Computational Education: A Data Opportunity?

Rakesh Agrawal
Microsoft Research
Mountain View, California
October 14, 2013
Thinking About Education

Three key questions:

• What is being taught
  – Curriculum, syllabus, educational material

• How it is being delivered
  – Teachers, classes, assessments

• How it is funded
  – Business models
Emergent Perfect Storm

• Electronic textbooks
  – Fast adoption of cloud-connected electronic devices (worldwide)
  – Open content (e.g. OpenStax, ck12.org, NCERT)

• Internet-based classes
  – MOOCs (e.g. Coursera, EdX, Udacity, Khan, TED-Ed)
  – Small virtual classes (e.g. Shankar Mahadevan Academy)
  – Electronic certification (e.g. Mozilla’s OpenBadges)

• New models of funding education
  – Recipients give back to the seed fund for future recipients at their pace (e.g. Dakshana)
  – Market for options on future earnings
Data Mining for Enriching Electronic Textbooks

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+13] 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
• Prototypes and research papers (see References)

Joint work with Sreenivas Gollapudi, Anitha Kannan, Krishnaram Kenthapadi, et al.
Some Data-Centric Research Questions

• Inferring learning units and dependence between them from current educational material (knowledge graph)
• Improvement in educational material based on data on student interactions with the material
• Personalized learning plans
• Dynamic formation of classes and study groups
• Performance evaluation methodologies and benchmarks

Meta Question

• Will we play or cede the space to others?
Data & Education: A Historical Perspective

• Readability Formulas (starting [Lorge 1939])
  – Coefficients of regression equations (e.g. over McCall-Crabbs Standard Test Lessons)

<table>
<thead>
<tr>
<th>Formula</th>
<th>Formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flesch Reading Ease Score</td>
<td>C = Number of words with three syllables or more</td>
</tr>
<tr>
<td>Flesch-Kincaid Grade Level</td>
<td>D = Number of words on the Dale Long List</td>
</tr>
<tr>
<td>Dale-Chall Grade Level</td>
<td>L = Number of letters</td>
</tr>
<tr>
<td>Gunning Fog Index</td>
<td>S = Number of syllables</td>
</tr>
<tr>
<td>SMOG Index</td>
<td>T = Number of sentences</td>
</tr>
<tr>
<td>Coleman-Liau Index</td>
<td>W = Number of words</td>
</tr>
<tr>
<td>Automated Readability Index</td>
<td></td>
</tr>
</tbody>
</table>

  

P(θi) = exp(-(θi - βj)) / (1 + exp(-(θi - βj)))

• Item Response Theory (starting 1950s, in use in ETS)

• Intelligent Tutoring Systems (starting [Pressy 1924])
  – Adapt tutoring strategies based on student actions
  – Biennial ITS conferences starting 1988, Also AIED, EDM Conf.


