

利用 CASSANDRA-I 神经计算机实现 有限词汇连续语音识别¹

顾晓东 余道衡 赵鹤鸣*

(北京大学电子学系 视觉与听觉信息处理实验室 北京 100871)

*(苏州大学通信与电子工程系 苏州 215006)

摘 要 该文提出并讨论了如何用小型神经计算机 CASSANDRA-I 进行有限词汇非特定人连续语音识别的方法,同时还提出了一种新的基于关键词的句法分析方法。利用 CASSANDRA-I 小型神经计算机,可满足有限词汇连续语音实时识别的要求;利用基于关键词的句法分析,可提高有限词汇连续语音识别的识别率与识别速度。

关键词 神经计算机, 句法分析, 语音识别

中图分类号 TP391.42, TN912.3

1 引言

当今信息时代,语音识别的研究对于通信、计算机、国防、办公室自动化等领域均有重大的现实意义。目前,大词汇表的连续语音识别是语音识别研究的热点。同时,有限词汇非特定人连续语音实时识别是语音识别研究的又一重点。有限词汇非特定人连续语音实时识别在许多领域都有其用武之地,如信用卡的认证、股市咨询、数据库检索、在高速行驶的汽车中进行语音拨号、旅游景点的查询等,对其的研究虽已取得一些成果,但目前仍处于研究之中。

语音实时识别时,识别速度与识别率是两个重要的性能指标。语音识别时,大部分时间花在音节模式匹配上,故若能提高音节模式匹配速度,则能明显地提高识别速度。本文提出使用 CASSANDRA-I 小型神经计算机^[1]进行音节模式匹配,这充分利用了 CASSANDRA-I 小型神经计算机的高速并行计算能力,从而使识别速度很快,实验表明,利用其进行有限词汇连续语音识别,可满足实时识别的要求,这将在本文第 2 节介绍。同时,本文针对有限词汇连续语音识别,提出了基于关键词的句法分析方法,实验表明,使用该方法后,语句识别率得到很大的提高。这将在本文第 3 节介绍。本文第 4 节将给出一采用了上述两种方法的有限词汇连续语音识别系统的测试结果。

2 用神经计算机 CASSANDRA-I 进行音节模式匹配

采用神经计算机进行音节模式匹配,这充分利用了神经计算机并行分布处理的能力与高效的学习算法,大大减少了识别时间,从而使有限词汇连续语音识别可满足实时识别的要求^[2-5]。本文使用的神经计算机是中国科学院半导体研究所研制的 CASSANDRA-I 小型神经计算机^[1],该机模拟 256 个神经元,有 256 个输入节点。为了便于其处理,可将输入的单个音节的特征数归整为 256。

CASSANDRA-I 小型神经计算机的识别原理:

训练及待识别的音节样本相当于 n 维空间中的一个点。将训练后得到的音节模板样本 $\bar{W} = (w_1, w_2, \dots, w_n)$ 和待识别的音节样本 $\bar{X} = (x_1, x_2, \dots, x_n)$ 作如下规整: $\sum_{i=1}^n w_i^2 = C^2$, $\sum_{i=1}^n x_i^2 = C^2$, 其中 C 为常数。

如此规整,即将模板样本点和待识别样本点均映射到模式空间的一个超球面上,见图 1。该图虽画成二维的情况,但表示多维的情况。

图 1 中 \bar{W} 和 \bar{X} 是半径为 C 的超球面上任意两点, θ 是 \bar{W} 和 \bar{X} 的夹角。 \bar{W} 和 \bar{X} 的欧氏距离为: $|\bar{W} - \bar{X}|^2 = \sum_{i=1}^n (w_i - x_i)^2 = \sum_{i=1}^n w_i^2 + \sum_{i=1}^n x_i^2 - 2 \sum_{i=1}^n w_i x_i = C^2 + C^2 - 2 \sum_{i=1}^n w_i x_i$ 。

¹ 2000-10-16 收到, 2001-02-05 定稿
国家自然科学基金资助项目

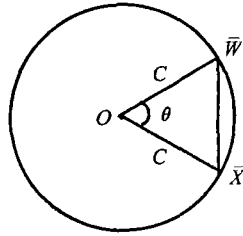


图 1 模式空间中的超球面

在图 1 的 $\triangle O\bar{w}\bar{x}$ 中, 对照三角形余弦公式有

$$C^2 \cos \theta = \sum_{i=1}^n w_i x_i$$

由上式可知, 如果将待识别音节样本作为神经网络的输入矢量 \bar{x} , 而音节模板样本作为神经网络的权值矢量 \bar{w} , 则各神经元输入端对乘积累加和的贡献正比于它的权值矢量与输入矢量之间夹角的余弦。因此, 神经网络输出的大小可以用来衡量输入样本与模板间的相似度。

CASSANDRA-I 小型神经计算机依此原理进行音节模式匹配, 66ms 可完成一个测试样本与 256 个 256 维的模板样本的比较。

3 基于关键词的句法分析

本文针对有限词汇连续语音识别, 提出了基于关键词的句法分析。有限词汇连续语音识别时, 在得到音节模式匹配的结果后, 对其进行基于关键词的句法分析, 从而得到最终的识别结果。连续语音识别中进行句法分析的目的是使识别结果不仅在声学、语音特性上具有最大似然特性, 而且在句法结构上也具有最大似然特性; 同时, 进行句法分析时, 利用句法结构知识可知输入连续语音在句法上的排列情况, 由此减少分割后的单个音节匹配时需考虑的匹配范围, 从而提高识别率与识别速度。有限词汇连续语音识别时, 音节参考模式库可根据句法结构模型建立。要进行句法分析, 首先需建立句法结构模型, 然后再据此模型采取一定的算法进行搜索, 得到识别结果。

3.1 句法模型的建立 句法模型的建立需依据具体的系统来建立。目前有一些方法(如位格文法、扩大转移文法等)可用来建立句法模型^[6-8], 用于连续语音识别的句法模型可采用上下文无关文法。

本文针对有限词汇连续语音识别中, 所需识别语句有限, 提出可对这些语句进行关键词分析, 找出关键词, 建立起基于关键词的句法模型。由于所建立的基于关键词的句法模型抓住了各条语句的关键词, 匹配时无需对输入的待识别连续语音的全部音节进行匹配。这样不但提高了识别速度, 还增强了系统的鲁棒性。

3.2 K -BEST-FIRST 语句识别算法 建立起句法结构模型后, 就可据此模型采取一定的策略^[6,8]进行搜索, 从而得到符合句法结构的识别结果。根据前面提出的基于关键词的句法模型, 我们提出了 K -BEST-FIRST 语句识别算法, 该算法的流程图如图 2 所示。

以上算法中, 第一次派生的结果为 K 个具有较高匹配似然度的局部结果, 对应 K 条局部派生路径, 后续的派生仅仅扩展具有最大匹配似然度的状态. K 值可以根据实验来确定. 匹配似然度的计算采用平均似然度计算方法. 有限词汇连续语音识别中, 待识别语句的声母特征的分布具有一定的模式, 因此在所采用的 K -BEST-FIRST 语句识别算法中, 并不直接将结果表中具有最大匹配似然度的候选结果作为最终识别结果, 而是采用声母段特征^[9]对候选结果根据匹配似然度大小依次进行检验, 从而得到具有最大匹配似然度且满足超音段特征的最终识别结果. 在图 2 的继续派生方框图内, 采取了措施以减少某些情况下由强连音引起的音节分割错误的影响. 具体做法是, 若派生得到的词汇在本层的匹配似然度小于阈值, 则忽略小于阈值的匹配似然度, 无条件跳转到所有可能的后继状态, 继续派生. 这样, 大大增强了系统的鲁棒性, 提高了识别率; 同时也大大减少了前面音节分割错误带来的影响, 改善了识别性能.

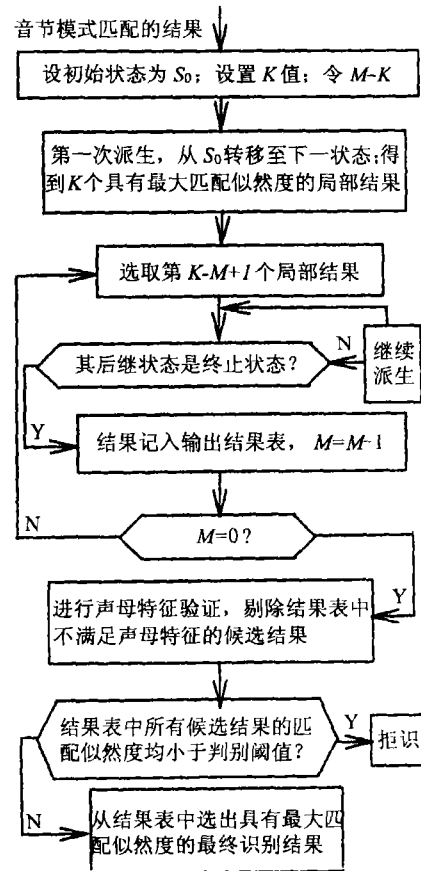


图 2 基于关键词的句法分析的语句识别算法的流程图

4 实验与讨论

实验表明, 在一共需识别 52 条语句的有限词汇非特定人连续语音识别系统中, 采用 CASSANDRA-I 小型神经计算机和基于关键词的句法分析后, 基本能满足实时识别的要求, 且识别率较未采用句法分析时有了很大的提高.

系统采用 CASSANDRA-I 小型神经计算机进行音节模式匹配时, 音节模式匹配耗时少, 单条语句所需的最长识别时间只需 0.8s, 基本能满足实时识别的要求.

采用 10 男 10 女共 20 人, 在系统采用基于关键词的句法分析和未用句法分析时, 分别进行测试; 该 20 人非系统训练者; 每人将 52 条语句用普通话以正常语速发音一遍; 每人测试两次, 间隔时间为三周. 测试结果见表 1.

系统的训练者 10 人 (5 男 5 女) 采用与前面同样的测试方法, 得到的测试结果见表 2.

表 1 与表 2 中未用句法分析的识别方法是指根据音节模式匹配的结果, 直接得到各候选语句的匹配似然度, 然后选取拥有最大匹配似然度且符合对应声母特征的语句为最终的识别结果. 从表 1 与表 2 的测试结果可知, 采用了本文提出的基于关键词的句法分析后, 系统的识别率较未采用句法分析时有了很大的提高.

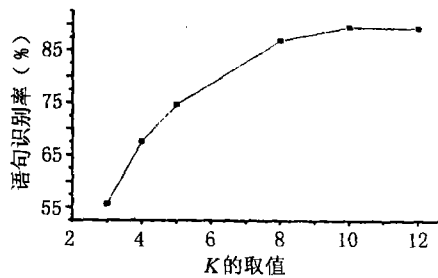
表 1 20 个非训练者的测试结果

类别	测试的语句数	正确识别的语句数	误识的语句数	拒识的语句数	误识率	拒识率	识别率
未用句法分析	2080	994	860	226	41.3%	10.9%	47.8%
采用本文句法分析	2080	1860	156	64	7.5%	3.1%	89.4%

表 2 10 个训练者的测试结果

类别	测试的语句数	正确识别的语句数	误识的语句数	拒识的语句数	误识率	拒识率	识别率
未用句法分析	1040	601	357	82	34.3%	7.9%	57.8%
采用本文句法分析	1040	966	56	26	5.4%	2.5%	92.9%

实验中还发现,采用语句识别算法 K -BEST-FIRST 时,第一次派生结果的个数 K 也会影响到语句的识别率.表 1 与表 2 中,采用句法分析后的语句识别率是在 $K = 10$ 的条件下得到的.系统的训练者 20 人(10 男 10 女)采用与前面同样的测试方法,得到 K 值对语句识别率的影响见图 3.

图 3 K 值对识别率的影响

5 结论

有限词汇连续语音识别时,采用小型神经计算机 CASSANDRA-I 进行音节模式匹配,可实现实时识别.进行大词汇表的连续语音识别时,亦可用 CASSANDRA-I 进行音节模式匹配,从而提高识别速度.有限词汇连续语音识别中,运用本文提出的基于关键词的句法分析方法,可极大地提高识别率,同时可提高识别速度,增强鲁棒性.

参 考 文 献

- [1] 王守觉,等,用神经计算机的说话人确认系统及其应用,人工神经网络与神经计算机应用研讨会,北京,1998,7,9-13.
- [2] R. P. Lippmann, An introduction to computing with neural nets, IEEE ASSP Magazine, 1987, 4(2), 4-22.
- [3] R. P. Lippmann, Review of neural networks for speech recognition, Neural Computation, 1989, 1(1), 1-38.
- [4] Xin Jiang, Zhengyu Gong, Fan Sun, Huisheng Chi, A hybrid speaker recognition system based on the auditory path model, Proc. of WCNN'93, Portland, June 1993, III598-601.
- [5] A. Waibel, et al., Phoneme recognition using time-delay networks, IEEE Trans. on ASSP, 1989, ASSP-37(12), 1888-1898.
- [6] 俞一彪,袁保宗,连续语音识别中句法结构知识的利用,电子学报,1990,18(6),68-74.
- [7] H. Ney, Dynamic programming parsing for context-free grammars in continuous recognition, IEEE Trans. on SP., 1991, SP-39(2), 336-340.

- [8] 徐近儒, 高枚, 汉语文本读入中音字转换的知识集成模型和时间搜索同步算法, 中文信息学, 1994, 8(1), 33-44.
- [9] 齐士铃, 张家禄, 汉语普通话辅音音长分析, 声学学报, 1982, 7(1), 8-13.

SPEECH RECOGNITION WITH LIMITED VOCABULARY BY USING CASSANDRA-I NEUROCOMPUTER

Gu Xiaodong Yu Daoheng Zhao Heming*

*(Dept. of Electron., Nat. Lab of Machine Perception and Center of Info. Sci,
Peking University, Beijing 100871, China)*

**(Dept. of Comm. and Electron. Eng., Suzhou University, Suzhou 215006, China)*

Abstract A new approach for continuous speech recognition based on neurocomputer CASSANDRA-I and syntactic analysis based on key words is proposed. By using CASSANDRA-I, the speed of the recognition is so rapid that recognition can be at real time. By using syntactic analysis based on key words, the performance of recognition improves greatly.

Key words Neurocomputer, Syntactic analysis, Speech recognition

顾晓东: 男, 1970 年生, 博士生, 从事信号与信息处理、通信、人工神经网络等领域的研究.

余道衡: 男, 1939 年生, 教授, 博士生导师, 中国电子学会会士, IEEE 高级会员, 从事电路与系统、信号与信息处理、现代通信等领域的教学与科研工作.

赵鹤鸣: 男, 1957 年生, 教授, 硕士生导师, 从事信号与信息处理、人工神经网络等领域的研究.