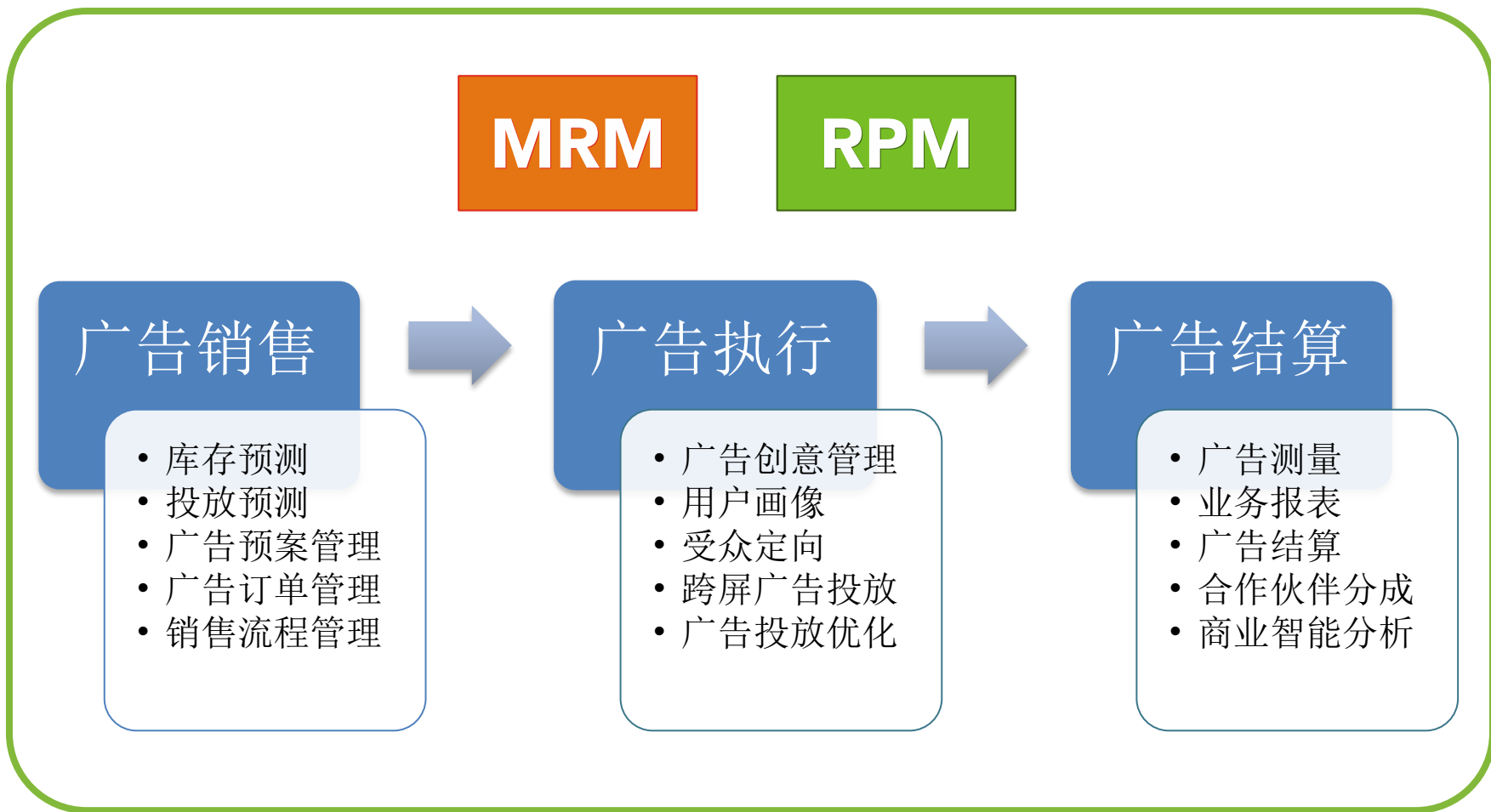




FreeWheel基于大数据的新兴 视频广告测量实践


李旻

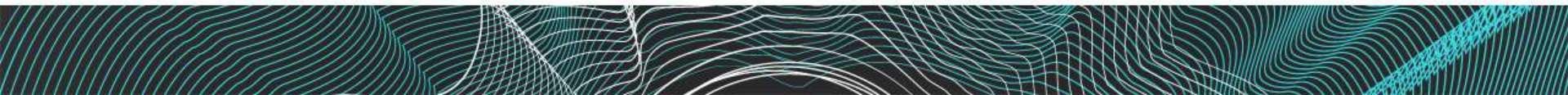


单日投放近10亿次广告，生成2TB广告投放数据

 新兴视频广告测量方式

 用户性别年龄预测

 广告完成率预测



Half the money I spend on advertising is wasted, the trouble is I don't know which half.



John Wanamaker

US TV* vs. Digital Video** Ad Spending, 2012-2018

billions and % change

	2012	2013	2014	2015	2016	2017	2018
TV*	\$64.54	\$66.35	\$68.54	\$70.59	\$73.77	\$75.98	\$78.64
—% change	5.4%	2.8%	3.3%	3.0%	4.5%	3.0%	3.5%
Digital video**	\$2.89	\$4.20	\$5.96	\$7.77	\$9.45	\$11.12	\$12.71
—% change	44.5%	45.3%	41.9%	30.4%	21.7%	17.6%	14.3%

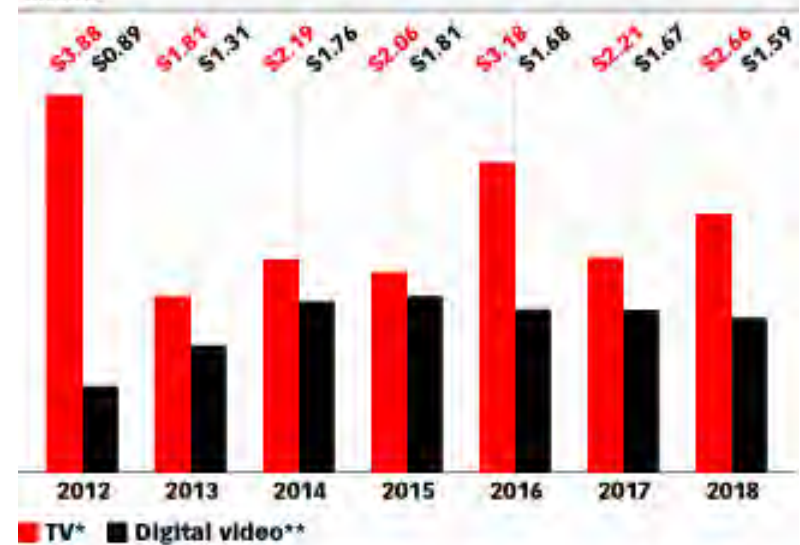
Note: *includes broadcast TV (network, syndication & spot) & cable TV;
 **data through 2013 is derived from IAB/PwC data; includes advertising that appears on desktop and laptop computers as well as mobile phones and tablets; includes in-banner, in-stream (such as pre-roll and overlays) and in-text (ads delivered when users mouse over relevant words)
 Source: eMarketer, June 2014

174128

www.eMarketer.com

US TV* vs. Digital Video** Ad Spending Annual Increases, 2012-2018

billions



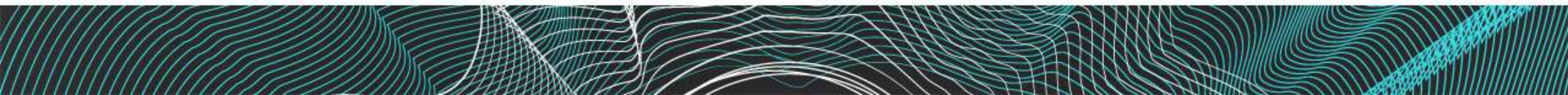
Note: *includes broadcast TV (network, syndication & spot) & cable TV;
 **data through 2013 is derived from IAB/PwC data; includes advertising that appears on desktop and laptop computers as well as mobile phones and tablets; includes in-banner, in-stream (such as pre-roll and overlays) and in-text (ads delivered when users mouse over relevant words)
 Source: eMarketer, June 2014

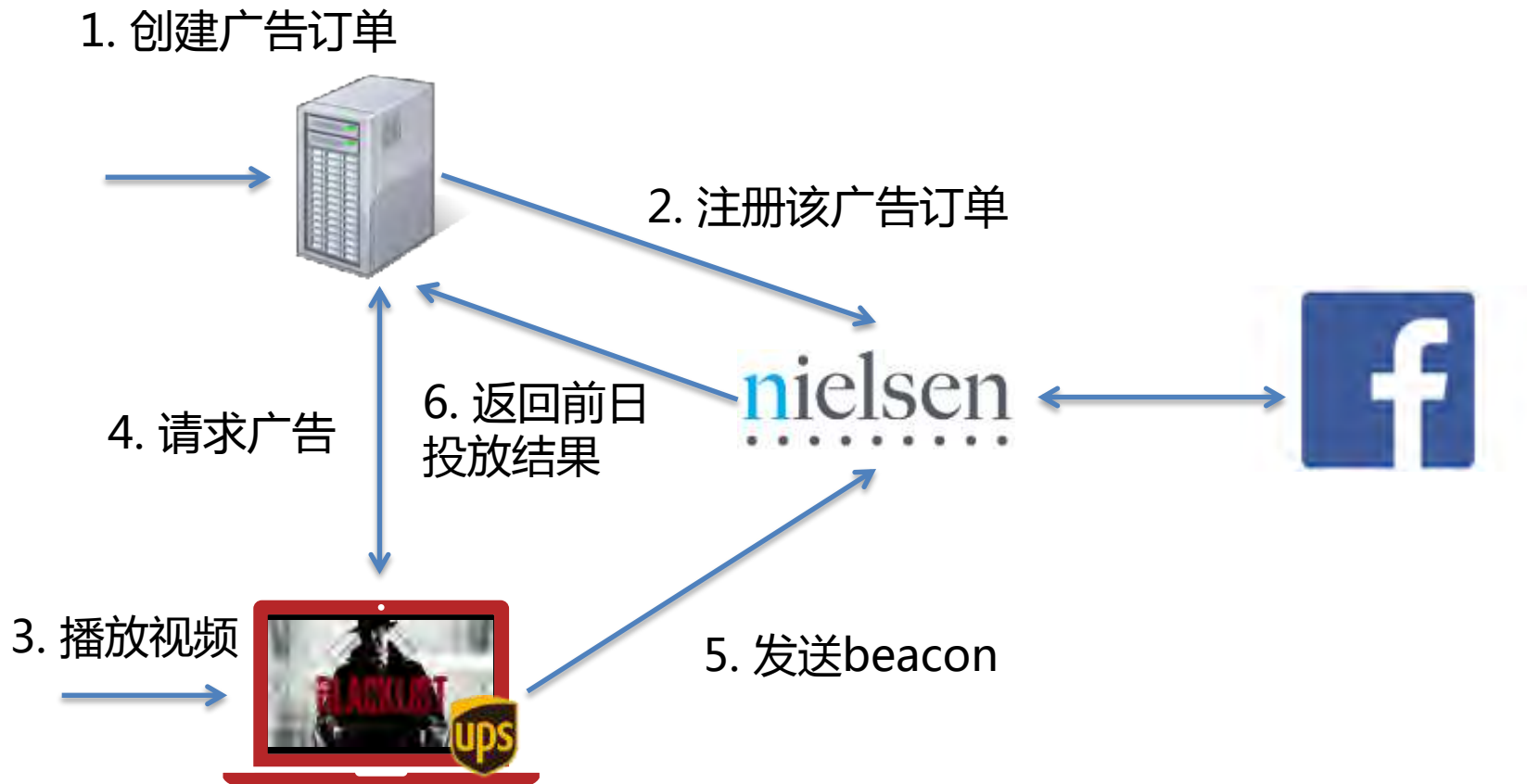
174128

www.eMarketer.com

引自: <http://www.emarketer.com/Article/US-TV-Ad-Market-Still-Growing-More-than-Digital-Video/1010923>

- 数字收视率 (Digital Rating)
 - 在特定性别年龄用户群体 (demographic) 上的广告曝光
- 广告完成 (Ad Complete)
 - 播放器发送IAB 4th Quartile响应的广告曝光
- 可见曝光 (Viewable Impression)
 - 视频广告50%的面积被持续可见不少于2秒的广告曝光





挑战：投放时预测用户年龄性别

观看历史



静态特征

Mountain View
CA, US

12:40 pm

UEFA Champions

Ubuntu 14.04

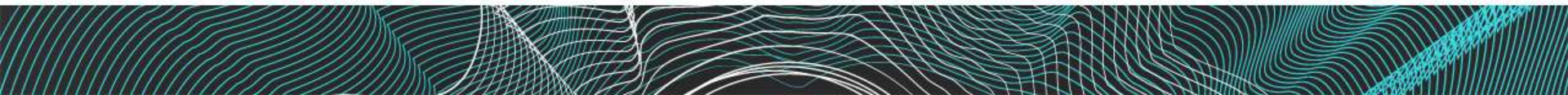
Chrome

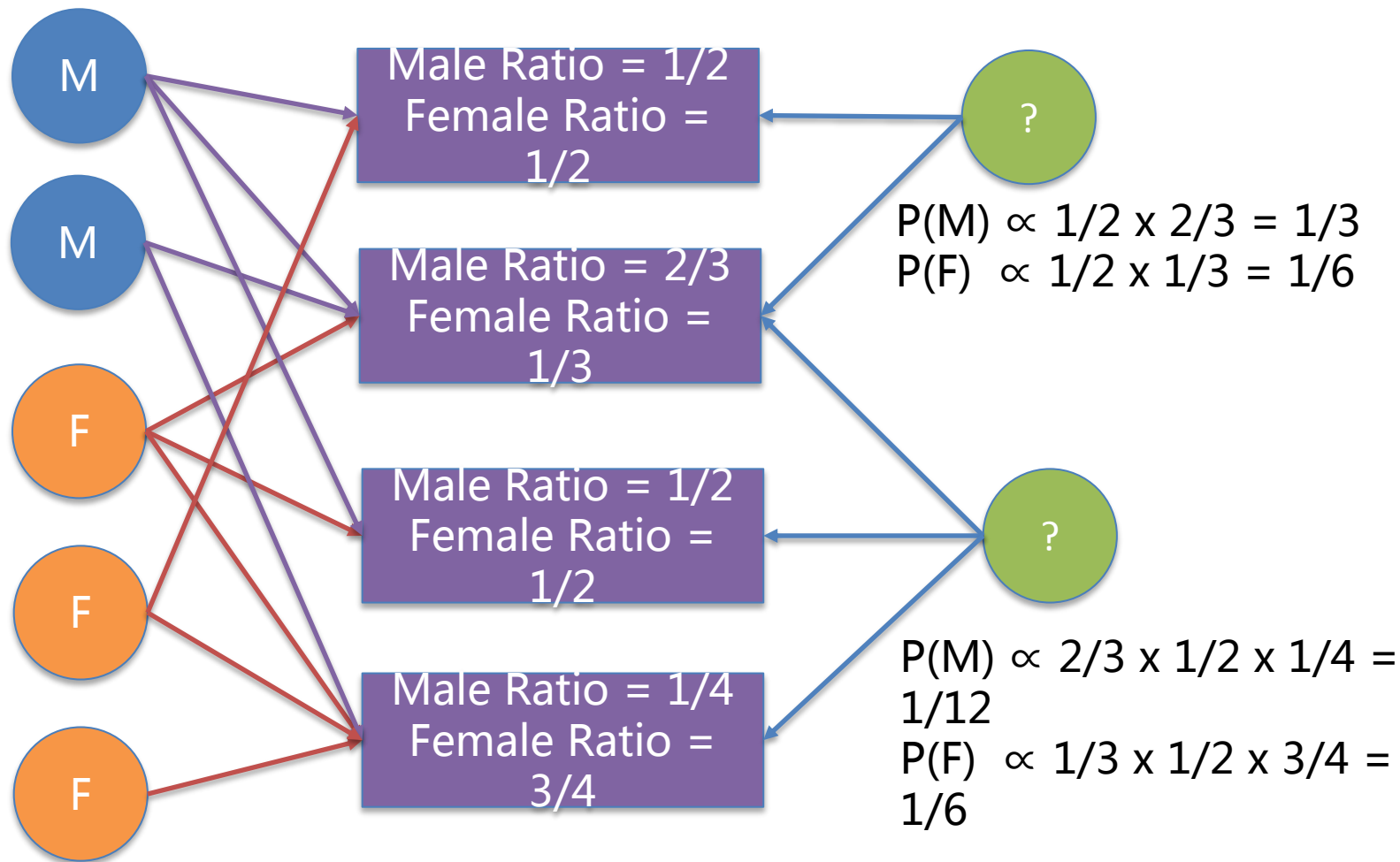


⇒ OR?



- 思路：有监督分类问题
- 挑战：缺少有效的样本标注手段（BlueKai标注数据的准确率只有60%）



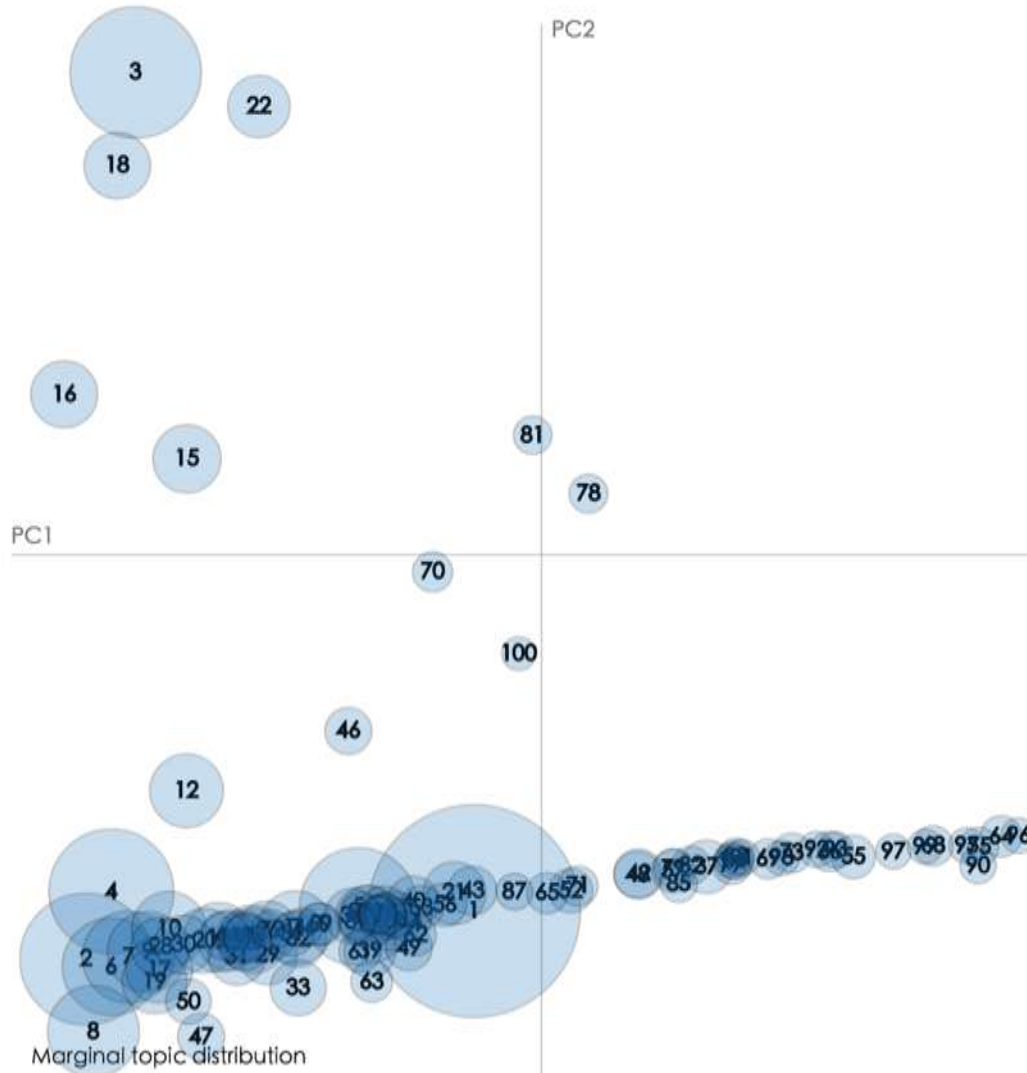


Top Male Page	Top Female Page
ESPN Soccer Game	ABC Family Drama Page
ESPN Front Page	Fox Bones Show
ESPN NBA Game	ABC Health News
ESPN NFL Game	ABC Lifestyle News
MSN Sports Video	ABC Entertainment News
Fox Family Guys Show	Fox Masterchef
Fox Gotham Show	Fox Empire Show
ABC Technology News	AOL Entertainment News
ABC Politics News	AOL Huffington Post

- 选取2000个流量较大的页面，获取Nielsen性别分布反馈
- 选取600W在这些页面上浏览较多的用户，并预测他们的性别
- 从中选取100W最可能是男性的用户(准确率90%)和100W最可能是女性的用户(准确率82%)作为训练样本

- 基于名称和简短描述，提取Genre + Topic
- Genre
 - Animation, Business, Comedy, Entertainment, News, Sports...
 - 对每个Genre训练一个LR模型
 - 特征为2-gram词表在正负样本中出现的次数的比值
- Topic
 - LDA提取了100个Topic

Intertopic Distance Map (via multidimensional scaling)



Topic #11

➤ NFL, NHL, Star...

Topic #31

➤ Vegas, Show, Hot...

Topic #32

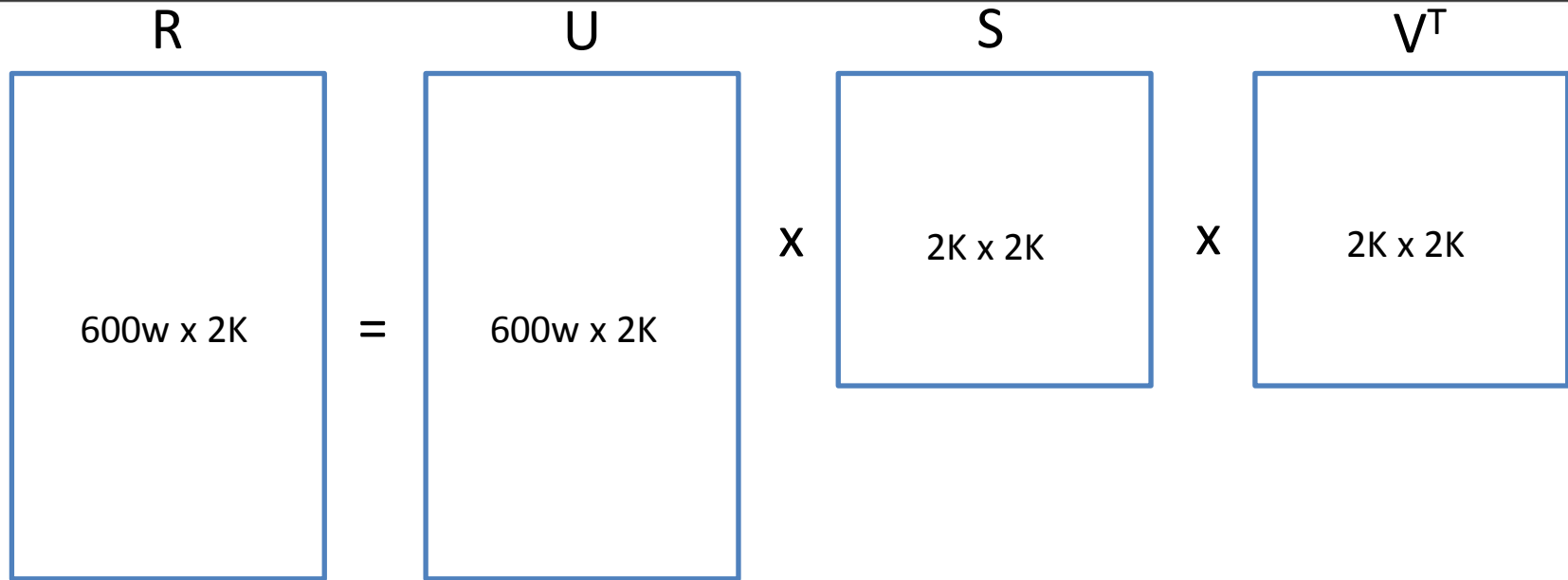
➤ Technology, Tennis, Jets...

Topic #33

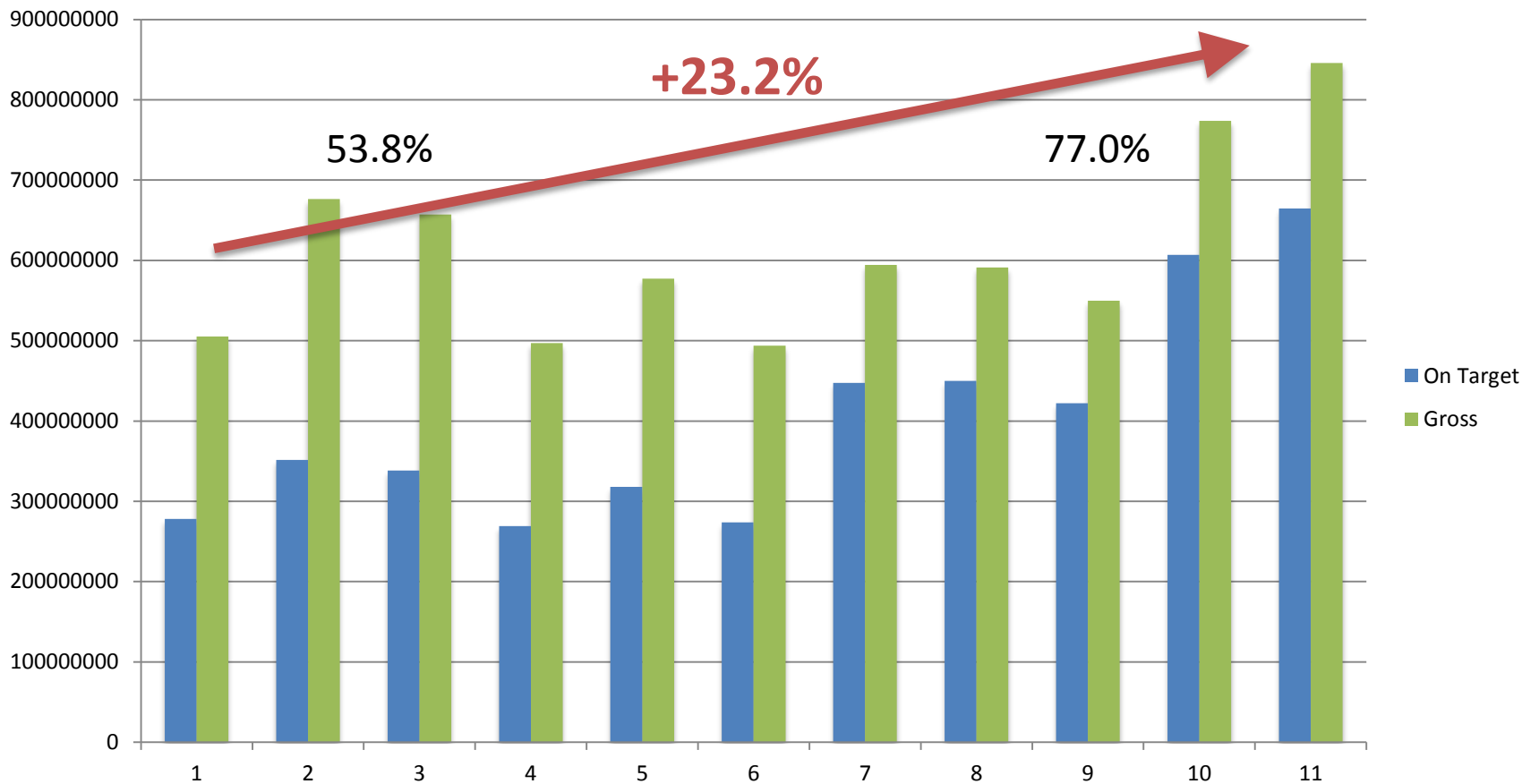
➤ Family, Weddings,
Queens...

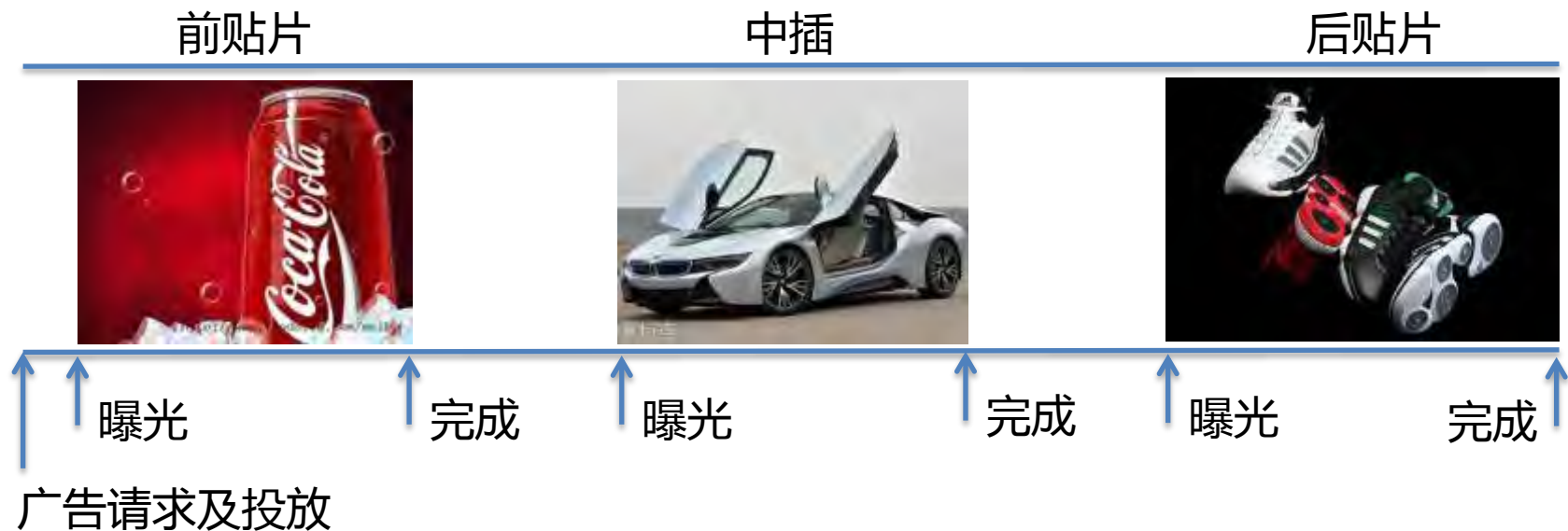
Topic #35

➤ Diner, Cafe, Bowling...



- SVD => 选取前200个特征值
- 根据新用户在2K个网页上的浏览记录映射到200维的特征特征空间
- 基于余弦距离选取T个相似用户
- $P_{\text{smooth}}(C|u) = \alpha \times P(C|u) + (1 - \alpha) / T \times \sum_{i=1 \dots T} P(C|NBi(u))$



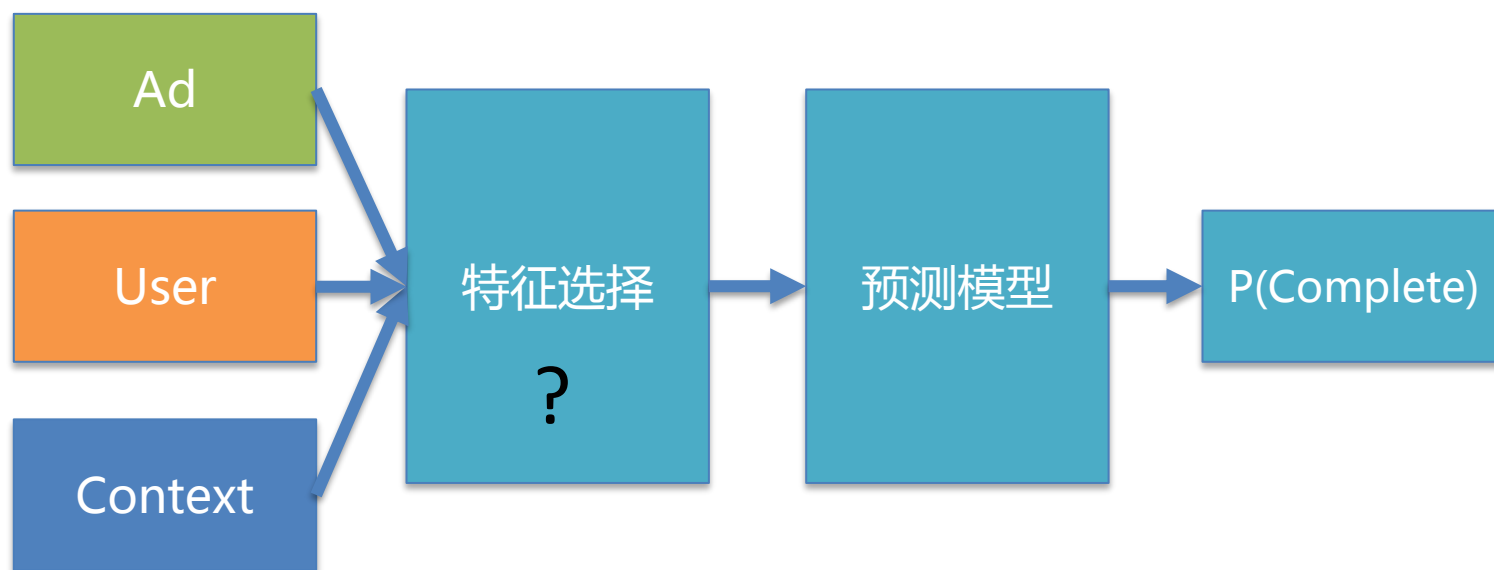


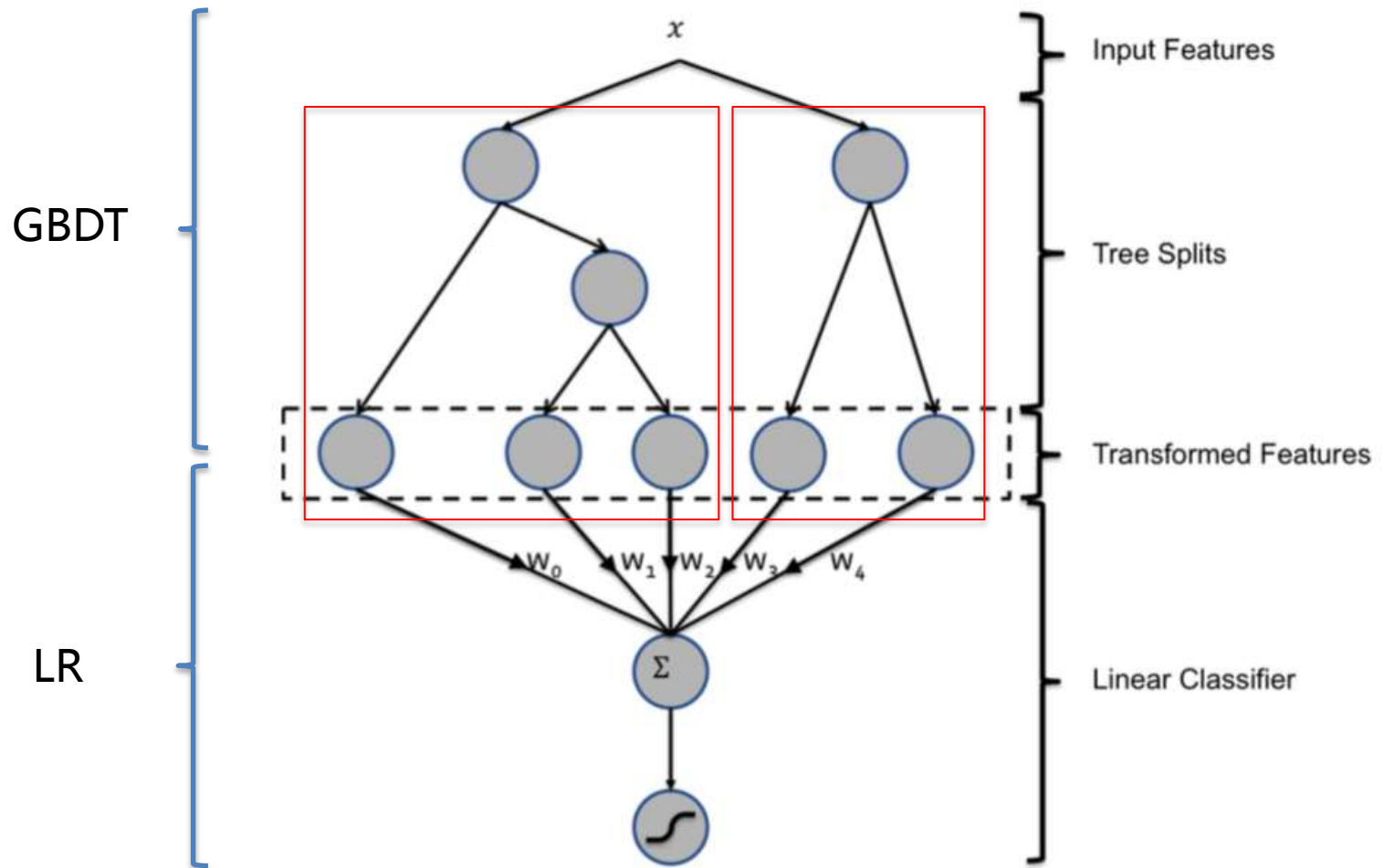
- 广告完成率 = (广告完成次数 / 广告被投放次数) x 100%

整体完成率	前贴片	31%
	中插	22%
	后贴片	4%

- eCPM = 广告完成价格 x 广告完成率 P(complete | ad, user, context)

- Ad: 广告、广告类型、品牌、广告主、广告代理
- User: 用户、地理位置、设备
- Context: 视频、页面、广告位、广告顺序
- 历史特征 + 静态特征 (500+)





From: "Practical Lessons from Predicting Clicks on Ads at Facebook" (ADKDD'14)

Ad Type	LR AUC	GBDT+LR AUC
前贴片	0.879	0.904
中插	0.826	0.843
后贴片	0.912	0.986

