Chapter 9:
Learner modeling
Learning Design
Orchestration
Graphs
Activities

Learning Analytics
Statistical models
Probabilistic inference

You are here

CS-411: Digital Education & Learning Analytics
\[ p \ (\text{Trump} \mid \text{Snow}) = \ ? \]
If the first snow comes on a Tuesday…

something will happen somewhere someday
Learner Modelling

\[ 5^2 = ? \]
Learner Modelling

From the learner’s behaviour, infer his/her learner’s knowledge state

\[ 5^2 = 25 \quad \Rightarrow \quad \text{knows } X^2 \]

\[ 5^2 \neq 25 \quad \Rightarrow \quad \text{doesn’t know } X^2 \]
## Learner Modelling

<table>
<thead>
<tr>
<th>Behavior (Answer)</th>
<th>$5^2 = 5^2$</th>
<th>$5^n = ...$</th>
<th>$n^2 = n \cdot N$</th>
<th>$x^n = x \cdot x \cdot ...$</th>
<th>$x^n = x \cdot x \cdot ...$ bad mult.</th>
<th>$x^n = x \cdot n$</th>
<th>$x^n = x \cdot n$</th>
<th>$x^n = x + n$</th>
<th>$x^n = ???$</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>0.25</td>
<td>0.125</td>
<td>0.125</td>
<td>0.125</td>
<td>0.125</td>
<td>0.125</td>
<td>0.125</td>
<td>0.125</td>
<td>1</td>
</tr>
<tr>
<td>35</td>
<td>0.125</td>
<td>0.125</td>
<td>0.125</td>
<td>0.125</td>
<td>0.125</td>
<td>0.125</td>
<td>0.125</td>
<td>0.125</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>0.125</td>
<td>0.125</td>
<td>0.125</td>
<td>0.125</td>
<td>0.125</td>
<td>0.125</td>
<td>0.125</td>
<td>0.125</td>
<td>1</td>
</tr>
<tr>
<td>27</td>
<td>0.125</td>
<td>0.125</td>
<td>0.125</td>
<td>0.125</td>
<td>0.125</td>
<td>0.125</td>
<td>0.125</td>
<td>0.125</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>0.125</td>
<td>0.125</td>
<td>0.125</td>
<td>0.125</td>
<td>0.125</td>
<td>0.125</td>
<td>0.125</td>
<td>0.125</td>
<td>1</td>
</tr>
</tbody>
</table>
Learner Modelling

From the learner’s *behaviour*, infer his/her learner’s *knowledge state*

\[
p (\text{state} = \text{knows} \mid \text{correct-answer}) = 1 - \text{Guess} \]
\[
p (\text{state} = \text{knows} \mid \text{incorrect-answer}) = 0 + \text{Slip} \]

*Factors that depend upon the response modality*

*Bayesian Knowledge Tracing, Corbezt & Anderson*
Learner Modelling

From the learner’s behaviours, infer his/her learner’s knowledge state

$x_i(s) \in X_i(S) = \{\text{fine, active, lost, drop}\}$

State “fine”: the learner is performing well
State “active”: the learner is working but does not seem to succeed well
State “lost”: the learner does not understand at all or did not complete the activities
State “drop”: the learned has dropped out (e.g. no login since N days)
From the learner’s **behaviours**, infer his/her learner’s **knowledge state**

- $b(s) = $ watch video with many pauses
- $b(s) = $ post a message “There is a mistake on the slide” (and there is one indeed)
- $b(s) = $ select correct definition of SD in a quiz with 5 possible definitions

$X(S) = \{\text{lost, active, fine, brilliant}\}$

**Normalized entropy of the diagnosis vector**

- $x(s) = [0.15, 0.40, 0.30, 0.15]$  $H_0 = 0.94$
- $x(s) = [0.05, 0.15, 0.25, 0.55]$  $H_0 = 0.80$
- $x(s) = [0.01, 0.02, 0.02, 0.95]$  $H_0 = 0.18$
The uncertainty of the diagnosis can be estimated by Shannon’s entropy applied to the vector of probabilities for the different states.

Since this value depends upon the number of states, we normalize it on a 0->1 scale by dividing it by the maximal entropy which $\log_2$ of the number of states.

The **diagnosis power** of a question can be measured by the entropy of the diagnosis vector.

$$H(X) = -\sum_i P(x_i) \log_b P(x_i)$$

$x(s)=[.15 .40 .30 .15]$

$H_0=0.94$
Write a question that

• determines if the learner understood the concept of standard deviation;

• has a high diagnosis power

• can be automatically graded
Learner Modelling

From the learner’s behaviour, infer his/her learner’s knowledge state

1. The basic approach
2. The good old AI approach
3. The data crunching approach
4. The Bayesian approach
5. The Markov approach
(1) Basic approach to learner Modelling

Decrease uncertainty by collecting multiple answers

<table>
<thead>
<tr>
<th>$5^2 = ??$</th>
<th>$7^2 = ??$</th>
<th>Knowledge States</th>
</tr>
</thead>
<tbody>
<tr>
<td>Behavior (Answer)</td>
<td>$5^2 = 25$</td>
<td>$5^n = \ldots$</td>
</tr>
<tr>
<td>25</td>
<td>49</td>
<td>0.125</td>
</tr>
<tr>
<td>25</td>
<td>21</td>
<td>0.125</td>
</tr>
<tr>
<td>35</td>
<td>49</td>
<td>0.125</td>
</tr>
</tbody>
</table>

If the learner makes more than $n \%$ errors in $a_i$, then (s)he is in state « low understanding »
To compute the length of the hypothenuse

1. Measure the length, L1 and L2
2. Compute $L1^2$ and $L2^2$
3. Sum them
4. Extract the square root

Learner Answer = 8.18
Correct answer = 8.54

What did he do wrong?
If, when bringing perturbation X to an expert system, it produces the same mistake as the learner, X is a good hypothesis of what the learner did not understand.
(3) Learner Modelling with data sciences

From the learner’s behaviours, infer his/her learner’s knowledge state

Behaviours

- Text entered
  - Item selected (button, menu,…)
- Area clicked
- Line drawn with mouse
- Response time
- …
- …

Behavioural ‘Dust’

(fragments of behaviour that do not have an explicit semantic value)

- Number of pauses
- Mouse path
- Gaze path
- Facial expressions
- Gestures
- …

« Social Signal Processing »
Example 1: From the learner’s gaze, infer the « withmeness »

Eye tracking experiment on MOOC Video

Following teacher’s references
Gaze of students’ watching Scala course by Prof. Martin Odersky (EPFL, Switzerland)

K. Sharma, P. Jermann, P. Dillenbourg
@ CHILI – http://chili.epfl.ch
Supported by the Swiss National Science Foundation
(Grants CR1211_132996 and PZ00P2_126611)
Example 1: From the learner’s gaze, infer the « withmeness » because it predicts learning gains

Kshitij Sharma, Patrick Jermann, Pierre Dillenbourg
EPFL Center for Digital Education
Example 2: From 2 learners gazes, infer the quality of collaboration

DUET - Dual Eye-Tracking
Pair programming experiment

Low gaze recurrence

High gaze recurrence

P. Jermann, M.-A. Nässli & P. Dillenbourg

Supported by the Swiss National Science Foundation
(grants #K-12K1-117909 and #PZ00P_126611)

P. Jermann, M.-A. Nässli & P. Dillenbourg

Supported by the Swiss National Science Foundation
(grants #K-12K1-117909 and #PZ00P_126611)
Example 2: From 2 learners gazes, infer the quality of collaboration.

The pairs that collaborate well tend to look ± at the same time at ± the same object.
Next Week

08:15 - 10:00  Eye tracking methods
               Kshitij Sharma

10:15 – 12:00 Learning analytics
            Try an eye tracker
Learner Modelling

From the learner’s behaviour, infer his/her learner’s knowledge state
**Example 3:** From the learner’s (co)movements, infer the class level of attention
<table>
<thead>
<tr>
<th>Kernel</th>
<th>Features</th>
<th>Score</th>
<th>Cohen’s kappa</th>
</tr>
</thead>
<tbody>
<tr>
<td>RBF(c=1.31, g=0.0211)</td>
<td>Distance, Head travel norm., Num. still periods</td>
<td>61.86%</td>
<td>0.30</td>
</tr>
<tr>
<td>RBF(c=1.21, g=0.11)</td>
<td>Period, Row, Head travel norm., Mean duration still</td>
<td>61.72%</td>
<td>0.32</td>
</tr>
<tr>
<td>RBF(c=1.11, g=0.061)</td>
<td>Head travel norm., Mean duration still</td>
<td>60.42%</td>
<td>0.28</td>
</tr>
<tr>
<td>RBF(c=1.4, g=0.04)</td>
<td>Period, Distance, Row, Mean duration still</td>
<td>59.23%</td>
<td>0.30</td>
</tr>
<tr>
<td>Behaviours</td>
<td>Behavioural ‘Dust’</td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------------------------------------------------------------------------</td>
<td>----------------------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1  Individual Plane</td>
<td>Video ‘Withmeness’</td>
<td></td>
<td></td>
</tr>
<tr>
<td>The learner answer to a quizz</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2  Team Plane</td>
<td>Gaze Recurrence</td>
<td></td>
<td></td>
</tr>
<tr>
<td>The concept map produced by a pair</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3  Class Plane</td>
<td>Head Co-Rotation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>The # messages in the forum</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
State A  State B

Classifier

\[ \Pi_1 \]
- Rate of back/cancel actions in a navigation task.
- Redundancy: Did the learner ask a question for which he already had an answer?
- “With-me-ness”: Did the learner look at the object mentioned by the lecturer in the video?
- Attention map: Which areas does the learner look at most often?
- .....
State A

State B

Classifier

Rich Data Set

Features

\( \Pi_2 \)

- **Balance** of participation: Did all team members do a fair share of the workload?
- **Task-distribution**: Do team members perform specific subsets of the tasks?
- **Rate of acknowledgement**: What percentage of utterances from a learner received acknowledgement—from a simple nod to an acknowledging action?
- **Transactivity**: Did team members build utterances upon the utterances produced by their peers?
- **Cross-recurrence**: Did team members look at the same object at (more or less) the same time?
- **Rate of redundancy**: Did the learner ask a question for which another team member already had the answer?
• **Conversation depth**: The average depth of conversation threads in forums.
• **Connectivity**: What is the minimal number of students that need to be removed from the social network to disconnect the other nodes from each other (Diestel, 2005)?
• **Homophily**: Do students form ties with similar versus dissimilar students? Ties can be forums postings; similarity is measured through students’ profiles.
• **Reciprocity**: If student A often replies to another student B in the forum, is the opposite true?
• **Propinquity**: The tendency for actors to have more ties with those who are geographically close (Kadushin, 2012).
• **Density**: The proportion of direct interactions between two students relative to the total number of possible interactions between all students (Xu et al., 2010).
Activity $a_5$. In order to reduce the variance of the set $[1 2 2 3 3 3 4 5 8]$, 3 numbers can be removed. Which ones?

a) Remove all occurrences of number 3
b) Remove the numbers that appear several times  
c) Remove 1, 5, and 8  
d) Remove 4, 5, and 8

$X_5(S) = \{\text{misunderstanding, good understanding}\}$

if $b_5(s)=c$, then $x_5(s)=g$
if $b_5(s)=a$, then $x_5(s)=m$

$P(x_5(s)=g | b_5(s)=c) = 1$
$P(x_5(s)=g | b_5(s)=a) = 0$
Activity a₅. In order to reduce the variance of the set [1 2 2 3 3 3 4 5 8], 3 numbers can be removed. Which ones?

a) Remove all occurrences of number 3
b) Remove the numbers that appear several times
  c) Remove 1, 5, and 8
  d) Remove 4, 5, and 8

\[ X₅(S) = \{\text{misunderstanding}, \text{good understanding}\} \]

if \( b₅(s)=c \), then \( x₅(s)=g \)

if \( b₅(s)=a \), then \( x₅(s)=m \)

\[
\begin{align*}
P \left( x₅(s)=g \mid b₅(s)=c \right) &= 75\% \text{ (he had 25\% to succeed by chance)} \\
P \left( x₅(s)=g \mid b₅(s)=a \right) &\approx 10\% \text{ (e.g. typing mistake)}
\end{align*}
\]
X₅(S) = {misunderstanding, good understanding}

P (x₅(s)=g | b₅(s)=c) = 75% (he had 25% to succeed by chance)

But if one knows a priori that this a difficult concept, e.g. that only 20% of students are usually in state « good understanding », one may apply Bayes Theorem

\[
P(A|B) = \frac{P(A)P(B|A)}{P(B)}
\]

\[
P (x₅(s)=g | b₅(s)=c) = \frac{P (x₅(s)=g) \cdot P (b₅(s)=c | x₅(s)=g)}{P (b₅(s)=c | x₅(s)=g) \cdot P (x₅(s)=g) + P (b₅(s)=c | x₅(s)≠g) \cdot P (x₅(s)≠g)}
\]

P (x₅(s)=g | b₅(s)=c) = 0.47
(4) Learner Modelling @ BayesianTimes

P (x_5(s)=g \mid b_5(s)=c) = 0.47

P (x_5(s)=m \mid b_5(s)=c) = 0.53

b_5(s) = c \Rightarrow x_5(s) = [0.47 \ 0.53]

The diagnosis power of this question is not great, close to 50/50. Entropy is very high!
(5) Learner Modelling @ MarkovTimes

Inferring the learner’s state from his previous state
The weight of edges
## State Transition Matrix

### Activity 2

<table>
<thead>
<tr>
<th></th>
<th>Lost</th>
<th>Active</th>
<th>Fine</th>
<th>Great</th>
<th>H</th>
<th>H0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lost</td>
<td>0.26</td>
<td>0.39</td>
<td>0.29</td>
<td>0.06</td>
<td>1.80</td>
<td>0.90</td>
</tr>
<tr>
<td>Active</td>
<td>0.19</td>
<td>0.34</td>
<td>0.26</td>
<td>0.21</td>
<td>1.96</td>
<td>0.98</td>
</tr>
<tr>
<td>Fine</td>
<td>0.11</td>
<td>0.28</td>
<td>0.45</td>
<td>0.16</td>
<td>1.81</td>
<td>0.90</td>
</tr>
<tr>
<td>Great</td>
<td>0.05</td>
<td>0.15</td>
<td>0.25</td>
<td>0.55</td>
<td>1.60</td>
<td>0.80</td>
</tr>
</tbody>
</table>

\[ H(X) = -\sum_i P(x_i) \log_b P(x_i) \]

\[ 1\cdot H0 = 0.10 \]
<table>
<thead>
<tr>
<th></th>
<th>M6</th>
<th>Lost</th>
<th>Active</th>
<th>Fine</th>
<th>H</th>
<th>H0</th>
<th></th>
<th>M7</th>
<th>Lost</th>
<th>Active</th>
<th>Fine</th>
<th>H</th>
<th>H0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lost</td>
<td></td>
<td>0.01</td>
<td>0.24</td>
<td>0.75</td>
<td>0.87</td>
<td>0.55</td>
<td></td>
<td>Lost</td>
<td>0.75</td>
<td>0.24</td>
<td>0.01</td>
<td>0.87</td>
<td>0.55</td>
</tr>
<tr>
<td>Active</td>
<td></td>
<td>0.01</td>
<td>0.24</td>
<td>0.75</td>
<td>0.87</td>
<td>0.55</td>
<td></td>
<td>Active</td>
<td>0.75</td>
<td>0.24</td>
<td>0.01</td>
<td>0.87</td>
<td>0.55</td>
</tr>
<tr>
<td>Fine</td>
<td></td>
<td>0.01</td>
<td>0.24</td>
<td>0.75</td>
<td>0.87</td>
<td>0.55</td>
<td></td>
<td>Fine</td>
<td>0.75</td>
<td>0.24</td>
<td>0.01</td>
<td>0.87</td>
<td>0.55</td>
</tr>
</tbody>
</table>

\[ \omega(M5) = 0.45 \]

\[ \omega(M6) = 0.45 \]
### State Transition Matrix: Utopy

<table>
<thead>
<tr>
<th></th>
<th>M8</th>
<th>M9</th>
<th>M10</th>
<th>M11</th>
<th>M12</th>
<th>M13</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
</tr>
<tr>
<td></td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
</tr>
<tr>
<td></td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
</tr>
<tr>
<td></td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
</tr>
<tr>
<td></td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>U(M)</th>
<th></th>
<th>U(M)</th>
<th></th>
<th>U(M)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>M8 1</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
</tr>
<tr>
<td>M9 1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.2</td>
<td>0.3</td>
<td>0.3</td>
<td></td>
</tr>
<tr>
<td>M10 1</td>
<td>0.5</td>
<td>0.1</td>
<td>0.2</td>
<td>0.1</td>
<td>0.1</td>
<td></td>
</tr>
<tr>
<td>M11 1</td>
<td>1.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>M12 1</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
<td></td>
</tr>
<tr>
<td>M13 1</td>
<td>0.8</td>
<td>0.2</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td></td>
</tr>
</tbody>
</table>

\[
\gamma(M) = \frac{1}{m(m-1)} \sum_{k=1}^{m-1} \sum_{l=k+1}^{m} (l-k) \alpha(l) - \sum_{k=2}^{m} (k-l) \alpha(l)
\]

\[
U(M) = 0.47
\]

\[
U(M) = 0.47
\]

\[
U(M) = -0.42
\]
For the exam, I don’t ask you to learn home-made formulas but to understand the principles. Formulas could be replaced by visualisations.
The strength relationship between activities will often degrade with time, e.g. even if $a_1$ is strong pre-requisite to $a_2$, the knowledge gained in $a_1$ won't remain activated for ever.
So far we treated them separately, but one may infer the learner’s state from **both** his behaviour **and** his previous state.
One step further: one may infer the learner’s state from his behaviour (depth), his previous state (horizontally) and the state of others (vertically)
The learning analytics cube: 3 axes of inference

A. John does probably not understand SD deviation because he removed the central values of the distribution

B. John does probably not understand SD deviation because he did not understand what is a mean and the mean is a pre-requisite

C. John does probably not understand SD deviation because most learners in that class failed and John is one of the weakest
This cube may allow second-level inferences
This cube may allow second-level inferences
This cube may allow second-level inferences
So far we use common sense to describe the learner state

\[ x_i(s) \in X_i(S) = \{\text{fine, active, lost, drop}\} \]

but educational research defined is much richer set of states
State ≠ Trait

- Anxious / Self-confident
- Risk-aversive / Risk-seeking
- Aural / visual / kinesthetic
- Deep / Surface
- Field-dependent / independent

Severe criticisms:
- Contextual rather than personal
- No clear effects of adaptation
- Should education mimic style or counterbalance them?
- Labels produce self-fulfilling prophecies
BEWARE OF
the medicalisation of Education !!!

- Learning disabilities, LD
- Attention-deficit disorder, ADD
- Attention-deficit hyperactivity disorder, ADHD
- Non-verbal learning disability, NVLD
- ...
- High-potential children
- ....

Labels help Sales