

# DL4J DEEPLARNING4J

(../cn/index.html)

深度学习教科书

(<https://www.amazon.com/Deep-Learning-Practitioners-Adam-Gibson/dp/1491914254>)

快速入门 >

教程 >

深度学习介绍 >

神经网络 >

数据与ETL >

模型训练和调参 >

部署 >

开源社区 >

自然语言处理 >

## Deeplearning4j中的Doc2Vec与段落向量

Doc2Vec的主要作用是在任意文档与标签之间建立联系，所以标签是必需的。Doc2Vec是Word2Vec的一项扩展，而Word2Vec的功能也是学习词与标签之间的关联，而非词与其他词之间的关联。Deeplearning4j中的Doc2Vec实现旨在服务Java、Scala (scala.html)和Clojure社区。

我们首先要获得表示一份文档的“意义”的向量，然后再将这个向量输入有监督的机器学习算法，确定与文档有关的标签。

在ParagraphVectors构建器模式中，`labels()`方法指明了定型中涉及到的标签。以下的示例中给出了与情感分析相关的标签：

```
.labels(Arrays.asList("negative", "neutral", "positive"))
```

以下是一个用段落向量进行分类 (<https://github.com/deeplearning4j/dl4j-examples/blob/master/dl4j-examples/src/main/java/org/deeplearning4j/examples/nlp/paragraphvectors/ParagraphVectorsClassifierExample.java>) 的完整实用示例：

# DL4J

## DEEPLARNING4J

(../cn/index.html)

深度学习教科书

([https://www.amazon.com/Deep-](https://www.amazon.com/Deep-Learning-Practitioners-Adam-Gibson/dp/1491914254)

[Learning-Practitioners-Adam-](https://www.amazon.com/Deep-Learning-Practitioners-Adam-Gibson/dp/1491914254)

[Gibson/dp/1491914254](https://www.amazon.com/Deep-Learning-Practitioners-Adam-Gibson/dp/1491914254))

快速入门 >

教程 >

深度学习介绍 >

神经网络 >

数据与ETL >

模型训练和调参 >

部署 >

开源社区 >

自然语言处理 ✓

```

1 public void testDifferentLabels() throws Exception {
2     ClassPathResource resource = new ClassPathResource("/labeled");
3     File file = resource.getFile();
4     LabelAwareSentenceIterator iter = LabelAwareUimaSentenceIterator.createWithPath(file.getAbsolutePath());
5
6     TokenizerFactory t = new UimaTokenizerFactory();
7
8     ParagraphVectors vec = new ParagraphVectors.Builder()
9         .minWordFrequency(1).iterations(5).labels(Arrays.asList("negative", "neutral", "positive"))
10        .layerSize(100)
11        .stopWords(new ArrayList<String>())
12        .windowSize(5).iterate(iter).tokenizerFactory(t).build();
13
14    vec.fit();
15
16    assertEquals(vec.lookupTable().vector("UNK"), vec.lookupTable().vector("negative"));
17    assertEquals(vec.lookupTable().vector("UNK"), vec.lookupTable().vector("positive"));
18    assertEquals(vec.lookupTable().vector("UNK"), vec.lookupTable().vector("neutral"));

```

## 扩展阅读

- 语句与文档的分布式表示 ([https://cs.stanford.edu/~quocle/paragraph\\_vector.pdf](https://cs.stanford.edu/~quocle/paragraph_vector.pdf))
- Word2vec教程 (word2vec)

Copyright © 2017. Skymind (<https://www.skymind.io>).  
 DL4J is distributed under an Apache 2.0 License.

Github (<https://github.com/deeplearning4j/>)

微博 (<https://weibo.com/dl4j>)

QQ交流群 (<https://shang.qq.com/idkey=5d3891e980a3a1d54f72dbe3e9831df9237be11634766440e8fbe2>)  
 English (../) 日本語 (../ja-index) 한글 (../kr-index)

ND4J (<http://nd4j.org>) 我们在Gitter聊天

# DL4J

## DEEPLEARNING4J

(../cn/index.html)

深度学习教科书

(<https://www.amazon.com/Deep-Learning-Practitioners-Adam-Gibson/dp/1491914254>)

快速入门 >

教程 >

深度学习介绍 >

神经网络 >

数据与ETL >

模型训练和调参 >

部署 >

开源社区 >

自然语言处理 >