Chapter 2
Computational Intelligence Techniques

The true sign of intelligence is not knowledge but imagination

Albert Einstein

Abstract The chapter explains the computational intelligence techniques utilized in the algorithms presented in the book. The fuzzy and rough sets, fuzzy-rough sets, genetic algorithm and, feature selection and classification using the fuzzy-rough sets are detailed. The biologically inspired feature extraction system utilized in the presented algorithms is explained.

Keywords Fuzzy sets · Rough sets · Fuzzy-rough sets · Genetic algorithm · Biologically inspired features · Classification

2.1 Fuzzy and Rough Sets

In classical set theory, an element of the universe either belongs or not belongs to a set. The belongingness of the element is crisp: either yes (in the set) or no (not in the set). In fuzzy sets, belongingness can lie between yes and no; for example, a set of *tall* persons. We cannot classify a person as tall into yes / no category, as there exists no well-defined boundary for the set *tall*. Attributes used for short and tall classifications namely height is a fuzzy attribute and the sets short and tall are fuzzy sets. Each fuzzy set is characterized by a membership function $m(x)$ defining the degree of membership of the element x, usually in the range [0, 1] [1].

Fuzzy inference systems have been successfully applied in control, pattern classification, decision analysis, expert systems, manufacturing and computer vision [2]. The concept of fuzzy sets is important in pattern classification, due to its similarity with human reasoning. Traditionally, the rules in a fuzzy inference system are generated from expert knowledge. If no expert knowledge is available, the usual approach is to identify and train fuzzy membership functions in accordance with the data clusters in a training set.

P. K. Pisharady et al., *Computational Intelligence in Multi-Feature Visual Pattern Recognition*, Studies in Computational Intelligence 556, DOI: 10.1007/978-981-287-056-8_2,
© Springer Science+Business Media Singapore 2014

The rough set theory was introduced in the early eighties as a tool to handle inconsistencies among data [3, 4]. A rough set is a formal approximation of a vague concept by a pair of precise concepts, called lower and upper approximations [5]. Rough sets handle uncertainty by computing the lower and upper approximations. Objects belonging to the same category characterized by the same attributes are indistinguishable and indiscernible.

Let $I = (U, A)$ be an information system, where U is the universe of discourse and A is a non-empty finite set of features such that $a : U \rightarrow V_a, \forall a \in A, V_a$ being the set of values for feature a. In most applications, the outcome of classification is known and represented by a special *decision attribute* set. Consider a decision support system, $A = C \cup D$ where C is a set of conditional attributes and D is a set of decision attributes. With any $P \subseteq A$ there associated is an equivalence relation $IND(P)$:

$$IND(P) = (x, y) \in U^2 | \forall a \in P, a(x) = a(y). \tag{2.1}$$

If $(x, y) \in IND(P)$, then x and y are indiscernible with the features from P. The equivalence classes of P-indiscernibility relation are denoted by $[x]_P$. For any $X \subseteq U$, X can be approximated using the information contained only in P by constructing *P-lower* and *P-upper approximations* of X, denoted by $\underline{P}X$ and $\overline{P}X$ respectively,

$$\underline{P}X = x | [x]_P \subseteq X \tag{2.2}$$

$$\overline{P}X = x | [x]_P \cap X \neq \phi \tag{2.3}$$

If P and Q are equivalence relations over U, then the positive region $POS_P(Q)$ is defined as,

$$POS_P(Q) = \bigcup_{x \in U/Q} \underline{P}X. \tag{2.4}$$

For pattern classification purposes, the positive region contains all objects of U that can be classified into classes of U/Q using the knowledge in attribute set P.

Several measures of uncertainty exist in rough set theory [6]. The most widely used is the quality of the lower and upper approximations. For a given set of data samples X, not necessarily definable by a set A of attributes, the quality of lower approximation is the ratio of the number of all elements in the lower approximation of X to the total number of samples. It is interpreted as the ratio of the number of all *certainly classified* samples by attributes from A as being in X to the number of all samples. Similarly, the quality of upper approximation is the ratio of the number of all elements in the upper approximation of X to the total number of samples. It is taken as the ratio of the number of all *possibly classified* samples by attributes from A as being in X to the total number of samples.

2.1.1 Fuzzy-Rough Sets

The concept of crisp equivalence class is the basis for rough set theory. A crisp equivalence class contains samples from different output classes. In addition, the various elements in an equivalent class may have different degrees of *belonging-ness* to different output classes. A combination of fuzzy and rough sets, namely fuzzy-rough sets [7, 8], is useful for decision making in such situations where both vagueness and indiscernibility are present. Fuzzy-rough set is a deviation of rough set theory in which the concept of crisp equivalence class is extended with the fuzzy set theory to form fuzzy equivalence class [8]. A fuzzy similarity relation replaces an equivalence relation in rough sets to form the fuzzy-rough sets. In fuzzy-rough sets the equivalence class is fuzzy in addition to the fuzziness of the output classes [9].

Let the equivalence classes be in the form of fuzzy clusters F_1, F_2, ..., F_H, which are generated by the fuzzy partitioning of the input set X into H number of clusters. Each fuzzy cluster represents an equivalence class containing patterns from different output classes. The definite and possible members of the output class are identified using the lower and upper approximations [3] of the fuzzy equivalence classes. The description of a fuzzy set C_c (output class) by means of the fuzzy partitions in the form of lower and upper approximations $\underline{C_c}$ and $\overline{C_c}$ is as follows [9] :

$$\begin{aligned} \mu_{\underline{C_c}}(F_j) &= \inf_x\{\max(1-\mu_{F_j}(x), \mu_{C_c}(x))\} \quad \forall x \in X \\ \mu_{\overline{C_c}}(F_j) &= \sup_x\{\min(\mu_{F_j}(x), \mu_{C_c}(x))\} \quad \forall x \in X \end{aligned} \tag{2.5}$$

The tuple $\langle \overline{C_c}, \underline{C_c} \rangle$ is a fuzzy-rough set. $\mu_{F_j}(x)$ and $\mu_{C_c}(x)$ are fuzzy memberships of sample x in the fuzzy equivalence class F_j and output class C_c respectively.

2.1.2 Feature Selection and Classification Using Fuzzy-Rough Sets

The concept of fuzzy discretization of feature space for a rough set theoretic classifier is provided in [10]. The merit of fuzzy discretization over crisp discretization in terms of classification accuracy is demonstrated when overlapping datasets are used. A novel fuzzification technique namely Modified Minimization Entropy Principle Algorithm (MMEPA) and an entropy-based fuzzy-rough approach for extracting classification rules are reported in [11].

The fuzzy-rough uncertainty is exploited to improve the classification efficiency of a conventional K-nearest neighbor (K-NN) classifier in [9]. The algorithm generalizes the conventional and fuzzy K-NN classifier algorithms. Another modification of the K-NN algorithm using fuzzy-rough sets is proposed in [12] for hand gesture recognition. Fuzzy-rough concept removes the training samples in the class boundary and overlapping regions, improving classification accuracy. The algorithm is applied to only one type of problem, the hand gesture recognition, whereas [9] applied the algorithm to different problems; speech classification, image based letter

classification, and medical diagnosis (head injury classification). [13] presented a fuzzy-rough nearest neighbour (FRNN) classification algorithm, as an alternative to the fuzzy-rough ownership function (FRNN-O) approach reported in [9]. In contrast to [9], the algorithm proposed in [13] utilizes nearest neighbors to construct lower and upper approximations of decision classes. The algorithm classifies test instances based on memberships in the lower and upper approximations. FRNN outperformed both FRNN-O and traditional fuzzy nearest neighbour (FNN) algorithms.

A novel concept namely *consistence degree* is proposed in [14] to use as a critical value to reduce redundant attributes in database. A rule based classifier using a generalized fuzzy-rough set model is reported. The classifier is effective on noisy data. A comparison between fuzzy-rough classifier and neural network classifier is provided in [15]. The fuzzy-rough classifier is reported as a better choice with lesser training time, lesser dependence on training data, and for being transparent.

A feature selection method with fuzzy-rough approach and ant colony optimization is provided in [16]. Shen et al. [17] proposed a classifier that integrates a fuzzy rule induction algorithm with a rough set assisted feature reduction method. The classifier is tested on two problems, the urban water treatment plant problem and algae population estimation. Fuzzy-rough approach is utilized in [18] for decision table reduction. Unlike other feature selection methods, this method reduces the decision table in both vertical and horizontal directions (both the number of features and feature dimensionality are reduced).

A robust feature evaluation and selection algorithm, using a different model of fuzzy-rough sets namely soft fuzzy-rough sets, is provided in [19]. This method is more effective in dealing with noisy data. [20] proposed a fuzzy-rough feature selection algorithm, with application to microarray based cancer classification. These works utilized standard classifiers (KNN, C 5.0) for classification.

2.1.3 Genetic Algorithm

Genetic Algorithms (GAs) are random search algorithms inspired by natural genetics to find optimum solutions to problems [21, 22]. The basic concept is to maintain a population of chromosomes representing candidate solutions. The population evolves over time through a process of competition and controlled variation. Each chromosome in the population has an associated fitness. Based on the fitness values, new chromosomes are generated using genetic operators such as *crossover* and *mutation*. The outline of a basic GA is given below:

1. Create an initial population;
2. Evaluate fitness of each chromosome in the population;
3. Based on fitness, select chromosomes for reproduction;
4. Apply genetic operations, *crossover* and *mutation* on selected chromosomes, to create new chromosomes;
5. Replace a part of the current population with newly generated chromosomes; and,
6. Terminate GA, if stopping criterion is satisfied, else return to step 2.

GAs have long been utilized in fuzzy inference systems for generating fuzzy rules and training membership functions [23–25]. There are there different approaches to utilize GA in learning, the Michigan, Pittsburgh and the Iterative Rule Learning (IRL) approaches. In Michigan approach, each chromosome corresponds to a classifier rule whereas in the Pittsburgh approach each chromosome encodes a complete set of classifier rules. In the IRL approach, each chromosome represents only one rule, but contrary to Michigan, only the best individual is taken as the solution while discarding the rest of the chromosomes (rules). Since a single rule provides only partial solution, GA is placed in an iterative scheme for generating a set of rules. In the iterative scheme, either the selected rules or the classified samples are penalized, by way of reduced weights, ensuring the search for new rules is focused on unclassified samples. A major issue in Michigan approach is the possible conflict between the objectives of collective rule set (the classifier) and individual rules. In Pittsburgh approach competition occurs among complete rule sets rather than among individual rules, avoiding the conflict. Pittsburgh approach is computationally more intensive due to maintenance and evaluation of complete rule sets. IRL approach is advantageous for being computationally less intensive (seeking only a single rule in each iteration), generating a cohesive classifier through a penalizing mechanism and avoiding conflict between the classifier rule set and individual rules [26].

2.2 Computational Model of Visual Cortex

Extracting features for visual pattern recognition is an ongoing research topic in computer vision. Over the past decades, there are several successful attempts in vision based pattern analysis. The mainstream computer vision has always been challenged by human vision, and the mechanism of human visual system is yet to understand, which is a challenge in both neuroscience and computer vision. The human visual system rapidly and effortlessly recognizes large number of diverse objects in cluttered natural scenes and identifies specific patterns which has inspired the development of computational models of biological vision systems. Recent developments in the use of neurobiological models in computer vision tries to bridge the gaps among neuroscience, computer vision and pattern recognition.

Hubel and Wiesel discovered the organization of receptive fields, and the properties of *simple* and *complex* cells in cat's primary visual cortex [27]. The cortical simple cell receptive fields are modeled [28] using a Gabor function (also known as Gabor wavelet or Gabor filter) described by 2.6.

$$F(x, y) = \exp\left(-\frac{({x_0}^2 + \gamma^2 y_0^2)}{2\sigma^2}\right) \times \cos\left(\frac{2\pi}{\lambda} x_0\right), \tag{2.6}$$

$$x_0 = x\cos\theta + y\sin\theta \text{ and } y_0 = -x\sin\theta + y\cos\theta. \tag{2.7}$$

where,

γ is the spatial aspect ratio of Gaussian function,
σ is the standard deviation of Gaussian function,
λ is the wavelength of the sinusoidal term,
θ is the orientation of Gaussian from x-axis.

Gabor wavelet based features have good discriminative power among different textures and shapes in images. Gabor filters resemble the receptive fields of neurons in the primary visual cortex of mammals [28]. Use of 2D Gabor wavelet representation in computer vision was pioneered by Daugman [29]. Riesenhuber and Poggio extended the approach and proposed a hierarchical model of ventral visual object-processing stream in the visual cortex [30]. Serre et al. implemented a computational model of the system and used it for robust object recognition [31, 32]. The features extracted by the model are known as the C_1 and C_2 standard model features (SMFs). The C_2 SMFs are used for hand writing recognition [33] and face recognition [34]. These features are scale and position invariant, and the feature extraction algorithm does not require image segmentation. The number of extracted features is independent of the input image size. Serre et al. used C_2 features for robust object recognition [31, 32]. The algorithms reported in this book utilize the C_2 features for multi-class recognition of hand postures and human faces [35–40].

2.2.1 Biologically Inspired Feature Extraction System

The computational model proposed by Serre et al. consists of four layers (Table 2.1). Layer 1 (S_1) consists of a battery of Gabor filters with different orientations (4) and sizes (16 sizes divided into 8 bands). This imitates the simple cells in the primary visual cortex (V1) which filters the image for the detection of edges and bars. Layer 2 (C_1) models the complex cells in V1, by applying a MAX operator locally (over different scales and positions) to the first layer results. This operation provides tolerance to different object projection sizes, and, to position and rotation variations in a 2-D plane of visual field. In layer 3 (S_2), radial basis functions (RBFs) are used to imitate the quaternary visual area (V4) and posterior inferotemporal (PIT) cortex. This aids shape recognition by comparing the complex features at the output of C_1 stage (corresponding to retinal image) with patches of previously seen visual images and shape features.[1] Finally, the fourth layer (C_2) applies a MAX operator (globally, over all scales and positions) to the output of layer S_2, resulting in a representation expressing the best comparison with previously seen images. The output of layer 4 are the C_2 SMFs, which are used for image classification.

Simple cells in the third layer implement an RBF combining bars and edges in image to more complex shapes. RBFs are a major class of neural network model, comparing the distance between input and a prototype [41]. Each S_2 unit response

[1] In humans these patterns are stored in synaptic weights of neural cells.

Table 2.1 Different layers in a C_2 feature extraction system

Layer	Process	Represents
S_1	Gabor filtering	Simple cells in V1
C_1	Local pooling	Complex cells in V1
S_2	Radial basis functions	V4 & posterior inferotemporal cortex
C_2	Global pooling	Inferotemporal cortex

depends on the Euclidean distance between a new input and a stored prototype in a Gaussian-like manner. The prototype patches of different sizes (centers of the RBF units) are drawn randomly (random image and position) from the training images at the second layer (C_1). Each patch contains four orientations. The third layer compares these patches by calculating the summed Euclidean distance between the patch and every possible crops (combining all orientations) from the image of similar size. This comparison is done separately with each scale-band representation in second layer.

The final set of shift and scale invariant C_2 response is computed by taking a global maximum over all scales and positions for each S_2 type, i.e., the value of the best match between a stored prototype and input image is kept and the rest is discarded. Each C_2 feature corresponds to a specific prototype patch with a specific patch size in layer 3. The more the number of extracted features the better is the classification accuracy. When more number of features are extracted, the computational burden (both for feature extraction and classification) will increase.

The feature selection and classification algorithms presented in Chaps. 4 and 5 extract the image features using the above feature extraction system. The focus of the chapters is on feature selection and classification aspects. The algorithms presented in Chaps. 7 and 8 modify the feature extraction system for hand posture recognition application.

References

1. A. Lotfi, Zadeh. Fuzzy Sets Inf. Control **8**(3), 338–353 (1965)
2. H.J. Zimmermann, *Fuzzy Set Theory and Its Applications* (Kluwer Academic Publishers, Boston, 1991)
3. Z. Pawlak, Rough sets. Int. J. Comput. Inform. Sci. **11**, 341–356 (1982)
4. Z. Pawlak, Rough classification. Int. J. Man Mach. Stud. **20**, 469–483 (1984)
5. Z. Pawlak, *Rough Sets: Theoretical Aspects of Reasoning about Data* (Kluwer Academic Publishers, Dordrecht, 1991)
6. Z. Pawlak, J. Grzymala-Busse, R. Slowinski, W. Ziarko, Rough sets. Commun. ACM **38**(11), 89–94 (1995)
7. D. Dubois, H. Prade, Rough fuzzy sets and fuzzy rough sets. Int. J. Gen. Syst. **17**, 191–209 (1990)
8. D. Dubois, H. Prade, in *Putting Rough Sets and Fuzzy Sets Together*, Theory ed. by R. Slowinski Intelligent Decision Support: Handbook of Applications and Advances in Rough Sets , Series D: System Theory, Knowledge Engineering and Problem Solving, vol. 11 (Kluwer Academic Publishers, Dordrecht, The Netherlands, 1992), pp. 203–232

9. M. Sarkar, Fuzzy-rough nearest neighbor algorithms in classification. Fuzzy Sets Syst. **158**, 2134–2152 (2007)
10. A. Roy and, K. P. Sankar, Fuzzy discretization of feature space for a rough set classifier. Pattern Recogn. Lett. **24**, 895–902 (2003)
11. Y.C. Tsai, C.H. Cheng, J.R. Chang, Entropy-based fuzzy rough classification approach for extracting classification rules. Expert Syst. Appl. **31**(2), 436–443 (2006)
12. X. Wang, J. Yang, X. Teng, N. Peng, Fuzzy-rough set based nearest neighbor clustering classification algorithm. Lect. Notes Comput. Sci. **3613**(2005), 370–373 (2005)
13. R. Jensen and C. Cornelis, *A new approach to fuzzy-rough nearest neighbour classification*, Proceedings of the 6th International conference on Rough sets and current trends in computing, (2008), pp. 310–319
14. S. Zhao, E.C.C. Tsang, D. Chen, X. Wang, Building a rule-based classifier-a fuzzy-rough set approach. IEEE Trans. Knowl. Data Eng. **22**(5), 624–638 (2010)
15. M. Juneja, E. Walia, P.S. Sandhu, and R Mohana, *Implementation and comparative analysis of rough set, artificial neural network (ann) and fuzzy-rough classifiers for satellite image classification*, International Conference on Intelligent Agent & Multi-Agent Systems, 2009. IAMA 2009, 2009, pp. 1–6
16. R. Jensen, Q. Shen, Fuzzy-rough data reduction with ant colony optimization. Fuzzy Sets Syst. **149**(1), 5–20 (2005)
17. Qiang Shen, Alexios Chouchoulas, A rough-fuzzy approach for generating classification rules. Pattern Recogn. **35**, 2425–2438 (2002)
18. E.C.C. Tsang, S. Zhao, Decision table reduction in kdd: fuzzy rough based approach. Trans. Rough Sets Lect. Notes Comput. Sci. **5946**, 177–188 (2010)
19. H. Qinghua, A. Shuang, Y. Daren, Soft fuzzy rough sets for robust feature evaluation and selection. Inf. Sci. **180**(22), 4384–4400 (2010)
20. F.F. Xu, D.Q. Miao, and L. Wei, Fuzzy-rough attribute reduction via mutual information with an application to cancer classification. Comput. & Math. Appl. **57**(6), 1010–1017 (2009)
21. D.E. Goldberg, *Algorithms in Search, Optimization, and Machine Learning* (Addison-Wesley, Reading, MA, 1989)
22. J.H. Holland, *Adaptation in Natural and Artificial Systems* (University of Michigan Press, Ann Arbor, 1975)
23. H. Ishibuchi, K. Nozaki, N. Yamamoto, H. Tanaka, Selecting fuzzy if-then rules for classification problems using genetic algorithms. IEEE Trans. Fuzzy Syst. **3**(2), 260–270 (1995)
24. T. Nakashima, G. Nakai, H. Ishibuchi, Improving the performance of fuzzy classifier systems by membership function learning and feature selection, in *Proceedings of IEEE International Conference on Fuzzy Systems* (Honolulu, USA, 2002), pp. 488–493
25. Y. Shi, R. Eberhart, and Y. Chen, implementation of evolutionary fuzzy systems. IEEE Trans. Fuzzy Syst. **7**(2), 109–119 (1999)
26. P.K. Pisharady, P. Vadakkepat, S. Ganesan, A.P. Loh, Boosting based fuzzy-rough pattern classifier, Trends in Intelligent Robotics, in *Proceedings of the 15th Robot World Cup and Congress, FIRA 2010*, Bangalore, 15–19 Sept 2010, vol. 103, pp. 306–313 (2010)
27. T.N. Wiesel, D.H. Hubel, Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. J. Physiol. **160**, 106–154 (1962)
28. J.P. Jones, L.A. Palmer, An evaluation of the twodimensional gabor filter model of simple receptive fields in cat striate cortex. J. Neurophysiol. **58**(6), 1233–1258 (1987)
29. J.G. Daugman, Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. J. Optical Soc. Am. A **2**(7), 1160–1169 (1985)
30. M. Riesenhuber, T. Poggio, Hierarchical models of object recognition in cortex. Nat. Neurosci. **2**(11), 1019–1025 (1999)
31. T. Serre, L. Wolf, S. Bileschi, M. Riesenhuber, T. Poggio, Robust object recognition with cortex-like mechanisms. IEEE Trans. Pattern Anal. Mach. Intell. **29**(3), 411–426 (2007)
32. T. Serre, L. Wolf, and T. Poggio, in *Object recognition with features inspired by visual cortex*, eds. by C. Schmid, S. Soatto, C. Tomasi Conference on Computer Vision and Pattern Recognition (San Diego, CA, 2005), pp. 994–1000

33. T. van der Zant, L. Schomaker, K. Haak, Handwritten-word spotting using biologically inspired features. IEEE Trans. Pattern Anal. Mach. Intell. **30**(11), 1945–1957 (2008)
34. J. Lai and W.X. Wang, in *Face recognition using cortex mechanism and svm*, eds. by C. Xiong, H. Liu, Y. Huang, Y. Xiong. 1st International Conference Intelligent Robotics and Applications (Wuhan, China, 2008), pp. 625–632
35. P.K. Pisharady, Computational intelligence techniques in visual pattern recognition. Ph.D. thesis, National University of Singapore (August, 2011).
36. P.K. Pisharady, Q.S.H. Stephanie, P. Vadakkepat, A.P.Loh, Hand posture recognition using neuro-biologically inspired features, in *Trends in Intelligent Robotics: Proceedings of the 15th Robot World Cup and Congress, FIRA 2010, Bangalore, India*, 15-19 Sep 2010, **103**, 290–297 (2010)
37. P.K. Pisharady, P. Vadakkepat, A.P. Loh, Graph matching based hand posture recognition using neuro-biologically inspired features, in *International Conference on Control, Automation, Robotics and Vision (ICARCV) 2010* (Singapore), 2010
38. P.K. Pisharady, P. Vadakkepat, and A.P. Loh, Attention based detection and recognition of hand postures against complex backgrounds. Int. J. Comput. Vision **101**(3), 403–419 (2013)
39. P.K. Pisharady, P. Vadakkepat, A.P. Loh, Fuzzy-rough discriminative feature selection and classification algorithm, with application to microarray and image datasets. Appl. Soft Comput. **11**(4), 3429–3440 (2011)
40. P.K. Pisharady, P. Vadakkepat, A.P. Loh, Hand posture and face recognition using a fuzzy-rough approach. Int. J. Humanoid Rob. **7**(3), 331–356 (2010)
41. C Bishop, *Neural Networks for Pattern Recognition* (Oxford University Press, New York, 1995)



http://www.springer.com/978-981-287-055-1

Computational Intelligence in Multi-Feature Visual Pattern Recognition
Hand Posture and Face Recognition using Biologically Inspired Approaches
Pisharady, P.; Vadakkepat, P.; Poh, L.A.
2014, XIII, 138 p. 50 illus., 25 illus. in color., Hardcover
ISBN: 978-981-287-055-1