Applied Soft Computing 11 (2011) 3643–3657

Contents lists available at ScienceDirect

Applied Soft Computing

journal homepage: www.elsevier.com/locate/asoc

Fuzzy cognitive map based approach for predicting yield in cotton crop production as a basis for decision support system in precision agriculture application

E.I. Papageorgiou[a,*], A.T. Markinos[b], T.A. Gemtos[b]

[a] Department of Informatics and Computer Technology, Technological Educational Institute (TEI) of Lamia, 3rd Km Old National Road Lamia-Athens, 35100 Lamia, Greece
[b] Laboratory of Farm Mechanization, Department of Agriculture, Crop Production and Rural Environment, University of Thessaly, Greece

ARTICLE INFO

Article history:
Received 11 February 2009
Received in revised form
21 September 2010
Accepted 30 January 2011
Available online 25 February 2011

Keywords:
Fuzzy cognitive maps
Modeling
Knowledge representation
Fuzzy sets
Decision making
Cotton
Yield

ABSTRACT

This work investigates the process of yield prediction in cotton crop production using the soft computing technique of fuzzy cognitive maps. Fuzzy cognitive map (FCM) is a fusion of fuzzy logic and cognitive map theories, and is used for modeling and representing experts' knowledge. It is capable of dealing with situations including uncertain descriptions using similar procedure such as human reasoning does. It is a challenging approach for decision making especially in complex processing environments. The FCM approach presented here was chosen to be utilized in agriculture because of the nature of the application. The prediction of yield in cotton production is a complex process with sufficient interacting parameters and FCMs are suitable for this kind of problem. Throughout this proposed method, FCMs designed and developed to represent experts' knowledge for cotton (*Gossypium hirsutum* L.) yield prediction and crop management. The developed FCM model consists of nodes linked by directed edges, where the nodes represent the main factors affecting cotton crop production such as texture, organic matter, pH, K, P, Mg, N, Ca, Na and cotton yield, and the directed edges show the cause–effect (weighted) relationships between the soil properties and cotton yield. The investigated methodology was evaluated for 360 cases measured during the time of six subsequent years (2001–2006) in a 5 ha experimental cotton field, in predicting the yield class between two possible categories ("low" and "high"). The results obtained reveal its comparative advantage over the benchmarking machine learning algorithms tested for the same data set for the years mentioned by providing decisions that match better with the real measured ones. The main advantage of this approach is its simple structure and flexibility, representing knowledge visually and more descriptively. Hence, it might be a convenient tool in predicting cotton yield and improving crop management.

1. Introduction

Early estimates of agricultural production are of great importance for agricultural policy and trade. Yield prediction based on the combination of the factors affecting it, is important for farm management. Cotton is a very important crop especially in Greece. There is a great need for good estimates of yield and total biomass production. The importance of this factor is more critical when site specific management is considered. Precision farming generates data which, due to their type and complexity, could not be efficiently analyzed by traditional methods. Crop management traditionally has been based mainly on the experience of the farmer for his field. But later with farm mechanization, the direct connection of the farmer with the field was lost and management was based on yield and soil mean properties. During the last years soil analysis data were available and were used for the crop management. The management was based on the mean values of the parameters under the assumption that the fields were homogeneous and yielded equally throughout their whole area. But farmers knew that their fields were not producing equally in all parts. When the first yield monitors were used in the 1990s it was revealed that yield was different in different parts of the field. The same findings were also established for soil properties and other parameters affecting yield. This variation should be managed properly in order to have better cropping results and reduce the adverse effects of agriculture to the environment. A new type of farm management was developed under the name of site specific management (SSM). SSM requires new analysis tools and models in order to achieve the maximum of the benefits both for the farmer and for the environment. The aim of this work is to apply this knowledge based approach in order to handle the large amount of data collected through sev-

* Corresponding author.
E-mail addresses: epapageo@ece.upatras.gr, epapageorgiou@teilam.gr (E.I. Papageorgiou), markinos@agr.uth.gr (A.T. Markinos).

1568-4946/$ – see front matter
doi:10.1016/j.asoc.2011.01.036

eral sensors and develop analysis tools for cotton management on a qualitative and more quantitative basis and some research has partly succeeded to do this. In the remaining paragraphs of this section, a literature review on methods and algorithms for cotton management and yield estimation is presented.

In most cases, the empirical methods for yield estimation include correlation and multiple regression models. They are commonly used to assess yield prediction and identify important factors influencing yield [1,2], but the results are not so encouraging due to the presence of polynomial and interaction terms [69]. Furthermore, they have an intensively local character and it is difficult for them to be generalized. On the other hand, there are more complicated models simulating the physiological processes of the plant growth, taking into account any factor affecting the crop and containing up to a thousand of equations [3].

A large number of approaches, models, algorithms and statistical tools have been proposed and used for assessing yield prediction. Many authors have used simple linear correlations of yield with soil properties but the results have varied from field to field and year to year [1,4,5]. Many other studies, using complex linear methods like multiple linear regression, have given similar results [1,2,4]. Some authors proposed non-linear statistical methods to investigate yield response [6,7].

Factors affecting crop yield are so complex that even statistical methods cannot give accurate results. Artificial neural networks (ANNs) were employed to model the nonlinear relationship between cotton yield and the factors influencing yield. Most of the reported studies have used ANNs and machine learning algorithms for setting target yields which is one of the problems in precision farming [8–10]. Schultz et al. [11] summarized the advantages of applying neural networks in agro-ecological modeling, including the ability of ANN to handle both quantitative and qualitative data, merge information and combine both linear and non-linear responses. Neural networks have been proposed for identifying important factors influencing corn yield and grain quality variability [10], for data analysis [12], for predicting crop yield based on soil properties [13], for setting target corn yields [8]. Shearer et al. [14] studied a large number of variables, including fertility, satellite imagery and soil conductivity for a relatively small number of observations in one site-year of data.

In the case of knowledge-based systems in the field of agriculture using fuzzy logic techniques, only a few studies have been undertaken (e.g. [15,16]). In a more recent study, Khan and Khor [70] proposed a framework of a fuzzy rule-based cognitive map. The simulation model based on fuzzy logic for the mapping of FCM input state space to the output state space has been implemented for corn yield estimation, considering only four yield factors (pH, potassium, phosphorus and organic matter). The model predicted the variations in corn crop yield with each one of the four factors using simulated data. The FCM simulations were conducted to study the effects of varying the membership grade of one yield factor, while keeping the other three factors constant at 0.1.

The aim of the work demonstrated in this paper is to present a methodology that can predict cotton yield behavior in precision farming, based on artificial intelligence techniques and particularly based on aspects related to knowledge representation. Fuzzy cognitive map (FCM) is a knowledge representation and management tool and it was chosen because of the nature of the application; yield prediction is a complex process with sufficient interacting parameters and there is available experience and accumulated knowledge from experts, farmers, and agriculture scientists.

FCMs are appropriate to explicitly encode the knowledge and experience accumulated on the operation of a complex system [17,18]. They can be viewed as an extension of cognitive maps, and their main advantages include flexibility and adaptability to a given domain [19,20]. Once constructed for a particular domain, an FCM allows a qualitative simulation of the system. In addition, FCMs represent knowledge in a symbolic manner, encoding the relations between the elements of a mental landscape so that the impact of these elements can be assessed. They are based on knowledge and experience for describing particular domains using concepts (variables, states, inputs, outputs) and the relationships between them, while taking into account the degree of uncertainty that may characterize these relationships in the real world using fuzzy logic [18]. This technique is particularly suitable to model qualitative rather than quantitative systems. Being a qualitative approach, FCMs are free from most of the drawbacks that are inseparable regarding quantitative modeling techniques [68].

FCMs were applied to a large number of diverse application areas such as engineering, medicine, political sciences, earth and environmental sciences, economics and management, etc. [19], and have already gained momentum due to their simplicity and easiness of use. A number of examples of specific applications of FCMs include: political developments [21], analysis of electrical circuits [22], failure modes effects analysis [66], B2B e-commerce decision making [23] management of relationships among organizational members of airline services [24], stock investment analysis [25], ecology and conservation [26,62,67], forest management [27], modeling of complex technological systems [28], modeling of software development project [29], time-series prediction [30], trust dynamics analysis in virtual enterprises [31], pattern recognition [32], business-to-consumer e-commerce web-based systems [33], management and organizational learning in support of organizational memory [34,35], nuclear power reactors [36], agile NPD process [37], identification of critical path in strategic domains [38], modeling educational software [39], cotton management in precision farming [44,61], medical diagnostics [40,41,71], medical decision making [42], decision-modeling for the assessment of the impact of contemporary human resource management (HRM) practices to the shareholder value and satisfaction [43], and many others. The scope and range of the applications demonstrate the usefulness of this method and motivate further research in this area.

In a recent work [44], the modeling approach of FCMs was investigated for the first time to help farmers by making decisions in precision agriculture. Through the proposed methodology, an FCM for describing cotton yield in central Greece, a main Greek agricultural area, was constructed and simulated giving useful and acceptable results to yield prediction. The success of precision agriculture depends on accurate and detailed knowledge of yield potential and crop response to specific conditions. The methodology of FCMs is evoked throughout this work from a different viewpoint to accomplish this task including more conditions and compared with other intelligent benchmarking techniques.

Therefore, this is the first step in the development of decision support system in agriculture that will help in decision making process, through the design of the knowledge representation and the design of reasoning with the aim of FCM in order to automate the decision. The next section provides an introduction to the FCM theory. This is followed by the reasons which strengthen the FCM approach to be considered appropriate for modeling this particular domain. The third section provides a description of the material and methods used to our approach, as well as a brief study on machine learning approaches for comparison purposes. Next the stages in the development of the FCM model are described, the resulting model is presented and its simulation analysis and results discussed. Finally, a comparison of the results with the benchmark computational intelligent methods and of the limitations of this study is performed and a conclusion is presented which explores the potential of the FCM as a dynamic model for making decisions.