

2018 AIoT+智慧城市峰会

BATJ巨头齐聚 · 11个精彩内容分享 · 共同探讨万物智联和智慧城市的现实应用

最新文章

清华大学韩旭：神经关系抽取模型 | AI研习社71期大讲堂

吸引 7198 支队伍参赛，看 Kaggle 信用预估比赛冠军方案

百度视觉团队斩获 ECCV Go AI 目标检测广冠冠军，获奖全解读 | ECCV 2018

大学王延森：如何利用拼增强开放领域对话系统互动性 | AI研习社66期大讲堂

如何从静态图像中识别“比心”动作

ICPR 图像识别与检测挑战赛方案出炉，基于偏旁部首来识Duang 字

雷锋网

读懂智能&未来

首页

AI研习社

AI影响因子

活动

专题

精选

爱搞机

申请专栏作者

业界

人工智能

智能驾驶

AI+

Fintech&区块链

未来医疗

网络安全

AR/VR

机器人

开发者

AI开发

正文

## 一文详解 Word2vec 之 Skip-Gram 模型（结构篇）

本文作者：AI研习社

2017-06-22 12:06

导语：这可能是关于 Skip-Gram 模型最详细的讲解。

雷锋网按：这是一个关于 Skip-Gram 模型的系列教程，依次分为结构、训练和实现三个部分，本文为第一部分：结构篇，后面两个部分我们将陆续发布，敬请期待。原文作者天雨粟，原载于作者知乎专栏，雷锋网已获授权。

这次的分享主要是对Word2Vec模型的两篇英文文档的翻译、理解和整合，这两篇英文文档都是介绍Word2Vec中的Skip-Gram模型。下一专栏文章将会用TensorFlow实现基础版Word2Vec的skip-gram模型，所以本篇文章先做一个理论铺垫。

原文英文文档请参考链接：

- Word2Vec Tutorial - The Skip-Gram Model

<http://t.cn/Rc5RfJ2>
- Word2Vec (Part 1): NLP With Deep Learning with Tensorflow (Skip-gram)

<http://t.cn/RoVEiUB>

### 什么是Word2Vec和Embeddings？

Word2Vec是从大量文本语料中以无监督的方式学习语义知识的一种模型，它被大量地用在自然语言处理（NLP）中。那么它是如何帮助我们做自然语言处理呢？Word2Vec其实就是通过学习文本来用词向量的方式表征词的语义信息，即通过一个嵌入空间使得语义上相似的单词在该空间内距离很近。Embedding其实就是一个映射，将单词从原先所属的空间映射到新的多维空间中，也就是把原先词所在空间嵌入到一个新的空间中去。

我们从直观角度上来理解一下，cat这个单词和kitten属于语义上很相近的词，而dog和kitten则不是那么相近，iphone这个单词和kitten的语义就差的更远了。通过对词汇表中单词进行这种数值表示方式的学习（也就是将单词转换为词向量），能够让我们基于这样的数值进行向量化的操作从而得到一些有趣的结论。比如说，如果我们对词向量kitten、cat以及dog执行这样的操作： $kitten - cat + dog$ ，那么最终得到的嵌入向量（embedded vector）将与puppy这个词向量十分相近。

#### 热门搜索

Oculus

4G

众筹

智慧城市

无线充电

隐私

苏宁

Win1

创客马拉松

AI

TC Disrupt

开源硬件

AI研习社

编辑

聚焦数据科学，连接AI开发者。

发私信

当月热门文章

猿桌会 | 人机交互技术探索

职博间 | 从产品经理到产品经理中的多角色学习 & 复旦大学NLP实验室介绍

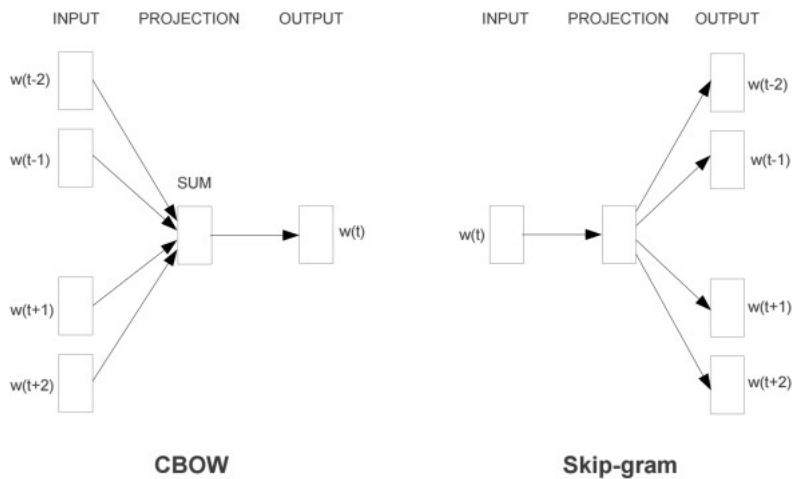
大讲堂 | 深度强化学习在电商推荐中的应用

猿桌会 | 从1到100进阶三部曲——关键知识、模型性能提升、产品落地

20:00大讲堂 | 机器学习在百度人岗匹配中的应用

## 模型

Word2Vec模型中，主要有Skip-Gram和CBOW两种模型，从直观上理解，Skip-Gram是给定input word来预测上下文。而CBOW是给定上下文，来预测input word。本篇文章仅讲解Skip-Gram模型。



Skip-Gram模型的基础形式非常简单，为了更清楚地解释模型，我们先从最一般的基础模型来看Word2Vec（下文中所有的Word2Vec都是指Skip-Gram模型）。

Word2Vec模型实际上分为了两个部分，**第一部分为建立模型，第二部分是通过模型获取嵌入词向量。** Word2Vec的整个建模过程实际上与自编码器（auto-encoder）的思想很相似，即先基于训练数据构建一个神经网络，当这个模型训练好以后，我们并不会用这个训练好的模型处理新的任务，我们真正需要的是这个模型通过训练数据所学得的参数，例如隐层的权重矩阵——后面我们将会看到这些权重在Word2Vec中实际上就是我们试图去学习的“word vectors”。基于训练数据建模的过程，我们给它一个名字叫“Fake Task”，意味着建模并不是我们最终的目的。

上面提到的这种方法实际上会在无监督特征学习（unsupervised feature learning）中见到，最常见的就是自编码器（auto-encoder）：通过在隐层将输入进行编码压缩，继而在输出层将数据解码恢复初始状态，训练完成后，我们会将输出层“砍掉”，仅保留隐层。

## The Fake Task

我们在上面提到，训练模型的真正目的是获得模型基于训练数据学得的隐层权重。为了得到这些权重，我们首先要构建一个完整的神经网络作为我们的“Fake Task”，后面再返回来看通过“Fake Task”我们如何间接地得到这些词向量。

接下来我们来看看如何训练我们的神经网络。假如我们有一个句子**“The dog barked at the mailman”**。

- 首先我们选句子中间的一个词作为我们的输入词，例如我们选取“dog”作为input word；
- 有了input word以后，我们再定义一个叫做skip\_window的参数，它代表着我们从当前input word的一侧（左边或右边）选取词的数量。如果我们设置skip\_window=2，那么我们最终获得窗口中的词（包括input word在内）就是**['The', 'dog', 'barked', 'at']**。skip\_window=2代表着选取左input word左侧2个词和右侧2个词进入我们的窗口，所以整个窗口大小span=2x2=4。另一个参数叫num\_skips，它代表着我们从整个窗口中选取多少个不同的词作为我们的output word，当skip\_window=2，num\_skips=2时，我们将会得到两组**(input word, output word)**形式的训练数据，即**(‘dog’, ‘barked’)**，**(‘dog’, ‘the’)**。
- 神经网络基于这些训练数据将会输出一个概率分布，这个概率代表着我们的词典中的每个词是output word的可能性。这句话有点绕，我们来看个栗子。第二步中我们在设置skip\_window和

### 最新文章

清华大学韩旭：神经关系抽取型 | AI研习社71期大讲堂

吸引 7198 支队伍参赛，看 Kagggle 信用预估比赛冠军方

百度视觉团队斩获 ECCV Go AI 目标检测竞赛冠军，获奖全解读 | ECCV 2018

清华大学王延森：如何利用增强开放领域对话系统互动性 AI研习社66期大讲堂

如何从静态图像中识别“比心”动作

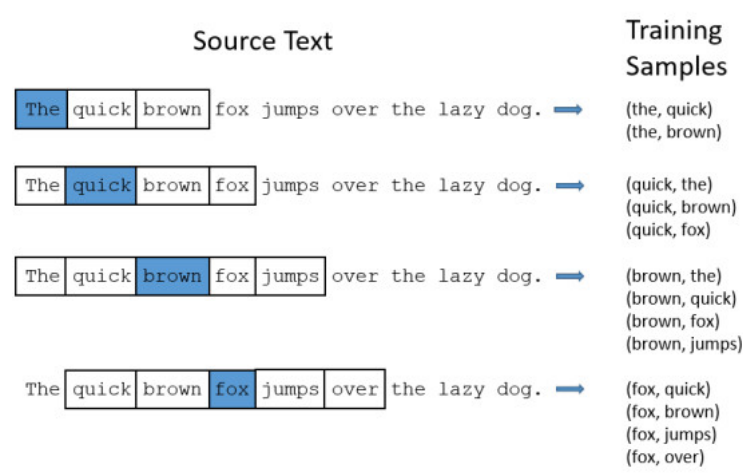
ICPR 图像识别与检测挑战赛方案出炉，基于偏旁部首来识Duang 字

### 热门搜索

- Oculus4G众筹
- 智慧城市无线充电
- 隐私苏宁Win1
- 创客马拉松TC Disrupt
- 开源硬件

num\_skips=2的情况下获得了两组训练数据。假如我们先拿一组数据 ('dog', 'barked') 来训练神经网络，那么模型通过学习这个训练样本，会告诉我们词汇表中每个单词是“barked”的概率大小。

模型的输出概率代表着到我们词典中每个词有多大可能性跟input word同时出现。举个栗子，如果我们向神经网络模型中输入一个单词“Soviet”，那么最终模型的输出概率中，像“Union”，“Russia”这种相关词的概率将远高于像“watermelon”，“kangaroo”非相关词的概率。因为“Union”，“Russia”在文本中更大可能在“Soviet”的窗口中出现。我们将通过给神经网络输入文本中成对的单词来训练它完成上面所说的概率计算。下面的图中给出了一些我们的训练样本的例子。我们选定句子“The quick brown fox jumps over lazy dog”，设定我们的窗口大小为2（window\_size=2），也就是说我们仅选输入词前后各两个词和输入词进行组合。下图中，蓝色代表input word，方框内代表位于窗口内的单词。



我们的模型将会从每对单词出现的次数中习得统计结果。例如，我们的神经网络可能会得到更多类似（“Soviet”，“Union”）这样的训练样本对，而对于（“Soviet”，“Sasquatch”）这样的组合却看到的很少。因此，当我们的模型完成训练后，给定一个单词“Soviet”作为输入，输出的结果中“Union”或者“Russia”要比“Sasquatch”被赋予更高的概率。

## 模型细节

我们如何来表示这些单词呢？首先，我们都知道神经网络只能接受数值输入，我们不可能把一个单词字符串作为输入，因此我们得想个办法来表示这些单词。最常用的办法就是基于训练文档来构建我们自己的词汇表（vocabulary）再对单词进行one-hot编码。

假设从我们的训练文档中抽取10000个唯一不重复的单词组成词汇表。我们对这10000个单词进行one-hot编码，得到的每个单词都是一个10000维的向量，向量每个维度的值只有0或者1，假如单词ants在词汇表中的出现位置为第3个，那么ants的向量就是一个第三维度取值为1，其他维都为0的10000维的向量（ants=[0, 0, 1, 0, ..., 0]）。

还是上面的例子，“The dog barked at the mailman”，那么我们基于这个句子，可以构建一个大小为5的词汇表（忽略大小写和标点符号）：("the", "dog", "barked", "at", "mailman")，我们对这个词汇表的单词进行编号0-4。那么“dog”就可以被表示为一个5维向量[0, 1, 0, 0, 0]。

模型的输入如果为一个10000维的向量，那么输出也是一个10000维度（词汇表的大小）的向量，它包含了10000个概率，每一个概率代表着当前词是输入样本中output word的概率大小。

下图是我们神经网络的结构：

最新文章

清华大学韩旭：神经关系抽取型 | AI研习社71期大讲堂

吸引 7198 支队伍参赛，看 Kaggle 信用预估比赛冠军方案

百度视觉团队斩获 ECCV Go AI 目标检测竞赛冠军，获奖全解读 | ECCV 2018

清华大学王延森：如何利用增强开放领域对话系统互动性 | AI研习社66期大讲堂

如何从静态图像中识别“比心”动作

ICPR 图像识别与检测挑战赛方案出炉，基于偏旁部首来识 Duang 字

热门搜索

Oculus

4G

众筹

智慧城市

无线充电

隐私

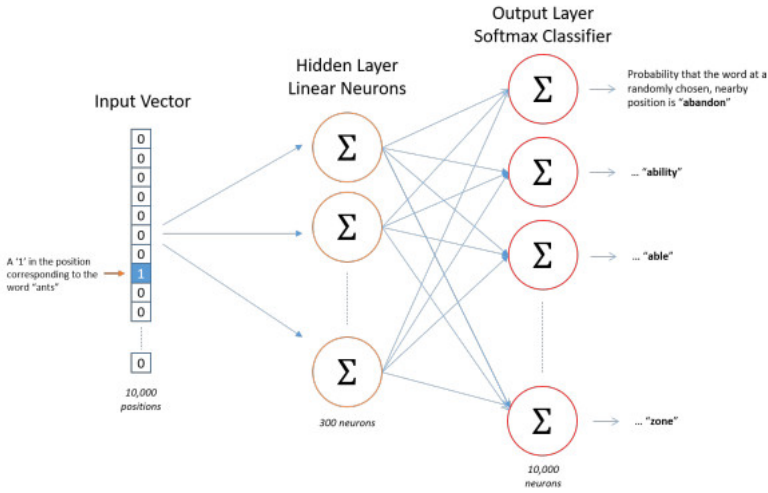
苏宁

Win1

创客马拉松

TC Disrupt

开源硬件



隐层没有使用任何激活函数，但是输出层使用了softmax。

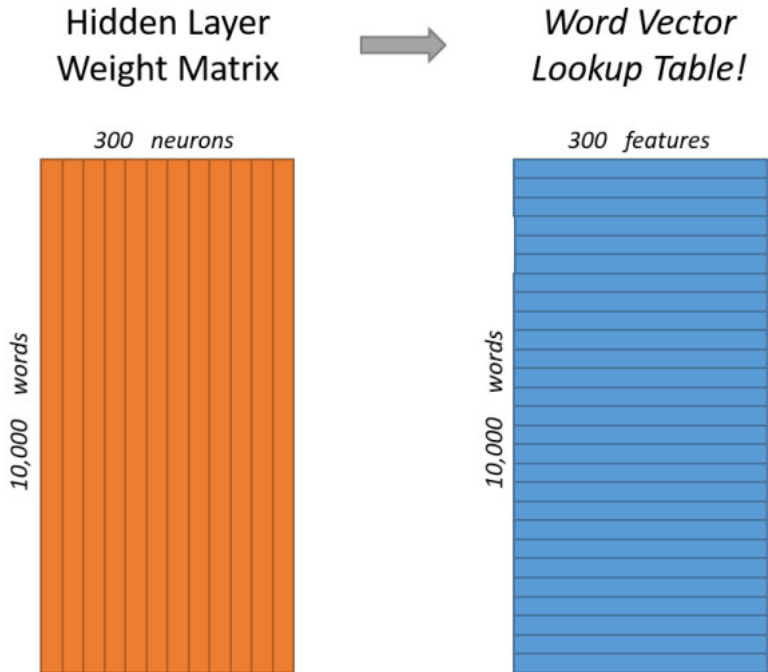
我们基于成对的单词来对神经网络进行训练，训练样本是（input word, output word）这样的单词对，input word和output word都是one-hot编码的向量。最终模型的输出是一个概率分布。

### 隐层

说完单词的编码和训练样本的选取，我们来看下我们的隐层。如果我们现在想用300个特征来表示一个单词（即每个词可以被表示为300维的向量）。那么隐层的权重矩阵应该为10000行，300列（隐层有300个结点）。

Google在最新发布的基于Google news数据集训练的模型中使用的就是300个特征的词向量。词向量的维度是一个可以调节的超参数（在Python的gensim包中封装的Word2Vec接口默认的词向量大小为100，window\_size为5）。

看下面的图片，左右两张图分别从不同角度代表了输入层-隐层的权重矩阵。左图中每一列代表一个10000维的词向量和隐层单个神经元连接的权重向量。从右边的图来看，每一行实际上代表了每个单词的词向量。



所以我们最终的目标就是学习这个隐层的权重矩阵。

我们现在回来接着通过模型的定义来训练我们的这个模型。

最新文章

清华大学韩旭：神经关系抽取模型 | AI研习社71期大讲堂

吸引 7198 支队伍参赛，看 Kaggle 信用预估比赛冠军方案

百度视觉团队斩获 ECCV Go AI 目标检测竞赛冠军，获奖方案全解读 | ECCV 2018

清华大学王延森：如何利用增强开放域对话系统互动性 | AI研习社66期大讲堂

如何从静态图像中识别“比心”动作

ICPR 图像识别与检测挑战赛方案出炉，基于偏旁部首来识 Duang 字

热门搜索

Oculus

4G

众筹

智慧城市

无线充电

隐私

苏宁

Win1

创客马拉松

TC Disrupt

开源硬件



上面我们提到，input word和output word都会被我们进行one-hot编码。仔细想一下，我们的输入被one-hot编码以后大多数维度上都是0（实际上仅有一个位置为1），所以这个向量相当稀疏，那么会造成什么结果呢。如果我们将一个1 x 10000的向量和10000 x 300的矩阵相乘，它会消耗相当大的计算资源，为了高效计算，它仅仅会选择矩阵中对应的向量中维度值为1的索引行（这句话很绕），看图就明白。

$$\begin{bmatrix} 0 & 0 & 0 & 1 & 0 \end{bmatrix} \times \begin{bmatrix} 17 & 24 & 1 \\ 23 & 5 & 7 \\ 4 & 6 & 13 \\ 10 & 12 & 19 \\ 11 & 18 & 25 \end{bmatrix} = \begin{bmatrix} 10 & 12 & 19 \end{bmatrix}$$

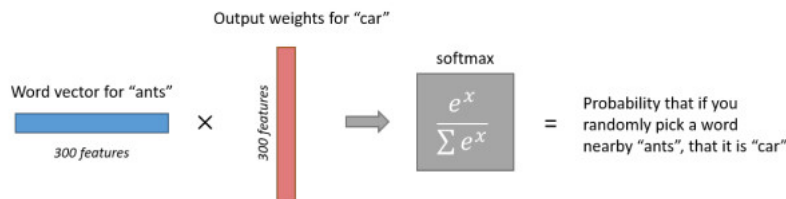
我们来看一下上图中的矩阵运算，左边分别是1 x 5和5 x 3的矩阵，结果应该是1 x 3的矩阵，按照矩阵乘法的规则，结果的第一行第一列元素为0 x 17 + 0 x 23 + 0 x 4 + 1 x 10 + 0 x 11 = 10，同理可得其余两个元素为12，19。如果10000个维度的矩阵采用这样的计算方式是十分低效的。

为了有效地进行计算，这种稀疏状态下不会进行矩阵乘法计算，可以看到矩阵的计算的结果实际上是矩阵对应的向量中值为1的索引，上面的例子中，左边向量中取值为1的对应维度为3（下标从0开始），那么计算结果就是矩阵的第3行（下标从0开始）—— [10, 12, 19]，这样模型中的隐层权重矩阵便成了一个“查找表”（lookup table），进行矩阵计算时，直接去查输入向量中取值为1的维度下对应的那些权重值。隐层的输出就是每个输入单词的“嵌入词向量”。

## 输出层

经过神经网络隐层的计算，ants这个词会从一个1 x 10000的向量变成1 x 300的向量，再被输入到输出层。输出层是一个softmax回归分类器，它的每个结点将会输出一个0-1之间的值（概率），这些所有输出层神经元结点的概率之和为1。

下面是一个例子，训练样本为 (input word: “ants”， output word: “car”) 的计算示意图。



## 直觉上的理解

下面我们将通过直觉来进行一些思考。

如果两个不同的单词有着非常相似的“上下文”（也就是窗口单词很相似，比如“Kitty climbed the tree”和“Cat climbed the tree”），那么通过我们的模型训练，这两个单词的嵌入向量将非常相似。

那么两个单词拥有相似的“上下文”到底是什么含义呢？比如对于同义词“intelligent”和“smart”，我们觉得这两个单词应该拥有相同的“上下文”。而例如“engine”和“transmission”这样相关的词语，可能也拥有着相似的上下文。

实际上，这种方法实际上也可以帮助你进行词干化（stemming），例如，神经网络对“ant”和“ants”两个单词会学得相似的词向量。

词干化（stemming）就是去除词缀得到词根的过程。

雷锋网(公众号：雷锋网)相关阅读：

一文详解 Word2vec 之 Skip-Gram 模型（训练篇）

### 最新文章

清华大学韩旭：神经关系抽取型 | AI研习社71期大讲堂

吸引 7198 支队伍参赛，看 Kaggle 信用预估比赛冠军方

百度视觉团队斩获 ECCV Go AI 目标检测竞赛冠军，获奖全解读 | ECCV 2018

清华大学王延森：如何利用增强开放领域对话系统互动性 | AI研习社66期大讲堂

如何从静态图像中识别“比心”动作

ICPR 图像识别与检测挑战赛方案出炉，基于偏旁部首来识 Duang 字

### 热门搜索

Oculus

4G

众筹

智慧城市

无线充电

隐私

苏宁

Win1

创客马拉松

TC Disrupt

开源硬件

一文详解 Word2vec 之 Skip-Gram 模型（实现篇）

25 行 Python 代码实现人脸检测——OpenCV 技术教程

雷锋网原创文章，未经授权禁止转载。详情见[转载须知](#)。

15人收藏

分享：

相关文章

Skip-GramWord2VecCBOW



一文详解 Word2vec 之 Skip-Gram 模型（实现



一文详解 Word2vec 之 Skip-Gram 模型（训练



大讲堂 | 计算机辅助词典编纂——以异体字词典为例



ImageNet 带来的预训练模型之风，马上就要吹进 NLP

文章点评：

我有话要说.....

☐ 同步到新浪微博

提交

最新评论

 小黑在此 10月09日 22:05

这三篇写的太好了，高产啊！！好想打赏你，太棒了

(0) 回复

最新文章

- 清华大学韩旭：神经关系抽取型 | AI研习社71期大讲堂
- 吸引 7198 支队伍参赛，看 Kaggle 信用预估比赛冠军方
- 百度视觉团队斩获 ECCV Go AI 目标检测竞赛冠军，获奖全解读 | ECCV 2018
- 清华大学王延森：如何利用增强开放领域对话系统互动性 | AI研习社66期大讲堂
- 如何从静态图像中识别“比心”动作
- ICPR 图像识别与检测挑战赛方案出炉，基于偏旁部首来识 Duang 字

热门搜索

- Oculus4G众筹
- 智慧城市无线充电
- 隐私苏宁Win1
- 创客马拉松TC Disrupt
- 开源硬件

腾讯优图贾佳亚在“腾讯·云+未来” AI大数据专场分享：计算机视觉有哪三种打开方式？

本文作者：奕欣 2017-06-22 12:04

导语：贾佳亚在今日“腾讯·云+未来” AI大数据专场所做的主题演讲《计算机视觉前沿与应用》，雷锋网AI科技评论对速记做了不改动原意的编辑和整理。

腾讯优图实验室杰出科学家贾佳亚，香港中文大学计算机科学工程系终身教授，于 2017 年 5 月 15 日公布消息，全职加入腾讯优图实验室，负责计算机视觉、图像处理、模式识别等人工智能领域的研究,及人工智能与各应用场景结合的深度探索。

雷锋网了解到，贾佳亚在加入腾讯后鲜少露面，本次在“腾讯云+未来” AI大数据专场做主题演讲，也是为数不多能一窥腾讯优图实验室及研究成果的公开场合。以下是贾佳亚在今日“腾讯云+未来” AI大数据



奕欣  
初心者

专场所做的主题演讲《计算机视觉前沿与应用》，雷锋网(公众号：雷锋网)AI科技评论对速记做了不改动原意的编辑和整理。



谢谢大家，刚才蒋杰和王龙都介绍了腾讯在整个腾讯云的部署中，其硬件和它的基础构架上的优势，我觉得这是一个非常重要的部分。

我今天给大家讲讲在 AI 的算法层面，我们能做什么事情。腾讯优图团队在世界上是做图像识别算法非常有优势的，我们有很强的研发团队，在深圳、香港、上海都有自己的研发队伍。这是腾讯这几年变化最大的地方，也就是我们会把新的技术、新的能力、新的硬件、新的软件全部提供给大家，希望给各位有需求的公司或者是业务人员有更加强大的助力。

计算机视觉是什么？它其实就是一种人的理解，我们看到东西的时候，不仅仅是看到红绿蓝三种颜色，这是我们最基本的对颜色的理解。但是当我们看一张图象的时候，我们不会分离看每一种颜色，我们会把它看成一个整体。这个图是我非常喜欢的，但是我想找到是谁创造的，在网上很难找到是谁发明的图片，这张图片是一张非常完美的正脸和侧脸的结合，从一个角度看，你可以说这是一张正脸，你也可以说是一个侧脸。

什么叫计算机视觉？计算机视觉的目标是让计算机程序能够解读和理解图片，不仅仅是颜色，而是能够更高层的理解它的语义、理解它的特征，从表面来看这个事情是很简单的，事实上很难。1983 年，华盛顿第一次召开计算机视觉大会（雷锋网 AI 科技评论按：即 CVPR，国际计算机视觉与模式识别会议），到现在已经超过了 30 年的历史。在这 30 多年中，对我们现在很多从业者，或者在工业界的朋友而言，其实他们并不了解这 30 年里研究者们到底干了什么事情，有什么发生了。但是到了云的时代，我们会发现能更加直接地对云的技术进行展示，我们也能更加多地接触到用户使用这种技术。

我们怎么看待这个世界？首先我们要理解计算机视觉为什么要在 1983 年开始？你想想 1983 年的时候大家在用什么样的计算机？我之前在别的地方进行过一个讲座，我问大家还记不记得 2000 年的时候用的是用什么样的计算机。那时候是没有 LED 和 LCD 的，那时候是用的很笨重的、很大的显示器，那时候用的手机是诺基亚，没有彩色屏幕，全部是一个单独的小机器，上面一个很小的屏幕。但在 1983 年的时候，我们这个领域开始建立起来的时候，那时候连计算机的基本能力都不具备，但是当时很多的科学家已经在开始想象，我们能不能在卫星图像上，比如说当时在军事运用上，或者是在一些非常高精尖的应用上，在卫星或者是载人飞船上面，能不能利用一些机器帮助人们解决问题，所以那时候就开始了这方面的研究。

这个研究是一种类比，我们之所以研究对视觉的理解过程，是因为我们人眼就是一种自然智能，我们能看到东西，当我们看到东西之后，我们就会自然想象，我能不能让一个机器也看到东西，这是很自然的。当

发私信

当月热门文章

吸引 7198 支队伍参赛，看 Kagle 信用预估比赛冠军方

最新文章

IEEE Fellow 又增一位华人学

中科院自动化所王亮研究员当

云从科技 OCR 新突破：端到

的深度学习文本检测框架 Pix

Anchor

李飞飞又有新动向，斯坦福 /

验室由 Christopher Mannir

接棒

怎样免费拿机械键盘、AI研

定制双肩包和保温杯？邀请组

学习「CMU 深度学习课程」

行！

不一样的论文解读：2018 KI

best paper「Embeddings i

Airbnb」

MSRA 成立 20 周年，北大、

华两校校长同台为其致辞

热门搜索

创客马拉松

TC Disrupt

Twitter

Apple

激光雷达

vivo

马化腾

新能源

ImageNet

锤子手机

奥迪

Alphabet

Disrupt 20

1



然我们现在能隔空打物，我也想设计一个新的方案让机器也隔空打物，显然我们现在还不行，所以也不会让机器也这样。

在大脑皮层里，超过 50% 的神经原细胞是处理视觉的，所以我们说计算机视觉是最重要的方向，它代表了信息的多样性和可用的信息。

计算机视觉有三个打开方式，**第一个是语义理解、识别、检测**。当我们看到一张图的时候，我们能不能像人一样分析这张图里有什么东西。比如说现在这样一个在开车的人，当然这明明是一个人，只是他戴了一个头套，计算机检测到这里面有一个狼，然后它有蓝色的眼睛，同时还在开车，这是一个理解过程、检测过程，这就是语义的理解、识别、检测，这是非常重要的一个计算机视觉的内容，在腾讯云上以后也会陆续输出这些部分。



其次是不受想象力约束的神奇效果。前几年有一个电影是讲你在一个梦境里的变化，就是你不停地下楼梯，每次下到最后都回到原点。这个事情我们作为科学家来研究到底是什么原理，我们用软件重构了这样一个系统，我们确实可以在视觉模式上把它搭建起来，虽然在语义逻辑上不是很好理解的，但是我们在视觉上确实可以体现出来，最终产生很酷炫的效果，比如说这个球，它还可以在这上面跳来跳去，跳到最后它就回到原点了，你看他每次都在下台阶，结果却回到了原来的起点，这是一个特殊的神奇效果，在视觉上可以做的事情。

还有一些固有需求的图像视频计算，比如说我们需要通过一张图像把一些最重要的内容提取出来。

第一个打开方式，理解识别。理解识别的应用是非常广泛的，这张图是 2006—2012 年我们这个领域里的一个非常大的竞赛，这个竞赛叫做 PASCAL，这个竞赛是给你超过 1000 张图像，希望你从这几千张图像里面找出其中 20 个物体的类别。从 2009 年到 2012 年，在这个领域里有超过 20 种方法解决这个问题，它的准确率不断提高，但是它始终是有瓶颈的，因为类别很少，我们发现大家对这个比赛的认可度也就慢慢降低了。

后来出现了这个领域里最出名的一个比赛叫做 IMAGE NET，这是斯坦福大学的几个教授发起的比赛，它推翻了原来所有一切的竞赛规则，他说现在在整个数据库里包含了超过 14000 万张图像，相比以前的几千张或者几万张图像，这是一个巨大的进步，在量级上是完全不可同日而语的，同时它有超过 2 万多种类别，你告诉我这张图象是属于 2 万个类别的哪一个，这一个非常大的匹配过程和检测过程，所以 IMAGE NET 推翻了当时对数据的理解，以前大家专注于做小数据，因为大家觉得反正我没有大数据，不如就做小数据，当你在学术界真正的把大数据放出来的时候，是有更多的研究人员愿意去跟上的。

这张图是告诉大家，我们从 2010 年到 2014 年的时候，开始在这个数据库上做事情，最开始 2010 年的时候，大家发现我们的错误率是 28%，也就是说有 100 张图像里面，有 28 张图像是分错的，就是说它

最新文章

清华大学韩旭：神经关系抽取型 | AI研习社71期大讲堂

吸引 7198 支队伍参赛，看 Kaggle 信用预估比赛冠军方

百度视觉团队斩获 ECCV Go AI 目标检测竞赛冠军，获奖全解读 | ECCV 2018

清华大学王延森：如何利用增强开放领域对话系统互动性 AI研习社66期大讲堂

如何从静态图像中识别“比心”动作

ICPR 图像识别与检测挑战赛方案出炉，基于偏旁部首来识 Duang 字

热门搜索

- Oculus4G众筹
- 智慧城市无线充电
- 隐私苏宁Win1
- 创客马拉松TC Disrupt
- 开源硬件



是找不到的，这个准确率已经是一个比较可以的准确率，大家会觉得，可能放在一些不太重要的岗位上或者应用上的时候就可以用了。但是到了 2011 年的时候，我们把它提高了 2%，也就是说现在可以多两张图像，能做得更好了。到 2012 年的时候，我们发现多了 10%，也就是说在 100 张图像里面多 10 张图片可以分析准确了。因为那一年出现了深度学习，它验证自己在整个领域里面，到底能不能推动 AI 的进程，就是通过这样一个竞赛，通过在那一年时间，把准确率提高 10% 这样一次事件，导致对于世界上所有的领域、所有的人，业界和学术界开始意识到，我们可以把这个事情做得更好。到现在可以做到 100 张图里只有一两张图是错误的，这个识别率远远高过人的识别率。人眼看一张图，你能分辨 2 万个种类吗？这是很难的事情。

另外一个就是检测，大家看到检测已经慢慢的变成这个领域的核心的内容或者是应用方向。大家可以想像以后我们的智能家居是什么样的，以后智能家居重要的应用可能是冰箱，你打开冰箱的时候，你会发现里面堆满了各种各样的东西，有苹果、梨、蔬菜、肉类，现在我们想进入智能家居的第一步，是不是说我打开一个冰箱，让电脑自动知道里面还有多少库存，如果库存不够的时候，我们要不要上京东去购买一些，或者我要去其它的电商买一些东西回来，这是一个很有趣的应用，我相信很多人或者是机构都在研发这些技术，其中一个重要的组成部分，在算法层面上，就是对物体的检测，也就是说我们希望发现中间到底这个是苹果还是西红柿，或者是其他的什么东西，它有多大，数量有多少，它能不能够用一个礼拜，或者说只够用 3 天，或者说你现在就必须买，否则明天就没饭吃了，这种状况下，我们希望通过一个检测来发现。

再往下走就是在最新的研究上，我希望能够分析到更加细颗粒度的图象识别，这个问题就对于智能驾驶、辅助驾驶，或者是大规模的城市理解和建设有着巨大的推动作用，因为在这样一个城市级的道路复杂环境下，你会看到每个东西都在动，上面一张图在下面会分割成不同的部分，机器会告诉你，我识别了这个车，识别了这个树，我也识别了这个电线杆和路，在智能驾驶或者辅助驾驶里需要有这样的颗粒度，越精细越好，以后希望通过计算机视觉帮助我们在自动驾驶这个新兴产业，这个市场规模可能是有几万亿美元，在这样的市场下，怎么体现我们在其中的技术实力，这是一个重要的部分。

**第二个打开方式：新视觉效果。**我给大家介绍一些好玩的东西。这是很早的一个连续剧，它当时有一个片头，就是这个人慢慢的从一张素描变成了一个状态，这个过场在当时是非常酷炫的。这是艺术家做出来的，他请了专门的人帮你画了一张素描的图，然后把这个图贴到视频里面慢慢的做转换，现在的 AI 已经自动实现这种功能了，比如说看到非常漂亮的自然场景的时候，我们团队研发的 AI 技术可以自动的产生非常有趣的一些效果，能够把它变成铅笔画、水彩画、油画，甚至是各种抽象的画法，我们在 2011 年的时候已经有这方面一系列的论文介绍这件事情，这个现在变成了一个可控的东西，每一个客户想去使用这样的功能的时候，通过我们的云，通过我们这样一些技术的扩展方向，我们可以把这件事情很容易的做到，以往你可能需要请一个艺术家坐在边上，大概花一天或者两天时间帮你设计这样东西，今天你只要打开电脑，接上我们的网络，用上我们这样一个 API 或者 SDK，然后你在一秒钟内就得到这样一个结果，这是一个非常大的进步和进展，这也是为什么在技术层面上很多东西可以用的，也就是用得更加舒服，比人的操作来得更加方便和直接。

这是另外一个例子，我们当时做的时候有一个初衷，我们看这张表，大家都有做图表的经验，当你做了这个表格很漂亮的时候，结果打印出来发现并没有做的时候这么好的效果，原因是我们在电脑做出来的效果很好，但是打印机打出来的颜色并没有那么炫。所以我们在前几年花了一两个月的时间做了一个非常小的工具，这个工具就是当我看到这样一张彩色图的时候，我能把它自动转换成一个黑白图象，这样大家要打印这样一张彩色图象的时候，你可以打印出一张非常漂亮的黑白图象，这样你就可以用一个普通的黑白打印机，打印出一个效果上可以媲美于彩色打印机的功能，当然这个功能我们当年是想推销给各种各样的打印机公司的，好像打印机公司也没有兴趣，最后我们就没有用上。

这是另外一个例子，我们经常会画饼状图，在 PPT 里画的很好看，结果打印出来给老板观察的时候发现分不出具体的颜色的细节。我们就通过一个算法把它直接变成这样一张黑白打印的效果，最后出来的时候，我们可以在 29 毫秒之内，把这张图转变成可以接受的效果，这也是一个技术的进步，当然我们做这个事情也就花了大概一个月不到的时间，这是我觉得有趣的地方，计算机视觉永远可以产生一些新的大家想象不到的效果。

最新文章

清华大学韩旭：神经关系抽取模型 | AI研习社71期大讲堂

吸引 7198 支队伍参赛，看 Kaggle 信用预估比赛冠军方

百度视觉团队斩获 ECCV Go AI 目标检测竞赛冠军，获奖全解读 | ECCV 2018

清华大学王延森：如何利用拼增强开放领域对话系统互动性 | AI研习社66期大讲堂

如何从静态图像中识别“比心”动作

ICPR 图像识别与检测挑战赛方案出炉，基于偏旁部首来识 Duang 字

热门搜索

- Oculus4G众筹
- 智慧城市无线充电
- 隐私苏宁Win1
- 创客马拉松TC Disrupt
- 开源硬件

**第三个打开方式：固有需求的图像视频计算。**比如说我要做视频，我希望做前景、背景分割，我需要做手势识别，我需要做人的操作，在图像里面做一些逆运算，比如说去模糊，我看到这样一张图，比如说你跟领导站在台上拍了一张照片，或者是在你领奖的时候，或者是好不容易见到一个你想见的朋友，大家在一块儿开心的照了一张照片，结果照模糊了，这时候你就面临一个非常尴尬的境地，因为你不可能再回到原来的场地再拍一张照片，所以我们当时在设计算法的时候就在想，能不能通过这张图回复其中的固有信息呢？我们通过一个非常高级的计算过程，可以把中间所存在的一些字或者是一些重要的信息重构出来，这样的结果是可以通过我们现在的算法完整实现的。

这是另外一个例子，这个图是什么东西，估计在座的没有几位能看得清楚，但是我们通过算法可以告诉你，这是一张城市方向的图像，我们甚至可以看得出来上面是「多伦多」，是在这个城市拍摄的图像。

还有一个是强迫透视。大家如果喜欢摄影，你会发现这个很有意思，这是一双筷子，好像是在夹光点，其实是在拍照的时候特意布置的场景。甚至我们可以做这样一个场景，前面有一个人在吹，后面的人全倒的状况，实际上前景、后景是有深度分别的，这也是在计算机运算里面非常有趣的效果。我们在前两年的时候开始研究这样一些效果上面，是不是背后有自己的原因，我们找找看它的几何的原因，我们分析对一张完全清晰的图像，我们也可以分析各个边缘的细小的分别，通过这些细小的分别，我们可以得到一张图像以后，然后我们可以在后期再去把它变成单反效果，可以把背景虚化，把前景凸显出来，甚至我们可以切换，把这个模糊的地方换得不一样，把聚焦点放在别的地方。我知道在所有的相机厂商采用这个方法之前，我们在学术界里面已经有了一个非常大的研究，我们在这上面是有一系列论文在解决这个问题的，当然现在也有一些厂商是在用我们这个技术，在实现自己的一个后期的单反效果的增强。

这是其中的一个例子，比如说在一张图像上，我先拍一张图像，后期在这个手机上我可以重新做虚化，显示哪个地方是你想要的，哪个地方是你不想看的，这也是技术发展的一些有趣的进展。每到一个时刻，我们总会产生一些新的大家想象不到的事情，这些都会自然而然的产生。

我们优图团队实验室着力于在做一些最前沿的研发，我们希望把视觉、自然语言处理和语音这样三个大的方向结合起来，在语音这一块，我们在小微的专场里面还有另外一个同事会介绍我们在语音合成、人声分离等等一系列的技术上的进展，所以大家有兴趣可以在小微专场看到我们的另外一个介绍。

**在视觉这一块，新的视觉体验、识别和视觉内容计算，这是我们不可分割的三个巨大的方向。**在接下来可能会有更多的方向，但是在这几个方向上，我们要做的事情，或者是我们的能力是能够得到充分体现的。其实我加入腾讯的时间并不长，但是我觉得我们整个优图团队在腾讯里面所产生的价值，因为有了云的部署，我们优图的能力会扩展得更加快一些。以后如果是我们的客户，一定会接触到更多更有趣的效果和应用。

更多资讯请持续关注雷锋网。

雷锋网原创文章，未经授权禁止转载。详情见[转载须知](#)。

5人收藏

分享：

1

相关文章

人工智能

腾讯AI Lab

计算机视觉



国内首例！乐普医疗人工智能心电产品获美国FDA批准



全球权威人脸识别竞赛成绩公布，中国人工智能算法继



机器学习教父 Tom Mitchell：人工智能将成为



义学教育松鼠AI联合创始人周伟：网络教育发展的六次

最新文章

清华大学韩旭：神经关系抽取型 | AI研习社71期大讲堂

吸引 7198 支队伍参赛，看 Kagggle 信用预估比赛冠军方

百度视觉团队斩获 ECCV Go AI 目标检测竞赛冠军，获奖全解读 | ECCV 2018

清华大学王延森：如何利用拼增强开放领域对话系统互动性 | AI研习社66期大讲堂

如何从静态图像中识别“比心”动作

ICPR 图像识别与检测挑战赛方案出炉，基于偏旁部首来识 Duang 字

热门搜索

Oculus

4G

众筹

智慧城市

无线充电

隐私

苏宁

Win1

创客马拉松

TC Disrupt

开源硬件

https://www.leiphone.com/news/201706/PamWKpfRFEI42Mcl.html

10/11



ApolloScape ECCV 2018  
挑战赛报名即将截止，三大



全球首个AI商用同传驻馆  
中关村国家示范区添“黑科



用深度学习技术，让你的眼睛  
可以控制电脑



银行业中的数据挖掘-贷款  
审批案例

文章点评：

我有话要说.....

☐ 同步到新浪微博

提交

最新文章

- 清华大学韩旭：神经关系抽取模型 | AI研习社71期大讲堂
- 吸引 7198 支队伍参赛，看 Kagggle 信用预估比赛冠军方
- 百度视觉团队斩获 ECCV Go AI 目标检测竞赛冠军，获奖全解读 | ECCV 2018
- 清华大学王延森：如何利用拟增强开放领域对话系统互动性 | AI研习社66期大讲堂
- 如何从静态图像中识别“比心”动作
- ICPR 图像识别与检测挑战赛方案出炉，基于偏旁部首来识 Duang 字

热门搜索

- Oculus4G众筹
- 智慧城市无线充电
- 隐私智苏宁智Win10
- 创客马拉松诗词TC Disrupt开源硬件Xperia z5 compact

热门关键字

热门标签 人工智能 机器人 机器学习 深度学习 金融科技 未来医疗 智能驾驶 自动驾驶 计算机视觉 激光雷达 图像识别 智能音箱 区块链 智能投顾 医学影像 物联网 IoT 微信小程序平台 微信小程序在哪 CES 2017 CES 2016年最值得购买的智能硬件 2016 互联网 小程序 微信朋友圈 抢票软件 智能手机 智能家居 360智能硬件 智能摄像机 智能硬件产品 智能硬件发展 智能硬件创业 黑客 白帽子 大数据 云计算 新能源汽车 无人驾驶 无人机 大疆 小米无人机 特斯拉 VR游戏 VR电影 VR视频 VR眼镜 VR购物 AR 直播 扫地机器人 医疗机器人 工业机器人 类人机器人 聊天机器人 微信机器人 微信小程序 移动支付 支付宝 创客马拉松 诗词 TC Disrupt 人脸识别 指纹识别 黑科技 谷歌地图 谷歌 IBM 微软 乐视 百度 三星s8 腾讯 三星Note8 小米MIX 小米Note 华为 小米 阿里巴巴 苹果 MacBook Pro iPhone Face ID GAIR IROS 双创周 云栖大会 优蒯 智能硬件公司 智能硬件 QQ红包 支付宝红包 敬业福 tensorflow ddos攻击 emoji表情代码 tile 大飞机 开源硬件 Xperia z5 compact 谷歌员工福利 黑客大战 停车场车位锁 vr手套 针孔录像机 卫星宽带 世界最大的黄色网站 马丁背包 更多

联系我们 关于我们 加入我们 意见反馈 投稿