Week 8 Video 1

Discovery with Models
Discovery with Models: Seems Tricky
Discovery with Models: Seems Tricky

- But, in fact, discovery with models is so simple that a 6-year old could explain it
Discovery with Models: Seems Tricky

- But, in fact, discovery with models is so simple that a 6-year old could explain it

- And now, here to do so, is today’s guest lecturer, Maria Baker
Discovery with Models: The Big Idea

- A model of a phenomenon is developed

- Via
  - Prediction
  - Clustering
  - Knowledge Engineering

- This model is then used as a component in another analysis
Thank you, Maria
Discovery with models can be used in Prediction

- The created model’s predictions are used as predictor variables in predicting a new variable

- E.g. Classification, Regression
Can be used in Relationship Mining

- The relationships between the created model’s predictions and additional variables are studied.
- This can enable a researcher to study the relationship between a complex latent construct and a wide variety of observable constructs.
- E.g. Correlation mining, Association Rule Mining.
Models on top of Models

- Another area of Discovery with Models is composing models out of other models

- Examples:
  - Models of Gaming the System and Help-Seeking use Bayesian Knowledge-Tracing models as components (Baker et al., 2004, 2008a, 2008b; Aleven et al., 2004, 2006)
  - Models of Preparation for Future Learning use models of Gaming the System as components (Baker et al., 2011)
  - Models of Affect use models of Off-Task Behavior as components (Baker et al., 2012)
  - Models predicting standardized exam scores use models of Affect and Off-Task Behavior as components (Pardos et al., 2013)
  - Models predicting college attendance use models of Affect and Off-Task Behavior as components (San Pedro et al., 2013)
When I talk about this, people often worry about building a model on top of imperfect models.

Will the error “pile up”? 
Models on top of Models

- May not be as big a risk as people worry about

- If the final model successfully predicts the final construct, do we care if the model it uses internally is imperfect?

- Systematic error at each step is taken into account at the next step!

- That said, there are some validity issues that should be taken into account — more on that in a minute
“Increasingly Important…”

- Baker & Yacef (2009) argued that Discovery with Models is a key emerging area of EDM.

- I think that’s still true, although it has been a bit slower to become prominent than I expected way back in 2009.
DWM analyses we’ve previously talked about in class

- San Pedro et al. (2013) – Week 1, Lecture 6
- Beck et al. (2008) – Week 4, Lecture 5
- Baker et al. (2010, 2012) – Week 4, Lecture 5
- Fancsali (2013) – Week 5, Lecture 2
- Gowda et al. (2012) – Week 6, Lecture 2
DWM analysis we’ll talk about in next lecture

Advantages of DWM
Advantages of DWM

- Possible to Analyze Phenomena at Scale

- Even for constructs that are
  - latent
  - expensive to label by hand
Advantages of DWM

- Possible to Analyze Phenomena at Scale
  - At scales that are infeasible even for constructs that are quick & easy to label by hand
    - Scales easily from hundreds to millions of students
    - Entire years or (eventually) entire lifetimes
      - Predicting Nobel Prize winners from kindergarten drawings?
Future Nobel Prize Winner?
Advantages of DWM

- Supports inspecting and reconsidering coding later
  - Leaves clear data trails
  - Can substitute imperfect model with a better model later and re-run
  - Promotes replicability, discussion, debate, and scientific progress
Disadvantages of DWM

- Easy to Do Wrong!
Discovery with Models:
Here There Be Monsters

“Rar.”
Discovery with Models: Here There Be Monsters

- It’s really easy to do something badly wrong, for some types of “Discovery with Models” analyses

- No warnings when you do
Think Validity

- Validity is always important for model creation

- Doubly-important for discovery with models
  - Discovery with Models almost always involves applying model to new data
  - How confident are you that your model will apply to the new data?
Challenges to Valid Application

- Many challenges to valid application of a model within a discovery with models analysis
Challenges to Valid Application

- Is model valid for population?
- Is model valid for all tutor lessons? (or other differences)
- Is model valid for setting of use? (classroom versus homework?)
- Is the model valid in the first place? (especially important for knowledge engineered models)
Next lecture

- Discovery with Models: Case Study