Few-Shot Cross Domain Battery Capacity Estimation

Zihao Zhou
zhouzh18@mails.tsinghua.edu.cn
Tsinghua-Berkeley Shenzhen Institute
Shenzhen, Guangdong, China

Aihua Ran
Tsinghua-Berkeley Shenzhen Institute
Shenzhen, China
rah19@mails.tsinghua.edu.cn

Shuxiao Chen
Tsinghua-Berkeley Shenzhen Institute
Shenzhen, China
csxiao@sz.tsinghua.edu.cn

Guodan Wei
Tsinghua-Berkeley Shenzhen Institute
Shenzhen, China
weiguodan@sz.tsinghua.edu.cn

Hongbin Sun
Tsinghua University
Beijing, China
shb@mail.tsinghua.edu.cn

Xuan Zhang*
Tsinghua-Berkeley Shenzhen Institute
Shenzhen, Guangdong, China
xuanzhang@sz.tsinghua.edu.cn

Yang Li*
Tsinghua-Berkeley Shenzhen Institute
Shenzhen, Guangdong, China
yangli@sz.tsinghua.edu.cn

ABSTRACT

Accurate capacity estimation is essential in a board range of battery applications. Because of the highly nonlinearity in the battery aging mechanism, recent works employ many supervised learning methods, which assume training and testing battery samples are generated from the same sample distribution. However, it is common for different battery data sets to have some extent of distribution shifts caused by different battery sizes, testing environments and historical load patterns. In this paper, we consider the scenario when only a few of labeled samples from the testing data set are available and formulate the battery estimation problem as a semi-supervised transfer learning problem. Inspired by JDOT, an unsupervised joint distribution domain adaptation algorithm based on optimal transport, we propose Semi-JDOT for regression problems where the source and the target label distributions have unequal supports. Our approach incorporates prior information of the labeled target samples as additional constraints and can be solved analytically. We conduct comprehensive experiments on a number of distinct battery data sets. The results show that the proposed approach outperforms existing supervised and semi-supervised methods by 10-30% under various few-shot experiment settings.

CCS CONCEPTS

• **Computing methodologies** → **Semi-supervised learning settings**; • **Applied computing** → **Electronics**.

KEYWORDS

optimal transport; battery capacity estimation; semi-supervised learning

UbiComp-ISWC '21 Adjunct, September 21–26, 2021, Virtual, USA

ACM ISBN 978-1-4503-8461-2/21/09…$15.00
https://doi.org/10.1145/3460418.3480409

ACM Reference Format:
Zihao Zhou, Aihua Ran, Shuxiao Chen, Guodan Wei, Hongbin Sun, Xuan Zhang, and Yang Li. 2021. Few-Shot Cross Domain Battery Capacity Estimation. In *Adjunct Proceedings of the 2021 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2021 ACM International Symposium on Wearable Computers (UbiComp-ISWC '21 Adjunct), September 21–26, 2021, Virtual, USA.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3460418.3480409

1 INTRODUCTION

Lithium-ion batteries are widely used in various kinds of applications from Electrical Vehicles (EV) to Energy Storage Systems in power grids [3, 9, 17]. Having a reliable estimation about the battery performance is essential in these applications. For example, an inaccurate estimation of the battery capacity results in overcharging, which may in turn accelerate battery aging or even cause spontaneous combustion [12]. Traditionally, the time-consuming Constant Current Constant Voltage (CC-CV) test is used to determine the battery capacity. Though it is the most accurate test for battery capacity, it costs several hours to test each battery, thus it is not suitable for real applications where hundreds of batteries need to be tested in a short amount of time.

Many scholars have used supervised learning to predict battery capacity from fast-to-obtain battery test signals, such as current, voltage and temperature during charging-discharging cycles [11, 13, 16, 22]. However, all these methods, relying heavily on the training data, can not generalize well to battery data collected from different experiment settings. As battery cells often experience different load patterns and have different aging paths, their training and testing data distributions often differ, also known as a domain shift. As a result, supervised learning methods trained on a battery data set from one specific experiment setting can not be used for prediction on other battery data sets with different experiment settings. Instead, these models need to be retrained from scratch for every different battery data set. What's worse, retraining the same model is quite time-consuming as the labels need to be generated from the CC-CV test. Hence it is critical to improve the cross-domain generalization ability in battery capacity estimation.

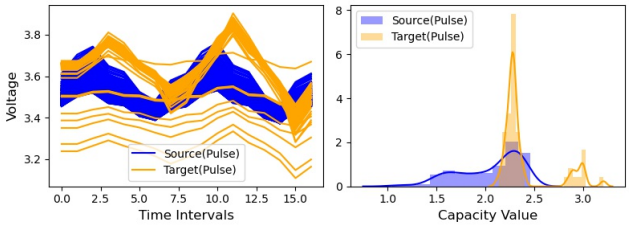

Figure 1: Empirical distributions for Pulse data set.

A common approach to deal with domain shift is transfer learning (domain adaptation), which aims at finding a model that works well on both the source data set and the target data set. Joint Distribution Optimal Transport (JDOT) proposed in [5] is an unsupervised domain adaptation method based on optimal transport. Optimal transport based approaches are suitable for the cross-domain capacity estimation problem, as they impose no distribution assumption on the data and can handle unequal support of the source and target feature distributions. Unlike earlier optimal transport methods that only consider the transport between marginal feature distributions [6], JDOT tries to minimize a total transport cost between the joint distributions from source and target domains. However, as JDOT is designed for unsupervised domain adaptation, it does not use any prior information of the target labels, which will be a problem when the target label distribution has a different support from the source. As shown in Figure 1, the source and target battery sets not only have differences in the marginal feature distribution but also in the marginal label distribution. Thus, merely using JDOT is not suitable for the cross domain battery capacity estimation.

To tackle the aforementioned challenges, we propose a semi-supervised joint distribution optimal transport model. To be specific, we regard cross domain battery capacity estimation as a semi-supervised transfer learning problem. And only a few labeled samples in the target domain are required in our proposed method. Note that it takes several hours to generate a required label, so that nearly dozens or hundreds of hours are needed to generate enough labels for training supervised learning models from scratch. We extend JDOT by adding additional constraints provided by the labeled target data and derive the optimal solution to the corresponding modified optimization problem.

The main contribution for this paper is summarized as following:

(1) A semi-supervised JDOT model is proposed. It needs only a few labels (4-8) in the target data set, which largely reduces the required label generation time and cost.
(2) The proposed semi-JDOT model can effectively address the domain shift problem in both feature and label spaces.
(3) Comprehensive experiments on two pairs of different battery data sets are conducted. The results demonstrate the effectiveness of our proposed methods under standard and few-shot settings.

2 METHOD

In this section, we begin with defining the cross-domain battery capacity estimation problem. Then we introduce the JDOT model for unsupervised domain adaptation. Finally we give the detailed explanation about our proposed semi-supervised extension for regression problems.

2.1 Problem Formulation

In cross-domain battery capacity estimation, we are given a source data set $S = \left\{\left(x_1^s, y_1^s\right), \cdots, \left(x_{n_s}^s, y_{n_s}^s\right)\right\}$ of n_s samples generated from controllable lab tests, such that their capacity values are known. We use $x_i^s \in \mathcal{X}$ and $y_i^s \in \mathbb{R}$ to denote the features (voltage curve) and the label (capacity value) of the i-th sample respectively. Our goal is to learn a prediction function for the target data set $T = \left\{\left(x_1^t, y_1^t\right), \cdots, \left(x_{n_t}^t, y_{n_t}^t\right)\right\}$ where x_i^t is known for each sample, but only $l << n_t$ of them have the corresponding capacity measurements y_i^t. For simplicity, we assume the target samples are ordered such that $y_1^t \ldots y_l^t$ are known and $y_{l+1}^t \ldots y_{n_t}^t$ are unknown. As a result, the training set includes the whole source data set S and the l samples with known labels $y_1^t \ldots y_l^t$ in the target data set T. While the test set is the remaining samples with unknown labels $y_{l+1}^t \ldots y_{n_t}^t$ in the target data set T.

Let $P_S = P(x^s, y^s)$ and $P_T = P(x^t, y^t)$ be the joint distributions that respectively generate the source and target data sets. Because of the differences in experiment conditions, historical load patterns and battery sizes [1, 25], we have that $P_S \neq P_T$. In addition, the marginal distributions $P(x^s)$ and $P(x^t)$ may have different support, and so are $P(y^s)$ and $P(y^t)$, as shown in Figure 1.

Further, we assume there is a natural probabilistic coupling[1] γ between joint distributions P_s and P_t due to the similar aging characteristics of the source and target batteries. The goal of Semi-JDOT is to find γ and the target prediction function $f : \mathcal{X} \to \mathbb{R}$ through solving the optimal transport between the joint distributions P_S and P_t, while making sure that f correctly predicts the labeled target samples, i.e. $y_i^t = f(x_i^t)$ for all labeled samples $(i = 1, \ldots, l)$.

2.2 Joint Distribution Optimal Transport

As proposed in [5], JDOT is a framework for unsupervised domain adaptation between joint distributions, i.e. all n_t target samples are unlabeled. Given a distance metric $d(x, x')$ on $\mathcal{X}$ and a loss function L, the overall minimization formulation is shown below:

$$\gamma^* = \underset{\gamma \in \prod(P_s, P_t)}{\arg\min} \int D\left(x^s, y^s; x^t, y^t\right) d\gamma\left(x^s, y^s; x^t, y^t\right). \quad (1)$$

where $D\left(x^s, y^s; x^t, y^t\right) := \alpha d\left(x^s, x^t\right) + L\left(y^s, y^t\right)$ is a joint distance measure between the source and target joint distributions, α is a positive parameter which balances the metrics in the feature space and the label space. And γ^* is the optimal transport plan which minimizes the whole cost function. Since $P(y^t)$ is unknown in the unsupervised setting, JDOT constructs a proxy target joint distribution $P_t^f = P(x_t, f(x_t))$ that depends on the estimated function f. Note that, to learn the optimal coupling from empirical data,

[1] Formally, $\gamma \in \Pi(P_s, P_t)$ where $\Pi(P_s, P_t)$ is the set of all randon coupling of P_S and P_T.

the problem can be further written in the following form:

$$\min_{f\in\mathcal{H},\gamma\in\Pi(P_s,P_t^f)} \sum_{i,j}\gamma_{i,j}\left(\alpha d\left(x_i^s,x_j^t\right)+L\left(y_i^s,f\left(x_j^t\right)\right)\right)+\lambda\Omega(f) \quad (2)$$

where f is assumed to be in a RKHS, $\Omega(f)$ is a regularization term and λ is a positive parameter which balances the loss and the complexity of function f. JDOT uses a Block Coordinate Descent method to gain the optimal f and γ. Specifically, when f is fixed, the problem boils down to a classical Optimal Transport (OT) problem that can be easily solved by existing solvers such as the Sinkhorn algorithm and stochastic algorithm [7, 10]. On the other hand, when γ is fixed, the problem can be simplified as:

$$\min_{f\in\mathcal{H}} \sum_{i,j}\gamma_{i,j}L\left(y_i^s,f\left(x_j^t\right)\right)+\lambda\Omega(f). \quad (3)$$

Furthermore, for regression problems which regard L as the squared-loss, Equation (3) can be further simplified into a regularized least-square problem:

$$\min_{f\in\mathcal{H}} \sum_{j}\frac{1}{n_t}\left\|\hat{y}_j-f\left(x_j^t\right)\right\|^2+\lambda\|f\|_{\mathcal{H}}^2 \quad (4)$$

where $\hat{y}_j = n_t\sum_i \gamma_{i,j}y_i^s$, that is, the predicted target label is expected to approximate the weighted average of source labels, with weights given by the optimal transport plan γ.

Though JDOT has a good performance on multiple open data sets under the unsupervised transfer learning setting, it can not handle distribution shift in the label space, as it is designed for an unsupervised domain adaptation problem. Without prior information of the target labels, it assumes that the labeling function (conditional probability $P(y \mid x)$) the same between source and target domains under an optimal coupling. However, this assumption may not hold in practice, as illustrated by Figure 2. Based on the original JDOT, the optimal coupling is a horizontal displacement, while given prior knowledge of the target labels, the coupling that gives the correct target prediction is a displacement to the upper right in Figure 2.

2.3 Semi-Supervised Joint Distribution Optimal Transport

To resolve the limitation caused by label distribution shift between the source and the target labels, we revise the basic JDOT model into a semi-supervised model.

Given l known labels $(x_1^t,y_1^t),\ldots,(x_l^t,y_l^t)$ in the target domain, the predictor f is expected to generate correct predictions on known samples. Hence equality constraints are added to Equation (1):

$$\begin{aligned}&\min_{f\in\mathcal{H},\gamma\in\Pi(P_s,P_t^f)} \sum_{i,j}\gamma_{i,j}\left(\alpha d\left(x_i^s,x_j^t\right)+L\left(y_i^s,f\left(x_j^t\right)\right)\right)+\lambda\Omega(f)\\ &\text{s.t. } y_j^t=f\left(x_j^t\right), \text{ for all } 1\leqslant j\leqslant l\end{aligned} \quad (5)$$

Similar to JDOT, we use Block Coordinate Descent to solve for the optimal coupling matrix γ and prediction function f. When f is fixed, the optimization with respect to γ mostly remains the same. We use the Sinkhorn algorithm [7] to solve the following entropy-regularized optimal transport problem:

$$\min_{\gamma\in\Pi(P_s,P_t^f)} \sum_{i,j}D_f(x_i^s,x_j^t,y_i^s,y_j^t)\gamma_{ij}-\lambda_h Ent(\gamma), \quad (6)$$

where $Ent(\gamma)$ is the Shannon entropy regularizer. When computing the cost function D_f, we replace the proxy target label $f(x_j^t)$ by the actual label y_j^t of the labeled samples:

$$D_f(x_i^s,x_j^t,y_i^s,y_j^t)=\begin{cases}\alpha d(x_i^s,x_j^t)+L(y_i^s,y_j^t) & \text{if } j\leq l\\ \alpha d(x_i^s,x_j^t)+L(y_i^s,f(x_j^t)) & \text{if } l<j\leq n_t\end{cases}$$

When the transport plan γ is fixed, equality constraints that restrict the prediction function f to approximate the real target label are added to the optimization problem for f:

$$\begin{aligned}&\min_{f\in\mathcal{H}}\sum_{i=1}^{n_t}\frac{1}{n_t}\left\|\hat{y}_i-f\left(x_i^t\right)\right\|^2+\lambda\|f\|_{\mathcal{H}}^2\\ &\text{s.t. } y_j^t=f\left(x_j^t\right), \text{ for all } 1\leqslant j\leqslant l.\end{aligned} \quad (7)$$

The assumption that f belongs to a RKHS with kernel function k allows us to rewrite Equation (7) as follows:

$$\begin{aligned}&\min_{c}\sum_{i=1}^{n_t}\frac{1}{n_t}\|\hat{y}_i-\sum_{j=1}^{n_t}c_j k\left(x_i^t,x_j^t\right)\|^2+\lambda\|c\|_{\mathcal{H}}^2\\ &\text{s.t.}\quad y_i^t=\sum_{j=1}^{n_t}c_j k\left(x_i^t,x_j^t\right) \text{ for all } 1\leq i\leq l.\end{aligned} \quad (8)$$

In practice, we choose k to be the RBF kernel $k\left(x_i^t,x_j^t\right)=e^{-\Gamma\|x_i^t-x_j^t\|_2}$ where hyperparameter Γ is learned by cross validation. And our goal changes to learn an optimal parameter vector c^* while satisfying the equality constraints. Notice that, this problem can be further written into a matrix form, with the Karush–Kuhn–Tucker conditions to gain an analytical solution. The corresponding Lagrangian function is expressed as:

$$L=\left(\hat{Y}-Kc\right)^\top\left(\hat{Y}-Kc\right)+\lambda c^\top Kc+\beta^\top(TKc-Y_t). \quad (9)$$

$$\text{where, } \hat{Y}=\begin{bmatrix}\hat{y}_1\\ \vdots\\ \hat{y}_{n_t}\end{bmatrix}_{n_t\times 1}, K=\begin{bmatrix}k(x_1,x_1) & \cdots & k(x_1,x_{n_t})\\ \vdots & \ddots & \vdots\\ k(x_{n_t},x_1) & \cdots & k(x_{n_t},x_{n_t})\end{bmatrix}_{n_t\times n_t}$$

$$T=\begin{bmatrix}I_{l\times l} & 0\\ 0 & 0\end{bmatrix}_{n_t\times n_t},\quad Y_t=\left[y_1^t\ldots y_l^t\ldots 0\right]_{n_t\times 1}^\top$$

where $I_{l\times l}$ is an identity matrix of size l. By setting the first order derivative to zero, we gain the optimal solution c^*:

$$\begin{bmatrix}c^*\\ \beta^*\end{bmatrix}=\begin{bmatrix}2(\lambda I+K) & T^\top\\ TK & 0\end{bmatrix}^{-1}\begin{bmatrix}2\hat{Y}\\ Y_t\end{bmatrix}. \quad (10)$$

Compared with the basic JDOT model, our semi-supervised JDOT model not only aims to find the optimal transport between the source and target joint distributions, but also includes important knowledge from the target domain labels. As shown in Figure 2, predictions on these a few target labels are strongly restricted to fit the real values, which in turn help the predictor generate the result out of the source label range. It is worth noting that the aforementioned formulation places infinite weight to the labeled target instances. In practice there may be a situation when those target labels have large empirical errors. In this case, it is better to

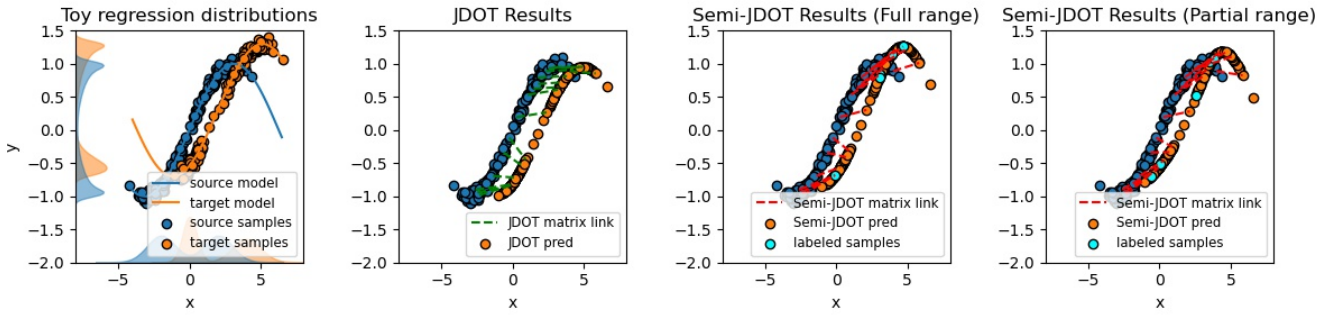

Figure 2: Illustration of JDOT and Semi-JDOT on a 1D sin curve regression problem introduced in [5]. Left: Source and target empirical marginal distributions. Middle left: JDOT prediction results and corresponding OT matrix links. Middle right: Semi-supervised JDOT prediction results and corresponding OT matrix links (labels are randomly sampled across the whole target label range). Right: Semi-supervised JDOT prediction results and corresponding OT matrix links (labels are randomly sampled across within 80% of the target label range). Five labeled target samples are used for both full range and partial range settings.

Algorithm 1: Learning algorithm of Semi-JDOT

Input: source data set S, target data set T;
Output: predictor $f(*)$ on the target data set;
Initialize kernel parameter γ, regularization coefficient λ, cutoff factor $rcond$, balance coefficient α;
$k = 1$;
while *not converged* **do**
  if $k == 1$ **then**
    $\gamma^k \leftarrow$ Solve OT problem only based on x^s and x^t;
  end
  $\gamma^k \leftarrow$ Solve OT problem with fixed f^{k-1};
  Calculate c^* by Equation (10);
  $f^k \leftarrow \quad f^k(x_i^t) = \sum_{j=1}^{n_t} c_j k\left(x_i^t, x_j^t\right)$;
  $k = k + 1$;
end
return f;

replace the equality constraint by a weighted penalty term in the objective function:

$$\min_{f\in\mathcal{H}} \sum_{i=1}^{n_t} \frac{1}{n_t} \left\|\hat{y}_i - f\left(x_i^t\right)\right\|^2 + \mu \sum_{i=1}^{l} \frac{1}{l} \left\|y_i^t - f\left(x_i^t\right)\right\|^2 + \lambda \|f\|_{\mathcal{H}}^2. \quad (11)$$

The choice of coefficient $\mu \geq 0$ and $\lambda \geq 0$ should be adjusted by cross validation. Similar to the equality constraint version, the analytical solution of Equation (11) can also be easily obtained using the kernel trick.

3 EXPERIMENTS

3.1 Data Sets & baseline methods

Self-Test Pulse Battery Data Set: The first data set we use in this paper is a self-test battery data set. While all samples from this data set are extracted from cylindrical 18650 lithium-ion batteries, there are two different groups of samples. One consists of 490 samples and is regarded as the source set. The other one consists of 54 samples and is regarded as the target set. For both of these sets, the capacity values are regarded as labels, while the voltage curves from the pulse test are regarded as features. We assume there are 8 known labels from the target set. Detailed information about this Pulse data set is available in Supplementary A.2.

NASA Public Battery Data Set: The second data set is obtained from the NASA Ames Prognostics Center of Excellence Battery Data Set [19]. The experiments consisted of applying different accelerated aging cycles to a number of commercially available 18650 lithium-ion batteries. Overall there are 41 samples in the source set and 40 samples in the target set. The detailed information is available in Supplementary A.2.

The baseline models can be roughly divided into two different groups, the first one contains supervised learning methods which only use known feature-label pairs to build a predictor on the target data set, such as Kernel Ridge Regression (KRR) and Support Vector Regression. The other one contains semi-supervised learning methods, which not only use known target samples, but also try to extract useful data structures from those samples without labels. As far as we know, there are still limited works that implement semi-supervised learning methods for the application of battery capacity estimation.

Here, for the supervised learning baselines, we use KRR for comparison, which gives state-of-the-art performance in existing works [22]. For the semi-supervised learning baselines, we use Semi-Supervised Regression with Co-training (COREG) [26], and Laplacian Regularized Least Squares (LapRLS) [2] . For domain adaptation methods, we use the basic JDOT model for comparison.

3.2 Experiment Setup

As briefly introduced in Section 3.1, the number of available labels in the target set is limited. In real applications, people still need to decide the way to choose the selected samples for their capacity values. Here we propose two different ways of sampling for the selection of known target labels. One is **Full range sampling** and the other is **Partial range sampling**.

For **Full range sampling**, the known label samples are randomly selected from the whole life range of the battery. Specifically,

Table 1: Results on Pulse data set.

Mean Square Error	100% range	90% range	80% range
Semi-JDOT	**0.048**	**0.069**	**0.074**
KRR	0.070	0.102	0.111
COREG	0.049	0.086	0.096
LapRLS	0.041	0.097	0.119
JDOT	0.202	0.202	0.202

Table 2: Results on NASA data set.

Mean Square Error	100% range	90% range	80% range
Semi-JDOT	**0.0026**	**0.0032**	**0.0024**
KRR	0.0696	0.0898	0.1203
COREG	0.0102	0.0129	0.0241
LapRLS	0.0039	0.0066	0.0557
JDOT	0.0574	0.0574	0.0574

both low capacity and high capacity samples may be selected. This is the normal situation when target set batteries are still in its initial service. These batteries could have either low or high capacity values. And people want to gain an accurate capacity estimation to help them decide later usage patterns. For **Partial range sampling**, the known label samples are randomly selected only from a limited life range of the battery. This range is set to be 90% or 80% of the nominal capacity. Thus, only samples with relatively low capacity values may be selected. This situation happens when target samples mainly come from retired batteries. Normally, batteries used in electrical vehicles are considered retired when their capacity values decrease to 80% of the nominal capacity, or these batteries have already been used beyond their designed period of usage, which should be sent directly for recycling. Thus the case with capacity under 80% is not common in battery capacity estimation and is not discussed here.

3.3 Results

Tables 1 and 2 summarize the results of Semi-JDOT and the baseline models under different data sets and experiment settings. Overall, the unsupervised domain adaptation method JDOT performs the worst since it does not use labeled target data. So we focus on the comparison among other supervised or semi-supervised methods.

Under the Full range sampling setting, all models reach relatively low MSE. Our proposed Semi-JDOT model reaches almost the same (on Pulse data set) or even more than 30% better performance (on NASA data set) compared with the best baseline model. When labeled samples are selected from the lower 80% of the capacity range, the performance of baseline models degrades significantly, while Semi-JDOT gives best results on both data sets.

Another important variable in our experiments is **number of labels in target data set**. We compare the performance of our algorithm with other baseline methods under different number of known target labels for both Pulse and NASA data sets in 3. The 80% partial range setting is used here since it is the more challenging scenario for cross-domain battery capacity estimation. The

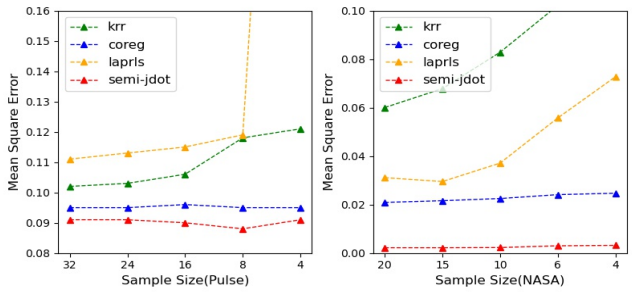

Figure 3: Results for different numbers of known target labels.

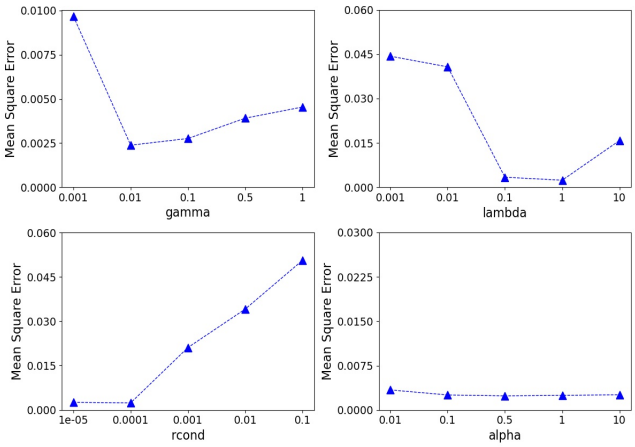

Figure 4: Parameters analysis of Semi-JDOT on NASA data set.

proposed Semi-JDOT method not only surpasses all baselines but also keeps a low and stable mean square error. Especially for supervised learning methods like KRR, it can not give comparable result even 8 times more target labels are available, which indicates that Semi-JDOT can largely reduce the required label generation time. Notice that, the optimal parameters for both Semi-JDOT and other baselines are available in Supplementary A.3. Finally, the determination of hyper parameters highly influence the performance of the proposed model. We analyze how α, Γ, λ and *rcond* would influence the performance of Semi-JDOT. Figure 4 shows the results when varying these parameters. Detailed discussion about parameter determination is available in Supplementary A.3.

4 CONCLUSION

In this paper, we have introduced the cross domain battery capacity estimation problem. Utilizing the common aging mechanisms in batteries, we propose a semi-supervised joint distribution optimal transport method, which can effectively deal with the distribution shifts problem and significantly reduces the time cost for collecting training labels. The experiments results show that our method outperforms baselines by 10-30% under various experiment settings. Possible future works include combining the proposed approach

with meta learning techniques for learning from multiple battery data sets.

Acknowledgement This work was supported in part by Tsinghua-Berkeley Shenzhen Institute Research Start-Up Funding; Shenzhen Science and Technology Program (KQTD20170810150821146); Tsinghua SIGS Scientific Research Start-up Fund (GD2021012C) and Natural Science Foundation of China (62001266).

REFERENCES

[1] Tobias C. Bach, Simon F. Schuster, Elena Fleder, Jana Müller, Martin J. Brand, Henning Lorrmann, Andreas Jossen, and Gerhard Sextl. 2020. Nonlinear aging of cylindrical lithium-ion cells linked to heterogeneous compression. *Journal of Energy Storage* 5 (Jan. 2020), 212–223. https://doi.org/10.1016/j.est.2016.01.003

[2] Mikhail Belkin, Partha Niyogi, and Vikas Sindhwani. 2006. Manifold Regularization: A Geometric Framework for Learning from Labeled and Unlabeled Examples. *Journal of Machine Learning Research* 7, 85 (2006), 2399–2434. http://jmlr.org/papers/v7/belkin06a.html

[3] Silvia Bobba, Fabrice Mathieux, Fulvio Ardente, Gian Andrea Blengini, Maria Anna Cusenza, Andreas Podias, and Andreas Pfrang. 2018. Life Cycle Assessment of repurposed electric vehicle batteries: an adapted method based on modelling energy flows. *Journal of Energy Storage* 19 (Aug. 2018), 213–225. https://doi.org/10.1016/j.est.2018.07.008

[4] Hicham Chaoui, Navid Golbon, Imad Hmouz, Ridha Souissi, and Sofiene Tahar. 2015. Lyapunov-Based Adaptive State of Charge and State of Health Estimation for Lithium-Ion Batteries. *IEEE Transactions on Industrial Electronics* 62, 3 (Aug. 2015), 1610–1618. https://doi.org/10.1109/TIE.2014.2341576

[5] Nicolas Courty, Rémi Flamary, Amaury Habrard, and Alain Rakotomamonjy. 2017. Joint distribution optimal transportation for domain adaptation. In *Advances in Neural Information Processing Systems*, Vol. 30. Curran Associates, Inc., 3730–3739. https://doi.org/abs/1705.08848

[6] Nicolas Courty, Rémi Flamary, Devis Tuia, and Alain Rakotomamonjy. 2017. Optimal Transport for Domain Adaptation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39, 9 (Sept. 2017), 1853–1865. https://doi.org/10.1109/TPAMI.2016.2615921

[7] Marco Cuturi. 2013. Sinkhorn Distances: Lightspeed Computation of Optimal Transport. In *Advances in Neural Information Processing Systems*, Vol. 26. Curran Associates, Inc., 2292–2300. https://proceedings.neurips.cc/paper/2013/file/af21d0c97db2e27e13572cbf59eb343d-Paper.pdf

[8] Bharath Bhushan Damodaran, Benjamin Kellenberger, Rémi Flamary, Devis Tuia, and Nicolas Courty. 2018. DeepJDOT: Deep Joint Distribution Optimal Transport for Unsupervised Domain Adaptation. In *Computer Vision - ECCV 2018 - 15th European Conference, Munich, Germany, September 8-14, 2018, Proceedings, Part IV (Lecture Notes in Computer Science, Vol. 11208)*, Vittorio Ferrari, Martial Hebert, Cristian Sminchisescu, and Yair Weiss (Eds.). Springer, 467–483. https://doi.org/10.1007/978-3-030-01225-0_28

[9] Bruce Dunn, Haresh Kamath, and Jean-Marie Tarascon. 2011. Electrical Energy Storage for the Grid: A Battery of Choices. *Science* 334, 6058 (July 2011), 928–935. https://doi.org/10.1126/science.1212741

[10] Aude Genevay, Marco Cuturi, Gabriel Peyré, and Francis Bach. 2016. Stochastic Optimization for Large-scale Optimal Transport. In *Advances in Neural Information Processing Systems*, Vol. 29. Curran Associates, Inc., 3440–3448. https://proceedings.neurips.cc/paper/2016/file/2a27b8144ac02f67687f76782a3b5d8f-Paper.pdf

[11] Peiyao Guo, Ze Cheng, and Lei Yang. 2019. A data-driven remaining capacity estimation approach for lithium-ion batteries based on charging health feature extraction. *Journal of Power Sources* 412 (2019), 442–450. https://doi.org/10.1016/j.jpowsour.2018.11.072

[12] Stephen J. Harris, David J. Harris, and Chen Li. 2017. Failure statistics for commercial lithium ion batteries: A study of 24 pouch cells. *Journal of Power Sources* 342 (Dec. 2017), 589–597. https://doi.org/10.1016/j.jpowsour.2016.12.083

[13] Chao Hu, Gaurav Jain, Craig Schmidt, Carrie Strief, and Melani Sullivan. 2015. Online estimation of lithium-ion battery capacity using sparse Bayesian learning. *Journal of Power Sources* 289, 6 (May 2015), 105–113. https://doi.org/10.3390/app6060166

[14] Benjamin Jenkins, Ashish Krupadanam, and Anuradha M. Annaswamy. 2020. Fast Adaptive Observers for Battery Management Systems. *IEEE Transactions on Control Systems Technology* 28, 3 (Feb. 2020), 776–789. https://doi.org/10.1109/TIE.2014.2341576

[15] Weihan Li, Yue Fan, Florian Ringbeck, Dominik Jöst, Xuebing Han, Minggao Ouyang, and Dirk Uwe Sauer. 2020. Model-Based Electrochemical Estimation and Constraint Management for Pulse Operation of Lithium Ion Batteries. *Journal of Power Sources* 476 (Aug. 2020), 228534. https://doi.org/10.1016/j.jpowsour.2020.228534

[16] Yuanyuan Li, Hanmin Sheng, Yuhua Cheng, Daniel-Ioan Stroe, and Remus Teodorescu. 2020. State-of-health estimation of lithium-ion batteries based on semi-supervised transfer component analysis. *Applied Energy* 277 (July 2020), 115504. https://doi.org/10.1016/j.jpowsour.2017.05.004

[17] Jeremy Neubauer and Ahmad Pesaran. 2007. The ability of battery second use strategies to impact plug-in electric vehicle prices and serve utility energy storage applications. *Journal of Power Sources* 196, 23 (June 2007), 10351–10358. https://doi.org/10.1016/j.jpowsour.2011.06.053

[18] Sinno Jialin Pan, Ivor W. Tsang, James T. Kwok, and Qiang Yang. 2017. Domain Adaptation via Transfer Component Analysis. *IEEE Transactions on Neural Networks* 22, 2 (Feb. 2017), 199–210. https://doi.org/10.1109/TPAMI.2016.2615921

[19] B. Saha and K. Goebel. 2007. Battery Data Set, NASA AMES Prognostics Data Repository. http://ti.arc.nasa.gov/project/prognostic-data-repository

[20] Kuniaki Saito, Donghyun Kim, Stan Sclaroff, Trevor Darrell, and Kate Saenko. 2010. Model-Based Electrochemical Estimation and Constraint Management for Pulse Operation of Lithium Ion Batteries. *IEEE Transactions on Control Systems Technology* 18, 3 (Sept. 2010), 654–663. https://doi.org/10.1109/TCST.2009.2027023

[21] Kuniaki Saito, Donghyun Kim, Stan Sclaroff, Trevor Darrell, and Kate Saenko. 2019. Semi-supervised Domain Adaptation via Minimax Entropy. *2019 IEEE/CVF International Conference on Computer Vision (ICCV)* (Feb. 2019), 8049–8057. https://doi.org/10.1109/ICCV.2019.00814

[22] Sheng Shen, Mohammadkazem Sadoughi, Xiangyi Chen, Mingyi Hong, and Chao Hu. 2019. A deep learning method for online capacity estimation of lithium-ion batteries. *Journal of Energy Storage* 25 (2019), 100817. https://doi.org/10.1016/j.est.2019.100817

[23] Fariborz Taherkhani, Hadi Kazemi, Ali Dabouei, Jeremy Dawson, and Nasser M Nasrabadi. 2020. Matching Distributions via Optimal Transport for Semi-Supervised Learning. *arXiv preprint arXiv:2012.03790* (2020).

[24] Yuguang Yan, Wen Li, Hanrui Wu, Huaqing Min, Mingkui Tan, and Qingyao Wu. 2018. Semi-Supervised Optimal Transport for Heterogeneous Domain Adaptation.. In *IJCAI*, Vol. 7. 2969–2975.

[25] Xiao-Guang Yang, Yongjun Leng, Guangsheng Zhang, Shanhai Ge, and Chao-Yang Wang. 2017. Modeling of lithium plating induced aging of lithium-ion batteries: Transition from linear to nonlinear aging. *Journal of Power Sources* 360 (Aug. 2017), 28–40. https://doi.org/10.1016/j.jpowsour.2017.05.110

[26] Zhi-Hua Zhou and Ming Li. 2005. Semi-Supervised Regression with Co-Training. In *Proceedings of the 19th International Joint Conference on Artificial Intelligence*. Morgan Kaufmann Publishers Inc., 908–913. https://doi.org//dl.acm.org/doi/10.5555/1642293.1642439

A SUPPLEMENTARY

A.1 Related Works

A.1.1 Battery Capacity Estimation. For the topic of battery capacity estimation, the existed literature can be roughly divided into model based methods and machine learning methods. For the model based methods, people use either electrochemical models [15, 20] or equivalent circuit models [4, 14] to represent the battery inner mechanism. Though these models have shown great success and enable some physical interpretation of the battery aging behaviour, these types of models are usually difficult to establish given the strict requirements for knowledge of the electrochemical parameters, aging mechanisms, and battery properties.

Machine learning methods are gaining more and more attention, as it has fewer requirements for deep theoretical knowledge about battery chemistry and aging mechanisms. A study [13] employed Relevance Vector Machine (RVM) as a probabilistic kernel regression method to learn the dependency of capacity on five selected characteristic features. Though this model gives good results on the continuous cycling data and still maintains a high interpretability, it needs relatively deep domain knowledge about battery to select effective features.

To avoid manually feature selection, a recent study [11] proposed a data-driven approach to achieve health feature extraction. Specifically, it utilized rational analysis and principal component analysis to extract useful features from the charging profiles. Then an adaptive RVM was implemented to gain the capacity estimation based on the extracted features. It emphasized that the used health features were effective under different battery operating conditions. Moreover, another recent study [22] used a deep convolutional neural network to automate the feature learning process on battery charge curves, which enabled reasonable outputs for new inputs that had never been seen during the training process.

However, as these methods are all supervised learning methods, they need to be retrained on every battery cell once its aging trajectory or experiment condition is different from the one used for former model training. Furthermore, retraining these models requires enough capacity values, which come from the time-consuming CC-CV test. It can be very difficult to generate a large amount of capacity values, which is especially challenging for deep learning models.

To overcome the drawbacks of supervised learning models, a recent work [16] proposed a battery health assessment model based on the semi-supervised transfer component analysis (SSTCA) algorithm. Specifically, it aimed to minimize the distance between sample distributions collected in different stages of the battery, so that a good predictor could be built only based on the combination of early usage data and several future usage samples. However, as in the setting of SSTCA [18], the labeled and unlabeled data are from different domains, and only the label values from source domain have influence on the final mapping of the new features, it can not address the domain shift in the label space between the source and target domains. In the proposed Semi-JDOT method, both the domain shift in feature and label spaces have been addressed.

A.1.2 Optimal Transport for Domain Adaptation. Optimal transport was first used for domain adaptation by Courty et. al. [6]. The proposed algorithm, Optimal transport domain adaptation (OTDA) regarded the domain adaptation problem as a graph matching problem, in which each source sample was mapped to target samples under the constraint of marginal distribution preservation. This method follows the covariate shift assumption, i.e. the conditional distributions of label (given input features) is the same between domains. Later works have relaxed this assumption by matching the joint distributions, such as the JDOT method [5] mentioned previously and its deep neural network extension, DeepJDOT [8]. The latter work introduces an end-to-end framework that can learn feature embeddings using CNN and perform joint optimal transport in embedded space. While DeepJDOT outperforms JDOT on image classification tasks, we still stemmed our method from its predecessor as capacity estimation data has a simpler structure that can be mostly represented by a shallow model. It is also easier to analyze analytically.

Table 3: Experiment Setting for NASA data set.

NASA	Source Set	Target Set
Battery ID	6/18/55	29/33
Discharge Pattern	2A/2.5V	4A/2V
Sample Quantity	41	40
Feature Dimension	50	50

A.1.3 Semi-Supervised Domain Adaptation. OTDA [6] can deal with the semi-supervised setting by constraining the transport matrix such that only source and target samples with same labels can be matched. This class-constrained idea is also used in other semi-supervised optimal transport methods to incorporate target labels. For example, [24] studied the semi-supervised transport between domains with heterogeneous feature spaces, while [23] uses optimal transport to produce pseudolabels in semi-supervised learning.

Moreover, a recent study [21] proposed a novel Minimax Entropy approach that adversarially optimized an adaptive few-shot model for the semi-supervised classification problem. Specifically, this model consists of a feature encoding network that computes the features' similarity to a set of estimated prototypes (representatives of each class). Then, adaptation is achieved by alternately maximizing the conditional entropy of unlabeled target data with respect to the classifier and minimizing it with respect to the feature encoder.

Though these methods gain state-of-the-art performance on many public data sets, they are designed for semi-supervised classification problem and thus can not be easily transformed to deal with regression problems.

A.2 Details about Data Sets

Self-Test Pulse Battery Data Set: Briefly speaking, the pulse test is an effective method for evaluating the consistency among batteries [20]. As it only takes 2 minutes for testing each battery cell and the test results still contain effective information about battery performance. Finding the corresponding mapping between the pulse test curve and capacity values can largely decrease the required time for battery capacity estimation.

Table 4: Experiment Setting for Pulse data set.

Pulse	Source Set	Target Set
Nominal Capacity	2.5Ah	3.2Ah
Sampling Interval	2 cycles	random
Sample Quantity	490	54
Feature Dimension	16	16

Table 5: Parameter choice for Semi-JDOT.

Parameters	Optimal values (Pulse)	Optimal values (NASA)
α	1	0.5
Γ	0.01	0.01
λ	0.1	1
rcond	1e-4	1e-4

For the source set, batteries with 2.5Ah nominal capacity have experienced predefined fast-aging cycles from new to 50% nominal. The pulse test and the CC-CV test are performed for every 2 aging cycles to gain feature-label pairs across the whole battery life range.

For the target set, batteries have experienced the same predefined fast-aging cycles. However, the nominal capacity is 3.2Ah. The pulse test and the CC-CV test are performed in random intervals to gain feature-label pairs across the whole battery life range.

NASA Public Battery Data Set: Specifically, for battery samples in NASA data set, charging patterns were kept identical for all batteries under the CC-CV setting, while discharging patterns were different. Here we use the No. 6,18,55 batteries to form the source data set, and the No. 29,33 to form the target data set. As for each battery, NASA provides data for the whole aging process at cycle level, and we take a sample every five cycles. As a result, there are 41 samples in the source set and 40 samples in the target set.

Notice that, normally, capacity values need to be derived under the same discharge pattern so that they can be compared to each other. In our design, while discharge patterns for source and target domains are different, within the target set, those capacity values are comparable. For both Pulse and NASA data sets, there exists domain shift problem in the marginal distributions between source and target samples.

A.3 Supplementary Results & Analysis

Effect of α. As mentioned in Section 2.3, α is a positive parameter in the joint cost measure term D, which balances the losses from the feature space to the label space. This parameter does not effectively influence the model performance on our two battery data sets. So we use the default setting discussed in the JDOT paper [5].

Effect of Γ. Γ is the bandwidth parameter in the chosen Gaussian kernel. As Γ defines how far the influence of a single training example reaches, a small Γ will lead in every single training example has a large influence on the prediction, while a large Γ only allows nearby sample points to influence the prediction. So, there exists a optimal range for Γ, beyond which, the model performance will decrease.

Effect of λ. λ is the coefficient of the regularization term. As for λ, it controls the trade-off between cost minimization and model complexity. A small λ gives a model that fits very well to the training samples but gain bad performance on unseen test samples. A large λ will cause the problem of underfitting, which gives bad predictions on both training and test samples. So, there exists a optimal vaule for λ.

Effect of *rcond*. Finally, *rcond* is the cutoff factor for small singular values in matrix pseudo-inverse calculation. Singular values less than (*rcond* $\times$ largest singular value) will be considered as zeros. Normally, this parameter should be the smaller the better. However, there exists a threshold value, below which the calculation accuracy is enough to give good results, while a smaller *rcond* causes extra calculation burden and do not improve the model performance. $rcond = 0.0001$ is enough to give a good result.

Notice that the source codes are avaliable at:
https://github.com/Zhouzihao914/Few-shot-Cross-Domain-Battery-Capacity-Estimation.

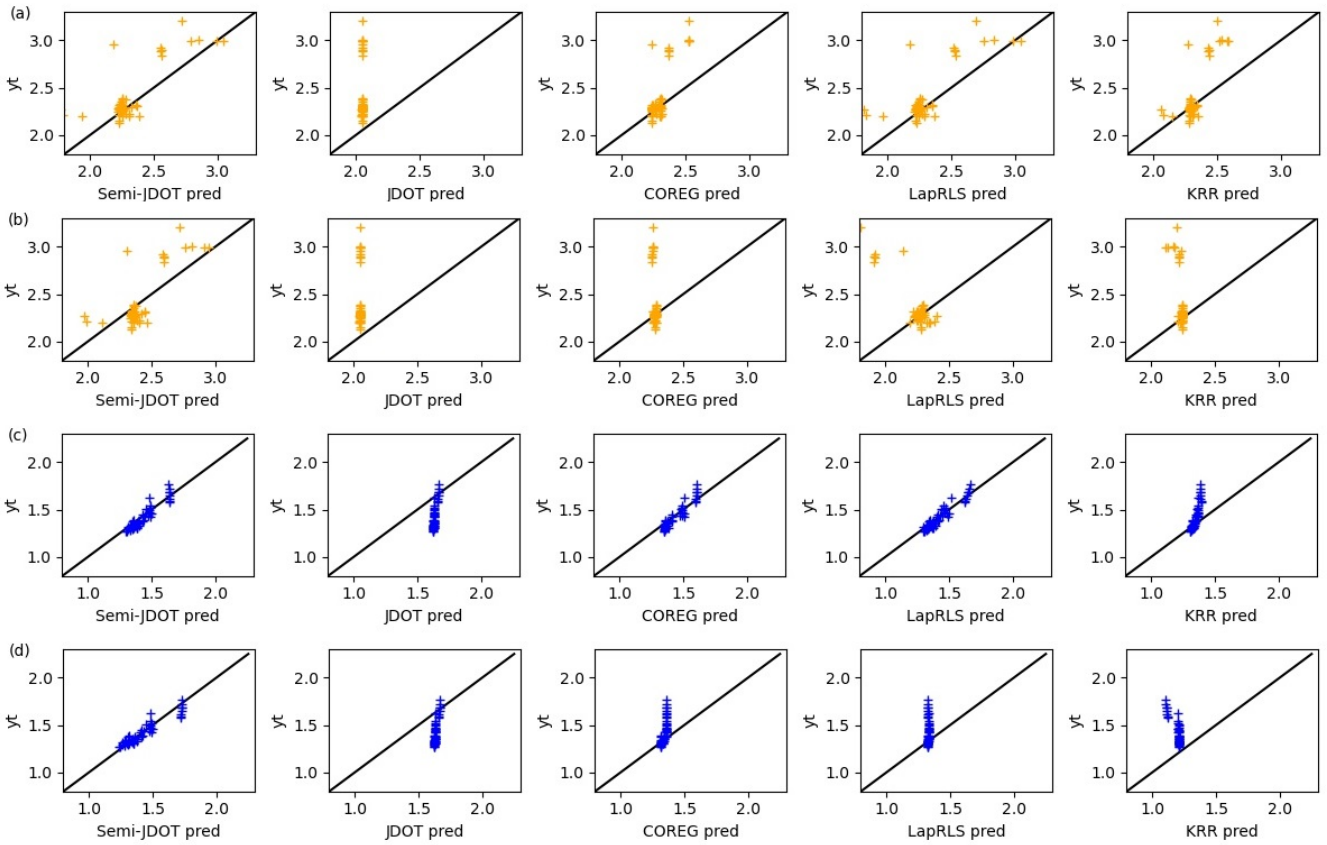

Figure 5: Visualization of typical results on both NASA and Pulse data sets. (a) Pulse data set results under random sampling on full label range. (b) Pulse data set results under random sampling on 80% label range. (c) NASA data set results under random sampling on full label range. (d) NASA data set results under random sampling on 80% label range.