COMPUTERS AND ELECTRONICS IN AGRICULTURE 65 (2009) 99–113

available at www.sciencedirect.com

ScienceDirect

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

Using fuzzy data mining to evaluate survey data from olive grove cultivation

G. Delgado[a,*], V. Aranda[b], J. Calero[b], M. Sánchez-Marañón[a], J.M. Serrano[c], D. Sánchez[d], M.A. Vila[d]

[a] Department of Pedology and Agricultural Chemistry, University of Granada, Spain
[b] Department of Geology, University of Jaén, Spain
[c] Department of Computer Science, University of Jaén, Spain
[d] Department of Computer Science and A.I., University of Granada, Spain

ARTICLE INFO

Article history:
Received 15 April 2008
Received in revised form
6 August 2008
Accepted 7 August 2008

Keywords:
Fuzzy data mining
Fuzzy database
User knowledge
Land evaluation
Olive grove cultivation

ABSTRACT

User knowledge about cultivation and soils is relevant in developing countries, given the lack of any other kind of information, and in developed countries where it may contribute to the sustainable and operational planning for farm production systems. But user data show a high degree of inaccuracy and uncertainty; hence the appropriate treatment of this kind of data in order to obtain potentially useful information requires suitable storage and processing techniques, such as fuzzy relational database and fuzzy data mining techniques. The user data were obtained from a survey with 126 variables carried out on 210 olive grove farms in the Province of Granada (southern Spain). A set of 34 variables were selected following a cleaning process, and 1420 fuzzy association rules relating handling, soil, or environment with olive fruit production (kg ha^{-1}), % of oil in olive fruit or acidity of fruit on tree (olive oil quality), were obtained. Several of these rules can be considered as *user evaluation rules* (UER) because they show a clear association between the corresponding variables. It is demonstrated that the greatest influence of the variables for handling, soil and environment on the production and the quality of the olive oil, arises when those variables limit or reduce the production or quality. Some UER are corroborated with the knowledge regarding olive cultivation found in the bibliography. Some others contradict the knowledge and others reveal relationships not described previously. Because of this, the proposed working method can be considered as mostly exploratory. The final objective of this work is to help decision-making in the field of olive cultivation in Andalusia. However, the working method described in this paper is applicable to other geographical areas and to other kinds of crops.

1. Introduction

In developing countries, in which scientific information about soils and crops is scant, the information supplied by farmers (user or traditional knowledge) is particularly useful (Grobben, 1992; Habarurema and Steiner, 1997).

In developed countries many research scientists and extension workers also recognize that farmers have a valuable

* *Corresponding author at*: Departamento de Edafología y Química Agrícola, Facultad de Farmacia, Universidad de Granada, Campus Cartuja, 18071 Granada, Spain.
E-mail address: gdelgado@ugr.es (G. Delgado).
0168-1699/$ – see front matter © 2008 Published by Elsevier B.V.
doi:10.1016/j.compag.2008.08.001

understanding of their resources (Thrupp, 1989), thus, there is a great challenge to listen and learn from the traditional knowledge (He et al., 2007). User knowledge about soil and cultivation is an important component in the development of effective specific management and useful policy. Traditional knowledge is developed and used over time by local people and is influenced by environmental and socio-economic realities (Saito et al., 2006; Beckford and Barker, 2007), and should be taken into account in rural development programs (Ali, 2003; Barrera-Bassols et al., 2006). For development planning and interventions to be successful it is necessary to fit external technologies and strategies to the local environmental and cultural context (Niemeijer and Mazzucato, 2003). User participation at all stages is an essential ingredient.

According to Stocking and Murnaghan (2001), user knowledge is difficult to process and interpret because its accuracy and reliability is not always clear. Furthermore, part of the information contained in user data is not always accessible to the common methods of classical statistics (Berzal et al., 2004).

Fuzzy data mining techniques are useful for extracting user knowledge from data (Delgado et al., 2003; Calero et al., 2003). These techniques are capable of handling information in which uncertainty, imprecision and ambiguity are a part, as often occurs in soil data (McBratney et al., 2000) or descriptions of cultivation information (Center and Verma, 1998).

Data mining techniques as prediction tools, and fuzzy logic, as a method of treating uncertainty and imprecise data, have been successfully used to solve some agricultural and environmental problems, such as improving cherry tomato cultivation (Hoshi et al., 2000), evaluating and minimizing the effect of pesticides and tillage in a sustainable mixed agriculture (Ferraro et al., 2003), predicting sugar production (Petridis and Karbulasos, 2003), optimizing the fertirrigation efficiency in a sustainable strawberry culture, seeking a balance between yield and environmental damage (de la Torre et al., 2005), minimizing the volume of herbicides used to treat weeds in a sustainable precision agriculture (Tellaeche et al., 2008), environmental evaluation of human activities (Kawano et al., 2005), etc. In this paper, we have used data mining techniques and fuzzy association rules (FARs) to evaluate data from a survey about olive groves in the Province of Granada (region of Andalusia).

Olive cultivation supports the population in many rural Mediterranean areas (Loumou and Giourga, 2003). Spain is the largest olive oil producing country (about 30% of worldwide production) and the region of Andalusia (southern Spain) accounts for 75% of olive oil production in Spain (de Graaff and Eppink, 1999).

The increasing demand for intensification of olive cultivation in the Province of Granada, especially in the areas with less favourable conditions such as mountains or arid lands, requires a detailed knowledge of the physical and socio-economic resources involved in the production system (Rallo, 1998). Moreover, they must be taken into consideration from the viewpoint of the environmental effects of modern agricultural practices. The need for optimal use of land in olive groves has never been greater than at the present time (Berzal et al., 2004).

The aim of this paper is to process the information, which was provided by a survey amongst olive grove farmers from the Province of Granada and which has a high degree of inaccuracy and uncertainty, using fuzzy data mining techniques, with the aim of making the user knowledge easier to handle.

From user knowledge, once suitably processed, we intended to obtain user evaluation rules (UER) for olive production in the Province of Granada. We did not intend to create a complete evaluation system for this crop, but rather to discover rules that will complement the classic evaluation systems by including experimental information provided by farmers. Furthermore, we intended to discover trends that mostly require specific research work in order to be ascertained.

The experience based on this advanced data processing methodology could be extended to the whole of the olive production in the Mediterranean basin and to other geographical areas around the world and even to other kinds of crops.

2. Materials and methods

2.1. Site description and sources of information

The Province of Granada (Fig. 1) in southern Spain, has a surface area of around 12,635 km^2. The climate is continental Mediterranean, with an average annual temperature between 15 and 20 °C. Rainfall at less than 500 mm year^{-1} is concentrated in winter months and spring, with an extended drought in the summer. Steep slopes (>30%) where machinery cannot be used cover over 31% of this area. Altitudes of between 800 and 1200 m a.s.l. cover 50% of the province.

Fig. 2 summarizes the methodological scheme followed in this work. The steps of the scheme can be described in detail as follows:

Fig. 1 – Olive-grove counties in the Province of Granada (southern Spain) and distribution of cultivation (%).

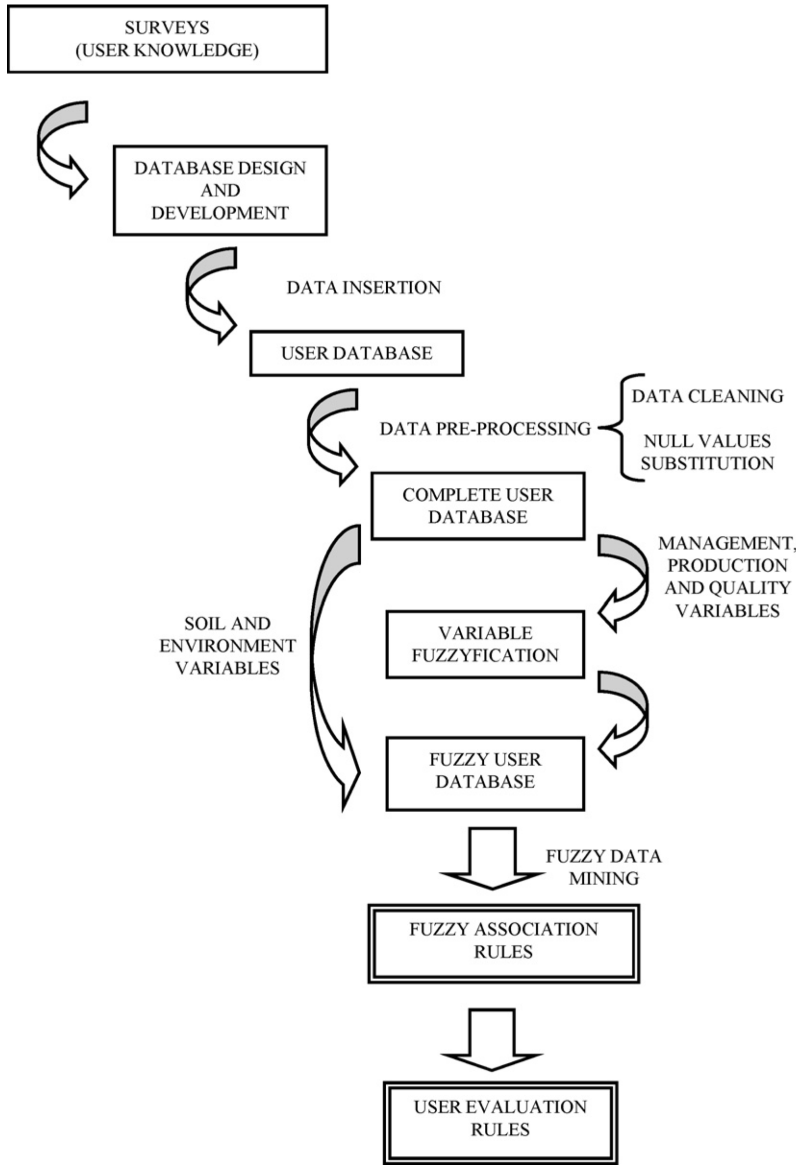


Fig. 2 – Methodological scheme.

Data used in this paper were obtained from surveys carried out in all the olive-growing areas of Granada (Fig. 1). The interviews were carried out in summer of 2000 and include 126 variables, 24 related to soils, covering aspects such as geographical location, management techniques, costs, production and soil characteristics. A total number of 210 farmers were surveyed. The number of participants needed to satisfactorily characterize the olive crop in the Province of Granada should be at least 148, proportionally distributed by the size of the olive growing districts (Loring-Miró, 1999).

The environmental variables were assigned to each survey according to the approximate location on the map for each farm surveyed. The *mean altitude* was obtained from the topographic map. The data for *mean annual rainfall* and *mean minimum temperature of the coldest month* were obtained from the network of weather stations belonging to the National Weather Service. The data for the closest station are assigned to each survey questionnaire.

2.2. *Fuzzy database design and development: information representation and storage*

Since one of our objectives was to obtain user knowledge at the end of the process, and that user knowledge is expressed usually in an imprecise way, the survey employed was designed to obtain such data and had to deal with imprecision (when users, farmers in our case, cannot return exact answers) and uncertainty (when it is not sure how true the information provided by the users is) coming from several sources. In order to handle the data gathered from the surveys and the inherent imprecise relations that can be defined between the terms employed, data were stored in a fuzzy relational database. Although several fuzzy database models can be found in the literature, see Medina et al. (1995) for a brief summary, at the time of this study, and given the nature of the data to be processed, we chose the GEFRED model (Medina et al., 1994). This model extends the classic relational database model (Codd, 1970). The main reason for our choice is that this model put together the different types of imprecision and uncertainty considered by other models (linguistic labels represented by fuzzy sets, possibility distributions, and resemblance relations) into a single model, together with the necessary operators to define, compare and, in general, deal with values expressed in different ways. Since these different types of representations are needed in the data we collected, GEFRED seemed to be the most appropriate and complete model for our problem.

FIRST architecture (Medina et al., 1995) implements the tools and structures proposed by GEFRED, which, in our case, is built under a database system provided by Oracle. Basically, it adds a number of new components to the existing structures, not only letting us to store and process non-fuzzy (crisp) data, but, according to our needs, fuzzy data. It also provides an extension to SQL language, named FSQL (*Fuzzy SQL*) (Galindo et al., 1998), which let us to perform flexible queries over the fuzzy database, i.e., a question like *"Which farms have a High oil production in Medium-depth soils"* in our database could be translated onto FSQL as:

SELECT surveycode FROM surveytable
WHERE production FEQ $High AND soildepth FEQ $Medium

where $*High* and $*Medium* are linguistic labels for the corresponding attributes (see Table 2), and FEQ (*Fuzzy Equal*) is an FSQL operator meaning "*approximately equal to*".

As FSQL is defined over SQL, by means of flexible queries we can retrieve both fuzzy and crisp data.

On the basis of this model, and with the help of experts (mainly soil scientists and agricultural technicians), a fuzzy database was designed and developed, so that the imprecision and uncertainty associated to the surveyed values for variables of the groups "Soil" and "Environment" (Table 1) could be managed in a satisfactory way. In particular, this was also possible because of the wide knowledge of soils and climate in this area, and the bibliographic information available (Aranda et al., 2001).

On the contrary, the existing knowledge to deal with imprecision in the agronomic variables ("Management, production and quality" in Table 1) was insufficient without considering previously the distribution of values in the surveys. Because of this, these were stored as crisp values, although they could be processed by fuzzy techniques later.

2.3. Data pre-processing of surveys

After creating the database, we pre-processed the data before the data mining step. This procedure consisted of data cleaning and null values substitution (i.e., values that are unknown or not present). Due to the uncertainty associated with the survey procedure, as in many cases the farmers were not able to give a proper answer, we should first ignore cases where null values are present and variables with more than 50% of missing values were eliminated, as classical clustering methods are not able to manage null values, and attributes with a high rate of missing values should lead to wrong results. The next step consisted of replacing the null values from the data matrix, so that the minimum amount of information would be lost in the process. We do not use previous user-defined typologies in this paper; we established a set of homogeneous clusters using the survey instead. The process of replacing null data was performed in two steps: first, a hierarchical clustering was applied over the set of "complete cases", that is, we considered only those variables where no null or missing values were present. We took into account expert criteria in order to choose the most appropriate number of clusters. Second, the resulting cluster set was refined, in order to reduce, if possible, the number of clusters, using a fuzzy *K*-means algorithm (Dubes and Jain, 1988; López de Mántaras and Valverde, 1988) to replace null or missing values in the original data matrix. This last process consisted in: (1) taking a training set (the clusters of "complete cases" previously obtained by means of the hierarchical clustering), (2) re-classifying cases, computing a classification error, and (3) repeating this process until the classification error is zero (or nearly zero), or a number of iterations is reached without any change in the classification error. Once we obtain the membership degree of each case to a cluster, we replace the null value with the value of the centroid for that variable, thus obtaining the initial matrix with 210×34, but now without null values. This process is explained in detail in (Serrano et al., 2002).

2.4. Fuzzyfication of agronomical variables

Finally, we had a certain number of crisp variables ("Management, production and quality" variables in Table 1) that needed to be fuzzified in order to be properly processed by our fuzzy data mining tools.

In this process, we distinguish between numerical (continuous) and scalar (nominal and ordinal) attributes.

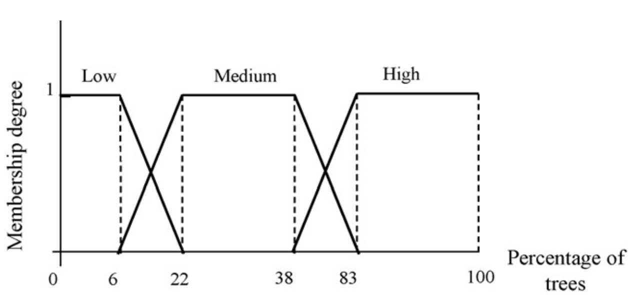
For each numerical attribute, we defined a set of linguistic labels {*high*, *medium*, *low*} over the numerical domain, in order to reduce the granularity level (that is, the total amount of values for each domain). A short set of values for a particular attribute domain allows us to represent and manage the user knowledge easier than continuous values (Hussain et al., 1999). This domain discretization was carried out following an unsupervised *splitting* algorithm that split a continuous domain into a set of categories previously specified. Previously, outliers are eliminated by means of a *binning-equal frequency* procedure (Hussain et al., 1999). Each category obtained is associated to a *linguistic label* by means of a heuristic *tuning* procedure, from which membership functions were obtained by expert knowledge. In this way, each numerical variable was represented by three trapezoid fuzzy numbers (Zadeh, 1978; Dubois and Prade, 1988), each one represented by a possibility distribution that can be defined by four parameters α, β, γ and δ (see Fig. 3 for an example). This definition allows us to work with linguistic terms instead of numerical values. This way, we can summarize information and manage it in a similar way to farmers.

In the same way, for each non-numerical attribute, we defined a fuzzy resemblance relation (Rundensteiner et al., 1989) between values. These relations let us manage the possible semantic overlapping within data and were provided by the domain experts.

Table 1 – Survey variables selected from the cleaning process

	Variable	Variable type	Mean value
Management, production and quality	Farming area (hectares)	Continuous	12
	Number of olive trees	Continuous	1429
	% *Picual* variety	Continuous	32
	% planting with traditional system[a]	Continuous	65
	% planting with modern system[b]	Continuous	30
	% trees with three trunks	Continuous	51
	% in planting frame of less than 10 m	Continuous	69
	% young olive trees (2 to 5–8 years)	Continuous	24
	% olive trees in full production (>5-8 years)	Continuous	55
	% area using conventional labour	Continuous	82
	% area harvested by knocking down	Continuous	88
	Pruning intensity	Ordinal	–
	Pruning frequency	Continuous	0.95
	Irrigation	Ordinal	–
	Irrigation with own water	Ordinal	–
	Phytosanitary treatments frequency (1/>1 year^{-1})	Ordinal	–
	Cost of management (€ ha^{-1})[c]	Continuous	675
	Cost of fertilizer applied (€ ha^{-1})[d]	Continuous	151
	Cost of phytosanitary treatments (€ ha^{-1})[e]	Continuous	121
	Total costs (€ ha^{-1})	Continuous	947
	Olive fruit production (kg ha^{-1})	Continuous	3230
	% oil in olive fruit	Continuous	22.8 (736 kg ha^{-1})
	Acidity of fruit on tree (acidity)	Continuous	0.93
Soil	Soil depth	Ordinal	–
	Hardened layer depth	Ordinal	–
	Workability	Ordinal	–
	Mean slope	Ordinal	–
	Stoniness	Ordinal	–
	Soil texture	Ordinal	–
	Erosion	Nominal	–
	Soil colour	Nominal	–
Environment	Mean altitude (m)	Continuous	846
	Mean annual rainfall (mm year^{-1})	Continuous	487
	Mean minimum temperature of the coldest month (°C)	Continuous	−4.2

[a] Cultivation of tree fragments without previous treatment.
[b] Cultivation of tree fragments previously treated and controlled in a greenhouse.
[c] Includes: cost of harvest by knocking down, cost of pruning and cost of labour.
[d] Includes: cost of fertilizer, cost of fertilizers for leaves and cost of fertirrigation.
[e] Includes: cost of insecticides and cost of weed-killers.
The mean value has not been included for the discontinuous variables due to its lack of statistical meaning.



For example: linguistic label "Medium" is associated to a possibility distribution defined by four parameters (α=6, β=22, γ=38, δ=83).

Fig. 3 – Trapezoidal fuzzy numbers (*low*, *medium*, *high*) associated to the variable % *Picual* variety.

2.5. *Fuzzy data mining tools: fuzzy association rules*

Data mining techniques can be highly effective when analyzing user knowledge. This is the case in particular of association rules. Formally, let I be a set of items and T a set of transactions, each transaction being a subset of I. Association rules (ARs) (Agrawal et al., 1993) are "*implications*" of the form $A \Rightarrow C$ that relate the presence of itemsets (sets of items) A (called *antecedent*) and C (called *consequent*) in transactions of T, assuming $A,C \subseteq I$, $A \cap C = \emptyset$ and $A,C \neq \emptyset$. In order to assess the accuracy and importance of ARs, *confidence* (noted $\mathrm{Conf}(A \Rightarrow C)$) and *support* ($\mathrm{Supp}(A \Rightarrow C)$) are usually employed (Agrawal et al., 1993). Both are based on the notion of and *support of an itemset*, defined as the percentage of transactions containing the itemset. Then, $\mathrm{Supp}(A \Rightarrow C) = \mathrm{Supp}(A \cup C)$ and $\mathrm{Conf}(A \Rightarrow C) = \mathrm{Supp}(A \Rightarrow C)/\mathrm{Supp}(A)$.

Several authors have pointed out that the measure of confidence can yield counterintuitive results. As proposed in (Berzal et al., 2001, 2002), we employ Shortliffe and Buchanan's Certainty Factors (CF) (Shortliffe and Buchanan, 1975). Certainty Factor takes values in [−1, 1], indicating the percentage of increment or decrement of our belief that the consequent is true when the antecedent is true.

A relational database (RDB) is transformed into a set of transactions by considering that each item is a pair <*attribute*, *value*> and each transaction is associated to a tuple (row) of the table, so a given item $i = <A, a>$ appears in a given transaction t if the corresponding attribute A takes the value a in the tuple corresponding to t. In order to obtain more meaningful rules, numerical values in items are replaced by linguistic labels defined by means of fuzzy sets. Since a linguistic label is compatible with a crisp value to a certain degree, the item pertains to the transaction to that degree, leading to the notion of fuzzy transactions as fuzzy subsets of items. There are some proposals to obtain (fuzzy) association rules (FARs) from fuzzy transactions. In particular, we follow that in (Delgado et al., 2003).

Fuzzy transactions can be also obtained by using fuzzy resemblance relations defined on a set of categorical values (i.e., if a and a' are resemblant to a certain degree α, then a transaction containing the item $<A, a>$ with degree 1 also contains $<A, a'>$ to a degree α). Again, this is useful since it allows us to introduce the knowledge of users (farmers) about the resemblance between terms employed in the database (Sánchez et al., 2004). We can incorporate these items to fuzzy transactions and obtain again rules as in (Delgado et al., 2003).

An association rule is considered strong iff both its support and accuracy are above user-defined thresholds *minsupp*, and (in our case) *minCF*. The computational procedure employed in this paper in order to extract strong association rules in a set of fuzzy transactions follows the adaptation of the classical Apriori algorithm (Agrawal and Srikant, 1994) as proposed in (Delgado et al., 2003). It consists of two steps:

(1) Obtain all frequent itemsets. An itemset B is said to be frequent if Supp(B) is higher than an user-defined threshold. This is the most computationally expensive part of the procedure. We follow the Apriori algorithm, in which the lattice of itemsets with respect to set inclusion is explored level by level (i.e., first singletons, then itemsets containing two items, etc., ...). Making use of the property that all the subsets of a frequent itemsets must be frequent, we employ the usual Apriori candidate generation in order to discard as possible frequent itemsets those that have a non-frequent subset.
(2) Generate all possible association rules from the set of frequent itemsets. Given a two frequent itemsets B and $B \cup \{i\}$ the rule $B \Rightarrow \{i\}$ is strong iff its Certainty Factor is greater or equal than *minCF*. This is a very fast computation.

The only necessary modifications to this procedure with respect to the usual Apriori algorithm are: (a) calculation of CF instead of confidence, and (b) dealing with the degrees in fuzzy transactions when calculating the support. In the first case, the complexity of the algorithm is the same since CF is calculated from confidence and the support of the consequent of the rule in O(1) time. In the second one, an array of k elements, corresponding to a selected set of alpha-cuts, is employed to calculate the support in time O(1). A typical value for k is 20.

In our experiments, we implemented the algorithm in Java (since we wanted to use a programming language without restrictions on machine architecture or operating system), which let us to connect to our database system under Oracle through a JDBC bridge. In practice, time and space employed are similar to those of a classical Apriori algorithm, even taking into account the connection to the fuzzy database and the internal computation.

3. Results and discussion

3.1. *Data pre-processing and analysis of the survey*

Of the 126 variables in the survey, only 34 remained following the cleaning process (Table 1). The 34 variables cover a wide range of characteristics of the olive agropedological system considered in this study. Consequently, the variables initially eliminated do not affect the essence of the research work. This statement is based on some variables eliminated being represented implicitly in the variables kept, or else variables with a specific interest to soil science and/or artificial intelligence, but which are useless for evaluating the agricultural system.

Data for the average area of the farms show that the olive grove in the survey (12 ha) is similar to the one given in statistical yearbooks for all of Andalusia (13.9 ha), downloaded from: http://www.juntadeandalucia.es/agriculturaypesca/portal/opencms/portal/portada.jsp. In terms of olive production, the data are consistent with those of intermediate size olive groves (between irrigated and mixed), with an *olive fruit production* of 3230 kg ha^{-1} (higher than that of any non-irrigated olive grove, although it should be noted that 31% of them are irrigated). The mean production of olive oil is 736 kg ha^{-1} (22.8% of *olive fruit production*), which matches the figure given by the *Department of Agriculture and Fisheries* (web page cited before) for the region of Andalusia (southern Spain). The percentage of *oil in the olive fruit* (22.8%) is agronomically optimum. The quality is also within the optimum range since the average *acidity of fruit on tree* is less than 1, which is a requirement for extra virgin olive oil. Thus, the variables for production and quality (variables *P/Q*) in the survey give an olive grove consistent with the average in Andalusia.

The mean production costs (the sum of the partial costs of tillage, pruning, harvest, pest-control substances and fertilizers) are slightly lower than those estimated in the bibliography (Loring-Miró, 1999).

3.2. *Fuzzy association rules*

Table 2 shows the values α, β, γ and δ for the linguistic labels (*high*, *medium* and *low*) of the numerical variables used in the fuzzy data mining. Approximately 1420 FARs with positive Certainty Factor were obtained with the linguistic labels in which the consequent is a variable of production (production in the strict sense or yield) or of quality of the olive oil (acidity). Those considered to be important were selected from this total of

Table 2 – Possibility distributions for analyzed variables

	Variable	Linguistic labels parameters (α, β, γ, δ)							
		Low				High			
				Medium					
Agronomical management techniques and production	Farming area (hectares)	0.1	0.1	0.9	1.6	4	12	241	241
	Number of olive trees	28	28	90	186	413	1500	25000	25000
	% Picual variety	0	0	6	22	38	83	100	100
	% planting with traditional system	0	0	7	32	62	86	100	100
	% planting with modern system	0	0	9	36	56	86	100	100
	% trees with three trunks	0	0	13	38	57	81	100	100
	% in planting frame of less than 10 m	0	0	7	29	49	83	100	100
	% young olive tress (2 to 5-8 years)	0	0	2.5	20	38	86	100	100
	% olive trees in full production (> 5-8 years)	0	0	20	52	71	81	100	100
	% area using conventional labour	0	0	7	37	48	76	100	100
	% area harvested by knocking down	0	0	33	50	33	50	100	100
	Cost of management (€ ha^{-1})	0	0	54	89	120	197	1288	1288
	Cost of fertilized applied (€ ha^{-1})	0	0	276	396	588	926	3748	3748
	Cost of phytosanitary treatments (€ ha^{-1})	0	0	23	58	100	180	773	773
	Total costs (€ ha^{-1})	82	82	436	659	874	1250	4161	4161
	Olive fruit production (kg ha^{-1})	208	208	1500	2334	3200	5000	25000	25000
	% oil in olive fruit	17	17	20	22.5	23.5	25	35	35
	Acidity of oil (°acidity)	0.1	0.1	0.4	0.5	0.6	0.9	1	1
Soil	Soil depth (m)	0.1	0.25	0.5	0.55	1	1.1	1.5	1.5
	Hardened layer depth (m)	0.1	0.25	0.5	0.55	1	1.1	1.5	1.5
Environment	Mean altitude (m)	Low				High			
				Medium					
		300	300	663	800	900	1023	1200	1200
	Mean annual rainfall (mm yr^{-1})	150	150	450	494	523	551	601	601
	Mean minimum temperature of the coldest month (°C)	-5.5	-5.5	-4.9	-4.6	-4.2	-3.5	-1	-1

	Variable										
Soil	Mean slope (%)	Flat				Sloping					
				Low sloping				High sloping			
		0	0	6	8	13	15	25	27	90	90

rules (Table 3), either because the value of its Certainty Factor was high or because the rule was interesting from the point of view of expert knowledge. These selection criteria agree with Berzal et al. (2001, 2002).

In the study carried out, the consistency of the method for obtaining FARs by means of fuzzy data mining was confirmed with the following rule that is not included in Table 3: [*Mean minimum temperature in the coldest month* = high] → [*Mean altitude* = low] Certainty Factor = 1. The association between these two variables has universal validity.

From an analysis of the rules shown in Table 3 we conclude that we can only interpret the associations in which both the antecedent and the consequent appear with extreme linguistic labels: low or high. When the medium level appears, the interpretation is ambiguous except in the cases in which there is also a complete relationship between the three linguistic labels of the two associated variables. This is the case of *% of young olive trees* (2 to 5–8 years old) with the production variable: high *% of young olive trees* is associated with low production, medium with medium and low with high production; the trend in this case remains clear.

3.3. *User evaluation rules*

The variables that favourably affect (high production, high yield and low acidity) or unfavourably (low production, low yield and high acidity) the production and the quality of the oil (Tables 4–6) were selected. These rules may be deemed to be user evaluation rules since they indicate the suitability for olive cultivation according to the different attributes and they have been drawn up on the basis of user knowledge. This would be an evaluation method based on empirical data (Dent and Young, 1981). Nonetheless, as we mentioned before, this study is merely of an exploratory nature; the UER indicate relationships between variables that may be *obvious* (direct, that do not need interpretation, simple), *contradictory* (that contradict the actual knowledge, either empirical or scientific), *interesting* (that yield interesting information), etc. In other words, they allow trends to be discovered that mostly require specific research in order to be verified.

A first analysis of Tables 4–6 tells us that for the variables of production and quality considered (*P/Q* variables), the number of UER in which the different variables are associated with an unfavourable *P/Q* is much greater than when the *P/Q* is favourable. Even in the case of yield they do not seem to be favourable UER (Table 5). This fact might be interpreted as a “law of the minimum”, similar to Liebigs’ Minimum Law for the mineral fertilization of soils, or else considering the concept of “limiting factor” used in evaluation (Dent and Young, 1981). That is to say, the greatest influence that the different variables exercise on the *P/Q* is when they have minimum or threshold values that limit the *P/Q*. When they are

Table 3 – Fuzzy association rules used in the process for obtaining the user evaluation rules

Antecedent	Consequent		
	Olive fruit production (production)	% oil in olive fruit (yield)	Acidity of oil (acidity)
Farming area (hectares)	High → low (0.49) Medium → medium (0.62)	High → low (0.50) High → medium (0.49)	Medium → low (0.36) High → medium (0.50)
Number of olive trees	Low → high (0.40) High → medium (0.42)	High → low (0.44) High → medium (0.27)	High → low (0.32)
% *Picual* variety	Medium → low (0.67)	High → low (0.43) Medium → medium (0.44)	Medium → low (0.60) High → high (0.31)
% planting with traditional system	Low → low (0.58) Medium → medium (0.23) High → medium (0.24)	Medium → low (0.53) Low → medium (0.30)	Low → low (0.42)
% planting with modern system	High → low (0.65)	Medium → low (0.48) High → medium (0.37)	High → low (0.43)
% trees with three trunks	High → low (0.35) Medium → medium (0.43)	Medium → low (0.59) Medium → medium (0.27)	Low → high (0.41) Medium → medium (0.30)
% in planting frame of less than 10 m	Medium → low (0.60) Low → medium (0.26)	Medium → low (0.60) Medium → medium (0.30)	Medium → high (0.44)
% young olive trees (2 to 5–8 years)	High → low (0.70) Medium → medium (0.30) Low → high (0.16)	High → low (0.48) Medium → low (0.47)	High → low (0.47) Low → high (0.33)
% olive trees in full production (>5–8 years)	Low → low (0.44) Medium → medium (0.35) High → high (0.24)	Medium → low (0.66) Medium → medium (0.24)	Medium → high (0.50) High → high (0.47)
% area using conventional labour	Low → low (0.26) Medium → high (0.50)	High → low (0.50) High → medium (0.50)	High → high (0.74)
% area harvested by knocking down	Medium → medium (0.32)	Low → low (0.58) Low → medium (0.32)	Low → low (0.32)
Pruning intensity	Low → low (1.0) Medium → medium (0.24)	Low → low (0.45) Low → medium (0.28)	Low → low (0.36)
Pruning frequency (2 years/>2 years)	>2 years → low (0.50)	>2 years → low (0.60) >2 years → medium (0.50)	2 years → low (0.40) >2 years → high (0.50)
Irrigation (with/without irrigation)	With → low (0.35)	With → medium (0.20) Without → low (0.36)	With → low (0.37)
Irrigation with own water (yes/no)	Yes → low (0.28)	–	–
Phytosanitary treatments frequency (1/>1 year^{-1})	1 → low (0.41) >1 → medium (0.35)	1 → low (0.45) >1 → medium (0.35)	>1 → high (0.41)
Cost of management (€ ha^{-1})	Low → low (0.52) Medium → medium (0.30) High → medium (0.30)	High → low (0.45) High → medium (0.30)	High → high (0.39)
Cost of fertilization applied (€ ha^{-1})	Low → low (0.38) Medium → medium (0.24) High → high (0.46)	Medium → low (0.48) Low → medium (0.28)	High → high (0.62)
Cost of phytosanitary treatments (€ ha^{-1})	Medium → low (0.35) High → medium (0.33)	High → low (0.47) Medium → medium (0.28)	High → high (0.37)
Total costs (€ ha^{-1})	Low → low (0.42) Medium → medium (0.28) High → high (0.46)	High → low (0.53) Low → medium (0.28)	High → high (0.57)
Olive fruit production (kg ha^{-1})	–	High → low (0.54) Medium → medium (0.25)	High → high (0.63)
% oil in olive fruit	High → low (0.40) High → medium (0.30)	–	Medium → low (0.56) High → high (0.58)

Table 3 (*Continued*)

Antecedent	Consequent		
	Olive fruit production (production)	% oil in olive fruit (yield)	Acidity of oil (acidity)
Acidity of oil (acidity)	Low → low (0.49) Medium → medium (0.33) High → high (0.27)	Medium → low (0.56) High → low (0.47) Low → medium (0.47)	–
Soil depth	Medium → low (0.45)	Medium → low (0.50) Medium → medium (0.30)	Medium → high (0.30) High → medium (0.45)
Hardened layer depth	Medium → low (0.45) High → medium (0.51)	Low → low (0.66) High → medium (0.34)	Low → high (0.66)
Workability	Low → low (0.40) Medium → high (0.26)	Low → low (0.51) Low → medium (0.28)	Low → high (0.50)
Mean slope	Medium → low (0.42) High → medium (0.32) Low → high (0.18)	High → low (0.65) Medium → medium (0.30)	High → high (0.32)
Stoniness	High → low (0.55) Medium → medium (0.23)	Medium → low (0.41) High → medium (0.41)	High → high (0.34)
Soil texture	Sandy → low (0.51) Sandy → medium (0.30) “Balanced” → high (0.18)	Sandy → medium (0.32) “Balanced” → low (0.50)	Sandy → medium (0.39) “Balanced” → high (0.38)
Erosion (yes/no)	Yes → low (0.30)	Yes → low (0.42) Yes → medium (0.22)	Yes → high (0.34)
Soil colour	Dark grey → low (0.59) Brown → low (1) Light grey → medium (1) Light red → medium (0.50) Greyish → high (1)	Very dark brown → low (0.74) Greyish → low (1) Light grey → medium (1) Light red → medium (0.50) Greyish white → medium (0.49)	Dark grey → low (0.50) Very dark brown → high (0.50) Light red → high (0.50) Greyish → high (1)
Mean altitude (m)	High → low (0.45) Low → medium (0.26)	High → low (0.52) Medium → medium (0.25)	High → high (0.50)
Mean annual rainfall (mm year^{-1})	High → high (0.40) Medium → medium (0.31)	High → low (0.69) Medium → medium (0.29)	Low → high (0.37)
Mean minimum temperature of the coldest month (°C)	High → low (0.4)	High → low (0.54) Medium → medium (0.27)	Low → high (0.42) High → medium (0.31)

Antecedent → consequent (Certainty Factor).

not limiting factors, they seem to be indifferent and there are other variables which condition the production and the quality.

There are interesting FARs amongst the three P/Q variables (Table 3). We consider these association rules to be UER, when production is an antecedent of yield and acidity, and yield is an antecedent of acidity (Tables 5 and 6). The inverse relationships only have a predictive value, an aspect that we do not deal with in this paper. The high production of olives is associated with a low yield and high acidity. It is also noted that high yield is associated with high acidity. Therefore, when we want to optimize the cultivation to produce quality oil, this incompatibility between optimizing production, yield and acidity should be taken into account.

3.4. *User evaluation rules from management variables*

Some UER are *obvious*, as is the case of *% of young olive trees* or *% of olive trees in full production* with the production (Table 4). Others seem to be *contradictory* according to current knowledge. This is the case of the negative relationship between irrigation and production (*irrigation* with low production, Table 4) that should be positive, as described in the bibliography (Pastor et al., 1998), even more if we bear in mind the strong summer drought in the region. This contradictory association might be explained using a third variable: the Granada olive grove system is undergoing a phase of modernization, and irrigation and the rejuvenation of the olive trees are included amongst the improvements (Calero et al., 2005); therefore, the olive trees with irrigation are usually the youngest ones, which are not yet in full production.

There are also UER that reveal knowledge that is difficult to interpret, although the information that they hide is *interesting*. For example: the rule that relates high production with a low *number of olive trees* and its complementary rule that is the one that associates low production with a large *farming area* (Table 4). It should be borne in mind that the larger the *farming area*, the larger the *number of olive trees*. How should these two rules be interpreted? Does it mean that the greater the number of olive trees, the less care taken by the farmer?

Table 4 – User evaluation rules (UER) for olive fruit production (production)

Favourable (high production)	Unfavourable (low production)
Management	
Low number of olive trees	High farming area
Low % young olive trees	Low % planting with traditional system
High % olive trees in full production	High % planting with modern system
High cost of fertilization	High % olive trees with three trunks
High total cost	High % young olive trees
	Low % olive trees in full production
	Low % area using conventional labour
	Low pruning intensity
	More than two year of pruning frequency
	Irrigation
	Irrigation with own water
	Only one phytosanitary treatment by year
	Low cost of management
	Low cost of fertilization
	Low total cost
Soil	
Low mean slope	Low workability
Balanced soil texture	High stoniness
Greyish soil colour	Sandy soil texture
	Erosion
	Dark grey and brown soil colour
Environment	
High mean annual rainfall	High mean altitude
	High mean minimum temperature of the coldest month

The values for the linguistic labels, high and low for each variable, showed in Table 2.

Or perhaps it means that the small farms are concentrated on the best soils and the large ones include good and bad soils? More research would be required to provide clearer answers.

Low total or partial management costs are related to low production (Table 4). These UER are *interesting* and seem to indicate that olive cultivation in the Province of Granada is carried out on worn or marginal soils (through a long history of cultivation and intense erosion processes) and therefore the production requires financial inputs. The complementary rule (high cost of management ⇒ high production) does not appear; we may deduce that there is a minimum financial input that limits the production, but once the latter has been overcome, there is no directly proportional relationship between production and financial input and there are other variables responsible for the production.

Two UER associate high *% of Picual variety* with low yield (Table 5) and high acidity (Table 6). The two dominant varieties of olive in the Province of Granada are *Picual* and *Hojiblanca*. The *Picual variety* is expanding since it has good agronomical characteristics with an early start for production, high productivity and yield and good adaptation to different edapho-climatic conditions (Tous et al., 1998). However, according to Barranco et al. (1999) for the Province of Granada it is not clear which are the most ideal species since there is a different culture for olive production from other places and this has given rise to the current unique distribution. Our results fit in with the statements from the aforementioned authors. Is the expansion of this variety that is admired in the Province of Granada appropriate then?

There are UER that associate the tree planting technique and the final shape thereof with the *P/Q* variables. For example, the shape of olive trees with three trunks, traditionally deemed to be productive, is now associated with low production (Table 4). At the present time, the trend is to obtain trees with a single stump. This shape is deemed to be the ideal one, not just through its productivity, but also because of the ease of mechanising its collection (Barranco, 1998).

The UER that associate the variable for *% area using conventional labour* with the *P/Q* variables are *interesting*. Low *% area using conventional labour* is associated with low production (Table 4), although the Certainty Factor for this rule is low (Table 3). But a high *% area using conventional labour* with a high Certainty Factor is related to low yield and high acidity (Tables 5 and 6). It may be deduced that conventional labour is perceived by the farmer as having a negative impact on the *P/Q*. This rule matches the idea that points to conventional labour (practices widely used by olive farmers who keep the soil free from vegetation all year long by means of ongoing work) as the system that most favours the loss of soil through erosion and the splitting of roots, with the latter creating an imbalance in the growth and fructification of the tree (Barranco et al., 1999; de Luna et al., 2000).

The frequency and the intensity of pruning, are variables of handling that generate *interesting* UER. Low *pruning intensity* is associated with low production (Table 4) and a low yield (Table 5). Low intensity in pruning seems to have a negative effect on production and the yield. Nonetheless in modern olive growing, it is not recommendable in any event to carry out severe pruning that might remove a great deal of leafage

Table 5 – User evaluation rules (UER) for % oil in olive fruit (yield)

Favourable (high yield)	Unfavourable (low yield)
Management	
–	High farming area High number of olive trees High % *Picual* variety High % young olive trees High % area using conventional labour Low % area harvested by knocking down Low pruning intensity More than two year of pruning frequency Without irrigation Only one phytosanitary treatment by year High cost of management High cost of phytosanitary treatment High total cost High olive fruit production
Soil	
–	Low hardened layer depth Low workability High mean slope Balanced soil texture Erosion Very dark brown and greyish soil colour
Environment	
–	High mean altitude High mean annual rainfall High mean minimum temperature of the coldest month

The values for the linguistic labels, high and low for each variable, showed in Table 2.

and little wood (Pastor and Humanes, 1996). As regards to the acidity, the UER are *contradictory* (Table 6), since low *pruning intensity* is associated with low acidity (favourable), whereas low *pruning frequency* (more than 2 years) is associated with high acidity (unfavourable).

3.5. *User evaluation rules from soil variables*

The associations of the P/Q variables with the soil variables are difficult to interpret and it is here that the method used is more exploratory than ever. We should not forget that this method involves variables that are neither measured nor estimated, but rather they are obtained by processing uncertain and imprecise information provided by the farmers. We shall now briefly discuss some of the most interesting UER obtained.

Erosion seriously affects 57% of the land in the eastern part of Andalusia (Moreira, 1991), where the Province of Granada is located. This is a highly accentuated problem since the rate of erosion due to the climate in this area is greater. We may state that most of the olive growing lands in the Province of Granada suffer from the problem of water erosion; in fact, the types of soils described by Pérez-Pujalte (1980) in this Province correspond to degraded soils to a greater or lesser extent. When the farmers detect *erosion*, the UER demonstrate an unfavourable situation: low production, low yield and high acidity (Tables 4–6). Such an important problem as *erosion* is well understood and soon detected by the farmers and is related to the production and the quality of the oil. This fact has not been described in the bibliography.

On the other hand, high *erosion* is related to high *mean slope*, *low hardened layer depth*, low *workability*, high *stoniness* and sandy *soil texture*, characteristics that are typical of degraded or badly developed soils and which partially or fully prevent certain handling practices or the development of the roots. These unfavourable conditions for olive cultivation are also detected by the farmer and the pertinent UER are unfavourable for the *P/Q* (Tables 4–6).

Soil colour generates interesting UER, although they are difficult to interpret except for the one that associates a greyish colour with high production, low yield and high acidity, the same trend as balanced *soil texture* (Tables 4–6). According to the soil map for the Province of Granada (Pérez-Pujalte, 1980), greyish soils with a balanced texture are Regosols and Cambisols over marls and marl-limestones, which are those typically used for olive cultivation, and which abound in the areas of Montefrío, Iznalloz and Vega (López, 1982), see Fig. 1.

These results confirm, on the one hand, that the knowledge extraction method used in this paper is coherent and, on the other hand, that the farmers have wide-ranging knowledge about the soils that they work and that this knowledge could be used to complement the scant scientific knowledge that might be held about the soils.

3.6. *User evaluation rules from environmental variables*

These FARs cannot be considered strictly as UER since the data for altitude, rainfall and temperature were not obtained from the survey amongst the farmers but rather from the topographic maps and from the rainfall monitoring stations in the Province of Granada. In any event, the UER which do appear (Tables 4–6) are coherent with the ecological requirements for olive cultivation (Sys et al., 1991).

The high *mean altitude* is a limiting factor for cultivation, since it conditions low temperatures. Bearing in mind the fact that Granada is a highly mountainous Province, the high value for the *mean altitude* is frequent and is associated with low production, low yield and high acidity (Tables 4–6). Olive cultivation at high mean altitudes is also risky, in fact in many areas of the Province of Granada their crops froze in the winter of 2004–2005.

A high *mean annual rainfall*, bearing in mind the lack of rainfall in the Province, is related with high production (Table 4), whereas it is unfavourable because it is related to low yield (Table 5). On the other hand, a low *mean annual rainfall* is related to the increase in acidity (loss of olive oil quality) due to unfavourable climatic conditions (Table 6).

The *mean minimum temperature of the coldest month* is related with all three P/Q variables. When it is high it is associated with low production (Table 4) and low yield (Table 5). The low *mean minimum temperature of the coldest month* is associated with high

Table 6 – User evaluation rules (UER) for acidity of oil (acidity)

Favourable (low acidity)	Unfavourable (high acidity)
Management	
High number of olive trees	High % *Picual* variety
Low % planting with traditional system	Low % olive trees with three trunks
High % planting with modern system	Low % young olive trees
High % young olive trees	High % olive trees in full production
Low % area harvested by knocking down	High % area using conventional labour
Low pruning intensity	More than two years of pruning frequency
Two years of pruning frequency	More than one phytosanitary treatment by year
With irrigation	High cost of management
	High cost of fertilization
	High cost of phytosanitary treatments
	High total cost
	High olive fruit production
	High % oil in olive fruit
Soil	
Dark grey soil colour	Low hardened layer depth
	Low workability
	High mean slope
	High stoniness
	Balanced soil texture
	Erosion
	Very dark brown, light red, and greyish soil colour
Environment	
–	High mean altitude
	Low mean annual rainfall
	Low mean minimum temperature of the coldest month

The values for the linguistic labels, high and low for each variable, showed in Table 2.

acidity (Table 6), i.e., olive cultivation requires a cold period, so that its production will be proper, but low temperatures make it difficult to grow olives and cut down the quality of the oil. The *mean minimum temperature of the coldest month* of around $-7\,^{\circ}$C, is considered to be the temperature that defines the geographical area for olive cultivation. The importance of this quality is great since, in general, prolonged temperatures of $-10\,^{\circ}$C cause the death of large branches and even that the whole of the part above ground (Barranco et al., 1999). It is deemed that the optimum level for this temperature comes within the interval ranging between -4 and $2\,^{\circ}$C (Sys et al., 1991).

4. Final considerations and conclusions

Fig. 4 roughly summarizes the information cycle in the agricultural system studied. The new technologies and knowledge generated by research centres is transmitted over different routes and finally comes to the farmers. Many farmers do not receive this information directly, but through other farmers in what we might call an "imitation chain". Transfer of this information on growing olives presents a severe problem for extrapolation of the experimental results: from experimental farms with controlled soils, environmental conditions, agricultural practices, type of plant, etc., the information is transferred to real farms, where these parameters are different and not as controlled. Short, mid and long-term results found for this crop may be different and even contradictory to the experimental results, and must generate "information feedback" in the other direction, which permits the results of research to be evaluated. This information feedback is fundamental for the agricultural production system to be effective and sustainable, since it enables errors to be detected quickly and thereby avoid obsolete or erroneous practices, unfounded beliefs, etc., typical of the agricultural systems.

This paper proposes the use of fuzzy data mining techniques to increase and filter that information feedback using adequate processing of the data supplied by the farmers in their surveys. The method proposed allows real farms to be considered experimental plots and the farmers (users) as "researchers" whose experiences are crossed with those of other farmers to generate information feedback (user knowledge) with general validity.

Information feedback is acquired in this paper by means of fuzzy data mining expressed as user evaluation rules. The UER found allow certain agricultural practices or certain growing conditions to be qualified as favourable or unfavourable for fruit production, oil yield or acidity of oil. The set of UER could make up a land evaluation system, since the purpose is, as shown by Beek (1978) to: "predict the input, outputs and other favourable as well as adverse effects resulting from specified uses of the land that is being evaluated".

Nevertheless, the method proposed goes beyond a land evaluation system, because UER agree in many cases with those mentioned in the bibliography, with the expert rules being based on specific studies of a scientific nature. In other cases, user knowledge allows us to complete scientific knowledge, by providing aspects not yet studied in the bibliography available. Many of the UER are highly complex and are diffi-

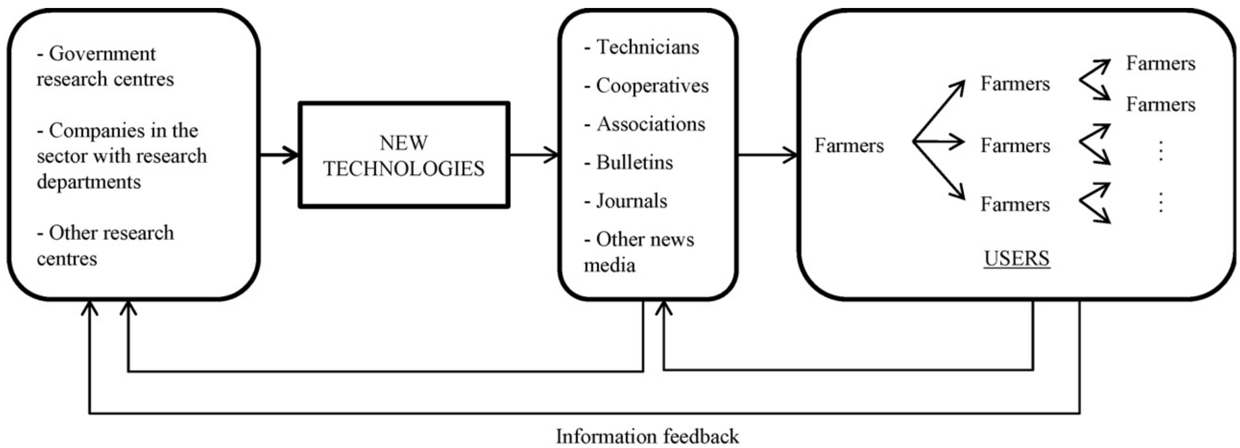


Fig. 4 – Information cycle in the agricultural system studied.

cult to interpret, due, on the one hand, to the lack of scientific information, and on the other hand, to the appearance of unexpected inconsistencies. This is why, within the rules found, a great number of hypotheses may be established, making user knowledge and fuzzy data mining techniques very effective exploratory tools complementing to scientific knowledge.

Specifically, the UER enable us to arrive at interesting conclusions in this paper with regard to olive growing in the Province of Granada. Among these conclusions we could emphasize: (1) "*Farming area*" and its complementary variable "*number of olive trees*" indicate that large farms (large number of olive trees) do not favour production. In the statistical yearbooks provide by the Andalusian Regional Government (http://www.juntadeandalucia.es/agriculturaypesca/portal/opencms/portal/portada.jsp), the tendency is to increase the size of the farm. Is this evolving in the right direction? (2) The current expansion of the *Picual variety* of olive is not justified by our results. (3) Conventional labours do not favour production or quality of the oil. They are obsolete practices sustained by tradition. (4) Soil erosion, the main environmental problem in Southeast Spain, is not only an environmental problem but is unfavourable for production and the quality of the olive oil, the main agricultural product in the area. (5) The farmers (users) have important information on the soils they cultivate and this information can be extracted and transferred to other fields (scientific or technological) by fuzzy data mining.

Finding UER by fuzzy data mining, as described in this paper, is not an isolated goal, but has the final aim of building an information system providing assistance in decision-making, to improve olive oil production and quality, and to determine the suitability of land for a sustainable culture of olive trees in Andalusia. It is intended to create an interactive system that will provide information to, and simultaneously receive information from, the different users of the system, using a flexible query procedure. The experience based on this advanced data processing methodology could be extended to olive growing in the whole Mediterranean basin and to other geographical areas around the world and even to other kinds of crops.

Acknowledgements

This work is part of the research projects 1FD97-0244-CO3-2 (financed with FEDER funds) and CGL2004-02282BTE (Spanish Ministry of Education and Science).

REFERENCES

Agrawal, R., Imielinski, T., Swami, A., 1993. Mining association rules between sets of items in large databases. In: Proceedings of the ACM SIGMOD International Conference on Management of Data, Washington, DC, USA, pp. 207–216.

Agrawal, R., Srikant, R., 1994. Fast algorithms for mining association rules. In: Proceedings of the 20th International Conference on Very Large Databases, Santiago, Chile, pp. 487–499.

Ali, A.M.S., 2003. Farmers' knowledge of soils and the sustainability of agriculture in a saline water ecosystem in Southwestern Bangladesh. Geoderma 111, 333–353.

Aranda, V., Delgado, G., Serrano, J.M., Vila, M.A., 2001. Using fuzzy relational databases to represent agricultural and environmental information. An example within the scope of olive cultivation in Granada. Mathware & Soft Computing 8, 275–289.

Barranco, D., 1998. Características y comportamiento de las variedades y patrones de olivo cultivados en España. Phytoma 102, 11–12.

Barranco, D., Fernández-Escobar, R., Rallo, L., 1999. El cultivo del olivo. Junta de Andalucía-Ediciones MundiPrensa, Madrid, Spain.

Barrera-Bassols, N., Zinck, J.A., Van Ranst, E., 2006. Symbolism, knowledge and management of soil and land resources in indigenous communities: ethnopedology at global, regional and local scales. Catena 65, 118–137.

Beckford, C., Barker, D., 2007. The role and value of local knowledge in Jamaican agriculture: adaptation and change in small-scale farming. The Geographical Journal 173, 118–128.

Beek, K.J., 1978. Land Evaluation for Agricultural Development. International Institute for Land Reclamation and Improvement/ILRI, Wageningen, The Netherlands.

Berzal, F., Blanco, I., Sánchez, D., Vila, M.A., 2001. A new framework to assess association rules. In: Hoffmann, F. (Ed.), Advances in Intelligent Data Analysis. Fourth International

Symposium (IDA'01), Lecture Notes in Computer Science 2189. Springer-Verlag, pp. 95–104.
Berzal, F., Blanco, I., Sánchez, D., Vila, M.A., 2002. Measuring the accuracy and interest of association rules: a new framework. Intelligent Data Analysis 6, 221–235.
Berzal, F., Calero, J., Cubero, J.C., Delgado, G., Sánchez, D., Serrano, J.M., Vila, M.A., 2004. A fuzzy data mining application over user knowledge involving fuzzy resemblance relations. In: Proceedings of Alternative Techniques in Data Mining and Knowledge Discovery Workshop, Fourth International Conference on Data Mining (ICDM'2004), Brighton, UK, pp. 1–8.
Calero, J., Delgado, G., Sánchez-Marañón, M., Sánchez, D., Serrano, J.M., Vila, M.A., 2003. Helping user to discover association rules. A case in soil colour as aggregation of other soil properties. In: Proceedings of the 5th International Conference on Enterprise Information Systems, ICEIS 2003, Angers, France, pp. 533–540.
Calero, J., Serrano, J.M., Aranda, V., Sánchez, D., Vila, M.A., Delgado, G., 2005. Analysis and characterization of olive tree cultivation system in Granada province (South of Spain) with optimal scaling and multivariate techniques. Agrochimica XLIX, 118–131.
Center, B., Verma, B.P., 1998. Fuzzy logic for biological and agricultural systems. Artificial Intelligence Review 12, 213–225.
Codd, E.F., 1970. A relational model of data for large shared data banks. Communications of the ACM 13, 377–387.
de Graaff, J., Eppink, L.A.A.J., 1999. Olive oil production and soil conservation in southern Spain, in relation to EU subsidy policies. Land Use Policy 16, 259–267.
de Luna, E., Laguna, A., Giráldez, J.V., 2000. The role of olive trees in rainfall erosivity and runoff and sediment yield in the soil beneath. Hydrology and Earth System Sciences 4, 141–153.
de la Torre, M.L., Grande, J.A., Aroba, J., Andújar, J.M., 2005. Optimization of fertirrigation efficiency in strawberry crops by application of fuzzy logic techniques. Journal of Environmental Monitoring 7, 1085–1092.
Delgado, M., Marín, N., Sánchez, D., Vila, M.A., 2003. Fuzzy association rules: general model and applications. IEEE Transactions on Fuzzy Systems 11, 214–225.
Dent, D., Young, A., 1981. Soil survey and land evaluation. Allen & Unwin (Ed.), London.
Dubes, R., Jain, A., 1988. Algorithms that cluster data. Prentice-Hall, Englewood Cliffs, NJ.
Dubois, D., Prade, H., 1988. Possibility Theory. An Approach to Computerized Processing of Uncertainty. Plenum Press, N.Y.
Ferraro, D.O., Ghersa, C.M., Sznaider, G.A., 2003. Evaluation of environmental impact indicators using fuzzy logic to assess the mixed cropping systems of the Inland Pampa, Argentina. Agriculture, Ecosystems and Environment 96, 1–18.
Galindo, J., Medina, J.M., Pons, O., Cubero, J.C., 1998. A server for fuzzy SQL queries. In: Andreasen, T., Christiansen, H., Larsen, H.L. (Eds.), Flexible Query Answering Systems. Lecture Notes in Artificial Intelligence (LNAI) 1495. Springer, pp. 164–174.
Grobben, P., 1992. Using farmers̆ Knowledge about soil. A case of a farming system research (FSR) project in Java. AT-Source 3/92, Wageningen.
Habarurema, E., Steiner, K.G., 1997. Soil suitability classification by farmers in southern Rwanda. Geoderma 75, 75–87.
He, X., Xu, Y., Zhang, X., 2007. Traditional farming system for soil conservation on slope farmland in southwestern China. Soil & Tillage Research 94, 193–200.
Hoshi, T., Sasaki, T., Tsutsui, H., Watanabe, T., Tagawa, F., 2000. A daily harvest prediction model of cherry tomatoes by mining from past averaging data and using topological case-based modelling. Computers and Electronics in Agriculture 29, 149–160.
Hussain, F., Liu, H., Tan, C.L., Dash, M., 1999. Discretization: An Enabling Technique. Technical Report. National University of Singapore (NUS), Singapore.
Kawano, S., Huynh, V.H., Ryoke, M., Nakamori, Y., 2005. A context-dependent knowledge model for evaluation of regional environment. Environmental Modelling & Software 20, 343–352.
López de Mántaras, R., Valverde, L., 1988. New results in fuzzy clustering based on the concept of indistinguishability relation. IEEE Transactions on Pattern Analysis and Machine Intelligence 10, 754–757.
López, A., 1982. Las comarcas olivareras andaluzas. INIA, Ministerio de Agricultura, Pesca y Alimentación, Madrid, Spain.
Loring-Miró, J., 1999. II anuario agrario de la provincia de Granada. ASAJA-Granada, Córdoba, Spain.
Loumou, A., Giourga, C., 2003. Olive groves: the life and identity of the Mediterranean. Agriculture and Human Values 20, 87–95.
McBratney, A.B., Odeh, I.O.A., Bishop, T.F.A., Dunbar, M.S., Shatar, T.M., 2000. An overview of pedometric techniques for use in soil survey. Geoderma 97, 293–327.
Medina, J.M., Pons, O., Vila, M.A., 1994. GEFRED: a generalized model for fuzzy relational databases. Information Sciences 77, 87–109.
Medina, J.M., Vila, M.A., Pons, O., Cubero, J.C., 1995. Towards the implementation of a generalized fuzzy relational database model. Fuzzy Sets and Systems 75, 273–289.
Moreira, J.M., 1991. Capacidad de uso y erosión de los suelos. Agencia de Medio Ambiente. Junta de Andalucía, Sevilla, Spain.
Niemeijer, D., Mazzucato, V., 2003. Moving beyond indigenous soil taxonomies: local theories of soils for sustainable development. Geoderma 111, 403–424.
Pastor, M., Humanes, J., 1996. Poda del olivo. Moderna agricultura. Editorial Agrícola Española S.A, Madrid, Spain.
Pastor, M., Humanes, J., Vega, V., Castro, J., 1998. Diseño y manejo de plantaciones de olivar. Junta de Andalucía (C.A.P.), Monografías 22/98, Sevilla, Spain.
Pérez-Pujalte, A., 1980. Mapa de Suelos de la provincia de Granada, E 1: 200000. Estación Experimental del Zaidín-CSIC, Granada, Spain.
Petridis, V., Karbulasos, V.G., 2003. FINkNN: a fuzzy interval number k-nearest neighbour classifier for prediction of sugar production from populations of samples. Journal of Machine Learning Research 4, 17–37.
Rallo, L., 1998. El olivar y la innovación tecnológica. Phytoma 102, 6–8.
Rundensteiner, E.A., Hawkes, L.W., Bandler, W., 1989. On nearness measures in fuzzy relational data models. International Journal of Approximate Reasoning 6, 267–298.
Saito, K., Linquist, B., Keobualapha, B., Shiraiwa, T., Horie, T., 2006. Farmers' knowledge of soils in relation to cropping practices: a case study of farmers in upland rice based slash-and-burn systems of northern Laos. Geoderma 136, 64–74.
Sánchez, D., Vila, M.A., Sánchez, J.R., Serrano, J.M., 2004. Association rules over imprecise domains applying fuzzy similarity relations. In: XII Spanish Conference of Fuzzy Logic and Technology (ESTYLF'2004), Jaén, Spain, pp. 129–134.
Serrano, J.M., Aranda, V., Calero, J., Delgado, G., Sánchez, D., Vila, M.A., 2002. Flexible land classification for olive cultivation using user knowledge. In: First International NAISO Congress on Neuro-Fuzzy Technologies (NF'2002), Habana, Cuba, p. 91.
Shortliffe, E., Buchanan, B., 1975. A model of inexact reasoning in medicine. Mathematical Biosciences 23, 351–379.

Stocking, M., Murnaghan, N., 2001. Handbook for the Field Assessment of Land Degradation. Earthscan Pubications Ltd., London-Sterling VA.
Sys, I.C., van Ranst, E., Debaeye, I.J., 1991. Land Evaluation. Agricultural Publications 7. University Ghent, Belgium.
Tellaeche, A., BurgosArtizzu, X.P., Pajares, G., Ribeiro, A., Fernández-Quintanilla, C., 2008. A new vision-based approach to differential spraying in precision agriculture. Computers and Electronics in Agriculture 60, 144–155.
Thrupp, L.A., 1989. Legitimising local knowledge: form displacement to empowerment for third world people. Agricultural and Human Values 6, 13–24.
Tous, J., Romero, A., Plana, J., 1998. Comportamiento agronómico y comercial de cinco variedades de olivo en Tarragona. Investigación Agraria (INIA) 13, 97–109.
Zadeh, L.A., 1978. Fuzzy sets as a basis for a theory of possibility. Fuzzy Sets and Systems 1, 3–28.