

# 关键词识别的基本技术和研究进展\*)

Techniques and Advances in Keyword Spotting

陈凯江 欧嘉致 黄董菁 吴立德

(复旦大学计算机系 上海200433)

**Abstract** Keyword Spotting (KWS) or Keyword Recognition is to recognize a set of given keywords in continuous, spontaneous speech. It is a very active branch of Speech Recognition. It differs from Continuous Speech Recognition (CSR) in terms of research focus, implementation techniques and application background. So they are often regarded as two different issues. This paper gives an overview of basic techniques and advances in recent years, including how to combine the traits of Mandarin.

**Keywords** Keyword spotting, Keyword recognition, Wordspotting, Continuous speech recognition, Language modeling

## 1. 引言

关键词识别(Keyword Spotting, 简称 KWS)的任务是从连续无限制的语音中, 识别给定的若干个词。它与通常所说的连续语音识别(Continuous Speech Recognition, 简称 CSR)有密切的联系, 但又有显著的区别。所以, 在语音识别的研究中, KWS 与 CSR 通常是作为两个不同的问题来处理<sup>[1,2]</sup>。KWS 与 CSR 的主要区别在于对输入语音的限制。CSR 的限制较强, 要求安静的环境, 较好的信道, 以取得较高质量的语音, 并且可以要求用户自己添加新词。但在噪音环境中, 性能显著下降<sup>[3]</sup>。而 KWS 允许在嘈杂的环境中使用, 可以通过诸如电话线等质量较差的信道, 要求系统能自动判断哪些是词汇表中没有的词 OOV(Out Of Vocabulary, 简称 OOV)。

随着多媒体技术的发展, 语音识别已经从单纯的人机输入界面扩展到包括信息处理、检索在内的广大领域, 从端坐于计算机前朗读规范文本, 扩大为在更广泛的场合, 输入更加自然的语音; 处理设备也不局限于一台 PC 为一个用户服务的模式, 可能是小巧的嵌入式设备, 也可能是通过电话向服务器传送语音。而相对于 KWS, CSR 资源耗费大, 速度慢, 抗噪能力不强, 这是 CSR 短时间内难以取得突破的问题<sup>[3]</sup>。因此, 许多应用领域不适合使用 CSR, 而要求用 KWS, 例如:

·电话应答/服务系统 只要求识别若干预定词语, 但可能有众多用户同时使用, 用户所在环境可能比较嘈杂。这种系统既没有能力, 也没有必要使用 CSR。由于不可能像 CSR 那样要求用户添加新词, 系统必须自动判断、处理 OOV。文[4]对电话总机收录的真实语音资料做了统计, 发现英文自然发音中, 夹杂有很多类似“after I”的插入语, 约每5秒钟有一句“um”之类的咕哝声, 这些声音的音素结构等特征与正常词语不同。

·视频/音频文档检索 由用户输入一个关键词, 系统则定位出这个关键词在哪些地方出现。视频/音频文档所包含的声音范围很广, 不仅有语音, 还有音乐等, 从而严重影响识别率, 难以直接应用 CSR。检索的另一个特点就是用户完成输入后, 系统必须在很短时间内做出反应。视频文档往往少则几

十小时, 多则几百小时。CSR 一般只有实时速度, 直接应用是不可能的。而 IBM 公司用 KWS, 达到了2400倍实时的检索速度<sup>[5]</sup>。

·电话录音、语音邮件分类 也只要求识别若干预定词语, 但用户可能没有意识到自己在与计算机交流, 发音较随意, 可能夹杂有很多的语气词; 而使用 CSR 的用户往往比较配合。并且, 在电话中安装庞大的 CSR 系统也是不现实的。

但 KWS 与 CSR 也有密切的联系。KWS 与 CSR 使用很多共同的技术, 有些基于音素的 KWS 技术的特征提取、识别单元的选取、训练等部分和 CSR 是类似的, 主要差别在于搜索、后处理阶段。

## 2. KWS 的基本模块

建立一个 KWS 系统, 需要几个基本模块: 特征提取, 建立声学模型(Acoustic Model)并进行训练, 搜索和识别, 对识别结果进行后处理, 结合语言模型来提高性能。如图1。

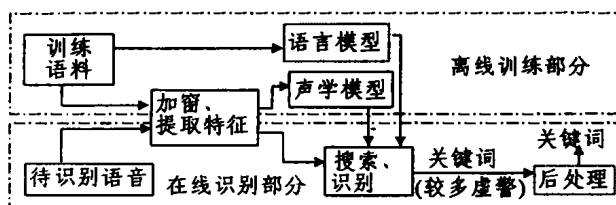


图1 KWS 的大致框架

图中, 语言模型、后处理部分是可选的。对于特征提取模块, 传统的倒谱特征的技术已经成熟, 只作简略介绍, 另外在 3.1 节介绍了日益被重视的韵律(Prosodic)特征; 其它几个模块, 尤其是搜索和识别、后处理模块, 与 CSR 有较大差异, 很多 KWS 的文献在这方面做了改进, 本文将用较多篇幅介绍。语言模型本身可作为一个研究领域, 但它在 CSR 和 KWS 中都有广泛的应用<sup>[6-8]</sup>, 并且很多 KWS 的研究者用它来提高识别率<sup>[9-11]</sup>; 近年来, 还有些研究者根据各种场合对 KWS 的不同要求, 致力于改进或应用新的语言模型, 以提高识别率, 例如文[8, 12]。因此, 本文对语言模型及其在 KWS 中的应用、

\* ) 本项目受国家自然科学基金资助(编号69935010和699.3011)。陈凯江博士, 主要研究方向为大规模文本处理、自然语言理解、多媒体信息检索。

硕士研究生, 研究方向为自然语言理解、多媒体信息检索。吴立德



进行训练,然后将它们组成一个类似图2的全局 HMM,如图3。

最基本的识别方法如下:对输入语音提取特征,将特征送入图3的全局 HMM,用 Viterbi 算法<sup>[14]</sup>找出最佳状态路径  $S = s_1s_2 \dots s_p$ ,如果  $S$  中含有子序列  $S' = s_{i+1} \dots s_{i+q}$ ,使得  $S'$  中的每个状态  $s_p$  都是关键词模型  $k$  中的状态,则认为  $S'$  对应的语音序列是关键词  $k$ 。

这种结构的系统构造简单,训练量相对较小,较早的研究者,如文[13,23],多应用这种结构。其性能比基于 CSR 的差一些,但有运行速度快、资源耗费小等特点,特别适合于电话应答系统,所以近年来仍有不少研究者使用这种结构,如文[40,41]。

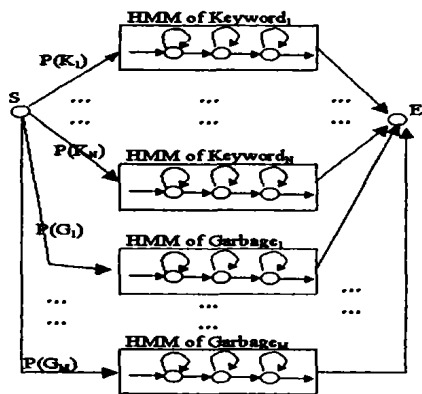


图3 基于关键词/垃圾模型的 KWS

关键词表改变时,大多数系统需要重新训练<sup>[5]</sup>,但也有些将关键词定义为子词或音素,使系统不经重新训练就可以加入新词<sup>[40]</sup>。

2) 基于 CSR 前  $N$  个识别结果<sup>[42]</sup> 这种技术先使用一个大词汇量的 CSR 识别输入语音,产生前  $N$  个最好结果列表及其置信度,然后再进行搜索。如果某个关键词存在于列表中,且置信度大于某个阈值,则认为找到该关键词。这种方法的优点是可以先进行预处理,然后在中间结果上进行快速搜索。但是 CSR 词表中没有的词将不会被查出来。

3) 基于音素网络<sup>[20]</sup> 有些 KWS 直接将输入语音看作音素的连接,并将待查询的关键词看作音素序列,在输入语音形成的音素网格(图4)中搜索。这类 KWS 先用一个音素识别器,对同一段语音输出多组结果及对应的置信度,形成一个音素网格(Phone Lattice),如图4。然后,用改进的 Viterbi 算法在网络上搜索。

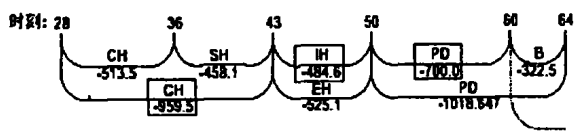


图4 音素网格

图中弧上所标的字母即是音素,下边的数字即是置信度(概率的对数)。图中3个方框指示的弧包含了关键词“CHIP”。

这种方法不受制于词典,也可以方便地加入新词,但是,其精确度要依赖于音素识别器,而音素识别器的正确率通常不够高<sup>[5]</sup>。

### 2.5 后处理

为了进一步提高识别率,很多系统在以上搜索之后,还对假想命中进行了后处理(或称语音验证, Utterance Verifica-

tion),常用的方法有:

1) 神经网络(Neural Network, 简称 NN)分类 如文[36],将假想命中的序列平均分为3部分,每部分用多个特征矢量的平均矢量来表示,并作为 NN 的输入。实验结果表明可以降低约16.4%的虚警。

2) 先分开关键词/非关键词 例如文[44],在第一阶段使用自组织的特征映射先分开关键词/非关键词,在第二阶段用模糊动态聚类再判断是哪个关键词。

3) 其它 有的研究者使用分段语音模型来进行后处理<sup>[43]</sup>。也有的利用前  $N$  个候选词的似然度之间的差距进行后处理<sup>[41,45]</sup>;或采用新的似然度对数比值(Log-Likelihood Ratio)<sup>[42]</sup>;对关键词、非关键词模型的对数似然度作仿射变换等<sup>[46]</sup>。

### 2.6 语言模型

语言模型在 CSR 和信息检索中有广泛的应用<sup>[6,7]</sup>,在基于音素的高性能 KWS 中,语言模型往往是重要的一部分<sup>[9-11]</sup>。较早的工作有文[9],当时仅仅使用了音素组成的 bi-gram,就可以使 FOM 由68%提高到76.4%。近年来,还有些 KWS 研究者将研究重点放在改进或应用新的语言模型<sup>[8,12]</sup>来提高识别率。

最基本的语言模型是用长度为  $k-1$  的历史  $(w_{i-1}, w_{i-2}, \dots, w_{i-k+1})$  来估计当前词的出现概率:

$$P(w_i | w_{i-1}, w_{i-2}, \dots, w_{i-k+1}) \quad (1)$$

其中,  $w$  可能是词,也可能是音素等识别单元; $k=2$ 时,称为 bi-gram 模型, $k=3$ 时,称为 tri-gram 模型。

理论上,只要有足够多的训练语料,(1)式可以用

$$\frac{c(w_i, w_{i-1}, w_{i-2}, \dots, w_{i-k+1})}{c(w_{i-1}, w_{i-2}, \dots, w_{i-k+1})}$$

( $c(\cdot)$ 表示出现频率)来估算。但实际上,由于可能的元组  $(w_{i-1}, w_{i-2}, \dots, w_{i-k+1})$  的数目非常巨大,大部分元组根本不在语料中出现,所以数据稀疏问题非常突出,要用概率平滑等方法来近似估计(1)式。文[47]给出了多种概率平滑、回推的方法。绝大多数的系统只能用到 tri-gram 模型<sup>[20]</sup>。

由于(1)式中, $k$  的值很小,所以(1)式也称为“短距离语言模型”。为利用更多的“历史”,研究者们提出了不少改进方案,如文[48]提出了可变长的  $N$ -gram 语言模型;文[49]提出了基于触发器的“长距离”语言模型;文[50]结合“长距离”、“短距离”语言模型各自的特点,提出了“多尺度”(Multispan)语言模型。文[51]则提出了一种新颖的思路:利用信息检索技术,根据较长的上下文,确定当前的主题,从而动态调整  $N$ -Gram 模型的概率。文[52]提出了相似的思路,但用最大熵方法实现。文[29]详细介绍了语音识别中常用的语言模型、统计方法。语言模型可以与全局 HMM 结合,图2就是 uni-gram 与全局 HMM 结合后的网络。bi-gram、tri-gram 则复杂得多,文[29]给出了较详细的介绍。tri-gram/HMM 网络搜索时需要专门的搜索策略,主要方法有2遍搜索<sup>[29]</sup>、多遍搜索<sup>[53]</sup>、动态规划<sup>[54]</sup>等。

## 3. 最新的进展

### 3.1 使用韵律特征

近年来,韵律特征(Prosodic Feature)正越来越被重视,特别是在中文语音识别中,韵律特征更为重要<sup>[19-21]</sup>。韵律特征包括:音长、能量、基音、音调。特别是音长、音调,受噪音、信道的影

一个是音调识别器,一个是音素/音节识别器。文[20]用归一化基音周期、归一化基音周期的差分、归一化基音能量对数、第一共振峰的能量对数组成韵律特征矢量,用中文语料测试,发现在相同的准确率下虚警率约下降3~8个百分点。再如文[22],使用韵律特征后,错误率约降低4.5个百分点。

### 3.2 结合中文特点

近年来,随着中文影响力的扩大,越来越多的学者研究如何针对中文特点,提高中文 CSR 和 KWS 的性能。相对于西方语言,中文有以下几个特点<sup>[20,55]</sup>:

1) 中文是由音节组成的,且音节要比英文简单得多——每个音节只由一个韵母组成,或由一个声母后面跟一个韵母组成,并且音节之间没有连读。

2) 中文有5个音调,同样的音节,不同的音调,代表不同的字和不同的意思。

所以,中文语音识别中,韵律信息更为重要,并且往往要求对音调特别处理<sup>[19,20]</sup>。文[19]使用了韵律信息后,虚警率下降约3~8个百分点。利用中文单音节、音节数目不多、音节结构简单的特点,一个比较直接的思路是将英文的音素网络<sup>[26]</sup>扩展为中文的音节网络。文[56]结合韵律信息,采用了多阶段的策略:

1) 利用韵律信息,如能量变化轮廓、过零率、停顿的长度等,先估计音节的大致边界。

2) 使用上下文无关的子音节模型,将语音序列分解为音节网格。

3) 使用该文提出的模糊搜索,找出可能的关键词,每个都对应若干个候选。

4) 用一个由上下文有关子音节构成的关键词模型,来给每个候选打分,选取最高分者作为输出。

该方法的特点是速度较快,能适应大词汇表的关键词识别。文[56]使用了2611个关键词,前10个候选至少有一个包含正确答案的概率是85.79%,在 Sparc20 工作站上只需要1.2倍实时的时间。

文[57]则精细地分析了中文声母、韵母的构成,为声母、韵母、无声母音节分别建立模型,使用动态规划在音节网格上进行搜索。

中文是一种有音调的语种,所以音调信息也较能反映中文的特点。文[58]先用一个不考虑音调特征的音节识别器给出音节的边界以及多个候选音节,然后再用一个音调识别器来给各个音调打分。文[19]比较了三种方法,第一种是不考虑音调特征,第二种是考虑韵母的不同音调,第三种还加入了基音特征。识别错误率分别为:7.32%,6.43%,6.03%。文[59]综合了多种方法的优点,比无音调的识别器降低了30%的错误。

另外,中文是以字为单位,词与词之间没有空格,先要经过自动分词处理。文[60]提出一套包括自动分词在内的系统的解决方案。自动分词也是一个研究课题,这里不作叙述,有兴趣的读者可参考文[47]等。

### 3.3 改进语言模型

大多数的语言模型依赖于语料的主题,有的系统为常见的主题各建一个语言模型,以提高精度。近年来的研究者注意到,在语音识别中,不仅要考虑该语音对应的主题,还要考虑到说话的风格(Speaking Style)。

文[8]认为,为每个任务、主题都建立收集足够多的语料,各自建立一个语言模型,是不现实的。在无约束的自然对话

中,每一场对话的主题都可能不一样。文[61]的实验表明,扩展训练集所包含的主题并不总是有害的;而对于语音识别使用的语言模型来说,说话的风格非常重要,并将说话风格分为正式书面用语、文学用语等几种风格。文[8]将语言模型根据说话风格分为对话风格、书面风格两种,并收集多个主题的语料,用信息论标准,找出与主题相关性较低的词构成词汇表,用于训练垃圾模型中的语言模型。其系统用于关键短语检测与确认的实验,在虚警率为2%时,错误拒绝率降低约8个百分点。

### 3.4 拒识 OOV

由于 KWS 通常在比 CSR 嘈杂的环境中使用,且一般不能要求用户加入新词,因此基于音素、CSR 的 KWS 为了进一步提高性能,还必须解决 OOV 的判断和拒识问题。如文[62]对其原有系统做了4方面的改进,其中包括 OOV 的拒识,使错误率减少约50%。

CSR、音素识别器只将似然度最高的结果输出,可能将 OOV,甚至非语音识别为词表中的某个词,而 HMM 计算出的似然度会随着环境的不同而有所改变,所以不能直接使用固定的阈值来决定接受还是拒绝,必须先经过似然度归一化处理。文[63]介绍了两种常用的归一化方法,一种是后验归一化,一种是假设检验。设有一段语音对应的特征序列  $O$ , CSR 识别器的最佳候选是  $M_c$ ,第一种方法定义  $O$  对应到词  $M_c$  的得分为:

$$S(M_c|O) = \frac{P(O|M_c)P(M_c)}{\sum_{i \in V} P(O|M_i)P(M_i)}$$

其中,  $V$  为词表,第二种方法是检验以下两个假设的似然比:

$$H_0: i(O) = c; \quad H_1: i(O) \in V - \{c\}$$

$$\text{似然比 } LR = \frac{L(H_0)}{L(H_1)}, L(\cdot) \text{ 表示似然度}$$

$L(H_0)$  可用  $P(O|M_c)$  来估计,  $L(H_1)$  通常用次佳候选的最佳似然度来估计。

文[64]在说话人确认的研究中,提出了基于竞争的归一化方法。文[63]将其用于 OOV 的拒识,可以降低30%的拒识错误。文[65]的实验表明,阈值与候选词的长度有关,音长较长的候选应该用较高的阈值。也有些系统将 OOV 作为词汇表中一个特殊的词,该词可以是音素的任意组合<sup>[66]</sup>。

**结束语** 本文对 KWS 的基本模块、相关技术和近年来的新进展做了较全面的综述,并介绍了国际上的评测方法。随着多媒体技术的进一步发展,语音识别应用领域的进一步扩大、普及,以及各种电子产品的进一步智能化, KWS 正受到日益的重视,其技术也在不断地进步、更新, KWS 将会在更多、更广泛的场合为人们提供便利。

### 参考文献

- Higgins Alan L, Wohlford Robert E. Keyword Recognition Using Template Concatenation. ICASSP-85, vol. 3 pp. 1233~1236
- Morgan David P, et al. A Keyword Spotter Which Incorporates Neural Networks for Secondary Processing. ICASSP-90, vol. 1, pp. 113~116
- O'Shaughnessy D, Gabrea M. Recognition of digit strings in noisy speech with limited resources. ICSLP-2000
- Picone, et al. Switchboard Statistics: Word Statistics. Institute for Signal and Information Processing, Mississippi State University, 1998. <http://www.isip.msstate.edu/projects/switchboard/>
- Dharanipragada S, Roukos S. A Fast Vocabulary Independent Al-

- algorithm for Spotting Words in Speech. ICASSP-98, pp. 233~236
- 6 Jelinek F. Self-organized Language Modeling for Speech Recognition. in Readings in Speech Recognition, A. Waibel and K. F. Lee, eds. Morgan-Kaufmann, San Mateo, CA, 1990. 450~506
  - 7 Zue V W. Navigating the Information Superhighway Using Spoken Language Interfaces. IEEE Expert, 1995, 10(5): 39~43
  - 8 Kawahara T, Doshita S. Topic Independent Language Model for Key-Phrase Detection and Verification. ICASSP-99, pp. 685~688
  - 9 Rohlicek J R, et al. Phonetic Training and Language Modeling for Word Spotting. ICASSP-93, vol. 2, pp. 459~462
  - 10 Weintraub M. Keyword-Spotting Using SRI's DECIPHER Large-Vocabulary Speech-Recognition System. ICASSP-93, vol. 2, pp. 463~466
  - 11 Meliani R E, O'Shaughnessy D. Accurate Keyword Spotting Using Strictly Lexical Fillers. ICASSP-97, pp. 907~910
  - 12 Rose R C, et al. Integration of Utterance Verification with Statistical Language Modeling and Spoken Language Understanding. ICASSP-98
  - 13 Rohlicek, et al. Continuous Hidden Markov Modeling for Speaker-Independent Word Spotting. ICASSP-89, pp. 627~630
  - 14 Rabiner L, Juang B-H. Fundamentals of Speech Recognition. Prentice Hall, 1993
  - 15 Davis S B, Mermelstein P. Comparison of Parametric Representations for Monosyllabic Word Recognition in Continuously Spoken Sentences. IEEE Trans. Acoustics, Speech, and Signal Processing, 1980, ASSP-28: 357~366
  - 16 Juang B H, Rabiner L R, Wilpon J G. On the Use of Bandpass Liftering in Speech Recognition, 1987, ASSP-35(7): 947~954
  - 17 Cox S. Speaker Normalization in the MFCC Domain. ICSLP-2000
  - 18 Zhen Bin, et al. On the Importance of Components of the MFCC in Speech and Speaker Recognition. ICSLP-2000
  - 19 Eric Chang, et al. Large Vocabulary Mandarin Speech Recognition with Different Approaches in Modeling Tones. ICASSP-2000
  - 20 Chen Yeou-Jiunn, et al. Utterance Verification Using Prosodic Information for Mandarin Telephone Speech Keyword Spotting. ICASSP-99. 697~700
  - 21 Gadde V R R. Modeling Word Durations. ICSLP-2000
  - 22 Wu Chung-Hsien, Chen Yeou-Jiunn, Hung Yu-Chun. Telephone Speech Multi-Keyword Spotting Using Fuzzy Search Algorithm and Prosodic Verification. ICSLP-98
  - 23 Wilcox L D, Bush M A. Training and Search Algorithms for an Interactive Wordspotting System. ICASSP-92. 97~100
  - 24 Rose Richard C, Paul Douglas B. A Hidden Markov Based Keyword Recognition System. ICASSP-90, vol. 1, pp. 129~132
  - 25 Rose R. Definition of Subword Acoustic Units for Wordspotting. Proc. EUROSPEECH-93, pp. 1049~1052
  - 26 James D A, Young S J. A Fast Lattice-Based Approach To Vocabulary Independent Wordspotting. ICASSP-94
  - 27 Manos A S, Zue V W. A Segment-based Wordspotter Using Phonetic Filler Models. ICASSP-97. 899~902
  - 28 Young S J, Odell J J, Woodland P C. Tree-Based State Tying for High Accuracy Acoustic Modeling. In: Proc. the Human Language Technology Workshop, March 1994. 307~312
  - 29 Jelinek, F. Statistical Methods for Speech Recognition. The MIT Press, 1998
  - 30 Bahl L B, et al. Decision Trees for Phonological Rules in Continuous Speech. ICASSP-91
  - 31 Ze'ev Rivlin, et al. A Phone-Dependent Confidence Measure for Utterance Rejection. ICASSP-96. 515~517
  - 32 Lee Kai-Fu. Automatic Speech Recognition. Kluwer Academic Publishers, 1989
  - 33 Chesta C, Laface P, Ravera F. Connected Digit Recognition Using Short and Long Duration Models. ICASSP-99
  - 34 Beaufays F, Weintraub M, Konig Y. Discriminative Mixture Weight Estimation for Large Gaussian Mixture Models. In: Proc. ICASSP, Phoenix, AZ, Mar. 1999. 337~340
  - 35 Juang B -H, Katagiri S. Discriminative Learning for Minimum Error Classification. IEEE Transactions on Signal Processing, 1992, 40(12): 3043~3054
  - 36 Lippmann R P, Singer E. Hybrid Neural-Network/HMM Approaches to Wordspotting. ICASSP-93. 565~568
  - 37 Sandness E D, Hetherington I L. Keyword-Based Discriminative Training of Acoustic Models. ICSLP-2000
  - 38 Wan E A, Nelson A T. Removal of Noise from Speech Using the Dual EKF Algorithm. ICASSP-98
  - 39 Takebayashi Y, Tsuboi H, Kanazawa H. A Robust Speech Recognition System Using Word-Spotting with Noise Immunity Learning. ICASSP-91. 905~908
  - 40 Guo Qing, et al. Keyword Spotting in Auto-Attendant System. ICSLP-2000
  - 41 Zheng Fang (郑方), et al. Harkman - A Vocabulary-Independent Keyword Spotter for Spontaneous Chinese Speech. J of Computer Science and Technology, 1999, 14(1): 18~26
  - 42 Weintraub M. LVCSR Log-Likelihood Ratio Scoring for Keyword Spotting. ICASSP-95, vol. 1 pp. 297~300
  - 43 Gish H, Ng K. A Segmental Speech Model with Applications to Word Spotting. ICASSP-93
  - 44 Tadj C, Poirier F. Keyword Spotting Using Supervised/Unsupervised Competitive Learning. ICASSP-95, vol. 1, pp. 301~304
  - 45 Chigier, Benjamin. Rejection and Keyword Spotting Algorithms for a Directory Assistance City Name Recognition Application. ICASSP-92, vol. 2, pp. 93~96
  - 46 Villarrubia, et al. Rejection Techniques for Digit Recognition in Telecommunication Applications. ICASSP-93, vol. 2, pp. 451~454
  - 47 吴立德,等. 大规模中文文本处理. 复旦大学出版社, 1997
  - 48 Niesler T, Woodland P. A Variable-Length Category-Based N-Gram Language Model. ICASSP-96, pp. 1164~1167
  - 49 Lau R, Rosenfeld R, Roukos S. Trigger-Based Language Models: A Maximum Entropy Approach. ICASSP-93, pp. 1145~1148
  - 50 Bellegarda J R. Large Vocabulary Speech Recognition with Multi-span Statistical Language Models. IEEE, Trans. Speech and Audio Processing, 2000, 8(1)
  - 51 Mahajan M, Beeferman D, Huang X D. Improved Topic-Dependent Language Modeling Using Information Retrieval Techniques. ICASSP-99
  - 52 Khudanpur S, Wu Jun. A Maximum Entropy Language Model Integrating N-Grams and Topic Dependencies for Conversational Speech Recognition. ICASSP-99
  - 53 Schwartz R, Nguen L, Makhoul J. Multiple-Pass Search Strategies. Automatic Speech and Speaker Recognition, Kluwer Academic Publishers, 1996. 429~456
  - 54 Ney H, Aubert X. Dynamic Programming Search: From Digit Strings to Large Vocabulary Word Graphs. Automatic Speech and Speaker Recognition, Kluwer Academic Publishers, 1996. 429~456

(下转第97页)

稳定的过程。

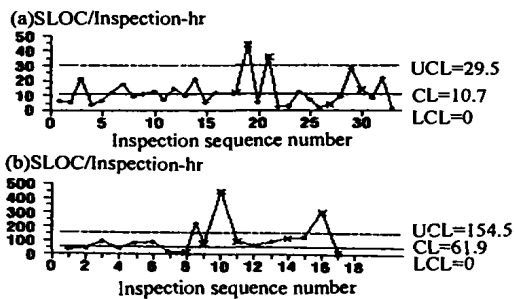


图5 (a) 审查率小于60 SLOC 的审查小组的数据  
(b) 审查率大于60 SLOC 的审查小组的数据

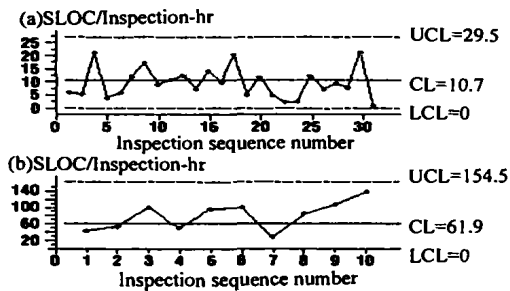


图6 排除再审查数据后控制图

需要注意的一点是,UCL、CL 和 LCL 的值都是通过计算得出的,增加了数据后应重新计算,以获得正确的结果。XmR 图的 UCL、CL、LCL 计算公式参见 5.1 节。

上文所举的 SPC 的应用实例是针对软件审查过程的,是通过 SPC 的方法判断审查过程的稳定性。这个例子有一定的特殊性。事实上,SPC 在软件质量管理中的应用远不止于此。例如,我们还可以用 SPC 方法来分析稳定的审查和测试过程产生的结果数据(如前文所述,可通过 SPC 方法判断过程的稳定性),以对产品质量进行评估,并可对将来正式发布版本的质量作出一定程度的预测。如果审查和测试活动发现的缺陷率保持稳定的话,说明开发过程应该是稳定的,产品的质量也保持稳定,这时我们可以根据中心线代表的值预测最终产品的质量;如果缺陷率出现异常,说明开发过程存在异常,可通过分析异常原因,对其进行改进。

**结论** 随着软件工程的发展,软件生产工程化的特点愈

来愈突出,像 SPC 这样的工程质量控制技术将会愈来愈多地用于软件工程领域。有效地实施 SPC 需要注意以下几点:

(1) 识别和定义软件工程的关键过程,确定可以实施 SPC 的比较稳定的软件过程,形式化并确定稳定过程的性能参数,坚持收集整理度量数据,理解并明确数据的含义。

(2) 数据的质量非常重要。将质量差的数据用于分析,不可能得到正确的结论。这要求度量技术能有进一步的发展。选择什么度量方法,度量时如何采样都关系到 SPC 实施的效果。

(3) 应用 SPC 方法需要对各种数据进行分析,虽然可以有相应的软件工具提供帮助,但发掘问题的原因,提出解决方案等工作只能是由有经验的管理人员来做。因此对于管理者的培训是十分重要的。

虽然,现在软件行业中 SPC 的应用仍不普遍,目前还只有有限的在高成熟度组织(CMM4级或5级)中的成功应用实例,但我们相信 SPC 技术可以在所有制定了度量程序的组织(2级以上)中实施并取得成效。

对实施 SPC 的收益应有正确的态度。SPC 只是用于帮助管理者保证当前定义的过程处于受控状态。它并不能保证产品满足顾客需要。如果一个受控过程生产的产品仍不符合顾客的需求,说明该过程定义不当。此时应从过程改进入手,采取更好的生产技术,提高过程的能力。

## 参考文献

- 1 Florac W A, et al. Practical Software Measurement: Measuring for Process Management and Improvement. [Tech Report CMU/SEI-97-hb003]. Software Eng. Inst., Carnegie Mellon Univ., Pittsburgh, 1997
- 2 Florac W A, et al. Statistical Process Control: Analyzing a Space Shuttle Onboard Software Process. IEEE SOFTWARE July/August 2000
- 3 Pressman R S. Software Engineering A Practitioner's Approach. McGraw-Hill, 1997
- 4 Weller E F. Practical Applications of Statistical Process Control. IEEE Software, May/June 2000
- 5 Paulk M, et al. Capability Maturity ModelSM for Software, Version 1.1. [Tech Report CMU/SEI-93-TR-024]. Software Eng. Inst., Carnegie Mellon Univ., Pittsburgh, 1993
- 6 Pressman R S 著. 黄柏素, 梅宏译. 软件工程——实践者的研究方法. 机械工业出版社, 1999

(上接第119页)

- 55 Li C N, Thompson S A. Mandarin Chinese: A Function Reference Grammar. University of California Press, 1981
- 56 Bai Bo-Ren, Tseng Chiu-Yu, Lee Lin-Shan. A Multi-Phase Approach For Fast Spotting Of Large Vocabulary Chinese Keywords From Mandarin Speech Using Prosodic Information. ICASSP-97. 903~906
- 57 Lin Chia-Hsien, Wang Hsiao-Chuan. Keyword Spotting by Searching the Syllable Lattices. ICSLP-2000
- 58 Lin C H, et al. Frameworks for Recognition of Mandarin Syllables with Tones Using Sub-Syllabic Units. In Speech Communication, Elsevier Science B. V., 1996. 18
- 59 Seide F, Wang N J C. Two-Stream Modeling of Mandarin Tones. ICSLP-2000
- 60 Gao Jianfeng, et al. A Unified Approach to Statistical Language

Modeling for Chinese. ICASSP-2000

- 61 Zhang Hong, Xu Bo, Huang Taiyi. How to Choose Training Set for Language Modeling. ICSLP-2000
- 62 Caminero J, López E, Hernández L. Two-Pass Utterance Verification Algorithm for Long Natural Numbers Recognition. ICSLP-98
- 63 Gu Yong, Thomas T. Competition-based Score Analysis for Utterance Verification in Name Recognition. ICSLP-2000
- 64 Gu Yong, Thomas T. A Hybrid Score Measurement For HMM-Based Speaker Verification. ICASSP-99
- 65 Jouvet D, Bartkova K, Mercier G. Hypothesis Dependent Threshold Setting for Improved Out-of-Vocabulary Data Rejection. ICASSP-99
- 66 Bazzi I, Glass J R. Modeling Out-of-Vocabulary Words for Robust Speech Recognition. ICSLP-2000