

机器学习测试探索 PPT

汇报人：孙高飞



架构迎接未来变化
IAS 2018

人工智能的定义

- 大数据+机器学习



A Venn diagram consisting of three concentric circles. The outermost circle is the largest and contains the text '人工智能' (Artificial Intelligence). Inside it is a medium-sized circle containing the text '机器学习' (Machine Learning). The innermost circle is the smallest and contains the text '深度学习' (Deep Learning). All circles are filled with a solid blue color.

人工智能

机器学习

深度学习

分类模型评估指标

- 混淆矩阵
- ROC与AUC
- 召回，精准, F1 Score

测试集混淆矩阵和评估指标

正负例判定阈值

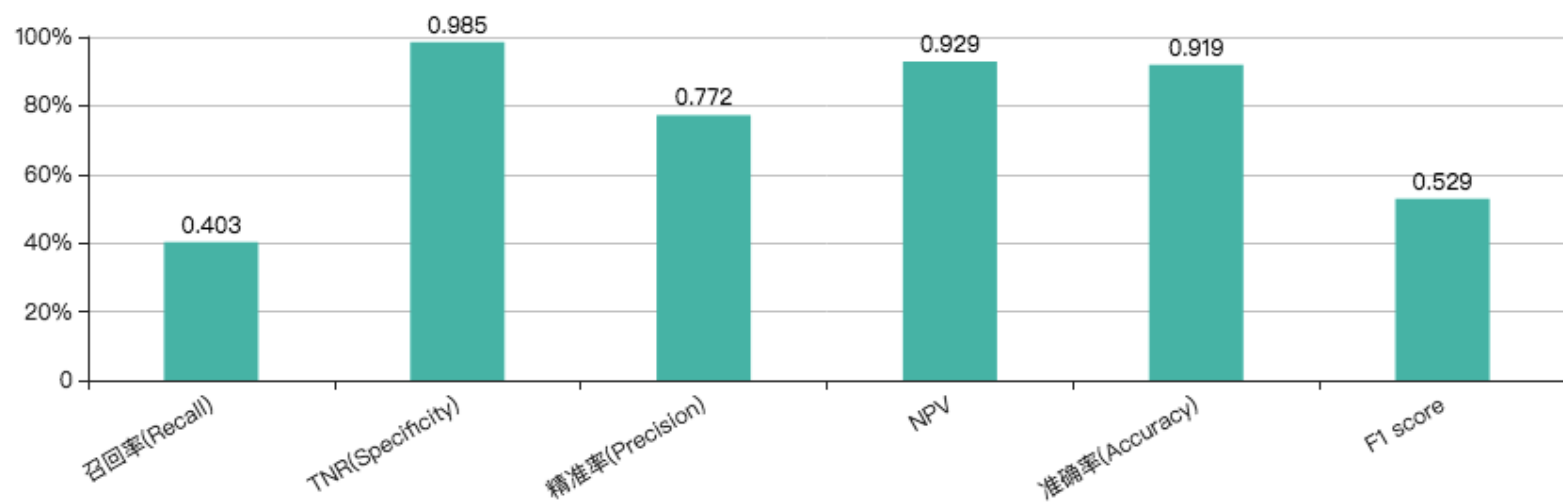


0.50

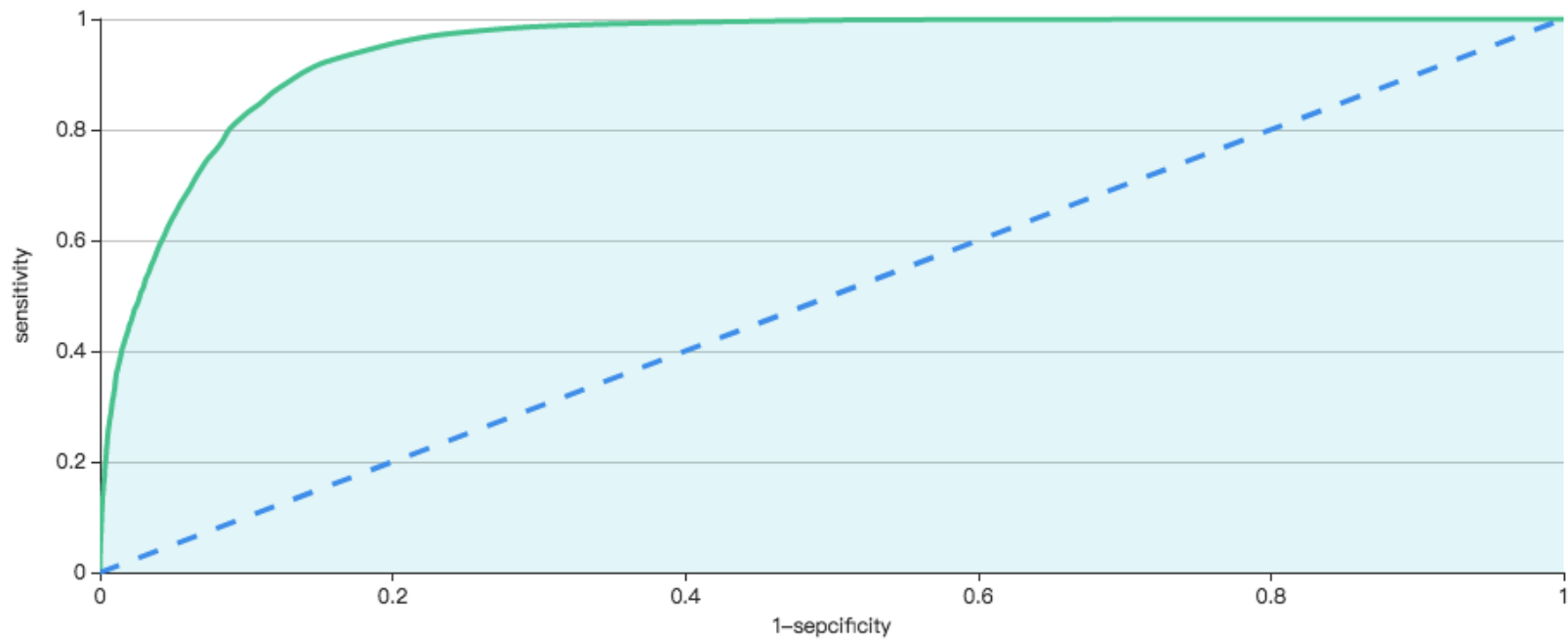
+

-

	模型预测（正）	模型预测（负）	总数
实际值（正）	1869 4.54%	2771 6.73%	4640 11.27%
实际值（负）	552 1.34%	35995 87.39%	36547 88.73%
总数	2421 5.88%	38766 94.12%	



测试数据 Area Under the Curve = 0.947



分组AUC

分组AUC

序号	education	样本数	Group_AUC	样本数占比
1	basic.9y	6045	0.054	0.147
2	high.school	9515	0.049	0.231
3	illiterate	18	0.071	0.000
4	professional.course	5243	0.051	0.127
5	basic.4y	4175	0.051	0.101
6	university.degree	12168	0.060	0.295
7	basic.6y	2292	0.046	0.056
8	unknown	1731	0.062	0.042

回归模型评估指标

- 均方误差

多分类模型评估指标

- 准确率
- TopN的准确率

测试数据的重要性

- 数据决定了模型效果的上限
- 阿尔法Go的案例

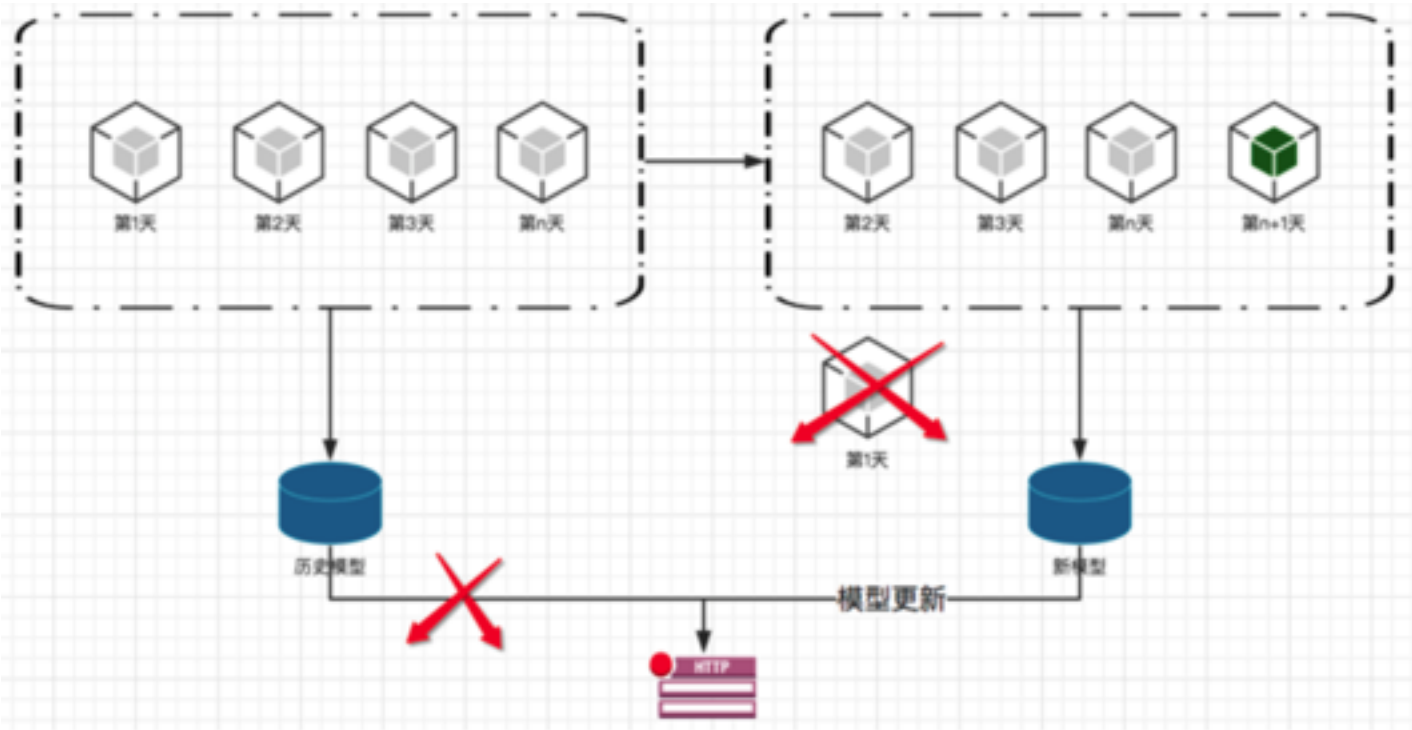
测试数据的采集

- 使用真实数据
- 注意特征呈现时序性
- 注意特征随字段呈现不均匀分布

基于业务场景的测试

- 自学习
- 数据流
- 结合外围系统的测试

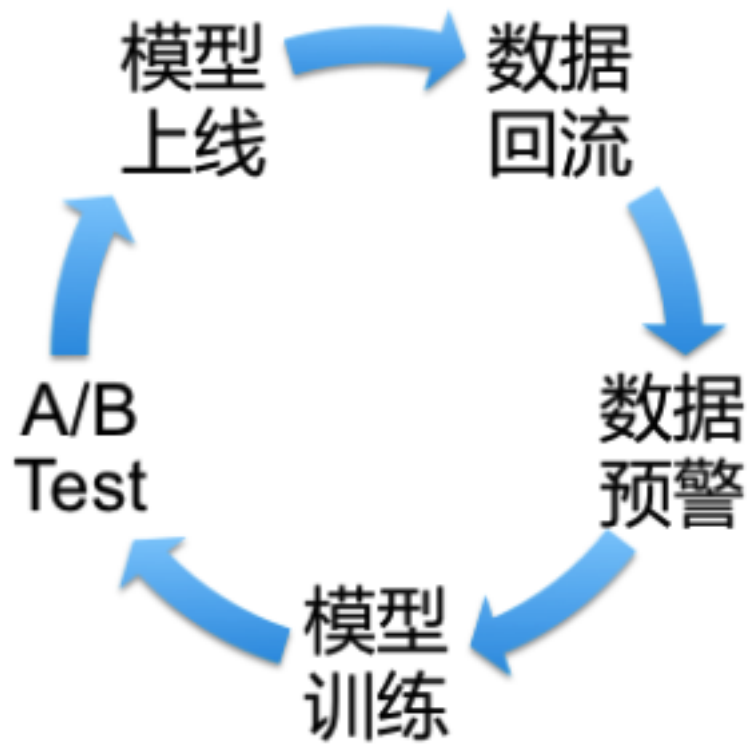
自学习



数据测试

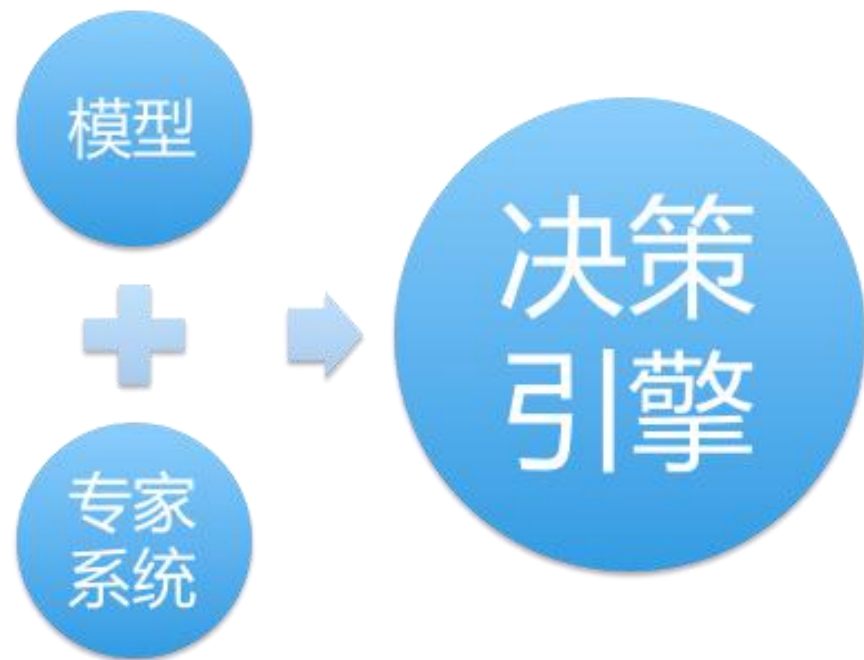
- 扫描数据中的异常值
- 统计数据中的异常分布
- 对比历史数据的差异

数据闭环



预测值的稳定性

score分组	样本占比	真正例	伪正例	伪负例	真负例	准确率	召回	Specificity	本组AUC	积累AUC
(0.900, 1.000]	0.003	25	2	0	0	0.926	1.000	0.000	0.000	0.000
(0.800, 0.900]	0.008	55	14	0	0	0.797	1.000	0.000	0.000	0.000
(0.700, 0.800]	0.013	79	29	0	0	0.731	1.000	0.000	0.001	0.001
(0.600, 0.700]	0.013	68	36	0	0	0.654	1.000	0.000	0.001	0.002
(0.500, 0.600]	0.014	65	48	0	0	0.575	1.000	0.000	0.002	0.004
(0.400, 0.500]	0.015	0	0	59	63	0.516	0.000	1.000	0.003	0.007
(0.300, 0.400]	0.023	0	0	86	105	0.550	0.000	1.000	0.007	0.014
(0.200, 0.300]	0.042	0	0	105	237	0.693	0.000	1.000	0.019	0.032
(0.100, 0.200]	0.120	0	0	199	777	0.796	0.000	1.000	0.080	0.112
(0.000, 0.100]	0.748	0	0	132	5963	0.978	0.000	1.000	0.798	0.910



性能测试

- 访问密集型
- 计算密集型

造数

- 不同的数据类型
- 不同的数据规模
- 不同的数据分布