

# 房地产的技术变革——使用 BigDL 创建基于图像相似性分析的房屋推荐系统

原创：IDZ英特尔开发人员专区2017-12-15

本文将介绍 MLSListings\* 和英特尔在 Microsoft Azure\* 上使用 BigDL 创建的基于图像的房屋推荐系统。使用英特尔 BigDL 分布式深度学习框架，该推荐系统经过精心设计，支持高效地索引和查询数百万张房屋图像，从而显著提高房屋购买体验。用户可选择喜欢的房屋照片，让系统推荐用户可能感兴趣的带有类似图像特征的照片列表。系统设计目标为：

- 根据标题图像的特征和相似性推荐房屋。大多数图像显示房屋正面外部结构，而其他图像可显示该房屋的代表性特征。
- 
- 
- 低延迟 API，支持在线查询（< 0.1 秒）。
- 

## 背景

MLSListings Inc. 是加州北部致力于提供房地产资源的上市服务 (MLS) 机构，他们与英特尔和微软合作，将人工智能 (AI) 集成至在线地产平台，以更好地为客户服务。通过集成房地产标准机构 (RESO) API 和英特尔面向 Apache Spark\* 的 BigDL 开源深度学习库，这些技术组合在一起，可使用视觉图像显著改进房屋购买搜索过程。这一项目将为高级分析应用创新在房地产行业中的应用铺平道路。

通过图像的相似性对其进行排名，可帮助解决计算机视觉领域中的大量问题。例如，电子零售商向客户展示与过去购买的商品相似的产品，提高线上销售量。几乎每个行业都将其视作颠覆性变革，包括房地产行业，因为过去十年该行业开始越来越多地使用数字技术。在寻找合适房地产的过程中，超过 90% 的购房者选择上网搜索。房主和房地产专业人士提供房屋特征信息，比如地理位置，大小、年代，以及许多关于房屋内外部结构的照片，以供房地产列表搜索。然而，由于技术上的限制，照片中的大量信息无法提取和编入索引，从而无法增强搜索或服务房地产列表结果。事实上，为我显示类似房屋 是用户进行查询时的最大心愿。现在，利用图像数据库来增强网络和移动数字体验，通过改善搜索相关性来提高用户满意度已成为现实。

作为一种快速崛起的分布式深度学习框架，BigDL 可为大型数据社区提供简单、集成的深层学习功能。BigDL 支持各种深度学习应用，开发人员可通过 BigDL 将深度学习应用编写为标准的 Spark 程序，这些程序可以直接在现有的 Apache Spark 或 Apache Hadoop 集群上运行。

### **图像相似性概述**

在学术领域中，图像相似性可以指语义相似性或视觉相似性。语义相似性指两个图像都包含同一类别的对象。例如，平房和传统住房属于同一类别（都是房子），但看起来可能完全不同。另一方面，视觉相似性并不注重对象类别，而是从视觉角度来衡量图像的外观；例如，公寓和传统住房在外形上可能非常相似。



语义相似性：



视觉相似性：

对于语义相似性来说，通常指的是图像分类问题，使用常见的图像感知模型，比如 GoogLeNet\*或 VGG\*来有效地解决这一问题。

对于视觉相似性来说，纵观历史，有许多技术可应用于这一领域：

-

## **SIFT、SURF、颜色直方图**

传统的特征描述符可用来比较图像的相似性。SIFT 特征描述符在均匀缩放、方向和光照改变过程中始终保持不变，适用于譬如在较大的图像中查找小图像的应用。

- 
- 

## **pHash**

这一数学算法可对图像内容进行分析，并使用 64 位数字指纹来表示其内容。如果两张图像的内容特征类似，那么它们的 pHash 值也会比较接近。

- 
- 

## **基于卷积神经网络 (convnet) 的图像特征 (embedding)**

通过卷积神经网络生成图像特征编码；通常为卷积和池化后的第一个线性层。

- 
- 

## **孪生网络或 Deep Ranking**

它是一种比较深入的深度学习解决方案，但结果模型很大程度上取决于训练数据，可能会失去通用性。

- 

## **基于 BigDL 的解决方案**

为了基于图像相似性推荐房屋，我们首先将所选查询图像与候选房屋列表的标题图像进行一一比较，接下来生成每个候选房屋的相似度得分并选取排名选择最顶端的结果。通过与领域专家合作，我们开发出了以下计算房屋图像相似度的衡量方法。

For each image in the candidates, compare with query image {

class score:Both house front?(Binary Classification)

tag score:Compare important semantic tags.(Multinomial Classification)

visual score:Visually similarity score, higher is better

final Score = class score (decisive) //~1  
+ tag score (significant) //~0.3  
+ visual score // [0,1]

}

该项目同时使用了语义相似性和视觉相似性。BigDL 提供多项功能支持训练或使用图像相似性模型，包括：

- 基于 Apache Spark 和 OpenCV\* 提供实用的特征提取和转换，以支持在 Spark 上并行预处理图像。
- 
- 支持 Spark ML\* Estimator/Transformer 接口，方便用户在 Spark ML Pipeline 内执行深度学习训练和预测。
- 
- 提供方便的模型调优支持，以及支持模型 fine tuning 的灵活编程界面。
-

- 用户可将预训练的 Caffe\*、Torch\* 或 TensorFlow\* 模型加载到 BigDL 之中，以支持调优或预测。

- **语义相似性模型**

就语义相似性来说，该项目要求使用三种图像分类模型。

```
val caffeModel =  
  Module.loadCaffeModel("/googlenet_places365.prototxt_modified",  
    "/googlenet_places365.caffemodel").asInstanceOf[Graph[Float]]  
val model = Sequential().add(caffeModel)  
  .add(View(Array(1024))).add(Linear(1024, classNum)).add(LogSoftMax())  
  
val dlClassifier = new DLClassifier[Float](model,  
  criterion = new ClassNLLCriterion[Float](),  
  featureSize = Array(3, imageSize, imageSize))  
  .setBatchSize(batchSize)  
  .setLearningRate(learningRate)  
  .setOptimMethod(new Adam[Float]())  
  .setMaxEpoch(maxEpoch)  
  .setFeaturesCol("features")
```

**模型 1.图像分类：确定是否为房屋正面图片。**我们需要区分标题图像是否为房屋的正面。该模型可通过 Places\* 数据集上的预训练 GoogLeNet v1 进行调优。我们在训练过程中使用 Places 数据集。

以下代码支持在 BigDL 中借助 DLClassifier\* 进行模型训练。我们加载 Places 数据集预训练而成的 Caffe 模型，从 Caffe 模型定义中删除了最后两层（线性（1024 -> 365 和 Softmax）。然后添加了一个包含 classNum 的新线性层，以帮助训练我们所需的分类模型。

**模型 2.图像分类：房屋风格（现代，农场，传统，西班牙）。**和模型 1 相同，该模型可通过 Places 数据集上训练好的 GoogLeNet v1 模型进行调优。我们使用 MLS Listings 拥有版权的房屋图片进行训练。

**模型 3.图像分类：房屋风格（单层，两层，三层或高层）。**和模型 1 相同，该模型可通过 Places 数据集上的预训练 GoogLeNet v1 进行调优。

### 视觉相似性模型

我们需要计算视觉相似性来获得排名得分。

每次查询时，用户输入图像，与数千张候选图像进行对比，在 0.1 秒内范围前 1,000 个结果。为了满足延迟要求，我们直接对比预先计算好的图片特征。

我们首先创建评估数据集，选择最佳选项来计算图像相似性。在评估数据集中，每个记录包含 3 张图像。

三联图像（查询图像，正面图像，反面图像），其中正面图像更加类似于查询图像。

```
if (similarity(query image, positive image) > similarity(query image, negative image))
```

```
correct += 1
```

```
else
```

```
incorrect += 1
```

在上述的四种图像相似性计算方法中，孪生图像或深度排名似乎比较精确，但由于缺乏训练数据支持有意义的模型，因此不用于本项目。在评估数据集的帮助下，我们尝试采用剩余的三种方法，SIFT 和 pHash

所生成的结果准确性较低。我们怀疑是因为这两种方法都无法代表房地产图像的本质特征。

利用在 Places 数据集上的训练的深度学习模型提取图像特征编码 ( embedding ) , 可达到预期的精度和准确性水平 :

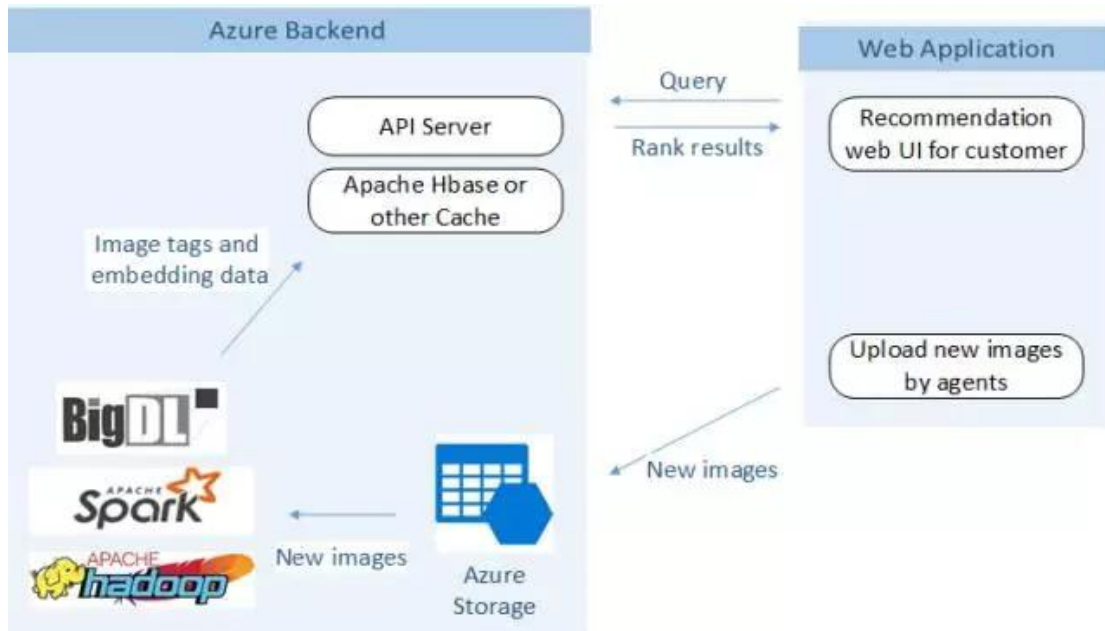
网络	特征	精度
Deepbit*	1024 二进制输出	80%
GoogLeNet*	1024 浮点数	84%
VGG-16	25088 浮点数	93%

$Similarity(m1, m2) = cosine(embedding(m1), embedding(m2))$ .

完成 L2 归一化后, 可以非常高效地计算出余弦相似性。虽然 VGG-16 嵌入有明显的优势, 但我们也尝试使用从评估数据集训练的 SVM 模型为每个嵌入特征分配不同的权重, 但这种方法所实现的改进有限, 而且我们担心 SVM 模型的通用型不足以覆盖真实世界的图像。

### 基于图像相似性的房屋推荐

完整的数据流和系统体系结构如下所示 :



在生产过程中，该项目可以分成三部分：

1.

### 模型训练（离线）

模型训练主要是指语义模型（在数据集上调优的 GoogLeNet v1），并查找相应的特征以支持视觉相似度计算。可能定期进行再训练，这取决于模型性能或要求的变化。

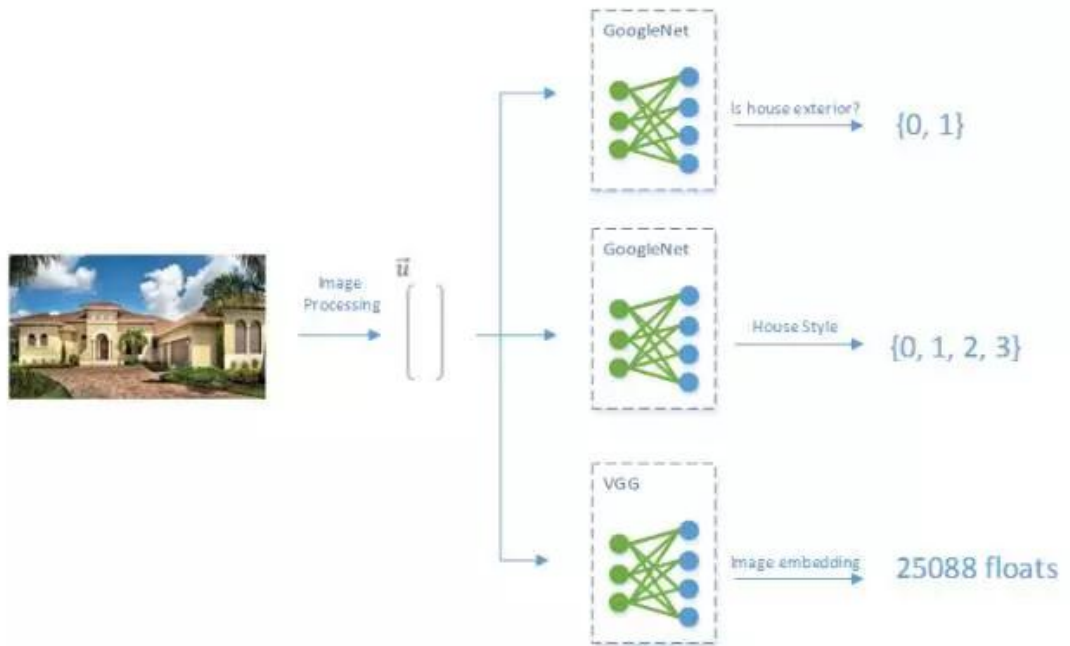
2.

3.

### 图像预测（在线）

借助第一步训练的语义模型（GoogLeNet v1）和预训练的 VGG-16，我们可以将图像转换为标记和特征，并将结果保存在键值高速缓存中。

（还可使用 Apache HBase\* 或 SQL\*）。



4.

所有现有图像和新图像都需要经过上述处理并转换为表格结构，如下所示:

lid	isExterior	styleTag	storyTag	simiFeature
6795554	1.0	4.0	3.0	[0.0,0.0,0.0,0.0,...]
6797479	1.0	2.0	1.0	[0.0,0.0,0.0,0.0,...]
6800859	1.0	2.0	3.0	[0.00122335643209...]
6802394	1.0	2.0	1.0	[0.0,0.0,0.0,0.0,...]
6807481	1.0	4.0	1.0	[0.0,0.0,0.0,0.0,...]
6817542	1.0	3.0	3.0	[0.0,0.0,0.0,0.0,...]
7920619	1.0	2.0	1.0	[0.0,0.0,0.0,0.0,...]
8009848	1.0	2.0	1.0	[0.0,0.0,0.0,0.0,...]
8011170	1.0	1.0	3.0	[0.02763030222620...]

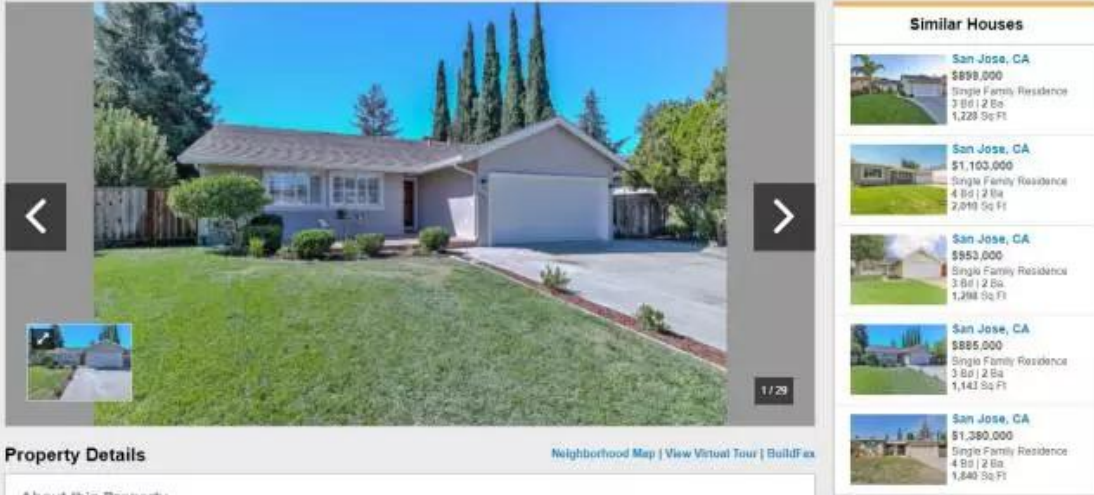
预测过程定期进行（例如一天一次），或者通过房地产列表条目上传新图像触发推理过程。每张生产图像只需要经历一次预测过程。借助索引图像标注和相似性特征，可在高并发环境中支持实现快速查询性能。

3.住宅推荐系统为向上游用户显示服务 API。每次查询都将查询图像和候选图像作为参数发送。上表显示索引图像信息后，我们可快速完成一至多项查询。关于余弦相似性的处理过程非常高效，而且可以扩展。





## 演示

我们提供从在线网站上找来的两个示例：

### 示例 1



**Similar Houses**

	San Jose, CA \$898,000 Single Family Residence 3 Bd   2 Ba 1,228 Sq Ft
	San Jose, CA \$1,103,000 Single Family Residence 4 Bd   2 Ba 2,010 Sq Ft
	San Jose, CA \$953,000 Single Family Residence 3 Bd   2 Ba 1,298 Sq Ft
	San Jose, CA \$885,000 Single Family Residence 3 Bd   2 Ba 1,143 Sq Ft
	San Jose, CA \$1,380,000 Single Family Residence 4 Bd   2 Ba 1,840 Sq Ft

### 示例 2



**Similar Houses**

	San Jose, CA \$1,000,000 Single Family Residence 2 Bd   1 1/2 Ba 1,216 Sq Ft
	San Jose, CA \$1,270,000 Single Family Residence 4 Bd   2 Ba 1,517 Sq Ft
	San Jose, CA \$835,000 Single Family Residence 4 Bd   2 Ba 1,530 Sq Ft
	San Jose, CA \$1,099,000 Single Family Residence 3 Bd   2 Ba 1,361 Sq Ft
	San Jose, CA \$799,000 Single Family Residence 3 Bd   2 Ba 1,350 Sq Ft

## 总结

本文介绍了如何在 Microsoft Azure 上集成英特尔 BigDL 深度学习库，从而帮助 MLSListings 构建基于图像分析的房屋推荐系统。我们通过基于 caffe 模型进行 fine tuning 生成了三个图片分类模型，以从房地产图像中提取重要的语义标记。我们进一步对比了不同的视觉相似度计算方法，发现通过 VGG 提取的图像特征最为有效。作为端到端行业示例，我们演示了如何利用基于 BigDL 的深度学习，以将更多图像识别创新应用到房地产行业之中。