15-780 - GRADUATE ARTIFICIAL INTELLIGENCE
AI AND EDUCATION I

Shayan Doroudi

April 24, 2017

Series on applications of AI to education.

Lecture	Application
4/24/17	Learning
4/26/17	Assessment
5/01/17	Instruction

Series on applications of AI to education.

Lecture	Application	AI Topics
4/24/17	Learning	Machine Learning + Search
4/26/17	Assessment	Machine Learning + Mechanism Design
5/01/17	Instruction	Multi-Armed Bandits

Series on applications of AI to education.

Lecture	Application	AI Topics
This Time	Learning	Machine Learning + Search
4/26/17	Assessment	Machine Learning + Mechanism Design
5/01/17	Instruction	Multi-Armed Bandits

- 1956: Dartmouth Workshop on AI.

> *The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it. An attempt will be made to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves. We think that a significant advance can be made in one or more of these problems if a carefully selected group of scientists work on it together for a summer.*

HISTORY OF AI AND EDUCATION AT CMU

- 1956: Dartmouth Workshop on AI.
- Herb Simon and Alan Newell continue this line of work for the rest of their lives. Newell develops SOAR model of human cognition.

- 1956: Dartmouth Workshop on AI.
- Herb Simon and Alan Newell continue this line of work for the rest of their lives. Newell develops SOAR model of human cognition.
- John Anderson joins CMU in 1978. Develops ACT-R theory of human cognition.

- 1956: Dartmouth Workshop on AI.
- Herb Simon and Alan Newell continue this line of work for the rest of their lives. Newell develops SOAR model of human cognition.
- John Anderson joins CMU in 1978. Develops ACT-R theory of human cognition.
- John Anderson and Albert Corbett develop LISPITS in 1983.

- 1956: Dartmouth Workshop on AI.
- Herb Simon and Alan Newell continue this line of work for the rest of their lives. Newell develops SOAR model of human cognition.
- John Anderson joins CMU in 1978. Develops ACT-R theory of human cognition.
- John Anderson and Albert Corbett develop LISPITS in 1983.
- Carnegie Learning founded in 1998 (including co-founders John Anderson and Ken Koedinger), which has taught math to over half a million students.

HISTORY OF AI AND EDUCATION AT CMU

What We Know About Learning*

HERBERT A. SIMON
Department of Psychology
Carnegie Mellon University

ABSTRACT

by copying their professors' lectures. In spite of the invention of printing not too long thereafter, students still continued to behave in their classes as copyists — assiduously taking notes, recording the deathless words of professors as if they didn't know printing had been invented and was available. I have heard that there are some universities where this happens even today.

Situated Learning and Education[1]

JOHN R. ANDERSON LYNNE M. REDER HERBERT A. SIMON

1987

Learning mathematics from examples and by doing

Xinming Zhu
Carnegie Mellon University

Herbert Alexander Simon

Artificial Intelligence and Psychology Project

Applications and Misapplications of Cognitive Psychology to Mathematics Education

John R. Anderson
Lynne M. Reder
Herbert A. Simon

Radical Constructivism and Cognitive Psychology

JOHN R. ANDERSON, LYNNE M. REDER, *and* HERBERT A. SIMON

THOSE WHO BELIEVE that education needs a foundation in the modern science of cognitive psychology sometimes feel that they are jousting with windmills. Virtually every educational movement, whatever its merits, claims to have a scientific basis. However, this is often not the case.

APPLICATIONS OF AI TO LEARNING

POWER LAW OF PRACTICE

- Power Law: $P = aT^b$
  - P = performance (error rate, reaction time)
  - T = number of trials/opportunities
  - a, b constants
- Log-log form: $\log P = b \log(T) + \log(a)$

(Content of these slides taken and modified from Ken Koedinger's slides

www.learnlab.org/opportunities/summer/presentations/2012/2.Learning-curves2.ppt)

- Newell and Rosenbloom (1981) tested fits of various models to learning curves and gave explanation for power law of practice.

POWER LAW OF PRACTICE

Data Set	Exponential $T = A + Be^{-\alpha N}$				Hyperbolic $T = A + B/(N + E)$				Power Law $T = A + B(N + E)^{-\alpha}$				
	A	B	α	r^2	A	B	E	r^2	A	B	E	α	r^2
Snoddy (1926)	27.01	38.80	.061	.916	24.49	243.6	1.3	.962	21.74	119.2	0.0	.71	.975
Crossman (1959)	7.19	4.59	3.1×10^{-7}	.842	7.10	2.4×10^{6}	151000	.983	6.91	20481	31000	.66	.990
Kolers (1975) - Subject HA	1.36	3.82	.018	.849	1.10	94.02	9.8	.915	.18	15.25	0.0	.46	.931
Neisser et al. (1963)													
Ten targets	.06	.83	.13	.905	.00	2.74	.9	.965	.00	2.35	.6	.95	.965
One target	.06	.44	.094	.938	.00	3.16	4.6	.951	.00	2.57	3.9	.94	.951
Card, English & Burr (1978)													
Stepping keys - Subj. 14	2.35	1.99	.011	.335	2.14	171.4	75.2	.338	.02	6.36	9.3	.14	.340
Mouse - Subj. 14	1.46	1.28	.028	.452	1.46	16.70	5.0	.603	.59	4.28	0.0	.33	.729
Seibel (1963) - Subject JK	.371	.461	.000055	.956	.328	3888.1	3042	.993	.324	2439.9	2690	.95	.993
Anderson (Note 1) - Fan 1	.487	.283	.00055	.774	.466	231.6	319.7	.902	.353	4.322	0.0	.39	.947
Moran (1980)													
Total time	13.80	6.66	.00073	.546	14.77	3335.9	474.6	.637	.03	30.24	0.0	.08	.839
Method time	11.61	3.11	.0010	.652	11.75	1381.8	360.0	.737	.26	19.35	0.0	.06	.882
Neves & Anderson (1980)													
Total time - Subject D	57.5	240.2	.019	.660	45.6	5000.2	7.3	.728	0.0	991.2	0.0	.51	.780
The Game of Stair													
Won games	476	319	.0052	.689	449	29800	40.1	.783	120	1763	0.0	.25	.849
Lost Games	152	326	.0016	.634	247	41270	124.1	.751	1	1009	2.5	.19	.841
Hirsch (1952)	2.76	4.35	.070	.819	2.34	37.05	4.9	.897	.00	10.01	0.0	.32	.932
General Power Law $T = 5 + 75(N + 25)^{-0.5}$	7.21	6.78	.0037	.983	6.41	1069.6	91.2	.997	5.00	74.85	24.9	.50	1.000
40 Term Additive Mixture	1.60	45.37	.0065	.904	.58	1231.2	10.2	.997	.19	753.1	7.2	.89	.998
Chunking Model Combinatorial TE	4.61	4.71	.0046	.957	4.35	365.7	55.3	.992	2.86	17.40	6.6	.33	1.000

Table 2: The General Learning Curves: Parameters from Optimal Fits in the *Log* Transformation Spaces

Newell, A., & Rosenbloom, P. S. (1981). Mechanisms of skill acquisition and the law of practice. Cognitive skills and their acquisition, 1, 1-55.

POWER LAW OF PRACTICE

- Newell and Rosenbloom (1981) tested fits of various models to learning curves and gave explanation for power law of practice.
- Heathcote, Brown, and Mewhort (2000) give alternative explanation:
    - Each student's practice is better fit by an exponential curve
    - Aggregation of them fit a power law curve

How can we apply learning curves to model a student's learning in an intelligent tutoring system?

How can we apply learning curves to model a student's learning in an intelligent tutoring system?

- There may be individual differences in students.

How can we apply learning curves to model a student's learning in an intelligent tutoring system?

- There may be individual differences in students.
- Students learn different skills at different rates.

additive factors model (afm)

How can we apply learning curves to model a student's learning in an intelligent tutoring system?

- There may be individual differences in students.
- Students learn different skills at different rates.
- Different problems may share some of the same skills.

How can we apply learning curves to model a student's learning in an intelligent tutoring system?

- There may be individual differences in students.
  (θ_i: Ability of student i)
- Students learn different skills at different rates.

- Different problems may share some of the same skills.

How can we apply learning curves to model a student's learning in an intelligent tutoring system?

- There may be individual differences in students. (θ_i: Ability of student i)
- Students learn different skills at different rates. (β_k: learning rate of skill k)
- Different problems may share some of the same skills.

How can we apply learning curves to model a student's learning in an intelligent tutoring system?

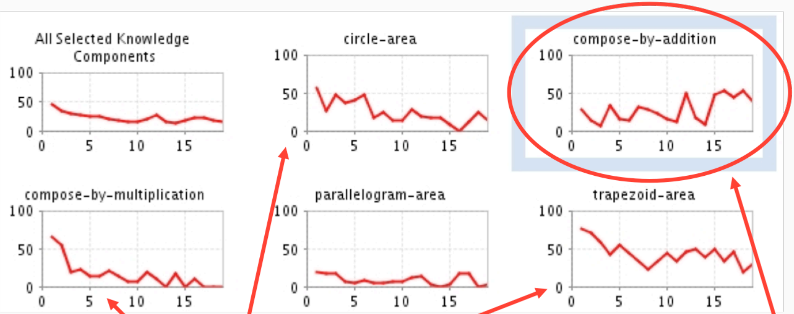
- There may be individual differences in students. (θ_i: Ability of student i)
- Students learn different skills at different rates. (β_k: learning rate of skill k)
- Different problems may share some of the same skills. (Q matrix: maps problems to skills)

	Skills			
Items	Add	Sub	Mul	Div
a*b	0	0	0	1
a*b + c	1	0	1	0
a*b - c	0	1	1	0
c + a*b	1	0	1	0

- $p_{ij,T}$: Probability that student i answers question j correctly at opportunity T.

- $p_{ij,T}$: Probability that student i answers question j correctly at opportunity T.
- AFM: $\log\left(\frac{p_{ij,T+1}}{1-p_{ij,T+1}}\right) = \theta_i + \sum_k Q_{jk}(\beta_k + \gamma_k T)$

- $p_{ij,T}$: Probability that student i answers question j correctly at opportunity T.
- AFM: $\log\left(\frac{p_{ij,T+1}}{1-p_{ij,T+1}}\right) = \theta_i + \sum_k Q_{jk}(\beta_k + \gamma_k T)$
- Poll: Which of the following is true about this model?
    - It is a linear regression model.
    - It is a logistic regression model.
    - It follows a power law of practice for $P = \log\left(\frac{p_{ij,T+1}}{1-p_{ij,T+1}}\right)$.
    - It follows an exponential law of practice for $P = \log\left(\frac{p_{ij,T+1}}{1-p_{ij,T+1}}\right)$.



Many curves show a reasonable decline

Some do not => Opportunity to improve model!

LEARNING FACTORS ANALYSIS (LFA)

- Method for automatically improving a cognitive model.

- Method for automatically improving a cognitive model.
- Inputs: a cognitive model (Q matrix), a model with hypothesized new skills (P matrix), and student log data.

- Method for automatically improving a cognitive model.
- Inputs: a cognitive model (Q matrix), a model with hypothesized new skills (P matrix), and student log data.
- Outputs: Cognitive models that fit the data best along with parameter estimates and model fits for those models.

Q Matrix

Items	Add	Sub	Mul	Div
	Skills			
a*b	0	0	1	0
a*b + c	1	0	1	0
a*b - c	0	1	1	0
c + a*b	1	0	1	0

P Matrix

Items	Multi-Step	Order of Ops
	Skills	
a*b	0	0
a*b + c	1	0
a*b - c	1	0
c + a*b	1	1

We refine our Q matrix by *adding* and/or *splitting* skills.

We refine our Q matrix by *adding* and/or *splitting* skills.

New Q Matrix

	Skills				
Items	Add	Sub	Mul	Div	Multi-Step
a*b	0	0	1	0	0
a*b + c	1	0	1	0	1
a*b - c	0	1	1	0	1
c + a*b	1	0	1	0	1

We refine our Q matrix by *adding* and/or *splitting* skills.

New Q Matrix

	Skills					
Items	Add	Sub	Mul-First	Mul-Second	Div	Multi-Step
a*b	0	0	1	0	0	0
a*b + c	1	0	1	0	0	1
a*b - c	0	1	1	0	0	1
c + a*b	1	0	0	1	0	1

1. Start with original Q matrix.
2. Apply all possible add and split operations using P matrix, evaluate model fit for each model, and add models to frontier.
3. Remove model from frontier with best fit, make that the new Q matrix.
4. Go back to step 2.

1. Start with original Q matrix.
2. Apply all possible add and split operations using P matrix, evaluate model fit for each model, and add models to frontier.
3. Remove model from frontier with best fit, make that the new Q matrix.
4. Go back to step 2.

What is the goal node?

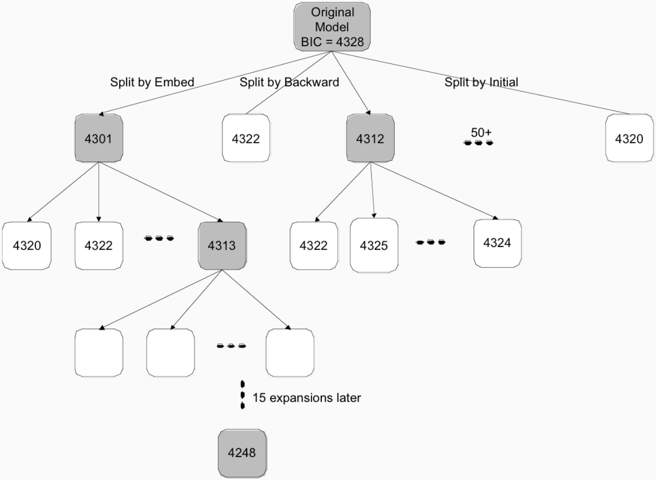
- Log likelihood $l(\theta)$?

- Akaike Information Criterion (AIC): $2k - 2l(\theta)$, where k is number of parameters.

- Akaike Information Criterion (AIC): $2k - 2l(\theta)$, where k is number of parameters.
- Bayesian Information Criterion (BIC): $Nk - 2l(\theta)$, where N is number of observations.

- Akaike Information Criterion (AIC): $2k - 2l(\theta)$, where k is number of parameters.
- Bayesian Information Criterion (BIC): $Nk - 2l(\theta)$, where N is number of observations.
- Cross-Validated Root Mean Squared Error
    - Ideal, but takes a lot longer to compute.

LEARNING FACTORS ANALYSIS (LFA)



Cen, H., Koedinger, K., Junker, B. (2006). Learning Factors Analysis: A general method for cognitive model evaluation and improvement. 8th International Conference on Intelligent Tutoring Systems.

LFA implements which of the following search algorithms?

- Uniform Cost Search
- Greedy (Best-First) Search
- A* Search
- None of the above
- Beats me

- Central advances in AI and cognitive psychology co-developed at CMU and have led to a rich history of research on AI and education.

SUMMARY

- Central advances in AI and cognitive psychology co-developed at CMU and have led to a rich history of research on AI and education.
- A combination of cognitive science/domain knowledge and machine learning can be used to model student learning.

SUMMARY

- Central advances in AI and cognitive psychology co-developed at CMU and have led to a rich history of research on AI and education.
- A combination of cognitive science/domain knowledge and machine learning can be used to model student learning.
- A combination of cognitive science/domain knowledge and AI can be used to automatically refine cognitive models.

- Central advances in AI and cognitive psychology co-developed at CMU and have led to a rich history of research on AI and education.
- A combination of cognitive science/domain knowledge and machine learning can be used to model student learning.
- A combination of cognitive science/domain knowledge and AI can be used to automatically refine cognitive models.
- Next time: how statistics/machine learning and AI has been used to model and improve assessment of student knowledge.