Prerequisite Concept Maps Extraction for Automatic Assessment

Lei Liu
HP Labs, USA
(with Shuting Wang from PSU)
Outline

• Motivation

• Prerequisite Concept Map Construction

• Prerequisite Concept Map Delivery

• Experiments

• Conclusion & Future work
Outline

• Motivation

• Prerequisite Concept Map Construction

• Prerequisite Concept Map Delivery

• Experiments

• Conclusion & Future work
Motivation

• Background:
  • Automated assessments are important
    • Fast changing technologies
    • Rapid growth of knowledge resources
  • Assessment tools: Concept maps based assessment systems, Rubrics ...
  • Advantages of prerequisite concept maps based assessments
    • Find user’s knowledge gap
    • Knowledge structure representation

• Goal:
  • Extracting domain prerequisite concept map and facilitating interactive assessments based on the extracted concept maps to find learner’s knowledge gap

• Application:
  • Classroom learning
  • Employee training
A Running Example

1. Start with Classes.
2. Choose Java Keywords.
3. Select Private.
4. Write a constructor with the following parameter.
5. Correct Answer: Constructor.
6. Private constructor.
7. Which of the following private constructors is correct?
8. Wrong Answer: Private.
9. A private method could be visited by...?
Existing Methods & Drawbacks

• Automated prerequisite concept map construction
  • Mining from structured textual data sources
  • Translating ontologies into concept maps [Kim et al.]
  • Drawbacks: Cannot be applied to areas without ontologies
  • Mining from unstructured text
  • Identify concept relations from content and context [Chen & Bai, Larranaga et al.]
  • Drawbacks: Did not consider articles’ structures

• Proposed method
  • Using knowledge from educational resources such as textbooks and slides, and web knowledge base such as Wikipedia and Freebase.
Outline

• Motivation

• Prerequisite Concept Map Construction

• Prerequisite Concept Map Delivery

• Experiments

• Conclusion & Future work
Prerequisite Concept Map Construction

• We propose a two-phase method to extract the prerequisite concept map
  • Phase 1: Node extraction - **Key concept extraction**
    • Extract domain key concepts from textbooks with the help of Wikipedia
  • Phase 2: Link extraction - **Prerequisite relation identification**
    • Identify the prerequisite relations among key concepts
Key Concept Extraction

- Wikipedia Concept Crawler
  - Construct a domain specific dictionary using Wikipedia by crawling Wikipedia using some root words: e.g., mathematics.

- Concept Matcher
  - Extract related Wikipedia concepts that appear in articles.

- Candidates Ranker
  - Rank Wikipedia concepts using similarity between the book content and Wikipedia concept page content and select top-N candidates.
Prerequisite Relation Identification

Two criteria to select pairs of concepts with prerequisite relations:

- First criteria: One concept is used in another concept’s definition

  *Example*: Distributed storage-> Apache Hadoop
  *(Apache Hadoop is an open-source software framework written in Java for distributed storage...)*

- Second criteria:
  - Content overlap (similarity)
  - Learning Level: fundamental concepts have high learning level and should be learn first
    - Link based: Wikipedia links structure (No. in-links / out-links)
    - Content based: entropy of topic distribution
Content based Learning Level

We leverage two types of relationships, Concept overlap and specificity.

- **Concept overlap**
- **Specificity**

focused Concept

broader Concept
Specificity Relationships

- **Specificity equivalence relation** $d_i \leftrightarrow d_j$:
  $d_i$ and $d_j$ cover the same topics at the same level

  $$d_i \leftrightarrow d_j \iff |g(d_i) - g(d_j)| \leq k \land o(d_i, d_j) \geq t$$

- **Specificity $d_i$ is a prerequisite to $d_j$** $d_i \rightarrow d_j$:

  $$d_i \rightarrow d_j \iff g(d_i) > g(d_j) \land o(d_i, d_j) > 0 \land (|g(d_i) - g(d_j)| > k \lor o(d_i, d_j) < t)$$

$g(d_i)$: learning level for $i$-th concept

$o(d_i, d_j)$: overlaps (similarity)
Outline

• Motivation

• Prerequisite Concept Map Construction

• Prerequisite Concept Map Delivery

• Experiments

• Conclusion & Future work
Top-k Concepts Delivery

- Questions: how to select top-k concepts to deliver to users
  - Connectivity:
    - Edge density: \[
    \frac{|E|}{|V|(|V| - 1)}
    \]
- Concept importance:
  - PageRank score \textit{Importance(.)} based on prerequisite concept graph

\[
Q(y) = \frac{1}{2} \sum_{ij} W_{ij} \left[ \frac{y_i}{\sqrt{D_{ii}}} - \frac{y_j}{\sqrt{D_{jj}}} \right] + \frac{\mu}{2} \sum_i (y_i - y_i^{(0)})^2
\]

- Objective Function:

\[
\arg \max_{v} \Psi = \alpha D + \beta \sum_{v \in V} \text{Importance}(v)
\]

where \(\alpha, \beta\) are parameters to tradeoff between subgraph connectivity and concept importance
Outline

• Motivation

• Prerequisite Concept Map Construction

• Prerequisite Concept Map Delivery

• Experiments

• Conclusion & Future work
Experiment Results

• Employee training: Big Data
  • Employee training: Big Data
  • Classroom learning: Mathematics

• Prerequisite Concept Map Evaluation
  • F - 1 Score

• Concept Map Delivery
  • Parameter tuning:
    • $\alpha$: graph connectivity
    • $\beta$: concept importance
Outline

• Motivation

• Prerequisite Concept Map Construction

• Prerequisite Concept Map Delivery

• Experiments

• Conclusion & Future work
Experiment Results – Prerequisite Concept Map

• Dataset:
  • Sampling two topics (Big data and Mathematics) from the domain dictionary and manually label the prerequisite relationships between the concepts

• Baseline
  • Using the first criteria of concept relationship extraction: whether one concept is used in another concept’s definition.

• Evaluation metric: F score:

\[
F1 = 2 \frac{pr}{p + r} \quad \text{where} \quad p = \frac{tp}{tp + fp}, \quad r = \frac{tp}{tp + fn}
\]
# Evaluation - Mathematics

Table 1: F-1 score for mathematics concept map extraction

<table>
<thead>
<tr>
<th>Content similarity</th>
<th>40%</th>
<th>60%</th>
<th>80%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learning level difference</td>
<td>Two phase</td>
<td>Baseline</td>
<td>Two phase</td>
</tr>
<tr>
<td>40%</td>
<td>0.31</td>
<td>0.11</td>
<td>0.46</td>
</tr>
<tr>
<td>60%</td>
<td>0.43</td>
<td>0.08</td>
<td>0.61</td>
</tr>
<tr>
<td>80%</td>
<td>0.42</td>
<td>0.13</td>
<td>0.55</td>
</tr>
</tbody>
</table>
# Evaluation – Big Data

Table 2: F-1 score for big data concept map extraction

<table>
<thead>
<tr>
<th>Content similarity</th>
<th>40%</th>
<th>60%</th>
<th>80%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learning level difference</td>
<td>Two phase</td>
<td>Baseline</td>
<td>Two phase</td>
</tr>
<tr>
<td>40%</td>
<td>0.35</td>
<td>0.16</td>
<td>0.43</td>
</tr>
<tr>
<td>60%</td>
<td>0.39</td>
<td>0.03</td>
<td>0.52</td>
</tr>
<tr>
<td>80%</td>
<td>0.39</td>
<td>0.2</td>
<td>0.51</td>
</tr>
</tbody>
</table>
Figure 1: Generated top-k concepts with Alpha (connectivity) = 0, Beta (concept importance) = 1 and k=11
Figure 2: Generated top-k concepts with Alpha (connectivity) = 0.4, Beta (concept importance) = 0.6 and k=10
Figure 1: Generated top-k concepts with Alpha (connectivity)= 1, Beta (concept importance) =0 and k=10
Outline

- Motivation
- Prerequisite Concept Map Construction
- Prerequisite Concept Map Delivery
- Experiments
- Conclusion & Future work
Contributions

• Proposed a two-phase method to extract prerequisite concept map which considers both the concept similarity and learning level difference

• Utilized prerequisite concept map to represent knowledge structure and help users to find their knowledge gaps

• Investigated a concept selection algorithm for concept map delivery
Future Work

➢ Different learning methods
  ➢ Bottom up
  ➢ Top down

➢ Different types of questions
  ➢ Programming language
  ➢ Procedure based

➢ Different domains
  ➢ Social science
  ➢ History
Thank you!

lei.liu2@hp.com

Try: www.hpmetis.com
Demo: Youtube + “HP METIS”