

BigDL : 运行在Apache Spark上的分布式深度学习类库

近日，Intel开源了基于Apache Spark的分布式深度学习框架BigDL。有了BigDL之后，用户可以像编写标准的Spark程序一样来编写深度学习（deep learning）应用程序，编写完的程序还可以直接运行在现有的Spark或者Hadoop集群之上。BigDL主要有以下三大特点：

[BigDL GitHub地址](#)

丰富的深度学习算法支持

BigDL的设计吸取了Torch框架许多方面的知识，为深度学习提供了全面的支持；包括数值计算（numeric computing）和高级神经网络（high level neural networks）；除此之外，用户可以使用BigDL将事先训练好的Caffe 或 Torch模型加载到Spark程序中。

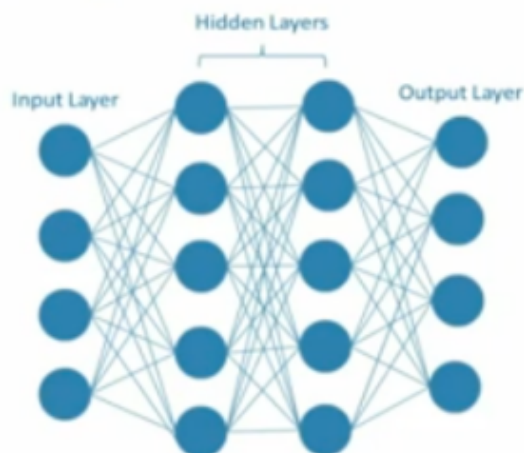
极高的性能

为了实现极高的性能，BigDL使用了Intel MKL和在每个Spark task中进行了多线程编程。所以，BigDL比其他开源框架Caffe、Torch和TensorFlow有显著的性能提升。BigDL速度可与主流GPU匹敌，而且能够扩展到数十个Xeon服务器。

强大的扩展能力

BigDL 可以很高效地进行规模扩展并执行大规模的数据。通过利用Spark计算框架，以及同步SGD（随机梯度下降，Stochastic gradient descent）的高效实现，全面减少了在Spark上的通信。（这段话翻译的可能不太准确，英文原句：BigDL can efficiently scale out to perform data analytics at "Big Data scale", by leveraging Apache Spark (a lightning fast distributed data processing framework), as well as efficient implementations of synchronous SGD and all-reduce communications on Spark.)

Deep Learning



如果想及时了解Spark、Hadoop或者Hbase相关的文章，欢迎关注微信公共帐号：iteblog_hadoop

为什么选择BigDL

如果有以下的场景，你很可能会使用BigDL来编写你的深度学习项目：

1、你的数据存储存储在HDFS, HBase, Hive上，并想使用同一大数据集群（Hadoop/Spark）来分析数据；

2、在Spark程序中/工作流中加入深度学习功能；

3、利用现有的 Hadoop/Spark 集群来运行深度学习程序，然后将代码与其他的应用场景进行动态共享，例如ETL、数据仓库（data warehouse）、特征工程（feature engineering）、经典机器学习（classical machine learning）、图表分析（graph analytics）等。

如何使用BigDL

你可以到[这里](#)学习如何安装和构建BigDL（Linux 和 macOS）；

你可以到[这里](#)学习如何运行BigDL应用程序，其中包括了本地Java程序和Spark程序；更多关于BigDL的使用，可以参见官方 [wiki](#)

本博客文章除特别声明，全部都是原创！
原创文章版权归过往记忆大数据（[过往记忆](#)）所有，未经许可不得转载。
本文链接：[【】（）](#)