**《机器学习与经济学实证应用》**

**阅读文献**

(2023年秋季)

**（文末注\*文献为课程汇报时可以认领的文献）**

**1、算法教材**

1. Burkov A., *The Hundred-Page Machine Learning Book*. Quebec City, Can.: Andriy Burkov, 2019. （入门）
2. 陈强，《机器学习及Python应用》，北京：高等教育出版社，2021年3月。（中等）
3. James, G., Witten, D., Hastie, T., and Tibshirani, R., *An Introduction to Statistical Learning*, Springer, 2013. （中等）
4. Hastie, T., Tibshirani, R., and Friedman, F., *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Second Edition, Springer, 2017. （较难）

**2、大数据与机器学习综述与概论**

1. Athey, S., “The Impact of Machine Learning on Economics”, Chapter in NBER book The Economics of Artificial Intelligence: An Agenda, 2019, p.507-547, edited by Agrawal, A., Gans, J., and Goldfarb, A.（较难）
2. Athey, S., and Imbens, G., “Machine Learning Methods Economists Should Know About”, *Annual Review of Economics*, 2019, 11(1), 685-725.（较难）
3. Athey, S., and Luca, M., “Economists (and Economics) in Tech Companies”, *Journal of Economic Perspectives*, 2019, 33(1), 209-230. （初级）
4. Kleinberg, J., Ludwig, J., Mullainathan, S., and Obermeyer, Z., “Prediction Policy Problems”, *American Economic Review*, 2015, 105(5), 491-95. （中等）
5. Mullainathan, S., and Spiess, J., “Machine Learning: An Applied Econometric Approach”, *Journal of Economic Perspectives*, 2017, 31(2), 87-106.（中等）
6. Varian, H. R., “Big Data: New Tricks for Econometrics”, *Journal of Economic Perspectives*, 2014, 28(2), 3-28. （中等）
7. 黄乃静、于明哲，《机器学习对经济学研究的影响研究进展》，《经济学动态》，2018年第7期，第115-129页。（初级）
8. 洪永淼、汪寿阳，《大数据、机器学习与统计学：挑战与机遇》，《计量经济学报》，2021年第1期（创刊号），第17-35页。（初级）
9. 王芳、王宣艺、陈硕，《经济学研究中的机器学习：回顾与展望》，《数量经济技术经济研究》，2020年第4期，第146-164页。（初级）

**3、自然语言处理综述**

1. Berger, J., Humphreys, A., Ludwig, S., Moe, W., Netzer, O., and Schweidel, D., “Uniting the Tribes: Using Text for Marketing Insight”, *Journal of Marketing*, 2020, 84(1), 1-25.（初级）
2. Gentzkow, M., Kelly,T. B. and Taddy, M., “Text as data”, *Journal of Economic Literature*, 2019, 57 (3), 535-74..（较难）
3. Loughran, T., and McDonald, B., "Textual Analysis in Accounting and Finance: A Survey." *Journal of Accounting Research*, 2016, 54(4), 1187-1230. （中等）
4. 马长峰、陈志娟、张顺明，《基于文本大数据分析的会计和金融研究综述》，《管理科学学报》，2020年第9期，第19-30页。（初级）
5. 沈艳、陈赟、黄卓，《文本大数据分析在经济学和金融学中的应用：一个文献综述》，《经济学季刊》，2019年第4期，第1153-1186页。（初级）

**4、经济预测**

1. Bianchi, D., Büchner, M., Tamoni A., “Bond Risk Premiums with Machine Learning”, *Review of Financial Studies*, 2021, 34(2), 1046-1089. （多算法，中等\*）
2. Gu, S., Kelly, B., and Xiu, D., "Empirical Asset Pricing via Machine Learning”, *The Review of Financial Studies*, 2020, 33(5), 2223-2273. （资产定价，中等\*）
3. Ke, Z. T., Kelly, B. T., and Xiu, D., “Predicting Returns With Text Data”, NBER Working Paper No. 26186, 2019. （有监督学习，中等\*）
4. Leippold, M., Wang, Q., and Zhou, W., “Machine Learning in the Chinese Stock Market”, *Journal of Financial Economics*, 2022, 134(2), 64-82. (多算法，中等\*)

**5、变量生成**

1. Dubé, J., and Misra, S., “Scalable Price Targeting”, NBER Working Paper No. 23775, 2017.（Lasso回归，中等\*）
2. Gründler, K., and Krieger, T., "Democracy and Growth: Evidence from a Machine Learning Indicator", *European Journal of Political Economy*, 2016, *45*, 85-107.（用ML测度Democracy，中等\*）
3. Schöll, N., Gallego, A., and Le Mens, G., "Politician-Citizen Interactions and Dynamic Representation: Evidence from Twitter," Working Papers 1238, 2021, Barcelona Graduate School of Economics. (深度学习算法，男女差异，中等\*)
4. Miric, M., Jia2, N., and Huang, K., G., “Using Supervised Machine Learning for Large-scale Classification in Management Research: The Case for Identifying Artificial Intelligence Patents”, *Strategic Management Journal*, 2023, 44, 491-519. （有监督学习，AI专利识别，中等\*）

**6、文本相似度**

1. Iaria, A., Schwarz, C., and Waldinger, F., “Frontier Knowledge and Scientific Production: Evidence from the Collapse of International Science”, *Quarterly Journal of Economics*, 2018, 133(2), 927-991.（文本相似度计算，较难\*）
2. Kelly, B., Papanikolaou, D., Seru, A., and Taddy, M., “Measuring Technological Innovation over the Long Run”, *American Economic Review: Insights*, 2021, 3(3), 303-320. (文本相似度计算，中等)

**7、文本情感分类**

1. Cookson, J. A. and Niessner, M., “Why Don't We Agree? Evidence from a Social Network of Investors”, *Journal of Finance*, 2020, 75(1), pp.173-228. (中等\*)
2. Li, J., Chen, Y., Shen, Y., Wang, J. and Huang, X., “Measuring China's Stock Market Sentiment”, Working Paper, 2019. (卷积神经网络，中等\*)
3. Shapiro, A., H., Sudhof, M., and Wilson, D., “Measuring News Sentiment”, *Journal of Econometrics*, 2022, 228(2), 221-243. （媒体情绪，中等\*）

**8、LDA主题模型**

1. Bandiera, O., Hansen, S., Prat, A., and Sadun, R., “CEO Behavior and Firm Performance”, *Journal of Political Economy*, 2020, 128(4), 1325–1369. (中等\*)
2. Bellstam, G., Bhagat, S., and Cookson, J. A., “A Text-Based Analysis of Corporate Innovation”, *Management Science*, 2021, 67(7), 3985-4642. （分析师报告）
3. Brown, N. C., Crowley, R. M., and Elliott, W. B., “What Are You Saying? Using topic to Detect Financial Misreporting”, *Journal of Accounting Research*, 2020, 58(1), pp.237-291. （LDA主题模型）
4. Hansen, S., McMahon, M., and Prat, A., “Transparency and Deliberation Within the FOMC: A Computational Linguistics Approach”, *Quarterly Journal of Economics*, 2018, 133(2), 801–870. （央行沟通，中等\*）
5. Mueller, H., and Rauh, C., “Reading Between the Lines: Prediction of Political Violence Using Newspaper Text, *American Political Science Review*, 2018, 112(2), 358-375. （较难\*）
6. Wong, F., Wong, T. J., Zhang, T., “Politics and Idiosyncrasy of Information: Evidence from Financial Analysts’ Earnings Forecasts in a Relationship-based Economy”, Working Paper, 2018. (中等\*)
7. 王靖一、黄益平，《金融科技媒体情绪的刻画与对网贷市场的影响》，《经济学季刊》，2018年第17卷第4期，第1623-1650页。（多算法，较难\*）

**9、文本可读性**

1. 陈霄、叶德珠、邓洁，《借款描述的可读性能够提高网络借款成功率吗》，《中国工业经济》，2018年第3期，第174-192页。（文本可读性，中等\*）
2. 丘心颖、郑小翠、邓可斌，《分析师能有效发挥专业解读信息的作用吗？——基于汉字年报复杂性指标的研究》，《经济学季刊》，2016年第15卷第4期，第1483-1506页。（文本可读性，中等\*）

**10、无监督学习**

1. Athey, S., and Mobius, M., and Pál, J., “The Impact of Aggregators on Internet News Consumption”, NBER Working Papers 28746, National Bureau of Economic Research. (数据降维，中等\*)
2. Qin, B., Stromberg, D., and Wu, Y., “Media Bias in China”, *American Economic Review*, 2018, 108(9), 2442-2476. （主成分分析，中等\*）
3. Fern´andez-Dur´an1, J., Gregorio-Dom´ınguez, M., “Consumer Segmentation Based on Use Patterns”, *Journal of Classification*, 2021, 38, 72–88. (K-mean聚类)
4. Elkins, R., and Schurer, S., “Exploring the Role of Parental Engagement in Non-cognitive Skill Development over the Lifecourse”, *Journal of Population Economics*, 2020, 33, 957-1004.

**11、其他自然语言处理**

1. Gentzkow, M., Shapiro,J., and Taddy, M., "Measuring Group Differences in High‐Dimensional Choices: Method and Application to Congressional Speech", *Econometrica*, 2019, 87(4), 1307-1340. （国会演讲，中等\*）
2. Hassan, T., Hollander, S., van Lent, L., Tahoun, A., “Firm-Level Political Risk: Measurement and Effects", *The Quarterly Journal of Economics*, 2019, 134(4), 2135-2202. （政治风险，中等\*）
3. Hillert, A., Jacobs, H., and Müller, S., “Journalist Disagreement”, *Journal of Financial Markets*, 2018, 41, 57-76. (媒体分歧，词典法，中等\*)
4. Hoberg G., and Phillips, G., "Text-Based Network Industries and Endogenous Product Differentiation", *Journal of Political Economy*, 2016, 124(5), 1423-1465. (用文本对行业重新分类，中等\*)
5. Li, K., Mai, F., Shen, R., and Yan, X., “Measuring Corporate Culture Using Machine Learning”, *Review of Financial Studies*, 2021, 34(7), 3265-3315. (词嵌入，中等\*)
6. Manela, A., and Alan Moreira, A., “News Implied Volatility and Disaster Concerns”, *Journal of Financial Economics*, 2017, 123, 137-162.（隐含波动率，中等\*）
7. Zhong, W., and Chan J. T., “Reading China: Predicting Policy Change with Machine Learning”, AEI Economics Working Paper Series, 2018-11, 2018.（报纸隐含政策变动指数，词嵌入与循环神经网络；中等\*）

**12、机器学习与因果推断**

**12.1 综述**

1. Guo, R., Cheng, Lu, Li, J., Hahn, R., and Liu, H., “A Survey of Learning Causality with Data: Problems and Methods”, *ACM Trans*, 2020, 9(4), Articale 39. (综述，较难)
2. Kreif, N., and K. DiazOrdaz, “Machine Learning in Policy Evaluation: New Tools for Causal Inference”, arXiv preprint, arXiv: 1903.00402, 2019, Available at https://arxiv.org/abs/1903.00402.
3. 郭峰、陶旭辉，《机器学习与社会科学中的因果关系：一个文献综述》，《经济学季刊》，2023年第1期。（综述，中等）

**12.2 工具变量**

1. Belloni, A., Chen D., Chernozhukov, V., and Hansen, C., “Sparse Models and Methods for Optimal Instruments with an Application to Eminent Domain”, *Econometrica*, 2012, 80(6), 2369-2429.（Lasso方法挑选工具变量，较难\*）
2. Gilchrist, D. S., and Sands, E. G., “Something to Talk About: Social Spillovers in Movie Consumption”, *Journal of Political Economy*, 2016, 124(5), 339-1382. (用Lasso挑选工具变量，中等\*)
3. Godzinski, A., and Castillo, M., “Disentangling the Effects of Air Pollutants with Many Instruments, *Journal of Environmental Economics and Management*, 2021, 109, 102489. (用Lasso挑选工具变量，中等\*)
4. Hansen, C., and Kozbur, D., “Instrumental Variables Estimation with Many Weak Instruments using Regularized JIVE”, *Journal of Econometrics*, 2014, 182(2), 290-308.（岭回归挑选工具变量，较难\*）

**12.3 匹配法**

1. Mozer, R., Miratrix, L., Kaufman, A. R., and Anastasopoulos, L. J., “Matching with Text Data: An Experimental Evaluation of Methods for Matching Documents and of Measuring Match Quality”, arXiv preprint, arXiv:1801.00644v7, 2019, Available at <https://arxiv.org/abs/1801.00644v7>. (构建对照组，文本数据匹配，中等\*)
2. Robertsy, M., Stewartz, B., and Nielsen, R., “Adjusting for Confounding with Text Matching”, *American Journal of Political Science, 2020, 64(4), 887-903*.（构建对照组，文本大数据的匹配问题, 中等\*）

**12.4 双重差分法**

1. Guo, F., Huang, Y., Wang, J., and Wang, X., “The Informal Economy at Times of COVID-19 Pandemic”, *China Economic Review*, 2022, 71, 101722.

**12.5 合成控制**

1. Arkhangelsky, D., Athey, S., Hirshberg, D., Imbens, G., Wager, S., “Synthetic Difference-in-Differences”, *American Economic Review*, 2021, 111(12), pp.4088-4118.
2. Kinn, D., “[Synthetic Control Methods and Big Data](https://ideas.repec.org/p/arx/papers/1803.00096.html)", arXiv:1803.00096, 2018. （构建对照组，合成控制，中等\*）
3. Mühlbach, N. N., “Tree-based Synthetic Control Methods: Consequences of moving the US Embassy”, Institut for Økonomi, Aarhus Universitet. CREATES Research Papers, No. 2020-04, 2020.（构建对照组，用随机森林进行合成控制，中等\*)
4. Doudchenko, N., and Imbens, G., W., “Balancing, Regression, Difference-In-Differences and Synthetic Control Methods: A Synthesis”, NBER Working Paper No. 22791, 2016.（构建对照组，DID、合成控制，中等\*）

**13、机器学习与异质性因果**

1. Knaus, C. M., M. Lechner, and A. Strittmatter, “Heterogeneous Employment Effects of Job Search Programmes: A Machine Learning Approach”, *Journal of Human Resources*, 2020, 0718-9615R1. (中等\*)
2. Seungwoo, C., M. E. Kahn, and M. H. Roger, “Estimating the Gains from New Rail Transit Investment: A Machine Learning Tree Approach”, *Real Estate Economics*, 2018, 48(3), 1-29. (中等\*)
3. Davis, J. M., and S. B. Heller, “Rethinking the Benefits of Youth Employment Programs: The Heterogeneous Effects of Summer Jobs”, *Review of Economics and Statistics*, 2017, 1-47. (中等\*)
4. 胡安宁、吴晓刚、陈云松，《处理效应异质性分析——机器学习方法带来的机遇与挑战》，《社会学研究》，2021年第1期，第91-114页。(中等\*)