A Robust and Accurate Load Forecasting Model for Noisy and Multi-Typed Load Data

Song Deng, *Member, IEEE*, Xia Dong, Junjie Wang, Di Wu, *Member, IEEE*, Yi He, *Member, IEEE*, Xin Luo, *Senior Member, IEEE*

This is the supplementary file for the paper entitled -A Robust and Accurate Load Forecasting Model for Noisy and Multi-Typed Load Data in IEEE Transactions on Systems, Man, and Cybernetics: Systems. Additional sections, figures and tables are put into this file and cited by the paper.

S1.SUPPLEMENTARY FIGURE

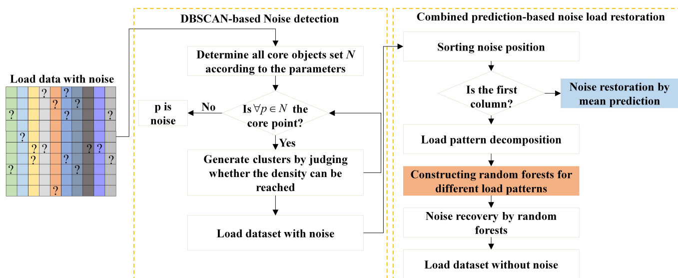

Figure 1: The logical flow of noise load detection and repair.

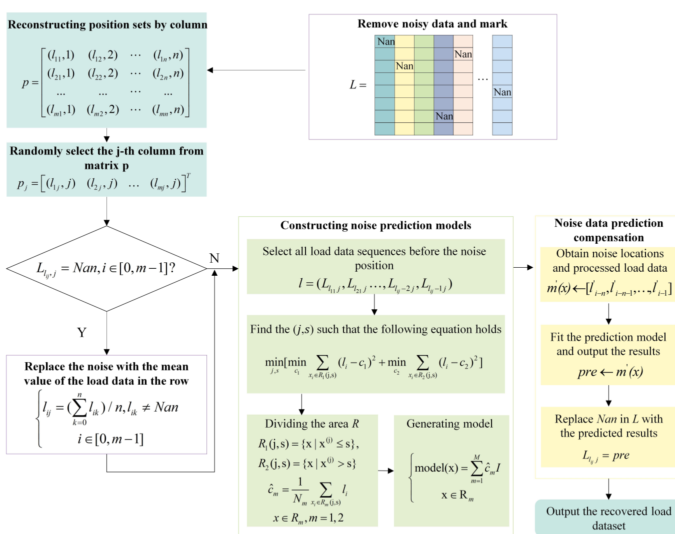

Figure 2: The flow of NLS-CP.

S2.SUPPLEMENTARY PROOF

Proof. According to *Definition 5* of the original paper, we have the following equation

$$\begin{aligned} & m_r(X,Y) = a_{v_k}(T_k(X)=Y) - \max_{j\neq Y} a_{v_k}(T_k(X)=j) \\ & = P_\Theta(T(X,\Theta)=Y) - \max_{j\neq Y} P_\Theta(T(X,\Theta)=j) \end{aligned} . \tag{1}$$

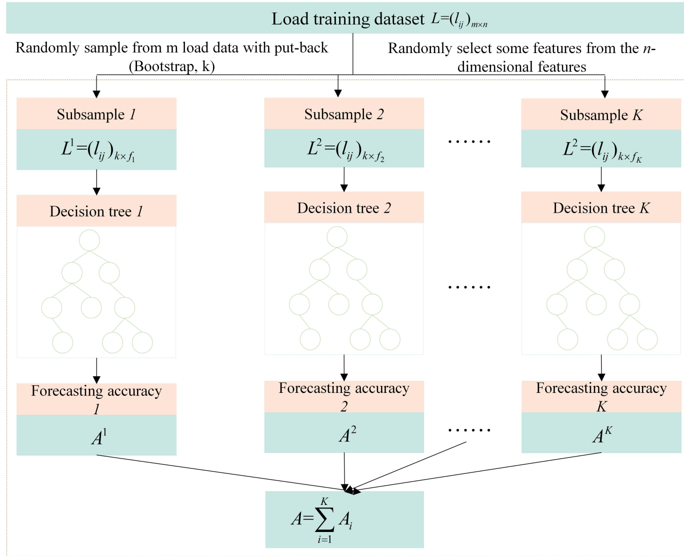

Figure 3: The schematic diagram of the LFR-MF.

Due to $GE = P_{X,Y}(m_r(X,Y) < 0) = P_{X,Y}(m_r(X,Y) - E_{X,Y}(m_r(X,Y)) < -E_{X,Y}(m_r(X,Y)))$, according to Chebyshev's inequality, we have the following inequality holds

$$GE \le \frac{Var(m_r(X,Y))}{(E_{X,Y}(m_r(X,Y)))^2} = \frac{Var(m_r(X,Y))}{s^2}. \tag{2}$$

Where $s = E_{X,Y}(m_r(X,Y))$. Let $\hat{j}(X,Y) = \arg\max_{j\neq Y} P_\Theta(T(X,\Theta) = j)$, then Eq. 1 can be changed to

$$\begin{aligned} m_r(X,Y) &= P_\Theta(T(X,\Theta) = Y) - P_\Theta(T(X,\Theta) = \hat{j}(X,Y)) \\ &= E_\Theta(I(T(X,\Theta) = Y) - I(T(X,\Theta) = \hat{j}(X,Y))). \end{aligned} \tag{3}$$

For a vector Θ, we have $E(r_{m_r}(\Theta,X,Y)) = m_r(X,Y)$. Since Θ and $\Theta^{'}$ are two independent and identically distributed random vectors, for any function f, the following equation holds

$$(E(f(\Theta)))^2 = E_\Theta(f(\Theta)) \cdot E_{\Theta'}(f(\Theta^{'})) = E_{\Theta,\Theta'}(f(\Theta)\cdot f(\Theta^{'})). \tag{4}$$

Thus, $(m_r(X,Y))^2 = (E(r_{m_r}(\Theta,X,Y)))^2 = E_{\Theta,\Theta'}(r_{m_r}(\Theta,X,Y)\cdot r_{m_r}(\Theta^{'},X,Y))$, then we have

$$\begin{aligned} Var(m_r(X,Y)) &= cov_{X,Y}(E(r_{m_r}(\Theta,X,Y)), \\ E(r_{m_r}(\Theta,X,Y))) &= E_{\Theta,\Theta'}(cov_{X,Y}(r_{m_r}(\Theta,X,Y), \\ r_{m_r}(\Theta,X,Y))) &= E_{\Theta,\Theta'}(\rho(\Theta,\Theta^{'})\cdot s_d(\Theta)\cdot s_d(\Theta^{'})) \\ &= E_{\Theta,\Theta'}(\rho(\Theta,\Theta^{'}))\cdot E_{\Theta,\Theta'}(s_d(\Theta)\cdot s_d(\Theta^{'})) \end{aligned} \tag{5}$$

Where $s_d(\Theta)$ is the standard deviation of $r_{m_r}(\Theta,X,Y)$. Then, Eq. 6 holds

$$\begin{aligned} Var(m_r(X,Y)) &= \bar{\rho}\cdot(E_\Theta(s_d(\Theta)))^2 \\ &\le \bar{\rho}(E_\Theta(s_d(\Theta)^2)) \\ &= \bar{\rho}(E_\Theta(Var(\Theta))) \end{aligned} \tag{6}$$

Where

$$\begin{aligned} Var(\Theta) &= Var(r_{m_r}(\Theta,X,Y)) \\ &= E_{X,Y}(r_{m_r}(\Theta,X,Y))^2 - (E_{X,Y}(r_{m_r}(\Theta,X,Y)))^2 \\ &= E_{X,Y}(r_{m_r}(\Theta,X,Y))^2 - s^2 \end{aligned} \tag{7}$$

And since $r_{m_r}(\Theta,X,Y) = I(T(X,\Theta) = Y) - I(T(X,\Theta) = \hat{j}(X,Y)) \le 1,$. From Eq. 7, we have $Var(\Theta) \le 1 - s^2$. Then, the following inequality holds

$$E_\Theta(Var(\Theta)) \le 1 - s^2. \tag{8}$$

We bring Eq. 8 into Eq. 6, and obtain

$$Var(m_r(X,Y)) \le \bar{\rho}(1 - s^2). \tag{9}$$

We then bring Eq. 9 into Eq. 2 to obtain the following inequality holds

$$GE \le \frac{Var(m_r(X,Y))}{s^2} \le \frac{\bar{\rho}(1-s^2)}{s^2}. \tag{10}$$

□

S3.SUPPLEMENTARY TABLE

Table 1: Experimental datasets

Abbreviations	Name	# Instance	# Features
D1	Nantong-Resident	159	96
D2	Nantong-Commerce	430	96
D3	Nantong-Mixed	159	96
D4	District1-Market	365	24
D5	District1-Resident	365	24
D6	District1-Restaurant	365	24
D7	District1-School	365	24
D8	District2-Resident	365	24
D9	District2-School	365	24
D10	District2-Warehouse	365	24
D11	District3-Hospital	365	24
D12	District3-Resident	365	24
D13	District3-Restaurant	365	24

Table 2: Experimental datasets with noise load data

Abbreviations	Name	# Instance	# Features	# Noise
D1	Nantong-Resident	159	96	768
D2	Nantong-Commerce	430	96	2112
D3	Nantong-Mixed	159	96	768
D4	District1-Market	365	24	456
D5	District1-Resident	365	24	456
D6	District1-Restaurant	365	24	456
D7	District1-School	365	24	456
D8	District2-Resident	365	24	456
D9	District2-School	365	24	456
D10	District2-Warehouse	365	24	456
D11	District3-Hospital	365	24	456
D12	District3-Resident	365	24	456
D13	District3-Restaurant	365	24	456

Table 3: Experimental data after multi-type load data fusion

Data source	Data name after fusion	#Instance after fusion	#Feature after fusion
Nantong-Resident Nantong-Commerce Nantong-Mixed	NT	600	96
District1- Market District1-Resident	D1-RM	584	24
District1- Resident District1-Restaurante	D1-RR	584	24
District1- Resident District1-School	D1-RS	584	24
District2-Resident District2-School	D2-RS	584	24
District2-Resident District2-Warehouse	D2-RW	584	24
District3-Resident District3-Hospital	D3-RH	584	24
District3-Resident District3-Restaurant	D3-RR	584	24

S4.SUPPLEMENTARY DEFINITION

Let $L = \{L_1, L_2, ..., L_m\}, L_i = \{l_{i1}, l_{i2}, ..., l_{in}\}, i \in [1, m], j \in [1, n]$ be load dataset, we have the following definitions.

Definition 1. *$\forall p \in L$, the neighborhood with p as the center and ε as the radius is called the ε-neighborhood of p, denoted as $N\_\varepsilon(p) = \{q \in L | d(p, q) \leq \varepsilon\}$.*

Definition 2. *$\forall p \in L$, $p_i, i \in [1, k]$ is in the ε-neighborhood of the load data p. If $k > Minpts$, then we call that the load data p is a core point.*

Definition 3. *Let the load data* p *be the core point. For any q in L, $q_i, i \in [1,k]$ is in the ε-neighborhood of the load data q. If $k < Minpts$ and $p_i \in N\_\varepsilon(p)$, then q is said to be a boundary point.*

Definition 4. *$\forall p, q \in L$, if $p \in N\_\varepsilon(q)$ and $|N\_\varepsilon(q)| \geq Minpts$, then q is said to be a direct density up to p, noted as $DDR(p, q)$.*

Definition 5. *$\forall p, q \in L$, $\exists p_1, p_2, ..., p_k$, if $p_1 = p$ and $p_n = q$, where p_i can be directly densely accessible to p_{i+1}. Then q is said to be densely accessible to p, denoted $DR(p, q)$.*

Definition 6. *$\forall p, q \in L$, $\exists r \in L$, If r can be densely connected up to p and q at the same time, then p and q are said to be density connected, denoted as $DC(p, q)$.*

Definition 7. *$\forall p, q \in L$, let C be load dataset, if p, q, C satisfy the following properties: (1) if $p \in C$ and $DC(p, q)$, $q \in C$; (2)$DC(p, q)$. Then $C \subset L$, we call that C is a cluster of L.*

Definition 8. *Let N be the set of load data, and for a given ε and $Minpts$, there exist k clusters $C_1, C_2, ..., C_k$, and if $N = \{p \in L | p \notin C_i\}$, then is called the set of noisy load.*

S5. SUPPLEMENTARY ALGORITHM

Algorithm 1: NLD-DBSPA

Input: $L = (L_1, L_2, ..., L_m)$,ε,$Minpts$
Output: Noise load dataset N

$L^{'} \leftarrow ReconCol(L)$;//Reconstructing the load dataset by column
Marking all load points in $L^{'}$ with an unprocessed status;
$(s_L, N_c) \leftarrow Init(L, \varepsilon, Minpts)$;
while *($|s_L - 1| > \sigma_1$ or $|1 - s_L| > \sigma_2$) and $|N_c - 1| > \sigma_3$* **do**
    for *each p in L* **do**
        For object p, check the ε-neighborhood $N\_\varepsilon(p)$;
        if *the number of objects contained in $N\_\varepsilon(p)$ is less than $MinPts$* **then**
            p is marked as a boundary point or noise point
        else
            p is marked as a core point, and a new cluster C is created, and all points in the neighborhood of p are added to C;
            for *all $q \notin N\_\varepsilon(p)$* **do**
                if *$N\_\varepsilon(p)$ contains $MinPts$ objects* **then**
                    we add to C the objects in $N\_\varepsilon(p)$ that are not classified;
for *each p in L* **do**
    if *$p \in L$ and $p \notin C_i$* **then**
        $N \leftarrow N.insert(p)$;
Return N;

S6. SUPPLEMENTARY FRIEDMAN AND NIMENYI TEST

There are N experimental datasets and k comparison algorithms, and let denote the average ordinal value of the i-th algorithm. In the experiment, we set the variable Γ_F to obey the F distribution with $k-1$ and $(k-1)(N-1)$ degrees of freedom. For the datasets after multi-type data fusion, this paper conducts Friedman test and Nemenyi post-hoc test on the prediction results of LFR-MF, LR, XGBoost, SVR, DRN_STLF and APLF. The null hypotheses (H_0) and opposite hypotheses (H_1) of the Friedman test on the experimental data are denoted as follows.

1) H_0: There is no significant difference in the performance of multiple algorithms on multiple datasets;
2) H_1: There are obvious differences in the performance of multiple algorithms on multiple datasets.

If the difference of average ordinal value between any two algorithms exceeds the critical value range, the following inequality holds.

$$(r_i - r_j) > q_\alpha \sqrt{\frac{k(k+1)}{6N}} \tag{11}$$

Where q_α is the test critical value. Then we reject the null hypothesis H_0.

Algorithm 2: NLS-CP

Input: $L = (L_1, L_2, ..., L_m)$, Noise Locations Set $P_N = [p_1, p_2, \cdots, p_N]$

Output: Recovered load dataset R_L

$P_N \leftarrow ReconCol(P_N)$;//Reconstructing position sets P_N by column

Mark all load points in $L^{'}.columns$ as unprocessed;

for $P_N[i]$ *in* P_N **do**

  $L^{'}[P_N[i]] = Nan$;

for $cols$ *in* $L^{'}.columns$ **do**

  $i = P_N[i][0], j = P_N[i][1]$;

  if $\exists L^{'}{}_{i,j} = Nan, i \in [0, n-1]$ **then**

    $L^{'}{}_{ij} = (\sum_{k=0}^{m} L^{'}{}_{ik})/m, p_{ik} \neq Nan$;

  else

    $l = (L^{'}{}_{i-n,j}, L^{'}{}_{i-n-1,j}, \ldots, L^{'}{}_{i-2,j}, L^{'}{}_{i-1,j})$;

    $l^{'} = \{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\} \leftarrow l$;

  Find the (j, s) such that the following equation holds $\min_{j,s}[\min_{c_1} \sum_{x_i \in R_1(j,s)} (l^{'}{}_i - c_1)^2 + \min_{c_2} \sum_{x_i \in R_2(j,s)} (l^{'}{}_i - c_2)^2]$;

  $(R_1(j, s), R_2(j, s)) \leftarrow Divid(R)$;

  $\hat{c}_m \leftarrow \frac{1}{N_m} \sum_{x_i \in R_m(j,s)} l^{'}{}_i$;

  $m(x) \leftarrow model(x) = \sum_{m=1}^{M} \hat{c}_m I$;

  $m^{'}(x) \leftarrow [l^{'}{}_{i-n}, l^{'}{}_{i-n-1}, \ldots, l^{'}{}_{i-1}]$;

  $pre \leftarrow m^{'}(x)$;

  $L^{'}{}_{ij} \leftarrow pre$;

  $R_L \leftarrow L^{'}$;

Return R_L;

Algorithm 3: MLF-CC

Input: Load dataset: $L^I_{m\times n}$, $L^C_{k\times n}$

Output: The fused load dataset L

for $i \leftarrow 1$ *to* m **do**

  for $j \leftarrow 1$ *to* k **do**

    $L^I_i \leftarrow (l^I_{i1}, l^I_{i2}, ..., l^I_{in})$;

    $L^C_i \leftarrow (l^C_{i1}, l^C_{i2}, ..., l^C_{in})$;

    $r_{L^I_i L^C_i} \leftarrow \frac{n\sum_{j=1}^{n} l^I_{ij} l^C_{ij} - \sum_{j=1}^{n} l^I_{ij} \sum_{j=1}^{n} l^C_{ij}}{\sqrt{n\sum_{j=1}^{n}(l^I_{ij})^2 - (\sum_{j=1}^{n} l^I_{ij})^2}\sqrt{n\sum_{j=1}^{n}(l^C_{ij})^2 - (\sum_{j=1}^{n} l^C_{ij})^2}}$;

    if $r_{L^I_i L^C_i} \approx 0$ **then**

      $L \leftarrow L.insert(L^I_i, L^C_i)$;

Return L;

If the test result rejects H_0, it means that the performance of the k algorithms is significantly different, and then the performance of the k algorithms is further distinguished based on the Nemenyi post-hoc test.

For the datasets after multi-type data fusion, the Friedman test results are shown in Table 4.

The experimental results in Table 4 demonstrate that p-values corresponding to RMSE, MAE, MAPE and R^2 are less than 0.05, respectively. Then we reject the null hypothesis H_0. In other words, the performance of each load forecasting algorithm is significantly different, and Nemenyi post-hoc test is required to further distinguish each algorithm. Fig. 4 shows the test results of six load forecasting algorithms based on Nemenyi post-hoc test for different evaluation metrics.

As can be seen in Fig. 4, our proposed LFR-MF algorithm has no overlapping regions with DRN_STLF, SVR and APLF in terms of MAE, RMSE and , which indicates that the LFR-MF is significantly different from the above three models. And compared with the LR, the LFR-MF has only a few overlapping regions with it on the RMSE. Although the results in Fig. 4 show that there is no significant difference in the performance of the LFR-MF compared with the XGBoost model. However, judging from the forecasting results in Fig.9 and Fig.10 of original paper, it still outperforms XGBoost.

Algorithm 4: LFR-MF

Input: Load dataset: $L^{I}_{m\times n}$, $L^{C}_{k\times n}$; Radius of cluster:ε;Density threshold: $Minpts$; Error: σNumber of regression decision trees: K

Output: Load Forecasting Accuracy A_c

$N^I \leftarrow NLD - DBSPA(L^I, \varepsilon, Minpts, \sigma)$;
$N^C \leftarrow NLD - DBSPA(L^C, \varepsilon, Minpts, \sigma)$;
$R_{N^I} \leftarrow NLS - CP(L^I, N^I)$;
$R_{N^C} \leftarrow NLS - CP(L^C, N^C)$;
$L \leftarrow MLF - CC(R_{N^I}, R_{N^C})$;
$\{L_1, L_2, ..., L_m\} \leftarrow SamplePutBack(L, m)$;
for $i \leftarrow 1$ *to* m **do**
    $T_i \leftarrow RDTree(L_i, k)$;
    $A_i \leftarrow CalAcc(T_i)$;
    $sum \leftarrow sum + A_i$;
$A_c \leftarrow \frac{sum}{m}$;
Return A_c;

Table 4: Friedman test results

Index	Test statistic value	p-value
RMSE	23.728	0.00127
MAE	27.485	0.00027
MAPE	14.626	0.0411
R^2	16.157	0.02372

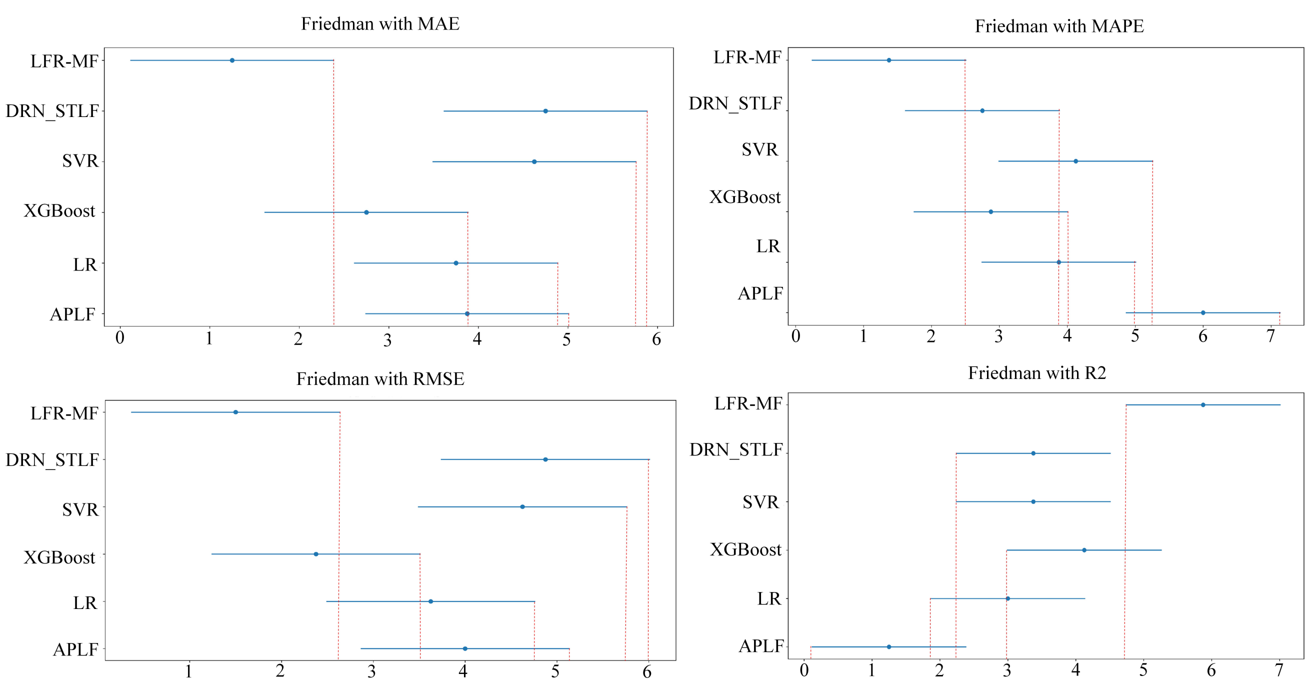

Figure 4: Comparison of test results of six load forecasting algorithms based on Nemenyi post-hoc test.