A Stochastic Approach towards Inclusive Rural Development Using Demographic Data

Mukundaram P
Department of Information Technology
Easwari Engineering College, Chennai,India
mukund.ram96@gmail.com

Pravien M
Department of Information Technology
Easwari Engineering College, Chennai, India
mpravien1996@gmail.com

***Abstract*— Technology has advanced immensely in India in the recent years but its advantage has not been wholly utilized by the rural parts of our country .Villages hold a diversified job environment where most jobs can be made highly profitable by establishing proper communication among various professions. Information awareness in cities is playing a major role in improving small scale businesses which is missing in villages. This can be achieved by using appropriate use of available technology in villages which will help us determine correlations in rural India. A demographic analysis using stochastic mechanisms focusing on specific aspects can be used to learn about villagers and target appropriate methods to improve their lifestyle by the application of technology. We attempt to extract the demographics (like age, gender, place, skill sets) of a person and find various overlaps of his/her job nature with other professionals in a specific geography. This analysis would give us insights on a lot of interrelations among the various professions which can be exploited to maximum extent so that we achieve higher levels of productivity. A stochastic approach on the demographic data can be used to achieve this and thereby we can expect an inclusive rural development. Since, there are high levels of uncertainty on a variety of parameters we go for a stochastic approach to carry out our analysis. If we take probability theory into consideration, a purely stochastic system is one which is randomly determined and also has a random probabilistic distribution. The pattern which is derived from this system can only be analyzed statistically but cannot be accurately predicted. The probabilistic patterns obtained finally would provide us insights of how rural development is being carried out in a village and what factors should we need to focus on a that village to ensure wholesome development.**

Keywords—stochastic approach, demography, rural development

I. INTRODUCTION

The contribution from villages to India's GDP has been constantly declining over the last decade. The primary reason is the enormous growth of service sector overshadowing the progress of agriculture and its related activities. A lot of villages are now moving towards non-agricultural opportunities due to lack of technological development. Previously technological innovations were not known to the villagers, as many are either illiterates or they were unaware as a result of poor communication mechanisms. Even after mobile connectivity extending to every nook and corner of India making information sharing easier across borders, in terms of growth of villages in comparison with urban areas there is a big difference. A large percentage of the technological applications are targeted only towards cities/towns and very few focus on rural development. With 60% of the total workforce located in rural India, the resources available are only being sparsely used. Experts may argue that the agricultural yield in India has been on increasing but on the contrast the contribution of this workforce towards our GDP has been declining. The agriculture growth rate in India's GDP has reduced over the years. Low productivity in rural areas is due to a number of factors such as illiteracy, insufficient finance, and inadequate knowledge or lack of appropriate information transfer methods. In India, the average size of farms is very small which stresses a need for using efficient methods to ensure higher productivity. Farming requires assistance from other sectors such as power to run motors, fertilizer for crops, agricultural equipments for mechanized farming and skilled labor. A lot of other professions in villagers such as animal husbandry, poultry farming has direct/indirect dependence on how farming spans out for a particular season. This all the more increases the importance of efficient farming in villages. The most interesting aspect which has not been efficiently captured is that there are immense overlaps when you analyze the work done by people in villages. If this can be identified, a more productive environment for growth can be given to our villages thereby ensuring wholesome development.

If you take the reason for the migration of villagers to urban areas, it is mainly due to lack of overall development in village like proper schooling and medical facilities, transportation facilities and lack of availability of consumer goods. This has created a lacuna in rural India where many skilled laborers are being forced to quit farming and resort to other professions to meet their expenses. This migration has resulted in overpopulated cities and has created a deficit of skilled laborers in villages. If the same trend continues, we will be losing a major chunk of skilled population in villages which in the years to come would result in major problems. By applying the findings of this stochastic analysis, we intend to find reasons for this migration and bridge the gap so as to ensure rural development happens on a balanced scale. We hope to create a profitable environment in rural areas by application of our analysis.

II. RELATED WORKS

One of the main applications of stochastic process is predicting Stock markets and exchange rate fluctuations .It is also used in random movement like Brownian motion. Another main application is the Google's Page rank algorithm which is completely based on stochastic matrices and its related approach. The PageRank will assign a rank to all possible search results. The higher the page score, the further up it will appear in the search list. Basically stochastic approach is used when we do not know about the next state of a process. Since take into account of a large sampling, we cannot accurately predict the end outcome we take a stochastic approach to do this. The stock market fluctuations which depends on a variety of factors is usually analyzed using stochastic methods as the end result is only a probabilistic measure and not an accurate prediction.

Stochastic approach is also increasingly used in the field of population genetics. Population genetics is the study of distributions and changes of allele frequency in a population, as the population depends on four main processes: natural selection, genetic drift, mutation and gene flow.

III. PROPOSED MODEL

In this section, we are going to present a conceptual model of our system and explain various aspects of it in a systematic manner.

A. Ensuring non duplicate entries

Since our system holds an enormity of data ensuring the uniqueness throughout the process is critical. Hence, we go for Aadhar Card as it's used for identity verification throughout the country. Every Aadhar card has a QR code embedded in it and by scanning the QR code we can get basic personal details of a person. This helps us to reduce the workload of collecting information for demographic analysis and this information can also help us locate his/her geography. This entry will serve as the unique ID in our database throughout all processes.

Fig.1 Sample Aadhar Card with QR Code

B. Questionnaire to get specific details

Questionnaire forms the core of our system as the way in which questions are framed will decide on how efficient it can be analyzed. It is a detailed questionnaire which focuses on different aspects that are needed to increase the productivity of the respective jobs. It is framed in an adaptive manner, i.e. depending on the responses of the previous question, the following questions will be asked and all answers are interdependent. We take into account of this interdependence for our stochastic analysis. The questions are framed keeping in mind of the result and hence are very specific and simple in nature. We've ensured that the responses for the questions cannot vary from our options and the respondents have to choose only from the options provided. This is done so as to give flexibility of converting the questionnaire into other languages in the future. This will also aid us in efficient analysis of the responses.

Questionnaire can be answered using the application or through a website. The information can also be collected at any government centers where computer facility is available if the customer is unable to send the information through home. All the responses will be collected, stored and sorted in a common database for stochastic analysis which will be explained in detail in the coming section.

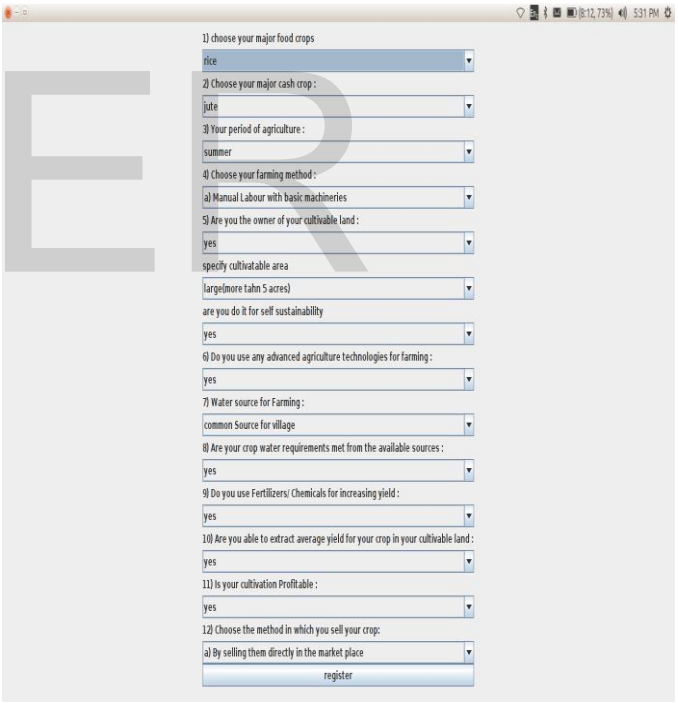

Fig. 2 Sample Questionnaire

Basic sorting of information will be done before we proceed to the analysis section. Sorted data will contain different job clusters from which we will break down into further groups according to our algorithm. The data which we obtain finally will give an outline of how various jobs are divided in a particular area and their interrelationships.

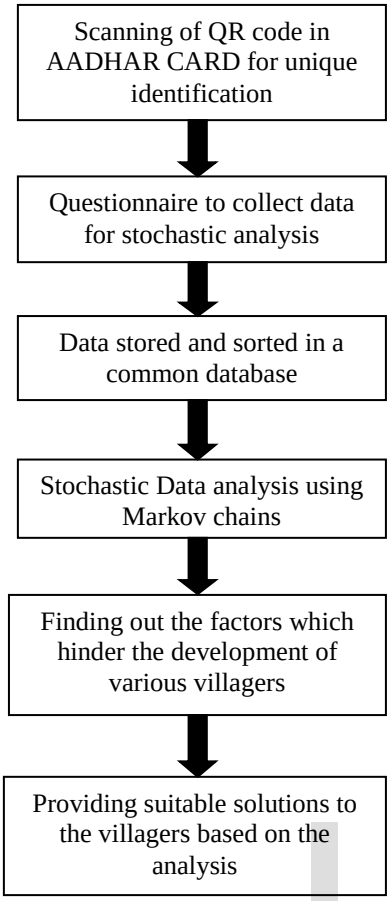

Fig.3 Flowchart of the Process

C. Stochastic Data Analysis using Markov Chains

In the stochastic data analysis, we utilize a method similar to the PageRank algorithm adopted by Google to find the important differentiating factors for profit and loss. Then, we apply Markov Chains to narrow down on the most critical factors.

Firstly, we segregate the responses depending on whether they are profitable or not. Then, we take into account of the profitable people and find out which answer for each of the factors has a maximum probability. We do the same non-profitable people. Based on these values we form a probability matrix. Then, we do a row-by-row comparison of the probabilities to find out which factors have the maximum deviation. In order to obtain a square matrix to apply Markov Principles, whenever we get an odd number of deviating factors, we include the next highest deviating factor. At the end of this process, we get a shortlisted square matrix which holds all important factors.

From this we construct a transition probability matrix for every non-profitable person with the obtained factors. For example, if there are four important factors we will construct a 4x4 transition matrix where both rows and columns have similar factors. The occurrence of factors is a Markov Process where the value of the current probabilistic state is dependent on the previous state. Considering there is a positive response for first factor, and then we calculate the negative response for this other three factors which will represent the first row of the matrix. Based on the values, we find the probabilities and based on these probabilities we formulate ratios so that we get a right stochastic matrix where the matrix is a real square matrix with each row summing to one.

After performing all these processes, we will add the column values of the stochastic matrix and find the column that has the maximum value. This will provide us the most inevitable factor which could be the major reason for loss. Based on this analysis, we may suggest to the villagers about the most important factor along with other factors which may also increase their profit margin. Since, this analysis is done as a whole and is not based on any individual responses we get a complete picture about the factors influencing various professions in a specific village. This will help the villagers to increase their productivity and thereby improve the quality of life of their respective villages.

D. Getting Time to Time Needs from the Villagers

At this moment, we have our systems ready which hold stochastically analyzed demographic information about the villagers with their various skill sets. Now, as and when there is a need for villagers it has to be communicated to the system. Then, the system routes an appropriate response to the requesting villager thereby filling the gap of information awareness that we want to bridge using this method.

IV. EXPERIMENTAL ANALYSIS AND RESULTS

Let us consider a sample of 1000 farmers. Let the number of profitable farmers 375 and the number of non-profitable farmers is 625. The data is obtained after extensive sorting and basic analysis. For every question, we find whether the response is positive or negative and find the cumulative amount of positive/negative responses for each question. This procedure is done for both profitable and non-profitable farmers.

We do a numerical comparison for the positive/negative responses for all questions asked to the profitable farmers. If the positive response is high, we consider it to be the optimal factor for profit. We repeat the same for all the questions. We find the probability of the highest response among the two responses for each question. The same is repeated for non-profitable farmers.

We construct a matrix taking into account of the probability of highest response for each question. Two matrix are constructed, one for profitable farmers and other for non-profitable farmers. Since there are 9 questions, we construct a 9x1 matrix.

Let the profitable matrix be P(X) and the non-profitable matrix be Q(X). Each value inside the matrix represents the ratio of number of respondents who use profitable factor for farming. Each row represents a factor. Out of 450 profitable farmers, 270 gave the positive response for first questions and so the probability of positive response for first question is 0.6. Similarly, we perform the same procedure and calculate the probability for remaining factors. The same is repeated for non-profitable farmers.

$$P(X)=\begin{pmatrix}0.6\\0.7\\0.75\\0.89\\0.90\\0.72\\0.88\\0.9\\0.78\end{pmatrix} \qquad Q(X)=\begin{pmatrix}0.66\\0.4\\0.3\\0.75\\0.91\\0.70\\0.17\\0.93\\0.20\end{pmatrix}$$

Fig. 4 Matrix containing positive response probabilities

Now, we compare both the matrices and the after row-by–row comparison for each factor, the factors which has high deviations are the considered to be critical factors for profit.

Comparing the above matrices, we infer that row 2, row 3, row 7 and row 9 have maximum deviation and the factors corresponding to these rows are considered.

Now we construct a 4x4 transition probability matrix. We will get a right stochastic matrix where the sum of rows will be equal to one since we transform the probabilities obtained into ratios. This matrix is constructed by assuming the first factor as yes for the first row and considering the negative response for other factors as a whole. This process is repeated for all other rows.

For the sample data used above, we get the following TPM:

	F1	F2	F3	F4
F1	0	0.33	0.58	0.09
F2	0.25	0	0.52	0.23
F3	0.29	0.48	0	0.23
F4	0.45	0.08	0.47	0

Fig. 5 TPM containing probability of important factors not used by non-profitable farmers

From this matrix, we can single out the most inevitable factor by finding out the summation of the column that has maximum value. In the above matrix, it is the third column, which infers that the factor 3 is the major reasons for profitability and in order to obtain profits that factor becomes an inevitable one.

Note: We take into account the summation of columns as it is a right stochastic matrix.

V. CONCLUSION

Thus, by implementing our system we can conclude on the most inevitable factor for profitability in each and every profession of a particular village by extensive analysis on the various methods adopted for their respective professions. This study will also reveal various interrelationships among the professions in villages which will help us devise appropriate methods to efficiently communicate to the villages in according to their needs. Since, we are taking into account of various factors and doing an analysis on a granular level, the results obtained will not be very deviating. Moreover, as this system completely depends on the data provided by the villagers the final analysis is not theoretical but conclusive.

We can dream of wholesome and inclusive rural development and ensure the 60% of workforce in villages are being efficiently utilized by adopting this method. The demographic data obtained as a result of this analysis would prove to be useful for other aided applications and can also be used for developmental activities as we also gather the information on people's skill sets.

VI. REFERENCES

[1] Dean L.Isaacson and Richard W.Madsen, Markov Chains Theory And Applications, John Wiley and Sons , 1976.

[2] Steven J.Luis and Dennis Mclaughlin, "A Stochastic Approach to Model Validation", Advances in Water Resources,vol.15,pp 15-32, 1992.

[3] Amy N.Langville and Carl D.Mayer,Google's Page Rank and Beyond:The Science of Search Engine Rankings, Princeton University Press, 2006.

[4] Christine M.Hunter, Hal Caswell, Michael C.Runge, Eric V.Regehr , Steve C.Amstrup and Ian Stirling, "Climate change threatens Polar Bear populations:A Stochastic Demographic Analysis", ESA Journals, vol.91 , pp 2883-2897, 2010.

[5] Aarti Kumar and Sujoy Das, " Rule Based Automatic Generation of Query Terms for SMS Based Retrieval Systems", International Journal of Computer Applications Technology and Research, vol. 3, no. 11, November 2014.

[6] http://en.wikipedia.org/wiki/Stochastic_matrix, Accessed Online March 5,2015.

[7] http://www.math.cornell.edu/~mec/Winter2009/RalucaRemus/Lecture3/lecture3.html, Accessed Online March 8,2015.