Large-Scale Modeling of Lexical Processes

Daragh E. Sibley
George Mason University
4400 University Drive
Fairfax, Virginia 22030
1-703-993-1207
Dsibley@gmu.edu

Christopher T. Kello
George Mason University
4400 University Drive
Fairfax, Virginia 22030
1-703-993-1207
Ckello@gmu.edu

Andrew Colombi
University of Illinois at Urbana-Champaign
201 N Goodwin Ave
Champaign, Illinois 61820
Colombi@uiuc.edu

Abstract

For about the last fifteen years, computational models have played a major role in developing and testing cognitive theories of language processing. The models have now reached the point where large-scale simulations are the next logical step. In the current study, a large-scale neural network model of word reading and word recognition is presented that consists of nearly thirty thousand processing nodes and nearly twenty million links among the nodes. The model is designed to account for behavioral data on tens of thousands of words and pseudowords. The data consist of reaction times and error rates in word naming and lexical decision tasks. Most of the model's links are associated with weight parameters that were learned gradually via gradient descent (i.e., back-propagation). Due to the size of the model, weight learning required the use of parallel computation on the NSF Teragrid. Preliminary comparisons between model and empirical data are presented.

Categories and Subject Descriptors

J.4 [**Social and Behavioral Sciences**]: *Psychology.*

General Terms

Measurement, Performance, Experimentation, Theory

Keywords

Large-scale connectionist modeling, computational cognitive modeling, artificial neural networks, lexical processing, word reading

1. Introduction

Cognitive processes such as perception, attention, memory, and language develop and change over the human lifespan, and operate roughly on the time scales of milliseconds to seconds. Cognitive scientists increasingly use computational models to aid in formulating and testing theories of cognitive processes as they develop, change, and operate. One of the most mature areas of computational cognitive modeling is lexical processing, i.e., the processes by which words are learned, represented, and used in spoken and written language [2, 4, 5, 6, 7, 8, 10]. These models typically take strings of sounds (e.g., phonemes like the 'b' sound in 'bat') or letters as inputs. The input strings are processed in order to activate stored lexical knowledge in the service of some language task. The two tasks used most commonly in both empirical and computational studies of lexical processing are lexical decision (i.e., determine as quickly as possible whether a given string of sounds or letters corresponds to a word in your language) and word naming (i.e., pronounce a letter string as soon as it is presented on a computer screen). The theories and models are typically held accountable for mean reaction times and error rates in these tasks, averaged over a set of participants in a given experiment.

Current models of lexical processing account for performance on thousands of monosyllabic words and pseudowords (i.e., pronounceable nonwords like BRECK). However, the lexicons of most languages (including English) are comprised of tens of thousands of words, most of them multisyllabic. The fact that current models cannot address the full breadth and scale of real lexicons severely limits their use and validity as tools for theory building and testing. These models are limited to monosyllabic words for three main reasons: 1) most empirical data is limited to monosyllabic words; 2) there are theoretical difficulties in representing the heterogeneity of structure that spans mono- and multisyllabic words, and 3) computational power has historically limited the size of lexical processing models to thousands rather than tens of thousands of words.

However, due to recent advancements in the field, the reasons for restricting the scale of lexical processing models are no longer applicable. First, several researchers across multiple universities have teamed up to create the eLexicon database [1] that includes over 1 million measurements of word naming performance, and over 2.5 million measurements of lexical decision performance. Thus models can now be tested against data for tens of thousands of mono- and multisyllabic words. Second, my colleagues and I recently developed a computational method that overcomes the theoretical difficulties in representing mono- and multisyllabic words [11]. Third, the NSF Teragrid was recently made available to the US-based research community. The Teragrid is a high-performance computing platform that can support very large-scale parallel computing applications, and the code for many models of lexical processing (neural network models in particular) can be easily parallelized. Thus it is now both theoretically possible and technically feasible to build large-scale models of lexical processing.

In the present study, dissertation work is presented on building such a model. It includes 28,032 of the 40,481 words in the eLexicon database, and can also process tens of thousands of pseudowords that are relevant to testing theories of lexical processing (but not in the database). The model has two main components, one that learns to recognize and represent English letter strings (analogous to areas of processing in visual cortex that develop during the course of reading acquisition), and another that links these orthographic representations with lexical nodes and their corresponding pronunciations.

2. Model Implementation

The orthographic component of the model consists of two simple recurrent networks [3] that are coupled by a "wordform" level of representation. The model learned wordform representations by taking letter strings that corresponded to English words (e.g., B, A, then T) as input sequences, holding them in memory, and then reproducing the strings as output sequences. The representations that held each sequence in memory served as orthographic wordform representations. The orthographic component consisted of 2559 processing nodes and 1,029,029 connections among the nodes, each one associated with a learned connection weight (for implementation details, see [11]).

The lexical component of the model was centered around 28,032 processing nodes, each one representing one word in the eLexicon database. A set of orthographic input nodes was connected to the word nodes according to the learned orthographic wordform representations. A given string of letters was processed by feeding it through the orthographic component to generate a wordform representation, and then feeding this representation to the orthographic input nodes to generate a pattern of activation over the word nodes. Word nodes were more activated to the extent that their corresponding wordforms were more similar to the input wordform representation. Finally, the word node activation pattern served as input to a simple recurrent network that learned to produce the corresponding sequence of phonemes (i.e., in order to simulation the word naming task). Including the word nodes, the lexical component consisted of 29,777 processing nodes and 1,999,444 connections among the nodes.

Learning the connection weights in order to output phoneme sequences was computationally the most demanding part of the model, and was estimated to require years of computer time if simulated using a single CPU. Fortunately, the backpropagation algorithm [9] used to learn these weights can be parallelized using a simple, synchronous client-server model that can be divided into four stages: 1) The server broadcasts the current connection weights to the clients; 2) each client processes a sample of the words to be learned; 3) each client sends the resulting weight changes back to the server, and 4) the server sums all the client feedback to update the current connection weights. When using N clients (where N was up to 100, depending on the simulation run), this parallelization scheme reduced the time needed to learn connection weights by nearly a factor of N.

3. Model Evaluation

The model was used to simulate both lexical decision and word naming tasks, where lexical decision performance was assessed by the distributional properties of the word node activation pattern, and word naming performance was assessed by the output phoneme sequence. Model performance was compared with mean reaction times and error rates for the respective tasks in the eLexicon database. Preliminary results show that the model can account for a significant amount of variance in the mono- and multisyllabic empirical data. These results go well beyond previous smaller-scale models that could only account for monosyllabic data.

4. References

[1] Balota, A. A., Cortese, M. J., Hutchison, K. A., Neely, J. H., Nelson, D., Simpson, G. B., et al. 2002. The English lexicon project: A web-based repository of descriptive and behavioral measures for 40,481 English words and nonwords. http://elixicon.wustl.edu:Washington University.

[2] Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. 2001. DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review, 108*, 204-256.

[3] Elman, J. L. 1990. Finding structure in time. *Cognitive Science, 14*, 179-211.

[4] Harm, M., & Seidenberg, M.S. 1999. Reading acquisition, phonology, and dyslexia: Insights from a connectionist model. *Psychological Review*, 106, 491-528.

[5] Kello, C. T. 2006. Considering the junction model of lexical processing. In S. Andrews (Ed.), *From Inkmarks to Ideas: Current Issues in Lexical Processing*. New York: Taylor and Francis

[6] McClelland, J. L., & Rumelhart, D. E. 1981. An interactive activation model of context effects in letter perception, Part 1: An account of basic findings. *Psychological Review*, 88, 375-405.

[7] Perry, C., Ziegler, J. C., & Zorzi, M. 2007. Nested incremental modeling in the development of computational theories: The CDP+Model of reading aloud. *Psychological Review*, 114(2), 273-315.

[8] Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. 1996. Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review*, 103, 56-115.

[9] Rumelhart, D. E., Hinton, G. E., & Williams, R. J. 1986. Learning Representations by Backpropagating Errors. *Nature, 323*(6088), 533-536.

[10] Seidenberg, M. & McClelland, J. 1989. A distributed developmental model of word recognition and naming. *Psychological Review*, 96, 523-568.

[11] Sibley, D. E., & Kello, C. T. 2006. Learning representations of orthographic word forms. In the *Proceedings of the 28th Annual Conference of the Cognitive Science Society.*