Learning *Matching* Representations for Individualized Organ Transplantation Allocation

Supplement

Can Xu*  
University of Cambridge  
ex243@cam.ac.uk

Ahmed M. Alaa*  
UCLA  
ahmedmalaa@ucla.edu

Ioana Bica  
University of Oxford  
The Alan Turing Institute

Brent D. Ershoff  
UCLA

Maxime Cannesson  
UCLA

Mihaela van der Schaar  
University of Cambridge  
UCLA  
The Alan Turing Institute

1 Dataset Description

1. **PLTSD** is a real-world dataset, in which pairs of recipients and donors are extracted from the Liver Transplant Standard Dataset. The dataset contains 4460 recipient-donor pairs. Recipients are described by 55-dimensional feature vectors, and donor feature vectors have a dimensionality of 28. Because this is a real-world dataset, and full counterfactual vectors are never observed in the real world, the dataset contains only factual outcomes. Thus, the evaluation metrics are different from those of other fully observed datasets, described in later sections. Besides, the dataset is biased because of ethical and other issues. For instance, about 98% of donors are paired with recipients having the same blood type, which means transplantation outcomes of recipient-donor pairs with different blood types are rarely observed in the dataset.

2. **FPLTSD** is a semi-synthetic dataset generated based on PLTSD. One limitation of PLTSD is that only factual outcome is observed, but entire vectors of counterfactuals are not observed in PLTSD. Because of the limitation, it is impossible to apply many evaluation metrics to evaluate models trained on PLTSD. To overcome the disadvantage, we create this semi-synthetic dataset, in which recipients’ and donors’ features are identical to PLTSD, but potential outcomes are generated. In order to generate potential outcomes, the dimensionality of counterfactual vectors has to be set. We assume there are three types of donors, which means the dimensionality of counterfactual vectors is set to be 3. We also assume that there are two types of donors. Recipients and donors are then clustered using the Expectation-Maximization algorithm. For each combination of recipient type and donor type, the mean and variance of the potential outcome are estimated using observed data in PLTSD. For each recipient, the counterfactual vector is then generated from a Gaussian distribution with corresponding estimated means and variances. In this way, the expected potential outcome for each combination of recipient type and donor type is kept invariant to observed data in PLTSD.

| Table 1: Donor features of donors in PLTSD |
|---|---|---|---|
| Donor Features | Donor type group | Donor BMI | Donor age at donation |
| Type of donor | Donor paediatric/adult indicator | Donor sex | Donor blood group | Donor cause of death |
| Donor CMV test result | Normal Anatomy | Left Hepatic Gastric | Right Hepatic Sma |
| Left And Right Hepatic | Number of hepatic arteries | Patch on liver | Common Hep Artery |
| Main Portal Vein | Common Bile Duct | Vena Cava | Iliac Artery present |
| Iliac Vein present | Lymph Node present | Spleen present | Steatosis |
| Capsular Damage | Kidney(s) retrieved | Heart retrieved | Lung(s) retrieved |
Table 2: Recipient features of recipients in PLTSD.

| Recipient Features                                                   |
|---------------------------------------------------------------------|
| age at transplant                                                   |
| paediatric/adult indicator                                          |
| sex                                                                 |
| 2010 IMD deprivation ranking order                                  |
| blood group                                                         |
| weight at registration (kg)                                          |
| height at registration (cm)                                          |
| ethnicity group                                                     |
| CMV test result at registration                                     |
| Most recent recipient HCV test result                               |
| NHS group status                                                    |
| Primary liver disease                                               |
| liver failure grading                                               |
| life-style activity score                                           |
| in patient status                                                   |
| ventilation status                                                  |
| renal support status                                                |
| clinically detectable anastomosis status                            |
| cirrhotic therapy status                                            |
| encephalopathy grade                                                |
| abdominal surgery indicator                                         |
| ICP monitor results                                                 |
| oesophageal varices                                                 |
| shunt for varices                                                   |
| pyrexial indicator                                                  |
| sepsis confirmed indicator                                          |
| site of sepsis                                                      |
| haemoglobin (gm/dl)                                                 |
| white blood cells (x10^9/l)                                         |
| platelets (x10^9/l)                                                 |
| creatinine (µmol/l)                                                 |
| albumin (g/l)                                                       |
| urea (mmol/l)                                                       |
| INR                                                                 |
| Icterus (µmol/l)                                                    |
| sodium (mmol/l)                                                     |
| UK End-stage Liver Disease score                                    |
| Model for End-stage Liver Disease score                             |
| potassium (mmol/l)                                                  |
| pH                                                                  |
| oxygen pressure (KPa)                                               |
| HBs AG antigen result                                               |
| HBs AG antigen result                                               |
| HCV-RNA (PCR) antigen result                                        |
| Anti HBclgM result                                                  |
| Anti HBe result                                                     |
| Anti HCV result                                                     |
| Anti DELTA result                                                   |
| Anti CMV result                                                     |
| Anti HSV result                                                     |
| Anti HIV result                                                     |

Table 3: Ratios of pairings of recipients and donors with different blood types in the PLTSD dataset. Rows are for fixed donor blood types, and columns are for fixed recipient blood types. As shown, about 98% of donors are paired with recipients having the same blood type, which means the dataset is biased.

| Blood Type | O     | A     | B     | AB    |
|------------|-------|-------|-------|-------|
| O          | 0.9863| 0.0026| 0.0106| 0.0005|
| A          | 0.0010| 0.9752| 0     | 0.0238|
| B          | 0     | 0     | 0.9933| 0.0067|
| AB         | 0     | 0.0138| 0     | 0.9862|

2 Additional Experiments: Merging Clusters

In this section, we demonstrate Matching rep.’s ability of merging similar clusters. Models are evaluated on a synthetic dataset, called GMixOverlap. The dataset is generated using Gaussian Mixture Models.

\[
r \sim \mathcal{N}(\mathbf{m}_r, \mathbf{v}_r); \mathbf{m}_r, \mathbf{v}_r \in \mathbb{R}^{128}
\]

\[
o \sim \sum_{i=1}^5 w_i \mathcal{N}(\mathbf{m}_{oi}, \mathbf{v}_{oi}); \mathbf{m}_{oi}, \mathbf{v}_{oi} \in \mathbb{R}^{64}
\]

\[
y \sim \mathcal{N}(\begin{bmatrix} 100 \\ 1000 \\ 100 \\ 1000 \\ 100 \end{bmatrix}, \begin{bmatrix} 10 & 0 & 0 & 0 & 0 \\ 0 & 100 & 0 & 0 & 0 \\ 0 & 0 & 10 & 0 & 0 \\ 0 & 0 & 0 & 100 & 0 \\ 0 & 0 & 0 & 0 & 10 \end{bmatrix})
\] (1)

Donor features are generated from 5 different Gaussian distributions, which means there are five generative classes for donors. However, donors from 3 out of 5 classes lead to similar outcomes, and donors from the other two classes lead to another type of outcome. Therefore, there are, in fact, two significant donor classes. In this experiment, we show that Matching rep. can merge similar donor types, independent from the hyperparameter setting of the number of donor types.

Models are trained with the number of clusters set to be 5 and 8. As shown in Table 4, Matching rep. outperforms all other models. Besides, as shown in the table, having more clusters does not have a significant impact on model performance. Different clusters can have similar estimation functions so that having extra clusters does not significantly affect model performance.

As shown in Figure 1, jointly trained DEC clusters in Matching rep. merges similar clusters. Extra clusters are often eliminated during training so that the setting of the number of clusters does not have significant effects on Matching rep.. With different settings of the number of initial clusters, the jointly trained clusters tend to approach the same optimal clustering. Although some times extra clusters are not eliminated completely (in
Table 4: Precisions of baseline models and Matching rep. evaluated on the GMixOverlap dataset. The average precision of the top 50 trails out of 100 trails is taken. The number of clusters is set to be 5 and 8 to evaluate the performance of models under different settings.

| Model                       | 5 clusters $\epsilon_F$ | 8 clusters $\epsilon_F$ |
|-----------------------------|--------------------------|--------------------------|
| K-Means/Linear              | 9.6768                   | 9.7185                   |
| EM/Linear                   | 11.0664                  | 9.7039                   |
| K-Means/Multi-Branched NN   | 8.7396                   | 8.7061                   |
| EM/Multi-Branched NN        | 8.7003                   | 8.7165                   |
| DEC/Multi-Branched NN       | 8.7078                   | 8.6890                   |
| Matching rep.               | **8.6157**               | **8.6489**               |

Figure 1: Results of clustering of K-Means, EM, DEC, and jointly train DEC in Matching rep.. Since clustering algorithms do not have class labels, colors are randomly assigned to each cluster, and the same color in different plots does not represent the same class. Sample points are mapped to dimension 2 using PCA Pearson (1901).

Figure 1, when initialized with 8 clusters, there are 3 clusters left, while the optimal number of clusters is 2), we can still see the intention of approaching the optimal merging. The ability to merge similar clusters does not only provide more robustness but also shows more interpretability of the data. In the real world, the number of types of donors is latent and unknown. The ability to merge similar types provides us insight into how many significant donor types there are, which is beneficial for further studies.

3 Additional Baselines

Table 5: Results of comparing to additional baseline ConfidentMatch (Yoon et al., 2017), evaluated on PLTSD.

| Policy             | Death rate | Avg. survival | Avg. benefit |
|--------------------|------------|---------------|--------------|
| Real               | 0.25       | 1280.41       | 703.23       |
| ConfidentMatch     | 0.30       | 1289.54       | 576.78       |
| Matching rep. (FCFS)| 0.24       | 1285.09       | 720.69       |
| Matching rep. (UF) | 0.24       | 1299.74       | 730.85       |
| Matching rep. (BF) | 0.23       | 1305.16       | 736.98       |
4 Additional Experiments: Distance Metrics

Table 6: Performance of using different distance metrics in representation learning.

| Distance Metrics         | PLTSD $\epsilon_F$ | UNOS-HR $\epsilon_F$ | UNOS-LU $\epsilon_F$ |
|-------------------------|---------------------|-----------------------|-----------------------|
| NONE                    | 14.52               | 7.04                  | 7.06                  |
| Euclidean Dist of means/variances | 14.51               | 6.88                  | 6.71                  |
| KL Divergence           | 14.50               | 6.91                  | 6.97                  |

5 Additional Experiments: Alternate Training

One limitation of Matching rep. is that Matching rep. tends to sacrifice clustering certainty to maximize the precision of estimation of expected factual outcomes. Although, in this paper, expected factual outcomes are emphasised, in other cases, deterministic factual outcomes could be crucial. Therefore, we introduce alternate training, with which Matching rep. shows better precision of estimated deterministic factual outcomes.

In the alternate training approach, we train one of the clustering component and counterfactual estimation component of Matching rep., with the other component frozen. Besides, when training the counterfactual estimator, an entry of the estimated counterfactual vector corresponding to the assigned donor cluster is taken to compute factual loss, instead of using the expectation of a counterfactual vector as estimated factual to compute factual loss. With the edited factual loss, the estimator is encouraged to maximize the precision of estimation of deterministic factual outcomes, rather than expected factual outcomes. Formally, the alternate training approach involves the following steps:

1. Use K-Means to fit donor feature vectors as the initialization of clustering targets of the clustering component.
2. Train a clustering model.
3. Train the counterfactual estimator using the edited factual loss function, while the clusters are fixed.
4. Take the index of the entry of each estimated counterfactual vector that is the closest to the ground truth factual outcome as the new clustering target of the clustering component.
5. Go back to step 2.

Table 7: Performance of Matching rep. with and without alternate training evaluated on the PLTSD dataset.

| Model                     | $\epsilon_F$ | $\epsilon_{WMSE}$ |
|---------------------------|--------------|-------------------|
| Matching rep.             | 13.0821      | 13.1192           |
| Matching rep. alternate trained | 13.1382      | 13.0708           |

As a result, alternate training improves Matching rep.’s precision of estimation of deterministic factual outcomes, but $\epsilon_{PEEF}$ decreases as the cost. Besides, we also test how alternate training effects Matching rep.’s ability to merge clusters. Matching rep. is trained on the GMixOverlap dataset with and without alternate training. Plots of clustering results is shown in Figure 2.

With alternate training, extra clusters are not eliminated but expanded to cover the whole subspace of the group of similar clusters. This is because, in step 3, clusters are fixed, so some branches of the branched neural network are forced to estimate the ground truth factual outcome. While without alternate training, these branches may not be updating significantly because their corresponding clusters may have been merged. Therefore, in step 4, some clustering targets may be set to be clusters that should have been merged without alternate training. Thus, clusters are not merged with alternate training.
Figure 2: Clusters of Matching rep. trained with and without alternate training. Matching rep. is trained on the GMixOverlap dataset. Colors are randomly assigned to each cluster, and the same color in different plots does not represent the same class.

6 Hyperparameter Settings

| Hyperparameter                          | Settings of hyperparameters       |
|----------------------------------------|------------------------------------|
| Layer initialization                   | random normal initialization        |
|                                        | for weights and biases              |
| Optimizer                              | Adam optimizer                      |
| Activation function                    | ReLU                                |
| Batch size                             | 256                                 |
| Hidden state dimension of representation network | 48/96/10                           |
| Hidden state dimension of branched network | 48/48/96/1                         |
| Hidden state dimension of clustering network | 48/48/96/5/#clusters               |
| $\alpha, \beta$                        | 0.2, 10                             |

References

Karl Pearson. LIII. on lines and planes of closest fit to systems of points in space. 2(11):559–572, 1901.

Jinsung Yoon, Ahmed M. Alaa, Martin Cadeiras, and Mihaela van der Schaar. Personalized donor-recipient matching for organ transplantation. In Thirty-First AAAI Conference on Artificial Intelligence, 2017.