Scientific Journal of Impact Factor (SJIF): 5.71
e-ISSN (O): 2348-4470
p-ISSN (P): 2348-6406
International Journal of Advance Engineering and Research Development
Volume 5, Issue 05, May -2018

RESULTS OF IMPLEMENTING HYBRID CAT USING IRT AND NAÏVE BAYES

Nikita Singhal [1]; Amitoz S Sidhu [2] *; Ajit Kumar Pandit [3]; Shailendra Pratap Singh Sengar [4]; Tutu Kumari [5]

1,2,3,4,5 ARMY INSTITUTE OF THECHNOLOGY, PUNE

Abstract: *With the rapid research happening in the field of education and testing many systems implementing the latest techniques like Item response theory and machine learning have been conceived. In this paper we will look at implementation of a system that aims at employing different fields of research to develop a comprehensive testing platform. The system sets itself apart by attempting to be self sufficient such that it requires little or no human intervention required. This is achieved through automatic question acquisition and classification through a community run forum and a dynamic database that automatically transforms itself in accordance with the trends in the test takers responses. It is based on Item response theory to get item characteristic classification for the question sets that allows for an efficient test generator that can effectively test users across a larger section on latent scale. It also provides a comprehensive result generation that informs the examinee about the various patterns in test thus allowing him to better prepare.*

***Keywords**: Item Response Model · Naive Bayes Model · CAT (Computer Adaptive Test) · 2 - Parameter Model · Recommendation System.*

I. INTRODUCTION

Many researchers and institutions have endeavored to provide advanced testing systems and platforms for informed learning. These systems while efficient suffer from drawbacks like they are difficult to maintain and require regular maintenance. They are often vulnerable to questions with outdated effectiveness i.e. while the questions are good at testing the targeted trait, over time these questions become known and their effectiveness is reduced. The proposed model in this paper tackles these problems while also providing the testing capabilities of other implemented system without any compromise. It also incorporates a comprehensive report generator [1] that analysis patterns of the examinee along with a comparative study against the top performers. The system is divided into the major modules namely the database manipulator module, the test module and the report module. The working of these modules will be explored further in the paper.

II. PROPOSED SYSTEM

The task separates itself by using both machine learning and IRT [2]. The paper proposes a framework, with an advancing database, which actualizes effective testing and scaling hypothesis while also implementing a complete report generator module. The undertaking intends to actualize a group run discussion that is utilized to populate the database where the inquiries are progressively scored. The proposed show gives a percentile score as well as an inside out analysis of the performance. It recognizes the patterns in the performance of the examinee thus providing an insight into how the examinee can improve his performance.

Actualizing of said system meets the objective of building up a framework with increased accuracy of estimating the learner's true ability while tending to the disadvantages of the current framework
The proposed system can be broadly divided into three main modules.

a) Database manipulator.
b) Adaptive test and report generator.
c) Forum.

Firstly, the forum module that employs the Bayes Model of classification. This module will be used to populate the database after correct classification of the questions picked from the forum.
Furthermore, we have the database control module, this module is in charge of the dynamic scaling of the inquiries in the database in light of their discrimination and difficulty. This scaling depends on the information about the response of various examinees on that inquiry. This information contains the quantity of right reactions recorded and the normal time to accomplish the said reactions.
The final module is the actual test that the examinee takes, this employs Item Response Theory Model and adaptive test principles to rate tested trait of the examinee and generate a comprehensive performance report.

III. SYSTEM ARCHITECTURE

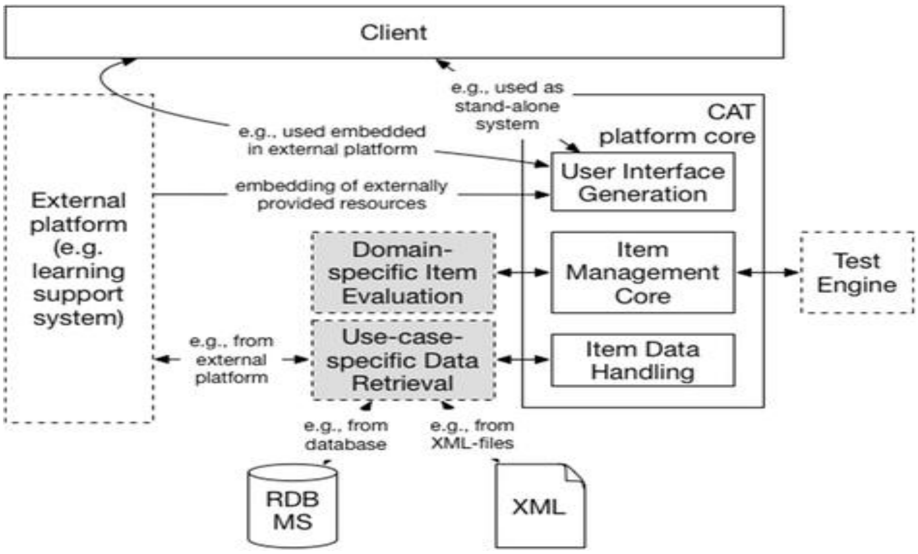

Figure 1. System Architecture

The above architecture of the proposed system shows the interactions between the main components of the proposed system. The Item Management core is responsible for the data regarding the questions it is responsible for the initial classification of the item/question into the correct dataset. Apart from this it is responsible for maintaining the metadata for the question (items) throughout their life in the dataset, this includes data like the total selections of the question and the total correct responses. This module basically maintains the parameters, the difficulty and the discrimination for the 2-parameter model employed in system, for the items that are required by the item management core for IRT application.
The Item management core is responsible for the activities under the IRT 2-parameter model. This includes the calculation and formulation of the item characteristic chart for the items in order to decide weather that item is applicable for testing a pre-decided trait level. The item management core works under the Test Engine as the engine sets the level against which the trait is to be judged based on this the item management core proceeds with its calculation.
The test engine is responsible for setting up the test paper based on the input from the user and using upon the services provided to it by the Item Management Core.
The External Platform shows the future scope where the system can be integrated with an external system which could be based on A.I or machine learning.
The client side accesses the platform as a web-based system while the heavy lifting is taken care of by the server-side processes.

IV. COMPARISON

A. Naïve Bayes vs other classification

Naive Bayes belongs to the family of probabilistic algorithms which works particularly well in text classification in NLP and spam filters. It easy to implement and provides extremely fast convergence especially when the NB conditional independence holds. In such cases it even out performs discriminative models like logistic regression. It is fast and easy to train as it requires very little training data as compared to other classifiers.
In our case we are using naïve Bayes to automatically classify a submitted question and place in the question bank in the database. It is important to note that in this case we are dealing with text classification of the questions and that the conditional independence holds. These ate ideal conditions for Bayes classifiers which is used in spam filters under same conditions. This allows the Bayes to fully show its capabilities. Since our dataset was also small high bias/ low variance classifier is bound to do better than a low bias/ high variance classifier like kNN, since the latter will tend to overfit

In the paper "On Discriminative classifiers: naive Bayes and logistic regression" [3] by Professor Andrew Ng and Professor Michael I Jordan provides a mathematical proof of error properties of both models. They conclude that when the training size reaches infinity the discriminative model: logistic regression performs better than the generative model Naive Bayes. However, the generative model reaches its asymptotic faster (O (log n)) than the discriminative model (O (n)), that is the generative model (Naive Bayes) reaches the asymptotic solution for fewer training sets than the discriminative model (Logistic Regression). This behavior is best represented by the experiment conducted by Ng & Jordan where they did predictions for 15 datasets from the UCI machine learning repository.

kNN on the other hand needs proper tuning of the K this includes model selection either with training-validation split or cross validation. kNN [4] also tends to be slower in case of prediction as compared to naïve Bayes. kNN is O(n) while Bayes is O (1) thus taking the response time into consideration naïve Bayes was selected

B. IRT models

The Item Response Theory (IRT) models are more based on the probabilistic models and are rooted in theoretical assumptions. Item Response Theory mainly focusses on the item level information in contrast to the CTT [5] which is based on test level information. This implies is that while the classical model rates the entire test the item response model rates individual items on that test. What this directly translates into is that IRT is more robust and flexible when it comes to grading the latent trait. Latent here implies that the responses are to be taken as a manifestation of hypothetical traits that the IRT uses to grade. Classical test theory is still the preferred for a smaller candidate size for example 60 examinees while IRT truly shines above 300 or so examinees.

Before comparing models let us see what the parameters [6] are that IRT is based on
Location: This is analogous to difficulty in the CTT. It can be understood as how many examinees are able to solve a particular problem. Lesser the number more the difficulty.
Discrimination: This indicates how strongly the item is related to a latent trait. This shows the level of understanding of a trait that would be needed to solve the problem.
Guess ability: This shows the probability that the answer was guessed.

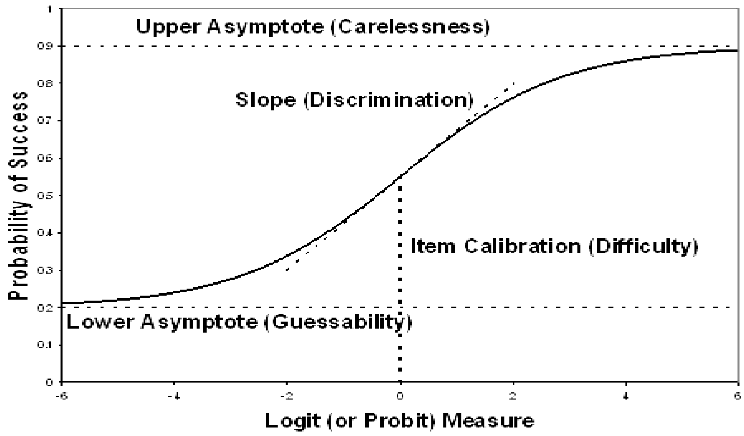

Figure 2. Item Characteristic curve

The reason we have chosen the two-parameter model is because it provides us with flexibility and economy while calculating the traits.
The one parameter model or the Rasch model uses only the difficulty as a trait this is quite similar to a CTT and thus it was not selected for the system. Since only difficulty is used it provides limited flexibility and the different items are differentiated based on the location on the graph.

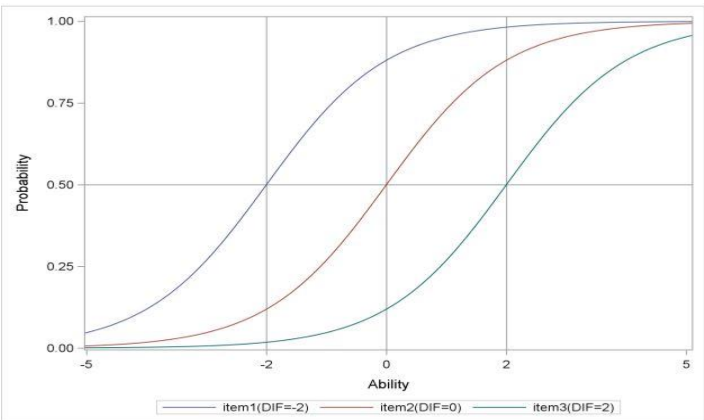

Figure 3. Different Difficulty of Item

On the other hand, the three-parameter model uses the 2 parameter model parameters but also considers guess ability. For our application we traded off this parameter for speed and instead tried to incorporate it by judging the speed of answering. Thus, the benefits of 3 parameter model with 2 parameters.
In itself the 2-parameter model is quite efficient as it uses both difficulty and discrimination as a parameter.

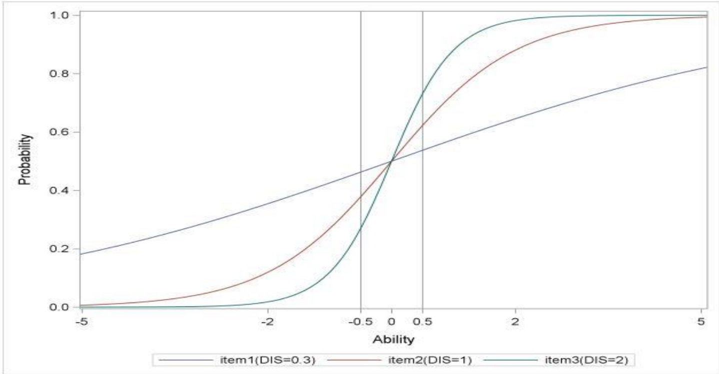

Figure 4. Different Discrimination curve

As seen above the item are not only differentiated based on the location as in Rasch model but also the slope this allows for better classification of the examinee on the trait scale.

V. SUMMARY/CONCLUSION

This study [7] proposes a new system that provides an efficient testing platform that can overcome the drawbacks of the existing systems while providing an adaptive environment that reduces the maintenance cost and changes with the trends observed in the examinees over time. Experiment results show that the proposed system provides precise results for examinees performance which is comparable to more costly systems. It is a comprehensive testing platform that employs a slew of different technologies to provide a low-cost testing system. This system brings together different technologies like Item response theory and machine learning methodologies to develop a system that evolves with every use thus adapting to the examinees whereby it can precisely and efficiently test examinees against a larger section of the trait scale. This system aims at providing the efficiency of high end costly systems while providing the robustness of a community run platform.

The project sets itself apart by harnessing the powers of both machine learning and the IRT. The project apart from scaling the tested latent trait of the examinee and providing a comprehensive performance report, also attempts to develop a truly adaptive system by making even the database adaptive. The project implements a community run forum that is used to populate the database where the questions are dynamically scored on a difficulty and discretion scale. The proposed model not only provides a percentile score but also an in-depth performance comparison with the other examinees including the time taken to answer the answer, allowing for a better understanding of the strengths and weaknesses of examinee and his current standing in the crowd. The proposed system has tremendous scope for future development by leaning towards the fields of Artificial Intelligence and Machine Learning. The system explored in this report was proposed after performing detailed research in the field of adaptive learning and studying the advantages and disadvantages of the various models of Item Response Theory.

Implementing the afore mentioned model meets the goal of developing a system with increased accuracy of estimating the learner’s true ability while addressing the drawbacks of the existing system.

VI. REFERENCES

[1] Owen Conlan, Cord Hockemeyer, Vincent Wade, Dietrich Albert Metadata Driven Approaches to Facilitate Adaptivity in Personalized eLearning Systems
[2] R.K. Hambleton and H. Swaminathan : IRT Principles and Applications:KLuwe Nijhoff,1985
[3] Professor Andrew Ng , Professor Michael I Jordan: “On Discriminative classifiers: naive Bayes and logistic regression”
[4] Mohammed J.Islam ,Q.M. Jonathan Wu : Investigating the performance of Naïve- Bayes Classifiers and K- Nearest Neighbor Classifiers
[5] Xitao Fan : Item Response Theory and Classical Test Theory: An Empirical Comparison of their Item/Person Statistic
[6] Deborah Harris: Comparison of 1-, 2- and 3- parameter IRT Models: DOI**:** 10.1111/j.1745-3992.1989.tb00313.x, (1984).
[7] Nikita Singhal, Amitoz S Sidhu, Ajit Kumar Pandit, Shailendra, Tutu Kumari: “Hybrid Cat using Bayes Classification and two parameter Model” ISBN 978-981-13-0616-7