Recovering HII bubble size distribution with artificial neural network

Hayato Shimabukuro (Tsinghua University) with Yi Mao
Dark Ages

No luminous object exists. (z>~30?)

Cosmic Dawn

First stars and galaxies form. (z~20-30?)

Reionization

UV photons by luminous objects ionize IGM. (z~6-15?)
**History of the Universe**

- **Dark Ages**
  - No luminous object exists. ($z > ~30$?)

- **Cosmic Dawn**
  - First stars and galaxies form. ($z ~ 20-30$?)

- **Reionization**
  - UV photons by luminous objects ionize IGM. ($z ~ 6-15$?)

---

[https://universe-review.ca/](https://universe-review.ca/)
21cm line radiation:

Neutral hydrogen emits the radiation due to the hyperfine structure.

Brightness temperature

\[
\delta T_b = \frac{T_S - T_\gamma}{1 + z} (1 - \exp(-\tau_\nu)) \\
\sim 27 x_H (1 + \delta_m) \left( \frac{H}{d\nu_r/dr + H} \right) \left( 1 - \frac{T_\gamma}{T_S} \right) \left( \frac{1 + z}{10} \frac{0.15}{\Omega_m h^2} \right)^{1/2} \left( \frac{\Omega_b h^2}{0.023} \right) \text{[mK]}
\]

Including both cosmological and astrophysical information
21cm power spectrum

We first aim to detect 21cm signal **statistically**.

21cm power spectrum (PS) : \( \langle \delta T_b(k) \delta T_b(k') \rangle = (2\pi)^3 \delta(k + k') P_{21} \)

(We use 21cmFAST)

Pober et al (2014)
Distinguish EoR models

We often discuss distinguishing EoR models on the basis of 21cm power spectrum. (e.g. EoR parameters, feedback, ionising sources)

Dixon et al. (2016)

LB1: only high mass halo $M > 10^9 \, M_{\odot}$
LB2: including low mass ($10^8-10^9 \, M_{\odot}$)
LB3: including low mass with radiative feedback
LB4: same with LB3 but mass dependent FB
21cm PS analysis is useful, but it is imperfect to describe 21cm fluctuations.

21cm PDF is highly non-Gaussian owing to astrophysical effects.

Higher order statistics!
- Bispectrum
  - HS+ 2016, Watkinson+ 2017, Majumdar+ 2018
- Skewness
  - HS+2015, Kubota + 2016
Morphology of ionised bubble

Kulkarni et al 2017
Bubble size distribution (BSD)

HII bubble size distribution is another method to break degeneracy in EoR models.

**However!!**

Difficult to determine bubble size and measure BSD.

Some methods are suggested to measure BSD from ionised map

- **Mean free path**
  - Mesinger & Furlanetto (2007)

- **Spherical average method**
  - Zahn+ (2007)

- **Friend-of-Friend (FoF)**
  - Friedrich+ (2011)

- **Watershed**
  - Lin+ (2016)
Comparison of methods

Yin+ (2016)

See solid line

Black: Reference
Green: Watershed (WS)
Blue: Mean free path (MFP)
Red: Spherical average (SA)

WS and MFP seem better
BSD from 21cm tomography

But, we practically measure BSD from 21cm tomography data. Different from the case of measuring BSD from ionisation map.

- ionised map
  - binary map (ionized or neutral)
- 21cm map
  - Not binary map

Therefore, we need to transform 21cm map into binary map (ionized or neutral).

We need to fix threshold to distinguish fields.

It is difficult because 21cm map depends on both $x, \delta$
K-means method

Giri+ 2017

Giri+ 2017 focuses on the fact that 21cm PDF is bimodal.

• K-means method
  unsupervised learning algorithm for clustering problem
Giri+ 2017 focuses on the fact that 21cm PDF is bimodal.
K-means method

Giri+ 2017 focuses on the fact that 21cm PDF is bimodal.
K-means method

Giri+ 2017

• K-means method is an unsupervised learning algorithm for clustering problems.

Giri+ 2017 focuses on the fact that 21cm PDF is bimodal.
K-means method

Giri+ 2017 focuses on the fact that 21cm PDF is bimodal.
K-means method

Giri+ 2017 focuses on the fact that 21cm PDF is bimodal.
K-means method

Giri+ 2017 focuses on the fact that 21cm PDF is bimodal.
Giri+ 2017 focuses on the fact that 21cm PDF is bimodal.

**K-means method**
unsupervised learning algorithm for clustering problem
BSD for different sources

Giri et al 2017

MFP-BSD

No smoothing

smoothing
Granulometry

Granulometry?  
Mathematical morphology = “Algebra of shapes”

→ Concept of sieving (Matheron 1975)

remaining

→ counts the objects of given sizes
Granulometry

Granulometry?

Mathematical morphology = “Algebra of shapes”

Granulometry aims to recover the underlying true size distribution in the unbiased way!!

→ counts the objects of given sizes
Our datasets consist of 21cm power spectrum as input data and bubble size distribution as output data.
Artificial Neural Network (ANN)

- ANN consists of input layer, hidden layer and output layer. Each layer has neurons.

- Training network with training dataset, ANN can approximate any function which associates input and output values.

\[ y = f(x) \]

- Applying trained network to unknown data in order to obtain expected value.

\[ y_{ANN} = f(x_{test}) \]
Some previous studies measure HII size distribution from 21cm 3D map directly.

From observational aspects, we require good angular resolution to make 21cm map.

Therefore, I attempt to recover HII size distribution from 21cm PS that does not require making 21cm map.
Setup

- 1000 EoR models
- 48000 training datasets (20% of which is used for validation)
- 2000 test datasets
- 21cm PS is ranged from $k=0.11/Mpc$ to $1.1/Mpc$ with 14 bins
- 5 hidden layers
- 212 neurons at each hidden layer
Training accuracy

MSE

iteration

10^{-2}

10^{-3}

10^{-4}

10^{-5}

0 250 500 750 1000 1250 1500 1750 2000

training
Recovered BSD

\[ x_{\text{HI}} = 0.39 \]

**Black:** Distribution obtained by 21cm map directly

**Red:** Distribution obtained by ANN

Using 21cm PS at \(0.1 < k/[\text{Mpc}] < 1\)
Different models

$x_{HI} = 0.39$

PDF

R[Mpc]

model1, true
model1, ANN
model2, true
model2, ANN
model3, true
model3, ANN
Different epoch of reionization

$x_{HI} = 0.30$

$x_{HI} = 0.51$

$x_{HI} = 0.67$

$x_{HI} = 0.80$
Scale dependence

$21\text{cm PS}$

$k_{\text{min}} < k < 1.1\text{Mpc}^{-1}$
Scale dependence

$\chi_{HI}=0.30$

$\chi_{HI}=0.51$

$\chi_{HI}=0.67$

$\chi_{HI}=0.80$
What does $k_{\text{min}}$ give best accuracy?
Thermal noise

21cm PS with thermal noises (SKA level)

Errors are estimated by 10 realizations thermal noises
Thermal noise

$x_{HI}=0.30$

$x_{HI}=0.51$

$x_{HI}=0.67$

$x_{HI}=0.80$
Past, current and future

Input

21cm power spectrum

Output

EoR parameters
bubble size distribution
global signal
Minkowski functionals

Other ideas?
Summary

• We applied artificial neural networks (ANN) to analysis of 21cm signal.
• Reconstructed EoR parameters and HII size distribution are good agreement with true values.
• (Future work) Are there other 21cm observables which we can apply machine learning to?
HII bubble size distribution

Black: Distribution obtained by 21cm map

Red: Distribution obtained by ANN

21cm PS at $0.1 < k [\text{Mpc}] < 1$
HII bubble size distribution

Black: Distribution obtained by 21cm map
Red: Distribution obtained by ANN

21cm PS at $0.1 < k[\text{Mpc}] < 1$
HII bubble size distribution

**Black:** Distribution obtained by 21cm map

**Red:** Distribution obtained by ANN

21cm PS at

$0.1 < k[/Mpc] < 1$
Sweet spot

The diagram shows the probability density function (PDF) of the comoving radius ($R_{\text{Mpc}}$) for different values of $k_{\text{min}}$: True, all_k, $k_{\text{min}}=0.15/\text{Mpc}$, $k_{\text{min}}=0.21/\text{Mpc}$, $k_{\text{min}}=0.31/\text{Mpc}$, and $k_{\text{min}}=0.45/\text{Mpc}$.
Sweet spot

![Graph showing PDF versus R[Mpc] with various kmin values.]

- True
- all_k
- kmin=0.15/Mpc
- kmin=0.21/Mpc
- kmin=0.31/Mpc
- kmin=0.45/Mpc
Sweet spot
Results

HII size distribution

Black: True distribution

Red: Obtained distribution

Shimabukuro & Semelin (2017)

EoR parameters

\( R_{mfp}, \zeta, T_{vir} \)

\( R_{mfp}^{\text{true}}, T_{vir}^{\text{true}} \)