

# 基于用户评论的自动化音乐分类方法<sup>①</sup>

郝建林, 黄章进, 顾乃杰

(中国科学技术大学 计算机科学与技术学院, 合肥 230027)  
(中国科学技术大学 安徽省计算与通信重点实验室, 合肥 230027)  
(中国科学技术大学 先进技术研究院, 合肥 230027)

**摘要:** 针对现有音乐平台分类类别固定、检索内容限制过多的问题, 本文提出了一种基于用户评论的自动化音乐分类方法. 首先, 通过 linear CRF 统计分词模型、 $n$  元取词和紧密度分析方法学习得到适合音乐语料分词的字典. 其次, 使用 linear CRF 在上述字典的基础上进行分词, 对分词结果进行分合测试, 修正分词结果. 然后, 使用优化后的 TFIDF 关键词提取算法进行标签提取, 再经过标签合并得到音乐的候选标签. 接着, 从全局角度出发对标签进一步筛选, 得到音乐的关联标签. 最后, 建立音乐和标签之间的概率分类模型, 对音乐进行分类. 实验结果表明, 该音乐分类方法准确率较高, 可以从用户评论中自动地获取音乐多个维度的分类标签, 为个性化的音乐检索提供了保障.

**关键词:** 音乐分类; 分词模型; 紧密度分析; 关键词提取; 关联标签

引用格式: 郝建林, 黄章进, 顾乃杰. 基于用户评论的自动化音乐分类方法. 计算机系统应用, 2018, 27(1): 154-161. <http://www.c-s-a.org.cn/1003-3254/6155.html>

## Automatic Music Classification Method Based on Users' Comments

HAO Jian-Lin, HUANG Zhang-Jin, GU Nai-Jie

(School of Computer Science and Technology, University of Science and Technology of China, Hefei 230027, China)  
(Anhui Province Key Laboratory of Computing and Communication Software, Hefei 230027, China)  
(Institute of Advanced Technology, University of Science and Technology of China, Hefei 230027, China)

**Abstract:** An automatic music classification method based on users' comments is presented in view of the few categories and limited search content for the existing music platforms. First of all, a linear CRF statistic segmentation model,  $n$ -gram word extraction and affinity analysis method are used to obtain a dictionary which can be adapted to music corpus segmentation. Secondly, we use linear CRF model to segment comments with dictionary above, and then we correct the segmentation result via split-merge testing. Thirdly, the optimized TFIDF keyword extraction model is applied to extract candidate tags, and we merge tags after that. Fourthly, candidate tags with fewer frequency are filtered from a global perspective. Finally, a probability classification network is established between the music and filtered tags to classify music. As the result shows, our music classification method achieves high accuracy. Furthermore, it can ensure the personality of music retrieval for generating music tags automatically in multiple dimensions according to users' comments.

**Key words:** music classification; word segmentation model; affinity analysis; keyword extraction; associated tags

## 1 系统概述

随着多媒体技术的发展, 音乐已经发展成人们日常生活中必不可少的一部分. 现有的音乐平台, 如网易

云、虾米、酷狗等为我们提供了更便捷的获取音乐的途径. 但由于分类模型的限制, 这些音乐平台的音乐检索方式仍局限于已有的音乐流派、艺术家、专辑等.

<sup>①</sup> 基金项目: 安徽省自然科学基金 (1408085MKL06); 高等学校学科创新引智计划项目 (B07033)  
收稿时间: 2017-04-14; 修改时间: 2017-05-02; 采用时间: 2017-05-08; csa 在线出版时间: 2017-12-22

用户无法通过输入个性化的内容检索到期望的音乐列表. 为了提高用户的检索体验, 则需要引入新的音乐分类模型.

现有的音乐分类模型主要从4个角度进行分类, 分别为流派、情感、乐器、注解.

第一类为基于流派的音乐分类, 这类方法将提取的音频特征作为流派分类的依据. Tzanetakis<sup>[1]</sup>提供了一个测试数据集, 将1000首歌音频分为10个对应的音乐流派. 此后出现的ISMIR<sup>[2]</sup>和Dortmund<sup>[3]</sup>数据集为这类研究的衡量提供了基础. Li等<sup>[4]</sup>提出了DWCHs模型, 通过计算音频的Daubechies小波系数的直方图对音乐进行自动化分类. Lidy等人<sup>[5]</sup>发现了心理声学变换对音频特征提取的影响, 使用两个特征表示统计频谱描述子和音乐韵律直方图特征, 提升了分类准确率.

第二类为基于情感的音乐分类方法, 这类方法将音乐分成开心、伤心、感动等多个类别, 主观性较强, 类别之间有交叉, 没有标准的测试数据集, 模型间的对比较为困难. Yang等人<sup>[6]</sup>使用人工和软件代理结合的方式确定音乐的情感类别, 减少人类的主观影响. Yang等人<sup>[7]</sup>尝试对情感进行量化, 建立美学情感中的评价值和唤起程度值平面, 平面上点的位置对应不同的情感类别.

第三类为基于乐器的分类, 这类方法为识别出音频中使用的乐器种类, 然后通过乐器种类对音乐进行分类. 早期的乐器识别主要对一件乐器独奏的音频进行识别, 比如Marques等人<sup>[8]</sup>基于高斯混合模型和SVM支持向量机、Agosfini等人<sup>[9]</sup>基于音频的频谱特征的乐器识别方法等. 现在研究方法已转到对复调音乐中乐器的识别. Essid等人<sup>[10]</sup>利用概率距离对音频进行分层聚簇, 每层簇的中心对应一类乐器.

第四类为基于注解的分类, 即为基于标签的分类. 此类方法为将音频关联相应的文本内容后, 再用于音乐分类, 最初由Slaney<sup>[11]</sup>提出. Wang等人<sup>[12]</sup>提出利用音频和与音频相关的社交信息将音乐关联对应的标签, 再按照音乐标签进行分类.

前三类为基于音频的分类模型, 第四类为基于文本和基于音频的分类模型. 基于流派的音乐分类, 现已在各大平台有了较为成熟的应用, 但均没有用于音乐检索. 基于情感的音乐分类, 分类类别之间有交叉, 分类界限模糊, 故分类模型的建立难度较大. 基于乐器的音乐分类, 对于多乐器合成的音乐, 难以准确的识别全

部乐器. 基于注解的音乐分类, 其分类准确率依赖于获取的音乐标签的准确率.

考虑以下需求: 用户希望获取某个音乐, 但是对这个音乐不了解, 通常会输入一些相关的个性化描述来检索期望的音乐. 前三类分类模型分类类别较为固定, 难以用于个性化的用户检索需求. 第四类分类模型在文本较为丰富的情况下, 如果能精确的挖掘出于其中蕴含的音乐标签, 那么用户的个性化检索需求就更可能得到满足.

本文为基于注解的音乐分类方法, 通过分析用户对音乐的评论信息, 提出了一种使用个性化标签对音乐进行分类的方法. 该方法的出发点为: 如果用户对音乐的较为熟悉, 那么当其对音乐进行评论时, 其对音乐的描述将更加的深入. 这些评论内容中将含有较多个性化的信息. 只要将这些信息挖掘出来, 将对个性化音乐检索和音乐推荐带来更好的用户体验. 该方法采用无监督的方式为音乐关联多个标签, 以这些标签进行音乐分类, 无需预先对音乐进行标签标注.

本文的组织结构如下: 第2节介绍了该音乐分类算法的相关背景; 第3节介绍提出的音乐分类算法的具体框架和实现; 第4节为实验与分析; 第5节对本文进行总结和下一步展望.

## 2 理论分析

### 2.1 标签提取

该音乐分类算法的核心为关联标签的提取. 关联标签提取过程即为关键词的抽取过程. 关键词提取主要包括关键词抽取和关键词分配两种方式.

关键词抽取, 先提取文档中所有的候选关键词再推荐关键词. 此类方式需要分词并选择合适的抽取方法. 抽取方法有很多, 如TFIDF<sup>[13,14]</sup>、TextRank<sup>[15]</sup>等.

关键词分配, 先预先定义一个受控词表, 然后分析文档, 再推荐受控词表中的部分关键词. 此种方式需要定义并扩充受控词表, 然后选择分配算法. 词表扩充的方式一般采用种子扩充.

一般来说, 关键词抽取要比关键词分配的提取准确率要高, 主要原因在于建立完善的受控词表难度较大, 并且分配的关键词可能不会出现在文档中, 分配算法的准确率难以保证. 因此, 本文的音乐分类方法中采用关键词抽取的方式获取音乐标签.

### 2.2 中文分词

现有使用的中文分词方法主要可以分为三类: 基

于匹配、基于统计和基于社交网络的分词。

基于字符串匹配的分词方法是將汉字串与词典中的词条做匹配,在匹配过程中可以加入匹配规则。分词方法包括:正向最大匹配(FMM)、逆向最大匹配(RMM)、MMSEG<sup>[16]</sup>等。该类分词模型的准确率依赖于字典的丰富程度。

基于统计的分词模型从词频出发进行分词,包括 linear CRF 模型<sup>[17]</sup>、隐马尔科夫模型<sup>[18]</sup>,等。该类模型准确率依赖于标注语料的丰富程度。

基于社交网络的分词模型是在  $n$  元取词的基础上进行分词。其核心在于分析  $n$  元串成词的可能性,通常使用紧密度分析判定。该类模型准确率依赖于紧密度分析方法和语料的丰富程度。

本文的分词方法综合了三类分词模型的优点,使用  $N$  元取词方法和 linear CRF 进行字典扩充,然后使用 linear CRF 进行分词,MMSEG 和紧密度分析进行分词修正。

### 3 音乐分类算法

本文提出的音乐分类方法主要分为以下几个部分,如图 1 所示。

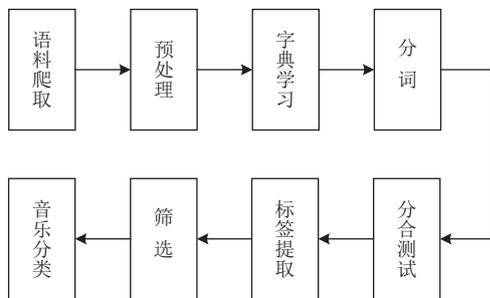


图 1 音乐分类算法框架

预处理为删除一些无用的音乐评论信息,减少对后续处理过程的干扰。字典学习是为了得到一个适合音乐语料分词的字典,提高分词准确率。分词是将音乐语料切分成单个词语,以便提取音乐标签。分合测试是用来修正误分词结果,同时提高对短语的识别率。标签提取是为了从分词结果中选取与音乐相关的候选标签。筛选的目的为过滤掉一些可能错选的音乐标签。最后使用关联的标签为音乐进行分类。

#### 3.1 语料爬取和预处理

本文爬取的音乐评论语料为网易云音乐平台(国

内最大的音乐平台)官方推荐专辑中的音乐评论,共 1459 个专辑,128 542 首歌曲,92 110 590 条评论。

对音乐评论语料按以下步骤进行预处理:

1) 含有脏话的评论参考价值较低,所以将含有脏话的评论删除。

2) 数字大多没有实际意义,所以将数字“233”、“666”等用空格代替,对只含有数字的评论直接删除。

3) 表情符一般表示心情,但是表情符过于多样,例如,网易云音乐自带表情和 emoji 表情编码方式不同,统一混用在评论中识别难度较大。因此将表情符用空格代替。将只含有表情符的评论删除。

4) 过多的重复评论也会影响提取的音乐标签的准确率。例如,一个音乐的评论中含有较多的“路过”,则“路过”很可能成为关键词,但是这种词应该忽略。为防止误判,相同的评论中保留一条。

5) 删除评论内容较少的音乐,评论内容过少则评论可参考性较低。

6) 删除评论数过少的音乐,评论数过少则评论的范围过于分散,提取的音乐标签可信度较低。

#### 3.2 字典学习

字典学习首先通过外部信息获取部分词汇,然后以种子生成的思想扩充字典,如图 2 所示。

具体步骤如下:

1) 从 1998 年和 2014 年的人民日报中文语料以及微软的中文分词标注语料库中提取部分词语,然后爬取百度百科和搜狗词库中的常用词汇加入到参照词典  $D'$ 。对于中国自然语言开源组织提供的 8 万多部小说,使用 ansj 分词工具进行分词,选出词频频率高于阈值  $\alpha$  的词汇加入到参照字典  $D'$  中。

2) 先使用  $n$  元分词对音乐语料做处理,获取所有的 2 字和 3 字词串。先使用紧密度分析,过滤掉明显不是词的串。找出剩余在参照字典  $D'$  中的词,加入到字典  $D$  中。

3) 基于上述获得的字典  $D$ ,统计其中每个词出现的频数。然后建立用于 linear CRF 学习模型的标注语料。最后用 linear CRF 学习后的模型进行分词。

4) 过滤掉词频低的词汇、非专有名词和单字。对剩余切分词语进行紧密度分析。如果词语紧密度高于阈值  $\beta$ ,则加入字典  $D$ 。

5) 重复步骤 3, 4, 直至字典  $D$  不再增大。

$\alpha$  和  $\beta$  值的确定可以通过 k-means( $k=2$ ) 算法聚类确定。

关于音乐标签, 我们关注的词性主要是一些专有名词或者名词性短语, 比如人名、地名、歌曲名、专

辑名、电影名、书名等, 而不关注其他的词性, 故需要尽量确保分词结果中所关注词性的词语的正确性. 可以预先根据它们在评论语料中出现的特点, 按特定的规则进行抽取, 加入字典  $D$ .

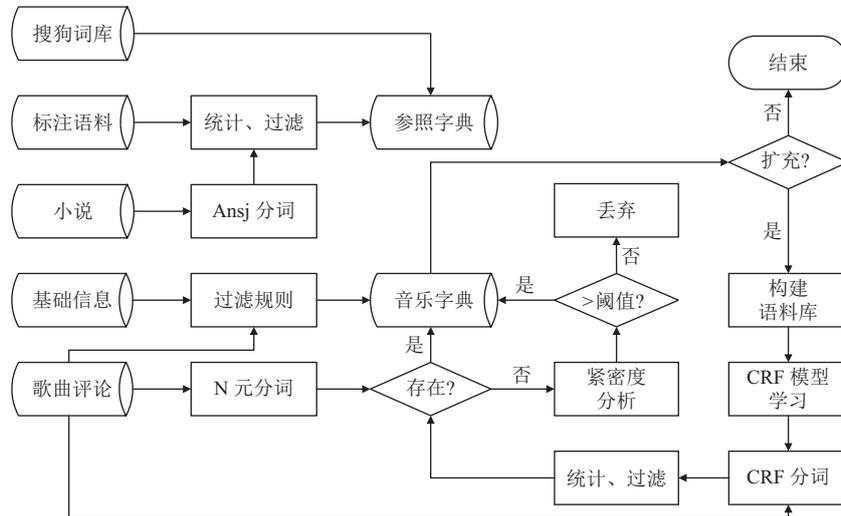


图2 字典学习算法

### 3.3 音乐语料分词

本文在使用过程中采用 linear CRF 和字典相结合的方式对中文进行分词。

由于 linear CRF 模型分词的准确率依赖于标注语料的规模和丰富程度, 适合音乐语料分词的标注语料并不存在, 因此使用 3.2 节获取的字典  $D$  进行 linear CRF 模型的训练, 使用 Viterbi 算法进行标注. 其分词设置的获取特征函数的特征模板如图 3 所示。

# Unigram	# Bigram
U01:%x[-1,0]	B
U02:%x[0,0]	B01:%x[0,0]
U03:%x[1,0]	
U04:%x[-1,0]/%x[0,0]	
U05:%x[0,0]/%x[1,0]	
U06:%x[-1,0]/%x[1,0]	

图3 linear CRF 特征函数模板

### 3.4 分合测试

分词后的结果对存在于字典中的词相对准确, 对于不存在的词需要进行分离, 对于分开的短语需要合并. 在此, 使用聚合度判定被切分为词语的汉字串是否应该分离, 使用自由度判定多个词语是否应该合并。

聚合度 (Degree of Polymerization, DoP) 用来衡量

词语内部组合的紧密程度. 字串  $w$  的聚合度用其中字  $a_i$  的方差表示, 具体如公式 (1) 所示。

$$DoP(w) = \frac{\sum_{i=1}^n (P(a_i) - avg(w))^2}{n} \quad (1)$$

自由度 (Degree of Freedom, DoF) 来衡量词语和上下文之间的关系. 此关系用该词左右的字的丰富程度来衡量. 字串  $w$  的自由度由其左边字  $l_i$  和右边字  $r_i$  的熵确定, 具体如公式 (2) 所示。

$$DoF(w) = \min\left(\sum_i -P(l_i)\log(l_i), \sum_i -P(r_j)\log(r_j)\right) \quad (2)$$

综合考虑词频、聚合度和自由度, 得到分合测试的衡量方法, 如公式 (3) 所示. 词频取对数是为了处理词频过大而无法拆分的伪词汇。

$$F = DoP(w) * DoF(w) * \log(Freq(w)) \quad (3)$$

对于未通过分合测试的句子使用 MMSEG 模型进行修正处理。

### 3.5 标签提取

标签提取类似于关键词提取, 从音乐的评论语料中提取与之关联程度最高的一个或者多个词或短语。

在 3.4 节分词后得到的词频统计结果, 其 Top  $N$  可

能并不是我们所需要的音乐标签. 下面取出周杰伦的歌曲“晴天”的分词结果中词频 Top 20 进行分析, 如表 1 所示. 左侧是 Top 10 频数的词语, 右侧为 Top 11 到 Top 20 的结果.

表 1 歌曲晴天评论分词结果 Top 20

词汇	词频	词性	词汇	词频	词性
评论	63958	n	歌词	4560	n
晴天	42736	n	个人	3951	n
周杰伦	27789	nr	初中	3836	n
青春	10546	nt	头像	3462	n
故事	9337	n	时间	3248	n
感觉	7924	n	专辑	3213	n
演唱会	7259	n	七里香	2961	n
音乐	6061	n	朋友	2930	n
前奏	5825	n	距离	2836	n
周董	5462	nr	高中	2802	n

对比关键词抽取算法, 实验发现 TFIDF 算法抽取标签准确率最高, 故选取 TFIDF 算法. TFIDF 算法如公式 (4) 所示.

$$TFIDF_{i,j} = TF_{i,j} * IDF_{i,j} = \frac{n_{i,j}}{\sum_k n_{k,j}} * \log \frac{|D|}{|\{j: t_i \in d_j\}|} \quad (4)$$

其中,  $n_{i,j}$  为该词在第  $j$  个文档中第  $i$  个词中出现的次数,  $|D|$  为语料中文档的个数,  $|\{j: t_i \in d_j\}|$  为包含该词的文档个数, 保证其不为 0.

在实验过程中, 直接使用 TFIDF 并不能急速降低表 1 中“评论”、“个人”这类词汇被选中的可能性, 原因为  $\log$  函数在  $n > 1$  时变化率太低.

故为模型选择变化率大的函数或对词汇的特点做一些限制来过滤掉这些无用的词汇.

根据标签的特性做以下两个假设:

假设一. 一个标签出现的频率在音乐评论中不能低于音乐评论总数的  $\alpha$  倍,  $0 < \alpha < 1$ .

假设二. 一个标签关联的音乐数量不能超过音乐总数的  $\beta$  倍,  $0 < \beta < 1$ .

假设一是为了降低词频过低的词语被选为标签的可能性. 例如, “刘德华”在音乐“狂风里拥抱”中出现了一次, 出现的内容为“@Andy 不是刘德华”, 该词不可能作为标签. 但是由于其在较多的音乐评论中出现, 导致逆文档频率较大, 容易被误选为标签. 为降低这种误选的可能性, 因此使用假设一过滤.

假设二是为了降低词频过高的常用词被选为标签

的可能性. 例如, “评论”在音乐“晴天”中出现了 63 242 次, 共出现在 89 297 首音乐的评论中, 这类的 TFIDF 值过大, 会被误选为标签. 显然, 其不应该作为音乐标签, 需要过滤掉. 为降低这种词被误选的可能性, 因此使用假设二进行过滤.

TFIDF 模型可以优化为如式 (5) 所示.

$$TF\_IDF_{i,j} = TF_{i,j} * IDF_{i,j} = \frac{n_{i,j}}{\sum_k n_{k,j}} \left[ n_{i,j} > \alpha \sum_k n_{k,j} \right] * \log \frac{|D|}{|\{j: t_i \in d_j\}|} \left[ |\{j: t_i \in d_j\}| < \beta |D| \right] \quad (5)$$

$\alpha$  和  $\beta$  值的确定可以通过 k-means( $k=2$ ) 算法聚类确定.

优化后的 TFIDF 降低了常用词汇被选中的概率, 提高了其它词汇被选中的概率. 优化后的 TFIDF 得到的候选标签, 如表 2 所示.

表 2 优化后 TFIDF 得到的候选标签 Top 20

词汇	TFIDF	词汇	TFIDF
晴天	0.1342	告白	0.0058
周杰伦	0.0699	叶惠美	0.0058
周董	0.0192	爱杰伦	0.0053
七里香	0.0150	下雨天	0.0052
黄花	0.0092	黄瓜	0.0051
告白气球	0.0085	阴天	0.0046
许嵩	0.0071	老周	0.0043
妖怪	0.0068	雨天	0.0040
距离	0.0061	兰花花	0.0039
小公举	0.0061	以父之名	0.0038

### 3.6 标签筛选

由于优化后的 TFIDF 算法获取的候选标签可能出现提取的标签依然出错的情况. 因此, 考虑删除这些错误的标签.

由于每个标签不可能只出现  $\gamma$  个文档,  $\gamma$  为 0 邻域范围内的值. 因此可以将出现的文档数小于  $\gamma$  的标签全部删除.  $\gamma$  值依然可以通过 k-means( $k=2$ ) 获取.

### 3.7 音乐分类模型

在此建立音乐和关联标签的网络模型. 图 4 给出了一个  $n$  标签关联的网络, 音乐  $S_i$  用正方形表示, 标签  $T_{i,j}$  用圆形表示, 音乐  $S_i$  和标签  $T_{i,j}$  的关联程度用边上的权重  $P_{i,j}$  表示,  $P_{i,j}$  可以通过标签频数归一化得到.

使用公式 (6) 进行相似度分析, 如果相似度高于某个阈值, 则认为两首歌曲可以归为一类.

$$Similar(S_m, S_n) = \frac{\sum_{i,j} (P_{i,m} P_{j,n} F(T_{i,m}, T_{j,n}))}{|D|} \quad (6)$$

其中 $|D|$ 为音乐  $S_m$  和  $S_n$  的关联标签数量之积, 用于归一化处理.  $F$  函数为两个标签的相似度, 可以标签关联的相同歌曲的数量的百分比乘积确定. 如公式 (7) 所示.

$$F(T_{i,m}, T_{j,n}) = \frac{Same(T_{i,m}, T_{j,n})^2}{|T_{i,m}| \times |T_{j,n}|} \quad (7)$$

其中,  $Same$  函数为两个标签关联的相同音乐的数量.  $|T_{i,m}|$ 和 $|T_{j,n}|$ 为对应标签关联的音乐的数量.

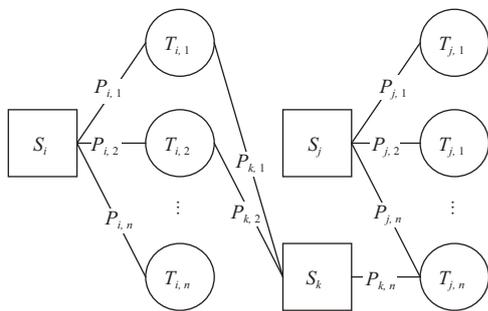


图4 音乐分类概率网络

## 4 实验和分析

本节将对第3节中提出的分类算法进行实验. 实验一是为了验证 MMSEG、HMM、CRF 模型分词效果的优劣. 实验二是为了选取合适关键词抽取算法并验证其优化算法的有效性. 实验三是为了验证该音乐分类算法分类结果的有效性.

### 4.1 分词模型准确率对比

选取9首歌曲23614条评论内容作为样本, 对其进行分词标注. 各种模型的分词准确率如图5所示.

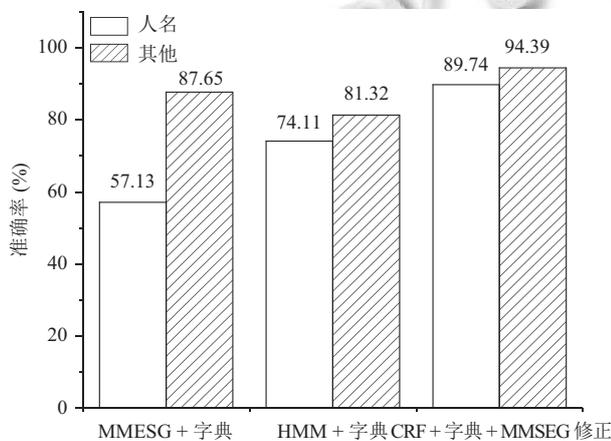


图5 分词模型准确率对比

从图5中可以看出 MMSEG 模型由于通用字典的局限性, 导致其对人名识别率不高, 远低于 HMM 模型和我们的混合模型. 由于不存在用于音乐语料的标注集合, HMM 模型未加入训练样本直接进行分词, 导致分出的单字较多, 因此对其它词语切分准确率略差. 使用 MMSEG+linear CRF+字典的分词结果, 其综合了 MMSEG 消除歧义的特点和 linear CRF 发现未登录词的特点. 词典的创建和分合测试提高了分词的准确率, 同时减小了 linear CRF 出现单字的可能性, 因此其准确率均比 MMSEG 和 HMM 模型都高.

### 4.2 标签提取算法对比

使用 TFIDF 算法和 TextRank 算法进行关键词抽取实验, 实验数据选用“周星驰版唐伯虎点秋香”背景音乐“勇往直前”的评论语料. 结果如表3所示. 可以看出, TFIDF 获取的标签的关联程度远高于 TextRank.

表3 TextRank 和 our TFIDF 候选标签 Top 10 对比

Top 10	TextRank	TFIDF
1	音乐	江南四大才子
2	江南	星爷
3	电影	胡伟立
4	星爷	鸡翅膀
5	画面	四大才子
6	时候	唐伯虎
7	经典	四大流氓
8	歌单	如花
9	微博	周星驰
10	评论	秋香

为了进一步验证优化方案的有效性, 随机取10首音乐, 每首音乐选取候选标签的 Top 10 进行准确率标注, 10首音乐标签准确率按评论数做加权平均, 实验结果如图6所示.

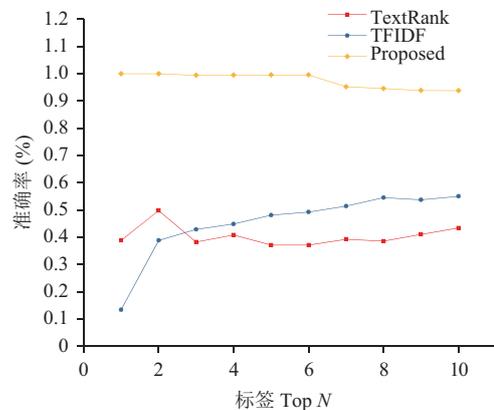


图6 标签提取算法准确率对比

容易看出,优化后的 TFIDF 算法的标签提取准确率要比 TFIDF 和 TextRank 都高. TFIDF 参考了其它音乐评论抽取的结果,虽然不能过滤掉所有频数过高的常用词语,但是能过滤掉部分. TextRank 由于没有停用词典,其抽取结果偏向于高频词汇.优化后的 TFIDF 考虑到高频词和逆文档词频之间的影响,因此,对高频词和低频词的过滤效果较好.

### 4.3 音乐分类结果对比

对最初爬取 66 198 首歌曲,再次爬取这些歌曲在网易云音乐平台推荐的相似音乐.得到一个相似音乐表 T,共 189 625 条相似记录.由于网易云音乐的推荐列表基于大量的用户数据,可以认为其推荐的音乐可信度较高.如果将相似的音乐可以分为一类,过滤掉其中不属于已爬取的音乐列表中的音乐,那么相似音乐表 T 就可以作为音乐分类标准测试集.

相似音乐表中每个音乐取 Top  $N$  标签,建立图 4 所示的音乐分类概率网络,使用公式 6 做相似度分析,测试算法分类准确率.对比 HiSVM<sup>[12]</sup>分类结果,如图 7 所示.

从图 7 中可以看出算法的在标签数在 25 后准确率逐步稳定,最终稳定在 87.96%.实验表明,关联标签数量低于 20 时,关联准确率高于 80%,因此认为 Top 20 标签可信度较高.而在标签数为 20 时,分类结果的准确率为 82.58%,而 HiSVM 的准确率不足 60%.因此,可以认为该音乐分类算法具有很高的有效性.

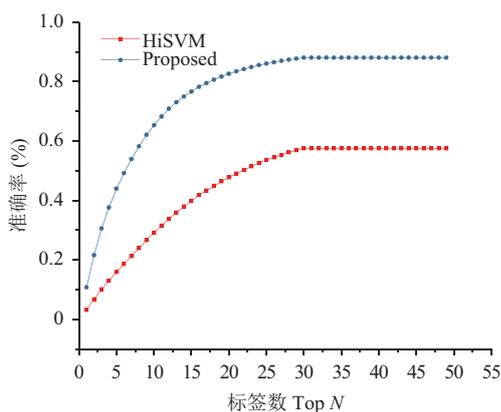


图 7 音乐分类算法准确率

## 5 结语

针对当前音乐平台音乐分类结果固定单一、搜索平台搜索效果差的问题,本文提出了一种基于用户评论的自动化音乐分类算法.该算法优化了已有的分词

模型和关键词提取算法 TFIDF,提升了分词的准确率和关键词提取算法的准确率,建立了基于关联标签的多标签音乐分类模型.实验结果表明,该音乐分类算法的准确率较高,获取的标签关联度高而且更加个性化,可以带来更好的音乐检索体验.

本文下一步工作为解决音乐评论过少或不存在的音乐分类问题,以提高该分类算法的扩展性.

### 参考文献

- 1 Tzanetakis G, Cook P. Musical genre classification of audio signals. *IEEE Transactions on Speech and Audio Processing*, 2002, 10(5): 293–302. [doi: 10.1109/TSA.2002.800560]
- 2 Cano P, Gómez E, Gouyon F, *et al.* ISMIR 2004 audio description contest. MTG-TR-2006-02, Stanford, CA: MTG, 2006.
- 3 Homburg H, Mierswa I, Möller B, *et al.* A benchmark dataset for audio classification and clustering. *Proceedings of the 6th International Conference on Music Information Retrieval*. London, UK, 2005. 528–531.
- 4 Li T, Ogiwara M, Li Q. A comparative study on content-based music genre classification. *Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Informaion Retrieval*. Toronto, Canada, 2003. 282–289.
- 5 Lidy T, Rauber A. Evaluation of feature extractors and psycho-acoustic transformations for music genre classification. *Proceedings of the 6th International Conference on Music Information Retrieval*. London, UK, 2005. 34–41.
- 6 Yang D, Lee WS. Disambiguating music emotion using software agents. *Proceedings of the 5th International Conference on Music Information Retrieval*. Barcelona, Spain, 2004. 218–223.
- 7 Yang YH, Lin YC, Su YF, *et al.* A regression approach to music emotion recognition. *IEEE Transactions on Audio, Speech, and Language Processing*, 2008, 16(2): 448–457. [doi: 10.1109/TASL.2007.911513]
- 8 Marques J, Moreno PJ. A study of musical instrument classification using gaussian mixture models and support vector machines. CRL 99/4, Cambridge, Massachusetts: Compaq Computer Corporation, 1999.
- 9 Agostini G, Longari M, Pollastri E. Musical instrument timbres classification with spectral features. *EURASIP Journal on Applied Signal Processing*, 2003, (2003): 5–14. [doi: 10.1155/S1110865703210118]
- 10 Essid S, Richard G, David B. Instrument recognition in

- polyphonic music based on automatic taxonomies. *IEEE Transactions on Audio, Speech, and Language Processing*, 2006, 14(1): 68–80. [doi: [10.1109/TSA.2005.860351](https://doi.org/10.1109/TSA.2005.860351)]
- 11 Slaney M. Semantic-audio retrieval. 2002 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP). Orlando, FL, USA. 2002. IV-4108–IV-4111.
- 12 Wang F, Wang X, Shao B, *et al.* Tag integrated multi-label music style classification with hypergraph. *Proceedings of the 10th International Society for Music Information Retrieval Conference*. Kobe, Japan. 2009. 363–368.
- 13 Luhn HP. A statistical approach to mechanized encoding and searching of literary information. *IBM Journal of Research and Development*, 1957, 1(4): 309–317. [doi: [10.1147/rd.14.0309](https://doi.org/10.1147/rd.14.0309)]
- 14 Jones SK. A statistical interpretation of term specificity and its application in retrieval. *Journal of Documentation*, 1972, 28(1): 11–21. [doi: [10.1108/eb026526](https://doi.org/10.1108/eb026526)]
- 15 Mihalcea R, Tarau P. TextRank: Bringing order into texts. *Proceedings of Empirical Methods in Natural Language Processing*. Barcelona, Spain. 2004. 404–411.
- 16 Tsai CH. MMSEG: A word identification system for Mandarin Chinese text based on two variants of the maximum matching algorithm. <http://www.geocities.com/hao510/mmseg.html> [2000-03-12].
- 17 Tseng H, Chang PC, Andrew G, *et al.* A conditional random field word segmenter for sighthan bakeoff 2005. *Proceedings of the Fourth SIGHAN Workshop on Chinese Language Processing*. Jeju Island, Korea. 2005. 161–164.
- 18 Eddy SR. Hidden markov models. *Current Opinion in Structural Biology*, 1996, 6(3): 361–365. [doi: [10.1016/S0959-440X\(96\)80056-X](https://doi.org/10.1016/S0959-440X(96)80056-X)]