

●韩圣龙

音乐旋律表示及匹配方法浅议

摘 要 音乐旋律的表示和匹配方法是基于旋律的音乐组织与检索的重要环节。音乐旋律的表示方法大致分为两种:一是描述旋律时忽略乐音在时间上的延续性,只描述乐音随时间的变化;二是同时考虑音乐旋律所包含的两方面的内容。从信息管理的角度看,对音乐旋律进行描述的最终目的是建立旋律数据库,对音乐信息进行组织和检索。进行音乐旋律描述时可以忽略旋律的时间属性,这个旋律表示法的基本思想是直接使用音符对应的绝对音高来表示音乐旋律。图1。参考文献12。

关键词 音乐旋律 音乐检索 旋律表示法 匹配方法。

分类号 G254

ABSTRACT Representation and matching methods of musical melodies is an important part of melody-based music organization and retrieval. The author proposes two methods for the representation of musical melodies, and thinks that from the point of view of information management, the ultimate objective for the representation of musical melodies is to develop melody databases and realize the organization and retrieval of music information. We can ignore the time attribute in the description of musical melodies, and use absolute pitches to represent musical melodies. 1 fig. 12 refs.

KEY WORDS Music melody. Music retrieval. Representation of melody. Matching method.

CLASS NUMBER G254

“乐音随着时间的延续、变化,构成旋律”^[1]。从这个表述中可以看出,一段音乐的旋律包含两方面的内容:一方面是乐音在时间上的延续,另一方面是乐音随时间的变化。要表示或描述一段音乐的旋律,就是要使用某种方法把这两方面的信息如实地表达出来。音乐旋律的表示方法和匹配方法是基于旋律的音乐组织与检索的很重要的一环,对音乐旋律的表示和匹配方法的探讨是进行音乐信息组织和检索研究的重要组成部分。

1 现有旋律表示和匹配方法

目前的音乐旋律表示方法大致可以分为两类:一类是描述旋律时忽略乐音在时间上的延续性,只对乐音随时间的变化进行描述;另一类表示法则同时考虑音乐旋律所包含的两方面内容,即乐音在时间上的延续性和乐音随时间的变化。

1.1 忽略乐音时间延续性的旋律表示和匹配法

(1)“旋律轮廓”表示法及其匹配方法。

使用这类旋律表示法的一个杰出代表是新西兰怀卡托大学(Waikato)的MEIDEX系统^[2]。它采用的旋律表示法完全忽略乐音在时间上的延续性,只关心乐音的音高随时间的变化。这种方法的基础是Parsons在1975年提出的“旋律轮廓”概念,以及Dowling在1978年做的一个实验结果:独立于音程大小的音

程方向是影响旋律识别的一个重要因素。“旋律轮廓”是指由一个音乐旋律相邻两个音的音程方向组成的音程方向序列。这个概念不仅完全忽略了乐音的时间延续性,而且忽略了音程的准确数值,只保留了音程的方向。对于一个具体的音乐旋律,它用*表示旋律开始的音符,用U表示上升的音程,用D表示下降的音程,用R表示0音程,也就是重复的音符。

这种方法的优点是允许用户以不同的基调哼唱检索提问旋律。“旋律轮廓”不受哼唱旋律的绝对音高的影响,使检索用户不必记住音符间准确的音程差,只要记住音程的方向,也就是“旋律轮廓”就行。

对旋律轮廓进行匹配最简单的办法是检测用户提问在被检索的旋律中出现的次数,出现次数越多,就说明该旋律所代表的音乐文件和检索提问越相关。但是当旋律数据库很大时,这种匹配算法的效率不是很高。为此,日本东京大学的学者提出了一种N字符串索引法^[3]。对一段有3个音符组成的旋律轮廓,一共有9种排列方式,如图1所示。

把这9种排列方式看做9个分量,数据库中的每个旋律就可以表示成为一个有9个分量的旋律向量,每个分量的数值就是该分量在该旋律中的出现频次,具体形式{3,1,7,0,0,3,5,5,0}。这样,在检索之前就可以先创建一个旋律索引。同样道理,用户输入的检索提问也可以表示成同样形式的向量。进行检索

时,把用户提问向量和旋律索引中的旋律向量进行比较,向量之间的距离就是它们的相似度。这种方法把大量处理工作放在标引阶段,可以大大提高检索效率。为了增加旋律索引的分辨力,可以增加音符的数量,但是这会导致运算次数的指数上升。比如,4个音符的旋律轮廓的可能排列就是27个,而5个音符的旋律轮廓的可能排列就会达到81个。东京大学的学者在实验中发现,当音符数超过4的时候,已经不能明显提高检索效果了。可见,并不是音符数越多越好。

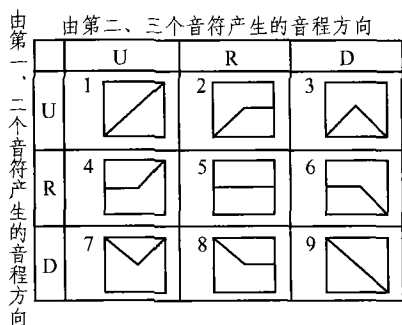


图1 三音符“旋律轮廓”排列方式示意图

尽管上述方法的检索效率有很大提高,还是不能解决用户提问中遗漏或错误添加音程的问题。MELDEX的研究人员采用了由Wagner和Fischer在1974年提出的一种文本字符串的模糊匹配算法^[4-7]。这种算法的思想是:对给定的两个字符串,寻找一组最经济的操作,把其中一个转化成另外一个。这组操作的成本就是两个字符串的相似度。转换成本越低,两个字符串的相似度就越高。转换操作包括插入、删除和替换。在上述算法思想的基础上,Sankoff和Kruskal于1983年给出了把串a转换成串b的成本的计算公式: $d_{ij} = \min[d_{i-1,j} + w(a_i, \varphi), d_{i,j-1} + w(a_i, b_j), d_{i,j+1} + w(\varphi, b_j)]$ 。

其中 $1 \leq i \leq \text{串 } a \text{ 的长度};$

$1 \leq j \leq \text{串 } b \text{ 的长度};$

$w(a_i, \varphi)$ 是向串b中插入 a_i 的成本(或权重);

$w(\varphi, b_j)$ 是在串b中删除 b_j 的成本;

$w(a_i, b_j)$ 是用 a_i 替换 b_j 的成本。

初始条件是:

$d_{00} = 0,$

$d_{i0} = d_{i-1,0} + w(a_i, \varphi),$ 其中 $(i \geq 1);$

$d_{0j} = d_{0,j-1} + w(\varphi, b_j),$ 其中 $(j \geq 1);$

这个算法常常被设计成对称的,也就是说,把串a转换成串b的成本应该和把串b转换成串a的成本

相同。这样,插入操作和删除操作的成本应该相同,并且 $w(a_i, b_j) = w(b_j, a_i)$,而对替换字符本身则没有成本,即 $w(a_i, b_j) = 0, (a_i = b_j)$ 。各种操作的成本可以根据实际情况赋值。

(2) 音程表示法及其匹配方法。

美国伊利诺依大学 Urbana-Champaign 分校的 J. Stephen Downie 在他的研究中使用了4种方法来表示音乐旋律^[8]。为了避免和 MELDEX 系统使用的“旋律轮廓”表示法相混淆,把这组表示方法称为音程表示法。Downie 提出的这组旋律表示法中,重复的音符都用小写字母 a 表示。

第一种表示法的思想和 MELDEX 系统使用的“旋律轮廓”表示法相同,也就是用音程方向来表示旋律;不同的是在这种表示法里,上升的音程用大写字母 B 表示,下降的音程用小写字母 b 来表示。

第二种表示法在表示了音程方向的同时使用了部分具体音程来表示音乐旋律,用大写字母 B、C、D 分别表示上升1、2、3个半音的音程,上升超过3个半音的音程用大写字母 D 来表示;用小写字母 b、c、d 分别表示下降1、2、3个半音的音程,下降超过3个半音的音程用小写字母 d 来表示。

第三种表示法在第二种表示法的基础上前进了一步:用从 B 到 H 的大写字母分别表示上升1到7个半音的音程,上升超过7个半音的音程用大写字母 H 来表示;用从 b 到 h 的小写字母分别表示下降1到7个半音的音程,下降超过7个半音的音程用小写字母 h 来表示。

第四种方法则直接使用音程的准确数值来表示音乐旋律。

和 MELDEX 系统的音乐旋律表示法相比,Downie 提出的旋律表示法中的后3种表示法,专指性更强,对旋律的分辨能力也更强。当然这3种表示法也有不足,就是用户在提出检索提问时必须部分或全部记住旋律的准确音程,这无疑增加了用户的负担。因此,在 Downie 的检索实验中使用的检索提问不是用户哼唱的,而是由 MIDI 文件中自动抽取的。

用准确音程来表示音乐旋律的音程序列,具有和文本信息相近的属性,可以把它们看成一个字符串。使用这种方法表示音乐旋律的研究大都采用了文本字符串的匹配方法来匹配用户的检索提问和数据库中的旋律。例如,Downie 的试验就采用了 SMART 系统的匹配算法和检索结果的排序方法。

这种检索匹配方法不适合基于旋律的音乐检索研究,因为音程序列和字符串毕竟不同。字符串“BBbabB”和“BbabBa”是不同的,文本检索会认为这不相关,而对于音乐来说,这两个音程序列则有着很

强的相关性。

(3) 简谱音符表示法及其匹配方法。

上海交大图书馆在他们的基于内容的音乐检索系统中使用简谱的音符来表示音乐旋律^[9]。

研究人员把旋律 $3\ 2\ 3\ |\ 5\ -\ | 6\ 5\ 1\ 6\ |\ 5\ -\ | 3\ 5$ 直接表示为 32356516535。作者把这种旋律表示法称为简谱音符表示法。

这种表示方法和 Downie 提出的第4种表示法有相似之处,都是由乐谱直接经过处理而得到,专指性非常强。但是这种表示法有一个很大弱点:它只用从1到7这7个数字来表示乐谱,不能区分音程为八度的整数倍的两个音。也就是说,5和 $\dot{5}$ 在这种表示法里是没有区别的。这势必造成检索上的误差。直接用简谱来描述旋律的简谱序列,用户提问和旋律数据库中的旋律的匹配方法完全借用文本信息检索的匹配算法,也就是把简谱序列完全当成了字符串来处理。

1.2 考虑乐音时间延续性的旋律表示法

(1) <TPB>表示法及其匹配方法。

美国 MIT 媒体实验室的研究人员在他们的研究中使用一个三元组 <TPB> 来表示一段旋律^[10]。T 表示旋律的拍号, P 表示用 U/D/R 表示的旋律轮廓, B 表示乐曲的节拍。用这种方法表示的旋律要对用户哼唱旋律的质量提出更高要求,因而在这个研究中使用的检索提问也是由 MIDI 数据库中随机产生的。

<TPB> 旋律表示法的匹配算法是这样的:首先被检索旋律和用户提问都被表示成 <TPB> 的形式,匹配时先比较用户提问和被检索旋律之间的拍号 T 是否相同,如果不同,匹配过程中止,相似度为0;如果相同,则逐小节比较用户提问和被检索旋律的“旋律轮廓”P,得到相似度(方法同前述“旋律轮廓”模糊匹配法);每次向后移动一小节,重复上一步,直至被检索旋律的结尾,每次得到一个相似度;在得到的相似度序列中找出最大值作为最终的相似度。

(2) 音程/节奏比率表示法及其匹配方法。

英国 Imperial College 的研究人员进行复调 MIDI 音乐文件的旋律处理时,同时考虑了旋律包含的两方面内容,用 [IRIR……IRI] 的形式表示音乐旋律。其中 I 表示音程, R 表示两个音程代表的时间的比率,也就是节奏比率。这种表示法有着和 <TPB> 表示法同样的弱点,因此在这项研究中,检索提问也是由数据库中随机产生的。

用音程/节奏比率表示法表示的音乐旋律只是在 Downie 提出的第4种表示法产生的音程序列中加入了节奏比率的信息,产生的音程/节奏比率序列实际

上和音程序列大同小异,因此,对音程/节奏比率序列的匹配可以直接采用对音程序列的匹配方法。

2 关于音乐旋律表示和匹配方法的讨论

从管理的角度来看,对音乐旋律进行描述的最终目的是建立旋律数据库,对数字音乐信息进行组织和检索。为了达到这个目的,作者认为,进行音乐旋律描述时基本上可以忽略旋律的时间属性值,也就是乐音在时间上延续性。这是因为:

(1) 同一首音乐作品由不同的艺术家来演奏(或演唱),结果可能是音乐旋律的时间属性值大不相同。例如,阿炳演奏的《二泉映月》和宋飞演奏的《二泉映月》在时间属性值上有着很大差异:阿炳演奏全曲总时长为 6'53", 而宋飞为 9'04", 后者长出前者 2'11", 已接近全曲总时长的三分之一^[11]。这种情况下,我们只能说艺术家的演奏(或演唱)风格不同,而不能说他们所表演的不是同一首作品。

(2) 用户用哼唱进行提问时,由于记忆模糊或哼唱技巧不高会导致哼唱的旋律出现错误。对用户哼唱错误的兼容也要求在进行旋律表示时忽略旋律的时间属性值。

(3) 人们在判断两个旋律是否相同或者相似时,也会不自觉地忽略旋律的时间属性值。人们能够很轻松地辨认出别人哼唱的旋律是什么乐曲,不管哼唱的节奏对与不对。

(4) 从对现有的旋律表示法回顾中我们可以看到,考虑旋律的时间属性值的表示方法无一例外地使用了从 MIDI 文件数据库中随机抽取的旋律作为检索提问。这是由于它们不能很好地兼容用户哼唱中的错误。这也从另一个侧面说明了进行旋律描述时忽略旋律时间属性值的必要性。

此外,前面讨论的所有的旋律表示法都是把音乐旋律表示成某种字符串的形式,这就决定了用这些旋律表示法进行组织的旋律数据库,检索时只能用字符串匹配技术来进行匹配。音乐旋律的高低起伏和数字有着天然的亲和性,用数值来表示音乐旋律有着无与伦比的优势。在上述讨论的基础之上,作者为尝试解决现有旋律表示法遇到的问题,提出了一个新的旋律表示法。

3 旋律的数值音名表示法及其匹配方法

这里提出的旋律表示法也忽略了旋律的时间属性值^[12]。它的基本思想是直接使用音符对应的绝对音高来表示音乐旋律。

音高表示法中有一种是把钢琴上音高最低的键

命名为1号键,连同白键、黑键,向右(即音高声高方向)顺序排列,依次用自然数1~88来命名。一般说来,各种乐器的基频音高都不会超过这个范围。我们就用从1到88这88个自然数来表示旋律中相应的音符,忽略休止符,作者把这种方法称为数值音名表示法。

数值音名表示法的优点是把旋律转化成一个直观的自然数序列,自然数的大小直接对应了音符的高低;进行检索时,可以把对两个旋律的比较转换成对数值的处理,它也可以很好地兼容用户哼唱错误。

对使用数值音名表示法表示的旋律进行检索匹配之前,用户提问首先被表示成为数值音名序列 $d_q = \{p_{q1}, p_{q2}, \dots, p_{qn}\}$ 。然后从被检索旋律开始位置取出和用户提问等长的音名序列 $d_m = \{p_{m1}, p_{m2}, \dots, p_{mn}\}$ 定义两个音名序列的距离为:

$$D = (\sum |p_{mi} - p_{qi}|) / n$$

D的值越小,就说明两个音名序列的相似度越大。每次向后移动一个音名,重复上述过程,直至被检索旋律结尾。在得到的距离集合中选取最小值作为用户提问和被检索的旋律之间的距离。

这个算法把字符串匹配转化成了数值计算,简化了检索匹配过程,提高了检索效率。这个算法还能很好地兼容用户输入的检索提问中的错误。用户输入的检索提问(以用户哼唱为例)中的错误一般有3种:哼唱的音高不对、漏唱音符和多唱音符。下面我们以前《二泉映月》(1=g)中的如下一段为例加以说明。

$$1 - \boxed{\dot{1} \underline{\underline{6}} \dot{1} \dot{3} \dot{3} \dot{2} | \dot{1} \underline{\underline{6}} | \dot{2} \dot{3} \dot{3} \dot{2} \dot{1} \dot{1} | \underline{\underline{6}} \dot{1} \dot{2} \dot{3} |}$$

这段旋律对应的数值音名序列为{35,47,44,47,51,51,49,47,44,47,49,51,51,49,47,47,44,47,49,51},共20个音符。假定用户要哼唱的是上图中矩形框内的部分,此时n=8。

(1) 没有错误的情况:

$d_q = \{47, 44, 47, 51, 51, 49, 47, 44\}$ 。进行计算得到距离集合{3.63, 0.00, 2.50, 3.38, 4.38, 3.88, 3.13, 1.13, 1.38, 1.63, 3.38, 3.88, 3.88},用户提问与被检索的旋律之间的距离是0,匹配成功。

(2) 出现音高错误的情况:

此时假定用户哼唱时把高音re唱成高音dol,即 $d_q = \{47, 44, 47, 51, 51, 47, 47, 44\}$ 。进行计算得到距离集合{3.88, 0.25, 2.25, 3.13, 4.13, 4.13, 3.38, 1.38, 1.63, 1.38, 3.13, 3.63, 3.63},用户提问与被检索的旋律之间的距离是0.25。

(3) 出现漏唱音符的情况:

假定用户哼唱时漏唱了高音re,即 $d_q = \{47, 44,$

$47, 51, 51, 47, 44\}$ 。进行计算得到距离集合{4.43, 0.71, 1.71, 3.29, 4.14, 4.14, 3.57, 1.57, 1.86, 2.00, 2.71, 3.86, 3.57, 3.57},用户提问与被检索的旋律之间的距离是0.71。

(4) 出现多唱音符的情况:

假定用户哼唱时多唱了一个高音re,即 $d_q = \{47, 44, 47, 51, 51, 49, 49, 47, 44\}$ 。进行计算得到距离集合{2.67, 0.89, 2.67, 3.67, 4.11, 3.44, 2.56, 0.78, 1.11, 2.33, 3.44, 4.11},用户提问与被检索的旋律之间的距离是0.78。

由此可以看出,用户哼唱中的错误对这种匹配算法的影响不是很大。

此外,对每个数值音名序列进行简单处理就能够得到相应的旋律轮廓,形成旋律数据库,为旋律数据库中的旋律创建旋律向量索引,就可以应用旋律轮廓匹配算法进行检索。也就是说,对于用数值音名表示法得到的音名序列也可以使用旋律轮廓匹配算法进行检索。

4 结语

本文对音乐旋律表示和匹配方法进行了较深入的分析 and 总结,并在分析结论的基础上提出了新的音乐旋律表示和匹配方法。文中提出的数值音名表示法和相应的匹配算法已经应用于中国艺术研究院音乐研究所承担的北京市科技计划项目“‘基于旋律的数字音频音乐信息检索系统’的研制与应用”,并取得了较好效果。由文中的讨论我们可以看到,音乐旋律的表示方法有很多种,并且各有特点。组织音乐数据库时,可以根据数据库的使用对象和使用目的,选择使用不同的旋律表示法并建立相应的索引。

参考文献

- 唐林等. 音乐物理学导论. 合肥: 中国科技大学出版社, 1991: 72
- Rodger J. McNab, Lloyd A. Smith, David Bainbridge, and Ian H. Witten. The New Zealand Digital Library ViELody index. [2007-01-14]. <http://www.dlib.org/dlib/may97/meldex/05witten.html>.
- Takashi Yanase, Atsuhiko Takasu, and Jun Adachi. Phrase based feature extraction for musical information retrieval. 1999 IEEE Pacific Rim Conference on Communications, Computers and Signal Processing. 1999. P. 396-9
- David Bainbridge, Rodger J. McNab, and Lloyd A. Smith. Melody based tune retrieval over the worldwide web. [2007-01-14]. <http://citeseer.nj.inet.com/271349.htm>.
- David Bainbridge, Craig G. Nevill-Manning, Ian H. Witten, Lloyd A. Smith, and Rodger J. McNab. Towards a Digital Li-

- brary of Popular Music. [2007-01-14]. <http://citeseer.nj.nec.com/bainbridge99towards.htm>.
- 6 Rodger J. McNab, Lloyd A. Smith, Ian H. Witten, Clare L. Henderson, and Sally Jo Cunningham. "Towards the Digital Music Library: Tune Retrieval from Acoustic Input". [2007-01-14]. <http://citeseer.nj.nec.com/mcnab96towards.htm>.
- 7 Rodger J. McNab, Ian H. Witten, and Stefan J. Boddie. A Distributed Digital Library Architecture Incorporating Different Index Styles. [2007-01-14]. <http://citeseer.nj.nec.com/mcnab97distributed.html>.
- 8 J. Stephen Downie, and Prof. Michael Nelson. Evaluation of a Simple and Effective Music Information Retrieval Method. [2007-1-14]. <http://citeseer.nj.nec.com/380216.html>.
- 9 薛锋, 杨宗英, 郑巧英, 黄敏. 基于内容的音乐检索. 大学图书馆学报, 1999, 17(4).
- 10 Youngmoo E. Kim, Wei Chai, Ricardo Garcia, and Barry L. Vercoe. Analysis of a Contour-based Representation for Melody. [2007-01-14]. <http://citeseer.nj.nec.com/387578.html>.
- 11 韩宝强, 赵文娟, 刘一青. 阿炳所奏《二泉映月》的音律研究. 中国音乐学, 2000(2).
- 12 韩圣龙. 数字音乐自动分析和检索. 中国教育网络, 2005(8).
- 韩圣龙 北京大学信息管理系教师. 通讯地址: 北京大学. 邮编 100871.
(来稿时间: 2007-01-23)

纪念《中国图书馆学报》创刊 50 周年座谈会召开

2007年9月20日上午在中国国家图书馆文会堂,《中国图书馆学报》创刊50周年纪念座谈会隆重举行。

出席座谈会的有国家图书馆名誉馆长任继愈先生,中国图书馆学会理事长、国家图书馆馆长詹福瑞教授、文化部办公厅办公室主任何晓娟女士、文化部社会文化图书馆司图书馆处处长陈胜利等领导。

出席座谈会的图书馆界资深学者和专家有北京大学教授周文骏先生、吴慰慈先生,武汉大学教授彭斐章先生,原文化部图书馆司副司长、北京图书馆副馆长鲍振西先生,南京大学倪波先生,吉林大学管理学院院长靖继鹏先生,原《中国图书馆学报》主编袁咏秋、孙蓓欣、黄俊贵和丘峰,中科院文献情报中心辛希孟、徐引篪教授,以及公共图书馆、高校和专业图书馆的馆长,图情院系的领导和专家教授,图书馆专业期刊的代表70多人。真可谓群贤毕至,高朋满座。

座谈会由国家图书馆张玉辉副馆长主持。他首先介绍了与会领导和来宾,然后宣读了文化部副部长、中国图书馆学会名誉理事长、原《中国图书馆学报》主编周和平先生的贺词和文化部办公厅的贺信。北京大学信息管理系主任王余光教授代表与会嘉宾致词之后,《中国图书馆学报》主编詹福瑞教授代表学报编辑部讲话。他首先简要介绍了学报50年的发展历史,接着介绍了学报的工作情况。他指出,长期以来,特别是近20年来,编辑部坚定地贯彻“坚守定位和宗旨、突出刊物特色、狠抓学术质量”的办刊方针,工作扎实、刊物发挥作用明显,取得了可喜的成

就。他说,学报长期保持优秀的制胜法宝是由于在工作上做到了四个“不放松”,即:坚守定位、贯彻宗旨,突出刊物特色不放松;抓稿件质量、提高刊物学术水平不放松;抓作者队伍建设,保持优质稿源不放松;抓编校作业,制作优良产品不放松。他同时指出,学报今后要努力的方面还很多。首先要不断提高刊物质量,其次要进一步提高刊物对专业学术研究的导向性,第三还要继续加强对外交流的力度。

任继愈先生、周文骏先生、彭斐章先生等来宾相继即席发言。他们充分肯定了学报50年来所取得的成就,并为学报未来的发展纷纷谏言献策,气氛活跃。原中科院文献情报中心主任徐引篪教授、中山大学图书馆馆长程焕文教授等提出了学报要加强对国外图书馆学研究前沿介绍的力度,加快刊物的网络化建设步伐,增强中国学术期刊的国际竞争力和影响力的建议。大家坚信,在作者、编者和读者的共同努力下,学报必将不断进步,更加发展,一定会拥有一个更加美好的明天。

为纪念学报50周年华诞,编辑部组织业内专家学者,从1979年改革开放后复刊,到2006年学报发表的论文中评选出37篇优秀论文。会上由学报常务副主编李万健先生宣读了优秀论文的作者获奖名单,并由詹福瑞主编为作者代表颁发了获奖证书和纪念品。现场展示了1957—2007年出版的全部学报样刊。

会议在隆重、热烈的气氛中圆满落幕。

(许欢)