

3.4 Numpy 的应用

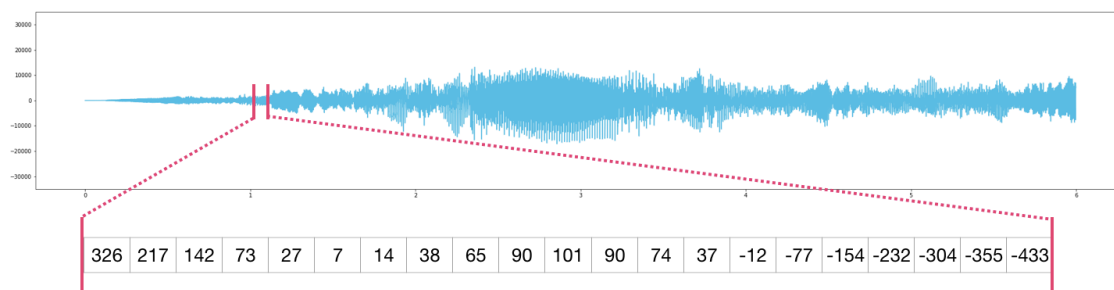
2023 年 9 月 8 日

3.4 Numpy 的应用

1. 音频数据

一个音频文件是一个一维数组的样本。每个样本都是一个数字，代表一小块音频信号。CD 的音频每秒可能有 44,100 个样本，每个样本是 -32767 到 32768 之间的整数。这意味着如果您有一个 10 秒的 CD 的 WAVE 文件，您可以将它装入一个长度为 441,000（也就是 $10 * 44,100$ ）的 NumPy 数组中。如果想提取音频的第一秒，只需将该文件加载到一个 NumPy 数组 `audio` 中，并使用 `audio[:44100]` 即可获取到。

下面是一个音频文件的一个切片：



2. 图像数据

一个图像就是一个像素矩阵，其维度就是高度 x 宽度。- 如果图像是黑白照片，也就是一个灰度图，每个像素可以用一个数字代表（通常是在 0（黑）和 255（白）之间）。如果想要裁切图像左上角 10 x 10 像素的部分，只需通过 NumPy 的 `image[:10, :10]` 函数即可获取。

下面是一个图片文件的切片：



- 如果是一个彩色图像，那么每个像素可以用 3 个数字表示——一个代表红色，一个代表绿色，一个代表蓝色。这种情况下，我们需要一个 3 维素组（因为每个单元仅可以包含一个数字）。因此，一个彩色图像可以被一个 ndarray 维度表示：（高度 * 宽度 * 3）：



3. 语言

如果我们处理的是文本的话，情况可能有点不同。要用数值表示一段文本需要构建一个词汇表（模型需要知道的所有的唯一词）以及一个词嵌入（embedding）过程。让我们看看用数字表示这个谚语的步骤：“Have the bards who preceded me left any theme unsung?” 翻译：在我之前的吟游诗人有没有留下任何主题？

我们可以使用一个词汇数据集来构建一个词汇表（71,290 个单词）：

词汇表

#	
0	the
1	of
2	and
...	...
71,289	dolophine

这个句子可以被划分为一系列词：

have	the	bards	who	preceded	me	left	any	theme	unsung
------	-----	-------	-----	----------	----	------	-----	-------	--------

然后用词汇表中单词的 ID 来替换它：

38	0	29104	56	7027	745	225	104	2211	66609
----	---	-------	----	------	-----	-----	-----	------	-------

然后进一步使用 word2vec 词嵌入、卷积向量等技术来处理文本，详见：
<https://jalammar.github.io/illustrated-word2vec/>

参考:

1. Numpy 和数据展示的可视化介绍, <http://www.junphy.com/wordpress/index.php/2019/10/24/visual-numpy/>
2. A Visual Intro to NumPy and Data Representation, <https://jalammar.github.io/visual-numpy/>