A Rule Based Classification for Vegetable Production Using Rough Set and Genetic Algorithm

R. Rathi, School of Information Technology and Engineering, VIT University, Vellore, India
Debi Prasanna Acharjya, School of Computer Science and Engineering, VIT University, Vellore, India

ABSTRACT

This article describes how agriculture is the main occupation of India, and how the economy depends on agricultural production. Most of the land in India is dedicated to agriculture and people depend on the production of agricultural products. Therefore, forecasting the accuracy of future events based on extracted patterns plays a vital role in improving agricultural productivity. By considering the availability of micronutrients and macronutrients of the soil and water in a particular place, the growth of a plant is determined. This helps people to determine the crops to be cultivated at a certain place. In this article, the forecasting is carried out using rough sets and genetic algorithms. Rough sets are used to produce the decision rules whereas genetic algorithms are used to refine the rules and improve classification accuracy. Accuracy of the classification rules is analyzed using different selection methods and crossover operators. Results show that genetic algorithms with a roulette wheel selection and single point crossover provides better performance when compared with other existing techniques.

KEYWORDS
Agriculture, Binary Relation, Classification, Crossover, Genetic Algorithm, Indiscernibility, Lower and Upper Approximation, Mutation, Prediction, Reduct, Rough Set

1. INTRODUCTION

In India, most of the geographical area is utilized for agriculture. Most of the people in India depend on agriculture and its productivity for their smooth running of day-to-day life. Indian economy today, depends on agriculture and its productivity. Agricultural productivity depends on the environmental factors, soil, and water. Many agricultural researches, surveys have tremendously increased the amount of production in the farmland that is capable of producing. Due to lack of knowledge, a farmer does not know which land is capable of producing more production. It is a very difficult task for many researchers to forecast the capability of land for a suitable plant growth. Many farmers do not get profit from the agriculture production though few farmers make profit with the agriculture by planting right plant in the right place. Therefore, a forecasting method has to be used to solve this problem.

Land suitability is the primary consideration for every farmer who wishes to cultivate. The major attributes which are essential for plant growth are soil, water and the climatic factors. These attributes

DOI: 10.4018/IJFSA.2018010106

Copyright © 2018, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.
should be analyzed before planting a plant so as to get more productivity. Soil has the essential nutrient for plant growth and the characteristics of soil determine the ability of the plant growth and also the absorption of essential nutrients by the plant. Many researchers and scientist are concentrating more on this area by applying soft computing techniques and data mining techniques for land suitability assessment so as to increase the agricultural productivity.

The agricultural data collected will not serve any purpose unless certain meaningful information is extracted from it. The difficult task lies in extracting knowledge from this huge data. This leads to decision rule mining using some well-known technique. Classical rules are used to handle classification in earlier days. Because the knowledge extracted with classical set is very limited and it could not able to process data having inconsistencies and uncertainties. Naturally, the objects in the information system contain uncertainties and imprecise information within it. For example, the rainfall at two different places may differ slightly. It leads to two different classes. Therefore, the concept of classical sets had been extended to fuzzy set (Zadeh, 1965) to handle these uncertainties. A fuzzy optimization model is proposed for analyzing the impact of human activities on ground water level change (Liu, Liu & Luo, 2015). Similarly, fuzzy partial and semi partial correlation rule mining for fuzzy data is also proposed (Sonia, Robinson & Rajesekaran, 2015). However, defining the membership function in fuzzy set is still critical. Later Pawlak (1982) invented rough set which handles the uncertainties among objects to model imperfect knowledge. The basic notion in this approach is an equivalence relation. The major advantage is that, it never uses any membership function to classify the objects. Further, rough set has been extended to many directions (Dubois & Prade, 1990; Acharjya, 2015). Additionally, rough set is hybridized with neural network (Anitha & Acharjya, 2015), rough set on fuzzy approximation space is hybridized with soft set (Das & Acharjya, 2014) etc.

The relationship between wheat yield and soil nutrients is analyzed using artificial neural network (He, Zhang, Zhang & Fang, 2005). Similarly, status of biomass and nitrogen in wheat crop and its discrimination using artificial neural network is also discussed (Junior, Caires & Guimaraes, 2014). Guo & Xue (2014) carried out their research work for forecasting crop yield using artificial neural network. They focused on comparing two models spatial and temporal pertaining to yield of crops. Furthermore, optimal use of fertilizers in crop production is also stressed upon (Ogasawara, Oliveira, Junior, Castaneda, Amorim, Mauro, Soares, Quadros & Bezerra, 2013). On the other hand genetic algorithm is used to maximize crop yield (Olakelehin & Omidiara, 2014). But, they failed to concentrate on uncertainties involved in the dataset. The improvements to classification of crop problems using principal concept analysis and genetic algorithm is also discussed (Cruz, Gerardo & Tanguilig III, 2014). Further in the literature it is found that support vector machine is used for crop classification (Campos-Valls, Gustavo, 2003). Various data classification techniques like neural networks, decision tree, and support vector machine were applied successfully over agricultural data. Biradar & Nigudgi (2012) discussed about the decision tree classifier and used statistical analysis over agricultural data. Bhargavi & Jyothi (2009) compared the performance of the three classification algorithms for soil classification. But, all these techniques fail to handle uncertainties and inconsistencies.

To overcome the limitations, we hybridize rough set and genetic algorithm. Rough set that handles uncertainties and inconsistencies. Further it computes reducts and generates decision rules. The decision rules generated by rough set are still refined using genetic algorithm to get an optimal and accurate solution. Genetic algorithm converges to the best chromosome, which is nothing but the optimal solution to the problem. Rough set is used in the initial-process whereas genetic algorithm is used as final-process in the proposed research design. In the final-process we have used various combinations of selection method and various combinations of crossover operator to improve the accuracy. Finally, we present a real-life case study on agriculture data collected especially for Tiruvannamalai district of Tamilnadu, India. The proposed hybrid model ends up with promising results when compared with standalone model of rough set theory and other existing techniques.
Ubiquitous Mediation and Critical Interventions: Reflections on the Function of Signs and the Purposes of Peirce’s Semeiotic
www.igi-global.com/article/ubiquitous-mediation-critical-interventions/56444?camid=4v1a