Crowdsourced Learning

Ashwin Venkataraman, Shiva Iyer, Kunal Gugale, New York University
Sepehr Vakil, UC Berkeley
Abhijai Garg, Jay Chen, New York University Abu Dhabi
Krishnaram Kenthapadi, LinkedIn
Lakshmi Subramanian, New York University

1. INTRODUCTION

The conventional education ecosystem in developing regions is plagued by the lack of good quality textbooks and educational resources, lack of skilled teachers and high variability across student skill and motivational levels [Crossley and Murby 1994; Glewwe et al. 2007]. This paper makes the case for establishing a crowdsourced learning ecosystem that leverages the collective intelligence of educators around the world to design a collaborative platform [Arias et al. 2000] to easily share, search, organize, rate and present educational materials for teachers and students around the world. The recent popularity of online learning platforms and Massive Open Online Courses (MOOCs) has made it possible for students to access high quality educational content from the comfort of their homes and enabled new forms of learning that were not possible before. Given the wealth of educational resources on the Web, this paper describes the vision of a crowdsourced learning platform that aims to integrate rich educational web content into an inquiry based framework using the 5E learning model [Bybee et al. 2006] for generating, sharing and rating web annotated lesson plans for school teachers and students. By sharing educational content across teachers, the crowdsourced learning platform should allow teachers to leverage content authored by other (potentially higher quality) teachers, rate the appropriateness of content and also provide feedback (in the form of votes/ratings) to promote high quality and relevant content as a function of student skill levels. Similarly, the platform should enable students to interact with their peers and the teachers to engage in discussions related to the course material (similar to existing forums in MOOCs [Mak et al. 2010]) and promote both a certain degree of peer learning as well as skill-based personalization. In essence, the goal of crowdsourced learning is to create an ecosystem that enables the collaboration amongst teachers and students of diverse backgrounds to improve the overall educational experience and lead to better learning outcomes for students. This paper specifically makes two important contributions: (a) **Modeling learning outcomes in crowdsourced learning**: We propose a mathematical framework that enables systematic modeling and comparison amongst different education paradigms. Our framework provides a means to quantify student learning under a given paradigm based on critical factors such as the student skill (or ability), quality of the reading material (or the teacher), etc. (b) **Crowdsourced Learning Platform**: We describe the design of YeSua, an initial prototype of our crowdsourced learning platform that uses an inquiry-based framework for generating annotated lesson plans for different subjects.

2. YESUA SYSTEM OVERVIEW

The YeSua platform leverages the 5E educational instructional framework which consists of 5 stages – *Engage*: which involves engaging the students and connecting the topic of instruction to the real experiences of students. *Explore*: where students are guided to explore a topic that emerged in the engagement phase. *Explain*: where teachers confirm student ideas or help clarify misconceptions revealed

Collective Intelligence 2016.

during earlier stages. *Elaborate*: promoting deeper inquiry into the topic of instruction by challenging students with complex problems and demonstrating real-world applications of the concepts discussed. *Evaluate*: provides an opportunity for teacher and students to assess the understanding and conceptual mastery of the course content. YeSua uses both crowdsourced content shared by teachers and educational resources available on the Web to enable easy and automated creation of subject-specific lesson plans, based on the 5E model. The relevant content is fetched from the Web using the YeSua search query interface, that automatically generates queries based on teacher provided keywords and concepts. Finally, the system enables teachers to sift through the collection of search results and select the pages that contain the relevant content they are looking for.

Preliminary Experience: We tested YeSua with mathematics and science teachers in Accra, Ghana. The participants expressed confidence that the tool assisted the process of mapping their existing curricular materials into a cohesive lesson plan. Some participants were observed to select web-based resources they had not originally conceived of, after reviewing the search results (as an anectode, one teacher used YouTube and Vimeo videos in the *Explore* stage that helped to explain related concepts not part of the original lesson plan). This illustrates how the YeSua tool can expand the realm of pedagogical possibilities for participants who may be unaware of the scope and nature of web-based material.

3. MATHEMATICAL MODEL

We consider a scenario where every student goes through a curriculum which involves a sequence of courses $C_1, C_2 \dots$ towards some goal (like a degree). At any point during this sequence, each student s has an underlying knowledge state $KS(s)$ [Corbett and Anderson 1994], where higher values indicate more skilled students. We assume that each student is equally motivated to learn and given enough time/practice, will eventually master any concept. Based on the 5E-driven lesson plan formulation above, we suppose that each course C_i consists of a sequence of *lessons*, which represent individual learning units of the course. Let $l_1, l_2, \dots, l_m$ denote the sequence of lessons for a particular course where the lesson ordering respects the dependence of concepts taught in the course. Every lesson contributes some "utility" Δ_i towards the student knowledge state and the total utility of a course is given by $\sum_{i=1}^{m} \Delta_i$.

Suppose that each lesson l_i consists of a *lecture* followed by p_i steps of *revision* or augmentation of the material taught in the lecture. These could include assignments, recitation sessions or in the case of online platforms, supplementary content in the form of videos [Goldman et al. 2014], PPT slides [Levasseur and Kanan Sawyer 2006] etc. The students progress through the lessons one after the other and the level of understanding (i.e. their knowledge states) across students is highly varied at any point in the course. For instance, while studying the lesson on Force in a Physics course, if the student did not (completely) understand prior lessons on Displacement, Acceleration etc. then her understanding of the Force concept will be weak, i.e. understanding of a lesson is dependent on the understanding of prerequisite lessons.

We quantify this intuition by assuming that every student "learns" a certain *fraction* of each lesson's content as she progresses through the course. In particular, for a student s and lesson l_i, let $0 \leq \alpha_s(i) \leq 1$ denote the *grasping coefficient* of student s; $\alpha_s(i)$ determines "how much" the student s is able to understand (or grasp) the concepts taught in a *single* step of lesson l_i (either the lecture itself or one of the revision steps). The parameter $\alpha_s(i)$ depends on the student knowledge state $KS(s)$, the *quality* of the educational content of the lesson $Q(i)$, as well as the understanding of the prerequisites of lesson l_i (based on our discussion above). Therefore, $\alpha_s(i)$ is a composite function of these different variables, and to keep the discussion simple we do not consider specific forms of the function, which can be chosen based on the context. However, $\alpha_s(i)$ does satisfy natural properties such as being monotonically increasing in the knowledge state $KS(s)$, the quality $Q(i)$ and the prerequisite understanding. In each step of the lesson, the student enhances her knowledge of the lesson material by a factor $\alpha_s(i)$. This relies on the fact that each lesson is a fundamental unit that contains one or a few main concepts, so that students can master the concepts through repeated practice. Intuitively, more the number of steps p_i, the better the student can understand the content and in the ideal scenario when $p_i \to \infty$ the student would perfectly master the lesson (this relies on our assumption that all students are motivated to learn). We capture this mathematically as:

$$\begin{aligned} F_s(i) &= \alpha_s(i) + \alpha_s(i) \cdot (1 - \alpha_s(i)) + \alpha_s(i) \cdot (1 - \alpha_s(i))^2 + \ldots + \alpha_s(i) \cdot (1 - \alpha_s(i))^{p_i} \\ &= \alpha_s(i) + (1 - \alpha_s(i)) \cdot [1 - (1 - \alpha_s(i))^{p_i}] \end{aligned}$$

where $F_s(i)$ represents the fraction of lesson l_i's content learnt by student s. The expression captures our intuition of iterative learning through the steps of the lesson and ensures that $F_s(i)$ is a valid fraction i.e. lies between 0 and 1 and $F_s(i) \to 1$ as $p_i \to \infty$. Additionally, the knowledge state of the student $KS_{\text{end}}(s)$ at the end of the course is given by: $KS_{\text{end}}(s) = KS_{\text{begin}}(s) + \sum_{i=1}^{m} F_s(i)\Delta_i$, where $KS_{\text{begin}}(s)$ represents the knowledge state at the beginning of the course. The student starts the next course in the sequence at this knowledge state.

3.1 Model Insights

We can use the above model to draw the following insights:

— **Comparing traditional and crowdsourced learning**: Our model enables one to derive conditions when crowdsourced learning can be better than traditional learning paradigms. For instance, we can compare the distribution of student knowledge states at the beginning and end of a course, in both a traditional and crowdsourced learning model, to evaluate which paradigm is better. In our context, this crucially depends on the grasping coefficient $\alpha_s(\cdot)$ for any student and the potential benefit (or detriment) of a crowdsourced learning model can be measured by the impact on the grasping coefficient, and thereby the student knowledge states.

— **Better quality content can lead to improved learning outcomes**: Since $\alpha_s(i)$ increases with increasing content quality $Q(i)$ (assuming other parameters being fixed), this results in an increased fraction $F_s(i)$ and eventually higher knowledge states $KS(s)$. Therefore, the crowdsourced learning platform should have suitable mechanisms to discover and promote (through feedback in the form of votes or ratings) high quality content that can be used to improve student learning. In particular, if the grasping coefficient of student s **increases** to $\beta_s(i)$ because of access to a crowdsourced platform, then the learnt fraction $F_s^{\text{cl}}(i)$ becomes

$$\begin{aligned} &\alpha_s(i) + \beta_s(i) \cdot (1 - \alpha_s(i)) + \beta_s(i) \cdot (1 - \alpha_s(i)) \cdot (1 - \beta_s(i)) + \ldots + \beta_s(i) \cdot (1 - \alpha_s(i)) \cdot (1 - \beta_s(i))^{p_i - 1} \\ &= \alpha_s(i) + (1 - \alpha_s(i)) \cdot [1 - (1 - \beta_s(i))^{p_i}] \end{aligned}$$

where the assumption is that the lecture component is unaffected. Since $\beta_s(i) > \alpha_s(i)$, we get that $F_s^{\text{cl}}(i) > F_s(i)$.

— **Choosing personalized revision steps**: For any lesson l_i, observe that the grasping coefficient of each student can be very different. As a result, showing the same sequence of p_i steps for each student might not be an optimal decision. If we have knowledge of the grasping coefficient of a student, then we can employ it to adaptively choose the augmentation steps that are shown, possibly focusing on prerequisite lessons if required. In practice, we can estimate $\alpha_s(i)$ by using the results from the *Evaluate* step of each lesson (which is part of the 5E model described above).

REFERENCES

Ernesto Arias, Hal Eden, Gerhard Fischer, Andrew Gorman, and Eric Scharff. 2000. Transcending the individual human mind-creating shared understanding through collaborative design. *ACM Transactions on Computer-Human Interaction (TOCHI)* 7, 1 (2000), 84–113.

Rodger W Bybee, Joseph A Taylor, April Gardner, Anne Westbrook, and Nancy Landes. 2006. The BSCS 5E instructional model: Origins and effectiveness. (2006).

Albert T Corbett and John R Anderson. 1994. Knowledge tracing: Modeling the acquisition of procedural knowledge. *User modeling and user-adapted interaction* 4, 4 (1994), 253–278.

Michael Crossley and Myra Murby. 1994. Textbook provision and the quality of the school curriculum in developing countries: Issues and policy options. *Comparative education* 30, 2 (1994), 99–114.

Paul Glewwe, Michael Kremer, and Sylvie Moulin. 2007. *Many children left behind? Textbooks and test scores in Kenya*. Technical Report. National Bureau of Economic Research.

Ricki Goldman, Roy Pea, Brigid Barron, and Sharon J Derry. 2014. *Video research in the learning sciences*. Routledge.

David G Levasseur and J Kanan Sawyer. 2006. Pedagogy meets PowerPoint: A research review of the effects of computer-generated slides in the classroom. *The Review of Communication* 6, 1-2 (2006), 101–123.

Sui Mak, Roy Williams, and Jenny Mackness. 2010. Blogs and Forums as Communication and Learning Tools in a MOOC. (2010).