Replication / Machine Learning

[Re] Measures for investigating the contextual modulation of information transmission.

Sepehr Mahmoudian[1,2, ID]

[1]Deptartment of Data-driven Analysis of Biological Networks, Campus Institute for Dynamics of Biological Networks, Georg August University Göttingen, Göttingen, Germany – [2]MEG Unit, Brain Imaging Center, Goethe University Frankfurt, Frankfurt, Germany

1 Introduction

Background – Systems such as neurons can have different types of input. The authors[1] propose dividing the input to a neuron into receptive (or driving) and contextual. They further present formulations of how contextual connections can amplify driving input. Finally, they use information theory to study their attributes.

Motivation – This work was published in 1996, but with the advances in computing and information theory, and increased emphasis on the importance of context[2], new studies[3,4] have started building on its foundations. Furthermore, there are many ways to write the activation functions and only a few simple formulations were analyzed. Therefore, the significance of this work is twofold: 1. To reproduce the study in a modern programming language and make it publicly accessible. 2. Enable new activation functions to be readily tested.

Reproduction – A successful attempt is made to qualitatively and quantitatively reproduce the results of the original paper by reproducing the three figures that comprise the results.

2 Methods

Input – The study is limited to information transmission from a neuron which has only one receptive (R) or contextual (C) connection as input and produces one output (X). There is input at every time step, +1 denotes firing and -1 being silent, which is then multiplied by the weight of their connection. probability distributions for **r** and **c** were created in the same way as original authors as described below, where for all figures P(C = 1 | R = 1) was set to 0.889972.

$$P(R = +1, C = +1) = 0.5P(C = 1|R = 1) \tag{1}$$

$$P(R = +1, C = -1) = 0.5[1 - P(C = 1|R = 1)] \tag{2}$$

$$P(R = -1, C = +1) = 0.5[1 - P(C = 1|R = 1)] \tag{3}$$

[Re] Measures for investigating the contextual modulation of information transmission.

UNDER REVIEW

$$P(R = -1, C = -1) = 0.5P(C = 1|R = 1). \tag{4}$$

Activation Functions – Three activation functions are studied by the authors. Based on how receptive and contextual information are formulated to interact, they are called A_a (additive), A_m (modulatory), and A_b (both additive and modulatory).

$$A_a(r, c) = r + c \tag{5}$$

$$A_m(r, c) = 0.5r[1 + e^{rc}] \tag{6}$$

$$A_b(r, c) = 0.5r[1 + e^{rc}] + c \tag{7}$$

The output of each activation function is passed into the sigmoidal activation function to generate a probability of firing:

$$p(X = 1|r, c) = \frac{1}{1 + e^{-A(r,c)}} \tag{8}$$

Information Theory – They use the three-way mutual information as a measure of the information shared between the output and the inputs and write it as the equation below.

$$I(X; R; C) = \sum_{x,r,c} p(x, r, c) log \frac{P(x|r)P(x|c)}{P(x)P(x|r,c)} \tag{9}$$

I(X;R|C) is used by the authors as a measure of information that is shared between the R input and the X output but not the C input and is written as:

$$I(X; R|C) = \sum_{x,r,c} P(x, r, c) log \frac{P(x|r,c)}{P(x|c)} \tag{10}$$

I(X;C|R) is used by the authors as a measure of information that is shared between the C input and the X output but not the R input and is written as:

$$I(X; C|R) = \sum_{x,r,c} P(x, r, c) log \frac{P(x|r,c)}{P(x|r)} \tag{11}$$

Dependencies – The reproduction was done using Ubuntu 19.10 with Intel® Core™ i7-7500U CPU with Anaconda Python-3.8.1, numpy-1.18.1, scipy-1.4.1, and matplotlib-3.1.3. There are no external libraries required for calculating the information theoretic terms as everything is provided in the accompanying code.

3 Results

Activation functions and transmitted information – Figure 1 shows the three-way mutual information and conditional mutual information terms for each of the activation functions and across a range of R input weights while C remains constant at 1. These results are produced in a similar way to the paper by using the exact probability distributions. They are qualitatively and quantitatively identical to the original paper.
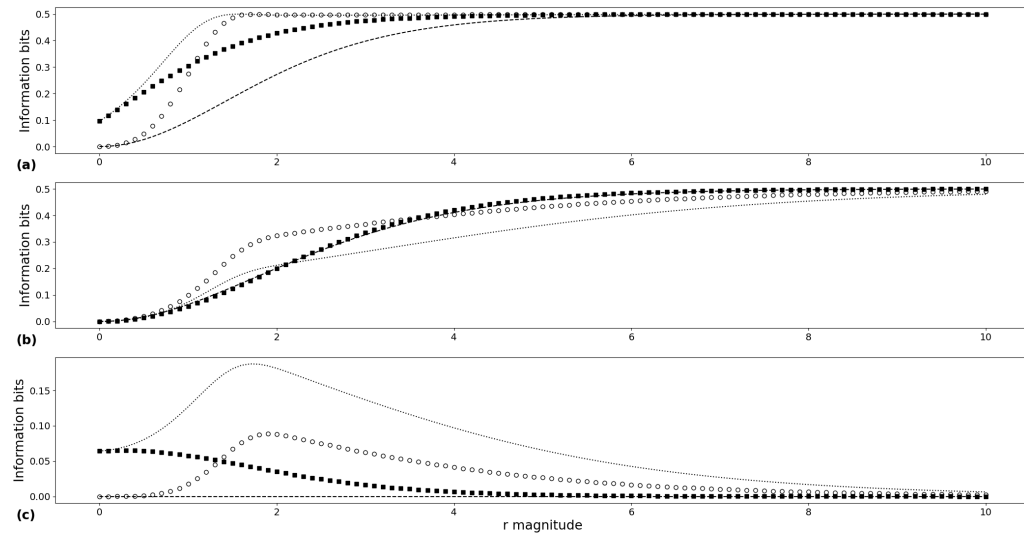
[Re] Measures for investigating the contextual modulation of information transmission.

UNDER REVIEW

Figure 1. Reproduction of the comparison of the activation functions for (a) I(X;R;C), (b) I(X;R|C) and (c) I(X;C|R). The activation functions are zero context (dashed), additive context (squares), modulatory context (circles), and both additive and modulatory combined (dotted).

Information surface plots for different input weights – Figure 2 shows the I(X;C|R) for each of the activation functions and across a range of r input weights while c remains constant at 1. These results are produced in a similar way to the paper by using the exact probability distributions. They are qualitatively and quantitatively identical to the original paper.
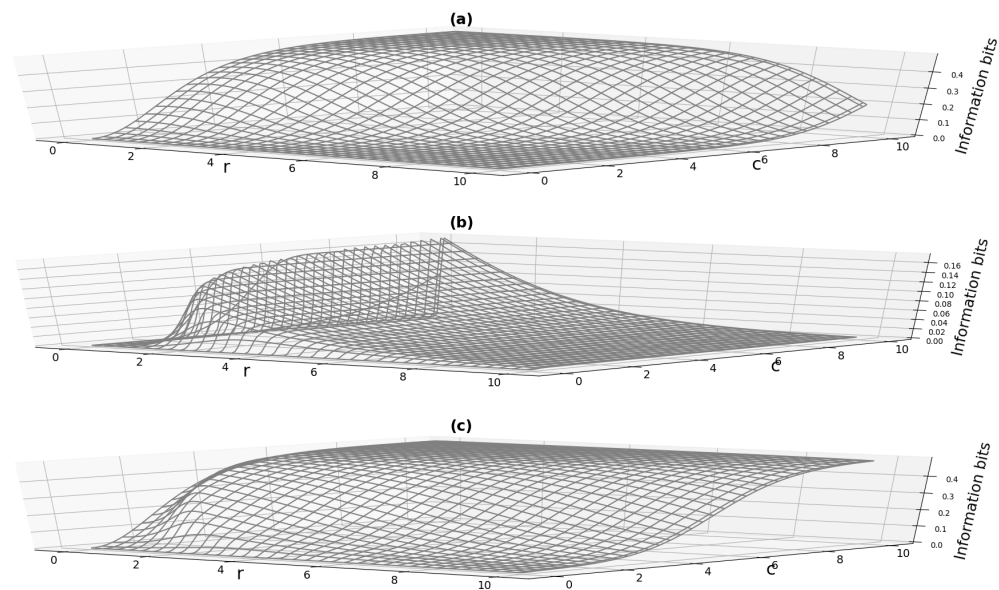

Figure 2. Reproduction of the comparison of the information surfaces of I(X;C|R) for (a) additive, (b) modulatory and (c) both over all R and C input weights. The conditional probability P(C = 1 |R = 1) = 0.889 972.

Sampling biases and variances – Figure 3 compares the mean and standard deviation of three-way measures sampled against the analytical measures for the modulatory activation function. The mean and deviations were computed from 100 trials per sample size and compared with the analytical measure across a range of sample sizes. As the

authors claim, sample size of 200 could be sufficient as the improvements beyond that are not much.
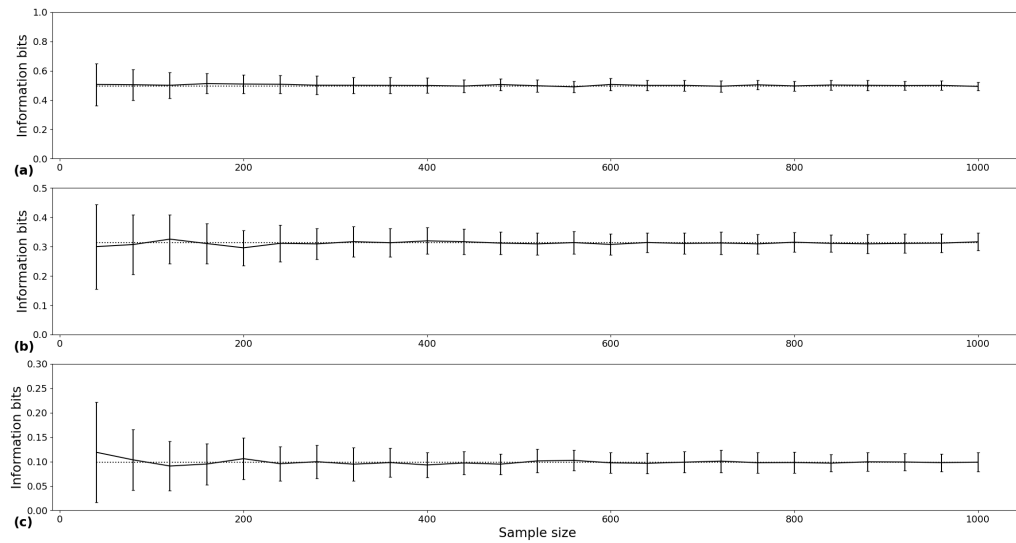

Figure 3. Reproduction of comparison of the analytical (dotted) and sampled (error bars) information measures for (a) I(X;R;C),(b) I(X;R|C) and (c) I(X;C|R). The error bars represent standard deviation about the mean information computed from 100 trials for each sample size. The R and C input weights were set to 2.

4 Discussion

The goal of the authors was to study contextual amplification in single binary neurons using information theory. In the same way, the same figures here can be used study the different activation functions and contextual amplification. Here are some examples:

If there is modulatory amplification as in (6), I(X;R;C) should increase more rapidly as a function of R. This is evident from Figure 1 (a) where the "zero context" function is the slowest to rise and the others differ in how they rise based on how they integrate R and C information.

Another example is that increasing C does not increase I(X;C|R) alone for the modulatory activation function as shown in Figure 2 (b). As expected, the role of R is to drive the neuron to fire and C is to amplify the firing. While increasing C alone has no effect on I(X;C|R), increasing R and C together does and a higher C with a constant R amplifies it. At low R values, C has a higher role in the output so I(X;C|R) begins to diminish at higher R values.

In conclusion, the main results of their work is successfully reproduced and the code is made available in a public repository which should enable other activation functions to be readily studied.

[Re] Measures for investigating the contextual modulation of information transmission.

UNDER REVIEW

References

1. D. Smyth, W. A. Phillips, and J. Kay. "Measures for investigating the contextual modulation of information transmission." In: **Network: Computation in Neural Systems** 7.2 (May 1996), pp. 307–316.

2. M. Larkum. "A cellular mechanism for cortical associations: An organizing principle for the cerebral cortex." In: **Trends in Neurosciences** 36.3 (2013), pp. 141–151.

3. J. Kay and W. A. Phillips. "Coherent Infomax as a Computational Goal for Neural Systems." In: **Bulletin of Mathematical Biology** 73.2 (2011), pp. 344–372.

4. M. Wibral, V. Priesemann, J. Kay, J. Lizier, and W. A. Phillips. "Partial information decomposition as a unified approach to the specification of neural goal functions." In: **Brain and Cognition** 112 (2017), pp. 25–38.