Artificial Intelligence for Business Research @Antai

Prediction Problems in Business Research

Renyu (Philip) Zhang

1

Why Do We Care About Predictions?

- Everyone cares about the prediction of **macro economic/political/natural outcomes**.
  - Population, elections, GDP, poverty, tax policy, market research, when will humans run out of fossil fuel, etc.

- Sometimes good predictions could directly lead to **good decisions/policies**.
  - Weather forecast, demand forecast, stock/asset return, recommendation system, user/patient LT(V), cancer screening, insurances, bail out, etc.

$$\frac{d\pi(X_0, Y)}{dX_0} = \frac{\partial \pi}{\partial X_0} \underbrace{(Y)}_{\text{prediction}} + \frac{\partial \pi}{\partial Y} \underbrace{\frac{\partial Y}{\partial X_0}}_{\text{causation}}.$$

American Economic Review: Papers & Proceedings 2015, 105(5): 491–495
http://dx.doi.org/10.1257/aer.p20151023

Prediction Policy Problems†

By Jon Kleinberg, Jens Ludwig, Sendhil Mullainathan, and Ziad Obermeyer*

Empirical policy research often focuses on causal inference. Since policy choices seem to depend on understanding the counterfactual—what happens with and without a policy—this tight link of causality and policy seems natural. While this link holds in many cases, we argue that there are also many policy applications where causal inference is not central, or even necessary.

causation and prediction; (ii) explain how machine learning adds value over traditional regression approaches in solving prediction problems; (iii) provide an empirical example from health policy to illustrate how improved predictions can generate large social impact; (iv) illustrate how "umbrella" problems are common and important in many important policy domains; and (v) argue that solving these

- Causal inference is all about **predicting the counterfactual outcomes**.
  - Causal ML, DML, honest tree, matrix completion, etc.

2

2

Macro Predictions

Science

Current Issue First release papers Archive About

RESEARCH ARTICLE

Combining satellite imagery and machine learning to predict poverty

NEAL JEAN, MARSHALL BURKE, MICHAEL XIE, W. MATTHEW DAVIS, DAVID B. LOBELL, AND STEFANO ERMON Authors Info & Affiliations

SCIENCE • 19 Aug 2016 • Vol 353, Issue 6301 • pp. 790-794 • DOI: 10.1126/science.aaf7894

Measuring consumption and wealth remotely

Nighttime lighting is a rough proxy for economic wealth, and nighttime maps of the world show that many developing countries are sparsely illuminated. Jean *et al.* combined nighttime maps with high-resolution daytime satellite images (see the Perspective by Blumenstock). With a bit of machine-learning wizardry, the combined images can be converted into accurate estimates of household consumption and assets, both of which are hard to measure in poorer countries. Furthermore, the night- and day-time data are publicly available and nonproprietary.

Science, this issue p. 790; see also p. 753

Abstract

Reliable data on economic livelihoods remain scarce in the developing world, hampering efforts to study these outcomes and to design policies that improve them. Here we demonstrate an accurate, inexpensive, and scalable method for estimating consumption expenditure and asset wealth from high-resolution satellite imagery. Using survey and satellite data from five African countries—Nigeria, Tanzania, Uganda, Malawi, and Rwanda—we show how a convolutional neural network can be trained to identify image features that can explain up to 75% of the variation in local-level economic outcomes. Our method, which requires only publicly available data, could transform efforts to track and target poverty in developing countries. It also demonstrates how powerful machine learning techniques can be applied in a setting with limited training data, suggesting broad potential application across many scientific domains.

(Almost) 200 Years of News-Based Economic Sentiment*

J. H. van Binsbergen† S. Bryzgalova‡ M. Mukhopadhyay§ V. Sharma¶

March 23, 2023

Abstract

Using the text of 200 million pages of 13,000 US local newspapers and state-of-the-art machine learning methods, we construct a novel 170-year-long time series measure of economic sentiment at the country and state levels, that expands the existing measures in both the time series (by more than a century) and the cross-section. We show that our measure predicts economic fundamentals such as GDP (both nationally and locally), consumption, and employment growth, even after controlling for commonly-used predictors, and materially predicts monetary policy decisions, particularly during recessions. Our measure is distinct from the information in expert forecasts and leads its consensus value. We use the text to isolate information about current and future events and show that it is the latter that drives our predictability results.

Keywords: Business cycle, macroeconomic news, economic sentiment, monetary policy, textual analysis, machine learning, big data, neural networks

JEL codes: G1, G4, E2.

3

3

Demand Forecasting

Machine Learning Methods for Demand Estimation

Patrick Bajari
Denis Nekipelov
Stephen P. Ryan
Miaoyu Yang

AMERICAN ECONOMIC REVIEW
VOL. 105, NO. 5, MAY 2015
(pp. 481-85)

Download Full Text PDF

Article Information

Abstract

We survey and apply several techniques from the statistical and computer science literature to the problem of demand estimation. To improve out-of-sample prediction accuracy, we propose a method of combining the underlying models via linear regression. Our method is robust to a large number of regressors; scales easily to very large data sets; combines model selection and estimation; and can flexibly approximate arbitrary non-linear functions. We illustrate our method using a standard scanner panel data set and find that our estimates are considerably more accurate in out-of-sample predictions of demand than some commonly used alternatives.

Citation

Bajari, Patrick, Denis Nekipelov, Stephen P. Ryan, and Miaoyu Yang. 2015. "Machine Learning Methods for Demand Estimation." *American Economic Review*, 105 (5): 481-85.

DOI: 10.1257/aer.p20151021

Shift from linear models to ML

More/Wider Data

PRODUCTION AND OPERATIONS MANAGEMENT POMS
Vol. 27, No. 10, October 2018, pp. 1749–1769
ISSN 1059-1478 | EISSN 1937-5956 | 18 | 2710 | 1749
DOI 10.1111/poms.12707

The Operational Value of Social Media Information

Ruomeng Cui
Kelley School of Business, Indiana University, Bloomington, Indiana 47405, USA, cuir@indiana.edu

Santiago Gallino
Tuck School of Business, Dartmouth College, Santiago, New Hampshire 03755, USA, gallino@tuck.dartmouth.edu

Antonio Moreno
Kellogg School of Management, Northwestern University, Evanston, Illinois 60208, USA, a-morenogarcia@kellogg.northwestern.edu

Dennis J. Zhang
John M. Olin Business School, Washington University in Saint Louis, Saint Louis, Missouri 63130, USA, denniszhang@wustl.edu

While the value of using social media information has been established in multiple business contexts, the field of operations and supply chain management have not yet explored the possibilities it offers in improving firms' operational decisions. This study attempts to do that by empirically studying whether using publicly available social media information can improve the accuracy of daily sales forecasts.We collaborated with an online apparel retailer to assemble a dataset that combines (1) detailed internal operational information, including data on sales, advertising, and promotions, as well as (2) publicly available social media information obtained from Facebook. We implement a variety of machine learning methods to forecast daily sales. We find that using social media information results in statistically significant improvements in the out-of-sample accuracy of the forecasts, with relative improvements ranging from 12.85% to 23.23% over different forecast horizons. We also demonstrate that nonlinear boosting models with feature selection, such as random forests, perform significantly better than traditional linear models. The best-performing method (random forest) yields an out-of-sample MAPE of 7.21% when not using social media information and 5.73% when using social media information is used. In both cases, this significantly improves the accuracy of the company's internal forecasts (a MAPE of 11.97%). Combining these empirical results, we provide recommendations for forecasting sales in general as well as with social media information.

Key words: social media; sales forecast; machine learning
History: Received: February 2016; Accepted: February 2017 by Ram Ganeshan, after 2 revisions.

informs
https://pubsonline.informs.org/journal/opre

OPERATIONS RESEARCH
Vol. 70, No. 5, September–October 2022, pp. 2597–2618
ISSN 0030-364X (print), ISSN 1526-5463 (online)

Contextual Areas

Data Aggregation and Demand Prediction

Maxime C. Cohen,[a] Renyu Zhang,[b,*] Kevin Jiao[c]

[a] Desautels Faculty of Management, McGill University, Montreal, Quebec H3A 1G5, Canada; [b] Department of Decision Sciences and Managerial Economics, Business School, The Chinese University of Hong Kong, Hong Kong, China; [c] Stern School of Business, New York University, New York 10012
*Corresponding author
Contact: maxime.cohen@mcgill.ca, https://orcid.org/0000-0002-2474-3875 (MCC); philipzhang@cuhk.edu.hk, https://orcid.org/0000-0003-0284-164X (RZ); jjiao@stern.nyu.edu (KJ)

Received: June 2, 2021
Revised: December 6, 2021; February 14, 2022
Accepted: March 11, 2022
Published Online in Articles in Advance: July 7, 2022

Area of Review: Operations and Supply Chains

https://doi.org/10.1287/opre.2022.2301

Abstract. We study how retailers can use data aggregation and clustering to improve demand prediction. High accuracy in demand prediction allows retailers to effectively manage their inventory as well as mitigate stock-outs and excess supply. A typical retail setting involves predicting demand for hundreds of items simultaneously. Although some items have a large amount of historical data, others were recently introduced and, thus, transaction data can be scarce. A common approach is to cluster several items and estimate a joint model for each cluster. In this vein, one can estimate some model parameters by aggregating the data from several items and other parameters at the individual-item level. We propose a practical method referred to as *data aggregation with clustering* (DAC), which balances the tradeoff between data aggregation and model flexibility. DAC allows us to predict demand while optimally identifying the features that should be estimated at the (i) item, (ii) cluster, and (iii) aggregate levels. We show that the DAC algorithm yields a consistent and normal estimate, along with improved prediction errors relative to the decentralized benchmark, which estimates a different model for each item. Using both simulated and real data, we illustrate DAC's improvement in prediction accuracy relative to a wide range of common benchmarks. Interestingly, the DAC algorithm has theoretical and practical advantages and helps retailers uncover meaningful managerial insights.

Funding: R. Zhang is grateful for financial support from the National Natural Science Foundation of China [Grant NSFC71802133] and the Shanghai Eastern Scholar Program [Grant QD2018053].
Supplemental Material: The online appendices are available at https://doi.org/10.1287/opre.2022.2301.

Keywords: retail analytics • demand prediction • data aggregation • clustering

More Efficient Use of Data

4

4

Recommendation (Business)

informs.
http://pubsonline.informs.org/journal/mnsc/

MANAGEMENT SCIENCE
Vol. 65, No. 7, July 2019, pp. 3131–3149
ISSN 0025-1909 (print), ISSN 1526-5501 (online)

Learning Preferences with Side Information

Vivek F. Farias,[a] Andrew A. Li[b]

[a] Sloan School of Management, Massachusetts Institute of Technology, Cambridge, Massachusetts 02142; [b] Operations Research Center, Massachusetts Institute of Technology, Cambridge, Massachusetts 02142

Contact: vivekf@mit.edu, http://orcid.org/0000-0002-5856-9246 (VFF); aali@mit.edu, http://orcid.org/0000-0002-9552-6421 (AAL)

Received: October 8, 2016
Revised: May 1, 2017; July 19, 2017
Accepted: August 14, 2017
Published Online in Articles in Advance: May 8, 2019

https://doi.org/10.1287/mnsc.2018.3092

Abstract. Product and content personalization is now ubiquitous in e-commerce. There are typically not enough available transactional data for this task. As such, companies today seek to use a variety of information on the interactions between a product and a customer to drive personalization decisions. We formalize this problem as one of recovering a large-scale matrix with side information in the form of additional matrices of conforming dimension. Viewing the matrix we seek to recover and the side information we have as slices of a tensor, we consider the problem of *slice recovery*, which is to recover specific slices of "simple" tensors from noisy observations of the entire tensor. We propose a definition of simplicity that on the one hand elegantly generalizes a standard generative model for our motivating problem and on the other hand subsumes low-rank tensors for a variety of existing definitions of tensor rank. We provide an efficient algorithm for slice recovery that is practical for massive data sets and provides a significant performance improvement over state-of-the-art incumbent approaches to tensor recovery. Furthermore, we establish near-optimal recovery guarantees that, in an important regime, represent an order improvement over the best available results for this problem. Experiments on data from a music streaming service demonstrate the performance and scalability of our algorithm.

History: Accepted by Noah Gans, stochastic models and simulation.
Supplemental Material: The e-companion is available at https://doi.org/10.1287/mnsc.2018.3092.

Keywords: personalization • e-commerce • online retail • recommender systems • collaborative filtering • matrix recovery • tensor recovery • side information • multi-interaction data

MIS Quarterly
RESEARCH NOTE

ON THE DIFFERENCES BETWEEN VIEW-BASED AND PURCHASE-BASED RECOMMENDER SYSTEMS[1]

Jing Peng and Chen Liang
Department of Operations and Information Management, School of Business, University of Connecticut
Storrs, CT, U.S.A. {jing.peng@uconn.edu} {chenliang@uconn.edu}

E-commerce platforms often use collaborative filtering (CF) algorithms to recommend products to consumers. What recommendations consumers receive and how they respond to the recommendations largely depend on the design of CF algorithms. However, the extant empirical research on recommender systems has primarily focused on how the presence of recommendations affects product demand, without considering the underlying algorithm design. Leveraging a field experiment on a major e-commerce platform, we examine the differential impact of two widely used CF designs: view-also-view (VAV) and purchase-also-purchase (PAP). We found several striking differences between the impact of these two designs on individual products. First, VAV is about seven times more effective in generating additional product views than PAP but only about twice as effective in generating sales due to a lower conversion rate. Second, VAV is more effective in increasing views for more expensive products, whereas PAP is more effective in increasing the sales of cheaper products. Third, VAV is less effective in increasing the views but more effective in increasing the sales of products with higher purchase incidence rates (PIRs). Finally, when aggregated over all products with the same levels of price or PIRs, VAV dominates PAP in generating views and the difference is more striking for products with higher prices or lower PIRs. Interestingly, PAP is more effective than VAV in increasing the sales of products with low prices or moderate PIRs, though VAV generates more sales than PAP overall. Our findings suggest that platforms may benefit from employing different CF designs for different types of products.

Keywords: Collaborative filtering, substitute, complement, price, purchase incidence rate, cross-sell, up-sell

5

5

Recommendation (CS)

RESEARCH-ARTICLE OPEN ACCESS

Deconfounding Duration Bias in Watch-time Prediction for Video Recommendation

Authors: Ruohan Zhan, Changhua Pei, Qiang Su, Jianfeng Wen, Xueliang Wang, Guanyu Mu, Dong Zheng, Peng Jiang, Kun Gai Authors Info & Claims

KDD '22: Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining • August 2022 • Pages 4472–4481 • https://doi.org/10.1145/3534678.3539092

Published: 14 August 2022 Publication History

ABSTRACT

Watch-time prediction remains to be a key factor in reinforcing user engagement via video recommendations. It has become increasingly important given the ever-growing popularity of online videos. However, prediction of watch time not only depends on the match between the user and the video but is often mislead by the duration of the video itself. With the goal of improving watch time, recommendation is always biased towards videos with long duration. Models trained on this imbalanced data face the risk of bias amplification, which misguides platforms to over-recommend videos with long duration but overlook the underlying user interests. This paper presents the first work to study duration bias in watch-time prediction for video recommendation. We employ a causal graph illuminating that duration is a confounding factor that concurrently affects video exposure and watch-time prediction---the first effect on video causes the bias issue and should be eliminated, while the second effect on watch time originates from video intrinsic characteristics and should be preserved. To remove the undesired bias but leverage the natural effect, we propose a Duration-Deconfounded Quantile-based (D2Q) watch-time prediction framework, which allows for scalability to perform on industry production systems. Through extensive offline evaluation and live experiments, we showcase the effectiveness of this duration-deconfounding framework by significantly outperforming the state-of-the-art baselines. We have fully launched our approach on Kuaishou App, which has substantially improved real-time video consumption due to more accurate watch-time predictions.

Deep Neural Networks for YouTube Recommendations

Paul Covington, Jay Adams, Emre Sargin
Google
Mountain View, CA
{pcovington, jka, msargin}@google.com

ABSTRACT

YouTube represents one of the largest scale and most sophisticated industrial recommendation systems in existence. In this paper, we describe the system at a high level and focus on the dramatic performance improvements brought by deep learning. The paper is split according to the classic two-stage information retrieval dichotomy: first, we detail a deep candidate generation model and then describe a separate deep ranking model. We also provide practical lessons and insights derived from designing, iterating and maintaining a massive recommendation system with enormous user-facing impact.

Keywords

recommender system; deep learning; scalability

1. INTRODUCTION

YouTube is the world's largest platform for creating, sharing and discovering video content. YouTube recommendations are responsible for helping more than a billion users discover personalized content from an ever-growing corpus of videos. In this paper we will focus on the immense impact deep learning has recently had on the YouTube video recommendations system. Figure 1 illustrates the recom-

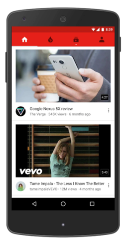

Figure 1: Recommendations displayed on YouTube mobile app home.

Table 3: Live experiments on Kuaishou App. We use VR as a baseline and show the relative performance of WLR and Res-D2Q with *#Groups* = 30. The square brackets represent the 95% confidence intervals for online metrics. Statistically-significant improvement (whose value is not in the confidence interval) is marked with bold font in the table.

Method	Main Metric.	Constraint Metrics.			
	Watch Time	Like	Follow	Share	Comment
WLR *v.s.* VR (baseline)	**+0.184%** [−0.16%, 0.16%]	**+1.012%** [−0.50%, 0.51%]	+0.214% [−0.4%, 0.4%]	+0.959% [−1.31%, 1.40%]	-0.137% [−0.75%, 0.73%]
Res-D2Q *v.s.* VR (baseline)	**+0.746%** [−0.15%, 0.15%]	+0.251% [−0.41%, 0.41%]	-0.167% [−0.6%, 0.6%]	-0.861% [−1.21%, 1.21%]	+0.271% [−0.85%, 0.86%]

6

6

Other Predictions

The Review of Financial Studies

Empirical Asset Pricing via Machine Learning*

Shihao Gu
Booth School of Business, University of Chicago

Bryan Kelly
Yale University, AQR Capital Management, and NBER

Dacheng Xiu
Booth School of Business, University of Chicago

We perform a comparative analysis of machine learning methods for the canonical problem of empirical asset pricing: measuring asset risk premiums. We demonstrate large economic gains to investors using machine learning forecasts, in some cases doubling the performance of leading regression-based strategies from the literature. We identify the best-performing methods (trees and neural networks) and trace their predictive gains to allowing nonlinear predictor interactions missed by other methods. All methods agree on the same set of dominant predictive signals, a set that includes variations on momentum, liquidity, and volatility. (*JEL* C52, C55, C58, G0, G1, G17)

A GPT-4 based stock selector: https://arxiv.org/pdf/2401.03737.pdf

nature medicine

Article https://doi.org/10.1038/s41591-023-02640-w

Large-scale pancreatic cancer detection via non-contrast CT and deep learning

Received: 9 February 2023

Accepted: 12 October 2023

Published online: 20 November 2023

Check for updates

Kai Cao[1,19], Yingda Xia[2,19], Jiawen Yao[3,4,19], Xu Han[5,19], Lukas Lambert[6,19], Tingting Zhang[7,19], Wei Tang[8,19], Gang Jin[9], Hui Jiang[10], Xu Fang[1], Isabella Nogues[11], Xuezhou Li[1], Wenchao Guo[3,4], Yu Wang[3,4], Wei Fang[3,4], Mingyan Qiu[3,4], Yang Hou[12], Tomas Kovarnik[13], Michal Vocka[14], Yimei Lu[8], Yingli Chen[9], Xin Chen[15], Zaiyi Liu[15], Jian Zhou[16], Chuanmiao Xie[16], Rong Zhang[16], Hong Lu[17], Gregory D. Hager[18], Alan L. Yuille[18], Le Lu[2], Chengwei Shao[1], Yu Shi[12], Qi Zhang[5], Tingbo Liang[5], Ling Zhang[2] & Jianping Lu[1]

Pancreatic ductal adenocarcinoma (PDAC), the most deadly solid malignancy, is typically detected late and at an inoperable stage. Early or incidental detection is associated with prolonged survival, but screening asymptomatic individuals for PDAC using a single test remains unfeasible due to the low prevalence and potential harms of false positives. Non-contrast computed tomography (CT), routinely performed for clinical indications, offers the potential for large-scale screening, however, identification of PDAC using non-contrast CT has long been considered impossible. Here, we develop a deep learning approach, pancreatic cancer detection with artificial intelligence (PANDA), that can detect and classify pancreatic lesions with high accuracy via non-contrast CT. PANDA is trained on a dataset of 3,208 patients from a single center. PANDA achieves an area under the receiver operating characteristic curve (AUC) of 0.986–0.996 for lesion detection in a multicenter validation involving 6,239 patients across 10 centers, outperforms the mean radiologist performance by 34.1% in sensitivity and 6.3% in specificity for PDAC identification, and achieves a sensitivity of 92.9% and specificity of 99.9% for lesion detection in a real-world multi-scenario validation consisting of 20,530 consecutive patients. Notably, PANDA utilized with non-contrast CT shows non-inferiority to radiology reports (using contrast-enhanced CT) in the differentiation of common pancreatic lesion subtypes. PANDA could potentially serve as a new tool for large-scale pancreatic cancer screening.

7

7

Predictions Interact with Decisions

JOURNAL ARTICLE

Human Decisions and Machine Predictions*

Jon Kleinberg, Himabindu Lakkaraju, Jure Leskovec, Jens Ludwig, Sendhil Mullainathan

The Quarterly Journal of Economics, Volume 133, Issue 1, February 2018, Pages 237–293,
https://doi.org/10.1093/qje/qjx032
Published: 26 August 2017

PDF Split View Cite Permissions Share ▾

Abstract

Can machine learning improve human decision making? Bail decisions provide a good test case. Millions of times each year, judges make jail-or-release decisions that hinge on a prediction of what a defendant would do if released. The concreteness of the prediction task combined with the volume of data available makes this a promising machine-learning application. Yet comparing the algorithm to judges proves complicated. First, the available data are generated by prior judge decisions. We only observe crime outcomes for released defendants, not for those judges detained. This makes it hard to evaluate counterfactual decision rules based on algorithmic predictions. Second, judges may have a broader set of preferences than the variable the algorithm predicts; for instance, judges may care specifically about violent crimes or about racial inequities. We deal with these problems using different econometric strategies, such as quasi-random assignment of cases to judges. Even accounting for these concerns, our results suggest potentially large welfare gains: one policy simulation shows crime reductions up to 24.7% with no change in jailing rates, or jailing rate reductions up to 41.9% with no increase in crime rates. Moreover, all categories of crime, including violent crimes, show reductions; these gains can be achieved while simultaneously reducing racial disparities. These results suggest that while machine learning can be valuable, realizing this value requires integrating these tools into an economic framework: being clear about the link between predictions and decisions; specifying the scope of payoff functions; and constructing unbiased decision counterfactuals.

JEL: C10 - General, C55 - Large Data Sets: Modeling and Analysis, K40 - General
Issue Section: Article

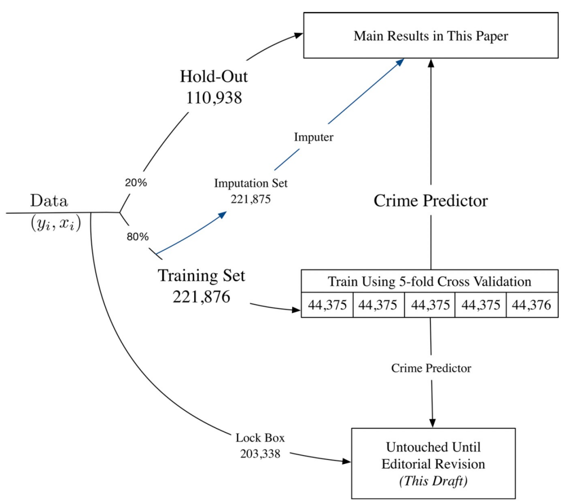

FIGURE I

Partition of New York City Data (2008–13) into Data Sets Used for Prediction and Evaluation

8

When Do Predictions Make No Sense?

- You are not predicting sufficiently important macro economic/political/natural outcomes.

- Your prediction is neither accurate nor causal for decision-making.

$$\frac{d\pi(X_0, Y)}{dX_0} = \frac{\partial \pi}{\partial X_0} \underbrace{(Y)}_{\text{prediction}} + \frac{\partial \pi}{\partial Y} \underbrace{\frac{\partial Y}{\partial X_0}}_{\text{causation}} .$$

Your prediction of Y is not accurate.

Your causal identification is not clean.

- Your predictions of the counterfactual outcomes are ungrounded because of the violation of unconfoundedness (a.k.a. CIA) and/or common support (a.k.a. overlapping condition) assumptions.

9

9