International Journal of Research in Advent Technology (IJRAT) Special Issue, January 2019
E-ISSN: 2321-9637
Available online at www.ijrat.org
International Conference on Applied Mathematics and Bio-Inspired Computations
10th & 11th January 2019

Plant Association of Bamboo in Pachaimalai Hills with Mining Fuzzy Bio-Statistical Rules

John Robinson P[1], Delphin Sonia M[2], Sebastian Rajasekaran[2]
1Department of Mathematics, Bishop Heber College, Trichy
[2]Department of Botany, Bishop Heber College, Trichy
robijohnsharon@gmail.com

Abstract: Forests provide the basic life support system to all the living entities of mother earth including mankind. Pachaimalai hills are known for its thick vegetation of sandalwood from the past, distributed throughout the hill area, and a manmade forest area comprising Bamboo cultivation. Association of different plant species in Pachaimalai hills with Bamboo will be the main focus of this research work. The integration of association rules and correlation rules with fuzzy logic can produce more abstract and flexible patterns for many real life problems, since many quantitative features in real world are fuzzy in nature. We present an algorithm for mining fuzzy association and correlation rules. The proposed mining algorithm is utilized for seeking close associations and relationships between a group of plant species clustering around Bamboo in Pachaimalai hills, Eastern Ghats, Tamilnadu.

Keywords: Pachaimalai Hills, Plant Association, Data Mining, Fuzzy sets, Bio-Statistics

1. INTRODUCTION

Plant association is defined as the grouping of plant species, or a plant community, that recurs across the landscape. Plant associations are used as indicators of environmental conditions such as temperature, moisture, light etc. It can be viewed as a collection of plant species within a designated geographical unit, which forms a relatively uniform patch, distinguishable from neighboring patches of different vegetation types. Frequent patterns are patterns that appear in a data set frequently. Finding such frequent patterns ([1], [2]) plays an essential role in mining associations, correlations, and many other interesting relationships among data. Frequent item set mining leads to the discovery of associations and correlations among items in large transactional or relational data sets ([6], [7]). Thorny bamboo (*Bambusa bambos*) is a species described in 1678, pre-Linnaeus time. Giant and dwarf ecotype are found in species. It is also known by the name of *B. arundinacea*, thus its nomenclature is complex. Present correct name is *B. bambos*. It has a gregarious flowering period of 43-49 years and high frequency reproducible in vitro regeneration protocol of a multipurpose Bamboo species Dendrocalamus hamiltonii Nees at Arn. As most of the time Bamboo seeds are not available, bamboos can be propagated vegetatively ([3], [4]). Bhol & Parida [5] studied on the effect of planting alignment and cutting size on propagation of *Bambusa vulgaris and the* Influence of growth regulators on propagation of culm- and branch cuttings of *Bambusa vulgaris*. Earlier in one of our works ([7]) we investigated the close association of some important species with sandalwood in Pachaimalai hills. In this work, the close plant association of Bamboo with other species in Pachaimalai hills, Eastern Ghats, Tamilnadu is studied.

2. MATERIALS AND METHODS

Study Area: The present study was conducted in a place called Keelakarai village in Pachaimalai hills which are a green hill range just 80 kms north of Tiruchirappalli via Thuraiyur, South India. The altitude of study is about 100 to 200metres where the plantations of Bamboo are promoted by the forest department, Tamilnadu. The hills is spread over an area of 13,500 square km and only the Bamboo plantation in the Keelakarai village were covered for the study.

Data Collection: The distribution of bamboo plantation in the foot of the hills is mostly even or predictable, and hence surveying the frequency of plant associations with bamboo is a not a tedious job as in the case of sandalwood. Hence we identified fifty potential spots in the region of bamboo plantations in the hills where the distribution of the same is considerably notable and the surveyed frequency data is converted into fuzzy numbers due to the irregular distribution of the plant communities. Fuzzy membership functions are used for the purpose of data extraction. Four different plant species are found to be closely associated to bamboo in Pachaimalai hills and they seem to recur in all the surveyed spots.

3. BASIC CONCEPTS OF DATA MINING AND BIO-STATISTICAL TOOLS

Data mining refers to extracting or "mining" knowledge from large amounts of data.

Frequent Fuzzy Item-Sets, Closed Fuzzy Item-Sets and Fuzzy Association Rules

The fuzzy item-sets which frequently occur together in large databases are found using fuzzy association rules. The fuzzy support and fuzzy confidence are used to identify the fuzzy association rules. Let $F=\{f_1, f_2,..., f_m\}$ be a set of fuzzy items, $T=\{t_1,t_2,...,t_n\}$ be a set of fuzzy records, and each fuzzy record t_i is represented as a vector with m values, $(f_1(t_i), f_2(t_i),..., f_m(t_i))$, where $f_j(t_i)$ is the degree that f_j appears in record t_i, $f_j(t_i)\in[0,1]$. Then a fuzzy association rule is defined as an implication form such as $F_X \Rightarrow F_Y$, where $F_X \subset F, F_Y \subset F$ are two fuzzy item-sets.

The fuzzy support and fuzzy confidence are given as follows:

$$fsupp(\{F_X, F_Y\}) = \frac{\sum_{i=1}^{n} \min(f_j(t_i) / f_j \in \{F_X, F_Y\})}{n}$$

$$fconf(F_X \Rightarrow F_Y) = \frac{fsupp(\{F_X, F_Y\})}{fsupp(\{F_X\})}$$

If the $fsupp(\{F_X, F_Y\})$ is greater than or equal to a predefined threshold, minimal fuzzy support (s_f), and the $fconf(F_X \Rightarrow F_Y)$ is also greater than or equal to a predefined threshold, minimum fuzzy confidence (c_f), then $F_X \Rightarrow F_Y$ is considered as an interesting fuzzy association rule.

Fuzzy Set: If X is a collection of objects denoted generically by x, then a fuzzy set Ã in X is a set of ordered pairs: $\tilde{A}=\{(x, \mu_{\tilde{A}}(x)) | x \in X\}$

$\mu_{\tilde{A}}(x)$ is called the membership function or grade of membership of x in Ã that maps X to the membership space M.

4. APPLICATION OF BIO-STATISTICAL METHODS AND DATA MINING FOR PLANT ASSOCIATION

The concepts of mining fuzzy correlation rules are mentioned in the following.

Mining Fuzzy Correlation Rules

Suppose there are two fuzzy itemsets A , B $\subset$ F , where F is a fuzzy space. A and B are defined on a crisp universal set X with membership functions μ_A and μ_B, and the fuzzy itemsets A and B can be expressed as follows:

$$A=(x, \mu_A(x)) | x \in X), \; B=(x, \mu_B(x)) | x \in X)$$

where $\mu_A, \mu_B \in [0,1]$. Assume that there is a random sample $(x_1, x_2,..., x_n) \in X$, alone with a sequence of paired data, $\{(x_i, \mu_A(x_i), \mu_B(x_i)) | i=1...n\}$, which correspond to the grades of the membership functions of fuzzy itemsets A and B defined on X. Then, the fuzzy correlation coefficient between the fuzzy itemsets A and B, $r_{A,B}$, is:

$$r_{A,B}=\frac{S_{A,B}}{S_A \cdot S_B}, S_{A,B}=\frac{\sum_{i=1}^{n}\left(\mu_A(x_i)-\overline{\mu_A}\right)\cdot\left(\mu_B(x_i)-\overline{\mu_B}\right)}{n-1},$$

$$\overline{\mu_A}=\frac{\sum_{i=1}^{n}\mu_A(x_i)}{n}, \; \overline{\mu_B}=\frac{\sum_{i=1}^{n}\mu_B(x_i)}{n},$$

$$S_A^2=\frac{\sum_{i=1}^{n}\left(\mu_A(x_i)-\overline{\mu_A}\right)^2}{n-1}, \quad S_B^2=\frac{\sum_{i=1}^{n}\left(\mu_B(x_i)-\overline{\mu_B}\right)^2}{n-1},$$

$$S_A=\sqrt{S_A^2}, \qquad S_B=\sqrt{S_B^2}.$$

5. NUMERICAL ILLUSTRATION

An experiment will be displayed in this section. Assume that $T=\{t_1,t_2,t_3,t_4,t_5,t_6,t_7,t_8,t_9,t_{10},t_{11},t_{12}\}$ is a random sample with 12 fuzzy records (of the four plant species clustered around Bamboo) shown in Table 1, and $F=\{f_1, f_2, f_3, f_4, f_5\}$ is the set of observed fuzzy items. Here s_f is set to 0.30; c_f is set to 0.80; r_f is set to 0.20; α is set to 0.1, and thus $t_{0.9,10}$ is equal to 1.372.

f_1 – Bamboo:

Kingdom	: Plantae
(unranked)	: Angiosperms
(unranked)	: Monocots
(unranked)	: Commelinids
Order	: Poales
Family	: Poaceae
Sub-family	: Bambusoideae

f_2 – Drypetes sepiaria:

Kingdom	: Plantae
(unranked)	: Angiosperms
(unranked)	: Eudicots
(unranked)	: Rosids
Order	: Malpighiales
Family	: Putranjivaceae
Genus	: *Drypetes*
Species	: ***D. sepiaria***

f_3 – Citrus sp:

Kingdom	: Plantae
(unranked)	: Angiosperms
(unranked)	: Eudicots
(unranked)	: Rosids
Order	: Sapindales
Family	: Rutaceae
Subfamily	: Aurantioideae
Tribe	: Citreae
Genus	: ***Citrus***L.

f_4 – Acacia pennata:

Kingdom	: Plantae
Phylum	: Magnoliophyta
Class	: Magnoliatae
Order	: Rosales
Family	: Mimosaceae
Genus	: ***Acacia***

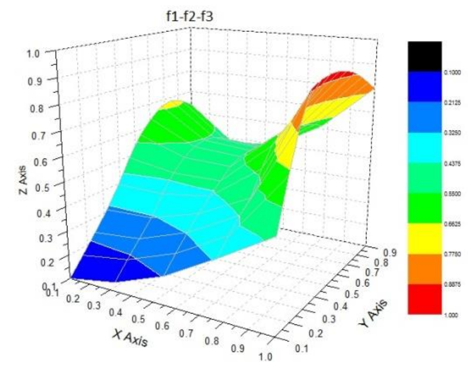
f_5 – Atalantia monophylla:

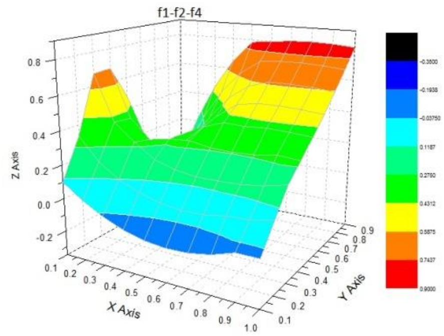
Familia	: Rutaceae
Subfamilia	: Citroideae
Tribus	: Clauseneae
Genus	: *Atalantia*
Species	: ***Atalantia monophylla***

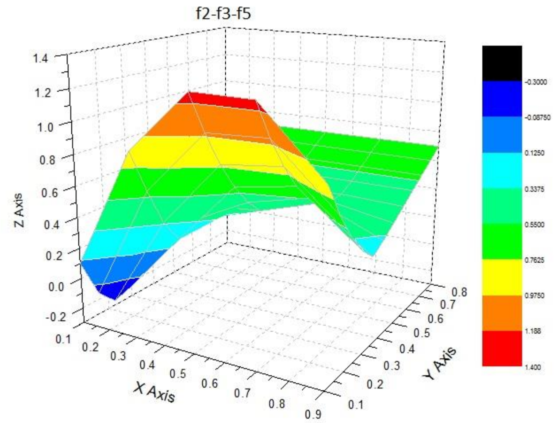
In this study, we observe five plant species occurring frequently almost in all transactions of a botanical survey conducted in Pachaimalai hills, Eastern Ghats for analyzing plant associations for Bamboo.

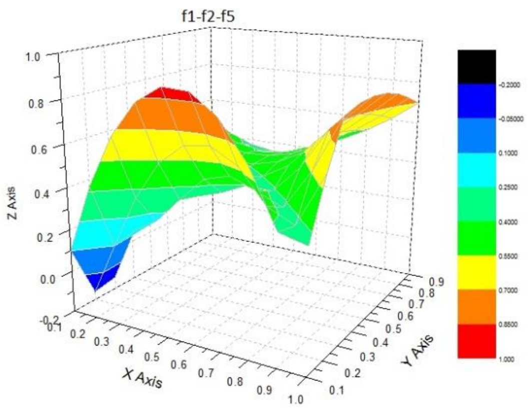
Table 1: A random sample with 12 fuzzy records.

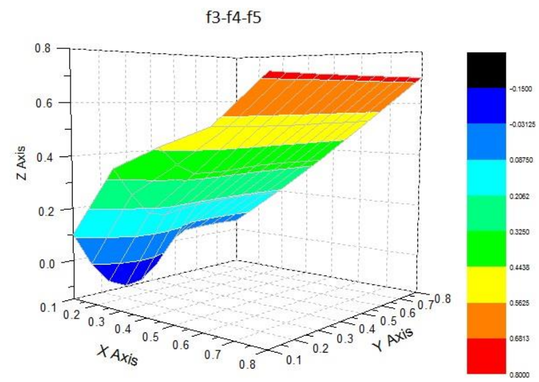
T \ F	f_1	f_2	f_3	f_4	f_5
t_1	0.4	0.4	0.4	0.3	0.3
t_2	0.2	0.2	0.2	0.2	0.1
t_3	0.6	0.5	0.5	0.4	0.4
t_4	0.7	0.7	0.6	0.6	0.5
t_5	1	0.9	0.8	0.8	0.7
t_6	0.9	0.8	0.7	0.7	0.6
t_7	0.3	0.3	0.3	0.3	0.2
t_8	0.9	0.8	0.8	0.7	0.7
t_9	0.7	0.6	0.6	0.5	0.5
t_{10}	0.1	0.1	0.1	0.1	0.1
t_{11}	0.6	0.5	0.5	0.4	0.4
t_{12}	0.7	0.7	0.6	0.6	0.5

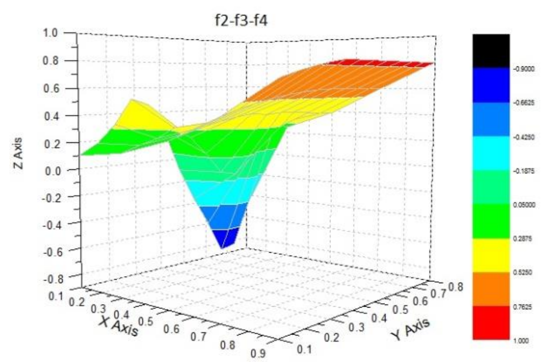


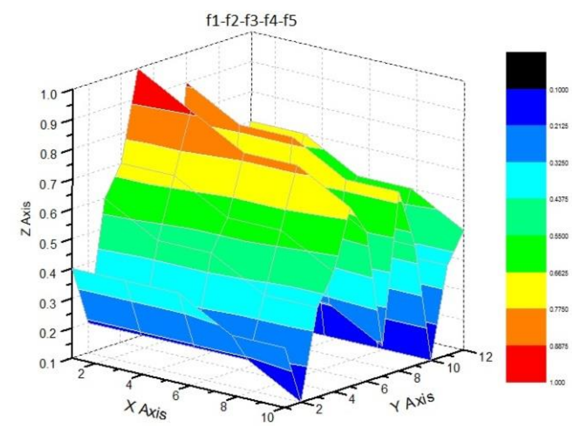












First, the fuzzy support of each fuzzy item of F is computed below and listed in Table 2. Because all $f\text{supp}(f_i)$, are greater than s_f, the set of the frequent fuzzy itemsets whose size is equal to 1 is $L_1 = \{f_1, f_2, f_3, f_4, f_5\}$.

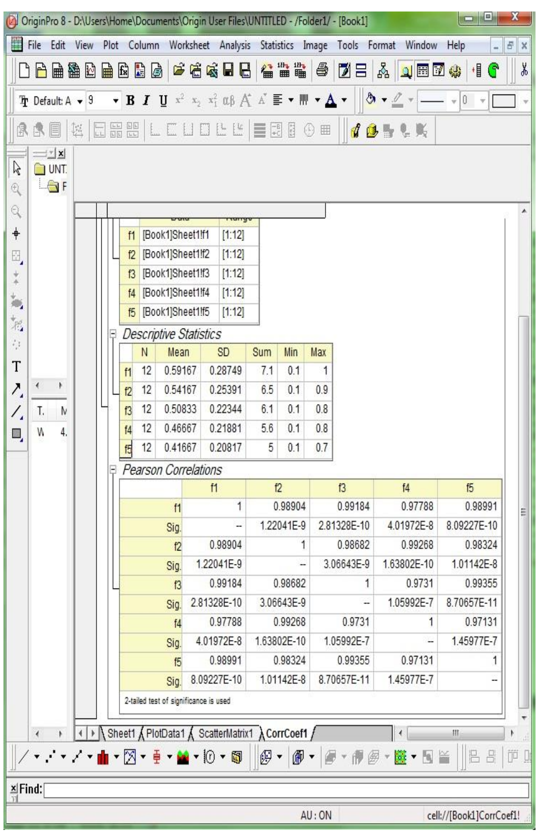
Table 2: The fuzzy support of each fuzzy item of F.

F	Fsupp
f_1	0.59
f_2	0.54
f_3	0.5
f_4	0.46
f_5	0.42

Next C_2, the set of all combinations of two elements of L_1, is generated by L_1 joint with L_1.

$$C_2 = \begin{Bmatrix} (f_1, f_2)(f_1, f_3)(f_1, f_4)(f_1, f_5)(f_2, f_3) \\ (f_2, f_4)(f_2, f_5)(f_3, f_4)(f_3, f_5)(f_4, f_5) \end{Bmatrix}.$$

Table 3: Pearson's Correlation coefficient for 5 variables



	Data	Range
f1	[Book1]Sheet1!f1	[1:12]
f2	[Book1]Sheet1!f2	[1:12]
f3	[Book1]Sheet1!f3	[1:12]
f4	[Book1]Sheet1!f4	[1:12]
f5	[Book1]Sheet1!f5	[1:12]

Descriptive Statistics

	N	Mean	SD	Sum	Min	Max
f1	12	0.59167	0.28749	7.1	0.1	1
f2	12	0.54167	0.25391	6.5	0.1	0.9
f3	12	0.50833	0.22344	6.1	0.1	0.8
f4	12	0.46667	0.21881	5.6	0.1	0.8
f5	12	0.41667	0.20817	5	0.1	0.7

Pearson Correlations

	f1	f2	f3	f4	f5
f1	1	0.98904	0.99184	0.97788	0.98991
Sig.	--	1.22041E-9	2.81328E-10	4.01972E-8	8.09227E-10
f2	0.98904	1	0.98682	0.99268	0.98324
Sig.	1.22041E-9	--	3.06643E-9	1.63802E-10	1.01142E-8
f3	0.99184	0.98682	1	0.9731	0.99355
Sig.	2.81328E-10	3.06643E-9	--	1.05992E-7	8.70657E-11
f4	0.97788	0.99268	0.9731	1	0.97131
Sig.	4.01972E-8	1.63802E-10	1.05992E-7	--	1.45977E-7
f5	0.98991	0.98324	0.99355	0.97131	1
Sig.	8.09227E-10	1.01142E-8	8.70657E-11	1.45977E-7	--

2-tailed test of significance is used

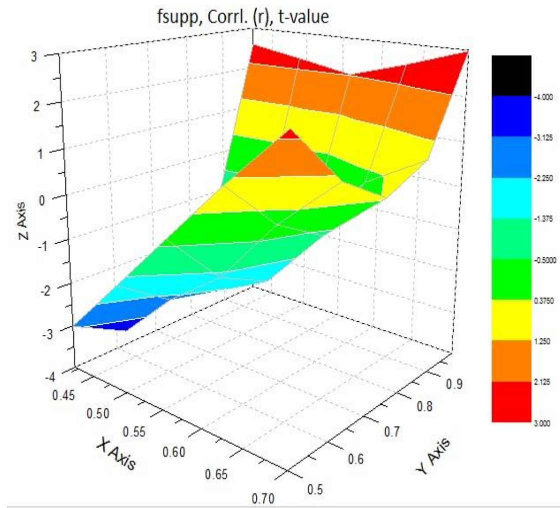
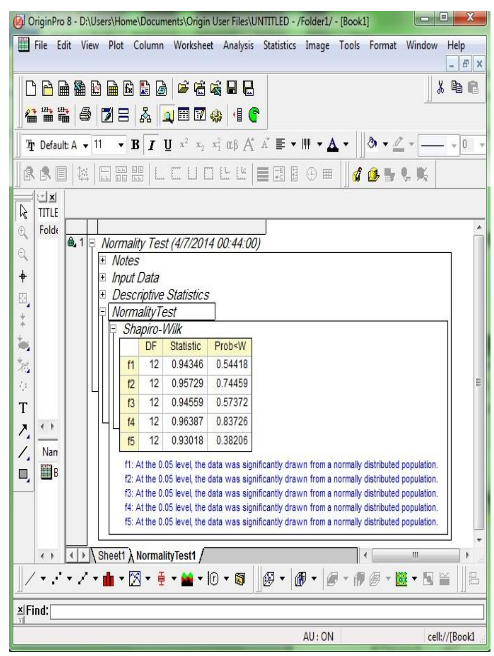


Table 4: Normality Test for 5 variables



Normality Test (4/7/2014 00:44:00)

Shapiro-Wilk

	DF	Statistic	Prob<W
f1	12	0.94346	0.54418
f2	12	0.95729	0.74459
f3	12	0.94559	0.57372
f4	12	0.96387	0.83726
f5	12	0.93018	0.38206

f1: At the 0.05 level, the data was significantly drawn from a normally distributed population.
f2: At the 0.05 level, the data was significantly drawn from a normally distributed population.
f3: At the 0.05 level, the data was significantly drawn from a normally distributed population.
f4: At the 0.05 level, the data was significantly drawn from a normally distributed population.
f5: At the 0.05 level, the data was significantly drawn from a normally distributed population.

Table 5: The fuzzy support, fuzzy correlation coefficient and t value of testing the fuzzy correlation coefficient of each element of C_2.

C_2	fsupp	r	t-value
({f_1},{f_2})	0.54	0.9890	1.9017
({f_1},{f_3})	0.5	0.9907	2.1093
({f_1},{f_4})	0.46	0.9571	0.6227
({f_1},{f_5})	0.42	0.9899	2.0067
({f_2},{f_3})	0.51	0.9859	1.6208
({f_2},{f_4})	0.47	0.9928	2.4550
({f_2},{f_5})	0.42	0.9829	1.4219
({f_3},{f_4})	0.47	0.9570	0.6223
({f_3},{f_5})	0.42	0.9935	2.5900
({f_4},{f_5})	0.42	0.9693	0.8907

In Table 6, an element whose fsupp is greater than or equal to $S_f(0.50)$ and t value is greater than or equal to $t_{0.9,10}$ (1.372) is considered an element of L_2. Thus, L_2 = {({f_1},{f_2}),({f_1},{f_3})({f_2},{f_3})}. When L_2 is obtained, C_3 is generated by L_2 joint with L_2.
C_3={ ({f_1},{f_2,f_3}), ({f_2},{f_1,f_3}), ({f_3},{f_1,f_2}) }.

Table 6: The frequent fuzzy elements of C_3.

F / T	{f_1}	{f_2,f_3}	{f_2}	{f_1,f_3}	{f_3}	{f_1,f_2}
t_1	0.4	0.4	0.4	0.4	0.4	0.4
t_2	0.2	0.2	0.2	0.2	0.2	0.2
t_3	0.6	0.5	0.5	0.5	0.5	0.5
t_4	0.7	0.6	0.7	0.6	0.6	0.7
t_5	1	0.8	0.9	0.8	0.8	0.9
t_6	0.9	0.7	0.8	0.7	0.7	0.8
t_7	0.3	0.3	0.3	0.3	0.3	0.3
t_8	0.9	0.8	0.8	0.8	0.8	0.8
t_9	0.7	0.6	0.6	0.6	0.6	0.6
t_{10}	0.1	0.1	0.1	0.1	0.1	0.1
t_{11}	0.6	0.5	0.5	0.5	0.5	0.5
t_{12}	0.7	0.6	0.7	0.6	0.6	0.7

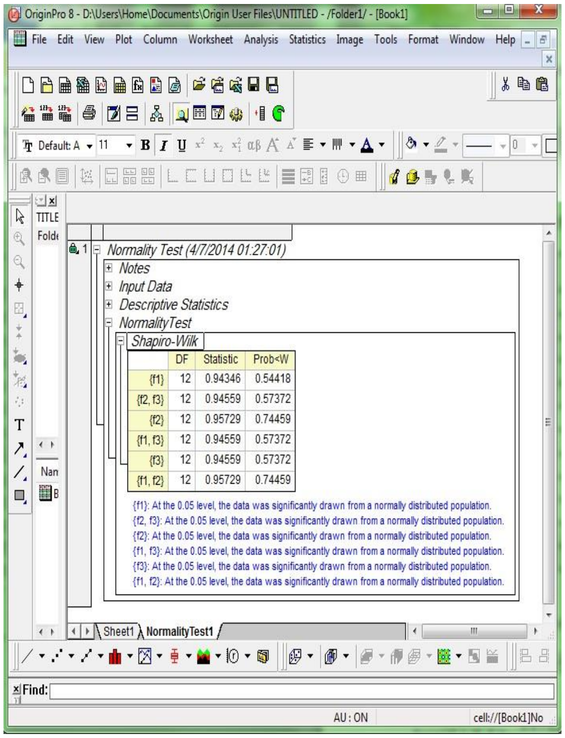
For each element of C_3, the value of t testing the fuzzy correlation is computed as:

$$t_{\{f_1\},\{f_2,f_3\}} = \frac{r_{\{f_1\},\{f_2,f_3\}} - r_f}{\sqrt{\frac{1-r^2_{\{f_1\},\{f_2,f_3\}}}{n-2}}} = \frac{0.9922-0.90}{\sqrt{\frac{1-0.9845}{10}}} = 2.3401$$

$$t_{\{f_2\},\{f_1,f_3\}} = \frac{r_{\{f_2\},\{f_1,f_3\}} - r_f}{\sqrt{\frac{1-r^2_{\{f_2\},\{f_1,f_3\}}}{n-2}}} = \frac{0.9877-0.90}{\sqrt{\frac{1-0.9756}{10}}} = 1.7753$$

$$t_{\{f_3\},\{f_1,f_2\}} = \frac{r_{\{f_3\},\{f_1,f_2\}} - r_f}{\sqrt{\frac{1-r^2_{\{f_3\},\{f_1,f_2\}}}{n-2}}} = \frac{0.9877-0.90}{\sqrt{\frac{1-0.9756}{10}}} = 1.7753$$

Table 7: Normality Test for frequent 3 itemsets



Normality Test (4/7/2014 01:27:01)

Notes

Input Data

Descriptive Statistics

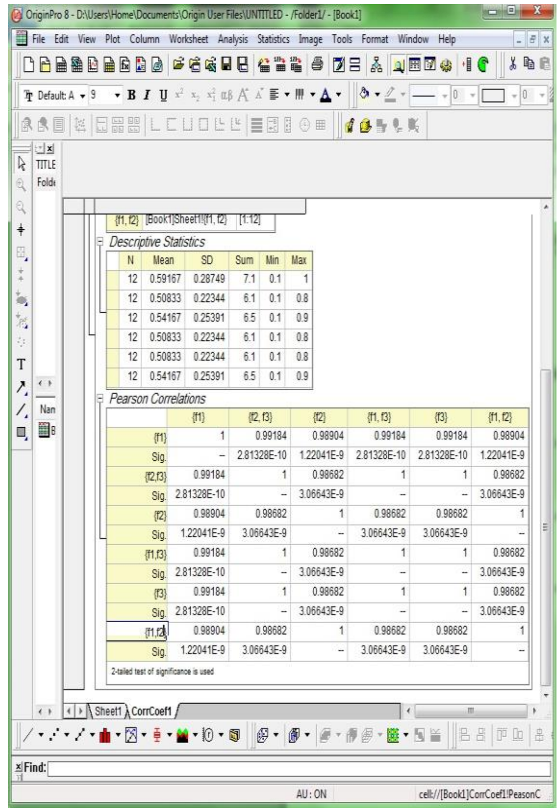
NormalityTest

Shapiro-Wilk

	DF	Statistic	Prob<W
{f1}	12	0.94346	0.54418
{f2, f3}	12	0.94559	0.57372
{f2}	12	0.95729	0.74459
{f1, f3}	12	0.94559	0.57372
{f3}	12	0.94559	0.57372
{f1, f2}	12	0.95729	0.74459

{f1}: At the 0.05 level, the data was significantly drawn from a normally distributed population.
{f2, f3}: At the 0.05 level, the data was significantly drawn from a normally distributed population.
{f2}: At the 0.05 level, the data was significantly drawn from a normally distributed population.
{f1, f3}: At the 0.05 level, the data was significantly drawn from a normally distributed population.
{f3}: At the 0.05 level, the data was significantly drawn from a normally distributed population.
{f1, f2}: At the 0.05 level, the data was significantly drawn from a normally distributed population.

Table 8: Correlation coefficient between the frequent 3 fuzzy itemsets



{f1, f2}	[Book1]Sheet1!{f1, f2}	[1:12]

Descriptive Statistics

	N	Mean	SD	Sum	Min	Max
	12	0.59167	0.28749	7.1	0.1	1
	12	0.50833	0.22344	6.1	0.1	0.8
	12	0.54167	0.25391	6.5	0.1	0.9
	12	0.50833	0.22344	6.1	0.1	0.8
	12	0.50833	0.22344	6.1	0.1	0.8
	12	0.54167	0.25391	6.5	0.1	0.9

Pearson Correlations

	{f1}	{f2, f3}	{f2}	{f1, f3}	{f3}	{f1, f2}
{f1}	1	0.99184	0.98904	0.99184	0.99184	0.98904
Sig.	--	2.81328E-10	1.22041E-9	2.81328E-10	2.81328E-10	1.22041E-9
{f2,f3}	0.99184	1	0.98682	1	1	0.98682
Sig.	2.81328E-10	--	3.06643E-9	--	--	3.06643E-9
{f2}	0.98904	0.98682	1	0.98682	0.98682	1
Sig.	1.22041E-9	3.06643E-9	--	3.06643E-9	3.06643E-9	--
{f1,f3}	0.99184	1	0.98682	1	1	0.98682
Sig.	2.81328E-10	--	3.06643E-9	--	--	3.06643E-9
{f3}	0.99184	1	0.98682	1	1	0.98682
Sig.	2.81328E-10	--	3.06643E-9	--	--	3.06643E-9
{f1,f2}	0.98904	0.98682	1	0.98682	0.98682	1
Sig.	1.22041E-9	3.06643E-9	--	3.06643E-9	3.06643E-9	--

2-tailed test of significance is used

Table 9: The fuzzy support, fuzzy correlation coefficient and t value of testing the fuzzy correlation coefficient of each element of C_3.

C_3	Fsupp	r	t-value
$(\{f_1\},\{f_2,f_3\})$	0.50	0.9877	1.7753
$(\{f_2\},\{f_1,f_3\})$	0.50	0.9877	1.7753
$(\{f_3\},\{f_1,f_2\})$	0.50	0.9922	2.3401

In Table 9, because of all elements of C_3 satisfy S_f and $t_{0.9,10}$, all elements of C_3 are elements of L_3. Thus $C_3 = L_3$.

No next C_4 can be generated by L_3 joint with L_3, so the mining procedure stops here. By using the elements of L_2 and L_3, 12 candidate fuzzy confidence of the candidate fuzzy correlation rules can be generated using $fconf\left(F_x \to F_y\right) = \frac{\text{fsupp}\left(\{F_x, F_y\}\right)}{\text{fsupp}\left(\{F_x\}\right)}$ and listed in the following, and only those fuzzy confidences that are greater than or equal to C_f (0.80) are only listed.

$$\{f_1\} \to \{f_2\},\ \{f_1\} \to \{f_3\},\ \{f_3\} \to \{f_1\}$$
$$\{f_3\} \to \{f_2\},\ \{f_1\} \to \{f_2, f_3\},\ \{f_2, f_3\} \to \{f_1\}$$
$$\{f_2\} \to \{f_1, f_3\},\ \{f_3\} \to \{f_1, f_2\},\ \{f_1, f_2\} \to \{f_3\}$$

6. CONCLUSION

Data mining techniques like Association Rule Mining and Correlation Rule Mining were used in decision making situations with large data sets. In the numerical illustration several interesting fuzzy rules were discovered. It is identified that the plant species f_1, f_2, and f_3 are the most frequently occurring among the five total species under study (*B. arundinacea*, - *D. sepiaria* – *Citrus sp*). These three plant species are found to be very closely associated with each other in the vicinities of Pachaimalai hills.

REFERENCES

[1] Agrawal, R, Imielinski, T., & Swami, A. Mining Association Rules between Sets of Items in Large Databases. *Proceedings of the ACM SIGMOD International Conference on Management of Data*, Washington D.C.(1993), 207-216.

[2] Agrawal.R, and Srikant,R. "Fast Algorithms for mining Association rules", *Proceedings of 20th international conf. on very large databases*, Santiago, Chile, 1994, pp.487–499.

[3] Bahadur, K.N. A note on the flowering of Bambusa nutans', *Indian Forester*, 106 (4), (1980), 314-316.

[4] Banik, R.L. Seed germination of some bamboo species', *Indian Forester*, 113(8), (1987), 578-586

[5] Bhol, N., & Parida, S. Influence of growth regulators on propagation of culm- and branch cuttings of *Bambusa vulgaris*. *Journal of Tree Sciences, 34*(1), (2015), 64–68.

[6] Robinson, J.P., Chellathurai, S., & Prakashraj, G.D. A Strategic Study of Mining FuzzyAssociation Rules Using Fuzzy Multiple Correlation Measures. *Journal of Algorithms and Computational Technology*, 6(3), (2012), 499-510.

[7] Sonia, M.D., Robinson, J.P., and Rajasekaran, C.S. Mining Efficient Fuzzy Bio-statistical Rules for Association of Sandalwood in Pachaimalai Hills, *International Journal of Agricultural and Environmental Information Systems*, 6(2), (2015), 40-76.