

# 心电节拍自动分类算法的研究

张如意<sup>1</sup> 廖京生<sup>1</sup> 李抱朴<sup>1</sup> 胡 超<sup>1, 2</sup>

<sup>1</sup> (中国科学院深圳先进技术研究院 深圳 518055)

<sup>2</sup> (浙江大学宁波理工学院信息学院 宁波 315100)

**摘 要** 世界卫生组织调查发现在全球范围内心血管、心脏疾病是导致死亡概率最高的疾病,心电图(ECG)是临床上广泛应用的预防、监护和诊断心血管及心脏疾病的重要工具之一。心电自动分析诊断技术可以大大减少心电医师的工作量,提高心电图的诊断效率,其中心电节拍(ECG Beat)分类是心电自动分析诊断技术的主要研究方向,是自动分析心律失常的一种重要分析手段,特别是在动态心电图或者长期心电记录领域发挥着重要的作用。本文提出一种心电节拍分类算法,该算法在聚类分析的基础上,结合线性分类器加权判断和心电医师对各聚类的抽样判断,获得心电节拍的最终分类。以MIT-BIH-AR<sup>[1]</sup>心律失常数据库作为原始数据,采用AAMI的ANSI/AAMI EC57:1998/(R)2003<sup>[2]</sup>标准规定的心电节拍分类种类及准确率的衡量方法,对该算法的检验,发现采用聚类分析和线性分类器加权判断的方法,分类的准确率达到86.60%;结合心电医师的抽样判断后,算法最终的准确率高达98.16%。

**关键词** 心电节拍分类;混合分类框架;聚类分析;辅助诊断

## Research on Automatic ECG Heartbeat Classification

ZHANG Ru-yi<sup>1</sup> LIAO Jing-sheng<sup>1</sup> LI Bao-pu<sup>1</sup> HU Chao<sup>1,2</sup>

<sup>1</sup>(Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China)

<sup>2</sup>(Ningbo Institute of Technology, Zhejiang University, Ningbo 315100, China)

**Abstract** The World Health Organization found that cardiovascular and heart disease causes the highest probability of death in the world. Electrocardiogram (ECG) is an important tool widely used in clinical prevention and diagnosis of cardiovascular and heart disease. Automatic analysis of ECG diagnostic technique can greatly reduce the workload of the cardiologists and improve the efficiency of diagnosis. The classification of ECG Heartbeat is the mainly research direction of ECG automatic analysis as automatic ECG heartbeat classification can improve the diagnostic quality of arrhythmia, especially in the area of dynamic electrocardiogram or the long-term ECG recording. This paper presents an ECG beat classification algorithm, The algorithm uses clustering analysis, mixed with linear classifiers, weighted judgment and physician-assisted classification. Using MIT-BIH-AR arrhythmia database as the raw data and ANSI / AAMI EC57: 1998 / (R) 2003 of AAMI as the standard of classification, the experiment results show that only using clustering analysis, mixed with linear classifiers and weighted judgment, the accuracy rate is 86.60%. After introducing cardiologists-assisted classification, the final accuracy rate is 98.16%.

**Keywords** heartbeat classification; hybrid classification framework; clustering analysis; assisted-diagnosis

## 1 研究背景

心血管疾病是危害人们健康的最常见的疾病之

一,世界卫生组织调查发现心血管疾病是排在全球范围内死亡率最高的疾病,并预测这种情况一直会持续到2030年。在我国,城镇居民因心血管疾病而导致的死亡率仅排在恶性肿瘤和脑血管疾病之后,是死亡率

**作者简介:** 张如意, 硕士研究生, 研究方向为心电智能诊断系统; 廖京生, 副研究员, 硕士生导师, 研究方向为无线医疗监护网络、低成本监护仪、智能监护系统; 李抱朴, 副研究员, 博士生导师, 研究方向为生物医学工程、图像处理 and 模式识别; 胡超, 博士生导师, 三江学者特聘教授, 研究方向为生物医学电子、传感器信息融合、机器人控制、图象处理和计算机视觉、光学与颜色测量、微计算机测控系统。

排名第三的疾病。心电图(ECG)是现代医学用以诊断心脏疾病的重要工具之一,它对心血管疾病的预防和诊断具有重要意义。由于ECG诊断可靠、方法简便、对病人无损害,因此在临床上广泛应用。

自1903年著名的心电图之父William Einthoven根据前人的许多研究成果,研制成了第一部用于监护心脏电位的心电图机,至今已有近百年的历史。体检测量的普通心电图只能观察某一段短时间的心电活动,医学诊断价值有限。长时间的心电监测可以连续实时观察并分析心脏的活动情况,在临床上对于心血管疾病的监护具有非常重大的意义。早期的心电监护系统要护理人员长时间地观察,往往会因为视力疲劳或注意力分散造成漏检。五十年代末以来,随着计算机技术的发展,人们开始了对心电自动分析技术的研究。许多学者进行了大量相关探讨,从噪声干扰的抑制,到特征点的识别、特征参数检测、波形的聚类分析、波形类别判断,以致到最终诊断方法的研究。对大量临床采集的心电信号进行迅速、准确、充分、全面的自动分析,从而使护理人员从繁重的数据处理中解脱出来,是研究人员开发心电自动分析仪的终极目标。心电节拍(ECG Beat)分类是心电自动分析诊断技术的主要研究方向,是自动分析心律失常的一种重要分析手段,特别是在动态心电图或者长期心电记录领域。

心电节拍(ECG Beat)分类算法的研究从工作模式上主要分为两类:自动心电分类和医生辅助诊断分类。自动心电分类的基本思路是基于MIT-BIH-AR数据库的部分数据及相应的人工标注,通过提取心电信号的特定特征集,对分类器进行训练,获得分类器的结构,并用剩余部分的MIT-BIH-AR数据库数据对获得的分类器进行检验。相应研究包括:Stanislaw Osowski<sup>[3]</sup>等人以QRS波的多阶累积量作为分类特征,采用混合模糊神经网络为全局分类器,将MIT-BIH-AR数据库的心电节拍分为正常节拍N,左束支传导阻滞L,右束支传导阻滞R,房性早搏A,室性早搏V,交界性早搏I,室性逸搏E等七类,与MIT-BIH-AR数据库的心电节拍注释作比较,其全局分类器的准确率达到96.06%。de Chazal P<sup>[4]</sup>等将MIT-BIH-AR分为DS1, DS2两个数据集,用DS1数据集来训练分类器,使用RR间期及归一化后的QRS波的抽样点作为特征,将心电数据按照AAMI<sup>[2]</sup>标准进行分类,然后用DS2数据集来测试分类器,最终得到的自动全局分类器的平均准确率为84.5%。类似的自动全局分类

的方法还有Sung-Nien Yu<sup>[5]</sup>。de Chazal P<sup>[4]</sup>和Hu Y H<sup>[5]</sup>等人都指出这类全局分类算法的缺点:由于不同病人的心电节拍形态上差异较大,此类全局分类器对训练过的心电数据准确率很高,但是对于未参与训练的心电数据,就很难做出准确的分类。针对这一问题, Hu Y H等<sup>[5]</sup>引入利用专家混合模型,他们将心电节拍重抽样作为特征,以自组网(SOM),学习矢量量化(LVQ)作为分类器,将MIT-BIH-AR心电数据库按照AAMI标准,分为N, V, F, Q四类。该方法由一个心电医师分析并标注每一个独立的病人五分钟的心电节拍(ECG Beat)用以进行病人个性化的分类器训练,最终得到的混合模型的平均准确率达到94%。类似的心电医生辅助诊断分类的方法还有Serkan Kiranyaz<sup>[6]</sup>等人提出对MIT-BIH Long-Term database心电数据库中的每个病人的Record进行K-means聚类,心电医师查看当前Record每个聚类种群中聚类中心的心电节拍(ECG Beat),给出其对应的AAMI标准的心电节拍(ECG Beat)类型,然后将聚类中心作为模板,对该病人的整个Record进行模板匹配,从而得到整个Record的分类结果,最终计算得到整个MIT-BIH Long-Term database心电数据库的平均分类准确率为99.04%,是上述所有心电分类算法中平均准确率最高的,但是S类心电节拍(ECG Beat)的真阳性率仅为16.61%。需要引入医生辅助诊断分类的依据是,心电数据的个体内差异性较小,因此非监督学习可以非常有效的对个体心电节拍进行聚类,但是由于心电数据的个体间差异性较大,因此需要心电医师对于聚类后心电节拍给予相应的诊断判别,从而得到更为准确的心电节拍分类结果。

## 2 方 法

### 2.1 心电数据

研究心电的自动分析诊断技术,需要有大量的经过心电医师注释过的心电数据,MIT-BIH-AR<sup>[1]</sup>是研究心律失常分类公认的权威心电数据库,由48组两导联心电记录组成,其中,每个心电记录(Record)长度为30分钟,采样率360 Hz,由头文件(扩展名为.heg)、数据文件(扩展名为.dat)、注释文件(扩展名为.atr)等三个文件组成。头文件详细说明了与它关联的数据文件的名称及其属性,存储方式为ASCII码字符;数据文件是以自定义的格式按二进制存储的信号原始数据,该文件所用的数据格式在头文件中已经进行了说明;注释文件是记录心电诊断专家对信号分

析的结果, 主要包括心电节拍的位置, 节律类型和信号质量等。

## 2.2 心电节拍分类种类及准确率评定标准

AAMI (The Association for the Advancement of Medical Instrumentation, 美国医疗仪器促进协会), 制定的ANSI/AAMI EC57: 1998/(R)2003, “Testing and reporting performance results of cardiac rhythm and ST-segment measurement algorithms”<sup>[2]</sup>中规定心电节拍可以分为五类: N(正常或者束支传导阻滞节拍), S(室上性异常节拍), V(心室异常节拍), F(融合节拍), Q(未能分类的节拍)。以及无法识读的U类, 辅助无法识读标志的X类和O类。具体的分类标准可以参考文献[2]。AAMI还规定了算法准确率的计算方法, 即混淆矩阵, 如图1所示。需要注意的是, AAMI提出F, Q类分类的准确率仅供参考, 而U, X, O类不是真正的心电节拍, 因此实际上我们需要分成N, S, V三类。

		Algorithm label							
		N	S	v	f	q	o	x	
Reference Label	N	Nn	Ns	Nv	Nf	Nq	No	Nx	
	S	Sn	Ss	Sv	Sf	Sq	So	Sx	
	V	Vn	Vs	Vv	Vf	Vq	Vo	Vx	
	F	Fn	Fs	Fv	Ff	Fq	Fo	Fx	
	Q	Qn	Qs	Qv	Qf	Qq	Qo	Qx	
	O	On	Os	Ov	Of	Oq			
	X	Xn	Xs	Xv	Xf	Xq			

图1 混淆矩阵

图1中, Reference Label是心电医师给出的心电节拍类型, 我们使用MIT-BIH-AR数据库中的注释文件中的心电节拍类型作为Reference Label, Algorithm Label是待测试的算法得到的心电节拍类型。由于AAMI规定的心电种类和MIT-BIH-AR心电数据库注释的心电节拍的类型不一致, 因此需要对MIT-BIH-AR数据库注释的心电节拍类型进行转换, 具体的转换表格参考文献[4]。

## 2.3 算法框架

心电节拍的分类是一个模式识别问题, 主要包括信号预处理、特征提取、特征选择、分类器设计及分类测试五个部分。

### (1) 心电信号的预处理

心电信号的预处理, 是指滤除心电信号中含有的基线漂移, 肌电干扰, 工频干扰等, 由于心电信号的噪声和有用信号在频域上较少重叠, 因此可以采用经典的基于频域的滤波, 设计一个2 Hz~25 Hz的带通滤波器就可以有效的滤除心电干扰噪声, 具体的方法可以参考文献[7]。

### (2) 特征提取

心电信号的特征主要分为时域特征和变换域特征, 其中时域特征主要包括: P波, T波, QRS波的幅值及宽度、各个间期及心电节拍的统计学信息和抽样点等特征。变换域的特征主要包括: 小波变换, 频域变换等的统计学信息和抽样点等特征。

时域特征的求解, 应首先求出图2中R点的位置, 即QRS波检测算法, 该算法非常成熟, 准确率达到99.3%以上, 我们采用[7]使用的方法来求解R点位置。

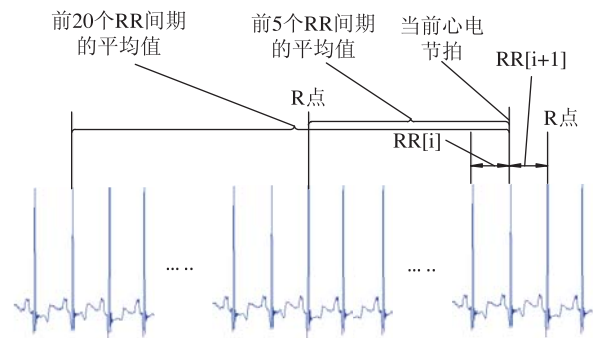


图2 心电的RR间期特征

考虑到特征的鲁棒性, 我们选择以下心电特征:

RR间期的时间长度包括 $RR[i]$ ,  $RR[i-1]$ ,  $RR[i+1]$ ,

$$RR[i]=R[i]-R[i-1] \quad (1)$$

前RR间期的变化量 $RR_{diff}[i]$ 。

$$RR_{diff}[i]=RR[i]-RR[i-1] \quad (2)$$

前num个RR间期的平均值 $RR_{mean\_num}[i]$ 。

$$RR_{mean\_num}[i]=\frac{1}{num} \sum_{j=0}^{num-1} |RR[i-j]| \quad (3)$$

截取滤波后的心电信号R点前50 ms (R-50 ms) 至R点后100 ms (R+100 ms) 作为表征QRS的数据段, 记为固定时间段T1, 截取R点后150 ms (R+150 ms) 至R点后500 ms (R+500 ms), 记为T2, 分别求T1, T2段心电信号的自相关函数 $R_{xx}$ 的均值及方差, 记为Cor\_ECG。

$$R_{xx}(l)=\sum_{n=i}^{N-|k|-l} x(n)x(n-l) \quad (4)$$

对T1段心电数据以60 Hz的采样率, 抽样9个点, 对T2段心电数据以20 Hz的采样率, 抽取7个点, 记为ECG\_resam。

$$x'(n)=x(Mn) \quad (5)$$

心电变换域的特征, 我们选择小波变换, 使用Harr小波对滤波后的ECG信号采用a'trous算法<sup>[8]</sup>进行五层小波分解, 选用a'trous算法的主要原因是我们要利用小波分解后分量的时间信息, a'trous算法分解得到的分量和原始数据时间上是对齐的。我们选择小波

分解后的第三层分量WT3和第四层分量WT4, 截取WT3和WT4 R点前80 ms (R-80 ms) 至R点后120 ms (R+120 ms) 时间段内的数据, 计算其期望和方差, 记为WT\_ECG, 并计算数据的自相关函数的期望和方差, 记为Cor\_WT\_ECG。

(3) 特征选择

为了节约资源, 降低分类算法的空间复杂度和时间复杂度, 我们使用主成分分析 (PCA) 来进行特征选择, PCA的主要原理可以参考文献[9]。

(4) 分类器的设计

考虑到心电节拍个体间差异较大, 个体内差异系较小的特点, 我们提出的心电节拍分类算法有两种工作模式, 自动分类, 心电医师辅助分类。其中使用到的分类器有LDC分类器、EM聚类、K-means聚类和模板匹配等四种, 它们分别所使用的心电特征如表1所示。

表1 各分类器所使用的特征

LDC	EM 聚类	K-means	模板匹配
ln(RR[i])	ln(RR[i])	ln(RR[i])	ECG_
ln(RR[i+1])	ln(RR[i-1])	ln(RR[i+1])	Resam
ln(RR[i-1])	RR_diff[i+1]	RR_mean_5	PCA
RR_mean_5	RR_diff[i]	RR_mean_20	
RR_mean_20	WT_ECG	Cor_ECG	
Cor_ECG	Cor_WT_ECG		

自动分类模式采用聚类分析和LDC分类器结合的方法来实现心电节拍的分类。为了简化模型的复杂度, LDC分类器采用简单的正态概率模型下的贝叶斯判别函数, 并假设他们的先验概率相等, 协方差矩阵

相同, 则LDC分类器的判别函数简化为:

$$g_i(x) = \mu_i^T \sum^{-1} x - \frac{1}{2} \mu_i^T \sum^{-1} \mu_i + \ln(P(w_i)) \quad (6)$$

自动分类模式具体的系统框图如图3所示, 其中, LDC分类器将心电节拍分为N, S, V, 三类。EM聚类的分类数为15, 判定是否将当前聚类种群全部分为LDC分类器最多的Labels的判别门限为70%。

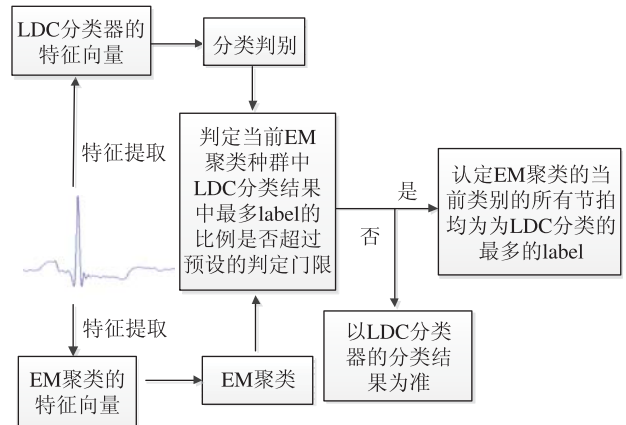


图3 自动分类模式的工作框图

辅助分类模式的系统框图如图4所示, 采用EM聚类, K-means聚类和模板匹配, 结合心电医师的辅助诊断。其中, EM聚类分为15类, 抽出距离EM聚类中心最近的5个心电节拍、最远的5个心电节拍, 这里的距离是根据EM聚类的原理, 即概率的大小来衡量。心电医师查看EM聚类后的当前种群的聚类中心, 和抽出的当前种群中的代表性心电节拍, 并给出相应的正确的分类标识。如果心电医师查看的当前种群的代表性心电节拍属于同一类型, 则判定当前聚类种群全部为该类型, 如果不属于同一类型, 则对当前种群进行K-means聚类, 类别数为4, 得到四个K-means聚类中心, 心电医师给出这四个聚类中心的节拍类型, 然

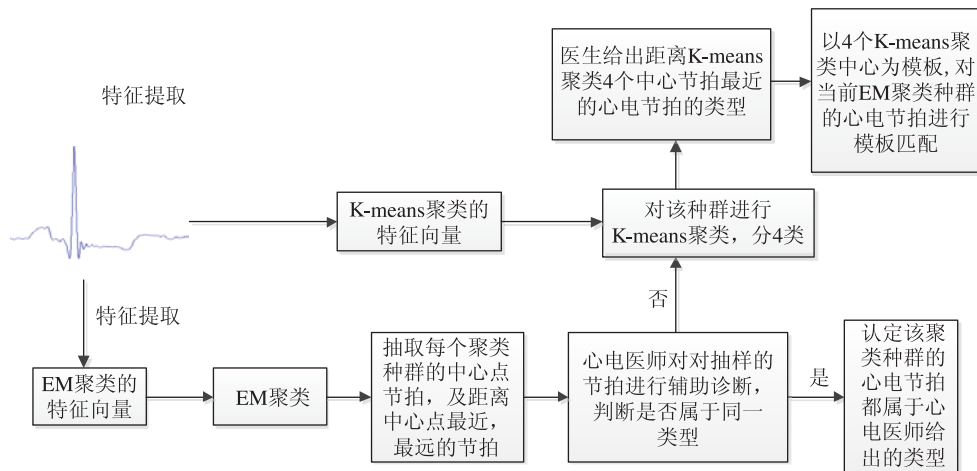


图4 辅助分类模式的系统框图

后以这四个聚类中心节拍为模板,进行模板匹配。在进行模板匹配时,先对ECG\_resample进行PCA,保留90%的分量,作为特征向量,当前EM聚类种群的心电节拍类型即为匹配的模板的节拍类型。心电医师的辅助诊断,判别心电节拍类型,是通过MIT-BIH-AR的注释文件中给出的心电节拍的类型仿真得到的。

### 3 实验结果

为了方便实验结果的对比,我们采用de Chazal P<sup>[4]</sup>等人提出的方法,将MIT-BIH-AR心电数据库中的Records分为DS1和DS2两部分,其中DS1中的Records作为训练,DS2中的Records作为测试,具体的划分细节见参考文献[4]。在自动分类模式下,得到的最终分类结果的混淆矩阵如表2所示,其平均准确率为86.6%。

表2 自动分类的结果

		Algorithm label		
		N	S	V
Reference Label	N	39479	1848	2931
	S	312	1046	479
	V	174	523	2524
	Other	292	2	101

在医生辅助诊断的辅助分类模式下,得到的最终的分类结果的混淆矩阵如表3所示。其平均准确率为98.16%。

表3 辅助分类结果

		Algorithm label			
		N	S	V	Other
Reference Label	N	44139	18	100	1
	S	316	1297	224	0
	V	584	509	2060	68
	Other	94	2	19	280

### 4 结论

本文提出了一种心电节拍分类算法,有两种工作模式,自动分类模式和医生辅助诊断模式。自动分类模式通过结合线性分类器和EM聚类的方法,将MIT-BIH-AR数据库中的心电节拍按照AAMI标准分为N, S, V三类,与de Chazal P<sup>[4]</sup>等人的心电自动分类方法的进行比较,在没有心电医师进行辅助诊断治疗的前

提下,他们得到的平均准确率为85.88%,而我们平均准确率为86.60%,主要是因为我们引入了聚类分析,充分利用心电个体内差异性较小的特性。

医生辅助诊断模式通过引入医生的辅助诊断将心电节拍分为N, S, V, Other四类,医生的辅助诊断是通过使用心电注释文件的心电节拍类型仿真得到。与同为辅助诊断模式的Kiranyaz S<sup>[6]</sup>等人的方法进行比较,他们使用MIT-BIH Long-Term作为原始心电数据库,得到的平均准确率为99.04%,但是S类的灵敏度只有40.15%,正检测率为16.61%。我们的平均准确率为98.16%,S类的灵敏度为70.6%,正检测率为71.03%。S. Kiranyaz的平均准确率比较高,是因为MIT-BIH Long-Term心电数据库中N和V类心电节拍所占比例为99.36%,而他们的算法对N类和V类的分类非常准确,我们使用的MIT-BIH-AR心电数据库中N类, S类, V类心电节拍所占比例分别为82.77%, 2.54%, 6.61%。所以我们的分类器算法在灵敏度及正检测率上优于S. Kiranyaz<sup>[6]</sup>等人的方法,主要原因是我们采用了不同的心电特征,进行不同类型的聚类分析,及模板匹配,使得算法普适性更好。

进一步工作的内容,主要是使用<sup>[1]</sup>中其他类型的心电数据库,来测试我们的算法,从而进一步改进算法,提高算法的普适性和鲁棒性,尤其是不同分类器使用哪些心电特征,使得整体性能最优,以及最优化聚类分析的分类数目。

### 参考文献

- [1] PhysioBank Archive Index [EB/OL]. <http://physionet.org/physiobank/database/>.
- [2] Association for the Advancement of Medical Instrumentation. Testing and Reporting Performance Results of Cardiac Rhythm and ST-Segment Measurement Algorithms [M]. American National Standard: USA, 1999.
- [3] Osowski S, Linh T H. ECG beat recognition using fuzzy hybrid neural network [J]. IEEE Transactions on Biomedical Engineering, 2001, 48: 1265-1271.
- [4] de Chazal P, O'Dwyer M, Reilly R B. Automatic classification of heartbeats using ECG morphology and heartbeat interval features [J]. IEEE Transactions on Biomedical Engineering, 2004, 51: 1196-1206.
- [5] Hu Y H, Palreddy S, Tompkins W J. A patient-adaptable ECG beat classifier using a mixture of experts approach [J]. IEEE Transactions on Biomedical Engineering, 1997, 44: 891-900.
- [6] Kiranyaz S, Ince T, Pulkkinen J, et al. Personalized long-term

- 
- ECG classification: A systematic approach [J]. *Expert Systems with Applications*, 2011, 38: 3220-3226.
- [7] Pan J, Tompkins W J. A real-time QRS detection algorithm [J]. *IEEE Transactions on Biomedical Engineerin*, 1985: 230-236.
- [8] Shensa M J. The Discrete Wavelet Transform-Wedding the a Trous and Mallat Algorithms [J]. *IEEE Transactions on Signal Processing*, 1992, 40: 2464-2482.
- [9] Zhang H, Zhang L Q. ECG analysis based on PCA and support vector machines [C] // *Proceedings of the 2005 International Conference on Neural Networks and Brain*, 2005, Vols 1-3: 743-747.