Doctoral Consortium: The Use of Student Confidence for Prediction & Resolving Individual Student Knowledge Structure

Charles Lang
Harvard Graduate School of Education
226 Longfellow Hall, 13 Appian Way,
Cambridge, MA, 01238
+1 (617) 501-3967
charles_lang@mail.harvard.edu

ABSTRACT

In this paper, I describe the beginnings of some research into the use of student confidence or certainty to predict student behavior and represent the structure of knowledge.

Keywords

Confidence, Bayes, certainty, online assessment, probabilistic multiple choice, partial knowledge, Systems Theory, confidence based assessment.

1. RESEARCH TOPIC

1.1 Background

The broader educational landscape is being altered by the ease with which new assessment formats can be administered through Internet-based applications. The workhorse of educational assessment, the multiple-choice question, can now be expanded and altered in ways that were not feasible even a decade ago. A popular expansion has been to collect information about what students think about their answers along with those answers; whether they think they have performed well or poorly, whether they are guessing, or how certain they are in their answer. The family of formats that utilize this strategy is large, including metacognitive assessment, certainty based assessment, and self-efficacy assessment. One common format change is to simply ask students how certain they are in a given multiple choice answer. This format, named a probabilistic multiple-choice question (PMCQ), has been of interest to educational research for at least 100 years.[1] Presently this format is being incorporated into several online assessment systems including the McGraw-Hill LearnSmart system.

Consensus is mixed as to whether the probabilistic multiple choice question adds value above and beyond the multiple choice format though. Indeed, interpretation of confidence is somewhat disputed. During the mid-1970s the PMCQ format was dismissed as flawed on the basis of experimental psychological research that had demonstrated that human beings suffered from overconfidence bias – the tendency for people to overestimate their own accuracy.[2] Furthermore, a reliable and interpretable scoring method was never agreed upon within the psychometric community despite increases in reliability.[3]

1.2 Topic

There are two aspects of Probabilistic Multiple Choice Questions that I have been pursuing. The first is whether student confidence data produces any improvement in the prediction of student performance when compared to student correct/incorrect data. The second is whether or not student confidence might provide a way of structuring representations of individual student knowledge.

2. PROPOSED CONTRIBUTIONS

2.1 Projection

With respect to the first contribution, I have preliminary data that supports the psychometric theory of [4–6]). The suggestion of which is that whether or not student confidence outperforms correct/incorrect may depend on the level at which the prediction is made.

We performed a test in which students were shown a multiple choice item, but instead of choosing a single, correct answer they reported their confidence in each of the possibilities. They were asked to do this four times for each item, but each time the item was shown two answers were removed.

In this test student confidence appeared to be better at predicting student level performance over time, but worse at predicting class level performance over time. The interpretation according to theory is that student confidence retains information peculiar to each student that is useful for predicting their individual behavior, but creates a very noisy signal when trying to predict the average behavior of the group.

Table 1. Prediction accuracy of student confidence vs. correct/incorrect at student and class level projecting first administration and second item administration.

	Confidence	Correct/Incorrect
Student Level	0.697	0.781
Class Level	0.956	0.853

$$Accuracy_i = \frac{\bar{x}_{correct} - \sqrt{(\bar{x}_{correct} - projection)^2}}{\bar{x}_{correct}}$$

2.2 Structure

If confidence is useful for predicting individual student performance we have some hope that confidence measurements may provide insight into the structure of knowledge for individual students. This makes sense at an intuitive level, if I am an expert in history I will likely be more confident in history than biology and this will be demonstrated in a test that includes both history and biology items.

But simply plotting out confidence levels seems to provide only a gross relationship and does not tell us the relationship between domains or topic or skills. For example, we can use a rudimentary social network analysis to map out items on a test according to a student's confidence in the correct answer. Edges represent the difference in confidence between different items and nodes represent items the image is iteratively resolved so that all nodes are the correct distance from each other but the structure is not necessarily meaningful:

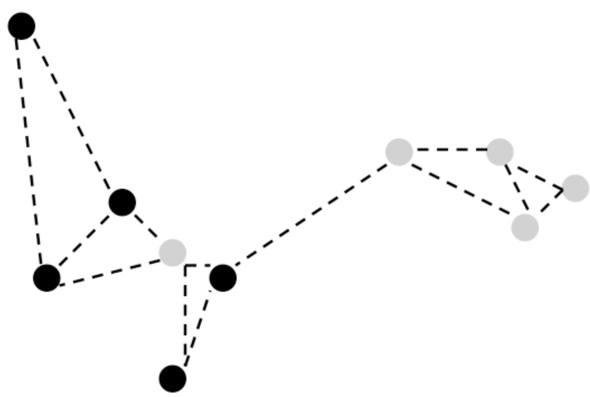

Figure 1. Social Network Analysis of one student's confidence in test items. History items are in black and biology items are in grey.

These structures seem to hint at something, but it isn't clear how to interpret the clustering. In an effort to bring structure to these diagrams I have developed an algorithm based on the Cognitive Bayesian work of Griffiths and Tennenbaum.[7]

2.3 Prediction

The fundamental idea behind applications of Bayes Theorem to people's thinking such as Decision Theory[8] and Cognitive Bayes [7,9] is to change the vantage at which it is applied. For example, instead of conditioning on the situation from the perspective of a researcher or an assessor (e.g. – the probability of the student being correct given the item) we condition on the situation from the perspective of the person being assessed (e.g. – what is her hypothesis, and on what data is she conditioning). For example, if we were studying a student as they answer the following item:

Koalas are:

A. Carnivores
B. Omnivores
C. Herbivores
D. Calmivores

We could devise a model for the way they approach each answer A, B, C & D:

$$P(koalas\ are\ herbivores) = \frac{P(koalas\ are\ herbivores).P(data|koalas\ are\ herbivores)}{P(data)}$$

In this model students weigh the likelihood of the data they have on hand against their prior beliefs, and as more data are presented, they are able to update those beliefs. For example, we might show a student pictures of koalas and every time we revealed a new picture we asked the student whether she thought the koala was a herbivore. We could model the process of the student's opinion as a Bayesian process where each new picture was a datum that changed the likelihood, generated a posterior and then that posterior became the new prior. This formalization is analogous to Snow's separation of internal and external factors: the internal factors are represented by the prior probability and the external factors are represented by the likelihood. The process whereby new data is incorporated into the prior is called Bayesian updating. Essentially, this allows us to directly account for different sources of data in a dynamic fashion, with the final iteration being the best estimate of student knowledge, accounting for external factors. The updating idea underlies features of Decision Theory and Cognitive Bayes, and is used in the classic student knowledge-tracing algorithm BKT. Where Decision Theory and Cognitive Bayes part ways though, is over the efficiency of that updating mechanism.

The Decision Theorist will assume that updating is efficient or *rational* [10] and that there is error in the individual's reporting of her posterior. Decision Theoretic questions tend to be along the lines of "Do financial analysts make rational decisions about market conditions?" The Cognitive Bayesian, however, presumes the individual can state his own posterior probability accurately, but that the incorporation of new information is rarely performed efficiently. Data may not be attended to, nor may they be wholly incorporated into a person's beliefs. A Cognitive Bayesian question tends to be drawn more from experimental psychology, asking questions such as "How do the following conditions impact peoples' prior probability in a specific task?"

The bottom line for the purposes of bringing structure to individual student confidence data is that the Cognitive Bayesian Model splits student confidence in two: the prior (what the student brought to the test inside their head) and the likelihood (the way the student is weighting new data during the test). This rudimentary but important categorization can be mapped onto the work of Snow [11] who conceived of student goal driven behavior as the interface between internal factors (cognitive, conative affective) and external factors (demand, opportunity). Ostensibly Snow's internal factors are represented by the prior probability, the posterior is the student behavior and the likelihood is how the student is mediating external factors.

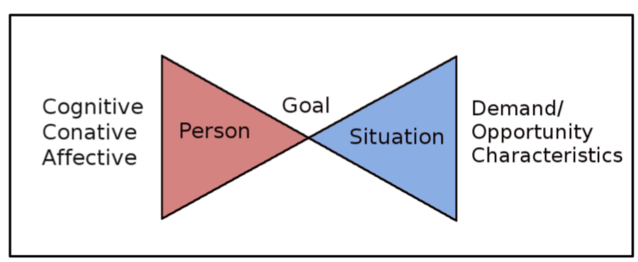

Figure 2. Snow's conception of the interface between internal (person) and external (situation) factors.

To investigate whether this algorithm is worth anything we plan to compare it to BKT and a variant of BKT developed by Wang & Heffernan[12] that has been successfully used in predicting partial knowledge (KTPC). Confidence data is currently being collected through the ASSISTments system.

Rudimentary results have been tested using Wang & Heffernan's partial knowledge data. This data is generated by scoring student performance based on how much assistance they receive (hints, trials, advice). The algorithm did not outperform KTPC in this test though partial knowledge generated in this way may be a poor proxy for confidence data.

3. Advice Sought

There are three areas I would like advice on. The first is that my background is in measurement and psychometrics. I would like to seek advice on how to adapt and change my approach and language to be appropriate for the EDM community. Second, but related, I am looking for advice on how to approach validity, in particular how to approach validity when using time series data. I can interpret the confidence data I will collect in terms of reliability, and compare the predictions of different models through correlation and standard error but I am quite adrift how this relates to a validity framework or whether it needs to?

Thanks in advance.

4. REFERENCES

[1] Williamson, G. F. Individual Differences in Belief, Measured and Expressed by Degrees of Confidence. *J. Philos. Psychol. Sci. Methods* **12,** 127–137 (1915).

[2] Langer, E. J. The illusion of control. *J. Pers. Soc. Psychol.* **32,** 311–328 (1975).

[3] Echternacht, G. The use of confidence testing in objective tests. *Rev. Educ. Res.* **42,** 217–236 (1972).

[4] Borsboom, D. *Measuring the mind : conceptual issues in contemporary psychometrics*. (Cambridge University Press, 2005).

[5] Ellis, J. & van den Wollenberg, A. Local homogeneity in latent trait models. A characterization of the homogeneous monotone irt model. *Psychometrika* **58,** 417–429 (1993).

[6] Molenaar, P. C. M. & Campbell, C. G. The New Person-Specific Paradigm in Psychology. *Curr. Dir. Psychol. Sci.* **18,** 112–117 (2009).

[7] Griffiths, T. L., Kemp, C. & Tenenbaum, J. B. in *Camb. Handb. Comput. Psychol.* (Sun, R.) 59–100 (Cambridge University Press, 2008).

[8] Schlaifer, R. & Raiffa, H. *Applied statistical decision theory*. (Division of Research, Graduate School of Business Administration, Harvard University, 1961).

[9] Tennenbaum, J. in *Adv. Neural Inf. Process. Syst.* (Solla, S. A., Leen, T. K. & Müller, K.-R.) 1098 (MIT Press, 2000).

[10] *The Probabilistic Mind: Prospects for Bayesian Cognitive Science*. (Oxford University Press, 2008).

[11] Shavelson, R. J. *et al.* Richard E. Snow's Remaking of the Concept of Aptitude and Multidimensional Test Validity: Introduction to the Special Issue. **8,** 77– (2002).

[12] Wang, Y. & Heffernan, N. Extending Knowledge Tracing to allow Partial Credit: Using Continuous versus Binary Nodes. (2011). at <http://web.cs.wpi.edu/~nth/pubs_and_grants/papers/2013/AIED2013/YuTaoContinousNodeSub.pdf>