The Online Supplementary Material

Baligh Al-Helali, Qi Chen, Bing Xue, and Mengjie Zhang

School of Engineering and Computer Science
Victoria University of Wellington, PO Box 600, Wellington 6400, New Zealand
Email: {baligh.al-helali, Qi.Chen, Bing.Xue, Mengjie.Zhang}@ecs.vuw.ac.nz

Supplementary Example S1. In this example, the high level steps of the proposed imputation method are clarified. Assume that the data $\mathcal{X}$ is given as in Table 1 where columns are features, rows are instances, and "?" refers to missing values.

Table 1. Data set with missing values

F1	F2	F3	F4	F5
3	?	3	1	?
3	1	2	2	2
-1	4	7	5	?
?	3	1	-3	5
2	?	-4	2	3
4	6	4	1	7
0	4	-2	9	1

Following the main steps of the proposed imputation method (GP-KNN) for the missing value at $i = 1, j = 2$:

1. From the 1st instance, extract the complete instance $V_{1,2}$.

$$V_{1,2} = \begin{pmatrix} 3 & 3 & 1 \end{pmatrix} \tag{1}$$

2. Obtain a sub data set $\mathcal{X}_{1,2}^T$ by excluding the features F2 and F5 and the instances 1 and 5 getting:

$$\mathcal{X}_{1,2}^T = \begin{pmatrix} 3 & 2 & 2 \\ -1 & 7 & 5 \\ ? & 1 & -3 \\ 4 & 4 & 1 \\ 0 & -2 & 9 \end{pmatrix} \tag{2}$$

3. Form a complete sub data set $\mathcal{X}_{1,2}$ by removing instances having missing values from $\mathcal{X}_{1,2}^T$.

$$\mathcal{X}_{1,2} = \begin{pmatrix} 3 & 2 & 2 \\ -1 & 7 & 5 \\ 4 & 4 & 1 \\ 0 & -2 & 9 \end{pmatrix} \tag{3}$$

4. Get k $(k = 2)$ nearest instances to $V_{1,2}$ as:

$$\mathcal{X}_{1,2}^{k} = \begin{pmatrix} 3 & 2 & 2 \\ 4 & 4 & 1 \end{pmatrix} \tag{4}$$

5. Build an SR function $f_{1,2}$ using $\mathcal{X}_{1,2}^{k}$ considering the 2nd feature of $\mathcal{X}$ as a target variable.

$$Target = \begin{pmatrix} 1 \\ 6 \end{pmatrix} \tag{5}$$

6. Impute the missing value $\mathcal{X}[1, 2]$ as:

$$\mathcal{X}^{C}[1, 2] = f_{1,2}(V_{1,2}) \tag{6}$$