Connectionist Model and Linguistics:

Model for Word Recognition

Lavina Sabhnani

University of North Carolina at Charlotte

Abstract

Connectionism is an extremely broad term that refers to a variety of ideas that approach cognition in an artificial intelligence aspect by creating numerous connections between nodes at different grain sizes or scales ranging from neurons to hidden behaviours that are results of large networks operating in parallel, broken down into simple processing units, grouped into hierarchies generating outcomes that gives a comprehensible way about of human cognition. There are many connectionist models specifically trained to recognize letter strings and compute its meaning have been designed over the years like the triangle framework by Seidenberg McCelland (1989), Plaut et al. (1996), and Harm and Seidenberg (1999, 2004). We will be reviewing the basic components that Connectionist models of Reading comprises of, triangle framework and the TRACE model developed by McCelland and Elman (1986) to understand the lexical access on them, their architecture and shortcomings of these models.

Keywords: Connectionist models, triangle framework, TRACE model.

Connectionist Model and Linguistics:

Model for Word Recognition

Reading is a complex skill that is a prerequisite to success in our society where a great deal of information is communicated in written form. Reading is a highly complex task, involving the rapid coordination of visual, phonological, semantic and linguistic processes Reading is also a process that has attracted the attention of many cognitive scientists because many fundamental cognitive processes are involved in reading. Questions that have been the central focus of cognitive scientists to understand reading that still are highly relevant and are the focus of a considerable amount of current empirical research are: how are written words identified, how does the system of oral language interact with word identification and reading, are words identified in text differently than in isolation. Word Recognition is the ability of a reader to recognize written words correctly and virtually effortlessly. Computational models facilitate reading by enabling research into the effects of specific hypotheses with respect to the representations and processes underlying reading acquisition. Connectionist or neural network modelling, a specific form of computational modelling, is even more beneficial, as it clearly depicts how such mechanisms might be implemented in the brain. Connectionist models extensively used for word recognition are Triangle Model and the TRACE Model which have been further studied to understand Lexical access**.**

Introduction to Connectionist Model

The core properties of neural computation are fairly estimated by a set of computational principles which are represented by Connectionist models. These models consist of a large network of simple neuron-like processing elements that learn to perform tasks such as reading words or recognizing objects. There are many different connectionists models developed over the years but they all abstract to similar framework that consists of Task Orientation, Processing, Distributed and Localist Representations, Learning, Hidden Units, Experience, and Network Architecture.

Task Orientation

Pronunciations and meanings of the input words are estimated. Main purpose of task orientation is to explain the theory regarding how the task will be carried out relating in contexts with cognitive, perceptual and learning capabilities. Task orientation helps understand how words are processed, understood by readers and with that knowledge, the model is trained to the input words so as to how will it act in respect to it. Reading models are constructed with different representations of lexical codes along with learning capacity. Lexical code is the syntax of programming language which is analyzed using a lexer and parser. Lexer is a program that performs lexical analysis - a sequence of characters converted into a sequence of tokens, and a parser analyzes this string of tokens. These lexical codes can be represented in orthographic form or phonological form or semantic form. In practice, performance of model was greatly affected by input and output properties. The focus is on these representations to exact the capabilities of the beginning reader's knowledge through their development. But these representations can be approached only in an implemented model. This implemented model worked by approximating the knowledge gained, limitations show up as deviations between the performance of model and human. For example, the early model by Seidenberg and McClelland (1989) the capacity to support generalization was limited by the model's phonological representation like pronunciation of non-words such as jinje. This limitation was overcome by Plaut et al. (1996) and harm and Seidenberg (1999) by addition of theoretical insights like the predictability of readers in phonological representations. The models later developed focused on these limitations simplifying these issues.

Processing

The input provided to the model is first computed by integrating all the available information, then by calculating the net input. Interactions of different such individual units helps calculate a pattern which helps further processing of the given input. Attractor is an important aspect of how the patterns evolve over the course of processing the given input, such that at any given instant, the pattern over a network is represented in terms of coordinates of a point in multi-dimensional state space, where each unit has a dimension. A degree of robustness to partially missing or noisy input or to the effects of damage is obtained by these attractor patterns.

Distributed and Localist Representations

The nature of the representations that are involved in network's response to the input is the way that inputs, outputs, and group of intermediate units that encode information with regards to pattern generation of activity. Distributed representations are ones where information is represented by a finite set of units, with each unit participating in many patterns. A model may have units that correspond to phonemes or phonetic features and these units are activated for all the words that contain that sound. Distributed representations are more difficult to understand and use but give better explanation about cognitive phenomena's by giving rise to unanticipated emergent properties. Patterns are assigned to entities such that similarities between the fundamental functional relationships amongst them are apprehended. All the inputs and outputs for each of these patterns are interpreted directly on the basis of these similarities. But for the tasks that are simple, extra intermediate units are required to interpret the similarities and it isn't reasonable to specify appropriate connection weights for these units by hand. Thus, learning is mostly used to determine these similarities. Localist representations are the ones where units correspond to higher-order entities such as word. They are easy to manipulate but too much flexibility is permitted to constraint theorizing sufficiently like in Plaut and McClelland (2000).

Harm and Seidenberg's (2004) model used localist representations of letters, phonetic and semantic features but distributed representation of the spellings, sounds and meaning of words. In McClelland and Rumelhart's model, localist representation is used at letter level, each unit corresponding to a letter and distributed representation is used with respect to words where each word corresponds to many letters. The contrast in these models isn't the use between localist versus distributed representations but the models that are committed to claim that there are localist representations of words and models for which there are no representations at this level. Localist representations are used by some connectionist models where individual units stand for known entities such as letters, words, concepts, and propositions, for example, localist model are used for early exploration of phenomena. Distributed representations are used by other connectionist models where entities are represented by a particular pattern of activity over numerous units than activity of a single unit.

Learning

A network is comprised of units linked to each other and when activated, it spreads along all the units. The linkages between them are weighted that helps determine how much activation is passed through the network. Learning is adjusting these weights on the basis of experience where a set of weights is found for a model to perform the task accurate and efficiently. Back propagation is used by reading models where the output produced by the models for a word is compared to the correct, target pattern where small discrepancy adjustments are made by adjusting weights efficiently. It suggests that when a person generates a response that is corrected by an omniscient teacher only then learning takes place. This was formulated by understanding basis in child's experience for the teaching signal provided by the learning algorithm, for example, learning to read. But in Harm and Seidenberg (2004) model where in conditions like a literal teacher, self-generating the target, remembering the word from previous exposure, context providing the target information; the child was able to determine correct output without explicitly being told but however, back propagation isn't similar as complete feedback was not provided about the correct target for every learning occasion.

Hidden Units

Hidden units allow network to encode more complex mappings and it also increases the computational power of the network i.e. the range and complexity of the problems a model can solve. Along with the power, hidden units also have theoretical interpretation: underlying representations that abstract away from surface features of input and output codes where developed by the hidden units. Non-linear distinction between the outputs is provided allowing generalizations to be learned forming intermediate codes giving the networks character.

Experience

The training of the model involves it learning from the experiences. These experiences are computed in many ways for example a child learning how to read. In Seidenberg and McCelland model (1989), corpus consists of 2900 monosyllabic words and while training the model, a word is selected, its phonological output is calculated and compared to the target and depending on that the weights are adjusted by learning algorithm. This process was repeated for many words. The words whose frequency was higher were sampled such as the word THE were presented more often than words such as SINGLE. Implicit structure was encoded in its weights. Correspondence between the sound and the spelling was learnt by selecting statistical structure of its mapping as instantiated across a large pool of words.

Network Architecture

The pattern within the linkages between the units symbolizes the architecture of the network that in-turn represents all kinds on information. Characterization helps in the effectiveness of learning in the connectionist network helps exploring the degree to which built-in structure is necessary to account for all the phenomena’s. Depending on the expression of the characteristics, a wide variety of architectures are modelled ranging from extensive built-in structures to minimal structures.

Lexical Access

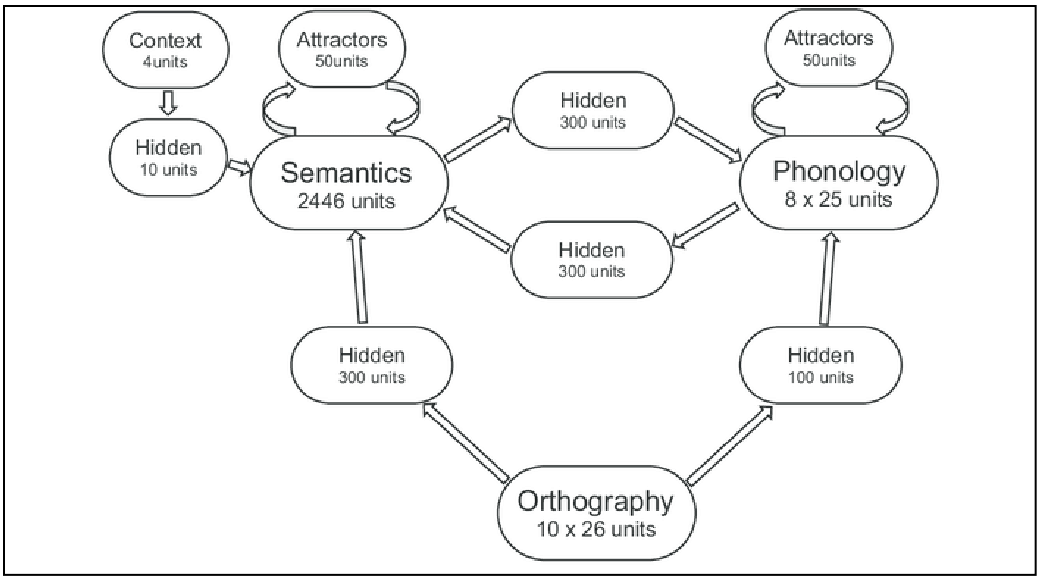
Lexicons are systematic organization of vocabulary that is stored in the mind in the form of individual lexical entries which are specific word stored in mind. Lexical access is the way individuals access words in mental lexicon and the factors affecting it are: frequency effect, the word/non-word effect, word superiority effect, the length and the image ability effect. Two models discussed below Triangle model of reading and TRACE model are studied to see the lexical access on them.

Triangle Model of Reading

The triangle model - Harm and Seidenberg (1999, 2004), Plaut et al. (1996), Seidenberg and McCelland (1989) can be explained on the basis of word identification and its architecture to understand how words are identified and represented in mental lexicon. During word identification soft constraints on the pronunciations and/or meaning generated by different types of lexical information is provided.

Architecture

The Triangle model consists of a set of phonological units, representing spoken word forms which are connected to, and from a set of semantic units representing the meaning of a word. Both the phonological and semantic units are self-connected to a set of attractor units, enabling stable phonological and semantic representations of words that are maintained.



Triangle Model of Reading.

There are 200 units in phonology layer, 2446 units in the semantics layer and 50 units in each in attractor layer. The orthography of words is represented by across 10 x 26 units of the orthography layer, which was connected to the phonological and semantic layers respectively via hidden layers each containing 300 units. Arrows in the figure indicate full connectivity between layers in the model.

Word Identification Model

The two basic fundamental assumptions of the triangle model are: First, a word's pronunciation is generated via propagating activation from processing units representing orthographic input along connections to other units representing phonological output. Knowledge that allows readers to identify printed words is enclosed in a single set of input-to-output connections allowing the sum of knowledge to influence the pronunciation of each word that is generated. Second, lexical information is represented in a distributed manner in the triangle models, with the dominant variants positing that orthographic input and phonological output are not represented by particular units per se but instead by specific patterns of distributed activity across the units. Thus the triangle model does not assume that the lexical information is represented by discrete processing units in the lexicon but instead assume that such information is contained in the connections that mediate between orthographic input and phonological output. The triangle model is thus able to predict that frequent words are pronounced more clearly and precisely than infrequent words by learning the strengths of the connections learned through repeated experience with words as these connections are more consistent for frequent words than that for infrequent words.

Predictions about response latencies, error rates, and the types of errors observed during lexical decisions and naming along with important phenomena's such as finding of frequent words faster than infrequent words and that the frequency effect is larger for irregular pronunciations, for example, words like pint, colonel, and yacht are explained by triangle model. Phonological dyslexia which can be characterized by difficulty pronouncing novel words and non-words, for example, pronounceable letter strings like burk, but not known words can be explained as, damage is reflected on some fraction of orthography-to-phonology connections to pronounce words that have been learnt previously along with little capability to be able to generalize pronunciation of new words or non-words. Surface dyslexia which can be characterized by difficulty pronouncing irregular words but not regular words, can also be explained, when the orthography-to-phonology connections become concentrated by pronunciation of consistent words.

Limitations

Several generations of models have been evolved and they continue to do so as inherent limitations of existing models and extend the range of phenomena addressed. The main limitations include: Quasi Regularity and Division of Labour should be addressed by adequate theory of word reading.

Quasi Regularity

Seidenberg and McClelland (1989), Plaut et al. (1996), and Harm and Seidenberg (1999) presented models that acquired spelling-sound knowledge where weights were accounted in rule-governed and exception cases and showed that the models accounted for phenomena's associating task of reading letters string aloud. The main aspects noted are: First, characterization of language as rule-governed as some aspect of underlying processing system of informal characterization is convenient but not accurate. Second, previously generalization required rules and evidence for its existence but its account in other domains is not known, although invites reconsideration of the kinds of evidence generally taken as evidence for rules. Third, approach to spelling-sound knowledge make different predictions than dual-route model.

Division of Labour

Computation of phonology has been a topic of discussion for many years, role of phonological information in silent reading and how meanings are computed. Trade-offs between direct orthography to semantics associations versus using phonological mediation is instantiated by model by having an efficient solution results if the output is determined jointly, with the division of labour determined by complementary computational properties.

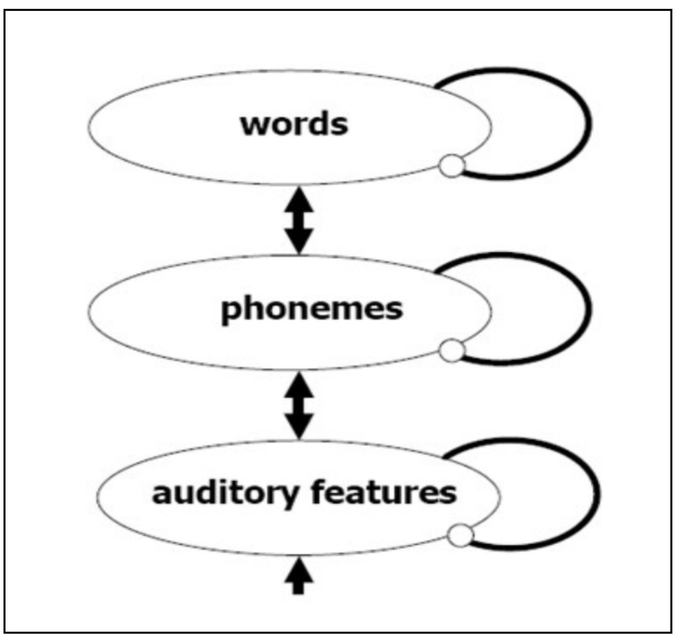
TRACE Model

The TRACE Model is based on interactive activation principles. It is named after the network of units forms a dynamic processing structure “the Trace” which basically is the perceptual processing mechanism and also as the system’s working memory.

The computational properties of simple recurrent networks trained to predict the next input in a stream of speech segments is the influential simulation and the result from these simulations is that in carrying out prediction task, the network displays sensitivity to the structure of lexical items. For example, consider the pair of words transmission and remission, the network may well be able to predict the next segment without error but it is not necessary for the system to produce a different representation depending on the onset of the word. Such representation is required if the network is to distinguish one word from another. Lexical effects obtained are small scale simulations which do not scale up to realistic training sets and one of the reasons it is observed in small scale is the problem with onset embedded words such as cap embedded in captain. For the recognition of these onset-embedded words, long lexical words that are carriers of the embedded words are to be ruled out. To do so may not be easy as word boundaries aren’t always distinct. But when TRACE is simulated, mismatching input along with lexical competition helps identify onset-embedded words (Frauenfelder & Peeters, 1990). For example, sequence ‘cap fits’, information after the offset of the embedded word ‘cap’, is fickle with words such as ‘captain’, thus, reducing the clash from carrier words, and allowing easy identification of embedded words.

Architecture

Information processing takes place through the excitatory and inhibitory interactions of a large number of simple processing units, each working continuously to update its own activation on the basis of the activations of other units to which it is connected. TRACE model is made up of large number of units, categorized into three levels- Feature, Phoneme and Word. There are several banks of feature detectors one for each of several dimensions of speech sound, at the feature level. Several successive moments in time are replicated by each bank. There are detectors for each of the phoneme at the phoneme level. One copy of each phoneme detector is over three time slices. Each unit spans six time slices. There are detectors for each word at the word level. One copy of word detector is over every three feature slices. Connections between the levels are bi-directional and excitatory. Connections within levels are inhibitory producing competition between alternatives. The entire network of units is called 'the TRACE' as the pattern of activation left by a spoken input is trace of analysis of the input at each of three processing levels.



TRACE Model

Limitations

Few of the deficiencies include decision mechanisms have not been fully elaborated, lack of explicit provision for variability in the activation and/or read out process. These deficiencies relate to simplifying assumptions of the stimulation model such as assumption that all phonemes are the same length, that all the features are equally salient, useful and overlap an equal amount from another phoneme. Two ways are required: Central representation that plays a role in processing every phoneme and word that is subject to learning, returning and priming. Dynamic trace of unfolding representation of the speech stream is required to continue to accommodate both left and right contextual effects.

Conclusion and Future Directions

Accomplishment of reading on the basis of Connectionist models is the foremost foundation of ideas. The approach incorporates ways of thinking about how knowledge is represented, acquired, and used that deviate in many respects from intuitive, traditional psychological accounts of cognitive phenomena. For example, people's knowledge of words is usually assumed to be stored in a dictionary-like mental lexicon with entries for individual words. Both the models reviewed in the paper: The Triangle Model of Reading and The TRACE Model are primary example of rudimentary connectionist models. Lexical effects on both the models vary on quite an extent. The Triangle model is capable of reading words from orthographic to phonological representation, its learning process mirrors a number of human characteristics, such as: difficulties with word processing, basic to expert reading transition, pronouncing uncommon words specially the ones not presented during training, and similar differences in performance on naming and lexical decision tasks. The TRACE model is a dynamic processing structure made up of a network of units, which performs as the system's working memory as well as the perceptual processing mechanism. The model has been made into a working computer program for running perceptual simulations. These simulations are predictions about how a human brain processes words in real time.

The cognitive processes that support reading are both varied and complex. All the connectionist models developed for word recognition are largely directed towards explaining one or two components of the reading process, for example, how printed words are identified, with little effort being directed towards explaining how these components interact with the remaining processes that are important in reading. Although this may have proven effective by focusing only on one problem, it fails to provide a narrow view of what actually transpires in the minds of readers. An alternative view of the role that computational models play in cognitive science is that it is better to develop large-scale models of the overall cognitive architecture than it is to develop smaller-scale models of specific cognitive tasks (Anderson & Lebiere, 1998; Newell, Rosenbloom, & Laird, 1989; Rumelhart, 1989). For example, specifying how words are identified across two or more different viewing locations within the perceptions of the human vision.

References

Anderson & Lebiere (1998). An Integrated Theory of the Mind

Allen Newell, Paul S. Rosenbloom, John E. Laird (1989). Symbolic architectures for cognition.

David C. Plaut (2004). Connectionist Approaches to Reading

David E. Rumelhart (1989). The architecture of mind: a connectionist approach

Frauenfelder, U. H., Segui, J., & Dijkstra, T. (1990). Lexical effects in phonemic processing: Facilitatory or inhibitory? *Journal of Experimental Psychology: Human Perception and Performance,* 16(1), 77-91

Harm MW, Seidenberg MS (1999). Phonology, reading acquisition, and dyslexia: Insights from connectionist models. *Psychological Review;* 106:491–528.

Harm MW, Seidenberg MS (2004). Computing the meaning of words in reading: Cooperative division of labor between visual and phonological processes. *Psychological Review*; 111:662–720.

James L. McCelland, David E. Rumelhart (1981). An Interactive Activation Model of Context effects in Letter Perception: Part 1 An Account of Basic Findings. *Psychological Review*; Vol. 88, No. 5, 375-407.

Mark S. Seidenberg (2005). Connectionist Models of Word Reading.

Mark S. Seidenberg, James L. McClelland (1989). A Distributed, Developmental Model of Word Recognition and Naming. *Psychological Review*, Vol. 96, No. 4, 523-568.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology, 18*(1), 1-86.

WIREs Cogn Sci 2015. doi: 10.1002/wcs.1340