Smart Tutoring using Statistical Modelling in the IoT

Sakchhi Mishra[1] Yamini Chouhan[2]
[1]M.Tech Student [2]Assistant Professor
[1,2]Department of Computer Science & Engineering
[1,2]Shri Shankaracharya Technical Campus, Bhilai, India

Abstract— Smart tutoring has established a new terminology to designate technological and societal improvements (e.g., Big and Open Data, IoT, RFID, and NFC) enable operative, skilled, attractive and custom-built learning. IoT is neither science fiction nor industry buildup however is rather in view of strong innovative advances and dreams of arrange universality that are energetically being figured out. Devices used in IoT produces huge amount of data which attract attention towards applying statistical modelling over data produced. In this paper we have proposed a smart tutoring which uses regression statistical model which will take the student class test marks and other activity marks as input dataset which is produced by IoT devices, our proposed statistical modelling based algorithm will predict the student performance, we have also passed the produced output to different classifiers to check the accuracy of prediction.
Key words: IoT; ML

I. INTRODUCTION

IoT, the word first presented by Kevin Ashton in 1998, is a future of Internet and pervasive computing. This technical revolution characterizes the future of reachability and connectivity. In IoT, 'things' state to some object on the face of the Globe, whether it is a collaborating device or a non- collaborating dumb object. In the whole world whatsoever can be part of Internet. The objects become connecting nodes over the Internet, through data communication channels, mainly from side to side Radio Frequency Identification (RFID) tags. IoT include some smart objects too. Objects those are are not only physical units, but also digital unit's ones and accomplish some jobs for humans and the atmosphere called Smart Objects. This is why, IoT is hardware as well as software prototype, but also include interaction and social features on top. Other than portraying the frameworks and things of IoT, various later reviews stressed that most things on the IoT should have knowledge, in this way are called "Smart Objects" (SO) and are expected fit for being recognized, detecting occasions, communicating with others, and settling on choices independent from anyone else.

Figure 1 shows the two distinct interaction modes that smartphones can enable in the IoT. Through direct interaction, a smartphone can query the state of an IoT device in its proximity and then provide a bridge between lowlevel peer-to-peer protocols, such as Bluetooth or Wi-Fi, and Internet protocols, such as HTTP and TCP. One example is the Fitbit ftness monitor, which uploads a user's step count through his or her phone over a 4G network to the user's account in the cloud. Through proxy interaction, mobile users who happen to be near an IoT-enabled object or device can look up associated information published by interested parties through a Web service using their smartphone, just as they would when performing a Web search [Roy Want et. al. IEEE 2015].

In IoT Sensors gather data, communication components relay the information gathered. In the future, enormous numbers of sensors will be deployed which will produce huge amount of data, contains hidden information (Knowledge) which can help to make better system. In this case Machine learning and data mining technique will play important role to extract knowledge from captured data form IoT system.

Earlier, a collection of hardware, software and online service comes into the market which claim of reforms to classrooms and tutoring methods. But the true commotion of education is yet to attain.

The institution goal is to develop a platform that provides real-time feedback and assistances online tutors become better at tutoring. For example, the system will perceive if a student's response to a concept follows a pattern of misinterpretation. By giving premature warning to teachers, the platform can help preclude problems further in the teaching learning.

"If we can aim to shape the performance of the teacher — the teacher being the significant input into a child's learning — then you're creating something truly powerful," says Tom Hooper, founder of Third Space Learning.

"With the increasing capabilities of machine learning, there is a unique opportunity to personalize learning to individual students," says Erik Choi, Principal Researcher at Brainly.

With the cumulative competences of machine learning (ML), there is an exceptional chance to personalize learning to each and every students. Even we can predict the performance of student so as to he/she can know, in which section they need improvement and can perform well.

Every student can access data that will help them along their one of a kind way of their expectation to learn and adapt. Later on, that implies that a student won't need to take in the same correct thing at the same correct pace. Figure-3 depicts how data gathered from IoT devices passed through preprocessing (Dimension Reduction) and Statistical Modelling i.e. ML (Machine Learning), outcome helps for decision making to organization.

Reading material, student's books and course material are generally custom-made for the normal understudy and imprinted in extensive numbers for a great many students to utilize. In any case, as we've all accomplished, not all educators and schools have a similar instructing style. By applying Statistical Modelling teachers and schools will be able to create textbooks and exercises that are made-to-order to the needs of their specific courses and students.

Content Technologies, Inc (CTI) is a company that is leveraging Deep Learning, to create assemble custom textbooks. After educators import their syllabus and material into CTI's engine, the system reads and masters the content,

and finds new patterns. The algorithms then use the gained knowledge to create textbooks and classroom material based on the core concepts. CTI is using the technology to provide publishing services to secondary and higher education institutions [www.thenextweb.com].

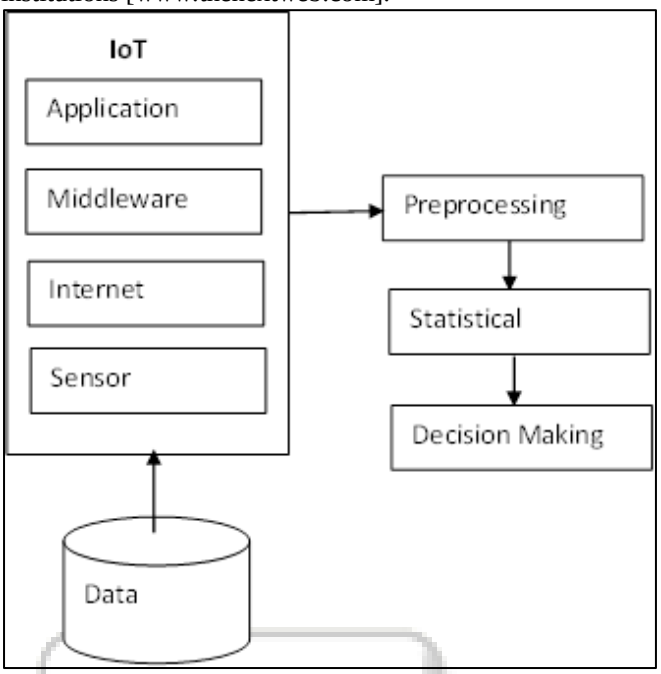


Fig. 1: IoT and Statistical Modelling

II. LITERATURE SURVEY

Chun-Wei Tsai, Chin-Feng Lai, Ming-Chao Chiang, and Laurence T. Yang [IEEE 2014] did review studies on applying data mining technologies to the IoT, which consist of clustering, classification, and frequent patterns mining technologies, from the perspective of infrastructures and from the perspective of services. The analysis and discussions on the scale of each mining technology and of the overall integrated system are also given. To make it easier for the audience of the paper to fully understand the changes brought about by the IoT, a discussion, which goes from the changes caused by the IoT in using mining technologies to the potentials to the open issues we are facing nowadays, is presented. Since the development of IoT is still at the early stage of Nolan's stages of growth model, it is often the case that the focus is first on the development of efficient preprocessing mechanisms to make the IoT system capable of handling big data and then on the development of effective mining technologies to find out the rules to describe the data of IoT. From the perspective of Nolan's stages of growth model, the issue to surface in the development of data mining for IoT is how to summarize and represent the mining results.

According to Najah Abu Ali et. al. IEEE 2012 discussed some of the data management solutions proposed for the Internet of Things, with a focus on mechanisms that target more efficient storage and communication for energy preservation. We have also outlined the challenges in providing comprehensive solutions for green IoT data management. Future work involves in-depth investigation and development of a data management solution that takes into consideration the sophisticated and heterogeneous architecture of IoT and integrates the needs flexible sources discovery and integration and better energy efficiency into its elements of design.

Feng Wang et. al. Hindawi 2014 Tese research angles come from other disciplines and are in the process of being adopted by the IoT. Our paper serves a key purpose: from the perspective of correlative technologies based on time, to review the evolutionary process of the IoT and depict the relations between the correlation techniques which are largely missing in current literature in which the focus has been more on the introduction and comparison of existing technologies and less on issues describing evolutionary process of the IoT. Author consider that the latter is crucial to understanding the evolution of the IoT. Trough generalizations of particular focus in different stages of each technology, author can better understand the current phase of the IoT and therefore predict future challenges. Tis paper aims to bridge this gap by providing guidance in terms of the evolutionary process of the IoT and gives readers a panoramic view of the IoT feld without repeating what is already available in existing literature so as to complement the existing IoT survey papers which have not covered the evolutionary process of the IoT.

Shen Bin et. al. IEEE 2010 author propose four data mining models for the Internet of Things, which are multi-layer data mining model, distributed data mining model, Grid based data mining model and data mining model from multi-technology integration perspective. Among them, multi-layer model includes four layers: 1) data collection layer, 2) data management layer, 3) event processing layer, and 4) data mining service layer. Distributed data mining model can solve problems from depositing data at different sites. Grid based data mining model allows Grid framework to realize the functions of data mining. Data mining model from multitechnology integration perspective describes the corresponding framework for the future Internet. Several key issues in data mining of IoT are also discussed.

Jayavardhana Gubbi et. al. Elsevier 2013 said that The proliferation of devices with communicating–actuating capabilities is bringing closer the vision of an Internet of Things, where the sensing and actuation functions seamlessly blend into the background and new capabilities are made possible through access of rich new information sources. The evolution of the next generation mobile system will depend on the creativity of the users in designing new applications. IoT is an ideal emerging technology to influence this domain by providing new evolving data and the required computational resources for creating revolutionary apps.

S.No.	Author/Title/Publication	Conclusion
1.	Chun-Wei Tsai, Chin-Feng Lai, Ming-Chao Chiang et. al./ Data Mining for Internet of Things: A Survey/IEEE 2014	In this paper, author review studies on applying data mining technologies to the IoT, which consist of clustering, classification, and frequent patterns mining technologies,

		from the perspective of infrastructures and from the perspective of services. The analysis and discussions on the scale of each mining technology and of the overall integrated system are also given. To make it easier for the audience of the paper to fully understand the changes brought about by the IoT.
2.	Najah Abu Ali et. al./ Data Management for The Internet of Things: Green Directions/ IEEE 2012	In this paper, author discussed some of the data management solutions proposed for the Internet of Things, with a focus on mechanisms that target more efficient storage and communication for energy preservation. Author have also outlined the challenges in providing comprehensive solutions for green IoT data management. Future work involves in-depth investigation and development of a data management solution that takes into consideration the sophisticated and heterogeneous architecture of IoT and integrates the needs flexible sources discovery and integration and better energy efficiency into its elements of design.
4.	Shen Bin, Liu Yuan et. al. / Research on Data Mining Models for the Internet of Things/ IEEE 2010	In this paper, author propose four data mining models for the Internet of Things, which are multi-layer data mining model, distributed data mining model, Grid based data mining model and data mining model from multitechnology integration perspective. Among them, multilayer model includes four layers (e.g. data collection layer, data management layer, event processing layer and data mining service layer). Distributed data mining model can well solve the problem arose from depositing data at different sites. At the same time, the complexity of problem is decomposed, and the requirements of highperformance, high storage capacity and high computing power for central nodes are reduced. Grid based data mining model.
5.	Jayavardhana Gubbi, Rajkumar Buyya et. al./ Internet of Things (IoT): A vision, architectural elements, and future directions/ Elsevier 2013	Presented here is a user-centric cloud based model for approaching this goal through the interaction of private and public clouds. In this manner, the needs of the end-user are brought to the fore. Allowing for the necessary flexibility to meet the diverse and sometimes competing needs of different sectors, we propose a framework enabled by a scalable cloud to provide the capacity to utilize the IoT. The framework allows networking, computation, and storage and visualization themes separate thereby allowing independent growth in every sector but complementing each other in a shared environment.
6.	Edmund W. Schuster,	This paper focuses on

	Sumeet Kumar et. al. /nfrastructure for Data-Driven Agriculture: Identifying Management Zones for Cotton using Statistical Modeling and Machine Learning Techniques/ IEEE 2011	the application of a clustering algorithm to field data with the goal to identify management zones. Author employ two sets of attributes, first yield and second field properties like slope and electrical conductivity to delineate the management zones. By definition, a management zone is a contiguous area defined by one or more features and may take on many different shapes. Building on the established machine learning approach of k-means clustering, author had successfully identify a near optimal number of management zones for a cotton field.
7.	Michail N. Giannakos, Demetrios G. Sampson and Łukasz Kidziński /Introduction to smart learning analytics: foundations and developments in video-based learning/ Springer 2016	The paper defines the subset of learning analytics that focuses on supporting the features and the processes of smart learning, under the term Smart Learning Analytics. This is followed by a brief discussion on the prospects and drawbacks of Smart Learning Analytics and their recent foundations and developments in the area of Video-Based Learning. Drawing from our experience with the recent international workshops in Smart Environments and Analytics in Video-Based Learning, author presented the state-of-the-art developments as well as the four selected contributions. The paper further draws attention to the great potential and need for research in the area of Smart Learning Analytics.

III. PROBLEM IDENTIFICATION

After going through several literature we came across following bottleneck of existing tutoring system as follows:

1) A prominent issue with any development programme is their monitoring, cost involved in programme and manpower availability.
2) Better implementation of student development programmes can be ensured only if those responsible for actual implementation are paid reasonably well, appropriately trained, and sufficiently motivated. But this has not been done as yet.
3) Value of MSE (Mean Square Error) is high while modelling between impendent and dependent parameters, hence accuracy decreases

A. Scope of smart tutoring

Connected devices can help make life easier for students with special needs. For instance, a visually impaired student who is given a special card that when registered by a computer, automatically enlarges font size. Rather than having to call a teacher over for help—costing both the student and the teacher time they could be using more productively—the student can take care of the issue, which also builds self-confidence and promotes independence.

Schools should be one of the places students feel most secure, but in light of recent events, it is easy to understand why, for some kids, that might not be the case. With the IoT, schools can create levers or buttons throughout the school that, when engaged, initiate a customized lockdown system. The lockdown can include, among other features: automatic perimeter security; immediate notification of authorities; and transmission of video to police so they can monitor intruder activity from headquarters and on the way to the building.

IV. PROPOSED METHODOLOGY

The institution goal is to develop a platform that provides real-time feedback and assistances online tutors become better at tutoring. For example, the system will perceive if a student's response to a concept follows a pattern of misinterpretation. By giving premature warning to teachers, the platform can help preclude problems further in the teaching learning.

We have proposed a smart tutoring based on regression statistical modelling, layout of our proposed tutoring scheme as follows:

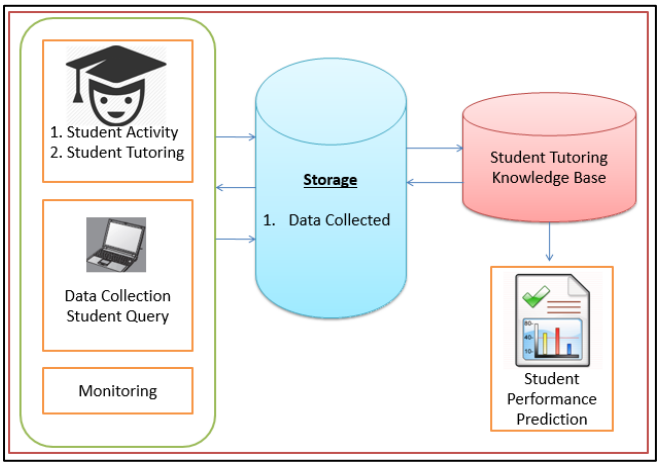


Fig. 2: Proposed Layout

A. Proposed Algorithm

1) Step-1. Input training dataset.
2) Step-2. Input test dataset.
3) Step-3. Process training dataset to find fitness function between dependent and independent variables of training dataset.
4) Step-4. Apply polynomial curve fitting.
5) Step-5. Find optimal fitness function as MSE (Mean square error) will be lesser.
6) Step-6. Apply regression statistical modelling as per optimal fitness function.
7) Step-7. Predicted data as output.
8) Step-8. Apply classifier to predicted output.

Curve fitting is the procedure of assembling a curve, or mathematical function that has the best fit to a series of data points, conceivably subject to constraints. Curve fitting can comprise either interpolation, where a strict fit to the data is essential, or smoothing, in which a "smooth" function is constructed that just about fits the data. Regression analysis, which focuses more on questions of statistical inference such as how much indecision is present in a curve that is fit to data observed with random errors. Fitted curves can be used as an aid for data visualization, to infer values of a function where no data are obtainable, and to encapsulate the relationships among two or more variables. Extrapolation mentions to the use of a fitted curve beyond the range of the observed data, and is subject to a degree of doubt since it may reflect the method used to construct the curve as much as it reflects the experiential data.

V. Result & Discussion

For implementation of our proposed smart tutoring we have used Matalb 2015b, dataset used in our experiment is data gathered from engineering college student dataset, snippet of dataset as follows:

A	B	C	D
RollNo	CT1	CT2	FINAL
NUMBER	NUMBER	NUMBER	NUMBER
3422214001	81.66666667	83.33333333	1
3422214002	80.83333333	80.83333333	1
3422214003	85.83333333	81.66666667	1
3422214004	79.16666667	72.5	1
3422214006	81.66666667	70.83333333	1
3422214007	80	81.66666667	1
3422214008	70.83333333	74.16666667	1
3422214010	82.5	87.5	1
3422214011	90.83333333	90	1
3422214012	91.66666667	89.16666667	1
3422214013	74.16666667	80.83333333	1
3422214014	86.66666667	79.16666667	1
3422214017	78.33333333	85.83333333	1
3422214018	70	70	0
3422214019	71.66666667	82.5	1

Fig. 3: Snippet of Input training dataset

Fig-4 shows the snippet of our proposed implementation GUI, through that we will apply curve fitting, for obtaining optimal fitness function.

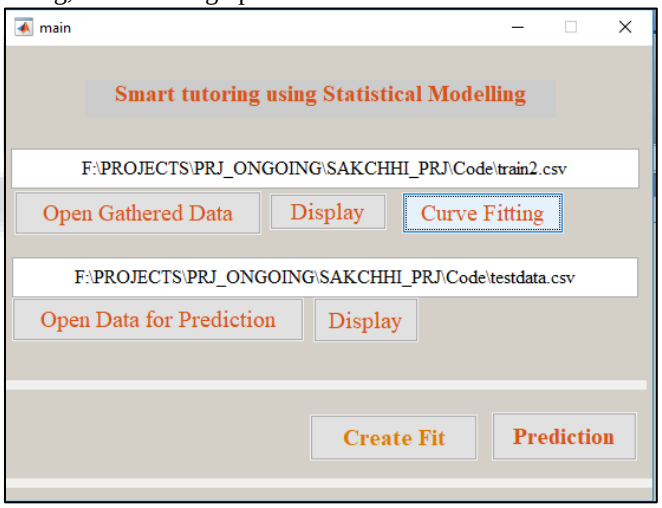


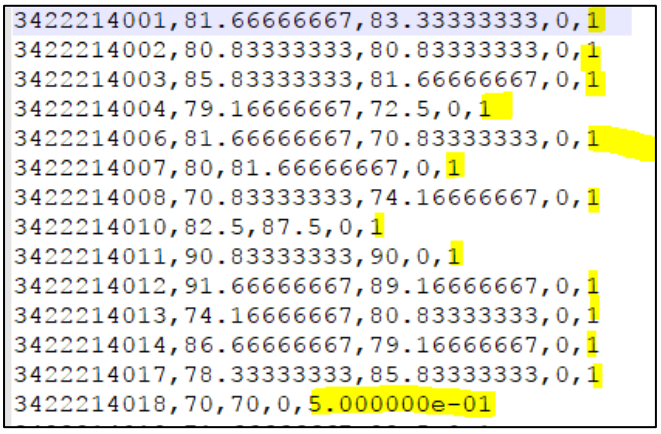
Fig. 4: Snippet of Main GUI

```
     Linear model Poly11:
     fitresult(x,y) = p00 + p10*x + p01*y
     Coefficients (with 95% confidence bounds):
       p00 =  -4.579e+08  (-1.496e+09, 5.802e+08)
       p10 =      0.1338  (-0.1695, 0.4371)
       p01 =      0.9114  (0.7448, 1.078)


gof =

           sse: 1.0719e+03
       rsquare: 0.8487
           dfe: 23
    adjrsquare: 0.8355
          rmse: 6.8266
```

Fig. 5: Fitness Function



```
3422214001,81.66666667,83.33333333,0,1
3422214002,80.83333333,80.83333333,0,1
3422214003,85.83333333,81.66666667,0,1
3422214004,79.16666667,72.5,0,1
3422214006,81.66666667,70.83333333,0,1
3422214007,80,81.66666667,0,1
3422214008,70.83333333,74.16666667,0,1
3422214010,82.5,87.5,0,1
3422214011,90.83333333,90,0,1
3422214012,91.66666667,89.16666667,0,1
3422214013,74.16666667,80.83333333,0,1
3422214014,86.66666667,79.16666667,0,1
3422214017,78.33333333,85.83333333,0,1
3422214018,70,70,0,5.000000e-01
```

Fig. 6: Predicted Final Result

VI. Conclusion

The institution goal is to develop a platform that provides real-time feedback and assistances online tutors, if institution having predicted data then, they can categories the student as Good, Average, Best etc. and by applying remedy can improve the student performance. Devices used in IoT produces huge amount of data which attract attention towards applying statistical modelling over data produced.

Classifier output shows that our proposed algorithm prediction accuracy is 92%. In future we can apply different efficient statistical modelling to increases the accuracy.

VII. References

[1] Chun-Wei Tsai, Chin-Feng Lai, Ming-Chao Chiang et. al. Data Mining for Internet of Things: A Survey IEEE 2014
[2] Najah Abu Ali et. al. Data Management for The Internet of Things: Green Directions IEEE 2012
[3] Feng Wang, Liang Hu, Jin Zhou et. al. A Survey from the Perspective of Evolutionary Process in the Internet of Things Hindawi Publishing Corporation International Journal of Distributed Sensor Networks Volume 2015
[4] Shen Bin, Liu Yuan et. al. Research on Data Mining Models for the Internet of Things IEEE 2010
[5] Jayavardhana Gubbi, Rajkumar Buyya et. al. Internet of Things (IoT): A vision, architectural elements, and future directions Elsevier 2013
[6] Edmund W. Schuster, Sumeet Kumar et. al. nfrastructure for Data-Driven Agriculture: Identifying Management Zones for Cotton using Statistical Modeling and Machine Learning Techniques IEEE 2011
[7] Michail N. Giannakos, Demetrios G. Sampson and Łukasz Kidziński Introduction to smart learning analytics: foundations and developments in video-based learning Springer 2016.

