A new penalisation term for image retrieval in clique neural networks

Romain Huet, Nicolas Courty, Sébastien Lefèvre

A new penalisation term for image retrieval in clique neural networks

Romain Huet, Nicolas Courty and Sébastien Lefèvre *

Univ. Bretagne-Sud, UMR 6074 IRISA
F-56000, Vannes, France
{firstname.lastname}@irisa.fr

Abstract. Neural networks that are able to retrieve store and retrieve information constitue an old but still active area of research. Among the different existing architectures, recurrent networks that combine associative memory with error correcting properties based on cliques have recently shown good performances on storing arbitrary random messages. However, they fail in scaling up to large dimensions data such as images, mostly because the distribution of activated neurons is not uniform in the network. We propose in this paper a new penalization term that alleviates this problem, and shows its efficiency on partially erased images reconstruction problem.

1 Introduction

While neural networks have been successfully used for decades, they have recently raised some new interest with the deep learning paradigm, where they achieve state-of-the-art results in various data processing tasks (e.g. in speech recognition or computer vision). Beyond the popular feedforward model, recurrent networks are appealing since their structure can be used a an associative memory. Such memories can be related to error correcting codes widely used in digital communications and signal processing [5, 9, 2]. We focus here on a specific clique-based model, namely the Gripon Berrou neural model (GBNN) [2], which uses the clique structure in the network adjacency matrix to efficiently retrieve the messages, and that has shown to outperform existing models like Hopfield [5] or Boltzman networks [6].

In this paper, we aim to apply the error correcting code principles brought by this associative memory to image data, as a way to recover partial occlusion of stored images. The GBNN architecture is a graph associated to a binary adjacency matrix indexing the connections in the network. While this binary architecture is of prime interest when considering hardware implementations, it faces ambiguity issues when applied to image data (i.e. pixel values) that are not independently and identically distributed: it is not able to recover properly the stored message.

We propose here to resolve this ambiguity problem by introducing some penalisation in the model, weighting the neurons based on their whole connectivity.

*The authors acknowledge the support of the CominLabs excellence cluster (SENSE project) and the Région Bretagne (SPANNVIS doctoral grant).

Results obtained on the standard dataset MNIST shows that the proposed strategy is able to reconstruct missing information conversely to the initial GBNN model.

2 Model

2.1 GBNN

We recall here the main principles of GBNN (for more details the reader is referred to [2]). The architecture is made of χ clusters, each of them containing l neurons. Information is stored through cliques, i.e. complete subgraphs (all edges are connected). Every message is thus mapped to a clique built by linking a single neuron for each active cluster (with c the number of such active clusters). While the initial GBNN model was defined as a full network with all clusters being active (i.e. $c = \chi$), it has been later shown [3] that using only a subset of clusters ($c \leq \chi$) allows storing more messages while keeping the network ability to recover the required information. We refer to this architecture as the sparse GBNN in the remainder.

A major drawback of the GBNN in the context of machine learning is related to its associative memory behavior, i.e. it lacks generalization capacity and can only retrieve and correct learnt messages. To illustrate, Fig. 1a shows how to learn a message $M = \{4, 4, 2, 3\}$ by filling the binary adjacency matrix W of a GBNN made of $c = \chi = 4$ clusters containing $l = 4$ neurons each.

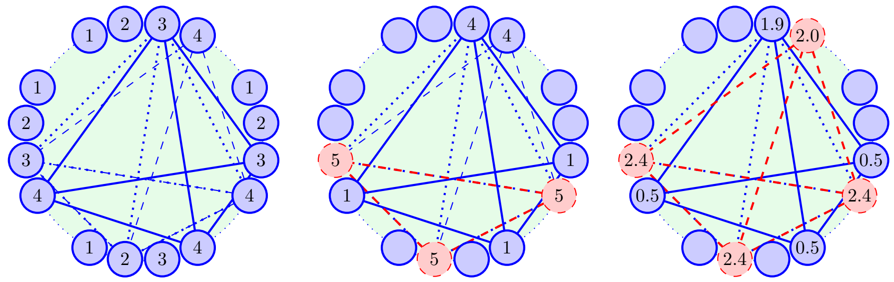
Correction occurs during the retrieval phase as an iterative process[1], made of two sucessive steps (called rules). First, a dynamical rule is applied to compute scores for all neurons involved in the retrieval process given an input message. Various rules exist and we consider here the *Sum-of-Sum* (SoS) rule, that assign to each neuron the sum of their common connections with the input message. This score is noted λ_i with $i \in [1, n = \chi \times l]$ the neuron index in the network, and is computed as:

$$\lambda_i = \gamma v_i + \sum_{j=1}^{n} W_{(i,j)} T_j v_j \tag{1}$$

with γ the memory effect (usually set to 1) controlling the weight of the neuron whose score is computed, v_i the state of the neuron i defined as 1 if activated, 0 otherwise. T is a binary representation of the input message M based on v_i codes (e.g. $T = 0001\ 0001\ 0100\ 0010$ for $M = \{4, 4, 2, 3\}$).W is the adjacency matrix and can be defined as the binary union of all the learnt messages T_i adjacency, *i.e.* $W = \bigcup_i T_i^t T_i$.

The second step consists in applying an activation rule, the goal of which is to select the neurons to be active in the next iteration of the retrieval process. Among the existing rules, we consider here the *Winner Take All* (WTA) one that will select within each cluster the neuron with the highest score

$$v_i = \begin{cases} 1, \text{if } i = \underset{j \in \text{cluster}(i)}{\arg\max}\ \lambda_j \\ 0,\ \text{otherwise} \end{cases} \tag{2}$$



(a) Trained GBNN (b) Original SoS scores (c) SoS + penalisation

Fig. 1: Illustration of the ambiguity occurring within a GBNN ($\chi = 4$, $l = 4$). 3 messages (dotted, dashed and solid: $\{4,4,2,3\}, \{3,3,4,4\}, \{3,4,2,3\}$) randomly generated are learnt (left), the red dashed one is to be recovered (middle), based on scores returned by the initial SoS rule (middle) and using proposed penalisation with $\beta = 0.05$ (right). The input message is $\{\_, 4, 2, 3\}$

Here lies the ambiguity problem, because in one cluster several neurons can share the same maximum score. In the case of a sparse network, the *Global k Winners Take All* (GkWTA) rule is preferred. It consists in selecting the k neurons with the highest scores through the network (with $k = c$).

The aforementioned ambiguity issues raised when multiple neurons have the same (best) score within a cluster, as illustrated by Fig. 1b. Such an issue is common for non iid data such as digital images. Furthermore, applying GBNN to images also requires to extend the initial model to deal with grayscale values, as discussed below.

2.2 Quantification

The naive strategy to store an image containing p pixels taking v possible grayscale values in a GBNN would be to define a network of $\chi = p$ clusters, each of them containing $v = l$ neurons. It is completely intractable, even for small image (e.g. for a greyscale 28×28 image, the size of W will be 200704^2!). To limit both the number of clusters and neurons, it is possible to rely on feature extraction and/or image quantification.

The image quantification can also be explained biologically by the pooling considered as a max operation, with the passage of the information from a simple cell to a complex cell in the visual cortex (V4) [7]. The powerfull feedforward convolutionnal neural networks also biologically inspired are using the pooling as reduction dimension that we can see as quantification [8].

Here we perform a quantification of the greylevels using a K-means clustering, and we thus have $l = K \ll 256$. However, by doing so we also introduce more ambiguity since only a few neurons will likely to be activated. Nevertheless, it is the only way to store pixel-based image representations in a GBNN.

2.3 Penalisation

The retrieval process described previously is iterative. It thus relies on a stopping criterion, e.g. a maximal number of iterations, a returned clique with the same scores for all selected neurons [1]. The latter criterion define that a clique is found when the scores of the neurons representing it have all the same scores. In the presence of ambiguity, this final configuration is very unlikely to be reached. To overcome this issue, we propose a new penalisation term based on the clique definition. We aim at promoting neurons activation that are similar to the activations of its connected neurons, *i.e.* suggesting to promote a clique in the network. For a neuron i, we thus define an activation score similarity s_i as:

$$s_i = \sum_{j=1, j \neq i}^{n} W_{j,i} |\lambda_j - \lambda_i| \tag{3}$$

This score can be appropriately normalised by the sum over all scores of each neurons of the same cluster, *i.e.* $\hat{s}_i = s_i / \sum_{k \in c_i}^{l} s_k$, with c_i designing all the neurons of the cluster where neuron i belongs to. Finally, in order to incorporate this new prior in a new dynamic rule, we propose to use a sigmoid activation rule based on this score, reading:

$$\tilde{\lambda}_i = \frac{\lambda_i}{1 + e^{-\beta s_i}}, \tag{4}$$

where β is a parameter that governs the strength of this prior over the network convergence behaviour. As a result, this new penalisation term will tend to promote solutions at each iteration where the neurons are connected through the adjacency matrix to neurons that have a similar score. As will be shown in the experiments, this allows to alleviate the ambiguity problem up to a broader density of the information in the network. Using this penalty term, we show as an illustration in Fig. 1c the updated scores, that no longer present an ambiguity.

3 Experimentation

The evaluation of the proposed strategy for disambiguation is conducted with an experimental setup where learnt images are partially erased and are to be recovered by the GBNN. For this purpose, we used a publicly available implementation of the GBNN[1].

Experimental setup We consider the standard dataset MNIST, characterized with a low spatial resolution allowing to use the GBNN directly in the pixel space and not through a feature extraction step. It is composed of small images (28×28 pixels) representing handwritten digits, and divided in training and testing sets. We use here only the training set.

Result assessment is achieved through measuring the average RMSE over the partially erased images that are submitted to GBNN for retrieval purposes. We

[1] available at `https://github.com/lrq3000/gbnn-matlab`

compare the standard dynamic rule SoS and the proposed rule with penalisation (SoSP). The activation rule is WTA when the network is full, and GkWTA with $k = c$ for a sparse network. This level of parcimony was set to the $c = \chi = 196$ in our experiments. The results are reported in table 1. In the case of SoSP, the β parameter was determined empirically through cross-validation.

Results As expected, the higher the quantification level, the more messages the network is able to store efficiently. This is mostly due to the lesser probability of ambiguities. While we reach a maximum density in the full architecture of the network with $l = 10$ sooner than in the case $l = 20$ (Fig. 2a), we can observe that the average RMSE is better with SoSP until a point where both strategies lead to the same result. As for the sparse version of the network, since there is less neurons needed to represent the information the ambiguity is pushed back. This is why even when the full version can not handle more images, the sparse version appears to be still efficient.

While the proposed method allows to push a little further the limits in term of number of stored messages, it comes at a price of computing the extra penalisation term in the dynamic rule. However this extra charge is tempered by faster convergence rates. We will present these rates, along with supplementary tests on other images, in a technical report following the acceptance of the paper. The generalization capacity of the network (retrieval of known cliques from unknown images) endowed with this penalization term is also an interesting direction, but is out of the scope of this paper.

4 Conclusion

The GBNN is an associative memory with a promising information recovery ability. However, when data are not iid such as the case of digital images, the network quickly faces ambiguity issues, and thus fails to retrieve learnt information. We address this ambiguity issue by forcing distinct scores among neurons from the same cluster, through a penalisation based on the neuron probabilty to belong to stored messages or cliques. Experimental results conducted on the MNIST dataset shows that the proposed strategy greatly reduces ambiguity, and as such improves the reconstruction of partially erased images (Fig. 2b, 2c).

To foster the application of GBNN on images, we are considering reducing the number of neurons, and moving from a pixel-based representation to another feature space. Features extracted from an autoencoder are here of high interest since they will increase the generalization ability of the GBNN as well as make possible its application on higher resolution images.

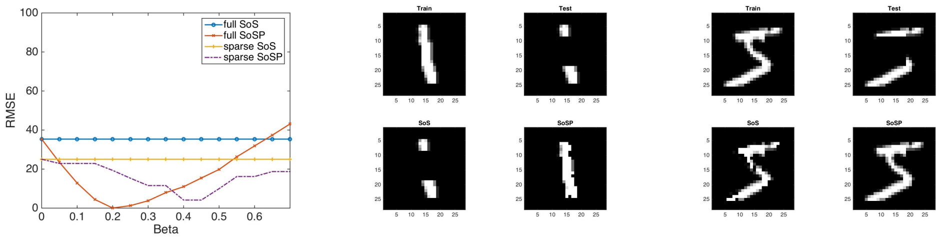
References

[1] A. Aboudib, V. Gripon and X. Jiang, A study of retrieval algorithms of sparse messages in networks of neural cliques, preprint arXiv:1308.4506 , 2013.

[2] V. Gripon and C. Berrou, A simple and efficient way to store many messages using neural cliques, *Computational Intelligence, Cognitive Algorithms, Mind, and Brain* (CCMB, 2011), pages 1-5, 2011.

Type	# Train / Quantif.	30		648		1265		1883		2500	
Full	l=10	35.4	**0.0**	49.4	**38.4**	46.3	46.3	47.3	47.3	48.3	48.3
	l=20	27.4	**26.5**	49.6	**2.7**	46.5	**23.9**	48.5	**47.5**	48.3	48.3
Sparse	l=10	25.0	**4.1**	57.6	**45.1**	51.4	**46.0**	52.7	**49.5**	53.0	**49.5**
	l=20	12.5	**4.1**	47.1	**30.2**	52.2	**48.9**	51.1	**43.3**	51.1	**45.9**
Method		SoS	SoSP	SoS	SoSP	SoS	SoSP	SoS	SoSP	SoS	SoSP

Table 1: Mean RMSE over test images (included in the training set) from the MNIST data set. Two configurations of the network are considered (full/sparse), with two different levels of quantification and number of training images



(a) Mean over all the test, $l = 10$ with 30 images learned

(b) Reconstruction with a full network, $l = 20$, $\beta = 0.05$

(c) Reconstruction with a sparse network, $l = 20$, $\beta = 0.15$

Fig. 2: Mean of reconstruction results over 15 MNIST training images partially erased by a fix hole, with a full network composed of $l = 20$ neurons for each of the $c = \chi = 784$ clusters in the full version (f) and $c = 196$ in the sparse version (s), in function of the number of messages learned. The retrieval is done with the dynamic rule SoS and activation rule WTA for the full network, and GkWTA for the sparse one

[3] BK. Aliabadi, C. Berrou, V. Gripon and X. Jiang, Learning sparse messages in networks of neural cliques, preprint arXiv:1208.4009, 2012.

[4] V. Gripon and C. Berrou, Sparse neural networks with large learning diversity, *Neural Networks*, 22.7:1087-1096, IEEE, 2011.

[5] JJ. Hopfield, Neural networks and physical systems with emergent collective computational abilities, *Proceedings of the national academy of sciences*, vol 79, no 8, pages 2554-2558, 1982.

[6] DH. Ackley, GE. Hinton and TJ. Sejnowski, A learning algorithm for boltzmann machines*, *Cognitive science*, vol 9, no 1, pages 147-169, 1985.

[7] T. Serre and T. Poggio, A neuromorphic approach to computer vision, *Communications of the ACM*, vol 53, no 10, pages 54-61, 2010.

[8] Y. LeCun, L. Bottou, Y. Bengio and P. Haffner, Gradient-based learning applied to document recognition, *Proceedings of the IEEE*, vol 86, no 11, pages 2278-2324, 1998.

[9] S. Hobson, Correlation Matrix Memories: Improving Performance for Capacity and Generalisation, 2011.