简析基于内容的音乐检索

黄　磊，冯　寅
（厦门大学 智能科学系 人工智能研究所　福建 厦门　361005）

【摘　要】：基于内容的音乐信息检索就是把音乐本身的内涵作为查询条件，对一个由不同格式音频媒体所构成的音乐数据库实施查询的过程。文章全面总结了这一领域所取得的主要研究成果，重点介绍了在音频符号序列音乐检索和波形音乐数据检索中所使用的各种智能分析检索方法，并列举了当前较具规模和影响力的项目和系统。

【关键词】：计算机音乐 基于内容的音乐检索 特征提取

1. 引言

根据音乐片段从海量音乐数据库中快速找到对应的音乐，是一个具有很大的实用价值且极具挑战的研究课题。随着基于内容音乐检索系统逐步获得应用，将给广大喜爱音乐的听众带来很多方便。所谓基于内容（content-based）的音乐检索就是把音乐本身的内涵，比如：节奏、旋律片段、和弦、伴奏音型、甚至音色（声音的感觉特性。发音体的振动是由多种谐音组成，其中有基音和泛音，泛音的多寡及泛音之间的相对强度决定了特定的音色。例如，用不同的乐器演奏或人声唱出同一旋律就会对人形成相同旋律但不同音色的听觉）作为查询条件，对一个由不同格式的音频媒体（如：MIDI、MP3 或 WAV）所构成的音乐数据库实施查询的过程。

对于很多人来讲，记住音乐的旋律比记住曲名或演唱者要来得容易得多。许多耳熟能详的音乐，如儿歌或民歌等，我们都不知道曲名和演唱者，但我们往往能哼上一两句。显而易见，单单从书目数据、文本注解来进行音乐检索是远远不足的，也不够有弹性。此时便发展出了本文开始所述的基于内容的音乐检索方法。这个方法主要是针对音乐的内容（如旋律、节奏、响度、音符、乐器特征等）进行检索。使用者只需哼上或输入复制现有的一段音乐，系统便能对比音乐数据库中的所有音乐，根据该音乐片段找出最相近的音乐曲目。

2. 当前存在的一些问题和困难

基于内容的音乐检索要求能够使用记忆不全或表达不清的音乐片段进行音乐检索。然而在实际检索过程中往往面临查询不匹配、查询表达困难以及结果的表现形式等问题。

首先，一首曲子可以用任何一个音高起唱，而检索得到的必须是同样的曲子。这样的情形就像在 KTV 唱歌唱不上去了，便调整伴奏曲的调（key）一样。因此音乐检索系统必须能允许使用者以任何的调作哼唱查询，并依此能找到正确的曲调，达到"音调无关"[1]（key-independent）的查询。

其二，使用者查询时可能无法正确记忆曲调而输入了不完整的片段，甚至不完全正确的曲调。这种情形在传统的检索领域（如文本检索），被称为"语汇不匹配问题"[2]（vocabulary mismatch problem）。也就是说，对于同样的概念，使用者所使用的检索关键字与系统中所记录的检索词不同，从而造成了检索的失败。因此，"模糊比对"在音乐检索领域显得特别重要。

其三，现有的基于内容的音乐检索系统多集中在以音频序列（主要是以 MIDI 音频）文件格式构成的数据库上实施查询。虽然，作为查询条件的输入可以是具体的波形音频音乐文件，但大多都是单声部的音乐（如通过麦克，让人哼唱输入）的。而较少涉及在由多声部音乐混合构成的波形文件（如 MP3、WAV）格式。这主要是因为把由多声部音乐混合而成的波形文件的不同声部之音乐内容分开并进一步分析各声部的音色、音高、时值和节奏，在技术上其实是极其困难的事。

3.主要的研究方法和技术

我们可以大体上将基于内容的音乐检索分为两类：音频符号序列音乐的检索（searching symbolic data）和波形音乐数据的检索（searching audio data）。由于 MIDI 等作为音频序列的存储格式，相对来讲比较结构化，它比原始的波形音频数据来得容易处理，因此目前相关的研究多使用以音频符号序列为存储结构的音乐（如 MIDI）作为处理对象。但由于实际场合中广泛使用的是波形音乐数据，而它的内容和表现力也远远超过 MIDI 一类的文件所能描述的内容，因此检索波形音乐数据也正逐渐成为该研究领域的一个热点。我们将分别介绍当前音频符号序列音乐的检索和波形音乐数据检索中所使用的一些方法与技术。

3.1 音频符号序列音乐的检索

3.1.1 单声部旋律的基于串的检索算法

无论是以波形音乐或是 MIDI 音乐作为检索条件输入的单声部音乐，它最终都可以转换并被表示成一维的字符串，其中每个字符都描述了一个或一对连续的音符。而为了解决用户使用时任意起调的问题（如用户哼唱输入的情况），现有系统大多沿用文献[3, 4]的方法，采用特征变化的幅度来描述特征，从根本上来讲，这种方法的检索过程就是将输入的单声部音乐经转换后得到的查询字符序列和 MIDI 音乐数据库中的相应旋律声部所对应的字符序列相匹配，由于这与文本检索很相似，一些文本检索的串匹配算法和检索串算法都可以被使用。

一些检索系统在实现中采用了精确匹配的方法，这时输入的字符串必须是音乐数据库中相应串的子串。但由于音乐的演奏形式是经常变化的，因此精确的匹配在旋律的比较上会产生许多问题，例如，民族乐曲的演奏就有许多变体，流行歌曲的演奏变化就更多了。所以在检索中，很重要的一点就是允许模糊检索。通常来说，模糊匹配与精确匹配在效率上相差很大，当进行二叉树搜索时，耗时将随数据库的增长呈对数增长。实际使用时，为了提高检索速度，一般从两方面进行解决，其一是寻找更快的模糊匹配算法，其二是建立旋律特征索引数据库，用特征索引数据库对主数据库进行快速索引。

3.1.2 复调旋律的检索算法

复调音乐是两个或两个以上各自具相对独立意义旋律在运动中同时结合在一起的音乐。由于在同一时刻有多个音符，因此相对于单声部音乐，在技术上其旋律、音色、音高、时值和节奏等特征的分离与提取就变得较为困难。现阶段可以考虑下面两种方法对复调音乐进行处理和检索，但对这些方法和技术的研究仍处于初级阶段。

其一、采用 n 元文法模型（n-grams）。

n 元文法模型是被广泛用于文本检索和语音处理的一种技术。其原理是在计算词序概率时为了避免参数过多，而只考虑有限的记忆能力，仅仅记住前那 n 个词，即为马尔科夫近似。此时 n-1 阶的马尔科夫近似就被称为 n 元文法模型。理论上，n 越大，效果就越好，但 n 的增加会使所需的参数迅速增多，以致很难进行估计，所以一般根据经验为 n 取合适的值。

在音乐信息检索中，同样可以利用这一模型对音乐数据进行索引和匹配。我们将每首乐曲都被转化为关于旋律和音程的序列，使用滑动窗口将序列划分为固定长度为 n 的字串，最后将这些 n-grams 编码为文本词汇或音乐词汇。一首乐曲就可表示

为此种 n-gram的有序列。这样就可以较为容易地使用现有的文本检索引擎及其相应的 n 元文法模型算法对其进行检索。Doraisamy 等人采用了 n 元语法模型开发的一套复调音乐检索原型系统在健壮性和可操作性上都取得了不错的成果[5,6]，基于存储了 10000 首复调音乐的数据库进行了实验性的检索，其检索质量上并不低于单声部的音乐检索。

其二、采用基于集合的运移距离(Transportation Distance)。

不同于基于字符串的匹配，基于集合的方法并不假设音符是顺序排列的，相反它将所有的表示为带有权值的点的集合。每个音符就是一个具有权值的点，其坐标用发音时间和音高来表示，而权值可以是该音符的持续时间，也可以根据不同的情况用音强、旋律、和弦甚至是多种特征的组合来表示。通过相应的算法比较这些集合之间的距离来进行匹配。

3.1.3 基于统计模型的检索算法

概率匹配就是将输入查询的概率属性与音乐数据库中的候选项的对应属性进行比较。现阶段此类系统大多采用了隐马尔可夫模型 HMM(Hidden Markov Mode)，如 GUIDO 系统[7]，其优点在于能够处理用户哼唱的音高误差和时间误差，使系统更健壮，具有更好的容错性。在面对不同的查询质量时，不会有太大的波动。该模型包含两个随机过程，其中是一个隐藏的(不可观察的)具有有限状态的马尔可夫链，另一个是与马氏链状态相关联的随机函数集(可观察的)。

为了提高检索的效率，状态转移矩阵有时还被组织成一棵树，其中叶子节点表示音乐数据库中每个音乐片段的转移矩阵，而内部节点则是其子树表示的音乐片段串连后的所得到的转移矩阵。

3.2 波形音乐数据的检索

波形音乐不同于音频符号序列的音乐，无法将其直接结构化，因此如何提取乐曲的信息及提取哪些信息能较好的描述乐曲一直是研究热点和难点。而这一领域的研究也才刚刚起步，无论从技术还是研究进展而言，都较之基于音频符号序列音乐的检索来得缓慢。

典型的波形音乐检索系统主要包括两大部分：建库过程和查询过程。建库过程包括特征的提取和计算(如能量分布、音乐轮廓等)、特征的转换(如"乐纹")、特征库构建。为了快速查询，特征的存储采用 hash 表结构。把从音乐库的每首歌曲提取的特征按照 hash 表的数据结构保存至特征库中。最终完成把容量庞大的音乐库转换成特征库的工作。考虑音乐库里的每首歌曲的特征与用户所输入的样本片段的特征的交集，这个交集不为空集的所有歌曲都可以成为"匹配候选者"(matching candidate)。首先求出样本片段的匹配候选者集，然后利用一定的度量算法计算出每个候选者与输入片段之间的"距离"(distance)。经过排序以后选择距离最小的前 N 首歌曲作为检索结果的输出。

由于波形音乐内容不仅仅包含音乐的一些基本特征，还蕴含着演奏者的情感等因素，并且这些因素对于提高检索质量具有极大的作用，因此如何挖掘乐曲深层次的内容也正引起越来越多的关注。

3.2.1 基于感知特征的音乐检索

音乐是人类情感和精神生活的创造表现，任何音乐表现形式都包含着特定的情感和思想。因此一种自然的有意义的比较波形音乐数据的方法就是抽取出音频信号中抽象的描述，使其可以反映波形音频数据的感觉认知等相关信息，并能充分地代表曲目的基调和情感，然后对抽取的这些信息进行距离等相似度的比较，在比较的基础上分类或检索出相似的波形音频数据。Erling Wold 列举了几种可在 25 毫秒至 40 毫秒的音频数据中提取的信息[8]，如响度、音调、音色、带宽等

目前的检索系统一般先对音频波形进行加窗处理形成帧，并对每一帧作离散傅立叶变换(DFT)，实际上常用快速傅立叶变换(FFT)，得到傅立叶系数和频域能量，再应用不同算法计算提取出所需的相关特征。最后将音乐数据库中所有的音频与查询输入的音频进行相似度计算，选取若干最相似的音频返回给用户，完成检索。

3.2.2 基于"乐纹" 的检索算法

"乐纹"是能够代表一段音乐的有效特征序列[9]。用过"乐纹"我们可以将容量很大的音乐用有限长度的字符序列来表示。如同人的指纹一样，"乐纹"在一定程度上能够代表一段音乐，相同或相似的音乐具有相同的指纹。关于乐纹的提取，学者们已经进行了很多的研究工作。广泛被使用的方法是从经过短时-傅里叶变换(Short-time Fourier Transform)以后的频谱图里面选择一些特征序列为该片段的乐纹。

乐纹的提取过程一般分为预处理、傅立叶变化、帧级特征点的提取、组成特征点对几个阶段。预处理是归一化过程，这个过程将输入的音乐片段转换为特定格式。分帧后进行 STFT 或 FFT 变换，然后从频谱中选择峰值点(Peak)作为特征点。乐纹的质量在很大程度上取决于特征的选取，由于音乐具有很多的特征，包括物理特征和情感特征，实验表明，情感特征的作用往往大于物理特征，这也体现了情感特征提取研究的重要性。

3.2.3 基于集合的检索算法

这一方面的研究刚刚起步，典型的是 Clausen 和 Kurth 将基于集合的方法用于波形数据的检索中[10]，他们使用一种特征提取器将 PCM 信号转化为集合形式，使之处理起来与音符集合无异。这一研究也提醒人们基于乐谱的音乐检索中使用的一些算法可以加以改进后应用于波形音频数据的检索中。

3.2.4 自组织映射检索算法

自组织映射(SOM)是一种近来非常流行的在无监督学习中广泛使用的神经网络算法，已经被应用于聚类相似的音乐片段并将其分类，例如 Rauber 使用特征向量来描述音频中的节奏模式，并使用 SOM 技术对其加以聚类[11]。SOM 由一些被组织成二维矩形栅格的单元组成，而每个单元都被赋予一个高维空间的模型向量。在训练过程中，要使得这些模型向量与其对应的单元之间的距离最短。

4. 相关的系统

在基于内容的音乐查询方面，哼唱检索是研究的主要方向之一，但是目前的研究绝大部分都在音乐数据库中使用音频序列为存储格式(如 MIDI 格式)，只有少数研究关注于原始波形音乐数据。目前较有规模的项目有以下几个。

4.1 英国南安普顿大学的 QBH 系统

近些年来，几乎所有有关音乐内容检索方面的研究报告都引用了英国南安普顿大学在 1995 年 ACM 多媒体研讨会上发表的论文[12]及其开发的一套名为 QBH(Query By Humming)的系统，该系统可以让使用者透过麦克风哼唱来对音乐数据库进行检索。他们通过自相关计算(Auto-correlation)来求出输入声波的基频分布图(Pitch Contour)，并用 U(当前音高比前音高高)、D(当前音高比前音高低)、R(当前音高和前音高相等)来表示当前音高和前后音节之间的相对关系，进而形成字符串用以进行音乐数据库的检索。但该系统的缺陷是并未发展出一套完整的音符切割程序，在使用 QBH 时，使用者需要自行切割。但是他们的研究在直觉式歌唱输入音乐检索领域迈出了重要一步。

4.2 新西兰 Waikato 大学的 MT 系统

新西兰 Waikato 大学的 Rodger J.McNab 和新西兰数字音乐数据库合作开发了一套名为 MT(Melody Transcript)[13,14]的系统，采用 Gold-Rabiner Algorithm[15]找出输入声波的基频分布，并转成标准音符表示。接着，他们将 MT 结合数字音乐数据库，开发成为一套 MELDEX 的系统[16]，让使用者可以透过麦克风哼唱就直接达到检索音乐数据库的目的。他们同时也将音乐数据库的歌曲数目提升至 9400 首左右，正确辨识率约在 77%-89%之间。但是 MELDEX 无法正确地将音符切割开来，因此使用者在哼唱时，在音符与音符之间，必须自行留下小小的间断或多加入"滴答"声音，因此对于非专业人士仍然相当的不便，也相当的不自然。

4.3 CoMIRVA 系统

Markus Schedl[17,18]使用了 SOM 神经网络算法和基于文本特征的技术构造了分析音乐家之间相似性的系统，其中系统中 8000 多首歌曲，并使用不同形式，如柱状图、网状图、概率网络、色块图等，来表现检索结果，提供了一个相当直观和友好的用户界面，使普通用户也可以根据自己习惯的方式对结果进行分析。

5.结论

笔者在本文中尝试对音频符号序列音乐的检索和波形音乐数据的检索中所使用的各种主流技术方法进行总结归纳和分类，希望这些总结能为将来的研究工作提供有意义的参考，从而为我们进一步深入研究指出目标和方向。

基于内容的音乐检索，尤其是波形音乐数据的检索仍然是一个尚未开发的新的研究领域，存在许多可能的发展方向。笔者认为未来的研究重点还应涉及下面几个方面：

· 错误模型的改善，使相应系统能具备较好的健壮性和容错性；

· 确定适当的评价机制和检索结果质量的度量标准；

· 检索算法的复杂性分析。

同时随着人们对音乐检索查询的需求不断加强，音乐检索系统从少数研究者的使用走向大众化将会是一个必然趋势，这也迫切需要使音乐检索系统复合化，不仅用户界面更加友善和易于操作分析，并同时支持哼唱、键盘输入、乐器输入，复制用例等多种查询方式，以满足各种人群的不同需求，适应不同的应用场合。

总之，基于内容的音乐信息检索作为一个新兴的研究领域，同时由于其检索对象和范围的复杂性和多样性，在该领域的研究还只是刚刚开始，任重而道远。

参考文献：

1.Chang, Chuan- Wang, " A Practical Music Retrieval System Based on Sliding Melodic Contour" int. Computer Symposium, Dec. 15- 17, 2004, Taipei, Taiwan.

2. Yuen- Hsien Tseng, "Music Indexing and Retrieval for Digital Music Libraries," Proceedings of The First International Workshop on Intelligent Multimedia Computing and Networking (in The Fifth Joint Conference on Information Sciences), Feb. 27 to Mar. 3, 2000, Atlantic City, NJ USA, Vol. 2, pp.533- 536.

3.A J Ghias, Logan D Chamberlain, B C Smith, "Query by humming- musical information retrieval in an audio database". ACM Multimedia'95, San Francisco, 1995

4.R J McNab, L A Smith, I H Witten et al. "Towards the digital music library: True retrieval from acoustic input". The ACM Digital Libraries Conference(Digital Libraries'96), Bethesda Maryland, 1996

5.Doraisamy, S and Rilger, S. "Robust Polyphonic Music Retrieval with N- grams", Journal of Intelligent Information Systems, 21(1), 53- 70.2005

6.Doraisamy, S and Rilger,S. "A Polyphonic Music Retrieval System Using N- Grams",2006

7.HH Hoos, " GUIDO/MIR - an Experimental Musical Information Retrieval System based on GUIDO Music Notation". Symposium on Music Information Retrieval: ISMIR, 2001 - music- ir.org

8.Erling Wold, "Content Based Classification, Search, and Retrieval of Audio". IEEE MULTIMEDIA, 1996.

9. J. Haitsma and T. Kalker, "A Highly Robust Audio Fingerprinting System", 3rd Int. Symposium on Music Information Retrieval (ISMIR), Oct. 2002.

10.Michael Clausen, Frank Kurth, Roland Engelbrecht, "Content - based Retrieval in MIDI and Audio". ECDL WS Generalized Documents 2001.

11. A Rauber, E Pampalk, D Merkl, "The SOM- enhanced JukeBox: Organization and Visualization of Music Collections Based on Perceptual Models". Journal of New Music Research, 2003 - Taylor & Francis.

12.A. Ghias, J. Logan, D. Chamberlain, B. C. Smith, "Query by humming- musical information retrieval in an audio database", ACM Multimedia '95 San Francisco, 1995.

13.Roger J. McNab, Lloyd A. Smith, Jan H. Witten, "Towards the Digital Music Library: Tune Retrieval from Acoustic Input" ACM, 1996.

14. Roger J. McNab, Lloyd A. Smith, Jan H. Witten, "Signal Processing for Melody Transcription" Proceedings of the 19th Australasian Computer Science Conference, 1996.

15.Gold, B. and Rabiner, L. "Parallel processing techniques forestimating pitch periods of speech in the time domain," J. Acoust. Soc. Am. 46 (2), pp 442- 448, 1969.

16.Roger J. McNab, Lloyd A. Smith, "Melody transcription for interactive applications" Department of Computer Science University of Waikato, New Zealand, 1996.

17.Schedl, M., Knees, P., and Widmer, G, " Interactive Poster: Using CoMIRVA for Visualizing Similarities Between Music Artists" Proceedings of the IEEE Visualization 2005 (Vis'05), Minneapolis, Minnesota, October 2005.

18.Schedl, M, " The CoMIRVA Toolkit for Visualizing Music- Related Data" Technical Report, June 2006.

(上接第 34 页)

题库"模块，或通过"讨论交流"模块与同学协作解决，或通过"提问"模块直接向教师请教。例如，学生调试程序时经常出错，屡次修改仍无法解决，教师可鼓励学生通过讨论交流寻求帮助。在讨论交流过程中，学生通过相互纠错，编程能力和调试技巧均得到很好训练。另一方面，教师通过"课程分析"模块，了解学生作业完成情况和提问情况，调整教学内容和方法。此外，网络课程的"拓展资源库"，对学生课后的知识拓展有重要意义。

3.4 多样化的学习评价

评价是依据一定的标准对学习者的业绩所进行的一种鉴定或价值判断[1]。采用多样化的评价，有利于促教和促学。在基于网络课程的教学中，教学评价包括诊断性评价、形成性评价和总结性评价。评价的方法除了测试，还有学生自我评价、学生互评等多种形式。

4. 教学反思

笔者采用网络课程进行教学实践两年，不仅解决了教师以往课时紧任务重的困难，而且大部分学生学习兴趣浓，动手能力明显增强，计算机等级考试通过率明显提高，教学效果良好。但在教学实践初期，部分学生由于习惯传统的"以教为主"的教学模式，难于接受基于网络课程的自主学习活动，对基于网络课程的教辅活动重视不够，置身事外。为使网络课程更好的服务于教学，必须建立良好的激励机制。首先，教师在课堂教学中，精心设计基于网络课程的教学，培养和鼓励学生使用网络课程进行学习；其次，把学生利用网络课程学习的积极程度做为学习评价的一个指标；再次，教师可参与学生的讨论交流，用幽默亲切的语言鼓励学生，也起到不错效果；最后，加强网络课程的建设，增强网络课程的教学功能，使学生爱用乐用。

5. 结束语

通过开展基于网络课程的教学实践，不仅提高了教学效率和教学质量，更培养了学生利用网络进行自主学习和协作学习的学习习惯，对日后的学习生活有重要意义。

参考文献：

1. 谢幼如，柯清超 .网络课程的开发与应用[M] .北京：电子工业出版社，2005 .