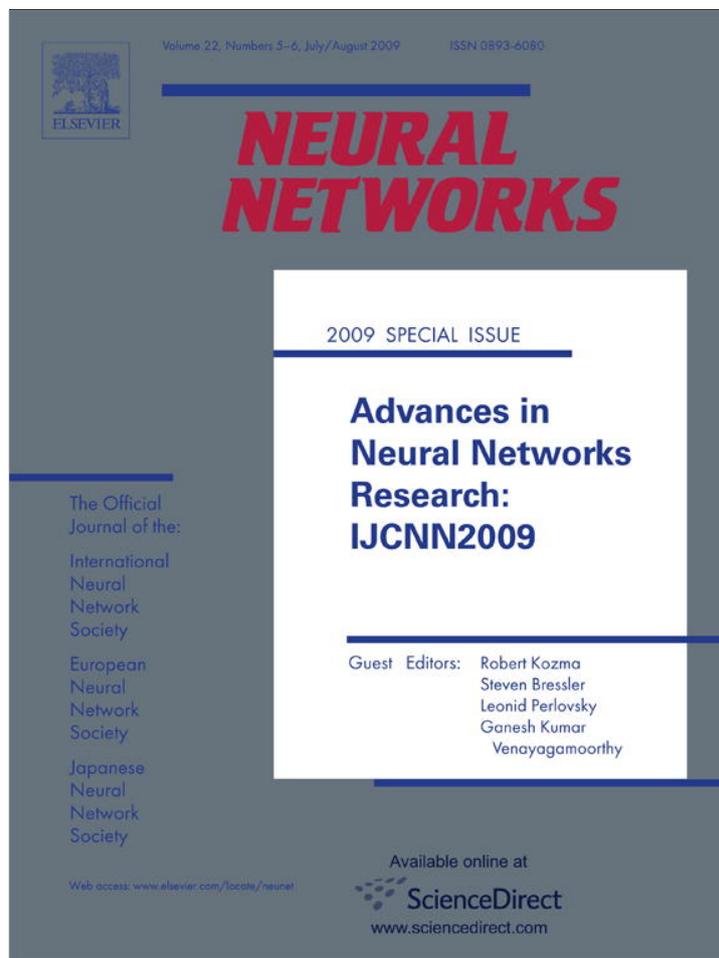


Provided for non-commercial research and education use.
Not for reproduction, distribution or commercial use.



This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

<http://www.elsevier.com/copyright>



Contents lists available at ScienceDirect

Neural Networks

journal homepage: www.elsevier.com/locate/neunet

2009 Special Issue

Cross-situational learning of object–word mapping using Neural Modeling Fields

José F. Fontanari^{a,*}, Vadim Tikhanoff^b, Angelo Cangelosi^b, Roman Ilin^c, Leonid I. Perlovsky^{c,d}^a Instituto de Física de São Carlos, Universidade de São Paulo, Caixa Postal 369, 13560-970 São Carlos, SP, Brazil^b Adaptive Behaviour & Cognition Research Group, University of Plymouth, Plymouth PL4 8AA, UK^c Air Force Research Laboratory, 80 Scott Drive, Hanscom Air Force Base, MA 01731, United States^d Harvard University, 33 Oxford Street, Rm 336, Cambridge MA 02138, United States

ARTICLE INFO

Article history:

Received 5 May 2009

Received in revised form 25 May 2009

Accepted 25 June 2009

Keywords:

Cross-situational learning

Language acquisition

Clustering algorithms

Neural Modeling Fields

ABSTRACT

The issue of how children learn the meaning of words is fundamental to developmental psychology. The recent attempts to develop or evolve efficient communication protocols among interacting robots or virtual agents have brought that issue to a central place in more applied research fields, such as computational linguistics and neural networks, as well. An attractive approach to learning an object–word mapping is the so-called cross-situational learning. This learning scenario is based on the intuitive notion that a learner can determine the meaning of a word by finding something in common across all observed uses of that word. Here we show how the deterministic Neural Modeling Fields (NMF) categorization mechanism can be used by the learner as an efficient algorithm to infer the correct object–word mapping. To achieve that we first reduce the original on-line learning problem to a batch learning problem where the inputs to the NMF mechanism are all possible object–word associations that could be inferred from the cross-situational learning scenario. Since many of those associations are incorrect, they are considered as clutter or noise and discarded automatically by a clutter detector model included in our NMF implementation. With these two key ingredients – batch learning and clutter detection – the NMF mechanism was capable to infer perfectly the correct object–word mapping.

© 2009 Elsevier Ltd. All rights reserved.

1. Introduction

Computational models play an important role in the investigation of language, both from an evolutionary perspective and from a developmental timescale. For example, simulation and robotic models of the evolution of language (Kirby, 2002; Parisi & Cangelosi, 2002) provide a complementary methodology that can help researchers to develop detailed and precise hypotheses on language origins and evolution, and to test these hypotheses in the virtual experimental laboratory of the simulation. Other models focus on the study of the developmental stages in language acquisition. This is the case of epigenetic robotics, where the design of the robot's behavioral and cognitive capabilities is directly inspired by developmental psychology theories (Cangelosi & Riga, 2006; Lungarella, Metta, Pfeifer, & Sandini, 2003). These can, for example, implement well-known phenomena in language development, such as vocabulary spurt and mutual exclusivity (Tomasello, 2003).

In this contribution we carry further the rather ambitious research program of integrating language and cognition within the Neural Modeling Fields Framework (NMF) (Cangelosi, Tikhanoff, Fontanari, & Hourdakis, 2007; Perlovsky, 2004, 2006a; Tikhanoff, Fontanari, Cangelosi, & Perlovsky, 2006). This is a task of enormous breadth that encompasses many unsolved (and, perhaps, unsolvable) problems such as object perception (Kellman & Spelke, 1983; Marr, 1982), symbol grounding which addresses the question of how physical signs can be given meaning (Cangelosi, Greco, & Harnard, 2002; Harnard, 1990), and the emergence of a common lexicon in a population of interacting agents (Cangelosi, 2001; Fontanari & Perlovsky, 2007, 2008a). This unifying perspective accords with the view that language is not an isolated capability of the individual and cannot be fully comprehended if one ignores its intrinsic relationships with the cognitive and social abilities (Cangelosi et al., 2002).

Although almost every statement about language can be considered controversial, one fact is noncontroversial, namely, that the lexicon must be learned from the active or passive interaction between children and language-proficient adults. Of course, the issue of whether this ability to learn the lexicon is due to some domain-general learning mechanism or is an innate ability, unique to humans, is still on the table (Bates & Elman, 1996). There is, nonetheless, a general agreement that lexicon learning

* Corresponding author. Tel.: +55 16 33739849; fax: +55 16 33739877.

E-mail addresses: fontanari@ifsc.usp.br (J.F. Fontanari), vadim.tikhanoff@plymouth.ac.uk (V. Tikhanoff), a.cangelosi@plymouth.ac.uk (A. Cangelosi), roman.ilin.ctr@hanscom.af.mil (R. Ilin), Leonid.Perlovsky@hanscom.af.mil (L.I. Perlovsky).

is accomplished by children chiefly through unsupervised learning (Bloom, 2000).

Here we explore one possible unsupervised learning scenario for lexicon acquisition, the cross-situational learning scenario, which is based on the very intuitive idea that one way that a learner can determine the meaning of a word is to find something in common across all observed uses of that word (Siskind, 1996). Borrowing an example from de Beule, de Vylder, and Belpaeme (2006), suppose one does not know the word 'banana' but one hears it often enough together with a situation where the fruit is actually present. As one hears the same sound across different situations, there will come a time when one makes the once and for all association between the word 'banana' and the yellow, sickle-shaped fruit. Although, the general notion of cross-situational learning has been proposed by many authors (see, e.g., Pinker (1989)) the instantiation of this rather abstract notion into a manageable mathematical or computational model is still under research (de Beule et al., 2006; Lenaerts, Jansen, Tuyls, & de Vylder, 2005; Smith, 2003). A rival learning scenario is the supervised learning, known generally as operant conditioning, which involves the active participation of a teacher who provides non-linguistic cues on the correctness of the learner inferences. Although this supervised learning scheme has been applied to the design of a system for communication by autonomous robots (Steels, 2003), it is unlikely to be the main mechanism used by children to learn the meanings of words (Bloom, 2000).

Regarding the program of integrating language and cognition, it is crucial to find whether a general purpose categorization algorithm such as NMF can be used as a mechanism to acquire a lexicon in an unsupervised learning scenario, this being the main motivation of this contribution. Essentially, language can be reduced to a mapping between sounds and meanings – the tenet of Chomsky's Minimalist Program (Chomsky, 1995). Here we address a part of this problem, in which language is viewed as a mapping between sounds (or words) and objects in the world. This is obviously a necessary first step to tackle more realistic problems, where the words may refer to particular sets of objects rather than to single objects only. In particular, we equip the learner with a NMF categorization mechanism sensitive to the frequency of co-occurrence of objects and words and show how the inclusion of a clutter detection module can identify and automatically discard the inputs representing wrong object–word associations. This is a major add-on to our previous attempt to apply the NMF categorization mechanism to the cross-situational lexicon acquisition problem (Fontanari, Tikhanoff, Cangelosi, & Perlovsky, 2009).

The rest of the paper is organized as follows. In Section 2 we present the cross-situational learning scenario and describe the structure of the input set for the batch learning mode. In Section 3 we describe at some length the NMF categorization mechanism within the context of the specific problem addressed in this paper. In Section 4 we present the results of the simulations of the NMF dynamics. Finally, in Section 5 we present some concluding remarks. A previous attempt to address the cross-situational learning of object–word mapping using the NMF algorithm was published in Fontanari et al. (2009).

2. Cross-situational learning scenario

Our minimal model of cross-situational lexicon learning involves two agents: the teacher, who has complete domain of the language, described by a one-to-one object–word mapping, and the pupil, equipped with a 'NMF-mind', and who has no knowledge of object–word relations. The pupil must infer that mapping from the examples provided by the teacher.

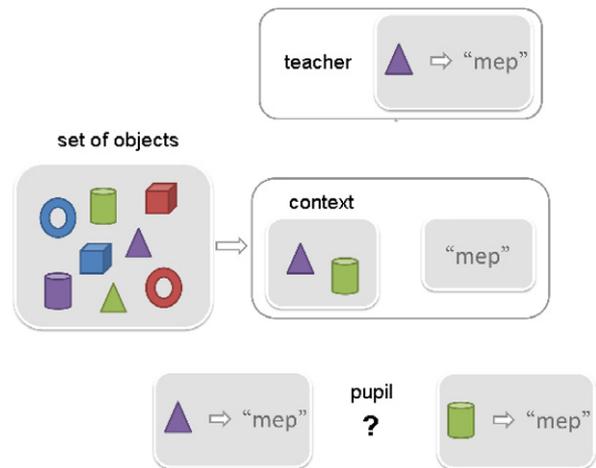


Fig. 1. Scheme of the cross-situational lexicon acquisition scenario. The pupil has access to a context comprised of two objects selected at random from the full set of N objects and to a word ('mep' in the figure) uttered by the teacher. The pupil's task is to infer the object–word mapping from the co-occurrence of objects in the context and the accompanying words.

We assume that there are N objects (and hence N words) in the agents' world. At each learning event, the teacher chooses two objects at random and without replacement from the fixed list of N objects. These two objects form the context. Then the teacher names one of the objects in the context. The pupil has access to the context as well as to the word emitted by the teacher. The pupil's task is to guess which of the two objects in the context the word refers to. Fig. 1 illustrates the cross-situational lexicon acquisition scenario.

The deterministic nature of the NMF algorithm requires that all input data (e.g., all possible examples of object–word pairs) be presented at once to the pupil. This sort of batch-mode learning setting differs considerably from the on-line learning of the typical guessing game scenario. Of course, real life learning is sequential as children are exposed to situations one at a time. However, learning is cumulative: a child looks around for several years and then very fast he/she learns to understand many situations. To simplify, we assume here that the data about all the situations are available at the same time. This assumption can be modified in the future.

Let us consider first the simple situation where there are three objects labeled by the integers 1, 2, 3 and three words labeled by the letters a , b , c . Once we understand the learning scenario in this case, we can generalize the notation for the case of an arbitrary number of objects and words. Without lack of generality, we assume that the correct one-to-one object–word mapping is such that object 1 corresponds to word a , object 2 to word b , and object 3 to word c . Consider the following learning event (see Fig. 1): objects 1 and 2 form the context and the teacher utters the word a . (Note that a and b are the only words that can accompany this context, as the teacher must refer either to object 1 or to object 2). From the pupil's perspective, this learning event leads to two possible object–word associations, namely, $(1, a)$ and $(2, a)$. In what follows we will refer to these object–word associations as inputs. There are 6 distinct learning events in this simple case: $\{(1, 2), a\}$; $\{(1, 2), b\}$; $\{(1, 3), a\}$; $\{(1, 3), c\}$; $\{(2, 3), b\}$ and $\{(2, 3), c\}$. In a batch learning scenario, these 6 learning events produce the 12 inputs $(1, a)$, $(2, a)$, $(1, b)$, $(2, b)$, $(1, c)$, $(3, a)$, $(1, c)$, $(3, c)$, $(2, b)$, $(3, b)$, $(2, c)$, and $(3, c)$. Among these 12 inputs, 6 represent correct object–word associations (e.g., $(1, a)$), and 6 represent incorrect associations (e.g., $(1, b)$). Note that any correct object–word association appears twice in the input set, whereas an incorrect association appears only once.

We consider now the general case where there are N objects and N words. Since the two objects that form the context are

chosen without replacement from a list of N objects without regard to order, there are exactly $N(N-1)/2$ different contexts. In addition, because each context can be associated to two different words to compose a learning event, there are $N(N-1)$ distinct learning events. Recalling that from the pupil's perspective a learning event generates two inputs (i.e., object–word associations), the total number of inputs to the pupil is

$$M = 2N(N-1). \quad (1)$$

To calculate how many times a correct object–word association, say $(1, a)$, appears in this input set, we note that the object 1 appears in $N-1$ contexts, and hence in $2(N-1)$ learning events, but in only half of these events, 1 is associated to the word a . So the correct input $(1, a)$ appears $N-1$ times in the input set. The total number of correct object–word associations in an input set of size M is then $M/2 = N(N-1)$. Of course, this result holds for the number of incorrect associations as well.

In summary, the input set consists of M object–word associations (o, w) in which the first entry represents the label of the object and the second entry the label of the word. In addition, the elements of the input set are not all distinct: any correct object–word association appears $N-1$ times in the input set, though any incorrect association appears only once. To facilitate the computational implementation of the input set, it is convenient to label both objects and words by integers, rather than by integers and letters as done in our didactic example, so that from now on an input will be represented by the pair (m, n) where $m, n = 1, \dots, N$. Since the first entry always refers to the object labels and the second, to the word labels, we do not need different sets of symbols to represent objects and words. The advantage of this scheme is that, without lack of generality, we can represent the correct object–word mapping by the pairs $(n, n); n = 1, \dots, N$. Hence an input (m, n) with $m \neq n$ corresponds to an incorrect object–word association.

The pupil attempts to model the input data through the object–word mapping (S_{1k}, S_{2k}) , with $k = 0, \dots, N$, where the components $S_{ek}, e = 1, 2$ – the so-called modeling fields – are real variables given by the NMF equations described in the next section. Those equations describe an unsupervised learning process in which the M inputs are grouped in a few classes that represent the correct object–word associations. The question is whether the pupil can recover the correct object–word mapping $(1, 1), (2, 2), \dots, (N, N)$ having access only to the information available in the cross-situational learning scenario summarized in Fig. 1. We note that there is a considerable loss of information involved in the procedure of replacing the sequential presentations of contexts plus words by the entire set of the M object–word examples, since one can easily imagine different situations which result in the same batch-mode learning set.

3. Neural modeling fields dynamics

The Neural Modeling Fields (NMF) algorithm proposed by Perlovsky (2001) is essentially an iterative, self-consistent, deterministic process designed to maximize the similarity between models and incoming signals. In this aspect, it shares some elements with the Hopfield–Tank neural network (Hopfield & Tank, 1985) and the mean-field annealing (Bilbro, Mann, & Miller, 1989). In fact, these two deterministic heuristics have been extensively used to search for optimal or quasi-optimal solutions of a variety of optimization problems, whereas NMF searches for the maximum of a global similarity function. The main feature that sets NMF apart from these heuristics, as well as from many other neural networks, is use of parametric models. The so-called fuzzy association variables, which can be thought of as weights of a neural network (Perlovsky & McManus, 1991) are expressed as functions

of these models. These variables give a measure of the probability of association between input data and concept models, although, as already pointed out, NMF is a deterministic algorithm.

As discussed in Section 2, each example (input signal) is described by the pair of integer variables (O_{1i}, O_{2i}) , with $O_{ei} = 1, \dots, N$ for $i = 1, \dots, M$. In most applications of the NMF algorithm, the aim is to maximize the similarity between input data and parametric models used to represent (and usually compress) that data. The difficulty with the cross-situational learning scenario is that some data must be ignored, as they represent wrong object–word associations. Let us assume that there are N concept models described by the pairs (S_{1k}, S_{2k}) with $k = 1, \dots, N$ that should ‘model’ the original set of M examples; hence the denomination ‘modeling fields’ to the mathematical quantities S_{ek} . In addition to these regular object–word pair models, we introduce a trash-can model $k = 0$ which is described by a constant value. Finally, to each model k we associate a *a priori* probability r_k such that $\sum_{k=0}^N r_k = 1$.

We begin the derivation of NMF algorithm, i.e., of the equations that govern the dynamics of the modeling fields, by introducing a measure for the similarity between input $i = 1, \dots, M$ and concept model $k = 1, \dots, N$, namely,

$$l(i | k) = \prod_{e=1}^2 [2\pi\sigma_{ek}^2]^{-1/2} \exp \left[-\frac{[d(O_{ei}, S_{ek})]^2}{2\sigma_{ek}^2} \right] \quad (2)$$

where $d(O_{ei}, S_{ek})$ is a measure of the distance between the components e of input i and model k . The fuzziness σ_{ek}^2 are sensitivity parameters of the similarity measure which are given *a priori* at this stage. For model $k = 0$, we define

$$l(i | 0) = 1/M \quad (3)$$

for all i . Clearly, model $k = 0$ plays the role of a trash can or clutter detector: it will collect any input i that is not close enough to one of the regular models.

The key issue here is the definition of the distance $d(O_{ei}, S_{ek})$ since we recall that in the notation for the inputs (m, n) introduced in Section 2 the integers m and n are simply labels for the objects and words, respectively. Of course, these labels tell us nothing about the physical nature of the objects and words. For example, in a realistic situation, the objects could be represented by pictures of geometric forms and the words by sound waves. In this case the component of the modeling fields associated to objects should be some complex parametric model designed for recognizing invariant geometric forms, and the component associated to words should be a set of Fourier frequencies, for instance. The important point is that, regardless of the physical nature of the inputs, it is always possible to define a distance between the inputs and the outputs of the model designed to describe them. Although such a realistic description of objects and words could give more ‘credibility’ to our virtual scenario, these complicating factors are peripheral to our primary goal of studying the learning of object–word associations within the NMF framework. Therefore, to avoid diversion by unnecessary complications here we stick to the simplest possible situation in which the integer input labels are the actual inputs (Fontanari & Perlovsky, 2007, 2008a), so that the input–model distance becomes simply $d(O_{ei}, S_{ek}) = O_{ei} - S_{ek}$.

The goal is to find an assignment between models k and inputs i such that the global similarity

$$L_0 = \sum_{i=1}^M \ln \sum_{k=0}^N r_k l(i | k) \quad (4)$$

is maximized with respect to the modeling fields S_{ek} and prior probabilities r_k . To take the normalization of r_k into account we maximize the quantity

$$L_\lambda = \sum_{i=1}^M \ln \sum_{k=0}^N r_k l(i | k) + \lambda \left(\sum_{k=0}^N r_k - 1 \right) \quad (5)$$

where λ is a Lagrange multiplier. Since

$$\frac{dL_\lambda}{dt} = \sum_{k=1}^N \sum_{e=1}^2 \frac{\partial L_\lambda}{\partial S_{ek}} \frac{dS_{ek}}{dt} + \sum_{k=0}^N \frac{\partial L_\lambda}{\partial r_k} \frac{dr_k}{dt} \quad (6)$$

we can guarantee that $dL_\lambda/dt \geq 0$ by defining the dynamics $dS_{ek}/dt = \partial L/\partial S_{ek}$ and choosing λ such that $\partial L/\partial r_k = 0$.

The calculation of $\partial L_\lambda/\partial S_{ek}$ is straightforward and yields

$$\frac{\partial L_\lambda}{\partial S_{ek}} = \sum_{i=1}^M \frac{r_k}{\sum_{k'=0}^N r_{k'} l(i | k')} \frac{\partial l(i | k)}{\partial S_{ek}}. \quad (7)$$

Using the identity $\partial y/\partial x = y \partial \ln y/\partial x$ and defining the fuzzy associations as

$$f(k | i) = \frac{r_k l(i | k)}{\sum_{k'=0}^N r_{k'} l(i | k')} \quad (8)$$

we can immediately write the equations for the modeling fields (Perlovsky, 2001)

$$\frac{dS_{ek}}{dt} = \sum_{i=1}^M f(k | i) \frac{\partial \ln l(i | k)}{\partial S_{ek}} \quad (9)$$

for $e = 1, 2$ and $k = 1, \dots, N$.

The condition $\partial L_\lambda/\partial r_k = 0$ is written as

$$\sum_{i=1}^M \frac{l(i | k)}{\sum_{k'=0}^N r_{k'} l(i | k')} + \lambda = 0. \quad (10)$$

Multiplying by r_k and summing over k yields $\lambda = -M$ so that

$$r_k = \frac{1}{M} \sum_{i=1}^M f(k | i). \quad (11)$$

We note that given $l(i | k)$ [see Eqs. (2) and (3)] we can obtain r_k by solving Eqs. (8) and (11) simultaneously. Hence the r.h.s. of our fundamental dynamic Eq. (9) is a function of the modeling fields S_{ek} only, as expected. In practice, the exact determination of r_k is not important provided they are approximately the same for all models. In our analysis we will fix r_k to the values corresponding to the optimal solution of the learning problem (see Section 4).

The fuzzy association variables $f(k | i)$, Eq. (8), play a fundamental role in the interpretation of the NMF dynamics by giving a measure of the correspondence between input i and model k relative to all other models k' . We note that the factor $f(k | i)$ in Eq. (9) couples not only S_{1k} and S_{2k} which is critical for producing a sensible object–word mapping but also the components of different modeling fields.

By construction, the dynamics (9) always converges to a (possibly local) maximum of the similarity L_λ for fixed fuzziness σ_{ek}^2 . A salient feature of the NMF is a match between parameter uncertainty and fuzziness of similarity. By properly decreasing the value of the fuzziness σ_{ek}^2 , a unique assignment between inputs and models is attained. In fact, for fixed σ_{ek}^2 we obtain the fuzzy logic limit, whereas for $\sigma_{ek}^2 = 0$ we obtain the usual crispy, Aristotelian logic limit. The basic idea of NMF is to reduce the fuzziness during the time evolution of the modeling fields and so, because of this interpolation, the algorithm is also referred as Dynamic Logic. Of course, this procedure is similar to the cooling schedule of simulated annealing (Kirkpatrick, Gelatt, & Vecchi, 1983), although, as pointed out before, a more appropriate comparison is with mean-field annealing (Bilbro et al., 1989) since both algorithms are deterministic.

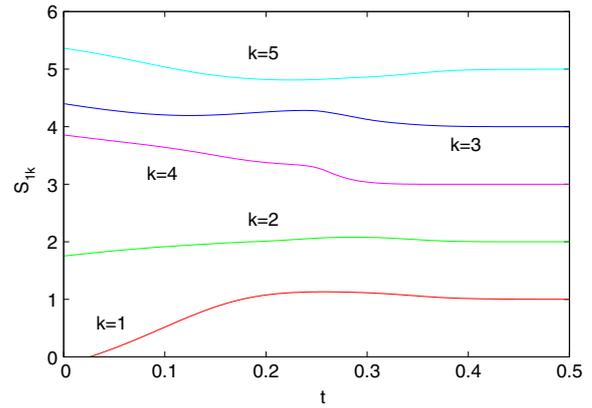


Fig. 2. Time evolution of the component S_{1k} of the modeling fields in the case of $N = 5$ objects. The indices k of the modeling fields are indicated in the figure. This component models the objects in the input data.

In what follows we decrease the fuzziness on the fly, i.e., simultaneously with the change of the modeling fields, according to the following prescription (Fontanari & Perlovsky, 2005)

$$\sigma_{ek}^2(t) = a_{ek}^2 + b_{ek} \exp(-\alpha_e t) \quad (12)$$

where $\alpha_e = 10$, $a_{ek} = 0.2$, and $b_{ek} = 1.0$ are time-independent control parameters. As a guideline for setting the values of these parameters, we note that b_{ek} must be chosen large enough such that, at the beginning, all category examples can be described by all modeling fields, whereas the baseline resolution a_{ek} must be small enough such that, at the end, a given modeling field will describe a single category. However, a_{ek} should not be set to a too small value to avoid numerical instabilities in the calculation of the partial similarities defined by Eq. (2). In practice, we replace the σ_{ek}^2 in Eq. (2) by the prescription (12) and solve the system of ordinary differential equation (9) using the Euler method with step size $h = 10^{-3}$.

4. Results

The desired or optimal performance of the categorization algorithm on the input set produced by the distinct learning events described in Section 2 is such that the $N - 1$ copies of input (1, 1) are associated to the same concept model, say model $k_1 > 0$, the $N - 1$ copies of input (2, 2) to concept model $k_2 > 0$, etc., and the remaining $N(N - 1)$ inputs corresponding to wrong object–word associations are associated to model $k = 0$. In other words, the algorithm must produce $N + 1$ distinct categories: all wrong object–word associations must go to category $k = 0$ (a trash can), and each distinct correct object–word association must go to a distinct category. The creation of these categories means that NMF mechanism has inferred the association between object label 1 and word label 1 (category k_1), object label 2 and word label 2 (category k_2), and so on.

In Figs. 2 and 3 we show the time evolution of the first and second entries, respectively, of the modeling fields in the case of $N = 5$ objects. These results were obtained by solving the ordinary differential equation (9) with the prior probabilities fixed at the values $r_k = 1/2N$ for $k = 1, \dots, N$ and $r_0 = 1/2$. As we will show below, this choice corresponds to the optimal solution of the categorization problem. In addition, the values of the modeling fields at $t = 0$ were chosen by modifying randomly the correct object–word mapping, (n, n) ; $n = 1, \dots, N$. Explicitly, for model $k = 1, \dots, N$ we set $S_{ek}(t = 0) = k + \eta$ for $e = 1, 2$ where η is a random variable uniformly distributed in the interval $[-1, 1]$. Despite this bias, the run illustrated in these figures shows an inversion between models $k = 3$ and $k = 4$; i.e., model $k = 3$

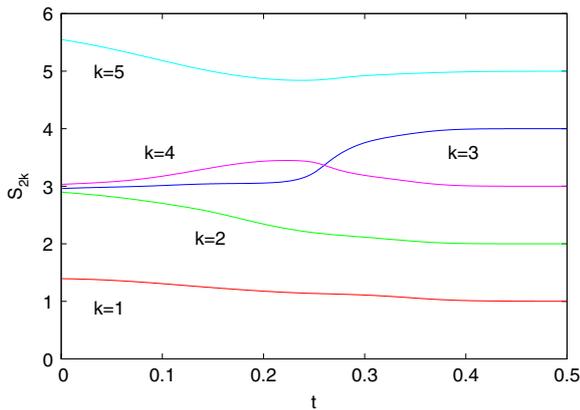


Fig. 3. Time evolution of the component S_{2k} of the modeling fields in the case of $N = 5$ objects. The indices k of the modeling fields are indicated in the figure. This component models the words in the input data.

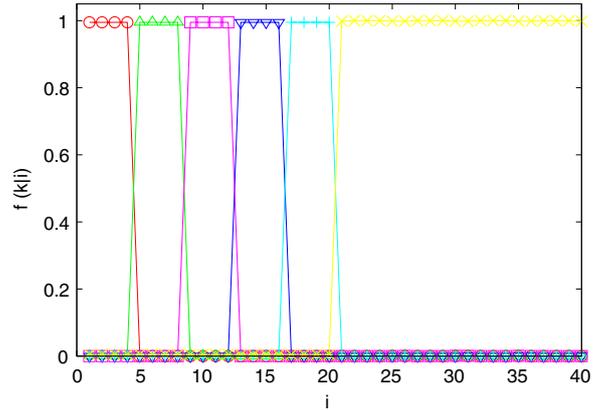


Fig. 5. Same as Fig. 4 but for the final values of the fuzziness association variables $f(k|i)$.

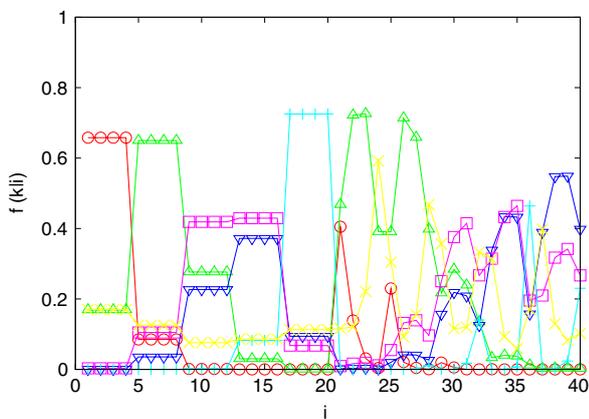


Fig. 4. The initial values of the fuzziness association variables $f(k|i)$ as function of the input labels $i = 1, \dots, 40$ for the same run of Figs. 2 and 3. The convention is $k = 0 (\times)$, $1 (\circ)$, $2 (\Delta)$, $3 (\nabla)$, $k = 4 (\square)$, and $k = 5 (+)$.

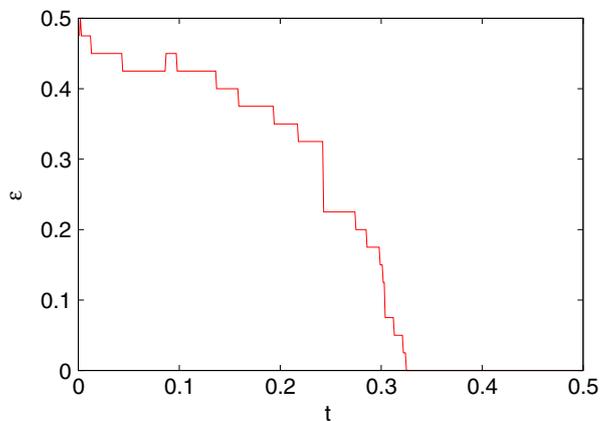


Fig. 6. Learning error ϵ as function of time t for the same run of the previous figures.

is associated to the object–word pair (4, 4) whereas model $k = 4$ is associated to the pair (3, 3).

Although the modeling fields S_{ek} are the fundamental quantities of the NMF algorithm, the fuzzy association variables $f(k|i)$ are more convenient to evaluate the categorization or learning performance of the algorithm. In terms of the fuzzy associations, the optimal solution is $f(k_1|i) = 1$ if input i is the pair (1, 1), and 0 otherwise; $f(k_2|i) = 1$ if input i is the pair (2, 2), and 0 otherwise; etc., and $f(0|i) = 1$ if input i is a wrong object–word association, and 0 if i is a correct object–word association. According to Eq. (11), the optimal solution yields the prior probabilities r_k used in the numerical solution of the modeling fields dynamics. To aid the visualization of the fuzzy associations, we set the inputs from $i = 1$ to $i = N - 1$ to (1, 1); from $i = N$ to $i = 2N - 2$ to (2, 2); ...; from $i = N(N - 1) - (N - 2)$ to $i = N(N - 1)$ to (N, N). Finally, the wrong object–word associations are distributed among inputs $i = N(N - 1) + 1$ and $i = M = 2N(N - 1)$.

Fig. 4 illustrates the initial values of the fuzzy association variables for the same run used to draw Figs. 2 and 3. It is interesting that, except for concept models $k = 1$ and $k = 5$ which are already well adjusted to their input subsets, the status of the other concept models are completely uncertain at the initial stage. Nevertheless, the NMF dynamics leads to the correct categorization of the input set, as illustrated in Fig. 5, which shows the values of the fuzzy associations in the stationary regime.

It would be desirable to define a measure of the learning error in terms of the fuzzy association variables. This error should be large in a situation depicted in Fig. 4 and zero in the situation depicted in Fig. 5. If the final assignment between models and inputs is known, then it is easy to define such measure: for each input i we calculate $k_m(t) = \max_k \{f(k|i;t); k = 0, \dots, N\}$, and increase the error by one unit whenever $k_m(t) \neq k_m(\infty)$. The following mathematical expression summarizes this procedure and yields a conveniently normalized learning error

$$\epsilon(t) = 1 - \frac{1}{M} \sum_{i=1}^M \delta(k_m(t), k_m(\infty)) \quad (13)$$

where $\delta(i, j)$ is the Kronecker delta. The results are summarized in Fig. 6. We note that a completely random assignment between models and inputs yields $\epsilon = 5/6 \approx 0.833$, the relatively low initial error ($\epsilon \approx 0.5$) is due to our biased choice of the initial conditions (see Fig. 4).

5. Conclusion

Interaction between language and cognition is among unknown neural mechanisms of the mind and a problem posing significant mathematical difficulties. These mechanisms have to explain several aspects of the mind functioning, which seem mysterious. How does language acquire meanings? Is cognition based on language, or vice versa? Do we think with words, or do we use

words to label already made up thoughts and decisions, and if so, then what are thoughts without words? Why children acquire language by the age of 5, but cannot act like adults, what exactly is missing in terms of neural mechanisms? How is it possible that we learn correct associations between words and objects, phrases and situations, among astronomically large number of possible incorrect associations? How phrases are formed from words so that they relate to the meanings of situations in the real world? An outline of possible mathematical and neural mechanisms has been described in Perlovsky (2006b, 2007a, 2007b, 2009a). The current contribution along with Fontanari and Perlovsky (2005, 2007, 2008a, 2008b) is a significant step in the mathematical implementation of that program.

In this contribution we address perhaps the simplest of the above questions, namely, how children learn the correct association between words and objects in an unsupervised learning scenario. Our results show that the NMF algorithm can find the optimal solution,¹ in the sense that the algorithm creates distinct categories for all correct object–word associations and dumps all wrong associations in a single category. The example demonstrated here is intentionally limited in scope and can be solved by mere statistical means. Its significance, however, is in that NMF is a scalable mathematical paradigm, its complexity grows linearly with the number of words and objects, and it has been demonstrated to solve problems, which could not have been solved by other methods due to combinatorial complexity (Perlovsky, 2006b, 2009a).

In this vein, we note that the convergence time of simple on-line learning statistical algorithms for detecting the co-occurrence of words and objects increases exponentially with the number of objects N in the case the context size C scales linearly with N (Fontanari & Cangelosi, unpublished). Although, as already pointed out, the convergence time of the NMF algorithm scales linearly with $N + 1$ (i.e., with the number of ordinary differential equations in the system (9)), the size of the input set of the batch-mode learning scheme increases with N^C , which becomes unwieldy when C grows linearly with N . Nevertheless, due to the high redundancy of the input set, applying the NMF algorithm to a relatively small random sample of the input set is likely to yield the correct object–word mapping. We note that also in the on-line learning scenario the pupil is exposed to only a small fraction of all possible contexts.

Future directions include multi-agent simulations (Fontanari & Cangelosi, unpublished), in which models are replaced by competing agents; mathematical methods implementing and demonstrating learning of phrases from words, situations from objects (Ilin, Fontanari, & Perlovsky, 2009), and extending these methods to abstract thinking in interacting language and cognition. These mathematical methods will be combined with perceptual symbol system (Barsalou, 1999) and related to neural mechanisms and brain imaging studies. Future research will address the role of language in rational and irrational thinking (Levine & Perlovsky, 2008); emotional mechanisms of language and roles of language emotionality and music in interaction of language and cognition (Perlovsky, 2008, 2009a); co-evolution of languages, cognition, and cultures (Perlovsky, 2007c, 2009b).

Acknowledgments

The research at São Carlos was supported in part by CNPq and FAPESP, Project No. 04/06156-3. AFRL research was supported in part by AFOSR PM Dr. D. Cochran, and by the NRC Research Associateship programs award.

¹ This is a great improvement compared to our previous results (Fontanari et al., 2009). The successful performance owes to the inclusion of the trash-can model $k = 0$, which absorbed all the wrong object–word associations.

References

- Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, 22, 577–609.
- Bates, E., & Elman, J. (1996). Learning rediscovered. *Science*, 274, 1849–1850.
- de Beule, J., de Vylder, B., & Belpaeme, T. (2006). A cross-situational learning algorithm for damping homonymy in the guessing game. In L. M. Rocha, M. Bedau, D. Floreano, R. Goldstone, A. Vespignani, & L. Yaeger (Eds.), *Proceedings of the Xth conference on artificial life* (pp. 466–472). Cambridge, MA: MIT Press.
- Bilbro, G., Mann, R., & Miller, T. (1989). Optimization by mean field annealing. In D. S. Touretzky (Ed.), *Advances in neural information processing systems I* (pp. 91–98). San Mateo, CA: Morgan Kaufmann.
- Bloom, P. (2000). *How children learn the meaning of words*. Cambridge, MA: MIT Press.
- Cangelosi, A. (2001). Evolution of communication and language using signals, symbols, and words. *IEEE Transactions on Evolutionary Computation*, 5, 93–101.
- Cangelosi, A., Greco, A., & Harnard, S. (2002). Symbol grounding and the symbolic theft hypothesis. In A. Cangelosi, & D. Parisi (Eds.), *Simulating the evolution of language* (pp. 191–210). London: Springer-Verlag.
- Cangelosi, A., & Riga, T. (2006). An embodied model for sensorimotor grounding and grounding transfer: Experiments with epigenetic robots. *Cognitive Science*, 30, 673–689.
- Cangelosi, A., Tikhanoff, V., Fontanari, J. F., & Hourdakis, E. (2007). Integrating language and cognition: A cognitive robotics approach. *IEEE Computational Intelligence Magazine*, 2, 65–70.
- Chomsky, N. (1995). *The minimalist program*. Cambridge, MA: Cambridge University Press.
- Fontanari, J. F., & Perlovsky, L. I. (2005). Meaning creation and modeling field theory. In *Proceedings of the IEEE conference on integration of knowledge intensive multi-agent systems* (pp. 405–410).
- Fontanari, J. F., & Perlovsky, L. I. (2007). Evolving compositionality in evolutionary language games. *IEEE Transactions on Evolutionary Computation*, 11, 758–769.
- Fontanari, J. F., & Perlovsky, L. I. (2008a). A game theoretical approach to the evolution of structured communication codes. *Theory in Biosciences*, 27, 205–214.
- Fontanari, J. F., & Perlovsky, L. I. (2008b). How language can help discrimination in the neural modeling fields framework. *Neural Networks*, 21, 250–256.
- Fontanari, J. F., Tikhanoff, V., Cangelosi, A., & Perlovsky, L. I. (2009). A cross-situational algorithm for learning a lexicon using Neural Modeling Fields. In *Proceedings of the IEEE international joint conference on neural networks IJCNN09*.
- Fontanari, J. F., & Cangelosi, A. (2009). A minimal model of cross-situational vocabulary learning. In *Proceedings of the IEEE conference on integration of knowledge intensive multi-agent systems*.
- Harnard, S. (1990). The symbol grounding problem. *Physica D*, 42, 335–346.
- Hopfield, J. J., & Tank, D. W. (1985). Neural computation of decisions in optimization problems. *Biological Cybernetics*, 52, 141–152.
- Ilin, R., Fontanari, J. F., & Perlovsky, L. I. (2009). Learning situations with multiple objects. (Unpublished).
- Kellman, P. J., & Spelke, E. S. (1983). Perception of partly occluded objects in infancy. *Cognitive Psychology*, 15, 483–524.
- Kirby, S. (2002). Natural language from artificial life. *Artificial Life*, 8, 185–215.
- Kirkpatrick, S., Gelatt, C. D., & Vecchi, M. P. (1983). Optimization by simulated annealing. *Science*, 220, 671–680.
- Lenaerts, T., Jansen, B., Tuyls, K., & de Vylder, B. (2005). The evolutionary language game: An orthogonal approach. *Journal of Theoretical Biology*, 235, 566–582.
- Levine, D. S., & Perlovsky, L. I. (2008). Neuroscientific insights on biblical myths: Simplifying heuristics versus careful thinking: Scientific analysis of millennial spiritual issues. *Zygon, Journal of Science and Religion*, 43, 797–821.
- Lungarella, M., Metta, G., Pfeifer, R., & Sandini, G. (2003). Developmental robotics: A survey. *Connection Science*, 15, 151–190.
- Marr, D. (1982). *Vision*. San Francisco: Freeman.
- Parisi, D., & Cangelosi, A. (2002). *Simulating the evolution of language*. London: Springer-Verlag.
- Perlovsky, L. I., & McManus, M. M. (1991). Maximum likelihood neural networks for sensor fusion and adaptive classification. *Neural Networks*, 4, 89–102.
- Perlovsky, L. I. (2001). *Neural networks and intellect: Using model-based concepts*. Oxford: Oxford University Press.
- Perlovsky, L. I. (2004). Integrating language and cognition. *IEEE Connections*, 2, 8–13.
- Perlovsky, L. I. (2006a). Integrated emotions, cognition, and language. In *Proceedings of the IEEE international joint conference on neural networks IJCNN06* (pp. 1570–1575).
- Perlovsky, L. I. (2006b). Toward physics of the mind: Concepts, emotions, consciousness, and symbols. *Physics of Life Reviews*, 3, 22–55.
- Perlovsky, L. I. (2007a). Symbols: Integrated cognition and language. In R. Gudwin, & J. Queiroz (Eds.), *Chapter in semiotics and intelligent systems development* (pp. 121–151). Hershey, PA: Idea Group.
- Perlovsky, L. I. (2007b). Neural networks, fuzzy models and dynamic logic. In R. Köhler, & A. Mehler (Eds.), *Chapter in aspects of automatic text analysis (Festschrift in Honor of Burghard Rieger)* (pp. 363–386). Berlin: Springer.
- Perlovsky, L. I. (2007c). Evolution of languages, consciousness, and cultures. *IEEE Computational Intelligence Magazine*, 2, 25–39.
- Perlovsky, L. I. (2008). Music and consciousness. *Leonardo, Journal of Arts, Sciences and Technology*, 41, 420–421.
- Perlovsky, L. I. (2009a). Language and cognition. *Neural Networks*, doi:10.1016/j.neunet.2009.03.007.

- Perlovsky, L. I. (2009b). Emotions, language, and Sapir–Whorf hypothesis. *Neural Networks*, 22(5–6), 518–526.
- Pinker, S. (1989). *Learnability and cognition*. Cambridge, MA: MIT Press.
- Siskind, J. M. (1996). A computational study of cross-situational techniques for learning word-to-meaning mappings. *Cognition*, 61, 39–91.
- Smith, A. D. M. (2003). Semantic generalization and the inference of meaning. *Lecture Notes in Artificial Intelligence*, 2801, 499–506.
- Steels, L. (2003). Evolving grounded communication for robots. *Trends Cognition Science*, 7, 308–312.
- Tikhanoff, V., Fontanari, J. F., Cangelosi, A., & Perlovsky, L. I. (2006). Language and cognition integration through modeling field theory: Category formation for symbol grounding. *Lecture Notes in Computer Science*, 4131, 376–385.
- Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Cambridge: Harvard University Press.