



Applying K-Means Clustering Algorithm in Item Response Theory

¹B. Venura Lakshman*, ²J. V. Wijayakulasooriya, ³Manjula Sandirigama

¹Department of Practical Buddhist Studies, Bhiksu University of Sri Lanka, Sri Lanka

²Department of Electrical & Electronic Engineering, University of Peradeniya, Sri Lanka

³Department of Computer Engineering, University of Peradeniya, Sri Lanka

Abstract—Data Clustering is used to extract meaningful information and to develop significant relationships among variables. We use Item response Theory to calculate probabilities of students & data clustering technique named k-means clustering is applied to divide students into groups. The selected sample was students of Teachers' College studying Primary Mathematics. It is an essential part in a proposed system of Agent based E Learning System .In the system this operation will be executed automatically.

Keywords: Agent based E Learning System, Item response Theory, k-means clustering.

I. INTRODUCTION

The North Central Province, the largest province in Sri Lanka, has encountered with various problems related to the education field. Gaining poor results in Mathematics in national level exams, especially in the Ordinary level, is one of them. Although the knowledge of Mathematics is very important to the Implementation of every stream of subjects, statistical data of the Department of Examinations reveals that more than fifty percent of candidates who sat for the Ordinary level Mathematics got poorer (W) grade. [1] According to the survey conducted by the author, it reveals that the poor knowledge of Primary Mathematical Concepts is one of main reasons for this situation. Introducing Agent based computer solutions is one of remedies for this. In order to introduce such a system, it is necessary to make psychosomatic measurements of the students. Item Response Theory (IRT) is a commonly used for such measurements. [2] To adopt Item Response Theory into Agent programming, it is necessary to introduce a new model with Item Response theory. In this paper, it shows how to introduce such model with Item Response Theory & use to develop Agent based algorithm on modified Item Response theory.[3]

In this research, it is essential to collect data of primary students' Mathematics knowledge relevant to subject areas. In order to collect such information, the author conducted several surveys in schools selected in the North Central Province of Sri Lanka.

At first, initially, the author conducted a survey covering all subject topics in Primary Mathematics in selected schools in the North Central Province. Afterwards, the second survey was conducted by selecting lower graded topics in the previous survey.

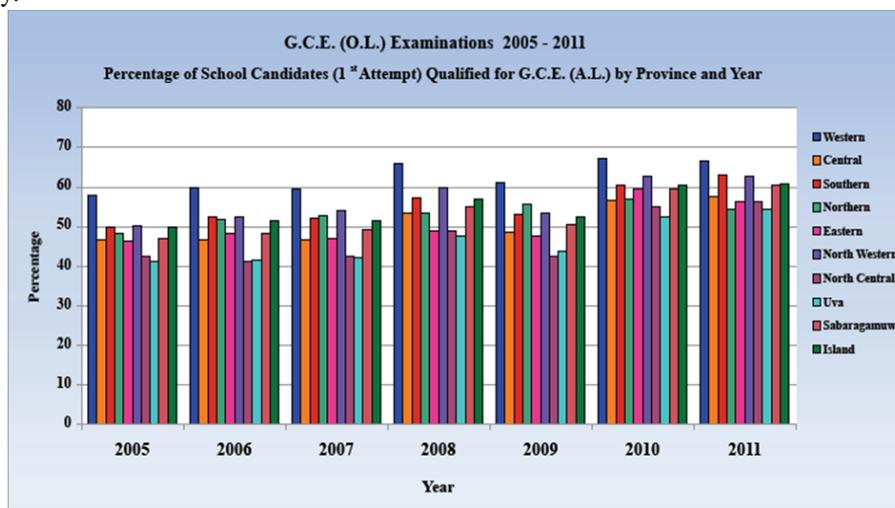


Chart 01- Performance of Students in Provincial Vice

II. DATA COLLECTION AND PRE-PROCESSING

Chart 01 shows that students of the North Central Province got lower results in Ordinary Level Examination compared to other provinces.

Chart 02 shows that students of the North Central Province got low marks for subject like Geometry, measurements... etc.

Chart 03 shows the same subject areas that are in primary level students' scores in the survey conducted by the Author

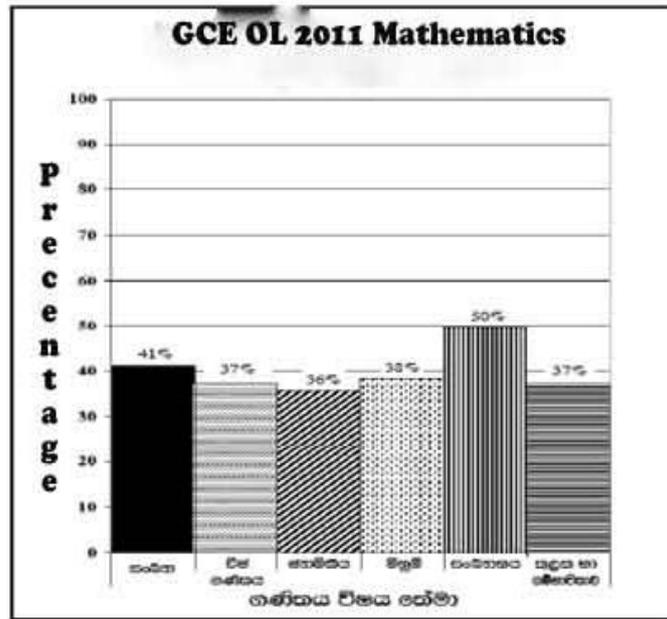


Chart 02 – Performance of students in Subject Vice

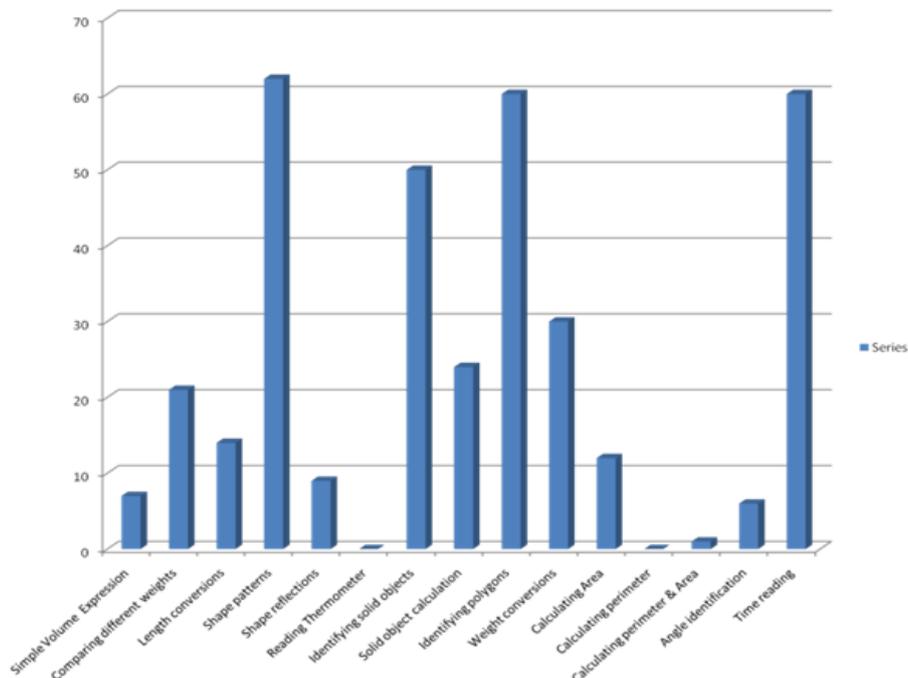


Chart 03 Performance of students in Subject Vice in the survey

III. RESULTS OF SURVEYS

It is clear that weaker Mathematics topics for students in Ordinary Level Examination & those in Primary Level Mathematics are almost similar. Therefore, this research reveals that results for the lower grade in Mathematics especially in the North Central Province is a root cause of lower knowledge in Primary level Mathematics. As such, it is essential to enhance Primary Mathematics knowledge in Primary Students in the North Central Province.

In addition, it is understood that there is a severe dearth of primary Mathematics teachers in the North Central province. Therefore, introducing Agent based system for primary Mathematics may be a great contribution to overcome the shortage issue of primary Mathematics teachers.

This can be performed with a computer System. Such system is called Computer based Adaptive Testing (CAT). In developing Agent based system with IRT; it is used Computer based Adaptive testing method.

IV. IRT MODEL

In this paper, agent based system is proposed to develop basic mathematical knowledge of primary level students. Before developing such a system, it is essential to develop model of the system. A brief description of the system as follows.

At the initial level, the question bank is prepared in three levels of questions. They are Hard (H), Middle (M) and Low (L). These categories & questions are prepared according to the results of initial surveys. At first, for any other student, the system fired 09 questions as a mix of all three categories randomly selected. A student belongs to H the probability of answering question q1 correctly given by $P(\xi q/H)$.

As the difficulty of each question is previously calculated, probability of $P(\xi q/H)$, $P(\xi q/M)$, $P(\xi q/L)$, can be calculated. This is shown by following tree diagram. Then we can categorize the student who has faced to the initial test, belongs to which group i.e. Hard (H), Middle (M) and Low (L) by the final conditional probability of each student. Then the system presents relevant lessons to the student according to students' level.

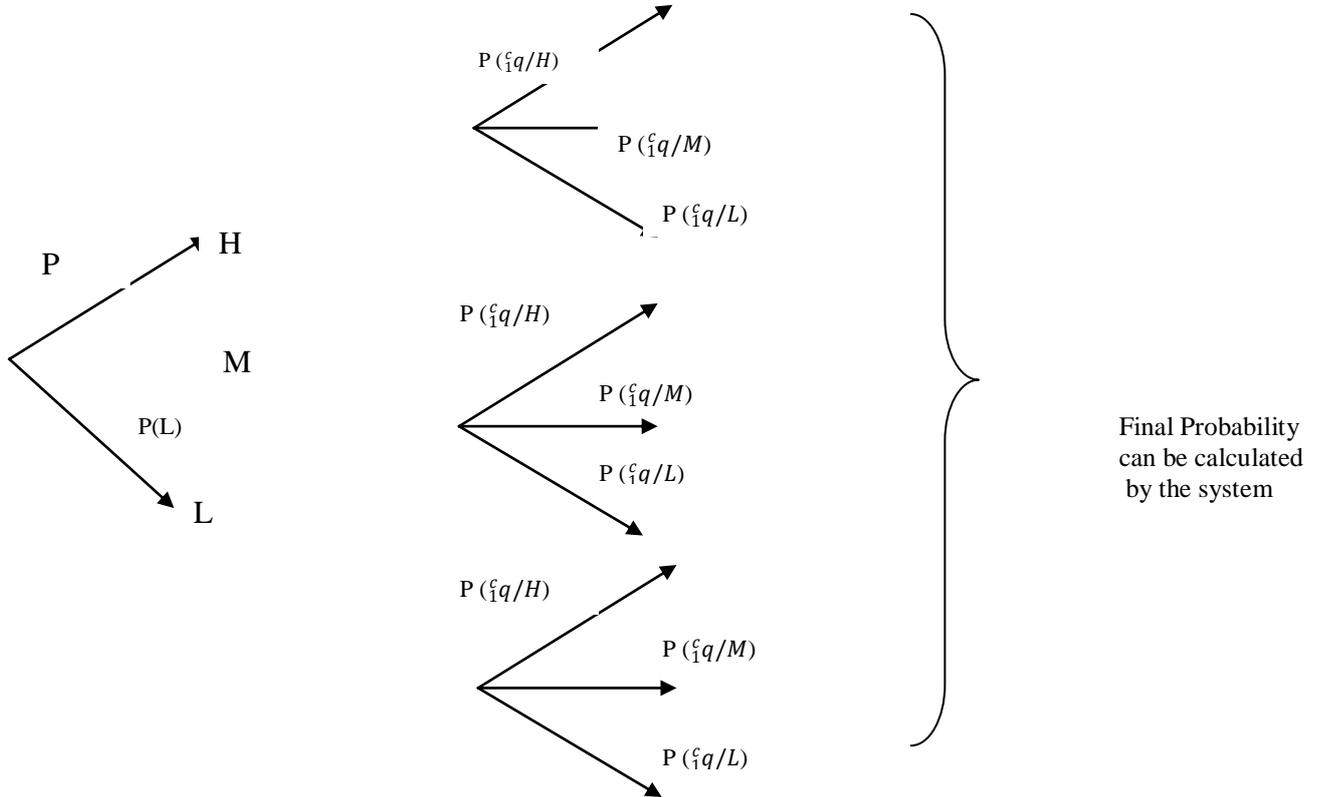


Fig.01 –Calculating Probability

V. CLUSTERING STUDENTS ACCORDING TO PROBABILITY

Data clustering is a process of extracting previously unknown, valid, positional useful and hidden patterns from large data sets.[4]Data Clustering is unsupervised and statistical data analysis technique. It is used to classify the same data into a homogeneous group.

In this paper, it is manually calculated Probability according to Item Response Theory using Microsoft Excel & k means algorithm was implemented by using Visual Basic. Results are shown below.

1	Student no	Item1	Item2	Item3	Item4	Item5	Item6	Item7	Item8	Item9	Item10	Correct	Incorrect	Ability	P
2	1	1	0	1	1	0	0	0	0	1	0	4	6	0.4	0.88079
3	2	1	1	1	1	0	0	0	0	0	0	4	6	0.4	0.82532
4	3	1	1	1	1	0	0	0	0	1	0	5	5	0.5	0.85452
5	4	1	1	1	0	0	0	0	0	1	0	4	6	0.4	0.76002
6	5	1	0	1	1	0	1	0	0	0	0	4	6	0.4	0.69989
7	6	1	0	1	1	0	0	0	0	1	0	4	6	0.4	0.7557
8	7	1	1	1	1	1	0	0	0	1	0	6	4	0.6	0.86446
9	8	1	0	1	1	0	0	0	0	1	0	4	6	0.4	0.78994
10	9	1	0	1	1	0	1	0	0	0	0	4	6	0.4	0.72402
11	10	1	1	1	1	0	0	0	1	1	0	6	4	0.6	0.81755
12	11	1	1	1	0	0	0	0	0	1	0	4	6	0.4	0.74265
13	12	1	1	1	0	0	0	0	1	1	0	5	5	0.5	0.88624
14	13	1	0	1	1	0	0	0	0	1	0	4	6	0.4	0.80085
15	14	1	0	1	1	0	0	0	0	1	0	4	6	0.4	0.75461
16	15	1	1	1	0	0	0	0	0	1	0	4	6	0.4	0.78778
17	16	1	1	1	1	0	1	0	0	1	0	6	4	0.6	0.78221
18	17	1	0	1	1	0	1	0	0	1	0	5	5	0.5	0.77368
19	18	1	0	1	1	0	0	0	0	1	0	4	6	0.4	0.77058
20	19	1	1	1	1	0	0	0	0	1	0	5	5	0.5	0.75569
21	20	1	1	1	1	0	0	0	0	1	0	5	5	0.5	0.86027
22	21	1	0	1	1	0	0	0	0	1	0	4	6	0.4	0.88445
23	22	1	0	1	1	0	1	0	0	1	0	5	5	0.5	0.78482
24	23	1	1	1	0	0	0	0	0	1	0	4	6	0.4	0.77058
25	24	1	0	1	1	0	0	0	0	1	0	4	6	0.4	0.88324

Fig2 Calculating Probability using Microsoft Excel

After applying k means algorithms to the data set

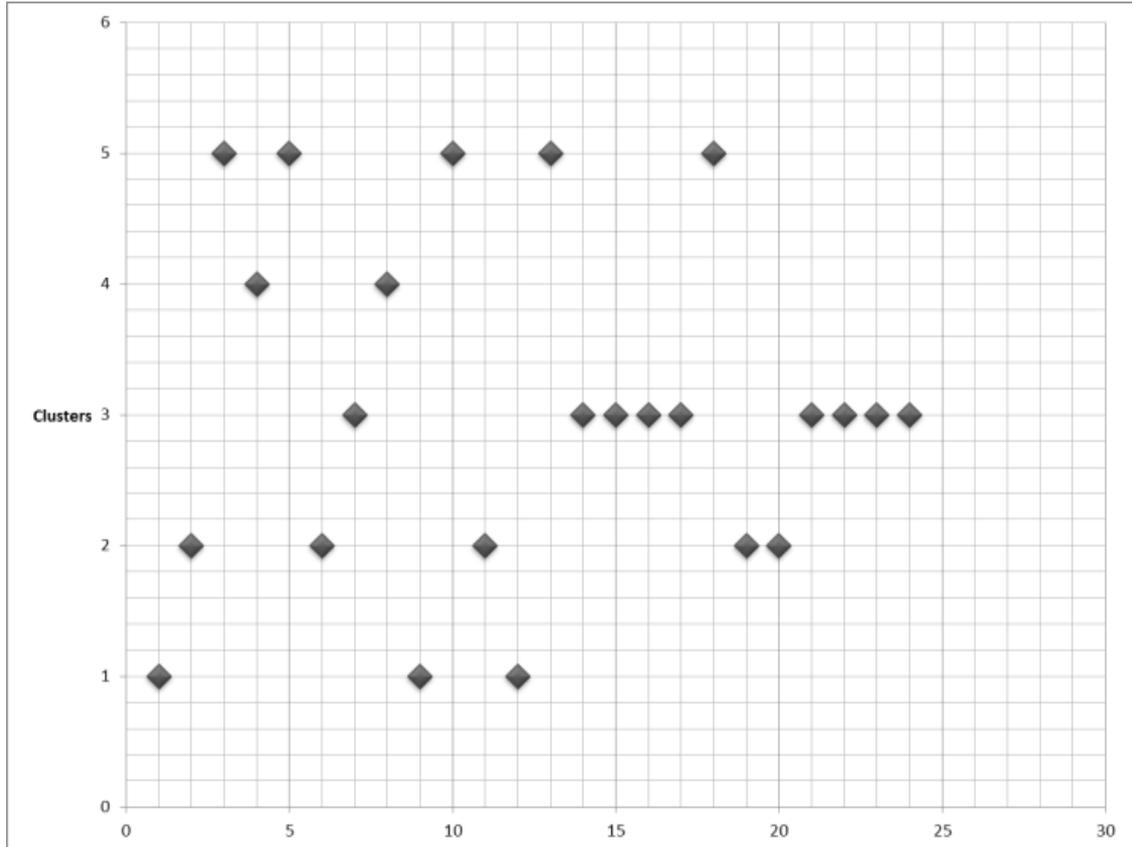


Chart 04 – After Clustering students

VI. CONCLUSION & FUTURE DIRECTION

In this paper it is shown that we are able to use k means algorithm in clustering students according to their probability which is implemented by manually in this case. But in proposed system, we can do it automatically. A Flow chart of proposed system is shown in annex 1. Therefore in proposed system, we are able to implement k means to cluster students according to the probability that they gained in the basic examination.

REFERENCES

- [1] "Department of Examination, 2011. Statistical handbook, 2008 – 2010", Department of Examination, Colombo: Department of Examination.
- [2] Chong Ho Yu, "a Simple Guide to the Item Response Theory (IRT) and Rasch Modeling:" (<http://www.creative-wisdom.com>), December 10, 2010.
- [3] "Item Response Theory", 2013. Available at: < <http://echo.edres.org:8080/irt/>> [Accessed 10 October 2013].
- [4] Han, J. and Kamber, M., (2006) "Data Mining: Concepts and Techniques", 2nd edition. The Morgan Kaufmann Series in Data Management Systems, Jim Gray, Series Editor.
- [5] Connolly T., C. Begg and A. Strachan (1999) Database Systems: A Practical Approach to Design, Implementation, and Management (3rd Ed.). Harlow: Addison-Wesley. 687

Annex 01

