Information-theoretic framework for unsupervised activity classification

Frédéric Kaplan[1] Verena V. Hafner[2]

(1) Sony CSL Paris, 6 rue Amyot 75005 Paris, France, kaplan@csl.sony.fr

(2) Technische Universität Berlin, Fakultät für Elektrotechnik und Informatik, Berlin, Germany, vvh@ieee.org

Abstract

This article presents a mathematical framework based on information theory to compare multivariate sensory streams. Central to this approach is the notion of *configuration*: a set of distances between information sources, statistically evaluated for a given time span. Because information distances capture *simultaneously* effects of physical closeness, intermodality, functional relationship and external couplings, a configuration can be interpreted as a signature for specific patterns of activity. This provides ways for comparing activity sequences by viewing them as points in an *activity space*. Results of experiments with an autonomous robot illustrate how this framework can be used to perform unsupervised activity classification.

keywords: activity classification, information metrics, unsupervised clustering

1 Introduction

This article presents a method for performing unsupervised activity classification. It can in principle be applied to any multisensory stream of data, such as the ones produced by an autonomous robot or other kinds of embodied context-aware devices. The method is based on a mathematical framework adapted to the comparison of temporally-extended multimodal sequences of data. It provides ways for comparing activity sequences by viewing them as points in an *activity space*. In this article, we consider more specifically the relevance of this method for mapping embodied sensorimotor organizations in a unified space. This issue is central in the challenging quest for autonomous development [WMP+01] as the possibility to find structural similarity between sensorimotor schemas is thought to be crucial for the emergence of higher-level forms of cognition.

The starting point of this approach is to suppose that behavioral complexity can be captured by studying the active organization in time of information coming from several *sources*. Sensors mounted on a robot or recording of neural activity are examples of information sources for artificial and living systems, respectively. Using information sources as a basic modelling unit has several advantages.

Models can be framed using the mathematically well-defined tools of information theory. Data coming from various sources (symbol, numerical) can be blended in an unified framework. And finally, the same approach can be used to study both living and artificial systems.

Information theory has historically been mainly concerned with information transmission between a sender and a receiver through a channel [SW62, CT91]. However more recently several lines of research have focused on defining theoretical measures for addressing information integration [TES98, SP03], information distance between information sources [Kul68]. Crutchfield in particular has shown that the space of information sources can be equipped with a simple *metric* [Cru90]. Therefore, it is possible to consider a form of *spatial* relationship between sources and to adapt some of the vocabulary and tools of *geometry* to the domain of information theory[1].

The fact that two information sources are related in terms of information (i.e. that they are close in the information space) means, in an informal way, that knowing the state of one permits to know things about the state of the other. This happens when there is a mutual causal relationship between the two sources or when information coming from both sources result from common causing factors. In our context, this can mean several things. The sources can be physically related and activated by the same localized stimuli. They can be functionally related and activated as a result of a particular control pattern. They can be only related in time, but still informationally related as the organism interacts with a slowly changing continuous environment. They can be related as a result of an external coupling, like in the case where the organism engages in reciprocal interaction with peers. Information is a common currency that permits to blend these multiple factors[2].

A set of distances between information sources, statistically evaluated for a given time span, specifies a *configuration*. Because information distances capture *simultaneously* effects of physical closeness, intermodality, functional relationship and external couplings, a configuration can be interpreted as a signature for specific patterns of activity. Configurations themselves can be viewed as points in a higher-dimensional space, the space of all possible configurations. In this space, related configurations are grouped together and independent ones are further apart.

The geometrical approach presented in this article is directly inspired by several methods concerning unsupervised map building recently described in the field of artificial intelligence and autonomous robotics. Pierce and Kuipers present a method for building maps of a sensory apparatus out of raw uninterpreted sensory data [PK97]. This so-called sensory reconstruction method is based on various distances between sensors such as a normalized Hamming distance metric and a frequency metric. Sensors are clustered into subgroups based on their relative distance. The dimensionality of each subgroup can then be computed, related sensors can be projected to form a sensor map. Building on this sensory reconstruction method, Olsson, Nehaniv and Polani [ONP04b, ONP04a] have suggested to use the in-

[1] It should be noted that although we call our approach *geometrical* it is not directly related to Amari's work on information geometry [AN00], which is based on investigations about the geometrical structures of the manifold of probability distributions.

[2] Using the standard terminology from classical statistics, information sources are said to be dependent, independent or conditionally independent on each other.

formation metric defined by Crutchfield [Cru90] as a more interesting measure of the distance between two information sources. They have conducted experiments with various sensor sets including visual and proprioceptive sensors on an AIBO robot. Related approaches were also investigated by Kuniyoshi's research team [KYO+04]. Most of these approaches interpret such sensory reconstruction methods as a way of building maps of sensors in an unsupervised manner. Some of these works draw an analogy with somatosensory maps discovered in the brain.

The approach described in this article extends and, more importantly, reinterprets this method. The sensory reconstruction method is well-adapted to address processes underlying the emergence of behavioral complexity, but it may be misleading to interpret it as a formation of a body map. A particular configuration captures not only aspects of an agent's embodiment, but also reflects the agent's current activities and the situated nature of its interaction with the environment. The rest of the paper describes the approach in a more formal manner, discusses the possible interpretations of the abstractions introduced and presents results of preliminary experiments showing how this framework can be used to capture the behavioral organization of an autonomous robot.

2 Definitions

2.1 Information sources

Let Ω be a system equipped with a set of n information sources $\{X_i\}$. Information sources can be proprioceptives (corresponding to internal states of Ω), heteroceptives (corresponding to information about Ω's environment), or both. Measurements are obtained out of each information source. These measurements typically correspond to elements belonging to an arbitrary number of bins[3] . At each time t, an element x_i corresponds to the information source X_i. The following notation will be used: $X_i(t) = x_i$.

At any time t, the state of Ω is captured by the vector $X(t)$.

$$X(t) = (X_1(t), X_2(t), \ldots, X_n(t)) \tag{1}$$

Values of $X(t)$ can potentially depend on the environmental context in which the system Ω is placed, the current activity of Ω, as well as its physical and structural organization.

2.2 Information distance between two information sources

The conditional entropy for two information sources X_i and X_j can be calculated as

$$H(X_j|X_i) = -\sum_{x_i}\sum_{x_j} p(x_i, x_j) \log_2 p(x_j|x_i) \tag{2}$$

[3]As usual with such a kind of framework, the choice of the bins is an important issue [Sch00]. For instance, in the case of sensors with a non-linear response, it could be important to increase the sampling around the inflexion point of the curve. A good solution to avoid this problem is to introduce adaptive binning [ONP05]. In such a case, the size of the bins is variable and chosen in a way that maximizes the entropy for each sensor.

where $p(x_j|x_i) = p(x_j, x_i)/p(x_i)$.

$H(X_j|X_i)$ is traditionally interpreted as the uncertainty associated with X_j if the value of X_i is known.

Crutchfield defines the normalized information distance between two information sources as:

$$d(X_j, X_i) = \frac{H(X_i|X_j) + H(X_j|X_i)}{H(X_i, X_j)} \tag{3}$$

d is a metric for the space of information sources [Cru90][4]. This means that it has the three properties of symmetry, equivalence and triangle inequality.

- $d(X, Y) = d(Y, X)$ follows directly from the symmetry of the definition
- $d(X, Y) = 0$ if and only if X and Y are recoding-equivalent (in the sense defined by Crutchfield [Cru90]).
- $d(X, Z) \leq d(X, Y) + d(Y, Z)$

As $H(X_i, X_j) = H(X_i) + H(X_j|X_i)$, $d \leq 1$.
$d = 1$ means that the two sources are independent.

The existence of this metric implies that the space of information has a topological structure. This permits interesting development such as the continuity of functions on information sources or the convergence of sequences of information sources. However, these properties are not central for the issues discussed in this article.

2.3 Configuration

Let us define a *configuration* as the information distance matrix **D** corresponding to the different distances between the information sources X_i over a temporally extended sequence τ.

$$\mathbf{D}_\tau = \left\{ \begin{array}{c} d(X_1, X_1) \; ... \; d(X_1, X_n) \\ d(X_2, X_1) \; ... \; d(X_2, X_n) \\ \dots\dots\dots\dots\dots\dots \\ d(X_n, X_1) \; ... \; d(X_n, X_n) \end{array} \right\} \tag{4}$$

As $d(X_i, X_i) = 0$, elements of the diagonal are all zero. As $d(X_i, X_j) = d(X_j, X_i)$, $\mathbf{D}_\tau$ is symmetrical.

[4] This is its main advantage compared to mutual information $MI(X_i, X_j) = H(X_i) + H(X_j) - H(X_i; X_j)$. Other information metrics exist like Fisher information used on statistical manifolds ([AN00], see also [LL05]). Those metrics are usually defined locally. To obtain the metric distance between two points on an information manifold, one needs to integrate over geodesics. In our context, Crutchfield's metric is simpler and more appropriate

$\mathbf{D}_\tau$ summarizes some important aspects about the organization of the information sources of the system Ω over a given period of time, by specifying which sources are related in terms of information and which ones are independent.

An attractive way of representing the information captured by a configuration is to draw a map showing the relative position of points $\{\mathbf{p_i}\}$ standing for the different sources. Multidimensional scaling methods can be used for that purpose [CC94]. Each couple of points $\mathbf{p_i}$ and $\mathbf{p_j}$ should satisfy:

$$||\mathbf{p_i} - \mathbf{p_j}|| = d_{i,j} \tag{5}$$

where $||\mathbf{p_i} - \mathbf{p_j}||$ is the Euclidean distance between the position of the ith and jth point and $d_{i,j}$ the corresponding distance in the matrix $\mathbf{D}_\tau$. There are $\frac{n(n-1)}{2}$ equations to satisfy. A set of n points of dimension $n-1$ permits to solve these equations given this set of constraints optimally, but in order to get a lower dimension representation approximation must be taken. A standard method can be used to determine a good dimensionality for projecting a given set of data [PK97]. In the rest of the article, two-dimensional projections are used for illustrative purposes although they may not be the optimal ones.

The information contained in $\mathbf{D}_\tau$ can be represented in two dimensions using a relaxation algorithm. The algorithm is an iterative procedure of positioning points in a two-dimensional space in such a way that the metric distance between two points in this map is as close as possible to the distance in the distance matrix (other algorithm exist but they use additional information like the relative orientation of connections between points [Haf00, DMS02]).

The algorithm used consists of an iteration of two simple steps. First, each information source X_i is randomly assigned to a point $\mathbf{p_i}$ on a two-dimensional plane.

1. The force f_i on each point $\mathbf{p_i}$ is computed as:

   $$f_i = \sum f_{ij}$$

   where

   $$f_{ij} = (||\mathbf{p_i} - \mathbf{p_j}|| - d(X_i, X_j))\frac{(\mathbf{p_j} - \mathbf{p_i})}{||\mathbf{p_j} - \mathbf{p_i}||}$$

2. Each point $\mathbf{p_i}$ is moved according to the force f_i:

   $$\mathbf{p_i} = \mathbf{p_i} + \eta f_i$$

   where $\eta = 1/n$.

The resulting map partly depends on the initial conditions of the iteration. Several examples of such maps are presented in the following sections.

3 Interpretations

3.1 Configurations and embodiment

Several authors have argued that distance between information sources could be interpreted as a kind of body map capturing the spatial organization of the different sensors considered [PK97, ONP04b, ONP04a]. To illustrate this point, let us consider an artificial retina made of $n = l^2$ spatially related information sources $\{X_1, X_2, ..., X_n\}$. The retina is organized as a $l \times l$ squared grid. A source X_j (not placed in a border of the grid) has potentially four adjacent sources X_{j-1}, X_{j+1}, X_{j-l}, X_{j+l}. The retina moves over a grayscale image (see Figure 1).

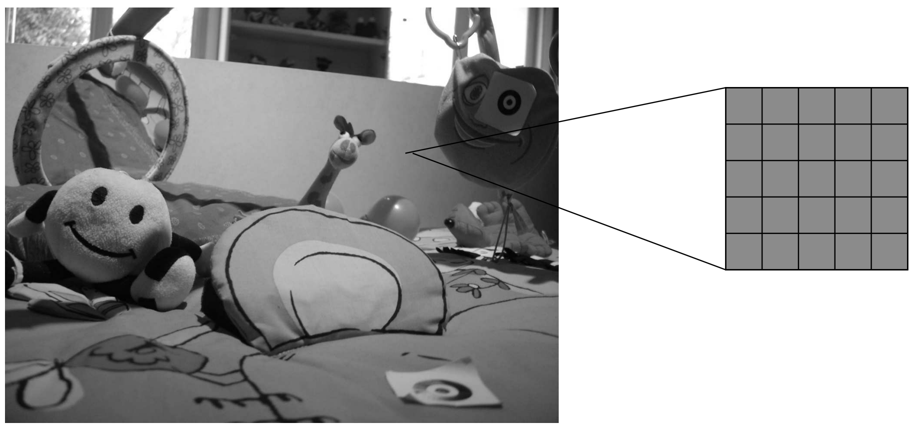


Figure 1: A 5×5 artificial retina scanning a grayscale picture

Figure 2 presents the distance matrix and corresponding two-dimensional map for $l = 5$, $\delta = 1$. In such an environment and using this simple control of the movement, a configuration emerges where a grid that characterizes the structure of the system is reproduced in the information space. Adjacent sources in the system are also the closest in the terms of information distance. This result supports the interpretation of the emerging configuration as a body map.

3.2 Configurations and environment structure

Let us now consider a second example, even simpler than the previous one. Information sources of this system are a set of n sensors $\{X_1, X_2, ..., X_n\}$ arranged sequentially. The system reads a linear tape made of a sequence of symbols $v(t)$ (Fig 3). At each timestep, the tape is shifted, so that :

$$X_1(t) = v(t) \tag{6}$$

and for $1 < i \leq n$

$$X_i(t) = X_{i-1}(t-1) \tag{7}$$

Information sources of this system are related in time: two adjacent sources perceived the same stimuli with an interval of one timestep. This means that sources will be related in terms of information,

a) b)

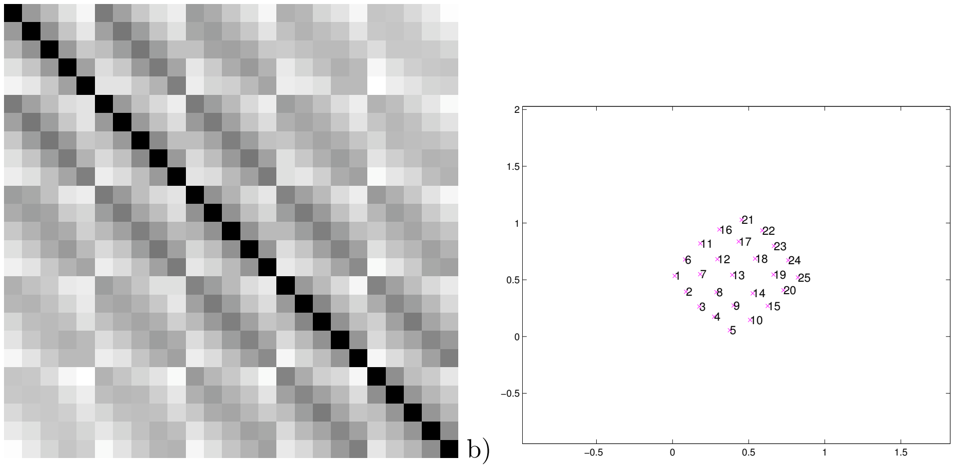


Figure 2: Artificial retina example. (a) Information distance matrix obtained for $l = 5$, $\delta = 1$ Small distances are plotted in blue, big distances in red. (b) Corresponding two-dimensional map

if the environment (i.e. the tape) with which the system interacts possesses a structure in which spatially close symbols are related.

Figure 3: A sequential reader

Two situations will be considered. In the first one $v(t) = sin(\alpha \cdot t)$ is a function that varies in a continuous manner. Figure 4 (a) represents the distance matrix obtained for $n = 25$ in this environment. Temporally close information sources are characterized by small information distances, whereas more temporally distant sources have large information distances. In such a case information distance directly correlates with temporal distance. Figure 4 (b) represents the two-dimensional map associated with the configuration specified in the distance matrix. In this case, like in the previous one, the sequential topology of the information sources of this system is captured by a configuration matrix.

In contrast, let us consider the case where the tape is structured in a random manner. In such a case, spatially close symbols bear no relation. The resulting distance matrix is shown in figure 5 (a) . Every source is equidistant from the others. The system has no particular structure. In the corresponding map shown on figure 5 (b), information sources are spatially distributed in order to be as distant from one another as possible.

This simple example illustrates how a given configuration captures neither the structure of the system, nor the structure of the environment but the *relationship* between the two. The emergent geometry corresponds in this case to the situated nature of the system.

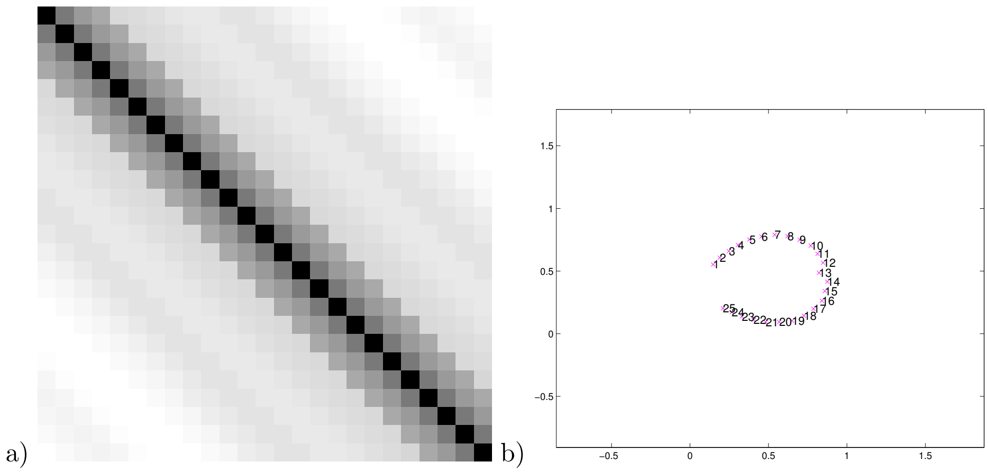


Figure 4: Sequential reader. (a) Distance matrix obtained for $n = 25$ in a sinusoidal environment. Small distances are plotted in blue, big distances in red. Temporally close sources are related in terms of information distance. (b) Corresponding two-dimensional map

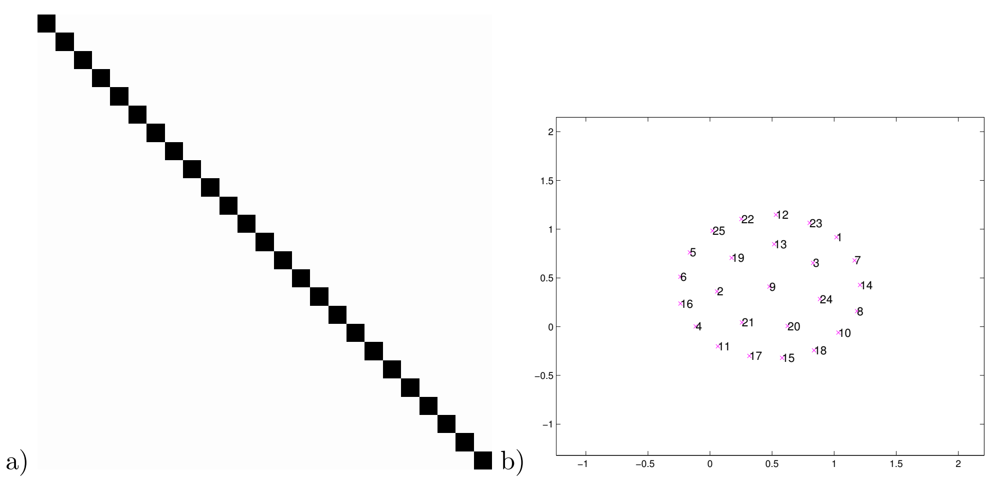


Figure 5: Sequential reader. (a) Distance matrix obtained for $n = 25$ in a random environment. Small distances are plotted in blue, big distances in red. No structure is present. (b) Corresponding two-dimensional map

3.3 Configurations and activity

The central hypothesis of this article is that configurations can also be used as signature for specific temporally-extended activities. If this is indeed the case, a relevant way for comparing two sequences $\tau 1$ and $\tau 2$ would be to compare their respective configuration. Various distances for comparing configuration matrices can be envisioned. A simple one is the following:

$$d(\mathbf{D}_{\tau 1}, \mathbf{D}_{\tau 2}) = \sqrt{\sum_{kl} (m_{kl} - n_{kl})^2} \tag{8}$$

where m_{kl} and n_{kl} are the components of the kth line and the lth column of respectively the $\mathbf{D}_{\tau 1}$ and $\mathbf{D}_{\tau 2}$ matrix.

More generally for a given set of sequences, it is therefore possible to iterate the process and consider the distance matrix of configurations

$$\mathbf{\Delta} = \left\{ \begin{array}{c} d(\mathbf{D_1}, \mathbf{D_1}) \; ... \; d(\mathbf{D_1}, \mathbf{D_n}) \\ d(\mathbf{D_2}, \mathbf{D_1}) \; ... \; d(\mathbf{D_2}, \mathbf{D_n}) \\ \dots\dots\dots\dots\dots\dots \\ d(\mathbf{D_n}, \mathbf{D_1}) \; ... \; d(\mathbf{D_n}, \mathbf{D_n}) \end{array} \right\} \tag{9}$$

As $d(\mathbf{D_i}, \mathbf{D_i}) = 0$, elements of the diagonal are all zero. As $d(\mathbf{D_i}, \mathbf{D_j}) = d(\mathbf{D_j}, \mathbf{D_i}) = 0$, $\mathbf{\Delta}$ is symmetrical.

In the next section we will study how the configuration of configurations $\mathbf{\Delta}$ capturing the relative organization of a set of sequences is a useful abstraction for characterizing a collection of related activities.

4 Experiments with an autonomous robot

Experiments of this section involve an autonomous four-legged robot (Sony AIBO ERS-7, dimensions: 180 (W) x 278 (H) x 319 (D) mm). A set of 18 information sources $\{X_1, X_2, ...X_{18}\}$ is used in these experiments. They correspond to distance sensors and proprioceptive position sensors (Table 1. Each leg has 3 degrees of freedom, as well as the head. Infrared distance sensors are mounted on the head and on the main body[5] (see table 1 and figure 6 for details of the 18 sensors used in this set of experiments).

The robot can be programmed to do various kinds of behavior, that range from simple motor skills like walking to integrated forms of behaviors involving more complex sensorimotor coordination like chasing a ball. In its regular autonomous behavior the robot can switch between these various kinds of behavior depending on the evolution of its internal drives and opportunities present in the environment

[5] The robot has a colour camera mounted above its mouth, electro-static touch sensors, paw sensors, LED lights, all of which are not used in the present experiment but have been exploited in other research conducted with this robot (e.g. [SK00, HK05b])

Table 1: Sensors used on the robot

Number	Name
1-3	IR distance sensors (head far, head near, chest)
4-6	head pan,tilt,neck (proprioceptive sensors)
7-9	right front leg (position of the 3 joints)
10-12	right hind leg (position of the 3 joints)
13-15	left front leg (position of the 3 joints)
16-18	left hind leg (position of the 3 joints)

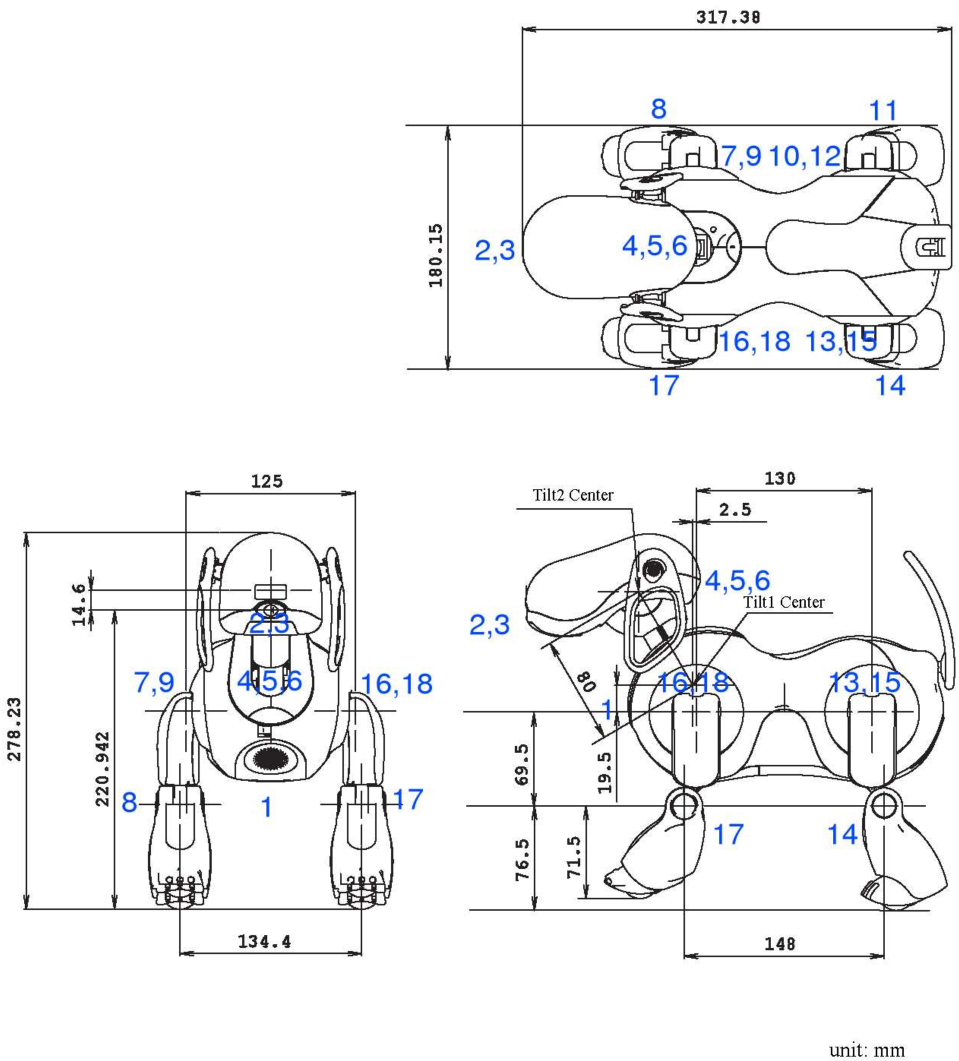


Figure 6: Top, front, and side view of an AIBO robot with blue numbers indicating the indices for information sources used in the experiment (Figure adapted from the Sony AIBO technical manual)

[FK98, FKD00]. Although a major issue is to design algorithms permitting to bootstrap new forms of behavior, this section only considers collections of already programmed skills.

The configurations **D** for 11 different types of robot behavior have been calculated. They consisted of five different walking behaviors (walk forward, walk right, ...) and six less oscillating behaviors (cheer, swing, ...). For each behavior, three samples of length 1000 were taken, which corresponds to about seven seconds each. The robot was placed back to a constant initial position for each experiment. No obstacles blocked the movements of the robot. The names and types of the 33 behaviors are listed in table 2.

Table 2: Names of the 33 examples of various types of behavior

Instance number	Behavior class	Behavior type
1-3	walking	forward walk
4-6	walking	backward walk
7-9	walking	walk to the right
10-12	walking	turn left
13-15	walking	turn right
16-18	other	swing
19-21	other	cheer happy 1
22-24	other	cheer happy 2
25-27	other	cheer happy 3
28-30	other	cheer sad 1
31-33	other	cheer sad 2

The relationship between these 33 configurations has been captured by computing the distance matrix Δ of this configuration of configurations. Corresponding maps have been obtained using a relaxation algorithm (see figure 7 b). In the distance matrix and maps of figure 7 configurations for walking behaviors and non-walking behaviors have been differentiated and appear as almost independent subgroups in the maps. At a finer level, configurations of same types of behavior are usually close but not always.

Another view of these data can be obtained by performing principal component analysis (PCA) in the configuration space and projecting the data onto its first few principal components. Figure 8 plots the points in the configuration space corresponding to the 33 recordings considered using the three first principal components. The two global clusters corresponding to walking and non-walking types of behavior can again be easily spotted.

In order to compare with the predefined hierarchical organization of behaviors used to generate the sequences, we have run two types of unsupervised clustering algorithms (Expectation-Maximization and K-means (see [WE00] for details) using configurations characterized only by the first four principal components in the configuration space. In a first series of trials, algorithms were set to produce two clusters. Each cluster was then assigned with a particular class (walking or non-walking) and, for each configuration belonging to the cluster, a comparison was made with the actual type of the behavior. For 91 % of the cases, instances of walking and non-walking types of behavior were classified in the right cluster. In a second series of trials, parameters were changed in order to produce 11 clusters and labels were assigned based on the behavior type. Best results obtained with the EM algorithm corresponded to 61 % correctly classified instances (see summary in Table 3 which compares the results of EM, K-means and random classification). The organization of configurations in the configuration space reflects the hierarchical structure of the behavior types, although it differs in its finer details.

Table 3: Correctly clustered instances in comparison with predefined behavior classes and types, using cluster algorithms with the four first principal components of the configuration space

Algorithm	Correctly categorized instances
EM (Expectation Maximization) (2 categories)	91 % (walking/other)
K-means (2 categories)	91 % (walking/other)
Random (2 categories)	50 % (walking/other)
EM (Expectation Maximization) (11 categories)	61 % (behavior type)
K-means (11 categories)	55 % (behavior type)
Random (11 categories)	9 % (behavior type)

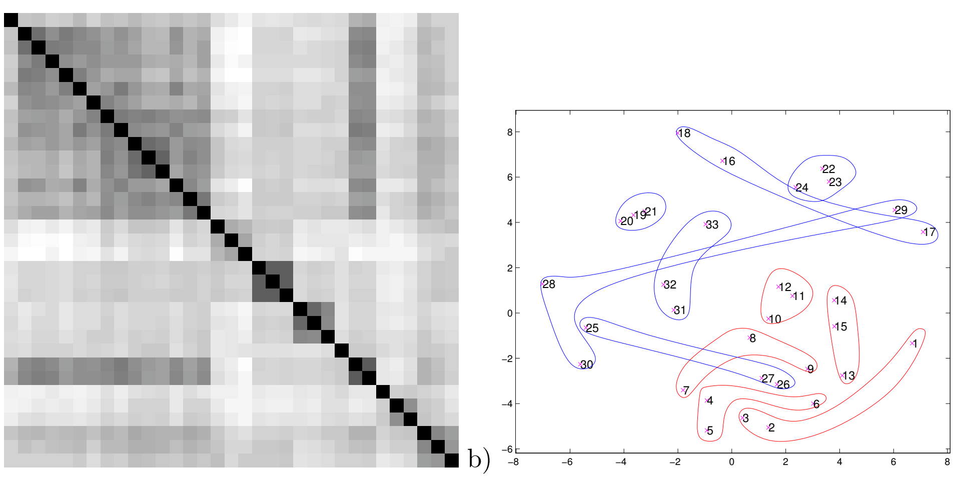
a)  b)

Figure 7: (a) Distance matrix corresponding to the configuration of the 33 configurations. Small distances are plotted in blue, big distances in red. (b) Map of configurations based on the relaxation algorithm. The triples represent same types of behavior. Walking behaviors are marked in red, all other behaviors are marked in blue.

Although PCA based methods are efficient when the set of instances is closed, they are not easy to

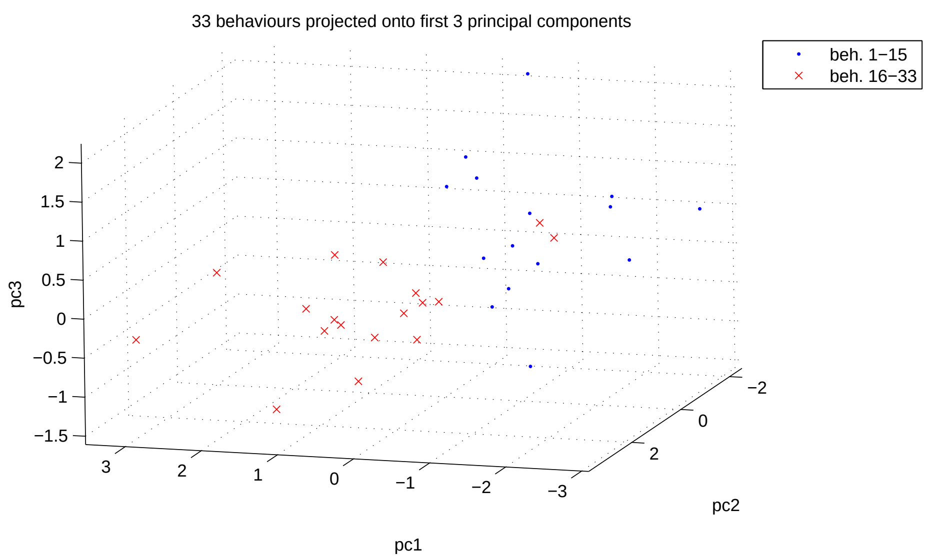


Figure 8: Representation of the points in the configuration space corresponding to the 33 recordings considered using the first three principal components.

extend to open systems. Other methods exist in such cases. For instance, self-organizing maps [Koh01] can be used for situations where the number of elements to project is open and potentially increasing over time (however, if the number of examples is not restricted, examples should be presented in a repetitive manner). As an illustration, figure 9 shows the results obtained with a 10×10 Kohonen self-organizing map after presentation of the configurations used in the previous experiments. Any novel run of the experiment should produce a different map where similar activities will be clustered close to one another.

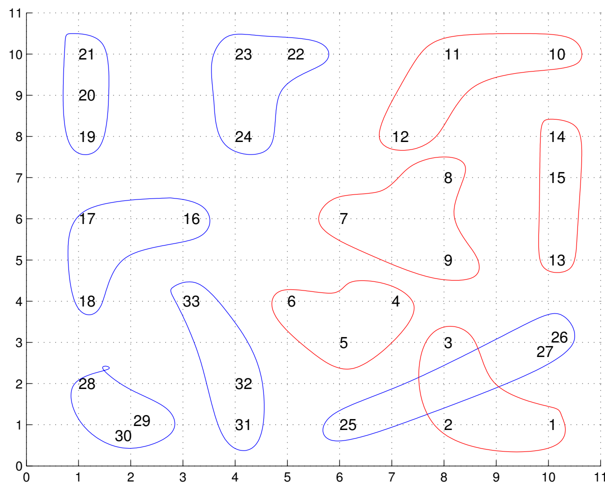


Figure 9: Map of configurations based on a 10×10 Kohonen self-organizing map (SOM). The triples represent same types of behavior. Walking behaviors are marked in red, all other behaviors are marked in blue.

5 Discussion

This article presents a mathematical framework for comparing multivariate sequences of data. It is based on two abstractions characterizing information dependencies in a set of information sources. Configurations reflect context dependent embodied sensorimotor organizations. They capture in a single format information about a physical body structure, particular coordinated actions performed and environmental context. At a second level, configurations of configurations give an integrated view of a collection of skills. This article provides simple illustrations of configurations for a collection of preprogrammed skills.

Clustering different kinds of behavior in a hierarchical manner plays an important role for many robotic applications (e.g. [KOKM02]). This issue is central in the challenging quest for autonomous development [WMP+01] as the possibility to find structural similarity between sensorimotor schemas is thought to be crucial for the emergence of higher-level forms of cognition. In particular they permit to consider possibilities of transfer for know-how developed in sensorimotor contexts to more abstract spaces [LJ98, Rob96]. Important literature exists on how to compare explicit symbolic structure (e.g. [GHK01]), but many authors have argued that generalization and transfer of skills could also be (maybe even more) efficient in the absence of symbolic representation [PJ96]. Given the variability of possible structures that can potentially underly the formation of sensorimotor schemas ([Arb03] (p.36–40), [Dre91, Min75, SA77, SPS99]), the approach taken in this article has been made as general as possible.

From a developmental perspective, the next step is to "close the loop" and show how the notion of emergent configurations can be used to structure behavior in return. The present article has explored some forms of bottom-up building of abstract knowledge, computation performed in the configuration space could now result in top-down influences. One possible way is to look at configurations that result from particular types of behavior in certain environments as an *emergent context*. As these contexts can be compared with one another, they can be used in a relevant manner during decision processes.

It should be noted that this approach can easily be extended to account for skills involving couplings between agents. Proprioceptive as well as heteroceptive information can be considered as information sources. In fact, for most organisms, no clear line can be drawn between both. Another set of experiments conducted in the same framework showed how in cases of strong couplings between agents, a "we-centric" space can emerge in which the agent's body structure can be directly mapped onto the structure of an observed body [HK05a].

Although these preliminary results are promising, more work needs to be done to specify the relevance of this framework for developmental robotics. Key issues that will be addressed in future work are first, to compare this approach with other ways of making analogies between sensorimotor trajectories and second, to investigate how emerging structures like configurations play a role in an overall developmental picture.

Acknowledgements

Research funded by Sony CSL Paris with additional support from the ECAGENTS project founded by the Future and Emerging Technologies programme (IST-FET) of the European Community under EU R&D contract IST-2003-1940. The authors would like to thank Martin Loetzsch for his precious help in collecting data for the robotic experiment as well as James Crutchfield, David Philipona, Pierre-Yves Oudeyer and Luc Steels for precious comments on the ideas discussed in this article.

REFERENCES

[AN00] S. Amari and H. Nagaoka. *Methods of information geometry*, volume 191. Oxford University Press, 2000. Translations of mathematical monographs.

[Arb03] M. Arbib. *The handbook of brain theory and neural networks.* MIT press, Cambridge, MA, 2003.

[CC94] T.F. Cox and M.A.A. Cox. *Multidimensional scaling.* Chapman and Hall, London, 1994.

[Cru90] J. P. Crutchfield. Information and its metric. In L. Lam and H. C. Morris, editors, *Nonlinear Structures in Physical Systems – Pattern Formation, Chaos, and Waves*, pages 119–130. Springer Verlag, 1990.

[CT91] T. Cover and J. Thomas. *Elements of information theory.* John Wiley and Sons, Inc., 1991.

[DMS02] T. Duckett, S. Marsland, and J. Shapiro. Fast, on-line learning of globally consistent maps. *Autonomous Robots*, 12:297–300, 2002.

[Dre91] G. L. Drescher. *Made-up minds.* The MIT Press, Cambridge, MA., 1991.

[FK98] M. Fujita and H. Kitano. Development of an autonomous quadruped robot for robot entertainment. *Autonomous Robots*, 5:7–20, 1998.

[FKD00] M. Fujita, H. Kitano, and T. Doi. Robot entertainment. In A. Druin and J. Hendler, editors, *Robots for kids : exploring new technologies for learning*, chapter 2, pages 37–70. Morgan Kaufmann, 2000.

[GHK01] D. Gentner, K. Holyoak, and N. Kokinov. *The analogical mind: perspectives from cognitive science.* MIT Press, 2001.

[Haf00] V. V. Hafner. Cognitive maps for navigation in open environments. In *Proceedings of the 6th International Conference on Intelligent Autonomous Systems (IAS-6)*, pages 801–808, Venice, Italy, 2000.

[HK05a] V. Hafner and F. Kaplan. Interpersonal maps and the body correspondence problem. In Y. Demiris, K. Dautenhahn, and C. Nehaniv, editors, *Proceedings of the Third International Symposium on Imitation in animals and artifacts*, pages 48–53, Hertfordshire, UK, 2005.

[HK05b] V.V. Hafner and F. Kaplan. Learning to interpret pointing gestures: experiments with four-legged autonomous robots. In S. Wermter, G. Palm, and M. Elshaw, editors, *Biomimetic Neural Learning for Intelligent Robots. Intelligent Systems, Cognitive Robotics, and Neuroscience*, Series: Lecture Notes in Computer Science. Subseries: Lecture Notes in Artificial Intelligence, Vol. 3575. Springer Verlag, 2005.

[Koh01] T. Kohonen. *Self-Organizing Maps. Extended edition.* Springer, Berlin, 2001.

[KOKM02] F. Kaplan, P-Y. Oudeyer, E. Kubinyi, and A. Miklosi. Robotic clicker training. *Robotics and Autonomous Systems*, 38(3-4):197–206, 2002.

[Kul68] S. Kullback. *Information theory and statistics.* Dover, 1968.

[KYO+04] Y. Kuniyoshi, Y. Yorozu, Y. Ohmura, K. Terada, T. Otani, A. Nagakubo, and T. Yamamoto. From humanoid embodiment to theory of mind. In *Embodied Artificial Intelligence*, pages 202–218. Springer Verlag, 2004.

[LJ98] George Lakoff and Mark Johnson. *Philosophy in the flesh: the embodied mind and its challenge to Western thought*. Basic Books, 1998.

[LL05] J. Lafferty and G. Lebanon. Diffusion kernels on statistical manifolds. *Journal of Machine Learning Research*, 6:129–163, 2005.

[Min75] M. Minsky. A framework for representing knowledge. In P. Wiston, editor, *The psychology of computer vision*, pages 211–277. Mc Graw Hill, New York, 1975.

[ONP04a] L. Olsson, C. Nehaniv, and D. Polani. Sensory channel grouping and structure from uninterpreted sensor data. In *NASA/DoD Conference on Evolvable Hardware*, pages 153–160, Seattle, Washington, USA, 2004.

[ONP04b] L. Olsson, C.L. Nehaniv, and D. Polani. The effects on visual information in a robot in environments with oriented contours. In *Proceedings of the Fourth International Workshop on Epigenetic Robotics*, pages 83–88, Boston, USA, 2004.

[ONP05] L. Olsson, C. Nehaniv, and D. Polani. Sensor adaptation and development in robots by entropy maximization of sensory data. In *Proceedings of the 6th IEEE International Symposium on Computational Intelligence in Robotics and Automation (CIRA 2005)*, pages 587–592, Espoo, Finland, 2005. IEEE Computer Society Press.

[PJ96] L. Pratt and B. Jennings. A survey of connectionist network reuse through transfer. *Connection Science*, 8(2):163–184, 1996.

[PK97] D. Pierce and B. Kuipers. Map learning with uninterpreted sensors and effectors. *Artificial Intelligence*, 92:169–229, 1997.

[Rob96] A. Robins. Transfer in cognition. *connection science*, 8(2):185–204, 1996.

[SA77] R. Schank and R. Abelson. *Scripts, plans, goals and understanding: An inquiry into human knowledge structures*. Lawrence Erlbaum Associates, Hillsdale, NJ., 1977.

[Sch00] T. Schreiber. Measuring information transfer. *Physical Review Letters*, 85(2):461–464, 2000.

[SK00] L. Steels and F. Kaplan. Aibo's first words: The social learning of language and meaning. *Evolution of Communication*, 4(1):3–32, 2000.

[SP03] O. Sporns and T. Pegors. Information-theoretical aspects of embodied artificial intelligence. In F. Iida, R. Pfeifer, L. Steels, and Y. Kuniyoshi, editors, *Embodied artificial intelligence*, LNAI 3139, pages 74–85. Springer, 2003.

[SPS99] R.S. Sutton, D. Precup, and S. Singh. Between mdpss and semi-mdps: A framework for temporal abstraction in reinforcement learning. *Artificial Intelligence*, 112:181–211, 1999.

[SW62] C. Shannon and W. Weaver. *The mathematical theory of communication*. University of Illinois Press, 1962.

[TES98] G. Tononi, G. Edelman, and O. Sporns. Complexity and coherency: integrating information in the brain. *Trends in cognitive sciences*, 2(12):474–484, 1998.

[WE00] Ian.H. Witten and Frank Eibe. *Data mining*. Morgan Kaufmann Publishers, 2000.

[WMP+01] J. Weng, J. McClelland, A. Pentland, O. Sporns, I. Stockman, M. Sur, and E. Thelen. Autonomous mental development by robots and animals. *Science*, 291:599–600, 2001.