

# 信誉系统和信用评分

Ratings, Reviews & Karma to Keep Your Community Healthy



Building

# Web Reputation Systems

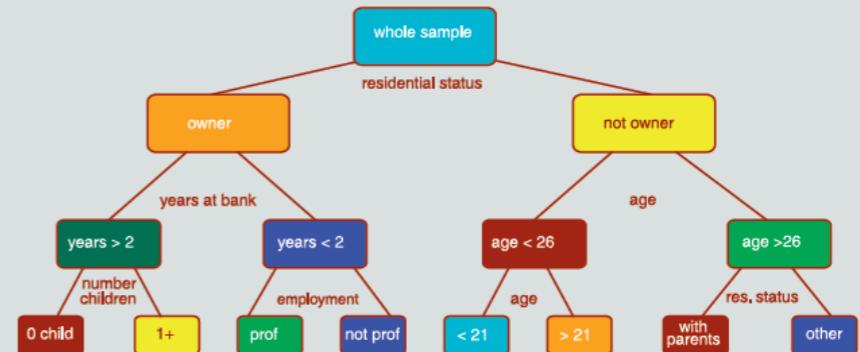
O'REILLY®

YAHOO! PRESS

F. Randall Farmer  
& Bryce Glass

## CREDIT SCORING and Its Applications

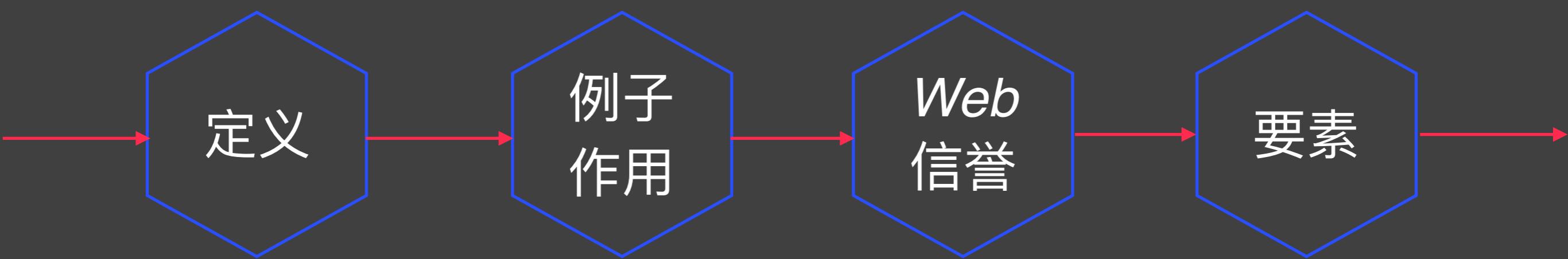
Second Edition



LYN THOMAS  
JONATHAN CROOK  
DAVID EDELMAN

• MATHEMATICS IN INDUSTRY •

# 信誉



- A reputation system **collects, distributes** and **aggregates feedback** about participants's **past behaviour**.  
→ **punish bad and reward good behaviour.**

---
- Reputation is **Information** used to make a **value judgement** about an **object or person**.

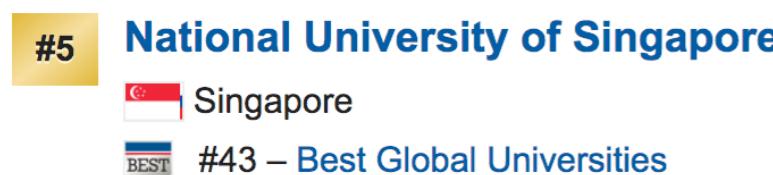
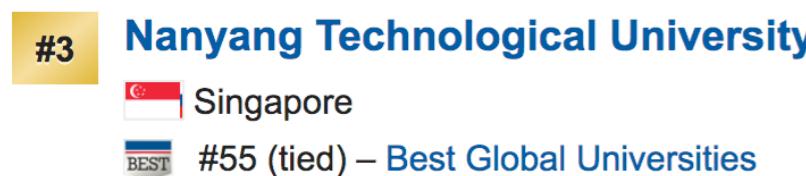
---
- Reputation is what is generally said or believed about a person's or thing's **character or standing**.

---
- Reputation **perception** that an agent create through past actions about its **intentions and norms**.

# Reputation

# 大学排名是一种信誉

## Best Global Universities for Computer Science



<https://cra.org/cra-statement-us-news-world-report-rankings-computer-science-universities/>

100  
Subject Score

99.3  
Subject Score

96.4  
Subject Score



#6

Stanford University

United States Stanford, CA

BEST #3 – Best Global Universities

#7

Harvard University

United States Cambridge, MA

BEST #1 – Best Global Universities

#8

University of California--Berkeley

United States Berkeley, CA

BEST #4 – Best Global Universities

#9

Huazhong University of Science and Technology

China Wuhan, Hubei

BEST #282 (tied) – Best Global Universities

#10

Zhejiang University

China Hangzhou, Zhejiang

BEST #159 (tied) – Best Global Universities

#22

Carnegie Mellon University

United States Pittsburgh, PA

BEST #77 (tied) – Best Global Universities

91.3  
Subject Score

90.7  
Subject Score

86.9  
Subject Score

85.5  
Subject Score

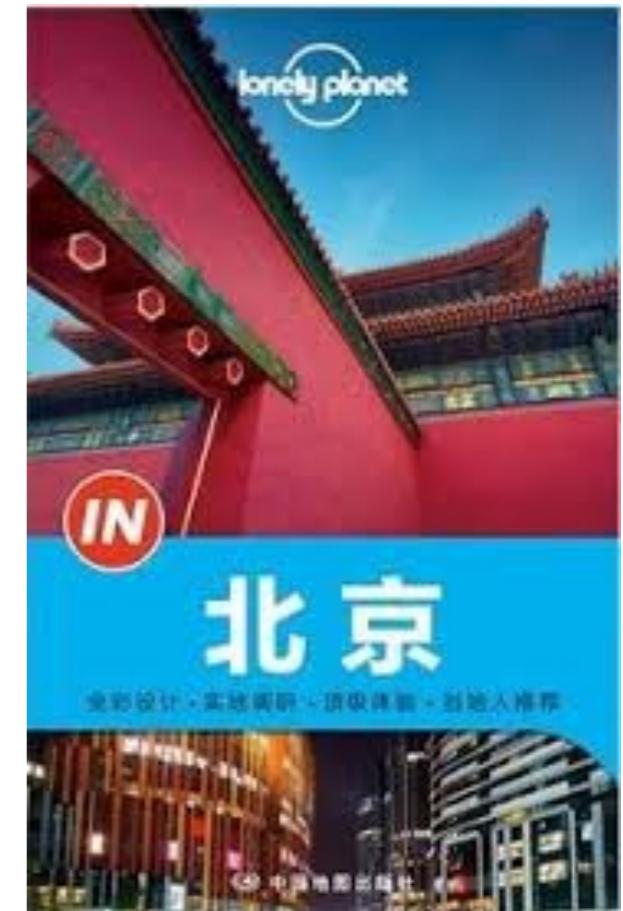
84.1  
Subject Score

78.2  
Subject Score

Anyone with knowledge of CS research will see these rankings for what they are — nonsense — and ignore them. But others may be seriously misled.

*Reputation*

# 旅游推荐也是一种信誉



现实生活中还有哪些是信誉？

- **GPA**
- **211/985/双一流**
- **CCF学术推荐**
- **论文引用**
- **...**
- **绩效考核KPI**
- **Top Ten**
- **尼尔森收视率**
- **道琼斯公司指数**
- **消费者信心指数**
- **影评**
- **...**

# *Reputation*

## 为什么需要信誉

陌生人

类型

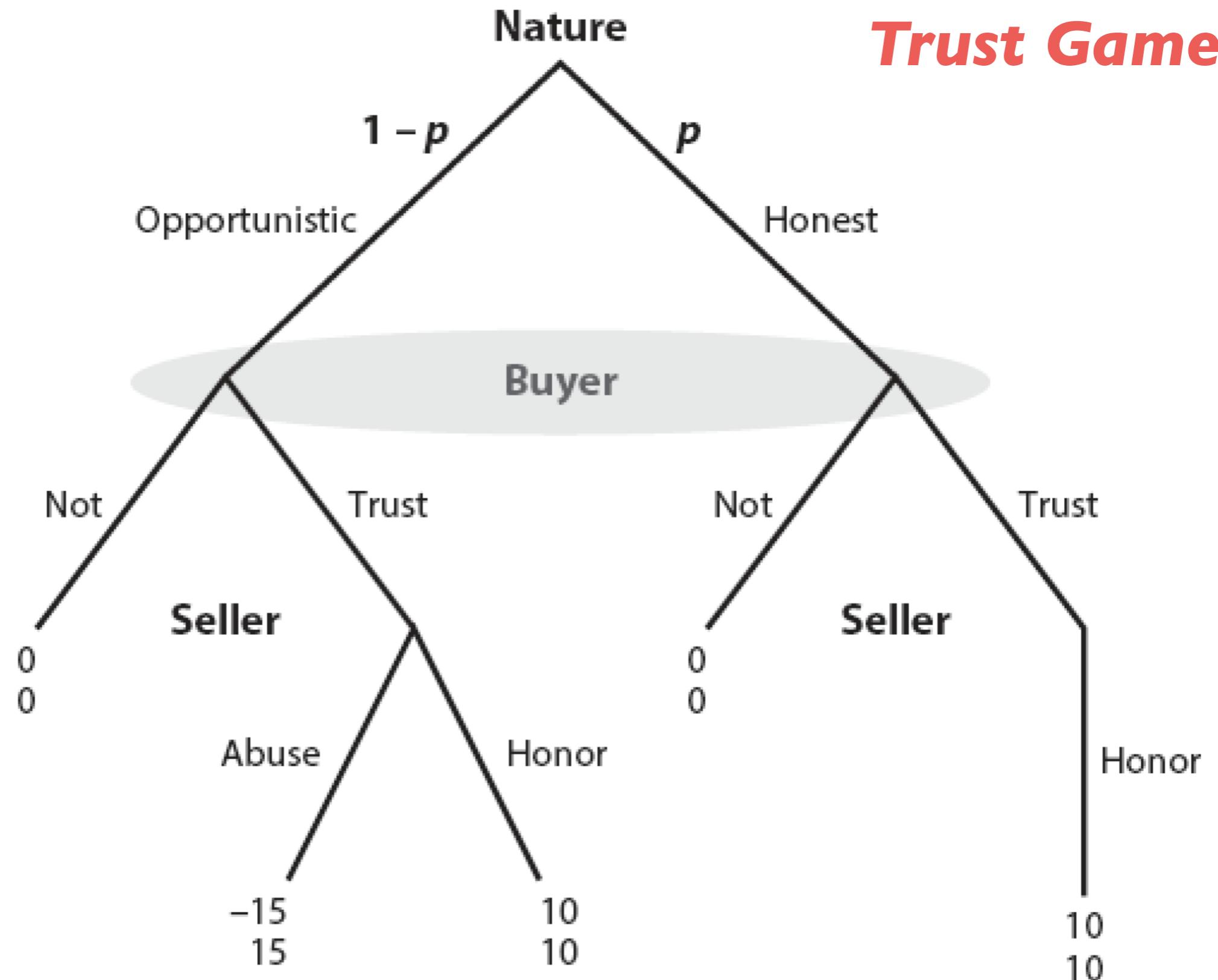
信任

保险

托管

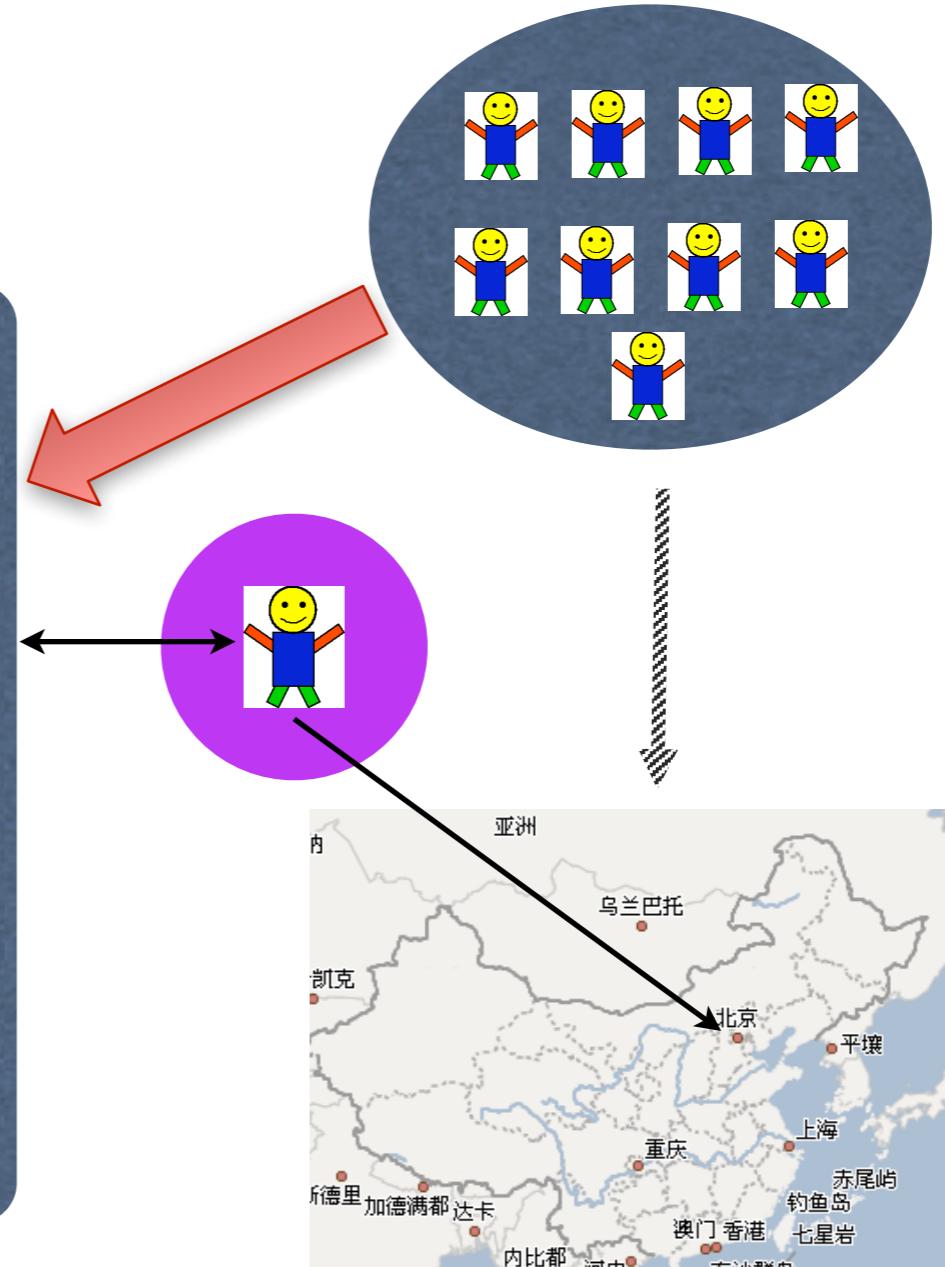
诉讼

## *Trust Game*



*Reputation*

# 现实中的信誉



小山村



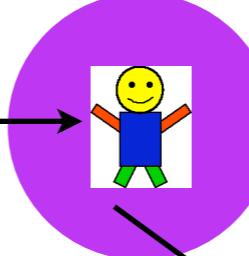
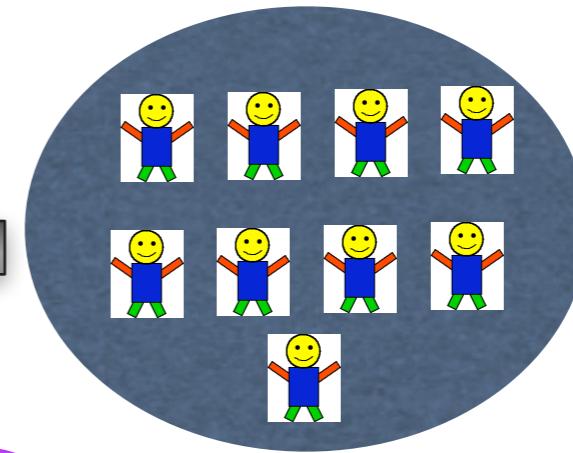
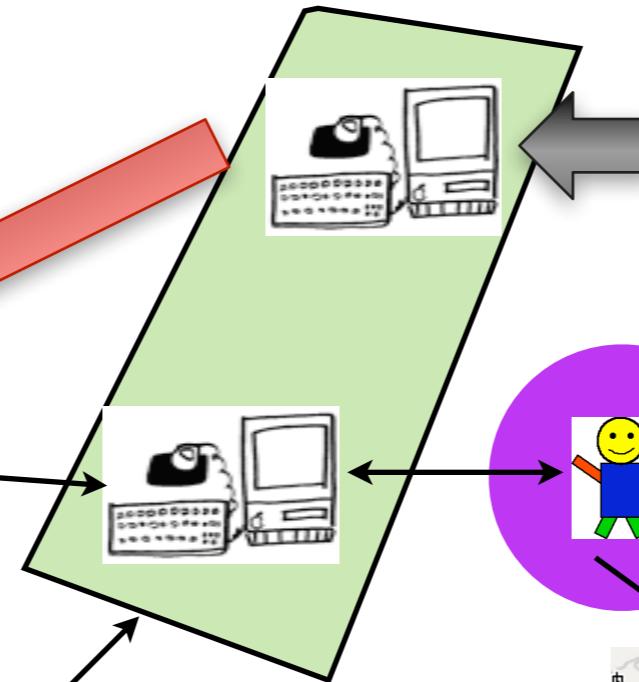
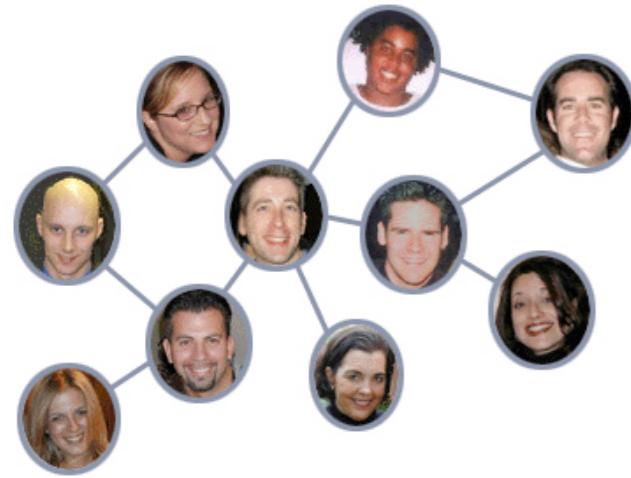
*Reputation*

# 挑战 vs. 机遇

挑战

机遇

大都市



## 商品评价

好评度

99%

就是快(4220)

物流很快(3213)

货真价实(1603)

是真品(1416)

价格优惠(1360)

很漂亮(1096)

性价比高(878)

够给力(728)

服务态度好(429)

挺不错(404)

自动关机(46)

有杂音(21)

反应迟钝(11)

全部评价(26万+)

晒图(500)

追评(1200+)

好评(26万+)

中评(1100+)

差评(2300+)

只看当前商品评价

推荐排序▼

全部评价(26万+)

晒图(500)

追评(1200+)

好评(26万+)

中评(1100+)

差评(2300+)

只看当前商品评价

推荐排序▼



飞\*\*\*哥

京享值10971

机器暂时没啥问题，物流很快，但是，强调，强调，苹果京东的客服的智商真的有问题，懂？智商有问题。个个某某亲戚一样，效率，等等，异常，异常，异常，看不懂，真的看不懂。现在京东客服高大威猛，我花7000买台手机还要求着它们，哎，世风日下！



银色

256GB

公开版

2017-11-11 10:17

举报

44

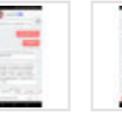
16



木头小卷儿

京享值11012

2017年10月30日拍的8P64G的，31日上午到货，打开包装后手机就是脏的（左上角有个黑点，怎么也擦不掉，左下角缝里面都是灰），见图二图三，当时联系客服，见图四五六七答复，也没说不明白，就说是正品。图一最不能理解，才用一天，就出现卡顿现象，怎么点也不好用，直接关机重启，难道这就是所谓的8P吗？买的朋友请注意了，而且今天11月4又便宜了几百，这价格波动也太快了，我不是喷子，人格担保，只是把买这个手机真实经历说出来，供果粉参考。



金色

64GB

2017-11-04 20:05

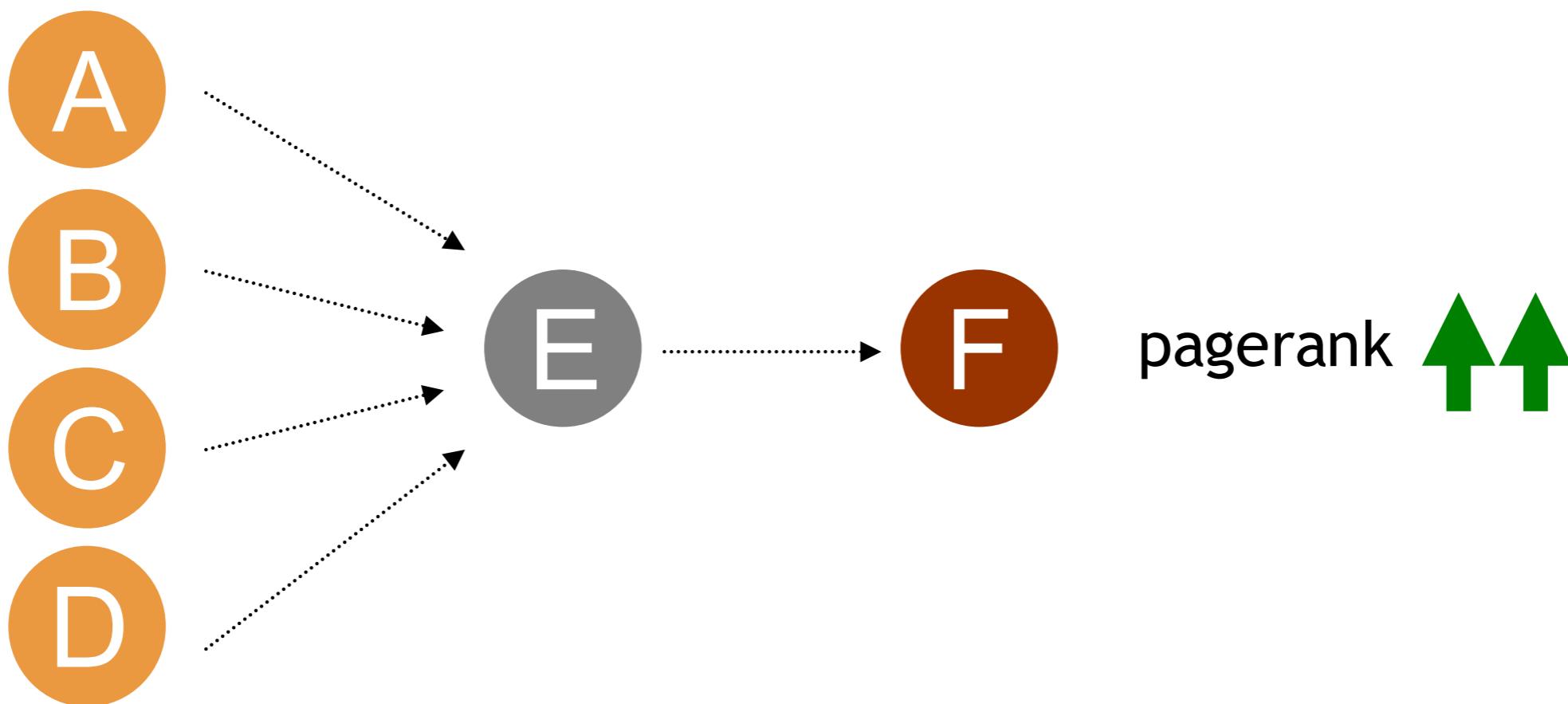
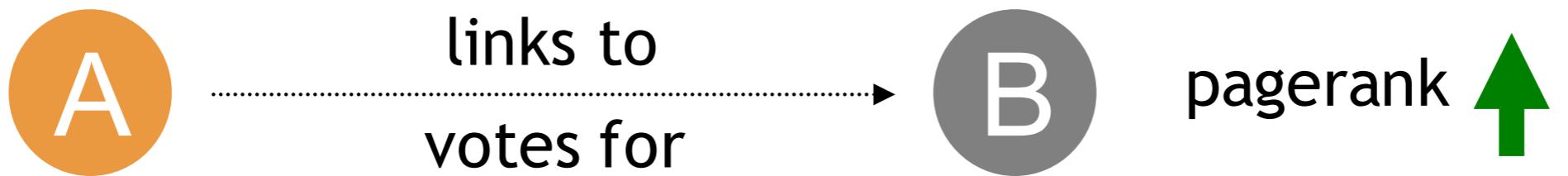
举报

250

14

# *Reputation*

# Pagerank



- 源、目标、声明
- *Explicit vs. Implicit*
- 上下文相关
- *Local vs. Global* ←
- 个人喜好
- 陌生人
- 负面 *vs.* 正面

信息

激励

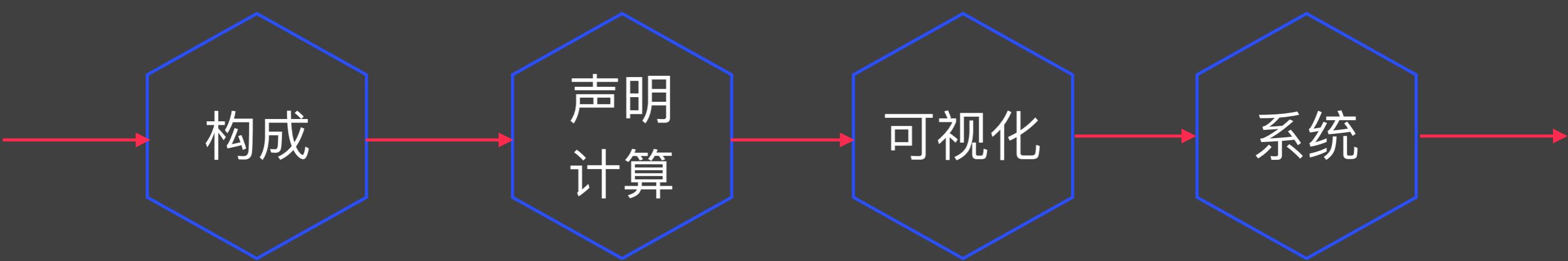
自选

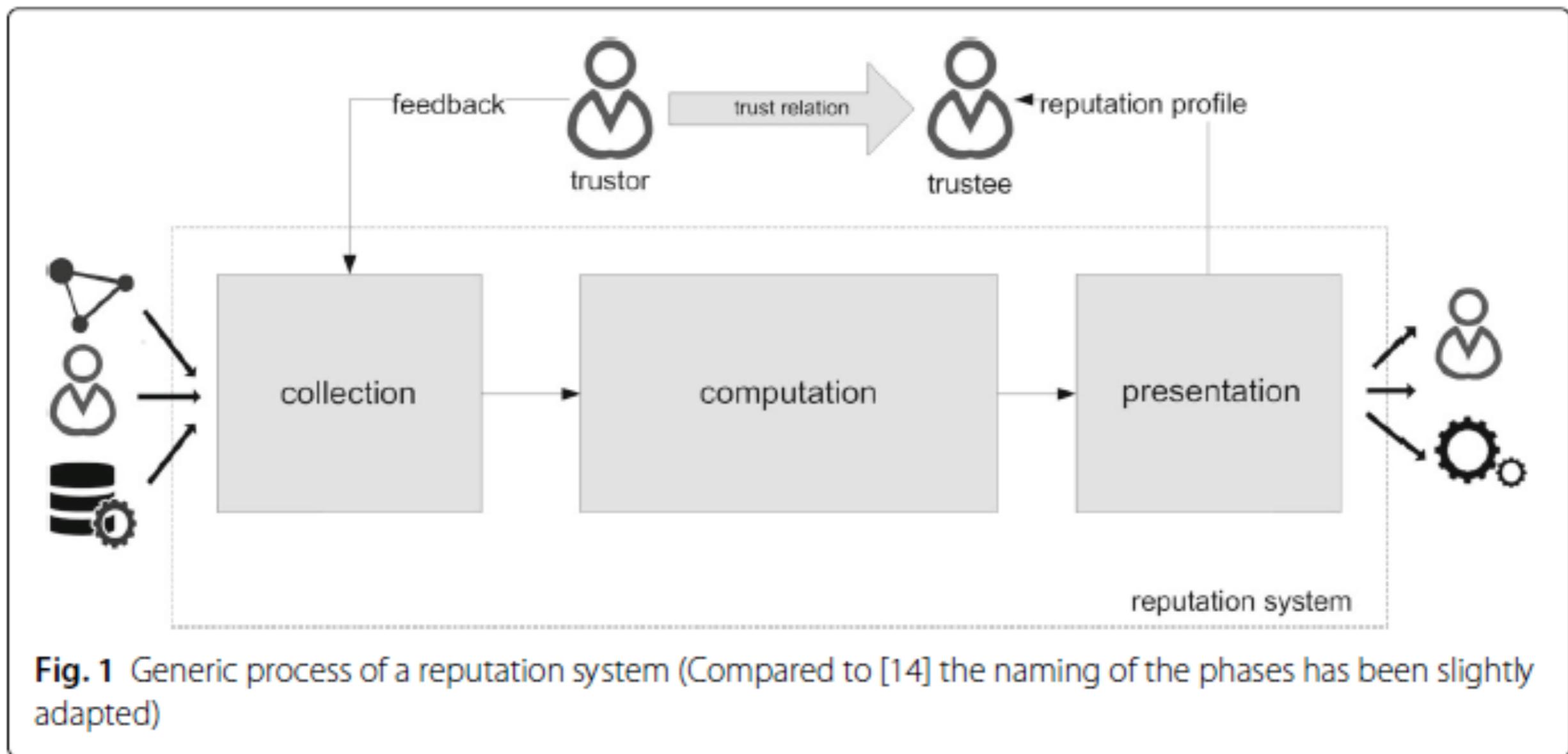
匿名

名字

租赁

# 信誉系统





收集

汇聚

分布

# Reputation

## 信誉语法

### ● 源

- \* ID
- \* 用户/实体
- \* 信息源

### ● 目标

- \* 用户/实体
- \* 信誉陈述

### ● 声明

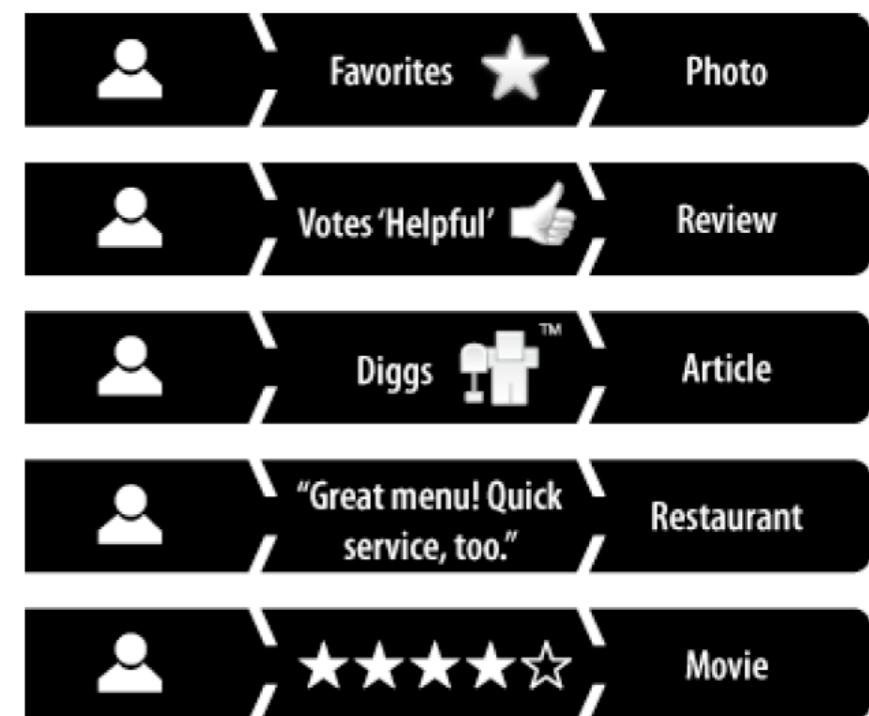
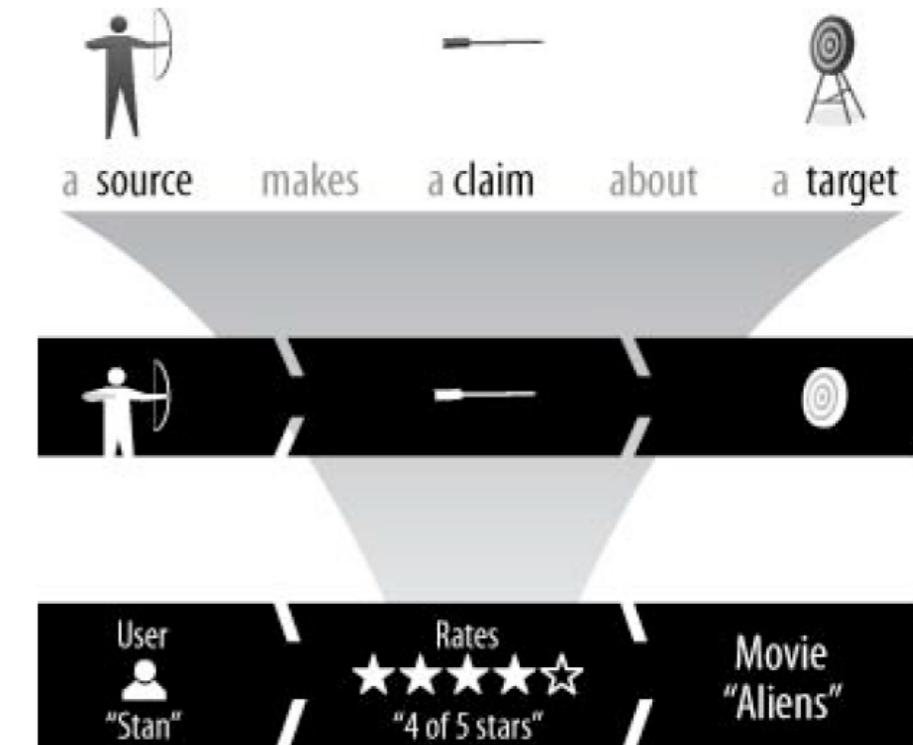
- \* 数字
- \* 定量
- \* 定性

\* 原始Score

\* 规范化Score

\* 文本视频

\* ...



# Reputation

## 声明

2,897 条评论 = 排序方式

Who

How

What



这个视频太好了！！！>

CANCEL

评论

LEAVE BRITNEY ALONE!

0:53 / 2:12 HQ

202,225 ratings 25,558,086 views

Favorite Share Playlists Flag

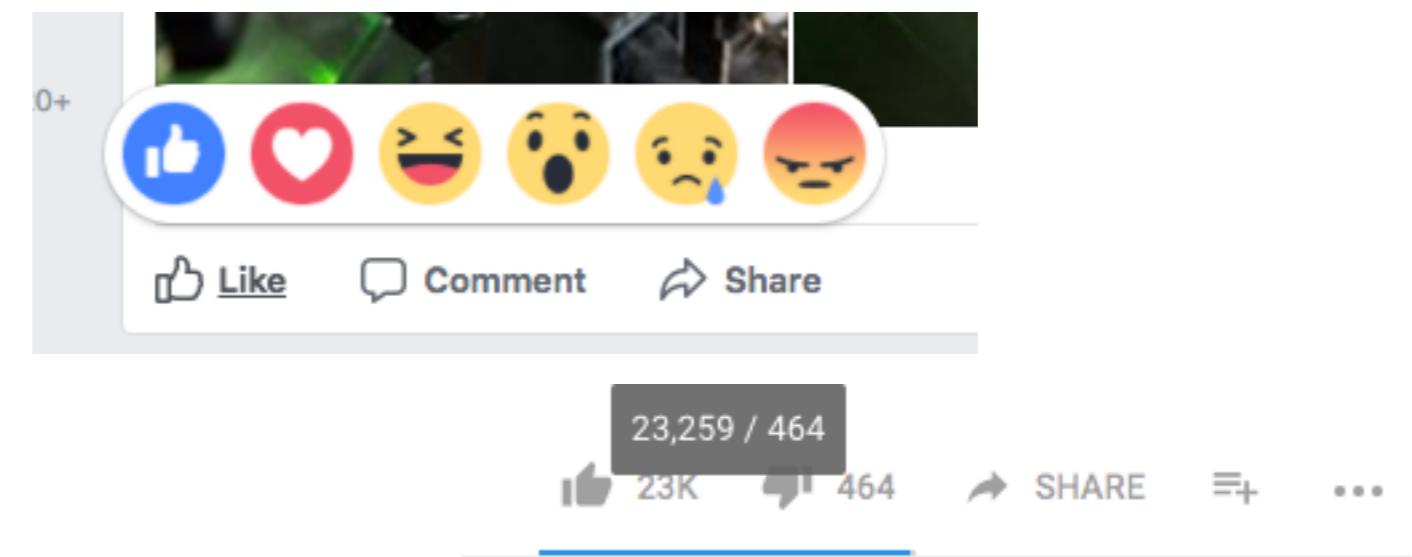
MySpace Facebook Send Video (more share options)

Statistics & Data

Video Responses (2,212)

Post a Video Response

1:09 synweap3... 0:37 amannul... 1:20 rpVerlaine 1:26 SarahAnd...

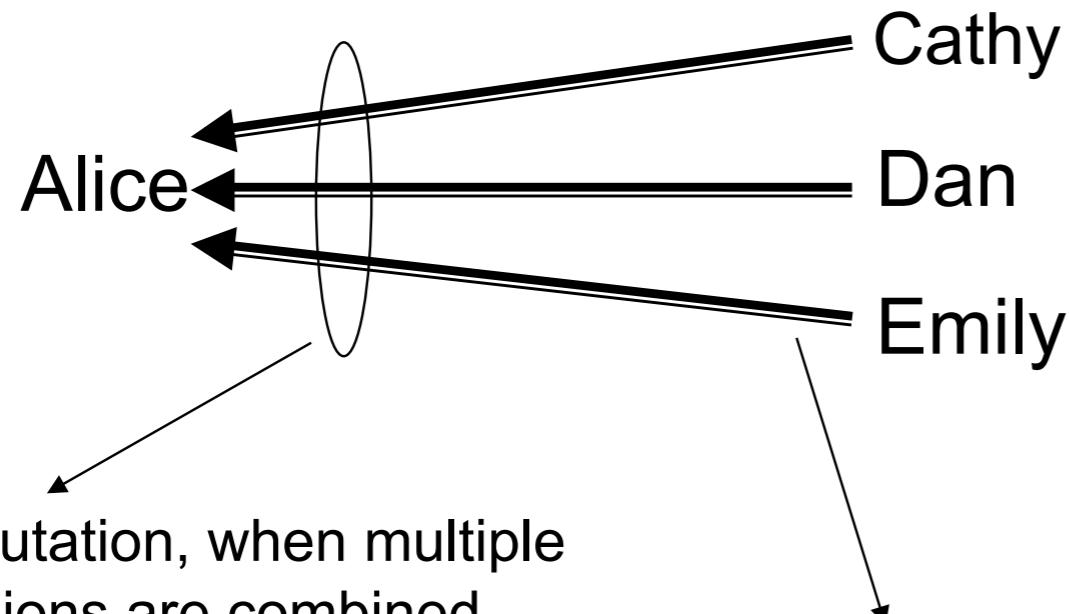
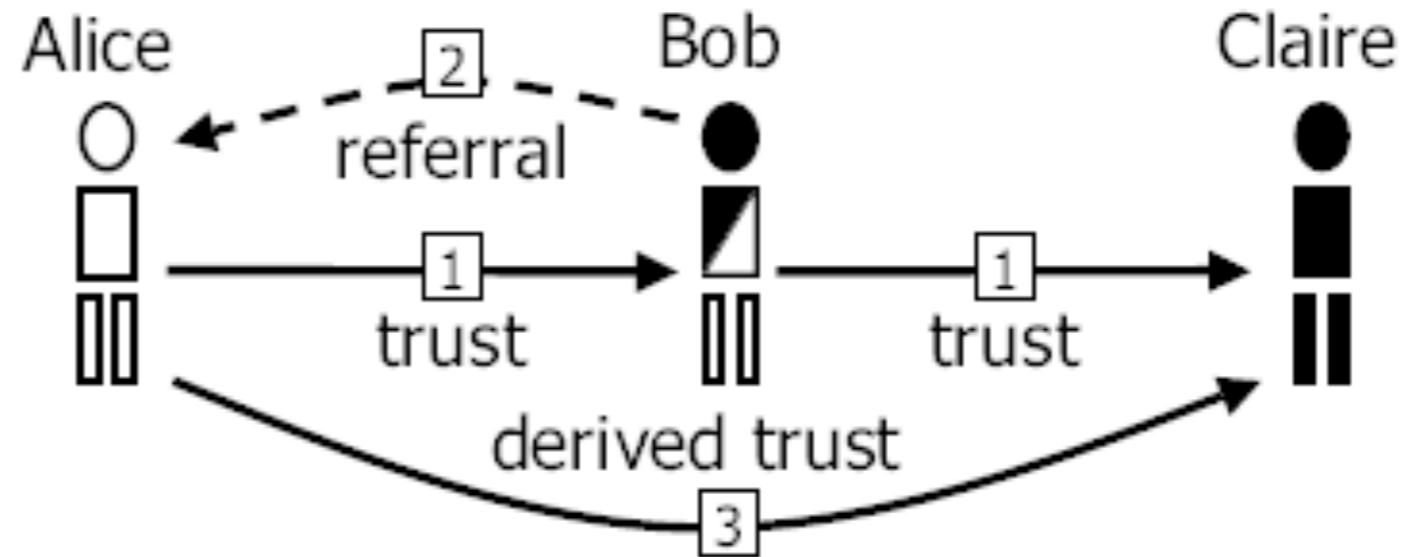


### Detailed seller ratings (last 12 months) ?

Criteria	Average rating	Number of ratings
Item as described	★★★★★	6176
Communication	★★★★★	6802
Shipping time	★★★★★	6673
Shipping and handling charges	★★★★★	7028

# Reputation

# 推荐



Reputation, when multiple  
opinions are combined

Recommendation, containing  
an opinion: "I trust Bob,  
context c, value v"

- 计数
- 累加
- 平均
- 比率

## ● 数据转换

- \* 标准化
- \* 加权
- \* 信息源



- 容易理解
- 便携
- 容易转换
- *bias*

### 商品评价

好评度

99%

就是快(4220) 物流很快(3213) 货真价实(1603) 是真品(1416) 价格优惠(1360) 很漂亮(1096) 性价比高(878)  
够给力(728) 服务态度好(429) 挺不错(404) 自动关机(46) 有杂音(21) 反应迟钝(11)

全部评价(26万+)

晒图(500)

追评(1200+)

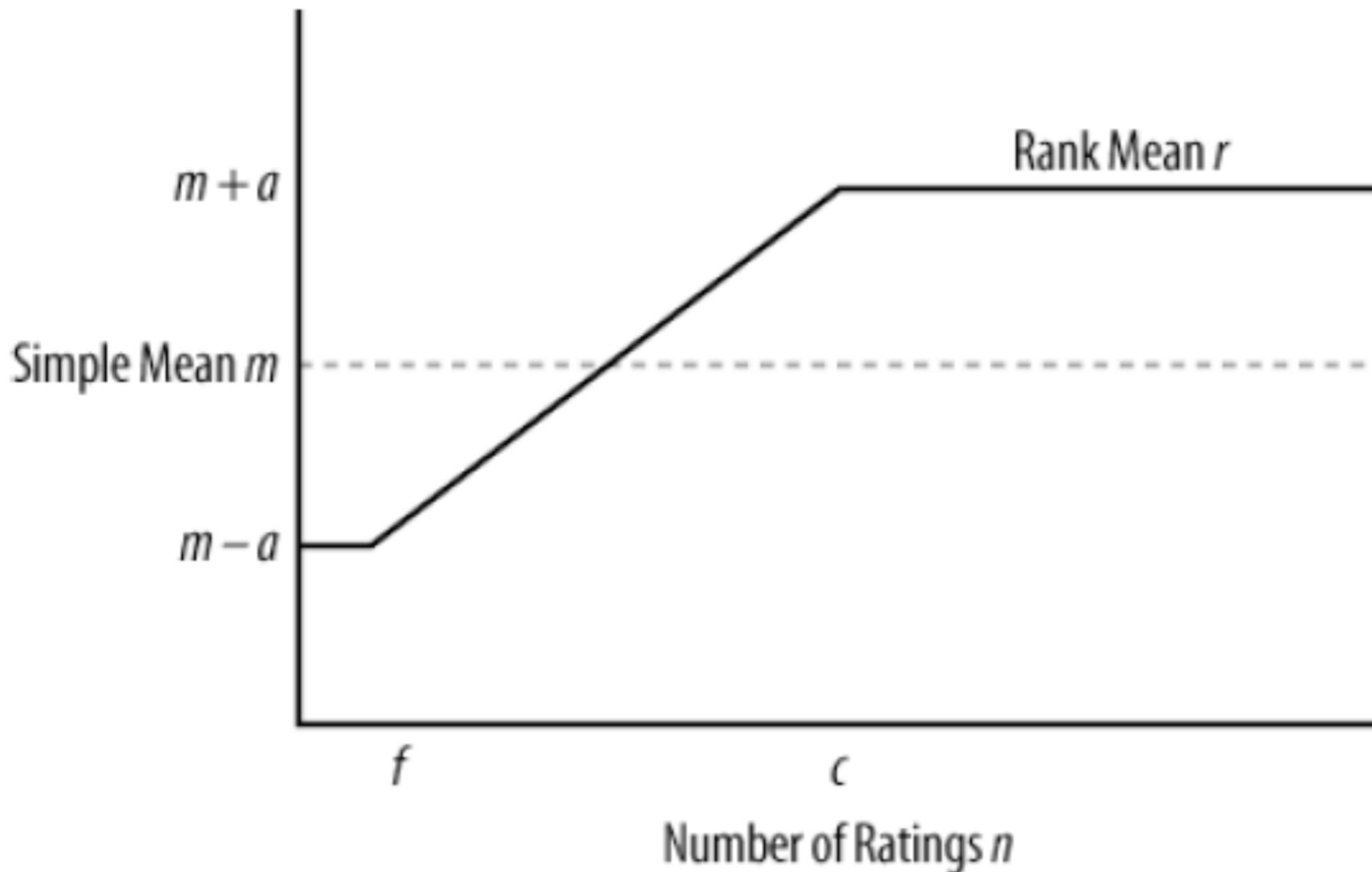
好评(26万+)

中评(1100+)

差评(2300+)

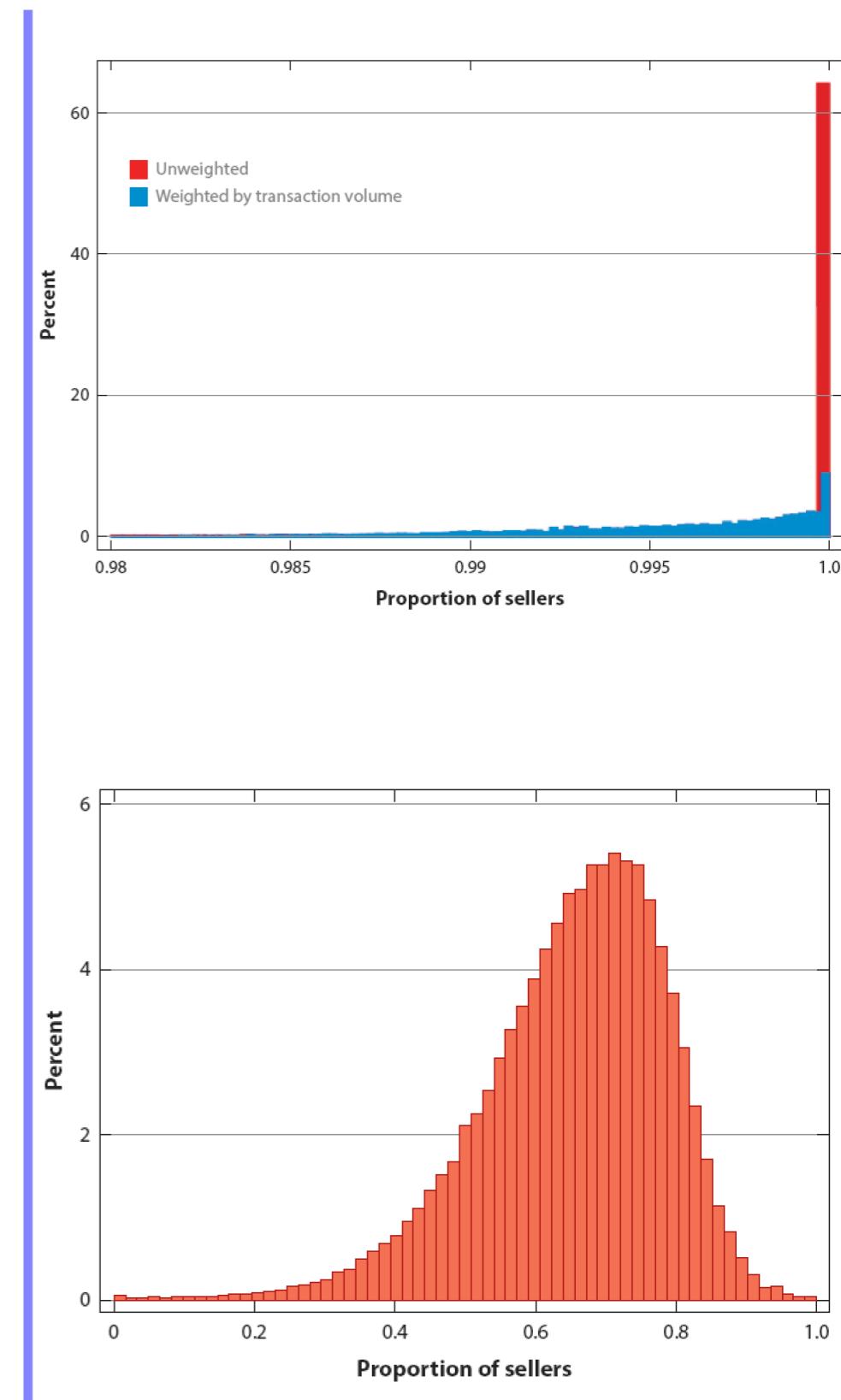
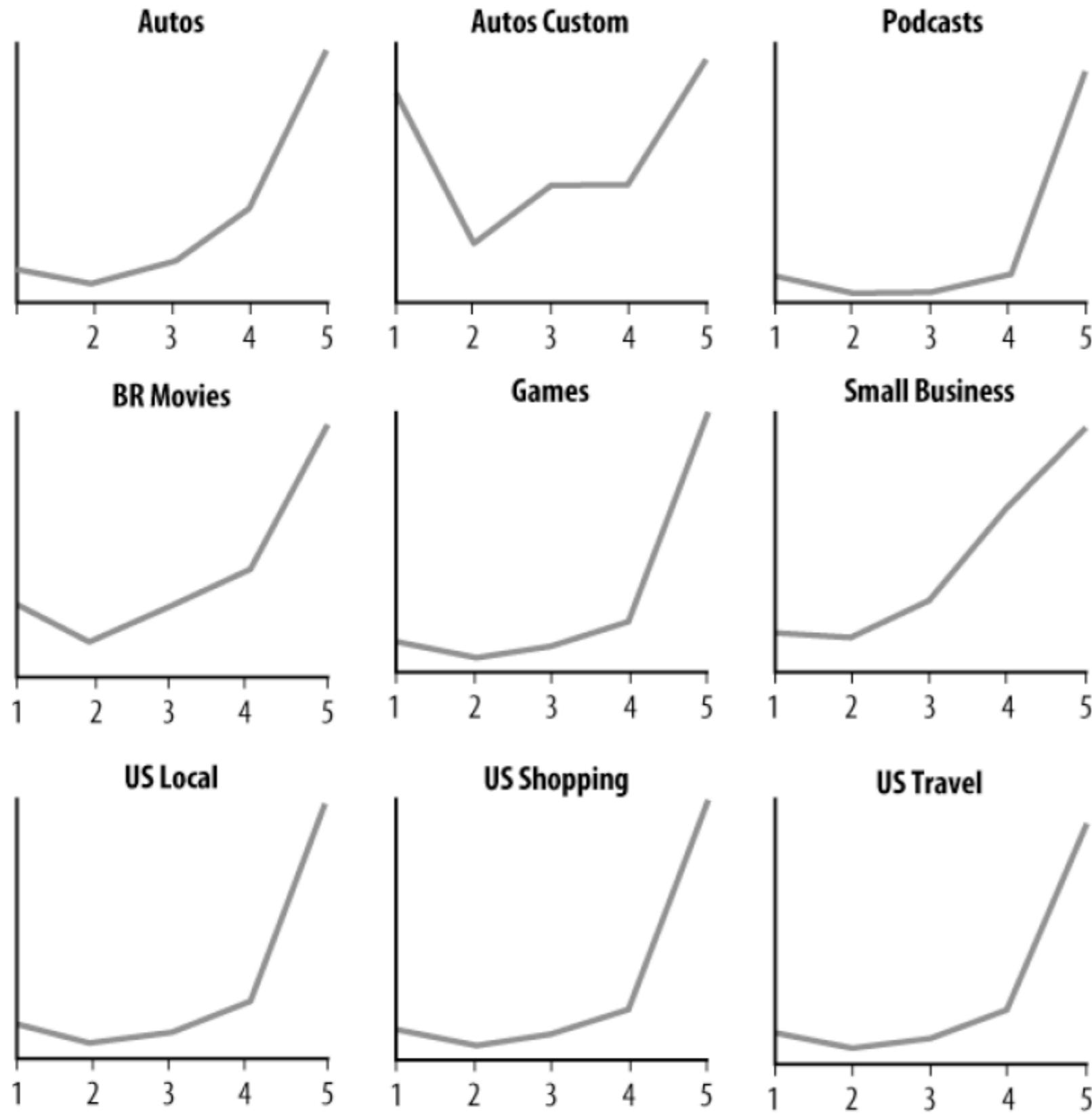
只看当前商品评价

推荐排序 ▼



# *Reputation*

# Bias

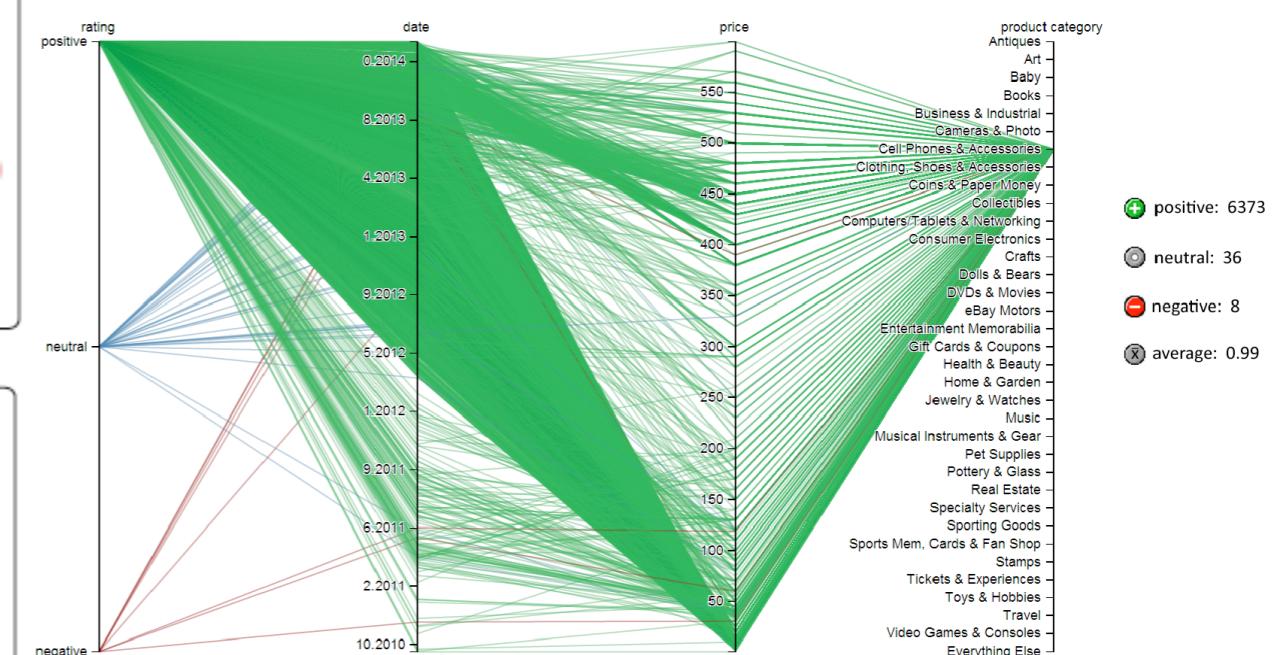
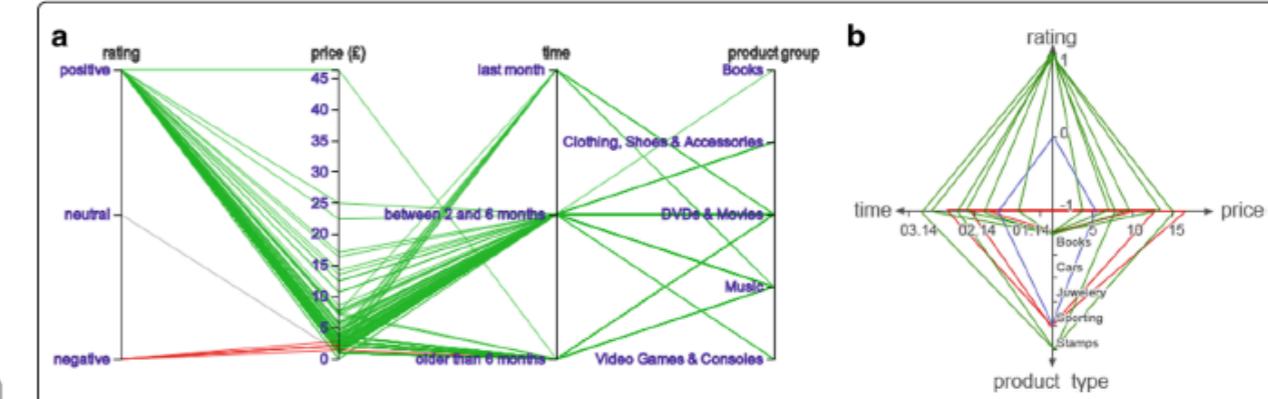
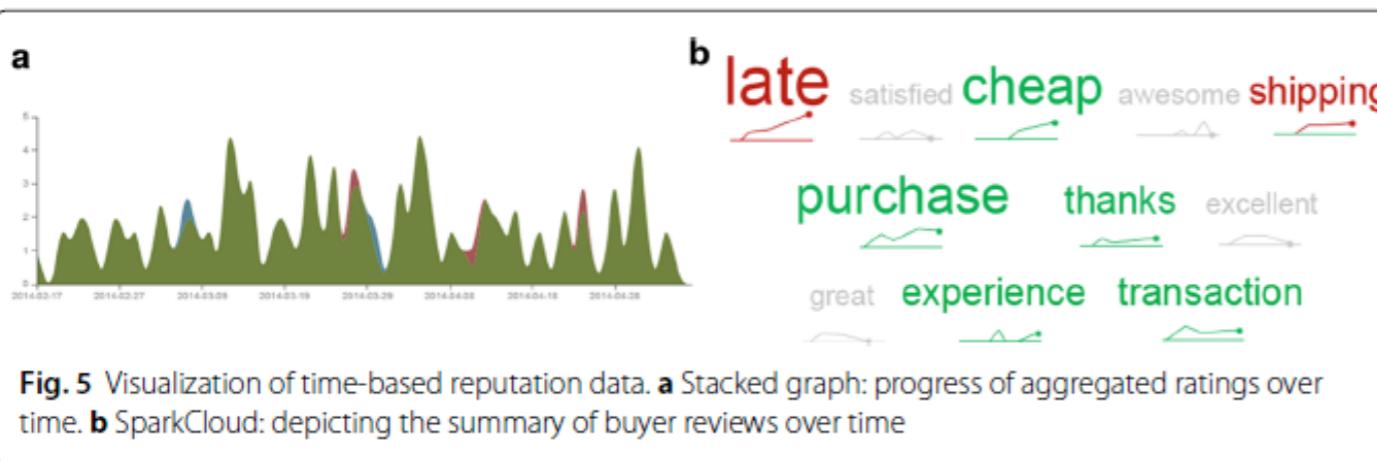
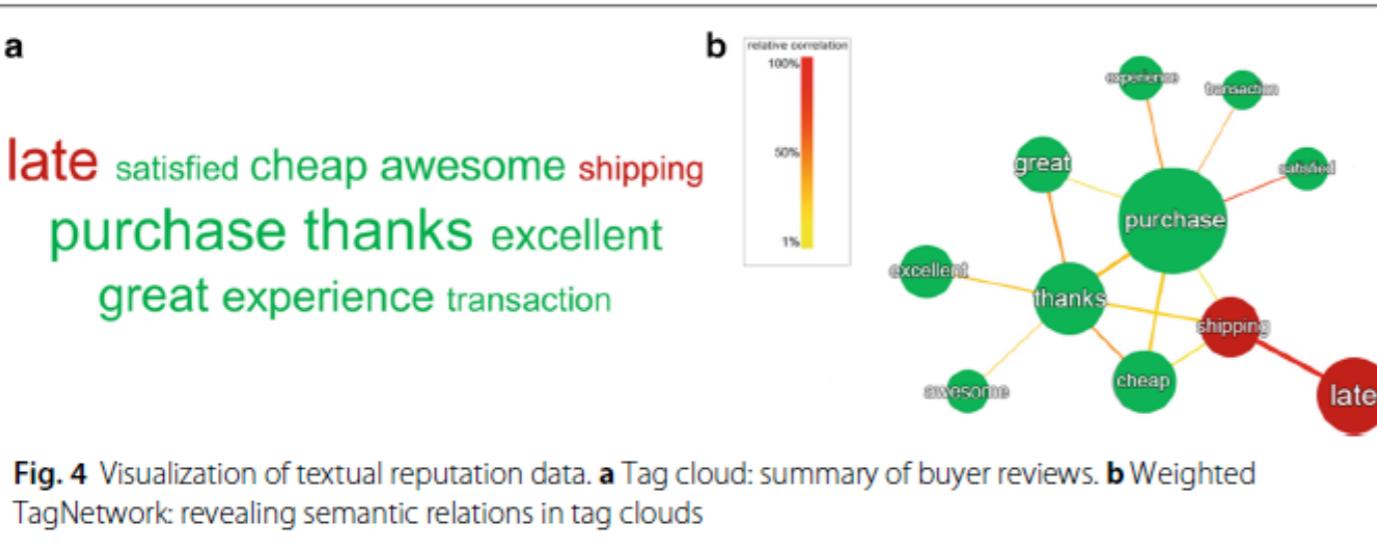
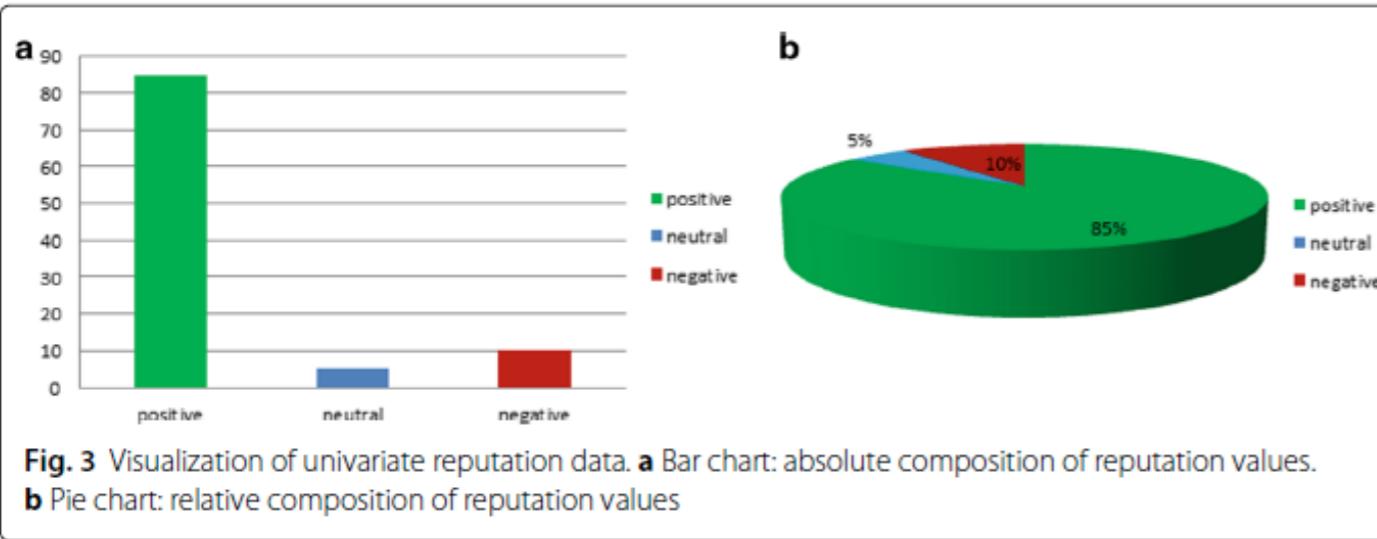


## 再看声明

- *ID*
- 先行者影响
- 冷启动
- 激励
- 错误
- 更新
- 删除
- 声明衰减
  - \* 线性衰减
  - \* 动态衰减
  - \* 时间窗衰减
  - \* 时间限制衰减

# Reputation

# 可视化



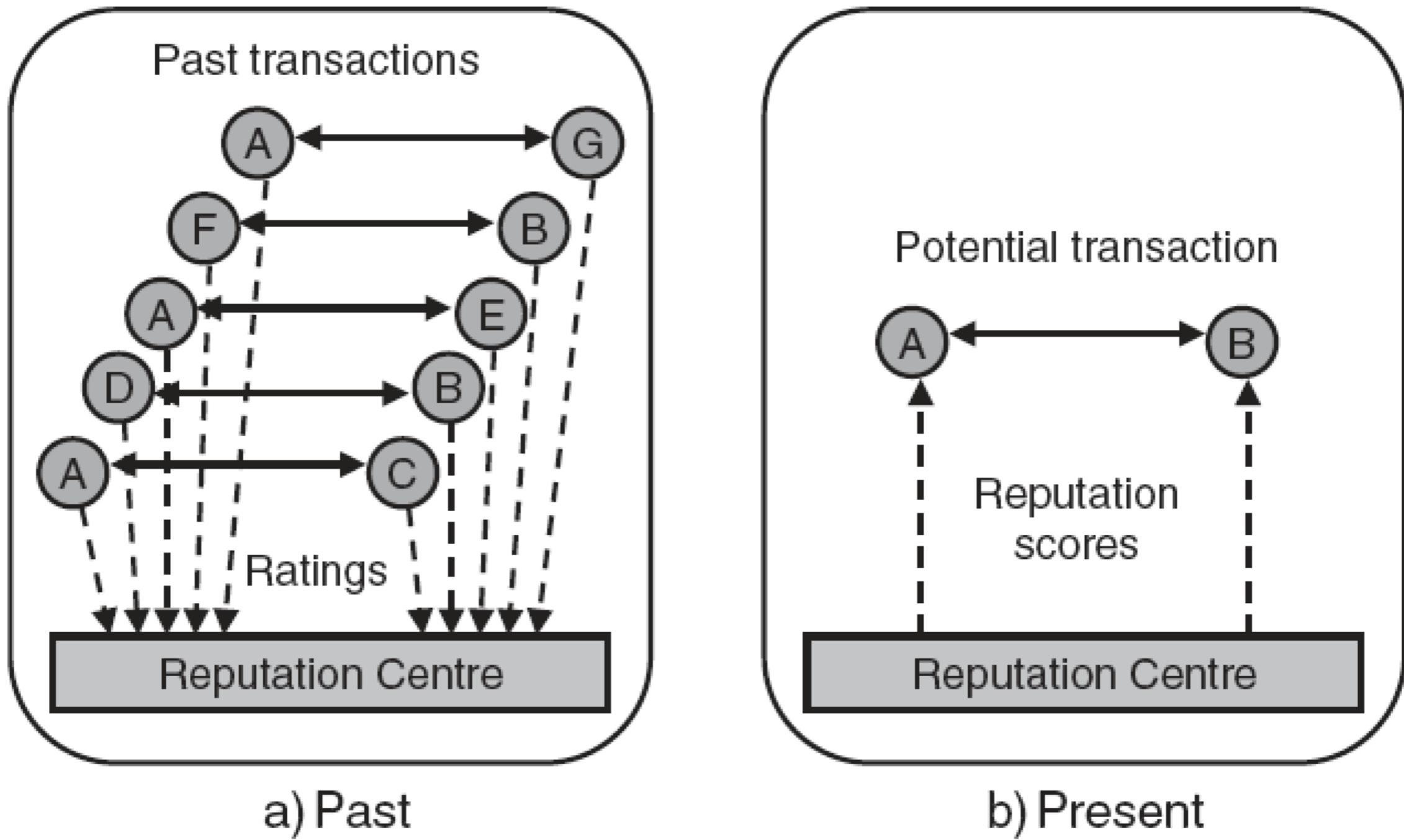


Fig. 3. General framework for a centralised reputation system.

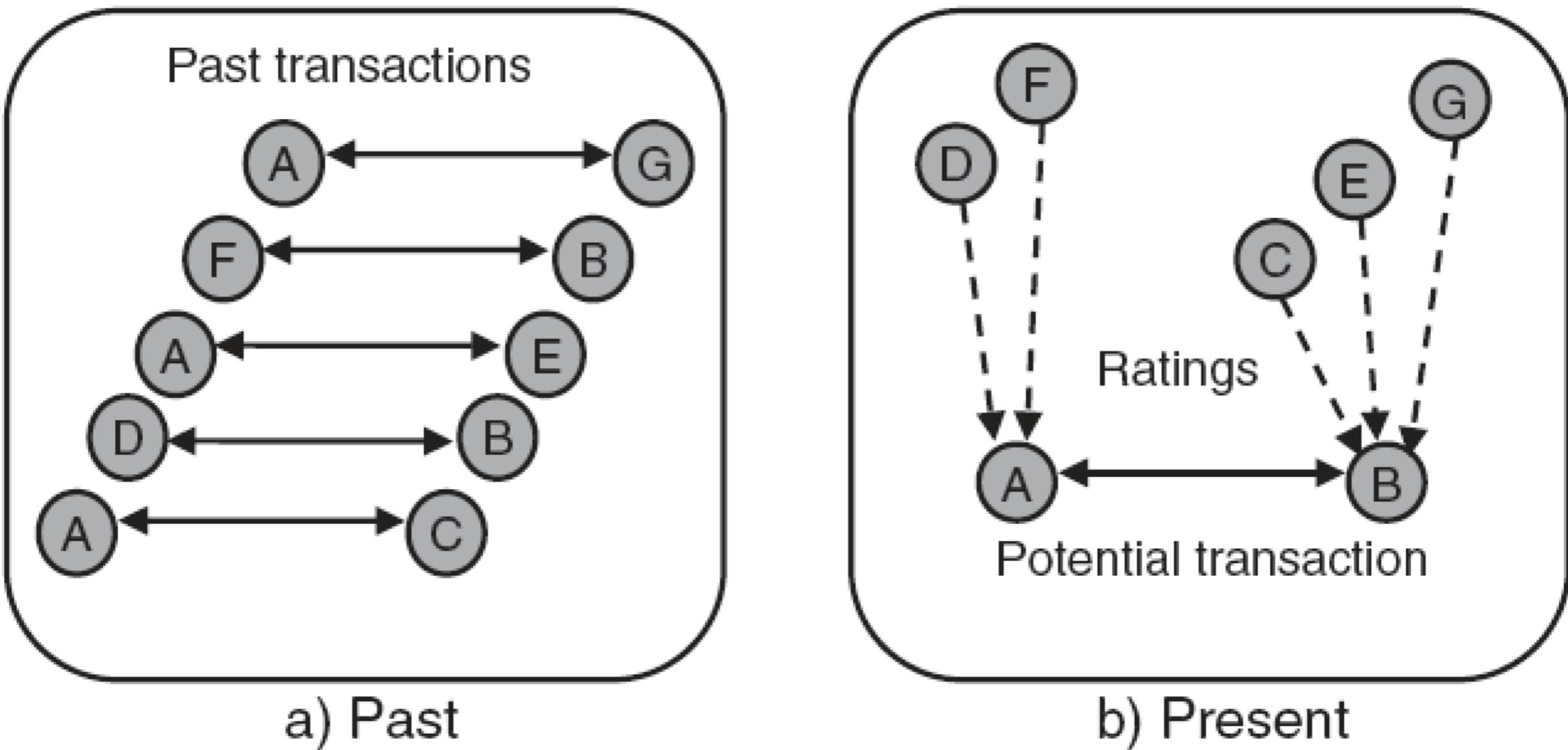
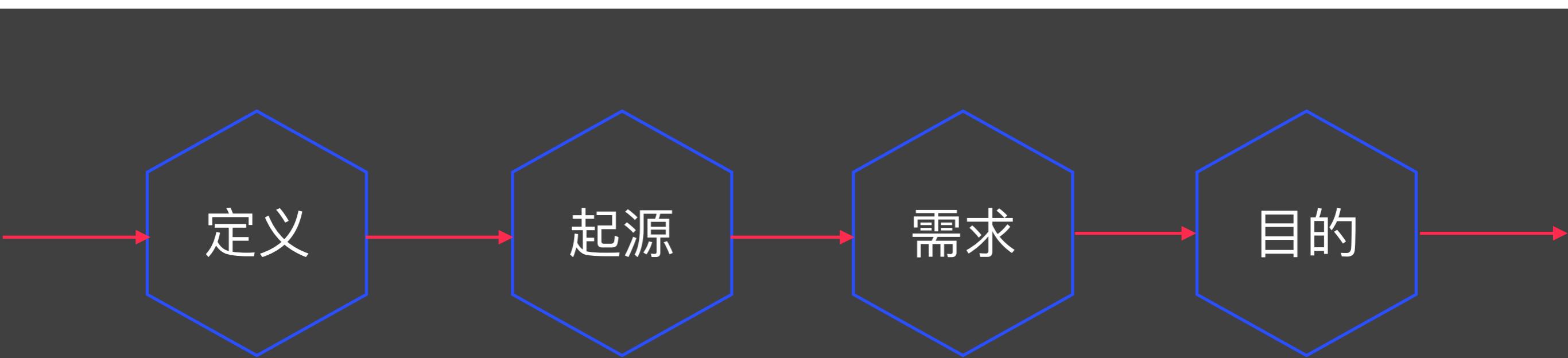


Fig. 4. General framework for a distributed reputation system.

# 信用评分简介



- Credit Scoring is decision support systems used in consumer credit aims at assessment of potential borrowers and existing borrowers.
- Default probability is predicted from observed borrowers characteristics on the basis of the analysis of known performance pf previous customers.
- Risk / creditworthiness is usually measured by default probability.

# Credit Scoring

# 信用评分起源



Character

Capacity

Collateral

Capital

Conditions

自动化

快速

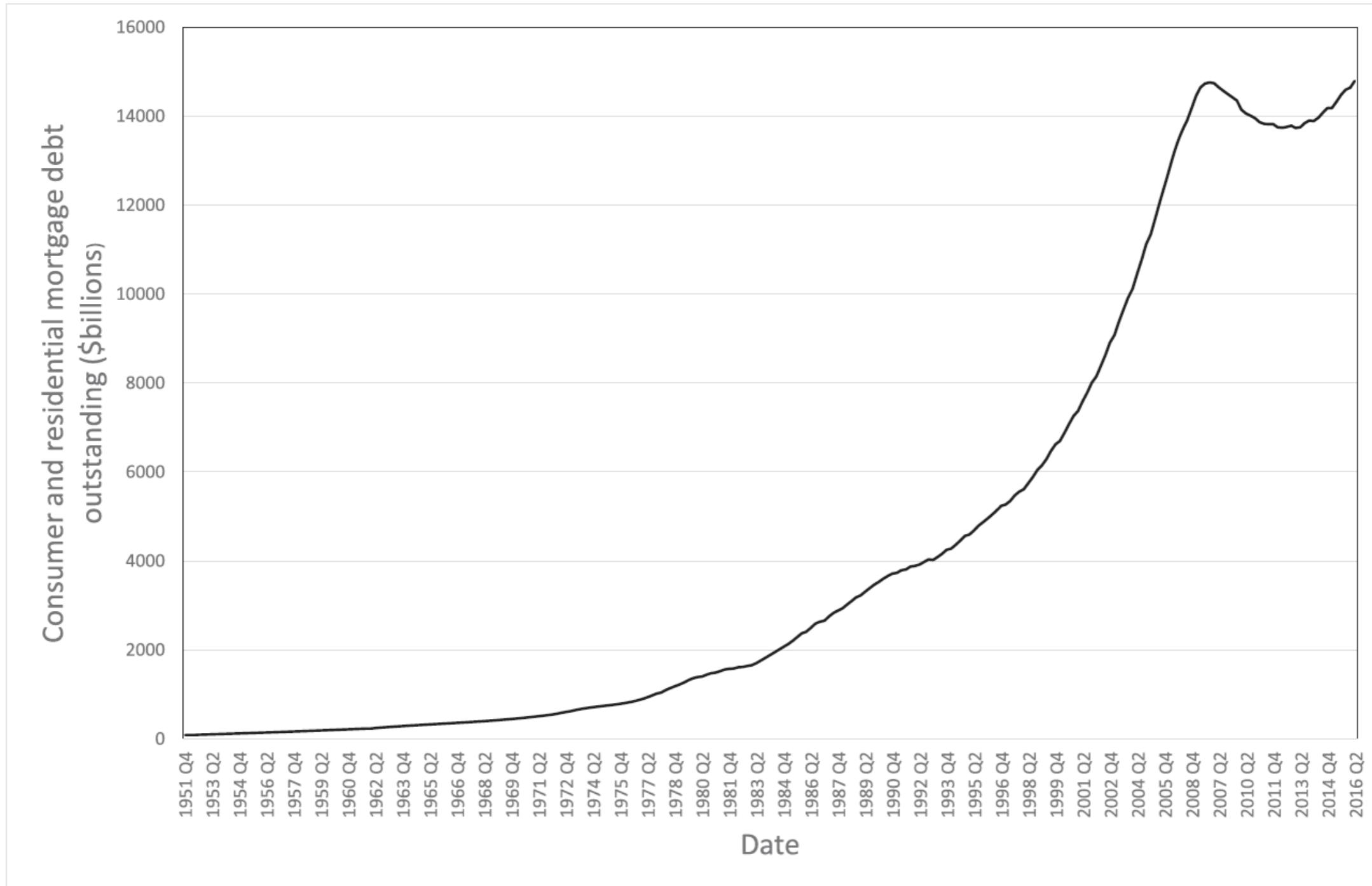
一致

客观

不安全

# Credit Scoring

# 信用评分发展



**Figure 1.1.** *U.S. household debt. Source: Board of Governors of Federal Reserve System.*

## 信用评分需求

### 个人贷款

贷款额度小

贷款客户多

主要是预测

研究较少

教科书少

管理 + 数据

### 企业贷款

贷款额度大

贷款客户少

主要是因果

大量研究

大量教科书

金融 + 会计

### 中小微企业贷款

贷款额度?

贷款客户?

关注?

模型?

研究?

学科?

## 信用评分目的

申请客户

债务违约

产品使用

用户流失

已有客户

信用更新

交叉销售

再次申请

问题客户

预警

催收

坏账

风险定价

抵押担保

利润评分

客户评价

资本充足

风险度量

IFRS9

# 为什么需要信用评分

风险  
评估

凭直觉

凭关系

凭信誉

封闭环境

担保抵押

偿还能力

借方特征

使用目的

长期训练

经验丰富

本地Office

风险保守

信用  
评分

销售产品

业务拓展

利益最大化

业务数量

电话公司

电话购物

电力公司

供水公司

信用卡

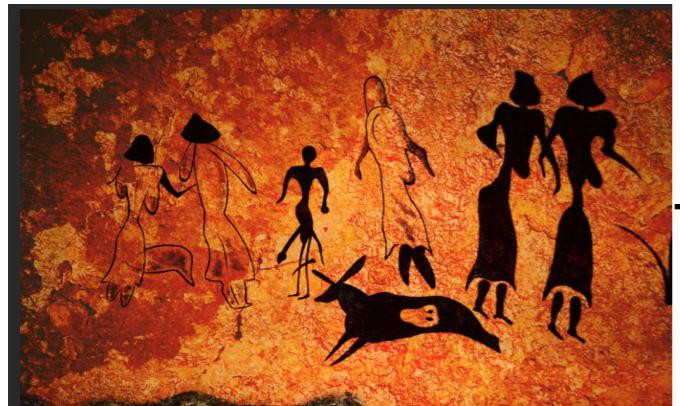
分期

抵押

透支

# Credit Scoring

# 信用历史



...

.....

第一个专家系统

1950

FICO,

1975

Equal Credit Opportunity Acts

1980

Bank, Logistic regression

1992

Credit Scoring Conference, CSCC

2000

巴塞尔协议, 1988, 2005, 2010

2008

次级房贷危机



评分卡

神经  
网络

随机森  
林

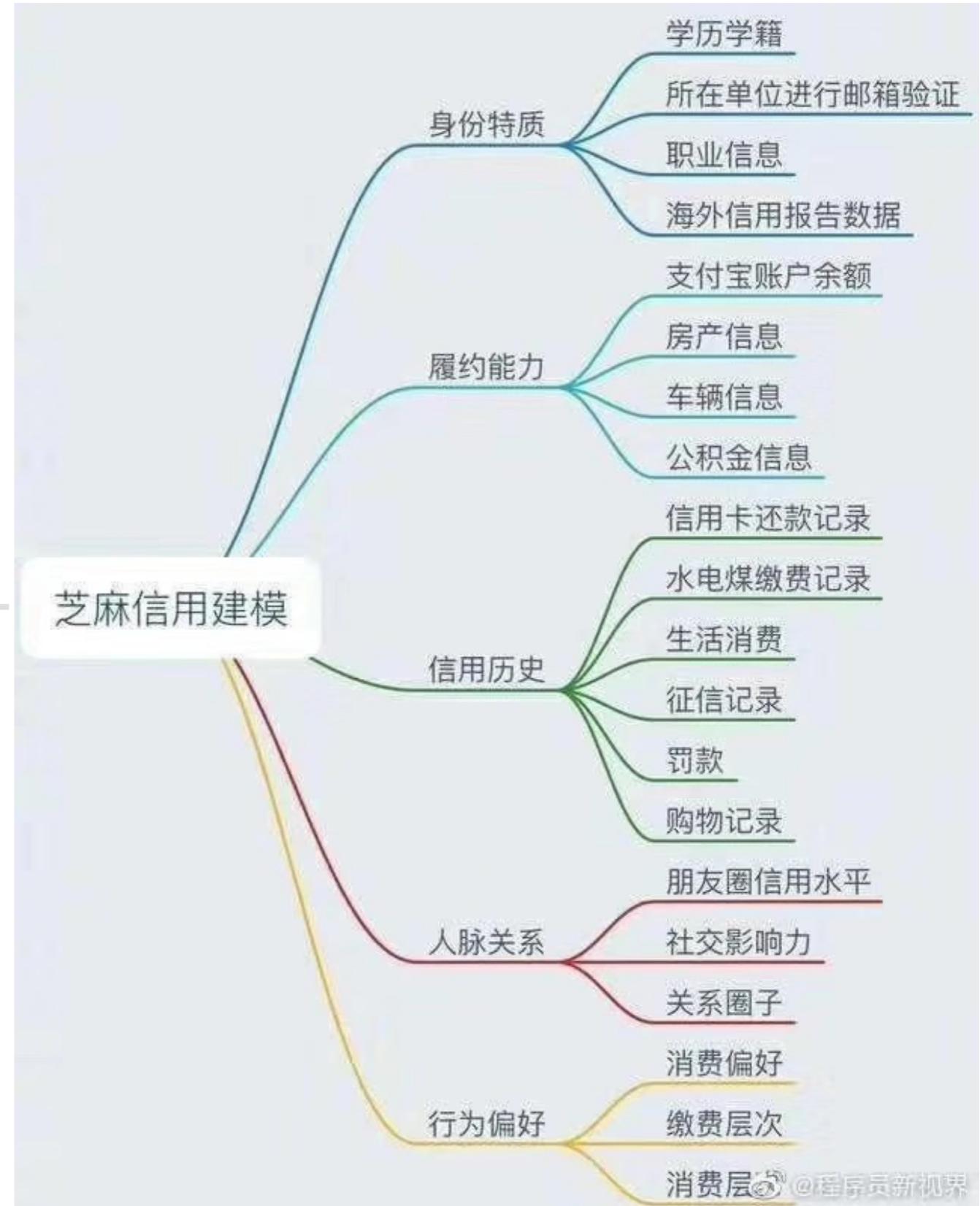
支持向  
量机

种族 地域 性别 年龄

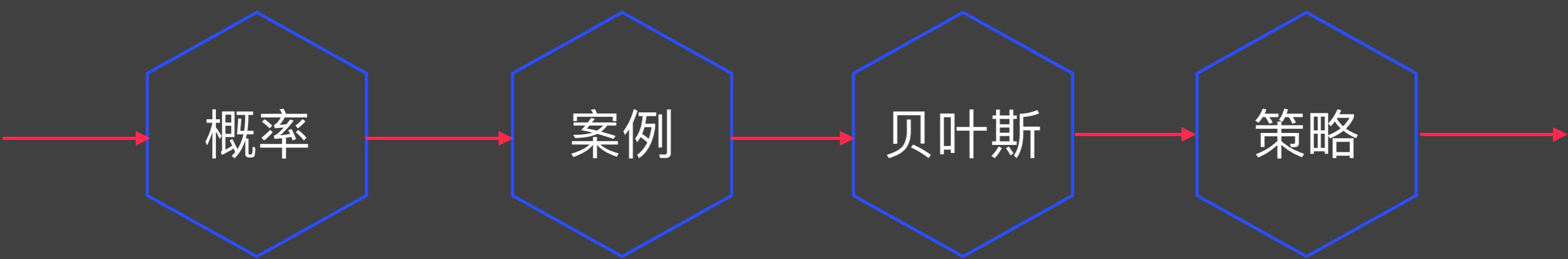
健康 保险 工作 房屋

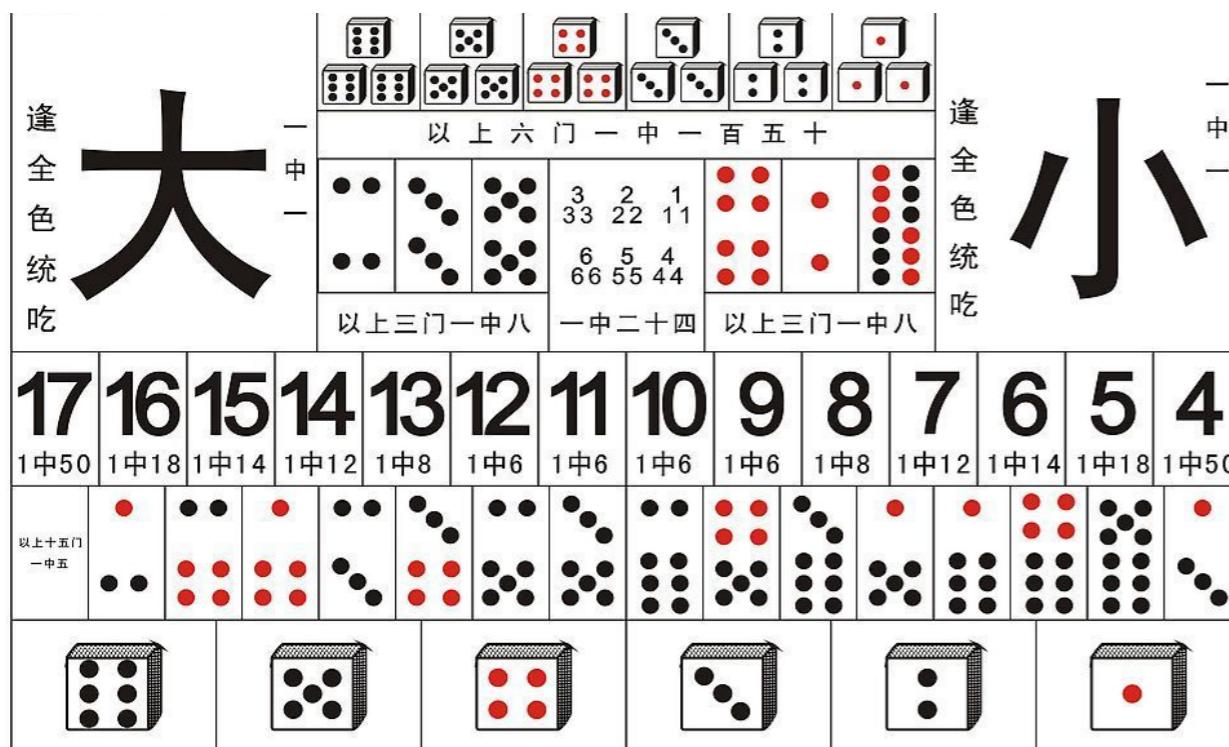
籍贯 孩子 婚姻 负担

现金 账户 支票 负债



# 评分卡





重复实验N次

R次结果是Good

$P(G)$ 、 $P(B)$

$$P = G/(G+B)$$

$$G: B = P/(1 - P)$$

$$O(G) = P(G):P(B)$$

银行有8000客户申请贷款，一年后7000是好的，1000是坏的，好客户平均收益一千，坏客户平均损失一万

好客户弥补坏客户、损益平衡点：10:1

群体好坏概率：7:1

Marital Status:

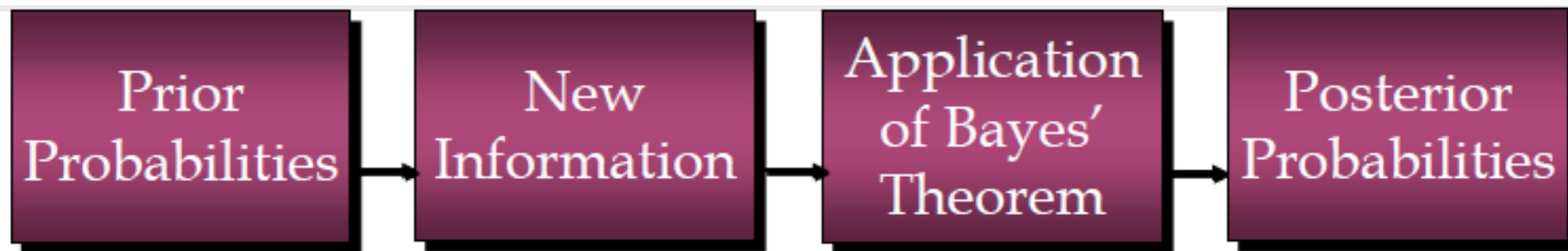
	Good	P(x G)	Bad	P(x B)	Marginal Odds
Married	4900	0.7	400	0.4	49 : 4 12.25:1
Not married	2100	0.3	600	0.6	21: 6 3.5:1
Total	7000	1	1000	1	

$$\text{Marginal Odds of Married} = \frac{0.7}{0.4} \times 7:1 = 12.25$$

$$\text{Marginal Odds of NM} = \frac{0.3}{0.6} \times 7:1 = 3.5$$

Information Odds

NB: Marginal Odds = Information Odds × Population Odds



Let  $\mathbf{X} = (X_1, X_2, \dots, X_m)$  be characteristics (variables) of the borrower such as age, marital status, etc.

$\mathbf{x} = (x_1, x_2, \dots, x_m)$  be outcomes/ attributes of characteristics.

$P(G)$  and  $P(B)$  are prior probabilities.

Posterior probabilities:

$P(G|\mathbf{x})$  is the probability of being Good given certain attributes

$P(B|\mathbf{x})$  is the probability of being a Bad customer given certain attributes

# Credit Scoring

## 工作时间

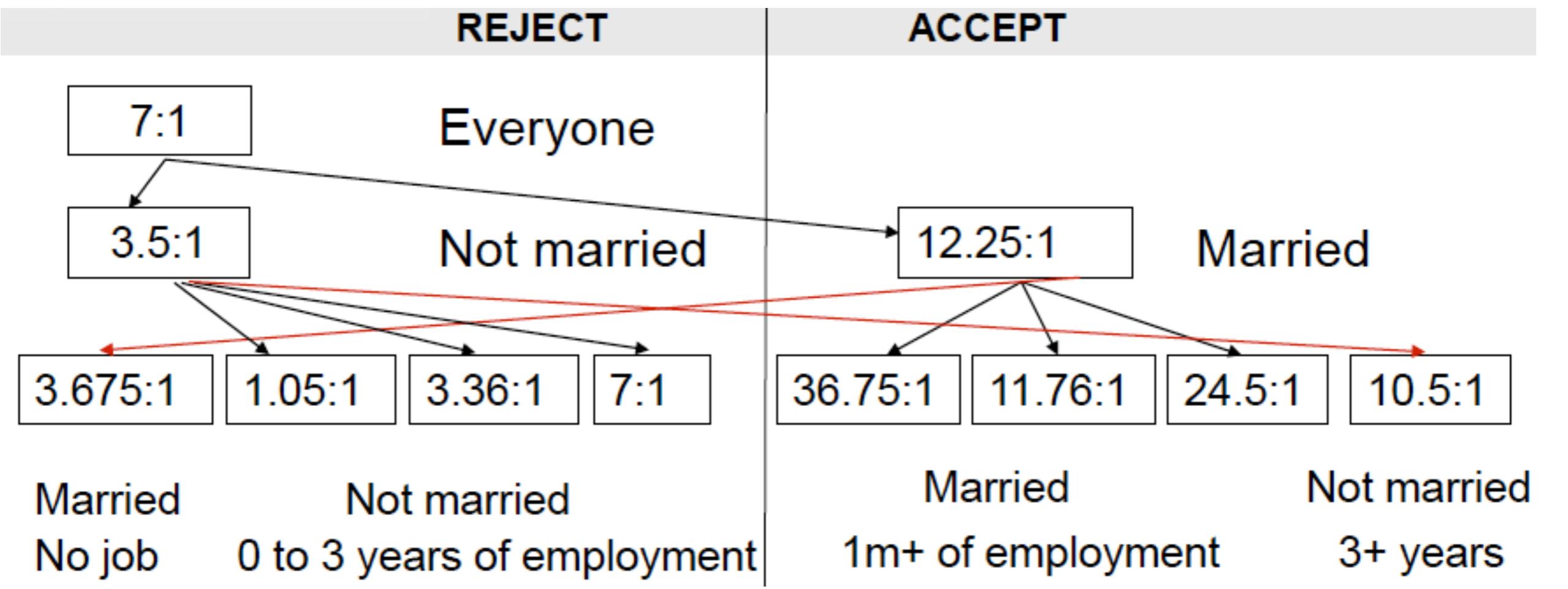
### Marital Status:

	Good	$P(x G)$	Bad	$P(x B)$	Marginal Odds, $O(G x)$	
Married	4900	0.7	400	0.4	49 : 4	12.25:1
Not married	2100	0.3	600	0.6	21 : 6	3.5:1

### Time in Employment :

0	1050	0.15	500	0.5	105 : 50	2.1:1
up to 6 m	1680	0.24	250	0.25	168 : 25	6.72:1
6m - 3y	1960	0.28	140	0.14	196 : 14	14:1
3y+	2310	0.33	110	0.11	231 : 11	21:1
<b>Total</b>	<b>7000</b>		<b>1000</b>			

$\text{Pop Odds} \times \text{Info Odds}(\text{Char 1}) \times \dots \times \text{Info Odds}(\text{Char } n)$



$$\begin{aligned}\text{Odds of Married and No Job} &= 7/1 \times 0.7/0.4 \times 0.15/0.5 = \\ &= 7 \times 1.75 \times 0.3 = 3.675\end{aligned}$$

独立性  
假设

$$\begin{aligned}\text{Odds of Not Married and 3+ years of employment} &= ? \\ &= 7/1 \times 0.3/0.6 \times 0.33/0.11 = 7 \times 0.5 \times 3 = 10.5\end{aligned}$$

# Credit Scoring

# 贝叶斯评分卡

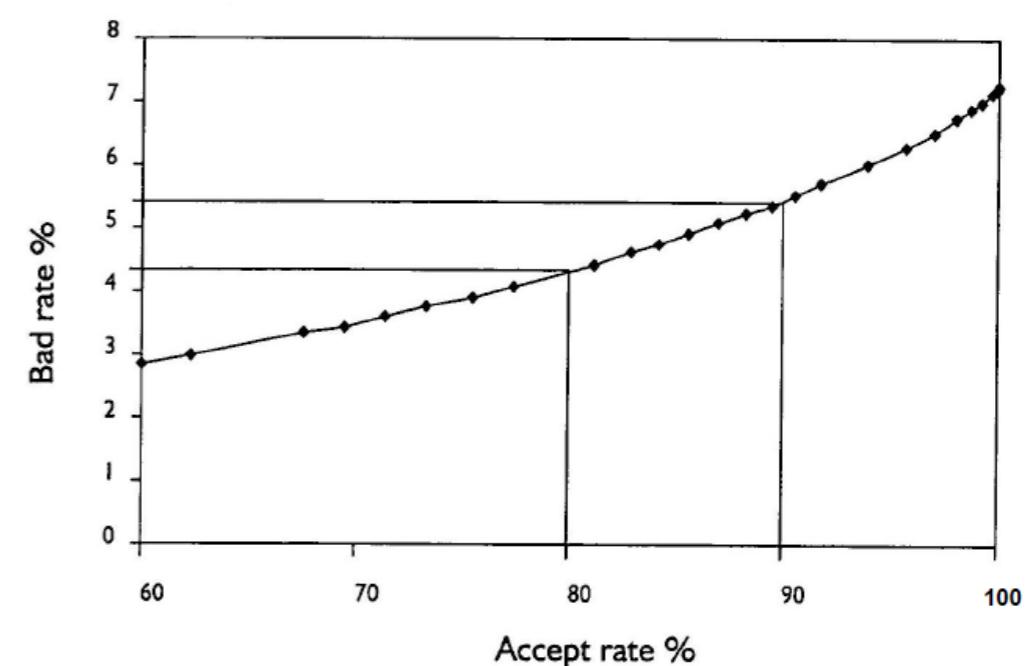
Time at current address	Less than 6 months	6m – 2 years	2 – 6 years	6 - 10 years	10 + years	Unknown		
	0	3	6	13	25	0		
Residential Status	Owner	Tenant	With parents	Unknown		属性	性能	风险
	15	5	2	0				
Banking	Current account	Saving account	Current and saving	No account	Unknown	历史		评分
	5	10	14	0	0			
Occupation	Retired	Full-time	Part-time	Self-employed	Student	Other	Un-known	
	21	16	7	6	5	10	0	
Age	18-25	26-31	32-40	41-54	55+	Unknown		
	5	10	15	20	25	0		

36.75:1	Married	3+ years of employment
24.5:1	Married	6m – 3y of employment
11.76:1	Married	up to 6m of employment
10.5:1	Not married	3+ years of employment
7:1	Not married	6m – 3y of employment
3.675:1	Married	No job
3.36:1	Not married	up to 6m of employment
1.05:1	Not married	No job

数据  
易获取  
自动化  
廉价

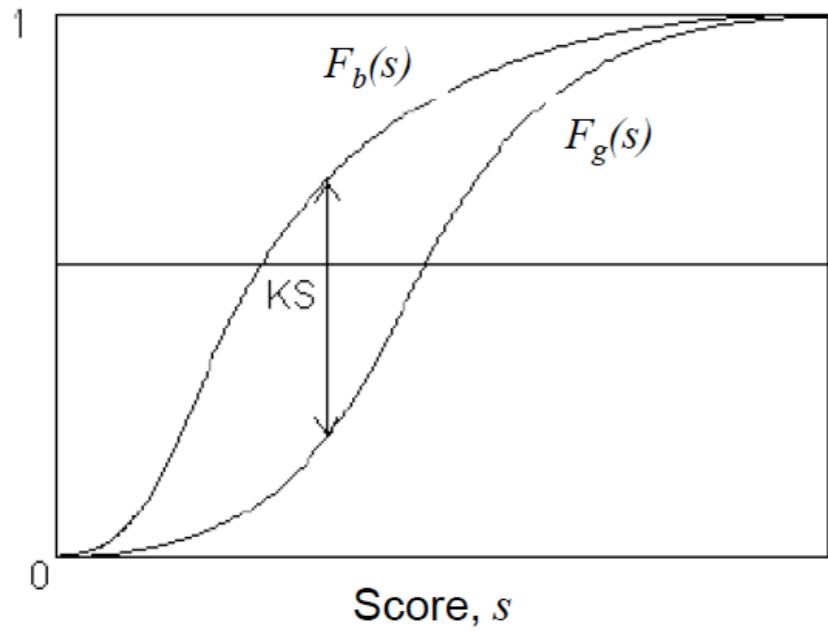
假设：未来和过去相似

信用评分是预测，不是可解释的



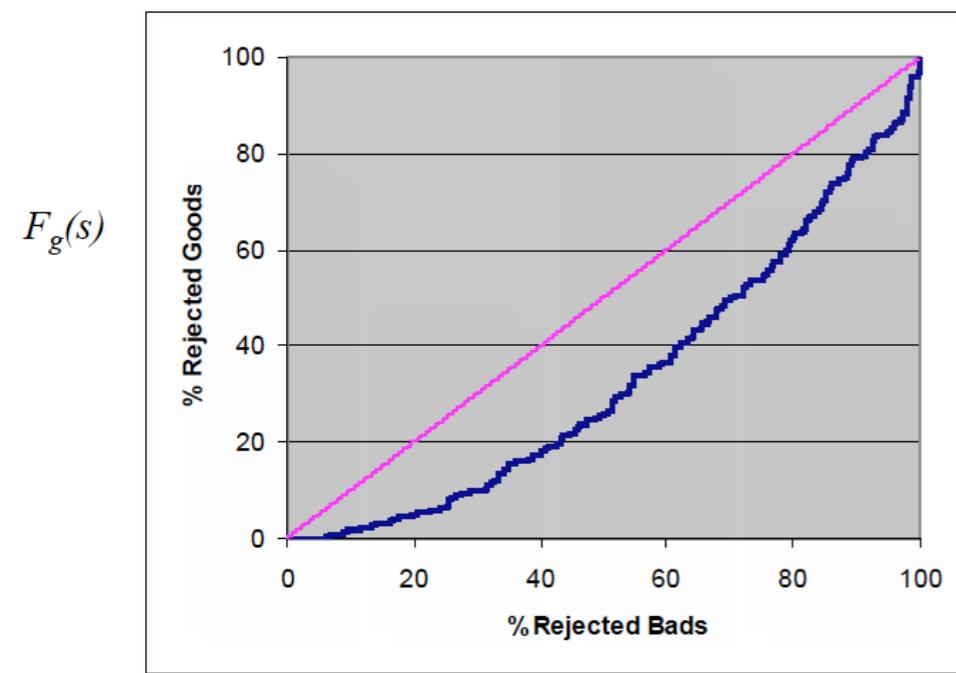
# Credit Scoring

## 评价



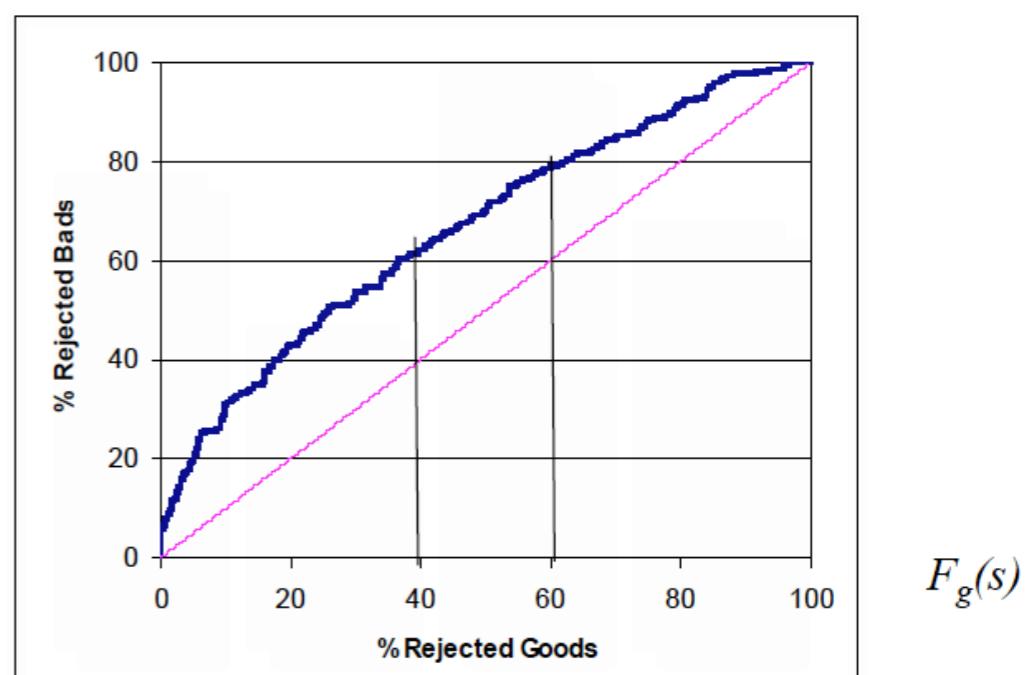
KS

基尼系数



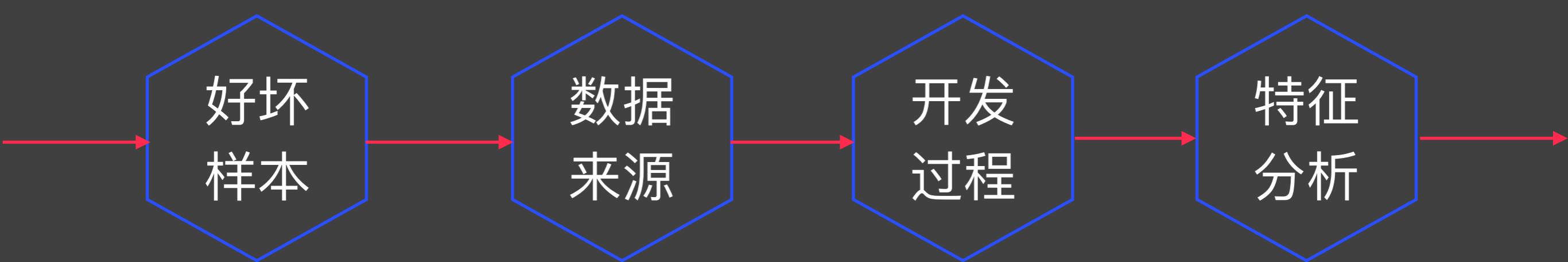
ROC

$F_b(s)$



or Lorenz diagram

# 评分卡



Good

各自1500

Bad

GD?

最好20000-50000, 所有的坏用户

随机性

正确性

实时性

完整性

合法性

6 -12 months

9 – 24 months

季节

Acceptance/ Sample  
period

Outcome/ Performance  
period

长度

# Credit Scoring

## 数据来源

申请表

年龄

婚姻

地址

工作

社交媒体

手机使用

账户

短信使用

抵押

地理位置

司法

银行纪录

借贷

账户

流水

贷款

新数据

# Credit Scoring

## 评分卡例子

Residential status	
Attribute	Score
Owner	30
Tenant	17
Living with Parents	20
Other	0



Age	
Attribute	Score
18–25	5
26–35	10
36–43	15
44+	20

20岁+和父母住+买车+2年居住: **43** (5+20+9+9)  
 55岁+自有住房+女儿婚礼+17年居住: **68** (30+20+0+18)

Loan purpose	
Attribute	Score
New Car	31
Used Car	9
Home Improvement	14
Other	0

SuperPass      SuperFail  
 风险定价      准入条件

Time at present address (years)	
Attribute	Score
< 2	4
2–5	9
6–11	16
12+	18

提高风险管理      减少业务花费  
 丰富客户服务      获取一致性

**Table 2.2.** Some reasons for data collection.

Purpose	Examples
To identify customer	Name, address, date of birth
To be able to contract with customer	Name, address, date of birth, loan amount, repayment schedule, interest rate
To process/score the application	Scorecard characteristics
To get a credit bureau report	Name, address, date of birth, previous address
To assess marketing effectiveness	Campaign code, date of receipt of application, application channel, loan amount, gender, date of birth, address
To effect interbank transfers of money	Bank account number, bank branch details
To develop scorecards	Any information legally usable in a scorecard (laws vary from country to country)

申请数据

征信数据

自有数据

第三方数据

新数据源

准确性

可用性

法律要求

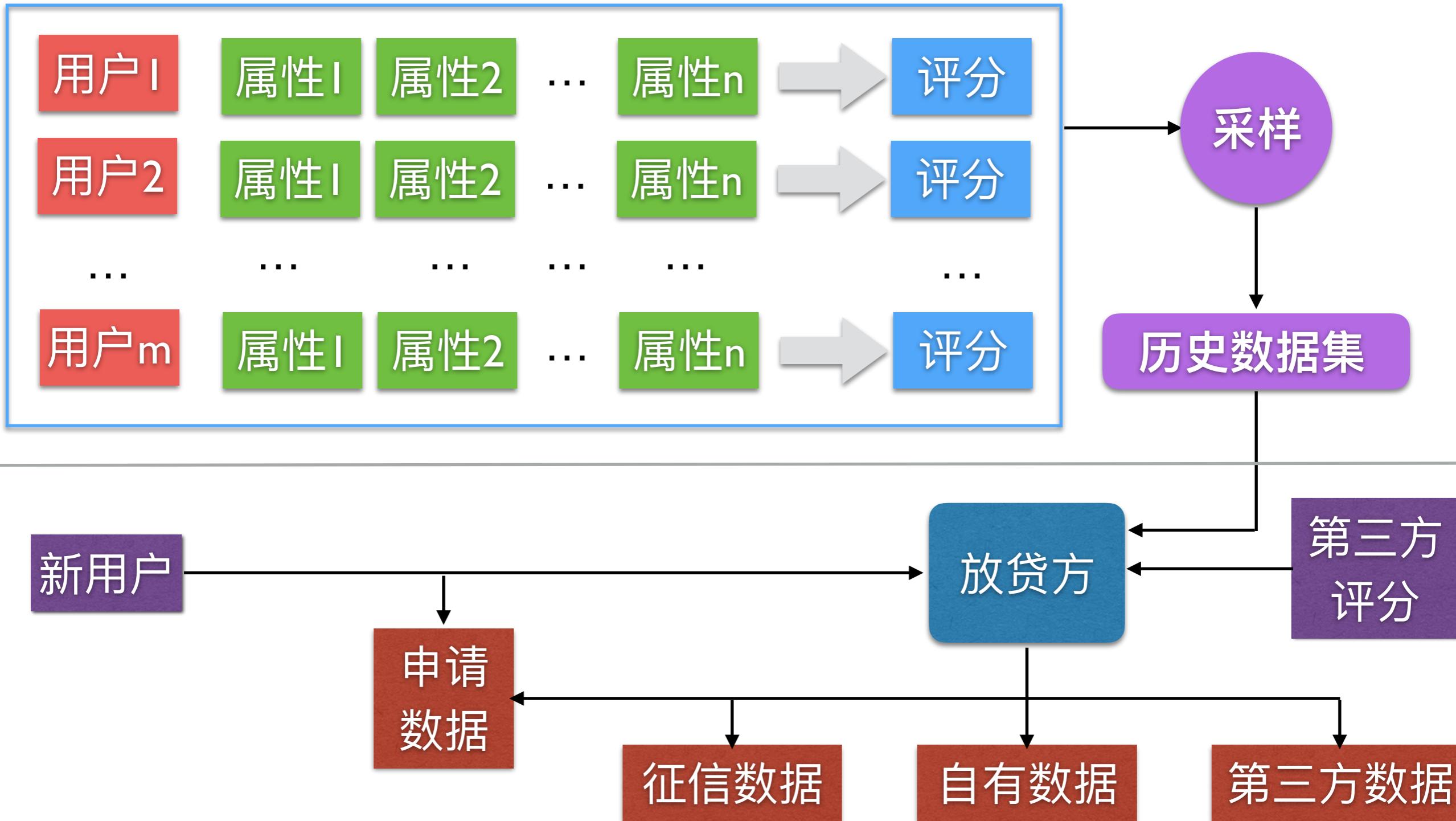
文化影响

数据保护

# Credit Scoring

## 数据处理

历史信息



## 好坏客户

**Good**

一次未还款

不能确定好  
坏的客户

**Bad**

没有足够经  
历的客户

小于6个月还款记录

三次以上未还款  
两次连续未还款

抵押  
贷款

**Good**

**Bad**

损失

**Bad**

**Good**

损失

$$p(\mathbf{x}|G) = \frac{\text{Prob}(\text{applicant is Good and has attributes } \mathbf{x})}{\text{Prob}(\text{applicant is Good})}.$$

$$p(G|\mathbf{x}) = \frac{p(\mathbf{x}|G)p_G}{p(\mathbf{x})}.$$

$$p(G|\mathbf{x}) = \frac{\text{Prob}(\text{applicant has attributes } \mathbf{x} \text{ and is Good})}{\text{Prob}(\text{applicant has attributes } \mathbf{x})},$$

$$s(\mathbf{x}) = \ln \left( \frac{p(G|\mathbf{x})}{p(B|\mathbf{x})} \right) = \ln \left( \frac{p_G p(\mathbf{x}|G)}{p_B p(\mathbf{x}|B)} \right) = \ln \left( \frac{p_G}{p_B} \right) + \ln \left( \frac{p(\mathbf{x}|G)}{p(\mathbf{x}|B)} \right)$$

or  $s(\mathbf{x}) = s_{pop} + \text{woe}(\mathbf{x})$ .

$$p(\mathbf{x}|G) = p(x_1|G)p(x_2|G)\dots p(x_p|G) \text{ and } p(\mathbf{x}|B) = p(x_1|B)p(x_2|B)\dots p(x_p|B).$$

$$\begin{aligned} s(\mathbf{x}) &= \ln \left( \frac{p(G|\mathbf{x})}{p(B|\mathbf{x})} \right) = \ln \left( \frac{p_G p(\mathbf{x}|G)}{p_B p(\mathbf{x}|B)} \right) = \ln \left( \frac{p_G}{p_B} \right) + \ln \left( \frac{p(x_1|G)}{p(x_1|B)} \right) \\ &\quad + \ln \left( \frac{p(x_2|G)}{p(x_2|B)} \right) + \dots + \ln \left( \frac{p(x_p|G)}{p(x_p|B)} \right), \end{aligned}$$

假设  
变量  
独立

	Owner		Not owner		Total	
Age	G	B	G	B	G	B
30-	100	10	200	40	300	50
30+	500	10	100	40	600	50
<b>Total</b>	<b>600</b>	<b>20</b>	<b>300</b>	<b>80</b>	<b>900</b>	<b>100</b>

$$s_{pop} = \ln(900/100) = 2.20,$$

$$\text{woe}(30-) = \ln \left( \frac{300/900}{50/100} \right) = \ln(2/3) = -0.41,$$

$$\text{woe}(30+) = \ln \left( \frac{600/900}{50/100} \right) = \ln(4/3) = 0.29,$$

$$\text{woe(owner)} = \ln \left( \frac{600/900}{20/100} \right) = \ln(10/3) = 1.20,$$

$$\text{woe(not owner)} = \ln \left( \frac{300/900}{80/100} \right) = \ln(5/12) = -0.88,$$

$$s(\mathbf{x}) = s_{pop} + \text{woe}(x_1) + \text{woe}(x_2).$$

$$w_0 + w_1 X_1 + w_2 X_2 + \cdots + w_p X_p = \mathbf{w}^* \cdot \mathbf{X}^{*T},$$

where  $\mathbf{w}^* = (w_0, w_1, w_2, \dots, w_p)$ ,  $\mathbf{X}^* = (1, X_1, X_2, \dots, X_p)$ ,

$$p_i = w_0 + x_{i1}w_1 + x_{i2}w_2 + \cdots + x_{ip}w_p \quad \text{for all } i.$$

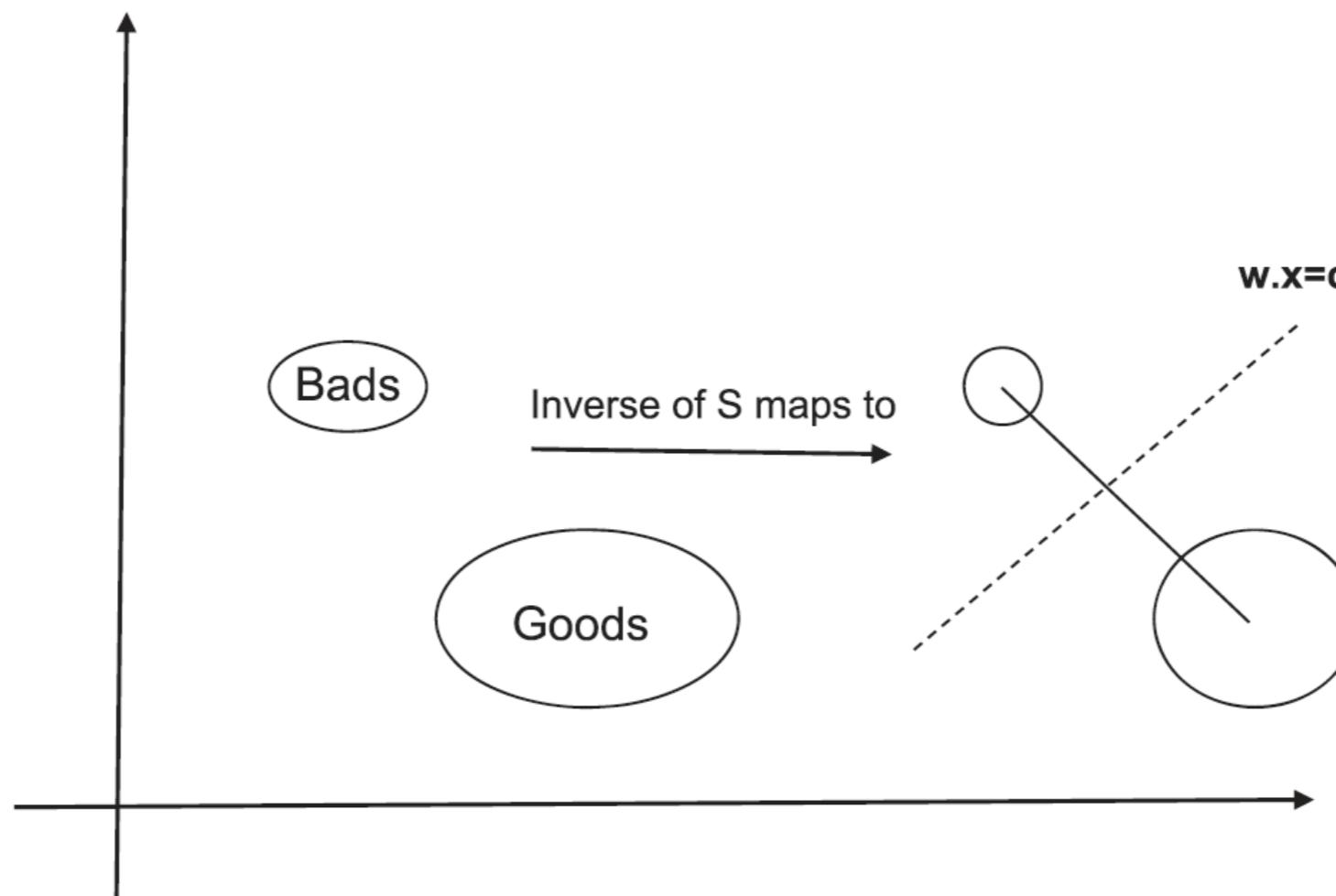
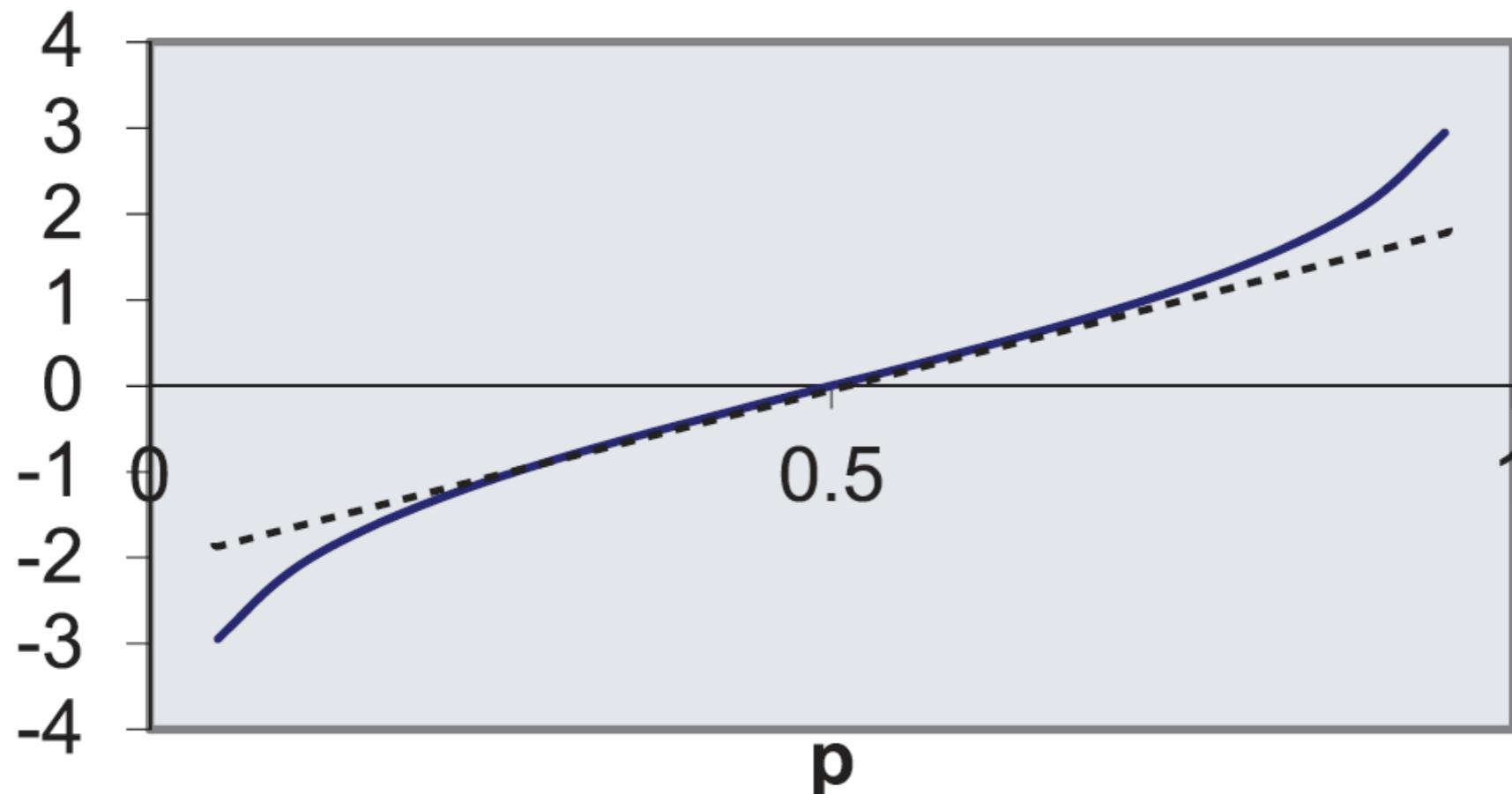


Figure 3.2. Line corresponding to scorecard.

$$s(\mathbf{x}) = \log\left(\frac{p}{1-p}\right) = w_0 + w_1x_1 + w_2x_2 + \cdots + w_p x_p = \mathbf{w} \cdot \mathbf{x}^T.$$



—  $\log(p/(1-p))$  .....  $p$

离散  
变量

Figure 3.3. Graph of  $\log(p/(1-p))$  and  $ap + b$ .

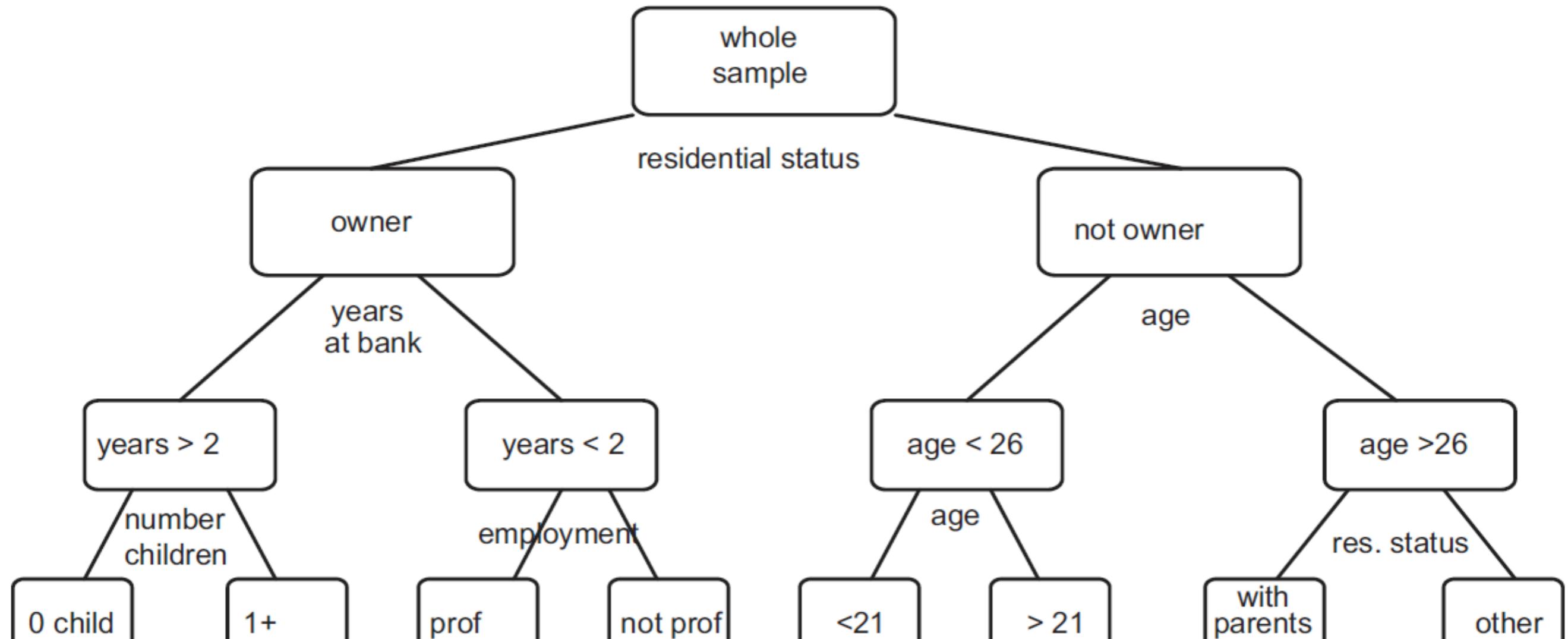


Figure 3.4. Classification tree.

划分  
规则

停止  
规则

分配  
规则

# Credit Scoring

# 其余方法

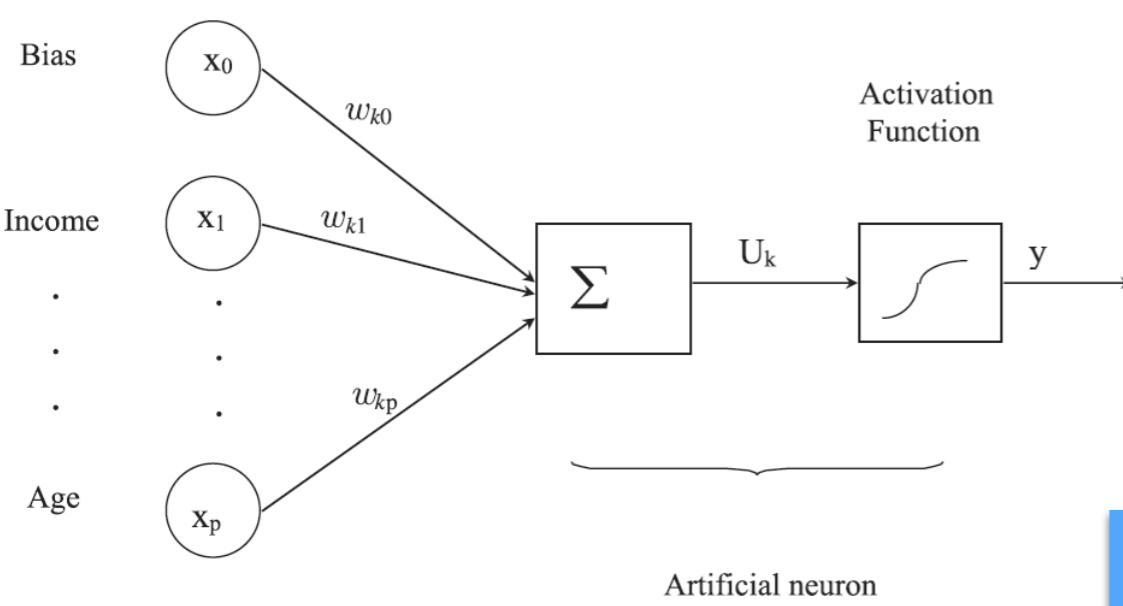


Figure 4.3. A single-layer neural network.

神经  
网络

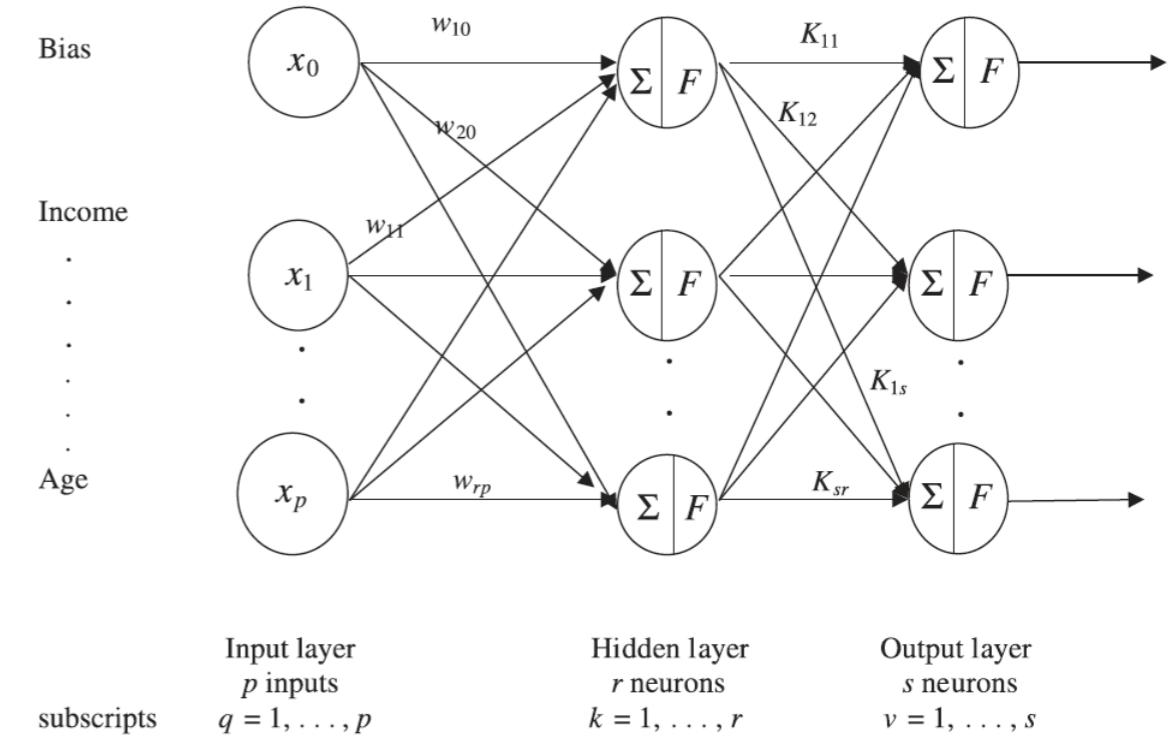


Figure 4.5. A multilayer perceptron.

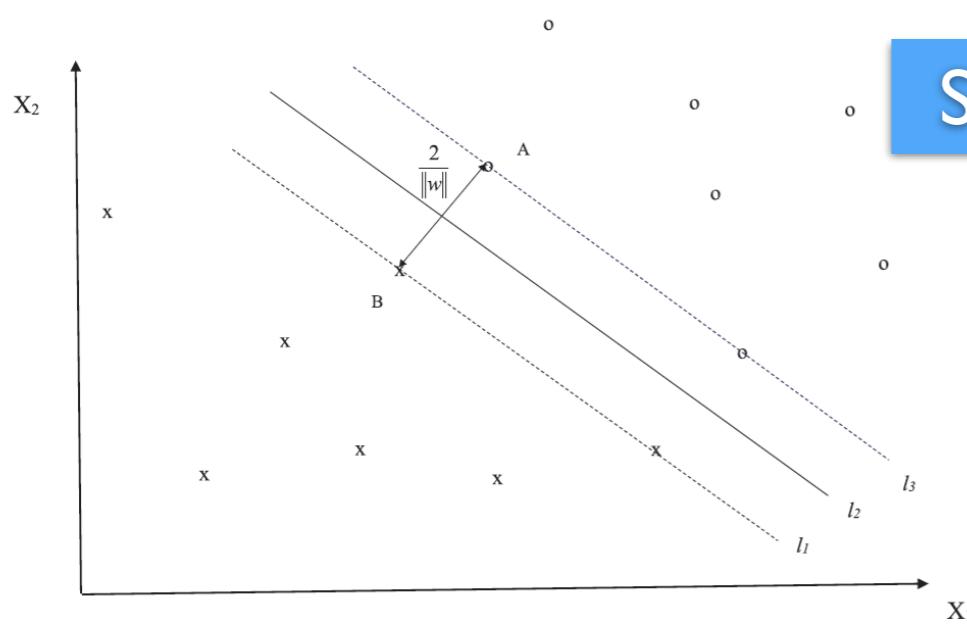


Figure 4.10. Support vectors for separable classes.

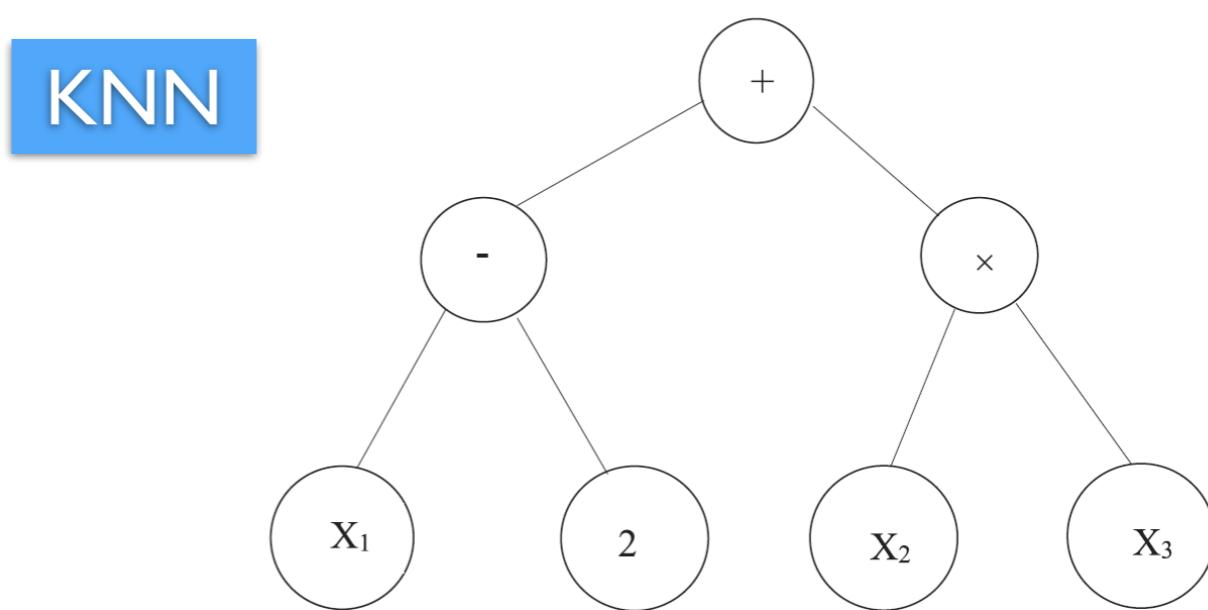


Figure 4.13. An example of a genetic program tree.

Table 6.1. Characteristics in three application forms.

Characteristic	Finance house	U.S. credit card	U.K. credit card
Zip code/postal code	X	X	X
Time at address	X	X	X
Residential status	X	X	X
Occupation	X	X	X
Time at employment	X	X	X
Appl. monthly salary	X	X	X
Other income	X	X	
No. dependents	X	X	
No children	X	X	X
Checking account/current account	X	X	X
Savings account	X	X	
Credit card	X	X	X
Store card	X	X	X
Date of birth			X
Telephone		X	X
Monthly payments	X		
Total assets	X	缺失	核实
Age of car	X		

歧视：种族、肤色、宗教、血统、性别、婚姻状态、年龄

CREDIT REFERENCE CENTER,  
THE PEOPLE'S BANK OF CHINA

# 企业信用报告 (自主查询版)

## 评分

名称：报告样本公司

机构信用代码：G11110108116779\*\*\*

中征码：11010800000000\*\*\*

报告日期：2016-08-12

NO.B201608120010323514

公开  
信息查询  
信息贷款  
信息违约  
信息账户  
信息汇总  
信息司法  
信息其余  
信息

### 信贷记录

这部分包含您的信用卡、贷款和其他信贷记录。金额类数据均以人民币计算，精确到元。

#### 信息概要 逾期记录可能影响对您的信用评价。

	资产处置信息	保证人代偿信息
笔数	1	2

	信用卡	住房贷款	其他贷款
账户数	7	3	4
未结清/未销户账户数	4	2	3
发生过逾期的账户数	4	1	1
发生过90天以上逾期的账户数	4	0	0
为他人担保笔数	0	0	1

#### 资产处置信息

- 2010年11月8日东方资产管理公司接收债权，金额400,000。最近一次还款日期为2011年1月8日，余额20,000。

#### 保证人代偿信息

- 2008年10月5日富登融资租赁担保公司进行最近一次代偿，累计代偿金额400,000。最近一次还款日期为2011年1月8日，余额20,000。
- 2009年6月21日平安保险公司进行最近一次代偿，累计代偿金额200,000。最近一次还款日期为2011年4月5日，余额135,000。

#### 信用卡

发生过逾期的贷记卡账户明细如下：

- 2004年8月30日中国工商银行北京分行发放的贷记卡（人民币账户）。截至2010年10月，信用额度10,000，已使用额度500，逾期金额500。最近5年内有11个月处于逾期状态，其中5个月逾期超过90天。
- 2003年4月1日中国民生银行信用卡中心发放的贷记卡（人民币账户），2009年12月销户。最近5年内有7个月处于逾期状态，其中3个月逾期超过90天。

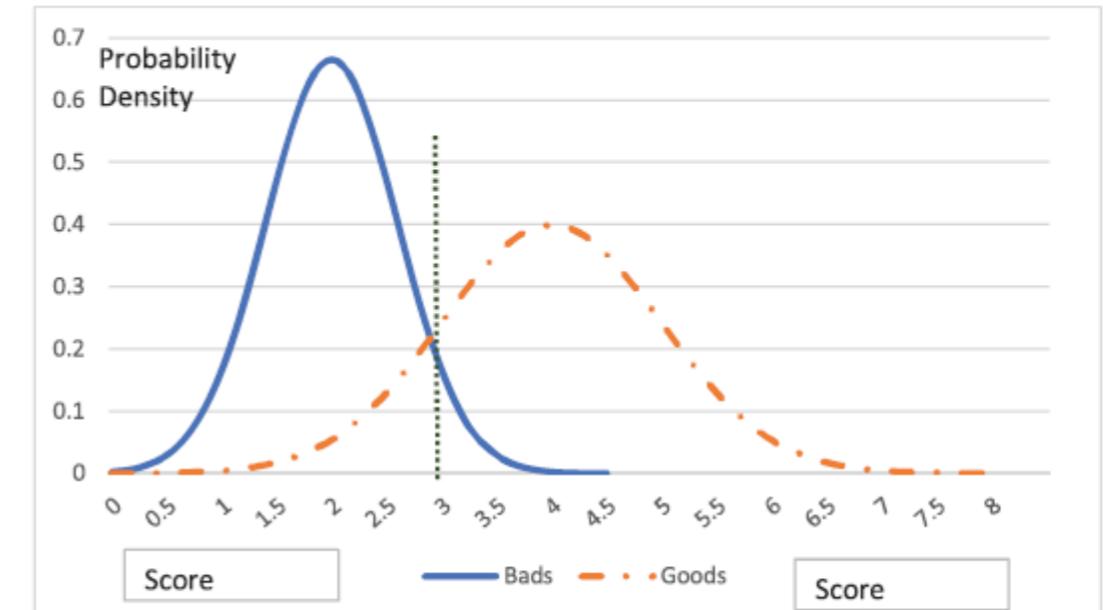
2010年3月，该机构声明：该客户委托XX房地产开发公司偿还货款，因开发公司不按时还款导致出现多次逾期。

透支超过60天的准贷记卡账户明细如下：

- 2007年6月30日中国银行北京分行发放的准贷记卡（人民币账户）。截至2010年10月，信用额度10,000，透支余额5,000。最近5年内有6个月透支超过60天，其中3个月透支超过90天。
- 2006年3月10日上海浦东发展银行北京分行发放的准贷记卡（人民币账户），2009年12月销户。最近5年内有20个月透支超过60天，其中16个月透支超过90天。



(a)



(b)

**Figure 8.1.** (a) *Means of Goods and Bads are apart.* (b) *Means of Goods and Bads are close.*

## 拒绝推断

Good

Bad

被拒绝

核准

三组

增补

外推

坏客户

增加样本

改变策略

新产品

好客户

## 其余应用

预先审核

预先批准

防范欺诈

住房贷款

小企业

风险定价

交易授权

债务偿还

坏账

出口担保

直销

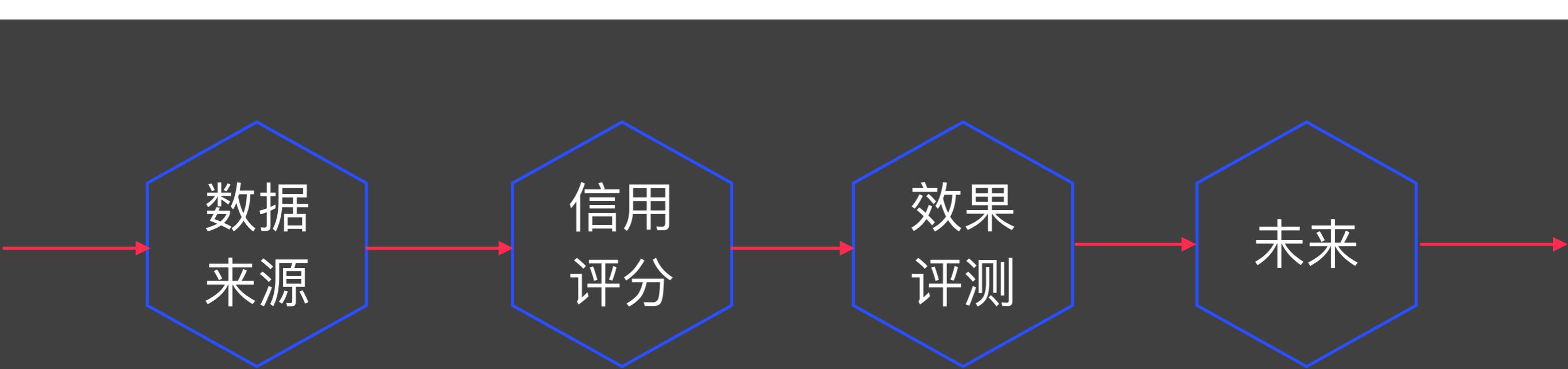
利润评分

税务检查

罚款

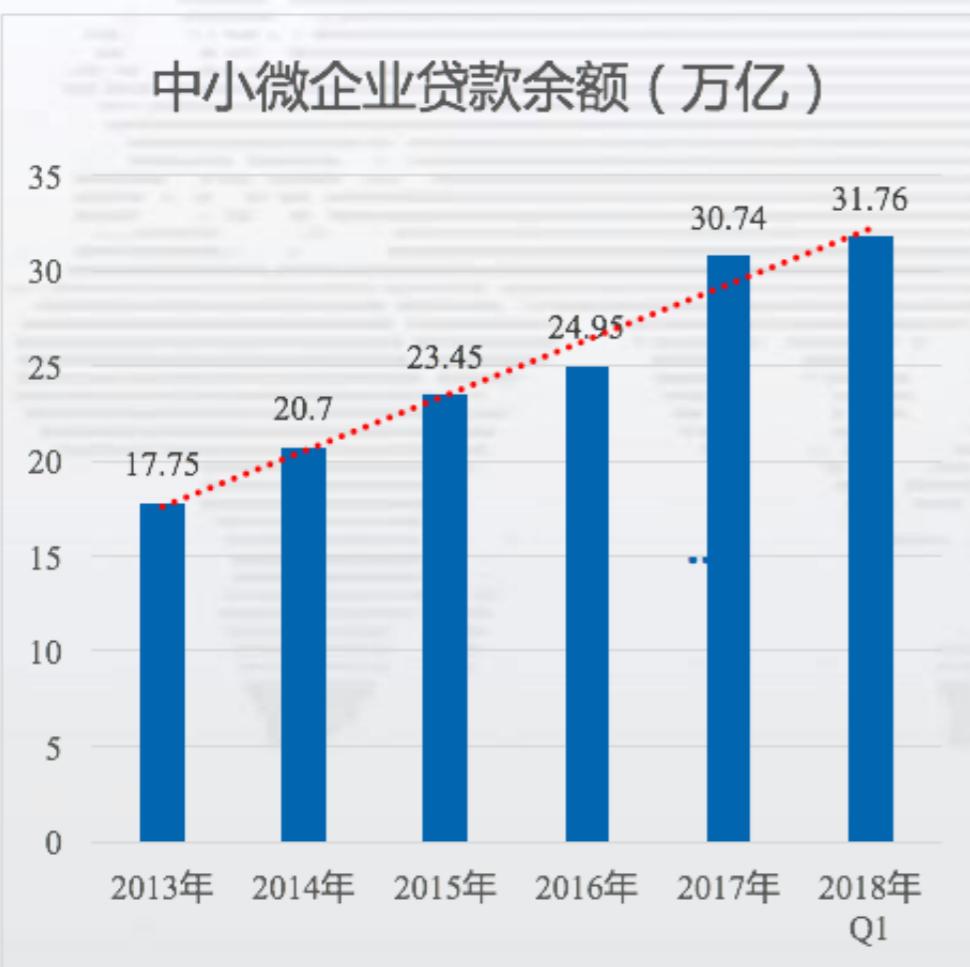
假释

# 中小微企业 信用评分



## 中小微企业

### 中国中小微企业贷款规模快速增长

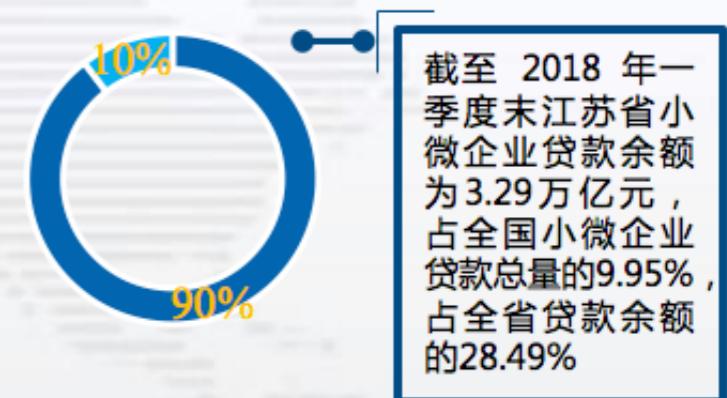


### 中小微企业贷款需求高



### 中小微企业风控服务市场潜力巨大

### 江苏中小微企业贷款余额占比高



### 国家频繁发布政策，加大对中小微企业的扶持和支持力度

#### 2018年11月9日主持召开国务院常务会议

- 国务院总理李克强要求加大金融支持缓解民营企业特别是小微企业融资难融资贵。
- 从大型企业授信规模中拿出一部分，用于增加小微企业贷款。

#### 2018年6月25日，五部委联合发布银发〔2018〕162号)

- 加大金融科技等产品服务创新。银行业金融机构要加强对互联网、大数据、云计算等信息技术的运用，改造信贷流程和信用评价模型，降低运营管理成本，提高贷款发放效率和服务便利度。

- 企业信息类8个特征变量**

企业经营年限、企业注册资本、企业所在区县、一般纳税人资质、企业类型、所属行业、法人持股比例、商变更情况

- 实际控制人6个特征变量**

法人年龄、婚姻状况、子女情况、户籍种类、住房情况、申请人本行业从业年限

- 经营发票数据14个特征变量**

销项发票计算的销售额、主要销售地区省内、主要下游客户经营年限、主要下游客户企业类型、主要下游客户行业、红冲发票比例、无效发票比例、专票占比、近24个月月波动率、近24个月季度波动率、近24个月交易方一致性、近24个月集中度、销售额全国企业中排名、销售额行业排名

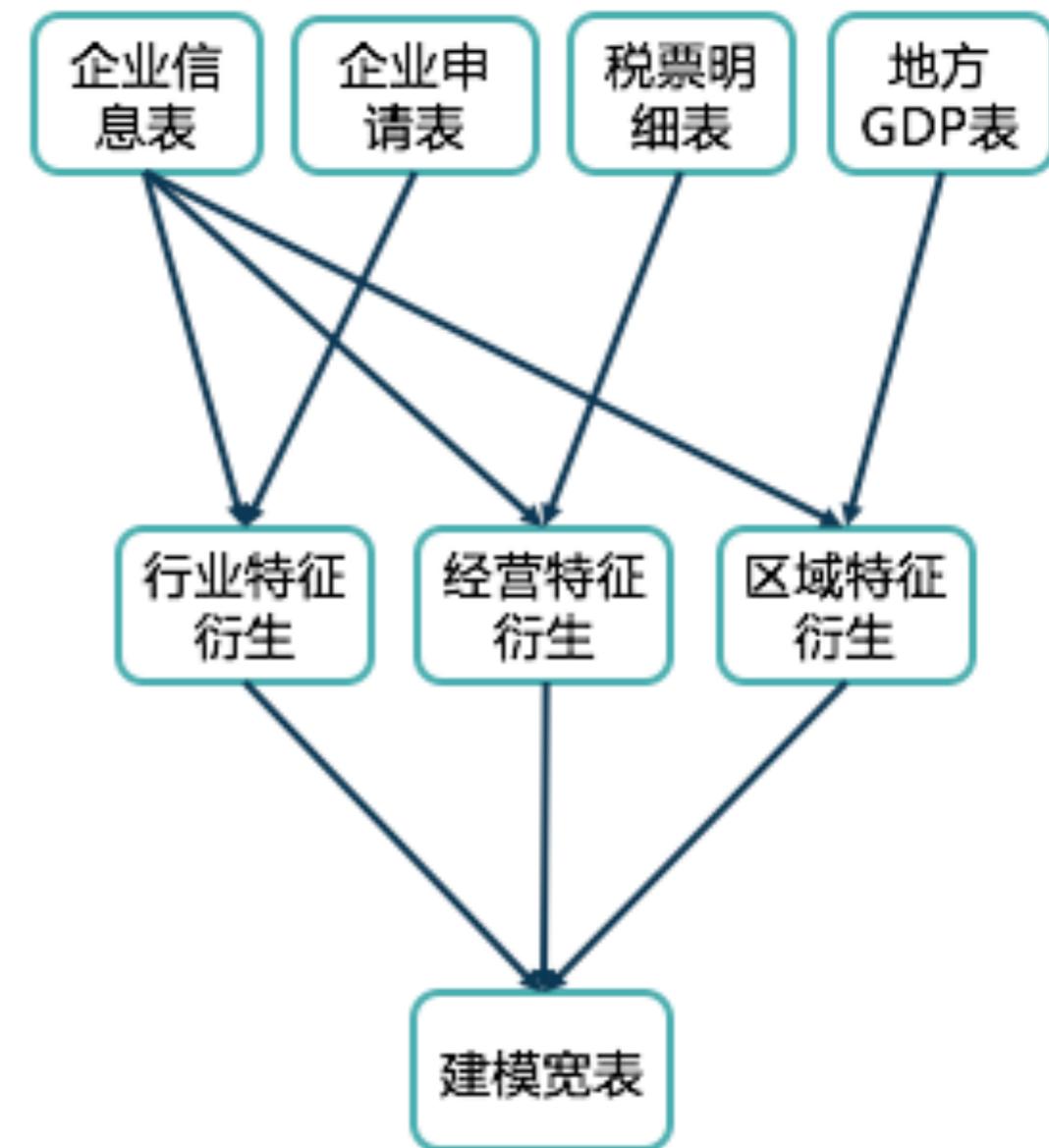
- 企业风险信息类2个特征变量**

企业有不良声誉记录的（如上过有关机构或部门的黑名单），

企业实际控制人有赌博、吸毒等不良嗜好的

- 综合评价类2个特征变量**

企业一致性指数(企业销售客户稳定性(下游))、企业授信倾向分。



# Credit Scoring

# 信用评分

行业  
规模  
注册年份  
股东及出资信息

地区  
企业性质  
法定代表人  
.....

诉讼信息  
失信被执行人  
行政处罚  
环保处罚  
税收违法  
统计失信  
环保失信  
食品药品抽检不合格  
环保违法  
税票机开机情况  
.....

行业排名  
规模排名  
经营年限排名  
地区排名  
增长率排名  
.....

基于计量经济分析，通过大数据以及机器学习算法，计算中小微企业信用评分



行业增长率  
地区GDP  
地区经济增长率  
地区PPI

行业利润率  
地区人口  
地区CPI  
.....

年税票总额  
主营商品  
平均月税票额  
年度同比税票额增长  
季度数票额方差  
最大季度税票额  
最小季度税票额  
最大月度税票额  
最小月度税票额  
有税票月份数  
无税票月份数  
专用增值税票金额  
普票金额  
红冲税票金额  
无效税票金额  
下游企业家数  
.....

- **模型超参数：**

学习速率：0.05

决策树最大深度：3

gbdt中决策树的棵数：76

- **模型效果：**

准确率：0.664

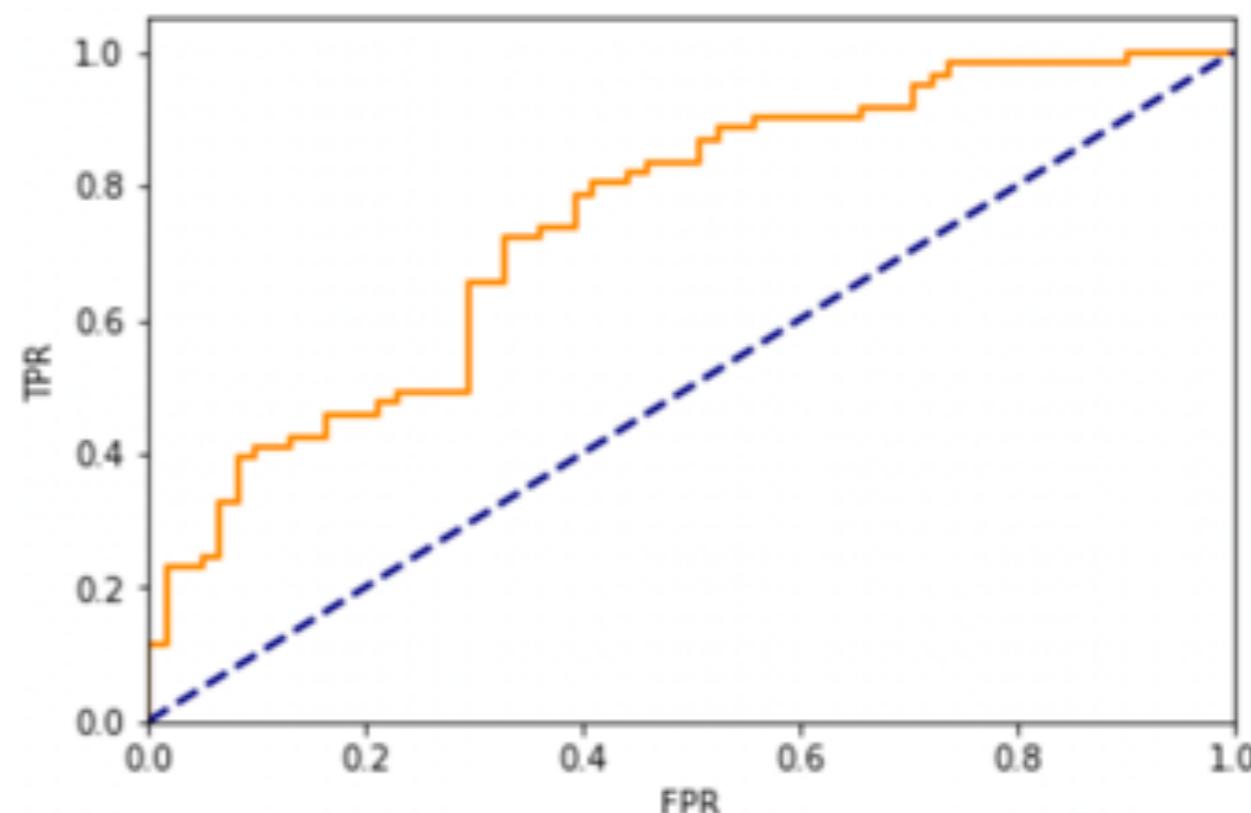
AUC：0.747

正样本精确率：0.61,

正样本召回率：0.90,

负样本精确率：0.81,

负样本召回率：0.43

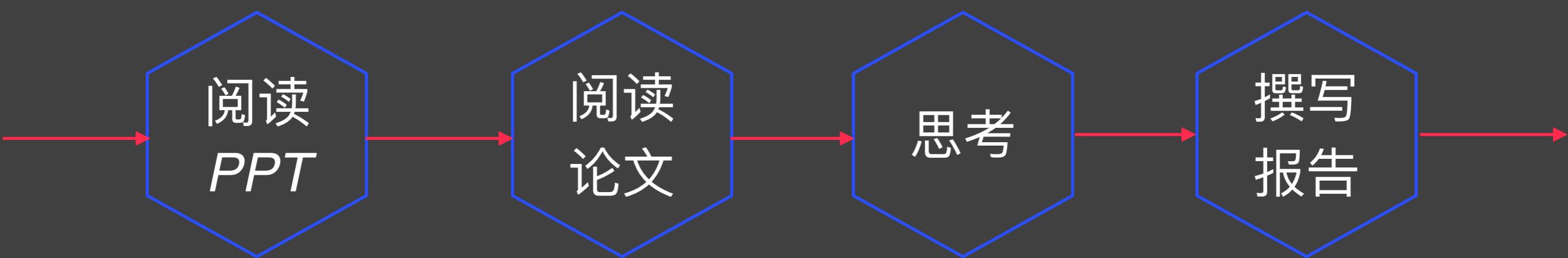


# Credit Scoring

未来

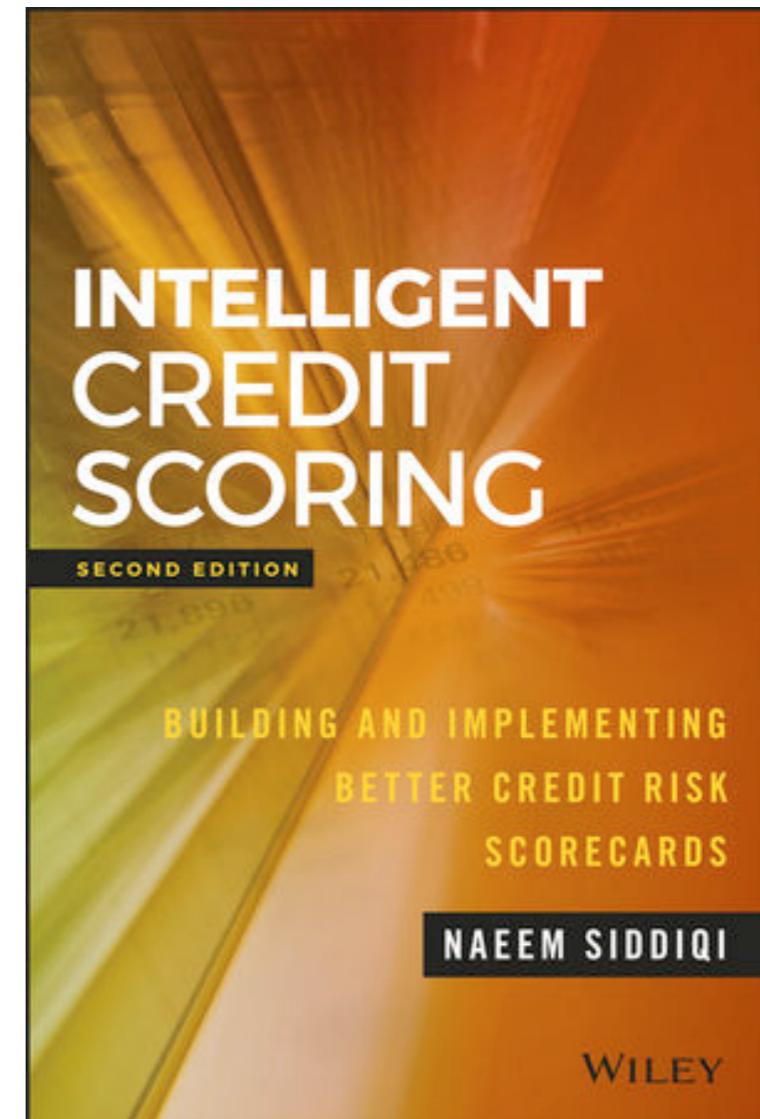
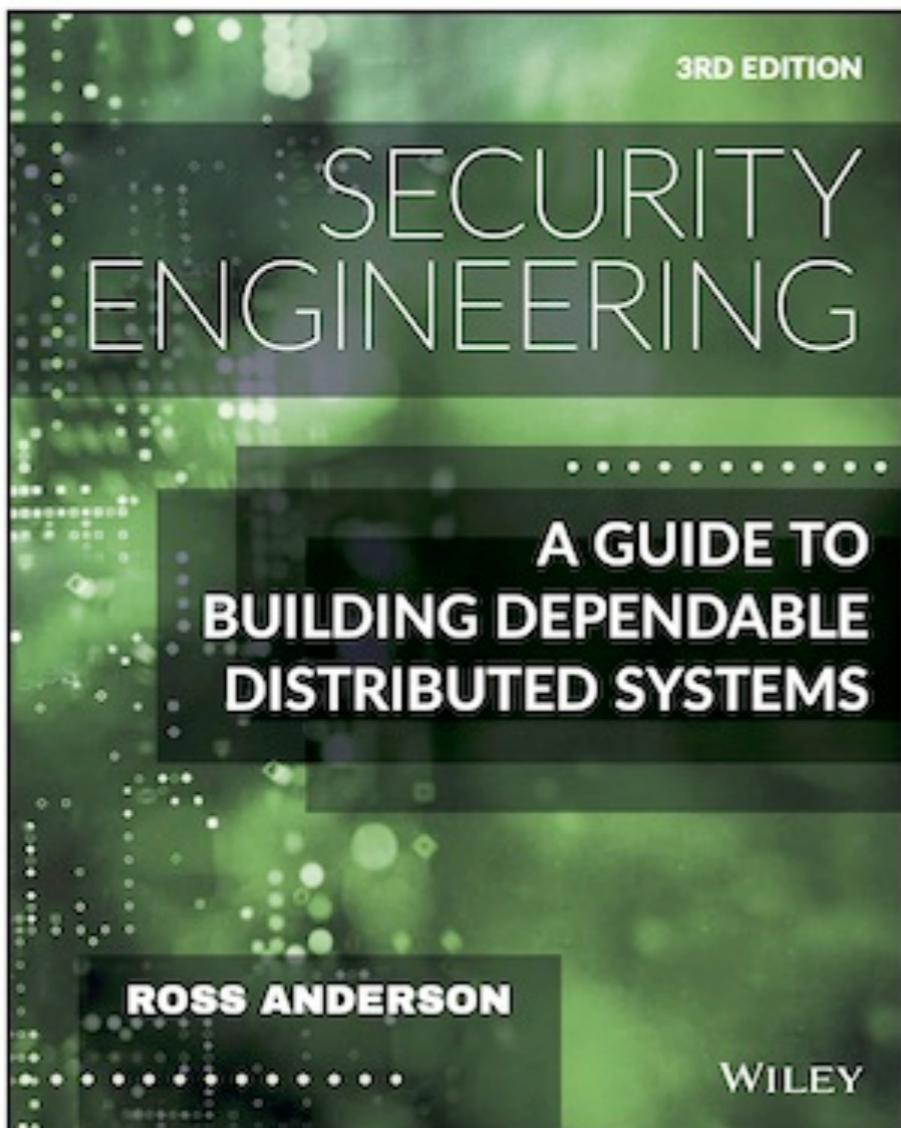


# 课后作业



# Homework

## 阅读教材



第6、13、16、18章

第1-4、6、8章

要求阅读如下文章，写阅读报告

自己选择和课程项目相关的  
一篇论文

说明课程项目涉及哪些论文  
这篇论文的作用

- 1、文章概述
- 2、主要收获
- 3、存在疑问
- 4、所思所感
- 5、一篇论文

周日晚上12点  
前提交

謝謝 !

*Huijing Sun*

[sunhp@ss.pku.edu.cn](mailto:sunhp@ss.pku.edu.cn)

<https://huijingsun.github.io>