Recommender Systems and Education (with Report on Practical Experiences)

Radek Pelánek
Warm-up Quiz
This Lecture

educational applications with focus on

- relation to topics discussed so far (collaborative filtering, evaluation, ...)
- specific examples
- connections between seemingly different problems/techniques
- personalization and different types of recommendations
- my experience
You are a member of a team developing a personalized learning system, e.g.:

- learning new language
- improving English vocabulary (advanced words)
- learning (advanced) math and machine learning
- (your favourite topic that everybody should learn)

You are responsible for the recommendation part of the project.
How would you approach the problem?
Designing New System: Questions

1. requirements:
   - What is the target group? Who are users?
   - What are user needs?
   - What should it do?

2. techniques, solutions:
   - How should it work?
   - What data you need?
Motivation: Personalization in Education

- each student gets suitable learning materials, exercises tailored to a particular student, adequate for his knowledge (mood, interests, ...)
- **mastery learning** – fixed outcome, varied time (compared to classical education: fixed time, varied outcome)
Motivation: Flow, ZPD

Vygotsky, zone of proximal development
Adaptation and Personalization in Education

... gets lot of attention:

- Khan Academy
- Duolingo
- MOOC courses
- Carnegie Learning
- Pearson
- ReasoningMind
- and many others
Technology and Education

e-learning, m-learning, technology-enhanced learning, computer-based instruction, computer managed instruction, computer-based training, computer-assisted instruction, computer-aided instruction, internet-based training, flexible learning, web-based training, online education, massive open online courses, virtual education, virtual learning environments, digital education, multimedia learning, intelligent tutoring system, adaptive learning, adaptive practice, . . .
Adaptive, Intelligent, and Personalized: Navigating the Terminological Maze Behind Educational Technology
<table>
<thead>
<tr>
<th>Tasks</th>
<th>Description</th>
<th>Generic recommender</th>
<th>TEL recommenders</th>
<th>New requirements</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Existing User Tasks supported by Recommender Systems</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. <strong>Annotation In Context</strong></td>
<td>Recommendations while user carries out other tasks</td>
<td>E.g. predicting how relevant the links are within a web page</td>
<td>E.g. predicting relevance/usefulness of items in the reading list of a course</td>
<td>Explore attributes for representing relevance/usefulness in a learning context</td>
</tr>
<tr>
<td>2. <strong>Find Good Items</strong></td>
<td>Recommendations of suggested items</td>
<td>E.g. receiving list of web pages to visit</td>
<td>E.g. receiving a selected list of online educational resources around a topic</td>
<td>None</td>
</tr>
<tr>
<td>3. <strong>Find All Good Items</strong></td>
<td>Recommendation of all relevant items</td>
<td>E.g. receiving a complete list of references on a topic</td>
<td>E.g. suggesting a complete list of scientific literature or blog postings around a topic</td>
<td>None</td>
</tr>
<tr>
<td>4. <strong>Recommend Sequence</strong></td>
<td>Recommendation of a sequence of items</td>
<td>E.g. receive a proposed sequence of songs</td>
<td>E.g. receiving a proposed sequence through resources to achieve a particular learning goal</td>
<td>Explore formal and informal attributes for representing relevancy to a particular learning goal</td>
</tr>
</tbody>
</table>

Recommender Systems in Technology Enhanced Learning
<table>
<thead>
<tr>
<th>5. JUST BROWSING</th>
<th>Recommendations out of the box while user is browsing</th>
<th>E.g. people that bought this, have also bought that</th>
<th>E.g. receiving recommendations for new courses on the university site</th>
<th>Explore formal and informal attributes for representing relevance/usefulness in a learning context</th>
</tr>
</thead>
<tbody>
<tr>
<td>6. FIND CREDIBLE RECOMMENDER</td>
<td>Recommendations during initial exploration/testing phase of a system</td>
<td>E.g. movies that you will definitely like</td>
<td>E.g. restricting course recommendations to ones with high confidence/credibility</td>
<td>Explore criteria for measuring confidence and credibility in formal and informal learning</td>
</tr>
</tbody>
</table>

### TEL User Tasks that could be supported by Recommender Systems

<table>
<thead>
<tr>
<th>1. FIND NOVEL RESOURCES</th>
<th>Recommendations of particularly new or novel items</th>
<th>E.g. receiving recommendations about latest additions or particularly controversial items</th>
<th>E.g. receiving very new and/or controversial resources on covered topics</th>
<th>Explore recommendation techniques that select items beyond their similarity</th>
</tr>
</thead>
<tbody>
<tr>
<td>2. FIND PEERS</td>
<td>Recommendation of other people with relevant interests</td>
<td>E.g. being suggested profiles of users with similar interests</td>
<td>E.g. being suggested peer students in the same class</td>
<td>Explore attributes for measuring the similarity with other people</td>
</tr>
<tr>
<td>------------------------</td>
<td>------------------------------------------------------</td>
<td>----------------------------------------------------------------</td>
<td>----------------------------------------------------------------</td>
<td>----------------------------------------------------------------</td>
</tr>
<tr>
<td>3. FIND GOOD PATHWAYS</td>
<td>Recommendation of alternative learning paths through learning resources</td>
<td>E.g. receive alternative sequences of similar songs</td>
<td>E.g. receiving a list of alternative learning paths over the same resources to achieve a specific learning goal</td>
<td>Explore criteria for the construction and suggestion of alternative (but similar) sequences</td>
</tr>
<tr>
<td>------------------------</td>
<td>------------------------------------------------------</td>
<td>----------------------------------------------------------------</td>
<td>----------------------------------------------------------------</td>
<td>----------------------------------------------------------------</td>
</tr>
</tbody>
</table>

Recommender Systems in Technology Enhanced Learning
<table>
<thead>
<tr>
<th>Name</th>
<th>Short description</th>
<th>Advantages</th>
<th>Disadvantages</th>
<th>Usefulness for TEL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Collaborative Filtering (CF) techniques</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>User-based CF</td>
<td>Users who rated the same item similarly probably have the same taste. Based on this assumption, this technique recommends the unseen items already rated by similar users.</td>
<td>No content analysis</td>
<td>New user problem</td>
<td>Benefit from experience</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Domain-independent</td>
<td>New item problem</td>
<td>Allocate learners to groups (based on similar ratings)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Quality improves</td>
<td>Popular taste</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Bottom-up approach</td>
<td>Scalability</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Serendipity</td>
<td>Sparsity</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Cold start problem</td>
<td></td>
</tr>
<tr>
<td>Item-based CF</td>
<td>Focus on items, assuming that the items rated similarly are probably similar. It recommends items with the highest correlation (based on ratings for the items).</td>
<td>No content analysis</td>
<td>New item problem</td>
<td>Benefit from experience</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Domain-independent</td>
<td>Popular taste</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Quality improves</td>
<td>Sparsity</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Bottom-up approach</td>
<td>Cold start problem</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Serendipity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stereotypes or demographics</td>
<td>Users with similar attributes are matched, then it recommends items that are preferred by similar users (based on user data instead of ratings).</td>
<td>No cold start problem</td>
<td>Obtaining information</td>
<td>Allocate learners to groups</td>
</tr>
<tr>
<td>CF</td>
<td></td>
<td>Domain-independent</td>
<td>Insufficient information</td>
<td>Benefit from experience</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Serendipity</td>
<td>Only popular taste</td>
<td>Recommendation from the beginning of the PRS</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Obtaining metadata information</td>
<td></td>
</tr>
</tbody>
</table>
**Content-Based (CB) techniques**

<table>
<thead>
<tr>
<th>Case-based reasoning</th>
<th>Assumes that if a user likes a certain item, s/he will probably also like similar items. Recommends new but similar items.</th>
<th>No content analysis</th>
<th>New user problem</th>
<th>Keeps learner informed about learning goal</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Domain-independent</td>
<td>Overspecialisation</td>
<td>Useful for hybrid RS</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Quality improves</td>
<td>Sparsity</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>No cold start problem</td>
<td>Cold start problem</td>
<td></td>
</tr>
<tr>
<td>Attribute-based techniques</td>
<td>Recommends items based on the matching of their attributes to the user profile. Attributes could be weighted for their importance to the user.</td>
<td>No new user/new item problem</td>
<td>Does not learn</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Sensitive to changes of preferences</td>
<td>Only works with categories</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Can include non-item-related features</td>
<td>Ontology modelling and maintenance is required</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Can map from user needs to items</td>
<td>Overspecialisation</td>
<td></td>
</tr>
</tbody>
</table>

---

Personal recommender systems for learners in lifelong learning networks: the requirements, techniques and model
many techniques applicable in principle, but application more difficult than in “product recommendation”

- longer time frame
- pedagogical principles
- domain ontology, prerequisites
- types of knowledge and learning processes (declarative vs procedural knowledge)
- learning outcomes not directly measurable
Evaluation

- evaluation even more difficult than for other recommender systems
- compare goals:
  - product recommendations: sales
  - text (blogs, etc) recommendations: clicks (profit from advertisement)
  - education: learning
- learning can be measured only indirectly
- hard to tell what really works
Examples of Techniques

- adaptive educational hypermedia
- learning networks
- intelligent tutoring systems
Adaptive Educational Hypermedia

- adaptive content selection ~ recommendations
  - most relevant items for particular user
- adaptive navigation support
  - navigation from one item to other
- adaptive presentation
  - presentation of the content
Adaptive Educational Hypermedia

Knowledge Representation Layer
- Learning Resources (e.g. learning objects, media files, description model)
- Domain Ontology (e.g. covered concepts, pre-requisites, curriculum requirements)
- User Model (e.g. learner preferences, previous knowledge, learning objectives, actions log)

Adaptation Layer
- Adaptation Mechanism and Rules

Interface Layer
- Adaptive Content Selection, Navigation and/or Presentation

Recommender Systems in Technology Enhanced Learning
Fig. 2. Evolution of a learning network (left: starting phase with a first learner moving through possible learning activities; right: advanced phase showing emerging learning paths from the collective behavior of all learners)
Intelligent Tutoring Systems

- interactive problem solving
- behavior
  - outer loop – selection/recommendation of “items” (problems, exercises)
  - inner loop – hints, feedback, ...
- adaptation based on learner modeling
- knowledge modeling more involved than “taste modeling” (domain ontology, prerequisites, ...)
Learner Modeling

Bayesian Knowledge Tracing, Logistic Models, and Beyond: An Overview of Learner Modeling Techniques
Graph the inequalities $y = -5x + 12$ and $y = 3x - 2$ to find their solution set. Also, find their point of intersection.

Choose a graphing method for $y = -5x + 12$.

Slope-Intercept

Slope: $-5$

Y-Intercept: $(0, 12)$

Graph Shade

Choose a graphing method for $y = 3x - 2$.

Slope-Intercept

Slope: $3$

Y-Intercept: $(0, -2)$

Graph Shade

Shade the Graph

Shade for the Equation $y = -5x + 12$

Shade for the Equation $y = 3x - 2$

Shade

No Intersection

Select Intersection Region

or

No Intersection Region

X Interval: 1.0

Y Interval: 2
**Scenario**

In metric units, 

- \( v \) is the initial velocity of the object,
- \( h \) is the initial height above the ground.

Suppose that a bottle rocket is shot from ground level with an initial upward velocity of 80 feet per second. Further imagine that the bottle rocket turns out to be a dud. That is, it does not explode in mid-air but simply travels upward to some maximum height then falls back to the ground.

Use the formula above to write an expression for the height of the bottle rocket in terms of the time after it was shot.

Note: Since the bottle rocket is being shot from ground level, its initial height is 0 feet.

**Worksheet**

<table>
<thead>
<tr>
<th>Expression</th>
<th>( -16t^2 + 80t )</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Question 1</strong></td>
<td>1</td>
</tr>
<tr>
<td><strong>Question 2</strong></td>
<td>4.5</td>
</tr>
<tr>
<td><strong>Question 3</strong></td>
<td>2</td>
</tr>
<tr>
<td><strong>Question 4</strong></td>
<td>3</td>
</tr>
<tr>
<td><strong>Question 5</strong></td>
<td>2.5</td>
</tr>
<tr>
<td><strong>Question 6</strong></td>
<td>1.25</td>
</tr>
<tr>
<td><strong>Question 7</strong></td>
<td>( t )</td>
</tr>
</tbody>
</table>

**Grapher**

Graph the height of the bottle rocket as a function of the time since it was shot.

**Questions**

1. How high will the bottle rocket be 1 second after it was shot?
2. How high will the bottle rocket be 4.5 seconds after it was shot?
3. How many seconds after it was shot will the bottle rocket first be 96 feet high?
4. How many seconds after it was shot will the bottle rocket next be 96 feet high?
5. What is the maximum height that the bottle rocket will reach?
6. When is the first time that the bottle rocket will be 76 feet high?
7. How many seconds after being launched will the bottle rocket hit the ground?
Student Modeling and Collaborative Filtering

user \sim student

product \sim problem, question

rating \sim student performance
(correctness of answer, problem solving time, number of hints taken)
Case Studies

- our projects (FI MU) – “adaptive practice”
  - Problem Solving Tutor
  - “Slepé mapy” (Map Outlines) – geography
  - Umíme (česky, anglicky, matiku, ...) – umimeto.org
- Wayang Outpost – math
- ALEF – programming
- CourseRank – course recommender
Problem Solving Tutor

- math and computer science problems, logic puzzles
- performance = problem solving time
- focus: predictions of times
- recommendations – problems of similar difficulty
Problem Solving Tutor

PROBLEM SOLVING TUTOR

Informatické

Interaktivní Python
Konečné automaty
Programování v C
Regulární výrazy
Robot Karel
Robotanik
Želví grafika

Matematické

Binární křížovka
Graf a (nová verze)
Matematické pezso 2
Obrazce
Rozbitá kalkulačka
Transformace
Výpočetní stromy
<table>
<thead>
<tr>
<th>Tutor: predictions</th>
</tr>
</thead>
<tbody>
<tr>
<td>tutor.fi.muni.cz</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Kuželosečky - hyperboly</th>
<th>Komplexní čísla - násobení</th>
<th>Logaritmy a mocniny - vzorečky</th>
<th>Komplexní čísla - mocniny i</th>
<th>Vlastnosti funkci</th>
<th>Kuželosečky 2</th>
<th>Zlomky</th>
<th>Komplexní čísla - absolutní hodnoty</th>
<th>Logaritmy - hodnoty 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neřešeno Předpověď 1:12</td>
<td>Výřešeno Čas 0:58</td>
<td>Neřešeno Předpověď 1:14</td>
<td>Výřešeno Čas 1:24</td>
<td>Neřešeno Předpověď 1:17</td>
<td>Výřešeno Čas 1:05</td>
<td>Neřešeno Předpověď 1:19</td>
<td>Výřešeno Čas 0:45</td>
<td>Neřešeno Předpověď 1:23</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Kvadratické rovnice - řešení</th>
<th>Vzdálenosti</th>
<th>Kuželosečky</th>
<th>Množiny - základní operace</th>
<th>Kombinační čísla</th>
<th>Kvadratická funkce 2</th>
<th>Definiční obory a obory hodnot</th>
<th>Logaritmy - vzorečky</th>
<th>Množiny</th>
</tr>
</thead>
<tbody>
<tr>
<td>Výřešeno Čas 1:16</td>
<td>Neřešeno Předpověď 1:35</td>
<td>Neřešeno Předpověď 1:42</td>
<td>Neřešeno Předpověď 1:42</td>
<td>Neřešeno Předpověď 1:44</td>
<td>Neřešeno Předpověď 1:44</td>
<td>Neřešeno Předpověď 1:46</td>
<td>Neřešeno Předpověď 1:56</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Směs</th>
<th>Derivace - goniometrické funkce</th>
<th>Součty</th>
<th>Kombinační čísla - vzorečky</th>
<th>Komplexní čísla</th>
<th>Úněj 2</th>
<th>Nerovnosti</th>
<th>Kuželosečky - kružnice</th>
<th>Limity funkci</th>
</tr>
</thead>
</table>
Model of Problem Solving Times
data: student $s$ solved problem $p$ in time $t_{sp}$
we need to estimate:
- student skills $\theta$
- problem parameters $a, b, c$

stochastic gradient descent

very similar to the “SVD” collaborative filtering algorithm
Evaluation of Predictions

- 20 types of problems
- data: 5,000 users, 8,000 hours, more than 220,000 problems
- difficulty of problems: from 10 seconds to 1 hour
- offline evaluation: train, test set
- metrics: RMSE
- results:
  - significant improvement with respect to a baseline (mean times)
  - more complex models do not bring much improvement
same basic difficulty

high discrimination  high randomness  "safe" problem

\[ \text{abs(log(abs(x))))} \]
\[ \text{sin(x^3)} \]
\[ -(x+4)^2+2 \]

**Graphs:**

1. \[a = -1.36\quad b = 5.48\quad c = 0.6\]
2. \[a = -0.77\quad b = 5.17\quad c = 0.93\]
3. \[a = -0.73\quad b = 5.11\quad c = 0.55\]
Geography: Map Outlines

- adaptive practice of geography knowledge (facts)
- focus on prior knowledge
- choice of places to practice ~ recommendation (forced)
Jak se jmenuje stát zvýrazněný na mapě?

Finsko  Norsko  Švédsko  Nevím
Geography – Difficulty of Countries
Geography – Model

Model (prior knowledge):
- global skill of a student $\theta_s$
- difficulty of a country $d_c$

Probability of correct answer = logistic function (difference of skill and difficulty):

$$P(\text{correct}|d_c, \theta_s) = \frac{1}{1 + e^{-(\theta_s-d_c)}}$$
Logistic Function

\[
\frac{1}{1 + e^{-x}}
\]
Elo rating system (originally from chess)

\[ \theta := \theta + K(R - P(R = 1)) \]

\[ d := d - K(R - P(R = 1)) \]

- magnitude of update \( \sim \) how surprising the result was
- related to stochastic gradient descent, “SVD” algorithm in collaborative filtering (but only single latent factor)
Geography – Current Knowledge

- estimation of knowledge after sequence of answers for a particular place
- extension of the Elo system
- short term memory, forgetting
question selection (based on predicted probability of correct answer) \sim item recommendation (based on predicted rating)

based on students’ history of answers, we want to create a new question

how?
Geography – Question Selection

question selection (based on predicted probability of correct answer) $\sim$ item recommendation (based on predicted rating)

scoring function – linear combination of several factors:
- predicted success rate, target success rate
- viewed recently
- how many times asked
Geography – Multiple Choice Questions

- number of options – based on estimated knowledge
- choice of options – ???

Example:
- correct answer is Hungary
- we need 3 distractors
- which countries should we use?
choice of options (distractors) – confused places (~ collaborative filtering aspect)

realization: roulette wheel selection (as used in genetic algorithms)
Geography – Evaluation

- evaluation of predictions
  - offline experiment
  - comparison of different models (basic Elo, extensions, …)
  - issue with metrics: RMSE, AUC (⇒ “Metrics for Evaluation of Student Models” paper)
- evaluation of question construction (“recommendations”)
  - online experiment, AB testing
  - issue with metrics: enjoyment vs learning
AB Testing

4 groups:

<table>
<thead>
<tr>
<th>Target item</th>
<th>Options</th>
</tr>
</thead>
<tbody>
<tr>
<td>adaptive</td>
<td>adaptive</td>
</tr>
<tr>
<td>adaptive</td>
<td>random</td>
</tr>
<tr>
<td>random</td>
<td>adaptive</td>
</tr>
<tr>
<td>random</td>
<td>random</td>
</tr>
</tbody>
</table>
Measuring Engagement – Survival Analysis

Stay curve

Users having at least 11 answers

Variant of algorithm for question construction
Measuring Learning

- we cannot measure knowledge (learning) directly
- estimation based on answers
- adaptive questions – fair comparison difficult
- use of “reference questions” – every 10th question is “randomly selected”
- problem with attrition (different number of answers per student)
Measuring Learning – Learning Curves
Other AB Experiments

- difficulty of questions
- choice of distractors (competitive vs adaptive)
- maximal number of distractors
- user control of difficulty
AB experiments

- \( \sim 1000 \) users per day
- Sometimes minimal or no differences between experimental conditions (in the overall behaviour)
- Reasons:
  - Conditions not sufficiently different (differences manifest only sometimes)
  - Disaggregation (users, context) shows differences, which cancel out in overall results
What is suitable target difficulty of questions?
Target success rate:
- 50 %
- 65 %
- 80 %
- 95 %
Difficulty and Explicit Feedback

Out-of-school usage

In-school usage

- Too Easy
- Appropriate
- Too Difficult
Umíme to

- and more... https://www.umimeto.org/
Czech Grammar – Project Evolution

- initial version
  - target audience: adults
  - single exercise type
  - coarse-grained concepts
  - focus on adaptive choice of items

- current version
  - target audience: children
  - many exercise types
  - fine-grained concepts
  - focus on mastery learning
  - several domains
Grammar – Basic Exercise

You _ playing tennis.

was  were
Personalization: Mastery Learning

- skill of the learner – estimated based on the performance, taking into account:
  - correctness of answers
  - response time
  - time intensity of items (median response time)
  - probability of guessing
- mastery criterion – comparison of skill to threshold
- progress bar – visualization of skill
# Umíme to – Skills

## Grammar

### Be, have, do

<table>
<thead>
<tr>
<th>Skill</th>
<th>Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>To be in present simple</td>
<td>lehké</td>
</tr>
<tr>
<td>To do, to have, to be in present simple</td>
<td>1 lehké</td>
</tr>
<tr>
<td>To do, to have, to be: questions and negatives</td>
<td>lehké</td>
</tr>
<tr>
<td>To do, to have, to be in past simple</td>
<td>3 lehké</td>
</tr>
<tr>
<td>Be, have, do: mix</td>
<td>lehké</td>
</tr>
</tbody>
</table>

## Tenses

### Talking about the present

<table>
<thead>
<tr>
<th>Skill</th>
<th>Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Present simple tense</td>
<td>1 lehké</td>
</tr>
<tr>
<td>Present tense: questions and negatives</td>
<td>lehké</td>
</tr>
<tr>
<td>Present simple vs. present continuous</td>
<td>3 lehké</td>
</tr>
</tbody>
</table>
Umíme to – Domain Model

“knowledge components”
- abstract concepts: “capitalization rules”, “addition of fractions”, “present simple tense”
- taxonomy (tree)
- prerequisite relation
- mapping to school grades

“problem sets”
- specific exercise type, set of items
- mapping to knowledge components
Umíme to – Recommendations

the system contains thousands of problem sets ⇒ recommendations are useful

types of recommendations:

- front page dashboard
- “default” practice within a selected exercise or topic
- “follow up” after reaching mastery within some problem set
- spaced repetition recommendations
- (recommendations for teachers)
Umíme to – Recommendation Approach

- IF-THEN rules with priorities
- IF condition THEN recommendation
- conditions:
  - student state / previous activity
  - problem set features / relations
Rule examples:

- X mastered easily and Y follows X with respect to difficulty ⇒ recommend Y, high priority
- X mastered weakly and Y precedes X with respect to topic ⇒ recommend Y, middle priority
- X mastered weakly T days ago ⇒ recommend Y, priority depends on T
- user grade G, X is popular for G ⇒ recommend X, low priority
Umíme to – Recommendation Rules

- different situations use different rules
  - homepage
  - follow-up recommendations
  - spaced repetition recommendations
- evaluation / monitoring with respect to rules: displays, clicks, click through rates
“design adaptation”, “avoiding stupidity”

data $\Rightarrow$ analysis $\Rightarrow$ insights $\Rightarrow$ revision of items or system behaviour

- difficulty of items
- survival analysis, length of practice
- response times
- item similarities
Item Similarities and Clustering

closely related to item-item collaborative filtering
item similarities: Pearson correlation of answers
clustering: $k$-means
visualization: tSNE
key issue: do we have enough data?
Note on Different Approaches

using data, models for:

- “automatic” interventions
  - recommendations
  - personalization choices
  - mastery learning

- support for “manual” interventions
  - items behaviour
  - system behaviour
  - user behaviour

“asking right questions” often more important than “using sophisticated methods”
A Multimedia Adaptive Tutoring System for Mathematics that Addresses Cognition, Metacognition and Affect

adaptive tutoring system for math

Wayang Outpost → MathSpring, http://mathspring.org/

specific feature: focus on affect and metacognition
Dion wants to earn a minimum quiz average of 92% in his biology course. His grades so far are 89%, 95%, and 85%. Which inequality below represents the possible scores for his next quiz which will allow Dion to achieve his goal?

\[
\frac{\text{Sum of the values}}{\text{Number of values}} \geq 92
\]

\[
\frac{89 + 95 + 85 + x}{4} \geq 92
\]

Solve for \(x\).

- A. \(\{x \mid x > 99\}\)
- B. \(\{x \mid x < 99.5\}\)
- C. \(\{x \mid x \geq 99\}\)
- D. \(\{x \mid x \leq 99.5\}\)

\[
269 + x \geq 368
\]

\[
269 + x - 269 \geq 368 - 269
\]

Fig. 1 The Wayang Outpost Math Tutor interface. An animated companion provides individualized comments and support.
The open student model in Wayang is called the Student Progress Page (SPP). It encourages students to reflect on their progress for each topic (column 1). The plant (column 2) demonstrates the tutor’s assessment of student effort, while the mastery bar (column 3) records presumed knowledge (according to Bayesian Knowledge Tracing). The tutor comments on its assessment of the student’s behavior (column 4) and offers students the choice to continue, review or challenge themselves and make informed decisions about future choices (column 5).
Wayang Outpost: Affect, Metacognition

Fig. 11  a. Progress Charts in Wayang show students the accuracy of their answers. b. Tips in Wayang encourage good learning habits
Wayang Outpost: Affective Learning Companions

Fig. 14 Animated pedagogical agents display a range of emotions. Companions act out their emotion and resolve negative ones, expressing full sentences of affective and metacognitive nature, to support growth of mindset towards the view that intelligence is a state (and thus changeable)
Effort Based Tutoring

Table 1 The effort-based tutoring algorithm informs pedagogical moves and affective decisions (last two columns) for each student on each problem. The algorithm first infers a reason for students’ behavior (fourth column) based on the number of incorrect student answers, hints requested and the amount of time spent (first three columns). Then the algorithm decides which pedagogical action the tutor should take (last two columns). The algorithm encourages transfer of student knowledge to subsequent questions of similar difficulty (rows 2, 4, 9), encouraging students to transfer skills and “fade” their need for help.

<table>
<thead>
<tr>
<th>Observed behavior and inferred reason for this behavior</th>
<th>Pedagogical Model Moves Cognitive or Affective or Metacognitive</th>
</tr>
</thead>
<tbody>
<tr>
<td>Incorrect Hints Time</td>
<td>Most Likely Reason Decision Affective/Metacog. Decisions</td>
</tr>
<tr>
<td>1 $&lt; E (b) - \delta_{hl}$ $&lt; E (hi) - \delta_{ih}$ $&lt; E (ti) - \delta_{th}$</td>
<td>Mastery without effort Increase Problem Difficulty Show learning progress</td>
</tr>
<tr>
<td>2 $&lt; E (b) - \delta_{hl}$ $&lt; E (hi) - \delta_{ih}$ $&gt; E (ti) + \delta_{th}$</td>
<td>Mastery with high effort Maintain Problem Difficulty Affective feedback: Praise Effort</td>
</tr>
<tr>
<td>3 $&lt; E (b) - \delta_{hl}$ $&gt; E (hi) + \delta_{ih}$ $&lt; E (ti) - \delta_{th}$</td>
<td>Hint abuse, low effort Reduce Problem Difficulty Deemphasize importance of immediate success</td>
</tr>
<tr>
<td>4 $&lt; E (b) - \delta_{hl}$ $&gt; E (hi) + \delta_{ih}$ $&gt; E (ti) + \delta_{th}$</td>
<td>Towards mastery, effort Maintain Problem Difficulty Praise effort</td>
</tr>
<tr>
<td>5 $&gt; E (hi) + \delta_{ih}$ $&lt; E (hi) - \delta_{ih}$ $&gt; E (ti) + \delta_{th}$</td>
<td>Quick guessing, low effort Reduce Problem Difficulty Deemphasize importance of immediate success</td>
</tr>
<tr>
<td>6 $&gt; E (hi) + \delta_{ih}$ $&lt; E (hi) - \delta_{ih}$ $&gt; E (ti) + \delta_{th}$</td>
<td>Hint avoidance and high effort Reduce Problem Difficulty Offer hints upon incorrect answer in the next problem</td>
</tr>
<tr>
<td>7 $&gt; E (hi) + \delta_{ih}$ $&gt; E (hi) + \delta_{ih}$ $&lt; E (ti) - \delta_{th}$</td>
<td>Quick guess and hint abuse Reduce Problem Difficulty Deemphasize importance of immediate success</td>
</tr>
<tr>
<td>8 $&gt; E (hi) + \delta_{ih}$ $&gt; E (hi) + \delta_{ih}$ $&gt; E (ti) + \delta_{th}$</td>
<td>Low mastery and High Effort Reduce Problem Difficulty Emphasize importance of effort and perseverance</td>
</tr>
<tr>
<td>9 Otherwise Expected Behavior Maintain Problem Difficulty</td>
<td></td>
</tr>
</tbody>
</table>

Note: Expected response (correct, hints, time) based on answers of other students $\sim$ collaborative filtering
Fig. 4 Massachusetts Statewide Standardized Test (MCAS) passing rates for experimental groups (using Wayang, dark grey) and control groups (in regular math class, light grey), within the same school, same grade and same teachers. Passing rates include several ratings above warning/failing.
Fig. 5 Area chart comparison of performance for a 7th grade of students on the Massachusetts Comprehensive Assessment System (MCAS), for students using vs. not using Wayang Outpost. Students represented by the yellow/green polygon used Wayang Outpost and students represented by the blue polygon did not use the tutor. Distribution of students using Wayang Outpost shifts to the right indicating that more students passed the exam and received a grade of “proficient” or “advanced” when using Wayang Outpost. Groups of students were matched in terms of teacher of seventh grade students.
Fig. 7 Mean improvement (and standard deviations) on hardest items of the math pre/posttest. The thick line represents students who received both the Wayang Tutor and math facts retrieval training software; all other groups did not really improve on these harder multi-step items.
Table 2  Students in the experimental group (last row) received tips and charts every 6 problems. Means and standard deviations in performance measures before and after tutoring for the three groups

<table>
<thead>
<tr>
<th>Group</th>
<th>Math Pretest</th>
<th>Math Posttest</th>
<th>Passing Rate in State Standard Exam</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Tutor Control</td>
<td>40 % (20) (N=40)</td>
<td>40 % (28)* (N=40)</td>
<td>76 % (N=38)</td>
</tr>
<tr>
<td>Tutor Control</td>
<td>33 % (19) (N=36)</td>
<td>42 % (22)* (N=36)</td>
<td>79 % (N=34)</td>
</tr>
<tr>
<td>ProgressTips Tutor</td>
<td>33 % (19) (N=36)</td>
<td>42 % (22)* (N=36)</td>
<td>92 % (N=24)</td>
</tr>
</tbody>
</table>

Fig. 12  High gaming students improve math performance when they receive progress tips and interventions (left) but not when they don’t receive interventions (right)
ALEF

- PeWe (Personalized Web) Group at UISI FIIT STU, Bratislava
- adaptive education (mainly) for programming exercises
ALEF: A Framework for Adaptive Web-Based Learning 2.0, Šimko, Barla, Bieliková
ALEF: A Framework for Adaptive Web-Based Learning 2.0, Šimko, Barla, Bieliková
ALEF: A Framework for Adaptive Web-Based Learning 2.0, Šimko, Barla, Bieliková
CourseRank

- recommendations of whole courses
- course evaluation and planning social system
- ranking of courses, grade distribution, other statistics
- originally Stanford, later many (US) universities, out of order now
- similar features e.g. in Coursera
- some attempts done in IS MU (but hard to practically apply in real university setting)
Summary

personalized education ↔ recommender systems

- many similarities
- specific challenges
- difficult evaluation