Neural Network Nonlinear Factor Analysis of High Dimensional Binary Signals

Dušan Húsek and Hana Řezanková and Václav Snášel
Institute of Computer Science
Academy of Sciences of the Czech Republic
Pod Vodárenskou věží 2
Prague 8, Czech Republic
dusan@cs.cas.cz, rezanka@vse.cz, vaclav.snasel@vsb.cz

Alexander A. Frolov and Pavel Polyakov
IHNA Russian Acad. of Sci.
117 485 Moscow, Russia
aafrolov@mail.ru

Abstract

Possible application of a new neural network suitable for binary factorization of signals of large dimension and complexity is introduced. We developed the new recall procedure of Hoppfield-like associative memory which allows search all attractors corresponding to factors (a true attractor). Necessary separation of spurious attractors is based on calculation of their Lyapunov function. Being applied to textual data the procedure allows to reveal groups of highly correlated words (factors) which frequently occur in documents jointly and represent topics of that documents.

1. Introduction

Factor analysis is one of the most efficient statistical method to overcome information redundancy of high-dimensional data set. Factors extraction is a procedure which maps objects from original space variables into the space of factors. One particular form of nonlinear factorization is a binary one, where a complex vector signal (pattern) has a form of the Boolean sum of weighted binary factors:

$$\mathbf{X} = \bigvee \mathbf{S}_l \mathbf{f}^l. \tag{1}$$

In this case, original signals $\mathbf{X}$, factor scores $\mathbf{S}_l$ and factor loadings $\mathbf{f}^l$ are binary, i.e. its components possess the values 0 or 1.

It was a challenge for us [Frolov et al., 2004] to utilize binary factorization of Hopfield-like? [Hopfield, 1982] neural network with parallel dynamics because it has a lot of similarities with the iterative procedure [Kinzel, 1985] for linear factorization. But there is a peculiarity that we have to solve. According to our paradigm, the network is learned by signals from original space. During learning phase attractors of two types are created in the energy landscape- true (corresponding to factors) and spurious (araising as a learning byproduct). So we had to develop a procedure to separate true attractors which are corresponding to factors from spurious ones. Another principal requirement on developed method is that factors are are supposed to be found sufficiently fast even if we have no a priory information about them, and much faster if we have. Starting from random initial state, network activity stabilizes in some attractor which corresponds to one of factor or spurious state.

This convergence is very fast: the most strong factors could be revealed for only several iteration steps. To separate true and spurious attractors we found procedure based on calculation of their Lyapunov function [Goles-Chacc and Fogelman-Soulie, 1985]. Unlearning of already found factors prevent against their repeated retrieval. Some background on this topic can be found in the work [Frolov et al., 2003].

2. Hopfield network

The neural network under consideration consists of N neurons of the McCulloch-Pitts type (integrate-and-fire binary neurons) with gradually ranged synaptic connections between them. Only a fully connected case is considered here.

Network is trained by a set of M patterns which are supposed to have the form $\mathbf{X}^m = \bigvee_{l=1}^{L} \beta_l^m \mathbf{f}^l$, where $\mathbf{f}^l \in B_n^N$ [1]

are L factors (N dimensional vectors) and for every m-th pattern $\beta_l^m \in B_C^L$ it is a corresponding factor scores vector. As follows from the definition every factor contains exactly $n = Np$ ones. Every complex pattern $\mathbf{X}^m$ contains in turn exactly the C factors, so it is quite natural to call the *complexity* of the pattern as C, i.e. as number of factors which are included in pattern. We assumed factors and factor scores to be statistically independent. In a limit case $C = 1$ patterns become pure factors and we obtain an ordinary Hopfield case.

2.1. Learning procedure

The connection matrix **J** of this network is a covariation matrix of input signals obtained by using the correlational Hebbian learning rule:

$$J_{ij} = \sum_{m=1}^{M} (X_i^m - q^m)(X_j^m - q^m),\ i{\neq}j,\ J_{ii} = 0, \quad (2)$$

where M is the number of patterns in the learning set and $q^m = \sum_{i=1}^{N} X_i^m / N$ is the total activity of the m-th pattern.

To describe recall procedure we must introduce several equations. The activity of neural network is determined by iterative procedure:

$$X_i(t+1) = \Theta(h_i(t) - T(t)),\ i = 1, \cdots, N \quad (3)$$

where Θ - step function, and $T(t)$ - activation threshold. The threshold $T(t)$ is chosen at each time step in such a way that the level of the network activity is kept constant and equal to n. Energy landscape can be described in terms of following Lyapunov function

$$\Lambda(t+1) = \mathbf{X}^T(t+1)\mathbf{J}\mathbf{X}(t). \quad (4)$$

Activity of Hopfield-like network with parallel dynamics converges not only to point attractors [Goles-Chacc and Fogelman-Soulie, 1985] but also to cyclic attractors of the length two.

Theoretical analysis and computer simulation performed by Frolov et al. [Frolov et al., 2004] completely confirmed the validity of Hopfield-like network for binary factorization. However, Hopfield-like network has one principal peculiarity. The network dynamics converges to one of the factors (true attractor) only when initial state falls inside its attraction basin. Otherwise it converges to one of the spurious attractors. Thus binary factorization requires special recall procedure to separate true and spurious attractors.

2.2. Recall procedure

To separate true and spurious attractors we developed two-run recall procedure. Its initialization starts by presentation of random initial pattern $\mathbf{X}^{in}$ with $k_{in} = r_{in}N$ active neurons. On presentation of $\mathbf{X}^{in}$, network activity $\mathbf{X}$ evolves to some attractor. The evolution is determined by equation (3). On each time step k_{in} “winners” (neurons with the greatest synaptic excitation) are chosen and only they are active on the next time step. When activity stabilizes at the initial level of activity k_{in}, $k_{in} + 1$ neurons with maximal synaptic excitation are chosen for the next iteration step, and network activity evolves to some attractor at the new level of activity $k_{in} + 1$. Then level of activity increases to $k_{in} + 2$, and so on, until number of active neurons reaches the final level $r_f N$. Thus, the whole procedure (one trial) contains $(r_f - r_{in})N$ iteration steps and several time steps inside each iteration step to reach some attractor for fixed level of activity.

At the end of each iteration step a relative Lyapunov function was calculated by formula: $\lambda = \Lambda/(rN)$ where Λ is given by (4). The relative Lyapunov function gives a mean synaptic excitation of active neurons. The time course of the relative Lyapunov function along the recall trajectory provides criterion for separation of true and spurious attractors which is based on identification of specific breaking point. Attractors with the highest Lyapunov function would be obviously winners in the most trials of the recall process. Thus, more and more trials are required to obtain new attractor with relatively small value of Lyapunov function. To overcome this problem attractors with high Lyapunov function should be deleted from the network memory. The deletion was performed according to Hebbian unlearning rule by substraction $\Delta J_{ij}, j \neq i$ from synaptic connections J_{ij} where

$$\begin{aligned}\Delta J_{ij} \quad = \quad & \frac{\eta}{2} J(\mathbf{X})[(X_i(t-1) - r)(X_j(t) - r) + \\ & + (X_j(t-1) - r)(X_i(t) - r), \end{aligned} \quad (5)$$

$J(\mathbf{X})$ is the average synaptic connection between active neurons of the attractor, $\mathbf{X}(t-1)$ and $\mathbf{X}(t)$ are patterns of network activity at last time steps of iteration process, r is the level of activity, and η is an unlearning rate. For point attractor $\mathbf{X}(t) = \mathbf{X}(t-1)$ and for cyclic attractor $\mathbf{X}(t-1)$ and $\mathbf{X}(t)$ are two states of attractor.

1 $B_n^N = \{X | X_i \in \{0,1\}, \sum_{i=1}^{N} X_i = n\}$

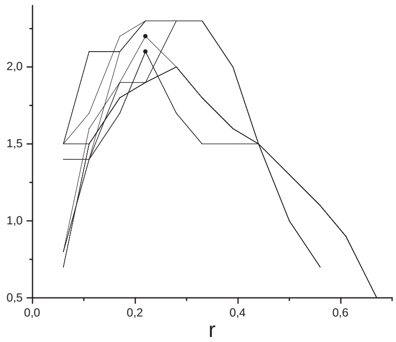

Figure 1. Relative Lyapunov function λ in dependence on the relative network activity r for 15 titles of medical articles. Circles are points of breaking which were identified as indexes of factors.

3. Computer simulation

We tested our procedure over different examples from literature and text collections. Binary factorization of text documents implies that we need to find groups of highly correlated words each of which represent one topic, and then describe documents in terms of revealed topics. So by means of binary factorization someone could mine from document collection its common properties and general features, construct rubrics and reduce dimension of documents representation.

First, we tested binary factorization over the list of titles of 15 medical articles presented in [Berry and Browne, 1999]. Constructed dictionary contains all words from the titles. The titles were transformed to binary vectors with 18 component according the rule: if a word occur in title, corresponding component equals to 1, otherwise 0. The obtained binary codes were stored in the network of 18 neurons according to (2). Each trial was initiated by activation of one of 18 neurons. Thus the total recall procedure includes only 18 trials. Only two factors were revealed according to the used criterion see Fig.1. The first factor contains words: blood, close, disease and pressure. The second: fast, rats, rise and pressure. It is interesting that the words “culture”, “discharge” and “patients” do not create a factor in spite of the fact that they are included into two first titles and, hence, one can expect that they should be tightly connected. However in these titles the word “culture” has different meaning and its banding with words “discharge” and “patients” is not reasonable. Thus we can conclude that our method could be sensitive to the context in which the words are used.

Second we applied our method to the set of 21000 messages of agency Reuters [Reuters, 2004, Rose et al., 2002] as well. The used vocabulary contained 5000 the most often words in the set (consequently network contained 5000 neurons). Each message was transformed to binary code dependently on presence or absence of words in the message. Each found factor was deleted from the network memory according to (5) with $\eta = 1$. Fig. 2 demonstrates the first 10 trials which were identified as true. Circles mark the points of curve breaking. All found factors happened to be reasonable and mirror the content of the corresponding messages.

Our method combines words in factors not only according to the frequency of their appearance together at the messages but mainly according to their appearance at the same context. We see that different factors reflect different contexts of word utilization and different topics of news messages, while messages with the same topics are connected with the same factors.

Two messages with highlighted words creating factors are shown below, as an example of the point. These factors may appear in different news messages. But if in several messages the same factors are revealed, then these messages should have the same topic. In particular, the topics of messages from example are *Japanese foreign commerce* and *activity of American administration*. Evidently, factors reflect mutual meaning of the messages quite right.

″ Message 1

U.S. ASKS **JAPAN** TO END AGRICULTURE IMPORT CONTROLS **TOKYO**, March 3

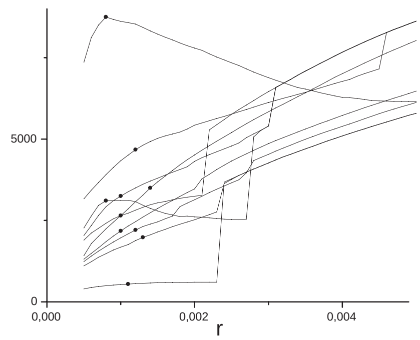

Figure 2. Relative Lyapunov function λ in dependence on the relative network activity r for 21000 messages of agency Reuters. Circles are points of breaking which were identified as indexes of factors.

The U.S. Wants **Japan**[1] to eliminate import controls on agricultural products within three years, visiting U.S. Under-Secretary of State for **Economic**[1] Affairs Allen Wallis **told**[2] Eishiro Saito, Chairman of the Federation of **Economic**[1] Organisations (Keidanren), a spokesman for Keidanren said. The spokesman quoted Wallis as saying drastic measures would be needed to stave off protectionist legislation by **Congress**[3] .Wallis, who is attending a sub-cabinet-level bilateral **trade**[1] meeting, made the remark yesterday in talks with Saito. Wallis was quoted as saying the **Reagan**[3] **Administration**[3] wants **Japanese**[1] cooperation so the **White House**[3] can ensure any U.S. **Trade bill**[1] is a moderate one, rather than containing retaliatory measures or antagonising any particular country. He was also quoted as saying the U.S. Would be pleased were **Japan**[1] to halve restrictions on agricultural imports within five years if the country cannot cope with abolition within three, the spokesman said. **Japan**[1] currently restricts imports of 22 agricultural products. A ban on rice imports triggered recent U.S. Complaints about **Japan's**[1] agricultural policy.

″ Message 2

U.S. COMMERCE SECRETARY QUESTIONS FUJITSU DEAL **WASHINGTON**, March 3

Commerce Secretary Malcolm Baldrige said he felt a proposed takeover by **Japan's**[1] <Fujitsu Ltd> of U.S.-based Fairchild Semiconductor Corp, a subsidiary of Schlumberger Ltd <SLB>, should be carefully reviewed. He **told**[2] the Semiconductor Industry Association the deal would soon be discussed by representatives of several different **government**[3] departments. The **Reagan administration**[3] has previously expressed concern that the proposed takeover would make Fujitsu a powerful part of the U.S. **market**[1] for so-called supercomputers at a time when **Japan**[1] has not bought any American-made supercomputers. In addition, U.S. defense **officials**[3] have said they were worried semiconductor technology could be transferred out of the United States, eventually giving **Japanese**[1]-made products an edge in American high-technology markets for defense and other goods. Treasury Secretary James Baker recently **told**[2] a **Senate**[3] committee the proposed takeover would be reviewed by the cabinet-level **Economic**[1] Policy Council.

Here terms marked [1] are contained in the first factor, terms marked [2] are common words - contained in both factors and terms marked [3] are words contained in the second factor. One can see that factorizations is really nonlinear as there is nonempty set of common words.

4. Conclusion

In this work we have shown next step in development of Hopfield based neural network capable of performing binary factorization of the signals of high dimension and complexity. Advantage of our NN attempt should be the possibility of incremental learning and capability to analyze large multidimensional data sets. This method is suitable for text collections analysis as shown in example. Being applied to textual messages of agency Reuters [Reuters, 2004], [Rose et al., 2002], result showed not only full applicability of this method but moreover sensitivity to the context in which the words were used. Therefore we see big future potential for this application.

Acknowledgement: This work was supported by the grants No. IM6840070004 and IM4674788502 awarded by the Ministry of Education of the Czech Republic .

References

[Berry and Browne, 1999] M.W. Berry and M. Browne. *Understanding Search Engines: Mathematical Modeling and Text Retrieval*. SIAM, NY, 1999.

[Frolov et al., 2003] A.A. Frolov, D. Husek, and P. Muravjev. Informational efficiency of sparsely encoded Hopfield-like autoassociative memory. *Optical Memory and Neural Networks (Information Optics)*, pages 177–198, 2003.

[Frolov et al., 2004] A.A. Frolov, A.M. Sirota, D. Husek, and P. Muravjev. Binary factorization in Hopfield-like neural networks: single-step approximation and computer simulations. *Neural Networks World*, pages 139–152, 2004.

[Goles-Chacc and Fogelman-Soulie, 1985] E. Goles-Chacc and F. Fogelman-Soulie. Decreasing energy functions as a tool for studying threshold networks. *Discrete Mathematics*, pages 261–277, 1985.

[Hopfield, 1982] J. J. Hopfield Neural network and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Science USA*, **79**, 1982, pages2544–2548.

[Kinzel, 1985] W. Kinzel Learning and pattern recognition in spin glass models. *Z Physik B]* **60**, 1985, pages 205-213

[Reuters, 2004] Reuters. *http://about.reuters.com/researchandstandards/corpus/*. Reuters, NY, 2004.

[Rose et al., 2002] T. Rose, M Stevenson, and M. Whitehead. The Reuters corpus volume 1 - from yesterday's news to tomorrow's language resources. In *Proceedings of the Third International Conference on Language Resources and Evaluation*, pages 397–402, 2002.