Big Data in Education

Assessment of the New Educational Standards

Markus Iseli, Deirdre Kerr, Hamid Mousavi
Big Data in Education

“Technology is disrupting education, expanding the education ecosystem beyond traditional lecture halls and classrooms to accommodate learners' preferences for time, place, style and previous levels of attainment.” [Gartner Technology Research, 2013]

Educational data is gathered through traditional assessments and increasingly also through educational games and simulations. => Lots of data!
The new Common Core State Standards (CCSS) in math and English language arts and the Next Generation Science Standards (NGSS) propose a shift from multiple-choice items towards short-answer response items or short essays.


Deep learning and application of knowledge, rather than memorization and rote learning are favored.
Consequences of New Standards

• The new standards require new assessment approaches.
• Tradeoff between problem realism and scoring difficulty: The more naturalistic the assessment item, the harder it is to score.
• Assessment of explanations, argumentation, and communication mostly require constructed responses (i.e. text)
• Approximately 25 million students will have to write at least two essays a year, not including the number of short answer constructed responses. [Bennett, 2013]
• Valid, informative, automated assessment scoring is needed.
Automated Scoring

• Machine scoring of constructed responses is generally not amenable to exact-matching approaches because the specific form(s) and/or content of the correct answer(s) is not known in advance.

• The state of the art of automated scoring depends highly on the domain that is scored.

• Some answer types that have been scored automatically to-date are essays, short-text, equations and expressions (math).

• We will focus on automated evaluation of text...
Why Automated Text Evaluation?

- To score student essays or short-answer responses.
- To understand textual information on the web, in books, or any other information resource.
- To organize information semantically so that it can easily be found and accessed.
- To make sense of Big Data.
Some Issues in Automated Essay Scoring

• Currently, automated essay scoring uses machine learning algorithms and models which are mainly trained on statistical text measures, such as word choice, text length, grammatical measures, and punctuation.

• The resulting models are not scalable, since they are highly domain/topic-dependent and require large amounts of human-scored essay data.

• The model outcomes are scores, numbers that are correlated with human rater scores but that contain no information about essay content and provide almost no clues on the underlying reasons of this number's value. (more diplomatic)
How We Addressed These Issues

• Focus on the evaluation of essay content and leave the evaluation of statistical text properties to other engines.

• Teach the computer how humans evaluate content: give it some rules and algorithms.

• Provide the computer with a reference text, which contains all the content that should be mentioned in the essay.

• These rules and algorithms can be applied to any other textual data. If scalable, even to Big Data.
Text to Graph

Text/Essay

Reference Text

convert

Graph

Reference Graph

Δ

Δ

convert
System Overview

The reader is transported through the ear canal to stimulate the ear. The nerve is activated, causing the ear to detect the sound waves. The sound waves travel through the ear canal to stimulate the ear. The nerve impulse is transmitted to the brain via the auditory nerve. The brain interprets the stimulus as sound.
The outer ear catches the sound waves and puts them through the ear canal.

Parser

NP
  DT  the
  JJ  outer
  NN  ear

VP
  VBZ  catches

NP
  DT  the
  NN  sound
  NNS  waves

CC  and

VBZ  puts

NP
  PRP  them

etc.
TextGraph

NP
  DT the
  JJ outer
  NN ear

VP
  VBZ catches
  NP
    DT the
    NN sound
    NNS waves

CC and
VBZ puts
NP
  PRP them

etc.

sound
  prop_of
  det_of
  the

catches
  obj_of
  waves
  det_of
  the

the
  det_of
  ear

and
  sub_of
  waves
  sub_of
  the

canal
  prop_of
  ear

through
  prop_of
  them

TD rules

National Center for Research on Evaluation, Standards, & Student Testing
The outer ear catches the sound waves and puts them through the ear canal.
Essay Evaluation Study

• Write rules and algorithms to extract propositions

• Apply rules and algorithms to 55 short essays about the human hearing process (4th and 5th grade students)

• Apply *same* rules and algorithms to 54 short essays about the vision process (4th and 5th grade)

• Provide reference essays and scoring rubrics in terms of propositions for both topics

• Compare performance of our rules and algorithms for the two topics and compare to human analysis
The outer ear (auricle) catches sound waves. Sound waves travel through the ear canal to vibrate the eardrum. The vibrating eardrum passes vibrations on to the middle ear, which is made up of the hammer, anvil, and stirrup. The middle ear passes vibrations to the inner ear, via the stirrup. The stirrup vibrates the oval window. The vibrating oval window causes the fluids in the cochlea to vibrate. The cochlea converts the sound waves to electrical impulses via the fibers in the cochlea. The electrical impulses from the cochlea travel to the brain via the auditory nerve. The brain interprets the vibration as sound.
Hearing Processing Sample Essay

When the phone rings, it goes to the outer ear. The outer ear catches the sound waves and puts them through the ear canal. Then the eardrum vibrates and makes the three bones to shake. Then the last bone of the three bones connects to the oval window and the cochlea makes waves like water that shakes in an earthquake and it sends it to a vein. Then it sends the message to the brain and you will hear it and you will pick it up.
Proposition Comparison

- **Sound waves**
  - Ear drum
  - Ear canal
  - Middle ear
  - Oval window
  - Inner ear
  - Cochlea
  - Vein
  - Brain
  - Message
  - Electrical impulses
- **Outer ear**
  - Phone ring
  - Telephone
  - Ear canal
  - Ear drum
  - Three bones
  - Last bone
  - Oval window
- **Sound waves**
  - Shake in water
  - Waves
  - Cochlea
  - Vein
  - Brain
  - Message
  - Electrical impulses
- **Outer ear**
  - Phone ring
  - Ear canal
  - Ear drum
  - Three bones
  - Last bone
  - Oval window
  - Inner ear
  - Cochlea
  - Vein
  - Brain
  - Message
  - Electrical impulses
- **Sound waves**
  - Shake in water
  - Waves
  - Cochlea
  - Vein
  - Brain
  - Message
  - Electrical impulses
# Study Results in Terms of Scores

**Comparison of SemScape to Human Raters for Hearing Essays**

<table>
<thead>
<tr>
<th>Human score</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>Percent matched</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>100%</td>
</tr>
<tr>
<td>2</td>
<td>-</td>
<td>7</td>
<td>3</td>
<td>-</td>
<td>-</td>
<td>70%</td>
</tr>
<tr>
<td>3</td>
<td>-</td>
<td>2</td>
<td>6</td>
<td>3</td>
<td>-</td>
<td>55%</td>
</tr>
<tr>
<td>4</td>
<td>-</td>
<td>-</td>
<td>8</td>
<td>15</td>
<td>2</td>
<td>60%</td>
</tr>
<tr>
<td>5</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>4</td>
<td>4</td>
<td>50%</td>
</tr>
</tbody>
</table>

\[ r = .798, \ qwk = .792 \]

**Comparison of SemScape to Human Raters for Vision Essays**

<table>
<thead>
<tr>
<th>Human score</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>Percent matched</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>100%</td>
</tr>
<tr>
<td>2</td>
<td>-</td>
<td>6</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>50%</td>
</tr>
<tr>
<td>3</td>
<td>-</td>
<td>-</td>
<td>8</td>
<td>9</td>
<td>2</td>
<td>42%</td>
</tr>
<tr>
<td>4</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>11</td>
<td>6</td>
<td>61%</td>
</tr>
<tr>
<td>5</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>4</td>
<td>80%</td>
</tr>
</tbody>
</table>

\[ r = .682, \ qwk = .6 \]
Study Results in Terms of Propositions

- Proposition Extraction: avg. precision 79%, avg. recall 65%
- The 35% of propositions that were not successfully extracted were split between Wrong (20%) and Missing (15%) propositions.
- Wrong propositions were largely due to incorrect pronoun resolution
- Missing propositions were largely due to uncommon or inaccurate sentence structures for which there were no applicable TextGraph rules
Issues and Possible Improvements

- Parse tree generation: Only extract information needed
- Rule generation is language and language style dependent and is manual: Use automatic rule generation strategies
- Graph comparison: Simplify graph structure and split into subgraphs to compare more complex patterns
- Ambiguous statements will still be ambiguous
- World knowledge ontologies required
Take Away

• No human-scored training data is required, only reference text or reference propositions

• Once the rules are established, they don’t need to be changed for different domains/topics since they are purely grammar-dependent

• Content analysis provides feedback on understanding and misconceptions

• Approach can be applied to Big Textual Data – once it has been engineered some more...

Thank You!
National Center for Research on Evaluation, Standards, & Student Testing

UCLA | Graduate School of Education & Information Studies

VISIT US ON THE WEB
cresst.org

iseli@cse.ucla.edu

Copyright © 2012 The Regents of the University of California
Text Domain (TD) Rules

Given:       Output:
0   NP       <0,1 prop_of 0,2>
0,0   DT
0,1   JJ|NN
0,2   NN|NNS

NP
  DT   the
  JJ   outer
  NN   ear

VP
  VBZ  catches

NP
  DT   the
  NN   sound
  NNS  waves

CC   and
VBZ   puts

NP
  PRP  them

etc.
Graph Domain (GD) Rules

Given: \(<a \text{ sub_of } b> \text{ and } <c \text{ obj_of } b>\)
Output: \(<a, b, c>\)