Machine Learning of Noise for LHD Thomson Scattering System

Keisuke Fujii, Kyoto univ.
LHD Thomson scattering data

Large helical device plasma

LHD-Thomson scattering system

(a)

Characteristics

High spatial resolution: ~2cm, ~100 points.

Steadily operated: measures almost all the LHD experiment.
LHD Thomson scattering data

Characteristics

High spatial resolution: ~2cm, ~100 points.

Steadily operated:
measures almost all the LHD experiment.

What is required for the data analysis

**Inference of the smooth latent function**
from the discrete measurement.
Random noise on LHD Thomson scattering data

Uniform Gaussian noise is not appropriate.

Noise scale have a dependence on latent $T_e$ or $N_e$ values.
(Diagnostic systems have a sweet spot.)
There is systematic noise due to calibration error. (Some channels show always larger values than the vicinity.)
Objectives

Random noise

Systematic noise

Latent functions

Estimate

- Random noise
- Systematic noise
- Latent functions

from vast amount of data.
Physics-driven or Data-driven?

**Physics-driven noise model**

1. List all the possible noise sources
2. Model its distributions
3. Propagate to the signal and marginalize (integrate) them

\[
A = \bar{A} \pm \Delta A \\
B = \bar{B} \pm \Delta B
\]

\[
\Delta \left( \frac{A}{B} \right) \approx \left| \frac{1}{\bar{B}} \right| \Delta A + \left| \frac{\bar{A}}{\bar{B}^2} \right| \Delta B
\]

Does not work if there is

- Unknown noise sources
- Mis-modeling of the noise property
Physics-driven or Data-driven?

Physics-driven noise model

List all the possible noise sources

Model its distributions

Propagate to the signal and marginalize (integrate) them

Does not work if there is

Unknown noise sources

Mis-modeling of the noise property

Uncertainty propagation usually underestimates the noise amplitude.

Uncertainty by the diagnostic team cannot be trusted!
Physics-driven or Data-driven?

Physics-driven noise model

- List all the possible noise sources
- Model its distributions
- Propagate to the signal and marginalize (integrate) them

Does not work if there is
- Unknown noise sources
- Mis-modeling of the noise property

Data-driven noise model

- Directly build a noise model without considering physics
- Estimate these parameters from (a huge) data

Free from unknown noise sources.

Uncertainty propagation usually underestimates the noise amplitude.
Bayesian inference for big data

Systematic noise
\(10^2 \times 1\)

Latent functions
\(\sim 10^2 \times 10^5\)

Hyper parameters

Data
\(\sim 10^2 \times 10^5\)

\[ p(f, \delta_y | y, \theta) = \frac{p(y|f, \delta_y, \sigma)p(f|\theta)p(\delta_y|\theta)}{p(y|\theta)} \]

Data:
Thomson scattering data \(T_e\) and \(N_e\) for LHD experiment in 2013.

Size: \(\sim 300,000\) sets of data
1 set: \(~100\) radial positions \(\times\) 2 kinds of values

Total size: \(> 10^7\) points.
Outlines

- Introduction
- Model
- Inference
- Result
- Future perspective
Model: likelihood

Random noise

Systematic noise

Latent functions

Noise scale

$T_e, N_e$

Signal

$f \exp(\delta_y)$

$f$

$T_e, N_e$

$p(y|f, \delta_y, \sigma) p(f|\theta) p(\delta_y|\theta)$

$p(y|\theta)$

$p(y|f) = \mathcal{St}(y|f \exp(\delta_y), \sigma_y(f), \nu)$

Student’s $t$-distribution

True value

Shift by systematic noise
Model: prior for the calibration error $\delta_y$

$$p(\delta_y) = St(\delta_y | 0, \sigma_{\delta_y}, \nu)$$

Calibration error may distribute around 0 scale $\sigma_{\delta_y} \in \theta$
Model: noise scale model

\[ p(y|f) = St(y|f \exp(\delta y), \sigma_y(f), \nu) \]

The noise scale dependence is approximated by N.N. (Densely connected layer)
Model: latent function model

Appropriate prior is not clear. We decided to determine the shape of prior from the data.

We adopted "low dimensional assumption" for the latent functions $f$.

$f$ (≈200 points for 1 data) is described by a few parameters $z$ ($n_z = 5$).

Prior:

$$z \sim p(z) = N(z \mid 0, I)$$
Outlines

- Introduction
- Model
- Inference
- Result
- Future perspective
Optimization

The integral is intractable and the integrand is super-high dimensional.

Variational Bayesian inference

\[ p(f, \delta_y | y, \theta) = \frac{p(y|f, \delta_y, \sigma)p(f|\theta)p(\delta_y|\theta)}{p(y|\theta)} \]

We need to maximize

\[ p(y|\theta) = \int p(y|z, \delta_y, \theta)p(z|\theta)p(\delta_y|\theta)dzd\delta_y \]

over \( \theta \) [10^2 x 10^5]

Latent parameters [10^2]

Correction factors [5 x 10^5]
Variational approximation

True posterior

\[ p(z, \delta_y | y, \theta) \approx \prod_i q(z_i | y_i) q(\delta_y) \]

Variational posterior (factorized)

\[ q(z_i | y_i) = N(z_i | \mu_i, \sigma_i) \]
\[ q(\delta_y) = N(\delta_y | \mu_\delta, \sigma_\delta) \]

Optimization target: Evidence Lower Bound (ELBO)

\[ \log p(y | \theta) > \sum_i \int q(z_i | y_i) \log p(y_i | z_i, \theta) dz_i - KL[q(z_i | y_i) || p(z_i | \theta)] \]
Neural network approximation

Task is now simplified; we only need to know $\mu_z, \sigma_z$ for each $y_i$ and $\mu_\delta, \sigma_\delta$ common for all the data.

This part is still computationally expensive because a large number of experimental data.

\[
q(z|y) = N(z|\mu_Z, \sigma_Z)
\]

Instead of evaluating $\mu_z, \sigma_z$ for all data separately, we constructed N.N. to directly estimate $\mu_z, \sigma_z$ from $y_i$. 
Full network structure

(a)

\[ q(z|y) = N(z|\mu_z, \sigma_z) \]

Estimate \( \mu_z, \sigma_z \) from \( y \)

(b)

\[ p(y|z) = St(y|\mu_y, \sigma_y) \]

Estimate \( \mu_y, \sigma_y \) from \( z \) and \( R \)

Inference network.

Prediction network.
Outlines

- Introduction
- Model
- Inference
- Result
- Future perspective
Noise scale dependence

\[ p(y|f) = St(y|f \exp(\delta y), \sigma_y(f), \nu) \]

- Large noise
- Small noise

S/N ratio get worse in lower \( N_e \) side
S/N ratio get worse in very low and high \( T_e \) region

Sweet spot

S/N ratio get worse in lower \( N_e \) side
Systematic noise

Inferred calibration correction factors

Now we have a correction factor of mis-calibration.

The original data can be Post-calibrated by this values.
Summary

We estimated
- Random noise
- Systematic noise
- Latent functions

based on the neural network variational Bayes method.

With the estimated noise properties, more accurate regression becomes available.
Outlines

- Introduction
- Model
- Inference
- Result
- Future perspective
Coordinate mapping

\[ p(y|z) = St(y|\mu_y, \sigma_y) \]

This network can be extended to include other physical constraints.

e.g. \( N_e \) and \( T_e \) should be a function of magnetic flux surface.
Laplace- and variational (KL[q||p]) approximation

Laplace  KL[q||p]

approximation

True posterior

http://prog3.com/sbdm/blog/nietzsche2015/article/details/43450853
\[ p(f) = \int p(f|y)p(y)dy \approx \frac{1}{N} \sum_{i}^{N} p(f_i|y_i) \]

Sample from data distribution \( p(y) \)

Unbiased

Biased

prior

posterior