

# 人工智能引领企业持续创新

AI POWER YOUR BUSINESS

演讲人：李英伟

跨界互联  
数聚未来

第四届中国数据分析师行业峰会

CHINA DATA ANALYST SUMMIT

北京 中国大饭店 2017.07

# 人工智能(AI)成为“新一轮产业变革的核心驱动力”



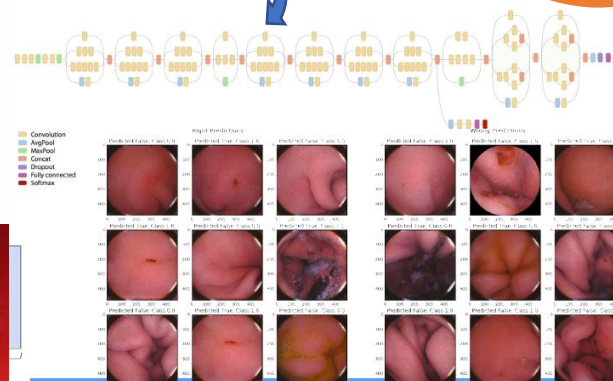
自动驾驶汽车

人工智能 (AI) 亦称机器智能，是指由人工制造出来的系统所表现出来的智能。通常人工智能是指通过普通电脑实现的智能。AI的核心问题包括推理、知识、规划、学习、交流、感知、移动和操作物体的能力等。

## New Tech in Fintech - Investopedia

在早期，个人和机构使用“看不见的手”，即价格信号，来做金融决策。但在新的科技条件下，像机器学习、预测性行为分析和数据驱动的市场营销，会取代金融决策里的猜谜和投机游戏。

## IBM Watson 在医疗行业开始“结果”



第一步：2020年，总体技术和应用与世界先进水平同步。核心产业规模超过1500亿元，带动相关产业规模超过1万亿元。

索引号: 000014349/2017-00142  
发文机关: 国务院  
标题: 国务院关于印发新一代人工智能发展规划的通知  
发文字号: 国发〔2017〕35号  
主 题 词:  
主 题 词:  
主 题 词:

### 国务院关于印发新一代人工智能发展规划的通知

各省、自治区、直辖市人民政府，国务院各部委、各直属机构：  
现将《新一代人工智能发展规划》印发给你们，请认真贯彻落实。

相关报道  
\* 国务院印发《新一代人工智能发展规划》  
图解  
\* 国务院印发《新一代人工智能发展规划》

无人超市



# IBM一直致力于人工智能领域并作出了卓越贡献

Learn  
学习能力



Understand  
理解各种数据  
的能力

Reason  
推理和决策  
能力



# IBM认知计算—中国已落地技术部分列表



## 问答

直接响应以主要文档来源为基础的  
用户查询



## 语音转文本

提供高度精确、低延迟的语音  
识别能力



## 概念洞察力

定位不能直接提到您的查询的相  
关文档



## 机器翻译

将文本从一种语言翻译为  
另一种语言



## 特征洞察力

特征洞察力有助于按用户自己的  
方式进行互动



## 消息共振

以适合特定用户的风格和言辞  
与之沟通



## 概念扩展

将委婉语或口语词与最常见的短  
语对应起来



## 关系提取

智能化地发现句子成分之间的关系  
(名词、动词、主语、宾语等)。



## 可视化呈现

数据分析的图形表示更易理解



## 文本转语音

从文本整合自然语音



## 权衡分析

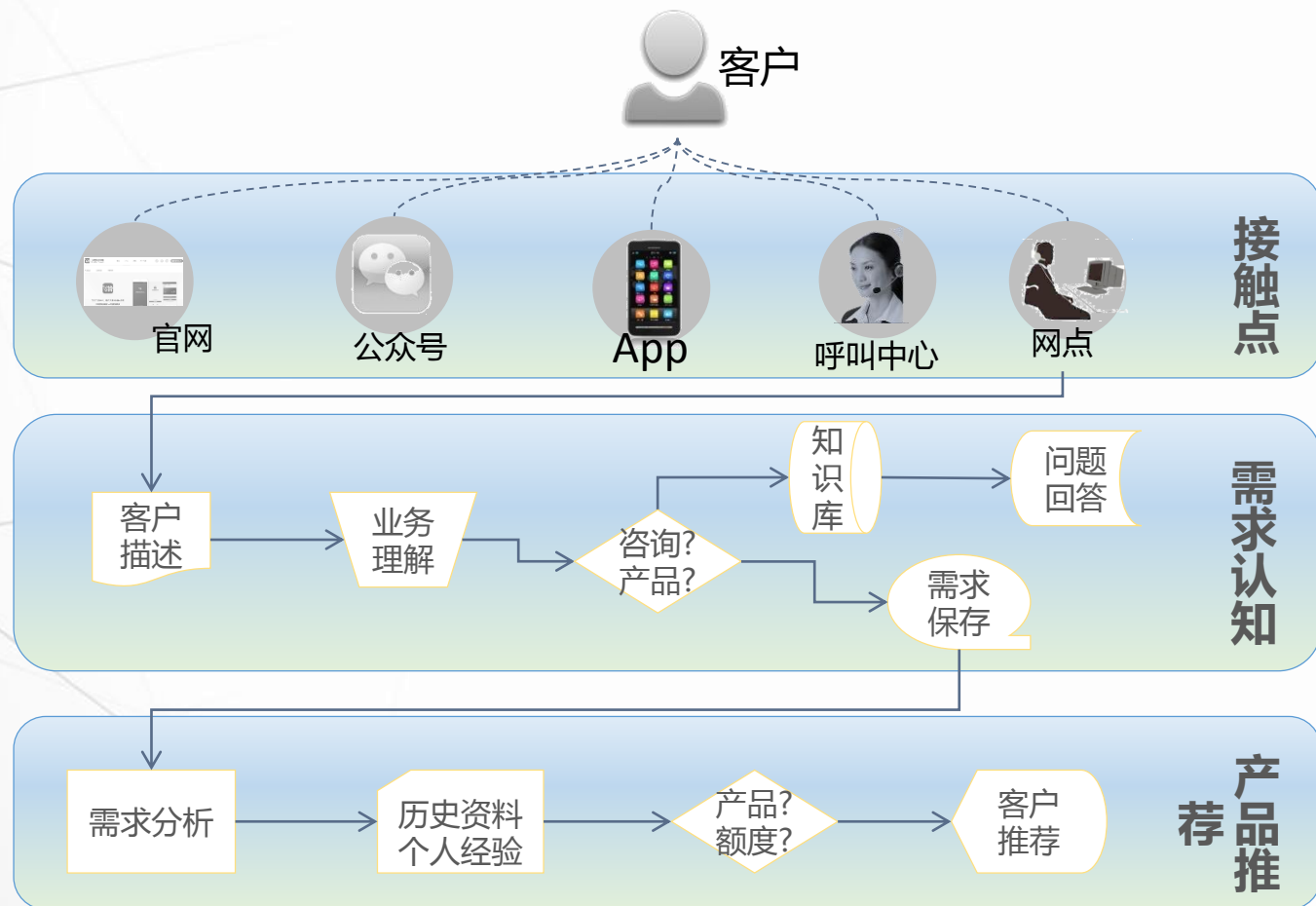
借助智能可视化和分析建议从多个相  
互冲突的目标中做出更好的选择



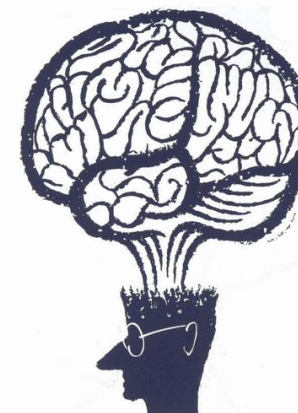
## 视觉识别

分析图像或视频帧的可视化外观，以  
便了解正在发生的事情

# 企业在多年信息化积累后亦然存有大量优化空间 – 以金融服务公司为例



信息化



人脑思考

# 人工智能与认知计算引领企业实现持续创新



客户

智能服务

智能助理

智能推荐

智能问答

智能辅助

增强的智能化应用提供全新的客户体验和交互

全面地认知理解并持续学习业务形成知识网络

利用知识网络识别、优化、推理完成业务决策



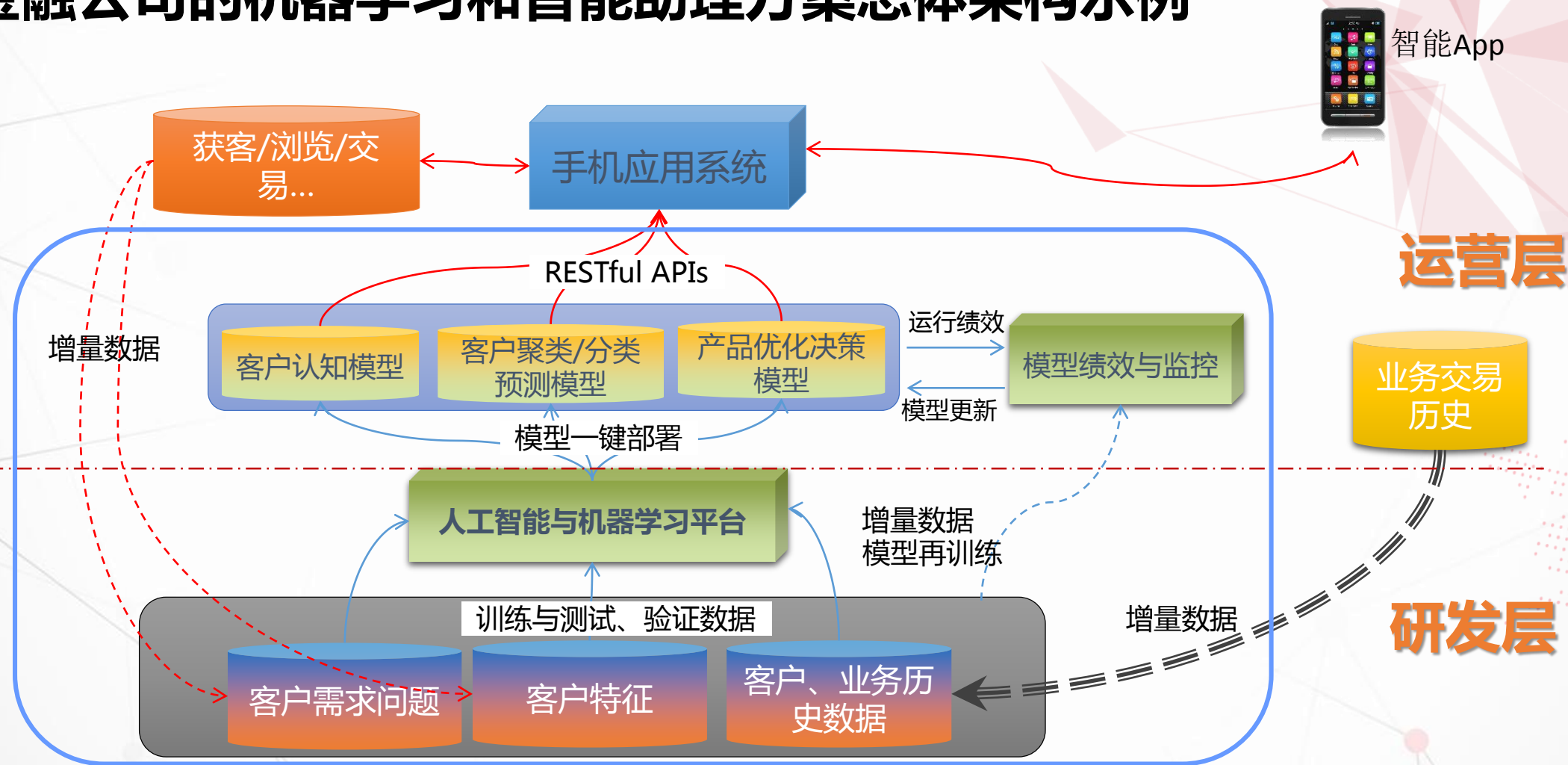
信息化增强智能



人工智能&机器学习



# 某金融公司的机器学习和智能助理方案总体架构示例



# 自然语言处理可以帮助实现与客户的认知交互，深入理解客户意图

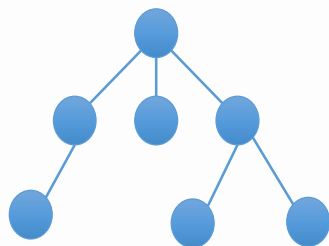
短信  
公众号  
手机App



自然语言处理



认知系统



输入：“赚得太少了”

- 语音转文本
- 上下文 = “产品：货币基金”

分类识别意图

- 意图 = “推荐更高收益率的产品”
- 置信度 = 0.876655900

对话系统实现

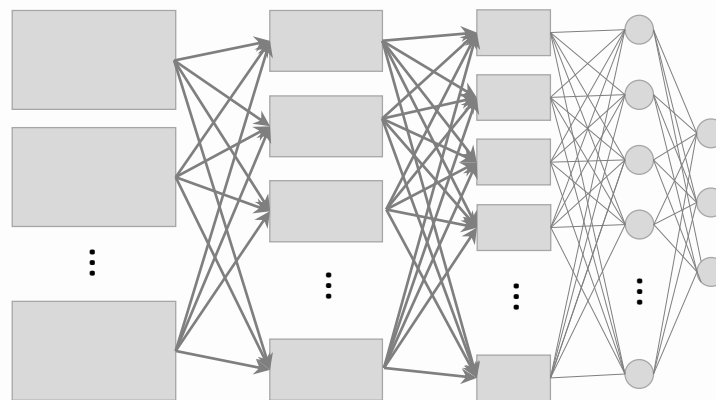
- 意图 = “推荐更高收益率的产品”
- 上下文 = “产品：偏股基金”
- 和后台数据交互



# 认知交互基础是近几年得到快速发展和应用的深度学习算法

涨的太少了,  
收益不好,  
赚的太差了,  
没赚多少钱啊,  
涨的不多啊,  
这个收益我不满意,  
不如存款呢,  
赚的不够多,  
... ..

训练语料库



基于深度学习神经网络(RNN)构建

# 之后，提供更好的服务需要深入了解客户的特征并能够准确预测客户行为

已有产品：时间-产品-利率  
交互事件：时间-方式-类别-内容-其它  
贷款还款：时间-产品-方式  
产品浏览：时间-产品

行为

购买需求：资金需求类别-产品推荐  
购买预测：购买时间-产品推荐-比较优化  
产品营销：产品关联  
客户转化：产品推荐

将来

过去

客户特征：  
• 客户分群  
• 风险类别  
• 客户类别标签  
• 客户层级  
• ... ..

K-Means  
聚类

风险计量

分类算法

RFM模型

客户偏好：  
• 资金需求模式偏好  
• 分类产品倾向  
• 营销影响  
• ... ..

客户购买偏好：  
• 产品需求偏好  
• 还款偏好  
• 恰当时间的产品推荐

基于内容的  
推荐

产品偏好客户：  
• 潜在高转化率客户  
• 新产品目标客户  
• 交叉营销客户

优化算法

特征



# 客户信息和已有的消费行为积累的数据往往是一个未能充分挖掘的金矿

中国移动 3G 16:57 91%  
< 信息 9XXXX 详细信息

短信/彩信  
今天15:55

【XX银行】您尾号为xxxx  
的银行卡收入1,3689.54  
元，类型为代发工资。



中国移动 3G 16:57 91%  
< 信息 9XXXX 详细信息

短信/彩信  
今天15:55

【XX银行】您尾号为xxxx  
银行卡收入7,000.00元  
整，类型入账还本(理财产  
品xxxxxx)

【XX银行】您尾号为xxxx  
银行卡收入673.41元，  
类型为入账分红(理财产  
品xxxxxx)



中国移动 3G 16:57 91%  
< 信息 9XXXX 详细信息

短信/彩信  
今天15:55

【XX银行】小秘书提醒：  
21日上午10:00发售低风险  
理财产品xxxxxx，预期收益  
5.1%，手机专属，10万起，  
建议购入。

## 变推销为贴身增值服务：

- 合适的条件(客户资金满足)
- 合适的时机(客户预期购买前点)
- 合适的内容(适合客户特点的产品)

# 为了更好、更全面的理解客户，为其刻画标签以方便管理和服

标签属性类型：实时类标签、挖掘模型类标签、预测类标签

标签分类

标签体系

客户基本属性：人口统计信息、位置信息、自定义信息

客户关联关系：金融关联信息、社交网络信息

客户兴趣偏好：产品偏好、渠道偏好、使用偏好

客户价值信息：用户自身价值、用户贡献

用户风险信息：用户风险评级、黑名单信息

客户兴趣偏好：近期需求、营销活动信息

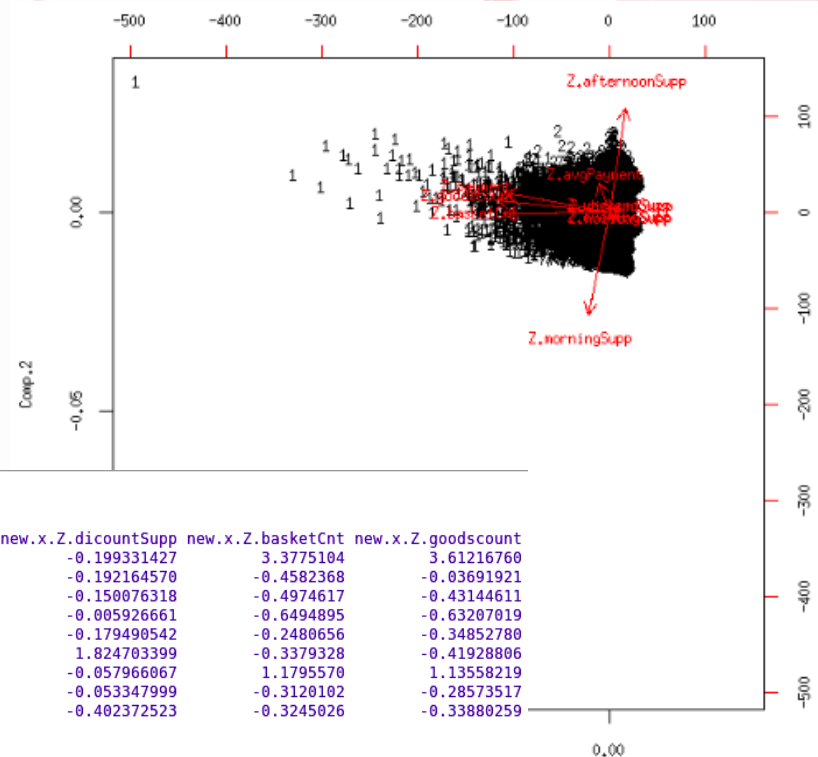
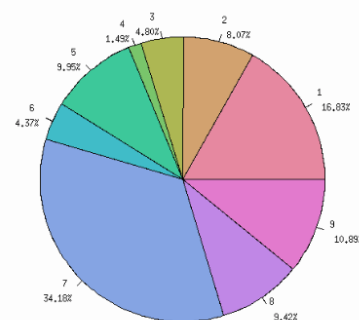
应用场景：营销增强、用户洞察、渠道优化、运营提升...



# 利用各种数据进行建模分析：如聚类分析可以帮助实现客户更好的分类

## • 采用主要特征

- 性别
- 生日
- 收入
- 风险分级
- 违约率
- 拥有产品数
- 资金期限比例
- 资金借入频率
- App登录频率
- ...



```
[1] 707 954 2316 1358 4389 2045 3811 3660 4592
> .cluster$centers # Cluster Centroids
new.x.Z.holidaySupp new.x.Z.weekendSupp new.x.Z.morningSupp new.x.Z.afternoonSupp new.x.Z.eveningSupp new.x.Z.dicountSupp new.x.Z.basketCnt new.x.Z.goodscount
1 -0.09322325 -0.16266709 0.33232172 -0.1588546936 -0.19593272 -0.199331427 3.3775104 3.61216760
2 -0.01850273 0.17561281 -0.01412079 0.1960903558 -0.20843332 -0.192164570 -0.4582368 -0.03691921
3 -0.36337497 1.48358180 -0.60004414 -0.9101416646 -0.35979485 -0.150076318 -0.4974617 -0.43144611
4 3.14966207 -0.36668520 0.04199112 0.0008155306 -0.04867011 -0.005926661 -0.6494895 -0.63207019
5 -0.23169817 -0.08573895 1.45978410 -0.9382180296 -0.58543564 -0.179490542 -0.2480656 -0.34852780
6 -0.20588147 -0.18866813 -0.23533309 0.4976428503 -0.30217424 1.824703399 -0.3379328 -0.41928806
7 -0.08065419 -0.06931664 0.10601531 -0.0285497289 -0.08783668 -0.057966067 1.1795570 1.13558219
8 -0.21059777 0.28537539 -0.71811421 -0.8815961358 1.82562905 -0.053347999 -0.3120102 -0.28573517
9 -0.18205409 -0.65520897 -0.56408198 0.9259252308 -0.41874666 -0.402372523 -0.3245026 -0.33880259
new.x.Z.payment new.x.Z.avgPayment
1 3.4619937 0.153196604
2 0.7524174 3.303090701
3 -0.3749726 0.048949125
4 -0.5467440 -0.007971684
5 -0.3405579 -0.246736027
6 -0.3637887 -0.282389434
7 0.8429205 -0.152968861
8 -0.2864821 -0.099233578
9 -0.3222365 -0.164509824
```

# 客户预测模型示例 – 实现客户购买时机和额度预测

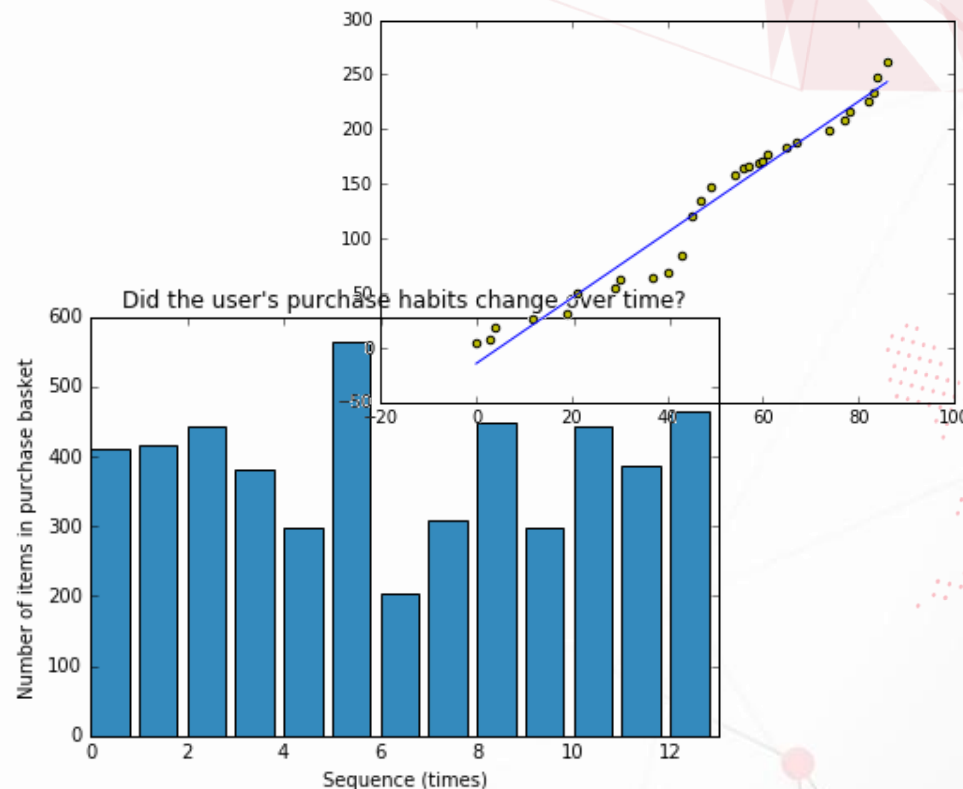
RFM模型的应用示例：

- **R**ecency – 最近一次购买？最近一次联系？
- **F**requency – 借入资金频率？
- **M**onetary Value – 资金额度？客户贡献？

客户，尤其是非偶然资金需求客户，一般对于资金需求有着内因驱动，该内容虽然隐藏，但可以通过其借入时间、金额、期限、还款等来认知。从而可以有效预测客户在当次借入资金时的需求。

客户的订购频率、联系频率符合泊松分布，根据各个客户的交互历史和订购历史，可以对客户未来购买时机作出预测；

```
In [99]: plt.scatter(df[['days']], df[['accomu']], c='y')
plt.plot(df[['days']], regr.predict(df[['days']]), color='blue')
plt.show()
```



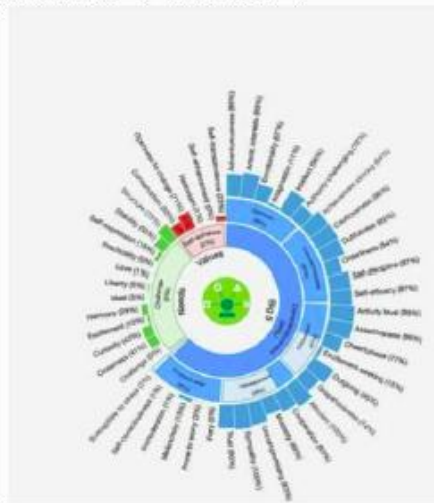
# AI技术的快速发展进一步增强客户标签体系 – 自然语言特征抽取示例

OBAMA		TRUMP
18:39	LENGTH OF SPEECH	16:30
2,420	NUMBER OF WORDS	1,116
2	MENTIONS OF "I"	3
60	MENTIONS OF "WE"	27
0	MENTIONS OF "ME"	0
12	PAUSES FOR APPLAUSE	26
Nation	MOST FREQUENT WORD	American
Sympathy	PRIMARY PERSONALITY TRAIT	Assertive
Analytical	SPEECH LANGUAGE STYLE	Confident

使用了4个watson应用程序 ( API ) 来帮助对比分析奥巴马和特朗普的总统任职演讲 :

- 演讲文本分析 ( Speech to Text )
- 情感分析 ( Sentiment Analysis )
- 语调分析器 ( Tone Analyser )
- 人格洞悉 ( Personality Insights )

Visualization of Personality Data



Watson对特朗普的人格总结 :

- 善于分析, 情感浮动并具有批判性
- 感他人之感受, 并具有同理心
- 很难被激怒
- 同时, 也是利他的, 帮助别人会感到满足, 并愿意竭力这么做
- 对追求成功和享受生活不太关心, 倾向于选择做更宏大的事

Data Behind Your Personality

Name	Value ± Sampling Error
<b>Big 5</b>	
Openness	97% (± 8%)
Adventurousness	89% (± 4%)
Artistic interests	98% (± 10%)
Emotionality	67% (± 4%)
Imagination	11% (± 6%)
Intellect	95% (± 5%)
Authority-challenging	69% (± 8%)
Conscientiousness	90% (± 7%)
Achievement striving	94% (± 9%)
Cautiousness	95% (± 9%)
Dutifulness	92% (± 5%)
Orderliness	63% (± 8%)
Self-discipline	87% (± 4%)
Self-efficacy	87% (± 9%)
Extraversion	47% (± 8%)
Activity level	99% (± 7%)
Assertiveness	99% (± 8%)
Cheerfulness	76% (± 10%)
Excitement-seeking	14% (± 8%)

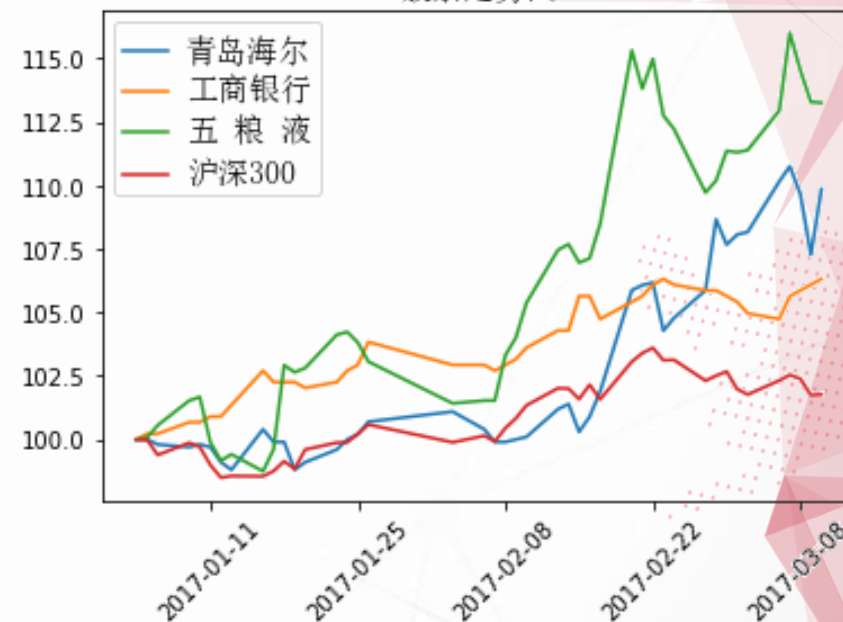
Watson通过几个方面的人格分析发现, 尽管川普的演讲得到了相对消极的评价, 但是奥巴马和川普的就职演讲获得了几乎一致的情感分析赋值。



# 业务的优化和增益：AI技术促进金融产品中风险的持续创新研究

- 如要完成对单个乃至组合的评价，需要量化风险及收益
  - 传统方式：对一段时期的收益率做平均，计算方差。
    - 选择时间阶段是关键(期限、间隔)
    - 国际常选择月收益率、三到五年
  - 传统方式的优点
    - 简单、快速
  - 传统方式缺点
    - 对变化考虑较少
  - 新尝试：根据近期的收益率变化，利用贝叶斯方法估算收益和方差
    - 一般选择周或天收益率
    - 同时考虑系统风险的转换
  - 优点
    - 考虑动态变化较多，更为精确
    - 计算相对复杂、研究论文较少

股票走势图





# 国内市场的实例计算尝试 – 利用共轭先验加速贝叶斯方法计算

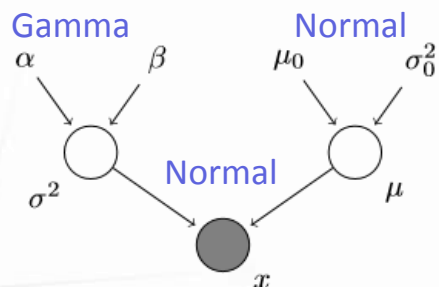
$$\begin{aligned} x_i | \mu, \tau &\sim \mathcal{N}(\mu, \tau) \quad i.i.d. \\ \mu | \tau &\sim \mathcal{N}(\mu_0, n_0\tau) \\ \tau &\sim \text{Ga}(\alpha, \beta) \end{aligned}$$

Then the posterior is:

$$\begin{aligned} \mu | \tau, x &\sim \mathcal{N}\left(\frac{n\tau}{n\tau + n_0\tau}\bar{x} + \frac{n_0\tau}{n\tau + n_0\tau}\mu_0, n\tau + n_0\tau\right) \\ \tau | x &\sim \text{Ga}\left(\alpha + \frac{n}{2}, \beta + \frac{1}{2}\sum(x_i - \bar{x})^2 + \frac{nn_0}{2(n+n_0)}(\bar{x} - \mu_0)^2\right) \end{aligned}$$

机器学习专家做出传统估计和贝叶斯分析估计算法和模型选择。并进行数据训练。

选择先验初始值，然后循环迭代，完成后验参数估计



以投资产品的收益和风险估计模型为例，投资与数学统计专家给出模型：

- 对数收益符合正态分布
- 收益和标准差都是变量
- 收益同样符合正态分布
- 精度(方差倒数)符合伽玛分布

均值( $\mu$ )、精度( $\eta$ , 方差倒数)都为变量的贝叶斯估计，使用共轭高斯伽马分布首先均值后验符合一个高斯分布如下：

$$\mu_* = \frac{\tau_0\mu_0 + n\bar{x}}{\tau_0 + n} \quad \text{且均值变量的精度(方差倒数)} \quad \eta_* = (\tau_0 + n)\eta_0$$

其中， $\mu$ 的条件分布是均值为 $\mu_0$ 、精度 $\tau_0\eta_0$ 的正态分布， $\bar{x} = \sum_{i=1}^n x_i/n$ 是样本的均值

而 $\eta$ 的边缘分布式具有正参数 $\alpha$ 和 $\beta$ 的伽马分布， $\eta$ 的后验分布是参数为 $\alpha + (n/2)$ 和 $\beta_*$ 的伽马分布

$$\beta_* = \beta + \frac{1}{2} \sum_{i=1}^n (x_i - \bar{x})^2 + \frac{\tau_0 n (\bar{x} - \mu_0)^2}{2(\tau_0 + n)}$$

伽马分布的均值 $\bar{\eta} = \frac{\alpha}{\beta}$ ，方差为 $\frac{\alpha}{\beta^2}$

```
def BAYESIAN_Estimator(arr, tau0=1.0, mu0=0.0, alpha0=2.0, beta0=0.5, eta0=1):
    n = len(arr)
    mean = arr.mean()
    sigma = arr.std()
    # eta0 = 1.0 / (sigma**2)
    mu = (tau0*mu0 + n*mean)/(tau0+n)
    eta = (tau0 + n)*eta0
    alpha = alpha0 + n / 2.0
    beta = beta0 + 1.0/2.0*(sigma**2*n) + tau0*n*(mean-mu0)**2 / (2*(tau0+n))
    # print alpha, beta
    # print alpha/beta, alpha/(beta**2)
    return mu, (1.0/eta)**0.5, (1.0/(alpha/beta))**0.5, tau0+n, alpha, beta
```



# 决策技术应用于智能投顾计算最优组合(客户风险下的收益最大化)比例

$$\text{最小化 } \sum_{i=1}^N (X_i^2 \sigma_i^2) + \sum_{i=1}^N \sum_{j=1, j \neq i}^N (X_i X_j \sigma_{ij})$$

约束条件为

$$(1) \sum_{i=1}^N X_i = 1$$

$$(2) \sum_{i=1}^N (X_i \bar{R}_i) = \bar{R}_P$$

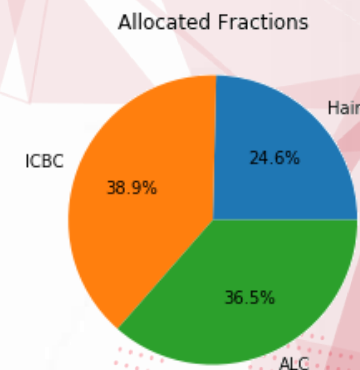
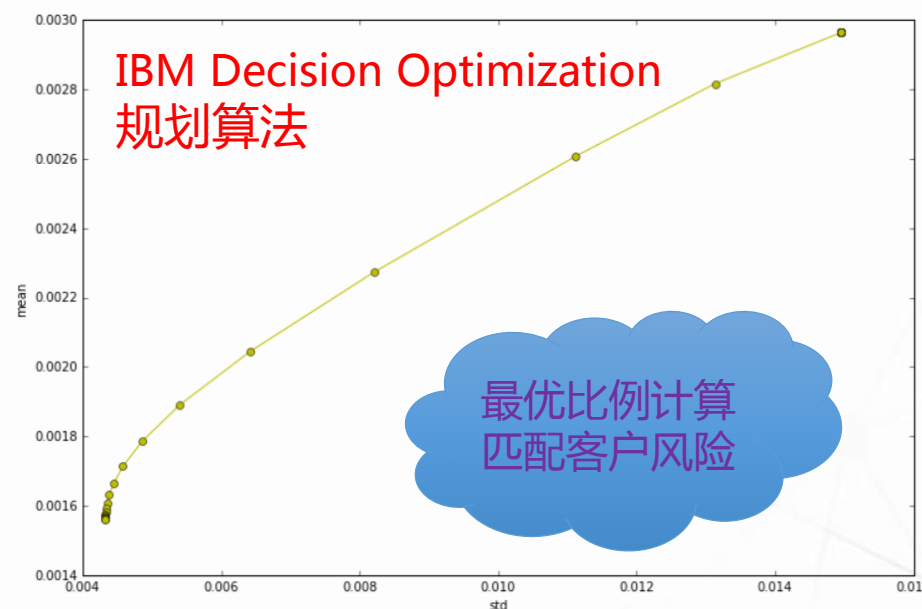
$$(3) X_i \geq 0, \quad i = 1, \dots, N$$

带有约束条件的二次规划模型-含无风险产品的投资组合，保持类别比例上限

从组合优化来看，要解决客户选择产品后的最佳配置是一个凸规划问题，计算组合边界

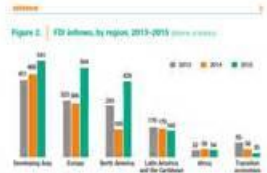
现代投资学理论可以具体量化组合模型。

针对各类产品，可以从收益、风险(波动)、期限、交易费用多方面进行评估。



# 自然语言处理技术还能从海量数据中提取有价值信息提升走势预测准确性

研究报告



Standard economies also led a rebound in FDI inflows

Pursuing three years of decline, FDI inflows from developed economies increased by 33 per cent to \$1.1 billion. As a result, developed economies accounted for 72 per cent of global FDI inflows in 2015, up from 61 per cent in 2014. The 14 percentage point increase in inflows from developed economies is the highest since 2007. Europe became the world's largest FDI inflow region in 2015, with FDI inflows of \$610 billion, up from \$590 billion in 2014. In contrast, FDI inflows declined in most regions. A combination of challenges, including slowing global growth and depreciating national currencies, contributed to the decline in FDI inflows.

社交媒体

银行王真 V  
余额宝收益率下降... 余额宝收益率下降... 余额宝收益率下降...

新华社金报  
余额宝收益率下降... 余额宝收益率下降... 余额宝收益率下降...

财经观察家  
余额宝收益率下降... 余额宝收益率下降... 余额宝收益率下降...

科技与金融将占主导!



自然语言理解  
特征提取

机器学习回归  
时间序列...

中国银监会办

银监办发(2016) 中国银监会办公厅 进一步完善新闻发言... 信托百伟汇 (ID: ...)

法规政策



全新头条栏目  
头条栏目增加实时行情  
财经新闻

财经报道

历史行情/  
经济数据  
CPI/GDP/  
利率/...

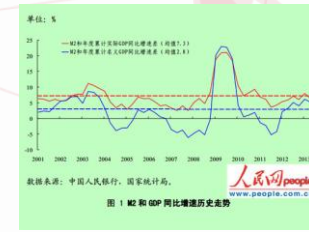
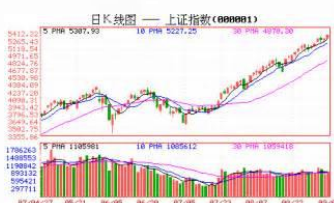


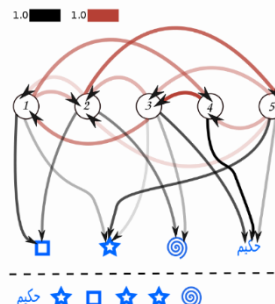
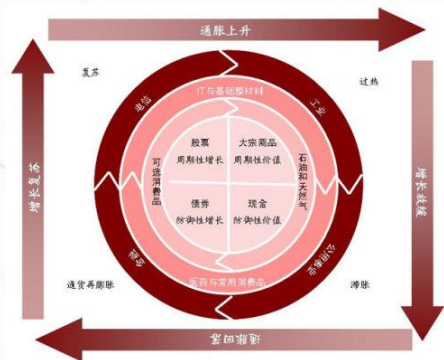
图 1: CPI 和 GDP 同比增速历史走势

行情走  
势判断



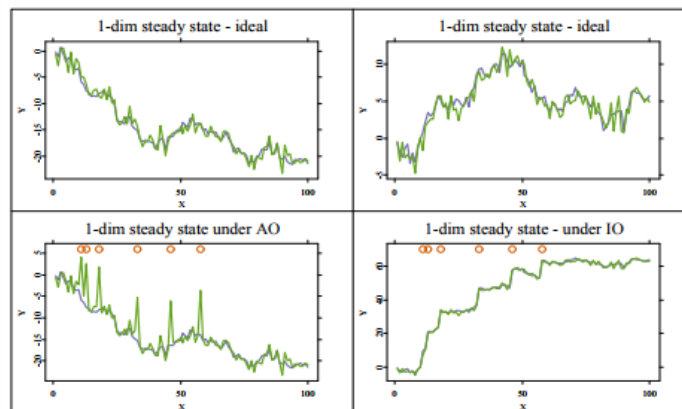
# 机器还可以去学习前人总结出来的投资经验并进一步提升精度

## 隐马尔科夫模型(HMM)



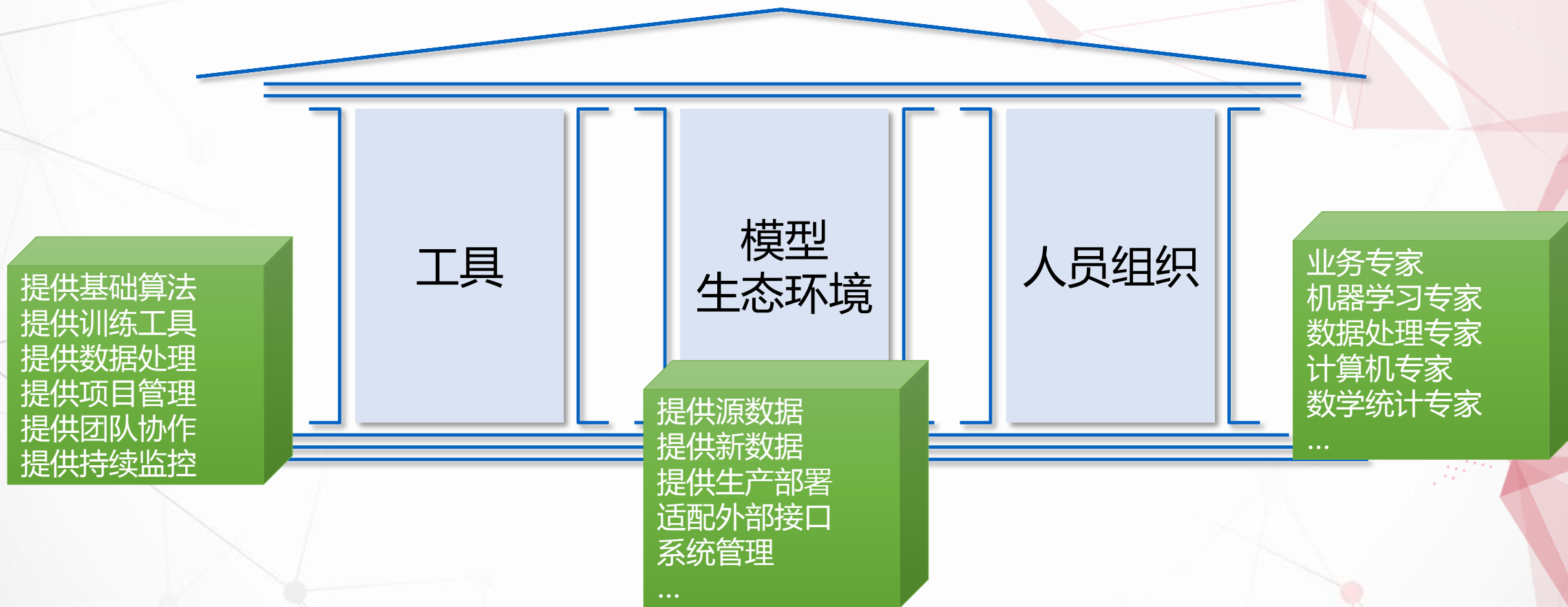
美林时钟在投资上的应用

历史数据给出的模式

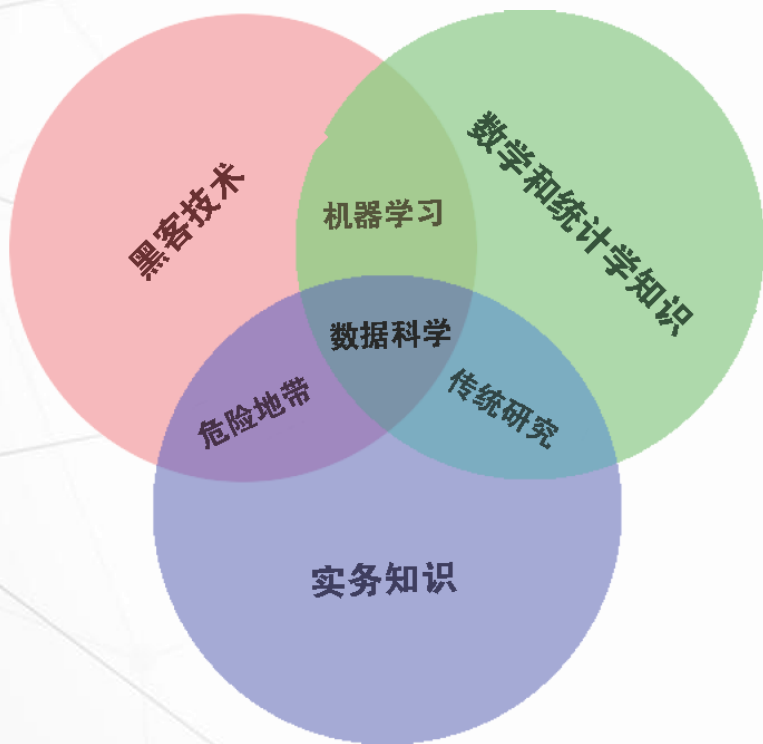


模式转换判断：  
投资方向调整

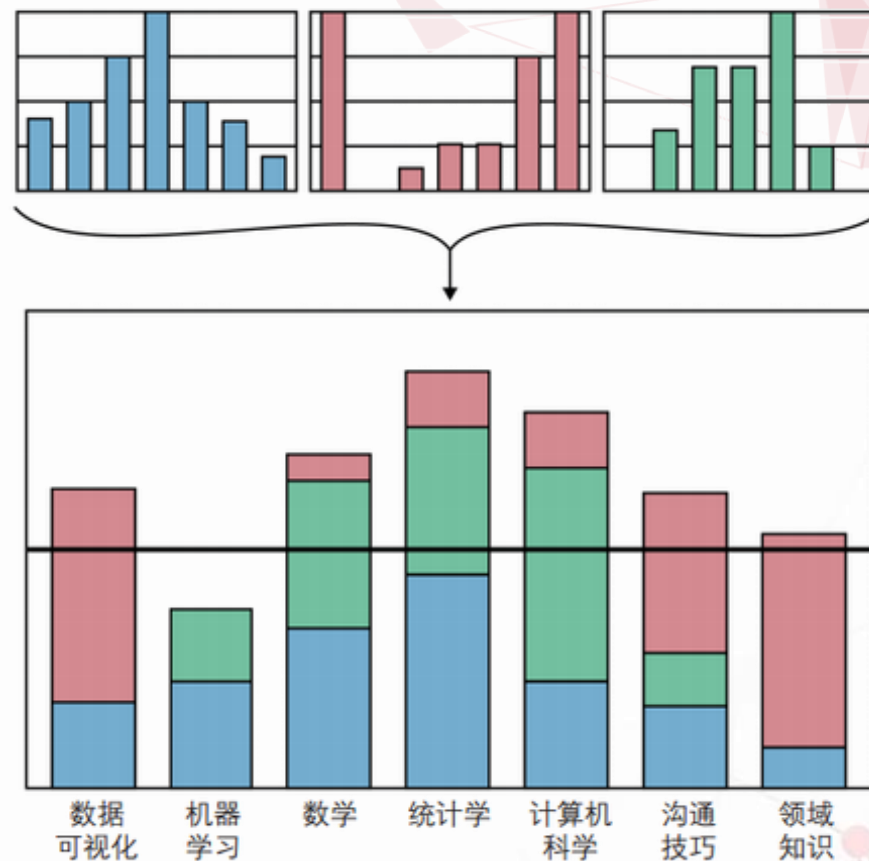
# AI驱动创新，不仅仅是技术平台，需要发挥三个支柱作用



# 过程中依赖数据科学家，数据科学是一门交叉学科，需要团队协作



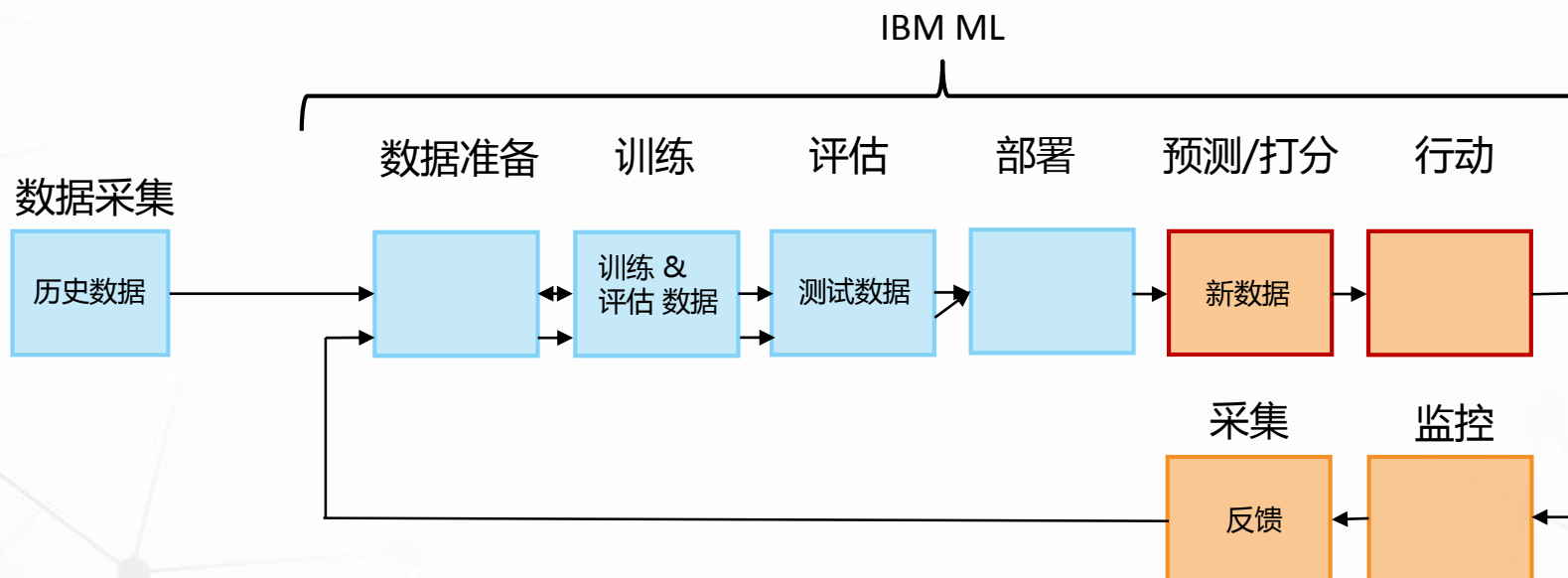
没有人会是完美的数据科学家，所以我们需要团队



# 而整体的目标是集合各类人员能力让企业从“机器学习”到“学习机器”

极少公司具备打分、行动、监控以及采集和反馈在一起的系统。

正是IBM Machine Learning的功能  
IBM ML 支持闭环的持续学习





# 从数据、运营环境与专业能力出发结合新技术、新思想去挑战新AI时代





CDA 数据分析师  
www.cda.cn

THANKS

# 跨界互联 数聚未来

第四届中国数据分析师行业峰会  
CHINA DATA ANALYST SUMMIT