# TensorFlow Hub：TensorFlow中可重用机器学习模块库

发布者，TensorFlow的开发倡导者

在软件开发中，最基本的一点是共享代码库的概念，这一点很容易被忽略。作为程序员，库立即使我们更有效。从某种意义上说，它们改变了编程的问题解决过程。在使用库时，我们经常考虑使用可以绑定在一起的构建块或模块进行编程。

图书馆如何寻找机器学习开发人员？当然，除了共享代码外，我们还希望共享预训练的模型。共享一个预先训练好的模型使开发人员可以为自己的域定制它，而不必访问计算资源或用于训练模型的原始数据。例如，培训花费了数千GPU小时。通过共享学习到的权重，模型开发人员可以使其他人更容易重用和构建他们的工作。

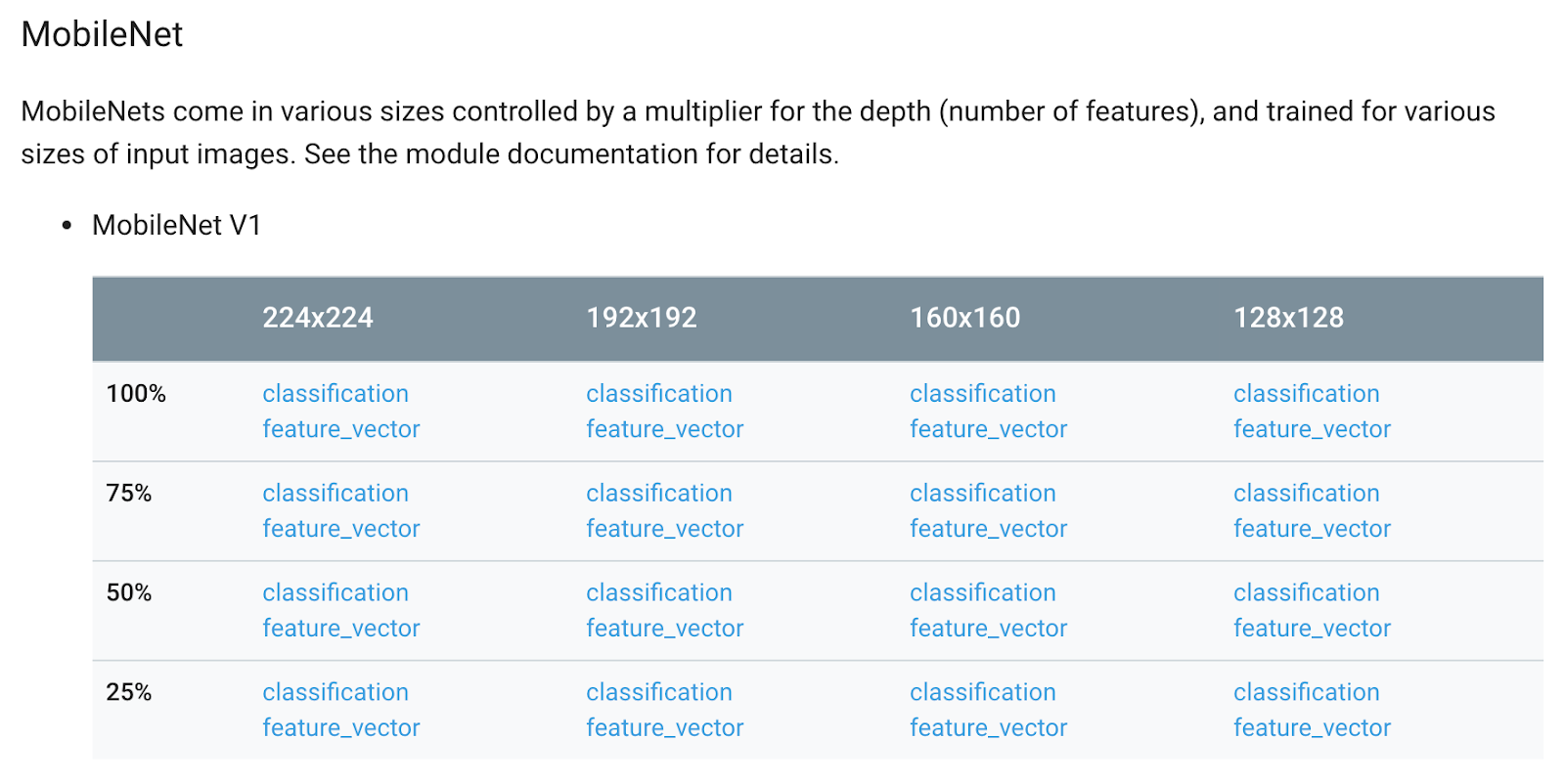


这是一个为机器学习开发人员提供的库的想法，它给了我们灵感，今天我们很高兴与社区分享它。TensorFlow Hub是一个发布、发现和重用TensorFlow中机器学习模块的平台。我们所说的模块，是指一个TensorFlow图的自包含部分及其权重，可以在其他类似的任务中重用。通过重用模块，开发人员可以使用较小的数据集训练模型，改进泛化，或者简单地加快训练速度。让我们看看几个具体的例子。

### 图像再训练

作为第一个例子，让我们看看一种可以用来训练图像分类器的技术，从少量的训练数据开始。现代图像识别模型有数百万个参数，而从头开始训练一个模型需要大量的标记数据和计算能力。使用名为的技术，可以使用更少的数据量和更少的计算时间来训练模型。以下是TensorFlow集线器的外观。

基本的想法是重用现有的图像识别模块从图像中提取特征，然后在这些图像上训练一个新的分类器。如上所示，在构造TensorFlow图时，可以从URL（或从文件系统路径）实例化TensorFlow集线器模块。有各种各样的on TensorFlow Hub供您选择，包括各种风格的NASNet（包括其最新的V2）、Inception、ResNet和其他。要使用模块，可以使用TensorFlow Hub，然后将模块的URL复制/粘贴到代码中。



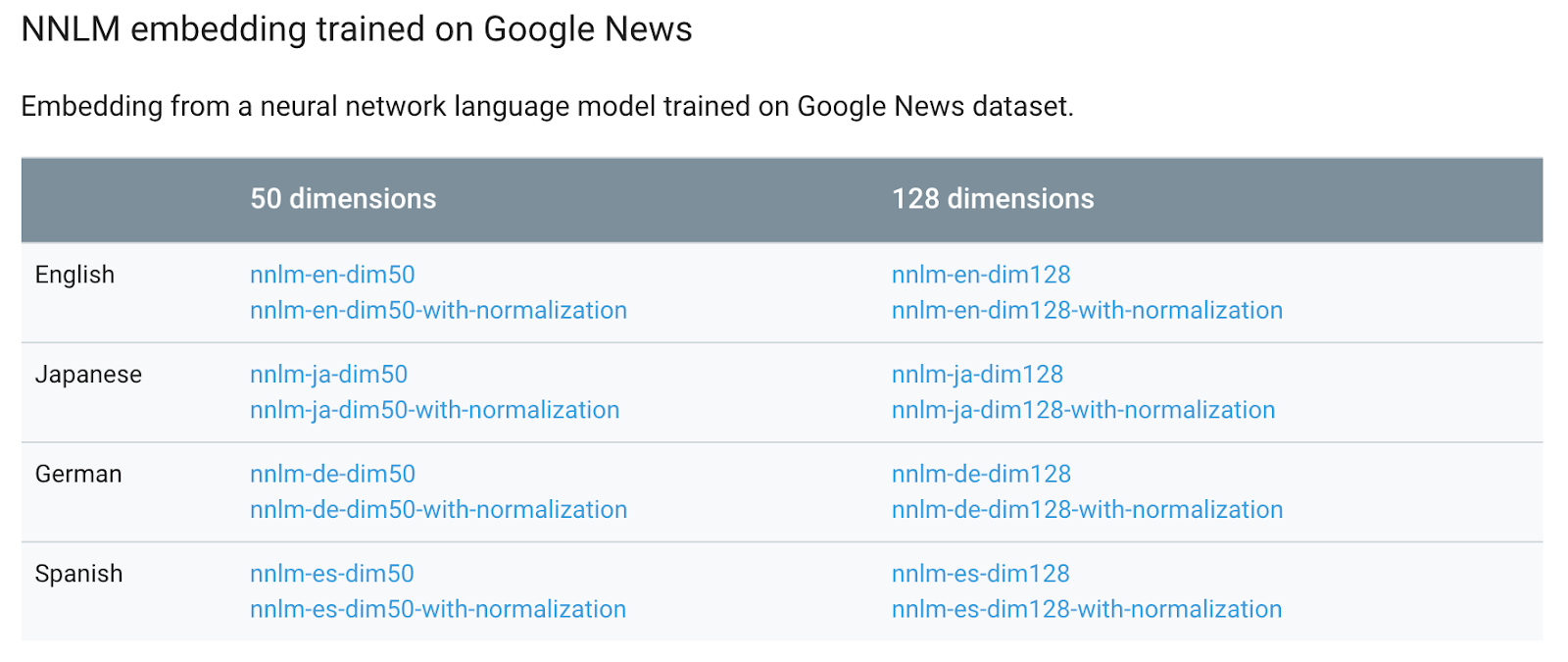
每个模块都有一个定义好的接口，允许以可替换的方式使用它，而对它的内部结构知之甚少或一无所知。在这种情况下，此模块有一个方法，您可以使用该方法来检索预期的图像大小。作为开发人员，您只需要提供一批形状正确的图像，并调用它们上的模块来检索特征表示。此模块负责为您预处理图像，因此您可以在一个步骤中直接从一批图像转到一个特征表示。从这里，您可以学习线性模型，或其他类型的分类器，在这些之上。

在本例中，请注意我们使用的模块是由Google托管的，并且是版本控制的（因此您可以依赖于在进行实验时不会更改的模块）。模块可以像普通的Python函数一样应用来构建图形的一部分。一旦导出到磁盘，模块是自包含的，可以被其他人使用，而无需访问用于创建和训练模块的代码和数据（当然，您也可以发布这些代码和数据）。

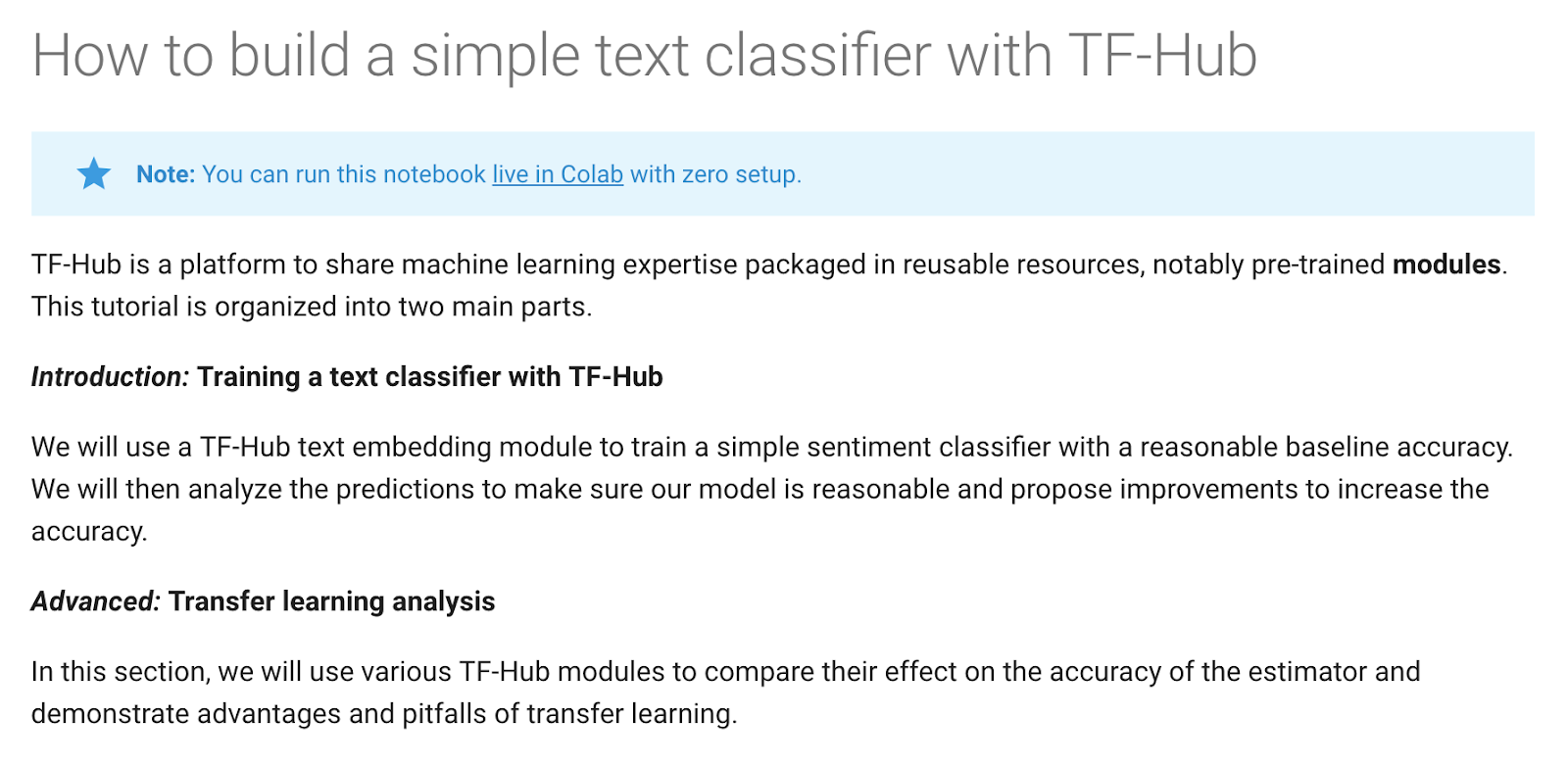
### 文本分类

让我们看看第二个例子。假设你想训练一个模型，将电影评论分为正面和负面，从少量的训练数据开始（比如，几百篇正面和负面的电影评论）。由于您的示例数量有限，因此决定利用以前在更大的语料库上训练过的单词嵌入数据集。下面是使用TensorFlow集线器时的外观。

和以前一样，我们从选择一个开始。TensorFlow Hub有多种文本模块供您探索，包括各种语言（EN、JP、DE和ES）中的神经网络语言模型，以及在Wikipedia上训练的Word2vec和在Google News上训练的NNLM嵌入。



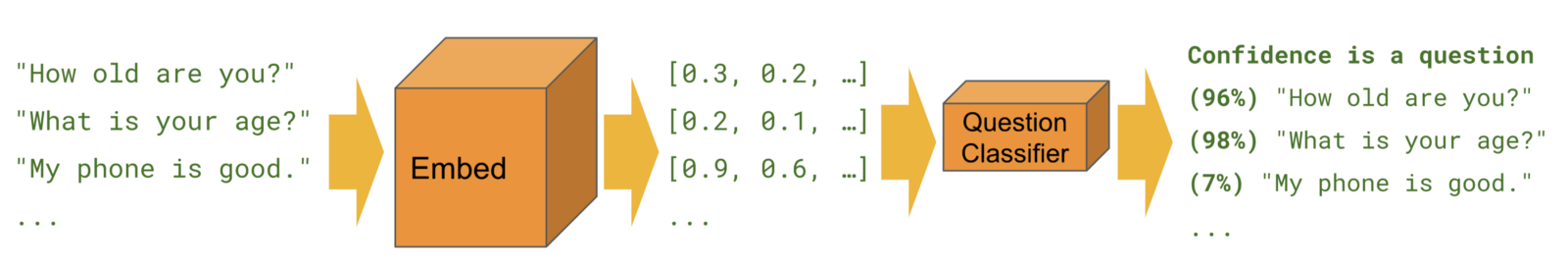
在本例中，我们将使用一个用于字嵌入的模块。上面的代码下载一个模块，用它来预处理一个句子，然后检索每个令牌的嵌入。这意味着您可以在一个步骤中将数据集中的句子直接转换为适合分类器的格式。模块负责将句子标记化，以及处理词汇表外单词等其他逻辑。预处理逻辑和嵌入都封装在一个模块中，这样就可以更容易地试验各种word嵌入数据集或不同的预处理策略，而无需实质性地更改代码。



如果你想试试这个方法，可以用它来进行旋转，并学习TensorFlow集线器模块如何与TensorFlow估计器一起工作。

### 通用句子编码器

我们还共享了一个TensorFlow集线器模块，以获得新的功能！下面是一个使用通用句子编码器的示例。它是一个句子级的嵌入模块，在各种数据集上进行训练（换句话说，“通用”）。它擅长于语义相似性、自定义文本分类和聚类。



在图像再训练中，为了使模块适应您自己的任务，需要相对较少的标记数据。例如，让我们在餐厅评论中试试。

查看此以了解更多信息。

### 其他模块

TensorFlow集线器不仅仅是图像和文本分类。在网站上，你还可以找到一对为进步甘和夫妇。

### 注意事项

使用TensorFlow集线器模块时有两个重要的注意事项。首先，记住模块包含可运行代码。始终使用来自受信任源的模块。第二，就像所有的机器学习一样，是一个需要考虑的问题。我们上面展示的两个示例都利用了大型的预训练数据集。当重用这样的数据集时，重要的是要注意它包含的数据（以及是否存在任何存在的偏差），以及这些可能如何影响正在构建的产品及其用户。

### Next steps

We hope you find TensorFlow Hub useful in your projects! To get started, head to [tensorflow.org/hub](http://tensorflow.org/hub). If you run into any bugs, you can file an [issue on GitHub](https://github.com/tensorflow/hub/issues). To stay in touch, you can star the [GitHub project](https://github.com/tensorflow/hub), and follow TensorFlow on [Twitter](http://twitter.com/tensorflow). Thanks for reading!