DEFINITION OF AN ARTIFICIAL INTELLIGENCE ENGINE FOR MATHEMATICS EDUCATION

Shankar Moni*, Jaya Swaminathan, Sheloney Moni* and Shalini Sinha
Moodeye Media and Technologies
shankar@myxyo.com, sheloney@myxyo.com

The National Education Policy Draft (2019) proposes that students across over a million schools have to achieve an adequate level of numeracy. While it may be a challenge to find quality teachers at such a large scale, certain facets of their wisdom and experience could be encapsulated in an Artificial Intelligence (AI) engine. The application of AI to education has been explored by various researchers. Here, we specifically investigate the considerations to develop an AI engine for school-level numeracy. Further, we incorporate these considerations into a probabilistic framework and propose the definition of such an AI engine.

Keywords: *Artificial Intelligence, Mathematics Education, Probabilistic Framework*

INTRODUCTION

Over the years, researchers have noticed that Artificial Intelligence (AI) engines can be trained to yield predictions that exceed human accuracy. This is achieved by applying a high quantum of computational effort to massive, and possibly ill-formed or poorly-defined data sets. Given this attractive feature, AI has been used for a variety of industrial applications (Nadimpalli, 2017). AI has been applied in education as well (Chassignol et. al., 2018), the aim being to have a program mimic the efforts of an expert teacher.

To build an engine capable of mimicking human effort, it is first necessary to understand how the learning process works, and how it is influenced by the teacher. At a middle-school level, an expert teacher may juggle limited resources to present different perspectives on a topic (Matic, 2019). At the secondary-school level, the teacher may need to respond to different understanding process used across students (Radmehr & Drake, 2019). In addition to methods of teaching and understanding, there are other factors that affect learning outcome. As pointed out in (Shah & Chandrashekaran, 2015), learning the sciences requires practice on the part of the student, and it could be beneficial to use the classroom as a practice session, rather than for just lecturing. Generations of teachers have educated students, and there is quite a challenge ahead to incorporate all of this perspective into an automated engine.

The focus of our effort has been to define a scalable education system for mathematics and numeracy that will help achieve the goals proposed by the Government of India's National Educational Policy (Kasturirangan 2019). While it is not possible to have an expert teacher in every school in the country, it may be possible to scale a software-based solution. To succeed in this endeavour, it is imperative to sensitivize ourselves to the issues faced in urban, rural and tribal areas. A common obstacle encountered in both urban (Matthews, 2018) and tribal (Panda, 2006) areas is closing the loop with some assurance of expected learning outcome for the given funds. Moreover, social inequities also play a role (Harper, 2019), and should be handled with sensitivity. Fortunately, a software-based AI engine is agnostic to human aspects like socio-economic level, caste, or race, and could also go a long way in encapsulating the insights of experienced educators.

What we have accomplished so far is as follows. We have developed an Android-based application for practice of all aspects of mathematics from grade 1 to 10 of the CBSE syllabus. Well-trained teachers use this application to create customized sessions for each student with a combination of questions from various topics at selected levels of difficulty. Over 35,000 students have used this application, and a vast quantity of data has been generated. The next step that we would like to take is to automate the process of selecting customized sessions by quantifying the data collected, building a statistical model, and developing the required AI engine. This is the area of research of the current paper.

The current paper focuses on quantifying three aspects of the learning process, and then moves on to propose a design for an AI system. Firstly, we need a statististical model of difficulty level of each topic. This is addressed in section 2. Second, we need to characterize whether a student has exhibited proficiency in a topic from short-term memory or is the internalization of the topic deeper than that. Section 3 focuses on this aspect. Third, given that the various topics of mathematics have intricate interdependencies, to what extent would exposure to a related topic assist in mastering the current topic. This is addressed in section 4. Finally, in section 5, we use these empirical probabilities to construct a statistical framework to optimize the probability that a student will make progress. Various real-world constraints are included in the model, including the fact that mathematical topics often have prerequisites that would need to be first be mastered, and the fact that a class period is usually of limited duration, and this time would need to be optimally used. We now get into the details.

STATISTICAL ESTIMATION OF DIFFICULTY OF A TOPIC

To estimate the level of difficulty of a topic, we not only need to measure the average marks in this topic, but also need to consider other factors including (i) quantifying the initial effort required to acquire a basic level of competence, (ii) the statistical variation of competence with stringency of requirement, and (iii) the variations in duration needed to attempt and succeed in various topics. Figure 1 sheds light on the first two considerations. The third consideration is addressed later in this section.

Initial effort to acquire competence

The y-intercept of the graphs of Figure 1 indicate the number of attempts required to attain an intial level of competence in various topics. By initial level, we mean that the student should correctly answer at least 3 consecutive questions of this topic. The x-axis of this graph denotes the required number “n” of consecutive correct responses required.

Certain topics (like “Quadrilaterals”) require a high level of initial effort to master the concept. In this topic, on an average, students have attempted 7.1 questions before they could correctly respond to 3 consecutive questions. On the other hand, in a topic like "Lines and Angles", students, on an average, needed to attempt only 4.2 questions before they exhibited the same level of competence.

Implication of Statistical Variation of Competence with Stringency

It may be necessary to vary the stringency of the requirement "n" depending upon how often a topic is needed as a prerequisite. As an example, a topic like "Decimals" (taught initially in 5^{th} grade) serves as a prerequisite for a variety of topics in future years. However, a topic like "Bodmas" is a prerequisite much less often. For a topic like "Decimals", the stringency requirement (i.e. the parameter "n" on the x-axis of Figure 1) may need to be increased.

Two features of the data stand out in Figure 1. The first is that on an average, some topics need more attempts to attain a given level of proficiency. For example, the graph for "Quadrilaterals" is higher than the graph for "Triangles". In the ensemble of students tested, questions in the topic "Quadrilaterals" were more challenging than those in "Triangles".

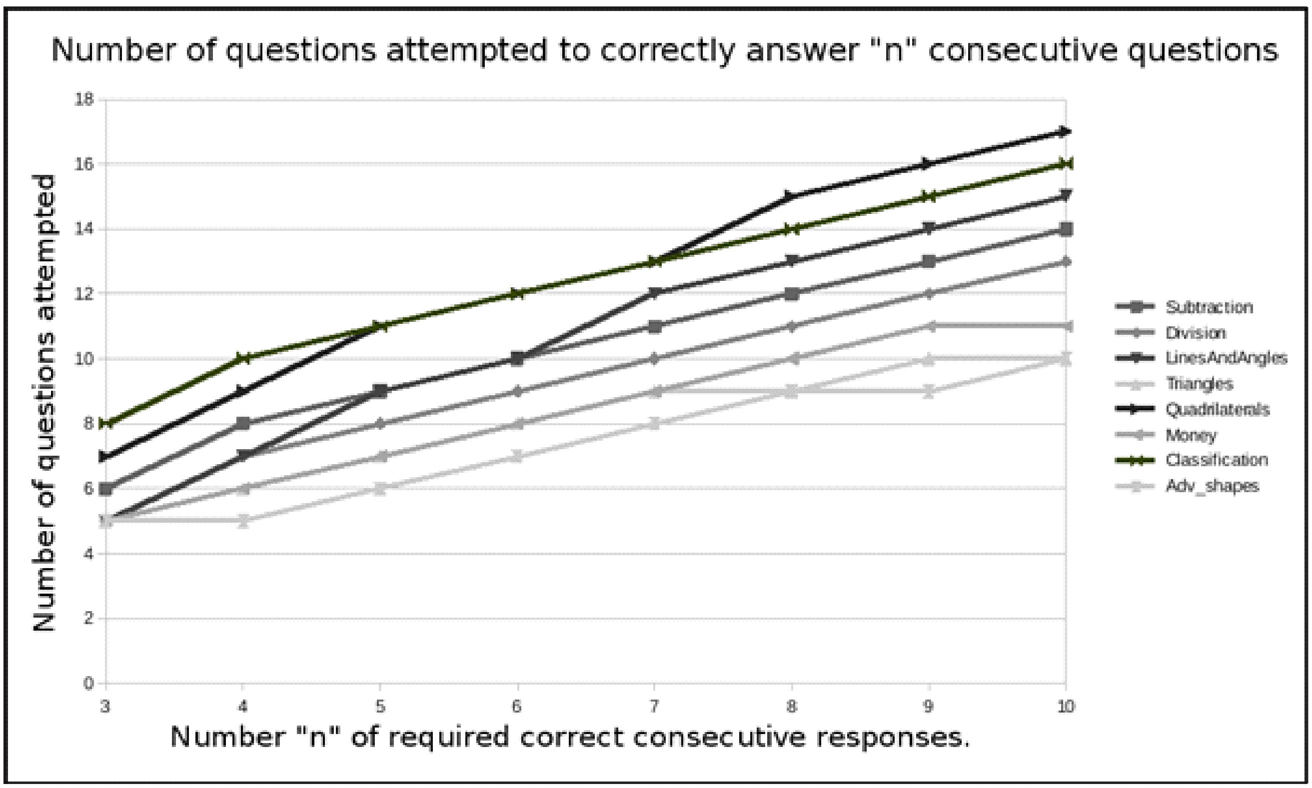

Figure 1: The number of questions required to internalize a topic increases with the parameter on the x-axis, and the variation is different for different topics

The second feature that stands out in the graph of Figure 1 is the shape of each curve, whether the second derivative is positive, zero, or negative. In some topics, the curve tends to have a positive second derivative, meaning it is accelerating upwards. The examples are "Quadrilaterals" and "Lines and Angles". In these topics, after the student has gained some proficiency, there is a higher level of challenge to continue to maintain that proficiency. This often happens in questions which require a high level of focus.

In some topics such as "Division" and "Subtraction", the second derivative is zero. Here, the level of focus required is just proportional to the number of questions given. While the duration required to solve a division question may go down with practice, the number of questions to be solved to meet a particular level of proficiency simply increases linearly with the required number "n" of consecutive correct responses.

Finally, some topics such as "Money" and "Advanced Shapes" exhibit a negative second derivative. Here, students may take a few attempts to determine how to solve the question. Once a level of proficiency has been attained, it is relatively easy to maintain that level.

Average Percentage and Duration

Given an ensemble of students, there is a variation of average percentage attained and mean duration required across topics. A particularly useful metric is the mean duration, because this is an indicator of how much effort the student is putting into this topic. These values are tabulated in Table 1.

In Table 1, we can see that topics can be grouped into three clusters: (i) Highly procedural topics like "Subtraction" and "Division", (ii) topics that initially require thinking but for which the student eventually derives a procedure, like "Lines and Angles", "Triangles", "Quadrilaterals" and "Money", and finally (iii) topics that are essentially logical reasoning (like "Word Classification" or "Advanced Shapes") for which there is no standard procedure.

Topic	Average Percentage	Mean Duration (Minutes)
Subtraction	95.13	3.10
Division	86.19	2.64
Lines and Angles	84.65	1.17
Triangles	82.84	1.07
Quadrilaterals	80.34	0.73
Money	79.69	1.08
Classification	63.79	0.49
Advanced Shapes	86.32	0.21

Table 1: The average percentage attained and mean duration required while solving questions in various topics

In highly procedural topics like "Subtraction" and "Division", students do take a long time to solve questions, but are not particularly flummoxed by the type of question or the juxtaposing of this question with questions from other topics (discussed in the next section and elucidated in Figure 2).

In newly introduced mathematical topics like "Lines and Angles", the student usually requires a while to determine how such questions are solved, and eventually develops his/her own procedure to solve these questions.

When we have topics focusing on logical reasoning, like “Word Classification” or “Shapes”, the student is able to quickly answer the question, but the frequency of getting the answer correct may or may not be high. This does not necessarily imply that the student is guessing (although that continues to be a possibility), but could mean that determining the classification or matching shape is a quick process without the labor of a standardized procedure.

All of these metrics to measure the difficulty of a topic need to be modulated by an orthogonal factor, which is whether the student is attaining competence using short-term memory, or is truly internalizing a topic for recollection at a later period in time.

Short-term versus Medium-term Proficiency

There are various time-frames in which to measure the idea of competence, roughly speaking, the short-term, medium-term, and long term periods (Norris, 2017). A short-term approach would be having a student attempt only one particular type of question until s/he starts getting it correct over a number of consecutive attempts (for example, “Fractions”). This offers both an intellectual and psychological benefit to the student. Once this level of proficiency has been attained, the next stage (i.e. medium-term horizon) would be to solve a question in a given topic interspersed with questions from other topics. In the longer-term, the student would be expected to answer questions from this topic many months after first attempting it, for example, in a final exam. In this research, we focus on the short and medium term.

In Figure 2, we show the empirical probability of being posed questions from mixed topics while correctly solving questions in various topics. The larger this number, the more ability students have, on an average, to correctly solve questions in this topic while switching context. Based on whether a topic is highly procedure-based or application-based, the ability to switch topics tends to vary. The foundation for the metrics and selection process (choosing topics, number of questions, measuring consistency) is discussed in (Swaminathan 2015).

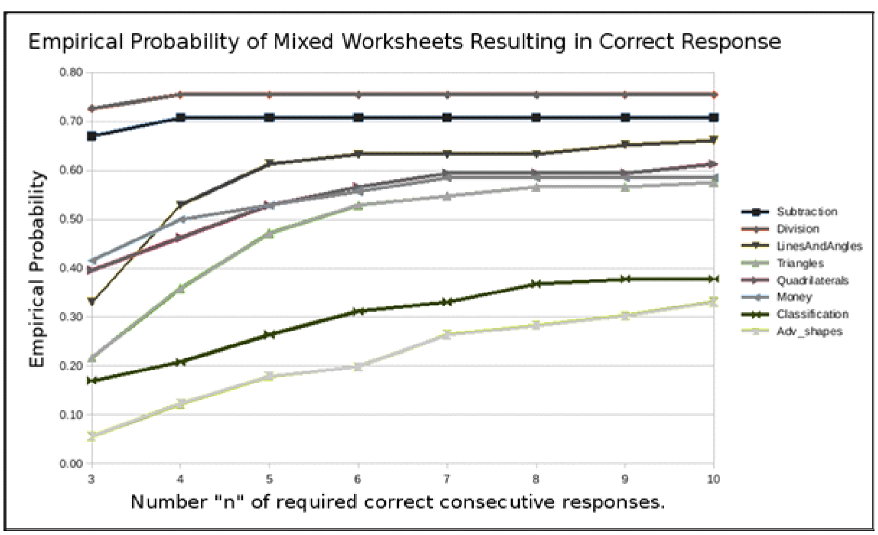

Figure 2: The empirical probability of allowing mixed topics while acquiring proficiency in any topic tends to increase as the parameter on the x-axis increases.

In highly procedural topics such as “Subtraction” and “Division” it is easy to gain proficiency and continue to maintain this proficiency despite the context being switched between different topics. Even with as high a context switch probability as 0.75 (i.e. 75% chance of context being switched), students, on an average, are able to attain and maintain proficiency.

Then, we have a topic like “Lines and Angles”, which starts off requiring a high level of single-topic focus (only 33% of the time does a switch in topic result in a correct response), but moving up to the possibility of a high level of multi-topic focus (can switch topics 67% of the time). Our teachers have observed that in such topics, the student needs single-topic focus to understand a new concept, and once this concept has been internalized, it is easier to handle this topic in conjunction with other topics.

Finally, we have topics in the realm of logical reasoning rather than conventional syllabus mathematics. Examples include “Advanced Shapes” (finding the matching shape) and “Classification” (word classification). In such topics, the data indicates that the student finds it difficult to retain proficiency in this topic while switching between topics.

The variation has dependence on whether the topic is highly procedural or more application-based. It is of interest to compare the data of Figure 2 to that in Table 1 since the clusters observed in both data are similar. In highly procedural topics such as “Subtraction” and “Division”, Figure 2 indicates that the student is able to handle the topic interspersed with other topics, and Table 1 indicates that the student requires a reasonably long time to perform this procedural task. In newly-introduced mathematical topics like “Lines and Angles”, “Triangles”, “Quadrilaterals” and “Money”, Figure 2 shows that the student initially needs to focus exclusively on this topic, but as competence is attained, is able to switch between other topics. Hand-in hand, Table 1 indicates that the student requires just about 1 minute to solve the question. The third cluster comprises topics focusing on logical reasoning, like “Word Classification” and “Advanced Shapes”. Figure 2 indicates that the student needs to focus exclusively on such topics to gain proficiency, and Table 1 indicates that the duration to solve these questions is quite low. There is no elaborate procedure to classify or select a matching shape, and questions can be answered quickly. However, it appears to be challenging to attempt such questions after answering questions of a completely different nature.

INTER-TOPIC DEPENDENCIES

Mathematics is a highly interconnected science. The solution to a problem may be inspired by working on an area of mathematics that bears but a subtle relation to the original problem. This observation inclines us to ask the question, “In order to correctly solve a problem in the current topic, is there another topic that could be presented prior to this that would perhaps stimulate a style of thinking conducive to solving this particular problem?”

To get a question correct in Division		**To get a question correct in Lines And Angles**		**To get a question correct in Quadrilaterals**	
Topic Given Earlier	Frequency	Topic Given Earlier	Frequency	Topic Given Earlier	Frequency
Division	1203	Lines & Angles	1343	Quadrilaterals	790
Multiplication	331	Area Volume	290	Area Volume	102
Place Value	269	Percentage	42	Circles	59
Subtraction	95	Decimals	31	Percentage	20
Fractions	62	Circles	19	Decimals	13
Mixed WP	55	Triangles	15	Lines & Angles	12
Factors Multiples	33	Fractions	11	Fractions	9
Measurements	21	Factor Multiples	10	Triangles	9
Percentage	16	Olymp Time	7	Directions	8
Area Volume	16	Word Classification	6	Mixed WP	7

Table 2: Frequency of topics attempted prior to getting a problem correct in the current topic

One can shed light on this matter by examining, for any given topic, the frequency with which various topics were attempted prior to correctly answering a question in this current topic. As can be seen in Table 2, prior to getting a question in "Quadrilaterals" correct, students had attempted questions in the same topic (i.e. "Quadrilaterals") 790 times, and had attempted "Area Volume" 102 times. To be clear, the students have correctly solved questions in "Quadrilaterals" 102 times after merely attempting a question in "Area Volume", and not necessarily getting the question in "Area Volume" correct.

In all cases, the best prior topic appears to be the topic itself, but this observation may have to be taken with a nuanced viewpoint. Often, the teacher makes a student continue to work on a particular topic until s/he attains a reasonable level of competence. Even under such a possibly suboptimal circumstance, it is noteworthy that working on other topics may have a significant desirable impact on gaining proficiency in the current topic.

Given that an Artificial Intelligence engine can handle mammoth quantities of data and run algorithms that are based on probabilities, the solution that suggests itself is to use each topic as a prior with a frequency proportional to its empirical probability, and continuously update the table of empirical probabilities.

DEFINITION OF AN ARTIFICIAL INTELLIGENCE ENGINE FOR MATHEMATICS EDUCATION

In this section, we define an Artificial Intelligence engine for mathematics education. The main objective of this section is to utilize the findings of the previous sections to create questions for the student that will be at the appropriate level and maximize the probability of the student making progress, given the time constraint of the class period.

We start by defining a few terms, then list out the empirically obtained quantities required. Then we build up the criterion that will be optimized to decide what question should be given to the student. Finally, we define the engine that uses a Bayesian model to generate a customized sequence of questions that will be optimal for the student.

Let us establish the terminology of "Level", "Prerequisite Factor" and "Accessable Level".

- Level: For each topic in mathematics (like "Addition", "Fractions", etc.), there would be several "Levels". Roughly speaking, a Level is a section of a chapter of the textbook. So if the level is 503, that might refer to the 3^{rd} section of a 5^{th} grade textbook. We define C_i to be the event that the student acquires consistency in the i-th level, and t_i to be the expected value of duration to correctly answer questions at this level.

- Prerequisite Factor: Each level in mathematics could serve as a prerequisite to a variety of other levels. For example, a level like one-digit addition would probably serve as a prerequisite to most of mathematics in the current and future years. The "Prerequisite Factor" is a measure of how many future levels in various topics depend on this particular level. We define F_i to be the prerequisite factor for the i-th level. Thus, F_i for one-digit addition would be a high number.

- Accessable Level: A level for which all prerequisites have been successfully completed is said to be an "Accessable Level". Let ***A*** be the set of all currently accessible levels. This will be set on a per-student basis. Now, for each level, the quantities required will be: (i) the required number "n" of questions to proficiency (the x-axis of Figures 1 and 2), which is decided by a panel of experts, (ii) the average number of questions to gain proficiency, which is obtained empirically, (iii) the average duration needed (obtained empirically), and (iv) the probability of having a mixed session (obtained empirically from Figure 2).

Using these quantities, we would like to define a criterion "J" to be optimized. Given a maximum time "T" and the constraint "M" that we want mixed topics if possible, we define

$$J=\sum_i F_i P\left(\frac{C_i}{M}\right) \text{such that } \sum_i t_i \leq T \text{ and } i \in A$$

as the criterion to be optimized. Essentially, we would like to maximize the expected value of prerequisite factors of competence, given that we are doing mixed topics and constrained to accessible levels, within the stipulated duration of a class period.

The value of $P(M/C_i)$ can be read from Figure 2, and emiprical values of $P(M)$ and $P(C_i)$ can be extracted from data. Using Bayes' Theorem, we get

$$P(C_i/M) = P(M/C_i)\ P(C_i)\ /\ P(M)$$

which can be used in the criterion J to be optimized. Further, the specific set of mixed topics can be selected probabilistically, by creating a probability density from the data of Table 2.

The levels selected will be determined by a greedy search algorithm, maximizing the values of F_i with $\notin \boldsymbol{A}$. Also, ***A*** is updated each time the required proficiency of any level is attained.

The above algorithm maximizes the expected value of prerequisite factors of competence using empirical probabilities from a global ensemble. Future work includes modifying these probabilities on a per-individual basis, and a multi-resolution approach for prerequisites.

CONCLUSION

It is a signifcant challenge to enable all of the schools of India (over a million in number) to have the benefit of an expert teacher for mathematics. However, some facets of such expertise could be codified into an Artifical Intelligence engine, and scaled across schools. The National Education Policy Draft (Kasturirangan, 2019) specifically mentions technology as an enabler of quality education at scale. The current paper has outlined the issues related to and the process of developing an AI engine specifically for mathematics education.

REFERENCES

Nadimpalli, M. (2017). Artificial Intelligence – Consumers and Industry Impact. *International Journal of Economics & Management Sciences*. 6(4) 439-441.

Chassignol, M., Khoroshavin, A., Klimova, A., & Bilyatdinova, A. (2018). Artificial Intelligence trends in education: a narrative overview, *Proc. 7th International Young Scientist Conference on Computational Science, Procedia Computer Science* (Vol 136, pp 16-24). Heraklion, Greece.

Matic, J., (2019). Pedagogical design capacity of lower-secondary mathematics teacher and her interaction with curriculum resources, *Journal of Research in Mathematics Education*, 8(1) 53-75.

Radmehr, F. & Drake M., (2019). Students' mathematical performance, metacognitive experiences and metacognitive skills in relation to integral-area relationships, *Teaching Mathematics and its Applications*, 38(2) 85-106.

Shah, P. & Chandrasekharan, S. (2015). Analyzing the role of digital media in science education: A distributed cognition approach, In S. Chandrasekharan, S. Murthy, G. Bannerjee & A. Murlidhar (Eds.), *epiSTEME6, Emerging Computational Media and Science Education*, (pp. 169-176), Mumbai, India.

Kasturirangan, K. & Committee for the Draft NEP. (2019). Draft National Educational Policy 2019, *Ministry of Human Resource and Development, Government of India*, New Delhi, India pp 1-484.

Panda, M. (2006). Mathematics and tribal children, *Economic and Political Weekly*, Vol 41, No 2, 2006, pp 117-120.

Matthews, L. E. (2018). Illuminating Urban Excellence: A Movement of Change within Mathematics Education, *Journal of Urban Mathematics Education*, 11(1&2), 7-10.

Harper, K. (2019). A Qualitative Metasynthesis of Teaching Mathematics for Social Justice in Action: Pitfalls and Promises of Practice, *Journal for Research in Mathematics Education*, 50(3) 268-310.

Norris, D. (2017). Short-term memory and long-term memory are still different, *Psychological Bulletin*, 143(9) 992-1009.

Swaminathan, J., Bhartari, R., Sinha S., Moni S. & Moni S. (2015). Quantified Closed-Loop Internalization of Mathematics, In S. Chandrasekharan, S. Murthy, G. Bannerjee & A. Murlidhar (Eds.), *epiSTEME6, Emerging Computational Media and Science Education*, (pp. 399-406), Mumbai, India.