



Cluster based Analysis and Consumption of Food Products in Targeted Public Distribution System

By Ms. P Shanmuga Priya & Dr. S Santhosh Baboo

D. G. Vaishnav College

Abstract - The Public Distribution System in India is 50 years old. At present it is being carried on as an anti-inflationary and antipoverty system. Tamil Nadu, the southernmost State in the country, is adopting the Universal Public Distribution System covering its entire population and supplying regularly rice, wheat, sugar, kerosene and other products like pulses, edible oil etc. The PDS is a centrally sponsored scheme that entitles beneficiaries to subsidized food grains every month. Several challenges have been identified in the implementation of PDS like (i) Targeting errors (ii) Large leakages or diversion (iii) The elimination of bogus cards and (iv) The problems in Fair Price Shops. This paper analyses and evaluates the problems and finds the possible solutions using the data mining techniques based on preprocessing and clustering. The K-means and K-harmonic means algorithms are combined to cluster the data based on the type of food commodities for rice and wheat.

Keywords : K-means, K-harmonic means, PDS, cluster based PDS.

GJCST-C Classification : D.2.8



Strictly as per the compliance and regulations of:



Cluster based Analysis and Consumption of Food Products in Targeted Public Distribution System

Ms. P Shanmuga Priya^α & Dr. S Santhosh Baboo^σ

Abstract - The Public Distribution System in India is 50 years old. At present it is being carried on as an anti-inflationary and antipoverty system. Tamil Nadu, the southernmost State in the country, is adopting the Universal Public Distribution System covering its entire population and supplying regularly rice, wheat, sugar, kerosene and other products like pulses, edible oil etc. The PDS is a centrally sponsored scheme that entitles beneficiaries to subsidized food grains every month. Several challenges have been identified in the implementation of PDS like (i) Targeting errors (ii) Large leakages or diversion (iii) The elimination of bogus cards and (iv) The problems in Fair Price Shops. This paper analyses and evaluates the problems and finds the possible solutions using the data mining techniques based on preprocessing and clustering. The K-means and K-harmonic means algorithms are combined to cluster the data based on the type of food commodities for rice and wheat.

Keywords : K-means, K-harmonic means, PDS, cluster based PDS.

I. INTRODUCTION

Data mining is defined as the discovery of interesting structure in data, where structure designates patterns, statistical or predictive models of the data, and relationships among parts of the data [3]. It is also the extraction of interesting non-trivial, implicit, previously unknown and potentially useful information or patterns from data in large databases. Preprocessing has been used to keep the data set ready for the process. Entity extraction has been used to automatically identify person, address and type of card. Filtering techniques has been used to filter the incomplete, noisy and inconsistent data.

a) Clustering

Clustering is the process of grouping data objects into similar classes. A cluster is a collection of data objects that are similar to one another within the same cluster and are dissimilar to the objects in other clusters. Clustering is an important area of application for a variety of fields including data mining. It is the unsupervised classification of patterns (observations, data items or feature vectors) into groups (clusters). The clustering problem has been addressed in many

contexts and by researchers in many disciplines. This reflects its broad appeal and usefulness as one of the steps in exploratory data analysis. There exist a large number of clustering algorithms in the literature. The choice of clustering algorithm depends both on the type of data available and on the particular purpose and application. The data mining clustering techniques are used to cluster the city ration card based on the type of cards. The preprocessed and clustered data are analyzed to evaluate the problems faced in the rationing system and the possible solutions are established.

In clustering the algorithms are applied to discover interesting data distributions in the underlying data space. The formation of clusters is based on the principle of maximizing similarity between patterns belonging to distinct clusters. Similarity or proximity is usually defined as distance function on pairs of patterns and based on the values of the features of these patterns [7] [8]. Many variety of clustering algorithms have been emerged recently. Different starting points and criteria usually lead to different taxonomies of clustering algorithms [9]. Some commonly used partitioning clustering methods are K-Means (KM) [10], K-Harmonic Means (KHM) [11], Fuzzy C-Means (FCM) [12] [13], Spectral Clustering (SPC) [14] and several other methods.

b) Public Distribution System

The PDS is a centrally sponsored scheme that entitles beneficiaries to subsidized food grains every month [6]. Currently, beneficiaries are divided into the following groups: Below Poverty Line (BPL), Above Poverty Line and Antodaya Anna Yojana. The above mentioned challenges have been identified in the implementation of PDS. Some of them are as follows:

i. Targeting Errors

Separating beneficiaries of the PDS into three categories requires their classification and identification. Targeting mechanisms, however, have been prone to large inclusion and exclusion errors. Targeting error is where the targeted group of Below Poverty Line were denied ration cards. In 2009, it was estimated that about 61% of the eligible population was excluded from the BPL list while 25% of non-poor households were included in the BPL list.

Author α σ : Research Scholar, P.G. & Research Department of Computer Science, D. G. Vaishnav College, Arumbakkam, Chennai, Tamil Nadu, India – 600 106. E-mails : shanmusekar@gmail.com, Santhos2001@sify.com

ii. *Large leakages and diversion of subsidized foodgrain*

Foodgrain is procured by the centre and transported from the central to state godowns. Last mile delivery from state godowns to the Fair Price Shop (FPS) where beneficiaries can purchase grain with ration cards, is the responsibility of the state government. Large quantities of foodgrain are leaked and diverted into the open market during this supply chain.

iii. *Bogus Cards*

The bogus cards or false cards where a person owns duplicate cards and card in the name of non existing person.

iv. *Fair Price Shops*

The common problems faced by the public in the Fair Price Shops run by the state governments are underweight, poor quality and non availability of the essential commodities etc.

Targeted Public Distribution System (TPDS) is operated under the joint responsibility of the Central and the State/Union Territory (UT) Governments. Central Government is responsible for procurement, allotment and transportation of foodgrains upto the designated depots of the Food Corporation of India. Tamil Nadu Government is implementing Universal Public Distribution System (UPDS) and no exclusion is made

The details of family cards in circulation in Tamil Nadu are as follows:

Sl. No.	Type of Card	Commodities entitled	No. of Cards
1.	Rice Cards	All Commodities	1,67,21,538
2.	Antoydaya Anna Yojana Cards	All Commodities	18,62,615
Total Rice Cards			1,85,84,153
3.	Sugar Cards	All Commodities except rice	10,76,552
4.	Police Cards	All Commodities	61,061
5.	No Commodity Cards	No Commodity	60,827
Total			1,97,82,593

II. CLUSTER BASED PDS

In spite of the effort taken by the central and the state governments, the struggle faced by poor public, which is a usual scenario in almost every fair price shop is the main motivation for this research work. The main aim is to develop analytical data mining methods that can systematically address the composite problems assisting to

- Perform fault analysis to detect error patterns.
- Formulate an approach for error prevention and reduction.
- Categorize and analyse the common error patterns and minimise further occurrences of similar incidence.

The present research work proposes the use of an amalgamation of various data mining techniques to achieve the following objectives:

based on the income criteria [1]. The state government has made the universal public distribution system 'poor friendly' by ordering rice at free of cost under public distribution system to all eligible card holders from 01.06.2011.

The public distribution system has been implemented through **33,222 fair price shops** comprising of **31,232** shops run by the **Cooperative Societies**, **1,394** shops run by the **Tamil Nadu Civil Supplies Corporation** and **596** shops run by **Women Self Help Groups**. Out of 33,222 fair price shops, **25,049** are **full time** shops and **8,173** are **part time** shops. Family cards are issued to the people of the State based on the option exercised by the individual family. The family cards are segregated as Rice Cards (rice with all other commodities), Antyodaya Anna Yojana Scheme cards, Sugar Cards (additional sugar in lieu of rice and all other commodities), and no ration commodity cards (cards for identification purposes). Besides, police personnel are issued with family cards in distinct colour. Transgender living in a house as a group is treated as a family and family cards are issued to them. As on 31.03.2012, **1,365 transgender family cards** have been issued [2]. This statistical information is taken from the website of Tamil Nadu civil supplies corporation.

- To develop a data cleaning/filtering algorithm that cleans the dataset.
- To explore and enhance clustering algorithms to identify the error patterns from historical data.
- To explore and enhance classification algorithms to predict the future trends.

a) *Methodology of Cluster based PDS*

Pre-processing has been used to keep the data set ready for the process. Entity extraction has been used to automatically identify person, address, and type of card. Filtering techniques has been used to filter the incomplete, noisy and inconsistent data. The data mining clustering techniques are used to cluster the city ration card based on the type of cards. The preprocessed and clustered data are analysed to evaluate the problems faced in the rationing system and the possible solutions are established.

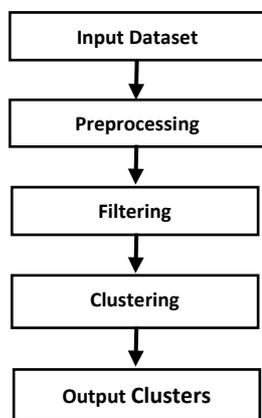


Figure 2.1 : The chart for cluster based PDS

i. Preprocessing

Most of the data collection techniques like survey studies, field experiments etc produce huge amount of information where the data are incomplete, noisy, inconsistent and contains missing values. Data preprocessing is a process that consists of data cleaning, data integration and data transformation which is usually processed by a computer program. It intends to reduce some noises, incomplete and inconsistent data. The results from preprocessing step can be used to proceed further by data mining algorithms.

ii. Filtering

Filtering is an initial step that is performed on the datasets in order to eliminate attributes that are unnecessary for the data mining process. This is an important task as some data mining tasks such as clustering (DB Scan, EM) and Association (Apriori) involving large amounts of data takes up large amount of time and memory. Filtering when not performed will slow down the application, to counter this problem the initial dataset instances which are populated by all the database attributes are filtered down to the most useful attributes depending on the applications requirements.

iii. K-Means Clustering

Given a set of objects, clustering is the process of class discovery, where the objects are grouped into clusters and the classes are unknown beforehand. Two clustering techniques, K-means and K-Harmonic means algorithms are considered for this purpose. The algorithm for k-means is given below [4].

$$V(D) = \sum_{i=1}^k \sum_{x_j \in C_i} (x_j - u_i)^2$$

Where V is the variance, C_i is a cluster, u_i is its mean, D is the dataset of all points x_j . Partition the dataset into k clusters such that intracluster variance is minimized.

The primary part of the algorithm is the standard K-means algorithm. Lets us assume that current partition of the N p -dimensional point into k clusters,

Compute the distances from each and every point to every cluster centroid and reassign. So for the simplest case of squared Euclidean distance, at every iteration, there is k computations of centroids, each one gets involved in p arithmetic means.

- k computations of centroids (Each of which involves p arithmetic means)
- $n * k$ distance computations, (Each of which involves p sums of squared differences)
- n minimums over k distances

K-means algorithm often has hierarchical clustering using LINKAGE concepts. Hierarchical clustering needs $n \times n$ distance matrix, while K-means requires only the $n \times p$ data matrix, p is often much smaller than n .

iv. K-Harmonic Means Clustering

K-Harmonic Means Algorithm is a center-based, iterative algorithm that refines the clusters defined by K centers [7]. KHM takes the harmonic averages of the squared distance from a data point to all centers as its performance function. The harmonic average of K numbers is defined as the reciprocal of the arithmetic average of the reciprocals of the numbers in the set.

$$HA((a_i | i = 1, \dots, k)) = K \left[\sum_{i=1}^k \frac{1}{a_i} \right]^{-1}$$

The KHM algorithm starts with a set of initial positions of the centers, then distance is calculated by the function $d_i, i = ||x_i - m_i||$, and then the new positions of the centers are calculated. This process is continued until the performance value stabilizes. Many experimental results show that KHM is essentially insensitive to the initialization. Even when the initialization is worse, it can also converge nicely, including the converge speed and clustering results.

III. EXPERIMENTAL DATASET

The dataset used in the present research work has been downloaded from the e-PDS portal Ministry of Consumer Affairs Food & Public Distribution, established to coordinate and integrate information resources created by the central government of India. The database is maintained by the central government of India. The dataset contains the allotted and off taken quantity of rice and wheat for all the states of India. This paper is concerned about only the state of Tamilnadu. The data are preprocessed, filtered and prepared for the clustering process with combined K- Means and K-Harmonic Means clustering techniques. The clustering process clusters the dataset based on the districts of Tamilnadu.

IV. RESULTS AND DISCUSSIONS

Experiments were conducted to analyse the efficiency of the clustering methods for the allocated

and off take quantity of food commodities like rice and wheat from the central government from 2001 to 2011 and the leakage or diversion of the commodities are calculated.

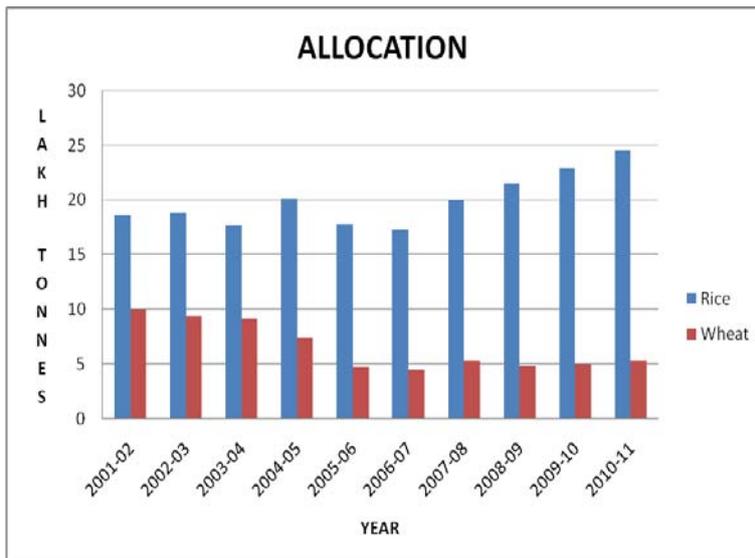


Figure 4.1 : Analysis of rice allotment

The graphs (Fig 4.1 & Fig 4.2) shows the allotted amount of rice and wheat from the central to the state government. The graph shows the amount of allocation and offtake of rice and wheat from the central government. There are ups and downs in the allotment from 2001 to 2005. The amount of rice allotted has been gradually increasing from 2006 to 2011. The offtake

quantity of rice and wheat has been fluctuating from 2001 to 2011. In Fig 4.3 the diversion of wheat and rice has been calculated and the graph was plotted accordingly, which shows there is a steep increase in the diversion of rice in the year 2005-2006 and again a steep hike in 2010-2011 and the offtake of wheat is decreasing since 2001.

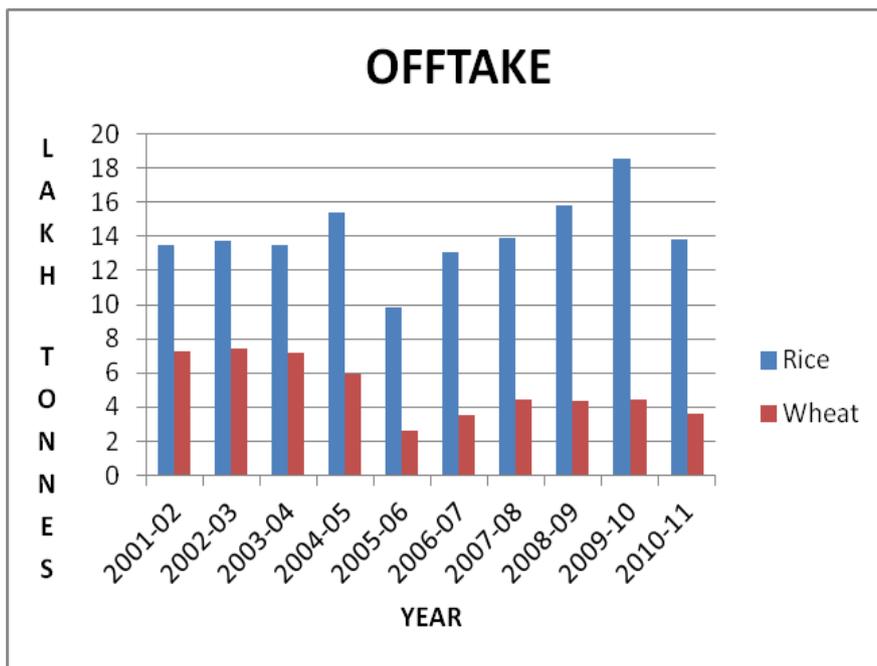


Figure 4.2 : Analysis of wheat allotment

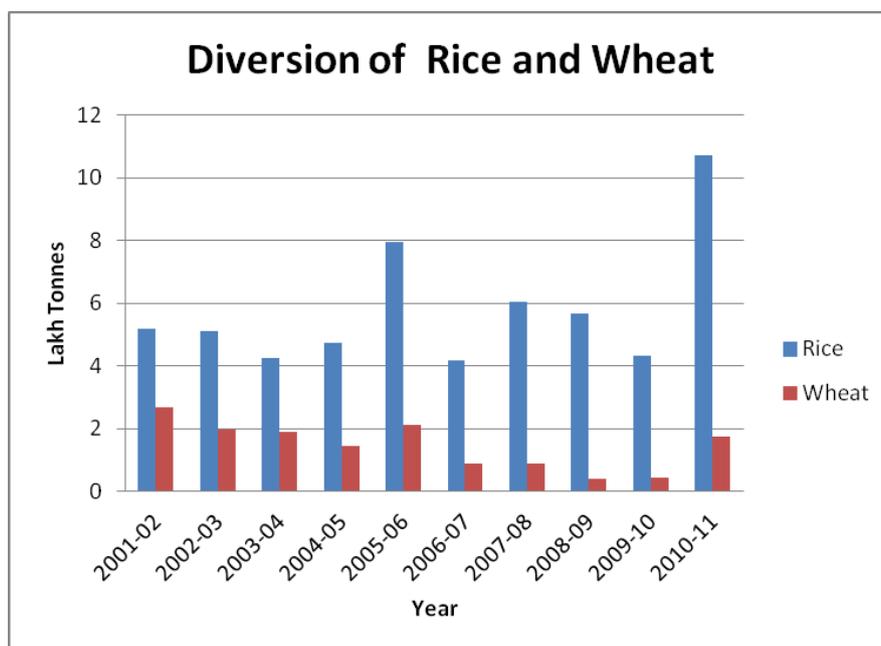


Figure 4.3 : Diversion of rice and wheat analysis

V. CONCLUSION AND FUTURE WORK

In spite of the steps taken from the central and the state government, eradication of the targeting problem can be established by the clustering techniques, where the different types of cards based on the food commodities of rice and wheat are clustered. The diversion of the food commodities can be tracked by the global positioning system installed in the transportation vehicles to the state godowns. This has been tested in the state of Tamilnadu and Himachal Pradesh and has been reported successively. The fair price shop management can be improved by sending SMS alert to the public about the food commodities availability and distribution on daily or weekly basis, so that the people can avoid the rush and repetitive visit to the fair price shop.

After Clustering the data can be classified to identify future fault that are emerging newly by using outlier detection on data. Experimental results prove that the algorithm is effective in terms of analysis speed, identifying common patterns and future prediction. The combined algorithm has promising value in the current changing scenario and can be used as an effective means by the Indian government PDS and Civil supplies corporation for fault detection and prevention.

REFERENCES RÉFÉRENCES REFERENCIAS

1. Policy note on Food and Consumer Protection, 2012-2013, <http://www.tn.gov.in/policynotes/>
2. Tamilnadu Civil Supplies Corporation, <http://www.tn.csc.tn.gov.in/html>
3. Fayyad, U.M., & Uthurusamy, R. (2002) Evolving data mining into solutions for insights. *Communications of the ACM*, 45(8), 23-31.
4. Machine Learning & Computational Biology Research Group (2011), <http://agbs.kyb.tuebingen.mpg.de/wikis/bg/2.pdf>, (July 1, 2011).
5. Malathi, A, Enhanced Algorithms to Identify Change in Crime Patterns, *International Journal of Combinatorial Optimization Problems and Informatics*, Vol. 2, No.3, Sep-Dec, 2011, pp. 32-38, ISSN: 2007-1558.
6. The PRS Blog The official blogsite of PRS Legislative Research <http://www.prsindia.org/theprsblog/>
7. A.K. Jain, R.C. Dubes, "Algorithms for Clustering Data," Prentice-Hall, Englewood Cliffs, NJ, 1988.
8. J.T. Tou, R.C. Gonzalez, "Pattern Recognition Principles," Addison-Wesley, Reading 1974.
9. Chonghui GUO, Li PENG, "A Hybrid Clustering Algorithm Based on Dimensional Reduction and KHarmonic Means," *IEEE* 2008.
10. R C Dubes, A K Jain, "Algorithms for Clustering Data," Prentice Hall, 1988.
11. B Zhang, M C Hsu, Umeshwar Dayal, "K-Harmonic Means-A Data Clustering Algorithm," *HPL- 1999-124*, October, 1999. K. Elissa.
12. X B Gao, "Fuzzy Cluster Analysis and its Applications," Xidian University Press, 2004(in china).
13. Wu K L, Yang M S., "Alternative fuzzy c-means clustering algorithm," *Pattern Recognition*, 2002, 35(10): 2267-2278.

14. A Y Ng, M I Jordan, Y Weiss. "On spectral clustering: Analysis and an algorithm," In: Dietterich, T G Becker S, Ghahraman; Z. Advances in Neural Information Processing Systems 14, Cambridge: MITPress, 2002, 849-856.

