



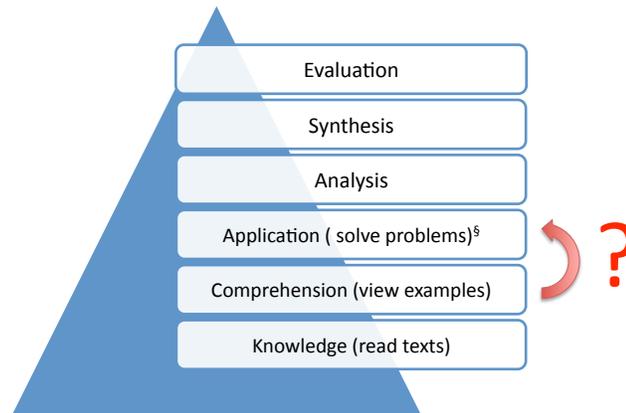
# Retrospective Evaluation of Blended User Modeling For Adaptive Educational Systems

Michael Yudelson, Sergey Sosnovsky  
University of Pittsburgh

## Retrospective User Model Evaluation

- Pros
  - No new user data needs to be collected
  - Verify/configure/compare user modeling approaches before deployment
- Cons
  - Actual, “in situ”, experimental deployment is necessary to finalize the evaluation
  - Not all aspects of user modeling mechanisms can be evaluated like this / not suitable for all UM approaches

## What Can Be Blended in a User Model



\* Bloom, B. S. (1956). Taxonomy of Educational Objectives, Handbook I: The Cognitive Domain. David McKay Co Inc., New York.

Michael V. Yudelson (C) 2009

3

## Agenda

- Modeling Users from Heterogeneous Sources
  - Without Blending
  - With Blending
- Experimental Results
- Discussion

Michael V. Yudelson (C) 2009

4

## User Modeling. Educational Domain

- Collection of learning objects (problems, examples, texts)
- Domain vocabulary (bag of words, taxonomy, ontology)
- LOs described in terms of the vocabulary
- User interaction with LOs update modeled user knowledge of concepts (overlay)

Michael V. Yudelson (C) 2009

5

## Modeling User Knowledge For Problem Solving. Example

- For every concept  $c$  of problem  $p$   
 $k_0=0$

result: 0 - error, 1 - correct

$$k_{n+1} = k_n + res \times (1 - k_n)^2 \times \begin{cases} k_n \leq .5 & w'/2 \\ k_n > .5 & w' \end{cases}$$

$$w' = \sqrt{\frac{w_{c,p}}{\sum_i w_{c,i} (succ_{att_p} + 1)}}$$

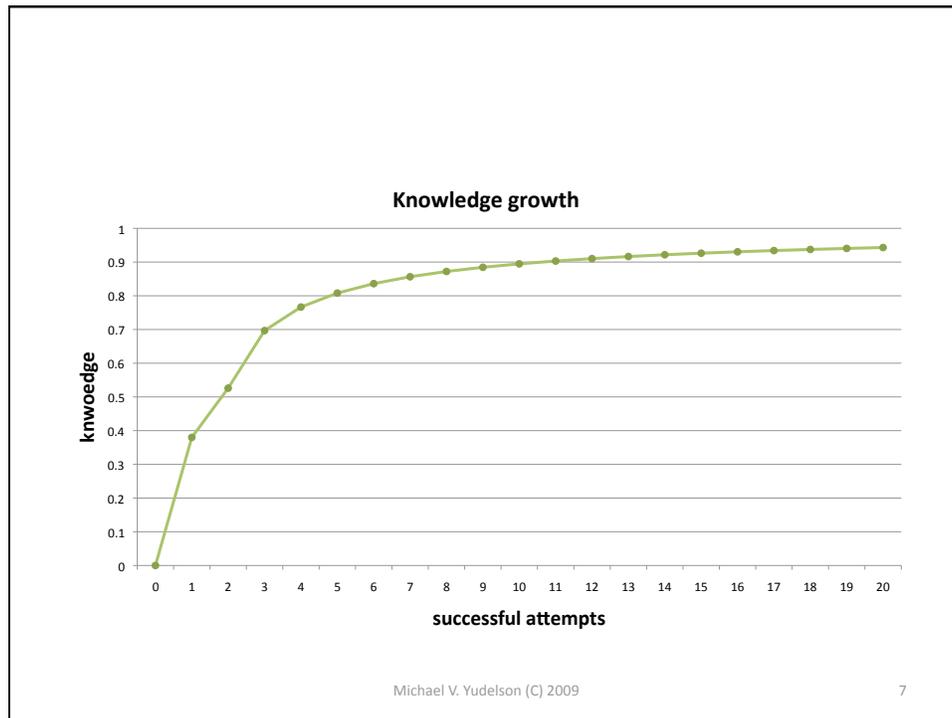
weight of concept  $c$  in problem  $p$

sum of all concept weights in problem  $p$

times problem was solved correctly

Michael V. Yudelson (C) 2009

6



- Problem solving knowledge is of a greater interest (can people apply their knowledge?)
- However, example browsing could also have a tangible impact (w.r.t. problem solving)
- Blend evidence of user problem solving and example browsing for the good of the UM

## Blended Modeling of User Knowledge

- For every concept  $c$  of problem (**example**)  $p$   
 $k_0=0$

$W = 1$  for problems  
 $W = [0, .1, .2, \dots, .8, .9, 1]$  for examples

$$k_{n+1} = k_n + W \times res \times (1 - k_n)^2 \times \begin{cases} k_n \leq .5 & w'/2 \\ k_n > .5 & w' \end{cases}$$

$$w' = \sqrt[4]{\frac{w_{c,p}}{\sum_i w_{c_i,p} (succ_{att_p} + 1)}}$$

Michael V. Yudelson (C) 2009

9

## Research Questions

- Using example browsing activity when modeling problem solving improves model accuracy?
- Different users benefit from different “blends”
- Find a single optimal blend for all users, OR
- Determine classes of users benefitting from different blends

Michael V. Yudelson (C) 2009

10

## Experimental Results



Michael V. Yudelson (C) 2009

11

## Experimental Setup

- Data collection
  - 4 database Courses (1 grad, 3 undergrad)
  - University of Pittsburgh: Fall 2007, Spring 2008
  - 114 students, 56 worked with our tools hard enough
  - 48 problems and 64 examples available
- Basic Per-Semester Statistics
  - 25-160 problem attempts per user (16-30 distinct problems attempted)
  - 45-260 clicks on examples per user (10-40 distinct examples explored)

Michael V. Yudelson (C) 2009

12

## Sample Problem

The screenshot shows a web browser window titled "Knowledge Tree - SELECT-FROM question1". The address bar shows the URL "http://adapt2.sis.pitt.edu/kt/content/Show?id=3532". The page content includes:

**Question:**  
Based on the tables below, write the required SQL expression.

**Task:**  
Show all the information contained in table "language".

A text input field contains the SQL query: `SELECT * FROM Language`. To the right, a message says "Well done! The answer is correct." with buttons for "Try a similar question" and "Go to SQL-Lab".

**Table NameSchema & Sample Data (click +/- to show/hide sample data)**

- `accident(+)` report\_number date location
- `actor(+)` actor\_id first\_name last\_name last\_update
- `address(+)` address\_id address district city\_id postal\_code phone last\_update
- `car(+)` license model year
- `category(+)` category\_id name last\_update
- `city(+)` city\_id city country\_id last\_update
- `language(+)` language\_id name last\_update

Michael V. Yudelson (C) 2009

13

## Sample Example

- Multiple Comparisons (2)
- `SELECT customer_no, borrower.loan_no`  
This query specify column 'loan\_no' should be why this column name is specified along with t 'loan' table have the column. If the table name the table where the column belongs to exclusiv
- `FROM borrower, loan`
- `WHERE borrower.loan_no = loan.`
- `'Perryridge';`

Michael V. Yudelson (C) 2009

14

## Experimental Procedures

- For each user
  - Compute 11 user models:
    - 0% of example activity impact problem-solving model (in terms of scale of each knowledge update)
    - 10%, 20%, ... 100%
  - 56 users \* 11 blends = 616 models
- For each model
  - Compute accuracy of predicting correct problem solution

Michael V. Yudelson (C) 2009

15

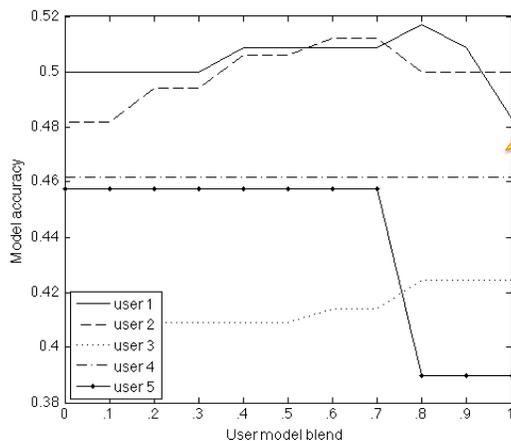
## Methods and Measures

- Accuracy (modeling problem-solving)
  - correct-predictions-of-outcomes/all-cases
- Existence of the blend preference
  - Check whether certain blends work for users best? (“best” blend), both individually and globally
- Classify users
  - Number of best blends

Michael V. Yudelson (C) 2009

16

## Blend vs. Accuracy



Different number of "best" blends

Michael V. Yudelson (C) 2009

17

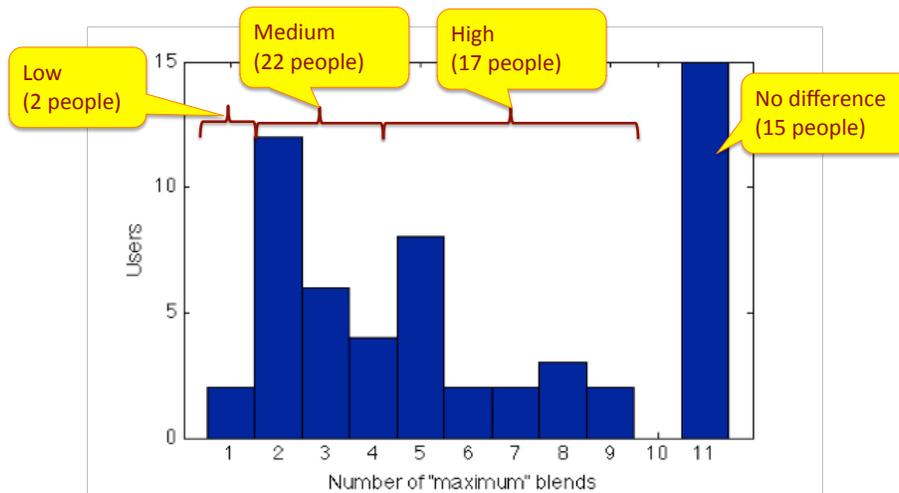
## Users and "Best" Blends

- Compare accuracy of 0% blend to the others
- Individual blend (0% vs. personal best)
  - Personal best is better: p-value=.001
- Community blend (0% vs. others)
  - 40% and 50% blends score significantly higher (p-value=.023)

Michael V. Yudelson (C) 2009

18

## Users and “Best” Blends (cont’d)



Michael V. Yudelson (C) 2009

19

## Users and “Best” Blends (cont’d)

- “Low” – 2 rare cases of just 1 blend preference  
– 80% and 100%
- “Medium” – inclined towards higher blends  
– 90% blend statistically superior to others  
– (p-value=.037)
- “High” – follow global trend (40,50% blends,  
p-value=.049)

Michael V. Yudelson (C) 2009

20

## Discussion

- Why accuracy is generally low?
  - Can we do something about it?
- Why the inter-group difference is so small?
  - Can we do something about that?
- The presence of adaptation potentially injects noise
- How about other approaches to user modeling?

Michael V. Yudelson (C) 2009

21

## Thank you!

- Questions?

Michael V. Yudelson (C) 2009

22