Using Contextual Factors Analysis to Explain Transfer of Least Common Multiple Skills

Philip I. Pavlik Jr., Michael Yudelson, and Kenneth R. Koedinger

Human Computer Interaction Institute,
Carnegie Mellon University, USA
{ppavlik,yudelson,koedinger}@cs.cmu.edu

Abstract. Transfer of learning to new or different contexts has always been a chief concern of education because unlike training for a specific job, education must establish skills without knowing exactly how those skills might be called upon. Research on transfer can be difficult, because it is often superficially unclear why transfer occurs or, more frequently, does not, in a particular paradigm. While initial results with Learning Factors Transfer (LiFT) analysis (a search procedure using Performance Factors Analysis, PFA) show that more predictive models can be built by paying attention to these transfer factors [1, 2], like proceeding models such as AFM (Additive Factors Model) [3], these models rely on a Q-matrix analysis that treats skills as discrete units at transfer. Because of this discrete treatment, the models are more parsimonious, but may lose resolution on aspects of component transfer. To improve understanding of this transfer, we develop new logistic regression model variants that predict learning differences as a function of the context of learning. One advantage of these models is that they allow us to disentangle learning of transferable knowledge from the actual transfer performance episodes.

Keywords: computational models of learning, educational data mining, transfer appropriate processing.

1 Introduction

Transfer of learning is often thought to be the sine qua non goal of education, and the field has now acquired more than one hundred years of experimental research in this area [4, 5]. One finding is that transfer of learning to new contexts is often a difficult feat to achieve [6]. Because of this difficulty, which is manifest in the often frequent tendency of changes in instruction to fail to result in changes to assessed performance, research that shows transfer and helps us understand its mechanisms is highly relevant to the goals of education.

Of course, different educational research approaches propose different mechanisms for transfer. One tradition uses task analysis to assigns skills to tasks within a particular domain [1, 7-10]. One main assumption of these models is that knowledge component (KC) transfer is a unitary process. This is explicit in the structure of the models. For instance, whether we use Bayesian knowledge tracing (BKT), AFM [3] or PFA [1], they all assume that when a knowledge component transfers, it transfer as

G. Biswas et al. (Eds.): AIED 2011, LNAI 6738, pp. 256–263, 2011.

a knowledge component unit that is the same for different contexts, where those contexts differ in features hypothesized to be irrelevant to the retrieval or application of the KC. So, if our model says that a certain problem step requires a KC for both least common multiple and equivalent fractions skills, this step is very literally taken to be the sum or product (depending upon the conjunction rule, i.e. additive factors (AFM) version or conjunctive factors version, CFM) of the two *unitary* skills as they function apart from particular learning or performance contexts. In other words, every KC is functionally independent of the presumed irrelevant aspects of the context of both learning and application and exists as a latent variable. This all-or-none Q-matrix skill assignment formalism is very convenient because every future performance can be considered as a simple formula of the prior experience with the KCs.

In contrast to this parsimonious assumption that each task can be cleanly categorized as involving some list of discrete KCs, we might speculate that if two problems/steps share a KC, these two problems may cause different learning of the shared KC or may react differently to the learning of the shared KC. Indeed, this idea is not completely new since different strategies causing different degrees of transferability has been shown before [11], so our main contribution here is a formal model to describe such situations. This model implies that the learning that occurs with practice cannot be simply described according to a list of the latent variables involved in the problem, but rather additional insight is gained when we assume that each item class or skill class causes more or less accumulation of latent skill strength and is affected more or less by the accumulation from other classes of items.

In this paper we explore the hypothesis that there is more to be learned about transfer than can be inferred from current discrete skill models (but see Pardos et al. [12, 13], which have related goals). To do this, we first analyze our data with the PFA model to provide a baseline, and then provide two variant models that provide deeper reflection on the transfer effects in our data by characterizing the independent effects of skills learned in different contexts. Because our intent is to understand the data rather than optimize fit to data, we do not compare our work with other models such as BKT. No doubt, if these other models were made sensitive to context (e.g., Bayesian knowledge tracing with different learning transition probabilities for different categories of future contexts), they might achieve similar explanations of the contextual transfer.

On a practical level, this modeling allows us to meticulously relate problems to see which are the most effective for creating transfer. For example, consider the task we will analyze, least common multiple (LCM). Some of the cases, like finding the LCM of 3 and 5, can be solved by providing the product, 15 (we call these problems "product" problems, or type A). Other problems cannot be solved by the product, since the LCM of 4 and 6 is not 24 but 12 (we call these problems "LCM" problems, or type B). To a student that has not clearly learned the meaning of LCM, feedback on attempts of the two types of problems may seem ambiguous, because the product knowledge component matches the answer for many problems. Our PFA model versions will primarily examine this distinction between question types.

A detailed analysis of situations like these can be very difficult using conventional experimental methods because conventional experiments tend to produce only a few data points during learning, and often are designed to contrast overall conditions rather than trial by trial transfer in the building of a cognitive skill. On the other hand, tutoring systems in the classroom typically do not deliver the sort of randomized practice needed for many of the most interesting analyses. Because of these limitations, we created an experimental design that merged the advantages of controlled design with the advantages of tutor based classroom delivery. With the help of Carnegie Learning Inc., we did this by placing our content within the Bridge to Algebra (BTA) tutoring system at Pinecrest Academy Charter Middle School.

2 Design

The data was collected in several sixth and seventh grade classes at Pinecrest Academy Charter Middle School as integrated “Warm-up” units that would come up in the course of student’s normal use of the Bridge to Algebra Cognitive Tutor. These warm-ups were given at 10 separate points with different content across the 62 sections of BTA. The LCM warm-up had 16 single step problems chosen randomly from a set of 24 problems, of which 14 were type A and 10 were type B. Correct responses were indicated and incorrect responses were followed by a review of the correct answer, which was presented on the screen for 18 seconds. While there were 4 conditions of practice that included some additional information for some of the problems, we did not find any reliable differences due to these conditions (which included providing some direct instruction or an analogy), so we will just be reporting on the effects during practice as a function of the text of the problem the student needed to solve.

The report below covers the results for the 1st problem set of LCM problems. 197 subjects completed 16 trials with this 1st problem set, and another 58 subjects were also included from a condition that had only 8 single step problems for this 1st problem set. Problems texts are shown in Table 1.

Table 1. Examples of the fixed factors conditions. Problem numerals (as shown below) were matched across the story or no-story question types (of which there were 12 each).

Problem Example	Story Item	Product Item
What is the least common multiple 4 and 5?	no	yes
What is the least common multiple 8 and 12?	no	no
Sally visits her grandfather every 4 days and Molly visits him every 5 days. If they are visiting him together today, in how many days will they visit together again?	yes	yes
Sally visits her grandfather every 8 days and Molly visits him every 12 days. If they are visiting him together today, in how many days will they visit together again?	yes	no

3 Performance Factors Analysis

The PFA model has been presented previously, so the following description is abbreviated. PFA owes its origin to the AFM model and the Q-matrix method, since it uses a Q-matrix to assign prior item types (or KCs in the typical Q-matrix) data to predict the future performance for these same types of items. Because it uses a logic of item categories rather than KC categories, PFA has a single intercept parameter for each item type that describes in-coming knowledge of that type of item. Given this configuration, the PFA model uses logistic regression to estimate item-type category performance as a function of all item types (or KCs) that transfer according to the Q-matrix.

PFA's standard form is shown in Equation 1, where *m* is a logit value representing the accumulated learning for student *i* (ability captured by α parameter) practicing with an item type *k*. The prior learning for this item type is captured by the β parameters for each KC, and the benefit of correctness (γ) or failure (ρ) for prior practice is a function of the number of prior observations for student *i* with KC *j*, (*s* tracks the prior successes for the KC for the student and *f* tracks the prior failures for the KC for the student).

$$m(i, j \in \text{KCs}, k \in \text{Types}, s, f) = \beta_k + \sum_{j \in \text{KCs}} (\gamma_j s_{i,j} + \rho_j f_{i,j}) \tag{1}$$

Together, the inclusion of both correctness and incorrectness in the model make it sensitive to not only the quantity of each event, but also the relative ratio of correct to incorrect. Data for success and failures counts always refers to events prior to the predicted event, consistent with creating a model that is predictive for the effect of learning [1].

4 Model Versions and Transfer Implications

The following section shows how the two similar types of LCM problems result in significantly different benefits to transfer. All of these models include fixed effect assumptions about both the LCM vs. product fixed effect and a fixed effect for the story problems as compared to the explicit LCM problems.

Note that while we are not fitting any fixed (optimized) student parameters for any of the following models, we are fitting students as a random effect that is estimated according to standard random effects modeling (`lme4` package in R, `lmer` function). A random effect is any effect that is sampled from a population over which statistical inferences are to generalize. Because, we want our models to generalize across students in general, not just those sampled, our subjects qualify as random effects. Furthermore, we often have practical problems (with large numbers of students) with producing continuous parameter distributions when we fit subjects as a fixed effect. In contrast, we find that when we include no subject parameter at all (a third alternative and the approach used in a prior PFA paper [1]), parameter values found by the model tend to settle on values that track student ability rather than learning, as evidenced by negative ρ values. Indeed, for both the fixed and random effect subject models, we find tend to find larger, usually positive, ρ values. Considering the wide distribution of student abilities, we believe that adding subject variance in the model "purifies" the γ and ρ parameters (yields more interpretable estimates of these parameters) by minimizing their role in tracking student prior differences (the subject variance does this) and better focusing their role on tracking learning.

4.1 Analysis of PFA Result (Full Q-matrix)

Table 2 shows the standard PFA model with a full Q-matrix assumption ([1,1],[1,1])[1] already makes a highly interpretable if simplistic prediction that type B (LCM<product) items cause far more learning than type A (LCM=product) items. That is, the success learning rate for LCM items (γ_B=.30) is greater than for product items (γ_A=.07). We offer this initial example to contrast with the following examples. Because we have modeled subject prior learning as random effects with mean 0, learning rates for failures (ρ), as well as success (γ), are positive (unlike prior model with no subject terms).

Table 2. Standard PFA parameters found. Product (subscript A) refers to LCM problems solved by the product of the two numbers. LCM (subscript B) refers to LCM problems where the LCM is less than the product. Random effect of subject prior learning had an SD of 0.71.

Influences	Parameter	Estimate	Z-score Est.	p-value	Factor
A & B	intercept	-0.33	-3.40	0.000671	overall prior learning
A	β	1.25	14.80	<2.0E-16	prior learning prod
story items	β	-1.03	-12.94	<2.0E-16	prior learning story
A & B	γ_A	0.07	2.07	0.0381	successes product
A & B	γ_B	0.30	9.22	<2.0E-16	successes LCM
A & B	ρ_A	0.05	1.01	0.313	failures product
A & B	ρ_B	0.07	2.47	0.0136	failures LCM

4.2 Analysis of Contextual AFM (CAFM) Result

Rather than assume some particular Q-matrix, we now introduce the CAFM model that instantiates each cell in the Q-matrix with a parameter. While a normal Q-matrix assumes that each Q-matrix *column* is controlled by 1(CAFM) or 2 (CPFA) parameters, our new contextual models assign 1 or 2 parameters per *cell*. In this case, we fit CAFM since we wished more clear comparison with the prior model with the ρ failure learning rate. Since we have dropped 2 parameters and added 2 parameters, the complexity of PFA and CAFM is equivalent.

While the model fit is slightly worse with the CAFM model (see Table 5), confirming the importance of capturing success and failures separately, we find that the model parameters in Table 3 enrich our understanding of student transfer, while not disagreeing with the PFA result that type B practice is more effective. The basic pattern that is being revealed is one that might be described as transfer appropriate processing (TAP) [14]. In the case of practice with type A or type B, we see that learning effects are much weaker when transfer is measured with the other type.

However, the story is not so simple because the parameters do indicate a significant transfer effect from type B practice to type A performance. This result conflicts with the simple TAP result and shows that the story is more complex. Indeed, this B to A transfer is dramatic since type B problems transfer about 6 times better to type A than the reverse (.09/.015). If, following Pennington, Nicolich, Rahm

[1] Q matrix specification is in matrix notation, row by row. Columns of the matrix correspond to the items (or KCs) that influence items in rows. Thus, a Q matrix defined as ([,X,Y], [X,1,1], [Y,0,1]) says X is influenced by both X and Y, while Y only influences itself.

[15], we advocate the idea that failure to transfer entails rote procedural learning, and that success at transfer involves declarative conceptual learning, we might suppose that type B problems provide more conceptual practice. This is plausible because, based on the very different demands of type B (a relatively complex back checking procedure that checks factors of the product, or a relatively complex sequence of steps starting with prime factors), we could easily expect different declarative learning effects for items that require these strategies, since these strategies tend to build an organized understanding of the factor structure of the specific numbers in addition to a general understanding of factoring. In contrast, it seems that type A problems may mostly review multiplication procedures, since any deeper factor search on these problems is not immediately productive; students fail to learn a transferable understanding of common factors from these problems.

Table 3. Contextual AFM parameters found. Random effect of subject prior learning had an SD of 0.99.

Influences	Parameter	Estimate	Z-score Est.	p-value	Factor
A & B	intercept	-0.47	-4.04	5.28E-05	overall prior learning
A	β	1.44	9.61	<2.0E-16	prior learning prod
story items	β	-1.06	-13.06	<2.0E-16	prior learning story
A	γ_A	0.18	3.83	0.000131	drill count product
A	γ_B	0.09	2.59	0.00962	drill count LCM
B	γ_A	0.015	0.43	0.670	drill count product
B	γ_B	0.24	8.62	<2.0E-16	drill count LCM

Table 4. Contextual PFA parameters found. Random effect of subject prior learning had an SD of 0.74.

Influences	Parameter	Estimate	Z-score Est.	p-value	Factor
A & B	intercept	-0.33	-3.06	0.00223	overall prior learning
A	β	1.29	8.81	<2.0E-16	prior learning prod
story items	β	-1.07	-13.15	<2.0E-16	prior learning story
A	γ_A	0.16	3.02	0.00254	successes product
A	γ_B	0.08	1.85	0.0639	successes LCM
A	ρ_A	0.19	2.52	0.0117	failures product
A	ρ_B	0.07	1.76	0.0779	failures LCM
B	γ_A	0.02	0.40	0.692	successes product
B	γ_B	0.43	10.53	<2.0E-16	successes LCM
B	ρ_A	-0.02	-0.30	0.763	failures product
B	ρ_B	0.07	2.17	0.0301	failures LCM

4.3 Analysis of Contextual PFA (CPFA) Result

The preceding model suggests that context of learning matters, and this final model provides further detail by joining CAFM and PFA to create CPFA. An interesting pattern in the learning of students picked up by this model is the strong effect of failures on type A to type A performance, and the similarly relatively weak effect of failures of type B on type B performance. This is interpreted by the likely much greater ease with which students can infer the method from the type A solution feedback. Indeed, one can imagine that students quite often note that the method is multiplication for type A problems after they fail. In contrast, type B solution feedback might provide an anchor for the ambitious student to build a useful conceptual structure for future problems, but knowledge of the answer allows no easy inferences about method such as for type A problems. This suggests that students may be particularly benefitted by instructional scaffolding following failure for these harder type B LCM problems.

5 Conclusions

The Q-matrix method of assigning each latent variable (or KC) a single parameter (or 2 for PFA) and then overlaying a binary matrix that assigns KC's to items is more parsimonious than the CFA method, but lacked the ability to provide as rich an understanding of how transfer was occurring. Notably, the best fitting Q-matrix models (R2s see below) predict no transfer, while the CAFM and CPFA models both find significant (1-way test in the CPFA case with $p<0.10$) transfer for parameters describing the effect of LCM items on product item performance. See Tables 3 and 4.

While the primary purpose of this paper was to show how artificial intelligence methods can be used to understand complex hypotheses about educational transfer, Table 5 shows some aggregate fit statistics. These statistics support the idea that contextual logistic regression models improve fit only slightly, highlighting the importance of their interpretive value. Table 5 shows AFM and PFA models in 4 Q-matrix variants for comparison, only PFA-F was described in detail above.

Table 5. Comparison of the fit of the 4 model versions. R1 Q-matrix – ([,A,B],[A,1,1],[B,0,1]). R2 Q-matrix – ([,A,B],[A,1,0],[B,0,1]). R3 Q-matrix – ([,A,B],[A,1,0],[B,1,1]).

Model	Obs.	*LL*	*MAD*	*r*	*A'*
AFM-F	3616	-2075	0.411	0.346	0.705
AFM-R1	3616	-2047	0.404	0.376	0.722
AFM-R2	3616	-2042	0.402	0.377	0.721
AFM-R3	3616	-2052	0.404	0.367	0.719
CAFM	3616	-2038	0.401	0.380	0.724
PFA-F	3616	-2038	0.392	0.430	0.754
PFA-R1	3616	-2037	0.394	0.422	0.751
PFA-R2	3616	-2020	0.388	0.434	0.755
PFA-R3	3616	-2032	0.391	0.420	0.750
CPFA	3616	-2017	0.387	0.440	0.759

Of course, the CFA method requires a number of parameters that scales with the number of KCs squared, while Q-matrix methods only increase parameters as a linear function of KCs. This clearly indicates that more data is needed to successfully fit a CFA model relative to a Q-matrix model. This does not diminish the fact that when enough data is available, and it is properly balanced and randomized, contextual models such as described in this paper will likely provide better quantitative fits and enhance opportunities to discover unexpected transfer effects.

Acknowledgements. This research was supported by the U.S. Department of Education (IES-NCSER) #R305B070487 and was also made possible with the assistance and funding of Carnegie Learning Inc., the Pittsburgh Science of Learning Center, DataShop team (NSF-SBE) #0354420 and Ronald Zdrojkowski.

References

1. Pavlik Jr., P.I., Cen, H., Koedinger, K.R.: Performance Factors Analysis – a New Alternative to Knowledge Tracing. In: Dimitrova, V., Mizoguchi, R. (eds.) Proceedings of the 14th International Conference on Artificial Intelligence in Education, Brighton, England (2009)
2. Pavlik Jr., P.I., Cen, H., Koedinger, K.R.: Learning Factors Transfer Analysis: Using Learning Curve Analysis to Automatically Generate Domain Models. In: Barnes, T., Desmarais, M., Romero, C., Ventura, S. (eds.) Proceedings of the the 2nd International Conference on Educational Data Mining, Cordoba, Spain, pp. 121–130 (2009)
3. Cen, H., Koedinger, K.R., Junker, B.: Learning Factors Analysis – A General Method for Cognitive Model Evaluation and Improvement. In: Ikeda, M., Ashley, K.D., Chan, T.-W. (eds.) ITS 2006. LNCS, vol. 4053, pp. 164–175. Springer, Heidelberg (2006)
4. Thorndike, E.L., Woodworth, R.S.: The Influence of Improvement in One Mental Function Upon the Efficiency of Other Functions (I). Psychological Review 8, 247–261 (1901)
5. Judd, C.H.: Special Training and General Intelligence. Education Review 36, 28–42 (1908)
6. Wertheimer, M.: Productive Thinking (1945)
7. Koedinger, K., McLaren, B.: Developing a Pedagogical Domain Theory of Early Algebra Problem Solving. CMU-HCII Tech. Report 02-100 (2002)
8. Kieras, D.E., Meyer, D.E.: The Role of Cognitive Task Analysis in the Application of Predictive Models of Human Performance. In: Schraagen, J.M., Chipman, S.F., Shalin, V.L. (eds.) Cognitive Task Analysis. Lawrence Erlbaum Associates Publishers, Mahwah (2000)
9. Barnes, T., Stamper, J., Madhyastha, T.: Comparative Analysis of Concept Derivation Using the Q-Matrix Method and Facets (2006)
10. Barnes, T.: The Q-Matrix Method: Mining Student Response Data for Knowledge. In: American Association for Artificial Intelligence 2005 Educational Data Mining Workshop (2005)
11. Simon, H.A.: The Functional Equivalence of Problem Solving Skills. Cognitive Psychology 7, 268–288 (1975)
12. Pardos, Z., Heffernan, N.: Detecting the Learning Value of Items in a Randomized Problem Set. In: Proceedings of the 14th International Conference on Artificial Intelligence in Education. IOS Press, Brighton (2009)
13. Pardos, Z., Heffernan, N.: Determining the Significance of Item Order in Randomized Problem Sets. In: Proceedings of the 2nd International Conference on Educational Data Mining, Cordoba, Spain, pp. 111–120 (2009)
14. Morris, C.D., Bransford, J.D., Franks, J.J.: Levels of Processing Versus Transfer Appropriate Processing. Journal of Verbal Learning and Verbal Behavior 16, 519–533 (1977)
15. Pennington, N., Nicolich, R., Rahm, J.: Transfer of Training between Cognitive Subskills: Is Knowledge Use Specific? Cognitive Psychology 28, 175–224 (1995)