

基于正交实验设计的语音识别特征参数优化

韩志艳¹ 王 健¹ 王 旭²

(渤海大学信息科学与工程学院 锦州 121000)¹ (东北大学信息科学与工程学院 沈阳 110004)²

摘 要 鉴于语音识别性能与所选用的语音特征参数密切相关,提出一种系统性的实用的特征参数优化方法——基于方差的正交实验设计法。首先进行因素(语音特征参数)和水平的选择,再根据数理统计与正交性原理,从大量的实验点中挑选适量的具有代表性、典型性的点构造正交表进行正交实验,最后通过计算对正交实验结果进行分析,找出最优的特征参数组合。与目前参数的简单组合方案相比较,新方法的误识率下降了 5.6%,响应时间减少了 181.37ms。实验结果表明,正交实验设计用于语音特征参数优化是有效的,对后续研究具有指导意义。

关键词 正交实验设计,语音识别,参数优化,正交性

中图分类号 TP391.9 **文献标识码** A

Parameter Optimization Based on Orthogonal Test Design in Speech Recognition System

HAN Zhi-yan¹ WANG Jian¹ WANG Xu²

(School of Information Science & Engineering, Bohai University, Jinzhou 121000, China)¹

(School of Information Science & Engineering, Northeastern University, Shenyang 110004, China)²

Abstract The performance of speech recognition depends on the choice of speech parameters. A systematic and practical method of the parameters selection based on the variance orthogonal test design was proposed. Firstly, chose factors (speech parameters) and levels. And then according to the principle of mathematical statistics and orthogonality, picked out proper and representative points from a large number experimental points to construct orthogonal table. Finally, calculated and analysed the experimental results, and the optimal set of process parameter values was discovered. Moreover, the word error rate is reduced 5.6% and response time is reduced 181.37ms when compared with that of the traditional parameter selection method. The experiment results show that the application of orthogonal test design is effective in parameter optimization. The conclusion provides guidance for the subsequent research.

Keywords Orthogonal test design, Speech recognition, Parameter optimization, Orthogonality

从 20 世纪 60 年代人类便开始了语音识别技术的研究,其目的就是使机器能够理解人类的自然语言。由于语言信号本身非常复杂,加上人类对语音学知识还很缺乏,因此目前的任何一种语音识别方法在识别率和实时反应速度上都还很难达到人类本身的听觉效果。而语音识别又分为特定人和非特定人识别,参加训练的人的未训练语音样本和未参加训练的人的样本,前者的识别率达到 95% 以上,而后的识别率只有 80% 左右。测试结果表明,系统对特定人(参加训练的人的样本)的识别率效果很好,对非特定人(未参加训练的人的样本)的识别效果有待进一步提高,这一点也是语音识别研究的关键所在,要想提高其性能,必须在非特定人的语音识别率上下功夫,即在语音信号特征提取上下功夫。由于语音识别设备可能用在各种环境中,因此要求系统具有较强的环境噪声的抗干扰能力和很好的鲁棒性,如果还坚持用单一参数来识别,那是不可能完成高鲁棒性这一要求的,所以笔者提出用组合特征的方法,但是如果把这些特征漫无目的地简单地组合在一起,不但不会提高系统性能,而且还会使系统负荷过

重,结果无法实现控制功能。如何在大量的特征参数中选择出少数具有互补作用的特征参数,既是一个理论问题,又是一个实际应用问题。目前特征参数选择主要有两种方法,一种是根据实验语音学和听觉方面的知识将特征参数分为若干类,通过对比实验在各类特征参数中选出一类。另一种方法是对一类特征参数通过实验确定特征向量的阶数,包括动态特征的阶数。这两种选择方法的共同之处是最终得到的特征参数是一类或者简单地加上其它几个特征参数的组合,不能找到不同类型特征参数的组合方法。在其他模式识别领域中有一些特征参数选择方法,例如,文献[1]给出了一种多目标优化的特征参数选择方法,但是,由于语音识别的特征参数选择时的实验量太大,因此,该方法是不可行的。以前的实验表明,在识别模型不变的条件下,识别器性能与特征向量维数之间没有单调上升的对应关系,简单地往特征维数上添加新特征参数并不能提高系统的识别性能。

针对这一问题,笔者提出一种基于方差的正交实验设计法^[2-4]来对特征参数进行优化组合。这种方法具有系统性、实

到稿日期:2009-02-25 返修日期:2009-05-06 本文受国家自然科学基金(50477015)资助。

韩志艳(1982-),女,博士,讲师,CCF 会员,主要研究方向为语音信号处理,E-mail: hanzyme@126.com;王 健(1982-),男,硕士,主要研究方向为信号处理;王 旭(1956-),男,博士生导师,主要研究方向为智能控制技术、交流调速技术、神经网络技术、语音信号处理等。

验效率高和当前的实验结果对将来的实验方案设计有指导意义等优点,从而消除了上述两个隐患。

1 语音特征参数正交实验设计

1.1 基本正交实验设计简介

正交实验设计法,是使用已经造好了的表格——“正交表”,来安排实验的。它能在很多的实验中选出代表性强的少数实验条件,并能通过少数实验条件,推断找到最好的工艺条件。正交实验设计在发达国家的农业实验、工艺流程设计中应用很广泛,在我国也有很多成功应用的例子^[1]。这种方法将影响实验结果的因素称为因子,因子的各个不同状态称为水平,为找到各因子较优的水平组合而进行的实验总次数比穷举法(所有因子的任一水平组合都进行搭配)要少很多倍。例如对2水平的30个因子,穷举法要做230次实验,而按照正交表安排实验只需做32次。工作量减少了很多,而在一定程度上代表了230次实验。正交实验法确定各因子的较优水平所需的步骤分为以下3个阶段^[5,6]。

第一阶段:构造正交表

构造一般正交表需要用到很高深的数学理论,但是当所有因子都是2水平时,正交表很容易构造,具体方法参见文献^[7]。这样构造出来的正交表有如下两个特点:

- 1) 每列中各水平都会出现,且出现的次数相等;
- 2) 任何两列之间各种不同水平的所有可能组合都会出现,且出现的次数相等。

第二阶段:正交实验

由于正交表有上述两条性质,因此用它来安排实验时,各因子的各种搭配是均衡的。正交表中每行表示一个实验方案,是各个因素所处状态的一种组合;每列数字表示对应因子所处的状态。按正交表对应的实验方案完成实验,将结果填入正交实验表中。

第三阶段:正交实验结果分析

通常在对正交实验结果进行分析时,都是采用极差法,因为极差法考虑简单易懂,计算量小,而且可以直观地描述。但是极差法没有把实验过程中由实验条件改变所引起的数据波动与由实验误差引起的数据波动区分开来,也没有提供一个标准来判断考察的因素的作用是否显著。为了弥补极差法的这个缺点,本文采用方差分析的方法,方差分析正是将因素水平的变化所引起的实验结果间的差异与误差的波动所引起的实验结果间的差异区分开来的一种数学方法。

利用“正交表”所作出的实验结果,根据方差分析理论,可以将因子水平的变化所引起的实验结果间的差异区分开来。如果因子水平的变化所引起的实验结果的变动落在误差范围以内,或者与误差相差不大,就可以判断这个因子水平的变化并不会引起实验结果的显著变动;相反,如果因子水平的变化引起的实验结果的变动超过误差范围,就可以判断这个因子水平的变化会引起实验结果的显著变动。其目的是通过数据分析,找出对该事物有显著影响的因子。这里应用了方差分析的数学理论,包括:

(1) 偏差平方和

$$S_k = \frac{1}{n} \sum_{i=1}^m I_i^2 - C \quad (1)$$

其中, n 为实验方案数, m 为水平数, r 为重复实验数, $I_i (i=1,$

$2, \dots, m)$ 是 i 水平对应的结果总和。 $C = \frac{T^2}{n}$ 为修正系数,其中

$T = \sum_{i=1}^m I_i$ 。偏差平方和反映了一组数据的分散和集中程度, S_k 大说明这组数据分散,反之说明数据集中。

(2) 平均偏差平方和和自由度

为合理地比较两组个数不同的数据分散和集中程度,应采用平均偏差平方和 $\bar{S}_k = \frac{S_k}{m-1}$ 。式中 $m-1$ 是 S_k 的自由度。

(3) F 比

因子水平的改变引起的平均偏差平方和与误差的平均偏差平方和的比值称为 F 比。

$$F = \frac{S_{\text{因子}}/f_{\text{因子}}}{S_e/f_e} \quad (2)$$

$$S_e = S - \sum S_{\text{因子}} \quad (3)$$

其中, $S = \sum_{i=1}^n y_i - C$, y_i 为识别结果, f 是自由度。

(4) 信度 α

在判断 F 比时,信度 α 是指对作出的判断大概有 $1-\alpha$ 的把握。

(5) 显著性

设因子 A 的 F 比为 F_A 。当 $F_A > F_{0.01}(n_1, n_2)$ 时,说明该因子水平的改变对实验结果有高度显著的影响,记作“*”;当 $F_{0.01}(n_1, n_2) > F_A > F_{0.05}(n_1, n_2)$ 时,说明该因子水平的改变对实验结果有显著影响,记作“*”;当 $F_{0.05}(n_1, n_2) > F_A > F_{0.10}(n_1, n_2)$ 时,说明该因子水平的改变,对实验结果有一定的影响,记作“(*)”。当 $F_{0.10}(n_1, n_2) > F_A > F_{0.25}(n_1, n_2)$ 时,称该因素不显著但有影响,记作“[*]”;当 $F_A \leq F_{0.25}(n_1, n_2)$ 时,称该因素无影响,不做记号。

1.2 因子和水平的选择

笔者选用的特征参数共有30个,其排列顺序为三个共振峰特征、一个短时过零率,一个谐振强度特征^[8,9]、12阶的LPCC特征和13阶的MFCC特征,故因子数为30。对每个特征参数而言,有两种可能的状态,即选用与不选用,1表示选用,2表示不选用,所以每个因子是2水平的。

1.3 正交表的选用及构造

对于2水平正交表 $L_{32}(2^{30})$,可采用哈达马矩阵^[6]构造:

设二阶哈阵 $H_2 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$ 为基本哈阵,

$$H_4 = H_2 \otimes H_2 = \begin{bmatrix} H_2 & H_2 \\ H_2 & -H_2 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix} \quad (4)$$

以此类推, $H_{32} = \begin{bmatrix} H_{16} & H_{16} \\ H_{16} & -H_{16} \end{bmatrix}$ 。去掉 H_{32} 中的第一列,

做30个因子的实验,只需30列,还需要去除1列,为简单起见,去除了后1列,并将-1替换为2,最后得到一个 32×30 的矩阵。

2 正交实验结果与分析

本文对50个语音命令在噪声为80分贝时进行了识别,选择的训练样本集和测试样本集分别来自不同人的发音,即

主要针对非特定人进行识别。采样频率为 11.025kHz,量化精度为 16bit,语音的帧长为 88 个采样点,数据采集前不做任何训练。

表 1 是按正交实验设计法设计的方案及实验结果,每一行表示一种组合方案,每一列对应一个特征参数。

表 1 正交实验结果

试验号	实验方案	识别率/%
1	11111111111111111111111111111111	92.74
2	212121212121212121212121212121	92.23
3	122112211221122112211221122112	91.35
4	221122112211221122112211221122	90.88
5	111222211112222111122221111222	92.45
6	212212112122121121221211212212	92.60
7	122221111222211112222111122221	90.33
8	221211212212112122121121221211	90.97
9	1111112222222111111112222222	90.72
10	21212122121211212121212121212	91.45
11	12211222211221112211222211221	92.39
12	221122121221122112211221212211	92.08
13	1112222222111111122222222111	90.04
14	212212121212121212122121212121	91.55
15	122221122111122112222112211112	91.90
16	221211221121221122121122112122	92.56
17	1111111111111222222222222222	91.20
18	21212121212121212121212121212	90.03
19	122112211221122221122112211221	92.44
20	221122112211221211221122112211	92.00
21	11122221111222222211112222111	91.45
22	212212112122121212121212212121	89.97
23	122221111222211221111222211112	92.49
24	221211212212112211212212112122	92.12
25	111111222222222222222211111111	92.32
26	21212122121212121212121212121	92.69
27	12211222211221122112211122112	90.12
28	221122121122112211221121221122	91.67
29	11122222221111122221111111222	91.55
30	21221212121212212112121212212	92.00
31	12222112211112222111221122221	89.78
32	221211221121221211212211221211	90.67

相应的特征参数的排列顺序为: $F_1, F_2, F_3, Z, H, LPCC_1, LPCC_2, LPCC_3, LPCC_4, LPCC_5, LPCC_6, LPCC_7, LPCC_8, LPCC_9, LPCC_{10}, LPCC_{11}, LPCC_{12}, MFCC_1, MFCC_2, MFCC_3, MFCC_4, MFCC_5, MFCC_6, MFCC_7, MFCC_8, MFCC_9, MFCC_{10}, MFCC_{11}, MFCC_{12}, MFCC_{13}$ 。水平 1 表示用该特征参数,水平 2 表示不用该特征参数,表的最后一列是针对非特定人的各实验方案的识别结果。

经查表得, $F_{0.01}(1,1) = 4052, F_{0.05}(1,1) = 161.4, F_{0.10}(1,1) = 39.86, F_{0.25}(1,1) = 5.83$ 。

对表 1 的数据进行了分析,其结果如表 2 所列。

表 2 正交实验结果

特征参数	自由度	F	显著性	理论最优方案
F_1	1	6.54	[*]	1
F_2	1	2.08		2
F_3	1	5.96	[*]	1
Z	1	20.35	[*]	1
H	1	9.86	[*]	1
$LPCC_1$	1	0.16		2
$LPCC_2$	1	0.11		2
$LPCC_3$	1	4.19		2
$LPCC_4$	1	122.03	(*)	1
$LPCC_5$	1	1.52		2
$LPCC_6$	1	9.42	[*]	1

$LPCC_7$	1	11.37	[*]	1
$LPCC_8$	1	0.18		2
$LPCC_9$	1	4.00		2
$LPCC_{10}$	1	2.57		2
$LPCC_{11}$	1	18.91	[*]	1
$LPCC_{12}$	1	9.14	[*]	1
$MFCC_1$	1	2.65		2
$MFCC_2$	1	31.41	(*)	1
$MFCC_3$	1	1.35		2
$MFCC_4$	1	25.47	[*]	1
$MFCC_5$	1	0.39		2
$MFCC_6$	1	1.35		2
$MFCC_7$	1	0.002		2
$MFCC_8$	1	20.99	[*]	1
$MFCC_9$	1	803.66	*	1
$MFCC_{10}$	1	6.54	[*]	1
$MFCC_{11}$	1	15.81	[*]	1
$MFCC_{12}$	1	0.14		2
$MFCC_{13}$	1	2.80		2

由表 2 可以看出,对识别率影响较大的特征参数为: $MFCC_9, LPCC_4, MFCC_2, MFCC_4, MFCC_8, Z, LPCC_{11}$, 这些特征参数在后面的实验方案中应该严格按照理论较优方案中的值来确定是否选用。对识别率影响较小的特征参数为: $MFCC_7, LPCC_2, MFCC_{11}, LPCC_1, MFCC_5$, 这些特征参数对在后面的实验方案中是否被选用影响不大。

由此可见,已做实验可以用于指导后续实验方案的设计,这正是正交实验法的价值所在。

图 1 为实验正交趋势图,由正交趋势图可以很容易地看出每个参数对识别率的影响程度,显著性大则表明该因素的存在对识别器性能影响较大。

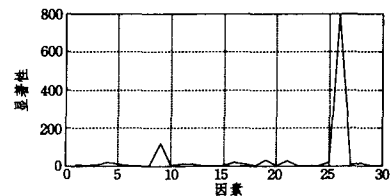


图 1 因素显著性趋势图

3 对比实验结果与分析

理论较优特征参数组合方案是经过分析得到的,为了检验是否真正较优,我们又做了一次实验,结果如表 3 所列。

表 3 正交实验结果

组合方案	维数	识别率(%)	总耗时(ms)
全特征	30	92.74	378.67
理论较优方案	15	93.15	560.04

由表 3 可以看出,在所做的 32 个正交实验中,选取全部 30 个特征参数的 1 号实验性能最好,达到了 92.74% 的正确识别率。但是经过理论分析得到的理论较优特征参数组合只选了 15 个特征参数,其识别性能却比 1 号实验好,得到了 93.15% 的识别率。另外,本文方法的速度较快,特征抽取和识别过程所消耗的总时间为 378.67ms,而传统的方法则需要 560.04ms。

结束语 本课题采用正交表设计的实验方案虽然对全体因素来说是一种部分实验,但是对任意两个因素来说实验条件是均匀搭配的,具有很强的代表性,因此,既可以减少实验

(下转第 250 页)

参考文献

- [1] The Special Issue on Computational Anaphora Resolution[J]. *Computational Linguistics*, 2001, 27(4)
- [2] Mitkov R. Anaphora resolution; the state of the art. Working paper(Based on the COLING'98/ACL'98 tutorial on anaphora resolution)[M]. University of Wolverhampton, Wolverhampton, 1999
- [3] Renata V, Massimo P. An Empirically-based System for Processing Definite Descriptions[J]. *Computational Linguistics*, 2000, 26(4): 525-579
- [4] Ge Niyu, John H, Eugene C. A Statistical Approach to Anaphora Resolution[C]// *Proceedings of COLING-ACL8*. Canada, 1998: 161-170
- [5] Soon W M, Ng H T, Lim D C. A Machine Learning Approach to Coreference Resolution of Noun Phrases[J]. *Computational Linguistics*, 2001, 27(4): 521-544
- [6] Strube M, Muller C. A Machine Learning Approach to Pronoun Resolution in Spoken Dialogue[C]// *Proceedings of 2003 ACL*. 2003
- [7] Orasan C, Evans R, Mitkov R. Enhancing Preference-Based Anaphora Resolution with Genetic Algorithms[C]// *Proceedings of NLP2000*. University of Patras, Greece, 2000: 185-195
- [8] Mitkov R. Anaphora resolution; a combination of linguistic and statistical approaches[C]// *Proceedings of the Discourse Anaphora and Anaphor Resolution*. Lancaster, UK, 1996
- [9] 李家治, 陈永明. 机器理解自然语言中有关代词处理的几个问题[C]// *自然语言理解年会论文集*. 1986
- [10] 许敏, 王能忠, 马彦华. 汉语中指代问题的研究及讨论[J]. *西南师范大学学报: 自然科学版*, 1999, 6: 633-637
- [11] 王厚峰, 何婷婷. 汉语中人称代词的消解研究[J]. *计算机学报*, 2001, 24(2): 6-13
- [12] 黄曾阳. HNC(概念层次网络)理论—计算机理解语言研究的新思路[M]. 北京: 清华大学出版社, 1998
- [13] 王厚峰, 梅铮. 鲁棒性的汉语人称代词消解[J]. *软件学报*, 2005, 16(5): 700-707
- [14] Mitkov R, Evans R, Orasan C. A new, fully automatic version of Mitkov's knowledge-poor pronoun resolution method[C] // *Proceedings of CICLing-2002*. Mexico, 2002: 168-186
- [15] Pawlak Z. Rough sets. Theoretical aspects of reasoning about data[M]. Kluwer Academic Publishers, 1991
- [16] Zhu W, Wang F Y. On Three Types of Covering Rough Sets [J]. *IEEE Transactions On Knowledge and Data Engineering*, 2007, 19(8): 1131-1144
- [17] Zhu W, Wang F Y. Reduction and Axiomatization of Covering Generalized Rough Sets[J]. *Information Sciences*, 2003, 152: 217-230
- [18] Dubois D, Prade H. Putting fuzzy sets and rough sets together. *Intelligent Decision Support*[M]// *Slowinski R*, ed. Kluwer Academic Publishers, 1992: 203-232
- [19] Wu Wei-zhi, Mi Ju-sheng, Zhang Wen-xiu. Generalized fuzzy rough sets[J]. *Information Sciences*, 2003, 151: 263-282
- [20] Chen D G, Wang X Z, Yeung D S, et al. Rough approximations on a complete completely distributive lattice with applications to generalized rough sets[J]. *Information Science*, 2006, 176: 1829-1848
- [21] <http://www.icl.pku.edu.cn>
- [22] 王厚峰. 汉语篇章的指代消解浅论[J]. *语言文字应用*, 2004, 4: 113-119
- [23] Halliday M, Hasan A K. *Cohesion in English* [M]. London, Longman Group, 1976: 27-28
- [24] Yeung D S, Chen D G, et al. On the Generalization of Fuzzy Rough Sets[J]. *IEEE Transactions on Fuzzy Systems*, 2005, 13(3): 343-361
- [25] Mitchell T M. *Machine Learning*[M]. McGraw-Hill Companies, Inc, 1997
- [26] Ng V, Cardie C. Improving Machine Learning Approaches to Coreference Resolution[C]// *Proceedings of 2002 ACL*. Philadelphia, 2002: 104-111
- [27] Pawlak Z. Rough sets, decision algorithms and Bayes' theorem [J]. *European Journal of Operational Research*, 2002, 136: 181-189
- [28] Grzymala-Buss J W. LERS(A system for learning from examples based on rough sets) *Intelligent Decision Support*[M]// *R. Slowinski*, ed. Kluwer Academic Publishers, 1992: 3-18
- [29] http://www.keenage.com/html/c_index.html

(上接第 216 页)

次数, 又能够比较全面地反映各因素的不同水平对实验指标的影响。对于因素复杂的实验问题, 它是一种行之有效的办法。在应用正交实验设计方法时, 除了依靠数学分析外, 还必须根据实践经验和专业知识, 对具体问题具体分析, 对实验结果进行验证, 这样才能得到正确可靠的结论。在语音识别的特征参数选择中, 采用正交实验设计法, 通过少数有代表性的几个试验, 得出了所需要的结论, 大大缩短了实验周期, 避免了盲目的实验, 提高了设计效率。

另外, 参数的简单组合方案特征维数较高, 直接影响随后的分类识别的速度和控制速度, 而按照本文所提出的方法组合后避免了维数的增加, 从而在提高识别率的同时不至于过多地增加时间开销。

参考文献

- [1] Emmanouilidis C, Hunter A. Multiobjective Evolutionary Setting for Feature Selection and a Commonality-Based Crossover Operator[C]// *2000 IEEE International Conference on Evolutionary Computation, CEC '2000*. 2000: 309-316
- [2] Ho S Y, Lin H, Liauh W H, et al. Orthogonal Particles Swarm Optimization and Its Application to Task Assignment Problems [J]. *IEEE Transactions on Systems, Man and Cybernetics*, 2008, 38(2): 288-298
- [3] Liang X B. Orthogonal Designs with Maximal Rates[J]. *IEEE Transactions on Information*, 2003, 49(10): 2468-2503
- [4] Seberry J, Finlayson K, Adams S S, et al. Orthogonal Designs with Maximal Rates[J]. *IEEE Transactions on Signal Processing*, 2008, 56(1): 256-265
- [5] 《现代应用数学手册》编委会. 现代应用数学手册—概率论与随机过程卷[M]. 北京: 清华大学出版社, 2000
- [6] 杨大利, 徐明星, 吴文虎. 语音识别特征参数选择方法研究[J]. *计算机研究与发展*, 2003, 40(7): 963-969
- [7] 陈魁. 实验设计与分析[M]. 北京: 清华大学出版社, 1996
- [8] Abu-Shikhah N, Deriche M. A Robust Technique for Harmonic Analysis of Speech[C]// *2001 IEEE International on Acoustics, Speech and Signal Processing, ICASSP '01*. 2001: 877-880
- [9] Virtanen T, Klapuri A. Separation of Harmonic Sounds Using Linear Models for the Overtone Series[C]// *2002 IEEE International on Acoustics, Speech and Signal Processing, ICASSP'02*. 2002: 1757-1760