A Symbolic Model of Cognitive Transition

William C. Schmidt
Department of Psychology, Dalhousie University
Halifax, NS. Canada. B3H 4J1
wcs@or.psychology.dal.ca

Charles X. Ling
Department of Computer Science
Hong Kong University, Hong Kong.
ling@cs.hku.hk

Abstract

Study of cognitive development on the balance scale task has inspired a wide range of human and computational work. The task requires that children predict the outcome of placing a discrete number of weights at various distances on either side of a fulcrum. The current project examined the adequacy of the symbolic learning algorithm C4.5 as a model of cognitive transition on this task. Based on a set of novel assumptions, our C4.5 simulations were able to exhibit regularities found in the human data including orderly stage progression, U-shaped development, and the torque difference effect. Unlike previous successful models of the task, the current model used a single free parameter, is not restricted in the size of the balance scale that it can accommodate, and does not require the assumption of a highly structured output representation or a training environment biased towards weight or distance information. The model makes a number of predictions differing from those of previous computational efforts.

Introduction

The balance scale task consists of showing a child a balance scale supported by blocks so that it stays in the balanced position. Next, a discrete number of weights are placed around one of a number of evenly spaced pegs on either side of the fulcrum (see the left side of Figure 1), and it becomes the child's task to predict which arm will go down, or whether the scale will balance, once the supporting blocks are removed.

The psychological task requires the integration of information from the dimensions of weight and distance through the course of development. Perfect performance on this task can be achieved by computing torques for both the left and right arms by multiplying weight by distance, and the side with the largest torque goes down. If torques are equal, then the scale will balance.

Siegler (1981) has partitioned the set of possible balance scale problems into the six sets of distinct problem types shown in Figure 1. Performance on the different problem types is used to gage the level of expertise that children have acquired and to gain insight into the types of information that children use to solve balance scale problems.

The first three types of problems are referred to as *simple* problems because the dimension of greater magnitude determines which side of the scale will tip. The final three problem types are referred to as *conflict* problems because the cue of weight conflicts with the cue of distance, and there is no simple way of determining the outcome. The side with the greater weight or distance drops respectively in *conflict-weight* and *conflict-distance* problems, while the scale balances for *conflict-balance* problems.

Siegler (1981) reported that children's performance on the balance scale task progresses through four distinct stages. In stage 1, children use only weight information to determine if the scale will balance. In stage 2, children emphasize weight information but use distance if weights on both sides of the fulcrum are equal. In stage 3 both weight and distance information is utilized for simple problems, but children seem to respond indecisively to conflict problems. By stage 4, there is a correct integration of weight and distance information resulting in the near flawless performance of the task. Figure 1 presents the predicted percentages of correct responses, broken down by problem type, for each of these four stages.

While orderly stage progression constitutes a major regularity of balance scale development, a second regularity can be observed by examining the predicted pattern of errors in Figure 1 for conflict-weight problems. In stages 1 and 2, children answer these problems correctly because of their early reliance on weight information. In stage 3 however, when weight and distance cues are in conflict, children often perform poorly on the same problems they had previously answered correctly. This situation is rectified by stage 4 however, at which point correct answers reoccur. This trend is referred to in the developmental literature as U-shaped development, reflecting the pattern of the longitudinal plot of performance.

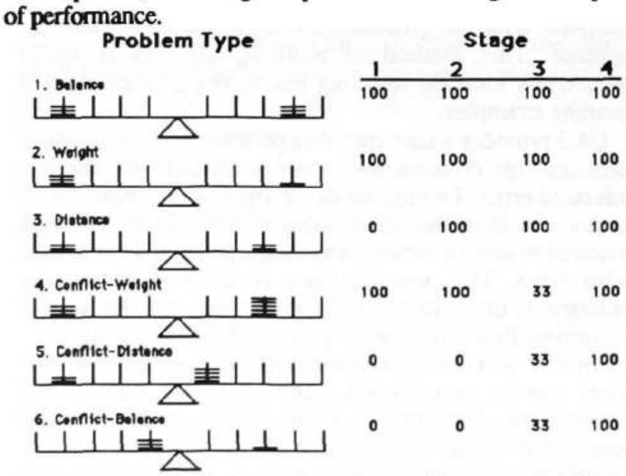

Figure 1. Predictions of percent problems correct for children using different rules.

A third major balance scale regularity was reported by Ferretti and Butterfield (1986). These researchers discovered that the rule classifications of many children systematically varied when assessed with different sets of testing problems drawn from the theoretically equivalent problem types. It was discovered that children's judgments about problems with a greater absolute difference in the amount of torque between the two arms (*torque difference*) were more often correct than similar types of problems with smaller torque differences. Therefore, Siegler's rule assessment procedure is systematically sensitive to the magnitudes of problems selected for use in stage diagnosis. This last phenomenon is dubbed the *torque difference effect* (TDE).

A number of computational models of the balance scale task exist. The most successful of these models, in terms of capturing the major developmental regularities, have utilized connectionist learning algorithms as mechanisms of cognitive transition (McClelland,1989; Shultz, Mareschal, & Schmidt, 1994; Shultz, Schmidt, Mareschal, & Buckingham, 1995). The goal of the current project was to investigate whether a popular symbolic learning algorithm could act as a transition mechanism for a successful model, in the hopes that the assumptions and predictions of such a model might provide alternative insight into the origins of the human data.

C4.5 - A Symbolic Classification System

Quinlan's (1993) C4.5 acted as a transition mechanism in our model. Given a set of training examples which vary along a set of attributes, C4.5 extracts rule-like regularity from the examples and builds a decision tree that classifies the examples with some degree of tolerated error. Like its connectionist cousins, C4.5 is a supervised learning algorithm.

C4.5 constructs a decision (sub)tree by computing the *information gain ratio* (IGR) for each of the possible attributes that could potentially be used to partition the data. The IGR is a heuristic method that evaluates an attribute's ability to reduce randomness in unclassified examples. The attribute with the greatest IGR is chosen as the root of a subtree. This method of building subtrees is applied recursively until the resulting tree fully classifies all of the training examples.

C4.5 provides a user specified parameter, m, which during decision tree construction, roughly controls the degree of tolerated error. To implement the transition component of a balance scale model, the number of cases (specified by m) required to merit a subtree branching operation was decreased with time. This manipulation resulted in the gradual emergence of an increasingly discerning decision tree. By assuming that what develops in children is an ability to assimilate more information over time, a series of decision trees can be constructed, each of which builds on its predecessor. Applying C4.5 with a large m yields smaller, less comprehensive decision trees because few attributes qualify to act as decision nodes. As m decreases, more attributes qualify to be split, more regularity in the training set is captured, and deeper and more complex trees are built.

Early in development, children have limited mental abilities. Their poor performance can be modeled with a large m value in C4.5. Performance and capacity improvement can be modeled by the gradual decrease of the m parameter. The following simulations demonstrate that the order of attributes C4.5 picks up in a series of decision trees with decreasing m coincides with the order of attributes children utilize during development, thereby demonstrating that C4.5 can provide a good model of developmental transition.

Simulation 1 - The Basic Model

Early in development, children rely more heavily on information derived from the weight dimension than the distance dimension, even though equal information from these dimensions is available. Any accurate model, therefore, requires some set of assumptions such that the transition mechanism relies more heavily on information from one dimension over the other. In McClelland's (1989) back-propagation model of this task, separate processing of the weight and distance dimensions were enforced architecturally, and the training environment gave the network more experience with weight information. These assumptions about cognitive architecture and the environment resulted in a realistic progression of the model, with weight information favored over distance information early in development. The Shultz, Mareschal, & Schmidt (1994) cascade-correlation model also required a strong environmental bias favoring weight information, but did not require the architectural assumption. The other successful connectionist model of this task, the Shultz, Schmidt, Mareschal, & Buckingham (1995) cascade-correlation model, removed the requirement for both a biased training environment and a modular separation of weight and distance processing, by biasing initial network weights such that early in development, weight but not distance information was favorably processed.

A further architectural assumption is made by all connectionist models of balance scale development, whereby their output information is encoded in a highly structured manner requiring algorithmic interpretation. Additionally, the models by McClelland (1988) and Shultz et al. (1994) required a further assumption regarding the level of training exemplar variability.

To produce results comparable to the connectionist simulations, we examined a five peg, five weight version of the balance scale task. For purposes of learning with C4.5, the set of 625 possible five peg, five weight problems needs to be represented in terms of a set of values on attributes with an associated classification. The set of attributes that yield a successful model provide at minimum, an existence proof about the types of information sufficient for producing the human data. Hence, attributes that yield a successful model, make predictions about the types of information that humans may use, or may be sensitive to, during development.

Although we experimented with a number of attribute sets, we found that few led to a successful model of the human data. Experimentation led to presenting C4.5 with seven attributes.

The first three attributes presented summary information about each problem that can be immediately derived from the visual input. Siegler's (1981) work suggested that children reason with information about which side of the balance scale has the greatest weight or distance, and whether the sides of the balance scale are equivalent for a given dimension. The first attribute concerned whether the problem presents an equal number of weights at equal distances on either side of the fulcrum, and took values of *yes* or *no*. The inclusion of this attribute was based on the perceptual salience of simple balance problems, the only problems of the set which are wholly symmetrical. The second and third attributes concerned the side of the scale with greater weight or distance respectively, and each took on one of three values: *left arm*, *neither arm*, or *right arm*. Siegler's (1981) rule models directly incorporated such information. Making this information primitively available to the learning algorithm presents it with the opportunity to capitalize on any informational value that such attributes may have for predicting problem outcomes. Because weight and distance information are equally predictive of problem outcomes, one dimension (i.e., weight) can be primarily relied upon by specifying it first. This order effect is equivalent to assuming that childrens' development internally relies on information from one dimension over the other.

The remaining four attributes were the actual number of weights and distances on either balance scale arm, and each of these was declared to be a continuous attribute taking on integer values ranging from 1 to 5. The inclusion of these attributes, again reflected that humans have such information readily accessible to them when confronted with balance scale problems.

Unlike many previous computational attempts, our model did not need to assume an explicit environmental bias favoring one input dimension. It did however, assume that simple balance problems (equal number of weights occur at equal distances on either side of the fulcrum) are particularly salient for the purposes of children's learning. This assumption was reflected in the choice of the first attribute, and by including three times as many simple balance problems as naturally occurs within the problem set, thereby giving C4.5 extra balance experience.

The C4.5 program was run, incrementally decreasing the *m* parameter which systematically resulted in the learning of increasingly complex decision trees. The gradual decrease in *m* corresponded to an assumption that the child's cognitive capacity increases in a gradual fashion yielding a processing structure in which successors build upon predecessors. Structures generated at each level of *m* were taken to represent the processing structures present for a discrete *era* of development.

Each era, the decision tree induced was used to classify the 425 examples which corresponded to the complete set of problems that could be classified into Siegler's six problem types. The responses to 24 problems (four from each of the six problem types), identical to those used to evaluate models by McClelland (1989) and Shultz et al. (1995) were then used in subsequent analyses to assess the model's success.

Results and Discussion

Figure 2 presents the stage classifications for each era of training, as diagnosed by using a procedure identical to that used with human children (Siegler, 1981). From the figure, it is apparent that the C4.5 model has captured the requirement of orderly stage progression.

Figure 3 presents the mean longitudinal performance of the simulation on the entire set of conflict-weight and conflict-distance problems. The model clearly exhibits U-shaped development on conflict-weight problems. By comparing the time of occurrence of this performance with the stage classification of the same simulation from Figure 2, it can be seen that the U-shaped developmental trend corresponds precisely with the period in which the simulation is classified at stage 3 (from approximately era 40 through era 80). The early reliance on weight information by C4.5 was interfered with during this period by the gradual integration and use of distance information on conflict problems. This can be verified by examining the longitudinal performance of the simulation on conflict-distance problems during stage 3 (Figure 3). At precisely the beginning of the decline in conflict-weight performance, distance information began to be assimilated. From Figure 3, it appears that there is a gentle vacillation between the learning algorithm's incorporation of weight and distance information with the inclusion of information from one of these dimensions conflicting with performance on the other. A negative correlation between conflict-weight and conflict-distance performance (r=-0.86; r^2=0.73) over the first 80 eras, confirms this discovery.

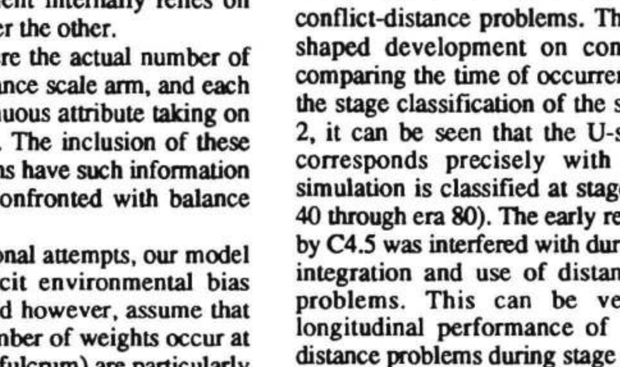

Figure 2. Longitudinal stage progression of Simulation 1 (left scale), and corresponding values of *m* (right scale).

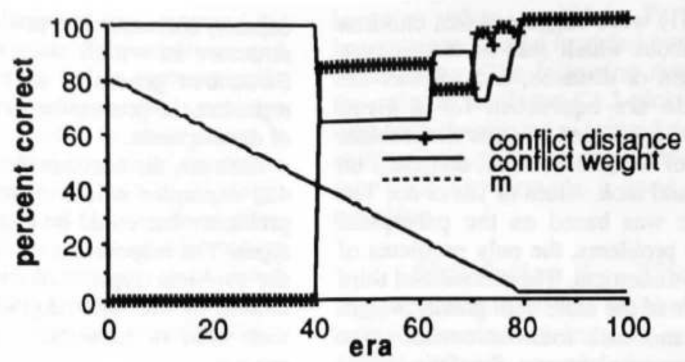

Figure 3. Performance on conflict-weight and conflict distance problems.

The final major effect characteristic of balance scale development, the TDE, was evaluated in the current simulation by classifying the model's performance using four different sets of testing patterns whose problems were drawn from four different ranges of torque difference. The TDE requires that the same set of simulation responses be classified at different stage levels depending upon the torque difference level of the testing problems used for stage evaluation. To correspond with the human data, testing sets with problems from larger torque difference levels should result in classifications at higher stages than testing sets with small torque difference problems.

Each testing set had the same balance and conflict-balance testing problems since the torque difference for these types of problems is always zero. The torque difference level for the other testing sets varied. Torque difference level 1 had problems with a torque difference of 1. Levels 2, 3 and 4 consisted of problems with torque differences in the range of 2-5, 6-9 and 10-20 respectively.

Only at stage 3, did stage classifications vary in accordance with the predictions of the TDE. Hence, the simulation was not capturing the TDE at all points in development.

An examination of the rule sets derived by C4.5 revealed that multiple rule sets mapped onto many of the stages. Stage 1 was achieved as the result of two distinct sets of rules while a single set of rules (identical to those derived by Siegler, 1981) mapped onto stage 2 performance. Stage 3 was accomplished through a set of five distinct rule systems, and two distinct decision trees resulted in stage 4 performance. No explicit computation and comparison of torques occurred. After the initial decision tree, each subsequent tree expanded upon previously derived structures.

Simulation 2 - The Expanded Model

In order to exhibit the TDE, a model must discriminate, and answer differently, problems from Siegler's theoretically equivalent problem types. If contingencies in the training data exist which distinguish problems based on information other than that used by Siegler (1981), then the TDE could arise if the learning algorithm were to pick up on such contingencies. Siegler's rule models, and our first simulation's stage 1 and 2 rules, all failed to distinguish problems with different input magnitudes. Instead, the induced rules considered only the side of the balance scale with greater weight or distance. Our model's stage 3 rules distinguished problems on the basis of their graded input levels, and its stage classifications did vary with torque difference levels.

It would appear that in order to get the TDE at all stages of development, C4.5 would be required to build rules which discriminated between problems with different levels of inputs. For our second simulation we augmented our model by changing only the representational format of the attribute specifying which side of the balance scale had greater weight or distance. Instead of classifications of left arm, right arm, or neither arm, these attributes took on values in the range of $-4 \leq x \leq 4$ (determined by subtracting the right side value from the left side value for each of the weight and distance dimensions). By doing this, we have prevented C4.5 from being able to consider the side of the balance scale with larger weight or distance information in an all or none fashion, and instead have forced it to consider the attribute in terms of a graded representation. No other conditions of the model were altered, and training and assessment were carried out as before.

Results and Discussion

An examination of the model's performance revealed that as in Simulation 1, every stage was classifiable, and orderly stage progression ensued. Longitudinal conflict-weight performance showed the characteristic U-shaped regression in performance that coincides with stage 3, however conflict-weight performance at the very earliest stage of development was slightly poorer. Nonetheless, the simulation exhibited the first two regularities required by a successful model.

To examine the model for the presence of the TDE, each era was independently assessed with four different sets of testing problems drawn from the four different torque difference intervals outlined earlier. Stage classifications varied on each of the first 77 eras. Beyond era 77, the simulation reached a saturation point, and all of the problem sets were classified at a stage 4 level of performance.

Performance on the entire set of problems in the four torque difference ranges was also examined by calculating the percentage of correct responses at the median era of each stage. This amounted to evaluating the model at eras 3, 25, 43 and 79 for stages 1, 2, 3 and 4 respectively. The mean increase in performance between torque difference levels at each of these points in development was 4%, 8%, 8%, and 4%. As dictated by the TDE, the model demonstrated superior performance on problems from larger torque difference intervals. From the results of these analyses for the TDE, it is clear that the model exhibits all of the regularities of the human data.

General Discussion

The C4.5 symbolic model was successful at capturing the three characteristic developmental findings of the balance scale task: orderly stage progression, U-shaped learning, and the TDE. The model assumed that balance problems are especially salient to children, and that the majority of children are internally biased towards processing the weight dimension over the distance dimension. In addition the model implicitly assumed that children have access to, and reason with, information about which side of the balance scale is larger for a given dimension. By implementing these assumptions and applying the C4.5 learning algorithm, these simulations provide an alternative developmental model, capable of successfully capturing many aspects of the human data.

The success of C4.5 in producing an accurate model of development demonstrates that a symbolic supervised learning algorithm can act as a mechanism for simulating cognitive transition. Like its connectionist cousins, graded representations seem to be a critical feature of the success of C4.5, as does the incrementality of the processing structures that it derives.

In contrast to the fragility of pilot work with the competing connectionist models regarding the size of the balance scale problems undertaken, the model that we report appears to be robust in this regard. The C4.5 model worked as well for smaller (4 peg, 4 weight) and larger (6 peg, 6 weight) balance scale simulations as it did for the five peg, five weight version. It is still an open empirical question whether balance scale data of other sizes can be easily accommodated using connectionist techniques.

C4.5 was also robust with respect to the format of its output encoding. While connectionist models' success hinges on the architectural assumption underlying the inclusion of a distributed encoding of two outputs (McClelland, 1989; Shultz et al, 1994; 1995), with the C4.5 model, alternative methods of representing the response yields identical results. Finally, in contrast to the vast space of possible connectionist implementations, which possess a large number of degrees of freedom and require the setting of a large number of free parameters, the C4.5 model varied only m. The C4.5 model makes a number of predictions that are different than, or opposed to, those made by previous approaches. First, while many connectionist accounts assume an environment strongly biased towards presenting information about the weight dimension (McClelland, 1989; Shultz et al., 1994), the C4.5 model predicts that the weight and distance dimensions are equally and symmetrically present in the natural world. Like Shultz et al. (1995), the current approach internalizes the early preference for information from a single dimension. If this characterization is correct, then these models suggest that despite sharing a common environment, there will be individual differences in the input dimension that children find most salient. Some support for this notion comes from related tasks requiring the integration of information from two dimensions, in which variability in the favored dimension exists (Siegler, 1981).

A second prediction of the C4.5 model is that simple balance problems are particularly salient and important in childrens' learning. Third, the model predicts that reasoning with primitive information derived from the initial presentation of the balance scale problem being solved is an important component of childrens' cognition. Fourth, the model predicts that stage 3 classifications of human performers masks a vacillation between relying more strongly on one dimension at the expense of integrating information from the other (see Figure 3).

It is our hope that the predictions derived from this alternative account will inspire further study of human development on the balance scale and related tasks, with an aim of determining the reasonableness of the assumptions that various models are based upon.

Acknowledgments

This research was supported by NSERC Canada Post Graduate funding and an Honorary Izaak Walton Killam Scholarship to William C. Schmidt. A more thorough and highly detailed examination of these simulations will appear in *Machine Learning*.

References

Ferretti, R. P., & Butterfield, E. C. (1986). Are children's rule assessment classifications invariant across instances of problem types?, *Child Development*, **57**, 1419-1428.

McClelland, J. L. (1989). Parallel distributed processing: Implications for cognition and development. In R.G.M. Morris (Ed.), *Parallel Distributed Processing: Implications for Psychology and Neurobiology*, pp. 8-45. Oxford: Clarendon Press.

Quinlan, J. R., (1993). *C4.5: Programs for machine learning*. San Mateo, California: Morgan Kaufmann.

Shultz, T. R., Mareschal, D., & Schmidt, W. C. (1994). Modeling cognitive development on balance scale phenomena. *Machine Learning*, , 57-86.

Shultz, T. R., Schmidt, W. C., Buckingham, D., & Mareschal, D. (1995). Modeling cognitive development with a generative connectionist algorithm. In T. Simon & G. Halford (Eds.) *Developing cognitive competence: New approaches to process modeling*, pp. 205-261. Hillsdale, NJ: Erlbaum.

Siegler, R. S. (1981). Developmental sequences between and within concepts. *Monographs of the Society for Research in Child Development*, **46** (Whole No. 189).