# Inferring Relevance from Eye Movements Using Generic Neural Microcircuits

**Tuomas Lepola**
Department of Computer Science, University of Helsinki
P.O. Box 68, FIN-00014, Finland
Tuomas.Lepola@cs.helsinki.fi

## Abstract

In the ”Inferring Relevance from Eye Movements Challenge 2005”, contestants were asked to apply machine learning techniques for predicting sentence relevancies based on eye movements of the readers. The two-part competition consisted of a classification problem and a time series analysis problem. In our winning solution to the time series analysis problem, we applied generic neural microcircuit as a nonlinear operator on time series with discriminative classifier as a readout. The results suggest that the model used has some practical merits as a generic time series analysis tool.

## 1 Introduction

The ”Inferring Relevance from Eye Movements Challenge 2005” [1] was organized in the form of a two-part data analysis competition. The contestant were given two competition data sets collected from the experimental setting where subjects were asked to identify correct answers to each assignment presented. Each of these assignments consisted of a question and 10 sentences of which five were irrelevant to the question, and of the remaining five relevant sentences one was the correct answer. In the test setting, the sentences were presented on a computer display and the eye movements and the pupil diameters of the subjects were recorded and stored as a form of time series. The contestants were asked to identify, on the basis of the eye movements, which sentences of each assignment were irrelevant, relevant, and correct. In Competition 1 the time series data were preprocessed into a form of traditional classification data. This preprocessing method segments the time series data in a manner typical to psychological research of reading. In contrast to this, in Competition 2, only the raw time series data was presented to contestants.

There were three main difficulties in the competition setting. Firstly, according to the description of the test setup, ”the subject was instructed to identify the correct answer” [1]. In particular, the subjects were not instructed to read through all sentences. In this sense, there was a slight discrepancy between the experimental setting and the challenge. This discrepancy creates the difficulty of how to classify unseen sentences, since many subjects read efficiently and completely skipped the remaining sentences after finding the correct answer. These unseen sentences were ignored while computing the prediction accuracy in Competition 1. However, in Competition 2 all sentences affected the accuracy score. Secondly, the irrelevant sentences may have contained some words which drew the attention of the subject in spite of the sentence not being relevant in the assignment. Hence the baseline methods applied by the organizers indicated that the separation of irrelevant and relevant sentences was the most difficult part of the prediction task. Thirdly, the eye-movement data is inherently noisy which in general makes data analysis always more challenging.

The winners of both competitions were selected based on prediction accuracy for the test sets. In the following, we present the winning solution in Competition 2: the model, the implementation and the results, and our conclusions based on the results.

## 2 The Model

The generic neural microcircuit model has been introduced recently as a realistic model of cortical columns (see [2] for a comprehensive introduction). The model encapsulates the generic and the stereotypical characteristics of cortical columns in a useful theoretical and practical framework. The generic neural microcircuit implements a nonlinear operator $L^M$ which transforms the input time series to the dynamic state $x^M(t)$ of the circuit at time $t$. As detailed in [2], the essential property of this operator is the pointwise separation property which informally means that different input signals lead to robustly separated dynamic states. The dynamic state $x^M(t)$ of the circuit is defined as the output of all the neurons in the circuit at time $t$. The readout of the circuit implements a memoryless map of the circuit state to the output time series. Putting all this together, we get the output of the circuit, given input $x(\cdot)$ at time $t$, to be

$$y(t) = f^M((L^M x)(t)),$$

where $f^M$ is the readout map. It has been shown in [2] that if the readout map $f^M$ has a certain approximation property, and assuming the pointwise separation property for $L^M$, then the whole circuit has "universal power for computation with fading memory on functions of time". More specifically, the proof technique using the Stone-Weierstrass theorem also applied in [3] shows that any time-invariant operator having fading memory can be approximated arbitrary closely by this circuit. Thus this model has the necessary flexibility to many time series processing applications. Our experiments with the eye movement time series data supports also the practical applicability of the model.

We implemented a generic neural microcircuit as a three-dimensional lattice of integrate-and-fire type spiking neurons as in [2]. Each circuit comprises a neural column of 135 neurons (dimensions 3 x 3 x 15). The connectivity structure of the network is highly recurrent and it is governed by a distribution which, roughly, renders connections (excitatory or inhibitatory) more likely between neurons close by than neurons far apart. The synaptic connections are using dynamical synapses (see details of the model further on). As shown in [2], the separation property is enhanced by combining parallel columns as a one circuit. Hence as whole we applied four parallel circuits in Competition 2 instead of the one used in Competition 1.

The neural column was constructed as follows. The membrane potential $u_i$ of each neuron $i$ is given by the standard integrate-and-fire model with the membrane time constant $\tau_m$ and the total membrane resistance $R$,

$$\tau_m \frac{du_i}{dt} = -u_i(t) + R \sum_j w_{ij}(t) \sum_{f \in \mathcal{F}_j} \alpha_j(t - t_j^{(f)}), \tag{1}$$

where the postsynaptic input current was modeled using

$$\alpha_j(s) = \frac{s - \Delta_j^{ax}}{\tau_s^2} \exp(-\frac{s - \Delta_j^{ax}}{\tau_s}) \Theta(s - \Delta_j^{ax}).$$

The parameter $\Delta_j^{ax}$ specifies axonal transmission delay, $\tau_s$ provides synaptic time constant, and $\Theta$ is the Heaviside function (see [4] for details). Additionally, the model was augmented by the standard firing threshold rule and the absolute refractory period. The threshold criterion specifies the set of spikes $\mathcal{F}_j$ and the firing times $t^{(f)}$ of Equation (1).

The synaptic efficacy function $w_{ij}$ was modeled according to the phenomenological model of frequency-dependent synaptic dynamics given in [5]. This model formulates synaptic facilitation and depression as a function of the absolute synaptic efficacy and the fraction of available and unavailable synaptic efficacy. See [5] for details of this intricate model. The parameter values of the whole model were selected for the computer simulation as in Appendix B of [2]. These parameter values could be argued to be biologically reasonable and the whole model is characterized by reasonable biological realism with the eye on feasible computer simulation (see Section 3 for discussion on computational issues).

The dynamic state of the circuit was read out as a vector of spike trains from all neurons of the circuit. Furthermore, this vector was transformed to time varying output currents with the effect of each spike upon current decaying exponentially, and 20 ms time-window was applied to discretize this output signal. Hence, we obtained for each input signal and for each column a 135-dimensional discrete output time series.

It should be emphasized that no learning is involved in applying the generic neural microcircuit. The circuit could be the same for each time series given as an input. Only the readout map is selected according to the task, for example, as a linear classifier. This means that the learning task is easy compared to e.g. adjusting the parameters of a nonlinear recurrent neural network with supervised learning. In terms of computational power of the circuit, the readout map only needs to posses some weak capabilities mentioned above. Moreover, the readout can be memoryless, that is, each discrete sample of the circuit output could be classified independently of the past of the series. Hence the circuit provides a kind of natural preprocessing of a time series data to a classification data.

However, instead of using a linear perceptron network as a readout map like in [2], we applied discriminative classification to the circuit output. More specifically, our goal was to directly estimate the parameters of a distribution $P(Y|X)$ where $Y$ is a binary variable specifying if the gaze of the subject is on relevant sentence or not given the output vector $X$ (correct answers were considered as relevant, see the reasoning given in the next section). We assumed that $P(Y|X)$ could be learned reasonably well using logistic regression and proceeded to maximize the conditional data log likelihood of the weight vector $W$:

$$l(W) = -\sum_t (Y^t(w_0 + \sum_i w_i X_i^t) - \ln(1 + \exp(w_0 + \sum_i w_i X_i^t))),$$

where the superscript $t$ denotes the variable at the time step $t$. We applied the standard conjugate gradient descent to optimize the weight vector. Due to the simplicity of the optimization task, the convergence to the global maximum is guaranteed. However, to avoid overfitting, we applied weight regularization to penalize weights too large and furthermore, we proceeded with data specific cross-validation and used the validation set (see next section) to ensure the generalization capability of our model.

Finally, we applied a Bayesian approach to compute, given the estimated distribution $P(Y|X)$, the probability that a sentence in an assignment is relevant. Due to nature of the Competition 2 data, more information was available concerning the eye movements of the subjects between fixations than in the Competition 1 data. The subjects may have visited some relevant sentences very briefly, and especially they have not fixated on some relevant sentences. Hence very little data was available of some sentences compared to others. We applied parameter smoothing (as in the $m$-estimate of probability, see e.g. [6]) with uniform prior to augment this weakness.

The reasoning behind the use of logistic regression as a readout processor is as follows. Firstly, by nature of the generic neural microcircuit, the nonlinear transformation removes the necessity of the readout to have memory. Hence we can classify the data as independent samples from some unknown distribution. All information is available in the dynamic state of the circuit at some specified moment of time to approximate the output signal at that time. The key issue is to assess whether this approximation can be learned easily and reasonably well, and it was assumed that logistic regression performs in this respect at least as well as the linear parallel perceptron applied in [2]. Secondly, informal testing with an advanced naive Bayes classifier B-Course (available as an online service, see [7] for details) seemed to indicate that the performance of the naive Bayes classifier was seriously hurt by the obvious violation of the independence assumption of the variables. Although, more generally, the gaussian naive Bayes model and logistic regression are intimately related, logistic regression is not as tightly constrained by the conditional independence assumption as the gaussian naive Bayes model. It seems that our classification problem, where data is in abundance, is an example of a case where logistic regression outperforms the gaussian naive Bayes model [8]. Due to time limitations of the challenge, we did not analyze this observation further.

## 3 Implementation and Results

The raw challenge data of Competition 2, consisting of the original horizontal and vertical eye movements and pupil diameter data, was preprocessed to six time series. All time series were normalized and injected as an input current to randomly selected neurons of the circuit. We applied very simple transformation in the spirit of feature extraction described in [1]. At the level of words and sentences we recorded cumulative visits and revisits, local movements inside a word relative to optimal viewing position [9], the relative movements between the words inside a sentence, and finally, the relative movements between the sentences of an assignment. Additionally, we registered the pupil diameter changes. Since some pupil diameter readings were clearly anomalously large, we set all pupil diameter readings above 6 mm to the maximum of 6 mm.

The simulation of the generic neural microcircuit involved numerical integration. We applied the standard Runge-Kutta-Fehlberg method (4th order, 5th order error estimate) with adaptive stepsize and a modification to detect firing condition to simulate the dynamics of the generic neural microcircuit model. All simulation software were implemented using C++ programming language in the Linux computing environment. Since we were able to distribute the computation of separate columns and readouts to different computing nodes and processors, this greatly enhanced the throughput of the data processing. Typically a two-way Intel Xeon processor node with hyperthreading support provided throughput as high as 900 spikes per second and a two-way 64-bit AMD Opteron node even higher. Most of the preprocessing and postprocessing of the data and the results were run in the Matlab environment.

Since our method is clearly oriented to time series processing and we had very limited time for the competition, our emphasis was on Competition 2. However, a modest attempt was also made in Competition 1 since we realized that we could estimate the original time series from the classification data to some degree of accuracy. Obviously, in the training setting we could have used the Competition 2 training set to train our model, but this would have created a problem with the test set since the Competition 2 test set was made available after Competition 1 was closed. Instead we estimated the original time series based on only the Competition 1 data with three main estimation methods. Firstly, we estimated the word lengths of each assignment by averaging and using the knowledge of first and last fixations to the word, and the optimal viewing position. Secondly, we added gaussian noise to fixation positions to simulate the measurement errors since the average accuracy of equipment was given in [1]. Thirdly, we interpolated the eye movements between fixations using the facts known about eye movement speed function and the dimensions of the display and positioning of the subject in front of the display [1].

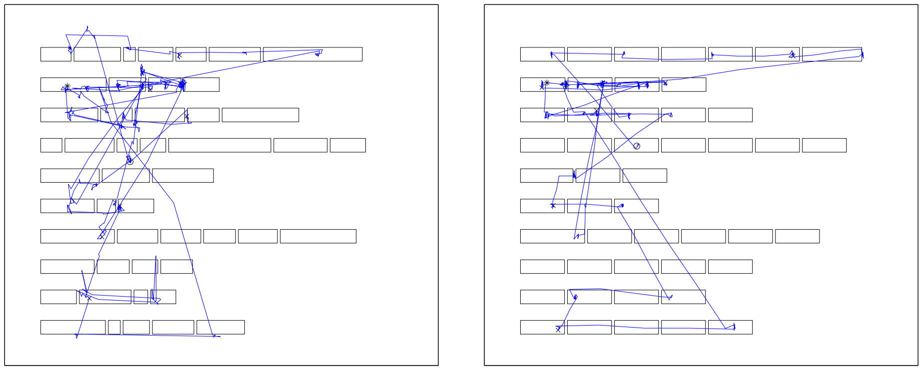

Figure 1: The actual eye trajectory and an estimation

Figure 1 shows one assignment from the training data. The actual eye movements are shown on the left and an estimation based on classification data is given on the right. As can be seen, qualitatively the eye movement trajectories look quite the same. However, some important trajectories are missing in the long gaps between some fixations. These short visits seemed to be crucial in separating some relevant and irrelevant sentences.

In Competition 1 we preprocessed the estimated time series as described above. The resulting six time series were given as an input to a single column of the generic neural microcircuit. The parameters of the readout map were estimated as specified in the previous section using the training data. After training with leave-one-assignment-out and leave-one-subject-out cross-validation, the model gave $60.9\%$ accuracy with the validation data. Since the test data turned out to give the accuracy of $60.7\%$, we can conclude that this simplified method performed robustly, although poorly compared to other competitors. We argue that this is mainly due to methodological discrepancy between the time series oriented model and the classification data.

In Competition 2 we applied the full model to predict the relevant sentences. After the time series preprocessing performed as above, we used the four-column generic neural microcircuit model to predict the probabilities of the sentences being relevant. The training with conjugate gradient descent optimization and cross-validation was essentially the same as described above. The most relevant sentence of the assignment was chosen to be the correct answer. If, instead of binary classification, we did three class classification (correct, relevant, irrelevant) we noticed a degrading performance in separating the relevant sentences from irrelevant ones. We concluded that we should concentrate on solving this harder separation problem, since this was in the spirit of the challenge. The test results proved that this approach was successful and overfitting of the noisy data was avoided.

Table 1: Confusion matrices of the Competition 2 validation and test sets

| Validation | $I(68.3\%)$ | $R(56.7\%)$ | $C(75.2\%)$ |
|---|---|---|---|
| $I(68.3\%)$ | 509 | 226 | 10 |
| $R(56.7\%)$ | 231 | 338 | 27 |
| $C(75.2\%)$ | 5 | 32 | 112 |

| Test | $I(68.7\%)$ | $R(63.9\%)$ | $C(77.8\%)$ |
|---|---|---|---|
| $I(68.7\%)$ | 618 | 276 | 6 |
| $R(56.4\%)$ | 282 | 406 | 32 |
| $C(78.9\%)$ | 0 | 38 | 142 |

Table 1 summarizes the results of the method we applied on the Competition 2 validation and test data sets. As can be seen in the confusion matrices, the strength of the method was exactly in separating the irrelevant and relevant sentences. The accuracies were $64.4\%$ on validation data and $64.8\%$ on test data. This indicates that the method is characterized by robust prediction performance, even in the presence of the unseen sentences.

## 4 Conclusions

Our solution in the eye movement data analysis challenge was based on generic neural microcircuit model with discriminative classification through logistic regression. Our solution won Competition 2 which involved predicting sentence relevancies using the raw time series data. We conclude that the neural microcircuit used exhibited practical genericity in time series analysis, since only very simple preprocessing was needed for the task at hand. We emphasize that the learning problem is easy with this model, since the microcircuit is completely generic and a linear classifier is sufficient to extract useful predictions from the model. Additionally, the test results indicate that the model is robust and does not overfit easily although the time series data was noisy. Furthermore, the results suggest that future work on at least three different areas would be interesting. Firstly, in the context of this modeling problem, it would be important to find methods for improving the classification of correct sentences without introducing overfitting. Secondly, this data analysis problem seems to be a case where logistic regression outperforms gaussian naive Bayes model and it might be of value to do more experimentation with the challenge data to compare the practical performances of the two models. Thirdly, the spatial structure of the generic neural microcircuit might benefit from improvements in biological realism.

### Acknowledgments

This work was supported in part by the IST Programme of the European Community, under the PASCAL Network of Excellence, IST-2002-506778. This publication only reflects the author's views.

### References

[1] Salojärvi, J., Puolamäki, K., Simola, J., Kovanen, L., Kojo, I. & Kaski, S. (2005) Inferring relevance from eye movements: Feature extraction. *Helsinki University of Technology, Publications in Computer and Information Science, Report A82*.

[2] Maass, W., Natschläger, T. & Markram, H. (2002) Real-time computing without stable states: A new framework for neural computation based on perturbations. *Neural Computation* **14**:2531-2560.

[3] Boyd, S. & Chua, L. O. (1985) Fading memory and the problem of approximating nonlinear operators with Volterra series. *IEEE Transactions on Circuits and Systems* **32**:1150-1161.

[4] Gestner, W. & Kistler, W. (2002) *Spiking Neuron Models.* Cambridge University Press.

[5] Markram, H., Wang, Y. & Tsodyks M. (1998) Differential signaling via the same axon of neocortical pyramidal neurons. *Proc. Natl. Acad. Sci.* **95**:5323-5328.

[6] Mitchell, T. M. (1997) *Machine Learning.* McGraw-Hill.

[7] Myllymäki, P., Silander, T., Tirri, H. & Uronen, P. (2002) B-Course: A web-based tool for Bayesian and causal data analysis. *International Journal on Artificial Intelligence Tools* **11**:369-387.

[8] Ng, A. Y. & Jordan, M. I. (2002) On discriminative vs. generative classifiers: A comparison of logistic regression and naive Bayes. In T. G. Dietterich, S. Becker and Z. Ghahramani (eds.), *Advances in Neural Information Processing Systems 14*, MIT Press.

[9] Rayner, K. (1998) Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin* **124**:372-422.