Ranking neurons for mining structure-activity relations in biological neural networks: *NeuronRank*

Tayfun Gürel[a,b,*], Luc De Raedt[a,b], Stefan Rotter[a,c]

[a]*Bernstein Center for Computational Neuroscience, Freiburg, Germany*
[b]*Institute for Computer Science, Albert-Ludwigs-University Freiburg, Germany*
[c]*Institute for Frontier Areas of Psychology and Mental Health, Freiburg, Germany*

Available online 6 November 2006

Abstract

It is difficult to relate the structure of a cortical neural network to its dynamic activity analytically. Therefore we employ machine learning and data mining algorithms to learn these relations from sample random recurrent cortical networks and corresponding simulations. Inspired by the *PageRank* and the *Hubs & Authorities* algorithms, we introduce the *NeuronRank* algorithm, which assigns a source value and a sink value to each neuron in the network. We show its usage to extract structural features from a network for the successful prediction of its activity dynamics. Our results show that *NeuronRank* features can successfully predict average firing rates in the network, and the firing rate of output neurons reflecting the network population activity.
© 2006 Elsevier B.V. All rights reserved.

Keywords: Structure–function relations; Link mining; Machine learning

1. Introduction

Most functions of our brain are mediated by the operation of complex neuronal networks. The relation between structure and function of the various types of networks has been subject of many theories and intense computational modeling. Fundamental questions, however, remain unanswered: How important is the structure of a network for its function? Is a certain type of structure essential for a particular function? Can one and the same structure support different functions? Can different structures support the same function? How does the repeated usage of a network change its structure and its function, respectively? How does the interaction between networks affect the function of the whole system?

We approach some of these questions by systematically exploring the relation between network structure and activity dynamics in network models of the cortex. The analysis of Brunel [3] and others showed how the complex dynamics of a random-topology cortical network is determined by various structural parameters. In particular, the influence of the relative strength of the inhibitory synaptic couplings in the network and the role of external inputs was elucidated. The question how structural variations contribute to variations in activity dynamics, however, was not tackled in this work. Several recent papers indicate that structural variations indeed influence the network dynamics [1,12].

Neural networks in the brain have, at the structural level, the same format as social networks, food webs, citation networks, the Internet, or networks of biochemical reactions: They can be represented by large graphs, linking many interacting elements to each other. Empirical data of this format are also called 'networked' data. Recently, mining networked data has gained a lot of interest and has resulted in a new subfield called *link mining* [6,7]. Kleinberg [8] proposed the *Hubs & Authorities* algorithm, which is able to detect authoritative sources of information on the web by exploiting its link structure. Page et al. [11] introduced the *PageRank* algorithm underlying the Google search engine, which successfully predicts the relevance of a web page to the user and ranks the page for him, by again exploiting link information.

---

*Corresponding author. Institute for Computer Science, Albert-Ludwigs-University Freiburg, Germany.

E-mail address: guerel@informatik.uni-freiburg.de (T. Gürel).

This note investigates the applicability of link mining techniques to reveal structure–activity relations in biological neural networks. In particular, we are interested in learning a function that maps structural features of neural networks to activity-related features. We introduce the *NeuronRank* algorithm, which yields structural features describing the level to which neurons are *functionally* excitatory and/or inhibitory within a recurrent network. *NeuronRank* is inspired by the *Hubs & Authorities* algorithm, and is shown to yield good predictions of network activity. We proceed by giving an overview of our approach in Section 2. In Section 3, we present our network model. We explain how we analyze the network activity in Section 4. We introduce our key contribution, the *NeuronRank* algorithm, in Section 5. We describe our structural feature extraction methodology in Section 6. In Section 7, we refer to the machine learning algorithms we employ. We finally present our experimental results in Section 8 and our conclusions in Section 9.

2. Overview of the method

Aiming at discovering structure–activity relations in recurrent cortical networks, we focus here on the following specific problem: can we extract meaningful structural features from a random-topology network and use these to predict the characteristics of its activity dynamics? As this problem cannot be solved with current analytical techniques, we tackle it with machine learning and link mining methods. These algorithms learn the desired mappings from a set of examples. In our case, an example consists of a set of values for structural features and the corresponding activity features. Fig. 1a depicts a schematic overview of our approach.

Various structural features of the networks were extracted, based on simple counting statistics and on the new *NeuronRank* algorithm. We also performed numerical simulations of the activity dynamics exhibited by these networks, and then measured the mean firing rates and other characteristic parameters describing the activity dynamics. Eventually, machine learning and data mining algorithms were applied to those data, allowing us to detect any relations between structure and dynamics. Our methods generated statistical models, which were able to predict the dynamics of unseen networks based on their structural features. We assessed the quality of these models by determining their predictive power.

3. The network model

We used the leaky integrate-and-fire neuron model with the following parameters: membrane time constant 20 ms, membrane capacitance 250 pF, spike threshold 20 mV, reset potential 10 mV, refractory period 2 ms. Synaptic currents were modeled as δ-pulses, delayed by 1.5 ms with respect to the inducing action potential, the amplitude of excitatory postsynaptic potentials was 0.1 mV, inhibitory postsynaptic potentials had an amplitude of -0.6 mV.

We created recurrent neural networks of $n = 1000$ integrate-and-fire neurons, according to a simple statistical characterization of the neocortex with respect to neuron types and synaptic connectivity [2]. Each of the $n(n-1)$ potential synapses was established with probability 0.1, independently of all the others. Neurons were inhibitory
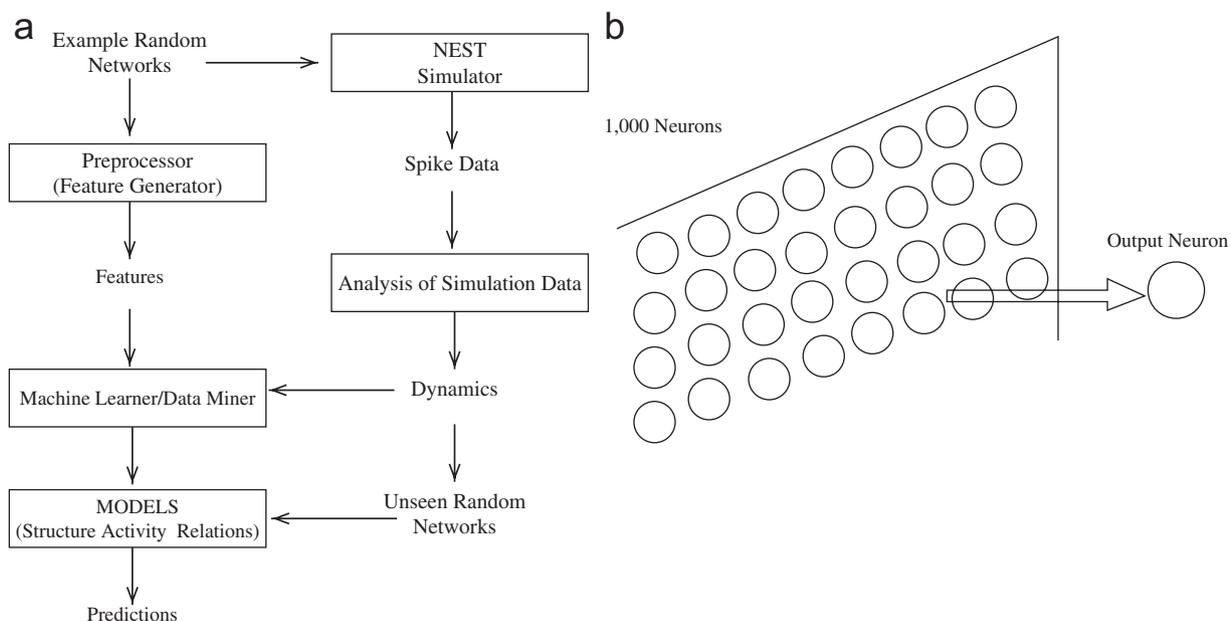

Fig. 1. (a) Mining structure–activity relations in biological neural networks. (b) Setup of the numerical simulations. We simulated recurrent cortical networks of 1000 neurons. Each neuron in the network received external input in the form of a excitatory Poisson spike train with mean rate slightly above the threshold for sustained activity. All neurons in the network projected to a single 'readout' neuron, which did not receive extra external inputs.

with probability 0.2 and excitatory otherwise. All synapses on the axons of any particular neuron had the same sign. All network simulations were performed using the NEST simulator [5,10].

4. Activity-related features

We studied steady-state dynamics of activity in the network, based on the spike firing characteristics of the neurons. We focused on two particular aspects of activity dynamics.

Average firing rate: The networks were simulated for 1.2 s. Spike counts and rates were determined and averaged over all neurons

$$v_{\text{avg}} = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{T}\int_{0}^{T}\sum_{k}\delta(t - t_{ik})\,\mathrm{d}t,$$

where n is the number of neurons in the network, T is the duration of simulation and t_{ik} is the time of the k-th spike in the i-th neuron.

Firing rate of a readout neuron: The cortex is composed of many interacting local networks. It is, therefore, an interesting question how the activity of a local network affects other neurons or networks it is connected to. Here we considered the case of a single readout neuron that receives input from all neurons of a network (Fig. 1b). We were particularly interested in how the firing rate of the readout neuron depends on the structural variations in the local network. The firing rate of the readout neuron was defined as

$$v_{\text{out}} = \frac{1}{T}\int_{0}^{T}\sum_{k}\delta(t - t_{k})\,\mathrm{d}t,$$

where T is the duration of simulation and t_k is the time of the kth spike that the readout neuron fires.

5. The *NeuronRank* algorithm

The *NeuronRank* algorithm, which is introduced below, assigns a *source value* α_i and a *sink value* ω_i to each neuron i, based only on structural information. The source value α_i of a neuron encodes the net effect on the network induced by a spike in that neuron. As a rule, excitatory neurons will have positive source values, whereas inhibitory neurons will have negative source values. Exceptions from this rule, however, may exist. Namely, if an excitatory neuron excites many inhibitory neurons, it may attain a negative source value. On the other hand, if an inhibitory neuron inhibits many other inhibitory neurons, it may attain a positive source value. The absolute source value of a neuron is an indicator for its total impact on network activity. The sink value ω_i, on the other hand, encodes the sensitivity of a neuron for activity somewhere else in the network. Neurons with higher sink values tend to be excited more by other neurons and therefore tend to have higher firing rates.

In a recurrent network, the source value of a neuron depends on the source values of all other neurons. In other words, the vector of all source values in a network recursively depends on itself, and the same holds for the vector of sink values. We propose here to use the *NeuronRank* algorithm to find a consistent set of source and sink values in a network. It iteratively updates the source value of a neuron according to the source values of its postsynaptic nodes. If A denotes the weighted adjacency matrix of the network

$$A_{ij} = \begin{cases} 1 & \text{for an excitatory synapse } j \rightarrow i, \\ -g & \text{for an inhibitory synapse } j \rightarrow i, \\ 0 & \text{otherwise}, \end{cases}$$

where $g > 0$ is a number that encodes the relative impact of inhibitory couplings relative to excitatory ones. The update rule for the row vector of source values $\alpha = (\alpha_1, \ldots, \alpha_n)$ is given by

$$\alpha \leftarrow \alpha A$$

starting with initial values $\alpha_i \pm 1$ depending on whether neuron i is excitatory or inhibitory. In contrast to source values, the sink value of a neuron is updated according to the sink values of its presynaptic nodes. The update rule for the column vector of sink values $\omega = (\omega_1, \ldots, \omega_n)^{\mathrm{T}}$ is therefore given by

$$\omega \leftarrow A\omega,$$

starting with initial values $\omega_i = 1$ for all neurons. In each step of the iteration both α and ω are normalized to unit length, and the iteration stops upon convergence. The detailed algorithm is depicted in Table 1.

6. Structural features

Upon convergence of the *NeuronRank* algorithm, statistical summary information about the source and sink

Table 1
The NeuronRank algorithm

Input: A directed labeled (inhibitory/excitatory) recurrent network N, represented by a weighted adjacency matrix A.
Output: Source (α) and sink (ω) values of all nodes in N
for each node i in N
    $\omega_i \leftarrow 1$
    if i is excitatory
        $\alpha_i \leftarrow 1$
    else if i is inhibitory
          $\alpha_i \leftarrow -1$
      endif
    endif
endfor
repeat
    $\alpha \leftarrow \alpha A$
    $\omega \leftarrow A\omega$
    normalize α and ω such that $\sum_i \alpha_i^2 = 1$ and $\sum_i \omega_i^2 = 1$
until convergence
return α and ω

values in a network is passed on to data mining algorithms. We considered in particular mean and variance of the source values, and separately of the sink values of all neurons. In addition, mean, sum and variance were computed separately for excitatory or inhibitory neurons only. This yielded a set of total 16 *structural features*.

In order to assess the quality of the features obtained by the *NeuronRank* algorithm, we compared them to features obtained by counting network motifs [13] of second and third order. There are 16 third-order network motifs ignoring the neuron type (excitatory vs. inhibitory), and 93 third-order network motifs that take the neuron type into consideration. We also computed the average clustering coefficient [14] in each network and used it as an additional feature in our machine learning algorithms.

7. Machine learning algorithms

We applied three well-known machine learning algorithms implemented in the WEKA workbench for data mining [15] to detect structure-activity relations of random cortical networks. In particular, we employed J48, a decision tree learner, which uses the information gain as its heuristic, the K2 algorithm for learning Bayesian Networks, and Support Vector Machines [4].

8. Experimental results

In order to answer the question *whether NeuronRank features are good predictors for the network activity*, we set up the following experiments.

Predicting the average firing rate: We generated 330 random networks and performed numerical simulations of their activity dynamics. For each network, we measured the firing rate averaged across all neurons. Firing rates above the median were labeled as 'high', below the median as 'low'. The task then was to predict the firing rate correctly ('high' vs. 'low'), based on the features extracted by motif counting (cc: clustering coefficient, inh: inhibitory neuron count, 3m: third order motifs, 2ie: second order motifs with signs, 3ie: third order motifs with signs) and by the *NeuronRank* algorithm (mean, variance and sum of source and sink values). Note that running NeuronRank took 25.2 s in average per network on a computer with a Pentium-4 3.2 GHz processor and Linux Suse 10.0 operating system, whereas running the third order motif counter (with signs) took 255.2 s in average per network on the same system. We tested the prediction accuracy of three machine learning algorithms, using 10-fold cross validation. The results are shown in Table 2.

Predicting the firing rate of a readout neuron: In the numerical simulations, a readout neuron was added to each of the networks described above. This neuron received input from all neurons in the network, but no external input. We considered the same structural features as in the previous setting (ignoring the readout neuron) and trained the machine learning algorithms to predict the firing rate of the readout neuron as 'low' or 'high' on unseen networks. The results are shown in Table 2.

The results clearly show the success of the *NeuronRank* features. For both experiments, *NeuronRank* features contributed significantly to the accuracy of the prediction, and hence they can be regarded as good indicators of network activity.

9. Conclusions

We showed that it is possible to gain knowledge about certain aspects of the activity dynamics in random cortical networks by employing machine learning and data mining techniques. Furthermore, we demonstrated that the *NeuronRank* algorithm, which is related to the *Hubs & Authorities* and *PageRank* algorithms, can successfully extract structural features that are relevant for predicting

Table 2

Accuracy of prediction in random networks using three well known machine learning algorithms: K2 Bayesian Network algorithm, J48 decision tree learner, and Support Vector Machines (SVM) using first and second order polynomial kernels

Features	Average firing rate			Readout firing rate		
	BayesNet-K2 (%)	J48 (%)	SVM* (%)	BayesNet-K2 (%)	J48 (%)	SVM* (%)
cc	48.5	48.5	49.4	50.6	50.6	50.0
inh	87.0	87.0	89.1	89.7	89.7	90.9
inh + cc	87.0	86.7	88.8	89.7	89.1	90.1
inh + 2ie	89.7	91.5	88.8	91.2	91.2	91.5
inh + 3m	87.0	85.8	88.8	89.7	88.5	91.5
inh + 3ie	86.7	90.1	93.3	92.1	91.5	93.4
inh + source values	92.7	94.8	94.2	94.5	93.0	95.8
inh + sink values	93.0	93.0	94.5	91.8	95.2	95.8
inh + source + sink values	92.1	93.0	94.8	92.7	93.6	95.5
Source values	92.4	93.0	93.3	92.4	92.1	93.0
Sink values	90.9	92.4	92.1	92.1	93.0	93.0
Source + sink values	92.1	93.3	93.6	92.4	93.0	94.2

Note that the presented SVM results are the better ones from both types of kernels. cc: clustering coefficient, inh: inhibitory neuron count in the network, 2ie/3ie: second/third order motifs considering signs, 3m: third order motifs without considering signs.

activity. We conclude that link mining methods can be successfully employed for the discovery of structure-activity and structure-function relations. Building on our experiences with simulated activity data, we currently adapt our algorithms to discover structure-activity relations in biological neuronal networks, like cell cultures grown on multi-electrode arrays [9].

Acknowledgments

References

[1] M. Barahona, L.M. Pecora, Synchronization in Small-World Systems, Phys. Rev. Lett. 89 (2002) 054101.

[2] V. Braitenberg, A. Schüz, Cortex: Statistics and Geometry of Neuronal Connectivity, Springer, Berlin, 1998.

[3] N. Brunel, Dynamics of sparsely connected networks of excitatory and inhibitory spiking neurons, J. Comput. Neurosci. 8 (2000) 183–208.

[4] N. Cristianini, J. Shawe-Taylor, An Introduction to Support Vector Machines, Cambridge University Press, Cambridge, 2000.

[5] M. Diesmann, M-O. Gewaltig, NEST: An Environment for Neural Systems Simulations, Göttingen: Ges. für wiss. Datenverarbeitung, Forschung und wiss. Rechnen, Beiträge zum Heinz-Billing-Preis 2001 58 (2002) 43–70.

[6] L. Getoor, Link mining: a new data mining challenge, SIGKDD Explorations 5 (2003) 1–6.

[7] L. Getoor, C. Diehl, Link mining: a survey, SIGKDD Explorations 7 (2005) 3–12.

[8] J. Kleinberg, Authoritative sources in a hyperlinked environment, in: Proceedings of the Ninth ACM-SIAM Symposium on Discrete Algorithms, San Francisco, California, 1998, pp. 668–677.

[9] S. Marom, G. Shahaf, Development, Learning and memory in large random networks of cortical neurons: lessons beyond anatomy, Q. Rev. Biophys. 35 (1) (2002) 63–87.

[10] A. Morrison, C. Mehring, T. Geisel, A. Aertsen, M. Diesmann, Advancing the boundaries of high connectivity network simulation with distributed computing, Neural Comput. 17 (2005) 1776–1801.

[11] L. Page, S. Brin, R. Motwani, T. Winograd, The PageRank Citation Ranking: Bringing Order to the Web, Technical report, Stanford Digital Library Technologies Project, 1998.

[12] R.J. Prill, P.A. Iglesias, A. Levchenko, Dynamic properties of network motifs contribute to biological network organization, PLoS Biol. 3 (2005) e343.

[13] S. Song, P.J. Sjöström, M. Reigl, S. Nelson, D.B. Chklovskii, Highly nonrandom features of synaptic connectivity in local cortical circuits, PLoS Biol. 3 (2005) e68.

[14] D.J. Watts, S.H. Strogatz, Collective dynamics of 'small-world' networks, Nature 393 (1998) 440–442.

[15] I.H. Witten, E. Frank, Data Mining: Practical Machine Learning Tools and Techniques, Morgan Kaufmann, San Francisco, 2005.

Tayfun Gürel was born in 1980 in Turkey. He obtained a B.Sc. degree in Electrical and Electronics Engineering (Middle East Technical University, Ankara, Turkey) and a M.Sc. degree in Computer Science (Albert-Ludwigs University Freiburg, Germany). Currently he is a Ph.D. student in the Machine Learning Lab at Albert-Ludwigs University Freiburg and at Bernstein Center for Computational Neuroscience, Freiburg. His research interests are computational neuroscience, statistical-relational machine learning and data mining, with a focus on applications of machine learning algorithms to structure-function relations of biological neural networks.

Luc De Raedt received his Ph.D. in Computer Science (Dr. Informatica) from the Katholieke Universiteit Leuven in Belgium in 1991, where he held positions as teaching assistant (1986–1991), post-doctoral researcher (1991–1999), part-time assistant professor (1993–1998) and part-time associate professor (1998–1999). Since 1999, he is a Professor at University of Freiburg and head of the Machine Learning and Natural Language Processing Lab research group. Luc De Raedt's research interests are in Inductive Logic Programming, Machine Learning and Data Mining, as well as their applications. In 2005, he was elected as an ECCAI fellow. Since 1 October 2006, Luc De Raedt has taken up a research professorship at the Department of Computer Science of the Katholieke Universiteit Leuven (Belgium).

Stefan Rotter was born in 1961 in Germany, where he obtained his M.Sc. in Mathematics (Universities of Regensburg and Hamburg, Brandeis University, Boston, USA) and Ph.D. in Physics (University of Tübingen). After associations with the Max-Planck-Institutes for Biological Cybernetics and for Developmental Biology in Tübingen, he was Assistant Professor at the Albert-Ludwigs-University Freiburg (Germany), where he also received his habilitation for Neurobiology and Biophysics. Currently, he is at the Institute for Frontier Areas of Psychology and Mental Health, Freiburg, and at the Bernstein Center for Computational Neuroscience, Freiburg. His research interests are in the field of theoretical and computational neuroscience, with a focus on analysis and modeling of anatomical structures and physiological processes in biological neural networks.