Charcterizing Navigated Learning

SRINATH SRINIVASA, Gooru Labs, IIIT Bangalore
PRASAD RAM, Gooru Inc

Navigated Learning is a new paradigm of learning spearheaded by Gooru, that aims to implement digital empowerment by balancing between conflicting requirements of scale, individualization and social elements of learning. This technical report introduces the underlying science behind navigated learning.

Additional Key Words and Phrases: Navigated Learning, Semantic Embedding, Social Learning, Personalized Learning

Reference format:
Srinath Srinivasa and Prasad Ram. 2019. Charcterizing Navigated Learning. *Gooru Tech. Rep.* 3, 1, Article 1 (March 2019), 16 pages.

1 INTRODUCTION

Education is a human right, and is one of the primary means for social empowerment. It is essential that sound educational practices be implemented at scale. Education cannot result in social changes and reforms if it is accessible by only a privileged few – indeed, such a system may even exacerbate social fissures.

World over, the most common paradigm for implementing anything at scale, has been the model of the *factory*, that finds it roots in the industrial revolution of 18^{th} century Europe. This is based on creation of mechanized processes, mass production of homogenized output, and standardization. Mainstream education in most countries is built around classrooms implementing standard curricula and pedagogic styles, representing the factory paradigm of mass production.

In a classroom, curricula and educational practices are designed for a hypothetical "average" student, having "average" abilities, and education itself is calibrated based on measurable assessments and outcomes. However, recent research on individualization [20, 21], have shown huge variability between individual traits and group averages. Modeling based on averages may be effective when the target beneficiary is the group itself. Some examples include upholding collective metrics like average time for check-in in an airport, or average journey time for metro commuters. However, when the *individual* is the target beneficiary, designing models based on group averages, are ineffective.

This is one of the main criticism of MOOCs (Massive Open Online Courses) as well [17, 30]. The MOOC extends the classroom model designed for the average, over large populations.

Figure 1 shows several SVM (support vector machine) models built for learners by observing activity data generated by learners on the Gooru platform for the subject of high school mathematics. These models were then clustered based on the coefficients of each of the models. The figure also depicts the "average" learner model built from all activity data from all learners. The average learner model is so different from individual models, that it clusters into a separate singleton cluster.

Individualization is an important element of effective learning practices. The problem of achieving individualization at scale, has been addressed using AI techniques as part of the literature on

2019. /2019/3-ART1 $
.

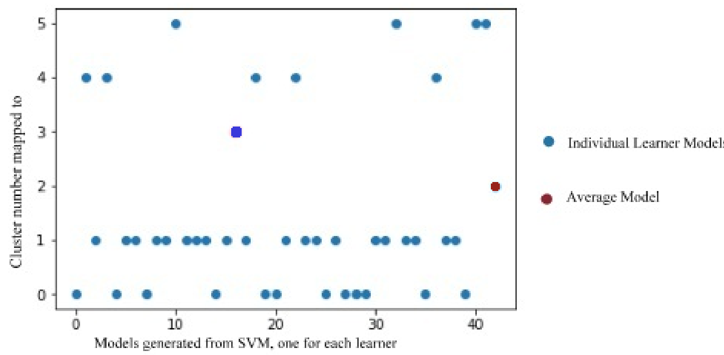

Fig. 1. Machine learning models built for math learners on Gooru, compared with the average model

Intelligent Tutoring Systems (ITS) [2, 26], and more recently Adaptive Educational Systems (AES) [7, 23] and Adaptive Educational Hypermedia Systems (AEHS) [8, 9].

The general design principle behind adaptive learning environments include a learner model and a knowledge model, which are used by an AI-based computational agent, to curate a customized learning experience. The agent continuously monitors learner progress based on their performance in intermediate assessments, and curates the learning pathway accordingly. Although individualized tutoring offered by a human tutor is more nuanced and subtle, studies have found that automated tutoring can be quite effective, and comparable to human tutoring, on several fronts [16].

However, the dearth of the human element is the biggest shortcoming of automated tutoring environments. Learning is an inherently *social* activity, that is continuously curated not just by the tutor, but also by peer influence from other learners. Indeed, an important element of learning is often based on copying practices and beliefs of other learners who are slightly ahead of us in the learning progression. Cognitive foundations of social learning date back to the 1960s in the works of Bandura [3, 4], where imitation of peers is seen as an important element of cognitive model building. Technology augmented pedagogy has incorporated elements of social learning in different ways, that are collectively called Computer Supported Collaborative Learning (CSCL), or more recently, Web-enabled Collaborative Learning [13, 18, 24].

A related field of study is that of educational practices around Distributed Cognition [22, 28], where learning and cognition is posed as a function of interaction between a human and his/her surroundings.

Figure 2 depicts the three strategic concerns of empowerment, and technology-enabled solutions that have predominantly focused on each of the concerns. The paradigm of Navigated Learning as introduced in this paper, aims to balance all three concerns to create a scalable, individualized, social learning platform.

2 SEMANTIC EMBEDDING AND PROGRESSION SPACE

The primary idea behind Navigated Learning that addresses all three concerns of scale, individualization and social elements of learning, is called *semantic embedding* in a *progression space.* We will describe both these terms in this section.

A *progression space* represents an area– specifically, a metric space, through which, several participants progress through time. The design of such spaces is meant to facilitate progressions for all participants, in as close to an optimal fashion as possible. Examples of progression spaces

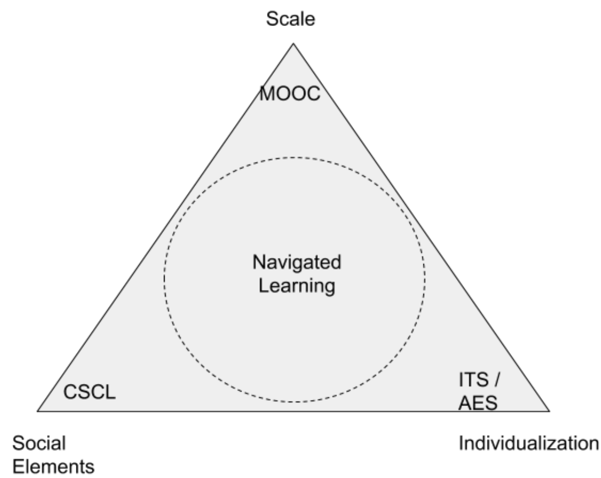

Fig. 2. Positioning Navigated Learning

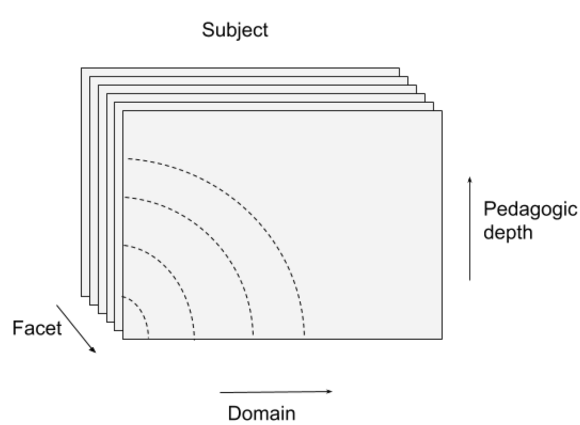

Fig. 3. Progression Space for a Subject

from other spheres of life include: airports, metro train stations, bus stations, fast-food outlets, etc. Such spaces are characterized by a floating population of individuals (also called *participants*, or, in our case, *learners*), each pursuing disparate goals, but of a largely similar nature. There need not be one common, overarching goal being pursued by the entire population. For instance, not all people in a bus station may be going to the same destination, but most of the participants in a bus station are there because they want to go somewhere or are arriving from somewhere.

Formally, a progression space is defined as follows:

$$S = (C, d, \preceq) \quad (1)$$

Here C is a set of points that make up the space, $d : C \times C \longrightarrow \mathfrak{R}$ is a "distance" function between any pair of points in the space. A distance function has the following characteristics:

$$\begin{aligned} Reflexivity \quad & \forall x \in C, d(x,x) = 0 \\ Symmetry \quad & \forall x, y \in C, d(x,y) = d(y,x) \\ Triangle\ inequality \quad & \forall x, y, z \in C, d(x,z) \leq d(x,y) + d(y,z) \end{aligned}$$

The term $\preceq \subseteq C \times C$ in Eqn 1 represents a "progression" relationship between pairs of points. A relationship of the form $c_1 \preceq c_2$ represents that a learner at point c_2 has progressed at least as much as learner c_1.

Actions and states of participants in a progression space are not necessarily independent of one another. The presence of some participant in some part of the space may (positively or negatively) affect the experience of other participants nearby.

Navigated learning characterizes learning as an aided navigation through a progression space comprising of several learning activities and other learners. The journey of a learner is curated by a navigator software, based on the learner's goals. The navigator not only aims to personalize the learning pathway for the learner's needs, it also tries to take advantage of the presence of other participants in the system.

The unit of activity in this space, is called a *learning activity*. A learning activity includes but is not limited to, consuming of resources like videos or lecture notes. A given learning activity may also involve several learners participating in a synchronous fashion, like in a group discussion or a debate. A learning activity also need not be a purely online activity. It can even be an offline activity, like attending a lecture in a conventional classroom. The only requirement for a learning activity is that it should *generate activity data* that helps the navigator locate and navigate the learner in the space.

Offline learning activities that are part of a navigated learning scheme, are designed such that they contain mechanisms to generate relevant activity data, and push them to the backend that is managing the navigated learning. These mechanisms include smartphone apps that learners install, which records and transmits their activity details, or even low tech mechanisms like a teacher collecting data from students after every learning activity, and uploading them in bulk.

Figure 3 depicts a semantic progression space for any given subject. Any subject matter of study is divided into one or more *facets* of study. Each of facet is represented by a 2-dimensional progression space.

A progression space for subject S and facet f, is represented as a 2-dimensional *competency map* that is formally defined as follows:

$$CM(S_f) = (C, P, Q, \gamma, d, \preceq) \quad (2)$$

Here the term C, which is the set of points forming the progression space, represents a set of *competencies* that are the basic unit of learning. The terms P and Q represent the horizontal and vertical axes of the progression space respectively. The horizontal axis represents a set of topics (also called "domains") that are addressed as part of this facet. The vertical axis represents the depth or "level" at which a particular topic is studied. Higher values of Q represents more depth, requiring a higher level of skill and comprehension. The terms d and $\preceq$ represent the distance and progression maps, as described in Eqn 1. The P and Q coordinates are organized such that progression is always from left to right, and bottom to top. Hence, for any pair of competencies $c_1, c_2 \in C$, $c_1 \preceq c_2 \Rightarrow p(c_1) \leq p(c_2)$ and $q(c_1) \leq q(c_2)$, and for at least one of the dimensions

$x \in \{p, q\}$, $x(c_1) < x(c_2)$. In the figure, this property is schematically depicted as concentric arcs, emanating from the bottom-left to the top-right.

The term $\gamma : C \longrightarrow P \times Q$ represents an embedding function for competencies that assigns a (p, q) coordinate for each competency. The actual form of the embedding function itself is not specified by the navigated learning framework. The only constraint for the embedding function is that it should preserve the progression property from bottom-left to top-right.

A *learning map* for a given progression space comprises of a set of *learning activities* that are embedded into the space. Formally:

$$LM(S_f) = (S_f, A, \delta, L) \tag{3}$$

Here A is a set of learning activities, and $\delta : A \longrightarrow S_f$ embeds each learning activity by mapping it onto a competency in S_f. A learning activity is an abstract container that encapsulates several kinds of learning engagements. For a given learning activity $a \in A$, the term $typeof(a)$ is used to ascertain what kind of learning activity it is.

As earlier, the navigated learning framework itself does not prescribe any specific technique for embedding learning activities into a competency space. The only semantic requirement is that a learning activity a mapped to a competency c should be useful for a learner in acquiring the competency c. In early implementations of navigated learning in K-12 settings, learning activities were manually embedded onto the space by experts. Later on, using this as a training dataset, machine learning algorithms were built to embed learning activities.

Each embedding of the form $\delta(a) = c$, represented by the tuple (a, c) is associated with it, several kinds of meta-data. Some of these are as follows:

Relevance: The relevance function $relevance : A \times C \longrightarrow [0, 1]$ measures the fit between the learning activity and the competency to which it is mapped to.

Engagement: This is a score of the form $engagement : A \times C \longrightarrow [0, 1]$ that rates the learning activity with respect to other learning activities mapped to the same competency c for relative popularity of it being used in any learning pathway.

Efficacy: This is a score of the form $efficacy : A \times C \longrightarrow [0, 1]$ that tries to estimate the probability that a given learner would obtain competency c, by performing learning activity a.

A set of data-driven interpretations for these metrics, and corresponding algorithms for computing these metrics are detailed in Ram and Srinivasa [19].

The term L in Eqn 3 represents a set of *learning pathways* that are part of the learning map. A learning pathway $l \in L$ is a sequence of competencies $l \in C^*$ that represents a coherent and progressive learning sequence. A learning pathway is a feature of the learning map, rather than that of a learner. That is, a learning pathway need not have a specific stated goal, and is not customized for a given learner. It only represents a coherent sequence of competencies. The formal notion of coherency of a learning pathway is called the "Narrative Arc" problem, and is detailed in a later section.

Principles of Navigated Learning. With the necessary definitions for semantic embeddings in a progression space, we now describe the principles of navigated learning. Navigated learning can be summarized by the following steps: *Locate*, *Curate*, *Mediate*.

Navigated learning is manged by a "Learning Navigator" with which every learner interacts. The Learning Navigator (or just, navigator), continuously interacts with the learning map and the learner to perform the following:

Locate: Based on data about their activities and outcomes from formal assessments, the "Locate" module of the navigator embeds learners in the space, and continuously updates their location. Unlike a geographical space, a learner may have acquired several competencies in the competency space. Thus, their location is not identified by a point, but by a data structure called a *Skyline*, that is detailed in a later section.

Curate: Once a learner's location is known, based on their stated goals or recently acquired competencies, a set of further candidate competencies are identified. Curating is based on competency modeling principles, that identifies complementary, supplementary and conflicting competencies.

Mediate: This is the logic by which the navigator navigates the learner by making suggestions. Mediation is based on computing an underlying "Narrative Arc" that computes a semantically coherent and meaningful learning sequence individualized for each learner. Mediation also involves suggesting connections with other learners as well as group learning activities.

In addition to the above, when learners interact with a learning map, they can explore the learning space as if it were a 2D shared physical space. They can browse through the space, and encounter learning activities, resources, as well as other learners at different points in the space. They can follow pre-computed learning pathways to go through a coherent sequence of activities.

3 COMPETENCY MAP CREATION

This section introduces the work in progress, meant to automate the creation of a competency map for a given subject and a facet. Currently, competency map is created with the help of domain experts who build the P and Q coordinate axes, and label the different topics that make up the competency map. A set of learning resources are also assigned by the experts, to appropriate positions in the map.

In this section, we discuss a methodology to automatically generate a competency map given just a representative set of learning resources. While any deployment of a competency map would likely need some form of manual intervention and vetting, the proposed method would greatly reduce the manual effort in creating the competency map. The proposed method does not require any prior knowledge of the subject matter, and also does not require learning resources to be designed in any specific manner. It considers each learning resources to be in the form of natural language text. (Other forms of learning resources like videos and slides, are first converted to text before being processed).

Given a set of learning resources R, the process of creating a 2D competency map, adopts the following steps:

(1) Identify the domains forming the P coordinate
(2) Organize the domains in increasing order of complexity with respect to the given competency map
(3) Identify the domain (P coordinate) for each learning resource
(4) Identify the height (Q coordinate) for each learning resource
(5) Create competency map (that includes competencies for which no learning resources have been mapped).

We first start with representing each learning resource as a text document, and creating a text corpus R. The corpus is then broken down into topical clusters using Latent Dirichlet Allocation (LDA) [5]. LDA is one of the most widely used algorithms for topic modelling. Here, a document is modeled as a distribution over a set of topics that are latent, and each topic in turn, is modeled as a distribution over terms, that are visible. Given a corpus of documents, LDA uses techniques like Gibbs sampling, to organize the visible terms from the corpus, into several topical clusters.

[TO DO: Explain the volume-based method for competency map creation]

4 PROFICIENCY ALGEBRA

Proficiency algebra refers to the framework of reasoning about the proficiency of a learner or a group of learners. For any given subject S, the proficiency profile of a learner l (denoted as $P_S(l)$) is defined as:

$$P_S(l) = (M, A, \psi) \tag{4}$$

Here M (represented as $M_S(l)$ when the context is not apparent) represents the set of all competencies mastered by the learner, A represents the set of all competencies that the learner has attempted but not shown mastery. The term ψ represents the learner "skyline" depicting the highest proficiency level obtained for any domain in the subject.

The skyline is a set of competencies across domains that represents the current "location" of the learner. Routing and mediation logic are based on interfacing with the skyline. Formally, the skyline is a set of competencies of the form:

$$\psi_S(l) = \{c \mid c \in C_S\} \tag{5}$$

Here, for any pair of competencies $c_1, c_2 \in \psi_S(l)$ if $\gamma(c_1) = (p_1, q)$ and $\gamma(c_2) = (p_2, q)$, then it follows that $p_1 = p_2$. That is, there is at most one element of any given domain that can be part of the skyline. Additionally, if $c(p, q)$ is part of the skyline, then for every p' such that $p < p'$, there is no competency for the form $c'(p', q)$ that is a member of $M_S(l)$. In other words, the skyline represents the highest competency in a given domain for which mastery has been achieved by the learner. For any higher competency in the same domain, the competency is either under progress, or not yet attempted by the learner.

Skyline dynamics. When a learner acquires a new competency $c(p, q)$ by demonstrating mastery, the skyline is modified accordingly. This done by removing any competency of the form $c'(p', q)$ such that $p' < p$, that is already in the skyline, and adding $c(p, q)$ to the skyline.

Sometimes, an acquired competency may have to be deleted from the learner's proficiency map. When a competency of the form $c(p, q)$ is deleted from the learner's proficiency map, the skyline is altered to include $c(p', q)$ in the skyline such that it represents the highest proficiency obtained in domain q.

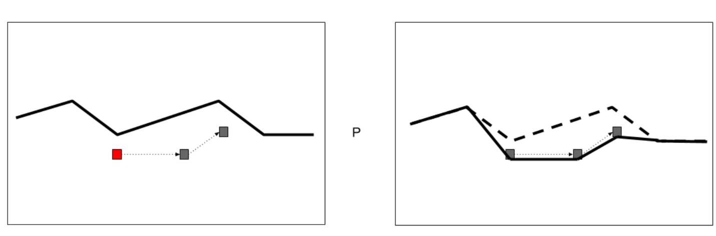

Fig. 4. Cascading effects of deleting a competency, on the learner's proficiency skyline

Note that the deleted competency $c(p, q)$ may not be on the skyline, but may just be part of the set of mastered competencies ($M_S(l)$) by the learner. When an acquired competency is removed, all other competencies for which this was an implied or stated pre-requisite, are also removed. Hence, removal of an acquired competency may involve adjustments in several parts of the skyline.

Figure 4 shows the cascading effects of deletion of a competency on the learner's skyline. The competency shown in red in the first figure is deleted. This competency has a chain of pre-requisites to competencies further in the map. When the competency is deleted, not only is the skyline adjusted for that domain, it is also adjusted for the rest of the domains affected by the competency.

Comparison of skylines. A learner's skyline $\psi_S(l)$ is said to "cover" the skyline of another learner $\psi_S(u)$, if for every competency from a given domain, the proficiency level of $\psi_S(l)$ is no lesser than the proficiency level of $\psi_S(u)$. Formally:

$$\psi_S(l) \sqsupseteq \psi_S(u) \Rightarrow \forall c_l(p_1, q) \in \psi_s(l), c_u(p_2, q) \in \psi_S(u), p_2 \leq p_1 \quad (6)$$

The skylines of two learners $\psi_s(l)$ and $\psi_S(u)$ are said to be equivalent (denoted as $\psi_s(l) \equiv \psi_S(u)$), if $\psi_s(l) \sqsupseteq \psi_S(u)$ and $\psi_S(u) \sqsupseteq \psi_s(l)$. The skyline $\psi_s(l)$ is said to "strictly cover" skyline $\psi_S(u)$ (denoted by $\psi_s(l) \sqsupset \psi_S(u)$) iff: $\psi_s(l) \sqsupseteq \psi_S(u)$ and $\psi_S(u) \not\sqsupseteq \psi_s(l)$. A pair of skylines $\psi_S(l)$ and $\psi_s(u)$ that do not cover one another, are said to be concurrent.

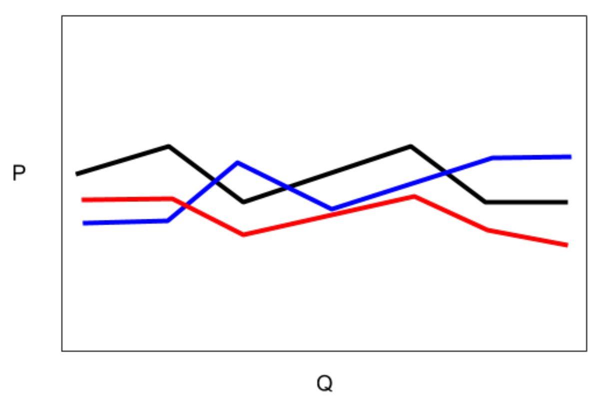

Fig. 5. Example of covering and concurrent skylines

In Figure 5, the black skyline is concurrent with the blue skyline, while it strictly covers the red skyline.

The difference between any two skylines is calculated using a family of distance functions, each of which represents a different interpretation of distance. They are detailed as follows.

Given two competencies for a given domain $c_1(p_1, q)$ and $c_2(p_2, q)$, the term $\Delta_q(c_1, c_2)$ represents the absolute difference between proficiency levels p_1 and p_2. The term $\delta_q(c_1, c_2)$ represents the directed difference from c_1 to c_2. Hence, $\delta_q(c_1, c_2) = -\delta_q(c_2, c_1)$.

Similarly, we also define the "improvement difference" between two competencies along a given domain q, as the levels of proficiency required to go from one competency to the other.

$$i_q(c_1, c_2) = \begin{cases} \delta_q(c_1, c_2) & if\ \delta_q(c_1, c_2) > 0 \\ 0 & otherwise \end{cases} \quad (7)$$

The improvement difference shows how much a learner has to improve along a given domain, to reach a target competency.

The *Manhattan distance*, or the "L_1 norm" between two skylines $\psi_s(l)$ and $\psi_S(u)$ is the sum of total differences between proficiency levels across all domains. Formally:

$$L_1(\psi_s(l), \psi_S(u)) = \sum_{\forall c_l \in \psi_s(l), c_u \in \psi_S(u), \forall q \in CM_S} \Delta_q(c_l, c_u) \tag{8}$$

The *grade difference*, or the "L_∞ norm" between two skylines is the maximum difference in their proficiency levels:

$$L_\infty(\psi_s(l), \psi_S(u)) = \max_{\forall q \in CM_S} \Delta_q(c_l, c_u), \forall c_l \in \psi_s(l), c_u \in \psi_S(u) \tag{9}$$

Given a skyline $\psi_s(l)$, we define a "difference vector" to a target skyline $\psi_S(u)$ as a set of directed distance functions between this and the target skyline, for each domain:

$$\vec{\delta}(\psi_s(l), \psi_S(u)) = \{(q, \delta_q(c_l, c_u)) \mid c_l \in \psi_S(l), c_u \in \psi_S(u), q \in CM_S\} \tag{10}$$

An "improvement vector" from a given skyline to a target skyline is also defined analogously:

$$\vec{i}(\psi_s(l), \psi_S(u)) = \{(q, i_q(c_l, c_u)) \mid c_l \in \psi_S(l), c_u \in \psi_S(u), q \in CM_S\} \tag{11}$$

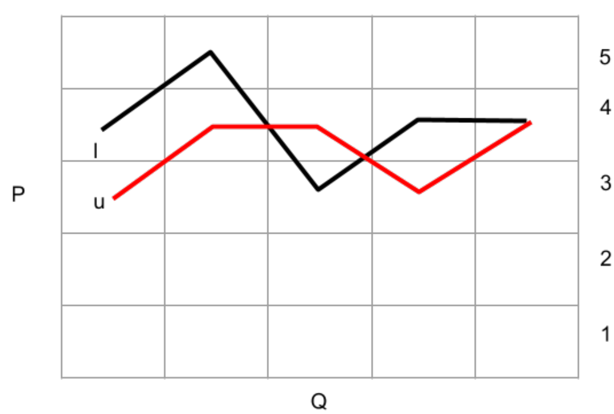

Fig. 6. Computing skyline distances

In Figure 6 two skylines l (in black) and u (in red) are shown. Distances between them can be calculated as follows:

$$L_1(l, u) = 4$$
$$L_\infty(l, u) = 1$$
$$\vec{\delta}(l, u) = (-1, -1, 1, -1, 0)$$
$$\vec{\delta}(u, l) = (1, 1, -1, 1, 0)$$
$$\vec{i}(l, u) = (0, 0, 1, 0, 0)$$
$$\vec{i}(u, l) = (1, 1, 0, 1, 0)$$

Skyline group operations. A set of learner skylines have to be considered as a unit, when we need to reason about the performance of a class as a whole. For this, we define a set of operators that operate on a set of skylines.

Let $\psi_S = \{\psi_S(l_1), \psi_S(l_2), \ldots, \psi_S(l_n)\}$ be the set of learner skylines for a class of n learners for subject S. (We will drop the subscript S whenever the context is clear). For a given skyline $\psi_S(i)$, let the term $p_q(\psi_S(i))$ denote the proficiency level (value on the P axis) for the skyline for domain q.

The "bottomline" $\perp_\psi$ of the class is represented as a virtual skyline showing the lowest competency obtained by the class, for every domain in the subject. The "topline" $\top_\psi$ is a virtual skyline showing the highest competency obtained by the class for every domain in the subject.

$$\perp_\psi = \{c(p,q) \mid \forall q \in Q_S, c(p,q) = \min_{\psi_S(i) \in \psi} p_q(\psi_S(i))\} \tag{12}$$

$$\top_\psi = \{c(p,q) \mid \forall q \in Q_S, c(p,q) = \max_{\psi_S(i) \in \psi} p_q(\psi_S(i))\} \tag{13}$$

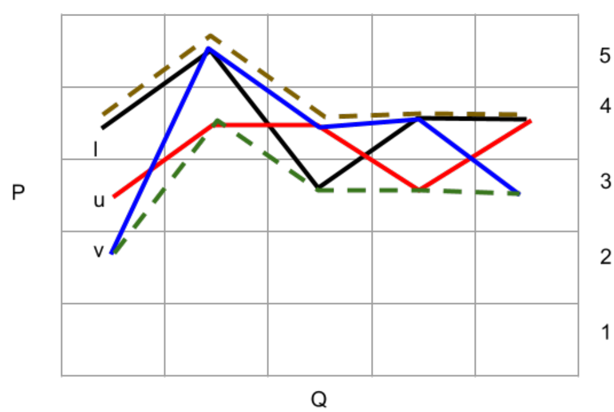

Fig. 7. Operations on groups of skylines

Figure 7 shows the proficiency profile of a class of three learners l, u and v, represented by their respective skylines. The figure also shows the bottomline and the topline, in the form of dashed skylines.

Given a class of learners ψ, the "grade disparity" Δ_ψ of the class, is defined as the total amount of competency units required to be executed by the class, to bridge between the bottomline and the topline. In other words, the grade disparity is the Manhattan distance between the topline and the bottomline.

$$\Delta_\psi = L_1(\top_\psi, \perp_\psi) \tag{14}$$

Similarly, given a class of learners ψ, the "grade separation" ∞_ψ of the class is the maximum disparity in proficiency levels in the class. This can be computed as the grade difference between the topline and the bottomline.

$$\infty_\psi = L_\infty(\top_\psi, \perp_\psi) \tag{15}$$

Given a class of learners ψ, the "mode" of the class Mo_ψ is a virtual skyline connecting proficiencies across domains where most of the learners are currently situated. For a given domain, if there is more than one mode, the lowest of them is chosen to represent the mode of the class.

In Figure 7, the mode of a class is a virtual skyline connecting the following proficiency levels across the first five domains: $(2, 5, 4, 4, 4)$.

The improvement vector from the bottomline to the mode of the class: $\vec{i}(\bot_\psi, Mo_\psi)$, represents the intervention profile suggested to the teacher, to help upskill learners with the lowest proficiency levels, to match with the rest of the class. Similarly, for any given learner l, the improvement vector between the learner's skyline and the mode of the class: $\vec{i}(\psi(l), Mo_\psi)$, represents the upskilling profile required by the learner to catch up with proficiency levels of the rest of the class.

Given any class of learners ψ, and a competency map for a subject C_S, the *weight* of any competency $c \in C_S$ (denoted as $w_S(c)$) is defined as the number of learners who have mastered the given competency. The "skyline weight" of competency c, denoted as $\hat{w}_S(c)$ is the number of skylines in the class that contain competency c. If domain of competency c is d, then the skyline weight of c represents how many learners in the class have c as the maximum proficiency that they have obtained in domain d.

Given any virtual skyline representing a collective characteristic (like $\top_\psi, \bot_\psi, Mo_psi$, etc.), the weight of the skyline is defined as the sum of the weights of all competencies that are part of the skyline:

$$w(\top_\psi) = \sum_{\forall c \in \top_\psi} w(c) \tag{16}$$

Let the Mode of a class at some time, be represented as Mo^0_ψ. After some intervention (named as s) by the teacher, let the mode of the class be represented by the skyline Mo^1_ψ. The "lift" obtained by the intervention is given by:

$$lift(s) = \frac{w(Mo^1_\psi) - w(Mo^0_\psi)}{N} L_1(Mo^1_\psi, Mo^0_\psi) \tag{17}$$

where $N = |\psi|$ is the total number of learners in the class. The lift score for an intervention, is a measure that combines the total gain obtained by the intervention, and the number of learners affected by the intervention.

5 LEARNING PATHWAYS AND NARRATIVE ARC

This section discusses the automatic generation of Learning pathways, Learning pathway network and personalized Narrative Arcs built over the learning pathway network.

A characteristic property of personalisation is "elasticity" in the system, even though designed for the mythical average student, it needs to adapt to the actual individual student to various degrees. In other words, first a common curriculum or a common sequence of competencies is designed which is basically a learning pathway. This learning pathway is then tweaked or personalized to each student according to their profiles and preferences and is called Narrative Arc.

Basically, learning pathways are created that connect different competencies in the competency map. These learning pathways together form a learning pathway network over the competency map. This is analogous to the road network on the physical map of locations, the learning pathways are like the roads formed on the competency map. Once there is a learning pathway network built, a learner takes the route comprising of the learning pathways to reach his learning goal from his current position(skyline). This route is called Narrative Arc. The analogy here is again the driver of a vehicle choosing a route comprising of different roads to reach his destination from his current position.

The learning pathways generated should be such that there is a semantic smoothness of transition from one competency to the next and overall learning pathway is semantically coherent. Even when learning resources address the same topic, their exposition styles may be different, covering different sets of other supporting topics. Hence it is important to ensure that topical dissonance is minimized as the learner progresses through a learning path. Further, as we progress through the learning pathway, the learner should learn new concepts or competencies. The learning pathway should thus provide novelty while maintaining the coherence. To create effective learning experiences, the learning pathway should have the right balance between coherence and novelty.

To generate the learning pathway networks, we create virtual learning resources at every competency in the competency map. The virtual learning resources are created using the different learning resources that are mapped to the competency such that the virtual learning resource correctly represents the topic on x-axis at the complexity level of that topic on y-axis in the competency map.

We propose a model to automatically generate the learning pathways and eventually a learning pathway network using these virtual learning resources. We use re-enforcement model to compute the learning pathways. This model has two components, a greedy generator that generates the learning pathways and a validator that validates if a learning pathway generated is valid or good for learning. The two components communicate with each other such that the validator provides feedback to greedy generator and the re-enforcement helps to improve the output of the greedy generator.

6 MEDIATION FOR NAVIGATED LEARNING

This section discusses meadiation and characterization of mediation based on cognitive and learning science principles that will be used in navigated learning.

The above proposed pedagogic model has two components. The first is the learning navigator, which helps a learner to navigate through the progression space (discussed in section 2) giving a smooth learning experience, and to get an overview of the learning map for any subject. The second component is the back-end community engine, which gives all users the opportunity to interact and contribute to the progression space on which the learning navigator operates.

While curating the content in the proposed system, a teacher can specify a learning path. The system also suggests a learning path based on a learner profile built by its deep-end combined with the information obtained through the *Event-Condition-Action (ECA)* rules. We define an *event* as any situation which the learner faces, either within the system or outside the system. The *conditions* are the information points obtained from the learner profile and learning activity vector. The *event trigger* is an element (a combination of events and conditions) that can trigger thinking or learning. *Suggest interventions* (i.e. *mediations*) can be seen as manipulations of pathways. The suggest system intervenes in several such workflows on a continuous basis, based on semantics extracted from systemic activities. Suggest algorithms are triggered by ECA rules embedded in different parts of the system.

Here is an example of an event: Leena, a 9th Grade student, takes a photo of a flower at Lalbagh botanical garden, and her phone GPS sends co-ordinates of her location to the system. Leena's profile information, such as *she has mastered flora identification; she has learned independently Botany-I with 60% proficiency; she has studied the history of Karnataka with 70% proficiency as a part of her classroom learning; she prefers videos and images*, could be considered as conditions. Combinations of any given event with given conditions could be considered an event trigger that can initiate one or more suggest interventions. The suggest trigger may intervene not just in the student's learning pathway, but also in the workflows of other pertinent stakeholders like the teacher, course content creator, etc. A system of Notifiers and Listeners manages ECA rules by subscribing to different forms of event notifications and checking the relevant condition to finally call a pertinent suggest method. The suggest of learning pathways, i.e. the mediation offered, will be determined by a thorough understanding of the learner – including their progress, performance, proficiency, preferences, portfolio, markers, history, and goals – and of the learning activity – including its relevance, engagement and efficacy – providing a consistent learning experience throughout.

The term distributed cognition refers to the idea that cognitive processes are distributed across members of social groups, i.e. the processes may be distributed among the coordination between internal structures (i.e. processing and understanding of information) and external structures (i.e. external environment). It is difficult to understand how an individual learns without looking at his/her interactions with other people and artifacts. Cognition happens through interactions between people and artifacts or tools situated in the sociocultural context. In distributed cognition, cognitive processes such as understanding of previous concepts are distributed through time as well, so that earlier events affect later events [12]. [25] portrayed organizational cognition as distributed cognition, where people get a rich representation of their knowledge and understanding by self-reflection and communication. The cognitive structures get formed and re-structured through communication and shared understanding. This concerns how the learner learns by using various resources and interacting with those resources, be they humans or artifacts or technology-based resources. [6] argued that information technology can support distributed cognition by offering ways to communicate, interact, and reflect to form the rich representation of their understanding and restructure the same by reflecting upon and sharing it with others and receiving feedback. The term mediated learning refers to learning with interventions by a mediator (for example, a human expert) and/or through an organized learning activity [11, 14]. An environment in which the learner's interactions with learning materials (such as readings, exercises, assignments, peers and/or instructors) are mediated through advanced information technologies is called technology-based mediated learning (TML) [1]. At the heart of the concept of mediated learning, there is a theory of structural cognitive modifiability (SCM) developed by [11], which says that a learner's cognitive structure (i.e. the way s/he learns or her/his intelligence) changes through expert interventions [15]. Hence, it is also characterized as learning to learn. Vygotsky [29] introduced the notion of the social origins of individual psychological functions. He defined the zone of proximal development (ZPD) as the distance between the actual developmental level as determined by independent learning and the level of potential development as determined by the expert's guidance. In other words, Vygotsky said that while learning, a learner has two performance levels: the level the learner can attain individually and the level the learner can attain with help of an expert. The latter (i.e. the learner's achievement with the help of an expert) is characterized as ZPD. The theory of ZPD specifies that the development of higher mental processes depends on the presence of the mediating agent that intervenes when a learner is interacting with a learning environment [14]. The assumption is that as a result of personal interaction in ZPD with the expert, the learner will eventually be able to create a functional system for independent learning. Smagorinsky [27] argued that Vygotsky's ZPD is wrongly interpreted as instructional scaffolding, and as the notion of ZPD is connected with development through social interaction, which is a long-term continuous process, ZPD should be accurately translated as "zone of next development" (ZND).

The basic prerequisites for achieving the vision discussed earlier include the capability to locate the learner, curate potential learning activities, and make appropriate suggestions based on their context. The suggest sub-system leverages key elements from learning science, cognitive science, neuroscience and data science to design the most precise learning pathway for every learner.

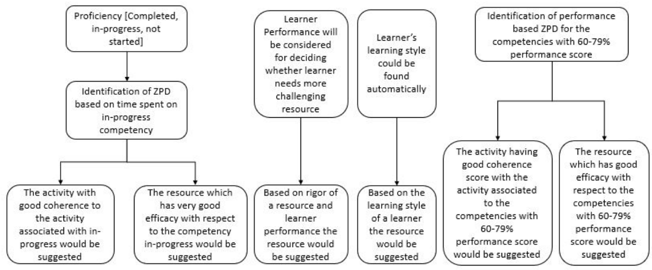

Fig. 8. Fully automated mediations

This requires multivariate optimization across different learning pathways based on learning, cognitive and data sciences. By using the available data about a learner and the available resources, her/his ZPD (set of competencies) would be identified, and the same would be used for mediation. Identification of ZPD automatically by using one or more data points is crucial for mediation. The very basic way to identify ZPD is through the immediate progressions (i.e. connected competencies) of a mastered (80% performance) competency. Competencies for which the performance of a learner is within the 60- 79% range would be identified as the learner's ZPD. The set of competencies associated with the resources that are most coherent with the resources associated with a mastered competency would be considered as a possible ZND for a learner. We can also identify the ZND of a learner based on time spent on the in-progress competencies in a learning map. Learner scores like authority, citizenship, and reputation, together with user preferences, would be considered for identifying a ZPD.

We categorize learning mediations into two major types: *fully automated mediation* and *semi-automated peer-driven mediation*. These two types of mediations can be characterized in detail based on elements such as learner context (i.e. the context in which the learner is), learner characteristics and learner profile. We consider the learning navigator as a hypothetical teacher who is observing the learner continuously, and also as a mediator that suggests learning resources automatically. This approach of mediations we name *fully automated mediation*. After determining the location of that learner for the domain on the learning map, the system would identify the ZPD/ ZND of the learner, which would be used for mediation. A few heuristics are given in figure 8. In these fully automated mediations, the system acts as mediator, a hypothetical teacher who understands the learner.

By considering the inputs from distributed cognition, we also consider the other users - i.e. the peers and even the teacher - as another set of mediators. The system will suggest to these mediators the learning resource and the point of mediation, i.e. the point where an intervention could be made. These mediators can use their own judgement to accept the suggestion or consider something else. These mediators would be decided automatically by the system based on their interaction among each other, and hence this kind of mediation by these mediators is named semi-automated peer-driven mediation. After identification of ZPD of a user automatically, the system would suggest a peer or a group of peers to a user for further interaction. A few heuristics are given in figure 9.

The essence of cooperative learning is students learning in groups, which increases motivation and improves interrelationships and as positive effects on achievements and higher-order thinking [10]. Considering group dynamics, the system can identify groups in which good students dominate and may thus affect the performance of the whole group in a positive manner, the system can

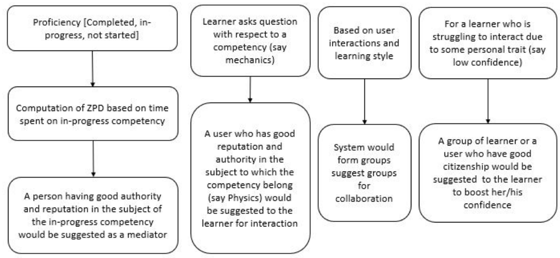

Fig. 9. Semi-automated peer-driven mediations

then suggest that a lower-performance learner in a subject should be part of a group with high-performing students in that subject who have good citizenship and reputation.

Even after the mediation with basic heuristics, the system would observe the learner to see the effect of the mediation. If the mediation is effective, then the system would remember that heuristic as a validated strategy, and if the mediation is found to be ineffective, then it would remember this and adjust the mediation with another heuristic. There could be pre-defined heuristics, or the system could learn the heuristics for mediation from user interactions. Thus, mediations by the system would be controlled by the user interaction with the system, i.e. learning resources and peers. Usually, adaptive systems are pre-designed for adaptation of the content and offering the content in a personalized manner to the user. So, an adaptive system's control is always only with the system and hence its design. In the proposed system, user interaction with the system is the prime driver for emergence of the network, and hence these medications are adjusted as per user interaction. So, mediations would not be controlled single handedly by a user or even by a system. In a technology- based mediated navigated learning platform, the mediation will be a balanced control between the system (the hypothetical teacher) and the users.

REFERENCES

[1] Maryam Alavi and Dorothy E Leidner. 2001. Research commentary: Technology-mediated learning—A call for greater depth and breadth of research. *Information systems research* 12, 1 (2001), 1–10.

[2] John R Anderson, C Franklin Boyle, and Brian J Reiser. 1985. Intelligent tutoring systems. *Science* 228, 4698 (1985), 456–462.

[3] Albert Bandura. 1969. Social-learning theory of identificatory processes. *Handbook of socialization theory and research* 213 (1969), 262.

[4] Albert Bandura. 1986. Social foundations of thought and action. *Englewood Cliffs, NJ* 1986 (1986).

[5] David M Blei, Andrew Y Ng, and Michael I Jordan. 2003. Latent dirichlet allocation. *Journal of machine Learning research* 3, Jan (2003), 993–1022.

[6] Richard J Boland Jr, Ramkrishnan V Tenkasi, and Dov Te'Eni. 1994. Designing information technology to support distributed cognition. *Organization science* 5, 3 (1994), 456–475.

[7] Christopher G Brinton, Ruediger Rill, Sangtae Ha, Mung Chiang, Robert Smith, and William Ju. 2015. Individualization for education at scale: MIIC design and preliminary evaluation. *IEEE Transactions on Learning Technologies* 8, 1 (2015), 136–148.

[8] Peter Brusilovsky. 2001. Adaptive educational hypermedia. In *International PEG Conference*, Vol. 10. 8–12.

[9] Peter Brusilovsky. 2003. Developing adaptive educational hypermedia systems: From design models to authoring tools. In *Authoring tools for advanced technology Learning Environments*. Springer, 377–409.

[10] Zoltán Dörnyei. 1997. Psychological processes in cooperative language learning: Group dynamics and motivation. *The Modern Language Journal* 81, 4 (1997), 482–493.

[11] Reuven Feuerstein, Mildred B Hoffman, Yaacov Rand, Mogens R Jensen, David Tzuriel, and David B Hoffmann. 1985. Learning to learn: Mediated learning experiences and instrumental enrichment. *Special Services in the Schools* 3, 1-2 (1985), 49–82.

[12] Edwin Hutchins. 2000. Distributed cognition. *International Encyclopedia of the Social and Behavioral Sciences. Elsevier Science* 138 (2000).

[13] Heisawn Jeong and Cindy E. Hmelo-Silver. 2010. An Overview of CSCL Methodologies. In *Proceedings of the 9th International Conference of the Learning Sciences - Volume 1 (ICLS '10)*. International Society of the Learning Sciences, 921–928.

[14] Alex Kozulin. 2003. Psychological tools and mediated learning. *Vygotsky's educational theory in cultural context* (2003), 15–38.

[15] Alex Kozulin, Jo Lebeer, Antonia Madella-Noja, Fernando Gonzalez, Ingrid Jeffrey, Naama Rosenthal, and Mennyi Koslowsky. 2010. Cognitive modifiability of children with developmental disabilities: A multicentre study using Feuerstein's Instrumental Enrichment—Basic program. *Research in Developmental Disabilities* 31, 2 (2010), 551–559.

[16] Douglas C Merrill, Brian J Reiser, Michael Ranney, and J Gregory Trafton. 1992. Effective tutoring techniques: A comparison of human tutors and intelligent tutoring systems. *The Journal of the Learning Sciences* 2, 3 (1992), 277–305.

[17] Libby V Morris. 2013. MOOCs, emerging technologies, and quality. *Innovative Higher Education* 38, 4 (2013), 251–252.

[18] Fan Qing and Lin Li. 2011. Web-Based Collaborative Learning. *Procedia Environmental Sciences* 11 (2011), 189 – 192. DOI: http://dx.doi.org/https://doi.org/10.1016/j.proenv.2011.12.029 2011 2nd International Conference on Challenges in Environmental Science and Computer Engineering (CESCE 2011).

[19] Prasad Ram and Srinath Srinivasa. 2017. *Designing Social Machines for Mediated Learning Environments*. Technical report. Gooru Labs.

[20] L Todd Rose, Parisa Rouhani, and Kurt W Fischer. 2013. The science of the individual. *Mind, Brain, and Education* 7, 3 (2013), 152–158.

[21] Todd Rose. 2016. *The end of average: How to succeed in a world that values sameness*. Penguin UK.

[22] Gavriel Salomon. 1997. *Distributed cognitions: Psychological and educational considerations*. Cambridge University Press.

[23] Valerie J Shute and Diego Zapata-Rivera. 2012. Adaptive educational systems. *Adaptive technologies for training and education* 7, 27 (2012), 1–35.

[24] Barry G. Silverman. 1995. Computer Supported Collaborative Learning (CSCL). *Computers Education* 25, 3 (1995), 81 – 91. DOI: http://dx.doi.org/https://doi.org/10.1016/0360-1315(95)00059-3

[25] Herbert A Simon. 1991. Bounded rationality and organizational learning. *Organization science* 2, 1 (1991), 125–134.

[26] Derek Sleeman and John Seely Brown. 1982. *Intelligent tutoring systems*. London: Academic Press.

[27] Peter Smagorinsky. 2018. Deconflating the ZPD and instructional scaffolding: Retranslating and reconceiving the zone of proximal development as the zone of next development. *Learning, culture and social interaction* 16 (2018), 70–75.

[28] Carole Steketee. 2006. Modelling ICT integration in teacher education courses using distributed cognition as a framework. *Australasian Journal of Educational Technology* 22, 1 (2006).

[29] Lev Vygotsky. 1978. Interaction between learning and development. *Readings on the development of children* 23, 3 (1978), 34–41.

[30] Li Yuan, Stephen Powell, JISC CETIS, and others. 2013. MOOCs and open education: Implications for higher education. (2013).