

基于 XGBoost 的低渗油田储层粒度预测^①



李建平¹, 张小庆¹, 李莹²

¹(东北石油大学 计算机与信息技术学院, 大庆 163318)

²(大庆油田有限责任公司 储运销售分公司, 大庆 163453)

通信作者: 李建平, E-mail: 18896169440@163.com

摘要: 针对低渗油田储层粒度预测问题, 本文提出利用机器学习中的极致剃度提升树 (extreme gradient boosting, XGBoost) 来对低渗油田储层粒度进行预测的方案. 首先, 根据问题构建合适的 XGBoost 模型, 然后根据已有的岩心储层粒度特征值与其余测井信息的关系, 选取适用于粒度预测的测井曲线建立样本库, 最后利用样本库数据对建立的 XGBoost 模型进行训练, 训练后的模型即可预测研究区域未知的储层粒度特征. 结果表明, 本文所设计的 XGBoost 模型对低渗油田的储层粒度预测方案在计算效率、预测准确率等方面均优于 BP 神经网络.

关键词: 机器学习; 神经网络; XGBoost; 储层粒度

引用格式: 李建平, 张小庆, 李莹. 基于 XGBoost 的低渗油田储层粒度预测. 计算机系统应用, 2022, 31(2): 241-245. <http://www.c-s-a.org.cn/1003-3254/8325.html>

Prediction of Reservoir Grain Size in Low Permeability Oilfield Based on XGBoost

LI Jian-Ping¹, ZHANG Xiao-Qing¹, LI Ying²

¹(School of Computer & Information Technology, Northeast Petroleum University, Daqing 163318, China)

²(Storage Sales Branch, Daqing Oilfield Co. Ltd., Daqing 163453, China)

Abstract: To address the prediction problems of reservoir grain sizes in low permeability oilfields, this study proposes a scheme for predicting reservoir grain sizes in low permeability oilfields with the extreme gradient boosting (XGBoost) in machine learning. First, a proper XGBoost model is built in consideration of the problems. Then, well logging curves suitable for grain size prediction are selected to create a sample database according to the established relationships of the characteristic values of the core reservoir grain size with other logging information. Finally, sample database data are employed to train the newly built XGBoost model. The trained model can predict unknown reservoir grain size characteristics in a study area. The results show that the XGBoost model designed in this study is superior to the back propagation (BP) neural network in calculation efficiency and prediction accuracy of reservoir grain sizes in low permeability oilfields.

Key words: machine learning; neural network; XGBoost; reservoir grain size

低渗透油气田在我国油气开发中有着重要意义^[1], 如何有效对低渗油田进行开发, 其中关键技术难题之一是如何摸清地层的粒度分布范围, 粒度分析的目的在于研究碎屑岩的粒度大小和分布, 粒度分析不仅有利于分析沉积水动力条件, 而且对于储层评价也有意义.

对于低渗油田^[2], 因渗透率低、丰度低、单井产能低, 比较难准确获取储层粒度, 在以往的测井解释中, 由于主要通过机械筛选、激光衍射、薄片分析等实验方法获得粒度参数, 而这些方法耗时, 且成本较高^[3-5]. 因此, 寻找一种不需要实际开采层位岩芯^[6], 利用声波、电阻

① 收稿时间: 2021-04-23; 修改时间: 2021-05-19, 2021-05-28; 采用时间: 2021-06-02; csa 在线出版时间: 2022-01-17

率等信息来预测储层粒度的方法显得尤为重要。目前粒度纵向剖面连续预测方法相关文献在国外较少, 2013年, Liu等^[7]通过 Sarvak 油田产能评价和物性对产能的影响分析, 研究了中东地区块状碳酸盐岩油藏油井生产模式及产能影响因素。2015年, Hossain等^[8]利用两个不同的油藏数据, 研究了人工蜂群(ABC)优化算法的效果, 为了提高神经网络的性能, 2019年, Abdelrigeeb等^[9]选取了中东地区挥发性油藏 2 445 个数据集和黑色油藏 766 个数据集, 将混合 PSONN 模型与现有的模糊逻辑(FL)模型进行了比较, 提出了一种基于粒子群算法的神经网络权值优化算法。同年, Li等^[10]开发了两种基于神经网络的机器学习技术, 长短期记忆(LSTM)网络和带有卷积层(VAEc)网络的变分自编码器, 用于处理易于获取的地层矿物和流体饱和度测井曲线。在国内, 相关文献更是稀少, 杨宁等^[11]提出利用伽马测井曲线小波变换提取粒度参数, 在前人研究的基础上, 扩大了粒度信息来源, 提高了测井资料使用率。2016年, 王利华等^[12]利用 BP 神经网络将粒度纵向剖面预测精度控制在 10% 以内, 缩短了完井周期, 但计算效率有待改进。多元分析的预测模型及数字化岩心照相技术也被许多研究人员用来预测储层砂粒度, 这些方法虽然能快速预测粒度, 但也存在问题, 如难以确定各类测井曲线与粒度之间的复杂关系, 因此无法用解析式表示目的储层粒度最优拟合模型, 故没有被广泛应用。近年来, 神经网络、机器学习在理论与实践上都取得了重大进展^[13-18], 这些技术在寻求测井数据与储层物性参数之间的线性关系也取得了一定进展, 已被用于岩性识别以及储层孔隙度、渗透率预测等。为节省测井的人力物力, 不用再人工测量, 故考虑将这些技术用于储层粒度预测。经过资料研究发现, 已有运用 BP 神经网络技术预测储层粒度的实例, 但在训练 BP 神经网络预测模型时需要上千甚至上万的数据, 而本次的测井点只有 80 多个。故考虑使用可用更少的数据解决相关问题的极致梯度提升, 即 XGBoost (extreme gradient boosting)。

本文通过研究低渗油田储层粒度与多种测井曲线的响应关系, 发现声波(AC)、自然伽马(GR)、密度(ZDEN)、电阻率(RS)及中子(CNC)与粒度中值具有很好的相关性。从而通过建立伽马测井、声波、密度、电阻率、中子与实测粒度特征值的样本库, 利用机器学习技术, 建立起储层粒度与测井曲线之间的非线性映射关系, 训练出满足工程需要的预测模型, 并运用该模型预

测研究区域未知的储层粒度特征。结果表明, 在使用较少训练数据集的情况下, 基于 XGBoost 的粒度预测方案在计算效率、预测准确率等方面优于普通 BP 神经网络。

1 方法原理

XGBoost 极致梯度提升, 是 Chen 在 2016 年提出的一个分布式通用 Gradient Boosting 库, 旨在高效、灵活且可移植^[19]。XGBoost 实现的是一种通用的 Tree Boosting 算法, 该算法以决策树作为基学习器的集成模型, 在梯度提升的基础上改善目标函数计算方式, 提高模型的精度。

因为 XGBoost 预测储层粒度建立在测井数据的基础上, 所以训练样本和预测样本数据集应由测井曲线和实测粒度数据构成, 可表示如下:

$$A = \{X_{mn}, Y_m\} \quad (1)$$

其中, X_{mn} 为模型的输入, 即训练测井数据集, m 为样本量, n 为测井曲线个数, Y_m 为粒度实值向量。在确定样本数据集后, 预测模型可表示为:

$$F(x_i) = f_0(x_i) + \sum_{k=1}^{K_X} \sum_{j=1}^{T_k} w_{j,k} \cdot \eta \quad (2)$$

其中, $f_0(x_i)$ 为基分类器(即初始模型), x_i 为第 i 个样本; $k = 1, 2, \dots, K_X$ 为循环迭代次数, $j = 1, 2, \dots, T_k$ 为 CART 叶节点, T_k 为第 k 次迭代的 CART 叶节点个数; $w_{j,k}$ 为第 k 次迭代的第 j 叶节点区域的所有样本的拟合值; η 为学习率。 $f_0(x_i)$ 一般由损失函数确定, 即:

$$f_0(x_i) = \arg \min \sum_{i=1}^m L(y_i, \alpha) \quad (3)$$

其中, y_i 为第 i 个样本目标值, α 为使损失函数达到最小的常数。损失函数 $L(y_i, \alpha)$ 确定目标值与预测值之间的差值, 表示为:

$$L(y_i, \alpha) = \sum_{i=1}^n l(y_i, \hat{y}_i) + \sum_{k=1}^K \Omega(f_k) \quad (4)$$

其中, $l(y_i, \hat{y}_i)$ 为样本 x_i 的训练误差, $\Omega(f_k)$ 表示第 k 棵树的正则项。 $w_{j,k}$ 由下式确定:

$$w_{j,k} = - \frac{\sum \partial L(y_i, \hat{y}_{i,k-1}) / \partial \hat{y}_{i,k-1}}{\sum \partial^2 L(y_i, \hat{y}_{i,k-1}) / \partial \hat{y}_{i,k-1}^2 + \lambda} \quad (5)$$

其中, $\hat{y}_{i,k-1}$ 为第 i 个样本在第 $k-1$ 次迭代的预测值; λ 为正则化系数。

结合 XGBoost 原理和本文所应用场景, 可设计如图 1 所示的实验流程图。首先是建立样本库, 选取和粒

度相关性较强的测井曲线作为模型的输入,已有的粒度实值作为模型训练的输出标签(式(1)中的 Y_m)。其中一部分数据用来训练,另一部分用于验证。为了防止过拟合,训练达到一定次数就停止迭代计算。

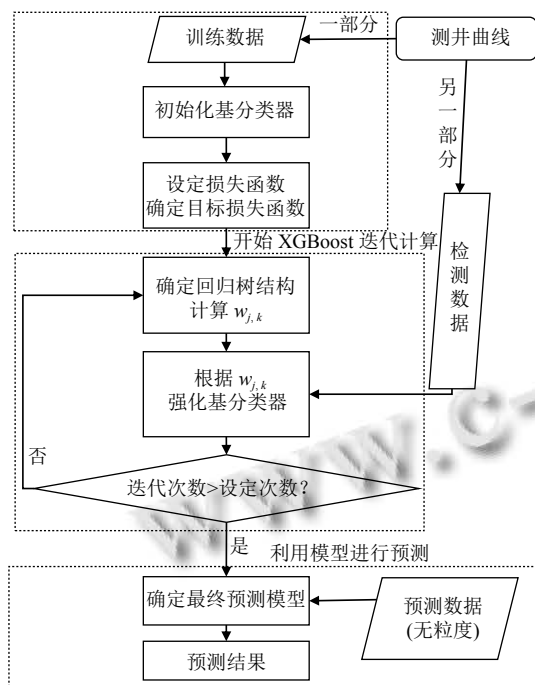


图1 XGBoost 计算流程图

2 储层粒度特征值预测

2.1 测井项与粒度相关性分析

根据已有的低渗油田储层测井资料可知,必须借助各类测井曲线才能实现粒度纵向连续预测,对储层响应比较明显的测井曲线通常包括声波(AC)、中子(CNC)、自然伽马(GR)、电阻率(RS)、密度(ZDEN)等。经作图(图2,红色曲线:伽马测井,蓝色曲线:粒度实值)比较分析可以看出低渗油田中伽马测井曲线较好地反映了粒度大小的变化趋势,这是因为伽马测井值主要由泥质含量决定,伽马测井值随泥质含量增加而变化。泥质含量越高,粒度越小,伽马测井值越高。故伽马测井可作为样本库的一个子项,以预测粒度分布特征。

在同一图里作密度和粒度测试实值变化趋势图(图3,红色曲线:粒度实值,蓝色曲线:密度测井),可以看出低渗油田中密度测井项较好地反映了粒度中值的变化趋势。这是由于密度越高,压实程度越好,孔隙度越小,孔隙度小正好可反映出储层粒度小。故可将密度测井作为样本库的一个子项,以预测粒度纵向分布特征。

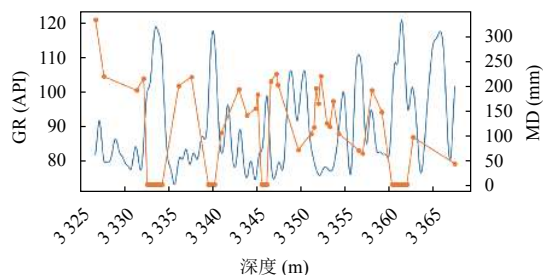


图2 伽马测井与粒度特征相关性分析图

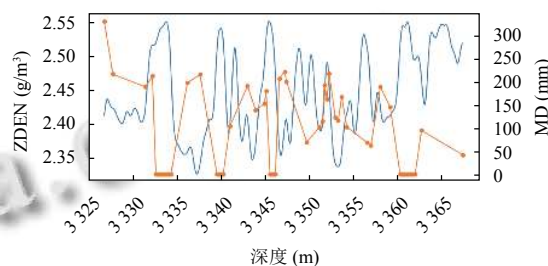


图3 密度测井与粒度特征相关性分析图

通过测井曲线与粒度实值相关性分析发现,AC、CNC、GR、RS、ZDEN测井曲线可以很好反应低渗油田储层粒度的大小变化,故选用这5条曲线作为训练模型的输入。

2.2 建立样本库

样本库的建立直接关系到XGBoost训练误差的精度以及最终粒度预测的精确度。因此,样本库的建立以测井资料与粒度实测数据为基础,其建立过程与XGBoost的学习过程是一个循环分析的过程,样本点的选取以训练学习后的误差为依据。一旦学习误差过大,重新选取样本点进行学习,直到误差控制在工程需要的精度内才建立最终预测模型。

以III井的实测粒度数据与其余测井资料为例建立样本库,共82个样本。选取样本库的80%数据作为训练数据集,剩余的20%作为验证数据集。最终选取AC($\mu\text{s}/\text{m}$)、CNC(g/m^3)、GR(API)、RS(OMM)、ZDEN(g/m^3)这5个测井参数来训练模型,分析误差,对比实验结果,建立测井资料预测纵向剖面粒度的方法。表1展示了部分训练样本数据。

2.3 仿真实验

由于相关性分析后选作特征值预测的参数有5个,故本次使用的对比BP网络(1个输入层,1个输出层以及1个隐藏层),输入层神经元个数为5个,输出层神经元个数为1个,隐藏层包含3个节点。为突出对比效果,本次对比实验使用同一的数据集,两种预测方式使用同样的误差函数—绝对误差(MAE)损失函数,同样

的学习速率、收缩步长、迭代次数. 两种预测模型的部分参数设置如表 2 所示.

表 1 粒度预测样本库 (82 个样本)

深度 (m)	MD (mm)	AC ($\mu\text{s/m}$)	CNC (g/m^3)	GR (API)	RS (OMM)	ZDEN (g/m^3)
3 326.6	333.262	77.881	0.164	83.422	5.945	2.419
3 327.6	219.822	75.511	0.125	79.886	5.657	2.423
3 331.3	192.458	74.619	0.129	83.434	5.715	2.429
⋮	⋮	⋮	⋮	⋮	⋮	⋮
3 345.4	5	79.048	0.148	83.078	4.252	2.549
3 345.5	5	79.538	0.154	84.386	4.276	2.544
3 345.6	5	80.018	0.161	87.313	4.374	2.537
⋮	⋮	⋮	⋮	⋮	⋮	⋮
3 361.8	5	80.608	0.1425	96.857	3.513	2.495
3 362.5	98.855	81.572	0.159	100.52	4.222	2.467
3 367.2	46.185	86.947	0.177	101.676	4.342	2.519

表 2 预测模型参数设置

BP神经网络	XGBoost
激励函数: Sigmoid函数	每棵树随机采样的样本比例: 0.8
误差函数: MAE	误差函数: MAE
学习速率: 0.3	学习速率: 0.3
收缩步长: 0.3	收缩步长: 0.3
最大循环次数: 1 000	迭代计算次数: 1 000
最小梯度值: 0.001	树的最大深度: 6

对归一化后的数据, 模型训练时记录相同训练次数下不同模型训练的损失, 从表 3 可以看出, 对于此次实验用得数据, XGBoost 在正确率方面约为 BP 神经网络的 2 倍. 模型训练完成后对全井段数据进行预测, 在两种模型预测开始时都添加计时器, 预测结束统计用时, 从表 3 可以看出, 预测同一批数据, 在使用相同的迭代次数的模型的情况下, 本文提出的基于 XGBoost 的粒度预测方案计算用时远远低于 BP 神经网络的计算用时, 计算速度方面约为 BP 神经网络的 60 倍. 并且, 在使用的模型训练次数达到 50 次后, 预测全井段用时波动不大.

表 3 BP 和 XGBoost 训练、预测结果统计表

训练模型	训练次数	预测平均耗时 (s)	训练MAE
BP	20	6.870 7	0.163
	50	6.860 7	0.053
	100	6.845 6	0.03
XGBoost	20	0.129 6	0.086
	50	0.106 5	0.025
	100	0.100 1	0.016

对全井段预测完成后, 将 BP 和 XGBoost 的预测结果放在同一图里进行对比 (红色: DUIBI_1.MD_MD_1 为粒度实值, 黑色: DUIBI.MD_XGBOOST_1 为 XGBoost 预测值, 蓝色: DUIBI.MD_BP_1 为 BP 预测

值), 图 4 为训练 50 次的模型预测的结果对比图, 图 5 为训练 100 次的模型预测的结果对比图. 以纵向表示井深度 (m), 横向表示粒度大小 (mm), 范围为 1–360 mm, 观察图 4、图 5 中的曲线, 可发现: 对于同一样本库, 在使用同样的误差函数、学习速率、收缩步长、迭代次数的情况下, 在边界值设置相同时, 当把 BP 预测值, XGBoost 预测值和粒度实值三者放在一起进行对比, 可明显发现, BP 预测的结果和 XGBoost 预测的结果, 在整体趋势变化上一致, 都和粒度实质走势吻合, 但是在具体的层位上, XGBoost 预测的粒度值比 BP 预测的粒度值和粒度实值的吻合度更高, BP 预测的结果偏低. 在某些层位 XGBoost 预测的结果突变较高, 而 BP 结果更趋于平缓. 使用训练 50 次和训练 100 次的模型预测结果波动不大. 表明在只有 82 个粒度真实值的情况下, 选取该井数据来进行模型训练, BP 神经网络训练结果达不到工程要求的精度, 不适用于工程预测, 而 XGBoost 具有良好的学习效率, 可以满足工程预测精度要求.

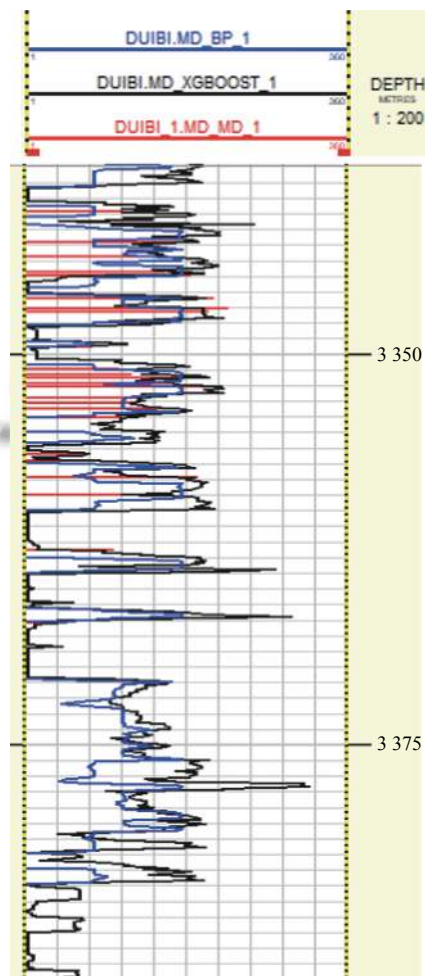


图 4 模型训练 50 次预测对比图

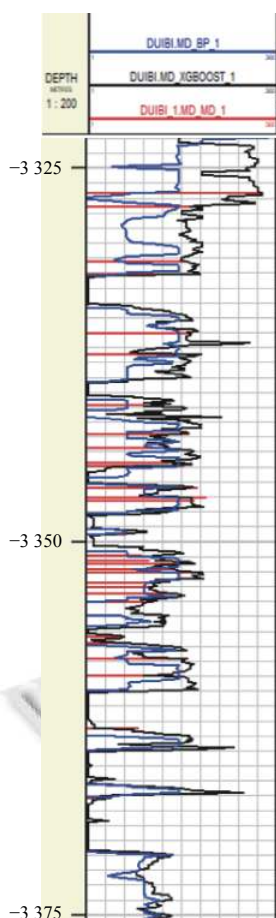


图5 模型训练 100 次预测对比图

3 结论

(1) 本文将 XGBoost 技术引入低渗油田储层测井解释, 实现了粒度特征值纵向剖面连续预测. 该方法克服了常规粒度测试方法岩芯数量少, 取芯成本高的缺点.

(2) 在真实数据较少的情况下, 粒度纵向剖面预测误差较小, 在可接受范围之内, 利用该方法对低渗油田进行开发, 降低了耗时, 节约了成本.

(3) 由于数据样本较少, 使用的模型在训练次数达到 50 次后, 预测结果波动不大.

参考文献

- 张金庆, 杨凯雷, 梁斌. 我国海上低渗油田分类标准研究. 中国海上油气, 2012, 24(6): 25–27.
- 童天喜, 刘敏. 成本控制的内生动力. 中国石油企业, 2014, (1–2): 100–101.
- 陈欢庆, 舒治睿, 林春燕, 等. 粒度分析在砾岩储层沉积环境研究中的应用——以准噶尔盆地西北缘某区克下组冲积扇储层为例. 西安石油大学学报(自然科学版), 2014, 29(6): 6–12, 34.

- 阎媛子. 红河油田长 6 段储层粒度分析. 中国石油和化工标准与质量, 2016, 36(12): 64–65. [doi: 10.3969/j.issn.1673-4076.2016.12.038]
- 田宇昕. 粒度分析及其地质应用. 西部探矿工程, 2020, 32(1): 159–160. [doi: 10.3969/j.issn.1004-5716.2020.01.052]
- 李月. 低渗油田提高单井产量技术策略探析. 化学工程与装备, 2020, (1): 146–147.
- Liu H, Guo R, Dong JC, *et al.* Productivity evaluation and influential factor analysis for Sarvak reservoir in South Azadegan oil field, Iran. *Petroleum Exploration and Development*, 2013, 40(5): 627–634. [doi: 10.1016/S1876-3804(13)60082-8]
- Hossain S, El-Shafie A, Wan Mohtar WHM. Application of intelligent optimization techniques and investigating the effect of reservoir size in calibrating the reservoir operating policy. *Water Policy*, 2015, 17(6): 1143–1162. [doi: 10.2166/wp.2015.023]
- Abedelrigeeb AG, Al-Khudafi AM, Baarimah SO, *et al.* Hybrid artificial intelligent approach for choke size estimation in volatile and black oil reservoirs. 2019 First International Conference of Intelligent Computing and Engineering. Hadhramout: IEEE, 2019. 1–5. [doi: 10.1109/ICICE48418.2019.9035198]
- Li H, Misra S, He JB. Neural network modeling of in situ fluid-filled pore size distributions in subsurface shale reservoirs under data constraints. *Neural Computing and Applications*, 2020, 32(8): 3873–3885. [doi: 10.1007/s00521-019-04124-w]
- 杨宁, 王贵文, 赖锦, 等. 应用伽马测井曲线小波变换计算粒度参数. 现代地质, 2012, 26(4): 778–783. [doi: 10.3969/j.issn.1000-8527.2012.04.020]
- 王利华, 楼一珊, 马晓勇, 等. 储层粒度神经网络预测模型研究. 西南石油大学学报(自然科学版), 2016, 38(1): 53–59.
- 房茂军, 曾祥林, 梁丹. 神经网络在储层粒度预测中的应用. 油气藏监测与管理国际会议论文集. 西安: 西安华线网络信息服务有限公司, 2011. 99–104.
- 邓浩阳. 高孔低渗碳酸盐岩储层孔隙结构及物性表征方法研究 [博士学位论文]. 成都: 西南石油大学, 2018.
- 李佳. 基于机器学习的多孔介质渗透率预测研究 [硕士学位论文]. 杭州: 浙江大学, 2019.
- Bagheri M, Rezaei H. Reservoir rock permeability prediction using SVR based on radial basis function kernel. *Carbonates and Evaporites*, 2019, 34(3): 699–707. [doi: 10.1007/s13146-019-00493-4]
- Zhang GY, Wang ZZ, Li HJ, *et al.* Permeability Prediction of isolated channel sands using machine learning. *Journal of Applied Geophysics*, 2018, 159: 605–615. [doi: 10.1016/j.jappgeo.2018.09.011]
- 李建平, 梁胜松, 范友贵. 基于量子神经网络的超深层储层评价. 计算机与数字工程, 2018, 46(12): 2499–2505. [doi: 10.3969/j.issn.1672-9722.2018.12.024]
- Chen T, Guestrin C. XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. San Francisco: ACM, 2016. 785–794.