Feature, Configuration, History
A Bio-inspired Framework for Information Representation in Neural Networks

Frédéric Alexandre, Maxime Carrere and Randa Kassab
Inria Bordeaux Sud-Ouest, 200 Avenue de la Vieille Tour, 33405 Talence, France
LaBRI, Université de Bordeaux, Institut Polytechnique de Bordeaux, CNRS, UMR 5800, Talence, France
Institut des Maladies Neurodégénératives, Université de Bordeaux, CNRS, UMR 5293, Bordeaux, France

Keywords: Information Representation, Computational Neuroscience, Pavlovian Conditioning.

Abstract: Artificial Neural Networks are very efficient adaptive models but one of their recognized weaknesses is about information representation, often carried out in an input vector without a structure. Beyond the classical elaboration of a hierarchical representation in a series of layers, we report here inspiration from neuroscience and argue for the design of heterogenous neural networks, processing information at feature, configuration and history levels of granularity, and interacting very efficiently for high-level and complex decision making. This framework is built from known characteristics of the sensory cortex, the hippocampus and the prefrontal cortex and is exemplified here in the case of pavlovian conditioning, but we propose that it can be advantageously applied in a wider extent, to design flexible and versatile information processing with neuronal computation.

1 INTRODUCTION

Artificial Neural Networks (ANNs) have proven to be an excellent formalism for adaptive information processing and have been successfully applied to an impressive range of application domains for such tasks as classification, control and prediction. One of the reasons for the undebatable success of ANNs is certainly its unique position between the fertile fields of Machine Learning and Computational Neuroscience. On the one hand, Computational Neuroscience irrigates ANNs with architectural and computational principles inspired from observation of the living brain. On the other hand, Machine Learning anchors in solid mathematical grounds the most prominent features of ANNs including learning algorithms.

Cross-fertilization from both domains explains most of the advances of ANNs in the past decades, even if some weaknesses remain. Among them, one of the most critical flaws of ANNs is certainly related to information representation. Basically, ANNs are mainly dedicated to unstructured information processing (their input is a 'flat' vector of data without structure), whereas most real-world applications require the elaboration and exploitation of complex structures for an adequate knowledge representation.

Several measures have already been proposed to overcome this problem. Firstly, input vectors can be built according to a structured representation of information proposed *a priori*, but this cannot be adapted by learning and cannot be applied to the inner representation of ANNs. Secondly, in opposition to the connectionist approach, the symbolic approach is generally more efficient at defining and processing elaborated knowledge representations and it has been proposed accordingly to associate both paradigms in the so-called Connectionist-Symbolic Integration approach (Sun and Alexandre, 1997) but here also, the two kinds of modules are generally too independent. Thirdly, some interesting attempts have been made to propose fully connectionist solutions to this problem.

Connectionist solutions to structured information processing are mainly centred on innovative architectures and propose to adapt existing learning algorithms to exploit them. One of the early models in this direction, the LRAAM network (Sperduti et al., 1997), explains how a recurrent elaboration of a network allows to encode graphs. More recently, deep neural networks (Bengio et al., 2013) renew studies in which complex representations can be extracted from many-layered neural networks, as suggested by the architecture of the visual system in the brain (Rousselet et al., 2004).

These solutions to build and manipulate complex representations in ANNs are diverse (recursive or iterative), but both are hierarchical : Complex represen-

Alexandre F., Carrere M. and Kassab R..
Feature, Configuration, History - A Bio-inspired Framework for Information Representation in Neural Networks.
DOI: 10.5220/0005156003160321
In *Proceedings of the International Conference on Neural Computation Theory and Applications* (NCTA-2014), pages 316-321
ISBN: 978-989-758-054-3

tations are seen as more and more abstract representations. In this position paper, we argue that these architectural solutions are not fully satisfactory because they miss an important dimension, the granularity of information representation. Taking inspiration from the brain, it can be remarked that hierarchical integration is not the only way to elaborate structured information representation : In the brain, specialized modules extract information in the data flow at various levels of granularity and mix them in heterogenous (and not only hierachical) systems for elaborated decision-making processes. As far as sensory information processing is concerned, we propose that these levels of granularity are of three kinds:

1. **Feature.** At a conceptual level, features can be extracted in the information flow and arranged in a hierarchical way;
2. **Configuration.** At the individual level, the agglomeration in a 'flat' vector of all the perceived sensory information allows for the representation of specific cases;
3. **History.** At the statistical level, overall trends can be extracted from the recent history of perception.

In the remaining of the paper, we first give some elements from neuroscience for the existence of such representations in the brain of mammals, then we propose to illustrate them and put them in situation in the case of pavlovian conditioning, before discussing the usefulness of such representations and of their association in heterogenous systems.

2 INFORMATION REPRESENTATION: A VIEW FROM NEUROSCIENCE

Information representation in a hierarchical way leading to abstraction of concepts has been extensively reported in the brain, for example in the visual cortex. This hierarchical view of knowledge representation has already inspired ANNs like deep networks, as evoked in section 1. From Hubel and Wiesel initial idea that neurons selective to orientation in the primary visual cortex are built from the activity of contrast-sensitive cells in a receptive field in the thalamus (Hubel and Wiesel, 1962), further experiments along the ventral visual pathway have identified neuronal areas responding to a hierarchy of concepts (Rousselet et al., 2004) captured in increasingly large receptive fields, from simple shapes to complex patterns like faces. In that respect, the ventral pathway of the cortex is proposed as the locus of semantic memory, with a cascade of neurons extracting a hierarchy of concepts. Here we will speak about **feature representation**.

Nevertheless, hierarchy is probably not the main characteristic to be evoked when characterizing information processing in the brain, but rather modularity and heterogeneity. Indeed, the brain is organized along complementary learning systems, as it has already been captured in early modeling approaches (Doya, 1999; Alexandre, 2000).

Particularly, learning mode and information representation in the hippocampus have been contrasted with that of the cortex described above (McClelland et al., 1995). Whereas the cortex learns slowly to elaborate a semantic memory as a hierarchy of concepts that will be exploited in generalization, the hippocampus receives as sensory input a summary sketch of current cortical activity and learns in one shot an episodic memory of a specific event in its spatial and temporal context. Contrarily to the cortical case, this memory process is not prone to generalization; instead, when faced to a similar episode, the recall process will rebuild the original episode: the representation in the hippocampus is consequently unitary, storing configural representations of specific events (O'Reilly and Rudy, 2001). Here we will speak about **representation of configuration**.

Though complementary, these memories are also in cooperation : The hippocampus is fed with sensory inputs encoded with cortical features and reciprocally participates to the slow cortical learning through the phenomenon of consolidation : Specific cases stored in the hippocampus during some task are sent back to the cortex for the extraction of new features, pertinent to the task (O'Reilly and Rudy, 2001). Whereas the cortex can work in generalization, the task is more difficult for the hippocampus : Faced to an episode similar to a stored one, the hippocampus can only choose between pattern completion (supposing that the episodes are equivalent, the stored one will be recalled) and pattern separation (supposing that the episodes must be distinguished one from the other, the new one will be stored as a different pattern) (O'Reilly and Rudy, 2001).

As far as task execution is concerned, the prefrontal cortex is certainly one of the most involved cerebral structures (Fuster, 2001), since it supports the temporal organization of behavior, i.e. selecting at each moment, the best action (or decision) to trigger, depending on the external (perceptive) and internal (emotional and motivational) sensory information. Most of the time this decision is not purely reactive (not depending directly on current stimuli) but deliberative (depending on a more complex evaluation, also based on past experience). To that end, the

prefrontal cortex builds task sets and maintains them active in a working memory mechanism, for comparison and selection (by a loop with subcortical structures, the basal ganglia) of the most adapted solution. In short, task sets are collections of recent cases of associations between actions (or decision) that have been triggered in certain sensory and temporal contexts, together with their performance in reaching the goal that was expected.

The prefrontal cortex is organized, from its premotor to its anterior part, in a hierarchy of levels of control (Koechlin et al., 2003) depending on the nature of the sensory and temporal context (and on the structure that sends this information). This cascade of control is made according to the perceptual context, the episodic context and the motivational and emotional context (Kouneiher et al., 2009). At each level of description, the decision to engage an action is taken, based on its recent history of success in this context. This statistical analysis makes us describe the prefrontal cortex as a locus for the **representation of history**.

To better understand the way these representations are elaborated in a concrete case, their association and their impact on cerebral processes, we propose an illustration in the case of pavlovian conditioning.

3 CASE STUDY: PAVLOVIAN CONDITIONING

Pavlovian or respondent conditioning is the learning process by which an animal is able to associate an Unconditional Stimulus (US = biologically significant stimulus announcing pain or pleasure, e.g. an electric shock) to a conditional stimulus (CS, a neutral event like a tone, that predicts the US). This learning has been extensively studied in so-called fear conditioning experiments and is reported to be acquired quickly if the electric shock is given subsequently to the tone (Herry et al., 2008). Later on, the animal exhibits fear behavior (ex: freezing) when solely exposed to the auditory CS. This can be explained as an anticipation or a preparation to the forthcoming pain. This response will disappear if the electric shock is no longer given (extinction process) and get re-activated by a new association (renewal process). As such, this simple description could assimilate pavlovian conditioning to an elementary associative learning where, for example, an hebbian learning rule could increase and decrease a weight between two neurons standing for the CS and the US. But reality is much more complex, as shown by many behavioral studies that required to refine our understanding on pavlovian conditioning.

One of the most famous paradigms is the blocking paradigm : In an early step, an association CS1-US is acquired. Then CS1 is paired with CS2 to announce the US. Here it could be said that CS2 becomes also a faithful predictor of US and, from a purely associative learning point of view, it should acquire also a predictive power. This is not confirmed by experiments: CS2 alone triggers no conditioned response. This paradigm has be interpreted as a parsimonious learning (CS1 is sufficient to predict US; no need to perform a new learning about CS2) and led to propose a non purely associative learning rule in the early 70's, the competitive Rescorla-Wagner learning rule (Rescorla and Wagner, 1972). This rule states that the modification of the strength of association (the weight, considering a neuronal implementation with an input layer of CS connected to an output layer of US) between CSi and US is proportional to the error of prediction, i.e. the difference between what actually happens (the US) and what was expected (the sum of the predictive values of all present CSi). Two other multiplicative terms in the rule are defined as the associability (saliency) of the CS and the effectiveness (behavioral importance) of the US and are often defined as constant (but cf. below).

Now comes the question of the nature of the CS and here also behavioral experiments are confusing: CS often corresponds to a salient stimulus (a tone, a flashing light), but sometimes as in the case of extinction and renewal, the pertinent element is the context in which conditioning occurs. This makes often modelers add a specific neuron in the CS layer standing for 'the context', even if this is not fully satisfactory, concerning representation of information. Another problem is about configural learning : In a prototypical case, CS1 or CS2 alone predicts US, whereas the occurrence of both predicts no US. This is difficult to explain with representation of single CSs because in this case CS1 and CS2 will acquire a strong predictive strength and the conjunction of both should add these strengths and predict the US. A practical solution to this classical problem is to create a new representation for CS1+CS2, seen as a new configural stimulus (Schmajuk and DiCarlo, 1992) but this raises again the question of building by hand the representation of sensory information during pavlovian conditioning.

All the models of pavlovian learning evoked so far are said US-processing models because their main goal is to predict the US from current CSs and consequently the same rule of predictive strength modification is applied to each CS, whatever its history, which is not realistic, as other behavioral experiments indicate (Le Pelley, 2004), hence the need for CS-processing models.

On the one hand, the Pearce-Hall model (Pearce and Hall, 1980) tries to capture the fact that a CS often associated with US learns less quickly than a not well known CS and proposes that the term of associabilility of the CS could change in time and depend on the level of surprise (absolute value of error of prediction) and also on its own current value.

On the other hand, the Mackintosh model (Mackintosh, 1975) maintains a high associability for reliable CS, to explain the fact that assigning a CS to a new US is quicker after an overtaining of the CS with the previous US. Both cases underline the need to take into account the recent history of conditioning. They also raise new problems since it is obvious that the Pearce-Hall and Mackintosh models are somewhat contradictory, respectively proposing to decrease and increase the associability of a frequent CS.

Confronting abstract models to behavioral observations is also an incitation to integrate more realistic constraints in the models. This is for example the case with uncertainty : Our word is stochastic (real CSs do not always reliably predict US) and non stationary (a CS-US association rule valid at one time can be obsolete the time after). Classical models of pavlovian conditioning generally address poorly this kind of constraints, which is a strong incentive to develop more realistic models, for example including bayesian approaches (Yu and Dayan, 2005; Deco et al., 2008).

In summary, this brief overview of pavlovian conditioning shows that, more than one hundred years after its identification, pavlovian conditioning is still not fully understood. We have mentionned several observations non consistent with the current understanding of this learning mechanism and, for each of them, a model yielding the corresponding behavior, but at the moment, there is no complete model integrating all these mechanisms. We think that going deeper in the understanding of the underlying cerebral substrate is an excellent way to design such a complete model.

4 CEREBRAL SUBSTRATE TO PAVLOVIAN CONDITIONING

The amygdala is the core cerebral structure for pavlovian conditioning (Holland and Gallagher, 1999) but it cannot be considered in isolation for two reasons. On the one hand, the amygdala receives from other cerebral structures, sensory information about the CS and the US, for their association, but also other kinds of information needed at various steps of the conditioning. On the other hand, pavlovian conditioning results in a series of effects, that are transmitted to other regions in the brain. These effects are of course the pavlovian response itself, but also other attentional and representational effects.

Amygdala is in fact a complex and heterogenous structure (LeDoux, 2007), composed of several subdivisions and it can be convenient to distinguish three regions with specific functions and connectivities :

1. The lateral amygdala (LA) receives sensory information about the US and the CS from the thalamus and the cortex. This region is responsible for learning the CS-US association.
2. The central nucleus of the Amygdala (CeA) is in charge of the three aspects of the pavlovian response: the motor aspect (e.g. in the case of fear, freezing), the autonomic aspect (e.g. physiological changes like heart beat or temperature increase) and the hormonal aspect (release of stress hormones and neuromodulators like acetylcholine, norepinephrine, dopamine, etc.).
3. The basal amygdala (BA) has a major representational role. It receives from LA information about CS-US association, modulates this information with contextual inputs coming from the hippocampus and prefrontal, to elaborate a representation about the sensory nature of the US (Cardinal et al., 2002). This information is sent to the CeA to produce the pavlovian response but also to the sensory cortex (attentional affect) and to the prefrontal cortex-basal ganglia system (emotional evaluation of stimuli has also strong effect in decision making and operant conditioning).

This very rough description of the amygdalar circuitry allows to put into context all the mechanisms and observations about pavlovian conditioning evoked in section 3. We posit a general computational principle with this circuitry : Sensory thalamic and cortical inputs to LA (generic features) should be sufficient to reliably predict US but other sensory information coming from the hippocampus to BA will possibly propose CS-US association rules corresponding to more specific cases. Later on, these specific cases could be 'compiled' from the hippocampus, in the creation of new features in the cortex.

The LA and BA regions project in CeA to trigger the good pavlovian response. The choice between the corresponding CS-US rules will be learned from a monitoring of their performance in predicting US, evaluated in the orbitofrontal cortex (OFC). This anterior region of the prefrontal cortex has very dense relations with all regions of the amygdala and builds an history of errors of US prediction, depending on the sensory context (in the cortex and in the hippocampus) in which they occurred (Pauli et al., 2011). These errors of prediction are of two kinds.

A positive error of prediction corresponds to the case where an US is received whereas it was not predicted by LA neither by BA : in this case BA will send a signal to the hippocampus to make this structure learn by heart the current episode (Paz et al., 2009). On the next occurrence of this episode the hippocampus will be able to recall the corresponding US and send a configural representation to BA. If this is often repeated, this will result in consolidation in the cortex and in the emergence of a feature representation in the cortex, leading to a reliable rule in LA.

When a negative error of prediction occurs (from the current sensory context, an US has been predicted and doesn't occur), several causes can be evoked :

1. The rule is valid but the specific context corresponds to an exception which must be stored in the hippocampus as associated to no US (case of an extinction). If this happens again in this context, the rule in BA based on this hippocampal input will be favored in this specific case with regard to the general rule coming from LA.
2. The rule is stochastic and not fully reliable. In this case the level of stochasticity of the rule has to be updated in OFC. This case and the previous case also result in an increase of the level of acetylcholine (Yu and Dayan, 2005), which favors learning in BA, in the hippocampus and in the cortex, with the aim of making more precise the rules currently under consideration.
3. The rule is no longer valid because the world is non stationary but it could be valid again in the future. The rule is consequently conserved without modification but inhibited by OFC that triggers a release of norepinephrine (Yu and Dayan, 2005). This neuromodulator acts on the thalamocortical inputs of LA, to look for a new rule (Johnson et al., 2011).

These causes have to be distinguished from the history of performance in US prediction stored in OFC. A corresponding algorithm has been proposed by (Yu and Dayan, 2005), to evaluate the decision threshold between stochasticity and non stationarity : When the rule is highly stochastic, many examples are needed before deciding for a new rule; conversely when the rule is very reliable, this change will be made more easily. This alternative is also useful to propose a way to release the seamingly contradictory effects between the Pearce-Hall and Mackintosh models : The Mackintosh model could apply only when the rule is judged valid and stable (no error of prediction), to evaluate its level of reliability, whereas the Pearce-Hall model could apply only in the early stages of selection of the rule, to explore and detect candidate CS.

5 DISCUSSION

In this paper, we have laid emphasis on the fact that different levels of granularity of information are represented and processed in different brain regions and can be manipulated in ANNs as well : Multi-layered ANNs can extract a set of **features**, as a hierarchy of concepts, with some analogy to the sensory cortex. Associative memories like the Hopfield model have been shown to learn by heart **configuration** of features representing specific cases in a recurrent architecture, as it is the case in the hippocampus. In addition to these cases, **history** of the statistical tendencies of certain information flows is believed to be captured in the prefrontal cortex. Though not yet mature in ANNs, early models implementing this latter kind of processing are proposed (Kouneiher et al., 2009).

In this paper, we have also shown, through the example of pavlovian conditioning, why mixing these levels of granularity is so powerful. On the one hand, the combination of associative rules involving features and configurations is of central importance in this learning and is more generally reminiscent of the principle of mixing generic rules and specific cases. On the other hand, it seems possible to reconcile different cases where pavlovian conditioning works differently by performing a statistical analysis that allows to distinguish different functioning modes. Selecting strategies and switching between procedures seem to be a major role of the prefrontal cortex (Fuster, 2001; Koechlin et al., 2003), fed by contextual and episodic information from the sensory cortex and the hippocampus respectively, and operated through its control on many cerebral structures, sometimes by the release of neuromodulators.

Pavlovian conditioning was chosen here to exemplify these principles but many characteristics of the cerebral system support the fact that computing with these different levels of representation is more general than this only case. In addition, we are deeply convinced that computing with these levels of representation might be advantageously exported to the domain of ANNs. It could provide this domain with an efficient way to combine generic rules and specific cases and to make decisions based on the evaluation of performances of such heterogenous modules, which seems to underlie an important part of our cerebral system and its emergent cognitive capabilities.

ACKNOWLEDGEMENTS

This work was partly supported by the Keops ANR project.

REFERENCES

Alexandre, F. (2000). Biological Inspiration for Multiple Memories Implementation and Cooperation. In *In V. Kvasnicka P. Sincak, J. Vascak and R. Mesiar, editors, International Conference on Computational Intelligence*.

Bengio, Y., Courville, A. C., and Vincent, P. (2013). Representation learning: A review and new perspectives. *IEEE Trans. Pattern Anal. Mach. Intell.*, 35(8):1798–1828.

Cardinal, R. N., Parkinson, J. A., Hall, J., and Everitt, B. J. (2002). Emotion and motivation: the role of the amygdala, ventral striatum, and prefrontal cortex. *Neuroscience & Biobehavioral Reviews*, 26(3):321–352.

Deco, G., Jirsa, V. K., Robinson, P. A., Breakspear, M., and Friston, K. (2008). The dynamic brain: From spiking neurons to neural masses and cortical fields. *PLoS Comput. Biol.*, 4(8).

Doya, K. (1999). What are the computations of the cerebellum, the basal ganglia and the cerebral cortex? *Neural Networks*, 12:961–974.

Fuster, J. M. (2001). The Prefrontal CortexAn Update : Time Is of the Essence. *Neuron*, 30(2):319–333.

Herry, C., Ciocchi, S., Senn, V., Demmou, L., Muller, C., and Luthi, A. (2008). Switching on and off fear by distinct neuronal circuits. *Nature*, 454(7204):600–606.

Holland, P. C. and Gallagher, M. (1999). Amygdala circuitry in attentional and representational processes. *Trends in cognitive sciences*, 3(2):65–73.

Hubel, D. and Wiesel, T. (1962). Receptive fields, binocular interaction and functional architecture in the cat visual cortex. *J Physiol*, 160:106–154.

Johnson, L. R., Hou, M., Prager, E. M., and LeDoux, J. E. (2011). Regulation of the Fear Network by Mediators of Stress: Norepinephrine Alters the Balance between Cortical and Subcortical Afferent Excitation of the Lateral Amygdala. *Frontiers in Behavioral Neuroscience*, 5.

Koechlin, E., Ody, C., and Kouneiher, F. (2003). The Architecture of Cognitive Control in the Human Prefrontal Cortex. *Science*, 302(5648):1181–1185.

Kouneiher, F., Charron, S., and Koechlin, E. (2009). Motivation and cognitive control in the human prefrontal cortex. *Nature Neuroscience*, 12(7):939–945.

Le Pelley, M. E. (2004). The role of associative history in models of associative learning: a selective review and a hybrid model. *The Quarterly Journal of Experimental Psychology*, 57(3):193–243.

LeDoux, J. (2007). The amygdala. *Current Biology*, 17(20):R868–R874.

Mackintosh, N. J. (1975). A theory of attention: Variations in the associability of stimuli with reinforcement. *Psychological Review*, 82(4):276–298.

McClelland, J. L., McNaughton, B. L., and O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. *Psychological review*, 102(3):419–457.

O'Reilly, R. C. and Rudy, J. W. (2001). Conjunctive Representations in Learning and Memory: Principles of Cortical and Hippocampal Function. *Psychological Review*, 108(2):311–345.

Pauli, W. M., Hazy, T. E., and O'Reilly, R. C. (2011). Expectancy, Ambiguity, and Behavioral Flexibility: Separable and Complementary Roles of the Orbital Frontal Cortex and Amygdala in Processing Reward Expectancies. *Journal of Cognitive Neuroscience*, 24(2):351–366.

Paz, R., Bauer, E. P., and Paré, D. (2009). Measuring correlations and interactions among four simultaneously recorded brain regions during learning. *Journal of neurophysiology*, 101(5):2507–2515.

Pearce, J. M. and Hall, G. (1980). A model for Pavlovian learning: variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychological review*, 87(6).

Rescorla, R. and Wagner, A. (1972). A theory of pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In *Classical Conditioning II: Current Research and Theory*, pages 64–99. Appleton Century Crofts.

Rousselet, G., Thorpe, S., and Fabre-Thorpe, M. (2004). How parallel is visual processing in the ventral path? *TRENDS in Cognitive Sciences*, 8(8):363–370.

Schmajuk, N. and DiCarlo, J. (1992). Stimulus configuration, classical conditioning and the hippocampus. *Psychological Review*, 99:268–305.

Sperduti, A., Sperduti, R., and Starita, A. (1997). Supervised neural networks for the classification of structures. *IEEE Transactions on Neural Networks*, 8:714–735.

Sun, R. and Alexandre, F., editors (1997). *Connectionist - Symbolic Integration: from Unified to Hybrid Approaches*. Lawrence Erlbaum Associates.

Yu, A. J. and Dayan, P. (2005). Uncertainty, neuromodulation, and attention. *Neuron*, 46(4):681–692.