

NRC Publications Archive Archives des publications du CNRC

Projection Learning vs. Correlation Learning: From Pavlov Dogs to Face Recognition Gorodnichy, Dimitry

This publication could be one of several versions: author's original, accepted manuscript or the publisher's version.
/ La version de cette publication peut être l'une des suivantes : la version prépublication de l'auteur, la version acceptée du manuscrit ou la version de l'éditeur.

Publisher's version / Version de l'éditeur:

The Correlation Learning Workshop [Proceedings], 2005

NRC Publications Archive Record / Notice des Archives des publications du CNRC :
<https://nrc-publications.canada.ca/eng/view/object/?id=4c4b24c4-66eb-41f2-b153-621661b69e3>
<https://publications-cnrc.canada.ca/fra/voir/objet/?id=4c4b24c4-66eb-41f2-b153-621661b69e3e>

Access and use of this website and the material on it are subject to the Terms and Conditions set forth at
<https://nrc-publications.canada.ca/eng/copyright>

READ THESE TERMS AND CONDITIONS CAREFULLY BEFORE USING THIS WEBSITE.

L'accès à ce site Web et l'utilisation de son contenu sont assujettis aux conditions présentées dans le site
<https://publications-cnrc.canada.ca/fra/droits>

LISEZ CES CONDITIONS ATTENTIVEMENT AVANT D'UTILISER CE SITE WEB.

Questions? Contact the NRC Publications Archive team at
PublicationsArchive-ArchivesPublications@nrc-cnrc.gc.ca. If you wish to email the authors directly, please see the first page of the publication for their contact information.

Vous avez des questions? Nous pouvons vous aider. Pour communiquer directement avec un auteur, consultez la première page de la revue dans laquelle son article a été publié afin de trouver ses coordonnées. Si vous n'arrivez pas à les repérer, communiquez avec nous à PublicationsArchive-ArchivesPublications@nrc-cnrc.gc.ca.



National Research
Council Canada

Conseil national
de recherches Canada

Institute for
Information Technology

Institut de technologie
de l'information

NRC - CNRC

Projection Learning vs. Correlation Learning: From Pavlov Dogs to Face Recognition *

Gorodnichy, D.
May 2005

* published at The Correlation Learning Workshop. May 8, 2005. Victoria, British Columbia, Canada. NRC 48209.

Copyright 2005 by
National Research Council of Canada

Permission is granted to quote short excerpts and to reproduce figures and tables from this report, provided that the source of such material is fully acknowledged.

Projection learning vs. correlation learning: from Pavlov dogs to face recognition

Dmitry O. Gorodnichy

Institute for Information Technology (IIT-ITI)
National Research Council of Canada (NRC-CNRC)
Montreal Rd, M-50, Ottawa, Canada K1A 0R6
<http://synapse.vit.iit.nrc.ca/memory>

Abstract—Face recognition in video is an example of the problem which is outstandingly well performed by humans, compared to the performance of machine-built recognition systems. This phenomenon is generally attributed to the following three main factors pertaining to the way human brain processes and memorizes information, which can be succinctly labeled as 1) non-linear processing, 2) massively distributed collective decision making, and 3) synaptic plasticity.

Over the last half a century, many mathematical models have been developed to simulate these factors in computer systems. This presentation formalizes the recognition process, as it is performed in brain, using one of such mathematical models, within which the projection learning appears to be a natural improvement to the correlation learning. We show that, just as the correlation learning, the projection learning can also be written in incremental form. By taking in account the past data and being non-local, this rule however provides a way to automatically emphasize more important attributes and training data over the less important ones.

The presented model, while providing a simple way to incorporate the main three factors of biological memorization listed above, is very powerful. This is demonstrated by incorporating it into a face recognition system which is shown to be capable of recognizing faces in video under conditions known to be very difficult for traditional Von-Neumann-type recognition systems.

I. MODELING ASSOCIATIVE PROCESS

Let us consider the task of associating one stimulus (R – receptor stimulus) to another (E – effector stimulus). One famous example of this task is the Pavlov dogs experiment, within which the dogs trained on a buzz (R) - food (E) stimuli pairs were shown to salivate at the sound of the buzz even when no food was presented.

Another example of this task, from our everyday life and which remains to be one of the most difficult tasks for the computer to do, is associating a person's face (R) to a certain feeling or knowledge (E): disgust, familiarity, race, level of IQ, name, etc.

To model this association process, let's consider synapses C_{ij} which, for simplicity and because we do not know exactly what is connected in the brain to what, are assumed to connect all attributes of stimuli pair R and E among each other. For a general model, several layers between R and E , each connecting to one another, may be considered, which can be simulated by adding extra attributes into the aggregated stimulus vector.

These synapses have to be adjusted in the training stage so that in the recognition stage, when sensing R , which is close to what the system has sensed before, based on the trained synaptic values a sense of the missing corresponding stimulus E is produced. Mathematically this can be written as follows. Let $\vec{V} = (R_i, E_i)$ be an aggregated N -dimensional vector made of all binary decoded attributes ($R_i, E_i \in \{-1; +1\}$) of the stimuli pair. The $N \times N$ synaptic matrix $\mathbf{C} = \{C_{ij}\}$ has to be computed so that when having an incomplete version of a training stimulus, the collective decision making produces the effector attributes most similar to those used in training, where the decision making process is based on summation of all input attributes weighted by the synaptic values, performed several times until the consensus is reached:

$$V_i(t+1) = \text{sign}(S_j(t)) \quad (1)$$

$$S_j(t) = \sum_{i=1}^N C_{ij} V_j(t), \quad \text{until} \quad (2)$$

$$V_i(t+1) = V_i(t) = V_i(t^*) \quad (3)$$

The last equation expresses the *stability condition* of the system. When it holds for all neurons, it describes the situation of the reached consensus. The obtained stimulus $\vec{V}(t^*)$, called the *stable state* or *attractor* of the network, is then decoded into receptor and effector components: $R_i(t^*)$ and $E_i(t^*)$ for further analysis of the result of the performed association.

The main question arises: How to compute synaptic values C_{ij} so that the best associative recall is achieved?

Ideally this should be done so that the computation of the synaptic values defined by a learning rule

- i) does not require the system to go through the already presented stimuli, i.e. there are no iterations involved, and
- ii) would update the synapses based on the currently presented stimuli pair only, without knowing which stimuli will follow, i.e. no batch mode is involved.

These two conditions represent the idea of *incremental learning*:

$$C_{ij}^m = C_{ij}^{m-1} + dC_{ij}^m. \quad (4)$$

Starting from zero ($C_{ij}^0 = 0$), indicating that nothing is learnt, each synaptic weight C_{ij} undertakes a small increment dC_{ij} , the value of which, either positive or negative, is determined by the training stimuli pair.

It is understood that for optimal memorization, the increments dC_{ij}^m should be functions of the current stimulus pair attributes (i.e. \vec{V}^m) and what has been previously memorized (i.e. \mathbf{C}):

$$dC_{ij}^m = f(\vec{V}^m, \mathbf{C}). \quad (5)$$

II. CORRELATION LEARNING RULES

Within the formalization of the problem described above, correlation learning, which updates C_{ij} based on the correlation of the corresponding attributes i and j of the training stimulus:

$$dC_{ij}^m = \alpha V_i^m V_j^m, \quad 0 < \alpha < 1 \quad (6)$$

is one of the most frequently used. It makes the stimuli used in training ($\vec{V}^m, m = 1 \dots M$) the minima locations of the Hamiltonian

$$E(\vec{Y}(t)) \doteq -\frac{1}{2} \vec{Y}^T(t) \vec{S}(t) = -\frac{1}{2} \vec{Y}^T(t) \mathbf{C} \vec{Y}(t), \quad (7)$$

which is known to govern the evolution of a dynamic system defined by the update rules of Eqs. 2-3. It can be seen however that this rule makes a default assumption that all training stimuli are equally important as are all attributes i , which is practically never true. As a result, the associative capability of this type of learning is poor.

Therefore, for better performance other rules taking into account the history of learning are used. One of the most known is the Widrow-Hoff delta rule:

$$dC_{ij}^m = \alpha V_i^m (V_j^m - S_j^m), \quad 0 < \alpha < 1 \quad (8)$$

or another version of it

$$dC_{ij}^m = \alpha (V_i^m - S_i^m)(V_j^m - S_j^m). \quad (9)$$

where S_j^m is the postsynaptic potential computed in 3. It can be seen that this rule does use the knowledge of the past experience (as it uses \mathbf{C}), but is not perfect either, since the learning rate α is the same for all training stimuli, regardless of whether the stimulus is useful or not.

Nevertheless, this rule is one of the most frequently used in neural computation, where it is used iteratively; rather than applying one-step increments to all synapses, this rule is executed several times on the entire training sequence, until dC_{ij}^m becomes sufficiently close to zero. The iterative nature of this rule however assumes that all training stimuli are always available or stored somewhere, which for many applications, such as real-time face memorization from live video, are not.

This is why another incremental learning rule, which would take in account the level of usefulness of each stimulus as well as the importance of the stimuli attributes, is needed to make the performance of the model better.

III. PROJECTION LEARNING

The *projection learning rule*, also known as the *pseudoinverse learning rule*, has been originally proposed in 1971 by Amari, then intensively explored by Kohonen, and then intensively used and refined in several institutes in Europe by such researchers as Gascuel and Weinfeld, Garder and

Derrida, Diederich and Opper, Personnaz and Dreyfus, Kanter and Sarnopolsky, Gorodnichy and Reznik.

It is obtained from the condition that the consensus, based on the collective decision making of the neural network, is reached for all training stimuli [1]. Using Eq. 4 this leads to the following system of equations needed to be resolved for unknown synapse values:

$$\mathbf{C} \vec{V}^m = \lambda_m \vec{V}^m, \quad (m = 1, \dots, M), \lambda_m > 0, \quad (10)$$

which can be rewritten in matrix form for $\lambda_m = 1$ as

$$\mathbf{C} \mathbf{V} = \mathbf{V}, \quad (11)$$

where \mathbf{V} is the matrix made of column prototype vectors (training stimuli). Resolving this matrix equation gives us the rule:

$$\mathbf{C} = \mathbf{V} \mathbf{V}^+, \quad (12)$$

where $\mathbf{V}^+ \doteq (\mathbf{V}^T \mathbf{V})^{-1} \mathbf{V}^T$ is the *pseudoinverse* of matrix \mathbf{V} .

The matrix defined by Eq. 12 is the orthogonal projection matrix into the subspace spanned by the prototype vectors $\{\vec{V}^m\}$, which explains the name of the rule.

At first, this rule may look like a non-incremental batch rule, which needs to know all prototypes prior to learning. As shown in [1] however, by using the Greville formula, one can rewrite this rule as a one-step incremental rule too as follows:

$$\mathbf{C}^0 = \mathbf{0} \quad (13)$$

$$\mathbf{C}^m = \mathbf{C}^{m-1} + \frac{(\vec{V}^m - \mathbf{C}^{m-1} \vec{V}^m)(\vec{V}^m - \mathbf{C}^{m-1} \vec{V}^m)^T}{\|\vec{V}^m - \mathbf{C}^{m-1} \vec{V}^m\|^2} \quad (14)$$

or in scalar form as

$$dC_{ij}^m = \frac{1}{D^2(V^m)} (V_i^m - S_i^m)(V_j^m - S_j^m), \quad \text{where} \quad (15)$$

$$D^2(V^m) = \|\vec{V}^m - \mathbf{C} \vec{V}^m\|^2 = N - \sum_{i=1}^N V_i^m S_i^m. \quad (16)$$

If $\vec{V}^m = \mathbf{C}^{m-1} \vec{V}^m$, which means that the prototype \vec{V}^m is a linear combination of other already stored prototypes, then the weight matrix remains unchanged. $D(V^m)$ in Eq. 16 is the projection distance, which indicates how far a new stimulus is from those already stored and which can be used to filter out identical visual stimuli.

For iterative training if required, the following approximation of the projective learning rule can be used:

$$C_{ij}^m = \frac{\alpha}{D^2(V^m)} (V_i^m - S_i^m)(V_j^m - S_j^m), \quad 0 < \alpha < 1 \quad (17)$$

It can be seen that the projective learning, as described by Eqs. 15, 17, looks very similar to the correlation learning described in the previous section. It however takes into account more of what has been already learnt and, as a result, provides a better associative recall for the model.

A. Processing and memory considerations

The associative model based on the projection learning guarantees convergence of a network to an attractor using synchronous dynamics, as long as the weight matrix is symmetric [2]. This makes the network fast not only in memorization but also in recognition. For this, as proposed in [2] and justified

clips showing the persons (several facial images extracted from the training video-clips are shown at the bottom of the figure). The resolution of the video is kept low at 160x120 pixels, which is just sufficient for humans to identify the persons and which is known to be difficult for conventional image-based face recognition approaches. When the entire 20-minute recording of the show is then presented to the system, the systems identifies in real-time (by firing the appropriate neuron) who of the four guests is being shown at the moment.

V. CONCLUSION

In recognition rate achieved for the video-annotation problem described above is around 90%. That is, out of ten video-frames, the face is not recognized properly in one of them. However, since the final recognition decision is based on several consecutive frames, rather than a single frame, the actual recognition rate is close to 95%. This is quite a good result for an automatic face recognition system, taking into account the unconstrained environment within which it is tested. What is most important however is that the presented neuro-associative model allows one to approach the video-based recognition problem within a new, non Von-Neumann, biologically inspired framework, which is very much needed for this problem, since it takes advantage of video over still imagery and can deal with the continuous flow of video data. Rather than storing a continuous amount of individual video frames, the approach uses the incoming flow of visual data to continuously tune the synaptic connections of a multi-connected neural network, similar to the process of associative memorization in the visual cortex. Then, when a new video stimulus is presented to the retina, the neural network converges to a state which is best described by the past experience.

The described neuro-biological model based on the projective learning can be also applied to other recognition and associative tasks. The care has only to be taken not too saturate the model (when some weights significantly dominate the others as a result of the limited size of the system), which can be done by either using the iterative approximation of the rule or the network desaturation technique described in this presentation.

The binaries of our demo programs as well as the cpp code of the implementation of the basic incremental projection learning are available from our website.

REFERENCES

- [1] D.O. Gorodnichy, "Investigation and design of high performance neural networks (in russian)," *PhD dissertation, Glushkov Cybernetics Center of Ac.Sc. of Ukraine in Kiev* (online at www.cv.iit.nrc.ca/~dmitry/pinn), 1997.
- [2] D.O. Gorodnichy and A.M. Reznik, "Static and dynamic attractors of autoassociative neural networks," in *Proc. Int. Conf. on Image Analysis and Processing (ICIAP'97), Vol. II (LNCS, Vol. 1311)*, pp. 238-245, Springer, 1997.
- [3] D.O. Gorodnichy and A.M. Reznik, "Increasing attraction of pseudo-inverse autoassociative networks," *Neural Processing Letters*, vol. 5, no. 2, pp. 123-127, 1997.
- [4] S. Ozawa, K. Tsutsumi, and Norio Baba, "A continuous-time model of autoassociative neural memories utilizing the noise-subspace dynamics," in *Neural Processing Letters, Vol. 10, Issue 2*, pp. 97-109, 1999.

- [5] Leslie G. Ungerleider, Susan M. Courtney, and James V. Haxby, "A neural system for human visual working memory," in *Proc Natl Acad Sci U S A. 95(3): 883890*, 1998.
- [6] M. Perus, "Visual memory," in *Proc. Info. Soc.'01 / vol. Cogn. Neurosci. (eds. D.B. Vodusek, G. Repovs)*, pp. 76-79, 2001.
- [7] T. Gisiger, S. Dehaene, and J. Changeux, "Computational models of association cortex," in *Curr. Opin. Neurobiol. 10:250-259*, 2000.
- [8] E.Gould, A.J. Reeves, M.S.A. Graziano, and C.G. Gross, "Neurogenesis in the neocortex of adult primates," in *Science Oct 15 1999: 548-552*, 1999.
- [9] Dmitry O. Gorodnichy, "Video-based framework for face recognition in video," in *Second Workshop on Face Processing in Video (FPiV'05) in Proceedings of Second Canadian Conference on Computer and Robot Vision (CRV'05), 9-11 May 2005, Victoria, BC, Canada, 2005*.
- [10] Dmitry O. Gorodnichy, "Fully connected neural network as an adequate tool for face recognition in video," in *International Joint Conference on Neural Networks (IJCNN'05), Montreal, July 31 - August 4, 2005*.