Adaptive Sequential Recommendation for Discussion Forums on MOOCs using Context Trees

Fei Mi, Boi Faltings

Artificial Intelligence Lab
École Polytechnique Fédérale de Lausanne

January 17, 2018
Outline

1 Introduction and Motivation
   • MOOCs & Discussion Forums
   • Why Adaptation Matters?

2 The Proposed Context Tree Recommender
   • Context Tree Structure
   • Recommendation using Context Tree
   • Adaptation Analysis

3 Experiment Results

4 Conclusion and Future Work
Development of MOOCs

- Coursera
- Udacity
- edX
- Open 2 Study
- Canvas

Growth of MOOCs

- 58M Students
- 700+ Universities
- 6850 Courses

Source: Class Central
• The only community to exchange ideas
• Boost Engagement and learning effectiveness

• Gaussian distribution → Mean & variance; BFS → DFS, A-star
• Recommend useful threads to students
Compared with Typical RecSys

- Items are static
  - Contents, features, ...
- Collaborative filtering; matrix factorization, ...

Frequently Bought Together

Customers Who Bought This Item Also Bought
1. Forum Threads are Evolving

- Threads are created during the course.
- Contents can be edited and updated frequently.
- A thread can even be superseded by another thread.

→ Recommendations need adapt to evolving threads
2. Drifting User Preference

![Distribution of Thread Views against Freshness](image)

**Figure**: Thread viewing activities against freshness

→ Recommendations need adapt to **drifting preference**
2. Drifting User Preference

![Distribution of Thread Views against Freshness](chart.png)

<table>
<thead>
<tr>
<th>General Threads</th>
<th>Specific Threads</th>
</tr>
</thead>
<tbody>
<tr>
<td>“Using GNU Octave”</td>
<td>“Homework Day 1 / Question 9”</td>
</tr>
<tr>
<td>“Any one from INDIA??”</td>
<td>“Quiz for module 4.2”</td>
</tr>
<tr>
<td>“Where is everyone from?”</td>
<td>“quiz -1 Question 04”</td>
</tr>
<tr>
<td>“Numerical Examples in pdf”</td>
<td>‘Homework 3, Question 11”</td>
</tr>
<tr>
<td>“How to get a certificate”</td>
<td>“Week 1: Q10 GEMA problem”</td>
</tr>
</tbody>
</table>
2. Drifting User Preference

- Mining sequential patterns among fresh & specific threads
• Originally used for data compression [3]
• Applied to news recommendation [1, 2]
• Running on largest French news website
Structure of Context (Suffix) Tree

Definitions:
- Suffix: $\xi = \langle n_3, n_1 \rangle \subset s = \langle n_2, n_3, n_1 \rangle$
- Context (node): all sequences end with the suffix

Properties:
- If $i$ is ancestor of $j$ then $S_j \subset S_i$
- From general to specific contexts
Local Expert for Each Context:
Experts Activation and Mixture of Experts:
**Efficient Computation:**

- Recursive recommendation and parameter update
• Build the CT incrementally (variable-order Markov model)
• Model parameters are updated online
• The CT structure itself
  • old patterns/contexts are kept
  • new patterns/contexts can be identified fast
  • fine-grained model v.s. interpolation
• Build the CT incrementally (variable-order Markov model)
• Model parameters are updated online
• The CT structure itself
  • old patterns/contexts are kept
  • new patterns/contexts can be identified fast
  • fine-grained model v.s. interpolation
Dataset:

- “Digital Signal Processing”
- “Functional Program Design in Scala”
- “Reactive Programming”

<table>
<thead>
<tr>
<th></th>
<th>Course 1</th>
<th>Course 2</th>
<th>Course 3</th>
</tr>
</thead>
<tbody>
<tr>
<td># of forum participants</td>
<td>5,399</td>
<td>12,384</td>
<td>13,914</td>
</tr>
<tr>
<td># of forum threads</td>
<td>1,116</td>
<td>1,646</td>
<td>2,404</td>
</tr>
<tr>
<td># of thread views</td>
<td>130,093</td>
<td>379,456</td>
<td>777,304</td>
</tr>
<tr>
<td># of sessions</td>
<td>19,892</td>
<td>40,764</td>
<td>30,082</td>
</tr>
<tr>
<td>avg. session length</td>
<td>6.5</td>
<td>9</td>
<td>25.8</td>
</tr>
<tr>
<td>avg. # of sessions per student</td>
<td>3.7</td>
<td>3.3</td>
<td>2.2</td>
</tr>
</tbody>
</table>

Evaluation Metric:

- **Succ@5**: MAP of predicting the immediately next thread view
- **Succ@5Ahead**: MAP of predicting the future thread views
### Overall Results of Sequential Methods

<table>
<thead>
<tr>
<th></th>
<th>Non-personalized</th>
<th></th>
<th>Personalized</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Succ@5</td>
<td>Succ@5Ahead</td>
<td>Succ@5</td>
<td>Succ@5Ahead</td>
</tr>
<tr>
<td><strong>Separated Sequences</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CT online-MF</td>
<td>[25, 23, 21]%</td>
<td>[48, 53, 52]%</td>
<td>[19, 14, 16]%</td>
<td>[41, 37, 42]%</td>
</tr>
<tr>
<td>Popular</td>
<td>[15, 12, 8]%</td>
<td>[33, 29, 23]%</td>
<td>[10, 7, 5]%</td>
<td>[27, 25, 20]%</td>
</tr>
<tr>
<td>Fresh_1</td>
<td>[15, 20, 16]%</td>
<td>[40, 61, 51]%</td>
<td>[9, 8, 8]%</td>
<td>[34, 31, 36]%</td>
</tr>
<tr>
<td>Fresh_2</td>
<td>[12, 14, 10]%</td>
<td>[37, 43, 41]%</td>
<td>[10, 10, 8]%</td>
<td>[33, 31, 37]%</td>
</tr>
<tr>
<td></td>
<td>[9, 8, 6]%</td>
<td>[31, 31, 29]%</td>
<td>[8, 7, 6]%</td>
<td>[30, 30, 28]%</td>
</tr>
<tr>
<td><strong>Combined Sequences</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CT online-MF</td>
<td>[21, 20, 20]%</td>
<td>[55, 55, 56]%</td>
<td>[16, 13, 14]%</td>
<td>[46, 39, 46]%</td>
</tr>
<tr>
<td>Popular</td>
<td>[9, 8, 7]%</td>
<td>[34, 27, 23]%</td>
<td>[7, 6, 6]%</td>
<td>[29, 24, 20]%</td>
</tr>
<tr>
<td>Fresh_1</td>
<td>[13, 14, 14]%</td>
<td>[52, 62, 58]%</td>
<td>[9, 8, 7]%</td>
<td>[45, 36, 43]%</td>
</tr>
<tr>
<td>Fresh_2</td>
<td>[10, 12, 9]%</td>
<td>[48, 44, 44]%</td>
<td>[8, 9, 8]%</td>
<td>[44, 34, 42]%</td>
</tr>
<tr>
<td></td>
<td>[7, 6, 6]%</td>
<td>[43, 34, 32]%</td>
<td>[6, 6, 6]%</td>
<td>[42, 32, 31]%</td>
</tr>
</tbody>
</table>

**Table:** Performance comparison of sequential methods
Adaptation to Fresh threads

**Quantity:**

- **Average CDF of Recommendation Freshness (Course 1)**
- **Average CDF of Recommendation Freshness (Course 2)**
- **Average CDF of Recommendation Freshness (Course 3)**

**Figure:** Distribution of recommendation freshness of CT and online-MF

**Quality:**

- **P(Success|Freshness) for Course 1**
- **P(Success|Freshness) for Course 2**
- **P(Success|Freshness) for Course 3**

**Figure:** Conditional success rate of CT and online-MF
Partial Context Matching (PCT)

- adding regularization to generalize to new patterns
- \( \langle n_1, n_2, n_4, n_6 \rangle \) v.s. \( \langle n_1, n_2, n_6 \rangle \)
- PCT - Skip one item

<table>
<thead>
<tr>
<th></th>
<th>Success@5</th>
<th>Success@5Ahead</th>
<th>Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCT-0.5</td>
<td>[+0.4, +0.6, +0.2]%</td>
<td>[+0.8, +0.9, +0.4]%</td>
<td>[4.9, 4.5, 3.3]</td>
</tr>
<tr>
<td>PCT-0.6</td>
<td>[+0.5, +0.8, +0.3]%</td>
<td>[+1.1, +1.3, +0.5]%</td>
<td>[4.4, 4.1, 2.9]</td>
</tr>
<tr>
<td>PCT-0.7</td>
<td>[+0.7, +0.9, +0.5]%</td>
<td>[+1.6, +1.9, +0.7]%</td>
<td>[3.7, 3.2, 2.5]</td>
</tr>
<tr>
<td>PCT-0.8</td>
<td>[+0.8, +1.1, +0.6]%</td>
<td>[+1.9, +2.4, +1.0]%</td>
<td>[3.2, 2.9, 2.1]</td>
</tr>
<tr>
<td>PCT-0.9</td>
<td>[+1.0, +1.4, +0.7]%</td>
<td>[+2.0, +2.7, +1.3]%</td>
<td>[2.4, 2.2, 1.4]</td>
</tr>
</tbody>
</table>

Table: Performance comparison of PCT against CT for three courses
Take-away:

- Apply the CT recommender to MOOCs discussion forum
- It adapts well to evolving threads and drifting preferences.
- Partial context matching technique further boosts performance.
- Adaptation issues in ML

Future work:

- Online evaluations (Learning Experiment)
- Online algorithms (RL, exploration)
