

联合主办: MEIZU | Flyme | msup[®] | IT大咖说
知识分享平台

数据洞察

魅族技术开放日第八期

金融大数据与人工智能实践

王健宗 博士 平安大数据高级产品总监

“阿法狗”带来的启示



此战成名的智能机器人 AlphaGo 利用“策略网络 (Policy Network)”和“值网络 (Value Network)”两大核心深度神经网络，分析棋盘局面，在可控计算量的范围内判断每步下子策略的优劣。

3月15日，持续七天的围棋“人机世界大战”落下帷幕，谷歌的人工智能机器人 AlphaGo 以4：1大比分战胜围棋世界冠军李世石九段，



人工智能进军金融界



智能投顾，是虚拟机器人基于投资者自身的理财需求，通过算法和产品来完成以往人工提供的理财顾问服务。用户无需掌握太多市场和金融产品知识就可以使用该服务。

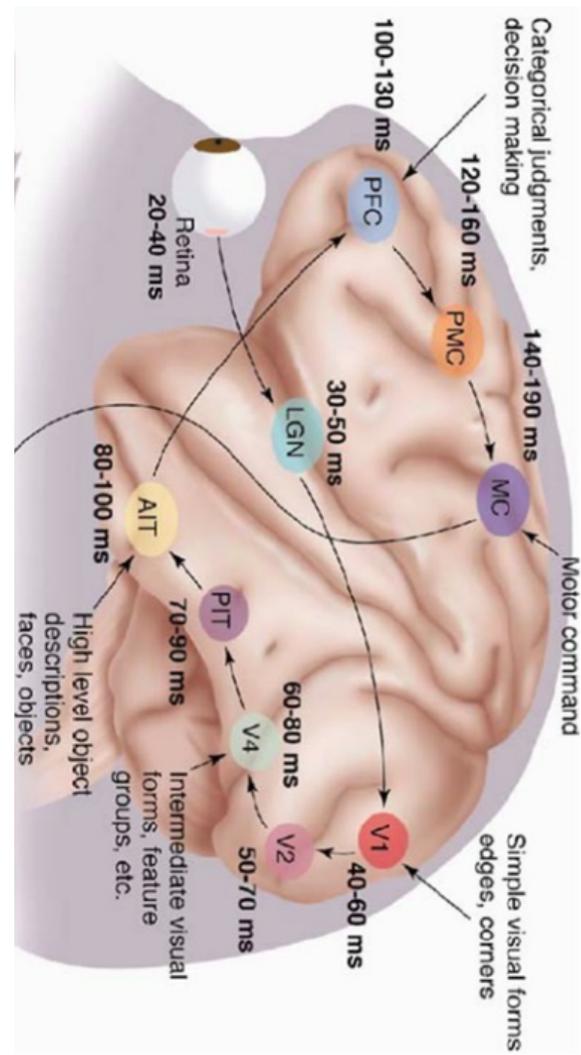
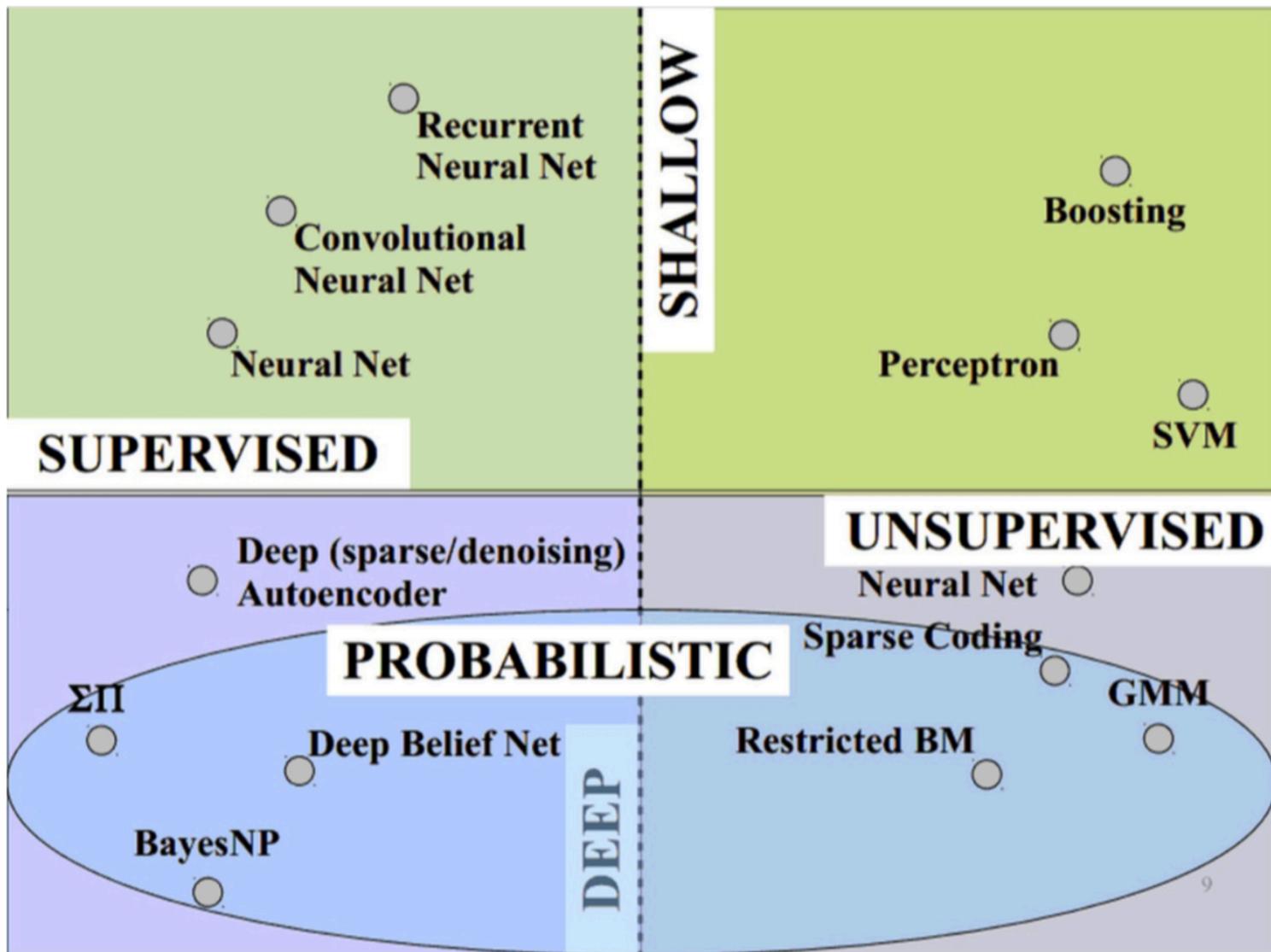
目前，此类人工智能产品已在金融理财崭露头角。

人工智能“脑”

- 各国“脑”计划：
 - 美国公司：谷歌，微软，脸书， ...
 - 中国公司：百度，讯飞，腾讯， ...
 - 政府：美国，欧盟，日本，中国， ...
- 目前主要工作方向：
 - 政府：收集脑活动数据，建模，模拟脑行为
 - 工业界：智能识别（语音，图像，文本， ...）
 - 广泛应用：搜索，广告，推荐，检测，监控，翻译，无人驾驶， ...
 - 核心技术：**深度学习**（Deep Learning）

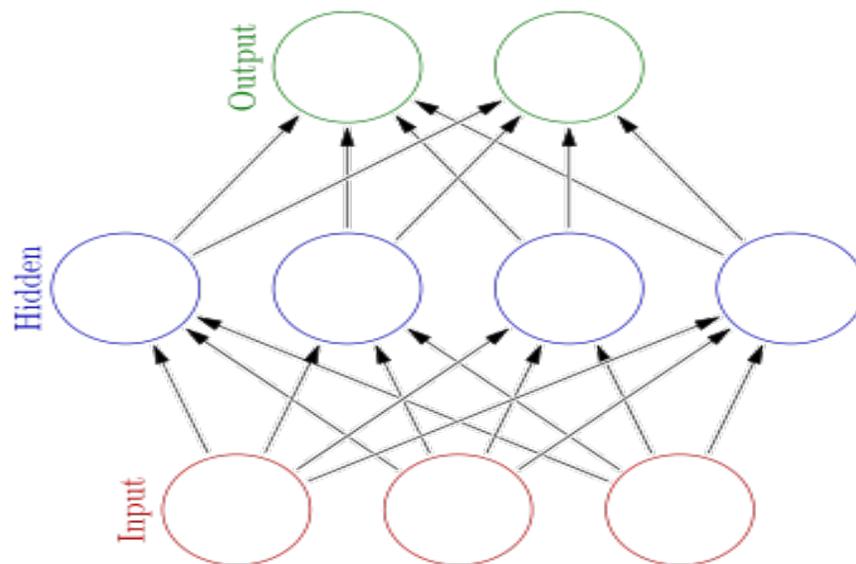


什么是深度学习技术？



人工智能“脑”

- 深度学习：多层人工神经网络
 - 80年代浅层神经网络很火
 - 可从数据中学习非线性模型
 - 90年代后逐渐被别的方法取代
 - 计算能力限制
 - 数据量限制
 - 黑盒子
 - 近5年深度学习炙手可热
 - 神经网络算法改进
 - 数据处理能力大大提高
 - 数据量大大增加
 - 3层 -> 64层



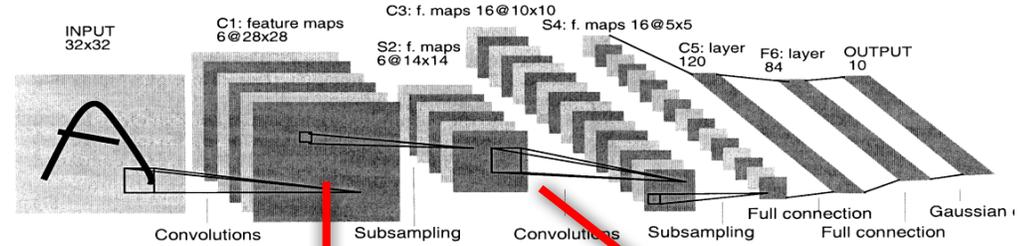
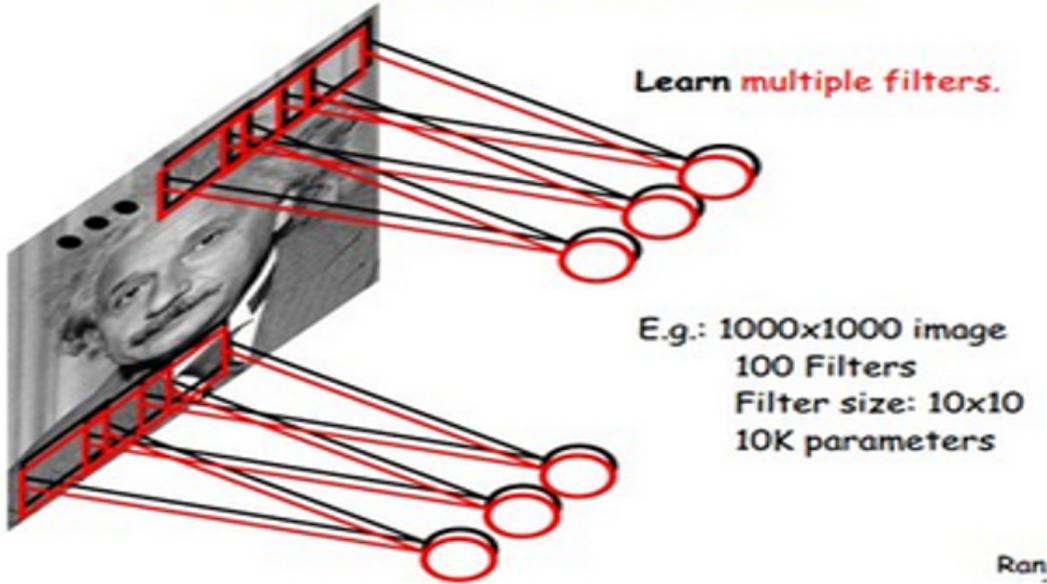
浅层神经网络



64层GoogleNet

基于卷积神经网络的深度学习技术

CONVOLUTIONAL NET

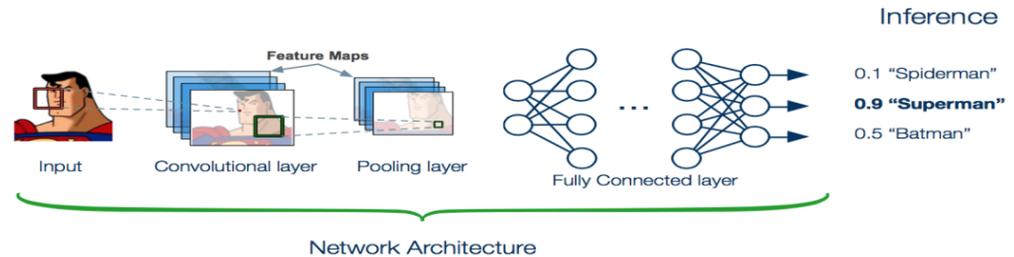


字符识别

卷积层（特征提取层）

降采样层

利用图像的空间联系是局部的，我们使每个神经元只感受局部的图像区域，然后在更高层中，将这些感受不同局部的神经元综合起来就可以得到全局的信息。



物品识别

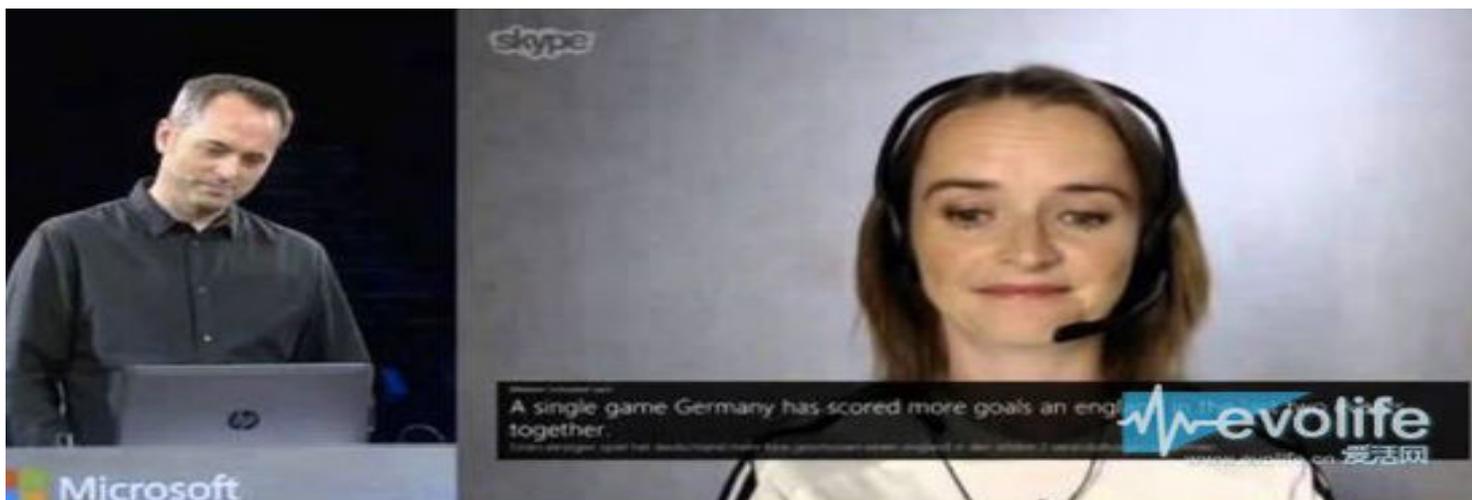
深度学习的应用

- 语音识别

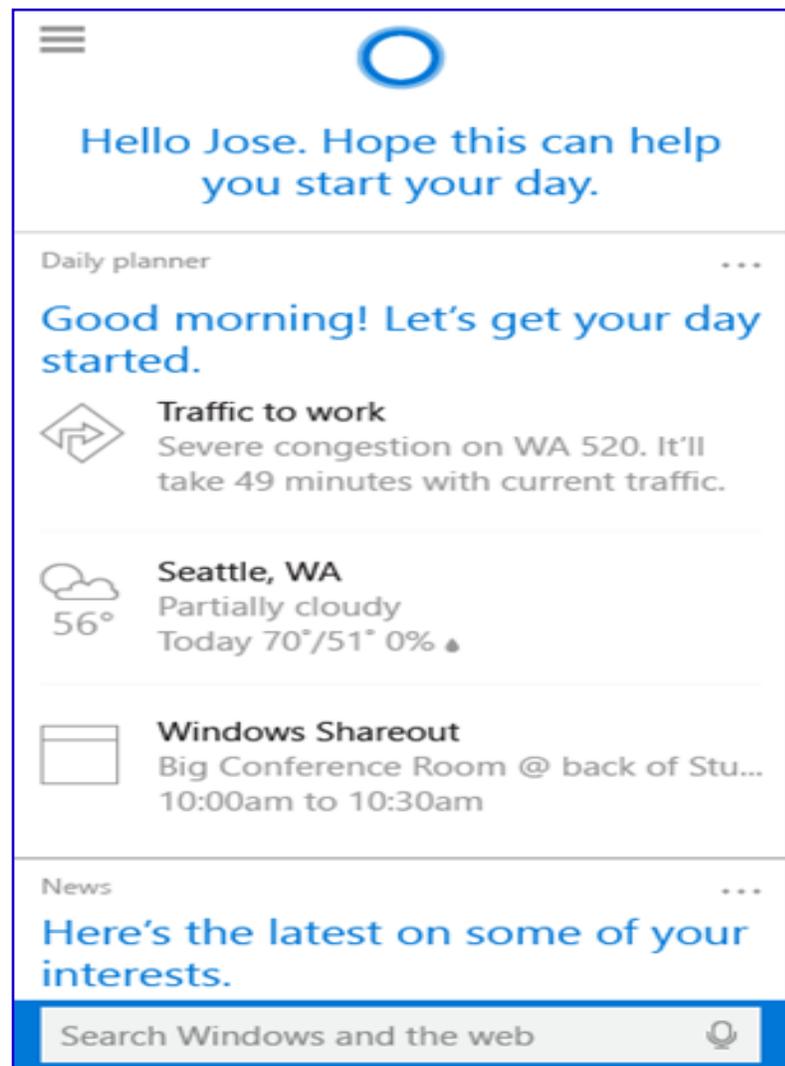
- 成功产品：Google Now，微软 Cortana，苹果Siri，百度语音助手，讯飞，出门问问，...
- 从隐性马尔科夫模型到深度学习：性能飞跃
- 利用成千上万小时的语音数据构建模型

- 语音识别+机器翻译+语音合成

- Skype语音聊天实时翻译
 - 英语，中文，西班牙语，意大利语



Skype语音聊天实时翻译



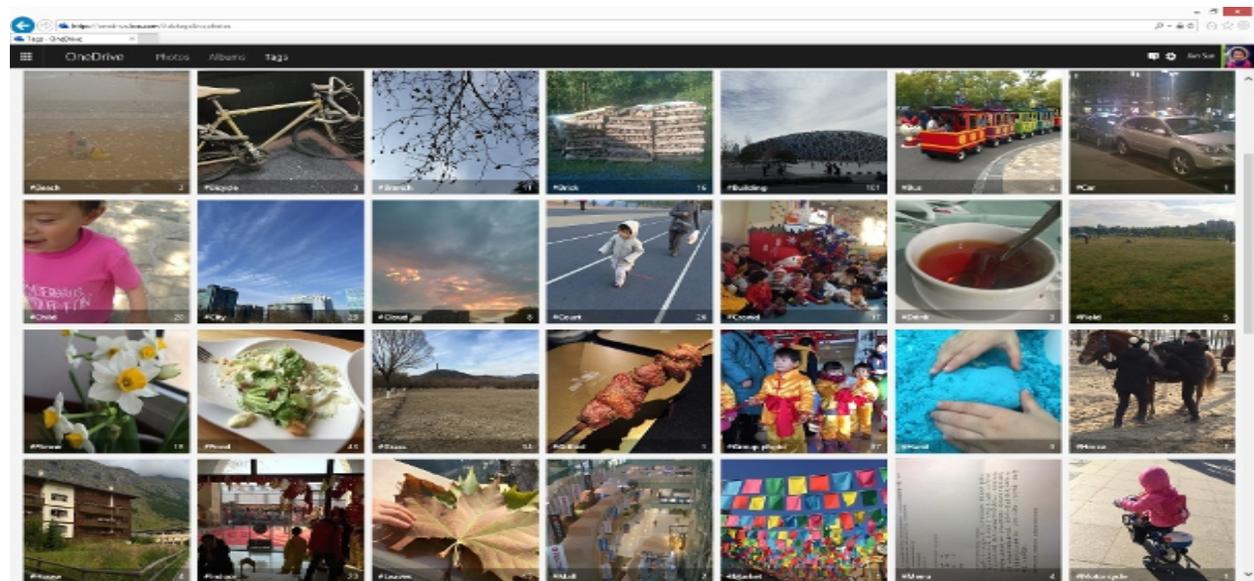
微软Cortana/小冰

深度学习的应用

- 计算机视觉
 - 人脸识别
 - LFW测试准确率近100%
 - 图像识别
 - ImageNet测试识别率超过人
 - 目标检测
 - 以图搜图
 - 看图说话
 - ...
- 各大公司“军备竞赛”：
 - 微软，谷歌，脸书，百度，...
 - 1000类目标图像识别：
 - 20% -> 10% -> 6% -> 5% -> 4%



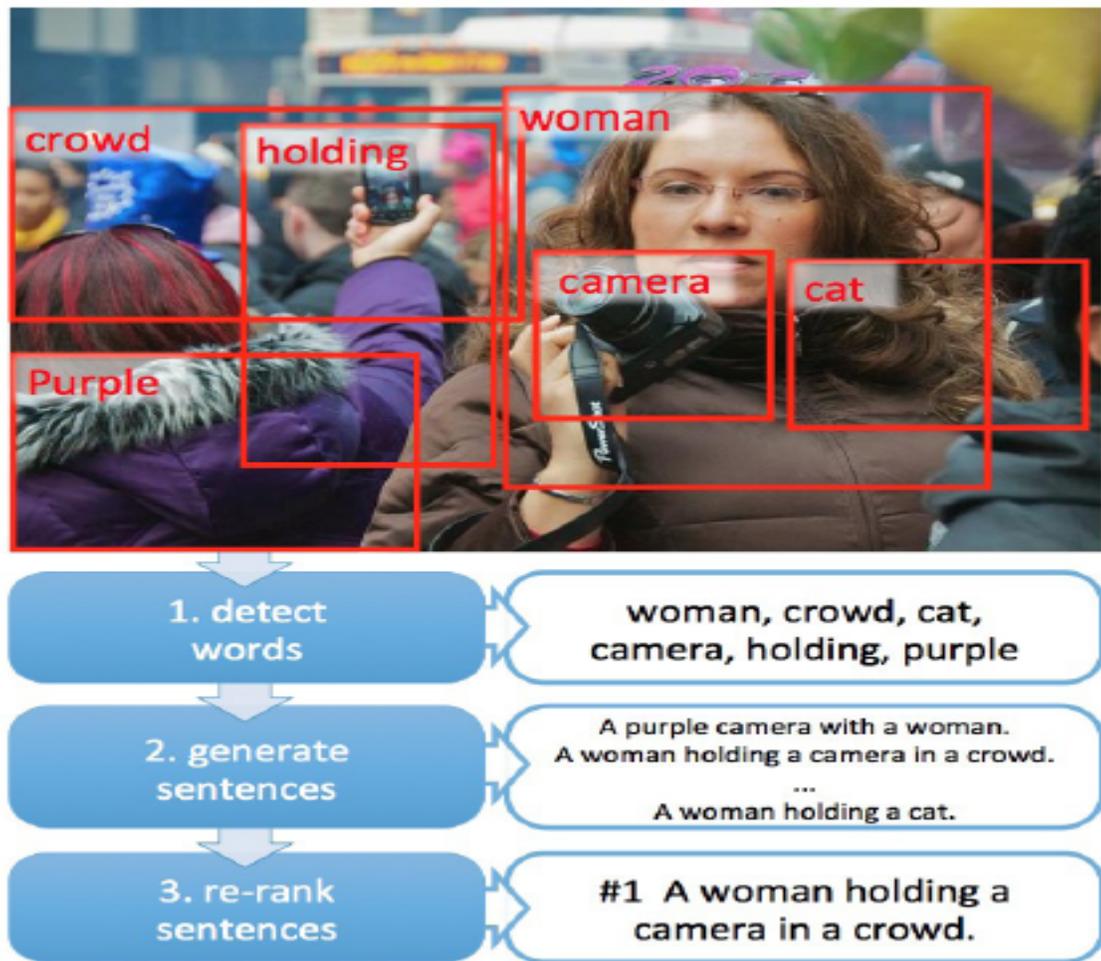
LFW人脸图像测试集



图像识别（1000类物体）

深度学习的应用

- 计算机视觉
 - 人脸识别
 - LFW测试准确率近100%
 - 图像识别
 - ImageNet测试识别率超过人
 - 目标检测
 - 以图搜图
 - 看图说话
 - ...
- 海量图像数据训练集+深度神经网络+并行计算 -> 强大的图像识别系统



自动“看图说话”

联合主办: MEIZU | Flyme | msup®

深度学习的应用

• 医学图像自动标识

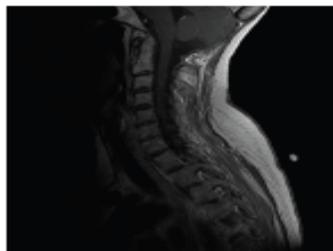
- 美国国家健康研究院和谷歌合作
- 10万多病人
- 20多万张医生标注的CT/MRI图像
- 构造一系列深度卷积神经网络模型
- 对新图像可以自动标识关键病症



**diameter
mass
kidney**

*avg distance:
0.33*

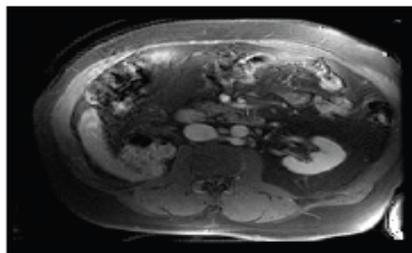
"... and solid lobulated **mass** arises from the anterior lower pole of right **kidney** and measures 1.6 cm in **diameter** ..."



**spine
chest
scoliosis**

*avg distance:
0.42*

" ... it measures a few mm in diameter and is best appreciated on series 3 image 6 in the thoracic **spine** no definite enhancing lesions are present. in the lumbosacral **spine** ..."



**diameter
lesion
kidney**

*avg distance:
0.33*

"... 2 apparently cystic **lesion** in the retroperitoneum adjacent to the crus 3 liver **lesions** for example series 17 image 22 series 16 image 172 and image 137 the lateral **lesion** ..."



**adenopathy
masses
lung**

*avg distance:
0.21*

"... dozens of **masses** of various sizes in or near right pleura and or peripheral **lung** without definite change ... by areas of right **lung** consolidation atelectasis and or confluent **masses** ..."

距离人脑还很遥远

- 人工神经网络连接数：
 - 目前：10亿
 - 几年前的1000倍
- 人脑：
 - 10亿的连接 = 1立方毫米的脑组织
- 奇点：人工智能超过人脑的时间
 - 2045？
 - 智能 ≠ 智慧：
 - 好奇心，情感，文化，...
 - 想象力/创造力：探索未知



Singularity

平安人工智能+大数据

平安脑



风险量化



反欺诈



智能推荐



健康医疗



智能运营

基础算法开发平台

深度学习平台

舆情监控与分析平台

智能推荐平台

可视化分析挖掘平台

信息安全监控平台

Hive

Hbase

Spark

Hadoop mr

Hdfs

R

Oozie scheduler

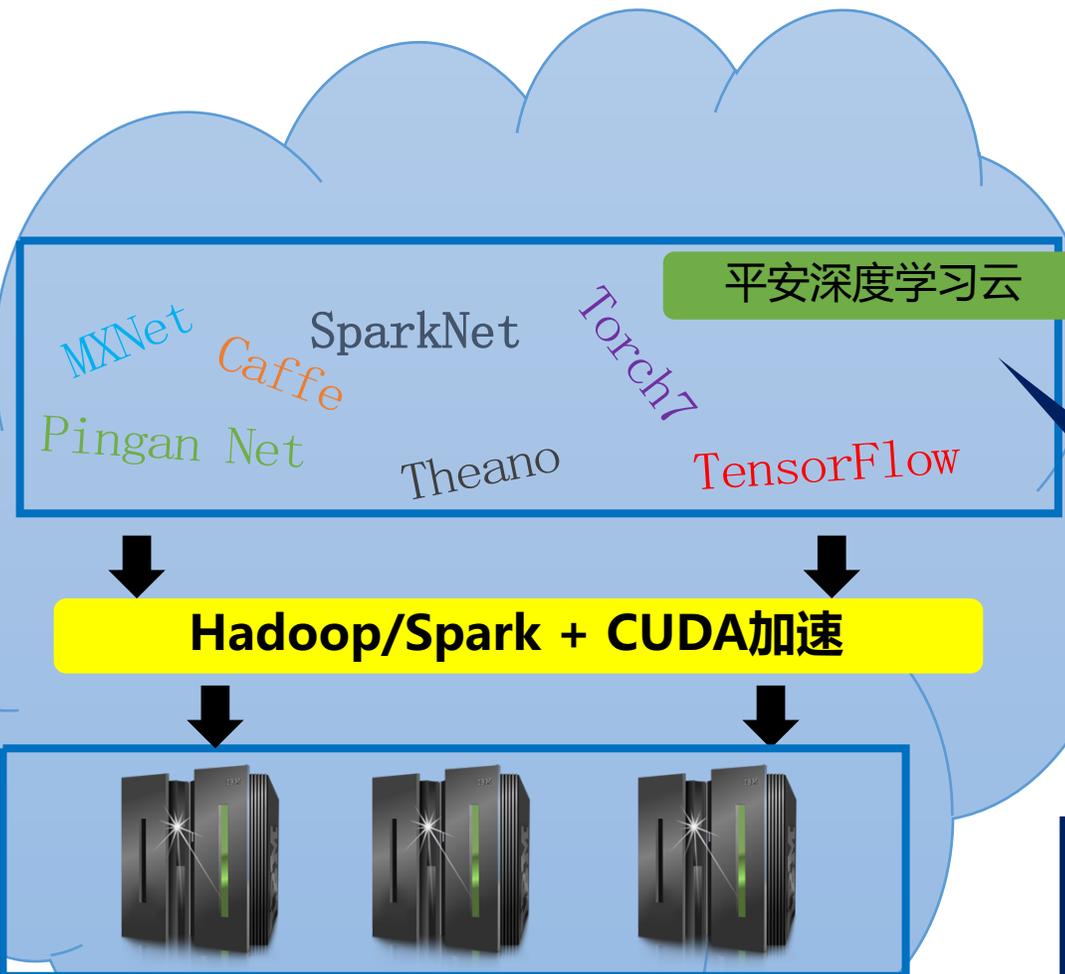
...

平安集团互联网+传统金融

- 互联网交易量破万亿
- 总用户超过3亿（金融+互联网）
 - 移动APP用户超过五千万
- 月均活跃用户达到千万级
 - 移动APP月活过八百万且持续快速增长
- 传统金融+互联网金融：
 - 大数据：
 - 静态信息，动态行为
 - 渠道：
 - 官网，移动APP，任意门，短信，邮件，电话，业务员，...



深度学习云平台



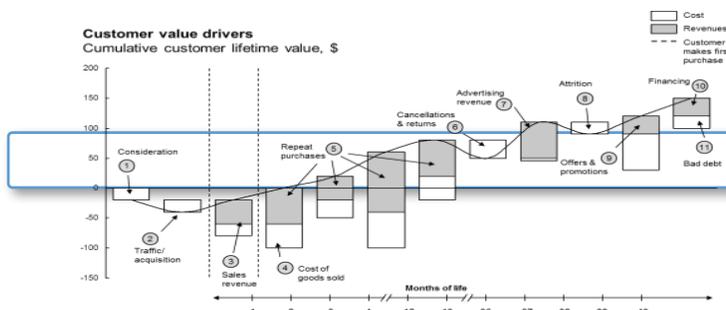
深度学习服务及软件栈。提供支持分布式计算以及单机版本的深度学习框架。

深度学习平台依托分布式于异构计算平台。利用强大的云计算能力，加快深度学习的计算，从而保证深度学习服务的高效与可靠。



平安脑应用示例

客户营销



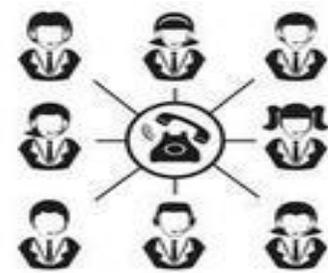
全生命周期

风险管控



全流程控制

运营管理

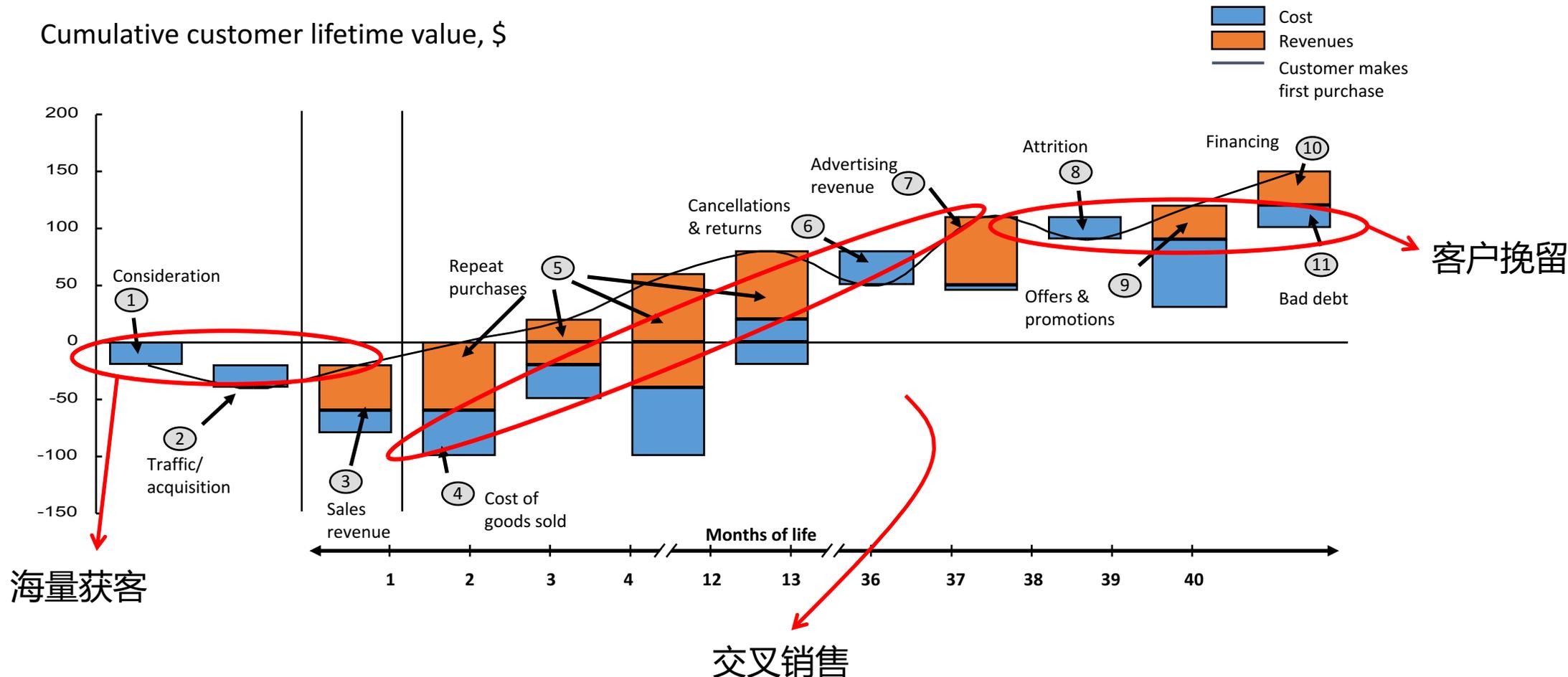


专项职能，精准优化



智能引擎

平安脑智能营销



海量获客

贯穿客户生命周期全过程

平安脑智能营销

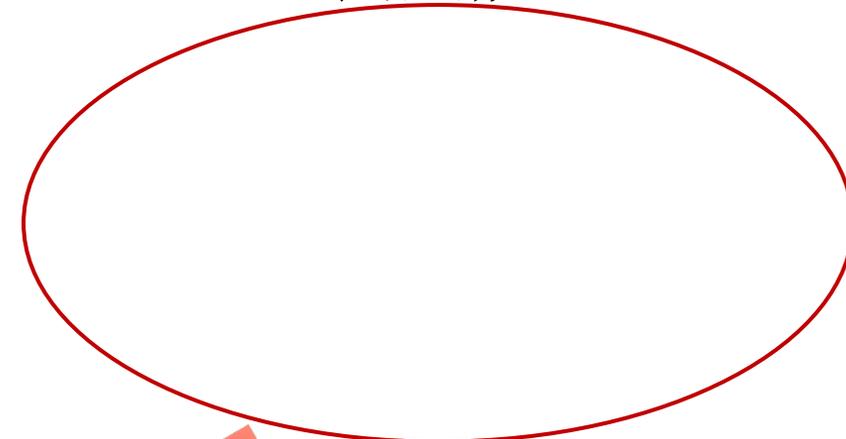
WHO
用户画像



WHEN & WHERE
业务场景



WHAT
产品画像



匹配模型

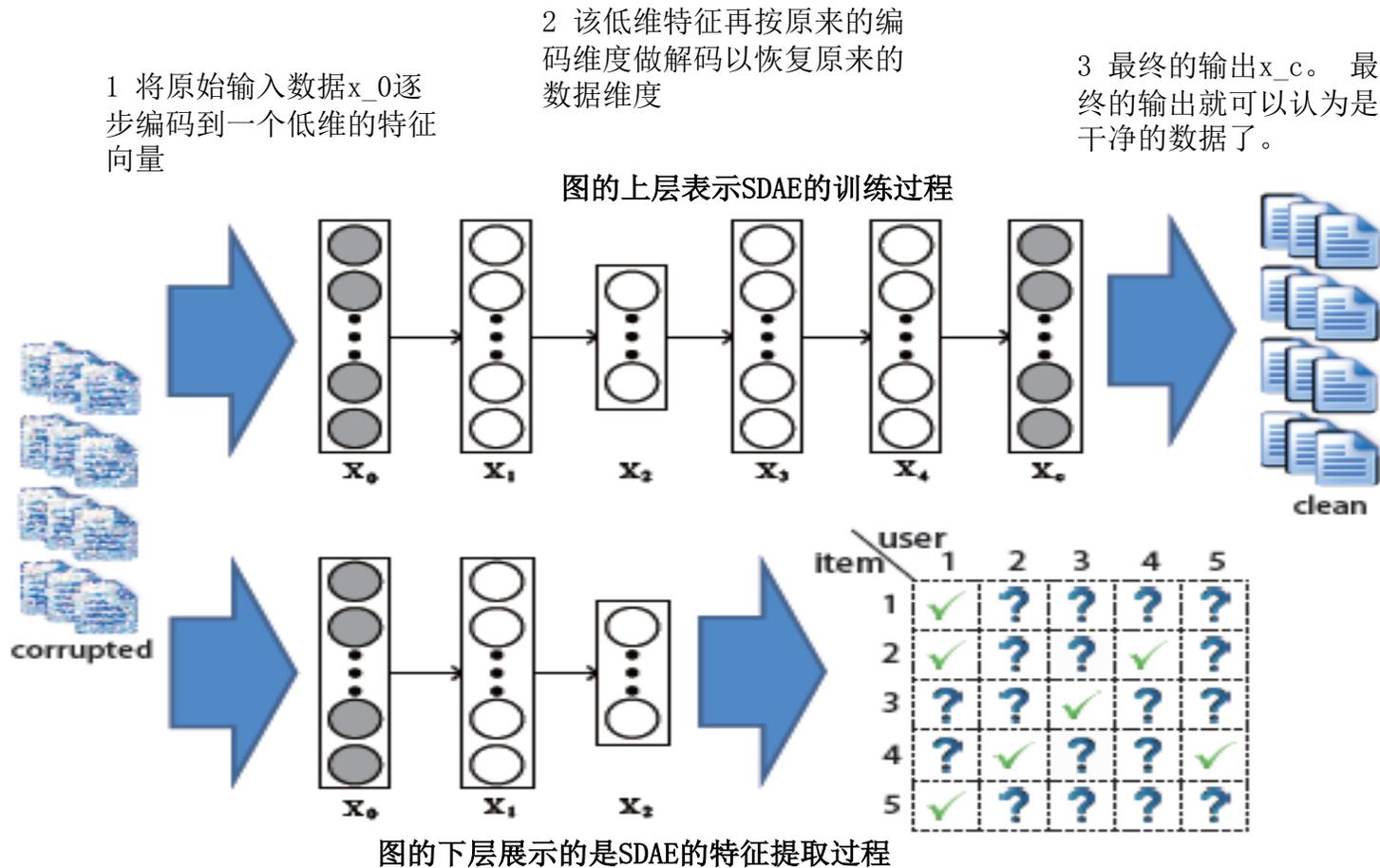
HOW/WHY 用户需求

个性化双向推荐

深度学习智能推荐

传统的协同过滤算法因噪声的干扰导致推荐效果不理想，而且数据的稀疏也进一步降低推荐的精准度。为解决推荐过程中，用户画像数据过于稀疏并存在大量噪声数据的问题，提出基于深度学习的推荐优化算法。

算法利用深度学习SDAE(栈化解噪音自动编码器)提取高鲁棒的判定特征，得到高质量的清洗数据，接着得到ITEM/USER矩阵完成精准推荐。

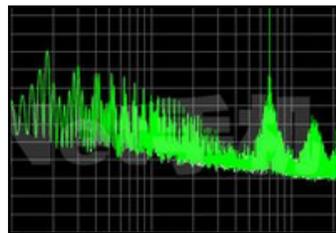


平安脑智能风控

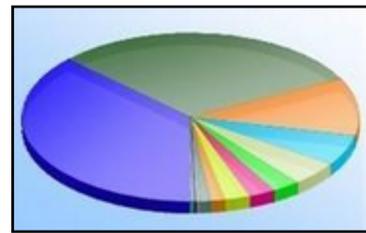
- 风险量化模型：
 - 广：多样化的数据因子
 - 深：个性化的数据因子
 - 机器学习辅以业务知识
- 保险，投资，健康



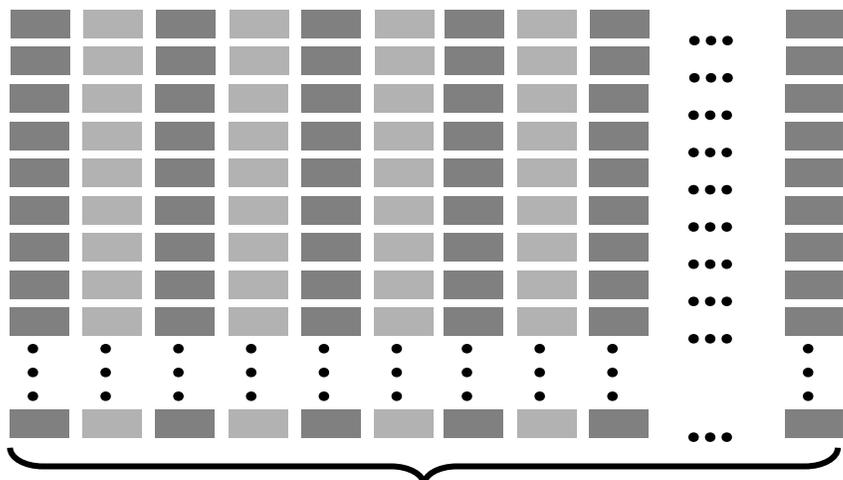
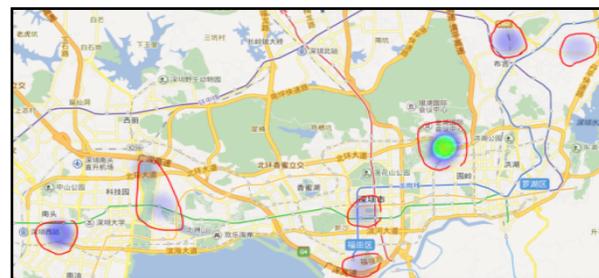
统计特征



频谱特征



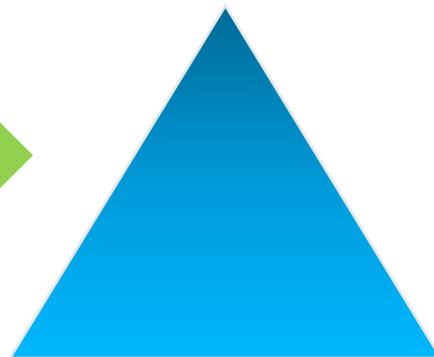
行为特征



多样的数据 → 风险因子



机器学习模型+业务经验规则



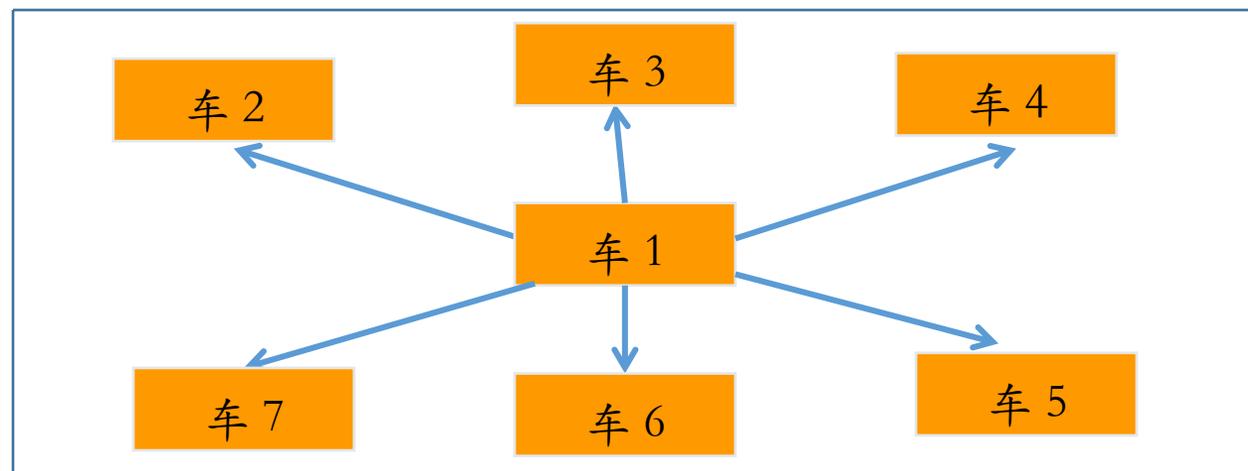
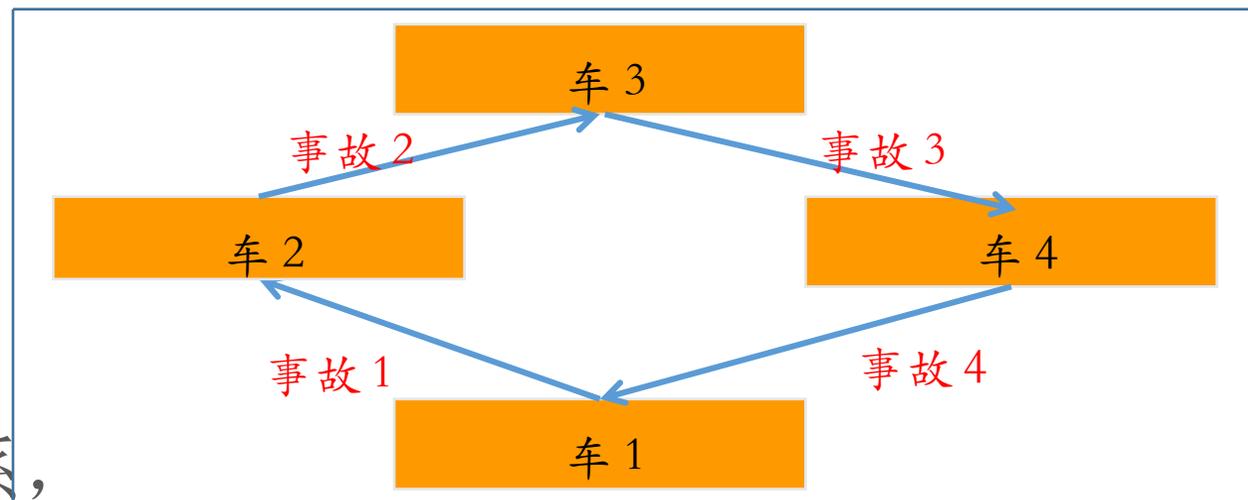
高



低

平安脑智能风控

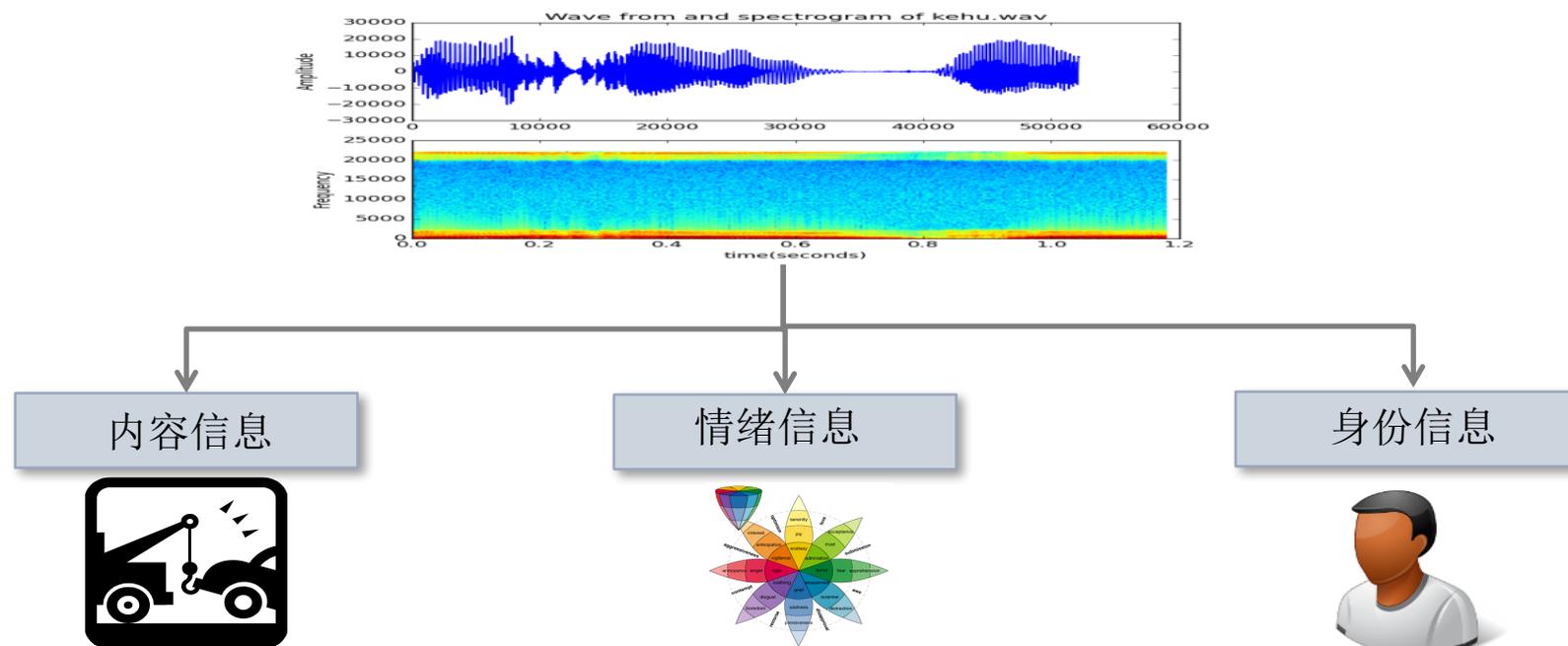
- 舆情分析：
 - 情感识别
 - 热点挖掘
- 关系网络分析：
 - 社交关系，亲属关系，任职关系，股权关系，...
 - 热点事件的影响程度和范围
 - 欺诈网络识别



疑似车险欺诈网络

平安脑智能风控

- 语音语义分析识别实现欺诈早期预警：
 - 报案人身份识别
 - 是否报假案



平安脑智能风控

- 医保欺诈检测：

数据模型评分



最高评分

中间评分

社保专家验证



可疑人员：王五

可疑度评分：A

就诊机构数 > 10

住院次数 > 20

药品种类数 > 40

个人现金金额 < 200

确属违规

行为正常

产品输出



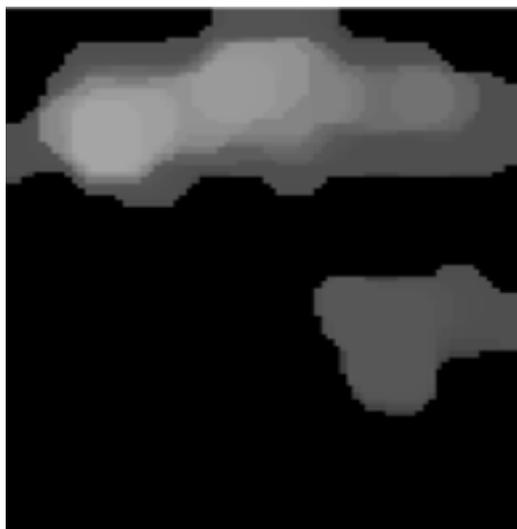
主动学习模型

平安脑智能风控

- PS照片自动检测及定位：
 - 防止欺诈
 - 提高效率



原始照片



定位结果

照片PS检测



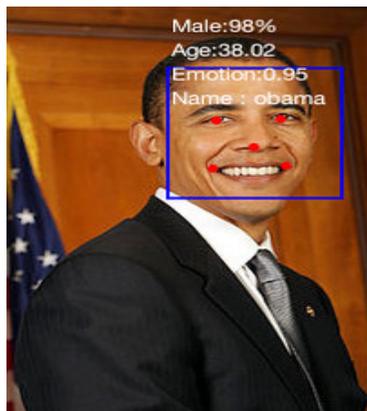
平安脑智能运营

■ 图像识别提升服务体验及效率

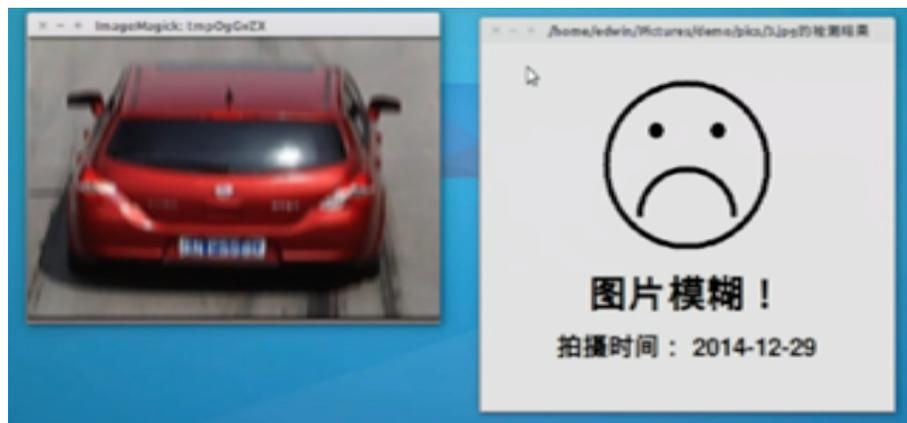
- 客户身份识别，车险查勘中的照片清晰度识别、证件识别、车辆信息及车损状况识别、场景三维重建及可视化、监控视频压缩及智能摘要

■ 自然语言理解完善机器人客服或家庭助手

- 结合专业知识库，实现拟人化实时交互，并可协助远程看护诊疗



VIP客户识别



照片清晰度自动检测



车辆信息自动识别

智能定损

- 问题情境：车险案件中会产生大量的车损查勘照片。目前平安查勘团队（含代理）每年的人力成本投入巨大，同时，对于车辆损失程度的判定仍缺乏验证机制。
- 借助深度学习技术，我们可以快速检测车损部位，识别车损程度。



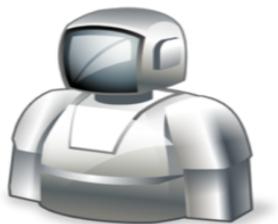
车牌分类

- **问题情境**：目前通用的车牌类型主要有五种，分别是：大型民用车牌、小型民用车牌、武警车牌、使领馆车牌、其他外籍车牌。
- **依靠深度学习技术**，我们可以准确定位车牌位置，并智能识别车牌所属类别。



问答机器人

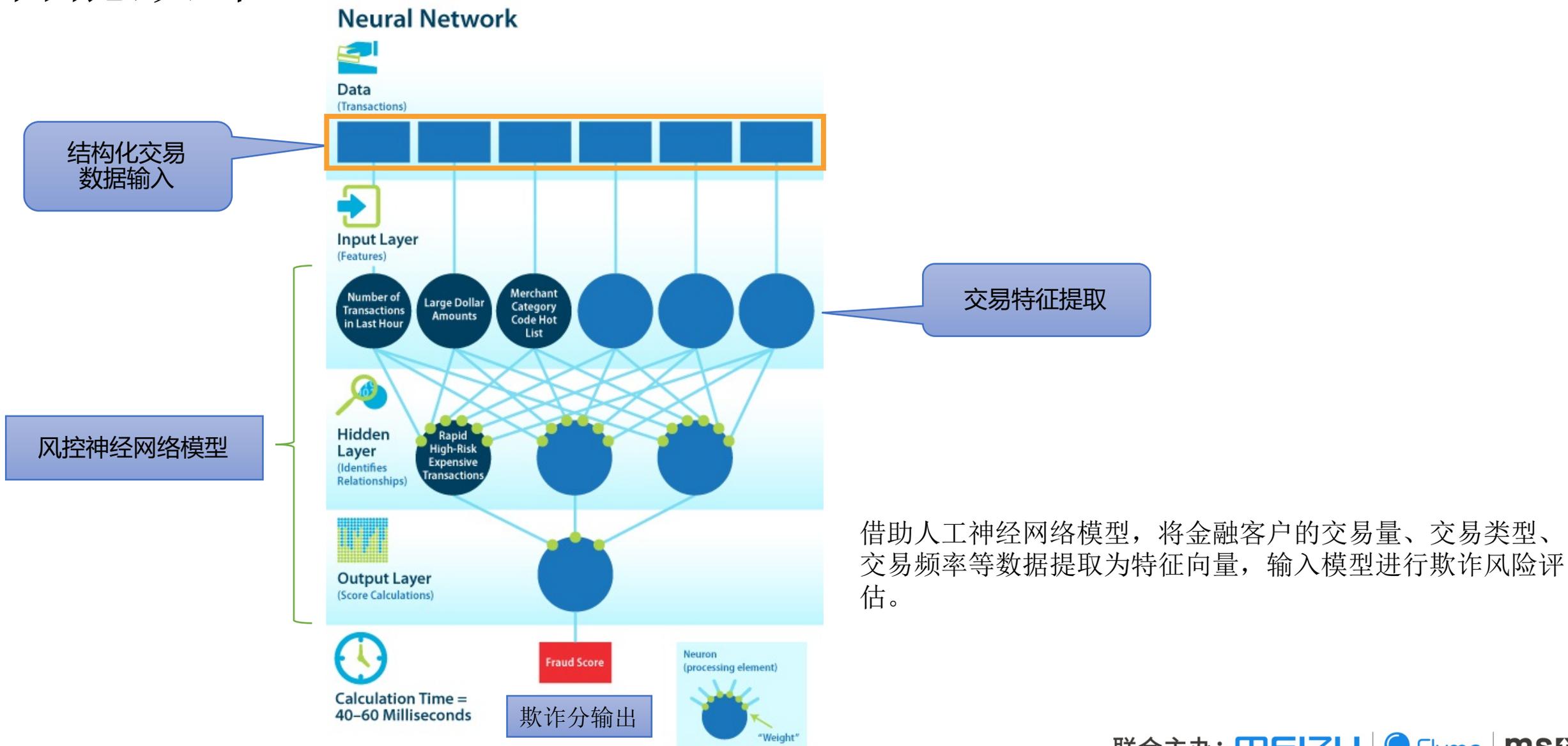
客服机器人产品以自然语言处理和人机交互等多种人工智能技术为基础，使用WEB、IM、WAP、SMS等多种表现形式，通过拟人化的方式与网络用户进行实时交互沟通。



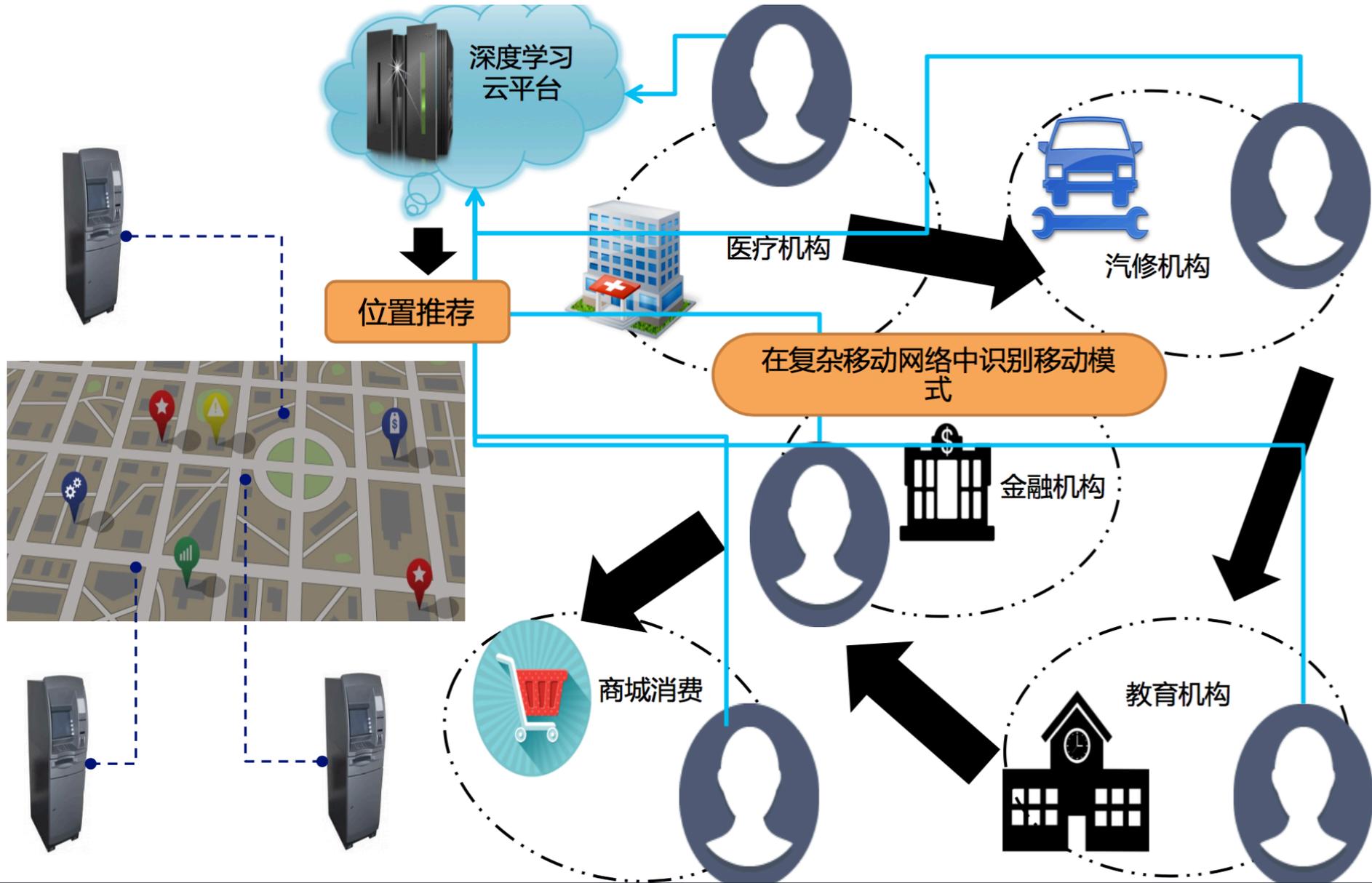
- 自动应答：无人工参与，不受工作时间、业务范围等外界环境影响。
- 智能记忆：支持上下文记忆、自动成长。
- 综合应用：可同时支持多种业务应用，不同的知识库，实际意义上的综合营销
- 可支持网页、手机页面/客户端



智能欺诈



LBS服务



物理机器人

智能迎宾 人脸检测，VIP客户识别，主动问候
询问客户需求，根据客户需求自主导航、引导至办理位置

语音查询咨询 可替换现有营业厅体积大、外观差的触摸屏交互查询系统，提供营业厅各类查询与业务咨询，并能够用语音加图文的方式告知客户各类业务办理流程



营销宣传 银行宣传片播放、理财产品宣传介绍等

远程业务指导 链接后台，远程与现场客户进行交流或业务指导



谢谢！

@BigData大数据