

深度学习开启社交电商 智能时代

杨骥
国美大数据中心副总监





目录

01

社交电商个性化场景

02

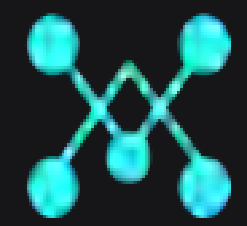
国美互联网机器学习实践

03

国美互联网深度学习的应用场景

04

未来的挑战



目录

01

社交电商个性化场景

02

国美互联网机器学习实践

03

国美互联网深度学习的应用场景

04

未来的挑战



国美 GOME PLUS.COM

418
伟霆ForU #
国美 Plus 烧屏趴 4.14-4.19

18亿优惠券免费送 抢418元大红包 抢10000个免单

排序问题→机器学习问题

搜索WOT 2017

首页 | WOT2017 – 全球创新技术峰会 - 51CTO.COM

WOT

WOT, World Of Tech 专注互联网IT技术、技术实战、技术前瞻的密集式干货集中营
wot.51cto.com/ - 百度快照

Perfect

WOT 2017全球架构与运维技术峰会(World Of Tech)

6天前 - WOT 2017全球架构与运维技术峰会
会议时间: 2017-04-14 08:00:00至 2017-04-14 16:00:00结束
地点: 北京 北京富力万丽酒店 朝阳...
https://www.huodongjia.com/eve... -

Excellent

WOT 2017嘉宾阵容 提前揭晓!

【导读】活动家与51CTO合作,今天活动家为大家提前揭晓WOT 2017的嘉宾阵容! WOT, World Of Tech 专注互联网IT技术领域 三大... 五大技术专场...
mt.sohu.com/it/d201702... - 百度快照

Good

WOT2016大数据技术峰会

查看WOT2017往届峰会 WOT2017...
10/26 北京 WOT2013大数据全球...
WOT2013云计算架构师峰会2013年10月26日北...
wot.51cto.com/2016bigd... - 百度快照

Fair

坦克世界2017_坦克世界2017官网_攻略_坦克世界2017礼包_安... 九游

坦克世界2017手机公测百度版是一...
一拥 [详情] 坦克世界2017手机公测...
游戏,游戏中一拥拥有超多二战的经...
www.9game.cn/tksj2017/ - 百度

Bad



Query	URL	标注	Page-Rank	BM25	...	特征_N
WOT 2017	51cto.com	P	0.6	0.9		100
WOT 2017	...	E	0.9	0.9		66
WOT 2017	...	G	0.8	0.7		75
WOT 2017	...	F	0.9	0.6		32
WOT 2017	...	B	0.2	0.65		120
国美	...	P	0.6	0.8		150
国美	...	P	0.5	0.7		50
...						

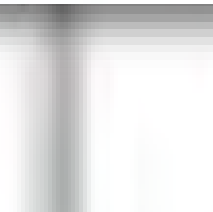
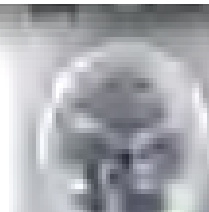
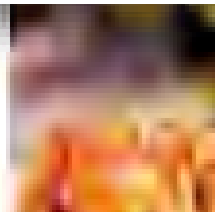
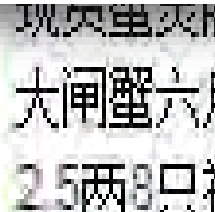
无处不在的个性化

51CTO WOTI



猜你喜欢 — 你的专属推荐 换一组

					
九阳 (Joyoung) DJ12B-A11豆浆 ¥189.00	九阳 (Joyoung) DJ12B-A11DEC ¥249.00	九牧Jomoo卫浴健康龙头厨房用水 ¥269.00	SIMEN不锈钢 手感好 不锈钢西餐 ¥28.00	联想 (Lenovo) Ideapad100 14.0 ¥2049.00	【国美自营】Lenovo/联想 天逸 ¥3598.00

				
国美在线 GOME.COM.CN	50LW(50551)FNA A3 2匹P立柜式 酷 ¥7770.0	XQG70- BS1286AM 7公斤 ¥2199.0		大闸蟹六月黄2.0- 2.5两8只礼盒装 鲜 ¥268.0



必应 WOT 2017 峰会

网页 图片 视频 学术 词典 网典 地图

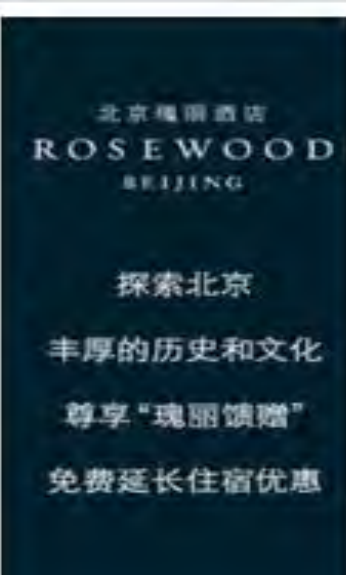
16,000 多结果 时间不限

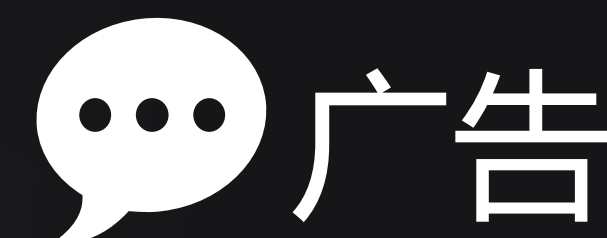
WOT 2017全球架构与运维技术峰会 (World Of Tech 2017 ...
WOT 2017全球架构与运维技术峰会 (World Of Tech 2017) 参会, 就上【活动家
www.huodongjia.com】。WOT 2017全球架构与运维技术峰会 (World Of Tech ...
<https://www.huodongjia.com/event-20462474.html> 2017-5-25

WOT 2017全球创新技术峰会 - 活动详情页 - 开源中国 ...
WOT 2017全球创新技术峰会, 北京 朝阳 北京富力万丽酒店, 2017-07-21 08:50 至 2017-
07-22 18:00, 活动专区, 开源中国社区 ... 为了获得更好的浏览体验, 我们 ...
<https://www.oschina.net/event/2243271> 2017-6-1

WOT 2017全球架构与运维技术峰会: 技术前瞻的密集 ...
WOT 2017全球架构与运维技术峰会: 技术前瞻的密集式干货集中营! - WOT, World Of
Tech专注互联网IT技术领域 三大宣节, 15大技术专场, 60+国内外一线互联 ...

People You May Know

 Ruotong (Walter) Ma Intern at ICBC ICFA level II Candidate seeking full-time...	 Jessica Min Consultant at Skanda Associates	 rhoda pang Research Associate at Hyper Growth Management...	 Wenbin Tang Software Engineer at Google	 北京瑰丽酒店 ROSEWOOD BEIJING 探索北京 丰厚的历史和文化 尊享“瑰丽馈赠” 免费延长住宿优惠
Connect	Connect	Connect	Connect	



人工规则



机器学习排序



✓Pointwise

排序问题转化为多类分类问题或者回归问题，相关度相同的为一类

✓Pairwise

排序问题转化为二分类问题，由两两之间的偏序关系得到全局的排序

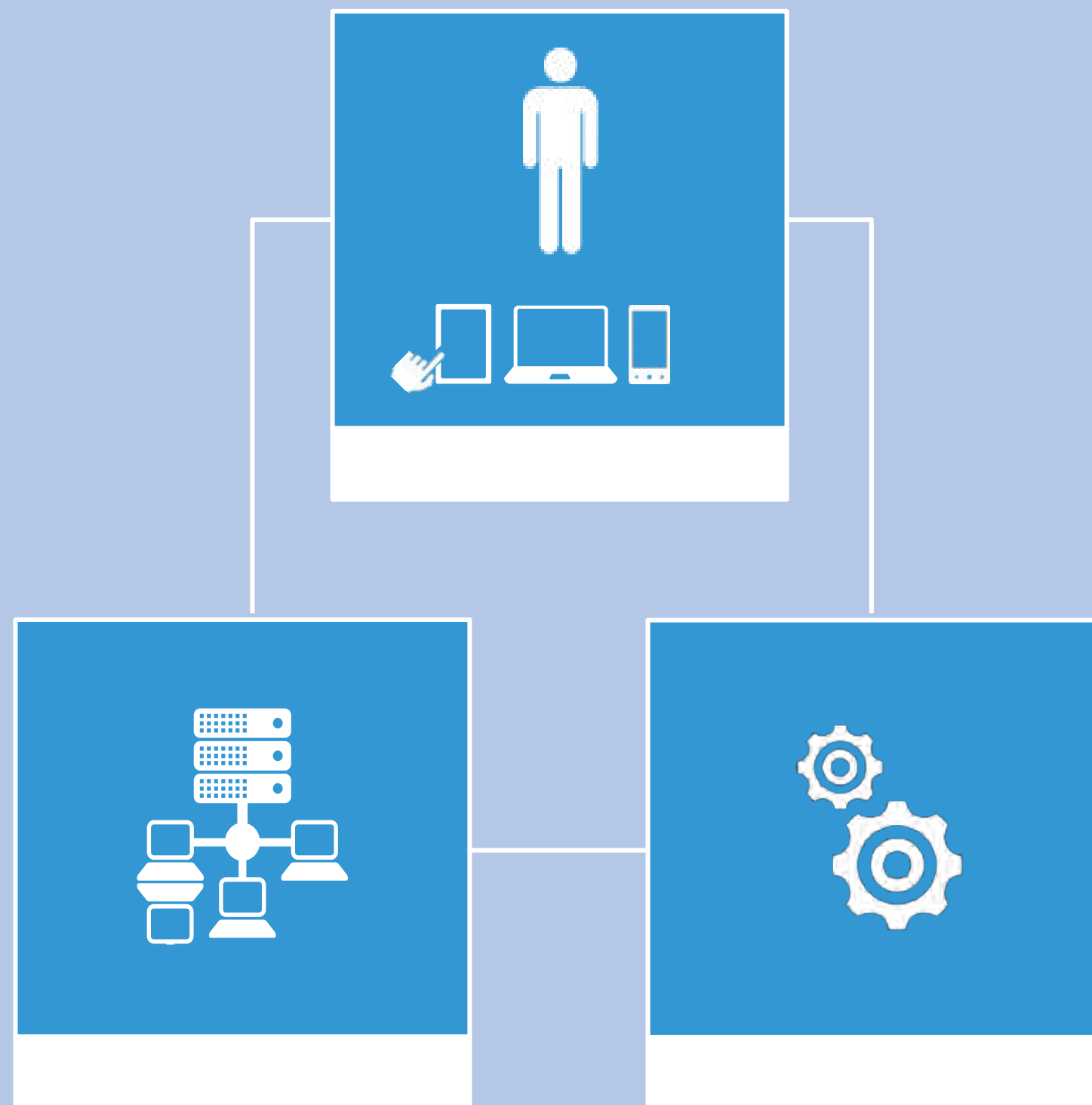
✓Listwise

直接对排序结果进行优化



人工规则 vs. 机器学习
(广告 vs. 搜索 vs. 推荐)

离线预处理



在线处理





目录

01

社交电商个性化场景

02

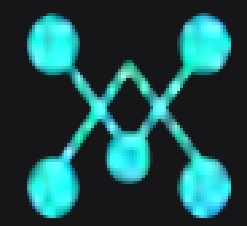
国美互联网机器学习实践

03

国美互联网深度学习的应用场景

04

未来的挑战



目录

01

社交电商个性化场景

02

国美互联网机器学习实践

03

国美互联网深度学习的应用场景

04

未来的挑战

训练数据

<Query, Doc> → 标注

A B C D E F G

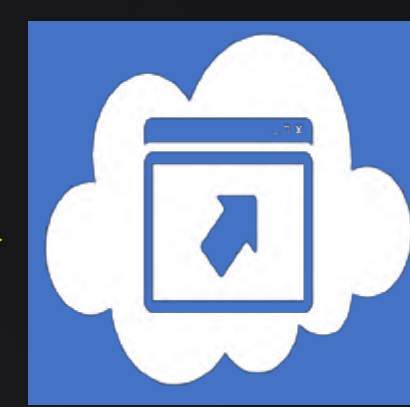
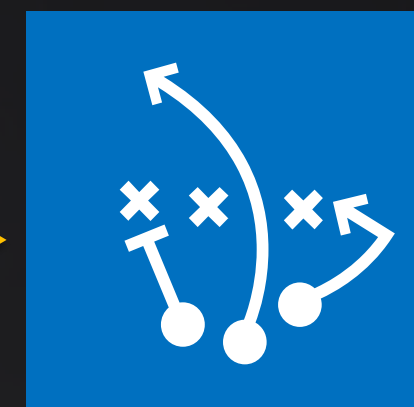
	标注		标注
B	+	B-A	+
E	+	B-C	+
A	-	B-D	+
C	-	C-E	-
D	-	D-E	-
		E-A	+

特征工程

模型训练
(& 离线测试)

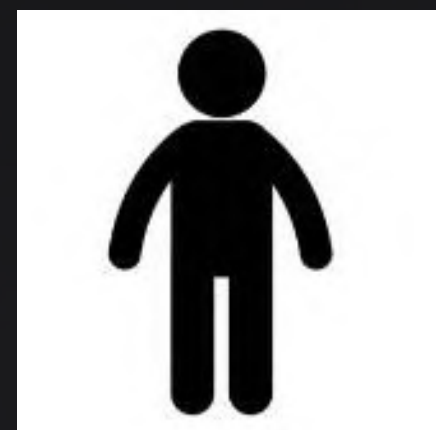
线上部署
(& A/B测试)

正式部署



特征工程-如何找特征

用户属性、
购买力、
品类倾向



用户

查询量、转
化率、品类
分



查询

UV、PV、
订单量、
转化率



商品

店铺评分、
店铺流量、
店铺评价



商家

当地温度、
污染指数、
人口数量



环境

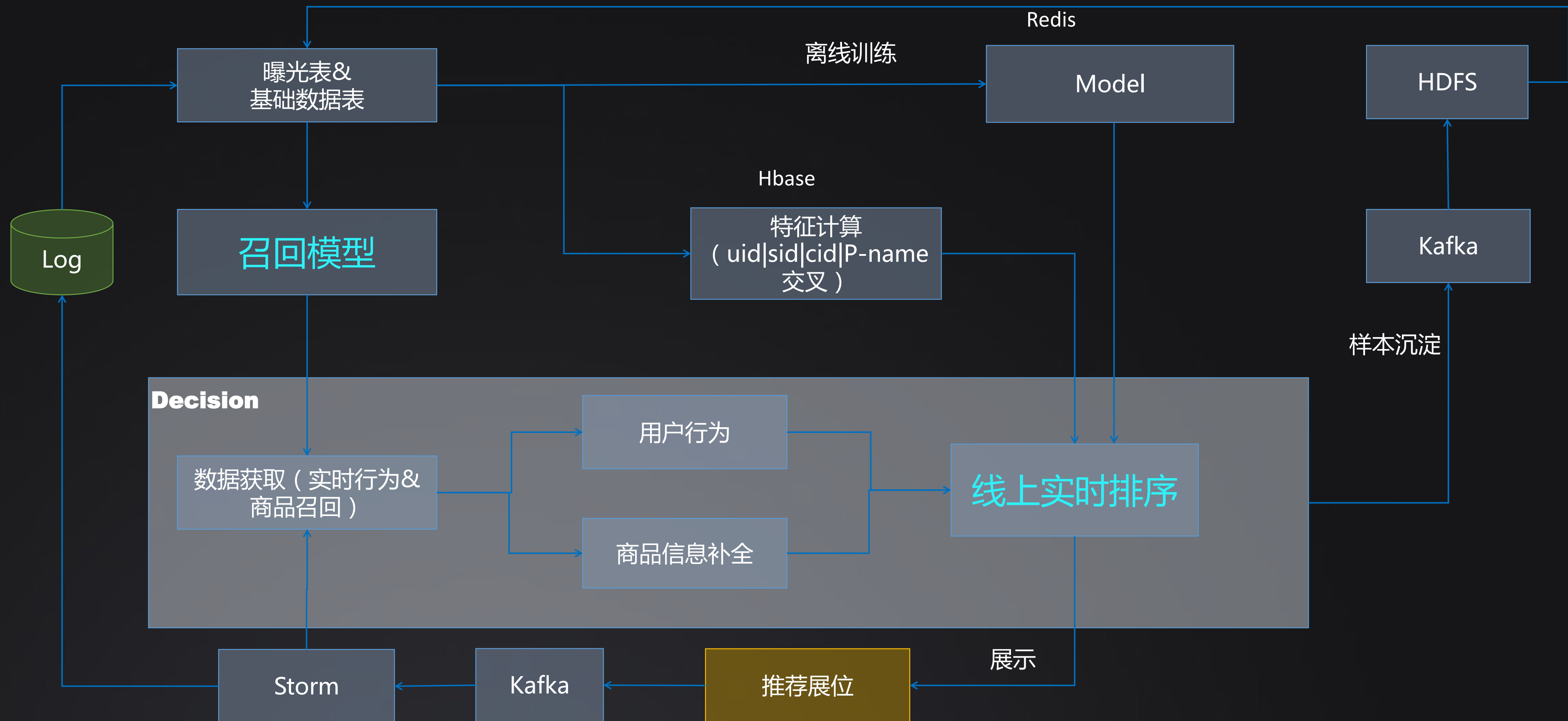
关键词匹配、价格匹配等

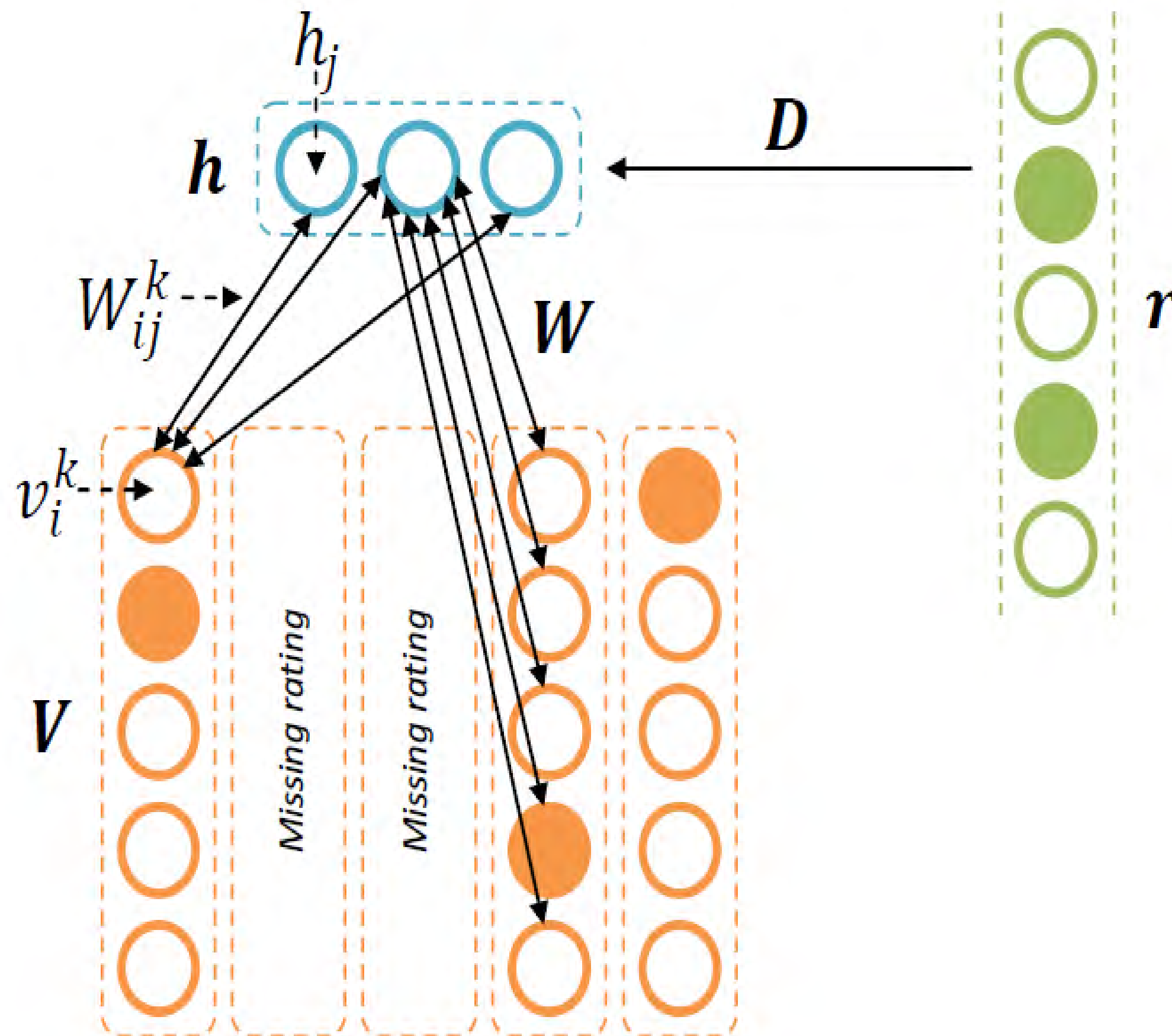
区域销量、商品天气匹配等

品类匹配、购买力匹配等

- 可量化目标：分类准确率、NDCG、点击率、转化率、GMV
- 假设→试验→结论







Algorithm 4.5 RBM - Learning algorithm

```

Initialize  $W_{ij}^k$ 's with small values sampled from a zero-mean normal distribution;
Initialize  $b_i^k$ 's to the log of their respective base rates;
Initialize  $b_j$ 's with zeroes;
Initialize  $D_{ij}$ 's with zeroes;
 $n := 0$ ;
Compute the prediction error  $Err_n$  at epoch  $n$ ;
repeat
  for all mini-batch of users in the training set  $S$  do
    for all user  $u$  in the current mini-batch do
      Clamp the ratings of  $u$  on the visible units;
      Compute  $p_j = p(h_j = 1 | V, r)$  for all the hidden units;
      Record samples  $v_i^k p_j, v_i^k, p_j$ ;
      Run the Gibbs sampler for  $T$  steps;
      Compute  $p_j = p(h_j = 1 | V, r)$  for all the hidden units;
      Record samples  $v_i^k p_j, v_i^k, p_j$ ;
    end for
    Average the first samples to get  $\langle v_i^k h_j \rangle^+, \langle v_i^k \rangle^+$  and  $\langle h_j \rangle^+$ ;
    Average the last samples to get  $\langle v_i^k h_j \rangle^T, \langle v_i^k \rangle^T, \langle h_j \rangle^T$ ;
    Update  $W_{ij}^k$ 's using equation 4.13 (augmented with momentum);
    Update  $b_i^k$ 's using equation 4.14 (augmented with momentum);
    Update  $b_j$ 's using equation 4.15 (augmented with momentum);
    Update  $D_{ij}$ 's using equation 4.18 (augmented with momentum);
  end for
   $n := n + 1$ ;
  Compute the prediction error  $Err_n$  at epoch  $n$ ;
until  $Err_{n-1} - Err_n > \epsilon$ 
    
```

Louppe G. **Collaborative filtering: Scalable approaches using restricted Boltzmann machine**[J]. English, 2010.



1. 收集用户行为，包括点击、加购、关注、下单等



2. 对行为进行过滤，比如：join(白名单)、统计截断、position-bias、多次加/删购等处理



3. 制定行为评分规则，生成评分矩阵



4. 训练矩阵分解模型



5. 导入缓存

评分规则

行为	评分
浏览1次	2.52
浏览2次	2.9
浏览3次	3.1
浏览4-5次	3.2
浏览>5次	3.3
加入购物车1次	3.9
加入购物车2次	4.1
加入购物车>2次	4.2
购买1次	4.7
购买2次	4.9
购买>2次	5.0

- 1.打分之前必须要对数据进行清洗，去除“点击狂人”、“用户误点”、“买后查单”等情况
- 2.对非常活跃的用户要对其行为进行降采样

实时排序模型部署流程

场景恢复

用户交互行为

用户特征

商品特征

环境特征

训练样本构建

样本标注

进行标记：
购买>购物车>关注>点击>展示

(购买力水平、长期偏好、性别、家庭属性、品类、品牌、产品词、修饰词、销量、点击量、转化率、浏览用户量、下单用户量等等特征)

模型训练

数据清洗

离线评测
(正确率、召回率、NDCG)

特征工程

参数调优

线上实验

用户实时行为

用户&商品画像

多模型召回商品

实时交互特征

长期交互特征

离线特征计算

在线特征计算

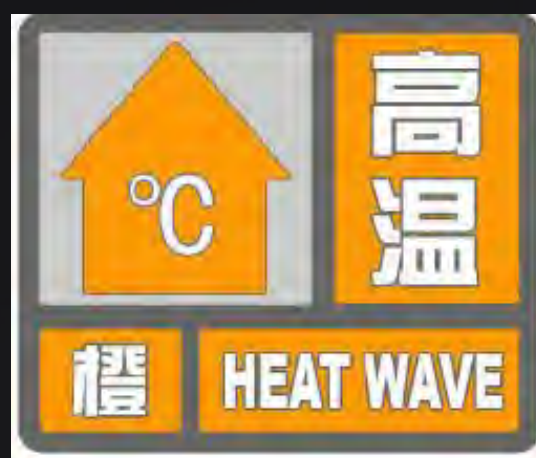
排序

A/B测试

推荐排序实践：环境特征

51CTO WOTI

某地高温
橙色预警



利用天气API定位某地高温橙色预警，作为排序特征。当用户浏览国美首页时，猜你喜欢展位推荐空调、风扇等商品：

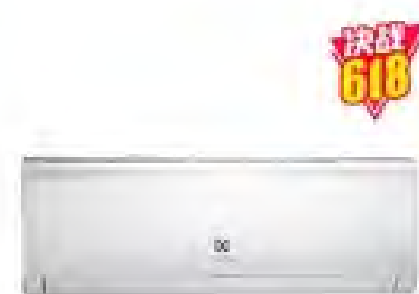
猜你喜欢—你的专属推荐

换一组



海尔(Haier) KFR-23GW/12N
WA13套机 小1匹P壁挂式定

¥1799.00



伊莱克斯(Electrolux)EAW25
FD13CA1 1匹P壁挂式定频

¥1299.00



美的 (Midea) AC120-16B
W 空调扇 (冷风机 加湿 净化

¥328.00



美的 (Midea) AC120-16BR
W 空调扇 (快速制冷 三档风

¥429.00



青岛啤酒 纯生330ml*24听 德
国风味 企业自营质量保障

¥110.00



青岛啤酒 经典500ml*12听 德
国风味 企业自营质量保障

¥69.00

某地雾霾
红色预警



利用天气API定位某地雾霾红色预警，作为排序特征。当用户浏览国美首页时，猜你喜欢展位推荐空气净化器、口罩等商品：

猜你喜欢—你的专属推荐

换一组



飞利浦 (PHILIPS) AC4076
空气净化器 (智能空气测控

¥2499.00



格力大松 (Gree) KJFC230
A-WG 空气净化器

¥2579.00



3M9001V防雾霾口罩9002V
带呼吸阀男女骑行PM2.5粉

¥90.00



美国 3M 9021 9022 KN90防
颗粒物防尘口罩工业粉尘 防

¥110.00



养生堂 天然维生素E软胶囊2
50mg*90粒 VE ve 维生素e

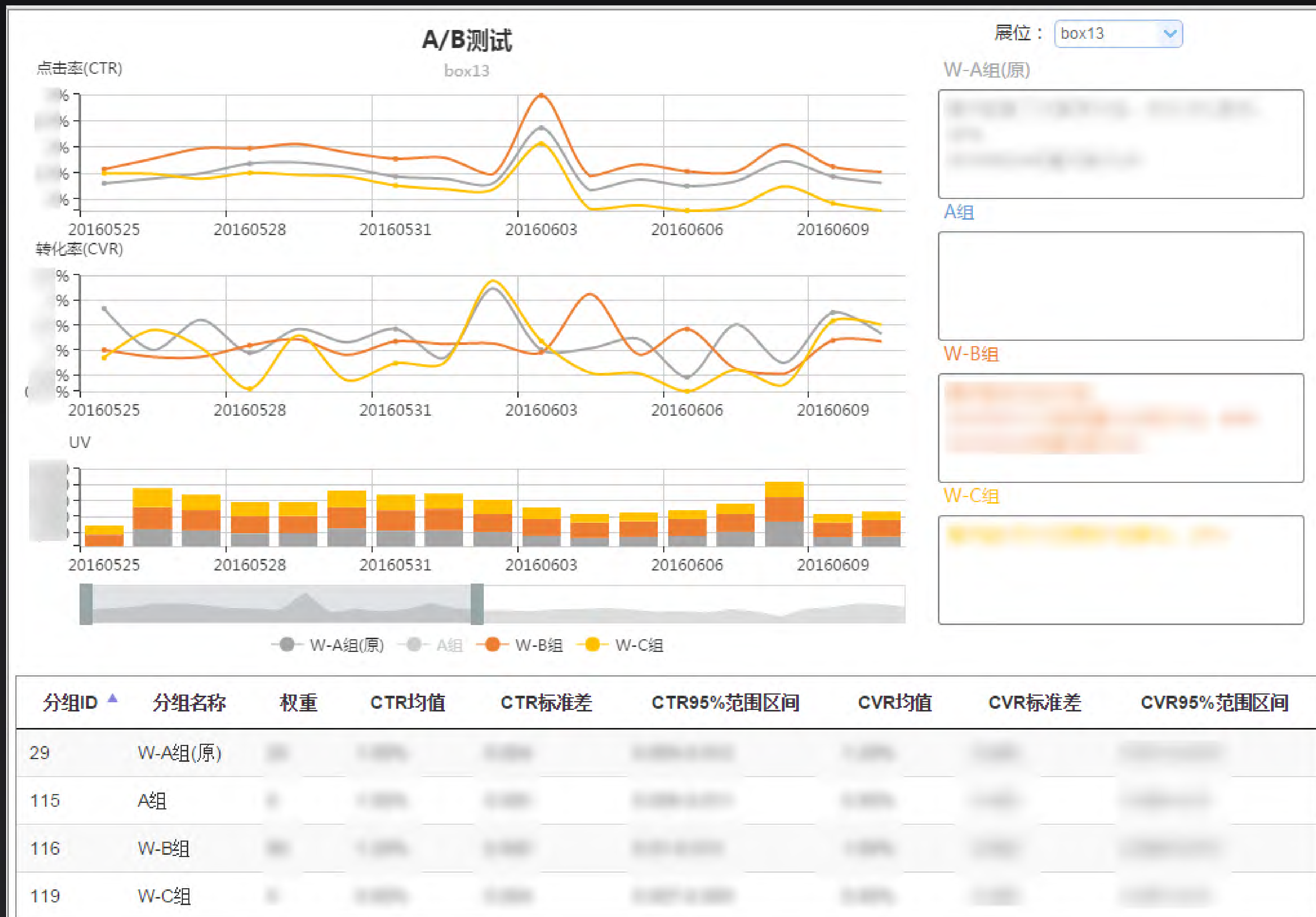
¥69.00



养生堂天然维生素C咀嚼片85
0mg*80片

¥82.00

推荐排序实践：A/B测试



实时排序

过滤：业务需求（无库存、成人用品、促销商品）

实时特征：人vs商品，实时行为交叉

多样性：多模型、多数据源融合

新颖度

KPI驱动：CTR&CVR&GMV&**粘性**

A/B测试，快速迭代

离线召回

隐语义模型数据：矩阵分解、RBM.....

协同过滤：UserBasedCF、ItemBasedCF

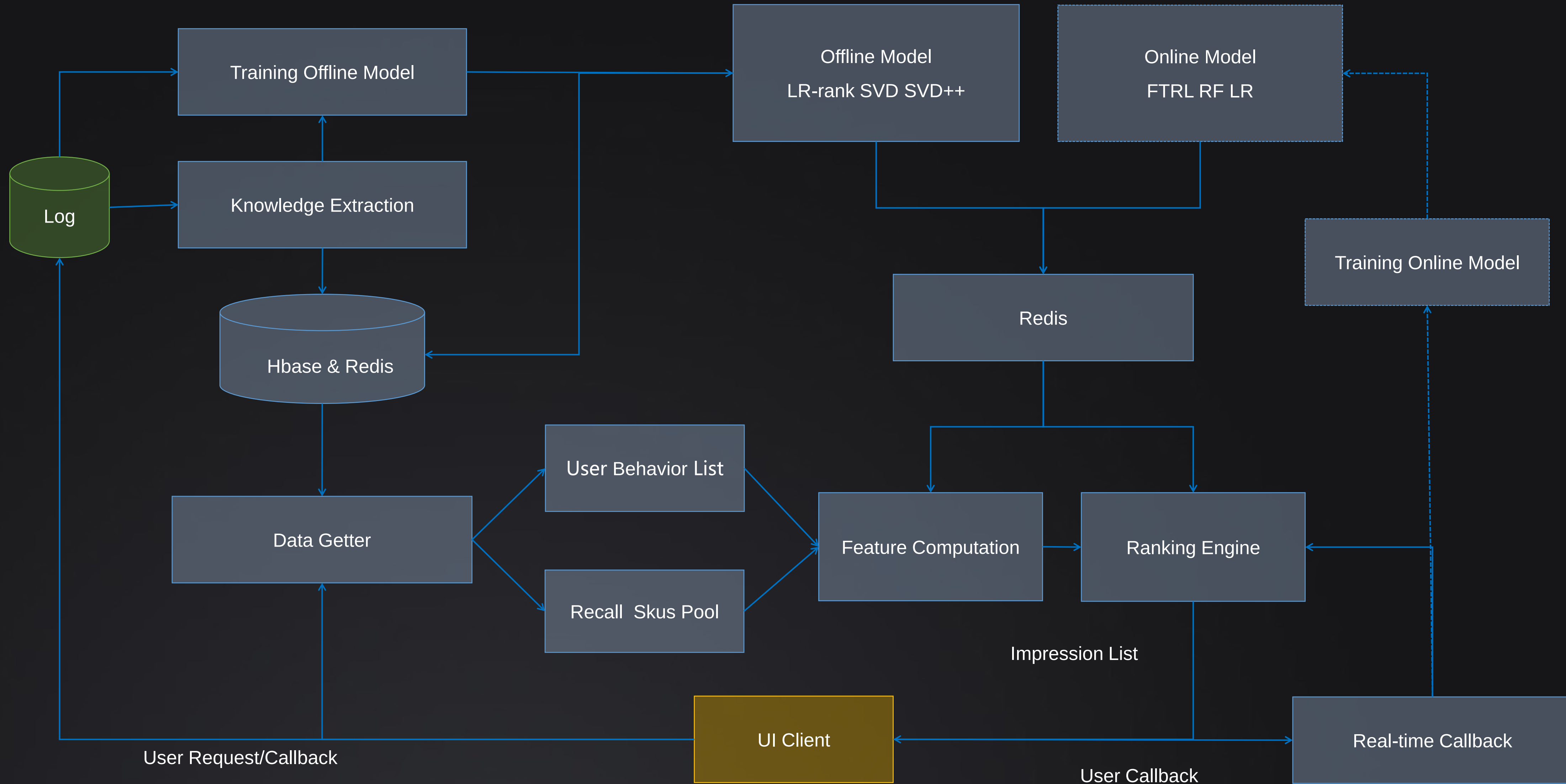
用户画像：品类&品牌兴趣、性别、年龄

商品画像：价格指数、人群、区域等

计算机视觉特征

离线测试







目录

01

社交电商个性化场景

02

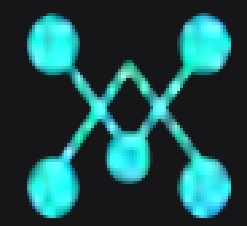
国美互联网机器学习实践

03

国美互联网深度学习的应用场景

04

未来的挑战



目录

01

社交电商个性化场景

02

国美互联网机器学习实践

03

国美互联网深度学习的应用场景

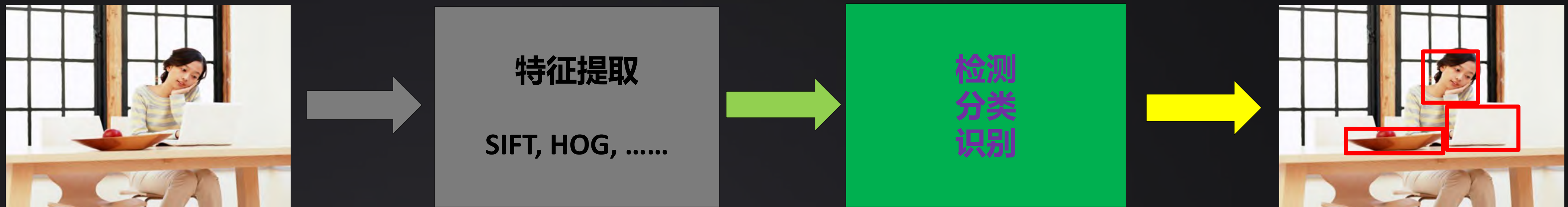
04

未来的挑战

计算机视觉专家

第一步：设计特征，比如SIFT，KAZE，HoG，SURF等

第二步：训练分类器，进行多目标检测、分类和识别



传统计算机视觉领域需要领域知识，耗费巨量时间

第一步：根据训练数据自动构建特征

第二步：将特征抽取和分类/检测任务同时进行

深度学习专家的任务：定义神经网络架构，并进行训练



深度学习会自动训练出好的特征，并且很多方法可以在各个领域中通用

我们的方案

第一步：同时抽取计算机视觉特征和DNN特征

第二步：进行多目标识别

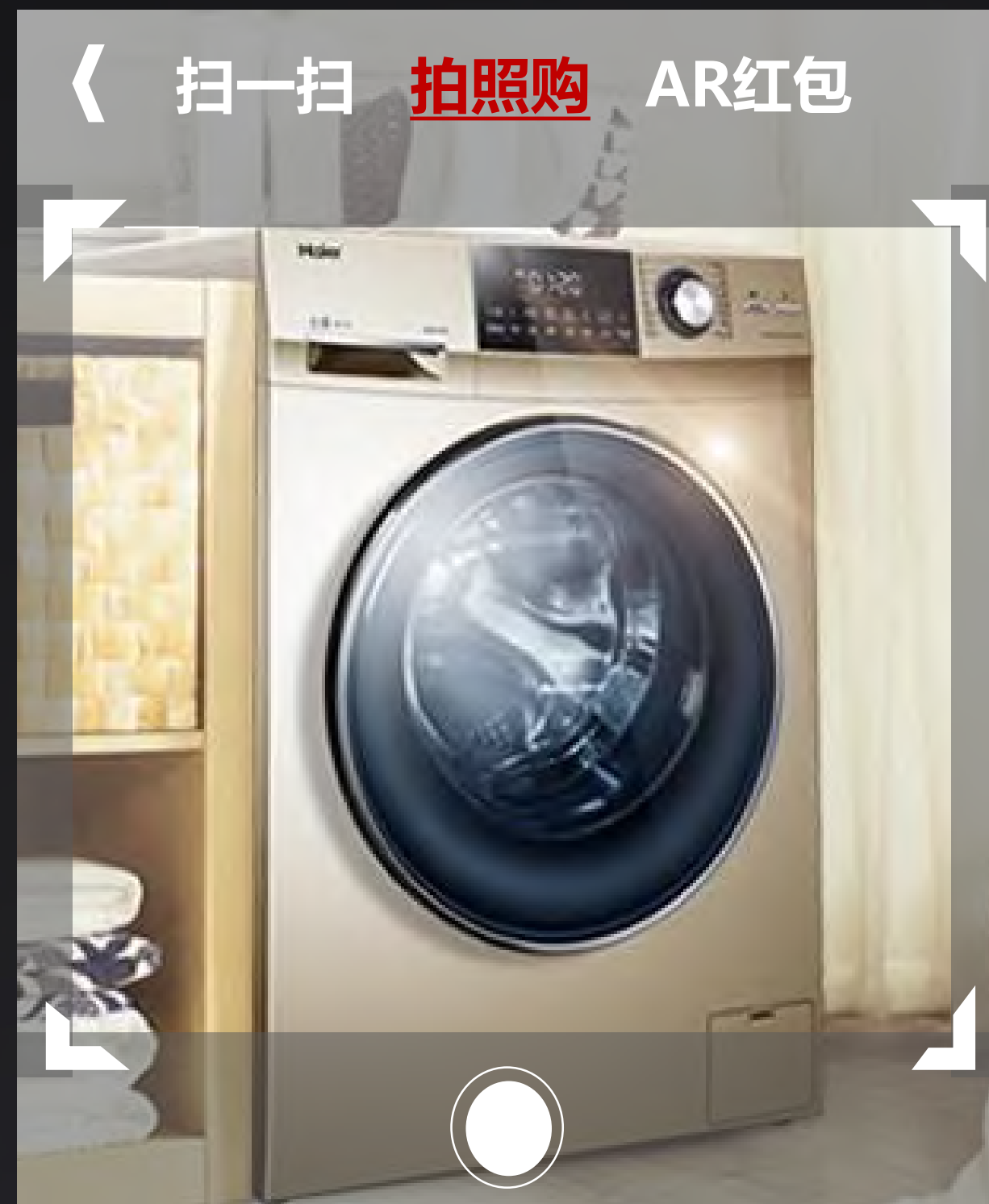


将计算机视觉特征和深度学习特征相结合

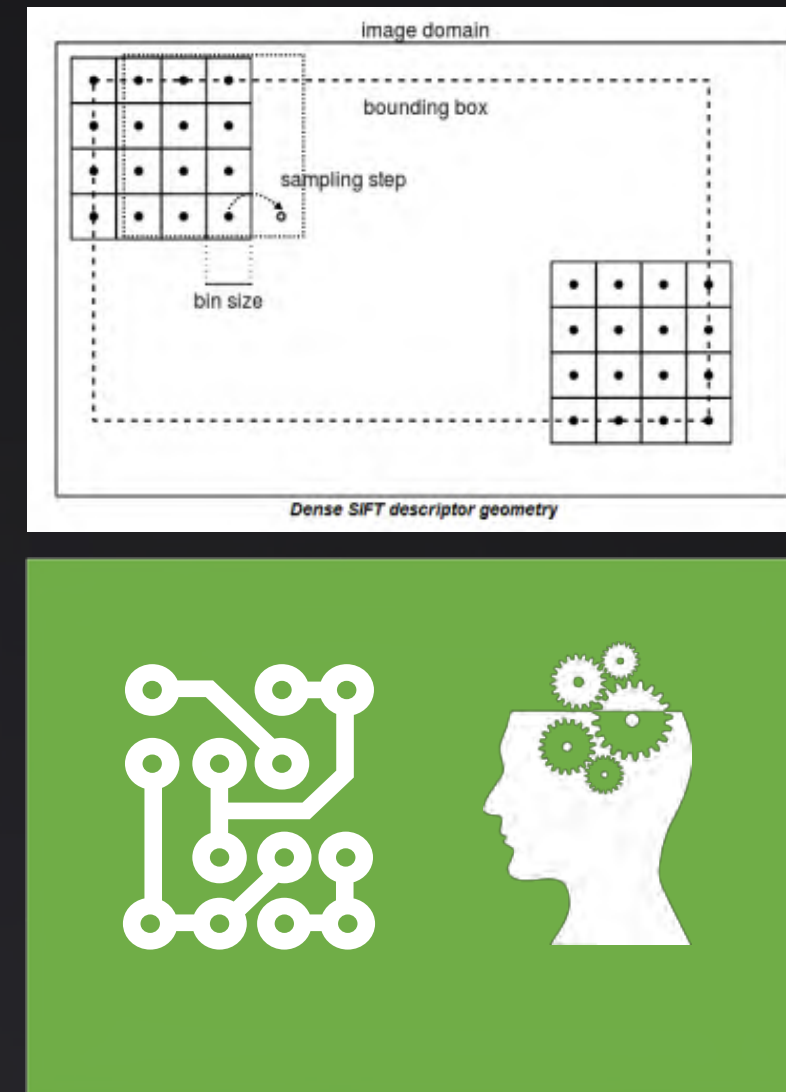
深度学习构建GomePlus新场景：“拍照购”51CTO WOTI

实现根据用户上传图片预测品类推荐相关商品

用户拍摄或选取商品图片上传



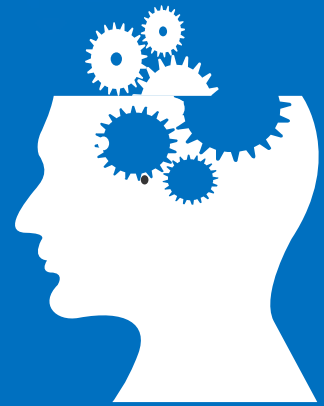
特征向量计算及匹配



根据商品图片库结果展示

猜你要找

西门子 (SIEMENS) WM12P2699W 9公斤 滚筒洗衣机 (缎光	西门子 (SIEMENS) XQG90-W M12P2689W 9KG变频滚筒洗衣
¥4599.00	¥4190.00
西门子 (SIEMENS) XQG80-WM1 2P2688W 8kg 滚筒洗衣机	西门子洗衣机WD12G4601W 8 公斤 洗烘一体机 智能 烘干
¥3599.00	¥4699.00



图像标注

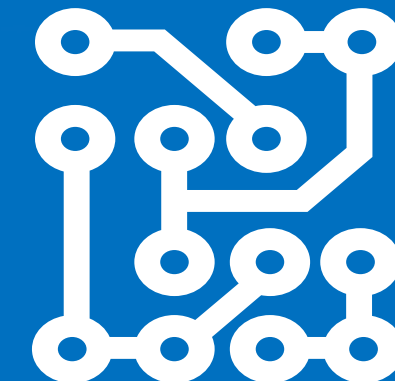
对国美全站的商品图像进行标注：

- 1.最主要的工作是数据清洗，即把不能和品类对应的图片删除或者重新进行品类校准
- 2.统计国美全站最近一年内各品类下商品总数的分布；然后按照分布进行图片抽样



提取特征

图片的特征分为两个部分
一是通过深度学习得到的特征，二是图像色局部特征：
利用TF/caffe等训练CNN将倒数第二层输出作为Feature Learning的结果提取出来 **(Deep Learning + Transfer Learning)**
利用局部特征算子(SIFT,kaze等)提取出图像的局部特征



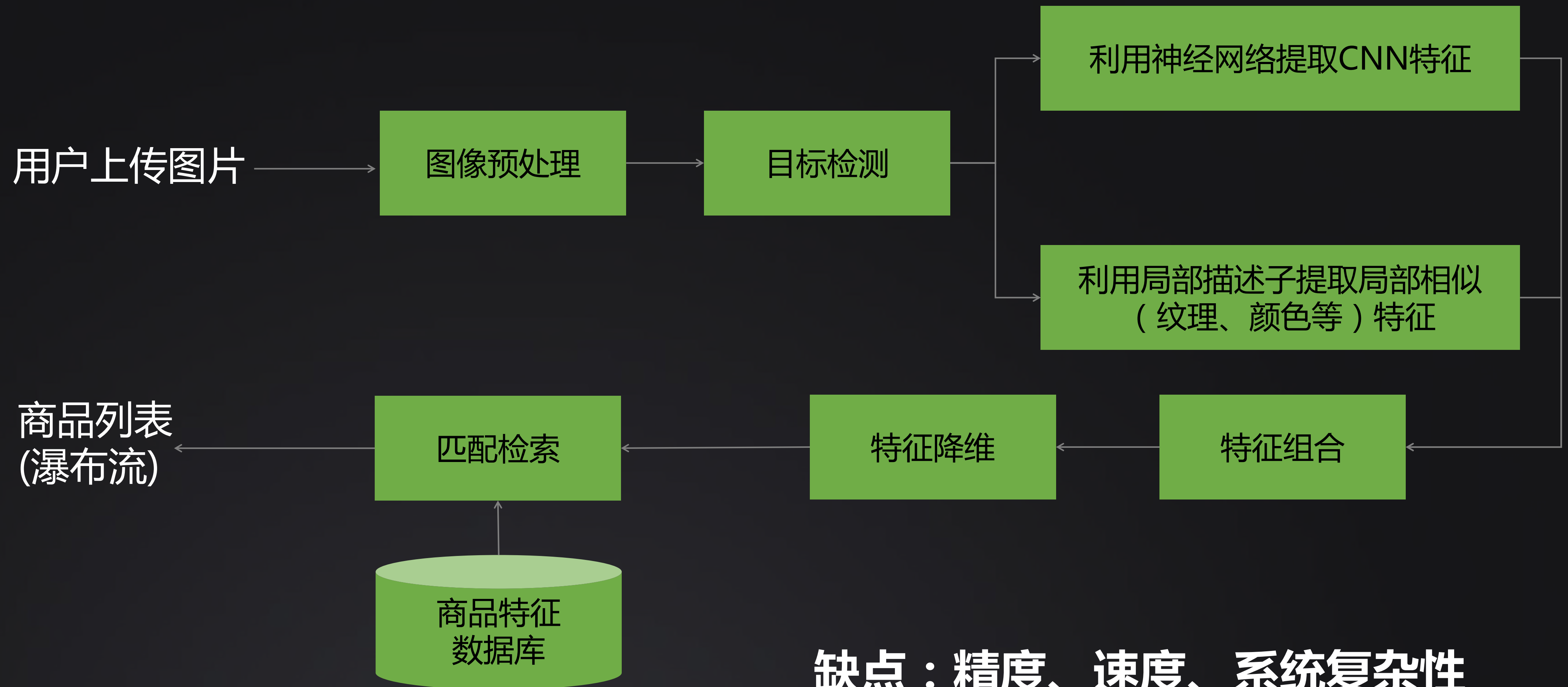
降维

分别对前面两种特征进行降维处理：
用积量化 (Product Quantization) 的方法对CNN特征进行降维
用VLAD对图像局部特征进行编码

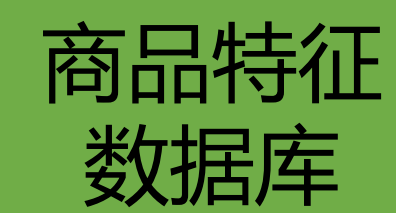


匹配检索

采用最近邻搜索的方法找出每一个商品的相似商品集合



缺点：精度、速度、系统复杂性



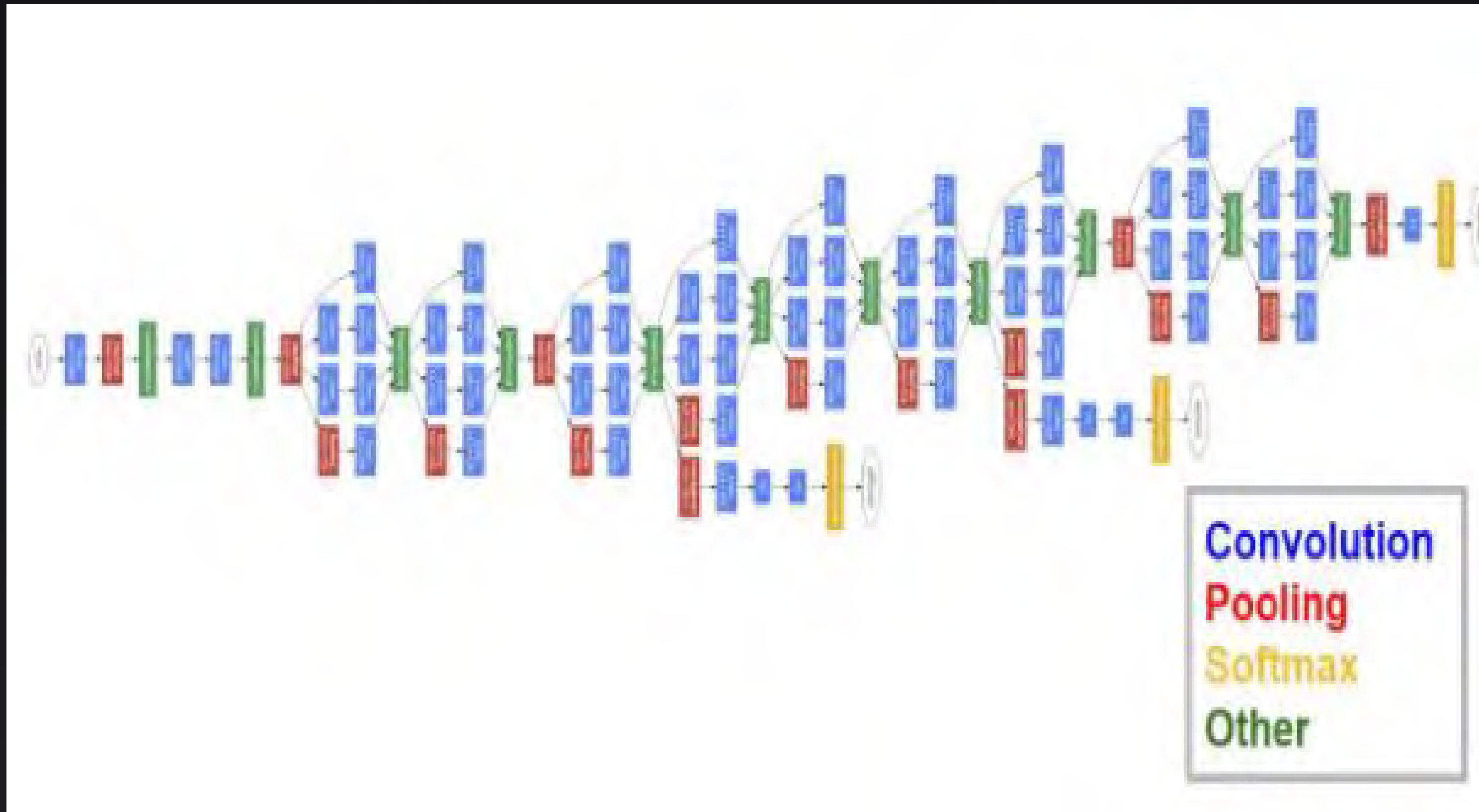


● 层数少

● 易于训练

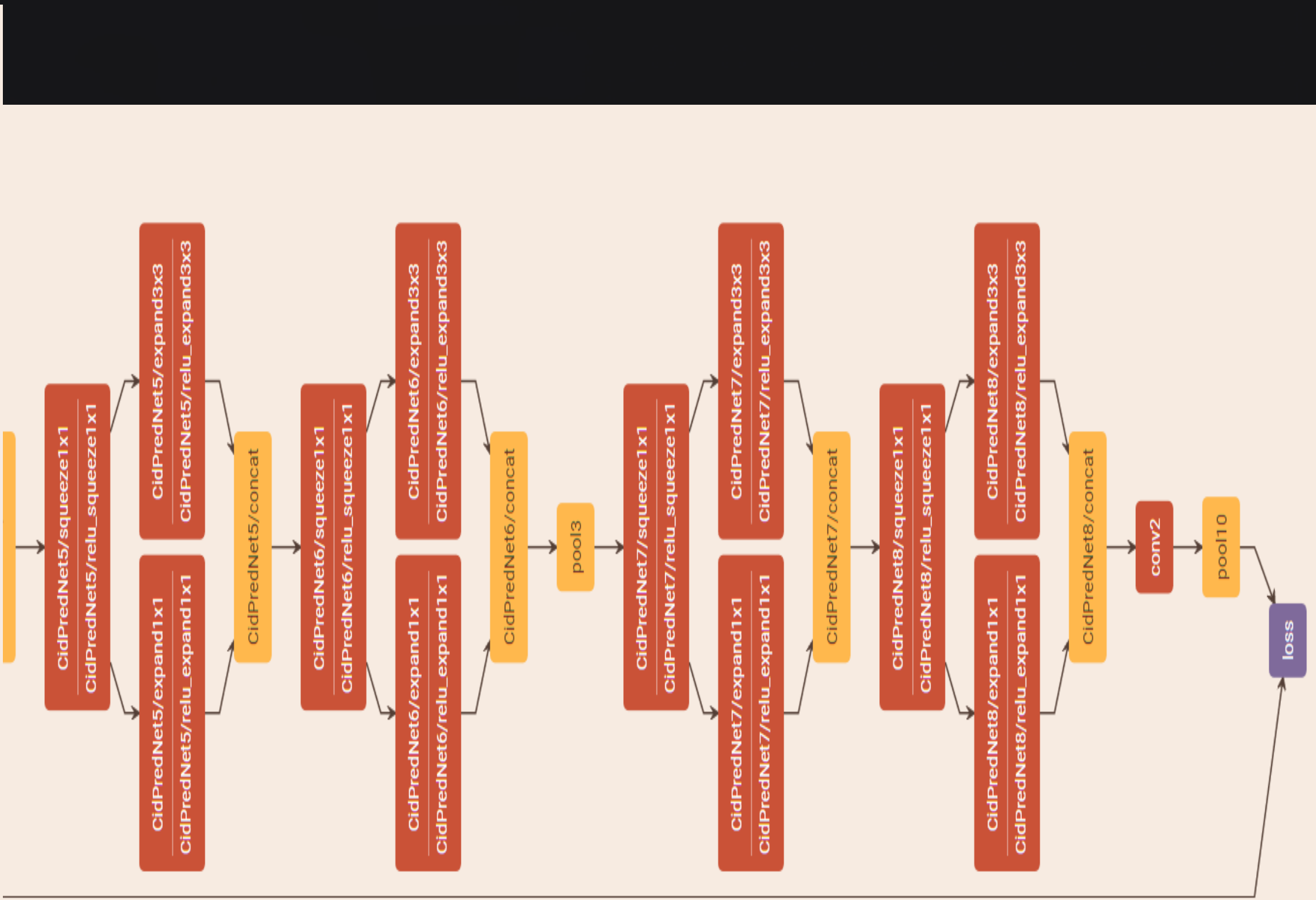
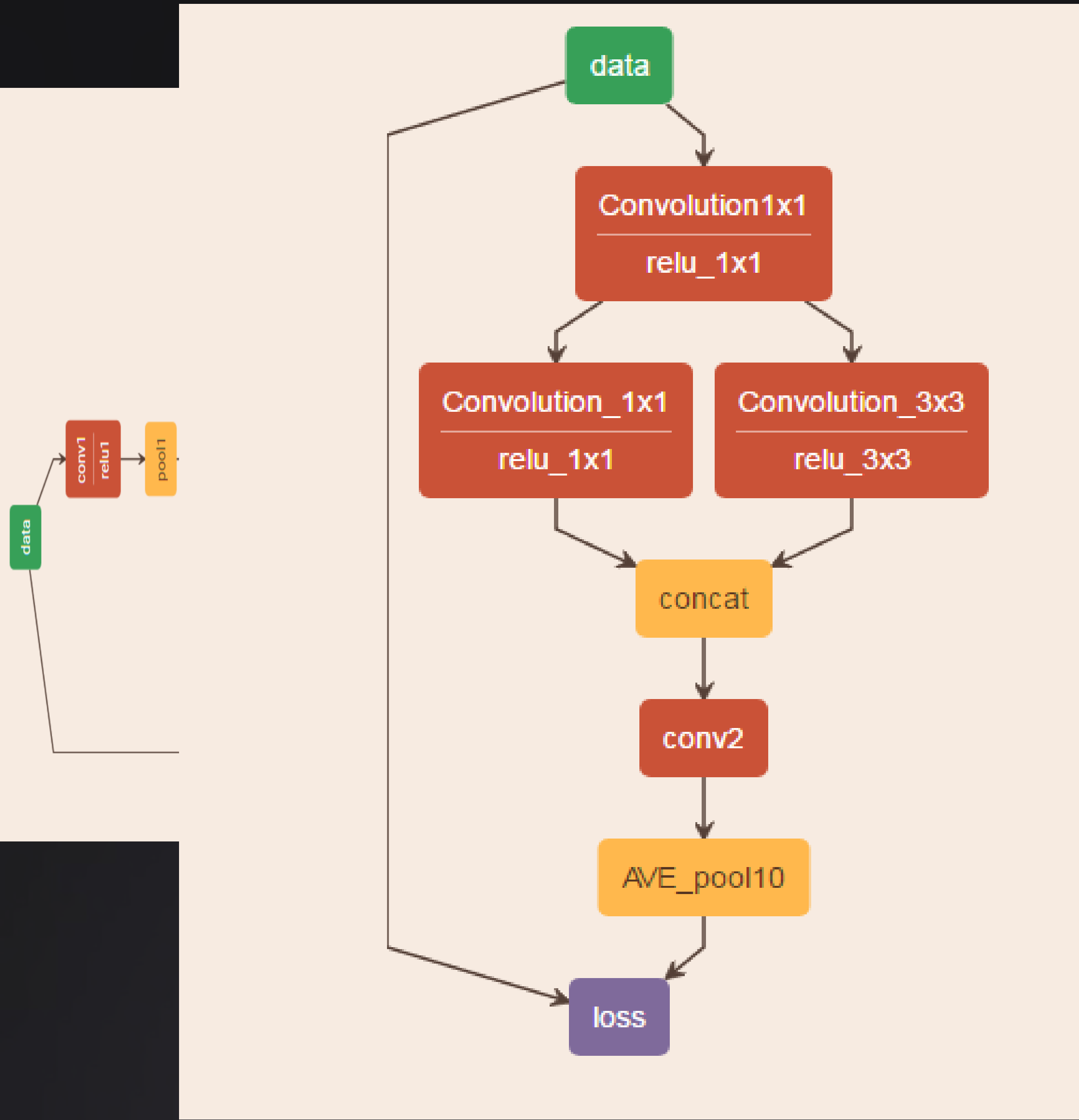
● 体积：200M

● **准确率：>94%**

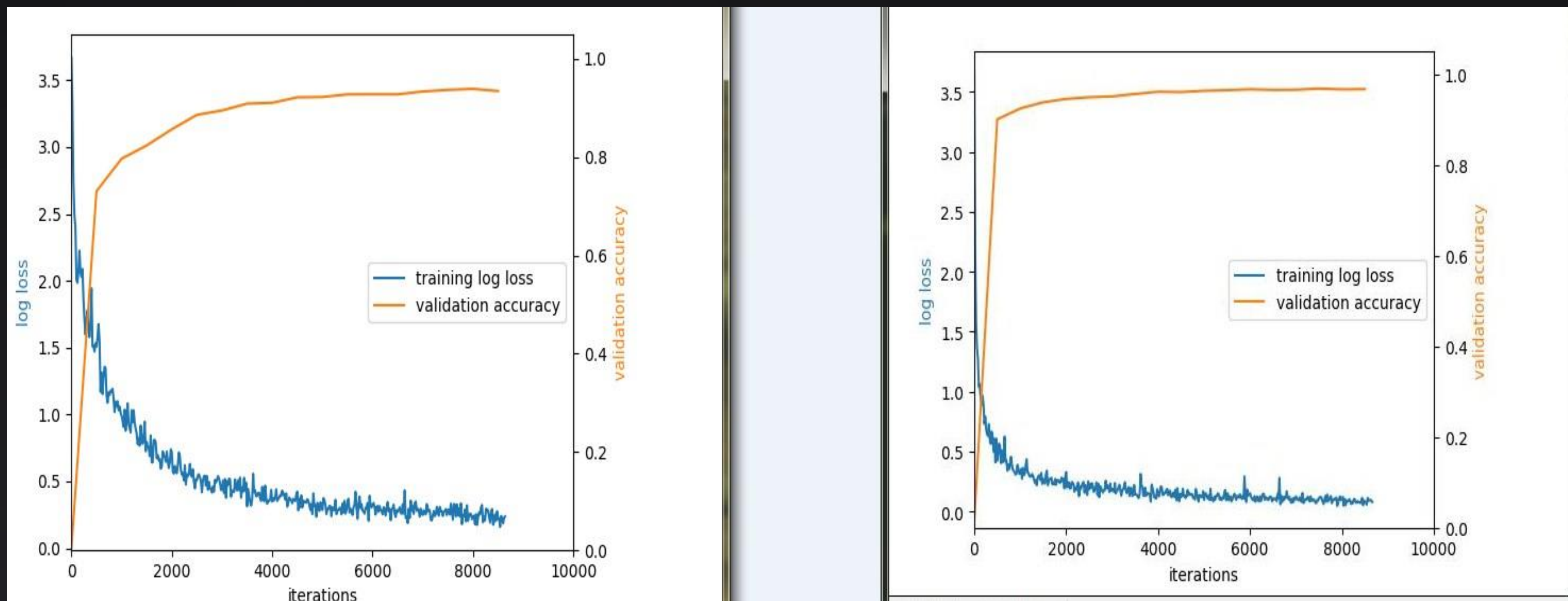


GoogleNet

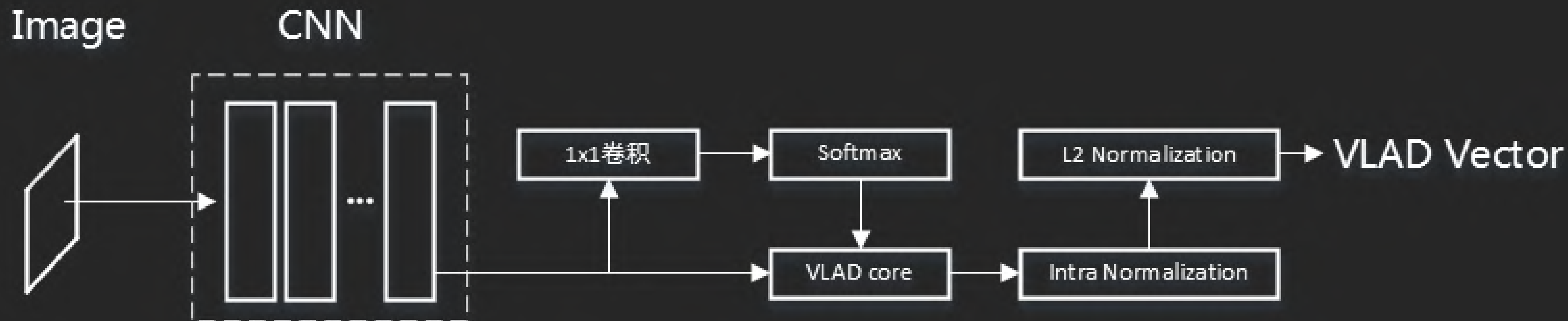
- 多尺度
- 1*1卷积
- Average Pooling
- 2*Softmax
- 1*1卷积
- 体积：50M
- 准确率：>96%
- 泛化能力好



准确率：>95%



利用Soft-assignment，把VLAD嵌入CNN模型，进行End2End的训练



Arandjelovic R, Gronat P, Torii A, et al. **NetVLAD: CNN architecture for weakly supervised place recognition.** [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2015, PP(99):1-1.

深度学习构建GomePlus新场景：相似推荐 51CTO WOTI

实现根据图像找相似的推荐应用

基于深度学习的检索技术可以得到一套通用的商品特征学习框架。在此基础上，针对某些特定场景(比如商品详情页的“找相似推荐”)，可以根据商品**主副图**，在同品类、同产品维度上，进行最近邻搜索，召回相似商品。即提高了召回商品的多样性，同时也加快了新上架商品由冷到热的速度，相当于一种基于内容的Explore-Exploit方法。



基于图像相似性的商品推荐

其他类似商品



男士短袖+恤衫大码宽松圆领
T恤衫半袖圆领打底衫 01白色

¥29.00



卡郎琳 男士夏季新款圆领修身
短袖T恤单穿打底衫 男青

¥58.00



夏装打底衫男短袖+恤韩版修身
圆领薄植个性上衣服男

¥79.00



本优品新款推荐男士莱卡
棉短袖T恤 3D骷髅头印花

¥39.00



【华康威】男士短袖+恤丁恤
青年纯棉修身半袖夏季款201

¥29.00



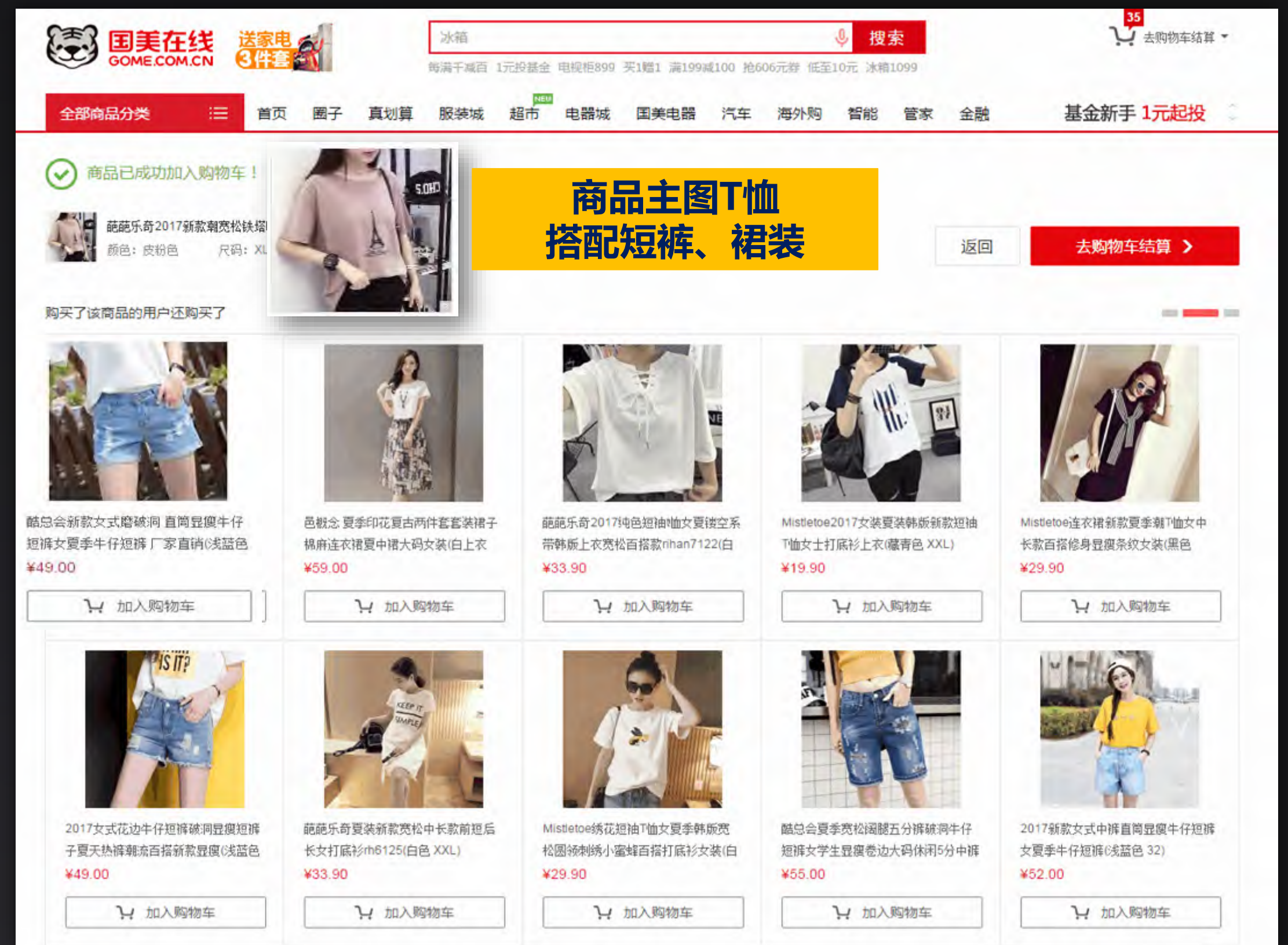
2017夏季新款圆领修身T恤男
百搭图案短袖 (白色 XXXL)

¥49.00

深度学习构建GomePlus新场景：搭配购推荐CTO WOTI

实现根据图像找搭配的推荐应用

搭配购是推荐领域非常重要的一个应用场景，即根据用户当前请求，推荐出能够与之搭配的商品。利用深度学习技术，通过对训练集图片中的不同主题进行检测，然后利用CNN学习出跨类目&跨产品最佳匹配的模式，然后进行相关推荐。搭配购一直是推荐中最难解决的问题之一，因为只依靠用户行为采样，进行类似购物篮分析之类的方法，很难做到精准。





目录

01

社交电商个性化场景

02

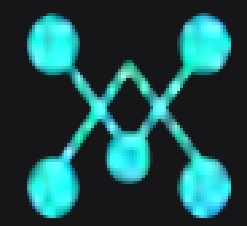
国美互联网机器学习实践

03

国美互联网深度学习的应用场景

04

未来的挑战



目录

01

社交电商个性化场景

02

国美互联网机器学习实践

03

国美互联网深度学习的应用场景

04

未来的挑战



图像描述的自动生成



实现各种场景下的图像合成和转换



2D图像转换为3D图像



模型压缩，Mobile端部署



分布式检索引擎

