Imputing KCs with Representations of Problem Content and Context

Zachary A Pardos\textsuperscript{1,2} \quad Anant Dadu\textsuperscript{3}

1. Graduate School of Education
2. School of Information
3. Computer Science & Engineering
Motivation

• The Importance of KSA (Knowledge, Skills, and Abilities) tags
Motivation

• The Importance of KSA (Knowledge, Skills, and Abilities) tags

- Enables adaptivity based on skill mastery
  (Corbet & Anderson, UMUAI 1994; Pardos & Heffernan, UMAP 2011)
Motivation

• The Importance of KSA (Knowledge, Skills, and Abilities) tags
  - Enables adaptivity based on skill mastery
    (Corbet & Anderson, UMUAI 1994; Pardos & Heffernan, UMAP 2011)
  - Used to generate reports for the teacher and learner
    (Duval, LAK 2011)
Motivation

• The Importance of KSA (Knowledge, Skills, and Abilities) tags
  - Enables adaptivity based on skill mastery  
    (Corbet & Anderson, UMUAI 1994; Pardos & Heffernan, UMAP 2011)
  - Used to generate reports for the teacher and learner  
    (Duval, LAK 2011)
  - Allows teacher and learner to find relevant content
Motivation

• What are KSAs (Knowledge, Skills, and Abilities)?

- Knowledge Component (KC) Mode
  (Cognitive Tutor)

- Skill Model
  (ASSISTments)

- Problem Metadata

- Domain Model

<table>
<thead>
<tr>
<th>Item</th>
<th>Skill 1</th>
<th>Skill 2</th>
<th>Skill 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Item 1</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item 2</td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Item 3</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Item 4</td>
<td></td>
<td></td>
<td>X</td>
</tr>
</tbody>
</table>

(Birenbaum, Kelly, & Tatsuoka, 1993)
Motivation

- The problem: KSAs missing meta-data in user sourced content

ASSISTments Platform
- Majority content from teachers
- Robust content builder
- 71% of problems are missing KCs

Assistments 2012-2013 dataset
Past approaches to impute KC meta-data

• Collaborative filtering using non-negative MF (Desmarais, 2012)

• Content based classification using problem text as Bag-of-Words (Rosé, Donmez, Gweon, Knight, Junker, Cohen, Koedinger, Heffernan, 2005; Mariheida, Córdova-Sánchez, Pardos, 2012)

• Context
Intuition for our context approach

Skills of observed problem set sequence:

Multiplying large integers  Square roots #2  Pythagorean Theorem

Observed problem set sequence with missing skills of second set

Multiplying small integers  ??  Pythagorean Theorem
Intuition for our context approach

Skills of observed problem set sequence:

- Multiplying large integers
- Square roots #2
- Pythagorean Theorem

Observed problem set sequence with missing skills of second set

- Multiplying small integers
- ??
- Pythagorean Theorem

- Sequence of problem sets answered by students but assigned by their teacher
- Exploit context to impute the skill of the middle problem set’s problems
Methodology

• **Skip-gram (word2vec) algorithm**  Mikolov, T., & Dean, J. (2013)

![NLP example diagram]

Hyper parameters

• Length of hidden layer
• Size of context window

*Previously only used for NLP*
Methodology

- Skip-gram (word2vec) algorithm  
  Mikolov, T., & Dean, J. (2013)

NLP example

\[
\text{[KING]} - \text{[MAN]} + \text{[WOMAN]} \approx \text{[QUEEN]}
\]

(Mikolov, Yih, & Zweig, 2013)
Methodology

- Skip-gram (word2vec) algorithm  
  Mikolov, T., & Dean, J. (2013)

NLP example

RQ: Can this method embed problems into a “skill space”?

(Mikolov, Yih, & Zweig, 2013)
Methodology

Student problem sequence example

1. Learn problem vectors
2. Classify held-out problem vectors using their similarity (cosine or Euclidian) to problem vectors averaged by skill
3. Alternatively, learn a vector classification model and apply it to held out vectors
Data

• ASSISTments (2012-2013)
  • 28,000 students’ problem answering sequences
  • 100 answers per student on average
  • Only problems with existing skills used
  • 5 fold CV used to train/test vector & BOW approach
Results

Do problems of the same KC cluster together in the vector space?

- Answer: Not substantially. Accuracy of 33.95% achieved with distance based classification methods

<table>
<thead>
<tr>
<th>Rank</th>
<th>Optimization</th>
<th>Token</th>
<th>Acc</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Cosine</td>
<td>Correctness</td>
<td>0.3395</td>
</tr>
<tr>
<td>2</td>
<td>Euclidian</td>
<td>Correctness</td>
<td>0.3348</td>
</tr>
<tr>
<td>3</td>
<td>Variance</td>
<td>Correctness</td>
<td>0.3335</td>
</tr>
<tr>
<td>4</td>
<td>Euclidian</td>
<td>no Correctness</td>
<td>0.3313</td>
</tr>
<tr>
<td>5</td>
<td>Cosine</td>
<td>no Correctness</td>
<td>0.3252</td>
</tr>
</tbody>
</table>
Results

Is vector distance based classification improved by optimizing to a validation set?

- Answer: Yes. Accuracy substantially increased to **55.97%** with validation set. Demonstrates word2vec models can significantly overfit.

<table>
<thead>
<tr>
<th>Rank</th>
<th>Distance</th>
<th>Token</th>
<th>CV</th>
<th>Acc</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Euclidean</td>
<td>C</td>
<td>P</td>
<td>0.5597</td>
</tr>
<tr>
<td>2</td>
<td>Cosine</td>
<td>C</td>
<td>P</td>
<td>0.5490</td>
</tr>
<tr>
<td>3</td>
<td>Euclidean</td>
<td>noC</td>
<td>P</td>
<td>0.5000</td>
</tr>
<tr>
<td>4</td>
<td>Cosine</td>
<td>noC</td>
<td>P</td>
<td>0.4917</td>
</tr>
<tr>
<td>5</td>
<td>Euclidean</td>
<td>C</td>
<td>B</td>
<td>0.2941</td>
</tr>
</tbody>
</table>
Results

Is additional KC information encoded in a distributed fashion in the vectors?

• Answer: Yes. Using a 100 node hidden layer FFNN to train a classification based on the vectors achieved a **86.43%** accuracy on the test set.

<table>
<thead>
<tr>
<th>Rank</th>
<th>Optimization</th>
<th>Token</th>
<th>CV</th>
<th>Acc</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Validation</td>
<td>noC</td>
<td>P</td>
<td>0.8643</td>
</tr>
<tr>
<td>2</td>
<td>Validation</td>
<td>C</td>
<td>P</td>
<td>0.8585</td>
</tr>
<tr>
<td>3</td>
<td>Variance</td>
<td>noC</td>
<td>P</td>
<td>0.8489</td>
</tr>
<tr>
<td>4</td>
<td>Variance</td>
<td>C</td>
<td>P</td>
<td>0.8418</td>
</tr>
<tr>
<td>5</td>
<td>Cosine</td>
<td>C</td>
<td>P</td>
<td>0.7086</td>
</tr>
</tbody>
</table>
How does this context vector approach compare to Bag-of-Words? Can prediction improve when combined?

- BOW approach achieves similar accuracy with 88.18% and is improved further by combining context vector and content approach, with an accuracy of 90.30%.

<table>
<thead>
<tr>
<th>Rank</th>
<th>Model</th>
<th>Acc (B)</th>
<th>Acc (P)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Combined</td>
<td>0.7322</td>
<td>0.9030</td>
</tr>
<tr>
<td>2</td>
<td>BOW</td>
<td>0.7260</td>
<td>0.8818</td>
</tr>
<tr>
<td>3</td>
<td>Supervised</td>
<td>0.6547</td>
<td>0.8643</td>
</tr>
<tr>
<td>4</td>
<td>Distance</td>
<td>0.2941</td>
<td>0.5597</td>
</tr>
</tbody>
</table>
Conclusions

• The same amount of information about KSAs exists in the problem content (well studied) as problem context (understudied)

• Teacher behaviors contain signal (assignment of problems)

• Representation learning (Word2vec) embedded problems into a vector space which encoded skill information
Computational Approaches to Human Learning (CAHL) Research Code and data: https://github.com/CAHLR/

Zachary A Pardos (zp@berkeley.edu)

Anant Dadu (anant.dadu.cse14@iitbhu.ac.in)