



2017第八届中国数据库技术大会

DATABASE TECHNOLOGY CONFERENCE CHINA 2017

基于深度学习的图片检索应用

朱剑锋

百度深度学习实验室

图片检索

query
猫



step1: 召回



step2: 排序



传统检索算法

➤ 召回

- 倒排索引

➤ 排序

- 特征提取

- 文本相关性: query与图片周边文本(标题, 描述等)的关系
- 图片质量: 分辨率、清晰度
- 站点信息: 权威度、热门度
- 历史行为: 点击率

- 排序算法

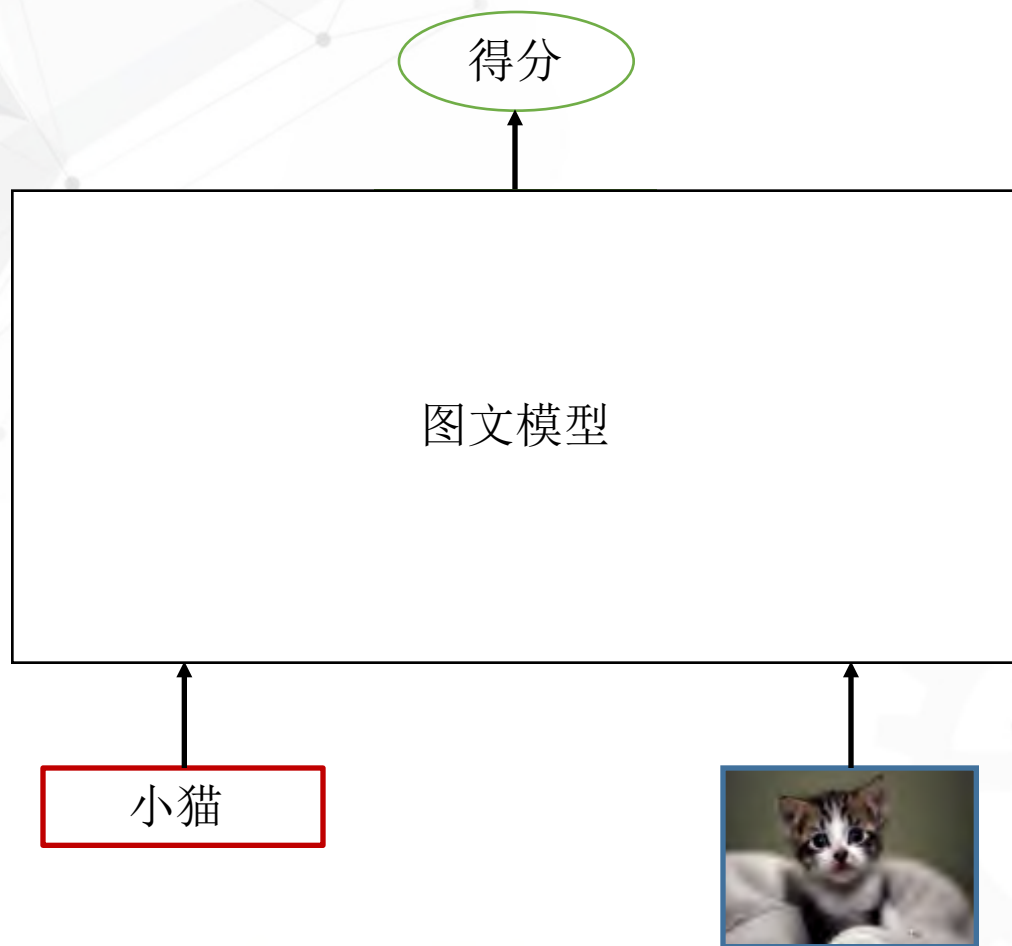
- LR, GBDT 等等

新的场景， 新的问题

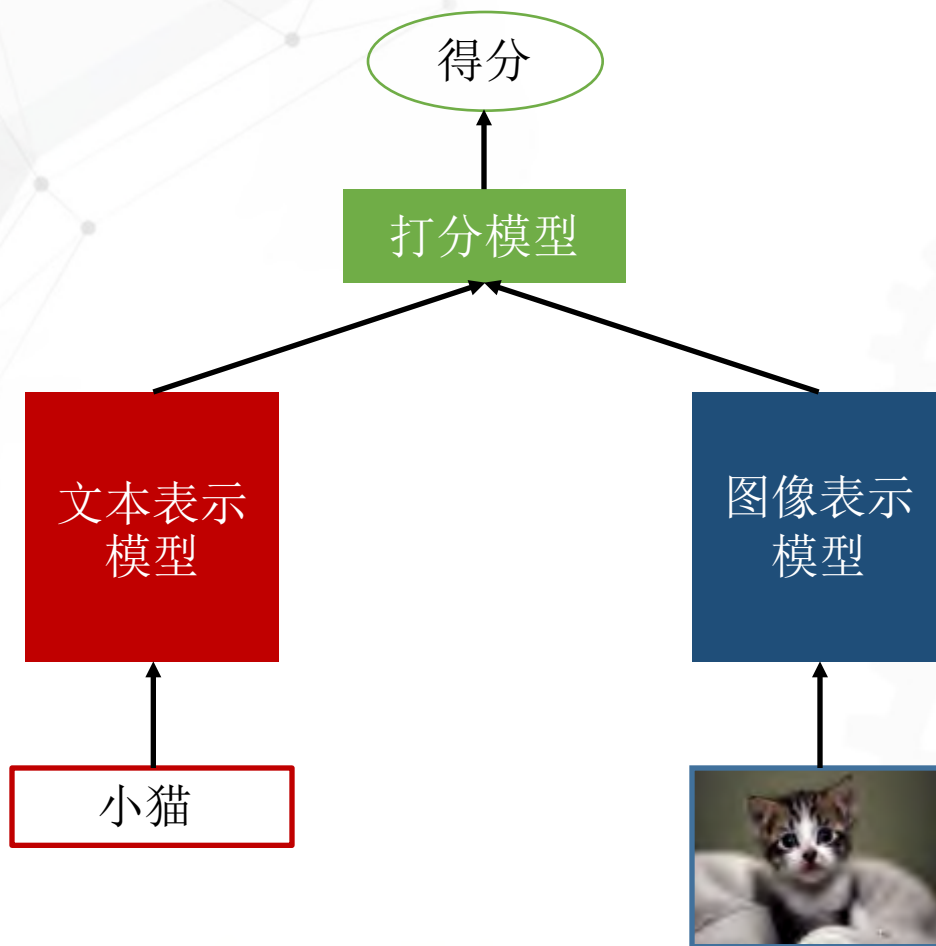
- 图片周边文本： 关键词准确率低（30%）
- 来源： 固定站点， 无先验偏好
- 用户行为： 冷启动， 无历史数据

能否训练模型直接对query和图片进行相关性打分？

图文模型框架



图文模型框架



新的工具—深度学习

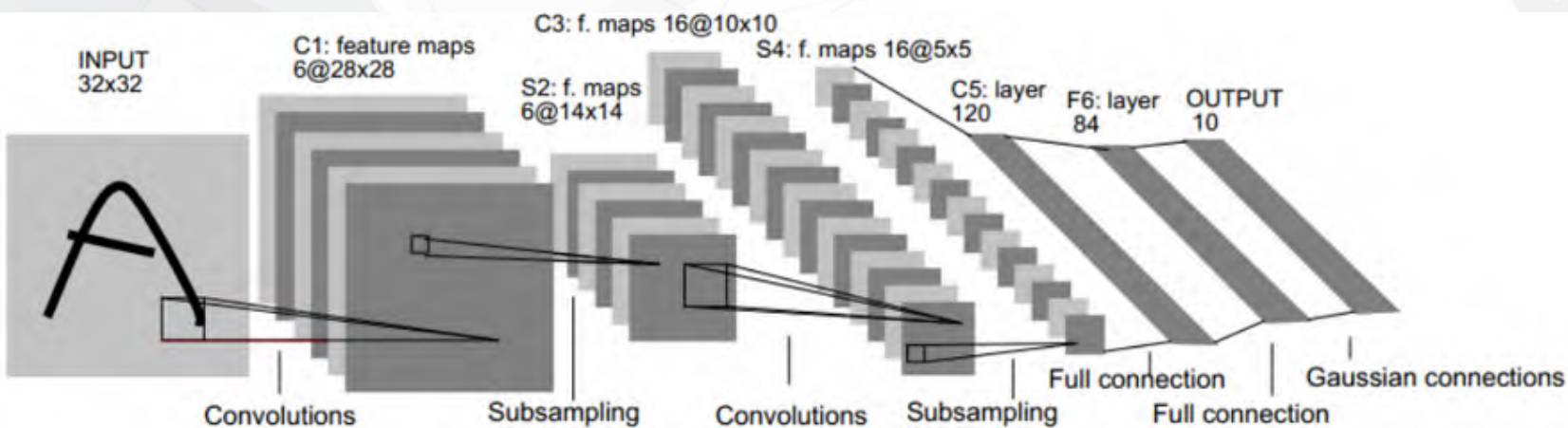
- 图像
 - 卷积神经网络（CNN）
- 文本
 - 词向量（Word2Vec）
 - 序列模型（RNN, LSTM）

手写数字识别

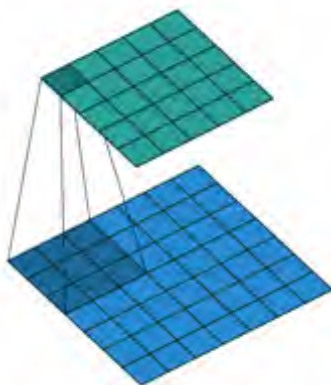
- MNIST数据集：6万图片，共10类



手写数字识别



错误率: <1%

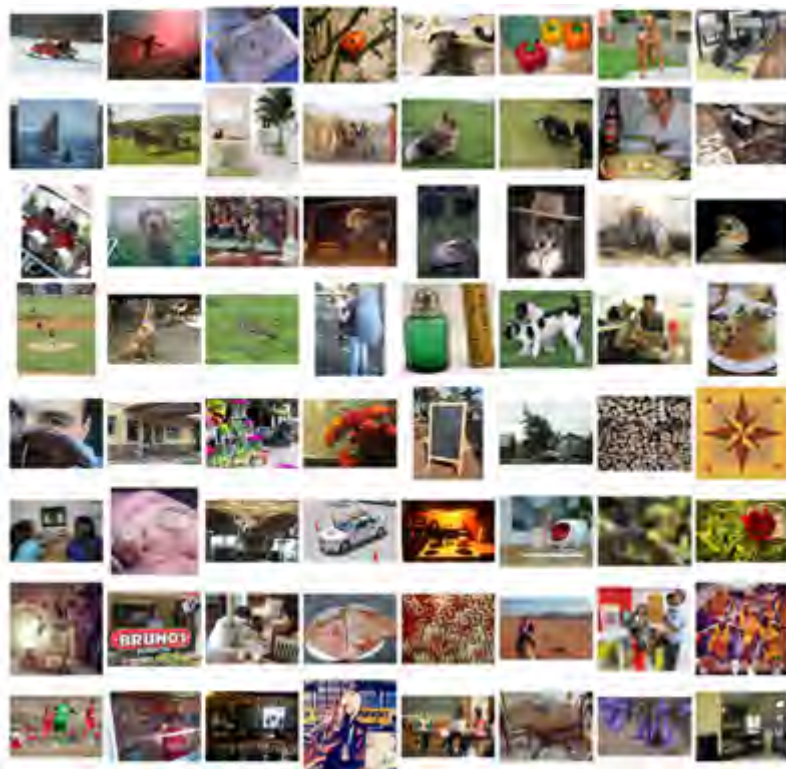


LeNet-5, 1998

<http://iamaaditya.github.io/2016/03/one-by-one-convolution/>

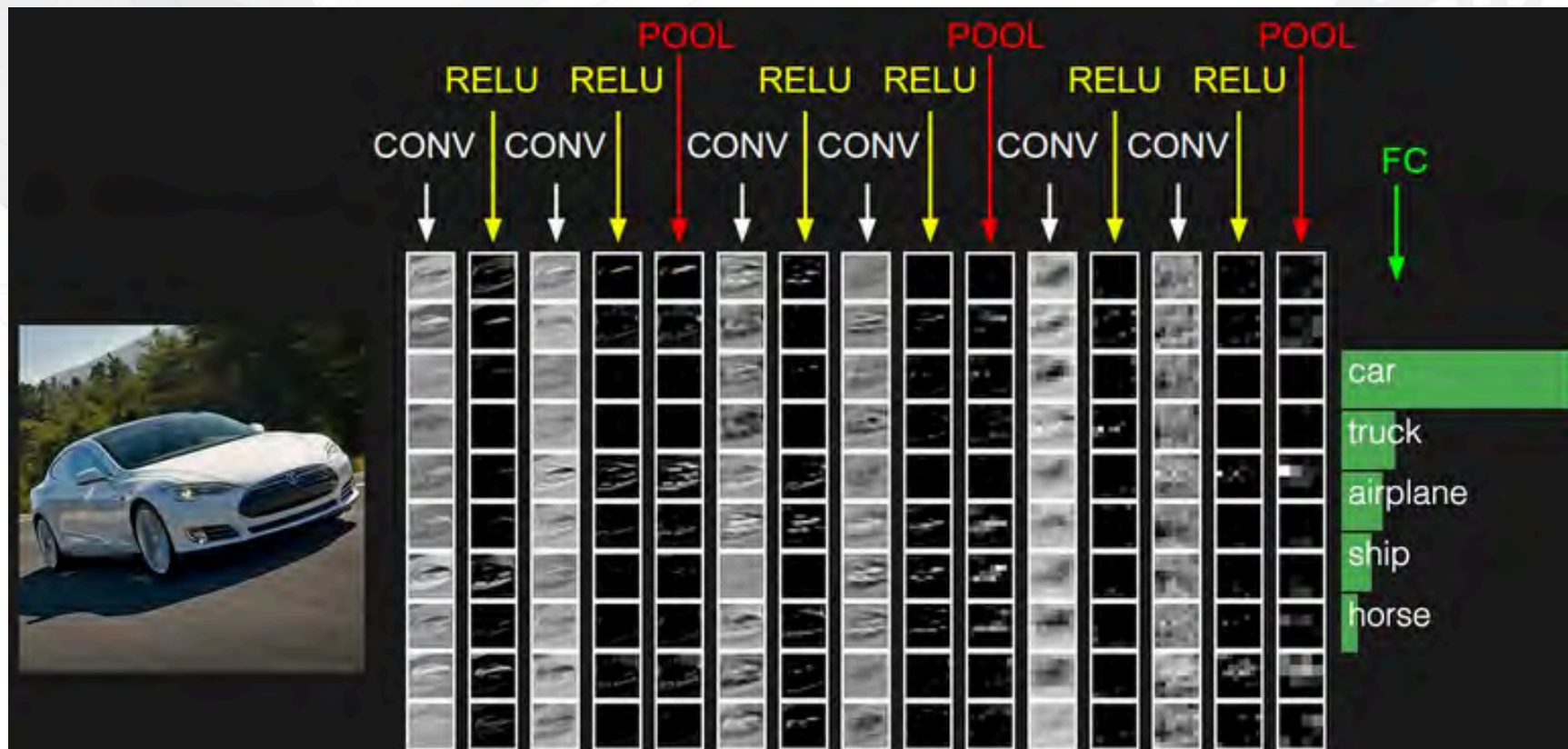
大规模图像识别

- ImageNet数据集：100万图片，共1000类



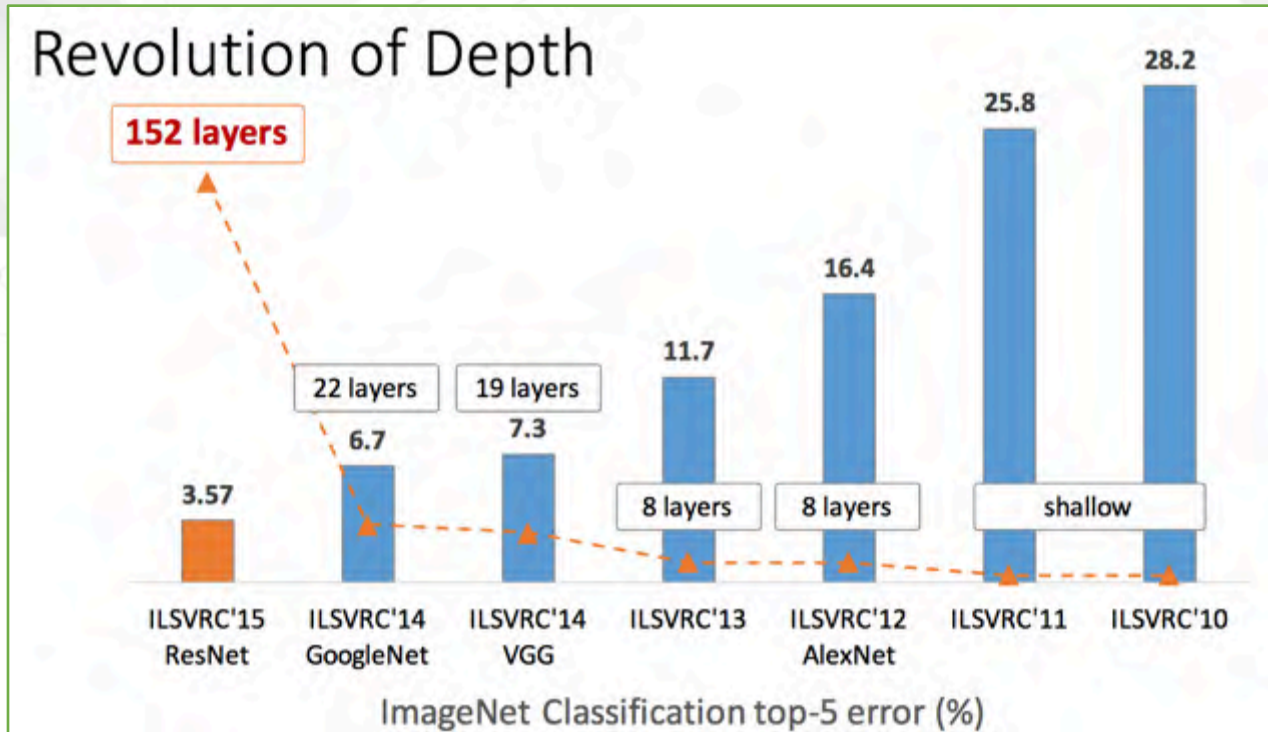
Olga Russakovsky. "ImageNet Large Scale Visual Recognition Challenge". 2014

大规模图像识别



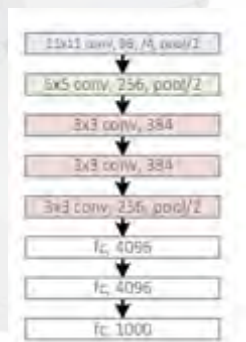
<http://cs231n.github.io/convolutional-networks/>

错误率 vs. 网络层数



Kaiming He. "Deep Residual Learning for Image Recognition". CVPR 2016

大规模图像识别



AlexNet, 8层



VGG, 19层



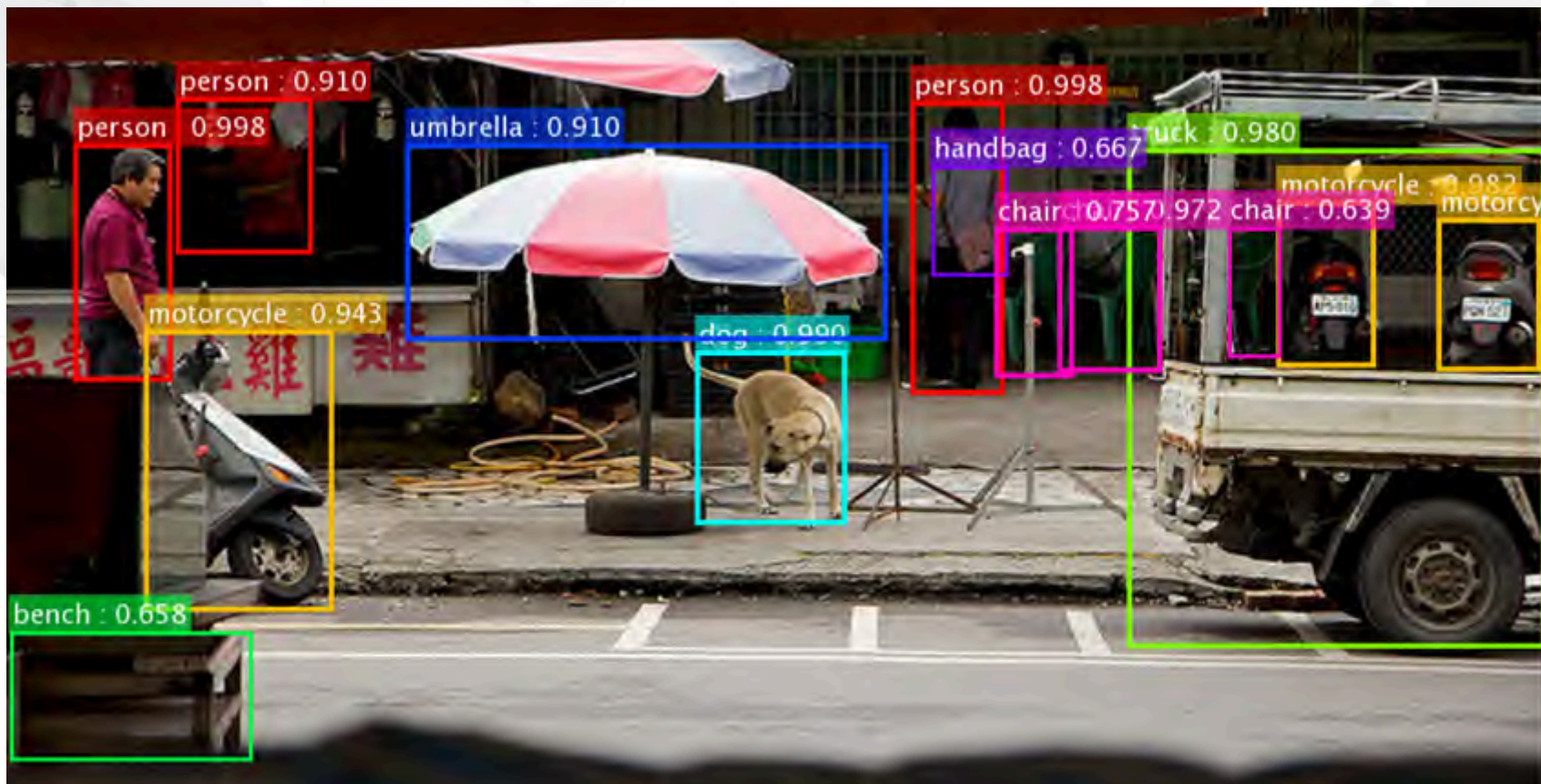
GoogleNet, 22层



ResNet, 152层

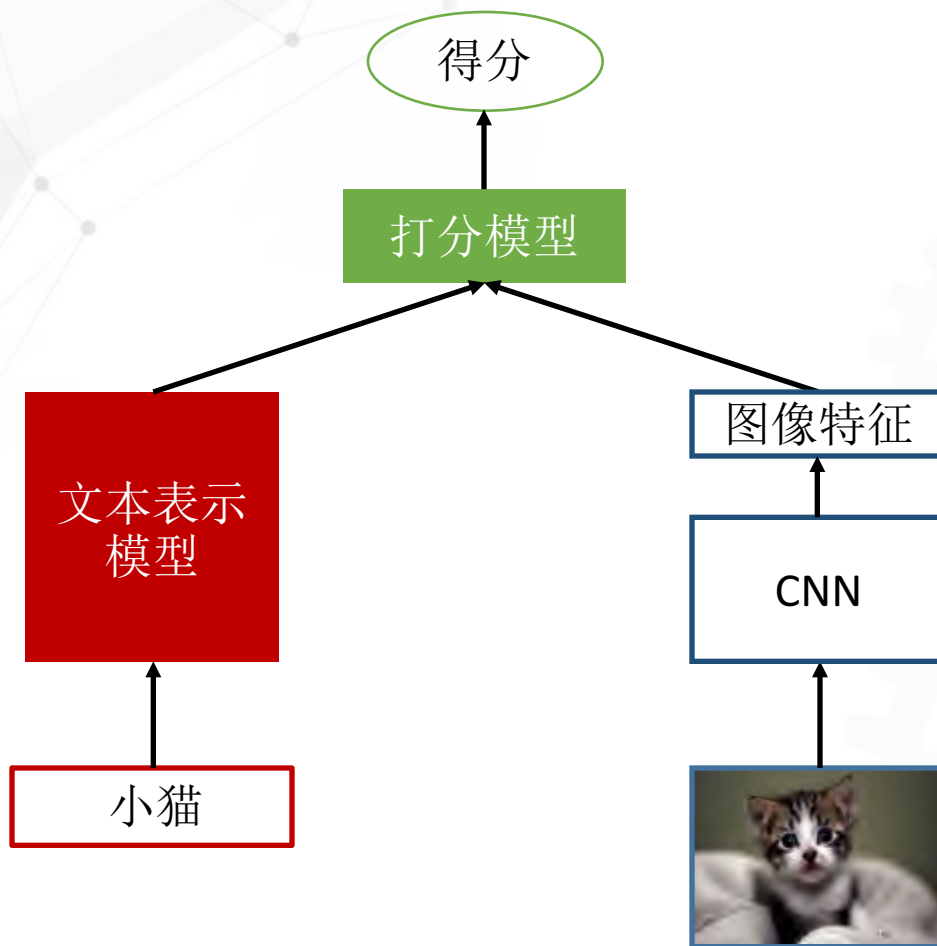
Kaiming He. "Deep Residual Learning for Image Recognition". CVPR 2016

大规模图像识别



Kaiming He. "Deep Residual Learning for Image Recognition". CVPR 2016

图文模型框架



文本向量化

- 离散表示 (One-Hot)

- 猫 [0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0]

- 狗 [0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0]

-

- 缺陷

- 无法度量词语之间的语义相似度

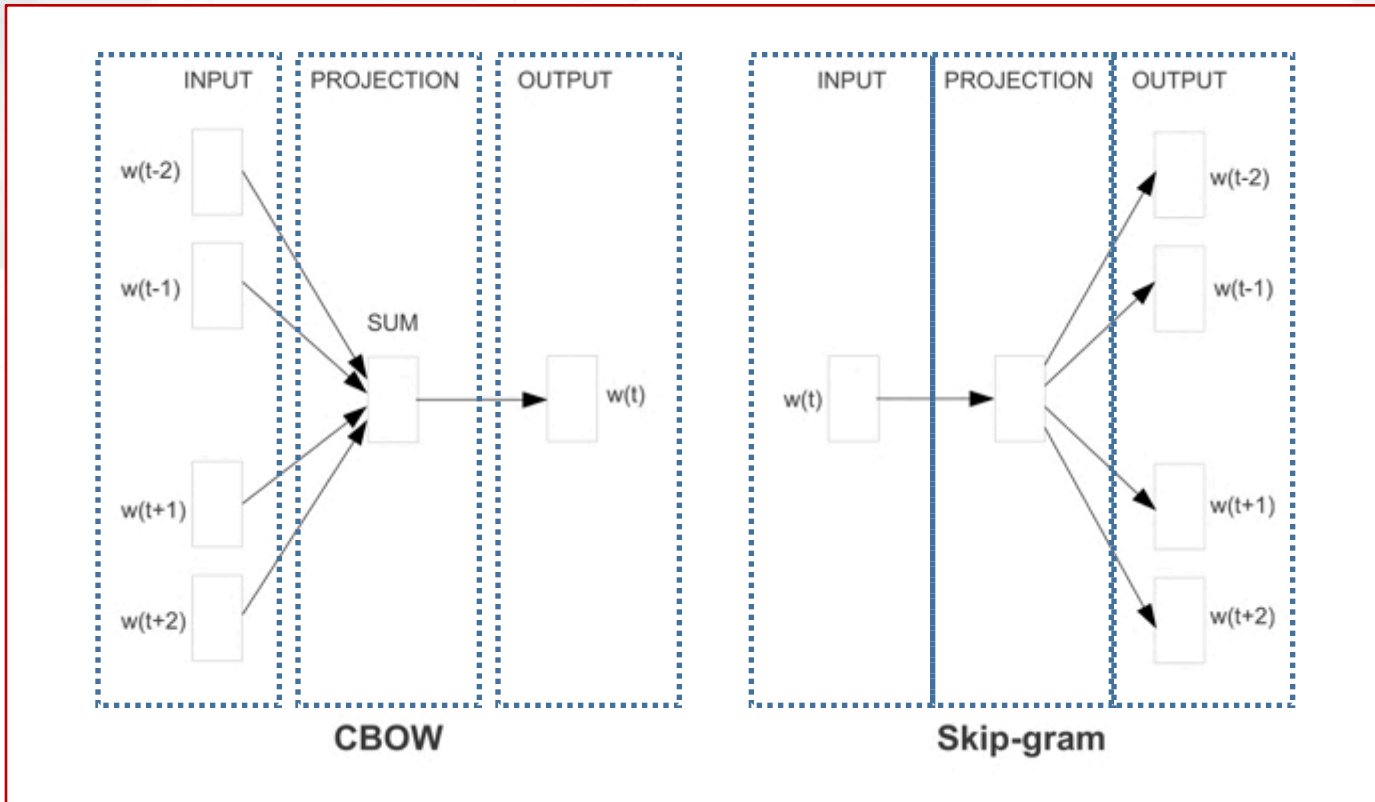
文本向量化

- 分布式假说（Distributional Hypothesis）

You shall know a word by the company it keeps
(J. R. Firth 1957)

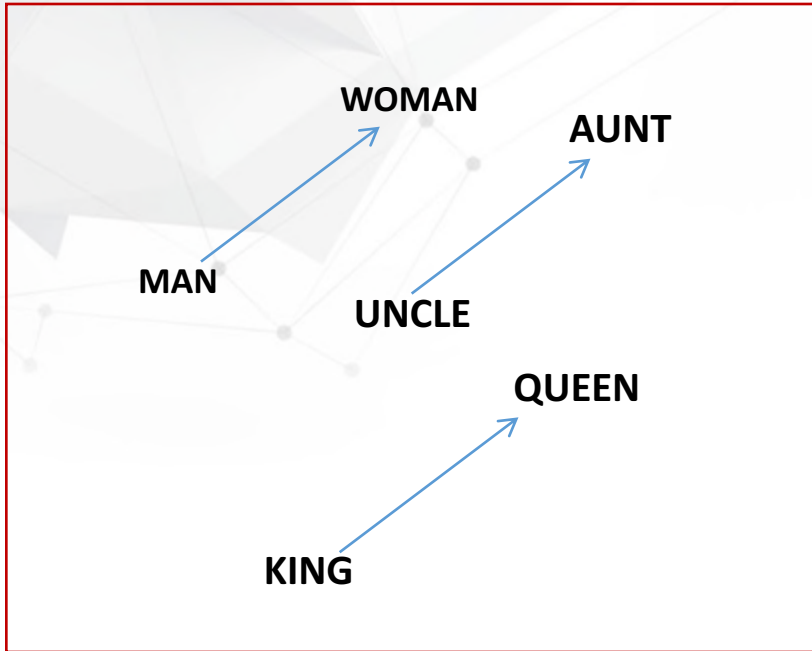
***Words which are similar in meaning
occur in similar contexts***
(Rubenstein & Goodenough 1965)

Word2Vec

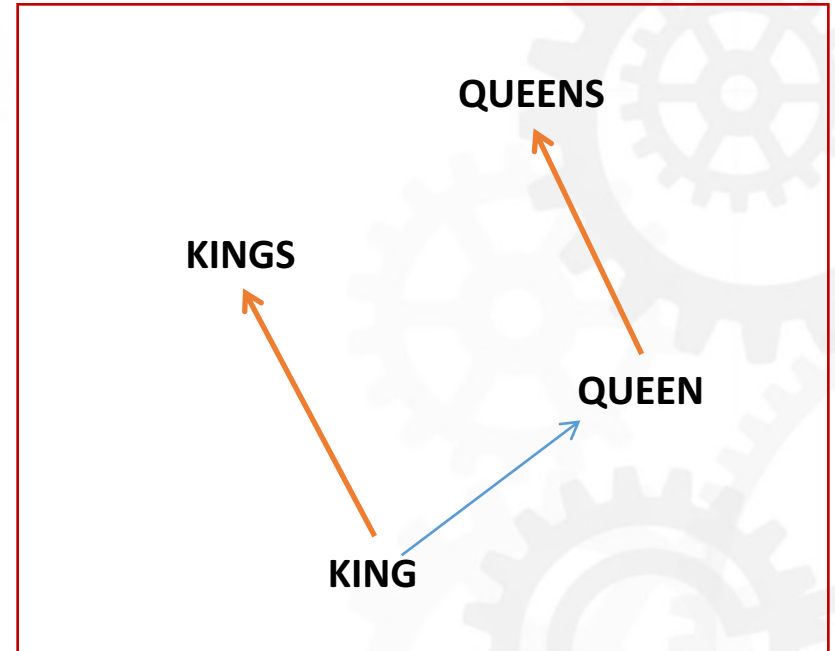


Tomas Mikolov. "Efficient Estimation of Word Representations in Vector Space". ICLR 2013

Word2Vec



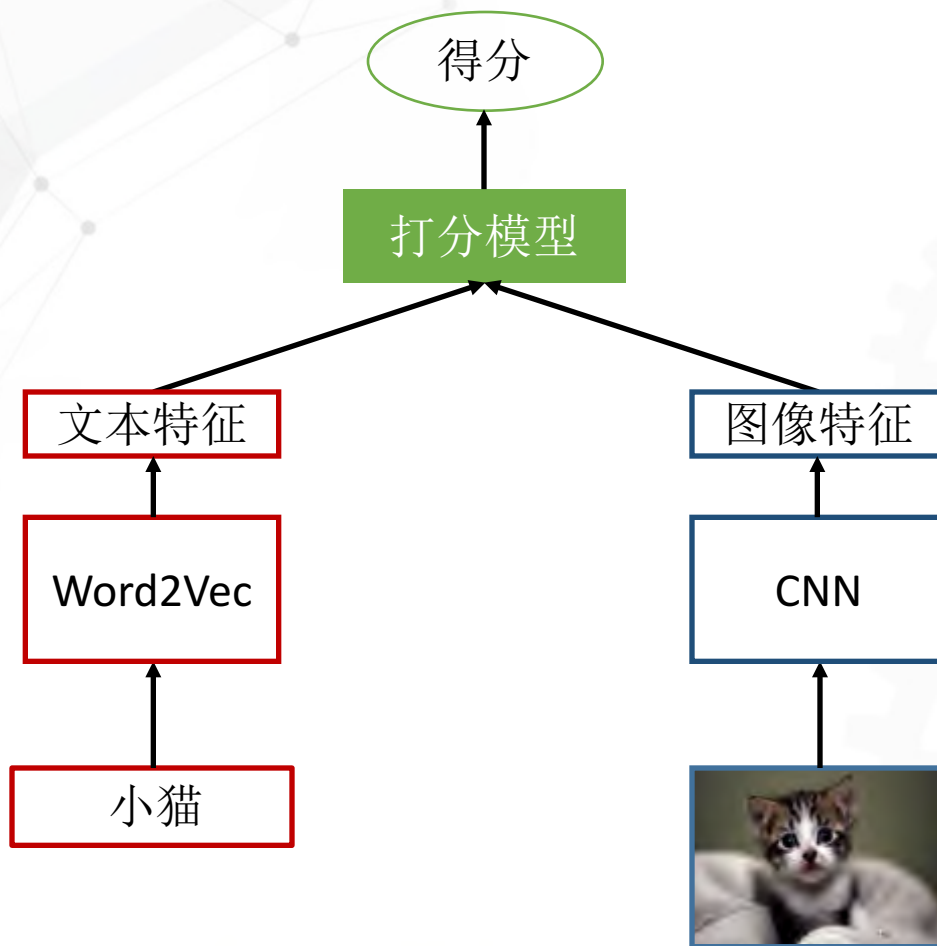
King – man + woman \approx Queen



Kings – King + Queen \approx Queens

Tomas Mikolov. "Efficient Estimation of Word Representations in Vector Space". ICLR 2013

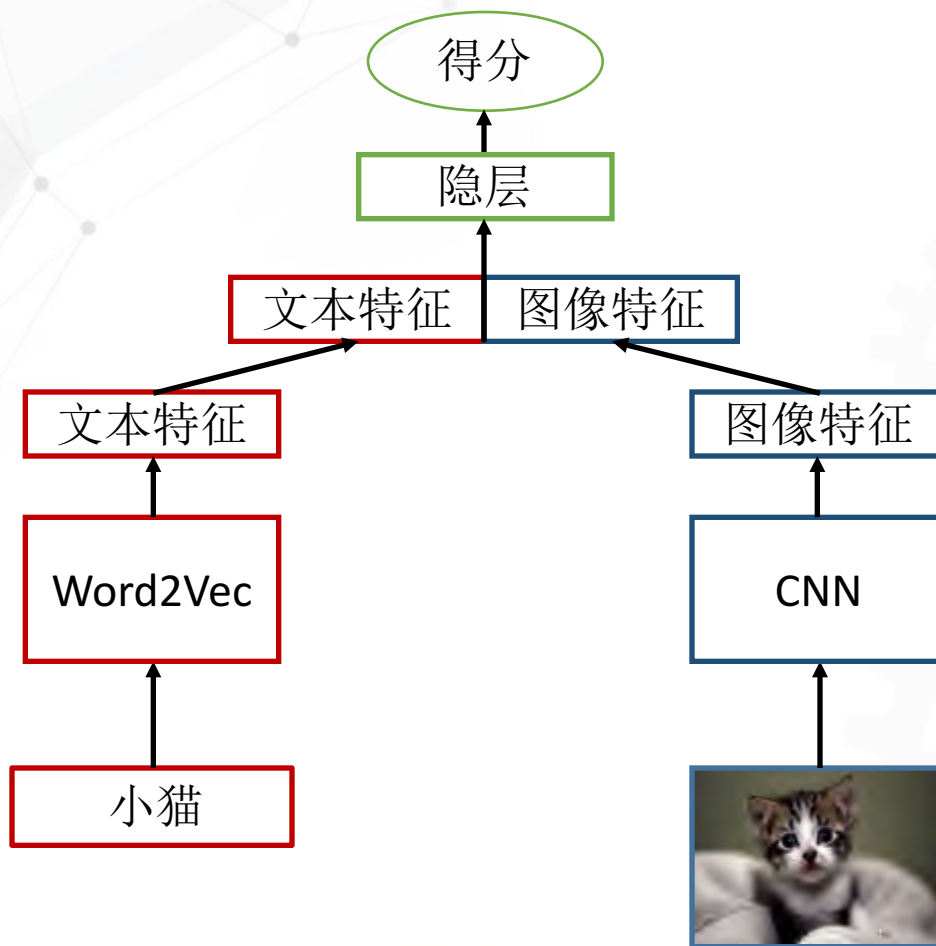
图文模型框架



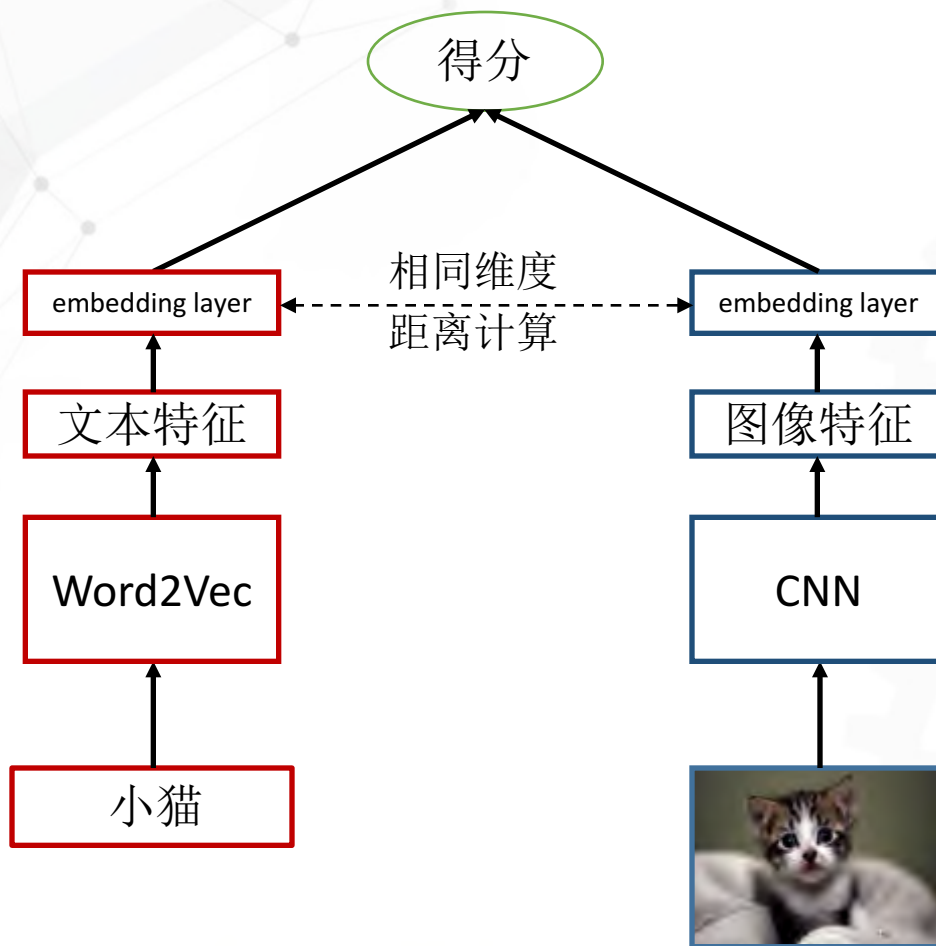
打分模型

- 特征拼接 (Concatenation)
- 特征嵌入 (Embedding)

特征拼接



特征嵌入



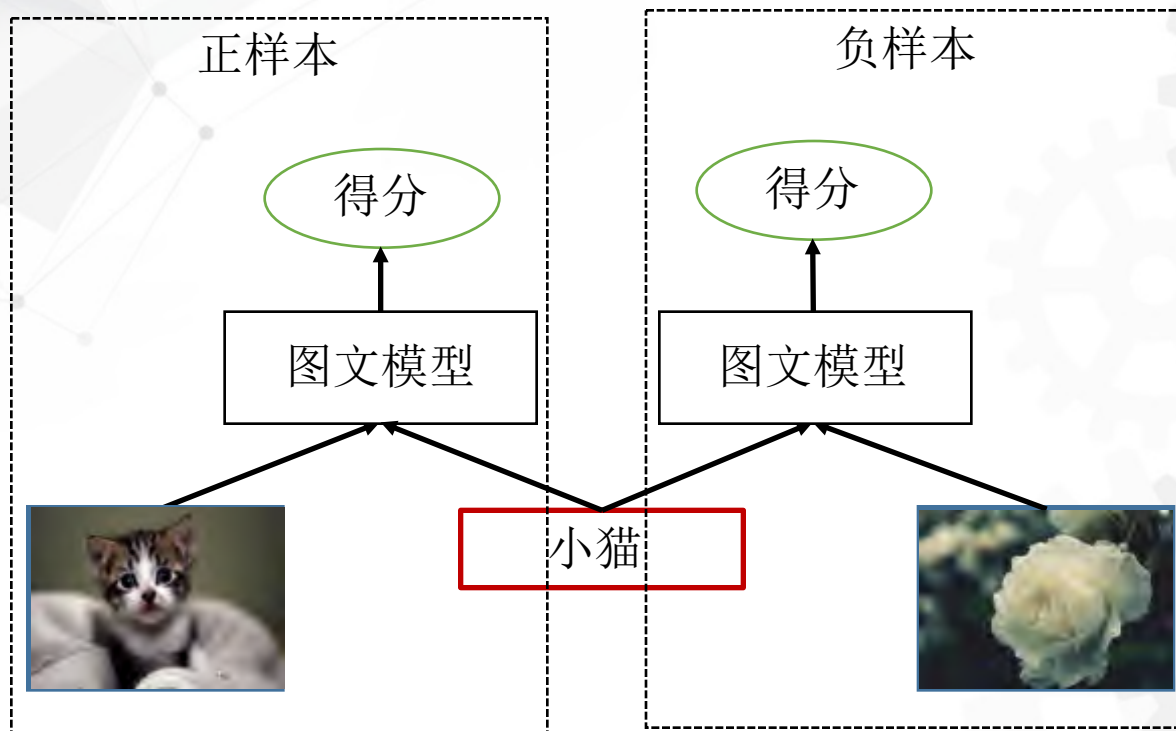
训练数据

- 来源：
 - 学术界：Flickr，MSCOCO
 - 工业界：百度图搜，Google图搜
- 形式：
 - 文本query，及对应的图片列表

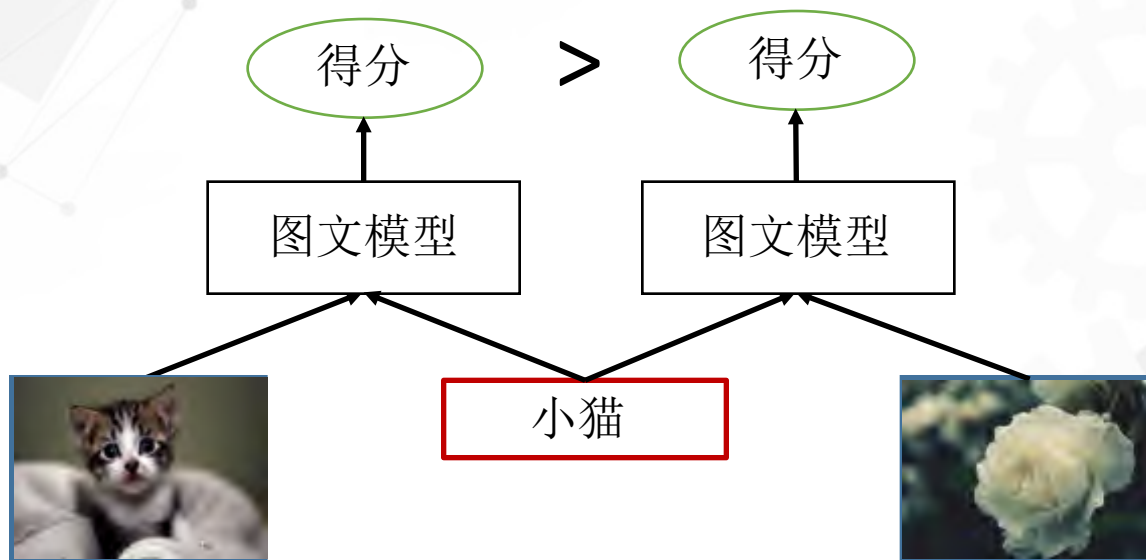
模型训练

- 分类模型
- 排序模型

分类模型



排序模型



训练平台——PaddlePaddle

- PaddlePaddle (PArallel Distributed Deep LEarning) 是百度研发的深度学习平台
 - 易用性：它为用户提供了直观且灵活的数据接口和模型定义接口
 - 灵活性：支持多种神经网络结构和优化算法，简单书写配置文件即可实现复杂模型
 - 高效性：在计算、存储、架构、通信等多方面都做了细致优化，性能优异
 - 可伸缩性：全面支持多核、多GPU、多机环境，优化的通信实现使高吞吐与高性能成为可能，轻松应对大规模数据训练需求
- <https://github.com/PaddlePaddle/Paddle>

特征嵌入+排序模型

```
1 """ 优化算法 """
2 settings(learning_rate=1e-3,
3           learning_method=AdamOptimizer(),
4           batch_size=1000)
5
6 """ 网络配置 """
7 query_feature = data_layer(name="query_feature", size=100)
8 query_embedding = fc_layer(input=query_feature, size=512)
9
10 pos_img_fea = data_layer(name="pos_img_fea", size=1000)
11 pos_img_embedding = fc_layer(input=pos_img_fea, size=512,
12                              param_attr=ParamAttr(name='img.w'),
13                              bias_attr=ParamAttr(name='img.b'))
14
15 neg_img_fea = data_layer(name="neg_img_fea", size=1000)
16 neg_img_embedding = fc_layer(input=neg_img_fea, size=512,
17                              param_attr=ParamAttr(name='img.w'),
18                              bias_attr=ParamAttr(name='img.b'))
19
20 pos_similarity = cos_sim(query_embedding, pos_img_embedding)
21 neg_similarity = cos_sim(query_embedding, neg_img_embedding)
22
23 label = data_layer(name="label", size=2)
24 outputs(rank_cost(pos_similarity, neg_similarity, label))
```

设置优化算法

query特征提取及变换

图片特征提取及变换

计算图文相似度

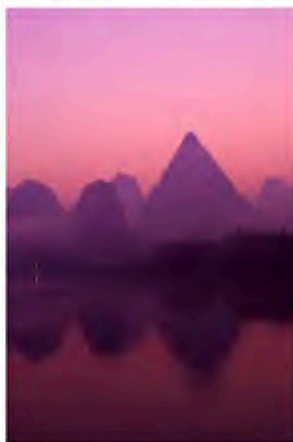
计算损失函数

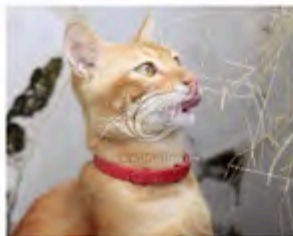
百度像素

- 百度像素是一款基于图像识别技术的正版图片搜索，管理等服务的产品。涵盖了国内主流资源方的海量高质量图片资源，是国内最大的版权图片聚合平台，设计师，运营，自媒体工作者找图必备利器。
- <http://pixel.baidu.com>

Baidu 像素

搜索关键字或试试以图搜图







THANKS

SequeMedia
威拓传媒

IT168.com

ITPUB

ChinaUnix