Distribution Shift in Airline Customer Behavior during COVID-19

Anonymous Author(s)
Affiliation
Address
email

Abstract

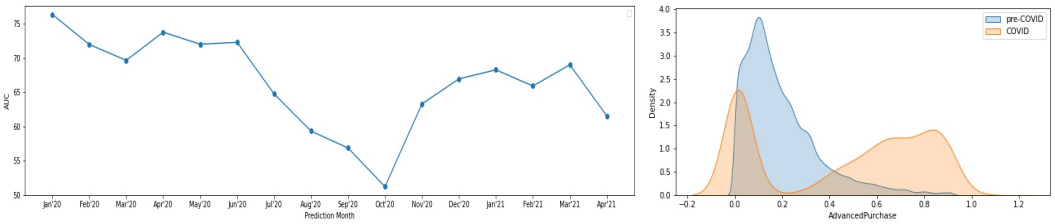
Traditional AI approaches in customized (personalized) contextual pricing applications assume that the data distribution at the time of online pricing is similar to that observed during training. However, this assumption may be violated in practice because of the dynamic nature of customer buying patterns, particularly due to unanticipated system shocks such as COVID-19. We study the changes in customer behavior for a major airline during the COVID-19 pandemic by framing it as a covariate shift detection problem. We identify which customers changed their travel and purchase behavior and the attributes affecting that change using (i) Fast Generalized Subset Scanning and (ii) Causal Forests. In our experiments with simulated and real-world data, we present how these two techniques can be used to detect covariate shifts through qualitative analysis.

1 Introduction

The novel coronavirus pandemic has had a seismic impact on many industries, including travel. During these unprecedented times, along with industrial operations, customer behavior also changed drastically [1]. Due to this, machine learning systems built on sequential decision making were affected the most. Machine learning applications often implicitly or explicitly assume that data sets are drawn from stationary distributions, and the sudden shift in underlying data makes the model prone to break. Pricing based on context is one such application that is prominently driven by customers' behaviour [2, 3]. The motivation of this work stems from examining the performance of machine learning models deployed to price add-on (ancillary) products at an international airline. To illustrate, the performance of one of the deployed models dropped from 75% during training to 50% during testing despite exhaustive offline analysis with K-fold cross validation as shown in Figure 1. Since the onset of COVID-19, one of the features used in the model started experiencing distribution change and needed further investigation.

Demand forecasting is frequently used for price determination given customers' context [4, 5]. Pricing applications that rely on forecasting need a distribution shift detection as the price that a customer is willing to pay can be affected by unanticipated events, such as COVID-19, that alter the system dynamics and change the data generating process. In this work, we (a) investigate covariate shift for a forecasting model used for dynamic pricing, where the distribution of the covariates between the training and test sets differ [6], (b) explore two potential techniques for detecting covariate shift for online use-case, and (c) discuss the proposal of a robust framework for contextual pricing applications in a real-world setting.

Submitted to 35th Conference on Neural Information Processing Systems (NeurIPS 2021). Do not distribute.



(a) AUC Trend Plot

(b) Density Plot

Figure 1: **(a)** AUC of the classifier used for purchase prediction observed from Jan 2020 to Apr 2021. **(b)** Distribution of Advanced Purchase feature capturing the number of days prior to departure the customer made a booking request for pre-COVID (Mar 2019 - Feb 2020) and COVID (Mar 2020 - Sep 2020) period.

2 Related Work

Distribution shifts have been extensively studied in the academic literature. From probability theory, Kullback-Leibler divergence and f-divergence are known to measure the differences in the distributions [7, 8]. Moreover, Wald-Wolfwitz test [9] and exponentially weighted moving average [10, 11, 12] based statistical approaches been proposed to detect the distrbution shifts. For small samples, techniques involving density estimation [13], margin-based approaches such as the one-class support vector machines [14], generative adversarial nets [15] have been explored. Two-sample-testing-based approach which uses pre-trained classifiers for dimensionality reduction has shown to perform well at detecting dataset shifts [16]. Literature also addresses distribution shifts in the context of domain adaptation. Recent works in *out-of-distribution* (OOD) sample detection have shown the possibility of explicitly adapting the loss to aid the task [17, 18, 19]. However, these methods can only be applied to a batch of observations and not in real-time. Jesson et al. [20] made one of the early efforts using causal inference for shift detection, but there is still scope for building methods to detect shift induced by an intervention.

3 Methods

In this section, we define how Fast Generalized Subset Scan [21] and Causal Forests [22] can be used for covariate shift detection due to the shocks, such as COVID-19, experienced by a system.

3.1 Fast Generalized Subset Scan

Fast Generalized Subset Scan (FGSS) is an unsupervised anomalous pattern detection technique, proposed by McFowland, Speakman, and Neill [21]. We use FGSS for covariate shift detection with the objective of finding shifted patterns in the test set. Under the null hypothesis that there is no anomalous pattern in the test set, we use FGSS to detect self-similar groups of observations in the test set for which some subset of features are shifted. If the null hypothesis is true, then the test set is generated from the same distribution as the training set. Otherwise, the training and test distributions are different with a significantly higher score $F(S)$ for a subset of observations in the test set, defined below.

$$F(S) = \max_{\alpha} F_{\alpha}(S) = \max_{\alpha} \phi(\alpha, N_{\alpha}(S), N(S)) \tag{1}$$

where $N(S)$ represents the size of the subset S and $N_{\alpha}(S)$ represents the total number of p-values (obtained by passing the observed values in the test through the inverse CDF of the training set) that are significant at level α in S. The function $\phi(\alpha, N_{\alpha}(S), N(S))$ is monotonically increasing w.r.t. N_{α}, monotonically decreasing w.r.t. N and α, and is convex. We use the Berk-Jones statistic [23] as ϕ in our experiments which satisfy these properties.

3.2 Causal Forests

Causal forests in a supervised method from Generalized Random Forests [22] that estimates heterogeneity in treatment effects. A treatment effect refers to a causal effect of a treatment or intervention on an outcome variable. If the treatment effect is heterogeneous, causal forests can be used to estimate

the Conditional Average Treatment Effect (CATE). This is useful in identifying the observations for which the treatment is positive and that benefit the most from a treatment; essentially an estimation of optimal policy assignment [24]. CATE cannot be directly observed for a unit because of the "fundamental problem of causal inference" [25], making it impossible to observe unit-level causal effects and the reason why we can never directly observe the counterfactual condition of a unit of observation. For each observation $X \in \mathbb{R}^m$ where m is the number of covariates, there are two potential outcomes Y_1 and Y_0 corresponding to the binary treatment variable $D \in \{0, 1\}$, but only one of them is observed. The conditional expectation of an outcome for the treatment or control, $\mu_d(x)$ is defined as:

$$\mu_d(x) = E[Y_i|X_i = x, D_i = d] \quad with \quad D \in \{0, 1\} \tag{2}$$

and CATE ($\tau(x)$) is the difference in expectation of the potential outcomes given x,

$$\tau(x) = E[Y_i^1 - Y_i^0|X_i = x] = \mu_1(x) - \mu_0(x) \tag{3}$$

We use causal forests to estimate the causal effect of COVID-19 intervention and detect the covariate shift in features. This requires some data observed after the intervention to be leaked into the training data so as to learn the unit-level interventional change in treatment.

4 Experiments

Ancillaries are optional products or services sold by businesses to complement their primary product [26]. In this work, we utilized the following datasets from the airline industry for an ancillary market: (1) simulated interaction of ancillary pricing, and (2) real-world ancillary booking requests data from a large airline containing price variability.

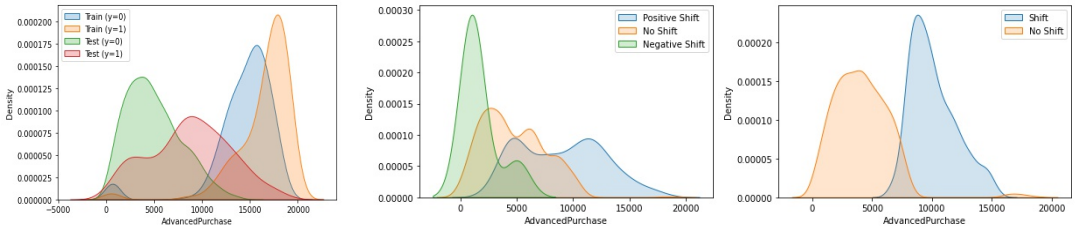
4.1 Simulated dataset

To test our approach, we generated a simulated dataset of customer and flight seat (ancillary) interactions using open-sourced flight simulator [1]. In this dataset, we artificially varied the arrival rates of the customers in train and test data. The details of the data-generation process can be found in Appendix A.1. Figure 2(a) shows the generated data distribution. Figures 2(b) and (c) show the results from Causal Forests and FGSS, respectively. FGSS is able to identify 26.01% points in the test set as shifted. The shifted points have *AdvancedPurchase* feature values greater than 5000 despite the test distribution having a peak around 4000. This is because the train distribution has few points in the values between 0 and 4000, which the FGSS model identifies as normal behavior. On the other hand, Causal Forests is able to identify 23.26% points with positive shift and only 0.56% points with negative shift, indicating a superior performance of Causal Forests model to identify shift on intervention. The positive shift occurs over the range of test set values, which aligns with the distribution of test set simulated significantly different from the train set.

4.2 Real-world dataset

The real-world dataset consists of customer booking requests from March 2019 to September 2020. We use this dataset to evaluate the performance of FGSS and Causal Forests in detecting distribution shift due to COVID-19 in real-world setting. The airline identifies the time period starting March 2020 as the "COVID" era because a large portion of scheduled flights started getting canceled/rescheduled and observed decreased ticket sales. We use five features in our experiments - *AdvancedPurchase*, *LengthOfStay*, *GroupSize*, *TotalDuration* and *TripType*. The description of the features can be found in Appendix: Table 1. Figure 3 and 4 show the results obtained on the hold-out test set. Both Causal Forests and FGSS results indicate that there is significant shift in all of the features. *LengthOfStay* feature has shift towards value 0 (indefinite stay at destination), indicating most customers are not opting for vacation/business travel but essential movement only. The same conclusion can be drawn from the shift observed in *TripType* feature with customers preferring to travel one-way. *GroupSize* feature has a shift towards value 1, indicating customers are traveling alone and not in large groups during the pandemic. *AdvancedPurchase* has a shift towards smaller values indicating customers are not booking tickets way ahead of the travel date due to the uncertain nature of the pandemic. *TotalDuration* has a shift towards smaller values as well, indicating the operational changes made by the airline to operate on shorter routes during the pandemic.

[1] https://github.com/deepair-io/flai



(a) Data Distribution **(b)** Causal Forests **(c)** FGSS

Figure 2: **(a)** Train and test distributions for the binary target of purchase/no-purchase. **(b)** Density plot of observations identified as positive, negative and not shifted by Causal Forests in the test set. Following Equation (3), *positive shift* identified by Causal Forests represent the observations more likely to have come from the population that experienced treatment (test) and similarly *negative shift* represent the observations more likely to have come from the control population (train), while *no shift* represent the observations that have similar distributions between treatment and control periods. **(c)** Density plot of observations identified as shifted and not shifted by FGSS in the test set. FGSS only tells whether an observation is shifted or not, and says nothing about it coming from the treatment or control period.

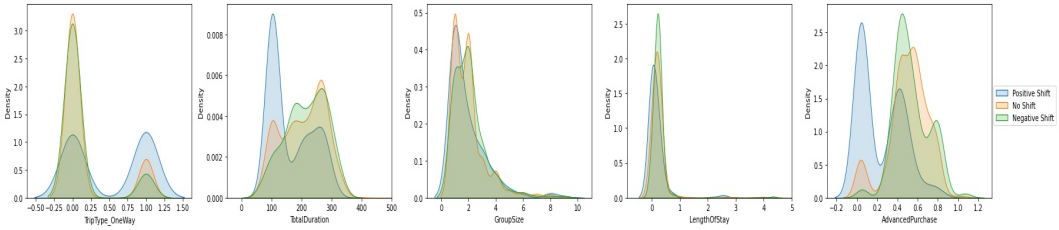


Figure 3: Density plot of observations identified as positive, negative and not shifted by Causal Forests.

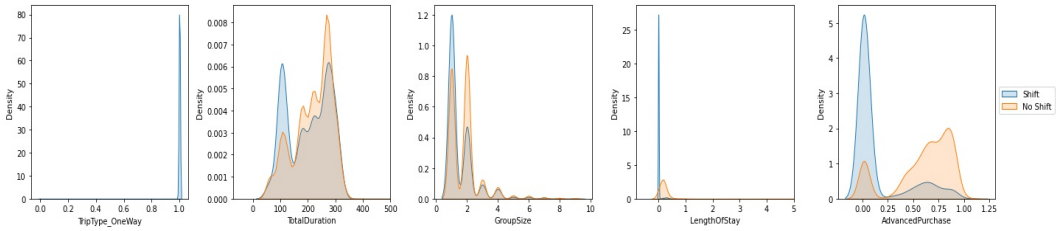


Figure 4: Density plot of observations identified as shifted and not shifted by FGSS.

5 Conclusion

We explored two techniques, Fast Generalized Subset Scan and Causal Forests, for covariate shift detection when a system experiences a shock. We applied these techniques to an airline ancillary purchase use case and saw that some of the features used for ancillary pricing experienced a shift during COVID-19. We identified the observations in the test set that follow a significantly different distribution compared to the training set. Causal forests, while able to detect a shift in real-time, need a significant amount of data post-intervention for training. On the other hand, FGSS can only detect shifts on a batch of data but does not need data post-intervention for training. Hence, there is a possibility to combine the two approaches, both addressing the limitations of one another, for a unified and robust covariate shift detection. At the same time, the construction of a good testing pipeline, along with an agreement by multiple models as well as a visual inspection, can jointly help identify shift patterns. These patterns can then be cross-checked with designers' hypothesis and domain experts' knowledge for validation and potential changes to the system before being corrected for enhanced model performance. In the future, we aim to extend this work to other applications, test multiple shift correction approaches and provide recommendations for adapting to a shift induced by sudden shocks to the system (see Appendix A.2 for details).

References

[1] Jagdish Sheth. "Impact of Covid-19 on consumer behavior: Will the old habits return or die?" In: *Journal of Business Research* 117 (2020), pp. 280–283. ISSN: 0148-2963. DOI: https://doi.org/10.1016/j.jbusres.2020.05.059. URL: https://www.sciencedirect.com/science/article/pii/S0148296320303647.

[2] Peng Ye et al. "Customized regression model for airbnb dynamic pricing". In: *Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining*. 2018, pp. 932–940.

[3] Naman Shukla et al. "Dynamic Pricing for Airline Ancillaries with Customer Context". In: *Proceedings of the 25th ACM SIGKDD Conference on ACM Knowledge Discovery and Data Mining*. ACM. July 2019, pp. 2174–2182.

[4] Kalyan T Talluri, Garrett Van Ryzin, and Garrett Van Ryzin. *The theory and practice of revenue management*. Vol. 1. Springer, 2004.

[5] Kris Johnson Ferreira, Bin Hong Alex Lee, and David Simchi-Levi. "Analytics for an online retailer: Demand forecasting and price optimization". In: *Manufacturing & Service Operations Management* 18.1 (2016), pp. 69–88.

[6] Joaquin Quionero-Candela et al. *Dataset Shift in Machine Learning*. The MIT Press, 2009. ISBN: 0262170051.

[7] S. Kullback and R. A. Leibler. "On Information and Sufficiency". In: *The Annals of Mathematical Statistics* 22.1 (1951), pp. 79–86. DOI: 10.1214/aoms/1177729694. URL: https://doi.org/10.1214/aoms/1177729694.

[8] A. Rényi. "On measures of information and entropy". In: *Proceedings of the fourth Berkeley Symposium on Mathematics, Statistics and Probability*. Berkeley, CA: University of California Press, 1961, pp. 547–561.

[9] Abbas Alhakim and William Hooper. "A non-parametric test for several independent samples". In: *Journal of Nonparametric Statistics* 20.3 (2008), pp. 253–261.

[10] Haider Raza, Girijesh Prasad, and Yuhua Li. "Dataset Shift Detection in Non-stationary Environments Using EWMA Charts". In: *2013 IEEE International Conference on Systems, Man, and Cybernetics*. 2013, pp. 3151–3156. DOI: 10.1109/SMC.2013.537.

[11] Haider Raza, Girijesh Prasad, and Yuhua Li. "EWMA Based Two-Stage Dataset Shift-Detection in Non-stationary Environments". In: *Artificial Intelligence Applications and Innovations*. Ed. by Harris Papadopoulos et al. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 625–635.

[12] Haider Raza, Girijesh Prasad, and Yuhua Li. "EWMA model based shift-detection methods for detecting covariate shifts in non-stationary environments". In: *Pattern Recognition* 48.3 (2015), pp. 659–669. ISSN: 0031-3203. DOI: https://doi.org/10.1016/j.patcog.2014.07.028. URL: https://www.sciencedirect.com/science/article/pii/S0031320314002878.

[13] Markus M Breunig et al. "LOF: identifying density-based local outliers". In: *Proceedings of the 2000 ACM SIGMOD international conference on Management of data*. 2000, pp. 93–104.

[14] Bernhard Schölkopf et al. "Support vector method for novelty detection." In: *NIPS*. Vol. 12. Citeseer. 1999, pp. 582–588.

[15] Thomas Schlegl et al. "Unsupervised anomaly detection with generative adversarial networks to guide marker discovery". In: *International conference on information processing in medical imaging*. Springer. 2017, pp. 146–157.

[16] Stephan Rabanser, Stephan Günnemann, and Zachary Lipton. "Failing Loudly: An Empirical Study of Methods for Detecting Dataset Shift". In: *Advances in Neural Information Processing Systems*. Ed. by H. Wallach et al. Vol. 32. Curran Associates, Inc., 2019.

[17] Shiyu Liang, Yixuan Li, and Rayadurgam Srikant. "Enhancing the reliability of out-of-distribution image detection in neural networks". In: *arXiv preprint arXiv:1706.02690* (2017).

[18] Kimin Lee et al. "Training confidence-calibrated classifiers for detecting out-of-distribution samples". In: *arXiv preprint arXiv:1711.09325* (2017).

[19] Alireza Shafaei, Mark Schmidt, and James J Little. "Does Your Model Know the Digit 6 Is Not a Cat? A Less Biased Evaluation of" Outlier" Detectors." In: (2018).

[20] Andrew Jesson et al. "Identifying Causal-Effect Inference Failure with Uncertainty-Aware Models". In: *Advances in Neural Information Processing Systems*. Ed. by H. Larochelle et al. Vol. 33. Curran Associates, Inc., 2020, pp. 11637–11649. URL: https://proceedings.neurips.cc/paper/2020/file/860b37e28ec7ba614f00f9246949561d-Paper.pdf.

[21] Edward III McFowland, Skyler Speakman, and Daniel B. Neill. "Fast Generalized Subset Scan for Anomalous Pattern Detection". In: *Journal of Machine Learning Research* 14.12 (2013), pp. 1533–1561. URL: http://jmlr.org/papers/v14/mcfowland13a.html.

[22] Susan Athey, Julie Tibshirani, and Stefan Wager. "Generalized random forests". In: *The Annals of Statistics* 47.2 (2019), pp. 1148–1178. DOI: 10.1214/18-AOS1709. URL: https://doi.org/10.1214/18-AOS1709.

[23] Robert H. Berk and Douglas H. Jones. "Goodness-of-fit test statistics that dominate the Kolmogorov statistics". English (US). In: *Probability Theory and Related Fields* 47.1 (Jan. 1979), pp. 47–59. ISSN: 0178-8051. DOI: 10.1007/BF00533250.

[24] Susan Athey and Guido W. Imbens. "Machine Learning Methods That Economists Should Know About". In: *Annual Review of Economics* 11.1 (2019), pp. 685–725. DOI: 10.1146/annurev-economics-080217-053433.

[25] Paul W. Holland. "Statistics and Causal Inference". In: *Journal of the American Statistical Association* 81.396 (1986), pp. 945–960. DOI: 10.1080/01621459.1986.10478354.

[26] Adam Bockelie and Peter Belobaba. "Incorporating ancillary services in airline passenger choice models". In: *Journal of Revenue and Pricing Management* 16.6 (2017), pp. 553–568.

[27] PA W Lewis and Gerald S Shedler. "Simulation of nonhomogeneous Poisson processes by thinning". In: *Naval research logistics quarterly* 26.3 (1979), pp. 403–413.

[28] Jerry Hausman and Daniel McFadden. "Specification tests for the multinomial logit model". In: *Econometrica: Journal of the econometric society* (1984), pp. 1219–1240.

A Appendix

A.1 Additional Information

The simulated dataset consists of interactions for selecting flight seat after the ticket is purchased. Customer arrivals are simulated using non-homogeneous poisson process [27]. Multinomial logit is used for customer choice model in the simulator [28]. The offered prices are randomly sampled from minimum to maximum allowed price for that flight. The intervention is artificially created by changing the customer arrival rate. We have simulated 2 datasets with different parameters of the poisson process for 10 long-distance flights as train and test (Fig. 2(a)). We use the arrival rate (*AdvancedPurchase*) and number of seats sold (*Sold*) as features, with *AdvancedPurchase* feature having an apparent covariate shift. For Causal Forests, we leak 40% points randomly sampled from the test set into training.

Table 1 presents the description of features discussed in Section 4.2. FGSS requires all features to be categorical. We discretize all numeric features by binning them at quartiles for FGSS training. We divide the real-world data into two parts: pre-COVID (March 2019 to February 2020) and COVID (March 2020 to September 2020). We use the pre-COVID period for training FGSS and Causal Forests model. Further, we leak March 2020 to August 2020 COVID data into training for Causal Forests, and leave September 2020 as the hold-out test set to report results.

Table 1: Feature Description

Feature	Type	Description
AdvancedPurchase	Numeric	Number of days prior to flight departure the request has been made
LengthOfStay	Numeric	Number of days between onward and return flight (zero for one-way ticket)
GroupSize	Numeric	Number of passengers on the customer booking
TotalDuration	Numeric	Duration of the flight in minutes
TripType	Categorical	Indicates whether the customer has booked "One-Way" or "Round-Trip"

A.2 Robust Pricing Model

Shukla et al. [3] proposed a two-stage pricing model that uses the purchase probability prediction for price recommendation. Figure 5 shows the proposed pricing framework with shift detection before the supervised model prediction. The shift detection layer first checks if there is a shift in the system or not. Causal Forests being a supervised technique can be used to detect shift "on-line". However, Causal Forests require data observed after intervention for training and identification of the start a shift induced by an intervention to the system is usually a difficult task. We propose to take advantage of the unsupervised nature of FGSS to achieve that. FGSS can learn the normal behavior from historical data and run in batch mode to detect anomalous patterns during testing. If the percentage of observed anomalous patterns breach a user-defined threshold for a significant period of time, Causal Forests can use that data as treatment for training.

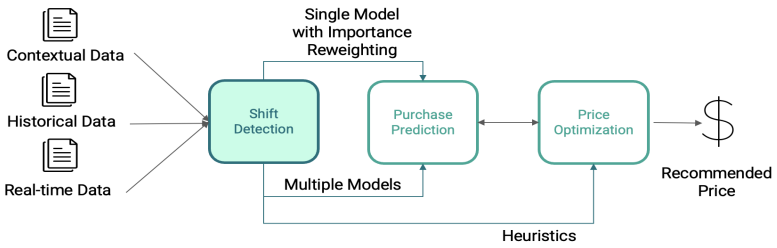


Figure 5: Proposed pricing framework.

Once the shift has been detected in "real-time", it can be handled in multiple ways: a single model trained after reweighting shifted points, multiple models - one for each shifted pattern, or heuristics based on domain knowledge to better price the product. We aim to perform further experiments in future to test the feasibility of these 3 approaches and propose an efficient strategy to handle covariate shift for decision-making under uncertainty.