一、代码相关

1. <https://zhuanlan.zhihu.com/p/294993724>

深度学习序列数据处理利器-tokenizer(快速建立词汇表字典，实现文本和序列之间的转换)：介绍了Tensorflow的tokenizer、Pytorch的torchnlp库和torchtext库里的两个tokenizer

2、[[python] spacy - 简书 (jianshu.com)](https://www.jianshu.com/p/e6b3565e159d)

关于spacy库的简单说明

3、[Pandas-DataFrame基础知识点总结 - 简书 (jianshu.com)](https://www.jianshu.com/p/8024ceef4fe2)

Pandas-DataFrame基础知识点总结：DataFrame是一种表格型数据结构，它含有一组有序的列，每列可以是不同的值。DataFrame既有行索引，也有列索引，它可以看作是由Series组成的字典，不过这些Series公用一个索引。

4、[PyTorch 71.Pytorch中nn.Embedding模块 - 知乎 (zhihu.com)](https://zhuanlan.zhihu.com/p/341176854)

Pytorch中nn.Embedding模块：一个保存了固定字典和大小的简单查找表。这个模块常用来保存词嵌入和用下标检索它们。模块的输入是一个下标的列表，输出是对应的词嵌入。相当于随机生成了一个tensor，可以把它看作一个查询表，其size为[num\_embeddings，embedding\_dim] 。

5、[pytorch技巧 五： 自定义数据集 torch.utils.data.DataLoader 及Dataset的使用\_董小姐～的博客-CSDN博客](https://blog.csdn.net/qq_40788447/article/details/114937779)

自定义数据集 torch.utils.data.DataLoader：该接口主要用来将自定义的数据读取接口的输出或者PyTorch已有的数据读取接口的输入按照batch\_size封装成Tensor。Pytorch提供两种数据集： Map式数据集 Iterable式数据集。其中Map式数据集继承torch.utils.data.Dataset，Iterable式数据集继承torch.utils.data.IterableDataset。

6、[【机器学习】模型融合方法概述 - 知乎 (zhihu.com)](https://zhuanlan.zhihu.com/p/25836678)

模型融合方法概述：想要在Kaggle比赛中提高成绩，一般要注意以下三点：特征工程、调参、模型融合。总结了几个stacking方法，从宏观上理解这它们。