

基于条件随机场的音乐共同语义标注

何晓梅

(河海大学 计算机与信息学院 南京 211100)

摘要: 音乐的语义标注旨在使用词语或标签自动将一段音乐标注为一个语义标注集。通常,人们将多标注学习转换为独立二进制分类问题解决,再给每一个语义标注单独建模。为了得到更好的分类结果,应考虑标注之间的依赖关系。文章中尝试共同的音乐语义标注,对单标注和具有高相关性的成对标注同时建立模型。使用多标注条件随机场(CRF)模型直接参数化多标注分类中的共现标注。用到两种 CRF 模型,一种是使用无条件标注相关的共同多标注分类(CML)模型,另一种是使用有条件标注相关的考虑特征的共同多标注分类器(CMLF)模型。实验表明,将这两种模型用到 CAL10K 数据集上,平均精确度、宏 F_1 分数和微 F_1 分数比用高斯混合模型(GMM)给单个标注建模要高。

关键词: 多标注分类;共现标注;CML 模型;CMLF 模型;CAL10K 数据集

中图分类号: TP311 **文献标识码:** A **国家标准学科分类代码:** TP391.42

Conditional random fields model for collective annotation of music

He Xiaomei

(Department of Computer and Information, Hohai University, Nanjing 211100, China)

Abstract: Music semantic annotation aims to automatically annotate a music signal with a set of semantic labels using words or tags. Usually, multi-label learning is done by transforming into multiple independent binary classification problems and model each semantic label individually. In order to produce better classification result, label correlations should be taken into account. In this paper, we try to use collective music semantic annotation, which not only builds a model for each semantic label, but also builds models for the pairs of labels that have significant correlations. We use multi-label conditional random fields(CRF) model to parametric the co-occurrence of multi-label classification. Two CRF models are proposed here, one is called collective multi-label classifier(CML) while utilized unconditional label correlation; while the other is called collective multi-label with features classifier(CMLF) which utilizes conditional label correlation. Experiments show that using these two models have higher average accuracy, macro-averaged F_1 score and micro-averaged F_1 score than using Guassian mixture model on the issue of semantic annotation while using CAL10K datasets.

Keywords: multi-label classification; co-occurrence annotation; CML model; CMLF model; CAL10K datasets

1 引言

单标注分类的任务是将一个目标确切地用一个类表示,而多分类问题^[1]是对传统的单分类问题的扩展,将一个目标同时分配给一个或多个类,被广泛应用于各种情境下,比如:文本分类、多媒体内容的自动标注、生物信息、信息检索等。

对音乐的语义标注^[2-3]已经成为在音乐信息检索领域一个重要的方向,将一首音乐与一个语义集关联,与未加工的音频或底层特征相比,是一个非常紧密而有效的表示方法。这种方法可以广泛地应用于音乐检索和推荐,因为对

用户来说,使用语义词汇描述一首歌更加自然,并且使用语义标注向量测试标注的相似性也更为灵活。

目前,自动语义标注的方法有很多。Dhanalakshmi 等人^[4]使用自动关联的神经网络模型获取声学特征向量的分布,再用基于高斯混合模型分类器训练这些类的模型。Turnbull 等人为每个语义词汇建立一个生成式模型 GMM,并且计算出每一首歌在词汇表中的一个多项式分布。Eck 等人使用 AdaBoost 预测每个社会标注在整个大的音乐数据集中的发生强度。Hoffman 等人^[5]提出使用码字伯努力平均(CBA)模型学习预测用某个词语代表一首歌的概率,高效地找到一个最能代表一首歌的特殊

词语。

上述方法通过给文本和音频的关系建模,都获得了很好的结果。但是,它们同时存在着一个问题,那就是语义标签都被单独建模,从而忽略了标注之间的相关性。但是,有些标注之间是有很强的相关性的,而且这种特性可以用于改善标注系统^[6]。比如说,“hard rock”和“electronic guitar”在一首歌的标注中很容易同时出现,而“happy”和“minor key”,“fast tempo”和“gloomy”就很少同时出现。使用标注间的相关信息后,当发现“hard rock”极可能出现在一首歌中时,那么就可以帮助预测“electronic guitar”的出现。同样地,由于“fast tempo”和“gloomy”这两个标注很少同时出现,“fast tempo”出现的情况下,“gloomy”就可以避免出现。

所以文章就是从这个方面出发,改善音乐自动语义标注模型,使标注结果更加具有可用性。接下来的部分就这样组织:第2节给标注的依赖关系做相关性分析;第3节描述如何使用条件随机场建模;第4节是实验章,具体解决语义自动标注的问题;第5节展示文章的实验结果;第6节总结全文。

2 相关性分析

2.1 多标注分类的定义

传统的二进制分类问题是通过将空间中的特征映射到一个二进制标注输出 y , 其中 $|y| = 2$, 也就是学习函数 $f: x \rightarrow y$ 。而多标注分类则是进一步归纳二进制分类: 给定一组标注 y (可以说两个或多个), 多标注分类的目的就是预测出是否每一个标注都该出现或不该出现, 也就是说, 输出结果是 y 的一个子集。因此, 要学习的映射函数可以表示为 $f: x \rightarrow 2^y$, 其中 2^y 可以表示 y 的幂集。事实上, 通常将特征表示成一个固定长度的实值向量, 因此多标注映射就成了 $f: \mathbf{R}^D \rightarrow 2^y$, 假设特征是 D 维特征。

假设有 L 个标注 ($|y| = L$), 所有的多标注结果可以表示为一个 L 维的二进制向量 $\mathbf{y} = [y^{(1)}, y^{(2)}, \dots, y^{(L)}]^T \in \{0, 1\}^L$, 其中若标注出现, 则 $y^{(i)} = 1$, 若标注不出现, 则 $y^{(i)} = 0$ 。从而, 多标注映射函数 $f: \mathbf{R}^D \rightarrow \{0, 1\}^L$ 就可以通过多标注训练数据 $T = \{(x_i, y_i)\}_{i=1}^N$ 学习得到, 这里, N 表示训练的歌曲数, $x_i \in \mathbf{R}^D$, $y_i \in \{0, 1\}^L$ 。多标注分类的目的是, 学习这样的映射关系, 能够根据一个具体的损失函数, 精确地预测一个未知的特征向量 $x \in \mathbf{R}^D$ 。

2.2 相关分析

总的来说, 利用标注的相关性可以提高多标注分类的结果。而很少有论文像 Ghamrawi 等人的研究一样, 为这样的相关性给出确切的定义, 他们把这样的概念分为有条件的标注相关和无条件的标注相关, 这是一个可以发展的方向。接下来, 将使用这个概念, 将标注之间的依赖关系以量化的方式描述出来。

2.2.1 无条件标注相关

对于多标注学习问题 $x \rightarrow 2^y$, 无条件标注相关^[7]描述了标注集 y 间的关系, 而不考虑相关的特征 x 。比如, 考虑了两个标注 $y = (Y_1, Y_2)$, 如果大多数歌曲都有 $(Y_1, Y_2) = (0, 0)$ 或 $(Y_1, Y_2) = (1, 1)$, 就可以认为 Y_1 和 Y_2 是正相关的, 否则就认为是负相关的。

为了简单化, 只考虑训练数据的成对标注而不是同时涉及多标注的相关性。最常用的描述相关性的方法有 Pearson's 相关系数法, 该方法将多标注预测问题看作一个回归问题。实验用到的方法是 χ^2 测试法, 用于测试两个随机标量的统计学特性, 通过下面的公式:

$$X^2(Y_1, Y_2) = \sum_{i,j \in \{0,1\}} \frac{(N_{ij} - E_{ij})^2}{E_{ij}} \quad (1)$$

式中: N_{ij} 表示标注满足 $(Y_1, Y_2) = (i, j)$ 的训练数量, 而 E_{ij} 表示假设 Y_1, Y_2 互相独立的情况下的期望频率, 可以由式(2)可得:

$$E_{ij} = N \times P(Y_1 = i) \times P(Y_2 = j) = N \times \frac{N_{i0} + N_{i1}}{N} \times \frac{N_{0j} + N_{1j}}{N} \quad (2)$$

那么, χ^2 分数就可以由式(3)可得:

$$X^2(Y_1, Y_2) = \frac{N(N_{11}N_{00} - N_{10}N_{01})^2}{(N_{11} + N_{01})(N_{11} + N_{10})(N_{10} + N_{00})(N_{01} + N_{00})} \quad (3)$$

上面的标注提供了任意两个标注相关程度的量化方法, 将该方法应用到所有的标注中, 可以建立一个相关矩阵 $C \in \mathbf{R}^{L \times L}$, 其中 C_{ij} 表示标注 i, j 的相关性。

2.2.2 有条件标注相关

与无条件标注相关相似, 不同的是: 有条件标注相关^[8]与给定固定特征 $X \in x$ 的概率条件依赖紧密相关。通过特征选择, 可以给条件相关性建模。比如说, 如果有 3 个标注 $y = \{Y_1, Y_2, Y_3\}$ 和 2 个特征 $x = \{X_1, X_2\}$ 。如果特征分析显示 $\sum_y P(y | x; w) = 1$ 都只与 X_1 有关, Y_3 只依赖 X_2 , 那么我们就可以得出结论 Y_1, Y_2 比 Y_1, Y_3 更相关。那么就可以通过每对标注特征集的相似度来计算条件相关。具体实现分两部分: 1) 计算标注与特征的相关度; 2) 计算每个标注对的相关系数。

首先, 计算标注与特征的相关度, 可以使用文献[6]中使用的 Fisher 标注分数方法: 考虑到第 f 个特征 $\{x_k^{(f)}\}_{k=1}^N$ 和第 l 个标注 $\{y_k^{(l)}\}_{k=1}^N$, Fisher 标准分数为 $r_f^{(l)} = \frac{(\mu_f^+ - \mu_f^-)^2}{(\sigma_f^+)^2 + (\sigma_f^-)^2}$, 其中 μ_f^+ 和 σ_f^+ 是 k 满足 $y_k^{(l)} = 1$ 的 $x_k^{(f)}$ 的均值和方差; 同样地, μ_f^- 和 σ_f^- 是 k 满足 $y_k^{(l)} = 0$ 的 $x_k^{(f)}$ 的均值和方差。

然后, 通过测试每个标注对与每个特征的相关程度, 计算相关系数。即通过 $\{r_k^{(f)}\}_{k=1}^D$ 和 $\{r_k^{(f)}\}_{k=1}^D$ 计算第 i 个标注与第 j 个标注的相关系数 C_{ij} 。

3 使用条件随机场实现多标注学习

条件随机场(CRF)^[9]是一个可以将无条件或有条件标注相关应用于多标注分类的框架,它是 1 种无向图模型,总体来说,是 1 个概率性对数线性模型,以 w 为条件概率参数:

$$P(y | x; w) = \frac{1}{Z(x)} \exp \left[\sum_k w_k f_k(x, y) \right] \quad (4)$$

式中: $f_k(x, y)$ 称为特征函数, $Z(x)$ 是分割函数,为了确保概率分布满足 $\sum_y P(y | x; w) = 1$, $Z(x) = \sum_{y'} \exp \left[\sum_k w_k f_k(x, y') \right]$ 。

在预测或解码阶段,确定在给定特征 x 的情况下,最可能的标注组合 y ,通过使用式(5):

$$\hat{y} = \underset{y}{\operatorname{argmax}} P(y | x; w) = \underset{y}{\operatorname{argmax}} \left[\sum_k w_k f_k(x, y) \right]$$

将 CRF 模型应用到多标注分类,选出两个适合的 CRF 模型: 1) 使用无条件标注相关的共同多标注分类器(CML); 2) 使用条件标注相关的考虑特征的共同多标注分类器(CMLF)。

3.1 共同多标注分类器(CML)

CML 模型引入两个特征函数,并利用无条件标注相关。第一个特征函数给每个特征 $x^{(i)}$ 对单个标注 $y^{(j)}$ 的影响建模,形式为: $f_k(x, y) = x^{(i)} I(y^{(i)} = 1)$,索引值 k 遍历了所有特征和标注的组合: $k \in \{(i, j): 1 \leq i \leq D, 1 \leq j \leq L\}$; 第二个特征函数给成对的无条件相关的标注建模,只需考虑每个标签对的正负相关性,所以设置特征函数为:

$$f_k = \begin{cases} I(y^{(i)} = y^{(j)}) & q = 1 \\ I(y^{(i)} \neq y^{(j)}) & q = 2 \end{cases} \quad (6)$$

式中: $k' \in \{(i, j, q): q \in \{1, 2\}; 1 \leq i < j \leq L\}$ 。

CML 的分布函数为:

$$P(y | x) = \frac{1}{Z(x)} \exp \left[\sum_k \lambda_k f_k(x, y) + \sum_{k'} \lambda_{k'} f_{k'}(y) \right] \quad (7)$$

图 1 是 1 个由 4 个标注表示的 CML 因子图。节点 Y_1, Y_2, Y_3, Y_4 是 4 个标注, X 表示特征随机变量。所有标注对都连接在一起,特征节点 X 也与每个标注节点相连。每个正方形代表 1 个与特征函数相关的因子,它连接着一个标注对。

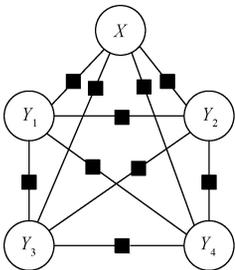


图 1 带有 4 个标注的 CML 因子

3.2 考虑特征的共同多标注分类器(CMLF)

与 CML 模型相似,第一个特征函数 $f_k(x, y)$ 相同。为了给条件标注相关建模,CMLF 介绍了另外一种特征函数,旨在获取 1 对标注和 1 个特征的组合。其特征函数为:

$$f_{k'} = \begin{cases} x^{(m)} I(y^{(i)} = y^{(j)}) & q = 1 \\ x^{(m)} I(y^{(i)} \neq y^{(j)}) & q = 2 \end{cases} \quad (8)$$

式中:索引值 $k' \in \{(i, j, m, q): q \in \{1, 2\}; 1 \leq m \leq D; 1 \leq i < j \leq L\}$ 。

CMLF 的分布函数为:

$$P(y | x) = \frac{1}{Z(x)} \exp \left[\sum_k \lambda_k f_k(x, y) + \sum_{k'} \lambda_{k'} f_{k'}(x, y) \right] \quad (9)$$

图 2 是 1 个由 4 个标注表示的 CMLF 因子图,特征节点连接着所有的标注。

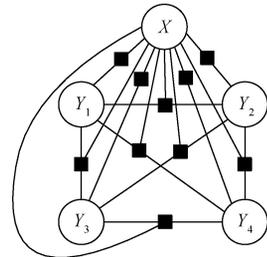


图 2 带有 4 个标注的 CMLF 因子图

4 实 验

4.1 特征选择和预处理

文章用到的数据集是 Computer Audition Lab 10K (CAL10K),由 UCSD 计算机听力实验室创建。总共 10 000 多首歌,由 4 597 个不同的歌手演唱。这里,选出其中的 1 500 首不同歌手演唱的歌曲,共提取出 155 个标注(保证每个训练集与测试集都有同一个标注的歌曲)。CAL10K 只有两种类型的标注,即体裁类标注和声学类标注。

特征提取,使用 MIRtoolbox 作为特征提取的工具包,提取出共 35 维长时特征,包含两种特征类型:节奏和拍子。

对特征向量预处理的步骤为:

1) 由于每首歌曲提取特征后,帧数过高,不能有效计算,首先使用 k-means 算法,得到 100 帧集群重心,即每首歌特征可描述为 $\mathbf{X} = \{x^{(1)}, x^{(2)}, \dots, x^{(35)}\}$, \mathbf{X} 是一个 100×35 的矩阵;

2) 使用二范数标准化特征向量 $\mathbf{X}, \mathbf{X}' = \mathbf{X} / \|\mathbf{X}\|_2$;

3) 尺度变换,将特征向量线性调整到 $[-1, +1]$ 。给定特征数据 $\{X_i\}_{i=1}^N$,如式(10)所示。

$$x^{(j)'} = \frac{2(x^{(j)} - \min_i x_i^{(j)})}{\max_i x_i^{(j)} - \min_i x_i^{(j)}} - 1, j = 1, \dots, 100$$

4) 求 Z 分数,将特征向量标准化,其值代表原始分数与平均值之间的距离, $x^{(j)'} = \frac{x^{(j)} - \mu^{(j)}}{\sigma^{(j)}}, j = 1, \dots, 100$ 。

5)分别计算无条件标注相关矩阵和有条件标注相关矩阵,使用 χ^2 测试方法和 Fisher 标注分数方法,分别得到两种相关系数矩阵 C_{ij} 和 C'_{ij} 。

4.2 建模

使用 CML 和 CMLF 模型,给单标注和成对标注同时建模。这里面,不需要给所有的标注都对建模,只需要通过相关系数矩阵,选出系数高于某一阈值的标注对,添加到模型中,这样可以大大减小模型的复杂度。建模过程,使用 UGM 工具包,并完成推断和优化。

4.3 实验结果评估方法

对于多标注分类问题,使用平均精确度(Acc),宏 F1 分数(MF_1)和微 F1 分数(μF_1)作为衡量标注,分别为式(10)~(12)所示:

$$Acc = \frac{1}{NL} \sum_{i=1}^N \sum_{j=1}^L I(y_i^{(j)} + \hat{y}_i^{(j)}) \quad (10)$$

$$\mu F_1 = \frac{2 \sum_{j=1}^L \sum_{i=1}^N \hat{y}_i^{(j)} y_i^{(j)}}{\sum_{j=1}^L \sum_{i=1}^N \hat{y}_i^{(j)} + \sum_{j=1}^L \sum_{i=1}^N y_i^{(j)}} \quad (11)$$

$$MF_1 = \frac{1}{L} \sum_{j=1}^L \frac{2 \sum_{i=1}^N \hat{y}_i^{(j)} y_i^{(j)}}{\sum_{i=1}^N \hat{y}_i^{(j)} + \sum_{i=1}^N y_i^{(j)}} \quad (12)$$

5 实验结果

5.1 标注的相关性分析

无条件标注相关的前 5 个最相关的标注对如表 1 所示。有条件标注相关的前 5 个最相关的标注对如表 2 所示。

表 1 无条件标注相关的前 5 个最相关的标注对

相关系数 C_{ij}	标注	对应的标注
0.898 6	portuguese lyrics	brazilian
0.870 4	brazilian roots	portuguese lyrics, brazilian
0.859 4	triple note feel	a_twelve_eigh time signature
0.846 5	blues roots	blues
0.824 6	a reggae feel	reggae

表 2 有条件标注相关的前 5 个最相关的标注对

相关系数 C'_{ij}	标注	相应的标注
0.987 4	brazilian roots	portuguese lyrics
0.980 9	hip hop	a dry recording sound
0.976 6	portuguese lyrics	brazilian
0.955 0	a gravelly male vocalist	hip hop
0.947 4	a dry recording sound	a tight kick sound

由表 1、表 2 的结果可以看到,在有考虑特征集与标注对之间关系的情况下,结果是有差异的,对最后的结果也必然有影响。

5.2 标注结果分析

表 3 为对音乐自动标注性能的比较,GMM 是使用 ReliefF-PCA 特征降维后改善了的结果,具体实现请参照文献[10]。

表 3 标准结果分析

	CML(给单标注和成对标注同时建模)	CMLF(给单标注和成对标注同时建模)	GMM (给单个标注建模)
Acc	76.82	79.42	60.12
μF_1	39.61	40.12	26.70
MF_1	23.53	25.61	19.82

通过表 3,可以得到:使用 CRF 模型给成单标注和成对标注同时建模的方法比使用高斯混合模型给单个标注建模的方法效果要好得多,可见考虑标注的依赖关系对语义自动标注的性能有很大的改善。

另外,对比 CML 模型与 CMLF 模型的结果可以知道,对大数据集,使用特征集比不使用特征集建模效果要好,但相应的,算法的复杂度也很高。

6 结论

多标注分类问题^[11-12]在许多领域都是一项很重要的任务,而在现实生活中,类之间通常都是不独立的。文章从标注相关性的角度,给单标注和具有高相关性的成对标注建立两种条件随机场模型 CML 和 CMLF,并与传统的生成式模型 GMM 进行比较,得到 CML 和 CMLF 的高标注性能。但是,这两种模型都是复杂度很高的模型,所以未来考虑根据标注对的相关矩阵,将标注集划分为多个类,分别对每一类中的标注对建模,期待能在标注分类问题上有所改善。

参考文献

- [1] AHSAN H, KUMAR V, JAWAHAR C V. Multi-label annotation of music [C]. 2015 Eighth International Conference on Advances in Pattern Recognition (ICAPR), IEEE, 2015: 1-5.
- [2] NAM J, HERRERA J, SLANEY M, et al. Learning sparse feature representations for music annotation and retrieval[C]. ISMIR, 2012: 565-570.
- [3] MIOTTO R, LANCKRIET G. A generative context model for semantic music annotation and retrieval[J]. IEEE Transactions on Audio, Speech and Language Processing, 2012, 20(4): 1096-1108.
- [4] DHANALAKSHMI P, PALANIVEL S, RAMALINGAM V. Pattern classification models for

- classifying and indexing audio signals[J]. Engineering Applications of Artificial Intelligence, 2011, 24(2): 350-357.
- [5] HOFFMAN M, BLEI D, COOK P. Easy as CBA: a simple probabilistic model for tagging music [J]. International Symposium/Conference on Music Information Retrieval, 2009.
- [6] WANG Y. The constrained Fisher scoring method for maximum likelihood computation of a nonparametric mixing distribution [J]. Computational Statistics, 2009, 24(1): 67-81.
- [7] ZHANG M L, ZHANG K. Multi-label learning by exploiting label dependency [C]. Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, 2010: 999-1008.
- [8] LI C X. Exploiting label correlations for multi-label classification[J]. 2011.
- [9] DELAYE A, LIU C L. Contextual text/non-text stroke classification in online handwritten notes with conditional random fields [J]. Pattern Recognition, 2014, 47(3): 959-968.
- [10] YIN J, YAN Q, LV Y, et al. Music auto-tagging with variable feature sets and probabilistic annotation[C]. 2014 9th International Symposium on Communication Systems, Networks & Digital Signal Processing (CSNDSP), IEEE, 2014: 156-160.
- [11] 田璟, 郭智, 黄宇, 等. 一种基于概率距离尺度学习的图像标注方法[J]. 国外电子测量技术, 2015 (7): 19-23.
- [12] 夏阁, 印晶晶, 陶秋雨. 最大熵模型在音乐自动语义标注中的应用研究[J]. 电子测量技术, 2014, 37 (12): 32-35.

作者简介

何晓梅, 1992 年出生, 硕士研究生, 主要研究方向为信号与信息处理、音频检索。

E-mail: hexiaomei. hhu@gmail. com

NI 发布高精度的 PXI 源测量单元

全新的低电流 SMU 提供了高达 10 fA 的电流灵敏度

新闻发布—2016 年 7 月 11 日—NI(美国国家仪器, National Instruments, 简称 NI) 作为致力于为工程师和科学家提供解决方案来应对全球最严峻的工程挑战的供应商, 近日宣布推出 NI PXIe-4135 源测量单元(SMU), 其测量灵敏度达 10 fA, 输出电压高达 200 V。工程师可以使用 NI PXIe-4135 SMU 来测量低电流信号, 并利用 NI PXI SMU 的高通道密度、高速的测试吞吐率和灵活性来实现晶圆级参数测试、材料研究以及分析低电流传感器和集成电路的特性等各种应用。

模块化 NI PXI SMU 可帮助工程师来构建紧凑型并行高通道数系统, 在单个 PXI 机箱中提供了高达 68 个 SMU 通道, 而且可扩展至数百个通道来满足晶圆级可靠性和并行测试需求。此外, 用户还可利用高速通信总线、确

定性硬件序列生成、以及使用数字控制回路技术对任意待测设备自定义调节 SMU 响应, 从而提高测试吞吐量。工程师还可以通过软件控制 SMU 响应, 避免了长时间等待 SMU 稳定, 并可借助 NI PXI SMU 的灵活性来最小化过冲和振荡, 即使是在带有高电容性的负载的情况下也是如此。

NI PXI SMU 不仅提供了台式 SMU 的易用性, 其交互式软件前面板也可用于进行基本的测量和调试自动化应用。驱动程序包含了帮助文件、文档描述和可立即运行的范例程序, 可辅助测试代码开发, 并包含了一个编程接口, 可支持 C、Microsoft .NET 和 LabVIEW 系统设计软件等各种开发环境。工程师还可结合 NI TestStand 测试管理软件使用 NI PXI SMU, 简化实验室或生产车间的测试系统开发和部署。