International Journal of General Systems
Vol. 37, No. 6, December 2008, 715–732

Neural memories and search engines

Eduardo Mizraji*

Group of Cognitive Systems Modelling, Biophysical Section, Faculty of Sciences, Universidad de la República, Montevideo, Uruguay

(*Received 15 March 2007; final version received 23 January 2008*)

In this article, we show the existence of a formal convergence between the matrix models of biological memories and the vector space models designed to extract information from large collections of documents. We first show that, formally, the term-by-document matrix (a mathematical representation of a set of codified documents) can be interpreted as an associative memory. In this framework, the dimensionality reduction of the term-by-document matrices produced by the latent semantic analysis (LSA) has a common factor with the matrix biological memories. This factor consists in the generation of a statistical 'conceptualisation' of data using little dispersed weighted averages. Then, we present a class of matrix memory that built up thematic blocks using multiplicative contexts. The thematic memories define modular networks that can be acceded using contexts as passwords. This mathematical structure emphasises the contacts between LSA and matrix memory models and invites to interpret LSA, and similar procedures, as a reverse engineering applied on context-deprived cognitive products, or on biological objects (e.g. genomes) selected during large evolutionary processes.

Keywords: neural memories; search engines; vector space models; context memories

AMS Subject Classifications: 92B20; 68T30; 68T35

1. Introduction

During the nineteenth century, thermodynamics produced a strange interaction between abstract scientific problems and practical technological objectives. The strangeness derives from the fact that deep and mysterious questions concerning the origin and the end of the universe, became blended with the search for optimal thermal engines able to perform mechanical work with maximal efficiency (see, for instance, Edsall and Wyman 1958, Feynman *et al.* 1963).

Nowadays, information sciences are provoking a similar kind of blending that involves the profound mysteries surrounding the nature and the potentialities of the human mind, and the very practical problems concerning the information retrieval from complex databases stored in computer memories (see Cooper 1973, 1980, Kohonen 1977, Landauer and Dumais 1997, Berners-Lee and Fischetti 1999).

The first objective of the present article is to show the formal convergence between the matrix models of biological memories and the vector space models designed to extract information from large collections of documents. The second objective is to propose an explanation for this convergence. The third objective is to introduce context-dependent matrix memories to model the organisation of the stored data in thematic clusters.

*Email: mizraj@fcien.edu.uy

ISSN 0308-1079 print/ISSN 1563-5104 online

DOI: 10.1080/03081070802037738
http://www.informaworld.com

This paper is organised as follows. We begin describing some aspects of neural databases, and we present the matrix models for associative memories. Then we show that the matrix representation of a collection of documents, the term-by-document matrix, admits the format of an associative memory model. In particular, we show that the dimensionality reduction of the term-by-document matrices using the latent semantic analysis (LSA) generates a mathematical object with suggestive conceptual similarities with the biological memory models. In Section 3, we propose an explanation for the mathematical similarities between biological memory models and LSA. After that, we describe matrix memories that allow thematic packing using multiplicative contexts. These packed memories permit the construction of modular networks acceded using passwords represented by the contexts. The emerged mathematical structure enhances the contacts between LSA and matrix memory models, and allows interpreting LSA as a reverse engineering applied over cognitive products, or over products (e.g. genomes) selected via large evolutionary processes.

2. Towards a cognitive dynamics

2.1 Neural databases

The human brain exhibits an extremely complex anatomical structure. A basic component of this structure is a class of nervous cell, the neuron, specialised in the transmission, the reception, and in some cases the storage, of coded information (for details, see Nauta and Feirtag 1986). The information is coded using electrochemical signals and is transmitted by the axon. The axon is a microscopic biophysical cable, and there exists only one axon per cell. The fundamental autopropagated electrochemical signal that travels across the axon is the action potential. Each axon contacts its effectors (for instance the dendrites, or the cell body of another neuron) in a complex cellular region: the synapse. Usually, in the synapses the incoming electrochemical signals are transduced into chemical signals carried by neurotransmitters. Inside the synapse, these chemical signals travel the space between two membranes, link the molecular receptors placed in the membrane of the target cell, and finally can be reconverted in a new electrochemical code. In the synapses, very complex biochemical and electrochemical events occur that promote and modulate the information transfer between cells. The global capacity of a synapse to transmit information can be characterised using a kind of global synaptic conductance. In many mathematical models, this synaptic conductance is represented by a single real number named ‘synaptic weight’. Synapses can enhance (excitatory synapses) or reduce (inhibitory synapses) the neural activity of their targets, a fact that mathematical models reflect in the signs of the synaptic weights.

In the human brain, the largest amount of neurons contact via synapses with other neurons placed in their vicinity, building up local neuronal (or neural) networks. At the same time, these local networks are connected with other local neural networks situated in other regions of the brain. It is currently accepted that the brain is organised as a modular network, each module being a local (yet large) neural network.

The human brain is a system open to information. Part of this information comes from the external world via the sensory organs; the other part comes from internal receptors that inform the brain about the state of many physiological variables, including the positions of the different parts of the body. The interaction of the brain with the external world occurs by means of motor actions, programmed into the brain and executed by skeletal muscles. These motor acts include speech and writing. The brain is also involved in the physiological regulations acting on endocrine or smooth muscular effectors.

Independently of the position of an animal in the evolutionary scale, the nervous system is always a central device for the adaptation to different environments (Ashby 1960). Both in very ancient species and in modern *Homo sapiens* we find repertories of neural circuits that pre-program rapid responses to usual stimuli and aggressions. These repertories are a set of programs rigidly installed in the 'neural hardware', and they are genetically determined. In modern and complex animals, the adaptive abilities become drastically enlarged with the emergence of neural memories. These memories depend on the learning capacities of some neural networks, and their physical support is the synaptic plasticity (Kandel and Schwartz 1985). This plasticity refers to the fact that some synapses are capable of modifying their global conductance (hence, their synaptic weights) as a function of their previous signal processing activities. In the biological realm, the human brain seems to display the higher learning capacity.

The large scale neural circuitry in human nervous systems is so regular, that it can be described in standard textbooks (e.g. Delmas and Delmas 1958, Kandel and Schwartz 1985, Nauta and Feirtag 1986). We mention in passing that into the cerebral cortex, this circuitry can be characterised by graphs having the properties of 'small-world' topologies (Sporns and Zvi 2004). However, in some neural regions, at the small-scale involving neuronal synapses, the connectivity is modifiable as a function of the experience, and this fact is behind the creation of specialised neural memories (Kandel and Schwartz 1985). Cognition depends entirely on these memories that install in the brain a complex network of information databases.

Different cognitive networks are built up by different human individuals, all of them sharing the same large scale brain circuitry. It is interesting to mention here that this fact suggests a metaphorical analogy between natural neuronal networks and some technological information networks. Thus, we can put in correspondence the neuro-anatomy with the Internet on the one hand, and the brain cognitive network with the World Wide Web (WWW) by the other. In the next sections, we show how, perhaps unexpectedly, this analogy is accompanied by structurally similar mathematical models concerning information retrieval in neural memories and in the WWW.

2.2 Matrix memories

In this section, we restrict our treatment to neural memories susceptible to being modelled using vector spaces and matrix algebra (Anderson 1972, Kohonen 1972, 1977, Cooper 1973, 2000). In this framework, a modular neural memory can be represented as an operator linking two vector spaces: a m-dimensional input space and a n-dimensional output space. The elements of the vectors are real numbers corresponding to the electrochemical signals transported by the neuronal axons (usually, frequencies of action potentials).

The vectorial representation is based on the fact that all kind of information processed and travelling inside the brain is coded in parallel by thousands or millions of electro-chemically active axons (Anderson 1995). Large-dimensional vectors are, consequently, natural mathematical variables to describe neural activity.

Consider, for instance, that a visual pattern (e.g. a face) is transported by photons that impact the retina. Here, the optical pattern is transduced into action potentials, and transferred inside the brain through thousands of axons. Each axon displays its own frequency of action potentials. This large set of different frequencies is the immediate neural codification. Then, this original vector is further processed in other neural centres and codified by other neural vectors. In this way, a memory that associates faces with names, receives as input a vector that results from many processing steps. The pronunciation (or writing) of the name associated with the original face, is also triggered by a neural vector that after the corresponding processing steps activates the muscular effectors responsible of speech (or writing).

Let $\hat{\mathbf{f}} \in \mathbb{R}^m$ and $\hat{\mathbf{g}} \in \mathbb{R}^n$ be, respectively, the input and the output column vectors processed by an associative memory Mem. The installation of a database of K associated patterns $(\hat{\mathbf{f}}_i, \hat{\mathbf{g}}_i)$ in a memory implies that

$$\hat{\mathbf{g}}_i = \text{Mem}(\hat{\mathbf{f}}_i), \quad (i = 1, 2, \ \ldots, \text{K}).$$

Neurons are non-linear devices, and the generation of action potentials involves thresholds. Nevertheless, the integrate-and-fire biophysical model of neuronal responses shows that in the frequency domain a linear region of behaviour can exist (Nass and Cooper 1975, Kohonen 1977, also see various papers in Anderson and Rosenfeld 1988). This linear region is the basis of the matrix models for associative memories. In the simplest case, the memory operator Mem can be represented by the matrix **M** given by

$$\mathbf{M} = \sum_{\text{i}=1}^{\text{K}} \hat{\mathbf{g}}_i \hat{\mathbf{f}}_i^{\text{T}}, \tag{1}$$

where the superindex T means transposition (Anderson 1972, 1995, Kohonen 1972, 1977). Let S_M be the subspace generated by the set of input vectors $\{\hat{\mathbf{f}}_i\}$ belonging to the memory M. Hence, $\forall \alpha_i \in \mathbb{R}$, is $\sum_{i=1}^{\text{K}} \alpha_i \hat{\mathbf{f}}_i \in \text{S}_\text{M}$.

The memory (1) behaves as follows:

$$\mathbf{M}\hat{\mathbf{f}} = \sum_{i=1}^{\text{K}} \langle \hat{\mathbf{f}}_i, \hat{\mathbf{f}} \rangle \hat{\mathbf{g}}_i, \tag{2}$$

where $\langle \mathbf{a}, \mathbf{b} \rangle = \mathbf{a}^{\text{T}}\mathbf{b}$ is the scalar product between column vectors **a** and **b**. If the set $\{\hat{\mathbf{f}}_i\}$ is orthogonal, and the input $\hat{\mathbf{f}} = \hat{\mathbf{f}}_k \in \{\hat{\mathbf{f}}_i\}$, then recollection of the memory M is exact (except for a scale factor):

$$\mathbf{M}\hat{\mathbf{f}}_k = \boldsymbol{\nu}_k \hat{\mathbf{g}}_k,$$

being $\boldsymbol{\nu}_k = \|\hat{\mathbf{f}}_k\|^2$ ($\|\mathbf{z}\| = \sqrt{\langle \mathbf{z}, \mathbf{z} \rangle}$ is the Euclidean norm of vector **z**. In general, if $\hat{\mathbf{f}} \in \text{S}_\text{M}$ the output is a linear combination of the $\hat{\mathbf{g}}_i$; if $\hat{\mathbf{f}}$ is orthogonal to S_M (we put: $\hat{\mathbf{f}} \perp \text{S}_\text{M}$) there is no recognition: $\text{M}\hat{\mathbf{f}} = 0$.

Matrix memory (1) can display acceptable recognitions when the stored vectors are large-dimensional and sparse. The reason is that sets of large-dimensional and sparse vectors, under many symmetrical statistical distributions of the scalar components, can be quasi-orthogonal (Mizraji *et al.* 1994). This fact implies that for inputs $\hat{\mathbf{f}}_k \in \{\hat{\mathbf{f}}_i\}$ the scalar products in Equation (2) are very small except for $i = k$. In this case, we have a noisy output

$$\mathbf{M}\hat{\mathbf{f}}_k = \boldsymbol{\nu}_k \hat{\mathbf{g}}_k + \text{Noise},$$

where (Signal/Noise) $\approx$ m/K (Anderson 1972, Kohonen 1972, 1977). Consequently, large dimensional input vector can produce accurate recollections.

Until now, we are concerned with non-normalised vectors $\hat{\mathbf{f}}_i$ and $\hat{\mathbf{g}}_i$. It is a natural consequence of most learning algorithms (e.g. Hebbian rules) that the iteration of pairs

$(\hat{\mathbf{f}}_i, \hat{\mathbf{g}}_i)$ enlarges the norms of the stored vectors. Hence, at the end of a learning process the vector norms measure the frequency of contacts with each stored pair. Nevertheless, usually the meaning of a pattern does not depend on its norm. For instance, the image of a human face can be mapped on a large vector whose numerical components are in the interval [0, 1], defining a grey scale with 0 corresponding to white and 1 to black. In this case, we recognise a known face independently of the intensity of the colour code employed. This means that to multiply a vector pattern by a scalar does not change the meaning (or the semantics) of the pattern.

The memory models lead us to assume that completely different patterns, with no overlaps, map on orthogonal vectors; on the contrary, identical patterns map on parallel vectors; and similar patterns form small angles in the corresponding hyperspace (Kohonen 1977). These facts explain the interest of using normalised (i.e. norm = 1) vectors in order to show clearly the structure of the matrix memory, the standardised structure of the patterns and the frequency of their presence during the learning process.

The normalised inputs and outputs of memory M are

$$\mathbf{f}_i = \hat{\mathbf{f}}_i \cdot \|\hat{\mathbf{f}}_i\|^{-1} \quad \text{and} \quad \mathbf{g}_i = \hat{\mathbf{g}}_i \cdot \|\hat{\mathbf{g}}_i\|^{-1}$$

Let

$$\boldsymbol{\mu}_i = \|\hat{\mathbf{g}}_i\| \cdot \|\hat{\mathbf{f}}_i\|.$$

We also define the input and the output sets In $= \{\mathbf{f}_i\}$ and Out $= \{\mathbf{g}_i\}$. Now, the matrix memory (1) can be expressed as follows:

$$\mathbf{M} = \sum_{i=1}^{\mathrm{K}} \boldsymbol{\mu}_i \mathbf{g}_i \mathbf{f}_i^{\mathrm{T}} \tag{3}$$

The numbers $\boldsymbol{\mu}_i$, measuring of the intensity of the experience of the memory with the pairs $(\mathbf{f}_i, \mathbf{g}_i)$, allow us to establish a rank for these pairs. We could, for instance, write matrix (3) according to the decreasing order of the $\boldsymbol{\mu}_i$. Expressing matrix memory (3) in terms of its coefficients, $\mathbf{M} = [\mathbf{M}_{\alpha\beta}]$, we obtain:

$$\mathbf{M}_{\alpha\beta} = \sum_{i=1}^{\mathrm{K}} \boldsymbol{\mu}_i \mathbf{g}_i(\alpha) \mathbf{f}_i(\beta), \tag{4}$$

being $\mathbf{g}_i\,(\alpha)$ and $\mathbf{f}_i\,(\beta)$ the components of vectors $\mathbf{g}_i$ and $\mathbf{f}_i$. This equation shows how the vector patterns are scattered and superimposed over all the matrix coefficients. This property provides a central biological argument to support the importance of these matrix models (see Cooper 1973, Anderson 1995). In addition, these memories are content addressable (Kohonen 1977): a cognitive query is a vector $\mathbf{f}$ able to trigger its output with no need of any specific address.

In a neural environment, quasi-orthogonality of vectorial patterns corresponds to separable perceptions, or to different concepts. In this sense, note that normal vectors show a direct relation between the Euclidean distance $d(\mathbf{f},\mathbf{f}')$ and the scalar product:

$$d^2(\mathbf{f}, \mathbf{f}') = \|\mathbf{f} - \mathbf{f}'\|^2 = 2(1 - \langle \mathbf{f}, \mathbf{f}' \rangle).$$

It is interesting to remark that if the input and output sets, In and Out, are orthonormal, the mathematical structure shown in Equation (3) is the singular value decomposition (SVD) of matrix **M** (Pomi-Brea and Mizraji 1999). Under these conditions, the scalars $\boldsymbol{\mu}_i$ are the singular values, and the input and output vectors $\mathbf{g}_i$ and $\mathbf{f}_i$ are the singular vectors of the rectangular matrix **M**. The orthonormality of In and Out implies that matrix **M** satisfies the SVD conditions:

(a) $\mathbf{M}\mathbf{f}_k = \boldsymbol{\mu}_k \mathbf{g}_k$

(b) $\mathbf{M}^{\mathrm{T}}\mathbf{g}_k = \boldsymbol{\mu}_k \mathbf{f}_k$

for $\mathbf{f}_k \in$ In and $\mathbf{g}_k \in$ Out.

In general, a matrix memory given by Equation (3) does not give us the real SVD. But if $\mathbf{g}_i$ and $\mathbf{f}_i$ are large-dimensional, sparse and quasi-orthonormal vectors, the formal similarity with the real SVD (and even the numerical proximity) can give us a useful point of contact with the vector space models used in the design of search engines, as we are going to see in the next section.

2.3 Cognitive dynamics in modular nets

In modern animals, the nervous system is well-designed to perform actions guided by objectives (or targets). This is especially clear for the neural circuits responsible for many reflexes; these circuits become pre-installed during the embryonic development and depend on the genetic program of the animal. Each reflex involves complex motor (or endocrine) actions oriented towards targets, and represents a basic survival strategy. This propensity to target-dependent actions is also evident in cognitive activities. Even the construction of a verse in a poem requires pursuing in parallel conceptual and acoustic targets. Perceptions are a primary source for the organisation of many of these target-guided cognitive behaviours. The emerging field of computing with perceptions (Zadeh 1999, Martin and Klir 2006) represents a technological counterpart of a very natural computing ability of the human brain.

Human cognitive activities depend, surely, on a large 'net of networks', a structure formed by a network of neural modules. Matrix associative memories can be embedded in these modules. Let us mention that Minsky's 'society of mind' (Minsky 1988) gives his personal vision about the nature and potentialities of this society of neural modules.

Among these modules there are specialised memories displaying different functions. Let us provide a highly schematic description of four important cases.

1. *Heteroassociative memories*. They associate vectorial patterns corresponding to different semantic categories (e.g. a face's image and a name). We analysed this case in the previous sections and its schematic form is

$$\mathbf{M} = \sum_i \mathbf{g}_i \mathbf{f}_i^{\mathrm{T}}.$$

2. *Autoassociative memories*. These memories (studied by Kohonen 1977 and Hopfield 1982, among many others) associate a pattern with itself:

$$\mathrm{S} = \sum_i \mathbf{a}_i \mathbf{a}_i^{\mathrm{T}}.$$

These autoassociative memories show the important ability of reconstructing incomplete patterns and filtering noise (Kohonen 1977, 1989).

3. *Reciprocal associative memories*. These memories establish reciprocal associations between patterns of the same dimension. The simplest representation is as follows. Given a particular set of orthogonal patterns reciprocally associated, $\mathrm{Rel} = \{(\mathbf{a}_i, \mathbf{a}_j), (\mathbf{a}_j, \mathbf{a}_i) : i \neq j\}$, and given the set of paired subindex that characterise the relation, $\mathrm{Ind} = \{(i, j) : (\mathbf{a}_i, \mathbf{a}_j) \in \mathrm{Rel}\}$, the reciprocal (or relational) memories can be represented by

$$\mathrm{R} = \sum_{(i,j)\in \mathrm{Ind}} (\mathbf{a}_i \mathbf{a}_j + \mathbf{a}_j \mathbf{a}_i)^{\mathrm{T}}.$$

Note that $\mathrm{R}\,\mathbf{a}_i = \mathbf{a}_j$ and $\mathrm{R}\,\mathbf{a}_j = \mathbf{a}_i$.

4. *Novelty filters (NF)*. These devices, deeply investigated by Kohonen (1977), are filters that select new data, non-stored previously in the memory database. Given a set of patterns $\mathrm{C} = \{\mathbf{a}_1, \ldots, \mathbf{a}_{\mathrm{K}}\}$ of orthogonal q-dimensional vectors, previously stored in a memory, NF are represented by

$$\mathrm{F} = \mathrm{I} - \sum_i \mathbf{a}_i \mathbf{a}_i^{\mathrm{T}},$$

being $\mathrm{I} \in \mathbb{R}^{\mathbf{q}\times\mathbf{q}}$ the identity matrix. Note that if $\mathbf{a}_k \in \mathrm{C}$ is $\mathrm{F}\,\mathbf{a}_k = 0$. Instead, if $\mathbf{a}_k \perp \mathrm{C}$, we have $\mathrm{F}\,\mathbf{a}_k = \mathbf{a}_k$. A recently updated pattern can be represented by $\mathbf{a}_{\mathrm{up}} = \mathbf{a}_{\mathrm{old}} + \mathbf{a}_{\mathrm{new}}$, with $\mathbf{a}_{\mathrm{old}} \in \mathrm{C}$ and $\mathbf{a}_{\mathrm{new}} \perp \mathrm{C}$; in these case, $\mathrm{F}\,\mathbf{a}_{\mathrm{up}} = \mathbf{a}_{\mathrm{new}}$. NF are matrix memories of the general class given by Equation (3), as can be shown expressing the identity matrix using unit vectors: $\mathrm{I} = \sum_i \mathbf{e}_i \mathbf{e}_i^{\mathrm{T}}$.

Using this kind of memory modules, among many other possible ones, the nervous system can deal with a variety of 'natural queries'. These queries can be proposed by the external world via sensorial events that catch our attention, inducing a behavioural response. But these queries can enter our nervous system via verbal codes created by other humans. And there are inner queries generated by one's own cognitive systems. Plausibly, the human curiosity is a basic stimulus to trigger these inner queries. In all the cases, the queries that penetrate the nervous system must travel across an intricate labyrinth of informational modules. The result of these processes is variable. In some cases, a query gets a satisfactory answer inside the modular net. In other cases, the query activates behaviours that conduce to search for the answer outside: consulting an encyclopaedia, connecting to the WWW, asking an expert, designing an experiment, etc.

3. Term-by-document matrices as memories

3.1 Vector space models for search engines

An important method of analysis of documents stored in computer memories is the vector space modelling, that requires the construction of a rectangular matrix $\mathbf{A} \in \mathbb{R}^{p\times q}$ denominated 'term-by-document matrix' (TD-matrix). Each column of the TD-matrix corresponds to a document, while each row shows the distribution of a particular term (generally a stemmed word, or a symbol) in the set of documents (Berry and Browne 2005). The dimension p of each column of $\mathbf{A}$ is the maximum size of the vocabulary employed in the stored documents. A document normally only uses a very small fraction of the full vocabulary, consequently the TD-matrix $\mathbf{A}$ is usually sparse. Dimension q is the number of stored documents. In the case of WWW is $q \gg p$, but this is not necessarily the cases in other databases. The vocabulary and the documents are pre-processed before constructing the TD-matrices (for details, see Berry and Browne 2005).

We can express the TD-matrix $\mathbf{A}$ as a partitioned matrix showing directly the column documents $\mathbf{d}_j$:

$$\mathbf{A} = [\mathbf{d}_1 \mathbf{d}_2 \cdots \mathbf{d}_q]$$

with $\mathbf{d}_j \in \mathbb{R}^p$. We can also express matrix $\mathbf{A}^{\mathrm{T}}$ using the terms vectors $\mathbf{w}_i \in \mathbb{R}^q$:

$$\mathbf{A}^{\mathrm{T}} = [\mathbf{w}_1 \mathbf{w}_2 \cdots \mathbf{w}_p].$$

A unit vector $\mathbf{e}_h$ of dimension t is a column vector having 1 in position h and 0 in the other $t - 1$ positions. Using unit vectors, we can represent a TD-matrix as an associative memory that associates each document $\mathbf{d}_j$ with an elementary vector $\mathbf{e}_j \in \mathbb{R}^q$:

$$\mathbf{A} = \sum_{j=1}^{q} \mathbf{d}_j \mathbf{e}_j^{\mathrm{T}}. \tag{5}$$

Hence, for $1 \leq k \leq q$ we have

$$\mathbf{A}\,\mathbf{e}_k = \mathbf{d}_k.$$

The unit vector $\mathbf{e}_k$ acts as a key that indicates the address (the k-column) of the document $\mathbf{d}_k$ in the TD-matrix $\mathbf{A}$. The expression (5) is symbolically interesting because it shows matrix $\mathbf{A}$ as an associative memory, but it is practically useless if we want to capture related documents in response to queries (and from databases containing thousand of documents). An equation formally identical to Equation (5) expresses matrix $\mathbf{A}^{\mathrm{T}}$ as a memory, associating terms $\mathbf{w}_i$ with p-dimensional unit vectors $\mathbf{e}_i$.

Documents vectors $\mathbf{d}_j$ are sparse but, in realistic cases, they are not orthogonal. One of the objectives of the representation of data using TD-matrices is to discover procedures capable of reducing the dimension of the document space and, at the same time, of producing a gain in our capacity to identify clusters of related documents. A leader idea was the construction of pseudo-documents representative of such clusters, and the proposed procedure begins obtaining the SVD for the TD-matrix (Deerwester *et al.* 1990, Berry *et al.* 1995). This decomposition is unique, and the result looks as follows:

$$\mathbf{A} = \sum_{i=1}^{r} \boldsymbol{\sigma}_i \mathbf{u}_i \mathbf{v}_i^{\mathrm{T}}. \tag{6}$$

$\mathbf{u}_i \in \mathbb{R}^p$ and $\mathbf{v}_i \in \mathbb{R}^q$ are the singular vectors, σ_i are the singular values, $r = \mathrm{rank}$ (A). The sets $\{\mathbf{u}_i\}$ and $\{\mathbf{v}_i\}$ of singular vectors are orthonormal, and the terms of matrix $\mathbf{A}$ are ordered according to the decreasing values of their scalar coefficients: $\boldsymbol{\sigma}_1 \geq \boldsymbol{\sigma}_2 \geq \cdots \geq \boldsymbol{\sigma}_r$.

Let us transform expression (7) using transposition:

$$\mathbf{A}^{\mathrm{T}} = \sum_{i=1}^{r} \boldsymbol{\sigma}_i \mathbf{v}_i \mathbf{u}_i^{\mathrm{T}}. \tag{7}$$

Now, we define a query as a vector $\mathbf{d}' \in \mathbb{R}^p$, that represents a document containing a small group of keywords. Asking for documents with a query $\mathbf{d}'$ means to operate as follows:

$$\mathbf{A}^{\mathrm{T}}\mathbf{d}' = \sum_{i=1}^{r} \boldsymbol{\sigma}_i \mathbf{v}_i \langle \mathbf{u}_i, \mathbf{d}' \rangle.$$

The output of this query $\mathbf{d}'$ has the dimensions of a term vector (equal to the total number of documents), and the final results can provide a ranked indication of the positions where the searched documents are placed.

The LSA uses the SVD of TD-matrix to obtain the previously mentioned reduction of dimensionality in the document space. The idea is to construct a rank-reduced matrix $\mathbf{A}_k$ that can be considered as a good approximation of the original TD-matrix matrix $\mathbf{A}$ retaining the first k larger terms of SVD (Deerwester *et al*. 1990, Berry *et al*. 1995, Berry and Browne 2005):

$$\mathbf{A}_k = \sum_{i=1}^{k} \boldsymbol{\sigma}_i \mathbf{u}_i \mathbf{v}_i^{\mathrm{T}}; \quad k \ll r. \tag{8}$$

The criteria for defining 'good approximation' is usually based on the Eckart and Young theorem, that gives the error of approximating $\mathbf{A}$ by $\mathbf{A}_k$ in terms of the Frobenius norm (Berry and Browne 2005):

$$\|\mathbf{A} - \mathbf{A}_k\|_F = \sqrt{\boldsymbol{\sigma}_{k+1}^2 + \boldsymbol{\sigma}_{k+2}^2 + \cdots + \boldsymbol{\sigma}_r^2}.$$

Consequently, if the singular values coming after the k-term are small then the Frobenius norm of error could be small.

Recently, Valle-Lisboa and Mizraji showed, using arguments based on Perron–Frobenius theorem and perturbation theory, that a selected set of non-adjacent singular values $\boldsymbol{\sigma}_{\varphi(\mathrm{i})}$ could generate more performant rank-reduced matrices $\mathbf{A}_L$ (Valle-Lisboa and Mizraji 2007). In this approach, the singular values $\sigma_{\varphi(i)}$ act as thematic labels in the document space, and allow capturing the virtual documents $u_{\varphi(i)}$ more relevant for each subject. The rank-reduced matrix $\mathbf{A}_L$ is given by

$$\mathbf{A}_L = \sum_{i=1}^{s} \boldsymbol{\sigma}_{\varphi(i)} \mathbf{u}_{\varphi(\mathrm{i})} \mathbf{v}_{\varphi(i)}^{\mathrm{T}}, \tag{9}$$

where $1 \leq \varphi(\mathrm{i}) \leq r$, $\varphi(i)$, being the position of the thematic label, $i = 1, \ldots, s$. The value of s measures the total amount of subjects, normally being $s \ll r$.

It is worthwhile to mention the similarity between the biological memory model given in Equation (3), and the TD-matrices expressed as in Equation (7) (and in the transposed versions of matrices (8) and (9)). The question now is to decide if there are some interesting conceptual reasons behind the formal convergence between the biological and the technological models.

3.2 Explaining the formal convergence between biological and technological models

Let us first mention that the biological and the technological models are concerned with the processing of large number of parallel variables, and that in both cases vectors are natural representations for these variables.

The matrix memory models assume that neurons are firing into a restricted linear range. For this reason, Equation (3) must be considered as an approximation with a restricted functional range of validity. Nevertheless, these matrices show important properties that are present in brain memories: the placement and distribution of information in synaptic coefficients, the robustness facing neuronal or synaptic deterioration, the noise suppression abilities, and the capacity to recognise partial patterns.

Plausibly, during the learning process the biological memories build up prototypes that emerge after the contact with similar patterns. 'Concepts' belong to this class of prototype. In the frameworks of the matrix memory models, a prototype is assumed to be an average taken over vectors learnt under similar contexts, and associated with similar responses. The theory (Cooper 1973, Kohonen 1977), and some beautiful numerical experiments (i.e. Kohonen *et al*. 1989) show that learning processes can self-organise prototypes as averages. Hence, any input vector of a matrix memory can be constructed from weighted averages. Imagine that each input vector $\hat{\mathbf{f}}_i$ results from the contact with N_i vectors $\hat{\mathbf{f}}_{ij}$. We can express $\hat{\mathbf{f}}_i$ as a function of the average $\langle \mathbf{f}_i \rangle$ in the following way:

$$\langle \mathbf{f}_i \rangle = \frac{1}{\mathrm{N}_i} \sum_{j=1}^{\mathrm{N}_i} \boldsymbol{\pi}_{ij} \mathbf{f}_{ij} \qquad (10)$$

and

$$\hat{\mathbf{f}}_i = \mathrm{N}_i \langle \mathbf{f}_i \rangle$$

In Equation (10) N_i is the number of similar normalised vectors $\mathbf{f}_{ij}$, and the $\boldsymbol{\pi}_{ij}$ are weight factors that absorb the modules of the original vectors $\hat{\mathbf{f}}_{ij}$: $\boldsymbol{\pi}_{ij} = \|\hat{\mathbf{f}}_{ij}\|$.

The possibility of installing into the memories concepts defined from averages, implies that separable concepts map on vectors with very low dispersion; hence, tightened around its averages. This geometry of the memory space is imposed by the properties of the external world, and requires very large-dimensional vectors. The signal-noise relation deduced by many authors (and commented in Section 2) shows that in this kind of matrix memories the amount of data stored (K vector pairs for memory (3)) must long be surpassed by vector dimensionality m ($\mathrm{K} \ll m$), an important difference with technological TD-matrices.

We now focus the SVD of TD-matrices. Using Equations (5) and (6), via post-multiplication by each one of the singular vectors $\mathbf{v}_h$, we can express the singular vectors $\mathbf{u}_h$ as a linear combination of the original documents $\mathbf{d}_j$ and then, multiplying by (q/q), express this linear combination as a weighted average:

$$\mathbf{u}_h = \frac{1}{q} \sum_{j=1}^{q} \left(\frac{\mathbf{q}}{\boldsymbol{\sigma}_h} \right) \mathbf{v}_h(j) \mathbf{d}_j, \quad (h = 1, 2 \ldots, r) \qquad (11)$$

where $\mathbf{v}_h(j)$ is the j-component of the singular vector $\mathbf{v}_h$. We recall that a singular vector $\mathbf{v}_h$ is associated with the distribution of word h on the set of documents. This equation shows that we can reasonably consider vector $\mathbf{u}_h$ as a pseudodocument, created as a weighted average of documents $\mathbf{d}_j$. The structure of the coefficients $[(\mathbf{q}/\boldsymbol{\sigma}_h)\mathbf{v}_h(j)]$ indicates that a document $\mathbf{d}_j$ does not influence the virtual document $\mathbf{u}_h$ if the j-component of vector $\mathbf{v}_h$ is zero. This means that document $\mathbf{d}_j$ does not contribute with words to the pseudoterm vector $\mathbf{v}_h$. Note that as documents $\mathbf{d}_j$ are not normalised, a large singular value in the denominator can be balanced by the norm of the documents. An expression dual of Equation (11) is obtained for the pseudoterm vectors $\mathbf{v}_h$, that become expressed as weighted averages of the term vectors $\mathbf{w}_h$.

What happen if now we approximate the complete TD-matrix using the rank-reduced matrix (8)? The same procedure described above lead to the same equations, but restricted to a small set of singular vectors:

$$\mathbf{u}_{h'} = \frac{1}{\mathrm{q}} \sum_{j=1}^{q} \left(\frac{\mathbf{q}}{\boldsymbol{\sigma}_{h'}} \right) \mathbf{v}_{h'}(j) \mathbf{d}_j, \quad (h' = 1, 2, \ldots, k). \tag{12}$$

We can guess that the extirpation in Equation (7) of terms corresponding to many singular vectors $\mathbf{u}_h (h > k)$ with small norms (and corresponding to the small singular values $\boldsymbol{\sigma}_{k+1}, \boldsymbol{\sigma}_{k+2}, \ldots, \boldsymbol{\sigma}_r$), means extracting noise: the noise in the query processing $\mathbf{A}^{\mathrm{T}}\mathbf{d}'$ could be significantly higher than in the reduced rank query $\mathbf{A}_k^{\mathrm{T}}\mathbf{d}'$ (but remark that this noise reduction depends on the accuracy of the rank-reduction procedure). The final result could be the increase of the specificity of the set of pseudodocuments $\mathbf{u}_{h'}$. These facts can help to understand the interesting performances showed by LSA and similar rank-reduction procedures.

Each one of the pseudodocuments from Equation (12) could be considered as a kind of conceptualised document, with some analogy with the empirical concepts described for neural models by the averages (10). The absence of many noise-generating pseudodocuments in Equation (12) is a way to reduce the (Noise/Signal) ratio using representative pseudodocument $\mathbf{u}_h$. In conclusion, the 'conceptualisation' via weighted averages with reduced dispersion seems to be a common feature, shared by the matrix models for biological memories and by the TD-matrices processed by LSA and other rank-reduction procedures.

4. Contexts and thematic packing

The human nervous system has a natural ability to generate thematic clusters with data coming from perceptual experiences. The biological mechanisms leading to this performance are not clearly understood yet, and have been focused by many theoretical approaches. In this section, we describe a procedure where the clustering of data is sustained by the existence of multiplicative vectorial contexts (Mizraji 1989, Pomi and Mizraji 2004).

4.1 Multiplicative contexts

Consider a memory D given by

$$\mathrm{D} = \sum_{\mathrm{i}} \mathbf{w}_i (\mathbf{y}_i \otimes \mathbf{z}_i)^{\mathrm{T}} \tag{13}$$

This expression uses the Kronecker product $\otimes$ that for arbitrary matrices $\mathbf{Z} = [\mathbf{z}_{ij}] \in \mathbb{R}^{m\times n}$ and $\mathbf{Z}' = [\mathbf{z}'_{ij}] \in \mathbb{R}^{p\times q}$ is defined as $\mathbf{Z}\otimes\mathbf{Z}' = [\mathbf{z}_{ij}\mathbf{Z}'] \in \mathbb{R}^{(mp)\times(nq)}$ (Graham 1981). We describe some important properties of the Kronecker products. Let $\mathbf{U}$, $\mathbf{U}'$, $\mathbf{V}$ and $\mathbf{V}'$ be matrices (including vectors) and $\boldsymbol{\alpha}, \boldsymbol{\beta}$ be scalars. Then we have:

(1) $$(\mathbf{U}\otimes\mathbf{V})^{\mathrm{T}} = \mathbf{U}^{\mathrm{T}}\otimes\mathbf{V}^{\mathrm{T}}$$

(2) $$(\mathbf{U}\otimes\mathbf{V})(\mathbf{U}'\otimes\mathbf{V}') = (\mathbf{U}\mathbf{U}')\otimes(\mathbf{V}\mathbf{V}')$$

(3) $$(\boldsymbol{\alpha}\mathbf{U})\otimes(\boldsymbol{\beta}\mathbf{V}) = \boldsymbol{\alpha}\boldsymbol{\beta}(\mathbf{U}\otimes\mathbf{V}).$$

(property (2) requires conformable matrices). As a consequence of (1) and (2), Q-dimensional column vectors **a**, **b**, **c** and **d** satisfy the following equations:

$$(\mathbf{a}\otimes\mathbf{b})^{\mathrm{T}}(\mathbf{c}\otimes\mathbf{d}) = (\mathbf{a}^{\mathrm{T}}\mathbf{c})(\mathbf{b}^{\mathrm{T}}\mathbf{d}) = \langle\mathbf{a},\mathbf{c}\rangle\langle\mathbf{b},\mathbf{d}\rangle.$$

This is a central property to understand how memories like (13) work (remark that the Kronecker product between scalars becomes the ordinary product).

The cardinal property of context-dependent memories like (13) is the improvement of the selectivity in matrix memories, due to the fact that these memories filter the incoming data using two scalar products. The double filtering can be illustrated easily. For an input $(\mathbf{y}\otimes\mathbf{z})$, the output is given by

$$\mathrm{D}(\mathbf{y}\otimes\mathbf{z}) = \sum_i \langle\mathbf{y}_i,\mathbf{y}\rangle\langle\mathbf{z}_i,\mathbf{z}\rangle\mathbf{w}_i \tag{14}$$

Note that even if $\mathbf{z}_i$ and $\mathbf{z}$ are parallel vectors (i.e.: the same pattern) the output is zero provided that $\mathbf{y}_i$ and $\mathbf{y}$ are orthogonal. Hence $\mathbf{z}_i$ and $\mathbf{z}$ are parallel, but $(\mathbf{y}_i\otimes\mathbf{z}_i)$ and $(\mathbf{y}\otimes\mathbf{z})$ are orthogonal. This implies that identical patterns with different meaning under different contexts can be separated by these memories. This improvement of the selectivity produces a rich functional versatility, which allows the use of these matrix memories in a variety of neural models (Mizraji 1989, Pomi and Mizraji 2001, 2002, Valle-Lisboa *et al.* 2005, Pomi and Olivera 2006). In addition, these context-memories are able to execute the dyadic logical gates, and provide an alternative algebraic representation for the Boolean functions that govern certain cellular automata (Mizraji 2006).

4.2 Input contexts

We show now a detailed version of memory (13). Let a vector $\hat{\mathbf{p}}_i \in \mathbb{R}^{n'}$ be the codified mapping of a thematic context. To each context $\hat{\mathbf{p}}_i$ we associate a set of vectors $\hat{\mathbf{f}}_{kj} \in \mathbb{R}^n$ that represent patterns. Let $\hat{\mathbf{d}}_{ij} \in \mathbb{R}^m$ be the output associated to the patterns $\hat{\mathbf{f}}_{kj}$ under the context $\hat{\mathbf{p}}_i$. Being $\mathbf{f}_{kj}$, $\mathbf{p}_i$ and $\mathbf{d}_{ij}$ their normalised versions, we can consider a matrix memory having the following structure:

$$\mathrm{H} = \sum_{i,j}\boldsymbol{\omega}_{ij}\mathbf{d}_{ij}\left(\mathbf{p}_i\otimes\sum_{k=1}^{\mathrm{N}_j}\boldsymbol{\varphi}_{kj}\mathbf{f}_{kj}\right)^{\mathrm{T}}. \tag{15}$$

The scalars are given by $\boldsymbol{\omega}_{ij} = \left\|\hat{\mathbf{d}}_{ij}\right\|\|\hat{\mathbf{p}}_i\|$ and $\boldsymbol{\varphi}_{ij} = \|\hat{\mathbf{f}}_{kj}\|$. Using the weighted average

$$< \mathbf{f}_j >= \frac{1}{\mathrm{N}_j}\sum_{k=1}^{\mathrm{N}_j}\boldsymbol{\varphi}_{kj}\mathbf{f}_{kj}, \tag{16}$$

we can express the memory H as

$$\mathrm{H} = \sum_{i,j}\boldsymbol{\omega}_{ij}\mathbf{d}_{ij}(\mathrm{p}_i\otimes\mathrm{N}_j < \mathbf{f}_j >)^{\mathrm{T}}.$$

Defining a normalised average $\mathbf{f}_j =< \mathbf{f}_j > \| < \mathbf{f}_j > \|^{-1}$, we finally obtain the expression

$$\mathrm{H} = \sum_{i,j}\boldsymbol{\lambda}_{ij}\mathbf{d}_{ij}(\mathbf{p}_i\otimes\mathbf{f}_j)^{\mathrm{T}}, \tag{17}$$

a normalised version of memory (13) with a weight given by

$$\boldsymbol{\lambda}_{ij} = \boldsymbol{\omega}_{ij}(\mathrm{N}_j \| < \mathbf{f}_j > \|). \tag{18}$$

We can reorder all the terms of Equation (17) according to the decreasing magnitude of the $\boldsymbol{\lambda}_{ij}$, and use this order to create a new subindex function $\tau = \tau(i, j)$, $\tau = 1, 2, \ldots, \mathrm{K}$. Defining $\mathbf{d}_\tau = \mathbf{d}_{ij}$, $\mathbf{h}_\tau = \mathbf{p}_i \otimes \mathbf{f}_j$, and $\boldsymbol{\lambda}_\tau = \boldsymbol{\lambda}_{ij}$ we obtain

$$\mathrm{H} = \sum_{\tau=1}^{\mathrm{K}} \boldsymbol{\lambda}_\tau \mathbf{d}_\tau \mathbf{h}_\tau^{\mathrm{T}}. \tag{19}$$

If the sets $\{\mathbf{d}_\tau\}$ and $\{\mathbf{h}_\tau\}$ are orthonormal, the Equation (19) is the SVD of matrix H. In this case, the singular values $\boldsymbol{\lambda}_\tau$ absorb the size of the stored thematic cluster, due to its proportionality to $\mathrm{N}_j \| < \mathbf{f}_j > \|$. In this framework, some important thematic clusters but with reduced frequency in the learning experience of memory (19), have reduced values of $\mathrm{N}_j \| < \mathbf{f}_j > \|$ and are small terms, located in the tail terms of matrix (19). Obviously, the full functionality of these memories databases needs not to neglect these low-frequency data. If the output set $\{\mathbf{d}_\tau\}$ is not orthogonal, memory (19) only can be considered as an approximated SVD provided that vectors $\mathbf{d}_j$ are large, sparse and quasi-random.

The Kronecker product is a useful device to represent the conceptual thematic packing carried out by the human brain. This packing can be the consequence of the nature of the sensorial channel transporting the information (e.g. music by hearing), and some of these packs map on relatively precise topographical brain territories because they are mainly organised according to the large-scale neuroanatomic connectivity. For other brain processes (symbolic reasoning, imagination of non-existent scenarios), the packing by thematic contexts can be the consequence of the semantic nature of the cultural database installed by learning.

The possibility of using techniques like LSA for the analysis of documents produced by humans (or even the analysis of DNA patterns produced by the biological evolution subject to selective constraints), implies the pre-existence of thematic packing. Let us mention in passing that Montemurro and Zanette (2002) studies some measurable consequences on the ensemble of Shakespeare's dramas when the plays are randomly mixed. Each one of the Shakespeare's plays represents a thematic partition labelled by very specific words (e.g. Macbeth). Valle-Lisboa and Mizraji (2007) uses the procedure described by Montemurro and Zanette (2002) to study the effects of word shuffling between documents, on the LSA performances. The TD-matrices in Valle-Lisboa and Mizraji (2007) were constructed from sets of well-partitioned documents; an interesting result is that when partitions are destroyed by shuffling the words, the profile of singular values only suffers small changes. This discovery shifted the attention towards the properties of singular vectors, and lead to approximate the TD-matrices by Equation (9).

4.3 Input–output contexts

During cognitive processes, like thought, neural modules engage in a dialogue to each other, in such a way that the output of a memory M1 is the input of a memory M2. Hence, it is interesting to explore the situation in which the output of a memory is a context-modulated vector with the structure

$$\hat{\mathbf{d}}_{ij} = \hat{\mathbf{p}}'_i \otimes \hat{\mathbf{g}}_{ij},$$

where $\hat{\mathbf{p}}_i'$ is the context and $\hat{\mathbf{g}}_{ij}$ is the associated output. The Kronecker product allows to factorise a pair of associated input–output vectors

$$(\hat{\mathbf{p}}_i' \otimes \hat{\mathbf{g}}_{ij})(\hat{\mathbf{p}}_i' \otimes \hat{\mathbf{f}}_{ij})^{\mathrm{T}} = \boldsymbol{\chi}_i \mathbf{p}_i' \mathbf{p}_i^{\mathrm{T}} \otimes \boldsymbol{\mu}_{ij} \mathbf{g}_{ij} \mathbf{f}_{ij}^{\mathrm{T}};$$

in the second member of this equation the vectors are normalised, being $\boldsymbol{\chi}_i = \|\hat{\mathbf{p}}_i'\|.\|\hat{\mathbf{p}}_i\|$ and $\boldsymbol{\mu}_{ij} = \|\hat{\mathbf{g}}_{ij}\|.\|\hat{\mathbf{f}}_{ij}\|$. If a given pair of contexts $(\mathbf{p}_i', \mathbf{p}_i)$ corresponds to a set of associations $\{(\hat{\mathbf{f}}_{ij}, \hat{\mathbf{g}}_{ij})\}$, a context-dependent matrix memory the matrix memory H can be written as

$$\mathrm{H} = \sum_i \left(\boldsymbol{\chi}_i \mathbf{p}_i' \mathbf{p}_i^{\mathrm{T}} \otimes \sum_j \boldsymbol{\mu}_{ij} \mathbf{g}_{ij} \mathbf{f}_{ij}^{\mathrm{T}} \right). \tag{20}$$

Vectors $\mathbf{f}_{ij}$ can be interpreted similarly as the weighted and normalised averages $\mathrm{f_j}$ of Equation (17), the subindex i labeling the context pair. Equation (20) illustrates the interesting fact that these contextualised inputs and outputs create a set of sub-modules inside memory H, each one of these sub-modules being a particular associative memory M_i with the structure

$$\mathrm{M}_i = \sum_j \boldsymbol{\mu}_{ij} \mathbf{g}_{ij} \mathbf{f}_{ij}^{\mathrm{T}}. \tag{21}$$

Usually, these memories M_i are scattered into matrix H in a complex way, due to the structure of the Kronecker products, but they are functionally dissected by the context. Therefore, the context can allow a given input to find its corresponding memory (21), hidden inside memory module H.

An interesting illustrative example is given when contexts $\mathbf{p}_i$ and $\mathbf{p}_i'$ are both the same unit n-dimensional unit vector $\mathbf{e}_i$ (we assume that there are s different memories M_i). Note that

$$\mathbf{e}_i \mathbf{e}_i^{\mathrm{T}} = [\boldsymbol{\delta}_{ki} \boldsymbol{\delta}_{ij}] \in \mathbb{R}^{n \times n}, \quad i,j = 1, \ldots, n$$

being $\boldsymbol{\delta}_{\alpha\beta} = 1$ iff $\alpha = \beta$ and $\boldsymbol{\delta}_{\alpha\beta} = 0$ iff $\alpha \neq \beta$. In this case, memory H is given by

$$\mathrm{H} = \sum_{i=1}^{s} \left(\mathbf{e}_i \mathbf{e}_i^{\mathrm{T}} \otimes \mathrm{M}_i \right),$$

a block-diagonal matrix having different associative memories placed in its diagonal:

$$\mathrm{H} = \begin{bmatrix} \mathrm{M}_1 & 0 & \ldots & 0 \\ 0 & \mathrm{M}_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \mathrm{M}_\mathrm{s} \end{bmatrix} \tag{22}$$

Matrix (22) is a rectangular matrix with rectangular submatrices (including the 0's rectangular submatrices). Note that in this extremely simple example, the contexts direct the different associated patterns towards different topographical regions. Thus, a pattern $\tilde{\mathbf{f}}$, eventually present in two different memory blocks M_i and associated with two different outputs, is oriented by the context to select the corresponding output. In the real world this kind of selection happens, for instance, in the case of the attribution of a meaning to a polysemous word.

The existence of context modulation of inputs and outputs has important consequences. The output of a memory, generated by a query **f**, can be distributed between many memory modules due to the anatomic connectivity. Nevertheless, if the output is $\mathbf{p}\otimes\mathbf{g}$, the context **p** acts as a password and it is only accepted by memories having this vectorial password in their database. In this way, a same query **f** can generate many associative trajectories according to the sequence of contexts involved. Within this framework, the interaction between contexts and patterns seems to light the 'tunnels' by which information moves forward into the brain.

5. Final comments

5.1 Clustering documents under contextual deprivation

The decoding of language by the brain is, plausibly, the result of the interaction of many neural modules. We can assume that one of the modules is a matrix memory that associates very specific words with thematic contexts, and that the context triggered by a specific word labels the subsequent flux of words until another specific word shifts the thematic context. This idea has been presented in a recent model (Valle-Lisboa and Mizraji 2005, preliminary communication). Let us assume now that the conceptual meaning of the words is stored in structured matrix memories similar to the memory shown in Equation (20). We can speculate that a verbal document is an 'echo' of a memory like (20), flattened by the disappearance of explicit contexts. Hence, the information becomes condensed in a matrix memory H' that keeps associations deprived of contexts:

$$\mathrm{H}' = \sum_{i'=1}^{\mathrm{K}} \boldsymbol{\mu}_{i'} \mathbf{g}_{i'} \mathbf{f}_{i'}^{\mathrm{T}}. \tag{23}$$

The subindex i' imply a reordering of the associations, with the possible loss of the original clustering imposed by the contexts in memory (20).

Under the situation illustrated by Equation (23), information retrieval from large collections of documents can be considered as a kind of reverse engineering, exerted on cognitive products in which contexts are absent or implicit. In memory (20), a context-embedded input $\tilde{\mathbf{f}}$ leads to its specific associated meaning **g**; instead in memory (23) the same $\tilde{\mathbf{f}}$ deprived of its contexts can produce a confused result $\boldsymbol{\mu}\mathbf{g} + \boldsymbol{\mu}'\mathbf{g}' + \boldsymbol{\mu}''\mathbf{g}'' + \cdots$. Similarly, a search engine acting on a collection of documents can get a confused retrieval if the query is vague and does not have clear thematic labels. The challenge faced by search engine designers is similar to the problem of reconstructing memory (20) from matrix (23). Surely, this is an ill-defined problem, but the brain displays some ability to deal with this kind of ill-defined data (e.g. three-dimensional interpretation from a two-dimensional retinal projection).

Let us emphasise an important qualitative difference between the biological memory model and the TD-matrix. An associative memory like (20) and the context-free version (23) are constructed from neural vectors whose dimensions are assumed 'anatomically' fixed, and the outputs representing words are large-dimensional vectors. Instead, each word in the TD-matrix is symbolically represented by a single position in the document vector (hence, by a scalar). A thematically clustered neural ensemble of words is represented in the TD-matrix by a subset of vector positions. Disjoint vocabularies imply disjoint positions in the corresponding TD-matrix. We have shown (Valle-Lisboa and Mizraji 2007) that LSA can produce block matrices for disjoint cases or noisy block matrices for non disjoint cases, being the block's size related to the size of the thematic cluster. We have a difference here with the neural models (20) and (22), where the blocks have the same dimension, and the thematic richness is shown by the number of terms of the sums. In neural models (20) and (22) non-disjoint vocabularies can be completely separated by the thematic contexts.

In the case of textual documents, highly specific keyword can help to recover a thematic context. Because different keywords can be associated with the same context, many documents can belong to the same thematic cluster even if the documents do not share keywords. Methods like LSA look for the best possible thematic clustering using keywords that belong to the documents. A possibility for further development of artificial information retrieval procedures is to create word-context databases that can be used together with the documents databases in order to perform a parallel evaluation: instead of [query $\rightarrow$ documents] alone, we can perform first [query $\rightarrow$ contexts] and then [(query + contexts) $\rightarrow$ documents]. This parallelism requires (as in idealised memories (15) and (20)) an adequate composition of documents and contexts.

5.2 Braining the Web

The technological information networks, and particularly the World Wide Web, transit nowadays an accelerated process of complexification. The topology of the network that connects web pages seems inextricable. Nevertheless, until recently the WWW seemed to maintain the properties of a small-world (for early reports see Watts and Strogatz 1998, Albert, Jeong and Barabasi 1999, 2000, Kleinberg 2000) and to display some scaling properties (scale-free within a range) (Barabasi and Albert 1999). The problem of the efficient search of targets in a network having the small-world connectivity lead Kleinberg (2000) to investigate a 'navigation' procedure that implies a decentralised search, with short delivery times if long-range links decay with distance as a power law. For a two dimensional lattice a critical exponent for decay's power law, that rends navigation optimal, has been reported (Kleinberg 2000).

Hyperlink structure of the WWW has been the basis for generating a hierarchical organisation of web pages using different approaches (e.g. the HITS Method or Page Rank, see Chapter 7 in Berry and Browne 2005). These topological procedures are pre-required in order to reduce the number of documents and, in this way, allow methods like LSA to become operative.

The navigation into the WWW requires a continuous man–machine interaction, where after the answer to an initial query the human operator responds with a sequence of decisions. These decisions depend upon the quality of the documents captured by the search engine. Advanced queries imply further refinements and different ways to put data in context. It is plausible that some of the strategies displayed by neural cognitive modules can inspire technological counterparts. As examples of 'neuralisation' of the search, let us mention the potential design of: (1) autoassociative memories to improve the queries and to clean noise, (2) novelty filters to only retain new information from recently modified documents, (3) reciprocal memories to establish relational information structures and (4) associative memories with the power to extract concepts from large collections of documents creating in this way thematic clusters (LSA is an important approach in this direction). Surely many other possibilities will arise, depending on our understanding of the neural strategies for information retrieval.

A pertinent question is if the WWW tends to self-organise into diversified modules, partitioned by different classes of contexts. Many approaches are currently investigated to detect the hidden thematic structure of the WWW (see Eckmann and Moses 2002).

If a tendency towards thematic modularisation is confirmed and remains stable, future search engines could be designed to display performances similar to those of the nervous system, with the capacity to automatically organise explorations across many modules without any man–machine interaction. The human operator will retain the fundamental initiative of introducing into the search engine contextualised queries. In addition, imitating neural actions, these engines should accept as data the long-term targets that motivate the search.

Let us end this speculative section considering some challenging issues. In this article, we assume that search engines are technological devices that allow the human mind to penetrate into a universe of information coded and stored in computers. One can ask if the eventual partial replacement of the human operator implies the existence of a close analogy between the search process in biological and artificial entities. We assume that the *a priori* answer must be negative. On the one hand, current neural models are largely provisional and we have a long way to go before arriving at a fully reliable neural theory (nevertheless, we believe that these neural models are conceptually sound and computationally powerful). On the other hand, it is well known that very different designs can produce objects with similar behaviours. But we cannot discard that the human-driven process of construction of search engines, that must retrieve information stored in very large distributed networks, finished creating procedures that mimic those generated during the natural developmental evolution of the human brain. In fact, the mathematical similarities between biological memory models and LSA suggest that this can be the case. In addition, the incorporation in modern search engines of 'neuromimetic' devices, based on neural models, creates the possibility of learning performances and adaptive behaviours, manifested for instance in personalised search preferences or adaptive spam filters.

Acknowledgements

I would like to thank Cristina Arruti, Andrés Pomi and Juan Carlos Valle-Lisboa for their comments and encouragement. This work was partially supported by PEDECIBA and CSIC (UdelaR). I wish to thank one of the anonymous reviewers for asking the challenging questions that motivate the last paragraph of this article.

Notes on contributor

Eduardo Mizraji is Professor of Biophysics of the Department of Cell and Molecular Biology, Faculty of Sciences, Universidad de la Republica (UdelaR), Montevideo, Uruguay. His research interests include information processing in extended neural systems. He received an MD degree in 1975 from the Faculty of Medicine, Universidad de la Republica, and a DEA in Applied Mathematics from the University of Paris V in 1977.

References

Ashby, W.R., 1960. *Design for a brain*. New York: Wiley.

Albert, R., Jeong, H. and Barabasi, A.-L., 1999. Diameter of the World Wide Web. *Nature*, 401, 130–131.

Albert, R., Jeong, H. and Barabasi, A-L., 2000. Error and attack tolerance of complex networks. *Nature*, 406, 379–381.

Anderson, J.A., 1972. A simple neural network generating an interactive memory. *Mathematical Biosciences*, 14, 197–220.

Anderson, J.A., 1995. *An introduction to neural networks*. Cambridge: The MIT Press.

Anderson, J.A. and Rosenfeld, E., eds, 1988. *Neurocomputing*. Cambridge: The MIT Press.
Barabasi, A.-L. and Albert, R., 1999. Emergence of scaling in random networks. *Science*, 286, 509–512.
Berners-Lee, T. and Fischetti, M., 1999. *Weaving the web: the original design and ultimate destiny of the World Wide Web*. San Francisco: Harper.
Berry, M.W. and Browne, M., 2005. *Understanding search engines: mathematical modeling and text retreval*. 2nd ed. Philadelphia: SIAM.
Berry, M., Dumais, S. and O'Brien, G., 1995. Using linear algebra for intelligent information retrieval. *SIAM review*, 37, 573–595.
Cooper, L.N., 1973. *A possible organisation of animal memory and learning*. Proceedings of the nobel symposium on collective properties of physical systems. Sweden: Aspensagarden.
Cooper, L.N., 1980. Sources and limits of human intellect. *Daedalus*, 109 (2), 1–17.
Cooper, L.N., 2000. Memories and memory: A physicist's approach to the brain. *International journal of modern physics A*, 15 (26), 4069–4082.
Deerwester, S., Dumais, S., Furnas, G., Landauer, T. and Harshman, R., 1990. Indexing by latent semantic analysis. *Journal of the american society for information science*, 41, 391–407.
Delmas, J. and Delmas, A., 1958. *Voies et centres nerveux-Introduction à la Neurologie*. Paris: Masson.
Eckmann, J-P. and Moses, E., 2002. Curvature of co-links uncover hidden thematic layers in the World Wide Web. *Proceedings of the national academy of sciences*, 99, 5825–5829.
Edsall, J.T. and Wyman, J., 1958. *Biophysical chemistry*. New York: Academic Press.
Feynman, R.P., Leighton, R.B. and Sands, M., 1963. *The Feynman lectures on physics*. Massachussets: Addison-Wesley.
Graham, A., 1981. *Kronecker products and matrix calculus with applications*. Chichester: Ellis Horwood.
Hopfield, J.J., 1982. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the national academy of sciences of the USA*, 9, 2554–2558.
Kandel, E.R. and Schwartz, J.H., 1985. *Principles of neural science*. New York: Elsevier.
Kleinberg, J., 2000. Navigation in a small world. *Nature*, 406, 845.
Kohonen, T., 1972. Correlation matrix memories. *IEEE transactions on computers*, C-21, 353–359.
Kohonen, T., 1977. *Associative memory: a system theoretical approach*. New York: Springer.
Kohonen, T., Oja, E. and Lehtiö, P., 1989. Storage and processing of information in distributed associative memory systems. In: G.E. Hinton and J.A. Anderson, eds. *Parallel models of associative memory*. New Jersey: Erlbaum, 129–167.
Landauer, T. and Dumais, S., 1997. A solution to Plato's problem: the latent semantic analysis theory of acquisition, induction and representation of knowledge. *Psychological review*, 104, 211–240.
Martin, O. and Klir, G.J., 2006. Retranslation in computing with perceptions. *International journal of general systems*, 35, 655–674.
Minsky, M., 1988. *The society of mind*. New York: Touchston, Simon & Schuster.
Mizraji, E., 1989. Context-dependent associations in linear distributed memories. *Bulletin of mathematical biology*, 51, 195–205.
Mizraji, E., Pomi, A. and Alvarez, F., 1994. Multiplicative contexts in associative memories. *Biosystems*, 32, 145–161.
Mizraji, E., 2006. The parts and the whole: inquiring how the interaction of simple subsystems generates complexity. *International journal of general systems*, 35, 395–415.
Montemurro, M.A. and Zanette, D.H., 2002. Entropic analysis of the role of words in literary texts. *Advances in complex systems*, 5, 7–17.
Nass, M.M. and Cooper, L.N., 1975. A theory for the development of feature detecting cells in visual cortex. *Biological cybernetics*, 19, 1–18.
Nauta, W.J.H. and Feirtag, M., 1986. *Fundamental Neuroanatomy*. New York: W.H. Freeman.
Pomi-Brea, A. and Mizraji, E., 1999. Memories in context. *Biosystems*, 50, 173–188.
Pomi, A. and Mizraji, E., 2001. A cognitive architecture that solves a problem stated by Minsky. *IEEE Transactions on systems, man, and cybernetics - Part B: cybernetics*, 31, 729–734.
Pomi, A. and Mizraji, E., 2004. Semantic graphs and associative memories. *Physical review E*, 70 (0666136), 1–6.
Pomi, A. and Olivera, F., 2006. Context-sensitive autoassociative memories as expert systems in medical diagnosis. *BMC medical informatics and decision making*, 6 (39), 1–11.
Shaw, G.L. and Palm, G., eds, 1988. *Brain theory*. Singapore: World Scientific.
Sporns, O. and Zwi, J.D., 2004. The small worlds of the cerebral cortex. *Neuroinformatics*, 2, 145–162.
Valle-Lisboa, J.C., Reali, F., Anastasía, H. and Mizraji, E., 2005. Elman topology with sigma pi units: an application to the modeling of verbal hallucinations in schizophrenia. *Neural networks*, 18, 863–877.
Valle-Lisboa, J.C. and Mizraji, E., 2007. The uncovering of hidden structures by latent semantic analysis. *Information sciences*, 177, 4122–4147.
Valle-Lisboa, J.C. and Mizraji, E., 2005. Un modelo neuronal de procesamiento de lenguaje basado en herramientas de búsqueda de información (Abstract). *XI Jornadas de la Sociedad Uruguaya de Biociencias*, Minas: Uruguay.
Zadeh, L.A., 1999. From computing with numbers to computing with words – from manipulation of measurements to manipulation of perceptions. *IEEE transactions circuits systems I fundamental theory applied*, 45, 105–119.
Watts, D.J. and Strogatz, S.H., 1998. Collective dynamics of 'small-world' networks. *Nature*, 393, 440–442.