Available online at www.sciencedirect.com

ScienceDirect

Cognitive Systems Research 52 (2018) 668–677

Cognitive Systems RESEARCH

www.elsevier.com/locate/cogsys

A visual auditory model based on Growing Self-Organizing Maps to analyze the taxonomic response in early childhood

Action editor: Dietmar Heinke

Valentina Gliozzi [a,b,*], Matteo Madeddu [b]

[a] *Center for Logic, Language, and Cognition, University of Torino, Italy*
[b] *Computer Science Department, University of Torino, Italy*

Received 22 December 2017; received in revised form 29 May 2018; accepted 31 July 2018
Available online 9 August 2018

Abstract

In this paper we present an extension of a visual auditory neural network model previously proposed by Mayor and Plunkett (2010) in order to explain the emergence of the taxonomic response in early childhood. The original model consists of two self-organizing maps (respectively, visual and acoustic) connected with Hebbian connections. With respect to the original model, our proposal adds two major features. First, our model follows a dynamic training regime, learning categories and word-object associations that evolve through time. Second, the visual and acoustic maps are Growing self-organizing maps that grow during training, when they are no longer able to consistently represent categories. With these two new characterizing features, our model replicates the performance of the original Mayor and Plunkett (2010)'s model, acquires psychological plausibility in the training regime, and avoids the risk of catastrophic interference.

Keywords: Neural networks; Taxonomic constraint; Growing self-organizing maps

1. Introduction

When a child learns a new word, she must decide what is the meaning of that word. Waxman and Kosowski (1990) suggest that preschool children approach the task of word learning equipped with implicit constraints that lead them to prefer some possible meanings over others. In a pioneering work Markman (1991) specifically examined in depth three constraints: the whole object constraint, the mutual exclusivity constraint, and the taxonomic constraint. Starting at about 18 months of age, children become remarkably capable of learning the vocabulary of natural languages. In order for children to acquire language as rapidly as they do, they must be able to eliminate many potential meanings of words (Markman, 1991). This is the role of the constraints. The *whole-object constraint* leads children to assume that terms refer to objects as a whole rather than to their parts, substance, color, or other properties. The *mutual exclusivity constraint* leads children to avoid two words for the same object: if a child already knows a word for an object, a new word for that object is at first rejected. The *taxonomic constraint*, on which we focus in this paper, leads children to extend words to taxonomically-related objects. In a typical labeling situation, the caregiver points at an object (e.g., Fido the dog) and says "Look, this is a dog!". In these circumstances, the infant has to rule out a huge number of possible meanings. Infants reliably interpret the word *dog* as a label that can be used for this dog and for *all* dogs.

* Corresponding author at: Center for Logic, Language, and Cognition, University of Torino, Italy.
E-mail address: valentina.gliozzi@unito.it (V. Gliozzi).

https://doi.org/10.1016/j.cogsys.2018.07.034
1389-0417/

The starting point of our work is the neurocomputational model proposed by Mayor and Plunkett (2010) to provide a mechanistic account of the taxonomic constraint, considered as the capacity of generalizing the meaning of a word to a whole category starting from a single labeling event. This notion of taxonomic constraint captures the formulation provided by Waxman and Markow (1995): "when infants embark upon the process of lexical acquisition, they are initially biased to interpret a word applied to an object as referring to that object and to other members of its kind". This ability to generalise from a single labelling event is closely related to the concept of *fast mapping* (Carey & Bartlett, 1978).

The intuitive grasp to the taxonomic constraint provided by Mayor and Plunkett (2010)'s model lies in the interplay between the topological organization of visual and acoustic categories (separately considered) provided respectively by a visual self-organizing map and an acoustic self-organizing map, on one side, and the Hebbian connections that develop between these two maps in a developmental as well as neurally plausible way, on the other side. These Hebbian connections establish a reference relation between words represented on the acoustic map and visual objects represented on the visual map. Given the topological organization of objects and words in the visual and acoustic map respectively, a single object-word co-occurrence event naturally generalizes to the whole category of the visual object and to the whole category of the word (i.e. to similar acoustic sequences, or phonological variants of the same word).

In this paper we address the following question: Can the model still account for the taxonomic constraint when trained in a more naturalistic way, by allowing the training stimuli to change over time, rather than being fixed once and for all at the beginning of training as in most neural network models? This question is important if one tries to model word learning and taxonomic responding in a naturalistically valid context. In such a context, new objects from previously unexperienced categories are continuously met, and the universe of known categories (what a model calls *training set*) expands through time. Furthermore, the activity of learning new visual categories is interleaved with the activity of learning word meanings: some words are associated to already known categories before new categories are learned. In a naturalistic setting these two phases are not separated, and in this paper we consider whether having this interleaved learning can be captured in a model.

As our results show, Mayor and Plunkett's (2010) model in its original formulation cannot cope with a training set evolving through time nor with a kind of learning alternating the acquisition of new categories and the acquisition of new word meanings.

In order to address this problem here, we extend the model in a few directions.

First of all, and most important, we allow a *flexible training regime*: the training set is not fixed once and for all at the beginning of training, and repeatedly presented in its entirety to the maps. On the contrary, training stimuli presented to the maps can evolve through training.

The flexible training regime is made possible by the fact that we use Growing self-organizing maps instead of standard self-organizing maps. In this way, the size of the maps augments when needed, and there is no a priori limitation on the number of categories that can be learned, nor risk for catastrophic interference. Indeed, if in standard self-organizing maps new learned categories can interfere with (and possibly override) previously learned categories, in Growing self-organizing maps this is avoided by expanding the map where and when needed to represent new information without overriding previously learned one. Our model based on Growing self-organizing maps does also pay attention not to disrupt previously learned word-object associations, by conservatively integrating new learned word object associations onto previously learned ones.

2. The starting Mayor & Plunkett model

Mayor and Plunkett (2010) present a neurocomputational model based on Self-Organizing Maps (SOM, for short) (Kohonen, 2001) that accounts for the emergence of taxonomic responding and fast mapping in early word learning. Fig. 1 is a drawing of the model.

The model is based on two SOMs. The first SOM is a visual map that processes visual input stimuli and represents the visual areas of the temporal cortex (inferior temporal, where object recognition takes place). The second map is an acoustic map that processes acoustic stimuli and represents auditory areas of the cortex involved in speech processing (as Wernicke area). The two SOMs are linked by Hebbian connections that develop when word-object pairs are jointly presented to the model.

The stimuli presented to the visual map are represented as distorted dot patterns, formed out of 9 points in a 30×30 matrix (similar to Posner, 1964; Posner, Goldsmith, & Welton, 1967) created starting from 100 prototypes. For each prototype, 24 distortions are created 8 low, 8 medium and 8 high variance distortions. Visual stimuli belong to the same category if they are distortions of the

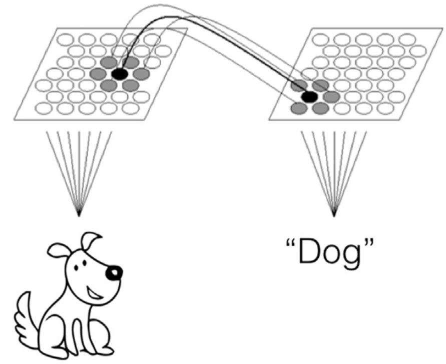

Fig. 1. Visual and auditory Hebbian training.