监控工作业务流程

监控工具业务流程
自己数据源
交易所行情接口
数据库读取
非自己数据源
新闻API
社交媒体爬虫
科研数据库
数据获取层
自己数据
输入：API/数据表连接
处理：
定时拉取行情
历史数据回填记录
多品种/多周期获取
输出：
原始OHLCV数据
时间戳 + 品种标识
数据质量：完整性/时效性/
非自己数据
输入：API/爬虫配置
处理：
定时抓取新闻
社交帖子采集
关键词/实体筛选
去重（URL哈希/内容指纹）
输出：
原始文本数据
时间戳 + 来源标识
数据质量：不相关
数据存储层
自己数据
输入：原始OHLCV
存储结构：
时序数据库（influxDB等）
按品种+周期分表
输出：可查询时序数据
非自己数据
输入：原始文本/结构化数据
存储结构：
文档数据库（MongoDB等）
按来源+类型分集合
输出：可查询时序数据
数据清洗与预处理层
自己数据
输入：原始OHLCV
处理：
缺失值填充（前值/插值）
异常值检测与修复
时间对齐（统一时区）
复权处理（股票）
数据质量评估
输出：
干净的OHLCV序列
数据质量标记
非自己数据
输入：原始文本
处理：
语义清洗（去HTML、特殊字符）
去重（相似内容合并）
时间标准化（统一时间格式）
来源权威评估（建立信誉子品种）
语言检测与过滤
事件交叉验证 ← 特殊处理
输入：
同一事件的多源报道
事件时间的大致窗口
来源元数据（发布者、平台、历史记录）
可信度评估：
来源权威性评分
官方渠道 vs 自媒体 vs 匿名
某些事件更可信或
多方独立报道的事件
多源交叉验证
多方独立报道
多源互动的一致性
跨实体共识
跨实体关联
内部可信度
某些有具体可验证细节
某些结果只有一种解释
某些引用可查证的数据/文件
逻辑矛盾检测：
事件内部矛盾
时间线顺序异常
数字/单位前后不一致
因果关系不合理
事件间矛盾
同一实体的多种表述
时间顺序冲突
与已知事实的矛盾
与自己数据的矛盾
事件时间对应行情无反应
声明的大幅变动但成交量变化不
输出：
可信度评分（多维度）+ 元数据
来源机制/时间线/笔
特征/指标计算层
自己特征
输入：干净的OHLCV
计算内容：
【预点特征】
高低点识别
趋势线拟合
结构特征
成交量比
【突破突破特征】
支撑位/阻力位
通道（ATR带）
突破确认指标
【情绪突破】
情绪加速度
持仓变化
量价共振
输出：
技术指标时间序列
每个时间点的特征向量
非自己特征
输入：干净的文本+元数据
计算内容：
【情绪特征】
情绪分析（正/负/中）
情绪强度
情绪变化率
【事件特征】
事件类型分类
事件重要性评分
时间衰减
与过去相似事件查询
【跨实体特征】
相关/一致性（多空比例）
可比性
联动强度
输出：
情绪/事件指标时间序列
每个时间点的特征向量
单类型信号生成层
自己信号
输入：自己特征向量
生成信号：
【预点信号】
信号类型：顶部/底部
置信度：0.00%
触发时间
【突破信号】
信号类型：向上突破/向下突破
置信度：0.00%
突破方向
【量价共识信号】
信号类型：多头共识/空头共识
置信度：0.00%
输出：
{信号类型, 方向, 置信度, 触发时间, 来源="自己"}
非自己信号
输入：非自己特征向量
生成信号：
【情绪信号】
信号类型：情绪驱动/情绪反转
置信度：0.00%
触发时间
【事件驱动信号】
信号类型：利好/利空
置信度：0.00%
事件类型
【跨实体信号】
信号类型：板块共识/领先信号
置信度：0.00%
输出：
{信号类型, 方向, 置信度, 触发时间, 来源="非自己"}
信号融合层
输入：
自己信号列表：[{预点信号}, {突破信号}, {量价共识信号}]
非自己信号列表：[{情绪信号}, {事件信号}, {跨实体信号}]
融合逻辑：
Step 1: 时间对齐
将不同信号的发生时间到统一时间窗口
Step 2: 同源信号聚合
自己内部：多重共识
非自己内部：情绪共识, 跨实体共识
自己 vs 非自：一致性奖励, 冲突惩罚
非自己内部：可信度加权
Step 3: 跨源交叉验证
方向判断：自己信号 vs 非自信号
置信度调整：自己 > 非自, 非自已内部, 权重)
置信度 = 加权后平均置信度, 冲突 = 惩罚)
冲突惩罚：min(各信号有效期) 或 基于历史统计
Step 4: 综合输出
自己信号 + 非自己信号, 非自已内部, 权重)
置信度 = 加权后平均置信度, 冲突 = 惩罚)
冲突惩罚：min(各信号有效期) 或 基于历史统计
输出：
{
"品种": "BTC/USDT",
"时间": "2026-02-05 10:00:00",
"方向判断": "多",
"方向概率": { "多": "72%", "空": "28%",
"有效期": "白天",
"置信度": 68%,
"主要依据": "情绪突破", "情绪转暖",
"风险提示": "置信不足", "宏观数据发布"}
}
结果验证/回测层
输入：历史信号 + 历史实际走势
验证内容：
方向准确率：信号方向与实际走势一致的比例
置信度校准：64%置信度的信号是否真的有64%准确率
特别关注：高置信度不利的情况(信号反转)
计数型验证：情绪信号更可信
输出：
{
准确率: 0.00%,
置信度校准: 0.00%,
高置信度不利情况: 0.00%,
计数型验证: 情绪信号更可信
}
分析结果/迭代层
输入：验证结果 + 问题分析
迭代内容：
调整特征计算参数
调整信号生成阈值
调整融合权重
调整/删除无效信号
输出：优化后的模型/参数 → 回到第4层重新运行