# Week 4 Video 3

## Knowledge Inference: Performance Factors Analysis

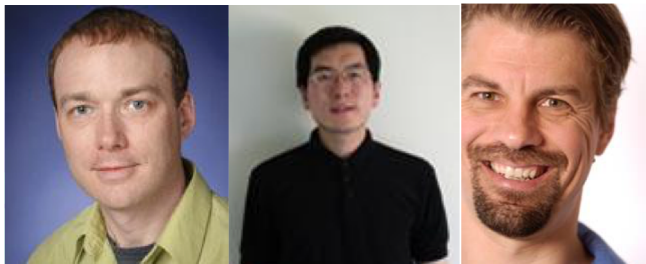
# Performance Factors Analysis

An alternative to BKT

Addresses some of the limitations of BKT

But doesn't have all of the nice features of BKT

Proposed in 2009 by Pavlik, Cen, & Koedinger



# PFA

- Measures how much latent skill a student has, while they are learning
  - But expresses it in terms of probability of correctness, the next time the skill is encountered
  - No direct expression of the amount of latent skill, except this probability of correctness

# What is the typical use of PFA?

Assess a student's knowledge of topic X

Based on a sequence of items that are dichotomously scored

- E.g. the student can get a score of 0 or 1 on each item

Where the student can learn on each item, due to help, feedback, scaffolding, etc.

# How does PFA differ from BKT?

# Key assumptions

Each item may involve multiple latent skills or knowledge components

- Different from BKT

Each skill has success learning rate $\gamma$ and failure learning rate $\rho$

- Different from BKT where learning rate is the same, success or failure

# Key assumptions

There is also a difficulty parameter β, but its semantics can vary – more on this later

From these parameters, and the number of successes and failures the student has had on each relevant skill so far, we can compute the probability P(*m*) that the learner will get the item correct

# PFA

# PFA

$$\mathrm{m}(i, j \in KCs, k \in Items, s, f) = \beta + \sum_{j \in KCs} (\gamma_j s_{i,j} + \rho_j f_{i,j})$$

$$\mathrm{p}(m) = \frac{1}{1 + e^{-m}}$$

# Reasonable Example

$\gamma = 0.2$, $\rho = 0.1$, $\beta = -0.5$

| Actual | m | P(m) |
|---|---|---|
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |

# Reasonable Example

$\gamma = 0.2$, $\rho = 0.1$, $\beta = -0.5$

| Actual | m | P(m) |
|---|---|---|
| | -0.5 | 0.38 |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |

# Reasonable Example

$\gamma = 0.2$, $\rho = 0.1$, $\beta = -0.5$

| Actual | m | P(m) |
|---|---|---|
| 0 | -0.5 | 0.38 |
| | -0.5+(0.1*1) | |
| | | |
| | | |
| | | |
| | | |
| | | |

# Reasonable Example

$\gamma = 0.2$, $\rho = 0.1$, $\beta = -0.5$

| Actual | m | P(m) |
|---|---|---|
| 0 | -0.5 | 0.38 |
| | -0.5+0.1 | |
| | | |
| | | |
| | | |
| | | |
| | | |

# Reasonable Example

$\gamma = 0.2$, $\rho = 0.1$, $\beta = -0.5$

| Actual | m | P(m) |
|---|---|---|
| 0 | -0.5 | 0.38 |
| | -0.4 | 0.40 |
| | | |
| | | |
| | | |
| | | |
| | | |

# Reasonable Example

$\gamma = 0.2$, $\rho = 0.1$, $\beta = -0.5$

| Actual | m | P(m) |
|---|---|---|
| 0 | -0.5 | 0.38 |
| 0 | -0.4 | 0.40 |
| | | |
| | | |
| | | |
| | | |
| | | |

# Reasonable Example

$\gamma = 0.2, \rho = 0.1, \beta = -0.5$

| Actual | m | P(m) |
|---|---|---|
| 0 | -0.5 | 0.38 |
| 0 | -0.4 | 0.40 |
| | -0.5+(0.1*2) | |
| | | |
| | | |
| | | |
| | | |

# Reasonable Example

$\gamma = 0.2$, $\rho = 0.1$, $\beta = -0.5$

| Actual | m | P(m) |
|---|---|---|
| 0 | -0.5 | 0.38 |
| 0 | -0.4 | 0.40 |
| | -0.5+0.2 | |
| | | |
| | | |
| | | |
| | | |

# Reasonable Example

$\gamma = 0.2$, $\rho = 0.1$, $\beta = -0.5$

| Actual | m | P(m) |
|---|---|---|
| 0 | -0.5 | 0.38 |
| 0 | -0.4 | 0.40 |
| | -0.3 | |
| | | |
| | | |
| | | |
| | | |

# Reasonable Example

$\gamma = 0.2, \rho = 0.1, \beta = -0.5$

| Actual | m | P(m) |
|---|---|---|
| 0 | -0.5 | 0.38 |
| 0 | -0.4 | 0.40 |
| | -0.3 | 0.43 |
| | | |
| | | |
| | | |
| | | |

# Reasonable Example

$\gamma = 0.2$, $\rho = 0.1$, $\beta = -0.5$

| Actual | m | P(m) |
|---|---|---|
| 0 | -0.5 | 0.38 |
| 0 | -0.4 | 0.40 |
| 1 | -0.3 | 0.43 |
| | | |
| | | |
| | | |
| | | |

# Reasonable Example

$\gamma = 0.2, \rho = 0.1, \beta = -0.5$

| Actual | m | P(m) |
|---|---|---|
| 0 | -0.5 | 0.38 |
| 0 | -0.4 | 0.40 |
| 1 | -0.3 | 0.43 |
| | -0.5+(0.1*2)+(0.2*1) | |
| | | |
| | | |
| | | |

# Reasonable Example

$\gamma = 0.2$, $\rho = 0.1$, $\beta = -0.5$

| Actual | m | P(m) |
|---|---|---|
| 0 | -0.5 | 0.38 |
| 0 | -0.4 | 0.40 |
| 1 | -0.3 | 0.43 |
| | -0.5+0.2+0.2 | |
| | | |
| | | |
| | | |

# Reasonable Example

$\gamma = 0.2$, $\rho = 0.1$, $\beta = -0.5$

| Actual | m | P(m) |
|---|---|---|
| 0 | -0.5 | 0.38 |
| 0 | -0.4 | 0.40 |
| 1 | -0.3 | 0.43 |
| | -0.1 | 0.48 |
| | | |
| | | |
| | | |

# Degenerate Example

$\gamma = 0.1$, $\rho = 0.2$, $\beta = -0.5$

| Actual | m | P(m) |
|---|---|---|
| 0 | -0.5 | 0.38 |
| 0 | -0.3 | 0.43 |
| 1 | -0.1 | 0.48 |
| | 0 | 0.5 |
| | | |
| | | |
| | | |

# “Negative Learning”

$\gamma = 0.1, \rho = -0.5, \beta = -0.5$

| Actual | m | P(m) |
|---|---|---|
| 0 | -0.5 | 0.38 |
| 0 | -1 | 0.27 |
| 1 | -1.5 | 0.18 |
| | -1.4 | 0.20 |
| | | |
| | | |
| | | |

# Note

Values of ρ below 0 don't actually mean negative learning

They mean that failure provides more evidence on lack of knowledge

Than the learning opportunity causes improvement

# Note

Parameters in PFA combine information from correctness with improvement from practice improvement

Makes PFA models a little harder to interpret than BKT

# Adjusting β

$\gamma = 0.2, \rho = 0.1, \beta = -0.5$

| Actual | m | P(m) |
|---|---|---|
| 0 | -0.5 | 0.38 |
| 0 | -0.4 | 0.40 |
| 1 | -0.3 | 0.43 |
| | -0.1 | 0.48 |
| | | |
| | | |
| | | |

# Adjusting β

$\gamma = 0.2, \rho = 0.1, \beta = -1.5$

| Actual | m | P(m) |
|---|---|---|
| 0 | -1.5 | 0.18 |
| 0 | -1.4 | 0.20 |
| 1 | -1.3 | 0.21 |
| | -1.1 | 0.25 |
| | | |
| | | |
| | | |

# Adjusting β

$\gamma = 0.2$, $\rho = 0.1$, $\beta = +3.0$

| Actual | m | P(m) |
|---|---|---|
| 0 | 3.0 | 0.953 |
| 0 | 3.1 | 0.957 |
| 1 | 3.2 | 0.961 |
| | 3.4 | 0.968 |
| | | |
| | | |
| | | |

# β Parameters

Pavlik proposes three different β Parameters

- Item
- Item-Type
- Skill

Result in different number of parameters

- And greater or lesser potential concern about over-fitting

# Fitting PFA

Typically Expectation Maximization is used

# Expectation Maximization

1. Starts with initial values for each parameter
2. Estimates student correctness at each problem step
3. Estimates params using student correctness estimates
4. If goodness is substantially better than last time it was estimated, and max iterations has not been reached, go to step 2

# Expectation Maximization

EM is vulnerable to local minima

Randomized restart typically used

# Is PFA better than BKT?

Approximately equal predictive power across a lot of studies (Pavlik et al., 2009; Gong et al., 2010; Baker et al., 2011; Pardos et al., 2011, 2012)

Different virtues and flaws – choose the one that better fits your goals

# Final Thoughts

PFA is a competitor for measuring student skill, which predicts the probability of correctness rather than latent knowledge

Can handle multiple KCs for the same item, a big virtue

# Next Up

Item Response Theory