [251] analyze the inclusion degree as a perspective on measures for rough set data analysis (RSDA). Other RSDA measures are the “accuracy of the approximation” and the rough membership function [120]; for example, in [108, 109], the rough membership function and other RST-based measures are employed to detect outliers (i.e., examples that behave in an unexpected way or have abnormal properties).

## 2.2 Descriptive and Predictive Knowledge Discovery

### 2.2.1 Decision Rule Induction

The knowledge uncovered by the different data analysis techniques can be either *descriptive* or *predictive*. The former characterizes the general properties of the data in the data set (e.g., association rules) while the latter allows performing inferences from the available data (e.g., decision rules). A decision rule summarizes the relationship between the properties (features) and describes a causal relationship among them. For example, IF Headache = Yes AND Weakness = YES THEN Influenza =



YES. The most common **rule induction** task is to generate a rule base  $R$  that is both consistent and complete.

According to [161], RST-based rule induction methods provide the following benefits:

- Better explanation capabilities
- Generate a simple and useful set of rules.
- Work with sparse training sets.
- Work even when the underlying data distribution significantly deviates from the normal distribution.
- Work with incomplete, inaccurate, and heterogeneous data.
- Usually faster execution time to generate the rule base compared to other methods.
- No assumptions made on the size or distribution of the training data.

Among the most popular RST-based rule induction methods we can cite LERS [67, 215], which includes the LEM1 (Learn from examples model v1) and LEM2 methods (Learn from examples model v2); the goal is to extract a minimum set of rules to cover the examples by exploring the attribute-value pairs search space of while taking into account possible data inconsistency issues. MODLEM [214, 215] is based on sequentially building coverings of the training data and generating minimal decision rule sets for each decision class. Each of these sets aims at covering all positive examples that belong to a concept and none from any other concept. The EXPLORE algorithm [216] extracts from data all the decision rules satisfying certain requirements. It can be adapted to handle inconsistent examples. The LEM2, EXPLORE and MODLEM algorithms rule induction algorithms are implemented in the ROSE2 software [3]. Filiberto et al. proposed the IRBASIR method [62], which generates decision rules using an RST extension rooted on similarity relations; another technique is put forth in [121] to discover rules using similarity relations for incomplete data sets. This learning problem in presence of missing data is also addressed in [80].

Other **RST-based rule induction algorithms** available in the literature using rough sets are [3, 14, 63, 110, 118, 129, 154, 179, 228, 229]. The use of hybrid models based on rough sets and fuzzy sets for rule induction and other knowledge discovery methods is illustrated in [2, 24, 41, 100, 123, 159, 201, 298, 300], which includes working with the so called “fuzzy decision information systems” [2].

One of the most popular rule induction methods based on rough sets is the so-called *three-way decisions* model [81, 260–263]. This methodology is strongly related to *decision making*. Essentially, for each decision alternative, this method defines three rules based on the RST’s positive, negative and boundary regions. They respectively indicate acceptance, rejection or abstention (non-commitment, denotes weak or insufficient evidence).

This type of rules, derived from the basic RST concepts, is a suitable knowledge representation vehicle in a plethora of application domains. Hence, it has been integrated into common machine learning tasks to facilitate the knowledge engineering process required for a successful modeling of the domain under consideration. The

three-way decisions model has been adopted in feature selection [106, 107, 133, 163, 265, 293], classification [273, 281, 282, 293], clustering [276, 277] and face recognition [132, 289].

### 2.2.2 Association Rule Mining

The discovery of **association rules** is one of the classical data mining tasks. Its goal is to uncover relationships among attributes that frequently appear together; i.e., the presence of one implies the presence of the other. One of the typical examples is the purchase of beer and diapers during the weekends. Association rules are representative of descriptive knowledge. A particular case are the so called “class association rules”, which are used to build classifiers. Several methods have been developed for discovering association rules using rough sets, including [49, 70, 94, 111, 127, 134, 211, 266].

### 2.2.3 Clustering

The **clustering** problem is another learning task that has been approached from a rough set perspective. Clustering is a landmark unsupervised learning problem whose main objective is to group similar objects in the same cluster and separate objects that are different from each other by assigning them to different clusters [96, 167]. The objects are grouped in such a way that those in the same group exhibit a high degree of association among them whereas those in different groups show a low degree of association. Clustering algorithms map the original  $N$ -dimensional feature space to a 1-dimensional space describing the cluster each object belongs to. This is why clustering is considered both an important dimensionality reduction technique and also one of the most prevalent Granular Computing [183] manifestations.

One of the most popular and efficient clustering algorithms for conventional applications is *K-means clustering* [71]. In the K-means approach, randomly selected objects serve as initial cluster centroids. The objects are then assigned to different clusters based on their distance to the centroids. In particular, an object gets assigned to the cluster with the nearest centroid. The newly modified clusters then employ this information to determine new centroids. The process continues iteratively until the cluster centroids are stabilized. K-means is a very simple clustering algorithm, easy to understand and implement. The underlying alternate optimization approach iteratively converges but might get trapped into a local minimum of the objective function. K-means’ best performance is attained in those applications where clusters are well separated and a crisp (bivalent) object-to-cluster decision is required. Its disadvantages include the sensitivity to outliers and the initial cluster centroids as well as the a priori specification of the desired number of clusters  $k$ .

Pawan Lingras [142, 145] found that the K-means algorithm often yields clustering results with unclear, vague boundaries. He pointed out that the “hard partitioning” performed by K-means does not meet the needs of grouping vague data. Lingras

then proposed to combine K-means with RST and in the so-called “Rough K-means” approach. In this technique, each cluster is modeled as a rough set and each object belongs either to the lower approximation of a cluster or to the upper approximation of multiple clusters. Instead of building each cluster, its lower and upper approximations are defined based on the available data. The basic properties of the Rough K-means method are: (i) an object can be a member of at most a lower approximation; (ii) an object that is a member of the lower approximation of a cluster is also a member of its upper approximation and (iii) an object that does not belong to the lower approximation of any cluster is a member of at least the upper approximation of two clusters. Other pioneering works on rough clustering methods are put forth in [78, 192, 235, 236].

Rough K-means has been the subject of several subsequent studies aimed at improving its clustering capabilities. Georg Peters [187] concludes that rough clustering offers the possibility of reducing the number of incorrectly clustered objects, which is relevant to many real-world applications where minimizing the number of wrongly grouped objects is more important than maximizing the number of correctly grouped objects. Hence in these scenarios, Rough K-means arises as a powerful and stronger alternative to K-means. The same author proposes some improvements to the method regarding the calculation of the centroids, thus aiming to make the method more stable and robust to outliers [184, 185]. The authors in [291] proposed a Rough K-means improvement based on a variable weighted distance measure. Another enhancement brought forward in [186] suggested that well-defined objects must have a greater impact on the cluster centroid calculation rather than having this impact be governed by the number of cluster boundaries an object belongs to, as proposed in the original method. An extension to Rough K-means based on the decision-theoretic rough sets model was developed in [130]. An evolutionary approach for rough partitive clustering was designed in [168, 189] while [45, 190] elaborate on dynamic rough clustering approaches.

Other works that tackle the clustering problem using rough sets are [35, 72, 76, 77, 122, 124, 135, 143, 144, 162, 177, 178, 213, 271, 272, 275, 292]. These methods handle more specific scenarios (such as sequential, imbalanced, categorical and ordinal data), as well as applications of this clustering approach to different domains. The rough-fuzzy K-means method is put forward in [88, 170] whereas the fuzzy-rough K-means is unveiled in [169, 188]. Both approaches amalgamate the main features of Rough K-means and Fuzzy C-means by using the fuzzy membership of the objects to the rough clusters. Other variants of fuzzy and rough set hybridization for the clustering problem are presented in [56, 126, 160, 171].

### 3 Special Learning Cases Based on RST

This section elaborates on more recent ML scenarios tackled by RST-based approaches. In particular, we review the cases of imbalanced classification, multi-label classification, dynamic/incremental learning and Big Data analysis.

### 3.1 *Imbalanced Classification*

The traditional knowledge discovery methods presented in the previous section have to be adapted if we are dealing with an *imbalanced dataset* [21]. A dataset is balanced if it has an approximately equal percentage of positive and negative examples (i.e., those belonging to the concept to be classified and those belonging to other concepts, respectively). However, there are many application domains where we find an imbalanced dataset; for instance, in healthcare scenarios there are usually a plethora of patients that do not have a particularly rare disease. When learning a normalcy model for a certain environment, the number of labeled anomalous events is often scarce as most of the data corresponds to normal behaviour. The problem with imbalanced classes is that the classification algorithms have a tendency towards favoring the majority class. This occurs because the classifier attempts to reduce the overall error, hence the classification error does not take into account the underlying data distribution [23].

Several solutions have been researched to deal with this kind of situations. Two of the most popular avenues are either resampling the training data (i.e., oversampling the minority class or undersampling the majority class) or modifying the learning method [153]. One of the classical methods for learning with imbalanced data is SMOTE (*synthetic minority oversampling technique*) [22]. Different learning methods for imbalanced classification have been developed from an RST-based standpoint. For instance, Hu et al. [91] proposed models based on probabilistic rough sets where each example has an associated probability  $p(x)$  instead of the default  $1/n$ . Ma et al. [158] introduced weights in the variable-precision rough set model (VPRS) to denote the importance of each example. Liu et al. [153] bring about some weights in the RST formulation to balance the class distribution and develop a method based on weighted rough sets to solve the imbalanced class learning problem. Ramentol et al. [194] proposed a method that integrates SMOTE with RST.

Stefanowski et al. [217] introduced filtering techniques to process inconsistent examples of the majority class (i.e., those lying in the boundary region), thereby adapting the MODLEM rule extraction method for coping with imbalanced learning problems. Other RST-based rule induction methods in the context of imbalanced data are also presented in [152, 243]. The authors in [218] proposed the EXPLORE method that generates rules for the minority class with a minimum coverage equal to a user-specified threshold.

### 3.2 *Multi-label Classification*

Normally, in a typical classification problem, a class (label)  $c_i$  from a set  $C = \{c_1, \dots, c_k\}$  is assigned to each example. However, in **multi-label classification**, a subset  $S \subseteq C$  is assigned to each example, which means that an example could belong to multiple classes. Some applications of this type of learning emerge from text clas-

sification and functional genomics, namely, assigning functions to genes [226]. This gives rise to the so-called *multi-label learning* problem. The two avenues envisioned for solving this new class of learning problems have considered either converting the multi-label scenario to a single-label (classical) scenario or adapting the learning methods. Examples of the latter trend are the schemes proposed in [47, 198, 227, 290]. Similar approaches have been proposed for multi-label learning using rough sets. A first alternative is to transform the multi-label problem into a traditional single-label case and use classical RST-based learning methods to derive the rules (or any other knowledge); the other option is to adapt the RST-based learning methods, as shown in [240, 278, 279, 288].

In the first case, a decision system can be generated where some instances could belong to multiple classes. Multi-label classification can be regarded as an inconsistent decision problem, in which two objects having the same predictive attribute values do not share the same decision class. This leads to the modification of the definition of the lower/upper approximations through a probabilistic approach that facilitates modeling the uncertainty generated by the inconsistent system. This idea gives rise to the so-called *multi-label rough set model*, which incorporates a probabilistic approach such as the decision-theoretic rough set model. Some RST-based feature selection methods in multi-label learning scenarios have been enunciated [131], where the reduct concept was reformulated for the multi-label case.

### 3.3 *Dynamic/Incremental Learning*

Data are continuously being updated in nowadays' information systems. New data are added and obsolete data are purged over time. Traditional batch-learning methods lean on the principle of running these algorithms on all data when the information is updated, which obviously affects the system efficiency while ignoring any previous learning. Instead, learning should occur as new information arrives. Managing this learning while adapting the previous knowledge learned is the essence behind *incremental learning*. This term refers to an efficient strategy for the analysis of data in dynamic environments that allows acquiring additional knowledge from an uninterrupted information flow. The advantage of incremental learning is not to have to analyze the data from scratch but to utilize the learning process' previous outcomes as much as possible [57, 73, 112, 176, 200]. The continuous and massive acquisition of data becomes a challenge for the discovery of knowledge; especially in the context of Big Data, it becomes very necessary to develop capacities to assimilate the continuous data streams [29].

As an information-based methodology, RST is not exempt from being scrutinized in the context of dynamic data. The fundamental RST concepts and the knowledge discovery methods ensuing from them are geared towards the analysis of static data; hence, they need to be thoroughly revised in light of the requirements posed by data stream mining systems [151]. The purpose of the incremental learning strategy in rough sets is the development of incremental algorithms to quickly update

the concept approximations, the reduct calculation or the discovered decision rules [40, 284]. The direct precursor of these studies can be found in [175]. According to [149], in recent years RST-based incremental learning approaches have become “hot topics” in knowledge extraction from dynamic data given their proven data analysis efficiency.

The study of RST in the context of learning with dynamic data can be approached from two different angles: what kind of information is considered to be dynamic and what type of learning task must be carried out. In the first case, the RST-based incremental updating approach could be further subdivided into three alternatives: (i) object variation (insertion or deletion of objects in the universe), (ii) attribute variation (insertion/removal of attributes) and (iii) attribute value variation (insertion/deletion of attribute values). In the second case, we can mention (i) incremental learning of the concept approximations [33, 139]; (ii) incremental learning of attribute reduction [52, 140, 237, 238, 250] and (iii) incremental learning of decision rules [59, 66, 148, 301].

Object variations include so-called object immigration and emigration [148]. Variations of the attributes include feature insertion or deletion [138, 287]. Variations in attribute values are primarily manifested via the refinement or scaling of the attribute values [32, 146]. Other works that propose modifications to RST-based methods for the case of dynamic data are [147, 149, 157].

The following studies deal with dynamic object variation:

- The update of the lower and upper approximations of the target concept is analyzed in [33, 137, 156].
- The update in the reduction of attributes is studied in [82, 250].
- The update of the decision rule induction mechanism is discussed in [4, 40, 59, 93, 148, 199, 230, 244, 269, 301].

If the variation occurs in the set of attributes, its effects have been studied with respect to these aspects:

- The update of the lower and upper approximations of the target concept is analyzed in [20, 36, 138, 139, 150, 287].
- The update of the decision rule induction mechanism is discussed in [39].

The effect of the variations in the attribute values (namely, via refinement or extension of the attribute domains) with respect to the update of the lower and upper approximations of the target concept is analyzed in [30–32, 50, 237, 308].

The calculation of reducts for dynamic data has also been investigated. The effect when the set of attributes varies is studied in [39]. The case of varying the attribute values is explored in [50, 69] whereas the case of dynamic object update is dissected in [199, 244]. Other studies on how dynamic data affect the calculation of reducts appear in [140, 204, 237, 238].

### 3.4 Rough Sets and Big Data

On the other hand, the accelerated pace of technology has led to an exponential growth in the generation and collection of digital information. This growth is not only limited to the amount of data available but to the plethora of diverse sources that emit these data streams. It becomes paramount then to efficiently analyze and extract knowledge from many dissimilar information sources within a certain application domain. This has led to the emergence of the *Big Data* era [25], which has a direct impact on the development of RST and its applications. Granular Computing, our starting point in this chapter, has a strong relation to Big Data [25], as its inherent ability to process information at multiple levels of abstraction and interpret information from different perspectives greatly facilitates the efficient management of large data volumes.

Simply put, Big Data can be envisioned as a large and complex data collection. These data are very difficult to analyze through traditional data management and processing tools. Big Data scenarios require new architectures, techniques, algorithms and processes to manage and extract value and knowledge hidden in the data streams. Big Data is often characterized by the 5 V's vector: Volume, Velocity, Variety, Veracity and Value. Big Data includes both structured and unstructured data, including images, videos, textual reports, etc. Big Data frameworks such as MapReduce and Spark have been recently developed and constitute indispensable tools for the accurate and seamless knowledge extraction from an array of disparate data sources. For more information on the Big Data paradigm, the reader is referred to the following articles: [25, 48, 60, 117].

As a data analysis and information extraction methodology, RST needs to adapt and evolve in order to cope with this new phenomenon. A major motivation to do so lies in the fact that the sizes of nowadays' decision systems are already extremely large. This poses a significant challenge to the efficient calculation of the underlying RST concepts and the knowledge discovery methods that emanate from them. Recall that the computational complexity of computing the target concept's approximations is  $\mathbf{O}(lm^2)$ , the computational cost of finding a reduct is bounded by  $\mathbf{O}(l^2m^2)$  and the time complexity to find all reducts is  $\mathbf{O}(2^lJ)$ , where  $l$  is the number of attributes characterizing the objects,  $m$  is the number of objects in the universe and  $J$  is the computational cost required to calculate a reduct.

Some researchers have proposed RST-based solutions to the Big Data challenge [191, 286]. These methods are concerned with the design of parallel algorithms to compute equivalence classes, decision classes, associations between equivalence classes and decision classes, approximations, and so on. They are based on partitioning the universe, concurrently processing those information subsystems and then integrating the results. In other words, given the decision system  $S = (U, C \cup D)$ , generate the subsystems  $\{S_1, S_2, \dots, S_m\}$ , where  $S_i = (U_i, C \cup D)$  and  $U = \bigcup U_i$ , then process each subsystem  $S_i$ ,  $i \in \{1, 2, \dots, m\}$ ,  $U_i/B, B \subseteq C$ . Afterwards, the results are amalgamated. This MapReduce-compliant workflow is supported by several theorems stating that (a) equivalence classes can be independently computed

for each subsystem and (b) the equivalence classes from different subsystems can be merged if they are based on the same underlying attribute set. These results enable the parallel computation of the equivalence classes of the decision system  $S$ . Zhang et al. [286] developed the PACRSEC algorithm to that end.

Analogously, RST-based knowledge discovery methods, including reduct calculation and decision rule induction, have been investigated in the context of Big Data [58, 256, 285].

### 3.5 Cost-Sensitive Learning

Cost is an important property inherent to real-world data. Cost sensitivity is an important problem which has been addressed from different angles. *Cost-sensitive learning* [252, 294, 303, 304] emerged when an awareness of the learning context was brought into Machine Learning. This is one of the most difficult ML problems and was listed as one of the top ten challenges in the Data Mining/ML domain [296].

Two types of learning costs have been addressed through RST: *misclassification cost* and *test cost* [253]. Test cost has been studied by Min et al. [163, 165, 166, 295] using the classical rough set approach, i.e., using a single granulation; a test-cost-sensitive multigranulation rough set model is presented in [253]. Multigranulation rough set is an extension of the classical RST that leans on multiple granular structures.

A recent cost-sensitive rough set approach was put forward in [115]. The crux of this method is that the information granules are sensitive to test costs while approximations are sensitive to decision costs, respectively; in this way, the construction of the rough set model takes into account both the test cost and the decision cost simultaneously. This new model is called *cost-sensitive rough set* and is based on decision-theoretic rough sets. In [132], the authors combine sequential three-way decisions and cost-sensitive learning to solve the face recognition problem; this is particularly interesting since in real-world face recognition scenarios, different kinds of misclassifications will lead to different costs [155, 294].

Other studies focused on the cost-sensitive learning problem from an RST perspective are presented in [84, 113, 253, 254]; these works have considered both the test cost and the decision cost. Attribute reduction based on test-cost-sensitivity has been quite well investigated [74, 86, 106, 114, 115, 133, 163, 164, 166, 296].

## 4 Reference Categorization

Table 1 lists the different RST studies according to the ML tasks they perform.



**Table 1** Rough sets in machine learning

Task	Subproblem/Approach	References	
Data preprocessing	Subtask		
	Discretization	[5, 6, 46, 173, 174]	
	Feature selection	Rough sets and Boolean reasoning	[26, 116, 128, 205, 219, 249]
		Other approaches	[37, 38, 85, 89, 98, 106, 107, 111, 133, 136, 163, 168, 196, 202, 208, 209, 221, 223, 239, 247, 248, 255, 265, 270, 293, 302]
		Reduct calculation	[97, 164, 302]
		Heuristic search	[8–11, 15, 64, 102, 119, 241, 242, 245, 246, 268, 274, 297]
	Instance selection	Metaheuristic search	
		Fuzzy and rough set hybridization	[13, 28, 42–44, 51, 75, 87, 90, 92, 101, 103–105, 125, 193, 197, 203, 225, 299]
		Rough sets	[16, 19]
	Meta-learning	Fuzzy and rough set hybridization	[99, 232, 233]
Quality of training set		[17, 18, 251]	
Outlier detection		[108, 109]	
Miscellaneous	Inter-attribute dependence/attribute importance	[120, 222]	
	Rough sets	[3, 14, 62, 63, 67, 110, 118, 129, 154, 179, 214–216, 228, 229]	
Knowledge discovery	Decision rule induction	Fuzzy and rough set hybridization	[2, 24, 41, 100, 123, 159, 201, 298, 300]
		Three-way decisions	[81, 260–263, 273, 281, 282, 293]
	Association rule mining	Rough sets	[49, 70, 94, 111, 127, 134, 211, 266]
		Rough clustering	[78, 130, 142, 145, 184–187, 192, 235, 236, 291]
	Clustering	Evolutionary/dynamic rough clustering	[45, 168, 189, 190]
		Other rough clustering scenarios	[35, 56, 72, 76, 77, 122, 124, 135, 143, 144, 162, 177, 178, 213, 271, 272, 275, 292]
		Fuzzy and rough set hybridization	[56, 88, 126, 160, 169–171, 188]
	Three-way decisions	[276, 277]	

(continued)

Table 1 (continued)

Task	Subtask	Subproblem/Approach	References
Special learning cases	Imbalanced classification	Rough set theory adaptation	[91, 152, 153, 158, 194, 217, 218, 243]
	Multi-label classification	Rough set theory adaptation	[131, 240, 278, 279, 288]
	Dynamic/incremental learning	Dynamic objects	[4, 33, 40, 58, 82, 93, 137, 146, 148, 156, 199, 230, 244, 250, 269, 301]
		Dynamic attributes	[138, 287]
		Dynamic attribute values	[30–32, 32, 50, 146, 237, 308]
		Incremental learning of concept approximations	[20, 33, 137–139, 150, 156, 287]
		Incremental learning of attribute reduction	[52, 82, 140, 237, 238, 250, 250]
		Incremental learning of decision rules	[4, 39, 40, 59, 66, 93, 146, 199, 230, 244, 269, 301, 301]
		Dynamic reducts	[39, 50, 69, 140, 199, 204, 237, 238, 244]
		Miscellaneous	[40, 147, 149, 151, 157, 175, 284]
	Big Data analysis	Rough set theory adaptation	[58, 191, 256, 285, 286]
	Cost-sensitive learning	Test-cost-aware	[74, 86, 106, 114, 132, 133, 163–166, 166, 253, 295, 296]
		Test-and-decision-cost aware	[84, 113, 115, 254]

## 5 Conclusions

We have reported on hundreds of successful attempts to tackle different ML problems using RST. These approaches touch all components of the knowledge discovery process, ranging from data preprocessing to descriptive and predictive knowledge induction. Aside from the well-known RST strengths in identifying inconsistent information systems, calculating reducts to reduce the dimensionality of the feature space or generating an interpretable rule base, we have walked the reader through more recent examples that show the redefinition of some of the RST's building blocks to make it a suitable approach for handling special ML scenarios characterized by an imbalance in the available class data, the requirement to classify a pattern into one or more predefined labels, the dynamic processing of data streams, the need to manage large volumes of static data or the management of misclassification/test costs. All of these efforts bear witness to the resiliency and adaptability of the rough set approach, thus making it an appealing choice for solving non-conventional ML problems.

## References

1. Abraham, A., Falcon, R., Bello, R.: *Rough Set Theory: A True Landmark in Data Analysis*. Springer, Berlin, Germany (2009)
2. Bai, H., Ge, Y., Wang, J., Li, D., Liao, Y., Zheng, X.: A method for extracting rules from spatial data based on rough fuzzy sets. *Knowl. Based Syst.* **57**, 28–40 (2014)
3. Bal, M.: Rough sets theory as symbolic data mining method: an application on complete decision table. *Inf. Sci. Lett.* **2**(1), 111–116 (2013)
4. Bang, W.C., Bien, Z.: New incremental learning algorithm in the framework of rough set theory. *Int. J. Fuzzy Syst.* **1**, 25–36 (1999)
5. Bazan, J.G.: A comparison of dynamic and non-dynamic rough set methods for extracting laws from decision tables. *Rough Sets Knowl Discovery* **1**, 321–365 (1998)
6. Bazan, J.G., Nguyen, H.S., Nguyen, S.H., Synak, P., Wróblewski, J.: Rough set algorithms in classification problem. In: *Rough Set Methods and Applications*, pp. 49–88. Springer (2000)
7. Bello, R., Falcon, R., Pedrycz, W., Kacprzyk, J.: *Granular Computing: At the Junction of Rough Sets and Fuzzy Sets*. Springer, Berlin, Germany (2008)
8. Bello, R., Gómez, Y., Caballero, Y., Nowe, A., Falcon, R.: Rough sets and evolutionary computation to solve the feature selection problem. In: Abraham, A., Falcon, R., Bello, R. (eds.) *Rough Set Theory: A True Landmark in Data Analysis*. *Studies in Computational Intelligence*, vol. 174, pp. 235–260. Springer, Berlin (2009)
9. Bello, R., Nowe, A., Gómez, Y., Caballero, Y.: Using ACO and rough set theory to feature selection. *WSEAS Trans. Inf. Sci. Appl.* **2**(5), 512–517 (2005)
10. Bello, R., Puris, A., Falcon, R., Gómez, Y.: Feature selection through dynamic mesh optimization. In: Ruiz-Shulcloper, J., Kropatsch, W. (eds.) *Progress in Pattern Recognition, Image Analysis and Applications*. *Lecture Notes in Computer Science*, vol. 5197, pp. 348–355. Springer, Berlin (2008)
11. Bello, R., Puris, A., Nowe, A., Martínez, Y., García, M.M.: Two step ant colony system to solve the feature selection problem. In: *Iberoamerican Congress on Pattern Recognition*, pp. 588–596. Springer (2006)
12. Bello, R., Verdegay, J.L.: Rough sets in the soft computing environment. *Inf. Sci.* **212**, 1–14 (2012)

13. Bhatt, R.B., Gopal, M.: On fuzzy-rough sets approach to feature selection. *Pattern Recogn. Lett.* **26**(7), 965–975 (2005)
14. Błaszczyński, J., Słowiński, R., Szelkag, M.: Sequential covering rule induction algorithm for variable consistency rough set approaches. *Inf. Sci.* **181**(5), 987–1002 (2011)
15. Caballero, Y., Bello, R., Alvarez, D., Garcia, M.M.: Two new feature selection algorithms with rough sets theory. In: *IFIP International Conference on Artificial Intelligence in Theory and Practice*, pp. 209–216. Springer (2006)
16. Caballero, Y., Bello, R., Alvarez, D., Gareia, M.M., Pizano, Y.: Improving the k-nn method: rough set in edit training set. In: *Professional Practice in Artificial Intelligence*, pp. 21–30. Springer (2006)
17. Caballero, Y., Bello, R., Arco, L., García, M., Ramentol, E.: Knowledge discovery using rough set theory. In: *Advances in Machine Learning I*, pp. 367–383. Springer (2010)
18. Caballero, Y., Bello, R., Arco, L., Márquez, Y., León, P., García, M.M., Casas, G.: Rough set theory measures for quality assessment of a training set. In: *Granular Computing: At the Junction of Rough Sets and Fuzzy Sets*, pp. 199–210. Springer (2008)
19. Caballero, Y., Joseph, S., Lezcano, Y., Bello, R., Garcia, M.M., Pizano, Y.: Using rough sets to edit training set in k-nn method. In: *ISDA*, pp. 456–463 (2005)
20. Chan, C.C.: A rough set approach to attribute generalization in data mining. *Inf. Sci.* **107**(1), 169–176 (1998)
21. Chawla, N.V.: Data mining for imbalanced datasets: an overview. In: *Data Mining and Knowledge Discovery Handbook*, pp. 853–867. Springer (2005)
22. Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P.: Smote: synthetic minority over-sampling technique. *J. Artif. Intell. Res.* **16**, 321–357 (2002)
23. Chawla, N.V., Cieslak, D.A., Hall, L.O., Joshi, A.: Automatically countering imbalance and its empirical relationship to cost. *Data Min. Knowl. Discovery* **17**(2), 225–252 (2008)
24. Chen, C., Mac Parthaláin, N., Li, Y., Price, C., Quek, C., Shen, Q.: Rough-fuzzy rule interpolation. *Inf. Sci.* **351**, 1–17 (2016)
25. Chen, C.P., Zhang, C.Y.: Data-intensive applications, challenges, techniques and technologies: a survey on big data. *Inf. Sci.* **275**, 314–347 (2014)
26. Chen, C.Y., Li, Z.G., Qiao, S.Y., Wen, S.P.: Study on discretization in rough set based on genetic algorithm. In: *2003 International Conference on Machine Learning and Cybernetics*, vol. 3, pp. 1430–1434. IEEE (2003)
27. Chen, D., Hu, Q., Yang, Y.: Parameterized attribute reduction with gaussian kernel based fuzzy rough sets. *Inf. Sci.* **181**(23), 5169–5179 (2011)
28. Chen, D., Zhang, L., Zhao, S., Hu, Q., Zhu, P.: A novel algorithm for finding reducts with fuzzy rough sets. *IEEE Trans. Fuzzy Syst.* **20**(2), 385–389 (2012)
29. Chen, H., Chiang, R.H., Storey, V.C.: Business intelligence and analytics: from big data to big impact. *MIS Q.* **36**(4), 1165–1188 (2012)
30. Chen, H., Li, T., Qiao, S., Ruan, D.: A rough set based dynamic maintenance approach for approximations in coarsening and refining attribute values. *Int. J. Intell. Syst.* **25**(10), 1005–1026 (2010)
31. Chen, H., Li, T., Ruan, D.: Dynamic maintenance of approximations under a rough-set based variable precision limited tolerance relation. *J. Multiple-Valued Log. Soft Comput.* **18** (2012)
32. Chen, H., Li, T., Ruan, D.: Maintenance of approximations in incomplete ordered decision systems while attribute values coarsening or refining. *Knowl. Based Syst.* **31**, 140–161 (2012)
33. Chen, H., Li, T., Ruan, D., Lin, J., Hu, C.: A rough-set-based incremental approach for updating approximations under dynamic maintenance environments. *IEEE Trans. Knowl. Data Eng.* **25**(2), 274–284 (2013)
34. Chen, Y.S., Cheng, C.H.: A delphi-based rough sets fusion model for extracting payment rules of vehicle license tax in the government sector. *Expert Syst. Appl.* **37**(3), 2161–2174 (2010)
35. Cheng, X., Wu, R.: Clustering path profiles on a website using rough k-means method. *J. Comput. Inf. Syst.* **8**(14), 6009–6016 (2012)
36. Cheng, Y.: The incremental method for fast computing the rough fuzzy approximations. *Data Knowl. Eng.* **70**(1), 84–100 (2011)

37. Choubey, S.K., Deogun, J.S., Raghavan, V.V., Sever, H.: A comparison of feature selection algorithms in the context of rough classifiers. In: Proceedings of the Fifth IEEE International Conference on Fuzzy Systems, 1996, vol. 2, pp. 1122–1128. IEEE (1996)
38. Chouchoulas, A., Shen, Q.: A rough set-based approach to text classification. In: International Workshop on Rough Sets, Fuzzy Sets, Data Mining, and Granular-Soft Computing, pp. 118–127. Springer (1999)
39. Ciucci, D.: Attribute dynamics in rough sets. In: International Symposium on Methodologies for Intelligent Systems, pp. 43–51. Springer (2011)
40. Ciucci, D.: Temporal dynamics in information tables. *Fundamenta Informaticae* **115**(1), 57–74 (2012)
41. Coello, L., Fernandez, Y., Filiberto, Y., Bello, R.: Improving the multilayer perceptron learning by using a method to calculate the initial weights with the similarity quality measure based on fuzzy sets and particle swarms. *Computación y Sistemas* **19**(2), 309–320 (2015)
42. Cornelis, C., Jensen, R.: A noise-tolerant approach to fuzzy-rough feature selection. In: IEEE International Conference on Fuzzy Systems, 2008. FUZZ-IEEE 2008. (IEEE World Congress on Computational Intelligence), pp. 1598–1605. IEEE (2008)
43. Cornelis, C., Jensen, R., Hurtado, G., Šle, D., et al.: Attribute selection with fuzzy decision reducts. *Inf. Sci.* **180**(2), 209–224 (2010)
44. Cornelis, C., Verbiest, N., Jensen, R.: Ordered weighted average based fuzzy rough sets. In: International Conference on Rough Sets and Knowledge Technology, pp. 78–85. Springer (2010)
45. Crespo, F., Peters, G., Weber, R.: Rough clustering approaches for dynamic environments. In: *Rough Sets: Selected Methods and Applications in Management and Engineering*, pp. 39–50. Springer (2012)
46. Dai, J.H., Li, Y.X.: Study on discretization based on rough set theory. In: 2002 International Conference on Machine Learning and Cybernetics, 2002. Proceedings, vol. 3, pp. 1371–1373. IEEE (2002)
47. De Comité, F., Gilleron, R., Tommasi, M.: Learning multi-label alternating decision trees from texts and data. In: International Workshop on Machine Learning and Data Mining in Pattern Recognition, pp. 35–49. Springer (2003)
48. Dean, J., Ghemawat, S.: Mapreduce: simplified data processing on large clusters. *Commun. ACM* **51**(1), 107–113 (2008)
49. Delic, D., Lenz, H.J., Neiling, M.: Improving the quality of association rule mining by means of rough sets. In: *Soft Methods in Probability, Statistics and Data Analysis*, pp. 281–288. Springer (2002)
50. Deng, D., Huang, H.: Dynamic reduction based on rough sets in incomplete decision systems. In: International Conference on Rough Sets and Knowledge Technology, pp. 76–83. Springer (2007)
51. Derrac, J., Cornelis, C., García, S., Herrera, F.: Enhancing evolutionary instance selection algorithms by means of fuzzy rough set based feature selection. *Inf. Sci.* **186**(1), 73–92 (2012)
52. Dey, P., Dey, S., Datta, S., Sil, J.: Dynamic discredution using rough sets. *Appl. Soft Comput.* **11**(5), 3887–3897 (2011)
53. Dougherty, J., Kohavi, R., Sahami, M., et al.: Supervised and unsupervised discretization of continuous features. *Machine Learning: Proceedings of the Twelfth International Conference* **12**, 194–202 (1995)
54. Dubois, D., Prade, H.: Twofold fuzzy sets and rough sets some issues in knowledge representation. *Fuzzy Sets Syst.* **23**(1), 3–18 (1987)
55. Dubois, D., Prade, H.: Rough fuzzy sets and fuzzy rough sets\*. *Int. J. Gen. Syst.* **17**(2–3), 191–209 (1990)
56. Falcon, R., Jeon, G., Bello, R., Jeong, J.: Rough clustering with partial supervision. In: *Rough Set Theory: A True Landmark in Data Analysis*, pp. 137–161. Springer (2009)
57. Falcon, R., Nayak, A., Abielmona, R.: An Online shadowed clustering algorithm applied to risk visualization in territorial security. In: IEEE Symposium on Computational Intelligence for Security and Defense Applications (CISDA), pp. 1–8. Ottawa, Canada (2012)

58. Fan, Y.N., Chern, C.C.: An agent model for incremental rough set-based rule induction: a Big Data analysis in sales promotion. In: 2013 46th Hawaii International Conference on System Sciences (HICSS), pp. 985–994. IEEE (2013)
59. Fan, Y.N., Tseng, T.L.B., Chern, C.C., Huang, C.C.: Rule induction based on an incremental rough set. *Expert Syst. Appl.* **36**(9), 11439–11450 (2009)
60. Fernández, A., del Río, S., López, V., Bawakid, A., del Jesus, M.J., Benítez, J.M., Herrera, F.: Big data with cloud computing: an insight on the computing environment, mapreduce, and programming frameworks. *Wiley Interdisc. Rev. Data Min. Knowl. Discovery* **4**(5), 380–409 (2014)
61. Filiberto, Y., Caballero, Y., Larrua, R., Bello, R.: A method to build similarity relations into extended rough set theory. In: 2010 10th International Conference on Intelligent Systems Design and Applications, pp. 1314–1319. IEEE (2010)
62. Filiberto Cabrera, Y., Caballero Mota, Y., Bello Pérez, R., Frías, M.: Algoritmo para el aprendizaje de reglas de clasificación basado en la teoría de los conjuntos aproximados extendida. *Dyna*; vol. 78, núm. 169 (2011); 62-70 *DYNA*; vol. 78, núm. 169 (2011); 62-70 2346-2183 0012-7353 (2011)
63. Gogoi, P., Bhattacharyya, D.K., Kalita, J.K.: A rough set-based effective rule generation method for classification with an application in intrusion detection. *Int. J. Secur. Netw.* **8**(2), 61–71 (2013)
64. Gómez, Y., Bello, R., Puris, A., Garcia, M.M., Nowe, A.: Two step swarm intelligence to solve the feature selection problem. *J. UCS* **14**(15), 2582–2596 (2008)
65. Greco, S., Matarazzo, B., Słowiński, R.: Parameterized rough set model using rough membership and bayesian confirmation measures. *Int. J. Approximate Reasoning* **49**(2), 285–300 (2008)
66. Greco, S., Słowiński, R., Stefanowski, J., Żurawski, M.: Incremental versus non-incremental rule induction for multicriteria classification. In: *Transactions on Rough Sets II*, pp. 33–53. Springer (2004)
67. Grzymala-Busse, J.W.: LERS—a system for learning from examples based on rough sets. In: *Intelligent decision support*, pp. 3–18. Springer (1992)
68. Grzymała-Busse, J.W.: Characteristic relations for incomplete data: A generalization of the indiscernibility relation. In: *International Conference on Rough Sets and Current Trends in Computing*, pp. 244–253. Springer (2004)
69. Grzymala-Busse, J.W., Grzymala-Busse, W.J.: Inducing better rule sets by adding missing attribute values. In: *International Conference on Rough Sets and Current Trends in Computing*, pp. 160–169. Springer (2008)
70. Guan, J., Bell, D.A., Liu, D.: The rough set approach to association rule mining. In: *Third IEEE International Conference on Data Mining, 2003. ICDM 2003*, pp. 529–532. IEEE (2003)
71. Hartigan, J.A., Wong, M.A.: Algorithm as 136: a k-means clustering algorithm. *J. Roy. Stat. Soc. Ser. C (Appl. Stat.)* **28**(1), 100–108 (1979)
72. Hassanein, W., Elmelegy, A.A.: An algorithm for selecting clustering attribute using significance of attributes. *Int. J. Database Theory Appl.* **6**(5), 53–66 (2013)
73. He, H., Chen, S., Li, K., Xu, X.: Incremental learning from stream data. *IEEE Trans. Neural Netw.* **22**(12), 1901–1914 (2011)
74. He, H., Min, F., Zhu, W.: Attribute reduction in test-cost-sensitive decision systems with common-test-costs. In: *Proceedings of the 3rd International Conference on Machine Learning and Computing*, vol. 1, pp. 432–436 (2011)
75. He, Q., Wu, C., Chen, D., Zhao, S.: Fuzzy rough set based attribute reduction for information systems with fuzzy decisions. *Knowl. Based Syst.* **24**(5), 689–696 (2011)
76. Herawan, T.: Rough set approach for categorical data clustering. Ph.D. thesis, Universiti Tun Hussein Onn Malaysia (2010)
77. Herawan, T., Deris, M.M., Abawajy, J.H.: A rough set approach for selecting clustering attribute. *Knowl. Based Syst.* **23**(3), 220–231 (2010)

78. Hirano, S., Tsumoto, S.: Rough clustering and its application to medicine. *J. Inf. Sci.* **124**, 125–137 (2000)
79. Ho, T.K., Basu, M.: Complexity measures of supervised classification problems. *IEEE Trans. Pattern Anal. Mach. Intell.* **24**(3), 289–300 (2002)
80. Hong, T.P., Tseng, L.H., Wang, S.L.: Learning rules from incomplete training examples by rough sets. *Expert Syst. Appl.* **22**(4), 285–293 (2002)
81. Hu, B.Q.: Three-way decisions space and three-way decisions. *Inf. Sci.* **281**, 21–52 (2014)
82. Hu, F., Wang, G., Huang, H., Wu, Y.: Incremental attribute reduction based on elementary sets. In: *International Workshop on Rough Sets, Fuzzy Sets, Data Mining, and Granular-Soft Computing*, pp. 185–193. Springer (2005)
83. Hu, H., Shi, Z.: Machine learning as granular computing. In: *IEEE International Conference on Granular Computing, 2009, GRC'09*, pp. 229–234. IEEE (2009)
84. Hu, Q., Che, X., Zhang, L., Zhang, D., Guo, M., Yu, D.: Rank entropy-based decision trees for monotonic classification. *IEEE Trans. Knowl. Data Eng.* **24**(11), 2052–2064 (2012)
85. Hu, Q., Liu, J., Yu, D.: Mixed feature selection based on granulation and approximation. *Knowl. Based Syst.* **21**(4), 294–304 (2008)
86. Hu, Q., Pan, W., Zhang, L., Zhang, D., Song, Y., Guo, M., Yu, D.: Feature selection for monotonic classification. *IEEE Trans. Fuzzy Syst.* **20**(1), 69–81 (2012)
87. Hu, Q., Xie, Z., Yu, D.: Hybrid attribute reduction based on a novel fuzzy-rough model and information granulation. *Pattern Recogn.* **40**(12), 3509–3521 (2007)
88. Hu, Q., Yu, D.: An improved clustering algorithm for information granulation. In: *International Conference on Fuzzy Systems and Knowledge Discovery*, pp. 494–504. Springer (2005)
89. Hu, Q., Yu, D., Liu, J., Wu, C.: Neighborhood rough set based heterogeneous feature subset selection. *Inf. Sci.* **178**(18), 3577–3594 (2008)
90. Hu, Q., Yu, D., Xie, Z.: Information-preserving hybrid data reduction based on fuzzy-rough techniques. *Pattern Recogn. Lett.* **27**(5), 414–423 (2006)
91. Hu, Q., Yu, D., Xie, Z., Liu, J.: Fuzzy probabilistic approximation spaces and their information measures. *IEEE Trans. Fuzzy Syst.* **14**(2), 191–201 (2006)
92. Hu, Q., Zhang, L., An, S., Zhang, D., Yu, D.: On robust fuzzy rough set models. *IEEE Trans. Fuzzy Syst.* **20**(4), 636–651 (2012)
93. Huang, C.C., Tseng, T.L.B., Fan, Y.N., Hsu, C.H.: Alternative rule induction methods based on incremental object using rough set theory. *Appl. Soft Comput.* **13**(1), 372–389 (2013)
94. Huang, Z., Hu, Y.Q.: Applying AI technology and rough set theory to mine association rules for supporting knowledge management. In: *2003 International Conference on Machine Learning and Cybernetics*, vol. 3, pp. 1820–1825. IEEE (2003)
95. Hüllermeier, E.: Granular computing in machine learning and data mining. In: *Handbook of Granular Computing*, pp. 889–906 (2008)
96. Jain, A.K., Murty, M.N., Flynn, P.J.: Data clustering: a review. *ACM Comput. Surv. (CSUR)* **31**(3), 264–323 (1999)
97. Janusz, A., Slezak, D.: Rough set methods for attribute clustering and selection. *Appl. Artif. Intell.* **28**(3), 220–242 (2014)
98. Janusz, A., Stawicki, S.: Applications of approximate reducts to the feature selection problem. In: *International Conference on Rough Sets and Knowledge Technology*, pp. 45–50. Springer (2011)
99. Jensen, R., Cornelis, C.: Fuzzy-rough instance selection. In: *2010 IEEE International Conference on Fuzzy Systems (FUZZ)*, pp. 1–7. IEEE (2010)
100. Jensen, R., Cornelis, C., Shen, Q.: Hybrid fuzzy-rough rule induction and feature selection. In: *IEEE International Conference on Fuzzy Systems, 2009. FUZZ-IEEE 2009*, pp. 1151–1156. IEEE (2009)
101. Jensen, R., Shen, Q.: Fuzzy-rough sets for descriptive dimensionality reduction. In: *Proceedings of the 2002 IEEE International Conference on Fuzzy Systems, 2002. FUZZ-IEEE'02*, vol. 1, pp. 29–34. IEEE (2002)
102. Jensen, R., Shen, Q.: Finding rough set reducts with ant colony optimization. In: *Proceedings of the 2003 UK Workshop on Computational Intelligence*, vol. 1, pp. 15–22 (2003)

103. Jensen, R., Shen, Q.: Fuzzy-rough attribute reduction with application to web categorization. *Fuzzy Sets Syst.* **141**(3), 469–485 (2004)
104. Jensen, R., Shen, Q.: Semantics-preserving dimensionality reduction: rough and fuzzy-rough-based approaches. *IEEE Trans. Knowl. Data Eng.* **16**(12), 1457–1471 (2004)
105. Jensen, R., Shen, Q.: New approaches to fuzzy-rough feature selection. *IEEE Trans. Fuzzy Syst.* **17**(4), 824–838 (2009)
106. Jia, X., Liao, W., Tang, Z., Shang, L.: Minimum cost attribute reduction in decision-theoretic rough set models. *Inf. Sci.* **219**, 151–167 (2013)
107. Jia, X., Shang, L., Zhou, B., Yao, Y.: Generalized attribute reduct in rough set theory. *Knowl. Based Syst.* **91**, 204–218 (2016)
108. Jiang, F., Sui, Y., Cao, C.: Outlier detection based on rough membership function. In: *International Conference on Rough Sets and Current Trends in Computing*, pp. 388–397. Springer (2006)
109. Jiang, F., Sui, Y., Cao, C.: Some issues about outlier detection in rough set theory. *Expert Syst. Appl.* **36**(3), 4680–4687 (2009)
110. Jiang, Y.C., Liu, Y.Z., Liu, X., Zhang, J.K.: Constructing associative classifier using rough sets and evidence theory. In: *International Workshop on Rough Sets, Fuzzy Sets, Data Mining, and Granular-Soft Computing*, pp. 263–271. Springer (2007)
111. Jiao, X., Lian-cheng, X., Lin, Q.: Association rules mining algorithm based on rough set. In: *International Symposium on Information Technology in Medicine and Education*, Print ISBN, pp. 978–1 (2012)
112. Joshi, P., Kulkarni, P.: Incremental learning: areas and methods—a survey. *Int. J. Data Min. Knowl. Manage. Process* **2**(5), 43 (2012)
113. Ju, H., Yang, X., Song, X., Qi, Y.: Dynamic updating multigranulation fuzzy rough set: approximations and reducts. *Int. J. Mach. Learn. Cybern.* **5**(6), 981–990 (2014)
114. Ju, H., Yang, X., Yang, P., Li, H., Zhou, X.: A moderate attribute reduction approach in decision-theoretic rough set. In: *Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing*, pp. 376–388. Springer (2015)
115. Ju, H., Yang, X., Yu, H., Li, T., Yu, D.J., Yang, J.: Cost-sensitive rough set approach. *Inf. Sci.* **355**, 282–298 (2016)
116. Jun, Z., Zhou, Y.H.: New heuristic method for data discretization based on rough set theory. *J. China Univ. Posts Telecommun.* **16**(6), 113–120 (2009)
117. Kambatla, K., Kollias, G., Kumar, V., Grama, A.: Trends in big data analytics. *J. Parallel Distrib. Comput.* **74**(7), 2561–2573 (2014)
118. Kaneiwa, K.: A rough set approach to mining connections from information systems. In: *Proceedings of the 2010 ACM Symposium on Applied Computing*, pp. 990–996. ACM (2010)
119. Ke, L., Feng, Z., Ren, Z.: An efficient ant colony optimization approach to attribute reduction in rough set theory. *Pattern Recogn. Lett.* **29**(9), 1351–1357 (2008)
120. Komorowski, J., Pawlal, Z., Polkowski, L., Skowron, A.: A rough set perspective on data and knowledge. In: *The Handbook of Data Mining and Knowledge Discovery*. Oxford University Press, Oxford (1999)
121. Kryszkiewicz, M.: Rough set approach to incomplete information systems. *Inf. Sci.* **112**(1), 39–49 (1998)
122. Kumar, P., Krishna, P.R., Bapi, R.S., De, S.K.: Rough clustering of sequential data. *Data Knowl. Eng.* **63**(2), 183–199 (2007)
123. Kumar, P., Vadakkepat, P., Poh, L.A.: Fuzzy-rough discriminative feature selection and classification algorithm, with application to microarray and image datasets. *Appl. Soft Comput.* **11**(4), 3429–3440 (2011)
124. Kumar, P., Wasan, S.K.: Comparative study of k-means, pam and rough k-means algorithms using cancer datasets. In: *Proceedings of CSIT: 2009 International Symposium on Computing, Communication, and Control (ISCCC 2009)*, vol. 1, pp. 136–140 (2011)
125. Kuncheva, L.I.: Fuzzy rough sets: application to feature selection. *Fuzzy Sets Syst.* **51**(2), 147–153 (1992)



126. Lai, J.Z., Juan, E.Y., Lai, F.J.: Rough clustering using generalized fuzzy clustering algorithm. *Pattern Recogn.* **46**(9), 2538–2547 (2013)
127. Lee, S.C., Huang, M.J.: Applying ai technology and rough set theory for mining association rules to support crime management and fire-fighting resources allocation. *J. Inf. Technol. Soc.* **2**(65), 65–78 (2002)
128. Lenarcik, A., Piasta, Z.: Discretization of condition attributes space. In: *Intelligent Decision Support*, pp. 373–389. Springer (1992)
129. Leung, Y., Fischer, M.M., Wu, W.Z., Mi, J.S.: A rough set approach for the discovery of classification rules in interval-valued information systems. *Int. J. Approximate Reasoning* **47**(2), 233–246 (2008)
130. Li, F., Ye, M., Chen, X.: An extension to rough c-means clustering based on decision-theoretic rough sets model. *Int. J. Approximate Reasoning* **55**(1), 116–129 (2014)
131. Li, H., Li, D., Zhai, Y., Wang, S., Zhang, J.: A variable precision attribute reduction approach in multilabel decision tables. *Sci. World J.* **2014** (2014)
132. Li, H., Zhang, L., Huang, B., Zhou, X.: Sequential three-way decision and granulation for cost-sensitive face recognition. *Knowl. Based Syst.* **91**, 241–251 (2016)
133. Li, H., Zhou, X., Zhao, J., Liu, D.: Non-monotonic attribute reduction in decision-theoretic rough sets. *Fundamenta Informaticae* **126**(4), 415–432 (2013)
134. Li, J., Cercone, N.: A rough set based model to rank the importance of association rules. In: *International Workshop on Rough Sets, Fuzzy Sets, Data Mining, and Granular-Soft Computing*, pp. 109–118. Springer (2005)
135. Li, M., Deng, S., Wang, L., Feng, S., Fan, J.: Hierarchical clustering algorithm for categorical data using a probabilistic rough set model. *Knowl. Based Syst.* **65**, 60–71 (2014)
136. Li, M., Shang, C., Feng, S., Fan, J.: Quick attribute reduction in inconsistent decision tables. *Inf. Sci.* **254**, 155–180 (2014)
137. Li, S., Li, T., Liu, D.: Dynamic maintenance of approximations in dominance-based rough set approach under the variation of the object set. *Int. J. Intell. Syst.* **28**(8), 729–751 (2013)
138. Li, S., Li, T., Liu, D.: Incremental updating approximations in dominance-based rough sets approach under the variation of the attribute set. *Knowl. Based Syst.* **40**, 17–26 (2013)
139. Li, T., Ruan, D., Geert, W., Song, J., Xu, Y.: A rough sets based characteristic relation approach for dynamic attribute generalization in data mining. *Knowl. Based Syst.* **20**(5), 485–494 (2007)
140. Liang, J., Wang, F., Dang, C., Qian, Y.: A group incremental approach to feature selection applying rough set technique. *IEEE Trans. Knowl. Data Eng.* **26**(2), 294–308 (2014)
141. Lin, T.Y., Yao, Y.Y., Zadeh, L.A.: Data mining, rough sets and granular computing. *Physica* **95** (2013)
142. Lingras, P.: Unsupervised rough set classification using gas. *J. Intell. Inf. Syst.* **16**(3), 215–228 (2001)
143. Lingras, P., Chen, M., Miao, D.: Rough cluster quality index based on decision theory. *IEEE Trans. Knowl. Data Eng.* **21**(7), 1014–1026 (2009)
144. Lingras, P., Chen, M., Miao, D.: Qualitative and quantitative combinations of crisp and rough clustering schemes using dominance relations. *Int. J. Approximate Reasoning* **55**(1), 238–258 (2014)
145. Lingras, P., West, C.: Interval set clustering of web users with rough k-means. *J. Intell. Inf. Syst.* **23**(1), 5–16 (2004)
146. Liu, D., Li, T., Liu, G., Hu, P.: An approach for inducing interesting incremental knowledge based on the change of attribute values. In: *IEEE International Conference on Granular Computing, 2009, GRC'09*, pp. 415–418. IEEE (2009)
147. Liu, D., Li, T., Ruan, D., Zhang, J.: Incremental learning optimization on knowledge discovery in dynamic business intelligent systems. *J. Glob. Optim.* **51**(2), 325–344 (2011)
148. Liu, D., Li, T., Ruan, D., Zou, W.: An incremental approach for inducing knowledge from dynamic information systems. *Fundamenta Informaticae* **94**(2), 245–260 (2009)
149. Liu, D., Li, T., Zhang, J.: A rough set-based incremental approach for learning knowledge in dynamic incomplete information systems. *Int. J. Approximate Reasoning* **55**(8), 1764–1786 (2014)

150. Liu, D., Li, T., Zhang, J.: Incremental updating approximations in probabilistic rough sets under the variation of attributes. *Knowl. Based Syst.* **73**, 81–96 (2015)
151. Liu, D., Liang, D.: Incremental learning researches on rough set theory: status and future. *Int. J. Rough Sets Data Anal. (IJRSDA)* **1**(1), 99–112 (2014)
152. Liu, J., Hu, Q., Yu, D.: A comparative study on rough set based class imbalance learning. *Knowl. Based Syst.* **21**(8), 753–763 (2008)
153. Liu, J., Hu, Q., Yu, D.: A weighted rough set based method developed for class imbalance learning. *Inf. Sci.* **178**(4), 1235–1256 (2008)
154. Liu, Y., Xu, C., Zhang, Q., Pan, Y.: Rough rule extracting from various conditions: Incremental and approximate approaches for inconsistent data. *Fundamenta Informaticae* **84**(3, 4), 403–427 (2008)
155. Lu, J., Tan, Y.P.: Cost-sensitive subspace analysis and extensions for face recognition. *IEEE Trans. Inf. Forensics Secur.* **8**(3), 510–519 (2013)
156. Luo, C., Li, T., Chen, H., Liu, D.: Incremental approaches for updating approximations in set-valued ordered information systems. *Knowl. Based Syst.* **50**, 218–233 (2013)
157. Luo, C., Li, T., Yi, Z., Fujita, H.: Matrix approach to decision-theoretic rough sets for evolving data. *Knowl. Based Syst.* **99**, 123–134 (2016)
158. Ma, T., Tang, M.: Weighted rough set model. In: *Sixth International Conference on Intelligent Systems Design and Applications*, vol. 1, pp. 481–485. IEEE (2006)
159. Maji, P., Garai, P.: Fuzzy-rough simultaneous attribute selection and feature extraction algorithm. *IEEE Trans. Cybern.* **43**(4), 1166–1177 (2013)
160. Maji, P., Pal, S.K.: RFCM: a hybrid clustering algorithm using rough and fuzzy sets. *Fundamenta Informaticae* **80**(4), 475–496 (2007)
161. Mak, B., Munakata, T.: Rule extraction from expert heuristics: a comparative study of rough sets with neural networks and ID3. *Eur. J. Oper. Res.* **136**(1), 212–229 (2002)
162. Miao, D., Chen, M., Wei, Z., Duan, Q.: A reasonable rough approximation for clustering web users. In: *International Workshop on Web Intelligence Meets Brain Informatics*, pp. 428–442. Springer (2006)
163. Min, F., He, H., Qian, Y., Zhu, W.: Test-cost-sensitive attribute reduction. *Inf. Sci.* **181**(22), 4928–4942 (2011)
164. Min, F., Hu, Q., Zhu, W.: Feature selection with test cost constraint. *Int. J. Approximate Reasoning* **55**(1), 167–179 (2014)
165. Min, F., Liu, Q.: A hierarchical model for test-cost-sensitive decision systems. *Inf. Sci.* **179**(14), 2442–2452 (2009)
166. Min, F., Zhu, W.: Attribute reduction of data with error ranges and test costs. *Inf. Sci.* **211**, 48–67 (2012)
167. Mirkin, B.: Mathematical classification and clustering: from how to what and why. In: *Classification, Data Analysis, and Data Highways*, pp. 172–181. Springer (1998)
168. Mitra, S.: An evolutionary rough partitive clustering. *Pattern Recogn. Lett.* **25**(12), 1439–1449 (2004)
169. Mitra, S., Banka, H.: Application of rough sets in pattern recognition. In: *Transactions on Rough Sets VII*, pp. 151–169. Springer (2007)
170. Mitra, S., Banka, H., Pedrycz, W.: Rough-fuzzy collaborative clustering. *IEEE Trans. Syst. Man, Cybern. Part B (Cybern.)* **36**(4), 795–805 (2006)
171. Mitra, S., Barman, B.: Rough-fuzzy clustering: an application to medical imagery. In: *International Conference on Rough Sets and Knowledge Technology*, pp. 300–307. Springer (2008)
172. Nanda, S., Majumdar, S.: Fuzzy rough sets. *Fuzzy Sets Syst.* **45**(2), 157–160 (1992)
173. Nguyen, H.S.: Discretization problem for rough sets methods. In: *International Conference on Rough Sets and Current Trends in Computing*, pp. 545–552. Springer (1998)
174. Nguyen, H.S.: On efficient handling of continuous attributes in large data bases. *Fundamenta Informaticae* **48**(1), 61–81 (2001)
175. Orłowska, E.: *Dynamic information systems*. Institute of Computer Science, Polish Academy of Sciences (1981)

176. Ozawa, S., Pang, S., Kasabov, N.: Incremental learning of chunk data for online pattern classification systems. *IEEE Trans. Neural Netw.* **19**(6), 1061–1074 (2008)
177. Park, I.K., Choi, G.S.: Rough set approach for clustering categorical data using information-theoretic dependency measure. *Inf. Syst.* **48**, 289–295 (2015)
178. Parmar, D., Wu, T., Blackhurst, J.: MMR: an algorithm for clustering categorical data using rough set theory. *Data Knowl Eng.* **63**(3), 879–893 (2007)
179. Pawlak, Z.: Rough sets. *Int. J. Comput. Inf. Sci.* **11**(5), 341–356 (1982)
180. Pawlak, Z.: Rough sets and intelligent data analysis. *Inf. Sci.* **147**(1), 1–12 (2002)
181. Pawlak, Z., Skowron, A.: Rough sets: some extensions. *Inf. Sci.* **177**(1), 28–40 (2007)
182. Pawlak, Z., Wong, S.K.M., Ziarko, W.: Rough sets: probabilistic versus deterministic approach. *Int. J. Man-Mach. Stud.* **29**(1), 81–95 (1988)
183. Pedrycz, W.: *Granular Computing: An Emerging Paradigm*, vol. 70. Springer Science & Business Media (2001)
184. Peters, G.: Outliers in rough k-means clustering. In: *International Conference on Pattern Recognition and Machine Intelligence*, pp. 702–707. Springer (2005)
185. Peters, G.: Some refinements of rough k-means clustering. *Pattern Recogn.* **39**(8), 1481–1491 (2006)
186. Peters, G.: Rough clustering utilizing the principle of indifference. *Inf. Sci.* **277**, 358–374 (2014)
187. Peters, G.: Is there any need for rough clustering? *Pattern Recogn. Lett.* **53**, 31–37 (2015)
188. Peters, G., Crespo, F., Lingras, P., Weber, R.: Soft clustering-fuzzy and rough approaches and their extensions and derivatives. *Int. J. Approximate Reasoning* **54**(2), 307–322 (2013)
189. Peters, G., Lampart, M., Weber, R.: Evolutionary rough k-medoid clustering. In: *Transactions on Rough Sets VIII*, pp. 289–306. Springer (2008)
190. Peters, G., Weber, R., Nowatzke, R.: Dynamic rough clustering and its applications. *Appl. Soft Comput.* **12**(10), 3193–3207 (2012)
191. Pradeepa, A., Selvadoss ThanamaniLee, A.: Hadoop file system and fundamental concept of mapreduce interior and closure rough set approximations. *Int. J. Adv. Res. Comput. Commun. Eng.* **2** (2013)
192. do Prado, H.A., Engel, P.M., Chaib Filho, H.: Rough clustering: an alternative to find meaningful clusters by using the reducts from a dataset. In: *International Conference on Rough Sets and Current Trends in Computing*, pp. 234–238. Springer (2002)
193. Qian, Y., Wang, Q., Cheng, H., Liang, J., Dang, C.: Fuzzy-rough feature selection accelerator. *Fuzzy Sets Syst.* **258**, 61–78 (2015)
194. Ramentol, E., Caballero, Y., Bello, R., Herrera, F.: Smote-rsb\*: a hybrid preprocessing approach based on oversampling and undersampling for high imbalanced data-sets using smote and rough sets theory. *Knowl. Inf. Syst.* **33**(2), 245–265 (2012)
195. Riza, L.S., Janusz, A., Bergmeir, C., Cornelis, C., Herrera, F., Šle, D., Benítez, J.M., et al.: Implementing algorithms of rough set theory and fuzzy rough set theory in the R package "roughsets". *Inf. Sci.* **287**, 68–89 (2014)
196. Salamó, M., López-Sánchez, M.: Rough set based approaches to feature selection for case-based reasoning classifiers. *Pattern Recogn. Lett.* **32**(2), 280–292 (2011)
197. Salido, J.F., Murakami, S.: Rough set analysis of a general type of fuzzy data using transitive aggregations of fuzzy similarity relations. *Fuzzy Sets Syst.* **139**(3), 635–660 (2003)
198. Schapire, R.E., Singer, Y.: Boostexter: a boosting-based system for text categorization. *Mach. Learn.* **39**(2–3), 135–168 (2000)
199. Shan, N., Ziarko, W.: Data-based acquisition and incremental modification of classification rules. *Comput. Intell.* **11**(2), 357–370 (1995)
200. Shen, F., Yu, H., Kamiya, Y., Hasegawa, O.: An online incremental semi-supervised learning method. *JACIII* **14**(6), 593–605 (2010)
201. Shen, Q., Chouchoulas, A.: Combining rough sets and data-driven fuzzy learning for generation of classification rules. *Pattern Recogn.* **32**(12), 2073–2076 (1999)
202. Shen, Q., Chouchoulas, A.: A modular approach to generating fuzzy rules with reduced attributes for the monitoring of complex systems. *Eng. Appl. Artif. Intell.* **13**(3), 263–278 (2000)

203. Shen, Q., Jensen, R.: Selecting informative features with fuzzy-rough sets and its application for complex systems monitoring. *Pattern Recogn.* **37**(7), 1351–1363 (2004)
204. Shu, W., Shen, H.: Incremental feature selection based on rough set in dynamic incomplete data. *Pattern Recogn.* **47**(12), 3890–3906 (2014)
205. Singh, G.K., Minz, S.: Discretization using clustering and rough set theory. In: *International Conference on Computing: Theory and Applications, 2007. ICCTA'07*, pp. 330–336. IEEE (2007)
206. Skowron, A., Rauszer, C.: The discernibility matrices and functions in information systems. In: *Intelligent Decision Support*, pp. 331–362. Springer (1992)
207. Skowron, A., Stepaniuk, J.: Tolerance approximation spaces. *Fundamenta Informaticae* **27**(2, 3), 245–253 (1996)
208. Slezak, D.: Approximate bayesian networks. In: *Technologies for Constructing Intelligent Systems 2*, pp. 313–325. Springer (2002)
209. Ślezak, D.: Approximate entropy reducts. *Fundamenta Informaticae* **53**(3–4), 365–390 (2002)
210. Slezak, D., Ziarko, W., et al.: The investigation of the bayesian rough set model. *Int. J. Approximate Reasoning* **40**(1), 81–91 (2005)
211. Slimani, T.: Class association rules mining based rough set method. arXiv preprint [arXiv:1509.05437](https://arxiv.org/abs/1509.05437) (2015)
212. Slowinski, R., Vanderpooten, D., et al.: A generalized definition of rough approximations based on similarity. *IEEE Trans. Knowl. Data Eng.* **12**(2), 331–336 (2000)
213. Soni, R., Nanda, R.: Neighborhood clustering of web users with rough k-means. In: *Proceedings of 6th WSEAS International Conference on Circuits, Systems, Electronics, Control & Signal Processing*, pp. 570–574 (2007)
214. Stefanowski, J.: The rough set based rule induction technique for classification problems. In: *In Proceedings of 6th European Conference on Intelligent Techniques and Soft Computing EUFIT*, vol. 98 (1998)
215. Stefanowski, J.: On combined classifiers, rule induction and rough sets. In: *Transactions on Rough Sets VI*, pp. 329–350. Springer (2007)
216. Stefanowski, J., Vanderpooten, D.: Induction of decision rules in classification and discovery-oriented perspectives. *Int. J. Intell. Syst.* **16**(1), 13–27 (2001)
217. Stefanowski, J., Wilk, S.: Rough sets for handling imbalanced data: combining filtering and rule-based classifiers. *Fundamenta Informaticae* **72**(1–3), 379–391 (2006)
218. Stefanowski, J., Wilk, S.: Extending rule-based classifiers to improve recognition of imbalanced classes. In: *Advances in Data Management*, pp. 131–154. Springer (2009)
219. Su, C.T., Hsu, J.H.: An extended Chi2 algorithm for discretization of real value attributes. *IEEE Trans. Knowl. Data Eng.* **17**(3), 437–441 (2005)
220. Su, C.T., Hsu, J.H.: Precision parameter in the variable precision rough sets model: an application. *Omega* **34**(2), 149–157 (2006)
221. Susmaga, R.: Reducts and constructs in classic and dominance-based rough sets approach. *Inf. Sci.* **271**, 45–64 (2014)
222. Świniarski, R.W.: Rough sets methods in feature reduction and classification. *Int. J. Appl. Math. Comput. Sci.* **11**(3), 565–582 (2001)
223. Swiniarski, R.W., Skowron, A.: Rough set methods in feature selection and recognition. *Pattern Recogn. Lett.* **24**(6), 833–849 (2003)
224. Tay, F.E., Shen, L.: Economic and financial prediction using rough sets model. *Eur. J. Oper. Res.* **141**(3), 641–659 (2002)
225. Tsang, E.C., Chen, D., Yeung, D.S., Wang, X.Z., Lee, J.W.: Attributes reduction using fuzzy rough sets. *IEEE Trans. Fuzzy Syst.* **16**(5), 1130–1141 (2008)
226. Tsoumakas, G., Katakis, I.: *Multi-label classification: an overview*. Aristotle University of Thessaloniki, Greece, Department of Informatics (2006)
227. Tsoumakas, G., Vlahavas, I.: Random k-labelsets: an ensemble method for multilabel classification. In: *European Conference on Machine Learning*, pp. 406–417. Springer (2007)
228. Tsumoto, S.: Automated extraction of medical expert system rules from clinical databases based on rough set theory. *Inf. Sci.* **112**(1), 67–84 (1998)

229. Tsumoto, S.: Automated extraction of hierarchical decision rules from clinical databases using rough set model. *Expert Syst. Appl.* **24**(2), 189–197 (2003)
230. Tsumoto, S.: Incremental rule induction based on rough set theory. In: *International Symposium on Methodologies for Intelligent Systems*, pp. 70–79. Springer (2011)
231. Vanderpooten, D.: Similarity relation as a basis for rough approximations. *Adv. Mach. Intell. Soft Comput.* **4**, 17–33 (1997)
232. Verbiest, N.: Fuzzy rough and evolutionary approaches to instance selection. Ph.D. thesis, Ghent University (2014)
233. Verbiest, N., Cornelis, C., Herrera, F.: FRPS: a fuzzy rough prototype selection method. *Pattern Recogn.* **46**(10), 2770–2782 (2013)
234. Vilalta, R., Drissi, Y.: A perspective view and survey of meta-learning. *Artif. Intell. Rev.* **18**(2), 77–95 (2002)
235. Voges, K., Pope, N., Brown, M.: A rough cluster analysis of shopping orientation data. In: *Proceedings Australian and New Zealand Marketing Academy Conference, Adelaide*, pp. 1625–1631 (2003)
236. Voges, K.E., Pope, N., Brown, M.R.: Cluster analysis of marketing data examining on-line shopping orientation: a comparison of k-means and rough clustering approaches. In: *Heuristics and Optimization for Knowledge Discovery*, pp. 207–224 (2002)
237. Wang, F., Liang, J., Dang, C.: Attribute reduction for dynamic data sets. *Applied Soft Computing* **13**(1), 676–689 (2013)
238. Wang, F., Liang, J., Qian, Y.: Attribute reduction: a dimension incremental strategy. *Knowl. Based Syst.* **39**, 95–108 (2013)
239. Wang, G., Yu, H., Li, T., et al.: Decision region distribution preservation reduction in decision-theoretic rough set model. *Inf. Sci.* **278**, 614–640 (2014)
240. Wang, X., An, S., Shi, H., Hu, Q.: Fuzzy rough decision trees for multi-label classification. In: *Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing*, pp. 207–217. Springer (2015)
241. Wang, X., Yang, J., Peng, N., Teng, X.: Finding minimal rough set reducts with particle swarm optimization. In: *International Workshop on Rough Sets, Fuzzy Sets, Data Mining, and Granular-Soft Computing*, pp. 451–460. Springer (2005)
242. Wang, X., Yang, J., Teng, X., Xia, W., Jensen, R.: Feature selection based on rough sets and particle swarm optimization. *Pattern Recogn. Lett.* **28**(4), 459–471 (2007)
243. Wei, M.H., Cheng, C.H., Huang, C.S., Chiang, P.C.: Discovering medical quality of total hip arthroplasty by rough set classifier with imbalanced class. *Qual. Quant.* **47**(3), 1761–1779 (2013)
244. Wojna, A.: Constraint based incremental learning of classification rules. In: *International Conference on Rough Sets and Current Trends in Computing*, pp. 428–435. Springer (2000)
245. Wróblewski, J.: Finding minimal reducts using genetic algorithms. In: *Proceedings of the Second Annual Joint Conference on Information Science*, pp. 186–189 (1995)
246. Wróblewski, J.: Theoretical foundations of order-based genetic algorithms. *Fundamenta Informaticae* **28**(3, 4), 423–430 (1996)
247. Wróblewski, J.: Ensembles of classifiers based on approximate reducts. *Fundamenta Informaticae* **47**(3–4), 351–360 (2001)
248. Wu, Q., Bell, D.: Multi-knowledge extraction and application. In: *International Workshop on Rough Sets, Fuzzy Sets, Data Mining, and Granular-Soft Computing*, pp. 274–278. Springer (2003)
249. Xie, H., Cheng, H.Z., Niu, D.X.: Discretization of continuous attributes in rough set theory based on information entropy. *Chin. J. Comput. Chin. Ed.* **28**(9), 1570 (2005)
250. Xu, Y., Wang, L., Zhang, R.: A dynamic attribute reduction algorithm based on 0–1 integer programming. *Knowl. Based Syst.* **24**(8), 1341–1347 (2011)
251. Xu, Z., Liang, J., Dang, C., Chin, K.: Inclusion degree: a perspective on measures for rough set data analysis. *Inf. Sci.* **141**(3), 227–236 (2002)
252. Yang, Q., Ling, C., Chai, X., Pan, R.: Test-cost sensitive classification on data with missing values. *IEEE Trans. Knowl. Data Eng.* **18**(5), 626–638 (2006)

253. Yang, X., Qi, Y., Song, X., Yang, J.: Test cost sensitive multigranulation rough set: model and minimal cost selection. *Inf. Sci.* **250**, 184–199 (2013)
254. Yang, X., Qi, Y., Yu, H., Song, X., Yang, J.: Updating multigranulation rough approximations with increasing of granular structures. *Knowl. Based Syst.* **64**, 59–69 (2014)
255. Yang, Y., Chen, D., Dong, Z.: Novel algorithms of attribute reduction with variable precision rough set model. *Neurocomputing* **139**, 336–344 (2014)
256. Yang, Y., Chen, Z., Liang, Z., Wang, G.: Attribute reduction for massive data based on rough set theory and mapreduce. In: *International Conference on Rough Sets and Knowledge Technology*, pp. 672–678. Springer (2010)
257. Yao, J., Yao, Y.: A granular computing approach to machine learning. *FSKD* **2**, 732–736 (2002)
258. Yao, Y.: Combination of rough and fuzzy sets based on  $\alpha$ -level sets. In: *Rough sets and Data Mining*, pp. 301–321. Springer (1997)
259. Yao, Y.: Decision-theoretic rough set models. In: *International Conference on Rough Sets and Knowledge Technology*, pp. 1–12. Springer (2007)
260. Yao, Y.: Three-way decision: an interpretation of rules in rough set theory. In: *International Conference on Rough Sets and Knowledge Technology*, pp. 642–649. Springer (2009)
261. Yao, Y.: Three-way decisions with probabilistic rough sets. *Inf. Sci.* **180**(3), 341–353 (2010)
262. Yao, Y.: The superiority of three-way decisions in probabilistic rough set models. *Inf. Sci.* **181**(6), 1080–1096 (2011)
263. Yao, Y.: An outline of a theory of three-way decisions. In: *International Conference on Rough Sets and Current Trends in Computing*, pp. 1–17. Springer (2012)
264. Yao, Y., Greco, S., Słowiński, R.: Probabilistic rough sets. In: *Springer Handbook of Computational Intelligence*, pp. 387–411. Springer (2015)
265. Yao, Y., Zhao, Y.: Attribute reduction in decision-theoretic rough set models. *Inf. Sci.* **178**(17), 3356–3373 (2008)
266. Yao, Y., Zhao, Y., Maguire, R.B.: Explanation oriented association mining using rough set theory. In: *International Workshop on Rough Sets, Fuzzy Sets, Data Mining, and Granular-Soft Computing*, pp. 165–172. Springer (2003)
267. Yao, Y., Zhou, B.: Two bayesian approaches to rough sets. *Eur. J. Oper. Res.* **251**(3), 904–917 (2016)
268. Ye, D., Chen, Z., Ma, S.: A novel and better fitness evaluation for rough set based minimum attribute reduction problem. *Inf. Sci.* **222**, 413–423 (2013)
269. Yong, L., Congfu, X., Yunhe, P.: An incremental rule extracting algorithm based on pawlak reduction. In: *2004 IEEE International Conference on Systems, Man and Cybernetics*, vol. 6, pp. 5964–5968. IEEE (2004)
270. Yong, L., Wenliang, H., Yunliang, J., Zhiyong, Z.: Quick attribute reduct algorithm for neighborhood rough set model. *Inf. Sci.* **271**, 65–81 (2014)
271. Yu, H., Chu, S., Yang, D.: Autonomous knowledge-oriented clustering using decision-theoretic rough set theory. *Fundamenta Informaticae* **115**(2–3), 141–156 (2012)
272. Yu, H., Liu, Z., Wang, G.: An automatic method to determine the number of clusters using decision-theoretic rough set. *Int. J. Approximate Reasoning* **55**(1), 101–115 (2014)
273. Yu, H., Su, T., Zeng, X.: A three-way decisions clustering algorithm for incomplete data. In: *International Conference on Rough Sets and Knowledge Technology*, pp. 765–776. Springer (2014)
274. Yu, H., Wang, G., Lan, F.: Solving the attribute reduction problem with ant colony optimization. In: *Transactions on Rough Sets XIII*, pp. 240–259. Springer (2011)
275. Yu, H., Wang, Y.: Three-way decisions method for overlapping clustering. In: *International Conference on Rough Sets and Current Trends in Computing*, pp. 277–286. Springer (2012)
276. Yu, H., Wang, Y., Jiao, P.: A three-way decisions approach to density-based overlapping clustering. In: *Transactions on Rough Sets XVIII*, pp. 92–109. Springer (2014)
277. Yu, H., Zhang, C., Hu, F.: An incremental clustering approach based on three-way decisions. In: *International Conference on Rough Sets and Current Trends in Computing*, pp. 152–159. Springer (2014)

278. Yu, Y., Miao, D., Zhang, Z., Wang, L.: Multi-label classification using rough sets. In: International Workshop on Rough Sets, Fuzzy Sets, Data Mining, and Granular-Soft Computing, pp. 119–126. Springer (2013)
279. Yu, Y., Pedrycz, W., Miao, D.: Multi-label classification by exploiting label correlations. *Expert Syst. Appl.* **41**(6), 2989–3004 (2014)
280. Zhai, J., Zhang, S., Zhang, Y.: An extension of rough fuzzy set. *J. Intell. Fuzzy Syst.* (Preprint), 1–10 (2016)
281. Zhai, J., Zhang, Y., Zhu, H.: Three-way decisions model based on tolerance rough fuzzy set. *Int. J. Mach. Learn. Cybern.* 1–9 (2016)
282. Zhang, H.R., Min, F.: Three-way recommender systems based on random forests. *Knowl. Based Syst.* **91**, 275–286 (2016)
283. Zhang, J., Li, T., Chen, H.: Composite rough sets. In: International Conference on Artificial Intelligence and Computational Intelligence, pp. 150–159. Springer (2012)
284. Zhang, J., Li, T., Chen, H.: Composite rough sets for dynamic data mining. *Inf. Sci.* **257**, 81–100 (2014)
285. Zhang, J., Li, T., Pan, Y.: Parallel rough set based knowledge acquisition using mapreduce from big data. In: Proceedings of the 1st International Workshop on Big Data, Streams and Heterogeneous Source Mining: Algorithms, Systems, Programming Models and Applications, pp. 20–27. ACM (2012)
286. Zhang, J., Li, T., Ruan, D., Gao, Z., Zhao, C.: A parallel method for computing rough set approximations. *Inf. Sci.* **194**, 209–223 (2012)
287. Zhang, J., Li, T., Ruan, D., Liu, D.: Rough sets based matrix approaches with dynamic attribute variation in set-valued information systems. *Int. J. Approximate Reasoning* **53**(4), 620–635 (2012)
288. Zhang, L., Hu, Q., Duan, J., Wang, X.: Multi-label feature selection with fuzzy rough sets. In: International Conference on Rough Sets and Knowledge Technology, pp. 121–128. Springer (2014)
289. Zhang, L., Li, H., Zhou, X., Huang, B., Shang, L.: Cost-sensitive sequential three-way decision for face recognition. In: International Conference on Rough Sets and Intelligent Systems Paradigms, pp. 375–383. Springer (2014)
290. Zhang, M.L., Zhou, Z.H.: MI-knn: A lazy learning approach to multi-label learning. *Pattern Recogn.* **40**(7), 2038–2048 (2007)
291. Zhang, T., Chen, L., Ma, F.: An improved algorithm of rough k-means clustering based on variable weighted distance measure. *Int. J. Database Theory Appl.* **7**(6), 163–174 (2014)
292. Zhang, T., Chen, L., Ma, F.: A modified rough c-means clustering algorithm based on hybrid imbalanced measure of distance and density. *Int. J. Approximate Reasoning* **55**(8), 1805–1818 (2014)
293. Zhang, X., Miao, D.: Three-way weighted entropies and three-way attribute reduction. In: International Conference on Rough Sets and Knowledge Technology, pp. 707–719. Springer (2014)
294. Zhang, Y., Zhou, Z.H.: Cost-sensitive face recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **32**(10), 1758–1769 (2010)
295. Zhao, H., Min, F., Zhu, W.: Test-cost-sensitive attribute reduction based on neighborhood rough set. In: 2011 IEEE International Conference on Granular Computing (GrC), pp. 802–806. IEEE (2011)
296. Zhao, H., Wang, P., Hu, Q.: Cost-sensitive feature selection based on adaptive neighborhood granularity with multi-level confidence. *Inf. Sci.* **366**, 134–149 (2016)
297. Zhao, M., Luo, K., Liao, X.X.: Rough set attribute reduction algorithm based on immune genetic algorithm. *Jisuanji Gongcheng yu Yingyong (Comput. Eng. Appl.)* **42**(23), 171–173 (2007)
298. Zhao, S., Chen, H., Li, C., Du, X., Sun, H.: A novel approach to building a robust fuzzy rough classifier. *IEEE Trans. Fuzzy Syst.* **23**(4), 769–786 (2015)
299. Zhao, S., Tsang, E.C., Chen, D.: The model of fuzzy variable precision rough sets. *IEEE Trans. Fuzzy Syst.* **17**(2), 451–467 (2009)

300. Zhao, S., Tsang, E.C., Chen, D., Wang, X.: Building a rule-based classifier—a fuzzy-rough set approach. *IEEE Trans. Knowl. Data Eng.* **22**(5), 624–638 (2010)
301. Zheng, Z., Wang, G., Wu, Y.: A rough set and rule tree based incremental knowledge acquisition algorithm. In: *International Workshop on Rough Sets, Fuzzy Sets, Data Mining, and Granular-Soft Computing*, pp. 122–129. Springer (2003)
302. Zhong, N., Dong, J., Ohsuga, S.: Using rough sets with heuristics for feature selection. *J. Intell. Inf. Syst.* **16**(3), 199–214 (2001)
303. Zhou, Z.H.: Cost-sensitive learning. In: *International Conference on Modeling Decisions for Artificial Intelligence*, pp. 17–18. Springer (2011)
304. Zhou, Z.H., Liu, X.Y.: Training cost-sensitive neural networks with methods addressing the class imbalance problem. *IEEE Trans. Knowl. Data Eng.* **18**(1), 63–77 (2006)
305. Zhu, W.: Generalized rough sets based on relations. *Inf. Sci.* **177**(22), 4997–5011 (2007)
306. Zhu, W.: Topological approaches to covering rough sets. *Inf. Sci.* **177**(6), 1499–1508 (2007)
307. Ziarko, W.: Variable precision rough set model. *J. Comput. Syst. Sci.* **46**(1), 39–59 (1993)
308. Zou, W., Li, T., Chen, H., Ji, X.: Approaches for incrementally updating approximations based on set-valued information systems while attribute values’ coarsening and refining. In: *2009 IEEE International Conference on Granular Computing* (2009)