

基于 SVM 的地震序列类型早期预测研究^{*}

李冬梅^{1,2}, 周翠英², 朱成林², 孙龙梅³, 吕子强², 程显洲¹

(1. 中国地震局兰州地震研究所, 甘肃 兰州 730000; 2. 山东省地震局, 山东 济南 250014;
3. 山东女子学院, 山东 济南 250300)

摘要: 在 Matlab 环境下, 通过构造 SVM, 建立地震序列特征参数与序列类型之间的一种非线性映射关系对地震序列类型进行早期分类预测。依据我国 1970 年以来的 $M_s \geq 5.0$ 地震序列资料, 使用 SVM 对震后 1、2、3、5、7 天 5 个时间尺度的地震序列类型进行早期预测, 识别效果较好, 处理速度快, 具有较强的实用性。

关键词: 地震序列; SVM; 统计模式识别; 序列类型; 早期预测

中图分类号: P315.7

文献标识码: A

文章编号: 1000-0666(2013)01-0069-05

0 引言

地震序列的类型判断及其成因机理研究是地震学的基本问题之一, 对于揭示地震孕育和发生发展过程物理本质具有重要理论意义。一次较大地震发生后, 后续的地震活动趋势如何发展? 其后是否有较大余震或者更大地震发生? 其发震时间和地点判断等, 都是政府和震区民众最为关心的问题。而要较好地回答上述问题, 地震序列类型的早期快速预测是基础和关键。追踪 2008 年汶川 8.0 级地震和最近 40 年来我国发生的历次大地震, 震后趋势快速判断对震后科学应急决策、及时的救助救援和稳定社会、安抚民心、灾后重建等工作均起着至关重要的作用。当前, 我国仍存在发生破坏性地震的危险, 尤其是在人口稠密、经济发达的华北地区, 这种危险在日益迫近, 因此震后趋势快速判断问题愈显重要。

对于一个完整的地震序列, 其类型的判别已有较为成熟的方法和参数 (吴开统等, 1976, 1990; 周惠兰等, 1980)。但在地震刚刚发生, 序列还极不完整的情况下, 进行序列类型的早期预测, 目前尚无统一的规范性的成熟方法或规则。

地震序列类型的划分, 可以将之转化为统计模式识别问题。SVM (Support Vector Machine, 支持向量机) 是一种基于统计学习理论、建立在 VC

维理论和结构风险最小原理基础上的方法, 其在解决小样本、非线性、过学习及高维模式识别中表现出许多特有的优势, 在模式分类问题上具有很好的泛化性能。它的主要思想是在特征空间中建构最优分割超平面, 使得正例和反例之间的隔离边缘最大化。支持向量机是处理高度非线性分类、回归等新问题的新方法, 近年来在地震学领域也有应用, 王炜等 (2005, 2006b) 开展了相关工作, 将支持向量机方法引用到对我国大陆强震预测及时间序列预测中, 效果较好, 王炜等 (2006b) 还介绍了支持向量机在地震预测其它可应用的领域。基于支持向量机计算简单, 且有很强的鲁棒性和泛化性能, 我们将其应用到地震序列类型早期预测。笔者是在 Matlab 环境下, 通过构造 SVM, 建立地震序列特征参数与序列类型之间的一种非线性映射关系, 从而对地震序列类型进行早期分类预测的尝试。

1 资料整理及分析

1.1 资料的选取

针对中国大陆地震序列类型所开展的研究中, 蒋海昆等 (2007) 搜集整理了 1966 ~ 2002 年中国大陆地区的纪录相对完备的 5 级以上地震序列 183 次。本文基于这些资料, 从中筛选出最大地震 $M_1 \geq 5.0$ 的中、强地震序列, 并删除余

^{*} 收稿日期: 2012-03-07.

基金项目: 国家科技支撑计划项目专题五 (2012BAK19B04-05) 资助.

震过少（<30次）的序列28个，补充了2003~2011年发生的地震序列25个，共计180个地震序列构成本文研究的基础数据，其中7.0级以上地震序列23个；6.0~6.9级序列66个；5.0~5.9级序列91个。进一步根据资料情况，确定研究样本的震级下限（刘正荣，1984），具体方法是绘制每个地震序列样本频度—震级图，统计相应震级的地震次数，根据最大频次相应的震

级确定下限震级，低于下限震级的地震被认为记录不完整，不予采用。

1.2 资料初始分类

吴开统等（1990）根据多年的观测经验提出了3条地震序列类型的划分原则，分别为序列记录中最大地震释放的能量与全序列地震释放的能量之比、最大地震与次大地震的震级差以及余震多少，具体规则（周翠英等，2010）见表1。

表1 地震序列类型划分标准
Tab.1 Classification standards of earthquake sequence type

序列类型	判别标准			备注
	E_m/E_Σ	ΔM	频度情况	
主余型	$90\% \leq E_m/E_\Sigma < 99.9\%$	$0.7 < \Delta M \leq 2.4$	余震数较多	序列中只存在1个突出大震，序列地震次数多
震群型（包括多震型和双震型）	$E_m/E_\Sigma < 90\%$	$\Delta M \leq 0.7$	余震数较多	序列中存在两个以上突出大震，且它们之间相差不大，序列地震次数多。其中最大地震 $M_1 \geq 5.0$ 为多震型
				序列， $M_1 < 5.0$ 为中、小震群型序列
孤立型	$E_m/E_\Sigma \geq 99.9\%$	$\Delta M > 2.4$	余震数很少	序列中只存在1个突出大震，序列余震次数较其它两类少

表1中震级差是以 M_s 为震级标度，若以 M_L 为标度，蒋海昆等（2007）做过推导，统一用序列主震与12个月内最大余震之间的震级差 ΔM 进行序列类型划分，划分标准为：当 $0.6 \leq \Delta M \leq 2.4$ 时，序列为主余型；当 $\Delta M \geq 2.5$ 且余震次数较少时，序列为孤立型；当 $\Delta M < 0.6$ 序列为震群型。本文综合地震波能量比和蒋海昆（2006）提出的震级差划分原则，对所要研究的180次地震序列进行了初始分类，结果见表2。

表2 不同主震震级地震序列类型初始分类统计
Tab.2 Initial classified statistics of earthquake sequence types with different magnitudes

类型	7 级(G)	6 级(S)	5 级(M)	总数	比例	
主余型	17	44	55	116	64.5%	
孤立型	0	7	13	20	11.1%	
震群型	多震型	3	6	11	20	11.1%
	双震型	3	9	12	24	13.3%

1.3 资料及预处理

在所搜集的180个地震序列资料中，分别选取序列第一次大震 M_1 （对于主余型序列，该震即为主震 M_{max} ，对于强震群型或双震型，则该震为序

列第一个大震）后1~7天的资料，对其后1、2、3、5和7天5个时间尺度进行序列分类早期预测研究（王伟等，2006b）。

2 选择序列单参数判据

根据地震序列分类规则和活动特征，序列参数的选择主要考虑以下2个方面的因素（蒋海昆等，2007）：（1）参考以往对单参数序列分类能力的统计评价结果；（2）对一些余震不十分发育的序列，许多参数由于余震数目的限制而无法计算，因此判别函数的建立主要考虑采用直接表征序列频次及应变能释放特征的简单参数。

对于序列参数的选取的研究前人已开展工作，本文没有再进行相关的参数选择，依据蒋海昆等（2007）、周翠英等（1996）、刘正荣和孔绍麟（1986）的研究成果，在广泛细致分析各种单项特征指标基础上，综合分析寻求有可能反映序列活动本质规律的整体特征，笔者选择了表3中所列的7个常用的表征序列特征的参数，蒋海昆等（2006）给出各参数详细物理含义。需要指出的是虽然上述参数很多是相关的，但是从不同侧面反映序列演化特征。进而分震后1、2、3、5、7天5个时间尺度计算各单项序列参数。

表 3 序列单项参数

Tab. 3 Single parameters of sequence

序号	参数名称	参数含义
1	M_2	统计时间段内，序列次大地震
2	$\angle M_1$	统计时间段内，序列最大地震 M_1 与次大地震 M_2 之差
3	$\angle M_2$	统计时间段内，序列最大地震 M_1 与第三大地震 M_3 之差
4	E1E	统计时间段内，最大地震 M_1 的能量与统计时间段内总能量之比
5	E2E	统计时间段内，次大地震 M_2 的能量与统计时间段内总能量之比
6	N2	统计时间段内震级 $\geq M_1 - 2.0$ 的地震数
7	K 值	归一化信息熵，描述地震序列能量分配均匀程度

3 构造 SVM 模型

SVM 模型建立主要有 6 个步骤（史峰等，2010）：首先从原始数据中提取训练集和测试集，然

后进行预处理（有时需特征提取），选择 svmtrain 的惩罚参数 c 和核函数参数 g （本文运用粒子群优化算法），之后用最佳参数对训练集进行训练，用得到的模型来预测测试集的分类标签，将其分类，得到分类准确率，算法流程如图 1 所示。

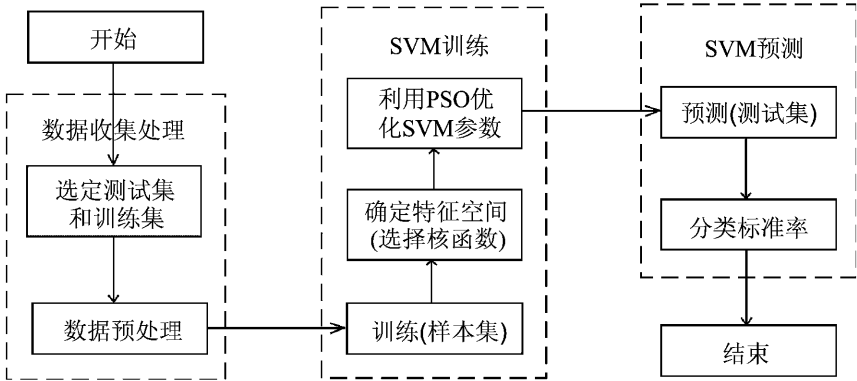


图 1 SVM 模型流程图

Fig. 1 Flow chart of SVM model

4 SVM 的 Matlab 实现

4.1 选定训练集和测试集

在所选取的 180 个地震序列样本中，每个样本含有 7 个特征分量（序列单参数判据），这些样本有 3 种序列类型，分别给予类别标签。其中 1 ~ 116 属于主余型（类别标签为 1），117 ~ 160 属于震群型（类别标签为 2），161 ~ 180 属于孤立型（类别标签为 3）。按照 3:1 将每个类别样本随机分为两组，重新组合数据，分别作为训练集和测试集。

4.2 数据预处理

对训练集和测试集资料进行归一化处理，使各样本元素的值在 $[0, 1]$ 之间，确保网络对样

本具有足够的输入敏感性和良好的拟合性。选用的归一化方法为

$$\hat{x} = \frac{x - x_{\min}}{x_{\max} - x_{\min}}.$$

(1)

在 Matlab 中，mapminmax 函数可以实现上述归一化，常用的实现命令（刘正荣，1984）为

$$[y, ps] = \text{mapminmax}(x).$$

(2)

其中， x 是原始数据， y 是归一化后的数据， ps 是个结构体，记录的是归一化的映射。

4.3 c&g 参数寻优

SVM 做分类预测需要调节相关参数（主要是惩罚参数 c 和核函数参数 g ），以达到比较理想的

预测分类准确率，避免过学习和欠学习状况发生。关于 SVM 参数的优化选取，国际上目前还没有公认统一的最好的方法，本文采用粒子群优化算法参数寻优（Particle Swarm Optimization，简称 PSO），其是通过粒子在解空间追随最优的例子进行搜索（史峰等，2010）。

4.4 训练和预测

采用 matlab 中的 libsvm 工具箱，运用训练函数 svmtrain，选择 C-SVC 类型，采用径向基函数作为核函数，它利用局部接收域完成函数映射，在 SVM 中应用最为广泛。进一步确定 SVM 其它参数，运用 PSO 算法寻优，使用训练集对 SVM 进行训练，用得到的模型分别对训练集和测试集进行内符及外推检验。

4.5 内符检验结果与分析

依照以上步骤，用所得到的地震序列参数判据，对 180 个已知样本进行内符检验，选择合适的 c&g 参数，180 个样本都能识别正确。说明 SVM 模型具有较好的内符识别效果，对训练集学习后能够高准确率的预测，学习能力较强。

4.6 测试集分类结果与分析

按照步骤 1 对训练集和测试集划分，进而对 SVM 进行学习训练，选择合适的 c&g 参数，用得到的模型对 1、2、3、5 和 7 天 5 个时间段的测试集进行预测，以可达到的分类准确率作为评价 SVM 的性能指标。将 5 个时段的判别结果按序列

类型统计列于表 4，可知：

（1）利用震后 1~7 天的序列资料进行序列类型划分，1 天的判别正确率达到 82.2%，具有相对较高的类型判别能力。并且资料长度（震后时间）越增加，3 种序列类型识别正确率总和越高。

（2）主余型序列识别正确率介于震群型和孤立型之间，其主要原因为主余型的训练样本是最多的，学习最为完善，识别正确率相对较高；序列自身活动特征也是影响因素。

（3）震群型序列的识别正确率基本上随着资料长度（震后时间）增加而增大。深入分析震群型序列识别正确率相对较低原因，除去资料样本的关系，也和该序列的形态和活动特征有关。

（4）除第一天外，其余 4 个时间段孤立型序列都完全识别出来，识别正确率基本保持在一个较高水平，其判别正确率与震后时间长度变化关系不大。

（5）具体分析识别错误的地震序列，发现 5 个时段错判对象有些是基本固定的，例如 1976 年 11 月 7 日宁蒗 6.7 级地震序列在 5 个时段错分 3 次，分析发现该序列从 1976 年 11 月 7 日 6.7 级地震后，序列一直延续到 12 月 13 日才发生 6.4 级次大地震，两次地震相距 36 天。该序列的活动特征影响了序列类型判定，在本文 7 天的统计时间段内未发生次大地震，在判别过程中将其误判为主余型。

表 4 分序列类型识别正确率统计
Tab. 4 Accuracy rate statistics of type recognition of subsequences

序列类型	1 天	2 天	3 天	5 天	7 天
主余型	0.931(27/29)	0.966(28/29)	0.897(26/29)	0.966(28/29)	0.931(27/29)
震群型	0.636(7/11)	0.636(7/11)	0.727(8/11)	0.727(8/11)	0.818(9/11)
孤立型	0.6(3/5)	1(5/5)	1(5/5)	1(5/5)	1(5/5)
总计	0.822(37/45)	0.867(39/45)	0.867(39/45)	0.911(41/45)	0.911(41/45)

5 结论与讨论

（1）笔者对提取的地震序列早期（1~7 天内）分类参量进行综合后，优于一般的单参数分类指标效果，有较高的实用价值。

（2）SVM 预报效果依赖于所选用的网络模型、核函数及所选用的各参数指标，如何设置更为有效需要进一步探索。

（3）从输出结果看，对于震群型序列，本文的识别正确率较低。SVM 的预测准确性与训练样本数量及选取有较大关系，进一步丰富地震序列样本，和进行分区建模试验，有望进一步改善预测模型，提高正确识别率。

（4）建模实验中采用的序列参数是基于以往研究结果，参数选取影响模型预测效果。为获得更为完美的预测模型，充分的参数选取试验和进一步挖掘新的序列参数是有待继续进行的工作。

综上所述, SVM 方法用于对地震序列类型预测, 效果较好、处理速度快、具有较强的实用性, 有广泛的应用前景。但是也存在一些要完善和检验的问题, 所以要提高模型的分类准确率还需要进一步完善 SVM 模型。

参考文献:

- 蒋海昆, 代磊, 侯海峰, 等. 2006. 余震序列性质判定单参数判据的统计研究[J]. 地震, 26(3): 17-25.
- 蒋海昆, 傅征祥, 刘杰, 等. 2007. 中国大陆地震序列研究[M]. 北京: 地震出版社.
- 刘正荣, 孔绍麟. 1986. 地震频度衰减与地震预报[J]. 地震研究, 9(1): 6-8.
- 刘正荣. 1984. 根据地震频度衰减预报地震的工作细则[J]. 地震, (1): 32-37.
- 史峰, 王小川, 郁磊, 等. 2010. Matlab 神经网络 30 个案例分析[M].

- 北京: 北京航空航天大学出版社.
- 王炜, 林命遇, 马钦忠, 等. 2006a. 支持向量机及其在地震预报中的应用前景[J]. 西北地震学报, 28(1): 78-84.
- 王炜, 刘悦, 李国正, 等. 2005. 中国大陆强震时间序列预测的支持向量机方法[J]. 地震, 25(4): 26-32.
- 王炜, 刘悦, 李国正, 等. 2006b. 我国大陆强震预测的支持向量机方法[J]. 地震学报, 28(1): 29-36.
- 吴开统, 焦远碧, 吕培苓, 等. 1990. 地震序列概论[M]. 北京: 北京大学出版社.
- 吴开统, 岳明生, 武宜英, 等. 1976. 海城地震序列的特征[J]. 地球物理学报, 19(2): 95-109.
- 周翠英, 耿杰, 王梅, 等. 2010. 山东地区地震分析预报手册—预测方法、指标、震例[M]. 济南: 山东科学技术出版社, 37-38.
- 周翠英, 张宇霞, 王红卫. 1996. 以模式识别方法提取地震序列早期判断的综合指标[J]. 地震学报, 18(1): 118-124.
- 周惠兰, 房桂荣, 章爱娣, 等. 1980. 地震震型判断方法探讨[J]. 西北地震学报, 2(2): 45-59.

Research on Early Judgment of Earthquake Sequence Types Based on SVM

LI Dong-mei^{1,2}, ZHOU Cui-ying², ZHU Cheng-lin², SUN Long-mei³,
LU Zi-qiang², CHENG Xian-zhou¹

(1. Lanzhou Institute of Seismology, CEA, Lanzhou 730000, Gansu, China)

(2. Earthquake Administration of Shandong Province, Jinan 250014, Shandong, China)

(3. Shandong Women's University, Jinan 250300, Shandong, China)

Abstract

In the environment of Matlab, we construct Support Vector Machine (SVM) to build a kind of nonlinear mapping relationship between seismic sequence characteristic parameter and sequence type, and do the early predictions of earthquake sequence types. On the basis of $M_s \geq 5.0$ earthquake sequences in China since 1970, we divide the data in 5 time scales according to one, two, three, five and seven days after the earthquake and apply the SVM to the early predictions of earthquake sequence types. The results show that it achieves good recognition and fast processing speed, and has a strong practicability.

Key words: seismic sequence; SVM; statistic pattern recognition; sequence type; early prediction