**基于用户协同的音乐个性化推荐挖掘模型的5W1H分析**

1. **成员介绍**

|  |  |  |
| --- | --- | --- |
|  | **OTQ4F[HO[$R$C7KAR~@3UTV** | **2019-05-06 23:28:56.640000** |
| 雷文瑶（组长） | 郭亚琪 | 夏馨宇 |
| 2016141093013 | 2016141093016 | 2016141093045 |

1. **场景分析**

小郭是一个很喜欢听歌的女大学生，但是她渐渐注意到身边的朋友都在用网易云音乐，自己却一直在用QQ音乐，好几次和同学之间互相分享喜欢的歌曲都不太方便。同学们都说很喜欢网易云的“日推”功能，所以小郭也下载了网易云音乐来用。由于自己所学的专业课程最近在学习个性化推荐，小郭决定去了解一下个性化推荐背后的原理。

**Why**：了解个性化推荐背后的实质，特别是目前中国市场上的音乐类平台是如何进行个性化推荐的。

**What**：从歌曲本身和平台两方面入手，了解个性化推荐的

**Where**：网易云音乐

**When**：小郭成为网易云音乐的新用户

**Who**：使用具有个性化推荐功能的音乐平台3用户。

**How**：从数据源头入手，找到符合项目需求的数据并进行清理和规范，选择合适的模型与算法进行学习和训练，得到结果后进行评估与优化。

1. **研究思路**

通过文献调研，当前主流的音乐的个性化推荐方法归纳如下：

我们倾向于从**基于用户协同过滤推荐**的角度进行研究。该方法基于用户对音乐的评分数据，找到具有与该用户最相似偏好的用户，进而推荐相似用户评分最高的音乐。这种方法善于发掘用户新的兴趣点，优点是处理的信息和对象简单，易分析，只需实时搜集即可，无需像基于内容的推荐，需要计算大量复杂的音频数据特征；缺点是冷启动问题和数据稀疏性，对用户过往产生的使用数据要求比较高。

在分析该问题时，我们倾向于将其理解为解决预测的回归问题，最后将要产生一个向用户个性化推荐的歌单。Word2Vec是一个基于神经网络的学习算法，应用在歌曲推荐中，可从大量流数据中提取歌曲潜入信息的神经网络方法；它不仅可以捕捉相似用户通常对哪些歌曲感兴趣，还能捕捉在相似环境下哪些歌曲频繁地一起出现。最后我们将得到一个模型，其中每首歌都用一个高维向量来表示，即每首歌在空间中都将有一个位置，距离越近的歌，说明越可能受到同一用户的喜爱。

基于以上思路，在数据源上我们将使用Million Songs Dataset，或者直接从网易云音乐上进行数据爬取，目的是获取用户听歌记录以及相关歌曲数据。对数据进行处理后，进行建模和编程，最后进行评估，找出不足和可优化的地方。

1. **成员分工**

|  |  |  |  |
| --- | --- | --- | --- |
| 建模 | 1. 获取数据 2. 学习Word2Vec模型原理 3. 结合数据建立模型 | | |
| 编程 | 1. 建立并规范数据集 2. 完成CBOW模型的编程 3. 完成skip-grow模型的编程 | | |
| 论文 | 1. 研究意义 2. 文献综述 3. 研究思路 | 1. 介绍模型算法 2. 数据预处理 3. 建立模型 | 1. 结果展示 2. 模型评估 3. 未来展望 |

1. **进度安排**

|  |  |
| --- | --- |
| 第11周 | 选择合适的数据源，并完成数据的预处理 |
| 第12周 | 筛选出合适的数据特征，完成基本的数据建模 |
| 第13周 | 完成Word2VEC中CBOW模型的编程 |
| 第14周 | 完成Word2VEC中skip-grow模型的编程 |
| 第15周 | 对所建立的模型进行评估 |
| 第16周 | 完成论文撰写 |
| 第17周 | 进行项目展示 |