Data-Driven Design of Study Plans

Rakesh Agrawal
President, Data Insights Laboratories
Rukmini Chair Professor, Indian Institute of Science
Synthesizing Study Plans

Imagine you are an instructor who wants to offer a new course. You know the concepts you want to teach in the course, but need help with formulating the study plan:

a. What concepts should you cover in one session
b. The sequencing of sessions

Joint Work with Behzad Golshan and Evangelos Papalexakis [EDM 2016]
Study Plans

Input

- Force
- Displacement
- Speed
- Mass
- Harmonic Motion
- Acceleration
- Waves
- Sound

Study Plan #1

- Speed
- Displacement
- Acceleration
- Harmonic Motion
- Waves
- Sound
- Force
- Mass

Study Plan #2

- Speed
- Acceleration
- Harmonic Motion
- Sound
- Mass
- Waves
- Force
Why is this problem important?

The increasing demand for many more courses of smaller duration on specialized topics

The trend accentuated by increasing availability of inexpensive devices connected to fast Internet

- 250 million smartphones in India within 2 years!
Outline

• Problem Statement
• Related Work
• Our Method
• Experimental Results
• Future Work & Conclusions
Outline

• Problem Statement
• Related Work
• Our Method
• Experimental Results
• Conclusions & Future Work
Axioms

• Learning Unit
  • A group of coherent concepts suitable to be covered together
  • Cohesion: Concepts within a learning unit must be closely related
  • Isolation: Concepts in different learning units must be independent
  • Unity: A concept should be covered in one unit

• Study plan
  • An ordering of some number of learning units
  • Prerequisite compliance: L1 < L2 => concepts in L2 not needed for L1
  • Locality of references: L2 builds upon L1 => L2 should come soon after L1
Problem Statement

Given a set of concepts,

- Partition them into a given number of learning units, and
- Provide a sequencing of learning units

such that an objective function $f$ is minimized

Problem 1 (Study Plan Design Problem). Given a concept graph $G = \langle V, E \rangle$ with $n > 0$ nodes, and the number of desired learning units $m$ ($m \leq n$), output an ordered vector of learning units $L = \langle L_1, L_2, \cdots, L_m \rangle$ to

Minimize: $f(L)$

s.t. $\forall i : L_i \subseteq V$, $L_i \neq \emptyset$, and

$\cup_i L_i = V$. 
Objective Function

\[ f(\mathcal{L}) = \sum_{\substack{\pi(u) < \pi(v) \\ (u,v) \in E}} (\pi(v) - \pi(u)) \ast C_r, + \sum_{\substack{\pi(u) > \pi(v) \\ (u,v) \in E}} (\pi(u) - \pi(v)) \ast C_p + \sum_{\substack{\pi(u) = \pi(v) \\ (u,v) \not\in E}} C_d. \]

- Prerequisite Compliance Violation
- Locality of Reference Violation
- Cohesion Violation (Also Isolation)
- Unity Violation Penalized by 1st two terms
Outline

• Problem Statement
• Related Work
• Our Method
• Experimental Results
• Conclusions & Future Work
Related work

On concept maps:

On learning spaces:
Outline

• Problem Statement
• Related Work
• Our Method
• Experimental Results
• Conclusions & Future Work
Concept Graph

- Directed graph $G = \langle V, E \rangle$
  - Nodes correspond to concepts
  - Edge $e = (u, v)$ implies
    - $u$ and $v$ are related concepts
    - $u$ is a prerequisite of $v$
Method

0) Creating the Graph
1) Finding Learning Units
2) Ordering Learning Units
3) Organizing Learning Units
Method

0) Creating the Graph
1) Finding Learning Units
2) Ordering Learning Units
3) Organizing Learning Units

On identifying prerequisites:

Inferring the graph using Wikipedia:
To be discussed along with experimental results
Method

0) Creating the Graph
1) Finding Learning Units
2) Ordering Learning Units
3) Organizing Learning Units

Partition the graph into dense communities
• Spinglass (Reichardt et al. 2006)
• Allows to adjust the cost of missing edges
• Available in the statistical package R
Method

0) Creating the Graph
1) Finding Learning Units
2) Ordering Learning Units
3) Organizing Learning Units

Ordering the learning units
• Shorter forward edges
• Fewer and Shorter backward edges

Theorem: This is NP-hard.
• Proof: Reducing the Minimum Linear Arrangement problem to our problem

Effective heuristics exist for MinLA problem
• Simulated annealing [SS98]
Method

0) Creating the Graph
1) Finding Learning Units
2) Ordering Learning Units
3) Organizing Learning Units

- Step 1: Find the Strongly Connected Components
- Step 2: Sort topologically the components
  - Not unique (This implies flexibility)
Outline

• Problem Statement
• Related Work
• Our Method
• Experimental Results
  • Creating the Concept Graph
  • Synthesizing the Study Plan
• Conclusions & Future Work
Input: 139 high school physics concepts from CK12.org

Question: Can we use Wikipedia to induce the concept graph?
- Map each concept to a Wikipedia page
- Find the Wikipedia graph between these pages
  - Edges are hyperlinks that link to another page
The CK12 Wikipedia Graph
Problem in using Wikipedia Graph as Concept Graph

Edges do not capture the prerequisite relationships

Two main sources of errors:

• Informative edge, but the direction is wrong
  “capacitors” → “Joule”
  Conventional capacitors provide less than 360 joules per kilogram of energy

• Unrelated edges
  “capacitors” → “lasers”
  Low-inductance high-voltage capacitors (capacitor banks) are used to supply huge pulses of current for many pulsed power applications. These include electromagnetic forming, Marx generators, pulsed lasers
Correcting Errors

• A machine learning approach
  • We manually labeled the edges.

• Features
  • In-degree
  • Out-degree
  • # of languages
  • # of categories
  • …

<table>
<thead>
<tr>
<th>ACTUAL EDGE TYPE</th>
<th>PREDICTION</th>
<th>OUTCOME</th>
</tr>
</thead>
<tbody>
<tr>
<td>prereq.</td>
<td>29</td>
<td>142</td>
</tr>
<tr>
<td>unrelated</td>
<td>9</td>
<td>668</td>
</tr>
<tr>
<td>reverse</td>
<td>0</td>
<td>110</td>
</tr>
</tbody>
</table>

Many misclassified as unrelated

But few are misclassified in the wrong directions
## Synthesized Study Plan

<table>
<thead>
<tr>
<th>Unit 1 (20 concepts)</th>
<th>Unit 2 (21 concepts)</th>
<th>Unit 3 (14 concepts)</th>
<th>Unit 4 (18 concepts)</th>
</tr>
</thead>
<tbody>
<tr>
<td>buoyancy</td>
<td>acceleration</td>
<td>atom</td>
<td>calorimetry</td>
</tr>
<tr>
<td>euclidean vector</td>
<td>angular momentum</td>
<td>bohr model</td>
<td>change of state</td>
</tr>
<tr>
<td>force</td>
<td>angular velocity</td>
<td>conservation of energy</td>
<td>combined gas law</td>
</tr>
<tr>
<td>free body diagram</td>
<td>centripetal force</td>
<td>elastic collision</td>
<td>conversion of units</td>
</tr>
<tr>
<td>friction</td>
<td>circular motion</td>
<td>inelastic collision</td>
<td>double-slit experiment</td>
</tr>
<tr>
<td>impulse</td>
<td>displacement</td>
<td>kinetic energy</td>
<td>energy</td>
</tr>
<tr>
<td>inclined plane</td>
<td>keplers laws of ...</td>
<td>mass versus weight</td>
<td>energy conversion</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Unit 5 (27 concepts)</th>
<th>Unit 6 (28 concepts)</th>
<th>Unit 7 (11 concepts)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ammeter</td>
<td>beat</td>
<td>doppler effect</td>
</tr>
<tr>
<td>capacitor</td>
<td>color</td>
<td>general relativity</td>
</tr>
<tr>
<td>capacitors in circuits</td>
<td>concave lens</td>
<td>half-life</td>
</tr>
<tr>
<td>electric charge</td>
<td>conduction</td>
<td>length contraction</td>
</tr>
<tr>
<td>electric current</td>
<td>curved mirror</td>
<td>mathematical physics</td>
</tr>
<tr>
<td>electric field</td>
<td>diffraction</td>
<td>newtons law of univer...</td>
</tr>
<tr>
<td>electric power</td>
<td>diffraction grating</td>
<td>rc time constant</td>
</tr>
</tbody>
</table>
Organizing Learning Units

<table>
<thead>
<tr>
<th>ID</th>
<th>Concepts</th>
<th>ID</th>
<th>Concepts</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>weightlessness</td>
<td>11</td>
<td>simple harmonic motion</td>
</tr>
<tr>
<td>2</td>
<td>projectile motion</td>
<td>12</td>
<td>acceleration</td>
</tr>
<tr>
<td>3</td>
<td>motion</td>
<td>13</td>
<td>centripetal force</td>
</tr>
<tr>
<td>4</td>
<td>moment of inertia</td>
<td>14</td>
<td>newton’s laws of motion</td>
</tr>
<tr>
<td>5</td>
<td>vector addition</td>
<td>15</td>
<td>kepler’s laws of ...</td>
</tr>
<tr>
<td>6</td>
<td>displacement</td>
<td>16</td>
<td>orbital motion</td>
</tr>
<tr>
<td>7</td>
<td>velocity</td>
<td>17</td>
<td>motion graphs</td>
</tr>
<tr>
<td>8</td>
<td>angular velocity</td>
<td>18</td>
<td>uniform acceleration</td>
</tr>
<tr>
<td>9</td>
<td>angular momentum</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>rotation around a ...</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>kinematics</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>pendulum</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Unit 2

Diagram:

```
  1  2  3  4  5  6  7  8  9  10  11  12  13  14  15  16  17  18  19
```

1. weightlessness
2. projectile motion
3. motion
4. moment of inertia
5. vector addition
6. displacement
7. velocity
8. angular velocity
9. angular momentum
10. rotation around a ...
11. kinematics
12. pendulum
13. simple harmonic motion
14. newton’s laws of motion
15. kepler’s laws of ...
16. orbital motion
17. motion graphs
18. uniform acceleration
19. circular motion
User Study

• Recruited 9 domain experts (Physics teachers, Graduate students)

• They were given the following tasks:
  1) *Count the number of odd concepts in each learning unit that you believe do not belong to the unit*
  2) *Without changing any of the learning units proposed, what order do you suggest?*
Results of the User Study

- Number of concepts that do not belong in the respective unit:

<table>
<thead>
<tr>
<th></th>
<th>Min</th>
<th>Max</th>
<th>Median</th>
<th>Mean</th>
<th># Concepts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unit 1</td>
<td>1</td>
<td>6</td>
<td>3.0</td>
<td>3.4</td>
<td>20</td>
</tr>
<tr>
<td>Unit 2</td>
<td>0</td>
<td>3</td>
<td>1.0</td>
<td>1.1</td>
<td>20</td>
</tr>
<tr>
<td>Unit 3</td>
<td>1</td>
<td>7</td>
<td>3.5</td>
<td>3.7</td>
<td>14</td>
</tr>
<tr>
<td>Unit 4</td>
<td>0</td>
<td>5</td>
<td>2.0</td>
<td>1.8</td>
<td>18</td>
</tr>
<tr>
<td>Unit 5</td>
<td>0</td>
<td>4</td>
<td>1.0</td>
<td>1.0</td>
<td>26</td>
</tr>
<tr>
<td>Unit 6</td>
<td>0</td>
<td>3</td>
<td>0.5</td>
<td>0.9</td>
<td>28</td>
</tr>
<tr>
<td>Unit 7</td>
<td>0</td>
<td>5</td>
<td>1.0</td>
<td>1.4</td>
<td>11</td>
</tr>
</tbody>
</table>

- Only two participants ordered the units somewhat differently
- The high school Physics teacher: our study plan was very clever
Outline

• Problem Statement
• Related Work
• Our Method
• Experimental Results
• Conclusions & Future Work
Recap

• We formalized the problem of synthesizing study plans automatically
• We provided a novel and pragmatic solution
• Our method did not use domain specific knowledge
  • Generalizing to other areas seems promising
• Our experimental results as well as the user study show that the problem of creating study plans is amenable to computational approaches
Further Work

• Incorporating user modeling into the system
  • Creating study plans that suit students background/interests/abilities
• Investigating how human input (implicit or explicit) can improve the quality of generated study plans
Bigger Picture: Datafication of Education

How to enhance the quality of the electronic textbooks?
How to form teams of students in a class?
How to create study plans for courses?
Diagnostic tools for identifying weaknesses in textbooks

Within section deficiencies
- Syntactic complexity of writing and dispersion of key concepts in the section [AGK+11a]

Across sections deficiencies
- Comprehension burden due to non-sequential presentation of concepts [ACG+12]

Algorithmic enhancement of textbooks for enriching reading experience

References to selective web content
- Links to authoritative articles [AGK+10], images [AGK+11b] and videos [ACG+14] based on the focus of the section

References to prerequisites
- Links to concepts necessary for understanding the present section, derived using a model of how students read textbooks [AGK+13]

Validation on textbooks from U.S.A and India, on different subjects, across grades

Joint work with Sreenivas Gollapudi, Anitha Kannan, Krishnaram Kenthapadi, et al.
Forming Beneficial Teams of Students

Joint work with Behzad Golshan and Evimaria Terzi [KDD 2014]
Research Opportunities in Data-Driven Education

- Validation of experimental results through deployment
- Synergies with crowd-sourcing approaches
- Use of logs of interactions data and personalization
- Performance evaluation methodologies and benchmarks
- Issues related to privacy, security, confidentiality, copyright, royalty ...

*Magic happens when what is desperately needed meets what is technically feasible*


Questions?