Pooled Wisdom, Unique Insights: Improving Customer Predictions with Cooperative Databases and Deep Transfer Learning

Sumon Chaudhuri
WU Vienna
Arnaud De Bruyn
ESSEC Business School

Cite as:
Chaudhuri Sumon, De Bruyn Arnaud (2025), Pooled Wisdom, Unique Insights: Improving Customer Predictions with Cooperative Databases and Deep Transfer Learning. *Proceedings of the European Marketing Academy*, 54th, (126418)

Paper from the 54th Annual EMAC Conference, Madrid, Spain, May 25-30, 2025

EUROPEAN MARKETING ACADEMY

Pooled Wisdom, Unique Insights:
Improving Customer Predictions with
Cooperative Databases and Deep Transfer Learning

Abstract — This paper explores the effectiveness of cooperative (co-op) databases in improving customer behavior prediction through transfer learning. Co-op databases are a recent concept that enables firms to pool anonymized data managed by third parties, enhancing insights without compromising data privacy. We study fourteen charitable organizations, utilizing transfer learning to create individual predictive models from shared data while ensuring confidentiality. Our findings reveal that pooling data can improve prediction accuracy and is particularly beneficial for firms with limited data or experiencing strategic shifts or market disruptions (i.e., models calibrated on pooled data are more resilient to covariate shifts and external shocks). This research underscores the importance of co-op databases in predictive analytics. It demonstrates how machine learning can be applied to maximize the benefits of data sharing across diverse operational contexts.

Keywords — *Cooperative databases; Predictive modeling; Transfer learning.*

Track — *Methods, Modelling & Marketing Analytics*

1. Introduction

Firms increasingly seek deeper customer insights by participating in cooperative (co-op) databases. These large, pooled datasets aggregate contributions from multiple firms, enabling more accurate predictions while safeguarding privacy through third-party intermediaries (e.g., Data Axle, Verisk). These intermediaries provide curated insights, models, or summary statistics, eliminating the need for firms to directly share sensitive customer data.

This paper shows how co-op databases can also solve some long-standing challenges in customer behavior prediction. For instance, it is very difficult for a company to predict how its customers will behave on a sales channel they have never relied on before. Companies with poor data (limited in time history or purchase volume) may experience challenges in building a predictive model. Likewise, if a firm considers radically shifting its targeting strategy or changing the number of solicitations it sends to its customers, past data will provide limited insights into how such a change in marketing pressure will affect customers. To solve these challenges, we train firm-specific models that harness the pooled data of co-op databases while preserving unique customer behaviors visible in the firm's own data, using transfer learning. While past behavior often predicts future outcomes, co-op databases offer a way to address situations where the future diverges from historical trends, using one firm's past to inform another's future.

2. Literature Review

2.1 Sequential targeting

Firms have always faced the challenge of targeting their customers in the best way possible. Hence, long-term sequential targeting has emerged as a significant stream of research over the years (for example, Simester et al., 2006; Durango-Cohen et al., 2013; Van Diepen et al., 2009). In recent years, deep learning has emerged as a methodology of choice that can improve long-term sequential targeting. Recurrent neural networks like the LSTM model (Hochreiter & Schmidhuber, 1997) can accurately predict long-term customer behavior from high-dimensional and complex data without manually-engineered features (Sarkar & De Bruyn, 2021; Valendin et al., 2022). The highly accurate (albeit opaque) prediction of customer behavior from these models can help firms gauge responses to marketing activity and find improved targeting strategies. Improvement in targeting can be achieved by applying managerial heuristics to the predictions or using a modeling approach, like deep reinforcement learning (DRL), to find optimal solutions (Wang et al., 2022).

A co-op database provides a unique solution to this age-old sequential targeting problem. By going beyond the data we have at our disposal, we can build models that are more robust to limitations in our data set. Moreover, by using transfer learning to marry the diversity of the co-op database with the nuances and idiosyncrasies of firm-specific customer behavior, we can build models less susceptible to covariate shifts than most approaches in the literature that rely exclusively on firm data. For example, if a firm decides to adopt a never-explored-before solicitation strategy, most models trained exclusively on firm data will not be able to predict long-term customer behavior accurately. However, a model that benefits from the shared knowledge of a co-op database may be more resilient to these covariate shifts. Our approach will enhance the robustness of recent sequential targeting optimization algorithms (Wang et al., 2022) that rely on accurate predictions of long-term customer behavior.

3. Modeling Framework

3.1 Overview

In this paper, all the competing models we will discuss share the same architecture, whether the model is calibrated on proprietary data, pooled data, or a mix of both, hence ensuring that any observed difference cannot be attributed to differences in model specifications or architecture.

We first discuss the general architecture of the model and then discuss the various calibration strategies, where transfer learning will be more specifically discussed.

3.2 LSTM model for long-term customer behavior prediction

At the heart of each model is a LSTM neural network. The LSTM has been previously shown to be particularly effective in predicting long-term customer behavior (Sarkar & De Bruyn, 2021; Valendin et al., 2022; Wang et al., 2022). Hence, we considered it to be the ideal choice for our tests.

Architecturally, the models comprise four components or layers, as described:

- L0: A fully connected layer that initializes the LSTM model based on donor characteristics.
- L1: An LSTM layer that processes donation sequences
- L2: A fully connected layer between the LSTM and the final layer, added specifically for the purpose of transfer learning
- L3: A final, fully connected layer whose output nodes feed into the loss function.

We use different activation functions for each layer. The outputs of L1 and L2 are passed through a sigmoid activation function. Our models, unlike traditional LSTM models in the marketing literature, are multi-task learning models, i.e., they simultaneously make predictions for donation probabilities *and* donation amounts. Hence, L3 consists of two output nodes. The first node is used to model donation probabilities, whereas the second node is used to model donation amounts. The former is passed through a sigmoid activation function, and the latter is passed through a ReLU activation function.

During the training process, loss is calculated as the sum of BCE loss from predicting donation probabilities, and the MSE loss from predicting log of donation amounts in instances where donations have been made.

To train the model, we use the Adam optimizer (Kingma & Ba, 2014) and tune hyperparameters of learning rate, weight decay, number of hidden nodes of the LSTM layer, and batch size using Bayesian optimization (Snoek et al., 2012). We use a combination of weight decay and early stopping to prevent overfitting during the training process.

Note that the first initialization layer is optional, whereas the fourth layer is specific to our multi-objective loss function. In the next section, we focus our attention on the two central layers of the model and how we calibrate them.

3.3 Estimation strategies

We calibrate three categories of models:

1) Pooled model: This model is trained on a co-op database, i.e., data pooled from several firms. There is only one pooled model, regardless of the number of participating firms. In this scenario, each participating firm receives predictions from the same model.

2) Fine-tuned model: In this model, we first estimate the pooled model on a co-op database, as in the previous scenario. We then extract the shared LSTM component and, in a second model calibration phase, use each participating firm's proprietary data to fine-tune L2. Our hypothesis is that such an approach will retain the "shared wisdom" of the co-op database (in the LSTM layer) while allowing each firm to fine-tune its predictions to the specificities of its unique customer database. While the LSTM component is shared across participating firms, there are as many fine-tuned models as there are partners.
3) Standalone model: This is the typical approach used by firms when they train a predictive model: they rely exclusively on the data they have at their disposal. There are as many standalone models as there are participating partners.

4. Data

We collected the data from 14 charities that operate all across the world, supporting a variety of causes. Names of the charities are presented in a codified manner for brevity. We randomly select 10 out of these 14 charities (VVA, RESTO, BA, AZV, MUCO, ASF, FBB, VSF, SENSO, and JDO) to build our co-op database and leave four additional charities on the side for testing and simulation purposes (LIBAN, ESPOIR, CREE, and SNDL). The charities are fairly homogeneous with respect to the geographical markets they operate from but vary substantially in terms of the size of the donor base and active timeline.

5. Empirical Results

5.1 Performance for participating firms

We randomly split the data into training (80%), validation (10%), and test (10%). We discard the last 24 months of the training and validation data to test predictive accuracy of long-term behavior. Since it takes time for charities to stabilize their customer base and develop consistent targeting strategies, predictions made during the customer acquisition phase are often noisy, i.e., high in variance. Consequently, we report in-sample fit of the models only on the last 24 months of the training data (i.e., the last 25-48 months of the original data).

We summarize the data strategy in Figure 1.

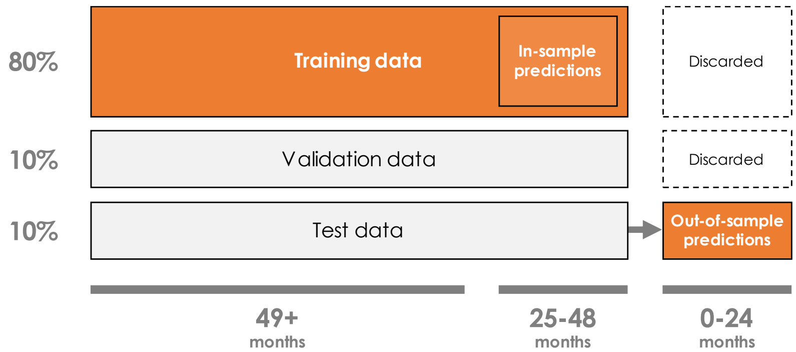

Figure 1 – Split of data into training, validation, and test, as well as in-sample and out-of-sample prediction periods.

Table 1 shows the confusion matrix and RMSE performance of the three models across the ten charities in the co-op database. The three models are quite similar in terms of predictive performance. For the standalone models, we observe that the percentage of false negatives is higher, and the percentage of true positives is lower in comparison to the pooled and the fine-tuned models. This is a result of the inadequate training data of the SENSO. After excluding the SENSO charity, the average true positive rate of the standalone models in the last 24 months is 3.66% (sd = 0.01) on training data and 3.69% (sd = 0.01) on test data. Similarly, the average false positive rates are 0.09% (sd = 0.00) and 0.1% (sd=0.00) on the training and test data, respectively.

Interestingly, even after we remove SENSO, the standalone models make more false negatives and fewer true positives than the pooled model and the fine-tuned models. This suggests that, by and large, it is somewhat beneficial for firms to enter into a co-op database. For firms with a limited number of repeat observations on their customers, however, the benefits of joining a co-op database can be even more substantial.

	In-sample predictions (25-48 months)			*Out-of-sample predictions (0-24 months)*		
	Standalone models	***Pooled models***	***Fine-tuned models***	***Standalone models***	***Pooled models***	***Fine-tuned models***
True Positive	3.58 *(0.01)*	3.62 *(0.01)*	3.63 *(0.01)*	3.32 *(0.01)*	4.15 *(0.01)*	4.15 *(0.01)*
False Positive	0.13 *(0)*	0.16 *(0)*	0.16 *(0)*	0.09 *(0)*	0.11 *(0)*	0.12 *(0)*
False Negative	0.06 *(0)*	0.02 *(0)*	0.02 *(0)*	0.83 *(0.02)*	0 *(0)*	0 *(0)*
True Negative	96.23 *(0.01)*	96.19 *(0.01)*	96.19 *(0.01)*	95.76 *(0.01)*	95.74 *(0.01)*	95.73 *(0.01)*
RMSE	38.17 *(33.85)*	38.4 *(33.63)*	38.29 *(33.74)*	30.2 *(18.16)*	30.15 *(18.07)*	30.14 *(18.12)*

Table 1 – Predictive performance on in-sample and out-of-sample data. Note: Standard deviations are provided in parentheses. All numbers (except RMSE) are expressed in percentages.

5.2 Performance for joining firms at different stages

It is also possible that firms approach a data aggregator solely for insights or models and refrain from contributing their data to the co-op database. To comprehensively analyze this situation, we performed a scenario analysis on three charities that did not contribute data to the co-op database.

We take the donor database of four external charities —that did not contribute their data to the co-op database— and vary the number of repeated observations those charities have on each customer, simulating the fact that the charity benefits from the co-op database model at different stages of its lifecycle. We then imagine each charity choses one of three modeling approaches for accurately predicting the behavior of their customers, given these data restrictions. In each simulated cutoff in the history of the database, we compare a standalone model trained on the truncated data, the pooled model, and a fine-tuned model that uses the pooled model as a starting point but learns idiosyncratic customer behavior from the truncated data.

The four charities we use are LIBAN, ESPOIR, CREE and SNDL. They are very different from each other in terms of the size of the donor base, as well as the number of repeated observations available for each donor. Furthermore, SNDL observes sudden changes of customer behaviour. In 2011, they suddenly increased marketing pressure by 50% —a quite significant strategy shift whose effects could not be anticipated based on their sole past experience—. In addition, approximately twelve months after that shift in strategy, the charity was hit by a natural catastrophe, which generated a wave of exceptional support and additional donations.

Error! Reference source not found.2 displays the RMSE calculated over the entire v alidation data set for every 6-month increment in training for LIBAN, ESPOIR, CREE, and SNDL databases, respectively.

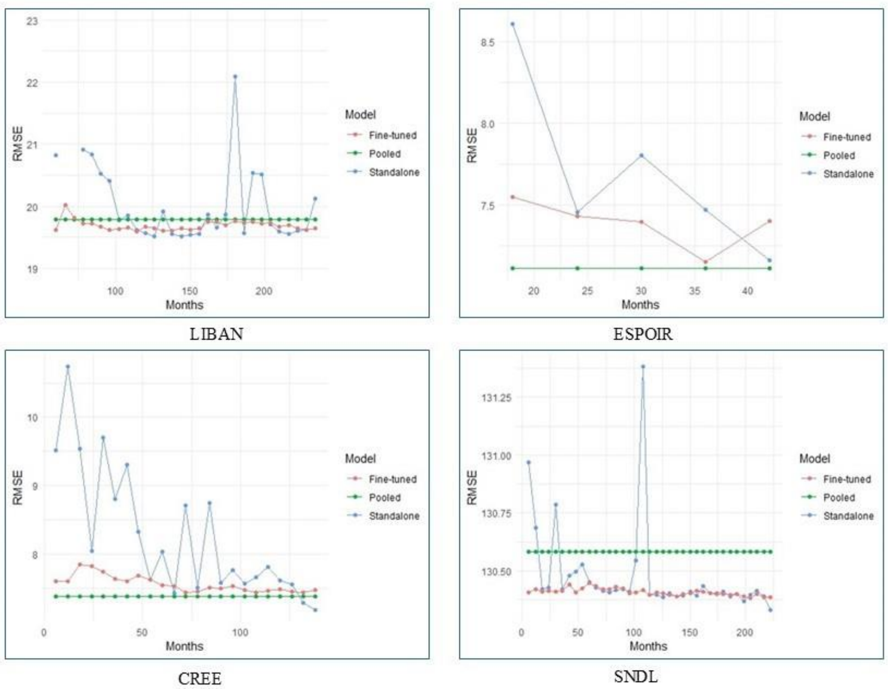

Figure 2 – RMSE (validation data) of the three models as the history of available data increases for the LIBAN, ESPOIR, CREE and SNDL databases. Note that two data points for LIBAN for the standalone model are so large that they are not displayed on the plot as they occur beyond the upper limit of y-axis.

We observe a few noteworthy trends. Firstly, standalone models perform very poorly with a limited number of repeated observations. As the charity increases the size of its proprietary data, the standalone models begin to perform better, and in some cases, even beat the pooled model. This is because the idiosyncratic behavior of customers specific to a firm cannot be systematically captured even by pooling data across several other firms (more on this later).

Secondly, the pooled model outperforms the standalone models for CREE and ESPOIR. This is because the pooled model has trained on sequence lengths much longer than the active timeline of these two charities and has more "experience."

Thirdly, the fine-tuned models outperform the standalone models and the pooled model for LIBAN and SNDL. We suspect this is due to the fact that the charity has a longer active timeline than the maximum length for which the pooled model has been trained. Hence, a model that combines the shared wisdom of the pooled model with idiosyncratic customer behavior performs the best.

We also observe that the errors made by the standalone model are not consistent. This phenomenon is very prominent in SNDL. As soon as the charity is subject to the aforementioned shocks, the standalone model starts making highly inaccurate predictions. Because the charity has never been exposed to shocks of such amplitude in the past, the standalone model has no reference point and simply assumes that whatever spikes in donations have been observed recently will become "the new normal." When we restrict data to the standalone model till right after the spike in solicitation and donation, the model expects the altered donation behavior to be more prevalent in the future, which in turn leads to an inflated RMSE. The fine-tuned model, which has been trained on a pooled database of charities with various histories and shocks, appears to make predictions that are much more resilient.

6. Conclusion and Discussion

In this paper, by combining transfer learning and model customization, we show that co-op databases can provide a creative solution to long-standing customer prediction challenges. Transfer learning on a co-op database allows firms to leverage pooled insights from multiple databases while customizing their model to the idiosyncratic behavior of their customers. This is especially useful when firms try to predict future behavior based on past data whenever past data does not accurately represent (or is not sufficient to predict) the future. In addition, sudden changes in targeting strategies, external events, and the limited number of repeated observations can all result in inaccurate long-term predictions. In such cases, co-op databases provide guidance as the unforeseen future of one firm could be the already observed history of another.

7. References

Durango-Cohen, E. J., Torres, R. L., & Durango-Cohen, P. L. (2013). Donor Segmentation: When Summary Statistics Don't Tell the Whole Story. *Journal of Interactive Marketing*, *27*(3), 172–184. https://doi.org/10.1016/J.INTMAR.2013.04.002

Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, *9*(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Kingma, D. P., & Ba, J. L. (2014). Adam: A Method for Stochastic Optimization. *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*. https://arxiv.org/abs/1412.6980v9

Sarkar, M., & De Bruyn, A. (2021). LSTM Response Models for Direct Marketing Analytics: Replacing Feature Engineering with Deep Learning. *Journal of Interactive Marketing*, *53*, 80–95. https://doi.org/10.1016/J.INTMAR.2020.07.002/ASSET/IMAGES/LARGE/10.1016_J.INTMAR.2020.07.002-FIG7.JPEG

Simester, D. I., Sun, P., & Tsitsiklis, J. N. (2006). Dynamic Catalog Mailing Policies. *Management Science*, *52*(5), 683–696. https://doi.org/10.1287/MNSC.1050.0504

Snoek, J., Larochelle, H., & Adams, R. P. (2012). Practical Bayesian Optimization of Machine Learning Algorithms. *Advances in Neural Information Processing Systems*, *25*.

Valendin, J., Reutterer, T., Platzer, M., & Kalcher, K. (2022). Customer base analysis with recurrent neural networks. *International Journal of Research in Marketing*, *39*(4), 988–1018. https://doi.org/10.1016/J.IJRESMAR.2022.02.007

Van Diepen, M., Donkers, B., & Franses, P. H. (2009). Dynamic and Competitive Effects of Direct Mailings: A Charitable Giving Application. *Journal of Marketing Research*, *46*(1), 120–133. https://doi.org/10.1509/JMKR.46.1.120

Wang, W., Li, B., Luo, X., & Wang, X. (2022). Deep Reinforcement Learning for Sequential Targeting. *Management Science*, *69*(9), 5439–5460. https://doi.org/10.1287/MNSC.2022.4621