Modeling Student Strategy Usage with Mixed Membership Models

April Galyardt
University of Georgia

NIPS 2012 Workshop
Personalizing Education with Machine Learning
December 8, 2012

Example: Addition

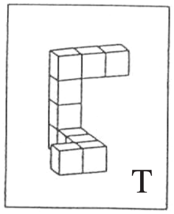
- Addition Strategies
  - **Retrieval** or Memorization
  - **Count-on:** to solve *7+2*, the child counts *7,8,9*
  - **Count-all:** to solve *7+2*, the child counts *1,2,3,4,5,6,7,8,9*
- Strategies differ in solution time, and accuracy
- Children switch between these strategies.
  - 99% of students use more than one strategy.
  - The mixture of strategies is different for different grade levels.

(Siegler, 1987)

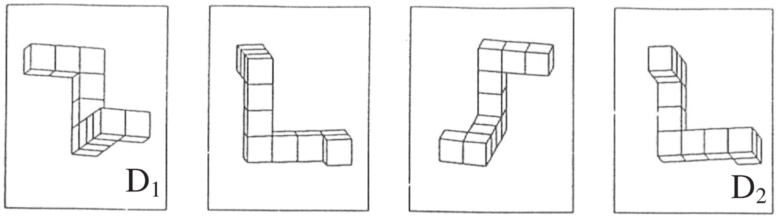
Example: Mental Rotation

Targets

Identify ALL solutions

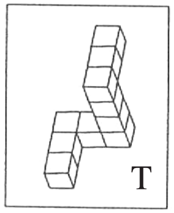


(a)

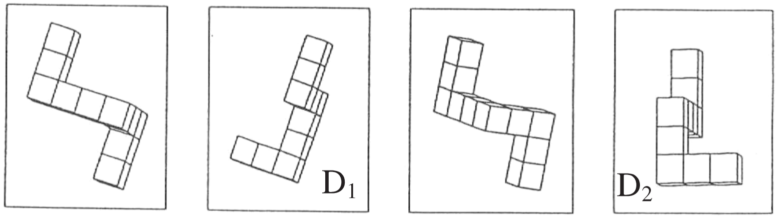


Mental rotation ✔

Analytic strategy ✘



(b)



Mental rotation ✔

Analytic strategy ✔

(Geiser, et al. 2006)

Example: Least Common Multiples

Problem	Correct Strategy	Multiplicative Strategy
{4,5}	4×5 = 20	4×5 = 20
{4,6}	2×2×3 = 12	❌ 4×6 = 24

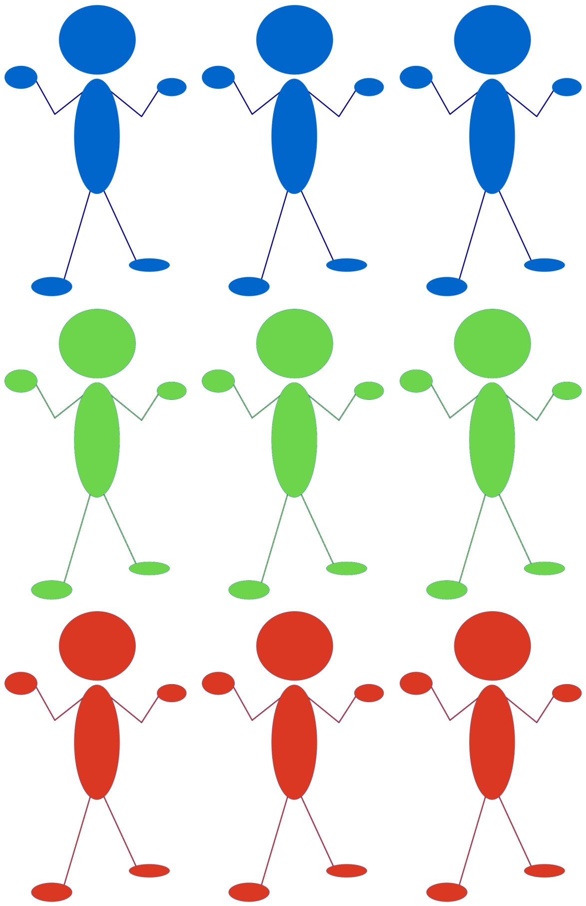
(Pavlik et al., 2011)

The Problem of Multiple Strategy Usage

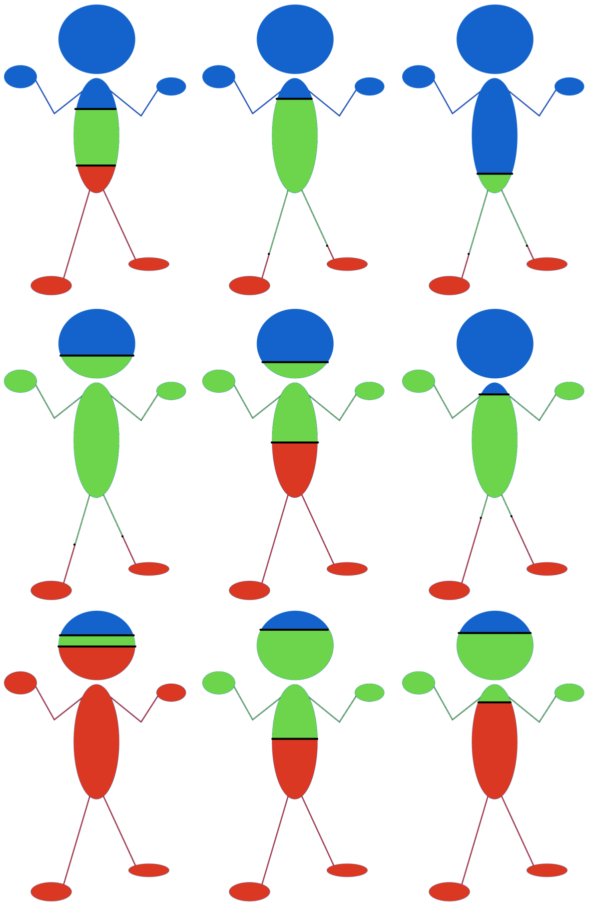
- Children switch strategies on even the simplest tasks. (Siegler, 1987)
- As students gain expertise, the mixture of strategies they use changes. (National Research Council, 2001)
- Four levels for psychometric modeling of multiple strategy usage. (National Research Council, 2001)
    1. No modeling of strategies
    2. Different people use different strategies.
    3. Individuals use different strategies from task to task.
    4. Individuals use different strategies within a task.

Mixed Membership Models

Latent Class Models



Mixed Membership



Mixed Membership Multiple Strategies Model

- Data for person i on item j includes any measured variable:

$$X_{ij} = (\underset{\substack{\nearrow \\ \text{accuracy}}}{C_{ij}}, \underset{\substack{\nwarrow \\ \text{time}}}{T_{ij}}, \ldots)$$

- Each strategy profile k defines a factorable distribution for these variables, a ***process signature***:

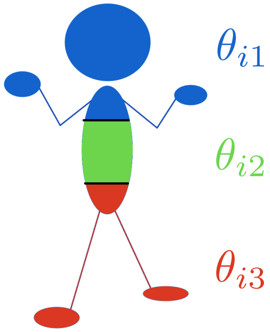
$$F_{kj}(X_{ij}) = F_{kj}(C_{ij}) \times F_{kj}(T_{ij}) \times \ldots$$

- Underlying Mixed Membership model allows for strategy switching.

Generative Model Definition

- For each individual *i*, draw a membership vector.

$$\theta_i \sim D(\theta)$$



1. For each item j: draw a strategy

$$Z_{ij} \sim \text{Multinomial}(\theta_i)$$

2. Draw the observed data X_{ij} from the strategy profile distribution.

$$X_{ij}|Z_{ij} = k \ \sim \ F_{kj}(x)$$

Problem!

- Mixed membership models are really complicated.
    - Typical data sets are 10K subjects and 100-1000 observations per subject.
- Educational data sets are comparatively tiny.
    - 100s of subjects and 10s of observations per subject.
- Is this mixed membership strategy idea even feasible?

Least Common Multiples Data

- Computer based assessment of Least Common Multiples
- N = 255 students
- J = 24 items total
  - Students were randomly assigned 16 items
  - 58 students received only 8 items
- Data for each student i on item j includes
  - correct/incorrect response C_{ij},
  - and the solution time T_{ij}.

$$X_{ij} = (C_{ij}, T_{ij})$$

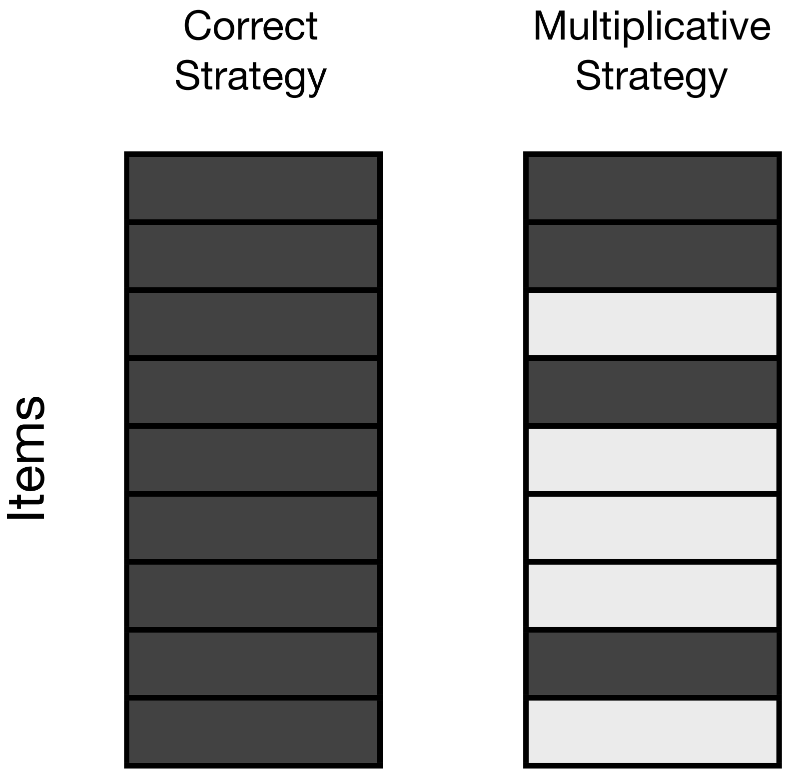
- An opportunity for learning followed each incorrect answer. This provides students additional opportunity to switch strategies.

(Pavlik et al., 2011)

Least Common Multiples Strategies

Problem	**Correct Strategy**	**Multiplicative Strategy**	**Other Strategies**
{4,5}	4×5 = 20	4×5 = 20	???
{4,6}	2×2×3 = 12	~~4×6 = 24~~	???

Theoretical Response Behavior



Darker cells indicate a higher probability of a correct response

Goal: Can the model uncover these strategies from the data?

Model Details for LCM Data

- Data

$$X_{ij} = (C_{ij}, T_{ij})$$

- Strategy distribution

$$F_{kj}(X_j) = \mathrm{Bernoulli}(\mathrm{C_j}; \lambda_{\mathrm{kj}}) \times \mathrm{Exp}(\mathrm{T_j}; \beta_{\mathrm{k}})$$

- λ_{kj} is probability of a correct response for strategy k on item j

$$p(\lambda_{1j}) = Beta(10, 1)$$ correct strategy

$$p(\lambda_{2j}) = Beta(1, 1)$$

$$p(\lambda_{3j}) = Beta(1, 1)$$

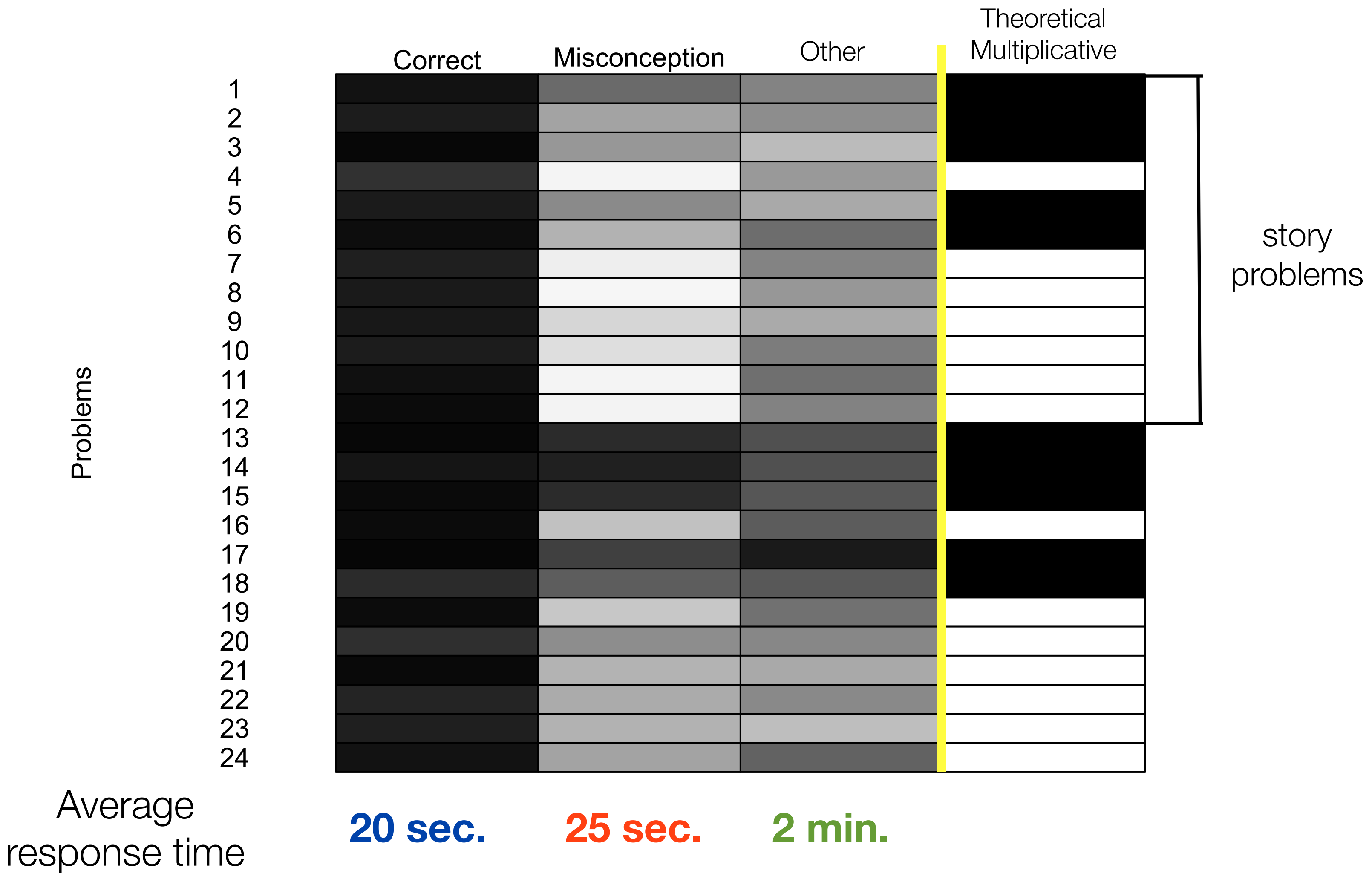
- $1/\beta_{\mathrm{k}}$ is mean response time for strategy k in milliseconds

$$p(\beta_k) = Gamma(1,\ 40000)$$

- Strategy membership parameter

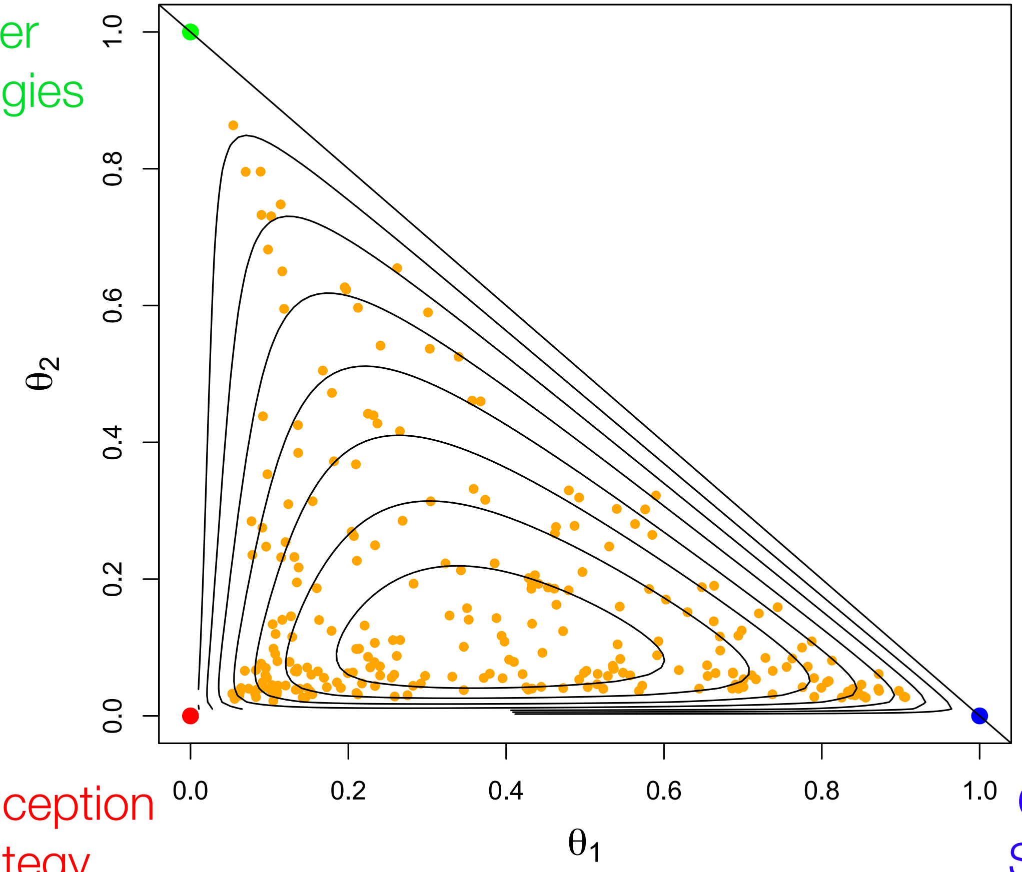
$$\theta_i \sim \text{Logistic-Normal}(\mu, \Sigma)$$

Posterior Probability of a Correct Response for Each Strategy



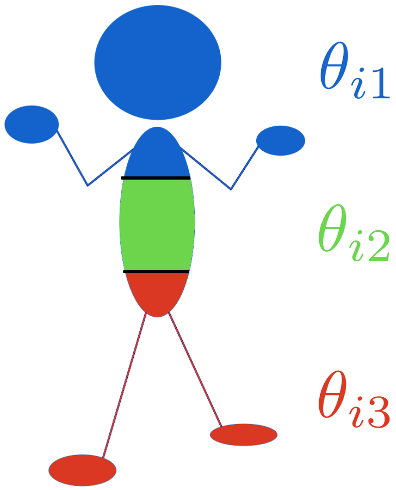
Posterior means of Strategy Membership Parameters

Other Strategies



Misconception Strategy

Correct Strategy



Conclusions

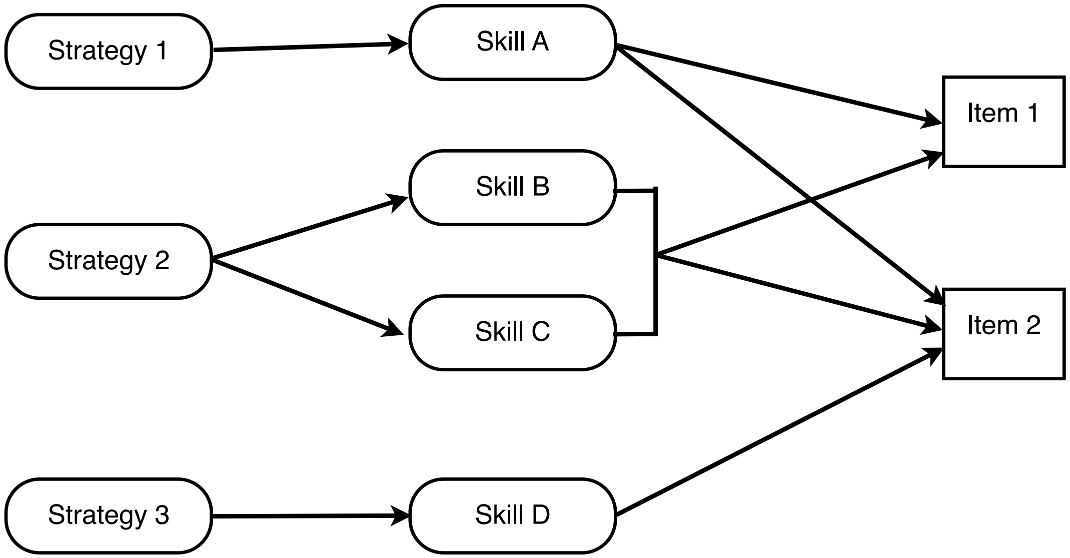
- It is possible to model strategy switching with mixed membership.
- We can recover both the strategies and how much students use each strategy with small data sets and very little prior information.
  - With 15 items/student - need prior information about 1 strategy
  - With 30 items/student - need no prior information

What’s novel here?

- Models each student using a mixture of strategies.
- Captures the mixture of strategies each student uses as an important measure of expertise.
- Models multiple student observations, including both accuracy and response time data.
- The conditional independence structure reflects that observed variables are outcomes of the same cognitive process.

Future Work

- A Multiple Strategies - Multiple Skill Model
  - Each strategy knowledge component may require a different set of skill knowledge components to execute it. (Koedinger et al, 2010)



Acknowledgements

This work was supported in part by Graduate Training Grant awarded to Carnegie Mellon University by the Department of Education # R305B040063.

Thanks to Philip Pavlik, Micheal Yudelson, and Ken Koedinger for generously sharing the Least Common Multiples data. This research was supported by the U.S. Department of Education (IES-NCSER) #R305B070487.

Thanks also to my dissertation advisors Brian Junker and Stephen Feinberg.

Thank You

References

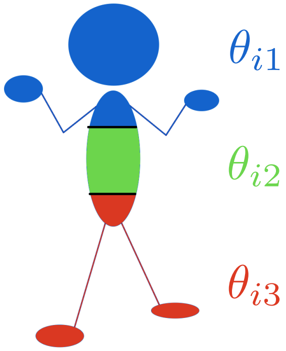
- Erosheva, Elena. (2002) Grade of Membership and Latent Structure Models With Application to Disability Survey Data. PhD thesis, Carnegie Mellon University, Pittsburgh, PA 15213.
- Erosheva, E., Fienberg, S.E., and Joutard, C. (2007). “Describing disability through individual-level mixture models for multivariate binary data.” *Annals of Applied Statistics*. 1(2) pp.502–537.
- Geiser, C., Lehmann, W., and Eid, M. (2006). “Separating ‘Rotators’ from ‘Nonrotators’ in the Mental Rotations Test: A Multigroup Latent Class Analysis.” *Multivariate Behavioral Research.* 41(3) pp.261-293.
- Junker, B. Some statistical models and computational methods that may be useful for cognitively-relevant assessment. Technical report, Committee on the Foundations of Assessment, National Research Council, November 1999.
- Koedinger, K.R., Corbett, A.T., Perfetti, C. (2010) The knowledge-learning-instruction framework: Toward bridging the science-practice chasm to enhance robust student learning. Technical Report CMU-HCII-10-102, Carnegie Mellon

References

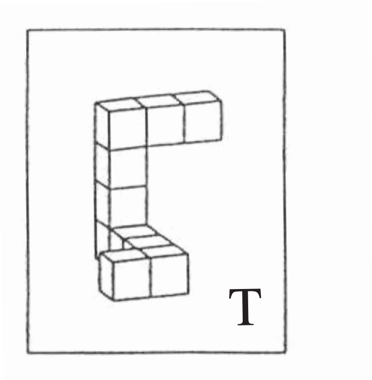
- Pavlik, P., Yudelson, M., and Koedinger, K.R. (2011) “Using contextual factors analysis to explain transfer of least common multiple skills.” In G. Biswas, et al. (eds) Artificial intelligence in education, vol. 6738 pp. 256-263. Springer Berlin / Heidelberg.
- Rouder, J.N., Sun, D., Speckman, P.L., Lu, J., and Zhou, D. (2003). “A hierarchical Bayesian statistical framework for response time distributions.” *Psychometrika*, 68(4): 589-606.
- Siegler, R.S. (1987). “The perils of averaging data over strategies: An example from children’s addition.” *Journal of Experimental Psychology: General.* 116(3) pp.250-264.
- National Research Council. (2001). *Knowing what students know: The science and design of educational assessment.* Washington, D.C.: National Academy Press.
- Wenger, M.J. (2005). “Models for the statistics and mechanisms of response speed and accuracy.” *Psychometrika*, 70(2).

Extra Slides

Foundational Ideas for the Multiple Strategies Model



- Each student uses a mixture of strategies. (Siegler, 1987)
- Each strategy knowledge component may require a different set of skill knowledge components to execute it. (Koedinger et al, 2010)
- For each item a student answers, we may observe several variables. These variables all depend on the same cognitive processes. (Wenger, 2005)



Formal Mixed Membership Model

1. **Assumptions/Definitions:**

k=1 k=2 k=3

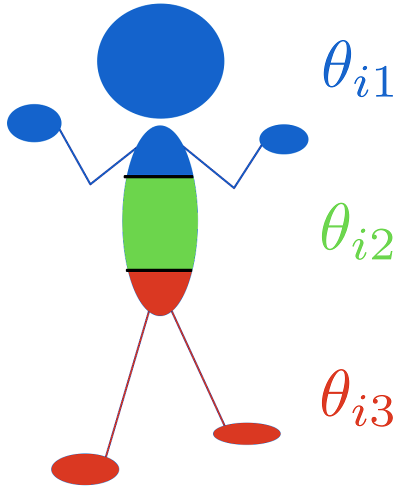
- N people
- K profiles
- J observed variables per person i $(X_{i1}, X_{i2}, ..., X_{iJ})$

$$X_j \sim F_{kj} \quad \text{for each profile}$$

2. **Subject level:**

- Individual membership in each profile is given by the vector θ_i
- Component θ_{ik} indicates the degree to which individual i belongs to profile k

$$\theta_{ik} \in [0,1] \qquad \sum_{k=1}^{K} \theta_{ik} = 1$$

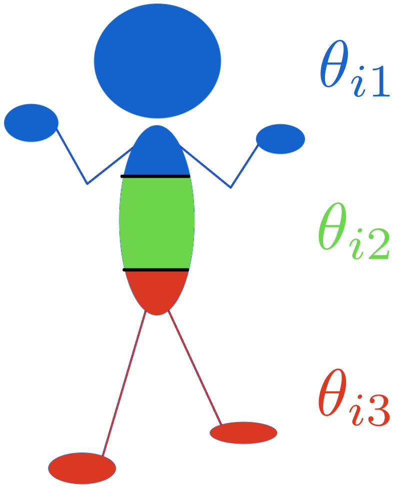


Formal Mixed Membership Model

2. **Subject level**:

- For each observed variable X_j, individual i's probability distribution is

$$F(x_j|\theta_i) \;=\; \sum_{k=1}^{K} \theta_{ik} F_{kj}(x_j)$$

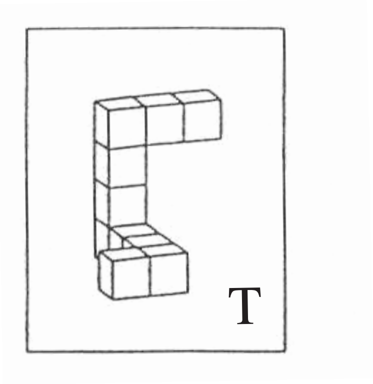


- Local Independence: Variables X_j are independent given membership vector θ_i

$$F(X_{i1}, X_{i2}, \ldots, X_{iJ}|\theta_i) = \prod_{j=1}^{J} \left[\sum_{k=1}^{K} \theta_{ik} F_{kj}(X_{ij}) \right]$$

Multiple Strategies, Multiple Skills Model

Skills and Strategies



- Each strategy may require a different set skills.
- Within the Knowledge-Learning-Instruction (KLI) Framework (Koedinger et al, 2010):
  - Skills are ‘atomic’ knowledge components.
  - Strategies are ‘integrative’ knowledge components.
- From a psychometric standpoint (Junker, 1999):
  - Strategies are disjunctive, a student can only use one strategy.
  - Skills are conjunctive, a student must possess all of the required skills to execute a particular strategy correctly.

Generalize to a
Multiple-Strategies, Multiple-Skills Model

- Data for person i on item j includes any measured variable:

$$X_{ij} = (C_{ij}, T_{ij}, \ldots)$$

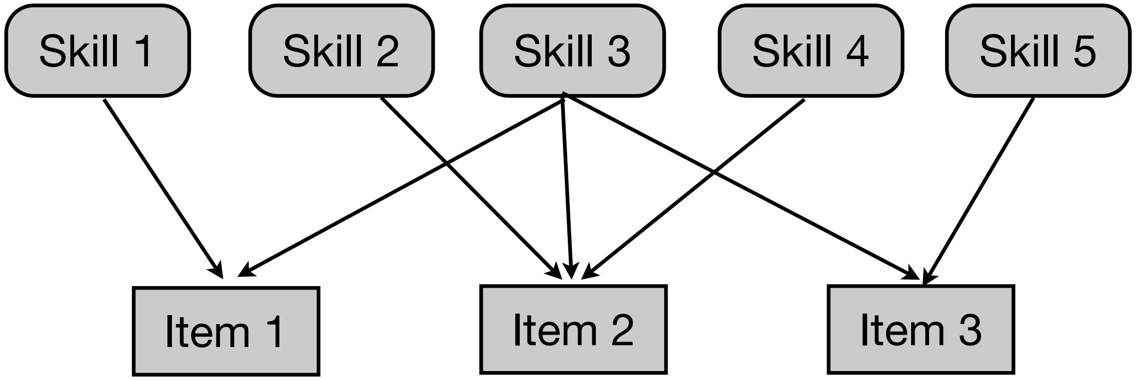
- Each strategy profile k defines a factorable distribution for these variables:

$$F_{kj}(X_{ij}) = \underset{\substack{\nearrow \\ \text{Cognitive Diagnosis Model}}}{F_{kj}}(C_{ij}) \times \underset{\substack{\nwarrow \\ \text{Response Time Model}}}{F_{kj}}(T_{ij}) \times \ldots$$

- Underlying Mixed Membership model allows for strategy switching.

Statistical Model for Accuracy Component Cognitive Diagnosis Models (CDM)

- In a CDM, the probability student i will correctly respond to item j depends on
  - q_j, the skills the item requires
  - α_i, the skills the student has mastered

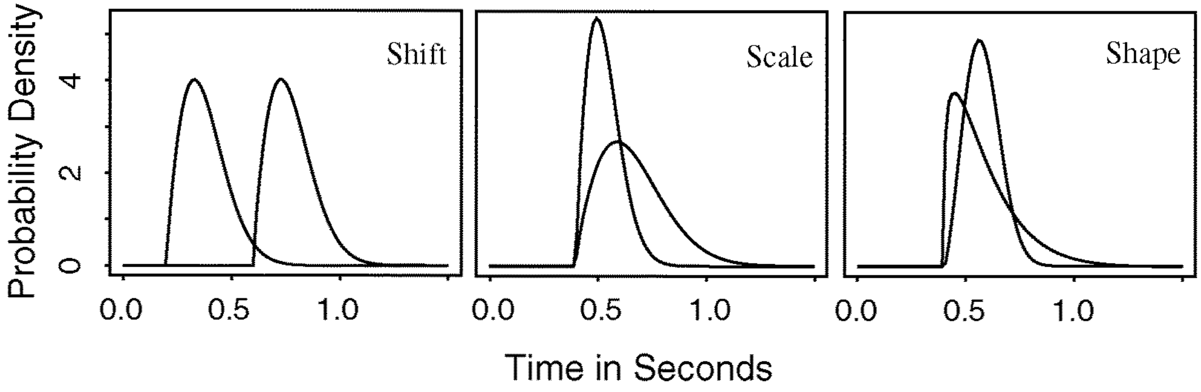


$q_1 = (1,0,1,0,0)$ $q_2 = (0,1,1,1,0)$ $q_3 = (0,0,1,0,1)$

Specifying q defines a strategy $F_{kj}(C_j) = Pr(C_j = 1|\alpha_i, q_j)$

Statistical Model for Response Time and Other Variables

- Example: Addition Strategies (Siegler, 1978)
  - Fast ● **Retrieval** or Memorization
  - Slower ● **Count-on:** to solve $7+2$, the child counts $8,9$
  - Very Slow ● **Count-All:** to solve $7+2$, the child counts $1,2,...,8,9$
- Each strategy has its own distribution of response times, $F_{kj}(T_j)$.
- Rouder et al., (2003) argue for a 3-parameter Weibull distribution.



Multiple-Strategies, Multiple-Skills Model

- Each strategy has factorable distribution for observed variables.

$$F_{kj}(X_{ij}) = F_{kj}(C_{ij}) \times F_{kj}(T_{ij}) \times \ldots$$

- The individual student distribution is the usual mixed membership distribution:

$$\begin{aligned} F(X_{i1}, X_{i2}, \ldots, X_{iJ} | \underset{\text{strategies}}{\theta_i}, \underset{\text{skills}}{\alpha_i}) &= \prod_j \left[\sum_k \theta_{ik} F_{kj}(X_{ij}|\alpha_i) \right] \\ &= \prod_j \left[\sum_k \theta_{ik} F_{kj}(C_{ij}|\alpha_i) F_{kj}(T_{ij}) \right] \end{aligned}$$

Theorems

Mixed Membership ⇔ Finite Mixture Model

THEOREM 1

A Mixed Membership Model with

- J observed variables and
- K basis profiles

can be represented as a Finite Mixture Model

- with K^J components indexed by

$$\zeta \in \mathcal{Z}^J = \{1, 2, \ldots, K\}^J$$

Erosheva (2004)

Multiple sets of MMM Profiles can generate the same FMM components

THEOREM 2

Let *F* and *G* be two sets of Mixed Membership profiles with

- J observed variables and
- K basis profiles

If $\forall\ k\ \exists\ k'$ such that $F_{kj} = G_{k'j}$

Then *F* and *G* generate the same Finite Mixture Model Components $F_{\zeta}(x)$

There are $K!^{(J-1)}$ such sets of basis profiles

Distinct basis profiles produce distinct probability constraints

THEOREM 3

Let *F* and *G* be distinct sets of Mixed Membership profiles with

- J observed variables and
- K basis profiles
- which produce the same set of components $F_\zeta(x)$

Then *F* and *G* induce distinct constraints on π_ζ

Galyardt

Main Identifiability Result

THEOREM 4

Let $A \subseteq \mathcal{Z} = \{1, \ldots, K\}$, and let $\mathbb{A}$ be the set of all bi-jections $a : \mathcal{Z} \rightarrow \mathcal{Z}$ s.t. $a(i) = i \quad \forall \quad i \in A^C$.

If

- Condition 1: $\forall \;\; a \in \;\mathbb{A} \quad D\left(\theta_z\right) = D\left(\theta_{a(z)}\right)$
- Condition 2: $\exists \;\; a \in \mathbb{A}$ s.t. $F_{kj} = G_{a(k)j} \;\; \forall \; j, k$

Then F and G generate the same Mixed Membership Model

There are $|A|!^{(J-1)}$ sets of basis profiles in the equivalence class.

Galyardt

Addition Strategies Example

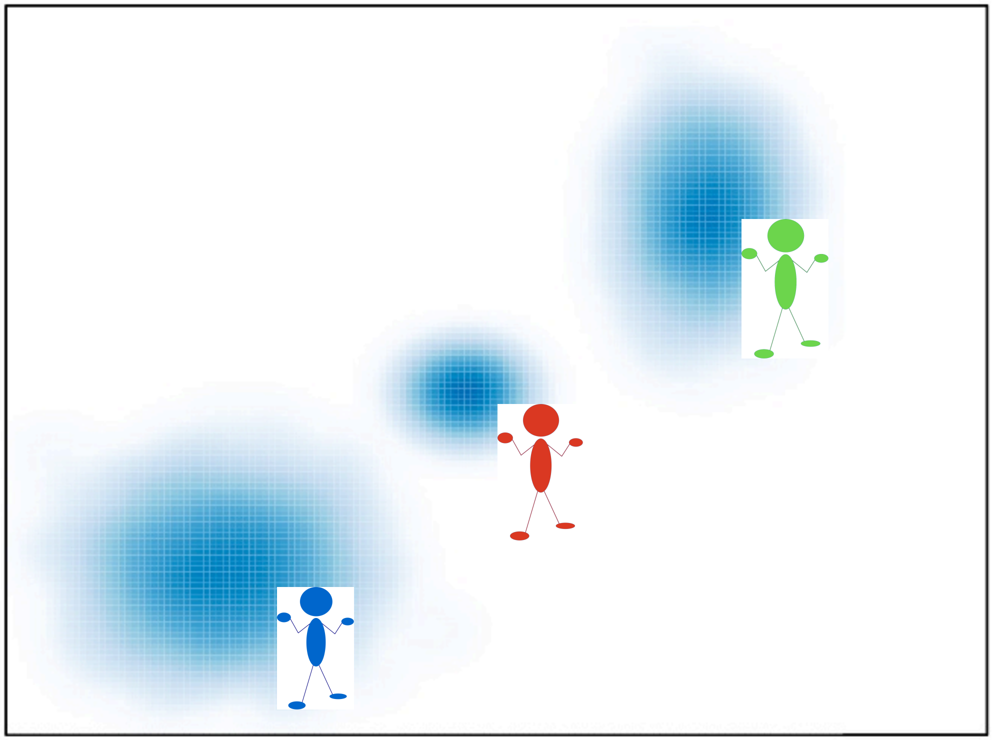
- Addition Strategies
  - Fast **Retrieval** or Memorization
  - Slower **Count-on:** to solve $7+2$, the child counts $8,9$
  - Very Slow **Count-All:** to solve $7+2$, the child counts $1,2,\ldots,8,9$
- Solution times distinguish strategies.
- Multiple problems to observe multiple strategies.
- 2 problems make a simple example.

(Siegler 1987)

Addition Solution Times

2 addition problems

3 strategies



Solution Time for addition problem 2

Solution Time for addition problem 1

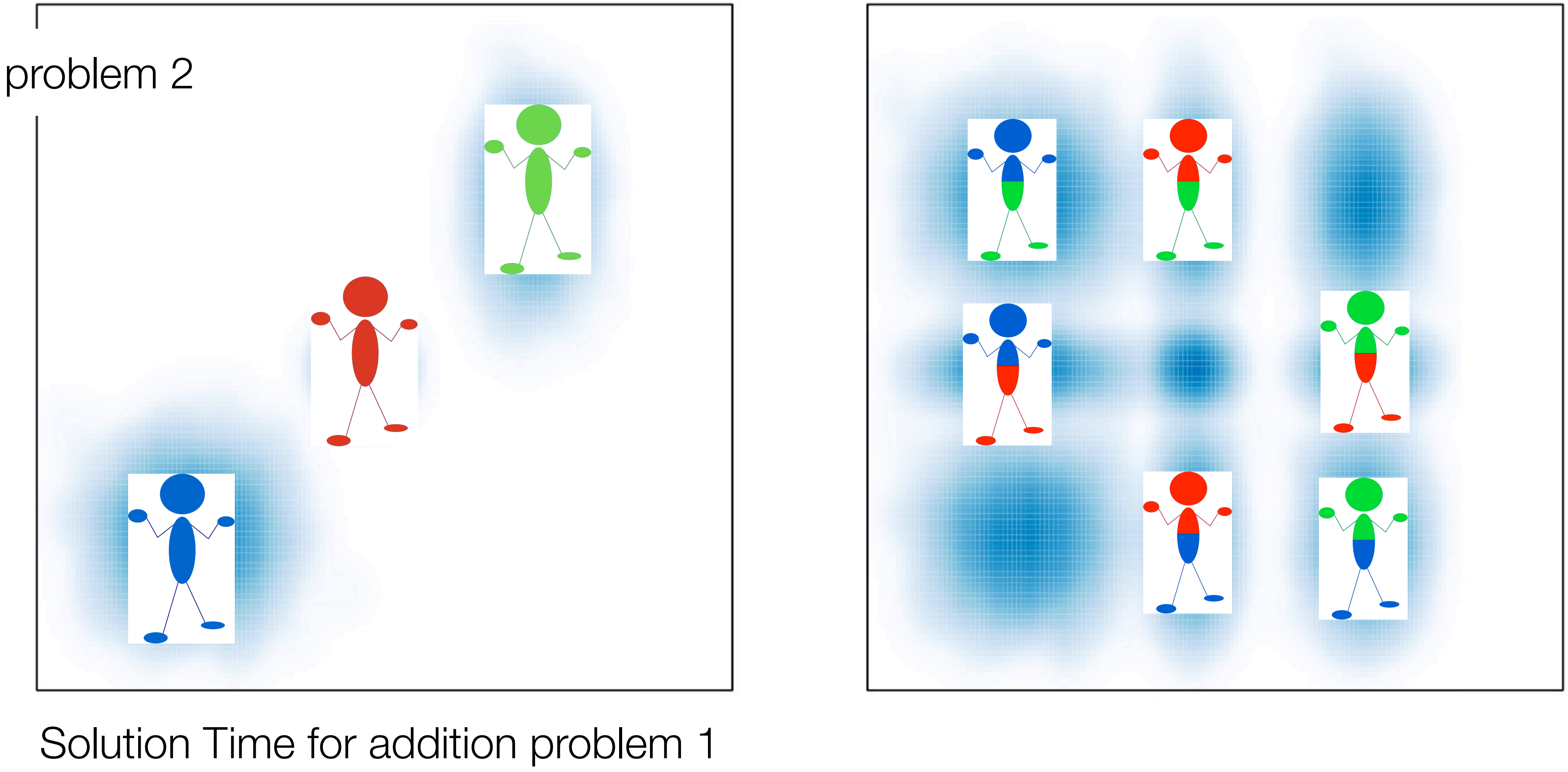
Count all, very slow

Count on, slow

Retrieval, fast

Every MMM can be written as an Latent Class Model with many more classes

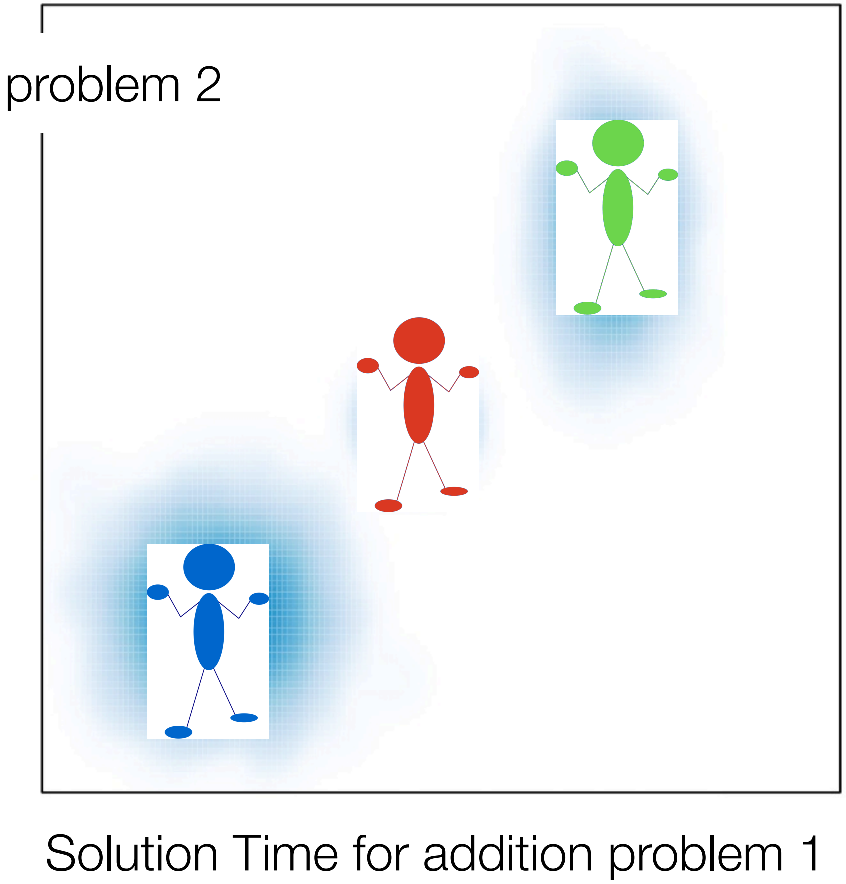
2 problems, 3 strategies ➙ 3^2 LCM classes

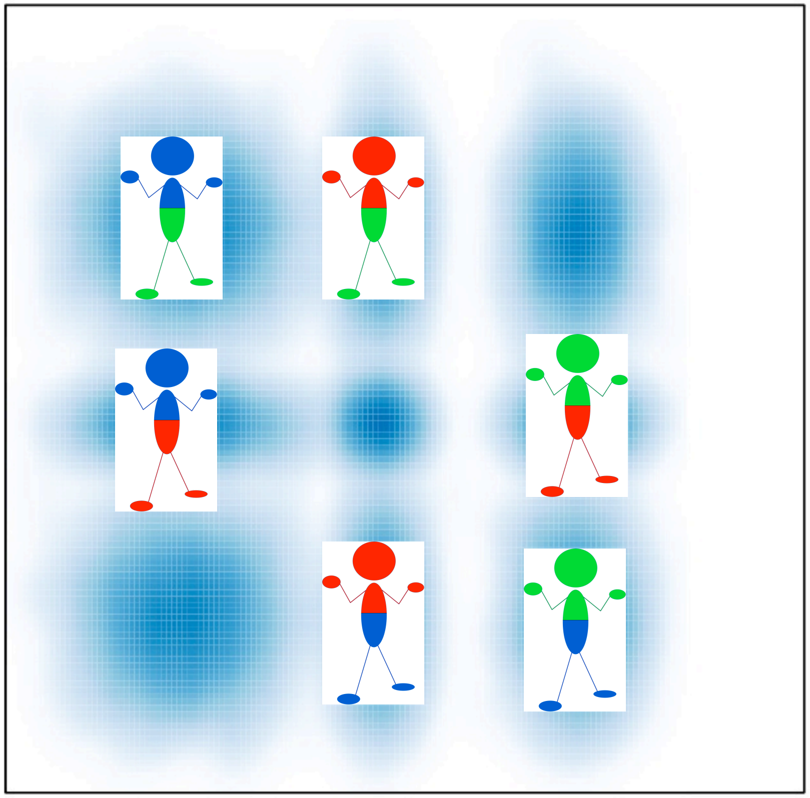


(Erosheva, 2007; Galyardt, 2012)

LCM Class probability constraints

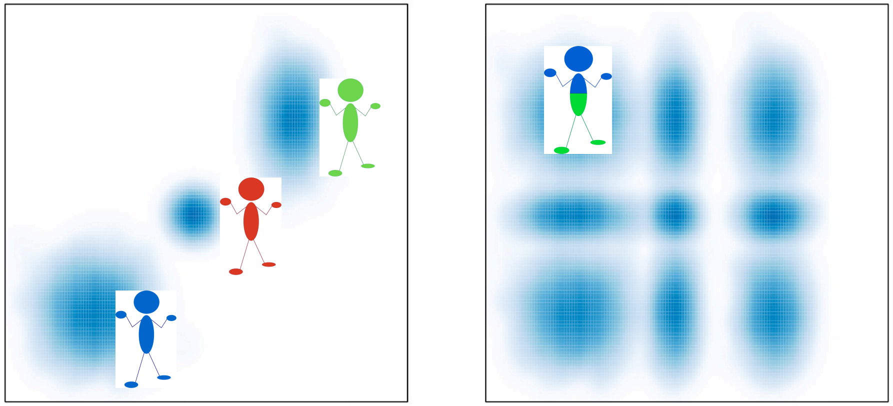
For these strategy profiles ➠ Blue-Red is equivalent to Red-Blue



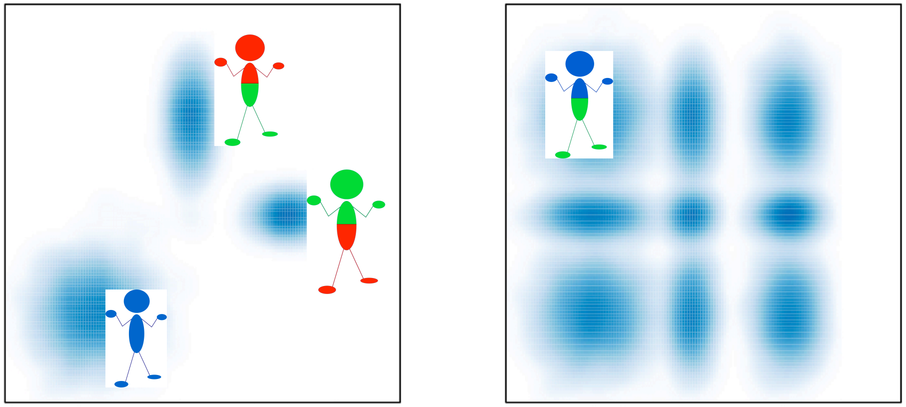


Different strategy profiles could generate the same data.

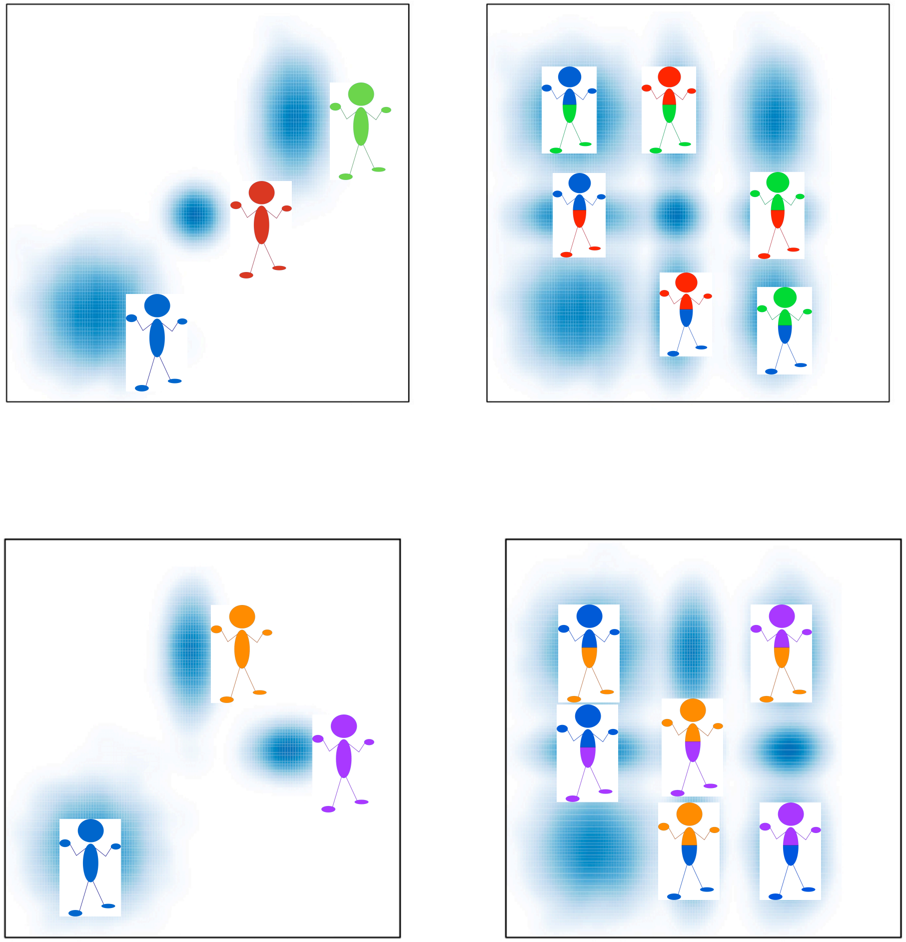
“Pure” strategies



Alternate profiles



Distinct strategy profiles produce distinct probability constraints



Cause for Concern

- Mixed Membership Models have serious potential identifiability problems analogous to
  - Latent Class Models
  - Factor Analysis

- This has implications for modeling multiple strategy use.
  - Addition Strategies

Main Identifiability Result

THEOREM

Let $A \subseteq \mathcal{Z} = \{1, \ldots, K\}$, and let $\mathbb{A}$ be the set of all bi-jections $a : \mathcal{Z} \to \mathcal{Z}$ s.t. $a(i) = i \quad \forall \quad i \in A^C$.

If

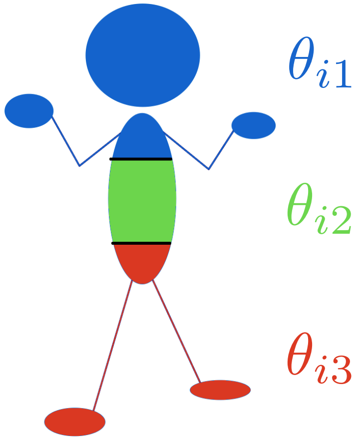
- Condition 1: $\forall \;\; a \in \; \mathbb{A} \quad D\left(\theta_z\right) = D\left(\theta_{a(z)}\right)$
- Condition 2: $\exists \;\; a \in \mathbb{A}$ s.t. $F_{kj} = G_{a(k)j} \;\; \forall \; j, k$

Then F and G generate the same Mixed Membership Model

There are $|A|!^{(J-1)}$ sets of basis profiles in the equivalence class.

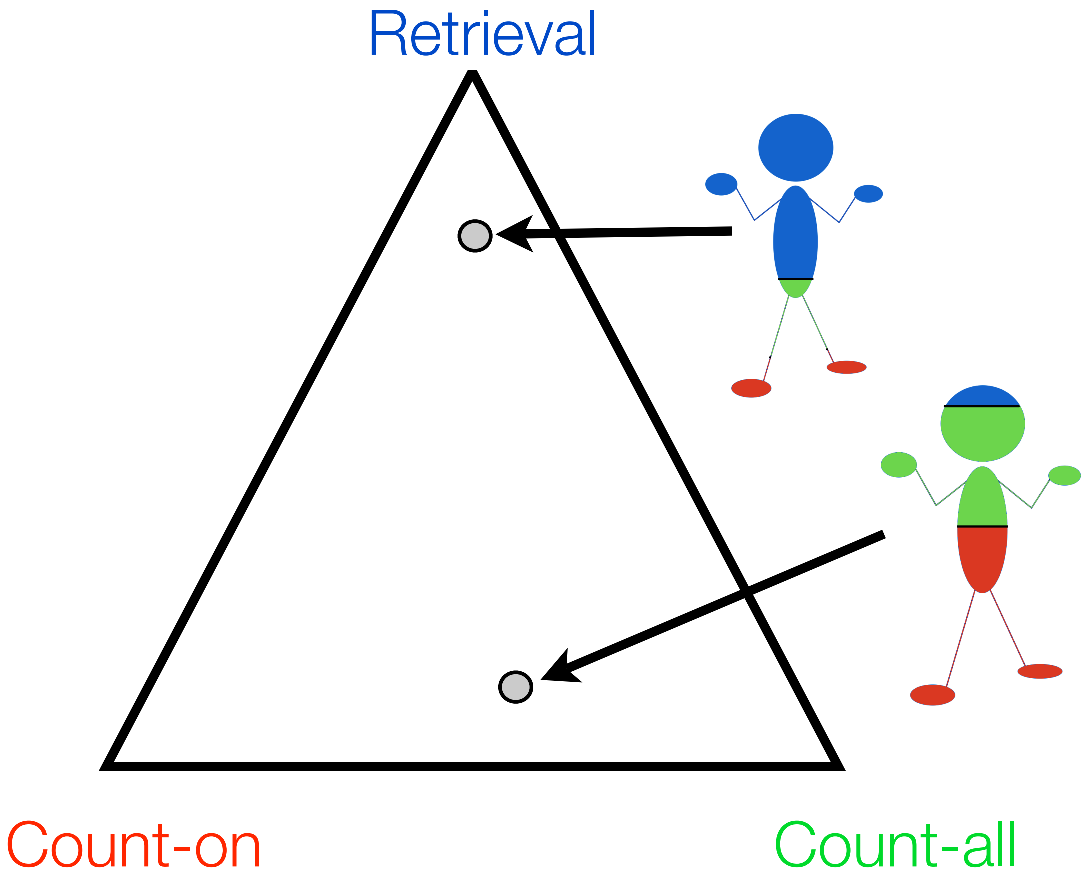
(Galyardt, 2012)

Distributions of Strategy Use

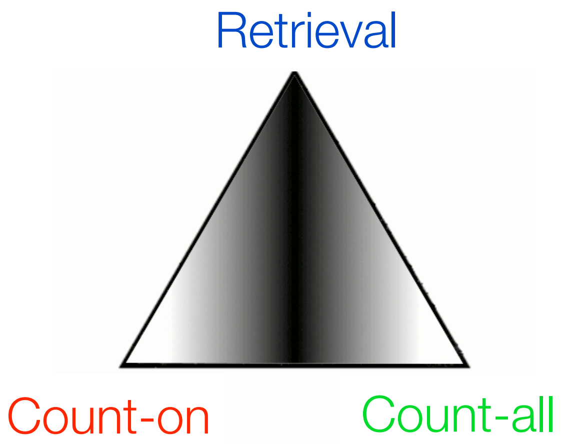


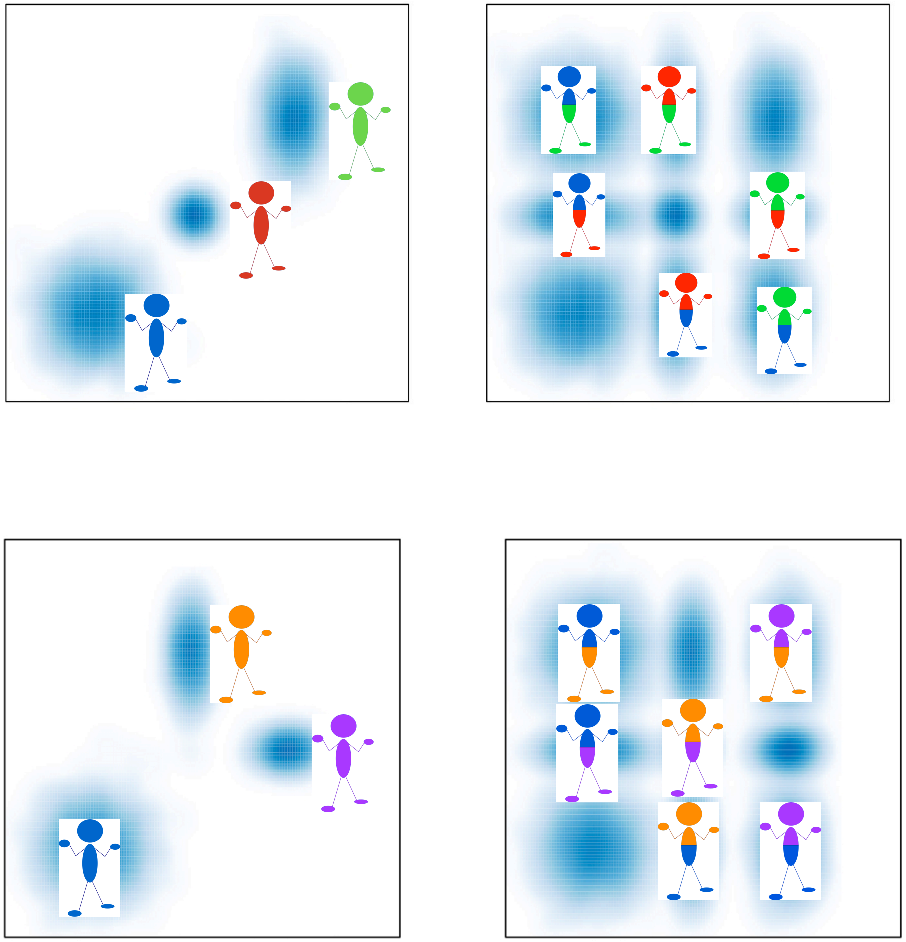
$$\theta_{ik} \in [0, 1]$$

$$\sum_{k=1}^{K} \theta_{ik} = 1$$



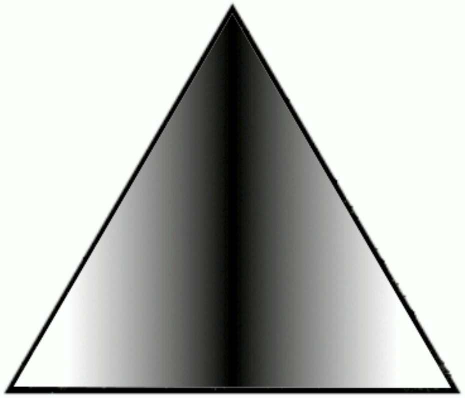
Two strategy profiles
and a particular strategy-use distribution





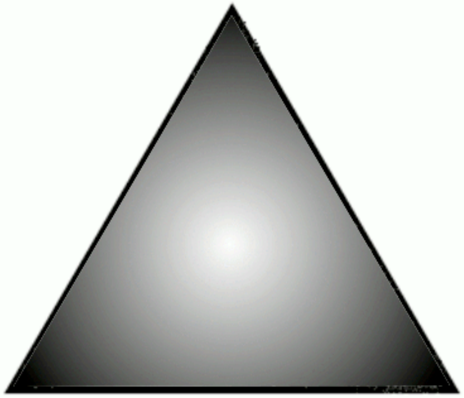
In this example,
these 2 profile sets
are the entire
equivalence class.

Example Distributions of the Strategy-Profile Membership Parameter



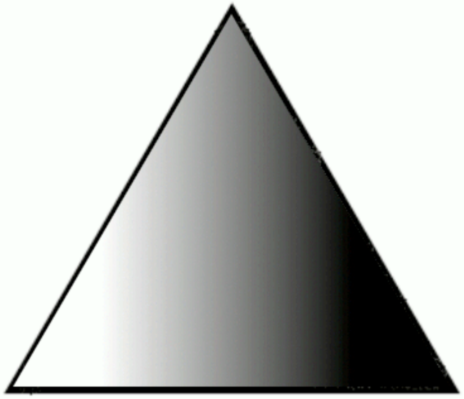
2-fold symmetry

Some equivalent sets of strategy profiles.



Complete symmetry

Equivalence class at maximal size.



No symmetry

Unique set of strategy profiles.

Continuous & Categorical Data

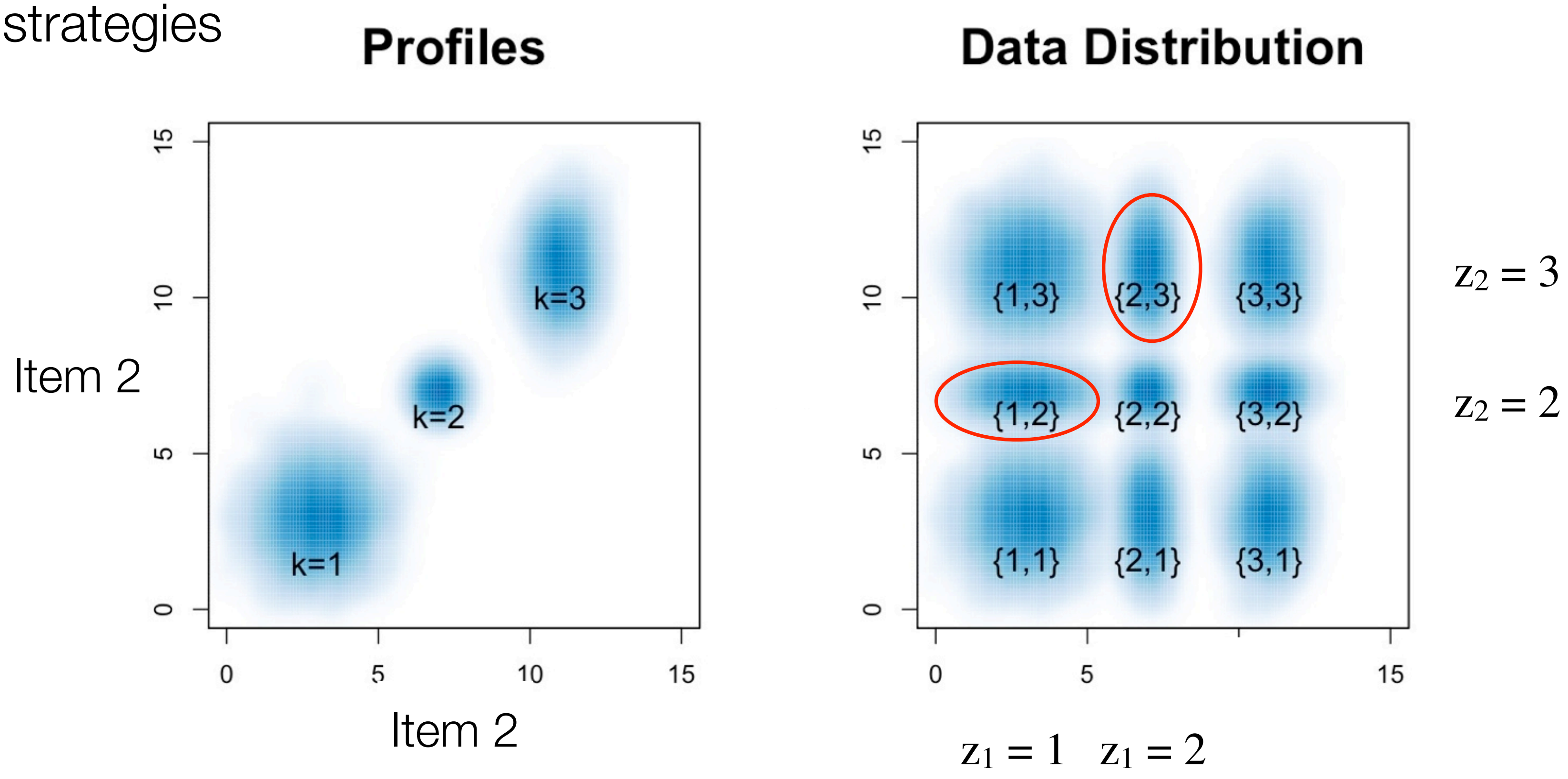
Implications

- When data is categorical, Mixed Membership is appropriate
  - IF Students switch strategies, OR
  - IF Students use a blend of profile strategies.

- When data is NOT categorical, Mixed Membership is appropriate
  - ONLY IF Students switch strategies.

General Interpretation: Switching

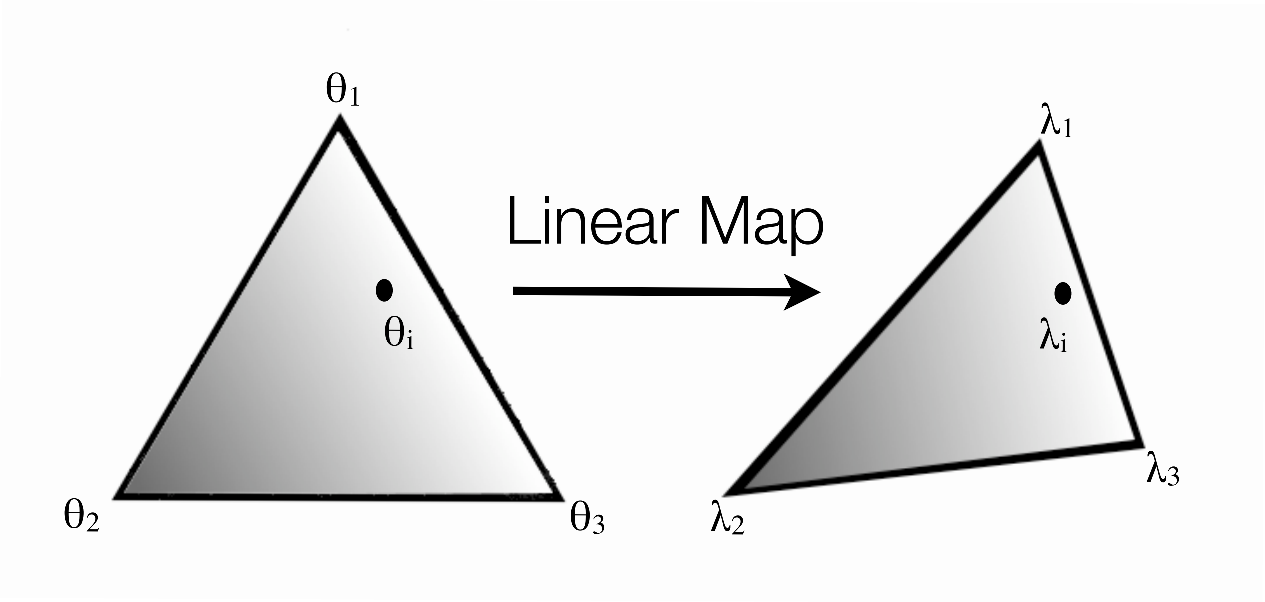
2 items
3 strategies



$$\zeta = \{z_1, z_2\} = \{2, 3\}$$

Categorical Interpretation: Between

- When data is categorical, we can interpret individuals as being “between” strategies.



Membership in Strategy Profiles

Categorical Probability