Causal Modeling to Understand the Relationship between Student Attitudes, Affect and Outcomes

Dovan Rai
Department of Computer Science
Worcester Polytechnic Institute
dovan@wpi.edu

Joseph E. Beck
Department of Computer Science
Worcester Polytechnic Institute
josephbeck@wpi.edu

Ivon Arroyo
Department of Computer Science
University of Massachusetts Amherst
ivon@cs.umass.edu

ABSTRACT

We are using causal modeling to analyze relationships between pedagogical intervention, students' attitudes, affective states, perceptions and outcomes, based on the data from a math tutor, Wayang Outpost. The causal model generated gives interpretable multi level interrelationships within the data variables identifying direct and indirect effects among them. We observed that among the four affective variables, confidence and frustration are more tightly linked with their performance and ability whereas interest and excitement are more related to their attitude and appreciation of math and tutor.

1. INTRODUCTION

In recent years, researchers have found that attending to students' motivational and affective characteristics are as crucial as cognitive aspects for effective learning. Since students' cognitive, affective and behavioral aspects are interconnected, analyzing their interrelationships give us clearer understanding of the learning process.

Causal models [1, 2] are graphical models that make the additional assumption that the links between nodes represent causal influence. By causal, we mean that a link A→B means that if we intervene and change the value of A, then B will change. Based on the conditional independencies within the data, causal modeling makes causal inferences among the variables.

We used TETRAD, a free causal modeling software package [2] to generate causal graphs. We also made use of an extension to Tetrad, developed by Doug Selent, a graduate student at WPI, that enabled us to restrict links on the basis of the magnitude of the relation between the pair of nodes. It measures the magnitude of the relation using R^2, the amount of the variability accounted for by the relationship between the nodes.

The data comes from students working with Wayang Outpost, an adaptive math tutoring system that helps students learn to solve standardized-test questions, in particular state-based exams taken at the end of high school in the USA.

We are using data from 94 middle school students, in the 7th and 8th grade (approximately 12 to 14 years old), who were part of mathematics classes in a rural-area public middle school in Massachusetts, USA. As part of the activity, students took a survey on the first day, which assessed their baseline achievement level, as well as affective and motivational factors related to mathematics problem solving. The students took identical survey at the end of the study. Student responses were collected in 5-point Likert scale.

Based on the survey data, we created different variables, which we will be explaining in the following paragraphs.

Attitudes and appreciation of Math: Students were asked various questions related to their attitude towards Math before they used the tutor.

Math Self Concept (Sample item: *How good would you be at learning something new in math?*)

Math Liking (Sample item: *How much do you like doing math?*)

Math Value (Sample item: *In general, how useful is what you learn in math?*)

Affect: Students were asked questions on four affective variables while using the tutor: Confident, Frustrated, Interested, Excited (Sample Item: How confident do you feel when solving math problems?)

Pedagogical intervention: MathFluencyTraining (training on basic math facts, e.g. multiplication tables, and retrieval speed) Students were randomly assigned to either receive or not receive math fluency training.

Perception of tutoring system: PerceptionWayang (Students' perception of the tutor)

Perception_LC (Students' perception of learning companion, an animated learning peer who offers encouragement to the student)

Pretest Score and learningGain: Students took MCAS (state standardized test) test before using the tutoring system. We are using this test score as pretest score.

We calculate the difference in test scores between the MCAS tests students took before and after using the tutor and designate that value as learningGain.

Gain and outcomes
Based on student's pre and post survey responses, gain outcomes were calculated. We are using the base and gain values for the constructs and not considering post values (which are redundant once we took the first the based and gain values).

2. CAUSAL MODELING

We inputted the data in TETRAD and performed a model search using the PC search algorithm and generated a graph as shown in figure 1. PC search algorithm assumes that there are no hidden common causes between observed variables in the input (i.e., variables from the data set, or observed variables in the input graph) and that the graphical structure sought has no cycles [2]. The causal links are color-coded (corresponding to the R^2 value: red >0.5, orange >0.1, yellow >0.05) and 'plus' sign denotes positive relationship and 'minus' sign denotes negative relationships. We also input domain knowledge based on temporal precedence. For example: gender is put on higher knowledge tier than math self concept so that there could be a causal link directed from gender to math self-concept but not the other way round.
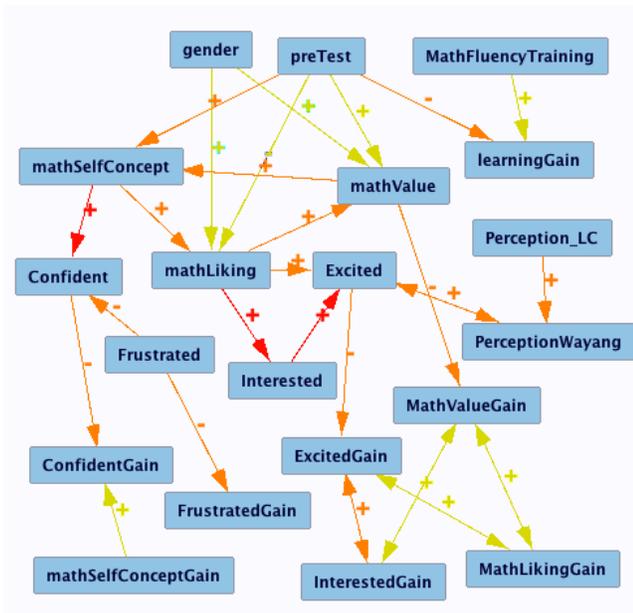
Figure 1 Causal model of attitude, affect and outcomes

We observe very strong relationships between student attitude towards math and her affective states within the tutor. Students who have higher self-concept in math report being more *Confident*. Students who like math report being more *Interested* and *Excited* while using the tutor. While *Interested, Excited and Confident* are more tightly coupled with attitude variables, *Frustrated* is relatively separate and connected to that web via the variable *Confident*.

In the correlation matrix, all three math attitude variables are related to *Interested, Excited,* and *Confident* and only *mathSelfConcept* and *MathLiking* are negatively correlated to *Frustrated* (*mathValue* has no relation to *Frustrated*). Three math attitude variables are also highly correlated among themselves. Causal model has teased apart this dense correlation web into sparser directed structure.

From causal model, we see that gender directly affects *mathLiking* and *mathValue*. Affect variables *Excited* and *Interested* are indirectly affected by gender mediating through *mathLiking*. The causal structure (gender→mathLiking→ Interested) asserts that female students like math more, which makes them more interested while using Wayang. There is no direct link gender→Interested which implies that the female students who like math as much as male students do not necessarily have any higher Interest level. But, this could be a case of multicollinearity since Correlation (mathLiking, Interest)=0.85** [4].

Math fluency training has a significant impact in improving learning gain of students as shown by the link Math Fluency (Training → learningGain.) This indicates that a group of students who got math fluency training achieved higher improvement in math tests after using the software. A paper has been published about this [3]; however, it is interesting to see the strength of this causality compared to other factors.

Pretest is only indirectly related to affect variables through attitude variables as mediators. This suggests that higher

knowledge has to be internalized as higher math self-concept and math liking to be reflected as higher enjoyment of the tutor.

In terms of affective variables, we can see the following two clusters:

Performance oriented (incoming math ability) student descriptors:

preTest, math self-concept→confidence and frustration

Students who have higher prior knowledge and better self concept in math reported higher confidence and lower frustration.

Liking and Appreciation:

Math value, math liking, perception of LC, perception of Wayang → interest and excitement

Also, students who reported gain in confidence also had higher gain in self-concept in math and those who gained in interest and excitement also ended up with higher liking for math and had greater value for math. We also observe that gender belongs to this cluster (related to mathValue, mathLiking, interest and excitement). Basically, among the four affective variables, confidence and frustration are more tightly linked with their performance and ability whereas interest and excitement are more related to their attitude and appreciation of math and tutor.

3. CONCLUSIONS

Since correlation underdetermines causality, there are multiple equivalent causal models that can be generated from the same set of data. Moreover, the stronger causal assumptions employed by this approach add more analytical power but also introduce higher chances of inaccuracy. Researchers have to, therefore, be careful about interpreting the causal relations of the model before making any causal claims. Most of the times, causal models are only able to summarize associations rather than uncover new causal mechanisms. It basically depends on whether we are able to observe all possible common causes in our data set. We do not recommend causal modeling to use as a tool to make strong causal conclusions, as we would do with randomized controlled studies. A thorough understanding of the domain is required before we make causal interpretations of the model. Causal models can only be as good as the variables that we are able to include in the modeling process. But given the data variables, causal modeling is the most superior exploratory tool available compared to the existing statistical approaches such as correlation and regression.

4. REFERENCES

[1] Pearl J. 2009. *Causality*. 2nd Edition. Cambridge University Press.

[2] Glymour, C. and Scheines, R. 2004. *Causal modeling with the TETRAD program*. Synthese. 37-64

[3] Arroyo, I., Royer, J. M., and Woolf, B.P. 2011. *Using an Intelligent Tutor and Math Fluency Training to Improve Math Performance*. IJAIED, Best of ITS 2010. Vol 21 (2), pp. 135-152. IOS Press.

[4] Rai, D. and Beck J. E. 2011. *Exploring user data from a game-like math tutor: a case study in causal modeling*. In Proceedings of the 4th International Conference on Educational Data Mining (EDM). 307-313