

基于非线性预测效果的癫痫脑电信号的特征提取方法^{*}

孟庆芳^{1,2,†} 周卫东²⁾ 陈月辉¹⁾ 彭玉华²⁾

1) 济南大学信息科学与工程学院, 济南 250022)

2) 山东大学信息科学与工程学院, 济南 250100)

(2009 年 2 月 25 日收到, 2009 年 5 月 5 日收到修改稿)

在非线性的时间序列预测研究的基础上, 提出了基于非线性预测效果的癫痫脑电信号特征提取方法, 从脑电信号中自动检测出癫痫脑电信号. 采用基于可预测性的选取嵌入维数的方法确定脑电信号序列的嵌入维数, 进行相空间重构. 实验结果表明, 基于非线性预测效果的特征提取方法提取的特征能明显地区分癫痫脑电信号与正常脑电信号, 该非线性特征提取方法适合小数据量的情况且对噪声的稳定性好.

关键词: 脑电信号, 癫痫, 非线性预测, 特征提取

PACC: 0545

1. 引言

脑电信号中包含了大量的生理与病理信息, 在临床医学与脑科学的研究中起着非常重要的作用, 是进行神经系统疾病和症状特别是癫痫病诊断的主要依据. 癫痫是一种危害健康的常见慢性脑部疾病, 其发作时表现为大脑神经元过度同步化放电, 并导致短暂性中枢神经系统功能失常. 临床一般通过脑电图(electroencephalograph, EEG)检查诊断癫痫. 脑电图分析主要是对异常活动的检测分析, 该工作目前主要由医疗工作者根据脑电图通过视觉检测来完成. 由于视觉检测费时、效率低, 因此癫痫脑电信号的自动检测在临床上有很重要的意义.

从 20 世纪 80 年代中期开始, 人们把非线性科学特别是非线性动力学的理论与方法应用于脑电信号的研究中^[1], 给脑电研究带来了新的希望. 越来越多的证据表明大脑是一个复杂的非线性动力学系统, 是由大量的非线性神经元通过广泛连接所构成的网络系统, 从而造成神经系统的功能表现出极大的复杂性. 因此, 人们已把非线性动力学方法广泛地应用于脑电信号的分析^[1-4]. 文献 [1] 分析了大脑在

听音乐、放松与正常状态下的非线性统计量, 如相关维数、最大 Lyapunov 指数、Hurst 指数与近似熵. 文献 [2] 分析了癫痫发作前后最大 Lyapunov 指数与近似熵. 文献 [3] 应用局部投影非线性降噪方法对脑电信号降噪.

非线性动力学方法提供了处理非线性不规则时间序列新的思路和方法^[5-23]. 从观测时间序列预测系统的时间演化序列是时间序列分析中最重要的经典问题之一. 除了预测本身的重要性, 系统的可预测性还有重要的方法论意义. 系统的可预测性是非线性时间序列分析的一个有力工具^[5]. 预测效果主要在以下三个方面有重要应用. 非线性预测效果通常被作为判断时间序列中是否存在非线性确定性成分的依据^[17-22]. 基于预测效果的方法是选取嵌入维数的一种重要方法^[12-14]. 数据的可预测性还被应用于检测时间序列的平稳性^[23].

在非线性的时间序列预测研究的基础上, 本文提出了基于非线性预测效果的癫痫脑电信号特征提取方法, 从脑电信号中自动检测出癫痫脑电信号.

现有用非线性动力学方法来分析脑电信号的研究, 都是把非线性统计量, 如相关维数、最大 Lyapunov 指数、近似熵, 作为特征^[1, 2]. 而本文首次提

^{*} 国家自然科学基金(批准号: 60573065), 山东省自然科学基金(批准号: Y2007G31, Y2007G33), 山东省优秀中青年科学家科研奖励基金(批准号: BS2009SW003), 山东省“十一五”重点学科研究方向团体建设(批准号: XTD0708)资助的课题.

[†] E-mail: qfmgeng@sohu.com

出把预测误差的归一化方差的对数值作为特征,来分析脑电信号,其中预测误差的归一化方差的对数值表征非线性预测效果与系统的可预测性.本文的研究将有助于对癫痫脑电信号有更进一步的认识.

2. 癫痫脑电信号的特征提取方法

2.1. 脑电信号的相空间重构

从标量时间序列重构与系统原相空间等价的相空间是非线性动力学方法的基础^[5].在 Whitney 拓扑嵌入定理的基础上,Takens^[6]证明了用延迟坐标重构的动力轨迹相空间与原动力系统保持微分同胚,即单变量时间序列在无限长且无噪声的情况下,延迟时间取任意值都能重构原系统相空间.但实际上,实测时间序列是有限长的,且不可避免的被噪声污染,因此延迟时间取任意值不能重构原系统相空间,嵌入定理也没有提供嵌入维数的选取方法.因此实测时间序列相空间重构的关键是其参数的选取.假设实测标量时间序列为 $\{x_i\}, i = 1, 2, \dots, N$, 由延迟坐标相空间重构法可得延迟矢量

$$X(n) = [x_n, x_{n-\tau}, \dots, x_{n-(m-1)\tau}]^T, \\ n = (m-1)\tau + 1, (m-1)\tau + 2, \dots, N, \quad (1)$$

其中 m 为嵌入维数, τ 为延迟时间.

如何从标量时间序列确定嵌入维数的问题已经得到了广泛的研究和讨论^[7-14].目前嵌入维数的选取方法主要有四类.其中基于统计不变量的方法^[7,8]、奇异值分解法^[9]和伪邻近点法^[10,11]存在一些缺点,例如不适合小数据量的情况,对噪声敏感,计算量大,判定准则主观性强等.而基于预测效果的方法^[12-14]有以下优点:适合小数据量的情况;对噪声的稳定性好,计算效率高;不包含任何主观引入的参数.本文采用基于预测效果的选取嵌入维数的方法^[14]来确定实测脑电信号序列的嵌入维数,从而对脑电信号序列相空间重构.

2.2. 癫痫脑电信号的特征提取方法

从观测时间序列预测系统的时间演化序列是时间序列分析的最重要的经典问题之一^[14,15].除了预测本身的重要性,系统的可预测性还有重要的方法论意义.系统的可预测性是非线性动力学方法的一个有力工具^[5].预测效果主要在以下三个方面有重要应用.非线性预测效果通常被作为检测时间序列

中非线性确定性成分的依据,大量的基于非线性预测效果的方法被提出来区别确定性混沌与随机噪声^[17-22].基于预测效果的方法也是选取嵌入维数的一种重要方法^[12-14].数据的可预测性还被应用于检测时间序列的非平稳性^[23].

把时间序列 $\{x_n\}, n = 1, 2, \dots, N$ 看作一个闭环动力结构,即输出 x_n 作为一个延迟输入.在该框架下,用离散 Volterra 自回归模型来得到预测时间序列 $\hat{x}_n$:

$$\hat{x}_n = a_0 + a_1 x_{n-1} + a_2 x_{n-1-\tau} + \dots + a_m x_{n-1-(m-1)\tau} \\ + a_{m+1} x_{n-1}^2 + a_{m+2} x_{n-1} x_{n-1-\tau} + \dots \\ + a_{M-1} x_{n-1-(m-1)\tau}^d = \sum_{k=0}^{M-1} a_k z_k(n), \quad (2)$$

其中基 $\{z_k(n)\}$ 由延迟矢量 $X(n-1)$ 的 d 阶的所有截断的组合项组成,总维数为 $M = C_{m+d}^d = (m+d)!/(m!d!)$.因此,每一个模型由两个参数 m 和 d 决定,它们分别对应于嵌入维数和模型的非线性阶数.系数 a_k 可通过 Gram-Schmidt 过程来递归估计得到.

线性或非线性模型的拟合程度可由预测误差的归一化的方差来表征:

$$\epsilon^2(m, d) = \frac{\sum_{n=M+1}^N [\hat{x}_n - x_n]^2}{\sum_{n=M+1}^N [x_n - \bar{x}]^2}, \quad (3)$$

其中 $x_n (n = 1, 2, \dots, L)$ 为实际值, $\hat{x}_n (n = 1, 2, \dots, L)$ 为由预测模型得到的预测值, $\bar{x} = \frac{1}{N-M} \sum_{n=M+1}^N x_n$.

模型的非线性阶数 d 是使下列 Akaike 信息准则最小化的值:

$$D(d) = \ln \epsilon^2(m, d) + 2r/N, \quad (4)$$

其中 $r \in [1, M]$ 是参数为 $\{m, d\}$ 的 Volterra 多项式的项数.

模型的嵌入维数 m 是使下列 Akaike 信息准则最小化的值:

$$\alpha(m) = \ln \epsilon^2(m, d) + 2m/N. \quad (5)$$

模型参数 m 和 d 确定后,模型即确定.

本文对脑电信号进行非线性预测,提出基于非线性预测效果的癫痫脑电信号特征提取方法,提取的特征值能区分正常脑电信号与癫痫脑电信号.本文提出的癫痫脑电信号特征提取方法的算法步骤为

1) 对一组脑电信号数据,首先把脑电信号数据分成等长的数据段.

2)对每段脑电信号数据进行非线性预测,得到每段脑电信号的预测误差,其中预测模型的参数 m 和 d 取固定值。

3)把每段脑电信号的预测误差的归一化方差的对数值,即 $\ln \epsilon^2(m, d)$,作为提取的特征值,该特征值能表征该段脑电信号是否存在癫痫发作脑电信号。

把提取的每段脑电信号的特征值组成该组脑电信号的特征向量.用该组脑电信号的特征向量训练某种分类器,可实现癫痫脑电信号的自动检测。

3. 实验结果及分析

本文以两组包含癫痫发作的脑电信号时间序列为分析数据,如图 1 所示.其中,对第一组脑电信号序列,癫痫发作从第 3020 个数据点开始,至第 3980 个数据点结束(图 1(a));对第二组脑电信号序列,癫痫发作从第 4379 个数据点开始,至第 5296 个数据点结束(图 1(b)).

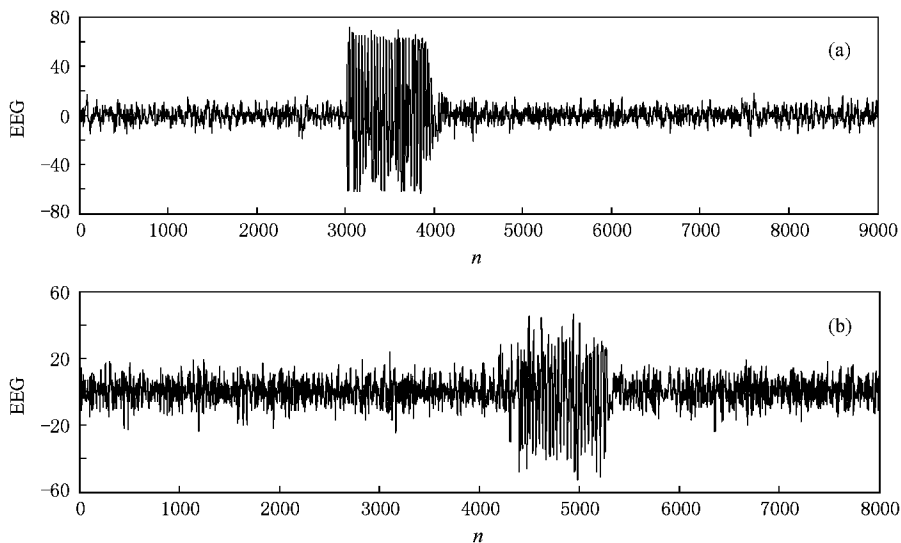


图 1 两组含癫痫发作的脑电信号

相空间重构是非线性动力学方法的基础与首要步骤.首先采用基于可预测性的选取嵌入维数的方法^[14]来分别选取正常脑电信号与癫痫发作脑电信号的嵌入维数.模型非线性阶数 d 取为 3,实验结果如图 2 所示。

从图 2(a)可见,对第一组正常脑电信号,当横坐标 $m=4$ 时,函数取到最小值,因此第一组正常脑电信号的最小嵌入维数应取为 4.从图 2(b)可知,第一组癫痫脑电信号的最小嵌入维数应取为 7.因此对第一组脑电信号提取特征时,嵌入维数统一取为 7,延迟时间取为 1。

图 2(c)所示,第二组正常脑电信号的最小嵌入维数应取为 4.图 2(d)所示,第二组癫痫脑电信号的最小嵌入维数应取为 7.因此对第二组脑电信号提取特征时,嵌入维数统一取为 7,延迟时间取为 1。

应用本文提出的基于非线性预测效果的特征提取方法对这两组脑电信号进行特征提取.对分析数据分段,分别计算每段数据的预测误差,把预测误差

的归一化方差的对数值 $\ln \epsilon^2(m, d)$ 作为提取的特征值.其中模型非线性阶数 d 取为 3,嵌入维数统一取为 7。

为了研究本文提出的基于非线性预测效果的特征提取方法是否适合小数据量的情况,是否对分段数据长度敏感,分段数据长度分别取 250, 500, 1000, 实验结果如图 3 和图 4 所示,其中横坐标为分段数据的段序号,纵坐标为提取的特征值 $\ln \epsilon^2(m, d)$ 。

从图 3(a)可见,对第一组脑电信号,当分段数据长度为 250 时,第 13, 14, 15, 16 段数据(对应第 3001 点数据至第 4000 点数据,即癫痫发作时的数据段)提取的特征值明显小于其他数据段提取的特征值.从图 3(b)可见,对第一组脑电信号,当分段数据长度为 500 时,第 7, 8 段数据(对应第 3001 点数据至第 4000 点数据,即癫痫发作时的数据段)提取的特征值明显小于其他数据段提取的特征值.从图 3(c)可见,对第一组脑电信号,当分段数据长度为 1000 时,第 4 段数据(对应第 3001 点数据至第 4000

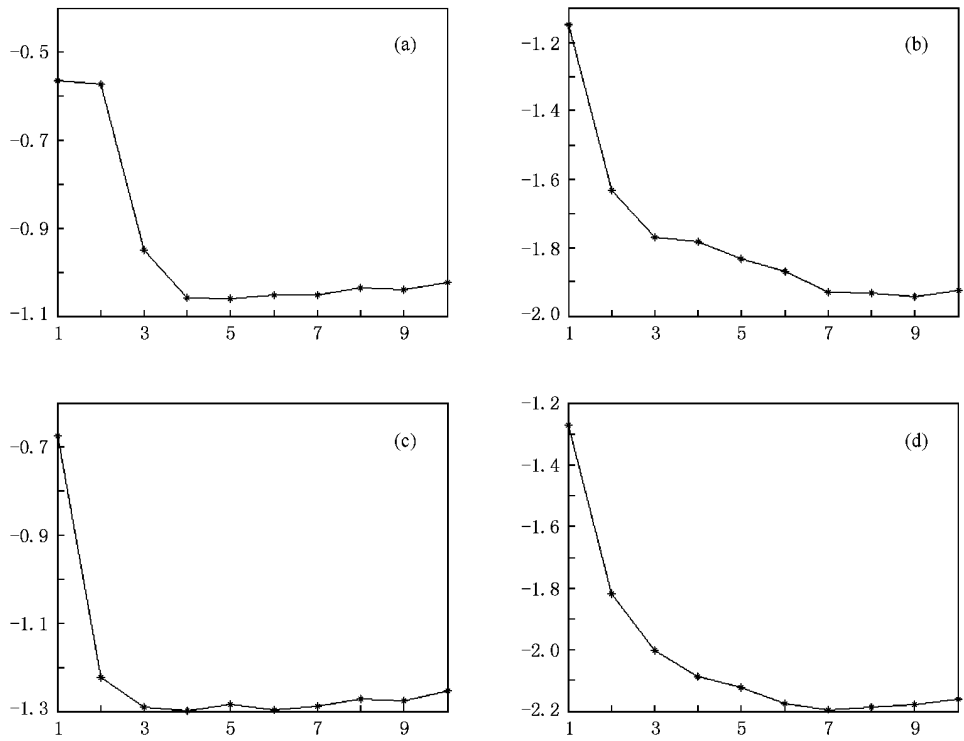


图 2 脑电信号的嵌入维数的选取结果 ($\alpha(m)-m$) (a) 第一组正常脑电信号 (b) 第一组癫痫发作脑电信号 (c) 第二组正常脑电信号 (d) 第二组癫痫发作脑电信号

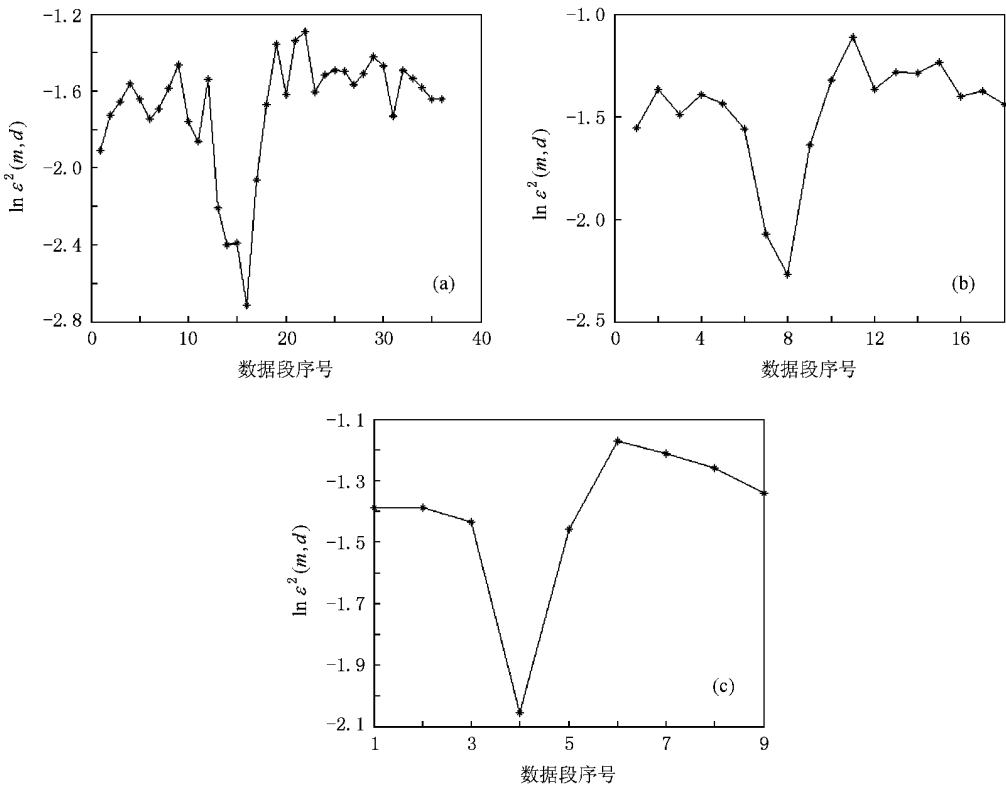


图 3 第一组脑电信号不同分段数据长度的特征提取的结果 ($\ln \varepsilon^2(m,d)$ 数据段序号) (a) 分段数据长度为 250 时 (b) 分段数据长度为 500 时 (c) 分段数据长度为 1000 时

点数据 ,即癫痫发作时的数据段)提取的特征值明显小于其他数据段提取的特征值 .表明 ,当分段数据的长度分别为 250 ,500 与 1000 时 ,癫痫发作时提取的特征值都明显小于正常脑电信号提取的特征值 .

对第二组脑电信号序列 ,从图 4(a)可见 ,当分段数据长度为 250 时 ,第 18 ,19 ,20 ,21 段数据(对应第 4251 点数据至第 5250 点数据 ,即癫痫发作时的数据段)提取的特征值明显小于其他数据段提取的特征值 ;从图 4(b)可见 ,当分段数据长度为 500 时 ,

第 9 ,10 ,11 段数据(对应第 4001 点数据至第 5500 点数据 ,即癫痫发作时的数据段)提取的特征值明显小于其他数据段提取的特征值 ;由图 4(c)可知 ,当分段数据长度为 1000 时 ,第 5 ,6 段数据(对应第 4001 点数据至第 6000 点数据 ,即癫痫发作时的数据段)提取的特征值明显小于其他数据段提取的特征值 .可见 ,当分段数据的长度分别为 250 ,500 与 1000 时 ,癫痫发作时提取的特征值都明显小于正常脑电信号提取的特征值 .

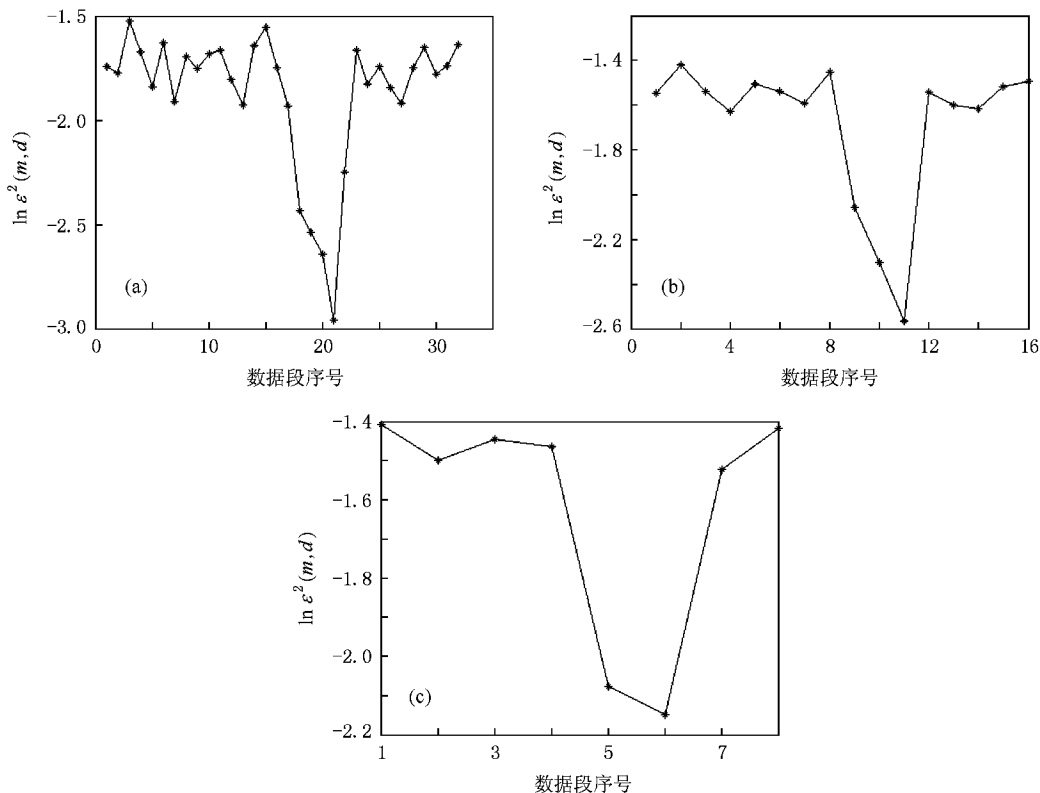


图 4 第二组脑电信号不同分段数据长度的特征提取的结果($\ln \varepsilon^2(m, d)$ -数据段序号) (a)分段数据长度为 250 时 (b)分段数据长度为 500 时 (c)分段数据长度为 1000 时

图 3 和图 4 的结果表明 ,基于非线性预测效果方法提取的特征值能明显地区分癫痫脑电信号与正常脑电信号 ,并且该非线性特征提取方法适合小数据量的情况 ,且对分段数据长度的稳定性好 .

为了研究基于非线性预测效果的特征提取方法对噪声的稳定性 ,把这两组脑电信号序列分别叠加高斯白噪声 ,信噪比为 0 dB ,然后分析叠加高斯白噪声的脑电信号 ,实验结果如图 5 与图 6 所示 ,其中横坐标为分段数据的段序号 ,纵坐标为提取的特征值 $\ln \varepsilon^2(m, d)$.

如图 5 所示 ,对第一组脑电信号序列 ,当信噪比

SNR= 0 dB 且分段数据的长度分别为 250 ,500 与 1000 时 ,癫痫发作时提取的特征值都明显小于正常脑电信号提取的特征值 .

对第二组脑电信号序列 ,如图 6 所示 ,当信噪比 SNR= 0 dB 且分段数据的长度分别为 250 ,500 与 1000 时 ,癫痫发作时提取的特征值都明显小于正常脑电信号提取的特征值 .

图 5 和图 6 的结果表明 ,基于非线性预测效果的特征提取方法对噪声的稳定性好 ,对低信噪比的脑电信号仍能明显地区分癫痫脑电信号与正常脑电信号 .

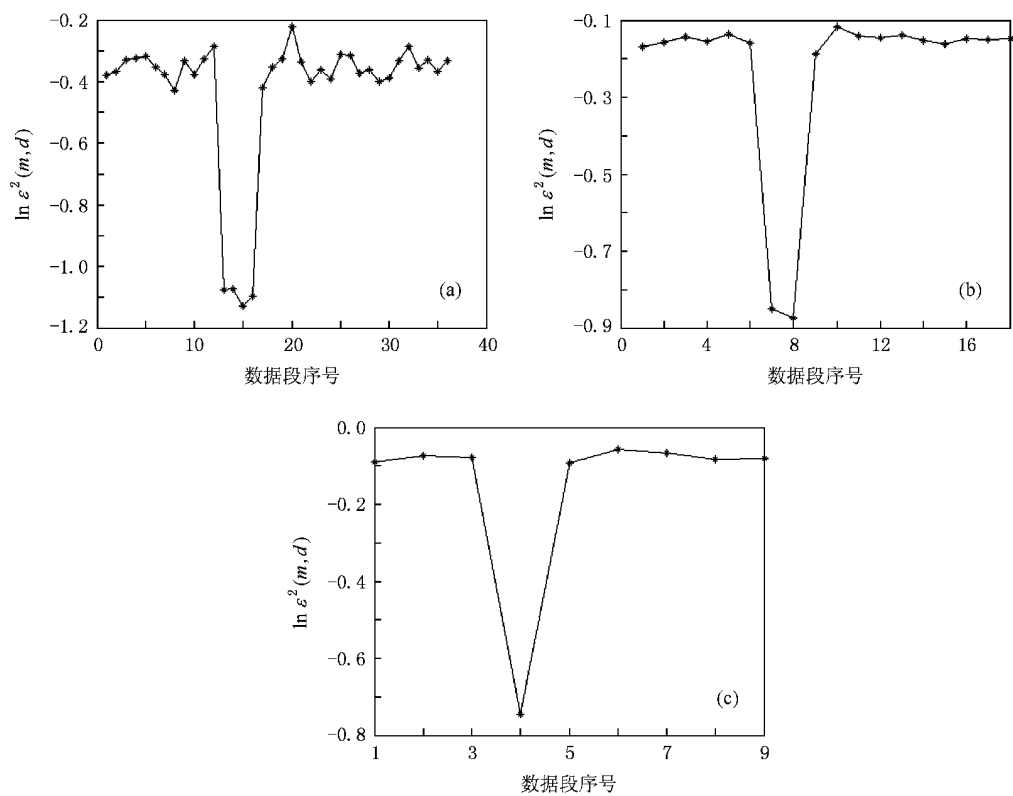


图 5 第一组脑电信号信噪比为 0 dB 时特征提取的结果($\ln \varepsilon^2(m, d)$ -数据段序号) (a)分段数据长度为 250 时; (b)分段数据长度为 500 时 (c)分段数据长度为 1000 时

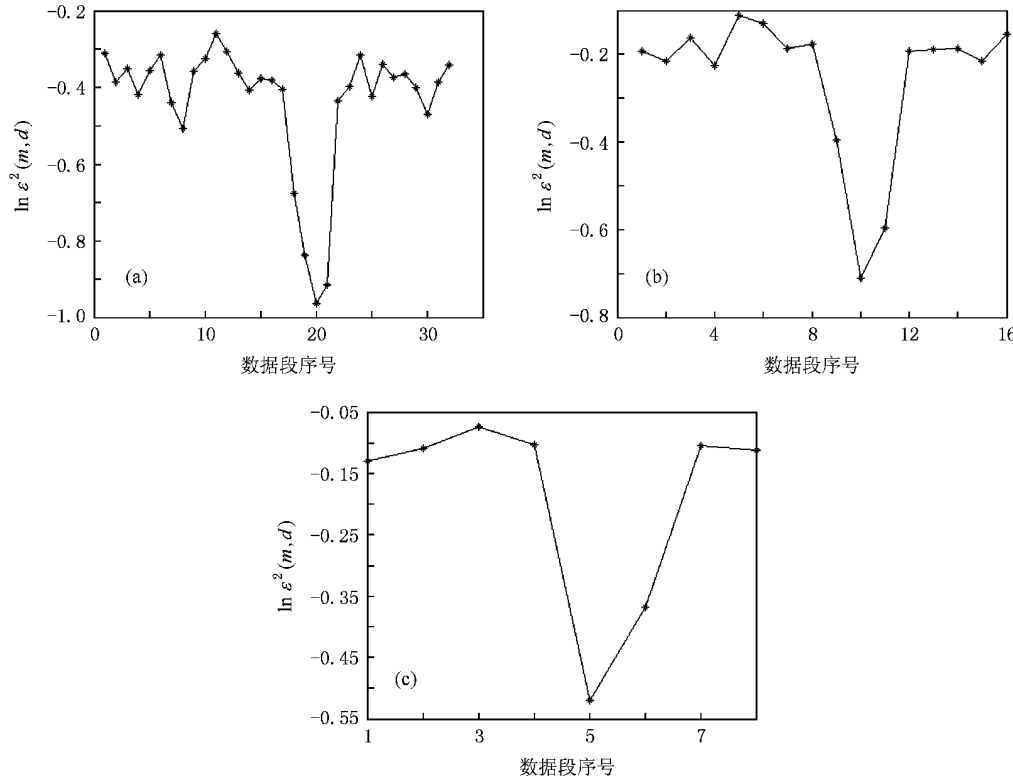


图 6 第二组脑电信号信噪比为 0 dB 时特征提取的结果($\ln \varepsilon^2(m, d)$ -数据段序号) (a)分段数据长度为 250 时; (b)分段数据长度为 500 时 (c)分段数据长度为 1000 时

4. 结 论

本文提出了基于非线性预测效果的癫痫脑电信号特征提取方法,从脑电信号中自动检测癫痫发作信号.本文的研究表明非线性动力学方法的成功应用将使研究者们对癫痫脑电数据的非线性特性有更进一步的认识.文中采用基于可预测性的选取嵌入

维数的方法来确定脑电信号的嵌入维数,进行相空间重构,然后通过非线性预测模型提取脑电信号的特征.实验结果表明:基于非线性预测效果的特征提取方法提取的特征值能明显地区分癫痫脑电信号与正常脑电信号,并且该特征提取方法适合小数据量的情况,对分段数据长度与噪声的稳定性好,对低信噪比的脑电信号仍能明显区分癫痫脑电信号与正常脑电信号.

[1] Natarajan K ,Acharya R ,Alias F ,Tiboleng T ,Puthusserypady S K 2004 *BioMedi . Engin . OnLine* **3** 1

[2] Xie Y ,Xu J X ,Yang H J ,Hu S J 2002 *Acta Phys . Sin .* **51** 205 (in Chinese) [谢 勇、徐健学、杨红军、胡三觉 2002 物理学报 **51** 205]

[3] Xie Y ,Xu J X ,Kang Y M ,Yang H J ,Hu S J 2003 *Acta Phys . Sin .* **52** 1121 (in Chinese) [谢 勇、徐健学、康艳梅、杨红奎、胡三觉 2003 物理学报 **52** 1121]

[4] Li X L ,Ouyan G X ,Guan X P ,Li Yan 2006 *Chin . Jour . of Biom . Eng .* **25** 346 (in Chinese) [李小俚、欧阳高翔、关新平、李 岩 2006 中国生物医学工程学报 **25** 346]

[5] Kantz H ,Schreiber T 2003 *Nonlinear Time Series Analysis* (Second Edition) (Cambridge :Cambridge University Press)

[6] Takens F 1981 *Dynamical Systems and Turbulence* (Berlin : Springer) p366

[7] Xiao F H ,Yan G R , HanY H 2005 *Acta Phys . Sin .* **54** 550 (in Chinese) [肖方红、阎桂荣、韩宇航 2005 物理学报 **54** 550]

[8] Grassberger P ,Procaccia I 1983 *Phys . Rev . Lett .* **50** 346

[9] Broomhead D S ,King G P 1986 *Physica D* **20** 217

[10] Kennel M ,Brown R ,Abarbanel H D I 1992 *Phys . Rev . A* **45** 3403

[11] Cao L 1997 *Physica D* **110** 43

[12] Casdagli M 1989 *Physica D* **35** 335

[13] Sugihara G ,May R M 1990 *Nature* **344** 734

[14] Meng Q F ,Peng Y H ,Xue P J 2007 *Chin . Phys .* **16** 1252

[15] Meng Q ,Peng Y 2007 *Physics Letters A* **370** 465

[16] Meng Q F ,Zhang Q ,Mu W Y 2006 *Acta Phys . Sin .* **55** 1666 (in Chinese) [孟庆芳、张 强、牟文英 2006 物理学报 **55** 1666]

[17] Tsonis A A ,Elsner J B 1992 *Nature* **358** 217

[18] Rubin D M 1992 *Chaos* **2** 525

[19] Koh T ,Powers E J 1985 *IEEE Trans . Aco . Spe . Sign . Proc .* **33** 1445

[20] Gautama T ,Mandic D P ,Hulle M M 2004 *IEEE Trans . on Biom . Eng .* **51** 728

[21] Barahona M ,Poon C S 1996 *Nature* **381** 215

[22] Ikeguchi T ,Aihara K 1997 *Phys . Rev . E* **55** 2530

[23] Schreiber T 1997 *Phys . Rev . Lett .* **78** 843

The feature extraction of epileptic EEG signals based on nonlinear prediction^{*}

Meng Qing-Fang¹✉† Zhou Wei-Dong² Chen Yue-Hui¹ Peng Yu-Hua²

¹✉ School of Information Science and Engineering, University of Jinan, Jinan 250022, China

²✉ School of Information Science and Engineering, Shandong University, Jinan 250100, China

(Received 25 February 2009 ; revised manuscript received 5 May 2009)

Abstract

Based on the nonlinear time series prediction, a feature extraction method for epileptic EEG signals using nonlinear prediction is proposed to automatically detect the epileptic EEG from EEG recordings. To reconstruct the phase space, the approach of determining the embedding dimension based on nonlinear predictability is used to determine the embedding dimension of the EEG signals. The experimental results show that the feature extracted with the method based on nonlinear prediction could clearly distinguish the epileptic EEG from the normal EEG, and the proposed nonlinear feature extraction method is fit for the small set time series and is stable to noise.

Keywords : EEG signals, epilepsy, nonlinear prediction, feature extraction

PACC : 0545

^{*} Project supported by the National Natural Science Foundation of China (Grant No. 60573065), the Natural Science Foundation of Shandong Province, China (Grant Nos. Y2007G31, Y2007G33), the Shandong Distinguished Middle-aged and Young Scientist Encourage and Reward Foundation, China (Grant No. BS2009SW003), and the Key Subject Research Foundation of Shandong Province, China (Grant No. XTD0708).

† E-mail : qfmeng@sohu.com