Chapter 6

Conclusions and Future Directions

6.1 Conclusions

One of the significant problems in pattern recognition and data mining, particularly given the explosive growth of available information, is the dimensionality reduction using feature selection. The use of user-supplied information plays essential role to many existing algorithms for feature selection in the literature. This is a significant drawback of these algorithms. Some of the feature selectors need noise levels to be specified by the user beforehand, some simply rank features leaving the user to select their own subset. There are those that require the user to state how many features are to be selected, or they must provide a threshold that determines when the algorithm should terminate. All of these require the user to make a decision based on their own (possibly faulty) judgement. This may lead to information loss. Rough set theory (proposed by Pawlak) can be used as a tool to determine data dependencies and to reduce the dimension of the dataset using the information from data alone without any additional information. Rough set based attribute selection approach has been successfully implemented to reduce the number of features with preserving the essence of the features. Rough set theory can produce the most informative subset of features from original attribute set of a dataset with discretized feature values. In case of real-valued datasets, rough set theory struggles as it can be applied to the datasets having features containing symbolic values only. So, discretization is applied to the real-valued datasets before using rough set based approach for feature selection, which usually leads to information loss. This problem was solved by using fuzzy rough set based approach for feature selection, which can tackle real-valued datasets in a better way.

D. Dubois and Henry Prade combined fuzzy set (proposed by Zadeh) and rough set (proposed by Pawlak) and presented the concept of fuzzy rough set. Fuzzy-rough set theory is evolved to address the two analogous and complementary concepts, viz., vagueness (for fuzzy set) and indiscernibility concept (for rough sets) with distinct notions and both the concepts are generated as the results of uncertainty in knowledge, which can be implemented for feature selection of datasets containing either discrete or continuous or heterogeneous attributes. The concept of a dependency function in a traditional rough set model into the fuzzy occurrence was proposed by Jensen and Shen and introduced a feature selection algorithm using fuzzy rough set concept and improved by many researchers.

Intuitionistic fuzzy set (proposed by Atanassov) is an extension of fuzzy set (proposed by Zadeh). Intuitionistic fuzzy sets have the better ability of narrating and describing ambiguities of the objective world than the traditional fuzzy sets as it considers the positive, negative and hesitancy degrees of an object simultaneously. In a fuzzy set, the membership degree of the element in a universe always takes a single value between 0 and 1 but those single values may not completely define about the lack of knowledge as the uncertainty is not found only in judgment but also in the identification. Therefore, some extensions of fuzzy sets are required to handle the latter uncertainty. Intuitionistic fuzzy sets have been efficiently applied to solve many of the decision problems with this latter uncertainty. In spite of the fact that rough sets and intuitionistic fuzzy sets both capture specific aspects of the same idea-imprecision, the combination of intuitionistic fuzzy set theory and rough set theory are rarely discussed by the researchers. This thesis consists of four studies, in which different intuitionistic fuzzy rough set model have been presented. Based on these models, different feature selection techniques have been presented using the concept of dependency function.

In the first study, we have given a novel approach for attribute reduction by using tolerance-based intuitionistic fuzzy rough set concept. We have defined lower and upper

approximations against a threshold value δ and presented a method to calculate degree of dependency of decision attribute over a subset of conditional attributes by using the tolerance-based intuitionistic fuzzy rough set for attribute reduction. Moreover, we have validated supporting theorems based on lower and upper approximations. Furthermore, we have applied our proposed algorithm to an example data set and a comparison has been presented with the tolerance-based fuzzy rough set method. We observed that with the previous algorithm the obtained reduct was $\{a, b\}$ and after applying our proposed method, the reduct was $\{b\}$. This clearly indicates the superiority of our proposed work. It is obvious from the given example that our model works fine when trying to discover the smallest reduct from a decision system. Moreover, reduct of the decision system can be improved by adjusting the parameter δ so that the ability of the model to handle tolerance for fault or noise may increase. It is also observed that the proposed algorithm is capable to handle uncertainty, vagueness as well as the noise of the information system.
In the second study,we introduced a novel intuitionistic fuzzy neighborhood rough set by defining a parameterized intuitionistic fuzzy relation and dependency of conditional features over decision. On the basis of intuitionistic fuzzy decision, we defined lower and upper approximations of decision attribute with respect to a subset of conditional attributes. We then defined positive region along with degree of dependency for attribute reduction. Using this model, a greedy attribute reduction algorithm is given. Finally, we applied our approach to an example data set with six attributes, i.e. $\{a, b, c, d, e, f\}$, to get a reduct set and obtained the reduct set as $\{a, b, d\}$.

In the third study, we have proposed a novel concept to feature selection for data set with labelled and unlabelled data. In this paper, we presented an intuitionistic fuzzy rough set model and generalized it for attribute selection for semi-supervised data. Furthermore, we proposed supporting theorems and an algorithm has been presented in order to demonstrate our approach. Finally, the proposed algorithm applied to a data set and observed that our proposed approach is performing better than previously reported semi-supervised fuzzy rough feature selection method in terms of selected attributes. An arbitrary example of semi-supervised fuzzy information system is taken with universe of discourse $X = x_1, x_2, x_3, x_4, x_5, x_6\}$, set of conditional attributes $C = \{a, b, c, d, e, f\}$ and one decision attribute $\{q\}$. In this information system, class labels of the objects x_1 and x_5 are missing while class labels for other objects are available. We applied already existing fuzzy rough set based and the calculated reduct was $\{b, c, e, f\}$. Finally we applied our proposed approach and obtained the more reduced dataset with reduct set as $\{b, d, e\}$. In the fourth study, a novel intuitionistic fuzzy rough set model based on (α, β) indiscernibility concept was introduced to cope with intuitionistic fuzzy information system. This model was validated by using supporting theorems. Furthermore, a positive region based feature selection technique was proposed by using this model. Moreover, a suitable algorithm was presented to calculate reduct for intuitionistic fuzzy information systems. Finally, we applied our approach on an intuitionistic fuzzy information system and calculate the reduct. Initially, the set of five conditional features, i.e. $\{A_1, A_2, A_3, A_4, A_5\}$, were given in the example dataset and after applying our proposed approach we obtained reduct set as $\{A_1, A_4, A_5\}$.

6.2 Future Directions

In the future, we intend to construct the intuitionistic fuzzy variable precision rough set model to tackle a large amount of noise in a data set by relaxing the conditions on membership grade for the lower approximation and tightening for the upper approximation. We can define vaguely quantified intuitionistic fuzzy rough set by giving different levels of membership grades to the lower as well as upper approximations. Furthermore, some robust models for intuitionistic fuzzy rough set based feature selection can be established to preserve the outcome while changing the involved parameters. Moreover, we can propose novel generalized methods for conversion of fuzzy data sets into intuitionistic fuzzy data sets in order to tackle real-valued data. All the above mentioned future endeavors can be very effective to handle uncertainty in the various fields of expert and intelligent systems. There is further scope of the above approach for feature selection based on discernibility matrix and interval intuitionistic fuzzy rough set models in order to tackle uncertainty in a much better way. In future, we want to propose the variable precision and type-2 intuitionistic fuzzy neighborhood rough set models for attribute redution. Moreover, we want to see the effect of parameter δ , since we need to set parameter in advance. All the above mentioned future endeavours can be very effectively implemented to handle uncertainty in the various fields of expert and intelligent systems.