

文章编号:1674-0386(2009)02-0061-04

基于音乐搜索排行榜测量的研究

杨悦, 陈常嘉, 陈一帅

(北京交通大学 电子与信息工程学院, 北京 100044)

摘要:在现代生活中,音乐、电影和美术已经是不可或缺的重要组成部分。音乐排行榜在音乐工业中也起着非常重要的作用。针对一种新型的排行榜,实时的搜索排行榜进行研究。通过数据的采集与分析,研究这一类型的排行榜的排名特征和用户行为。从研究结果中,发现了搜索排行榜与传统的商业性排行榜的差别就是这一新兴排行榜中存在一部分特殊用户。进而研究了这些特殊用户的用户行为特征以及给排行榜带来的影响。

关键词:搜索排行榜;特殊用户;实时排行榜

中图分类号:TP393 **文献标识码:**A

The Measurement of the Search Charts of Music

YANG Yue, CHEN Chang-jia, CHEN Yi-shuai

(School of Electric and Information Engineering, Beijing Jiaotong University, Beijing 100044, China)

Abstract: Nowadays, more and more singers publish various songs. The songs are refreshed everyday, so we need a real-time chart of music. In this paper, our study is based on the real-time search charts of music. We studied the algorithm of the charts and analyzed the ranking of the charts. According to the ranking of the charts, we got some special characters. Those characters can not be seen from other music charts, such as selling charts and other week's charts of music. While, there are also some problems about the real-time charts, like special users of charts. Those problems need us to solve. We will get more impersonality charts after solving those problems.

Key words: search charts; special users; real-time charts

在现代生活中,音乐、电影和美术已经是不可或缺的重要组成部分。音乐排行榜,也就是所说的音乐热榜,他在音乐工业中起着非常重要的作用,它就像是一个晴雨表,时刻反映着歌手与歌曲的流行程度。世界上有很多的音乐排行榜,其中UK排行榜和Billboard's^[1]排行榜是公认的两个最具权威的排行榜。除此之外,日本的“Oricon”排行榜也得到了音乐界的认可。Giles^[2]曾在他的文章中解释了人们重视音乐排行榜的原因。排行榜是根据专辑的销量进行排名的,排行榜的排

名显示了专辑销量的多少。因此,音乐排行榜就直接反映了音乐市场的信息,同时排行榜还能够影响人们的购买行为,进而影响整个音乐市场。总之,音乐排行榜对音乐工业发展产生了影响。

如今,随着因特网和多媒体技术的不断发展,音乐产品的形式也多种多样。除了购买CD之外,人们还可以通过网络下载、手机铃音、在线听歌等多种服务来欣赏音乐与歌曲。因此,单从专辑销量这一个方面已经无法准确的评价音乐尤其是歌曲的受欢迎程度,销量排行榜提供的市场信

收稿日期:2009-04-10

基金项目:中国科技部973项目(2007CB307101);中国自然科学基金项目(60772043,60672069)

作者简介:杨悦(1981-),女,博士研究生,研究方向:网络行为。

息已经不够全面。时下,各种排行榜层出不穷,其中包括以网络销量为基准的网络专辑销量排行榜,以铃音下载量为基础的铃音排行榜,以及以点击率和搜索率为基础的点击排行榜和搜索排行榜。以搜索排行榜为对象进行研究,通过数据采集和分析获得搜索排行榜的特征。

一、搜索排行榜

(一) 实时搜索排行榜

根据排行榜的更新频率,分为即时更新排行榜和非即时更新排行榜。即时更新排行榜,顾名思义,就是指排名的更新是随时间变化,而且更新周期短。在本文中,把更新周期小于 1 小时的排行榜都称为即时更新排行榜。比如“谷歌”(Google)的热榜,它的更新周期就是 1 小时。而非即时更新排行榜就是更新周期大于 1 小时的排行榜。大部分的以销售量为基准的排行榜,更新周期都是以天为单位。这些排行榜只显示前一天歌曲排名,而无法做到即时更新。

根据排行榜的排名标准,又可以将其分为商业排行榜和非商业排行榜两大类。商业排行榜是指以专辑的销量或者是其他的商业利润来进行排名的排行榜。首先,专辑销量榜就是典型的商业排行榜;再者,其他与音乐有关的收费服务的排行榜也是商业排行榜,如手机铃音的下载排行榜以及短信的支持率排行榜等;另外现在还出现了一些新型的排行榜,他们综合考虑专辑销量,用户点击率,短信支持率等因素进行排名,这些也都属于商业排行榜。非商业排行榜主要依据用户的点击率进行排名,比如谷歌的搜索排行榜就是根据用户搜索频度进行排名。

实时的根据用户的搜索进行排名的排行榜,就称为实时搜索排行榜。这一类的排行榜有两大特点,一个就是它不受商业环境的影响,第二个就是它能够及时反映用户的行为,反映音乐流行趋势的变化。

(二) “naver”排行榜

本文具体的研究对象是“naver”排行榜(直接称其为排行榜 N)。排行榜 N 就是一个实时的搜索排行榜。排行榜 N 是韩国著名网站 www.naver.com 的产物。这一网站占据了韩国市场份额的 70%,也就是说他拥有超过三千五百万的使用者。因为从这个网站上可以获得最新的韩国娱乐

资讯,所以在中国以及其他很多国家, www.naver.com 上的咨询都被翻译和转载。

排行榜 N 之所以受欢迎,有两个原因可以解释。一是它的用户数量很多,它直接反映了人们的兴趣与喜好,它是“韩流”娱乐的风向标。二是因为它特殊的排行机制。排行榜 N 是一个搜索排行榜,他的排名的依据就是用户对于关键词的搜索率。而且他是一个即时更新的排行榜,能够适时地反映新闻、歌曲、人物的受关心程度。据观测,他的更新频率最快可以达到 5s/次。

二、数据与分析

在本文中研究的数据都是来源于排行榜 N。排行榜 N 包括多个种类的排行,比如新闻、人物、电视剧、电影等等,而在本文中,选择音乐类排行榜进行研究。数据采集的时间是一周,从 2008 年 7 月 9 号到 7 月 17 号。采集频率是 20 秒/次。从数据中发现,星期六的数据结果和星期日的数据非常相似,所以就只选取了周一到周六这段时间的数据进行分析。数据从时间上分为两段,一段是从 7 月 9 号中午 12 点开始到 7 月 13 号凌晨零点结束;第二段是从 14 号的午夜零点开始到 17 号中午 12 点结束。由于应考虑到中国与韩国之间有 1 小时的时差,在数据采集时都是提前 1 小时进行,文中所显示的时间均为首尔时间。

由于所研究的排行榜 N 是一个搜索排行榜,所以音乐排行榜中的关键词句保健括了多个类别,比如歌手名、歌曲名、歌词等等。在表 1 中列出了关键词的分类情况。歌手数量为 55,当前流行歌手为 32,普通歌手为 23。在这些检索词中有一类特殊的检索词,就是“歌手名+歌曲名”。根据这一特性对歌手进行了分类,以“歌手名+歌曲名”被搜索的歌手称为当前流行歌手,因为这些歌手大都是最近发行专辑的歌手,他们的名字以及他们的歌曲都很受关注。另外一类为普通歌手。

表 1 检索词的分类情况

检索词类别	检索词数量
总数	758
歌手名	55
歌曲名	578
歌手名+歌曲名	81
歌词	48
其它	6

(一) 排行榜 N 的基本特征

排行榜 N 中即时的显示搜索量处于前 10 位的搜索关键词。根据音乐排行榜的排名变化,可以看出每个歌手以及每首歌曲的受欢迎程度。这里通过数据分析来显示这一信息。

首先来介绍两个歌手,第一个歌手给它命名为歌手 A(A = wonder girls),第二个歌手称之为歌手 B(B = MC Mong)。歌手 A 和歌手 B 都是当前流行歌手,他们都有“歌手名 + 歌曲名”这种检索词。将 A 的歌曲命名为 a(a = so hot),B 的歌曲称之为 b(b = circus,中文名为“马戏团”)。将“歌手”的排名变化和“歌手名 + 歌曲名”的排名变化进行比较。共比较了两组数据,这两组数据都关于“歌手名”和“歌手名 + 歌曲名”的排名变化情况。第 1 组的时间为 2008 年 7 月 9 号的中午 12 点到晚上 12 点,第 2 组数据的时间是从 2008 年 7 月 9 号到 7 月 17 号。先以 1 天的数据为例进行说明。对于歌手 A 而言,它的歌曲的上榜时间相对较短,只在下午 16 点到晚 20 点之间有过片刻的上榜,上榜总时间不过 1 小时。针对于歌手 A 的“A & a”的搜索,也就是“歌手名 + 歌曲名”的搜索却多次上榜,特别是在下午 14 点到 16 点之间,一直处在排行榜前 10 名的位置。关于歌手 B 的“歌手名 + 歌曲名”的搜索情况,两个歌手的情况非常相似,而且上榜时间也很接近。但是,如果只是比较两位歌手的歌曲,有很大差别。歌手 A 的歌曲直在榜上停留了不到 1 小时,但是歌手 B 的歌曲却是一直停留的榜单上。对于一周的数据,歌手 A 和歌手 B 的情况既有相同点也有不同点。相同点:关于“歌手名 + 歌曲名”的上榜情况,都只是出现在两个时段,一个是工作日的下午,一个是休息日的全天。不同点:关于“歌曲名”的上榜情况,歌曲名上榜并没有一般性的规律可循,只是歌手 B 的歌曲比歌手 A 的歌曲在排行榜上持续的时间更长,从时间的遍历性考虑,歌手 B 的歌曲 b 的上榜时间更广,从凌晨到午夜都有上榜,而歌手 A 的歌曲 a 相对上榜时间较为固定。

(二) “歌手名 + 歌曲名”(A & a)与“歌曲”(a)的相关性

在这里引入相关关系的概念。关系是指两个变量之间不确定的关系,这里指类目之间或检索词之间,除等同关系或等级关系以外的其他各种

关系。相关关系着重研究 X 与 Y 之间的相似程度,取 X 为“歌手名 + 歌曲名”,设定 Y 为“歌曲名”。那么 X 与 Y 的相关系数就表示“歌手名 + 歌曲名”和“歌曲名”的排名的相似程度,从而就可以得出歌曲对于歌手的影响。X 和 Y 的相关系数的表达式如式(1):

$$\rho(X, Y) = \frac{E(XY) - E(X)E(Y)}{D(X)D(Y)} \quad (1)$$

式中, X 表示“歌手名 + 歌曲名”的排名, Y 表示“歌曲名”的排名。E(X) 表示“歌手名 + 歌曲名”在一段时间内的平均排名, E(Y) 表示“歌曲名”在一段时间内的平均排名。D(X) 代表的是“歌手名 + 歌曲名”在一段时间内的排名的方差。D(Y) 则代表“歌曲名”在一段时间内的排名方差。P 就显示了“歌手名 + 歌曲名”和“歌曲名”的排名的相似程度。最终得到: $\rho(A \& a, a) = 0.0573$; $\rho(B \& b, b) = 0.2975$ 。显然, $\rho(B \& b, b)$ 要比 $\rho(A \& a, a)$ 大得多。因此看出一首歌曲越受欢迎,他的歌手的受关注程度就越多。也就是说歌曲的好坏直接影响到歌手的受欢迎程度。

(三) 搜索榜中的特殊用户

根据排行榜的时间特征还可以看出,“歌手名 + 歌曲名”这一检索词并不是一直出现在排行榜上的,它的出现时间有一定的间断性。对于 A 歌手而言,“歌手名 + 歌曲名”出现在下午 3 点到 8 点之间,其中 5 点到 7 点一直没有出现。而对于歌手 B 而言,他的“歌手名 + 歌曲名”这一类检索词也是出现在五点到八点之间这一范围内。尽管两个歌手的歌曲的受欢迎程度不同,但是“歌手名 + 歌曲名”的检索词出现的时间段却非常接近。其时间几乎是吻合的。

由此可以判断,使用“歌手名 + 歌曲名”进行搜索的用户为特定的人群,他们搜索时间相对固定,因此,称此类用户为特殊用户,称这一类用户为“粉丝”或者歌迷。他们的行为影响了排行榜的分布情况。

三、结语

从 50 年前音乐排行榜诞生至今,关于音乐排行榜的研究一直都没有停止。如何完善现有的排行榜使其能够反映音乐的受欢迎程度^[3],这一问题还需要继续研究下去。

参考文献:

- [1] David E. Giles. Increasing returns to information in the U. S. popular music industry[J]. Economics Letters, 2006, 92: 68-74.
- [2] David E. Giles. Survival of the Hippest: Life at the Top of the Hot 100 [J]. Applied Economics, 2007, 39:1877-1887.
- [3] Joe Futrelle, J. Stephen Downie. Interdisciplinary Communities and Research Issues in Music Information Retrieval [Z]. ISMIR 2002-3rd International Conference on Music Information Retrieval, 2002.

(上接第 6 页)

“批判性思维”理念关注的核心问题是逻辑知识与逻辑思维能力之间的关系,或者更一般地是知识和能力之间的关系。人的日常逻辑思维能力,实际上是一种相对独立于各种专门知识,包括逻辑专门知识的逻辑思维能力,即“批判性思维”能力。这种能力,第一是存在的;第二是有差异的;第三是可训练的;第四是可测试的。人的素质差异不在于他们所掌握的知识信息量的差异,而在于他们思维能力的差异。

批判性思维有很多含义。一方面,意味着从不想当然,意味着不唯书、不唯上;意味着不只是因为领导的要求,就相信这么做对学校的发展是有益的。另一方面,批判性思维是一个对问题、结论、以及其它论据仔细思考分析与检验的过程;它意味着校验一个概念的定义、检验论据背后的假设和前提,并仔细审察论据推理的逻辑关系。

贯彻上级的指示,服从领导要凡事多问几个为什么——即上级为什么这样决定?这样做了将有什么样的结果?这种结果将对教学工作产生什么样的影响?这种影响是正向的还是负向的?必须强调的是,服从领导要批判思维,它不是一种简单的对立思维,而是善意的提醒与修正,是刻意的补充与完善。

(七)分析问题要放大思维

放大思维不是要去任意夸大,危言耸听,而是防微杜渐,见微知著。

《韩非子·说林上》:“圣人见微以知萌,见端以

知末,故见象箸而怖,知天下不足也。”汉·袁康《越绝书·越绝德序外传》:“故圣人见微知著,睹始知终。”见微知著的“微”是微小的意思,但这种“微小”是有条件为前提的;“微”是在此条件下的微小,即对这“微”所在的整体的影响小,要“知著”,就需一个条件,在这个条件下,原来的“微”已经由原来处于矛盾的次要方面转换为主要方面了,因此能“知著”。

善于以小见大,要有独到的眼光,学会细腻观察。察己可以知人,察今可以知古,审堂下之阴,而知日月之行,明阴阳之变;见瓶中之冰,而知天下之寒,察鱼鳖之藏。

在高等学校管理工作中,要做到以小见大,重在发展性、全局性地分析问题。面对一位作弊的考生,似乎就是传递、夹带纸条的小事,但它是一所学校考风、学风的体现,将影响到一名学生、一所学校优良学风的养成,关乎一代社会主义事业接班人优良品质的培养。试想象一下:一所学风不良的学校将培养出多少学风不良的毕业生?他们走向社会又将成为什么样的人?为人则诤骗、为官则贪腐、为商则假冒,祸害无穷!今日之学风,即明日之党风、明日之政风、明日之社会风气!

总而言之,践行科学发展观,成功地管理一所大学,需要一个精诚团结的领导班子,一套精心设计的管理制度,一支精明强干的管理队伍,一种精细严谨的工作作风,缺一不可。

参考文献:

- [1] 覃伯军. 中共中央宣传部理论局科学发展观学习读本 [EB/OL]. <http://zgy.swu.edu.cn/teacher/article.php?tid=1800>. [2008-06-10]
- [2] 柏昌利. 高等教育管理导论[M]. 西安:西安电子科技大学出版社,2006.
- [3] 王教文. 教务工作文辑[M]. 北京:华文出版社,2003.