

基于关联规则的情感语音合成中 韵律参数基频研究

王敬华 刘建银 张国燕 赵新想
华中师范大学

摘要 情感语音合成是当前语音合成研究的热点，其中韵律参数描述的正确与否直接影响到情感语音合成的输出效果。为了攻克提高情感语音自然度这一难点，通过研究关联规则、改进数据挖掘Apriori算法并由此来获得韵律参数中基频变化规则，为情感语音合成的选音提供了指导和帮助。

关键词 情感语音合成 韵律 基频 关联规则 Apriori算法

1 引言

自20世纪60年代Klatt语音合成器诞生以来，语音合成技术进入了一个数字计算时代。伴随着计算机技术的飞速发展，人们对计算机提出了更人性化的要求，情感语音合成应运而生。Burkhardt依据韵律规则将中立语音用基于KLSYN88共振峰合成器产生的emoSny工具调整转换到情感语音。英国Dundee大学提出了基于规则的语音串联的情感语音合成技术。最近，人们又应用基音同步叠加（PSOLA）算法来对韵律参数进行研究，但合成的语音仍然具有很强的机器味，并没有形成一个被广泛认可的、有效的研究方案。本文通过研究韵律模型提出一种以关联规则为基础，改进Apriori算法探索情感语音合成中前后音节的韵律参数之间存在的关系并产生相应的基频变化规则，为今后情感语音合成的选音提供指导。

2 情感语音韵律模型

计算机语音合成技术经过几十年的发展已经在很多方面得到了广泛的应用，其中文本—语音转换系统（Text To Speech, TTS）的研究（如图1所示）已经达到了商用的水平。目前，语音合成技术合成的主要是没有感情色彩的朗读式语音，由于缺乏语调、语速、重音等丰富的韵律特征变化及情感的表达，因此听起来单调乏味。情感语

音的合成顺应了时代的发展，满足了人们的需求，它属于语音的规则合成，所以要合成能够表情达意的语音，韵律模型，特别是其中的基频模型急切需要得到改善。

2.1 韵律模型

韵律是指语音的超音段特征，在感知上的知觉韵律表现为语音的速度（speed）、音量（loudness）和音高（pitch）随时间的变化。韵律是语音的节奏（rhythm）、语调（intonation）和重音（stress, accent）所抽象出来的特性，在声学参数上韵律表现为基频（fundamental frequency）、音段时长（duration）和能量（energy）随时间的变化。研究各种方法建立韵律模型，以产生这些韵律参数，主要包括基频模型、时长模型等。

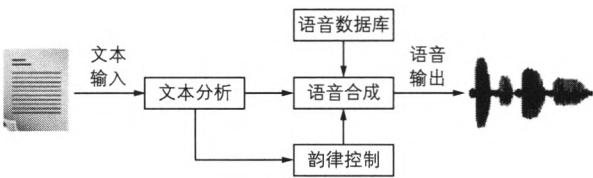


图1 文语转换系统

2.2 韵律特征参数选取

韵律参数主要是指音节的时长、能量、基频的变化以及适当的停顿等众多参数属性。对合成情感语音的自然度影响最显著的是基频变化和时长变化。本文主要选择基频作为研究对象,图2为一音节基频变化的基频曲线变化。

其中,横坐标 $n(ms)$ 表示音节中以时序为基础的不同位置的基频对应值,纵坐标表示音节中不同的基频频率 $f(Hz)$,曲线表示一个音节的基频曲线。 H 、 S 、 E 、 L 分别表示基频曲线的最大值、起始值、终止值、最小值, N_1 和 N_2 分别为最大值位置和最小值位置。

在选取韵律特征参数时,有的选择基频中值作为研究对象,并且取得了一定的成效。取基频中值固然方便,但情感语音合成的韵律系统不够准确,合成效果在有些情况下很难满足需要。情感语音合成技术要向前发展,就必须考虑到中值对高音特征点和低音特征点各自的作用模糊表现形式。当两个不同的音节基频中值相同时,低音部分和高音部分可能是不同的,这样就不能对音节更好地细分,在规则提取时就不能将有用的规则完全提取出来,因此本文选择基频曲线中的最高音频值(N_1)和最低音频值(N_2)作为研究对象,通过相同音节在大量语句中的音联现象进行数据挖掘,并以关联规则的形式给出前后音节之间的相互关系。

3 关联规则模型的构建

数据挖掘是从大型数据库的大量原始数据中提取人们感兴趣的、隐含的、具有潜在应用价值的信息和知识,被认为是目前解决数据丰富而信息贫乏的一种有效方法。寻找这种知识和信息的过程即是挖掘关联规则的过程。关联规则即为描述数据库中一组数据项之间的某种潜在关系的规则。

从前文所选韵律参数来看,每个音节选取其中的最大、最小值并对最值进行研究,设选取的韵律参数为 $P=$

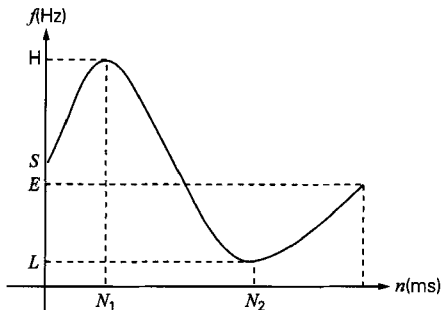


图2 基频曲线变化

(A_1, A_2) , A_1 和 A_2 分别表示基频和时长。 $A_1 = (S, N_1, N_2, E)$, 根据该表达式可知每个音节的起止位置及最大值情况,对某音节的下个音节进行选取时选取的规则基本形式为 $A \rightarrow B$ (A 、 B 为两前后相邻的音节)。由前面所选取的韵律参数来为每个文本信息单元选取韵律特征最优基频最值,通过设置一定大小的域值得到初始规则 $N_i \rightarrow NK$ 。由于对每个音节选取的韵律参数为两个最值,此时应再进一步考虑规则前后件中的相邻关系来筛选出扩展规则: $A[i] + N_1 \rightarrow N_k + B[k]$, 最终确定前后两音节,得到最后规则 $Y_i, k \rightarrow Y_i, k - S_{max}$, 其中 S_{max} 为最大支持度。该规则提取过程中需要在大量的数据中进行提取。

挖掘关联规则的问题描述如下。

定义1: 设 $I = \{i_1, i_2, \dots, i_m\}$ 是所有项目的集合, D 是所有事务数据库的集合,即 $D = \{T_1, T_2, \dots, T_n\}$, 每个事务 T 是一些项目的集合(即 $T \subseteq I$), 每个事务可以用唯一的标识符TID来表示。设 X 为某些项目的集合, 如果 X 包含在 T 中, 则称事务 T 包含 X 。

定义2: 关联规则是形如 $X \Rightarrow Y$ 的蕴涵式, 其中 $X \subseteq T, Y \subseteq T$, 且 $X \cap Y = \emptyset$, 若 D 中事务包含 $X \cup Y$ 的百分比为 S , 则关联规则 $X \Rightarrow Y$ 的支持度为 S , 它是概率 $P(X \cup Y)$, 如果 D 中包含 X 的事务同时也包含 Y 的百分比为 C , 则关联规则 $X \Rightarrow Y$ 的置信度为 C , 它是概率 $P(Y/X)$ 。通常将关联规则表示为 $X \Rightarrow Y (S\%, C\%)$ 。

挖掘关联规则一般分为两步: 首先, 找出所有的频繁项集, 即找出事务数据库中所有的数据项集必须满足支持度大于等于用户指定的最小支持度的, $Sup(X \cup Y) \geq \min_sup$; 其次, 根据频繁项集产生强关联规则, 这些规则必须满足最小置信度, $Conf(X \Rightarrow Y) \geq \min_conf$ 。

在以上两步骤中, 挖掘关联规则的关键步骤为第一步, 该步决定了关联规则挖掘的总体性能, 因此现有的研究都集中在第一步上, 也就是对挖掘频繁项集的处理。

4 产生候选项集的集合 C_k 的新算法

Apriori算法是一种处理无序数据集的算法, 它有一些无法克服的不足: 要对数据库频繁地扫描, 这对经常遇到的平均事务宽度很长的数据库或海量数据库来说, I/O 开销是非常大的, 会产生大量的候选2-项集, 并引发2-项集瓶颈问题。例如, 频繁1-项集的数目为 n , 则产生 C_n^2 个候选2-项集, 其中有很多是无效的2-项集, 这不仅占用了大量的空间, 而且增加了该步的工作量。由 k -项集连接生成的候选 $(k+1)$ -项集时, 该操作要判断是否前 $k-1$ 项相同而第 k 项不同, 且搜索空间较大, 效率较低。另外, 它

不具备时间序列，但语音数据是具有一定的时间顺序的，语句中语音音节之间先后顺序对数据挖掘的结果有直接的影响，针对音节间的关系及基于前面韵律参数选择本文对Apriori算法进行了修正，并采用了序列方式记录数据形式，得到适合情感语音数据的算法思想如下。

(1) 在 L_{k-1} 中的一个项集 I 与 L_{k-1} 中 I 之后所有的项集进行连接，由于各事务已经设定好按字典排序，对于其中的任何一个 k -项集 L ，有 $L[1] < L[2] < \dots < L[k]$ ，所以对于任意两个 $(k-1)$ -项频繁集 L_i 和 L_j ($i < j$)，如果 L_i 和 L_j 不符合连接条件，那么 L_i 和 L_j 之后的所有 $(k-1)$ -项集都不能满足连接条件。因此只要 L_i 和 L_j 不能连接，就无需判断 L_i 和 L_j 之后的所有 $(k-1)$ -项集是否可以连接，这样就减少了循环次数。

(2) 把连接得到的不同 k -项集按顺序存入TQ，存储方式：首先在存储前与已存入的TQ的 $k-1$ 项集进行对比，假如与某项集相同，那么相应 $k-1$ 项集计数值自动加1，若没有相同的 $k-1$ 项集，那就直接存进去，并且将计数值设为1。

(3) 确定包含项集 I 的所有符合剪枝后的候选 k -项集，即把TQ中所有计数值等于 $k-1$ 的 k -项集加入 C_k （若该 k -项集计数值不等于 $k-1$ ，则它一定不是频繁项集）。

(4) 循环执行(2)、(3)步，直到 L_{k-1} 中的每个项集都互相连接完为止。最终要生成的候选项集的集合就是 C_k 。这就省去了在 L_{k-1} 中寻找 $k-1$ 项集的所有 $(k-1)$ -子集的耗时剪枝操作，只需要对包含项集 I 产生的数据项集TQ进行扫描，这样剪枝部分扫描的集合减小，也就减小了总的运算量，因而提高了算法效率。

输入：事务数据库D，最小支持度计数阈值min_sup。

输出：D中的频繁项集L。

$L_1 = \text{find_frequent_1-itemsets}(D)$;

for($k=2$; $L_{k-1} \neq \Phi$; $k++$) //产生两字词、多字词的频繁项目集

$\{C_k = \text{apriori_gen}(L_{k-1}, \text{min_sup})$; //生成候选 k -项集

for each Transaction $t \in D$

$\{C_t = \text{subset}(C_k, t)$; //候选集 C_k 中提取包含事务 t 的候选 k -项集

for each Candidate $C \in C_t \{C.\text{Count}++\}$ }

$L_k = \{C \in C_t | C.\text{Count} \geq \text{min_sup}\}$; } //至此选出所有符合基本条件的规则

for($k=1$; $k \leq C.\text{Count}$; $k++$) //针对音节中的重音

进行判断筛选规则

$\{ \text{if}(\text{in}(N_i, N_k) N_i.\text{front} < \Phi) \text{put the rule in } L_i$;
else delete it; }

return $L = \bigcup_{i=1}^t (L_i)$; //求 L_i 的和，产生最终需要的规则

Procedure apriori_gen($L_{k-1}, \text{min_sup}$)

{for each itemset $L_i \in L_{k-1} \{TQ = \Phi$; //存储临时候选项集的集合

for each itemset $L_j \in L_{k-1}$ //其中($i < j$)

if ($(L_i[1] = L_j[1]) \wedge (L_i[2] = L_j[2]) \wedge \dots \wedge (L_i[k-2] = L_j[k-2]) \wedge$

$(L_i[k-1] < L_j[k-1])$) //其中，($L_i[k-1] < L_j[k-1]$)可以确保不产生重复的 k -项集

$\{C = L_i \oplus L_j$; // 连接步，产生候选项集

if($\text{length}(C) > k$) continue;

else if($\text{exists}(TQ, C)$ and $Q \in TQ$ and $Q == C$) //如果C在TQ中已经存在

$Q.\text{count}++$; /*计数初值加1*/ else // C在TQ中首次出现

$\{ \text{add } C \text{ to } TQ$; $Q = \text{find}(TQ, C)$; $Q.\text{count} = 1$;
//计数初值设置为1

}for each itemset $C \in TQ$ if ($C.\text{count} == k-1$) //TQ中出现 $k-1$ 次的项集加入到 C_k 中

add C to C_k ; } else break; } return C_k ; }

在关联规则的整个过程中，其核心问题是产生满足具有用户制定的最小支持度的所有频繁项目集。本算法之所以采用序列方式记录数据形式，主要考虑的是相邻音节间的相互关联关系。由于在选取韵律参数时是选取的基频曲线中的高音和低音的最值，默认产生规则中的前件为前一个音节中的第二个参数值（或是低音最值，或是高音最值），而后件默认为相邻后一个音节中的第一个参数值。假设A、B分别代表前后两个相邻的音节，考虑到音节之间的关联关系，可将其长度设为2，对两个音节得到的规则是 $A[2] \rightarrow B[1]$ ，最终采用哪一条规则要看音节A、B前后参数的基频值具体是多少，并且形如 $(A[1]+A[2]) \rightarrow B[1] + B[2]$ 的规则还有待进一步开发。

5 实验研究分析

实验研究：实验环境在Microsoft Visual Studio 2010、Microsoft SQL Server 2008环境实现；处理器为Intel(R) Core(TM) i5-2410M CPU @2.30Hz 2.30Hz，内存为8.00GB，64-bit Win7 Operating System。

实验数据由IBM数据生成器得到。此数据集大约有1000条事务，每个事务中平均有5个项集，在不同的最小支持度阈值下测得实验结果如图3所示。从图3可以看出，本文改进后的算法比原Apriori算法在时间性能上有明显优势，尤其在支持度比较小时效果更为明显。

本文所有实验语料均来自世界最专业的美国语音软件供应商Nuance公司旗下子公司环球语音技术提供商Loquendo语音技术研究实验室，其语音库提供了32种语音和76个声音，包括男性和女性，还可以发出外来词的读音。从中提取4800条情感语句，均来自Dave的语音库，采样率为32kHz，编码长度为16bit，单声道，以wave文件形式存储采样数据，一个wave文件作为一个独立单元、一个词或是一个句子。

库中所有音节利用语音分析软件praat进行音节边界和基频标注，在基频标注过程中均标注每个音节的最高和最低音频值。基频值一般在50~500Hz内，单位音节时长一般为几百毫秒。最大、最小值的位置用相对位置来表示，数值分布在0~1。实验对照为一青年男发音人在专业录音室录制完成，保存为32kHz，16bit量化的单声道波形文件，并对录音句子进行了基频曲线的提取和音节边界的划分。由于实验提取的基频曲线来自于真实的情感语音，因此我们可以认为此韵律参数是理想的，并且基频曲线和音节长度较好地反映了韵律参数中基频和语速的变化。图4展示了原始的人工朗读情感语音的波形和基频曲线（图4(a)）以及采用改进的Apriori算法合成的情感语音的波形和基频曲线（图4(b)）。可以看出，合成的基频曲线和音节长度基本和原始人工朗读的情感语音的基频曲线和音节长度基本是一致的，也就是说基于关联规则改进算法的韵律参数基频的研究方法是较为理想的。

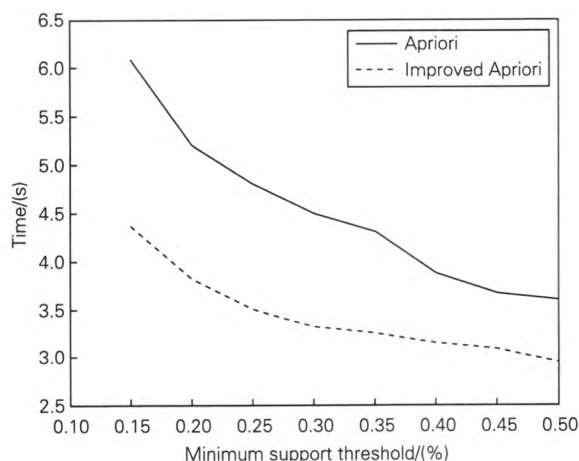


图3 在不同支持度阈值上两种算法的性能比较

6 结束语

研究表明，运用改进后的Apriori算法在挖掘情感语音频繁项目集的效率上有明显提高，为今后情感语音合成的选音问题提供了指导和帮助。虽然本文对情感语音韵律建模、利用提升算法性能来加快韵律参数选取效率，但由于时间和精力限制，本文对情感语音合成韵律的研究工作尚有很多不够细致深入的地方，仍然有很多问题有待进一步的深入研究和完善。情感语音合成中规则的制定肯定不是单一特征就能够达到完美的效果，如何考虑综合利用恰到好处语音特征，并制定出合适的情感语音合成规则将会是语音合成科学家们一个长期的、富有挑战性的课题。

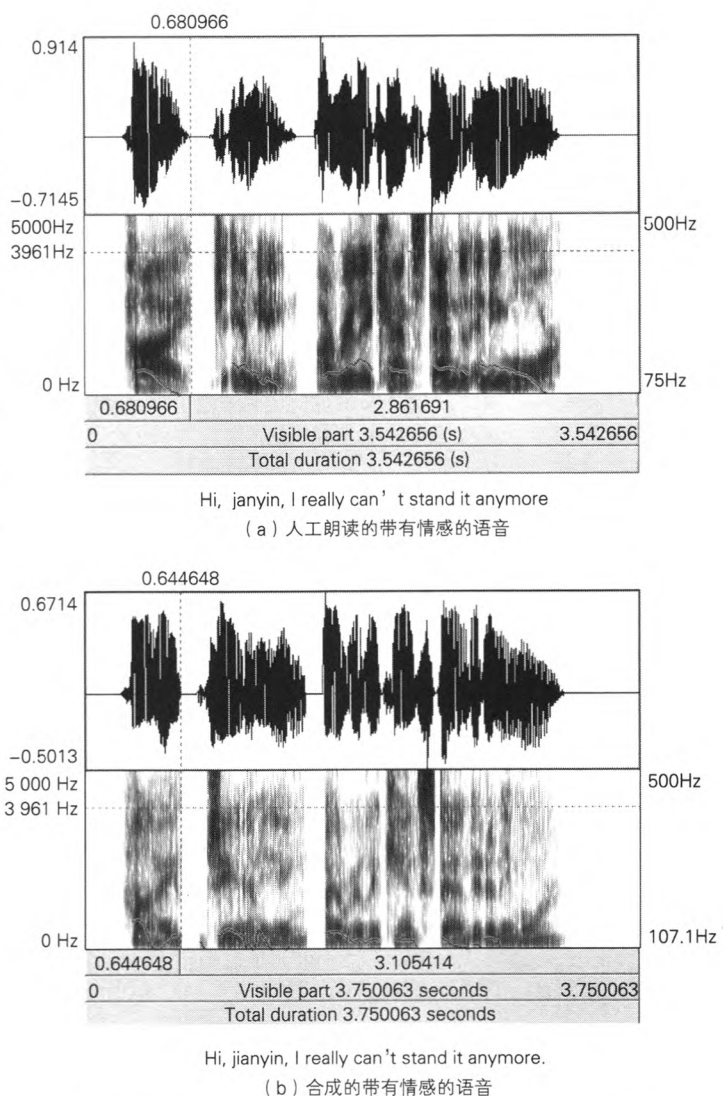


图4 句子"Hi, janyin, I really can't stand it anymore" 带有气愤的感情色彩时的波形和基频曲线

参考文献

- [1] Burkhart F. Verification of acoustical correlates of emotional speech using formant synthesis. Proceedings of the ISCA Workshop on Speech and Emotion. Northern Ireland, 2000
- [2] Murray I R. Emotion in concatenated speech. IEEE Seminar State of the Arts in Speech Synthesis Proceedings. London, 2000
- [3] 邵艳秋, 韩纪庆等. 韵律参数和频谱包络修改相结合的情感语音合成技术研究. 信号处理, 2007(4)
- [4] Gu W, Hirose K, Fujisaki H. A method for automatic tone command parameter extraction for the model of F0 contour generation for mandarin. IEEE Workshop on Automatic Speech Recognition and Understanding. Nara, Japan, 2004
- [5] 蔡连红, 黄德智, 蔡锐. 现代语音技术基础与应用. 清华大学出版社2003
- [6] 于剑, 黄力行, 陶建华. 汉语对话语气韵律建模方法. 清华大学学报(自然科学版), 2008, 48(S1)
- [7] 陈志泊. 数据仓库与数据挖掘. 清华大学出版社, 2009
- [8] Han J, Kamber M. Data Mining: Concepts and Techniques. Beijing: Higher Education Press, 2001
- [9] 刘明亮, 李雄飞, 孙涛等. 数据挖掘技术标准综述. 计算机科学, 2008, 35(6)
- [10] 毛宇星, 陈彤兵, 施伯乐. 一种高效的多层和概化关联规则挖掘方法. 软件学报, 2011, 22(12)
- [11] Pramudiono I, Kitsuregawa M. FP-Tax: Tree structure based generalized association rule mining. In: Proc. of the ACM Workshop on Research Issues in Data Mining and Knowledge Discovery(DMKD). 2004
- [12] Sriphaew K, Theeramunkong T. Fast algorithm for mining generalized frequent patterns of generalized association rules. IEICE Trans. on Information and Systems(TOIS), 2004, E87-D(3)
- [13] 陆丽娜, 陈亚萍. 挖掘关联规则中Apriori算法的研究. 小型微型计算机系统, 2000, 21(9)
- [14] Yu Wanjun, Wang Xiaochun, Wang Fangyi, et al. The research of improved Apriori algorithm for mining association rules. Proceedings of the 11th IEEE International Conference on Communication Technology(ICCT2008), 2008
- [15] 徐章艳, 张师超, 区玉明等. 挖掘关联规则中的一种优化Apriori算法. 计算机工程, 2003, 29(19)

作者简介

王敬华 男, 硕士, 副教授, 研究方向为数据库与数据挖掘、现代信息系统。
刘建银 男, 硕士, 研究方向为数据库与数据挖掘、情感语音识别和合成技术。
张国燕 女, 硕士, 研究方向为数据库与数据挖掘。
赵新想 女, 硕士, 研究方向为数据库与数据挖掘。

Prosodic Parameters Fundamental Frequency Research of Emotional Speech Synthesis based on the Association Rules

Jinghua Wang, Jianyin Liu, Guoyan Zhang, Xinxiang Zhao

Central China Normal University

Abstract: Emotional Speech Synthesis is the research focus of Speech Synthesis. Among them, the prosodic parameters described is correct or not that will have a direct impact on the output effects of the emotional speech synthesis. In order to improve the difficulty of natural degree which belongs to emotional speech, this paper by studying the association rules, and improved data mining Apriori algorithm, and thus to get the prosodic parameters of fundamental frequency changing rules to provide guidance for selecting sound in emotional speech synthesis.

Keywords: Emotional speech synthesis, Prosodic parameters, Fundamental frequency, Association rules, Apriori algorithm