A Psychometric PDP Model of Temporal Structure in Story Recall

Richard M. Golden, Sandra F. Golden, Joseph Strickland, and Inah Choi

University of Texas at Dallas, School of Human Development, GR41, Box830688, Richardson, Texas 75083-0688

Abstract

A new parallel distributed processing (PDP) model possessing a statistical interpretation is proposed for extracting critical psychological regularities from the temporal structure of human free recall data. The model is essentially a non-linear five parameter Jordan sequential network for predicting categorical time-series data. The model consists of five parameters: an episodic strength parameter (η), a causal strength parameter (β), a shared causal/episodic strength parameter (γ), a working memory span parameter (μ), and a number of items recalled parameter (λ). The "psychological validity" of the model's parameter estimates were then evaluated with respect to the existing experimental literature using children and adult free recall data from four stories. The model's parameter estimates replicated and extended several previously known experimental findings. In particular, the model showed: (i) effects of causal structure β, (ii) showed a decrease in (η+γ) while β remained constant as retention interval increased, and (iii) an increase in (η+γ) while β remained constant as subject age increased.

Introduction

Golden and Rumelhart (1991) have suggested that story comprehension may be viewed as a process where the reader understands a story by actively constructing an appropriate cognitive representation of the text. Moreover, a considerable amount of research in the experimental psychological literature has shown that the underlying cognitive representation of a text influences how the text is recalled from memory (Fletcher & Bloom, 1988; Graesser, Robertson, Lovelace, & Swinehart, 1980; Kintsch & Van Dijk, 1978; Trabasso, Secco, & Van Den Broek, 1984; Trabasso & Van Den Broek, 1985). Thus, recall data may provide a useful "psychological" window for revealing the detailed structure of the schemata used by people in the story comprehension process.

Fletcher and Bloom (1988) described a model of story comprehension which integrated the earlier modelling work by Kintsch and Van Dijk (1978) with recent findings regarding the importance of causal knowledge in text comprehension (e.g., Fletcher & Bloom, 1988; Graesser et al., 1980; Trabasso et al., 1984; Trabasso & Van Den Broek, 1985). The basic ideas behind the early Kintsch and Van Dijk model were that: (i) texts can be modelled as collections of propositions, (ii) groups of propostions are processed according to some known order, (iii) various "selection strategies" are used to decide which groups of propositions should be maintained in working memory, and (iv) the time a proposition spends in working memory affects how effectively the proposition will be stored and retrieved from long-term memory. More recently, Golden and Rumelhart (1991) placed considerable emphasis upon the temporal dimension of story schemata. Golden and Rumelhart (1991) began by defining a "situation state space" where a point in situation state space could be identified as a long list of binary features. A story was viewed as a "partially specified" trajectory (i.e., time-ordered sequence of points) in situation state space, while story understanding was defined as constructing the "most probable" trajectory which was consistent with the reader's world knowledge.

The goal of this research is to develop a fairly simple "minimal" model of story recall whose underlying cognitive assumptions can be empirically refined and tested in considerable detail. The proposed model is a specific instantiation of the more general theory of text comprehension proposed by Golden and Rumelhart (1991), but is also closely related to the class of processing models considered by Fletcher and Bloom (1988). The proposed model is essentially a parallel distributed processing (PDP) model with a statistical interpretation (see Golden, 1988, and White, 1989, for reviews of this type of interpretation), and may be briefly summarized as a five parameter Jordan sequential network (Jordan, 1992) for predicting categorical time series data. Alternatively the model may be for-

* We are grateful to the principals and teachers of the following California elementary schools for their continuing support: Cajon Park, Santee School District; Lindo Park, Lakeside School District; Naranca, Cajon Valley School District. We also wish to acknowledge assistance provided by the School of Education and Human Services, National University, San Diego, California.

The first author would also like to acknowledge communications with Dr. David Rumelhart regarding the use of neural networks with probabilistic interpretations which took place while the first author was an NIH post-doctoral fellow at Stanford University.

mally viewed as a non-linear constrained time-series path analysis statistical model of the temporal structure in free recall data. Although a long-term goal of this research is to develop a more realistic model of story comprehension and free recall *processes*, the model in its current form can directly aid researchers who wish to empirically evaluate and compare alternative detailed *representational assumptions* of their own theories of human knowledge and text representation.

Modelling Assumptions

Following Golden and Rumelhart (1991), consider the "situation state space" for a simplified version of the story *Jack and Jill* which is specified by the following feature table (Table 1).

Table 1. Example Feature Table for a Simplified "Jack and Jill" Story (d=3).

Text Feature Category	Sentence Fragment of Original Story
1. Go (J&J, hill)	Jack and Jill went up a hill
2. Desire (J&J,water)	to get a drink of water.
3. Ingest (Jack,water)	Jack had a drink of water.

Thus, the story is specified by the ordered set of text features: $\{f_1, f_2, f_3\}$.Suppose a subject recalls the story *Jack and Jill* as follows: *Jack and Jill wanted a drink of water so they went up a hill. Jack had a drink of water.* This recall protocol is represented using the situation state space coding scheme in Table 1 as the ordered set of features: $\{f_2, f_1, f_3\}$.

In addition to specifying a situation state space, it is also necessary to specify critical pair-wise relationships among the text feature elements following the work of Graesser et al. (1980) and Trabasso et al. (1984). In this paper only three types of relationships are considered but the model can be extended to include additional types of relationships. A *pure episodic relationship* is an ordered pair (f_j, f_i) of two text features such that: (i) text feature, f_i, immediately precedes text feature,f_j, in the original text which was presented to the subject, and (ii) text feature f_j is not a causal consequence of text feature f_i within the story's context. A *pure causal relationship* is an ordered pair (f_j, f_i) of two text features such that: (i) text feature, f_i, does not immediately precede text feature, f_j, in the original text which was presented to the subject, and (ii) text feature, f_j, is a causal consequence of text feature, f_i, within the story's context. Finally, a *shared causal/episodic relationship* is an ordered pair, (f_j, f_i), of two text features such that: (i) text feature, f_i, immediately precedes text feature,f_j, in the original text which was presented to the subject, and (ii) text feature, f_j, is a causal consequence of text featuref_i within the story's context.

The cognitive architecture which will be used to model the free recall process is based upon a type of *constrained categorical* Jordan sequential network (Jordan, 1992) which has *d* input units, *d* hidden units, and *d* output units where *d* is the dimensionality of the situation state space. Given the sequence of previous items recalled by the model, the network attempts to construct an activation pattern over the output units which indicates what item should be recalled next. Figure 1 shows the basic network architecture applied to the situation state space of Table 1 where *d=3*. The state of each unit is a real-valued number which is referred to as the unit's activation value. An "activation pattern" is the set of activation values associated with a particular group of units. A "situation activation pattern" indicates a situation where only the *i*th feature is active by a set of *d* activation values where the *i*th activation value is equal to one and the other activation values are set equal to zero. For example, feature f_2 in Table 1 is represented as a situation activation pattern by the *3*-dimensional vector: $[0\ \ 1\ \ 0]$.

The input unit activation pattern of the constrained categorical sequential Jordan network depicted in Figure 1 indicates the contents of the model's *working memory buffer*, $y(t)$, at the current instant in time. In particular,

$$y(t) = x(t-1) + \mu y(t-1)$$

where $0 < \mu < 1$. The vector $x(t)$ is a situation activation pattern which represents the *t*th item which was recalled by the model. If the *working memory span* constant μ is large, the model has difficulty discriminating past items which were recalled. Moreover, in this case the model's recall of the current text feature is functionally dependent upon the last group of items recalled by the model. On the other hand, if μ is small, the model's recall of the current text feature is only functionally dependent upon the last text feature recalled. The presence of a working memory span constant such as μ is not inconsistent with recent theories of text comprehension and retrieval processes (Fletcher & Bloom, 1988; Kintsch & Van Dijk, 1978).

The connections from the hidden unit layer to the output unit layer are fixed and not modifiable. These connections are chosen so that the ith output unit activation value may be interpreted as the probability that text feature i is recalled. This "response competition" assumption (see Murdock, 1974, pp. 82-88, for a discussion of the experimental evidence supporting this assumption) is formally instantiated by the following formula:

$$p_i(t) = exp(h_i(t)) / \sum_{j=1}^{d} exp(h_j(t))$$

where $h_i(t)$ is the activation of the ith hidden unit at time t and $p_i(t)$ is the activation of the ith output unit at time t.

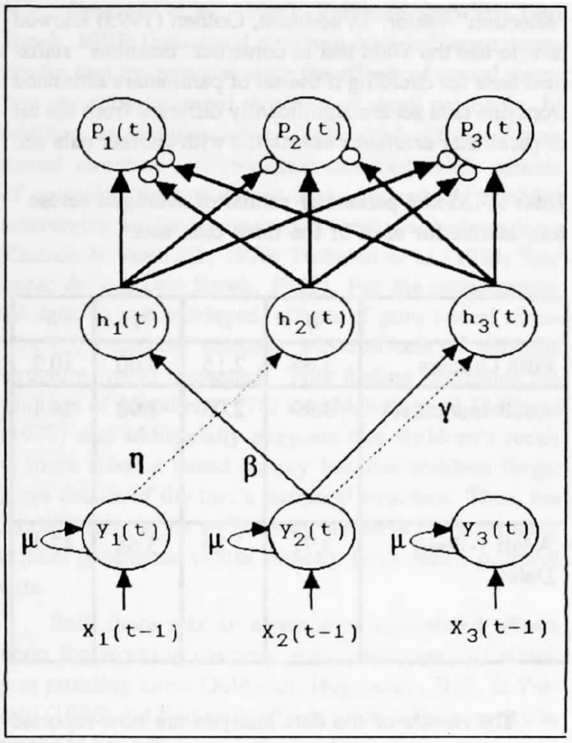

Figure 1. The constrained categorical Jordan sequential network for a text with *d=3* (see text for additional details).

The connections from the input unit layer to the hidden unit layer are now considered. These connections are modifiable, but their values are also highly constrained. Each "connection strength" or "weight" from the input unit layer to the hidden unit layer can be only one of three types: (i) "pure" episodic which is represented by a single dashed arrow, (ii) "pure" causal which is represented by a single solid arrow, and (iii) "shared" causal/episodic which is represented by both a dashed and solid arrow. All pure episodic connections have exactly the same connection strength value which is referred to as η. All pure causal connections have exactly the same connection strength value which is referred to as β. And all shared causal/episodic connections have exactly the same connection strength value which is referred to as γ. Thus, the values of all connections from the input unit layer to the hidden unit layer are identified if the free parameters η, β, and γ are known. Figure 1 shows how the causal and episodic connections for the situation state space in Table 1 are instantiated in the model. Formally, these assumptions are instantiated by the following equation:

$$h_i(t) = \sum_{j=1}^{d} [\eta e_{ij} + \beta b_{ij} + \gamma g_{ij}] y_j(t)$$

where $e_{ij} = 1$ indicates that episodic relationship (f_i, f_j) is present and $e_{ij} = 0$ indicates the episodic relationship, (f_i, f_j), is absent. The other constant digraph parameters b_{ij} and g_{ij} are specified in a similar manner for the pure causal and shared causal/episodic relationships. Thus, the constants e_{ij}, b_{ij}, g_{ij} precisely specify specific assumptions about the underlying causal chain representation of the story.

The model thus far indicates how the subject's previously recalled text features will influence how the next text feature will be recalled in terms of the four free parameters: β, η, γ, and μ. It is assumed that the probability that the subject recalls M text features from memory is given by a Poisson distribution whose mode and mean are identified by the fifth free parameter λ. Thus, λ is referred to as the parameter indicating the expected number of items recalled.

Application of the PDP Model to Recall Data Analysis

The proposed connectionist model was used to analyze children and adult free recall data. The purpose of this experiment was to see if estimated model parameters were "psychologically consistent" with respect to the existing experimental literature. There were several experimental questions which were of particular interest in this pilot test of the psychometric PDP model's performance. First, are there effects of the causal structure parameter (β)? Second, as retention interval increases, does the episodic component (η) of the temporal structure of the free recall protocol decrease while the causal component (β) increases or remains constant? Third, as age increases, does the episodic component (η) of the temporal structure of the free recall protocol increase while the causal component (β) decreases or remains constant? Fourth, as age increases, does the working memory span parameter (μ) increase in value indicating an increased working memory span?

Stimuli

The four stories which Trabasso et al. (1984) used to construct causal chain representations and evaluate children's recall performance were used. The dimensionality, d, associated with each story ranged from 22 to 25. The causal digraph representations used to decide which text features were causally related were empirically derived based upon rating data collected from college students but the results reported below are similar when causal digraphs based upon the original Trabasso et al. (1984) analyses are used.

Procedure

Two fifth grade classes of children (n=52) each read one of the four stories, and were then asked to write their recall of the four stories from memory. In addition, college students (n=24) read and recalled two of the same four stories in both an immediate recall condition and a one-week delayed recall condition.

Data analysis and results

The recall data was coded by two coders and an acceptable intercoder-reliability ($\kappa = 0.8$) was obtained.

Model selection. The model was fit to both the children and adult recall data using five different values for the working memory span parameter μ: $\mu_1 = 0/8, \mu_2 = 1/8, \mu_3 = 2/8, \mu_4 = 3/8, \mu_5 = 4/8$. For the adult immediate and delayed recall data, the "best-fitting" value of μ, μ_3=0.25 was significantly different at the experiment-wise significance level $\alpha_e = 0.05/4 = 0.0125$ from μ_1, μ_2, and μ_5. The difference in values between μ_3 and μ_4 was only marginally significant (p = 0.03) at the preset experiment-wise significance level.

For the children recall data, the "best-fitting" value of μ, μ_3=0.25 was significantly different at the experiment-wise significance level $\alpha_e = 0.05/4 = 0.0125$ from only μ_5.

Data analysis. The five model parameters were then estimated for each of the four stories for each of the three data sets (children, adult immediate recall, and adult delayed recall) using the "best-fitting" value of μ, μ_3=0.25. The average parameter estimates across stories which resulted from "learning" the recall data for each of the three groups are shown in Table 2. As previously noted, η refers to the link strength of the episodic links, β is the link strength of causal links, γ is the link strength of links which are both causal and episodic, and λ is the average number of items recalled per story.

Let $\hat{w}$ be a 12-dimensional vector containing the parameter estimates of some unobservable parameter vector w whose ith element is w_i in Table 2. Golden (1993) proved that under fairly general conditions, the parameter vector $\hat{w}$ is a quasi-maximum likelihood estimate which converges to a unique parameter vector as the number of subjects increases. Golden (1993) used White's (1989) asymptotic statistical theory to derive a formula for the variance, c_{ii}, associated with the ith parameter estimate and the covariance, c_{ij}, associated with the ith and jth parameter estimates. Golden (1993) showed how to use the Wald test involving the Wald statistic $\mathcal{W}$ recommended by White (1989) to test null hypotheses of the form: $H_o : \sum_i s_i w_i = 0$ where the coefficients s_i are the coefficients of a "contrast" or "selection" vector. In addition, Golden (1993) showed how to use the Wald test to construct "omnibus" statistical tests for deciding if the set of parameters estimated from one data set are significantly different from the set of parameter estimates associated with another data set.

Table 2. Model parameter estimates averaged across four stories for each of the three data sets.

	η	β	γ	λ
Fifth Graders	3.65	2.15	3.60	10.2
Adult Immediate	3.99	2.05	3.98	14.1
Adult 1-Week Delay	3.71	1.84	3.68	13.1

The results of the data analysis are now reported using the statistical tests derived by Golden (1993). The first of two planned omnibus comparisons between the adult immediate recall condition and adult delayed recall condition was significant ($\mathcal{W}(4) = 11.9, p < 0.02$). Post-hoc analyses of the data indicated that these differences were most likely due to effects of only the episodic digraph ($[\eta + \gamma]/2)(\mathcal{W}(1) = 7.26, p < .01$) and an effect of the number of items recalled (λ) ($\mathcal{W}(1) = 5.73, p < .02$). Additional post-hoc analyses showed that the causal strength parameter (β) was significantly different from zero ($W(1) = 765, p < 0.001$) as was the episodic strength parameter (η) ($W(1) = 3099, p < 0.001$).

The second omnibus comparison between the children recall data and the adult immediate recall data was highly significant ($\mathcal{W}(5) = 73.18, p < 0.001$). Post-hoc analyses of the data indicated that these differences were due to an effect of number of items recalled (λ) ($\mathcal{W}(1) = 67.8, p < 0.001$), and an effect of episodic structure ($[\eta+\gamma]/2)(\mathcal{W}(1) = 8.32, p < .005$). Additional post-hoc analyses showed that the causal strength parameter (β) was significantly different from zero ($W(1) = 546, p < 0.001$) as was the episodic strength parameter (η) ($W(1) = 2507, p < 0.001$).

Discussion

The positive contribution of the causal strength parameter (β) replicates and extends previous findings (e.g., Graesser et al., 1980; Trabasso and Van Den Broek, 1985) that causal structure does influence memory for text by demonstrating the effects of causal structure upon the temporal structure of recall protocols. In addition, as retention interval increased, effects of pure causal structure (β) remained constant while effects of episodic structure ($\eta+\gamma$) decreased which provides converging evidence supporting previous findings (e.g., Kintsch & Van Dijk, 1978; Trabasso et al., 1984; Trabasso & Van Den Broek, 1985). For the developmental data, as age increased, effects of pure causal structure (β) remained constant while effects of episodic structure ($\eta+\gamma$) increased. This finding replicates the findings of Mandler (1978) and Mandler and DeForest (1979) and additionally suggests that children's recall is more schema based simply because children forget more details of the text's temporal structure. Thus, the model seems to be sufficiently sensitive for estimating critical parameter values directly from recall protocol data.

Still, there was an absence of age-related effects upon the working memory span parameter (μ) which was puzzling since Goldman, Hogaboam, Bell, & Perfetti (1980) did find such effects using a memory probe task. On the other hand, the working memory span parameter did appear to be text-dependent which is consistent with Fletcher's (1986) post-hoc data analyses. Finally, the small number of texts and the developmental comparison between fifth graders and adults as opposed to second and fifth graders suggest that the above experimental findings should be cautiously interpreted.

References

Fletcher, C. R. 1986. Strategies for the allocation of short-term memory during comprehension. *Journal of Memory and Language, 25:* 43-58.

Fletcher, C. R. & Bloom, C. P. 1988. Causal reasoning in the comprehension of simple narrative texts. *Journal of Memory and Language, 27:* 235-244.

Golden, R. M. 1993. *Making correct statistical inferences using a wrong probability model with a PDP categorical time series application.* School of Human Development, University of Texas at Dallas.

Golden, R. M. 1988. A unified framework for connectionist systems. *Biological Cybernetics, 59:* 109-120.

Golden, R. M. & Rumelhart, D. E. 1991. A distributed representation and model for story comprehension and recall. In *Proceedings of the Thirteenth Annual Conference of the Cognitive Science Society,*7-12. Hillsdale, NJ: Erlbaum.

Goldman, S. R., Hogaboam, N. L., Bell, L. C., & Perfetti, C. 1980. Short-term retention of discourse during reading. *Journal of Educational Psychology, 72:* 647-655.

Graesser, A. C., Robertson, S. P., Lovelace, E. R., & Swinehart, D. M. 1980. Answers to why-questions expose the organization of story plot and predict recall of actions. *Journal of Verbal Learning and Verbal Behavior, 19:* 110-119.

Jordan, M. I. 1992. Constrained supervised learning. *Journal of Mathematical Psychology, 36:* 396-425.

Kintsch, W., & Van Dijk, T. A. 1978. Toward a model of text comprehension and production. *Psychological Review, 85:* 363-394.

Mandler, J. M. 1978. A code in the node: The use of a story schema in retrieval. *Discourse Processes, 1:* 14-35.

Mandler, J. M. & DeForest, M. 1979. Is there more than one way to recall a story? *Child Development, 50:* 886-889.

Murdock, B. M. 1974. *Human memory: Theory and data.* Potomac, Maryland: Erlbaum.

Trabasso, T. Secco, T., & Van Den Broek, P. 1984. Causal cohesion and story coherence. In H. Mandl, N. L. Stein, & T. Trabasso (Eds.), *Learning and comprehension of texts.* Hillsdale, NJ: Erlbaum.

Trabasso, T. & Van Den Broek, P. (1985). Causal thinking and the representation of narrative events. *Journal of Memory and Language, 24,* 612-630.

White, H. (1989). Learning in artificial neural networks: A statistical perspective. *Neural Computation, 1,* 425-464.