Neural Network Data Analysis in the Large Helical Device Thomson Scattering System[*]

Ichihiro YAMADA, Hisamichi FUNABA, Jong-ha LEE[1], Yuan HUANG[2] and Chunhua LIU[2]

National Institute for Fusion Science, Toki 509-5292, Japan
[1]*Korea Institute of Fusion Energy, Daejeon 34133, Korea*
[2]*Southwestern Institute of Physics, P.O. Box 432, Chengdu, 610041, China*

In Thomson scattering diagnostics systems, a combination of the lookup table and the minimum χ^2 methods has been widely used to determine electron temperature. The concept of the minimum χ^2 method is based on clearly defined mathematical statistics. However, the minimum χ^2 method calculation requires a large amount of time because all χ^2 values have to be calculated at all temperatures included in the lookup table. Thus, this method is unsuitable for the real-time data analysis required for the next generation of fusion devices, e.g., the International Thermonuclear Experimental Reactor in France. To establish real-time data analysis for Thomson scattering diagnostics, we have developed a neural network program for the large helical device (LHD) Thomson scattering (TS) system. First, we systematically studied the number of nodes and training cycles required to obtain satisfactory results, and then applied them to the LHD TS system. The calculation time was successfully reduced by approximately 1/50 - 1/100 of the χ^2 method calculation time. In addition, experimental error estimation has been performed according to the concept of the neural network method used in this study.

1. Introduction

In forthcoming thermonuclear fusion plasma research, real-time plasma diagnostics is an important issue. In current Thomson scattering (TS) diagnostics, the combined method of using a lookup table and the χ^2 method has been widely used to determine electron temperature (T_e). The concept of the traditional minimum χ^2 method is based on clearly defined mathematical statistics. However, this method requires a relatively long calculation time because all χ^2 values have to be calculated at all temperatures included in the lookup table. For example, the T_e range and the number of T_e points in the large helical device (LHD) TS lookup table are $T_e = 0.1\,\text{eV} - 50\,\text{keV}$ and 7000 points respectively [1, 2]. Therefore, this method is likely unsuitable for real-time data analysis.

To establish real-time data analysis in TS diagnostics, we applied the neural network (NN) method to determine T_e. In the Korea Superconducting Tokamak Research (KSTAR) and Huan-Liuqi-2A (HL-2A) TS systems, this method has already been successfully applied [3, 4]. We developed a new program for the LHD TS system that can be applied to the KSATR and HL-2A TS systems [5–8]. Section 2 describes the developed NN program, which includes an error estimation procedure according to the con-

cept of the NN method. The application of the developed program in the LHD TS diagnostic is presented in Section 3.

2. The Neural Network Method for Large Helical Device Thomson Scattering Diagnostics

2.1 Fundamentals

NN methods have been widely applied in various scientific fields. Once the training process has been completed in the NN method, calculation time is expected to be reduced even for complicated processes such as nonlinear processes with many input and output parameters. Although the basic concepts of the χ^2 method and the NN method are quite different, it is expected that the NN method will provide good results through proper training. Figure 1 shows a diagram of the NN program used in this study. In the LHD TS system, T_e is determined by five signals experimentally measured using a polychromator with five wavelength channels [1, 2]. Accordingly, the number of input parameters is five and the number of final results T_e is one. In this study, the depth of the hidden layer is fixed at one. The sigmoid function, $y = 1/[1 + \exp(-x)]$, is used as the activation function.

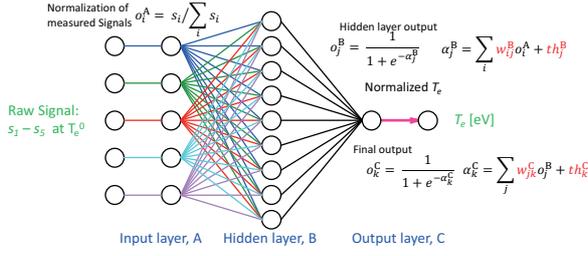In the NN program, input and output parameters are normalized to be non-dimensional values. The exper-

Fig. 1 Diagram of the NN program.

imentally obtained signal intensities are normalized as $o_i = s_i/\Sigma s_i$. Electron temperature T_e is normalized by a logarithm-type normalization as follows:

$$t = \frac{\log(T_e/T_e^{\min})}{\log(T_e^{\max}/T_e^{\min})}, \qquad (1)$$

where t means the temperature inside the program, and $T_e^{\min}$ and $T_e^{\max}$ are the minimum and maximum values of T_e in the used lookup table, respectively. The output from the j-cell in the hidden layer is given as follows:

$$o_j^B = \frac{1}{1 + e^{-\alpha_j^B}}, \quad \alpha_j^B = \sum_i w_{ij}^B o_i^A + th_j^B, \qquad (2)$$

where w_{ij} is the weight and th_j is the bias parameter. Similarly, the final output, o_k, is given as follows:

$$o_k^C = \frac{1}{1 + e^{-\alpha_k^C}}, \quad \alpha_k^C = \sum_j w_{jk}^C o_j^B + th_j^B. \qquad (3)$$

Once the training process is completed, T_e, can be easily calculated using the NN parameters determined in the training process.

In addition to determining T_e, estimating experimental error (ΔT_e) is also important in experimental studies. ΔT_e is calculated using the same NN parameters shown in Eqs. (4) and (5):

Error in the final output:

$$(\Delta o_k^C)^2 = \sum_j \left(\frac{\partial o_k^C}{\partial o_j^B} \Delta o_j^B\right)^2$$
$$= \sum_j \left\{\left[o_k^C(1 - o_k^C)W_{jk}^C\right] \Delta o_j^B\right\}^2. \qquad (4)$$

Error in the hidden layer output:

$$(\Delta o_j^B)^2 = \sum_i \left(\frac{\partial o_j^B}{\partial o_i^A} \Delta o_i^A\right)^2$$
$$= \sum_i \left\{\left[o_j^C(1 - o_j^C)W_{ij}^C\right] \Delta o_i^A\right\}^2, \qquad (5)$$

where Δo_k^C is the error of the final output, o_k^C, and Δo_j^B is the error of the hidden layer output, and Δo_i^A is the error of the input layer, i.e., the experimental errors included in the raw experimental signals. The error in the raw data propagates through Eqs. (4) and (5) to the error in the final result.
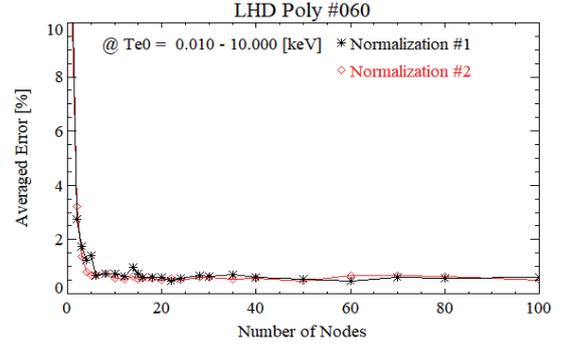
Fig. 2 Averaged error as a function of the number of nodes (N_{nodes}). The errors are estimated in the T_e range, $T_e = 10\,\text{eV}$ - $10\,\text{keV}$.

2.2 Training procedure

As a teacher data, we used a lookup table similar to that used in calculations employing the χ^2 method. In this study, we used a limited lookup table for which the T_e range and the number of T_e values are $T_e = 1\,\text{eV}$ - $50\,\text{keV}$ and 1000, respectively, to reduce training time. The initial values of the NN parameters, i.e., w and th, were set as random values between -1 and $+1$. Before performing the final training course, we determined the number of nodes, and the number of required trainings.

We first examined how many nodes were needed. The accuracy of the result provided by the NN program was studied systematically for the number of nodes: $N_{\text{nodes}} = 1$ - 100. Figure 2 shows the dependence of the averaged error for N_{nodes}, $\Delta T_e = |T_e - T_e^0|/T_e^0$ estimated for $T_e^0 = 10\,\text{eV}$ - $10\,\text{keV}$. In addition, we examined two types of normalization for experimental signals in the input layer: (1) $o_i^1 = s_i/\Sigma s_i$ and (2) $o_i^1 = s_i/\max(s_i)$, where o_i^1 is the normalized signal intensity and s_i is the raw signal ($i = 1$ - 5). As shown in Fig. 2, no significant difference was observed between the two normalization methods. Here, N_{nodes} of ~ 3 was insufficient and the deviation was large. However, this rapidly decreased from $N_{\text{nodes}} \sim 5$ and no significant improvement was observed for $N_{\text{nodes}} > \sim 10$. Accordingly, we concluded $N_{\text{nodes}} = 10$ to be sufficient for our case. Moreover, we examined the convergence property of the accuracy as a function of the number of training cycles. Figure 3 shows the results for $N_{\text{nodes}} = 3$, 6, 10, and 50. In the initial stage, the averaged error rapidly decreased as the number of training cycles increased. However, the convergence property for $N_{\text{nodes}} = 3$ is poor. After the first drop, this slowly decreased even when the number of training runs increased. For the other cases, i.e., $N_{\text{nodes}} = 6$, 10, and 50, the averaged error decreased to below 1% as the number of training cycles increased ($N_{\text{tr}} \geq 10^6$). No significant difference was seen among the results for $N_{\text{nodes}} = 6$, 10, and 30 at $N_{\text{tr}} \geq 2.5 \times 10^6$, as expected (see Fig. 2). Therefore, we consider 1×10^6 training cycles sufficient if $N_{\text{nodes}} \geq 6$.
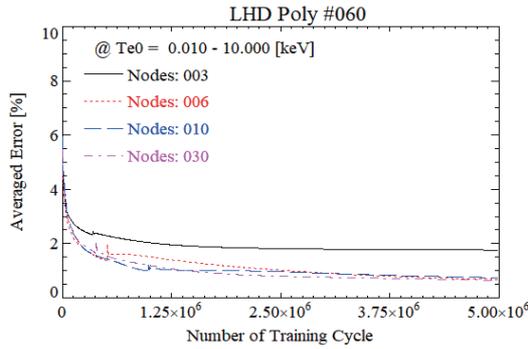
Based on the above results, we performed the training

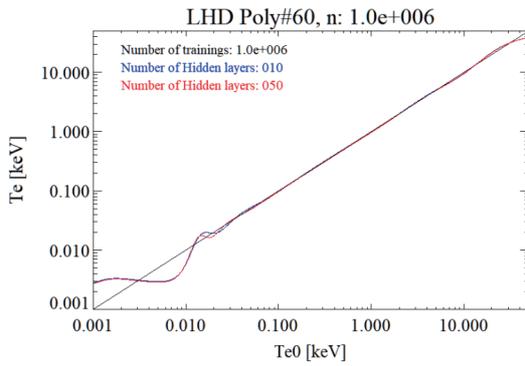Fig. 3   Averaged error as a function of the number of nodes (N_{tr}) for N_{nodes} = 3, 6, 10, and 30.

Fig. 4   Comparison between target temperature T_e^0 and final result T_e after completing the training course for N_{node} = 10 (blue) and 50 (red).
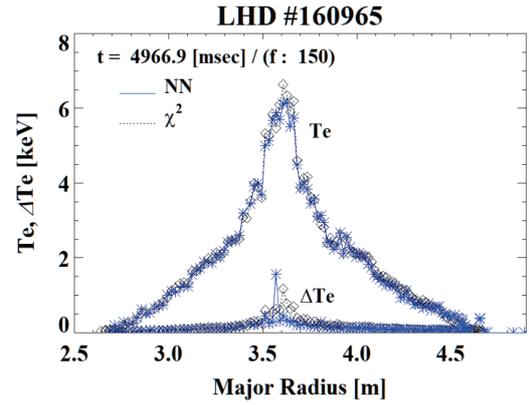

Fig. 5   (a) Comparison of T_e profiles obtained by NN and χ^2 methods. Estimated ΔT_e values are also plotted.

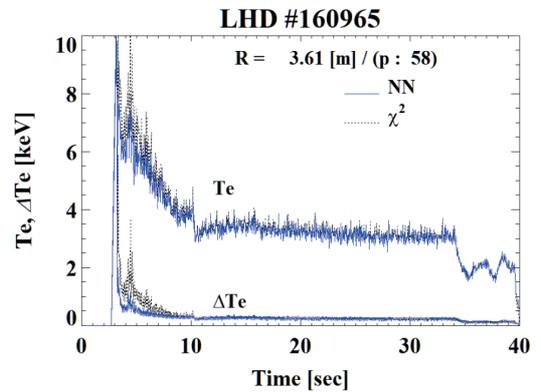Fig. 5   (b) Comparison of time histories of T_e at the plasma center obtained by the NN and minimum χ^2 methods.

procedure for all polychromators. Figure 4 shows an example of the results for N_{nodes} = 10 and 50. The number of trainings was $N_{tr} = 1 \times 10^6$. The horizontal and vertical axes show the target temperature, T_e^0, and the calculated temperature, T_e, respectively. As seen in the Fig. 4, the agreement of T_e^0 and T_e is worse in low and high T_e ranges. This deviation is due to the LHD polychromators are optimized to the temperature range $10\,eV \leq T_e \leq 10\,keV$. As expected (see Figs. 2 and 3), no clear difference was observed between the cases of N_{nodes} = 10 and 50. In low- and high-temperature regions, slight deviations between T_e^0 and T_e are observed. However, a positive agreement was indicated in the intermediate region. In regard to the normalization of T_e, a logarithm-type approach (Eq. 1) was used in this study. The KSTAR NN program uses a simple and linear normalization, i.e., $t = T_e/T_e^{max}$, where T_e is the electron temperature and T_e^{max} is the maximum temperature in the lookup table [3]. We compared results using the two normalization methods for the LHD TS. No clear or significant difference was observed between the results from the two methods.

3. Application in the LHD TS System

Using the NN parameters determined in the previous section, we calculated T_e and ΔT_e and compared them to those obtained by the traditional minimum χ^2 method. Figures 5 (a) and (b) show examples of the T_e profile and the temporal history of T_e at the plasma center, respectively. In both figures, ΔT_e values are also plotted (lower curves). As shown in the figures, the two T_e values derived using the NN and χ^2 methods show good agreement. For the ΔT_e, small deviations are observed near the center high-temperature region. However, the overall agreement is good. Therefore, we consider that the accuracy and reliability of the NN method is as good as those of the χ^2 method.

With regard to calculation time, we first divided the total calculation time, t_T, into four elements: data reading time from an analog-to-digital converter to a personal computer, t_R; preprocessing time (e.g., background signal subtraction), t_P; data analysis time, t_A; and the data writing time to a hard disk drive, t_W; $t_T = t_R + t_P + t_A + t_W$. In the case of the LHD TS system, the ratio is typically $t_R:t_P:t_A:t_W \approx 3:20:30:2$. The NN method was significantly faster than the χ^2 method, $\sim 1/50$ - $1/100$ of its t_A. The typical total calculation time of a T_e profile including 144 spatial points was ~ 30 msec and $<\sim 1$ msec for the χ^2 and

NN methods, respectively. The other three factors, t_R, t_P, and t_W were the same for both χ^2 and NN data analyses. In future work, we aim to reduce the preprocessing time, t_P, which is the most time-consuming process of the three factors, to further decrease total calculation time.

4. Conclusion

The NN method is one of the candidates for reducing data analysis time in TS diagnostics. We have developed a new program for the LHD TS system. The T_e and ΔT_e values calculated by the NN and minimum χ^2 methods show very good agreement. We consider that the accuracy and reliability of the NN method are as good as those of the χ^2 method. Data analysis time was reduced when the NN program is used for 1/50 - 1/100 that of the χ^2 method. This is an attractive feature for real-time data analysis in TS diagnostics. In addition to this merit, the data memory size required for the NN method is much smaller than that of the lookup table method. The NN method is thus suitable for calculation using a mini-board computer and GPU based system, in which memory size is limited [9].

Acknowledgments

[1] K. Narihara *et al.*, Design and performance of the Thomson scattering diagnostic on LHD, Rev. Sci. Instrum. **72**, 1122 (2001).

[2] I. Yamada *et al.*, Recent Progress of the LHD Thomson Scattering System, Fusion Sci. Technol. **58**, 345 (2010).

[3] S.H. Lee *et al.*, Development of a neural network technique for KSTAR Thomson scattering diagnostics, Rev. Sci. Instrum. **87**, 11E533 (2016).

[4] C. Liu *et al.*, Artificial neural network approach applied to data processing of Thomson scattering on HL-2A, High Power Laser Part. Beams **31**, 022003 (2019) (in Chinese).

[5] J. Lee *et al.*, Tangential Thomson scattering diagnostic for the KSTAR Tokamak, J. Instrum. **7**, C02026 (2012).

[6] J.H. Lee *et al.*, Development of prototype polychromator system for KSTAR Thomson scattering diagnostic, J. Instrum. **10**, C12012 (2015).

[7] C.H. Liu *et al.*, The progress in development of edge tangential Thomson scattering system on HL-2A tokamak, Rev. Sci. Instrum. **87**, 11E555 (2016).

[8] Y. Huang *et al.*, Multipoint vertical-Thomson scattering diagnostic on HL-2A tokamak, Rev. Sci. Instrum. **89**, 10C116 (2018).

[9] Seung-Ju Lee *et al.*, Design of GPU-based parallel computation architecture of Thomson scattering diagnostic in KSTAR, Fusion Eng. Des. **158**, 111624 (2020).