

Assessment of Financial Status of SHG Members: A Clustering Approach

Sajeev B. U
Research Scholar, Department
of Computer Science
PRIST University, Thanjore,
TamilNadu, India

K. Thangavel
Professor and Head,
Department of Computer
Science,
Periyar University, Salem-636
011 India.

ABSTRACT

Data mining has attracted a great deal of attention in information industry and in society as a whole in recent years, due to the wide availability of huge amount of data and for converting these data to useful information and knowledge. Clustering analysis is a key and easy tool in data mining and pattern recognition. In this paper K-Means and Fuzzy C-means clustering algorithms are used for evaluating the performance of various Self Help Groups (SHGs) in Kerala State, and suggestions are made to improve socioeconomic status. The necessary information about the members of SHG has been collected from 9 districts in Kerala State, Indi. The parameters chosen for the study are financial status, types of loan availed, improvement in assets before and after joining the group, effect of joining in more than one group. District wise analysis are also performed.

General Terms: Data mining, Clustering, Self Help Groups

Key words: K-Means, Fuzzy C-Means, financial status

1. INTRODUCTION

Data mining can be viewed as a result of the natural evolution of information technology. The database system industry witnessed an evolutionary path in the development of different functionalities like data collection, database creation, data management, and advanced data mining involving data warehousing. The term has been defined as the nontrivial extraction of implicit, previously unknown and potentially useful information from data [1], being considered as the science of extracting useful information from large data sets or databases [2]. It is also used in search of consistent patterns and/or systematic relationships between variables in business [3], evaluation of web-based educational programs [4], computer science [5], chemistry [6], engineering [7], and medicine [8] and in all domains where a large amount of data must be analyzed. The advantages associated with data mining are: (i) the accuracy of the data is constant, (ii) analysis work is done routinely, and (iii) large quantities of data can be processed

rapidly. Thus data mining is an essential process where intelligent methods are applied in order to extract patterns. Clustering is one of the core tools used by the data miner. It allows us to group entities in a generally unguided fashion, according to how similar they are. This is done on the basis of a measure of the distance between entities [9]. The aim of clustering is to identify groups of entities/patterns that are close together but as a group are quite separate from other groups. Cluster analysis leads the researcher to classification, definition of structure, new knowledge, and finally discovery. Now-a-days, cluster analysis is mainly conducted on computers to deal with very large-scale and complex datasets. With the development of computer-based techniques, clustering has been widely used in data mining, ranging from web mining, image processing, machine learning, artificial intelligence, pattern recognition, social network analysis, bioinformatics, geography, geology, biology, psychology, sociology, customers behavior analysis, marketing to e-business and other fields[10, 11]. To achieve different application purposes, a large number of clustering algorithms has been developed. However, there are no general-purpose clustering algorithms that fit all kinds of applications; thus, the evaluation of the quality of clustering results plays the critical role of cluster analysis. In this study, cluster has been applied to evaluate the performance SHG groups of nine districts in Kerala state, India. To the best of our knowledge, clustering has not been applied to analyze same kind of data.

1.1 About SHG

There were several efforts to mitigate poverty in the country. However, over the years the Community involvement and the concept of self-help were felt most required for the success of these efforts. In other words, Self-Help Groups are realized to be a proven model/methodology to address the issues of poverty alleviation and women empowerment. The statement of the Central Minister of Rural Development affirms this – “Solutions to the alleviation of poverty lie in generation of self-employment opportunities. Such opportunities can be created by bringing at least one member of every below poverty line family into the fold of SHGs.” In such context

of growing importance of SHGs and their federations as proven models to reach the poor.

A SHG is a group of about 20 people from a homogeneous class, who come together for addressing their common problems. They are encouraged to make voluntary thrift on a regular basis. They use this pooled resource to make small interest bearing loans to their members. The process helps them imbibe the essentials of financial intermediation including prioritization of needs, setting terms and conditions and accounts keeping. This gradually builds financial discipline and credit history for themselves, as the money involved in the lending operations on their own hard earned money saved over time with great difficulty. This is 'warm money'. They also learn to handle resources of a size that is much beyond their individual capacities. The SHG members begin to appreciate that resources are limited and have a cost. Once the groups show this mature financial behaviour, banks are encouraged to make loans to the SHG in certain multiples of the accumulated savings of the SHG. The bank loans are given without any collateral and at market interest rates. Banks find it easier to lend money to the groups as the members have developed a credit history. 'Cold (outside) money' gets added to the own 'warm money' in the hands of the groups, which have become structures, which are able to enforce credit discipline among the members. The members have experienced the benefits of credit discipline by being able to save and borrow regularly without many hassles. The groups continue to decide the terms of loans to their own members. The peer pressure ensures timely repayments and replaces the "collateral" for the bank loans.

1.2 Need for SHGs

Generally SHGs are expected to be platforms for women to help themselves for their social and economic empowerment. This suggests that SHGs basic philosophy lies in the principle of their self-management and self-regulation. The initial activities that groups generally do include community education, information dissemination on health and hygiene, mutual support by way of inter/intra loaning, income generation activities, services and advocacy. The social empowerment is generally seen in terms of influence on his/her own development as an individual, influence over economic resources and other decisions pertaining to general welfare of the family and participation in social, political and economic decision-making. Thus, especially in the socio-political context of Kerala, SHGs are not just needed but most required.

The potential of SHGs in developing as local financial intermediaries to reach the poor has gained recognition, mainly due to their community based participatory approach and Sustainability. This is done through Micro-finance, which is an innovative credit delivery mechanism that ensures viable financial services to the needy in the developing countries. It has the potential to address issues like actualizing equitable gains from the development on a sustainable basis and fight poverty. Thus, the Government of Kerala made financial assistance as integral component of the projects that were implemented through SHGs.

The micro-finance scene in the country is dominated by SHG-bank linkage. NABARD (National bank for agricultural and rural development) is the nodal apex

government financial institution to provide microfinance services to the very poor through the existing banking network and non government organizations. The pilot program aiming at building financial capabilities and self-confidence among the deprived, through sustainable development of SHG-Bank linkage was first initiated in 1992. Since then, the growth is significant, and benefits are widely acclaimed [12, 13].

The rest of the paper is organized as follows: Section 2 describes about the background of this paper and section 3 presents methods used for this research work. Experimental analysis and results are discussed in section 4. Section 5 concludes this paper with suggestion to improve the performance of SHG

2. BACKGROUND

Earlier SHG data evaluations were done using statistical tools [14]. As research methods, a mix of quantitative and qualitative tools is applied. Through a questionnaire quantitative data are collected. The qualitative information will enable verification of the quantitative findings as well as give more insight into the reasons behind these findings. The survey has been conducted through structured questionnaires, related to the socio-economic status of SHG members. Since the purpose of the study is to understand the trends of financial status within groups, the survey focused on group level information. At the individual level of members and the following information has been collected.

- Literacy and education status of group members
- Socio-economic composition of the groups
- Savings and credit related activities of the group
- Social issues taken up by the groups
- Linkage between the groups and bank
- Assets before and after being a member
- Loan taken and purpose of loan

There are 14 districts in Kerala state and this study has been restricted to 9 districts namely Kannur, Calicut, Malappuram, Palakkad, Wyanad, Trichur, Kottayam, Alleppy and Trivandrum. The above mentioned data has been collected from 3500 SHG members with 51 attributes/parameters. Majority members are female. For the better understanding of the financial, utilization of loan, educational and loan repayment status of the SHG members before and after availing the loans has been studied in detailed by applying clustering techniques [15].

The objectives of this study are as follows :

- To evaluate the performance of SHG group in Kerala State.
- To evaluate the nature and kind of loan most suitable for SHGs in different places.
- Evaluation of assets of members before and after joining SHG.
- Problems faced by SHGs due to the joining of members in more than one group.

3. MATERIALS AND METHODS

Clustering is the process of grouping of similar objects into clusters, where objects within each cluster have high similarity, but are dissimilar to the objects in other clusters.

Similarities are assessed based on the attribute value(s) that best describes the object. There are different types of clustering techniques such as Hierarchical clustering, Partitional clustering and Spectral clustering. Partitional clustering includes K-Means clustering and Fuzzy C-Means clustering. Among the various clustering algorithms, **K-Means (KM)** is one of the most popular methods used in data analysis due to its good computational performance. However, it is well known that KM might converge to a local optimum, and its result depends on the initialization process, which randomly generates the initial clustering. In other words, different runs of KM on the same input data might produce different results.

Fuzzy C-Means (FCM) is a method of clustering which allows one piece of data to belong to two or more clusters. This method was developed by Dunn in 1973 and improved by Bezdek in 1981 and it is frequently used in pattern recognition.

Fuzzy clustering in contrast to the usual methods does not provide hard clusters, but returns a degree of membership of each object to all the clusters. The interpretation of these degrees is then left to the user that can apply some kind of a thresholding to generate hard clusters or use these soft degrees directly.

3.1 K-Means Algorithm

K-Means [16-17] clustering technique creates a one level partition of data objects. We first chose K initial centroids, where K is a user specified parameter namely number of clusters desired. Each point is then assigned to the closest centroid and each collection of points assigned to a centroid is a cluster. The centroid of each cluster is updated based on the points assigned to the cluster. We repeat the assignment and update the steps until no point changes clusters or equivalently until the centroid remains the same.

A set of n vectors $X_j, j = 1, 2, \dots, n$, are to be partitioned into c groups $G_i, i = 1, 2, \dots, c$. The cost function, based on the Euclidean distance between a vector X_k in group j and the corresponding cluster center c_i , can be defined by :

$$J = \sum_{i=1}^c J_i = \sum_{i=1}^c \left(\sum_{k, x_k \in G_i} \|X_k - C_i\|^2 \right), \quad (1)$$

where $J_i = \sum_{k, x_k \in G_i} \|X_k - C_i\|^2$ is the cost function within group i.

The partitioned groups are defined by a $c \times n$ binary membership matrix U, where the element u_{ij} is 1 if the jth data point x_j belongs to group i, and 0 otherwise. Once the cluster centers c_i are fixed the minimizing u_{ij} for Equation (1) can be derived as follows:

$$u_{ij} = \begin{cases} 1 & \text{if } \|X_j - C_i\|^2 \leq \|X_j - C_k\|^2, \text{ for each } k \neq i \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

which means that x_j belongs to group i if c_i is the closet centre among all centers.

On the other hand, if the membership matrix is fixed, i.e if u_{ij} is fixed, then the optimal centre c_i that minimize equation (1) is the mean of all vectors in group i :

$$c_i = \frac{1}{|G_i|} \sum_{k, x_k \in G_i} X_k \quad (3)$$

where $|G_i|$ is the size of G_i , or $|G_i| = \sum_{j=1}^n u_{ij}$

The algorithm is presented with a data set $x_i, i = 1, 2, \dots, n$; it then determines the cluster centers c_i and the membership matrix U iteratively using the following steps.

Step 1: Initialize the cluster centre $c_i, i = 1, 2, \dots, c$. This is typically done by randomly selecting c points from among all of the data points.

Step 2: Determine the membership matrix U by Equation (2).

Step 3: Compute the cost function according to Equation (1). Stop if either it is below a certain tolerance value or its improvement over previous interaction is below a certain threshold.

Step 4: Update the cluster centers according to Equation (3). Go to step 2.

3.2 Fuzzy C-Mean Algorithm

The Fuzzy C-Means algorithm (FCM) [18-23], which is the best known unsupervised fuzzy clustering algorithm is also used in analyzing the SHG data. However, these FCM algorithms have considerable trouble in a noisy environment and inaccuracy with a large number of different sample sized clusters

It is based on minimization of the following objective function:

$$J_m = \sum_{i=1}^N \sum_{j=1}^C u_{ij}^m \|x_i - c_j\|^2, \quad 1 \leq m < \infty$$

where m is any real number greater than 1, u_{ij} is the degree of membership of x_i in the cluster j , x_i is the i th of d-dimensional measured data, c_j is the d-dimension center of the cluster, and $\|\cdot\|$ is any norm expressing the similarity between any measured data and the center. The algorithm is composed of the following steps

3.3 Data Cleaning

As data sets are not perfect, one can expect missing values for some attributes, some errors in transcription or data input, and duplicate entries[24-27]. Dealing with these issues is a topic of major study in itself. Sometimes, a

1. Initialize $U = [u_{ij}]$ matrix, $U^{(0)}$
2. At k -step: calculate the centers vectors
 $C^{(k)} = [c_j]$ with $U^{(k)}$

$$c_j = \frac{\sum_{i=1}^N u_{ij}^m x_i}{\sum_{i=1}^N u_{ij}^m}$$
3. Update $U^{(k)}, U^{(k+1)}$
4.
$$u_{ij} = \frac{1}{\sum_{k=1}^C \left(\frac{\|x_i - c_j\|}{\|x_i - c_k\|} \right)^{\frac{2}{m-1}}}$$

 If $\|U^{(k+1)} - U^{(k)}\| < \varepsilon$ then STOP;
 otherwise return to step 2.

received data set has already been 'cleaned' by filter method with the help of domain experts. Perhaps

‘scrubbed’ is a better term: missing values are sometimes filled in with average values, or values copied from similar looking records.

3.4. Feature selection

It is a preprocessing method of choosing a subset of features from the original ones. It has proven effective in reducing dimensionality, improving mining efficiency, increasing mining accuracy, and enhancing result comprehensibility[28]. Feature selection methods can broadly fall into the *wrapper* model and the *filter* model [29]. The wrapper model uses the predictive accuracy of a predetermined mining algorithm to determine the goodness of a selected subset. It is computationally expensive for data with a large number of features. The filter model separates feature selection from classifier learning and relies on general characteristics of the training data to select feature subsets that are independent of any mining algorithms. We have chosen filter method for the present study.

4.EXPERIMENTAL ANALYSIS

Survey was carried out among 3500 SHG members among 9 districts in Kerala. Detailed questionnaires were prepared. Qualitative information is gathered through semi-structured interviews with SHG members, SHG leaders, federation leaders, Bank officials, moneylenders and government officials. The selected SHG groups were found to be very stable for more than 3 years. From these groups we have collected 3500 objects with 51 attributes. The collected data have been cleaned with the help of domain experts and applying feature selection method. Finally we fixed the data set as 3434 objects with 17 attributes.

The selected attributes are given in the Table I

Table I Selected Attributes for the study.

No	Attribute
1	Age
2	Male/Female
3	Married/Unmarried
4	Literate / Illiterate
5	Total income of the family / month
6	Other income / month
7	Savings / month
8	Other income /extra income through SHG
9	Deposit in the group
10	Savings / month outside the group
11	Loan amount
12	Balance loan amount in the book
13	Amount taken
14	Assets before being a member
15	Assets increased after being a member
16	Membership in other groups (Y/N)
17	Economic Benefits gained

The K-Means and Fuzzy C-Means algorithm discussed in section 3.1 and 3.2 are applied for the SHG data collected from 9 districts in Kerala. The K-Means algorithm has been performed for different values of K and it was found that the best value for K is 2. After analysis the activities and functionalities of two different group members, one group

has been identified as performing group and other one is non-performing.

Clusters obtained by K-Means algorithm is dominated by the selection of initial seed or centroid. Hence K-Mean algorithm has been performed by selecting different set of initial seed and the result are tabulated in table II

Table II Results of K-Means for different centroids.

Number of runs of K-Means different seeds (Ri)	Number of patterns in cluster I (C1)	Number of patterns in cluster II (C2)	Number of runs of K-Means different seeds (Ri)	Number of patterns in cluster I(C1)	Number of patterns in cluster II(C2)
R1	3305	129	R14	591*	2843**
R2	2843**	591*	R15	2843**	591*
R3	591*	2843**	R16	591*	2843**
R4	2999©©	435©	R17	591*	2843**
R5	2843**	591*	R18	2843**	591*
R6	591*	2843**	R19	136	3297
R7	591*	2843**	R20	2999©©	435©
R8	591*	2843**	R21	591*	2843**
R9	2999©©	435©	R22	591*	2843**
R10	591*	2843**	R23	2843**	591*
R11	2999©©	435©	R24	435©	2999©©
R12	2843**	591*	R25	435©	2999©©
R13	2843**	591*			

*, **, ©, ©©, indicates same number of pattern in the clusters.

Each time when we apply the algorithm we obtained two clusters; of which one is performed cluster and other one is non- performing cluster. Since the numbers of objects in some clusters are same, which shows their stability; we have selected 8 clusters from 25 clusters for further studies, since the selected clusters have different number of patterns.

By applying K–Means algorithm for different centroids, we have obtained the patterns of each parameters like deposit, savings, total income of the family, balance loan amount in the book, other income, savings outside the group, deposit in the group, assets after being a member, economic benefits gained, amount taken (loan I and loan II).

Loan I is from Bank to SHG, and SHG to members and Loan II is from SHG (SHG is lending money from societies, money lenders etc to their members)

It is observed from the financial status has been improved for the members who have availed loans. Table III shows that the members who have taken loans showed improved financial status.

Table III Patterns obtained for different runs of K-Mean Algorithm

Attributes	R1C1	R1C2	R3C1	R3C2	R4C1	R4C2	R19C1	R19C2
Total income of family per month in Rs	2550.00	3520.00	3207.284	2457.2	2489.637	3252.658	3525.899	2547.55
Other income per month in Rs	150.00	440.00	307.6557	132.1418	137.3721	334.5397	469.6874	149.6753
Savings per month in Rs	40.00	90.00	59.29797	35.01337	35.84629	62.2647	90.81377	37.06429
Other per extra income through SHG in Rs	430.00	910.00	813.3695	377.5766	395.3318	847.2459	940.4924	432.4591
Deposit in the group	2640.00	3250.00	3708.391	2449.68	2512.475	3726.867	3293.335	2640.455
Savings per month outside the group in Rs	120.00	250.00	192.2848	110.9824	113.9493	200.9873	253.2564	119.6853
Loan I	20540.00	28660.00	48924.34	15012.79	17050.04	47040.48	28742	20523.61
Balance loan amount to be paid in Rs	4870.00	5910.00	11459.42	3546.038	3958.749	11452.01	6045.548	4861.048
Loan II	4920.00	127500.0	42813.59	2608.203	2588.903	57365.78	124040	4805.404
Assets before being a member (credit points)	20.00	20.00	18.96279	17.79107	17.85695	18.92876	19.12523	17.94603
Assets increased after being a member (credit points)	10.00	10.00	8.588854	6.420682	6.547182	8.494282	8.757773	6.712856
Membership in other groups	0	0	0.472083	0.462188	0.462154	0.475865	0.382374	0.467253
Economic-benefit	1430.00	1420	2039.599	1297.784	1361.554	1865.986	1444.925	1424.651

Table IV- Results of Fuzzy C-Means for different values on m

Differ ent runs	Number of iterations	Number of members in cluster 1	Number of members in cluster 2	m
R1	64	130	3304	1.25
R2	100*	158	3276	1.5
R3	100*	278	3156	1.75
R4	100*	2742	692	2
R5	100*	863	2571	2.25
R6	100*	901	2533	2.5
R7	100*	935	2499	2.75
R8	69	958	2476	3
R9	100*	958	2476	3.25
R10	68	991	2443	3.5
R11	59	2432	1002	3.75
R12	94	2416	1018	4
R13	51	2278	1156	10
R14	53	1173	2261	20
R15	46	1180	2254	30
R16	25	1186	2248	40
R17	3	1237	2197	50

* Indicates Maximum Number of Iterations

To perform a comparative study we have applied Fuzzy C-Means algorithm in the SHG data for different values on 'm'(weight exponent in the fuzzy membership) and the results are tabulated in table IV

We have applied different values for m and a total of 17 runs were performed. It was seen that when the values of m varies from 1 to 4 the number of iterations are high and as the values of m varies from 10 to 50, number of iterations are found to decrease. For further studies we have selected clusters with maximum number of iterations (100). The table V shows the patterns obtained for different runs of Fuzzy C-Means algorithm

On analysis of tables II, III, IV and V, we have found that R1C2 (second cluster in the first run), R3C1 (first cluster in the third run), R4C2 (second cluster in the fourth run), and R19C1 (first cluster in the nineteenth run), shows maximum savings, deposits, assets etc. Further investigations revealed that SHGs are effectively working in all districts with increased financial status for their members which is shown in the figure I and II

The table V shows the patterns obtained for different runs of Fuzzy C-Means algorithm

Attribute	R2C1	R2C2	R3C1	R3C2	R4C1	R4C2	R5C1	R5C2	R6C1	R6C2	R7C1	R7C2	R9C1	R9C2
Total income of the family per month in Rs	3360	2540	3155	2510	2448	3057	2995	2423	2968	2411	2948	2401	2388	2919
Other income per month in Rs	420	150	356	141	126	284	264	120	257	118	253	116	114	246
Savings per month in Rs	80	40	69	36	34	57	54	33	53	32	53	32	32	52
Other income per extra income through SHG in Rs	910	420	842	404	360	773	732	344	716	338	705	334	329	688
Deposit in the group in Rs	3370	2610	3442	2554	2364	3692	3681	2277	3649	2250	3620	2232	2211	3570
Savings per month outside the group in Rs	240	120	211	115	106	187	180	103	177	102	174	101	100	171
Loan amount in Rs	32330	19730	35197	18151	13604	45169	45173	12008	44165	11594	43269	11338	11053	41802
Balance loan amount in the book in Rs	7130	4650	8086	4251	3340	10226	9953	3067	9665	3001	9434	2960	2917	9082
Amount taken	104890	3970	74361	2832	2982	33380	23666	3292	21788	3244	20661	3189	3114	19274
Assets before being a member	20	20	19	18	18	19	19	18	19	18	19	18	18	19
Assets increased after being a member	10	10	8	7	6	8	8	6	8	6	8	6	6	8
Membership in other groups	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Economic Benefits gained	1560	1410	1674	1368	1190	2088	2182	1106	2173	1078	2158	1059	1036	2128

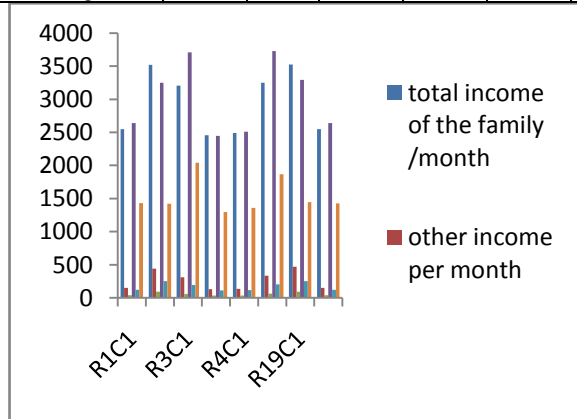


Figure 1 Financial status using K means

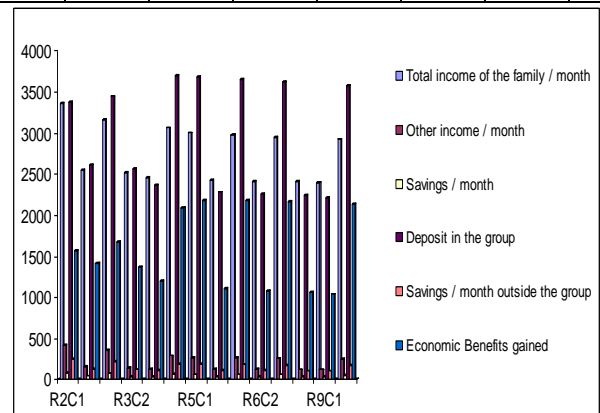


Figure II Financial status using Fuzzy C-Means

Figure III and IV shows the improvement in assets calculated before and after being an SHG member using K Means and Fuzzy C-Means algorithm respectively. Weightage was given to each type of assets and net weightage in terms of credit points was calculated and tabulated. For most of the SHG members the assets were obtained in terms of house, land, furniture, gold, gas connection, poultry, electrical equipments vehicles, agricultural implements etc.

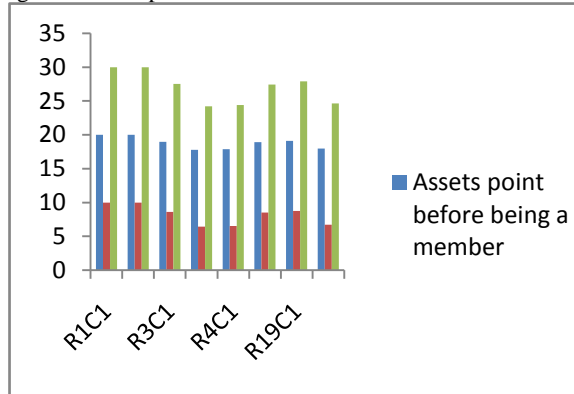


Figure III Clusters Vs Assets using K means

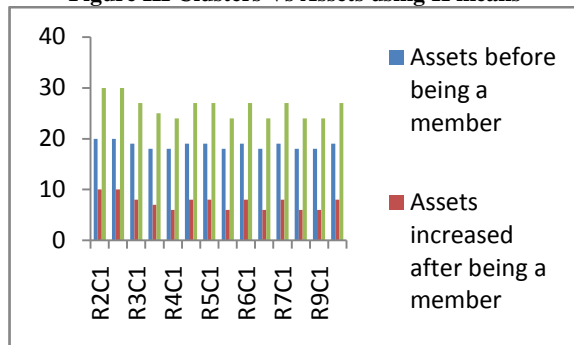


Figure IV Clusters Vs Assets using Fuzzy C-means

It was observed from the identified pattern of the members of the performing cluster, that the members joining in more than one group are having higher balance loan amount and they are having difficulty in repaying the loan amount, since they are availing loan from different groups (Figure V). In order to avoid this problem faced by members, it is not advisable for the members to join in more than one group.

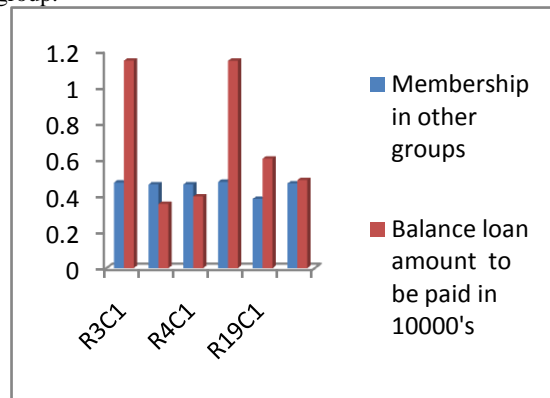


Figure V Members joining in other group Vs Balance loan amount

Based on the pervious analysis, we indentified the performing cluster such as R1C2, R3C1, R4C2 and R19C1. From this group, we have selected R4C2 randomly for further study about the purpose of loan pattern in the SHG group. In the performed clusters there is an increase in the savings, deposit in the group, total income etc. The effect of the membership in other groups could not be evaluated using Fuzzy C-Means since the values obtained was very low.

Out of around one billion people in India, 26 % are poor. At the bottom the poor need credit for small productive assets, working capital, housing, illness, and emergencies. The demand for credit here is not only large but heterogeneous as well. Table VI shows the loan number, type of loan and the number of persons availed these loans.

Table VI shows the loan availed by the number SHG members

Loan Number	Type of Loans	Number of persons availed this loan in the cluster
L1	Personal loan	45
L2	Educational loan for children	132
L3	Wedding	60
L4	Treatment	61
L5	Paying to money lenders	190
L6	Purchase of land	12
L7	Construction of house	173
L8	Digging well	14
L9	Toilet construction	0
L10	For buying domestic animals like cow	83
L11	For purchasing goat	22
L12	Poultry	9
L13	Agriculture	84
L14	Business	68
L15	Purchase of vehicle	11
L16	Purchase of home appliances	16

From the table it is clear that the most of the money which they get as loan is taken by money lenders from which they have borrowed it for various purposes. So it will be effective if bank linkages with the SHG members are more effective.

The SHG-bank linkage program is gaining increasing acceptance amongst bankers. The NABARD envisions covering one third of the rural population in India by establishing one million SHGs. The government of India has already made announcements for uplifting 3,00,000 SHGs by year 2011. The task force on microfinance sees the SHG-bank linkage program emerging as a major way of banking with the poor in coming years. Under the SHG-bank linkage program, banks interact with the poor, especially women, to form small homogenous groups. These small groups are encouraged to meet frequently and collect small thrift amounts from their members and are taught simple accounting methods to enable them to maintain their accounts. Although individually these poor

could never have enough savings to open a bank account, the pooled savings enable them to open a formal bank account in the name of the group. This is the first step in establishing links with the formal banking system. Groups then, meet often and use the pooled thrift to impart small loans to members for meeting their small emergent needs. This saves them from usurious debt traps and thus begins their empowerment through group dynamics, decision-making, and funds management. Gradually the pooled thrift grows and soon they are ready to receive external funds in multiples of their group savings. Bank loans enable the group members to undertake income generating activities. The loans availed by the SHG members are used for the construction of houses.

This study reveal that majority of the SHG members have taken loans for educational purposes. It is a fact that one of the major demands for loans within the Self Help Groups (SHGs) is for meeting the education expenses of the children of the women members. As and when their need increases the demand is met from outsourcing of loans by the groups. Most of the external loans of the groups continue to be targeted to meet the entire need of the group and seldom take into consideration the specific requirement of the members related to education of their children. However, the main demand for education loans is for those clients whose children are in colleges for getting higher education or pursuing courses like nursing, B.Ed courses, and technical diploma courses. Hence, it is important to give concession or free education to the children of SHG members, so that they can utilize the fund obtained for performing well by investing in business which has sufficient returns. The members who have availed loan L13 and L10 have made profit. So it is advisable to give loans for agriculture and for domestic animals like cow, goat, poultry etc since they are making profits and their repayment of loans are proper.

The district wise studies based on cluster analysis on the type of loan availed shows that Wyanad district availed maximum loan for L2, L5, L7 and L13. Loan L10 is availed maximum by Mallapuram district.

Figure VI shows the district wise studies on the type of loan availed by SHG members in a performed cluster R4C2.

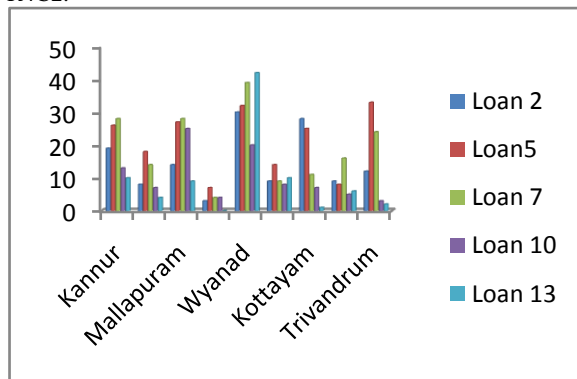


Figure VI shows the district wise study on types of loan
 After analyzing the cluster R4C2, it is clear that the maximum number of SHG members from Wyanad district

has availed maximum loan in L2, L5, L7 and L13. Loan L10 is maximum availed by Mallapuram district. Loan L13 is availed in minimum by Palakkad. The table V shows that maximum number of loan is availed by the members of SHG group in Wyanad and minimum number of loan is availed by the SHG members of Palakkad.

5. CONCLUSIONS

The potential of SHGs in developing as local financial intermediaries to reach the poor through bank has gained recognition and showed an improved performance. Conventionally statistical methods were used for the analysis for SHG data. In this paper we are introducing data mining method for easy and fast analysis. Based on clustering analysis using K-means and Fuzzy C-means, we have analyzed the performance of SHG groups in Kerala State. SHGs are effectively working in all selected districts. The clustering analysis reveals that the financial status of the members showed marked improvement on joining this group. Members joining in more than one SHG group showed a decreased performance rate due to the delay in repayment of loans. Most of the members are taking loan for educational purpose, which reduce their performance in the group. So our suggestion is to give free education or educational concession to the children of SHG members, so that they can use the fund for productive purpose. Most of members utilize their loan amount for repaying the money borrowed from money lenders, so effective bank linkage with the SHG group will produce marked increase in their performance rate. This reveals that loan availed for domestic animals and agriculture is gaining profit and their repayment is proper. From the model it is suggested to sanction more loans for agriculture and poultry in this area and there by involving a better linkage with nationalized bank. Cluster analysis reveals that Wayanad district availed maximum number of loans for education, for repaying money lenders, construction of house, agriculture. Analysis on SHG data's can be carried out effectively using K-means and Fuzzy C-means algorithm.

6. REFERENCES

- [1] Frawley, W., G. Piatetsky-Shapiro and C. Matheus, 1992. Knowledge Discovery in Databases: An Overview. *AI Magazine*, pp: 213-228.
- [2] Hand, D., H. Mannila and P. Smyth. 2001. *Principles of Data Mining*. MIT Press, Cambridge, MA.
- [3] Chen, Y.L., J.M. Chen and C.W. Tung, 2007. A data mining approach for retail knowledge discovery with consideration of the effect of shelf-space adjacency on sales. *Decision Support Systems*, 42: 1503-1520..
- [4] Romero, C. and S. Ventura. 2007. Educational data mining: A survey from 1995 to 2005. *Expert Systems with Application's*, 33: 135-146.
- [5] Lee, A.J.T., R.W. Hong, W.M. Ko, W.K. Tsao and H.H. Lin. 2007. Mining spatial association rules in image databases. *Information Science*. 177: 1593-1608. 2007.
- [6] Maran, U., S. Sild, I. Kahn and K. Takkis, 2007. Mining of the chemical information in GRID

- environment. Future Generation Computer Systems, 23, 76-83.
- [7] Yang, Q., J. Yin, C. Ling and R. Pan. 2007. Extracting actionable knowledge from decision trees. IEEE Transactions on Knowledge and Data Engineering, 19, 43-55.
- [8] Imamura, T., S. Matsumoto, Y. Kanagawa, B. Tajima, S. Matsuya, M. Furue and H. Oyama. 2007. A technique for identifying three diagnostic findings using association analysis. Medical and Biological Engineering and Computing, 45, 51-59.
- [9] The MathWorks, Inc. textbook online, <http://www.mathworks.com/>
- [10] A. Jain, M. N. Murty and P. J. Flynn, 1999. Data Clustering: A Review", ACM Computing Surveys, Vol. 31(3), pp. 264-323.
- [11] J. Hartigan. 1975. Clustering Algorithms. New York: Wiley.
- [12] B. Kumar. 2005. Impact of Microfinance through SHG-Bank Linkage in India: A Micro Study. Vilakshan, XIMB Journal of Management, July 9,.
- [13] Mahendra Varman P; 2005. Impact of Self-Help Groups on Formal Banking Habits; Economic and Political Weekly April 23, pp 1705-13.
- [14] B. Narayana swamy, K. Narayana Gowda and G. N. Nagaraj .2007. Performance of Self Help Groups of Karnataka in Farm Activities"; Karnataka J. Agric. Sci., 20(1):85 - 88.
- [15] J. Pena, J. Lozano, and P. Larranaga, 1999. An Empirical Comparison Of Four Initialization Methods For The K-Means Algorithm. Pattern Recognition Letters, Vol. 20 No. 10, pp. 1027-10.
- [16] J. Macqueen, 1967. Some Methods For Classification And Analysis Of Multivariate Observations. In proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, University of California Press, pp. 281-297.
- [17] Pang-Ning Tan, Michael Steinbach, Vipin Kumar, Text Book on Introduction to Data Mining" Pearson Education; PP 496-97; III edition.
- [18] J.C. Bezdek, 1981. Pattern Recognition with Fuzzy Objective Function Algorithms, Plenum, New York.
- [19] J.C. Dunn, 1974. A fuzzy relative of the ISODATA process and its use in detecting compact, well-separated clusters, J. Cybernet. 3, 32–57.
- [20] I. Gath, A.B. Geva, 1989. Unsupervised optimal fuzzy clustering, IEEE Trans. Pattern Anal. Mach. Intell. 11 773–781.
- [21] 21 R.N. Dave, 1990. Fuzzy-shell clustering and applications to circle detection in digital images, Int. J. General Syst. 16. 343–355.
- [22] M.S. Yang, C.H. Ko, 1997. On cluster-wise fuzzy regression analysis, IEEE Trans. Systems, Man, Cybern. 27, 1–13
- [23] Kuo-Lung Wu, Miin-Shen Yang; 2002. Alternative C-Means clustering algorithms, Pattern Recognition 35, 2267 – 2278
- [24] 24 H. Galhardas, D. Florescu, D. Shasha, and E. Simon. 2000. Ajax: An Extensible Data Cleaning Tool, Proc. ACM SIGMOD Int'l Conf. Management of Data,.
- [25] 25 H. Galhardas, D. Florescu, D. Shasha, E. Simon, and C. Saita, 2001 .Declarative Data Cleaning: Language, Model, and Algorithms, roc. 2001 Very Large Data Bases (VLDB) Conf.
- [26] 26 M. Hernandez and S. Stolfo, 1995. The Merge/Purge Problem for Large Databases," Proc. ACM SIGMOD Int'l Conf. Management of Data, pp. 127-138, May..
- [27] 27 M.L. Lee, T.W. Ling, and W.L. Low, 2000. Intelliclean: A Knowledge- Based Intelligent Data Cleaner. Proc. Sixth ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining,.
- [28] 28. L. Yu and H. Liu. 2003 . Feature selection for high-dimensional data: a fast correlation-based filter solution". In: Proceedings of the 20th International Conferences on Machine Learning, Washington DC.
- [29] 29. R. Kohavi and G. John. 1997 .Wrappers for feature subset selection. Artificial Intelligence. Vol. (1-2), pp. 273-324.