Applied Mathematics and Computation 245 (2014) 261–264

Contents lists available at ScienceDirect

Applied Mathematics and Computation

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

How much information should we drop to become intelligent?

Ophir Nave [a,*], Yair Neuman [b], Leonid Perlovsky [c], Newton Howard [d]

[a] Lev Academic Center (JCT), Department of Mathematics, Ben-Gurion University of the Negev, Israel
[b] Department of Education, Ben-Gurion University of the Negev, Israel
[c] School of Engineering and Applied Sciences, Harvard University, USA
[d] Synthetic Intelligence Lab, Massachusetts Institute of Technology, USA

ARTICLE INFO

Keywords:
Cognition and physics
Set partition
Entropy
Interdisciplinary research

ABSTRACT

Cognitive processing by intelligent systems involves the deletion of information in favor of higher level abstractions. This process can be addressed through the physics of computation but a formal model that explains this process has not been proposed yet. In this short paper, we propose a model that through physical constraints only generates optimal solution to the collapse of n objects into n sets. A numerical simulation of the model results in a logarithmic function of information loss and condensation that perfectly fits our knowledge of cognitive processes.

1. Introduction

Intelligence is usually conceived in simple additive terms. For instance, the more declarative-factual knowledge one possesses the more intelligent s(he) is considered to be.

However, a remarkable characteristic of intelligent systems is that they involve the erasure of information in favor of higher level abstractions. One who holds a large amount of information without being able to drop some of it in favor of abstractions is an *Idiot Savant*. This ‘Savant Syndrome’ is beautifully illustrated in Borges famous story “Funes the Memorious” [1] in which a person who has an incredible and unlimited memory cannot think beyond the myriad concrete details that populate his memory.

Cognition is mainly about producing categories and operating on categories [2] and categories are abstractions that involve the loss of information. When a child develops the category of Cats, for instance, he ignores many details of the particular cats that he has encountered in favor of the abstract category that has no tail or mustache. Despite the differences between particular cats they are grouped into a single category which is different from the Dogs category for instance.

The process of erasing information is favor of abstractions exists in a variety of living systems and for a good physical reason [3]. In the physics of computation, it has been argued by Bennett and Landauer [4] in one of their more philosophical papers that information should be equated with *differences* and that a process of computation, which is an irreversible process of producing an output from an input necessarily involves the loss of information/differences.

Given that any intelligent system necessarily involves the loss of information, one may ask how much information should we drop to become “intelligent” and more generally how can we model the loss of information in the formation of categories. This question has interesting consequences not only for our understanding of human and non-human cognition but for the design of intelligent systems specifically in the age of Big Data.

* Corresponding author.
E-mail address: naveof@bgu.ac.il (O. Nave).

http://dx.doi.org/10.1016/j.amc.2014.07.029

Nomenclature

n	number of objects
x_i	objects $i = 1, \ldots, n$
E	energy function of the system
T	temperature of the system
k	Boltzmann's constant
Z	partition function
Γ	phase space
S	entropy
ϕ	probability density function (PDF)
$S(,)$	Stirling number

2. The theoretical approach

In this paper, we try to address the above challenge by providing a formal model of categories formation and information loss. Although intensive work has been conducted on data categorization, our paper follows a totally different line by studying the mathematical constraints imposed on set partition regardless of real world applications.

Our starting point is by considering abstraction as the *collapse of several differences into a single set*. More specifically, we approach the problem of abstraction through the partitions of a set where each partition is considered to be a category.

The lattice formed through this process includes an upper and lower bound that are clearly uninformative. For instance, let's imagine a universe that includes only 4 objects. A universe in which all objects are members of the same set is an undifferentiated universe with no informative value, the same as the universe in which each and every object stands "in-and-for-itself" to use an old philosophical phrase coined by Husserl. Hence, cognition and intelligence exist *in-between* these upper and lower bounds of the lattice.

The total number of partitions of an n objects set is calculated through the Bell Number. The Bell Number gives us the problem space that covers all potential partitions. However, an intelligent system cannot use this mathematical space as it grows exponentially in a way which is far beyond the limits of efficient cognitive processing computation. For example, for a set of five elements there are 52 possible partitions. It is unreasonable to assume that an intelligent system that processes n objects, automatically computes the Bell Number for all possible partitions. Therefore, while mathematics gives us the potential space of set partitions, at least as an analytical frame of reference, we must move to the physical realm in order to impose real-world constraints on the number of possible partitions and the way information is lost while collapsing differences into categories. The next section develops the physical–mathematical model of this process with the aim of better understanding information loss in abstraction as evident in the formation of categories.

3. The model, the simulations and the results

We start with the basic assumption that one should "pay" in energetic terms for a refinement partition of a given set. In other words, for collapsing objects into subsets we pay a price. Hence, for a system processing n objects and categorizing them into n categories/sets the question is how to optimally behave. Our model draws on statistical mechanics [11,12].

We start with a set of objects $x_1, x_2, \ldots, x_n$. For these sets we define the following probability density function:

$$\phi(x_1, x_2, \ldots, x_n) = e^{-E(x_1, x_2, \ldots, x_n)/kT}/Z, \tag{3.1}$$

where the partition function Z is given by:

$$Z = \int_\Gamma e^{-E(x_1, x_2, \ldots, x_n)/kT} dx_1 dx_2 \ldots dx_n. \tag{3.2}$$

We define the partition function through the integral because our phase space, Γ, is defined by the Bell Number that may be approximated by the integral. Our next step is to define the Gibbs Entropy of the system for each group of subsets i.e., level in Hasse diagram as follows:

$$S_i = -k \int_{\Gamma_i} \phi_i(x_{i_1}, x_{i_2}, \ldots, x_{i_m}) ln(\phi_i(x_{i_1}, x_{i_2}, \ldots, x_{i_m})) dx_{i_1} dx_{i_2} \ldots dx_{i_m}, \tag{3.3}$$

where Γ_i is the phase space defined by the Stirling number of the second kind. Let S_{In} and S_{Out} be the Entropy associated with the upper and the lower nodes of the Hasse diagram. Because of the symmetry between Γ_i and Γ_{i+1} the partition function for the i level in the Hasse diagram is twice the partition function of the $i+1$ level. Hence $\phi_i = \phi_{i+1}/2$ and

$$S_i = -2k\int_{\Gamma_i}\frac{1}{2}\phi_{i+1}ln\left(\frac{\phi_{i+1}}{2}\right)dx_{i_1}dx_{i_2}\dots dx_{i_m}$$
$$= -k\int_{\Gamma_{i+1}}\phi_{i+1}ln(\phi_{i+1})dx_{i_1}dx_{i_2}\dots dx_{i_m} + kln(2)\int_{\Gamma_{i+1}}\phi_{i+1}dx_{i_1}dx_{i_2}\dots dx_{i_m} = S_{i+1} + kln(2) \Rightarrow S_i - S_{i+1} = kln(2). \quad (3.4)$$

It means that for each transition from the level i to the level $i+1$ in the Hasse diagram, the Entropy increases by $kln(2)$. This is the price the system pays for the refinement.

The total entropy of the system is therefore:

$$S_{In} - S_{out} = (n-2)\cdot kln(2). \quad (3.5)$$

Our aim is to minimize the Entropy of the system, hence we define the constrain:

$$F(x_{i_1}, x_{i_2}, \dots, x_{i_m}) = S_{In} - S_{out} - (n-2)\cdot kln(2) = 0. \quad (3.6)$$

Let us define a new phase space with the following axis: The *x-axis* is the objects' level in the Hasse diagram and the *y-axis* is the Stirling number of the second kind $S(n,k)$. For example, a set with 4 objects has 4 level: $\{1,2,3,4\}$ and 4 Stirling number of the second kind: $\{1,7,6,1\}$ respectively.

The next step is to find the optimal loss of information when we refine in the Hasse diagram. For this purpose, $\phi_i(x_{i_1}, x_{i_2}, \dots, x_{i_m}) < \phi_{i+1}(x_{i_1}, x_{i_2}, \dots, x_{i_{m_l}})$ iff $S(m,k) < S(l,k)$ where ϕ_i is the probability density function at the level i as normalized by the partition function. We define $\phi(x_1, x_2, \dots, x_n)$ by $\phi(x_1, x_2, \dots, x_n) = \sum_{i=1}^{n}\phi_i(x_{i_1}, x_{i_2}, \dots, x_{i_m})$. In order to optimally delete information, minimize the above function as follows:

$$min(\phi(x_1, x_2, \dots, x_n))\Big|_{\sum_i S(i,k)=Bell} = min\sum_{i=1}^{n}\phi_i(x_{i_1}, x_{i_2}, \dots, x_{i_m})\Big|_{\sum_i S(i,k)=Bell}. \quad (3.7)$$

In our analysis, we should minimized the average function ϕ as present in Eq. (3.7) with the constrains (3.6). For this purpose we applied Lagrange multiplier for the system.

For testing the above model, we used a computer simulation with 0–100 objects. In Fig. 1 we present the number of optimal partitions computed by the model as a function of the set's number of objects.

As we can see the number of categories/partitions produced by our model as a function of the number of objects is described by a logarithmic function found to be: $1.7\cdot Log(3x-0.5)$. This finding is in line with the psychological knowledge pointing to the centrality of the logarithmic function in a variety of cognitive processes [5].

Following [4] information is considered in terms of differences. The number of differences for a given number of objects is $(n^2-n)/2$. Fig. 2 presents the results of our model for the loss of information as measured in the number of differences, as we move from n objects to the optimal number of partitions. The loss of information was calculated as follows:

$$\rho(n) = \frac{\frac{n^2-n}{2} - \frac{(1.7\cdot Log(3x-0.5))^2 - 1.7\cdot Log(3x-0.5)}{2}}{\frac{n^2-n}{2}} \quad (3.8)$$

As we can see from Fig. 2 the ratio of information loss behave like logarithmic function and the loss of information quickly convergences to a complete loss of information. It means that as the number of objects increases, most of the information is lost in favor of a few categories.

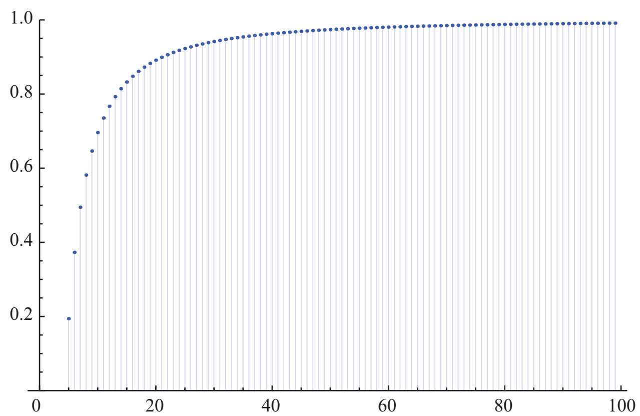

Fig. 1. The number of optimal set partition as a function of the given number of objects.

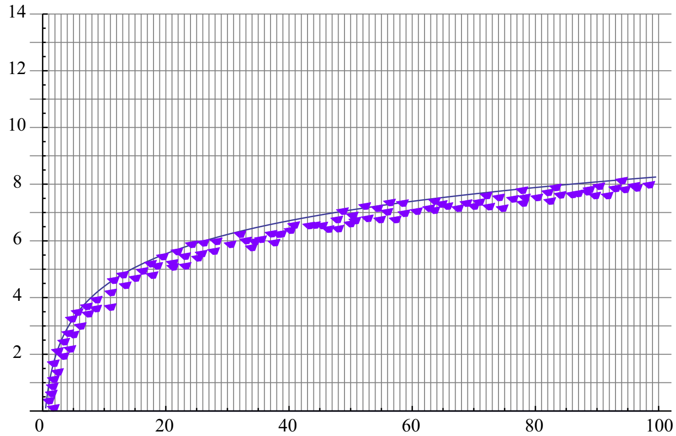

Fig. 2. Loss of information as a function of the set refinement. The x-axis is the number of objects, the y-axis is the ratio of information loss.

4. Discussion

In this short paper, we present a general physical model that explains the optimal information loss during categorization. By using minimal physical constrains our model produces a logarithmic function of information compression that adheres to our accommodated knowledge in cognition. Therefore the model can be used not only for explaining a variety of cognitive processes involving the loss of information in favor of higher-level abstractions but can be also used for the design of artificial intelligent systems and for Big Data processing.

The developed model seems to correspond with human brain mechanisms. Perception and cognition processes involve an approximate hierarchy of mental representations from objects, to situations, to more abstract concepts higher up in the hierarchy. The mental hierarchy is composed of vague and abstract categories-representations at the top and concrete categories-representations at the bottom. Every act of perception or cognition involves matching bottom-up (BU) and top-down (TD) signals between representations at adjacent layers [7,8,6]. The matching of BU and TD signals should take into account the energetic efficiency of the brain [9] and the need to optimally remove some information along the way. Things are much more complex as TD signals are intermingled with linguistic and higher-order cognitive processes [10] and therefore we should take into account that a more realistic model should take into account the real-world various constraints imposed on information loss. As the model we proposed is general and generic any future real world application of the model should take into account the contextual constraints facing the system under inquiry. Therefore this paper should be considered only as an invitation for further research and cooperation.

Acknowledgment

The authors would like to thank Yiftach Neuman for his constructive advice in the mathematical formulation.

References

[1] J.L. Borges, Funes the memorious. Labyrinths: Selected stories and other writings, New Directions Publishing Corporation Library of Congress Catalog, 1962, pp. 59–66.
[2] S. Harnad, To cognize is to categorize: cognition is categorization, in: C. Lefebvre, H. Cohen (Eds.), Handbook of Categorization in Cognitive Science, Elsevier, 2005.
[3] Y. Neuman, Reviving the Living: Meaning Making in Living Systems, Elsevier, Oxford, 2011.
[4] C.H. Bennett, R. Landauer, The fundamental physical limits of computation, Sci. Am. 253 (1985) 48–56.
[5] L.R. Varshney, J.Z. Sun, Why do we perceive logarithmically, Significance 10 (2013) 28–31.
[6] M. Bar, K.S. Kassam, A.S. Ghuman, J. Boshyan, A.M. Schmid, A.M. Dale, et al, Top-down facilitation of visual recognition, USA: Proc. Nat. Acad. Sci. 103 (2006) 449–454.
[7] L.I. Perlovsky, Neural Networks and Intellect: Using Model-Based Concepts, Oxford University Press, New York, NY, 2001 (3rd printing).
[8] L.I. Perlovsky, Toward physics of the mind: concepts, emotions, consciousness, and symbols, Phys. Life Rev. 3 (2006) 22–55.
[9] N. Howard, Energy paradox of the brain, Brain Sci. J. 1 (2012) 46–61.
[10] L.I. Perlovsky, Language and cognition, Neural Networks 22 (2009) 247–257, http://dx.doi.org/10.1016/j.neunet.2009.03.007.
[11] J. Sethna, Statistical Mechanics: Entropy, Order Parameters, and Complexity, Oxford University Press, 2006.
[12] R. Bowley, M. Sanchez, Introductory Statistical Mechanics, Oxford University Press, 2000.