Peering towards Cosmic Dawn
Proceedings IAU Symposium No. 333, 2018
V. Jelić & T. van der Hulst, eds.

doi:10.1017/S1743921318000455

Foreground Subtraction and Signal reconstruction in redshifted 21cm Global Signal Experiments using Artificial Neural Networks

Madhurima Choudhury and Abhirup Datta

Indian Institute of Technology, Indore
Simrol, Khandwa Road, Indore, M.P., India
email: `abhirup.datta@iiti.ac.in`

Abstract. Observations of HI 21cm transition line is a promising probe into the Dark Ages and Epoch-of-Reionization. Detection of this redshifted 21cm signal is one of the key science goal for several upcoming low-frequency radio telescopes like HERA, SKA and DARE. Other global signal experiments include EDGES, LEDA, BIGHORNS, SCI-HI, SARAS. One of the major challenges for the detection of this signal is the accuracy of the foreground source removal. Several novel techniques have been explored already to remove bright foregrounds from both interferometric as well as total power experiments. Here, we present preliminary results from our investigation on application of ANN to detect 21cm global signal amidst bright galactic foreground. Following the formalism of representing the global 21cm signal by 'tanh' model, this study finds that the global 21cm signal parameters can be accurately determined even in the presence of bright foregrounds represented by 3rd order log-polynomial or higher.

Keywords. Keyword1, keyword2, keyword3, etc.

1. Introduction

The cosmological 21cm signal is going to be an excellent probe and could be a powerful diagnostic to test cosmological theories. The 21 cm line of neutral hydrogen presents a unique probe of the evolution of the neutral intergalactic medium (IGM), and cosmic reionization. There are many advantages of using the HI line for this purpose Furlanetto *et al.* (2006). We always try and draw analogies and point out the differences from some already known studies. In this case, a constant comparision is made with the CMB studies, wherever applicable. The HI line provides rich information of the evolution of cosmic structure. The interplay between the CMB temperature, the kinetic temperature and the spin temperature, along with radiative transfer, lead to very interesting physics of the 21cm signal. The HI 21cm observations can be used to study evolution of cosmic structure from the linear regime at high redshift (i.e., density-only evolution), through the non-linear regime associated with luminous source formation Barkana and Loeb (2005).

The 21cm H1 line: The hyperfine transition line of atomic hydrogen (in the ground state) arises due to the interaction between the electron and proton spins. The spin temperature, T_S, primarily determines the intensity of the 21cm radiation. In radio astronomy, the measured quantity is the brightness temperature or more accurately called the differential brightness temperature $\delta T_b \equiv T_b - T_\gamma$. Substituting the various quantities into the equation for brightness temperature and rearranging, we obtain,

$$\delta T_b \approx 27(1 - x_i)\left(\frac{\Omega_{b,0}h^2}{0.023}\right)\left(\frac{0.15}{\Omega_{m,0}h^2}\frac{1+z}{10}\right)^{1/2}\left(1 - \frac{T_\gamma}{T_S}\right) \tag{1.1}$$

This equation is used to estimate the global 'all sky-averaged' redshifted 21cm signal from Cosmic Dawn and Epoch of Reionization.

Foregrounds: There are two major challenges faced while trying to detect the global signal. One involves foregrounds and the other involves calibration. Radio emission from our galaxy as well as terrestrial radio emission dominate the feeble signal we are trying to detect. This foreground includes galactic and extragalatic sources. The expected signal is about 10^4 times weaker than the foreground emission.

For global experiments, these foregrounds are well represented by a polynomial Harker (2015) in $ln(T)$-$ln(\nu)$.

$$ln\ T_{FG} = \sum_{i=0}^{n} a_i[ln(\nu/\nu_0)] \tag{1.2}$$

Here, ν_0 is a pivot scale and a_0 is recast as $a_0 = log\ T_0$ to emphasise that the zeroeth order coefficient has units of temperature. It has been shown that a polynomial of atleast of order 3 is necessary to remove the foreground.Pritchard *et al.* (2015)

2. Overview - Artificial Neural Network

Artificial neural networks(or ANNs) are machine learning algorithms which mimic the functioning of the human brain. The basic building blocks of the neural nets are the neurons. The most simple ANN usually consists of a multi-layer network comprising of the input layer, the hidden layer and the output layer. The number of neurons in the hidden layers and the number of hidden layers can be both chosen as per requirement-initially arbitrarily, but then deciding upon the final structure after validating the architecture of the ANN.

The basic neural network model is described as a series of functional transformations. Let us consider that there are D inputs $(x_1, x_2, ..., x_D)$. Each neuron in the input layer is connected to the next layer, that is the hidden layer and a weight and a bias (both are initially randomly chosen) is associated with the connection. The input to a single neuron in the hidden layer is a linear combination of all the input neurons with their respective assigned weights and biases, and is written as:

$$a_j = \sum_{i=1}^{D} w_{ji}^{(1)}\ x_i + w_{j0}^{(1)} \tag{2.1}$$

where, $w_{ji}\ 's$ are the weights and $w_{j0}\ 's$ are the biases associated with each input. The superscript (1) denotes that these corresponding parameters are in the first layer of the network. These $a_j's$, are known as activation. Each of the activations is transformed using a non-linear activation function $h()$, such that, $z_j = h(a_j)$. These quantities $z_j's$ correspond to the outputs of the neurons in the hidden layer. These values $(z_j's)$ are again linearly combined along with their weights and biases to give the inputs to each neuron in the output layer, also called the output unit activations, $a_k's$, given by:

$$a_k = \sum_{j=1}^{M} w_{kj}^{(2)}\ z_j + w_{k0}^{(2)} \tag{2.2}$$

here, the superscript (2) corresponds to the second layer of the network. Finally the output unit activations are again transformed using an appropriate activation function, to give a a set of network outputs y_k.These $y_k' s$ are the end outputs of the feed-forward process, where we had given a input vector X and obtained an output vector Y. Now,the outputs are compared with the inputs with which the network was fed. The purpose of training the ANN is to find out the set of weights that ensures that the output produced by the network is sufficiently close to the desired output values. Thus the weights needs to be appropriately adjusted to obtain the desired outputs. This adjustment of weights is done iteratively, by computing an error function or a cost function and minimizing it. The error function is usually chosen to be the standard mean squared error function, which is written as:

$$\sqrt{\frac{1}{N_{train}} \sum^{N} \left(\frac{y_{pred} - y_{ori}}{y_{ori}}\right)^2} \tag{2.3}$$

There are two more important steps: the parameter optimization and the other is the back-propagation step which heavily influence the success of the network output. Interested readers are encouraged to look into the references given in Shimabukuro and Semelin (2017)

3. Application of ANN to extract Global 21cm Signal

Here, we have used ANN to extract faint 21cm global signal in presence of significantly dominant foreground. The training process for an ANN is an iterative procedure that begins by collecting the data and preprocessing it to make training more efficient. In this type of network, the information flow is unidirectional that takes place with the help of an activation function between each layer.

First, we create a set of models for 21cm global signal using *ARES* (Mirocha *et al.* 2015) within frequency range of 20-160 MHz or redshift range of 8 to 68. The foreground model is created using equation 1.2 and added to the signal models. We have not included any instrumental white noise in the training datasets so that we can test the efficiency of the network by introducing increasing noise in the test dataset. The network then uses back-propagation as the learning technique where the output values are compared with the input values to compute the value of an error-function. The root mean-squared error (RMSE) is used as the predefined error-function here. The error is then fed back through the network. Using this information, the algorithm adjusts the weights of each connection in order to reduce the value of the error function iteratively. After sufficient iterations, the training process usually converges such that the network has learned a certain target function. To adjust weights properly, we use the *'adam'* (Kingma and Ba (2014)) optimizer for non-linear optimization. In this work, the data was normalized using min-max normalization and was split in 7:3 ratio in training and testing sets out of which the training set was further split into 7:3 ratio in actual training set and validation set. As the test data, we supplied a data-set comprised of a different signal model along with the foreground and relevant instrumental noise.

Here, we have used a network with a 1024 neurons in the input layer corresponding to the number of frequency channels we plan to work with. It is a feed-forward network, which means each neuron in a particular layer is connected to all other neurons in the next layer (along one particular direction). Next in the hidden layer we have chosen 14 neurons activated by the sigmoid function. Finally,in the output layer. we have 11 output neurons corresponding to each parameter we are interested to find out.

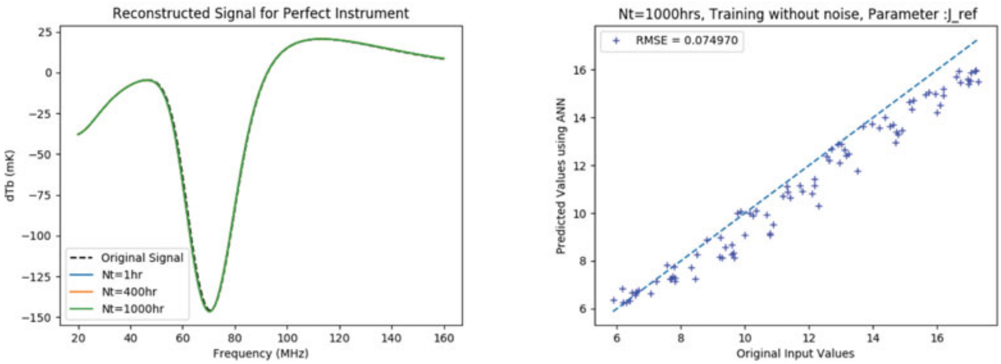

Figure 1. (Left) Quality of the reconstructed global 21cm signal with the change in integration time.(Right) RMSE varying with different test datasets. This confirms the robustness of the prediction for one of the Signal parameters (J_{ref}, Mirocha *et al.* (2015))

4. Implications

For the simplest case, we just take the 21cm signal along with the model foreground (assuming perfect instrument). The training datasets are constructed using the following relation: $T_{train}(mK) = (T_{21} + T_{FG})$. The network is trained using this data, and fitted for the 7 signal parameters and 4 foreground parameters. Next, we create a test dataset, which include thermal noise (depends on the observing time) for the instrument in addition to the components in T_{train}. Initial results are presented in figure 1.

Our preliminary results looks encouraging to continue our work with the ANNs in extracting global 21cm signal. We will extend this work for different instrumental response scenario and more realistic observing strategies. The final results of this work are expected to be reported in a refereed journal publication in early 2018. In future, we plan to extend this work for power spectrum estimation of 21cm fluctuations measurements with HERA and SKA.

References

Furlanetto, S. R., Oh, S. P., & Briggs, F. H. Cosmology at low frequencies: The 21 cm transition and the high-redshift Universe. *Phys. Rep.*, 433: 181–301, October 2006.

Barkana, R. & Loeb, A. A Method for Separating the Physics from the Astrophysics of High-Redshift 21 Centimeter Fluctuations. *ApJL*, 624: L65–L68, May 2005.

Harker, G. J. A. Selection between foreground models for global 21-cm experiments. *MNRAS*, 449: L21–L25, April 2015.

Pritchard, J., Ichiki, K., Mesinger, A., Metcalf, R. B., Pourtsidou, A., Santos, M., Abdalla, F. B., Chang, T. C., Chen, X., Weller, J., & Zaroubi, S. Cosmology from EoR/Cosmic Dawn with the SKA. *Advancing Astrophysics with the Square Kilometre Array (AASKA14)*, art. 12, April 2015.

Shimabukuro, H. & Semelin, B. Analysing the 21 cm signal from the epoch of reionization with artificial neural networks. *MNRAS*, 468: 3869–3877, July 2017.

Mirocha, J., Harker, G. J. A., & Burns, J. O. Interpreting the Global 21-cm Signal from High Redshifts. II. Parameter Estimation for Models of Galaxy Formation. *ApJ*, 813: 11, November 2015.

Kingma, D. P. & Ba, J. Adam: A Method for Stochastic Optimization. *ArXiv e-prints*, December 2014.