

寻找地震相关地区的时间序列相似性匹配算法*

吴绍春¹⁺, 吴耿锋¹, 王 炜², 蔚赵春¹

¹(上海大学 计算机工程与科学学院, 上海 200072)

²(上海市地震局, 上海 200062)

A Time-Sequence Similarity Matching Algorithm for Seismological Relevant Zones

WU Shao-Chun¹⁺, WU Geng-Feng¹, WANG Wei², YU Zhao-Chun¹

¹(School of Computer Engineering and Science, Shanghai University, Shanghai 200072, China)

²(Seismology Bureau of Shanghai, Shanghai 200062, China)

+ Corresponding author: Phn: +86-21-56335263, Fax: +86-21-56333061, E-mail: scwu812@staff.shu.edu.cn, <http://www.shu.edu.cn>

Wu SC, Wu GF, Wang W, Yu ZC. A time-sequence similarity matching algorithm for seismological relevant zones. *Journal of Software*, 2006,17(2):185–192. <http://www.jos.org.cn/1000-9825/17/185.htm>

Abstract: On the basis of analyzing the newly time sequence research achievement nowadays, several definitions on seismological zone relativity are put forward in this paper for integrating the large amount of history earthquake source data and the experimental expert knowledge in seismological field. At the same time, the time sequence similarity-matching model of the relevant seismological zone is presented, and then it is implemented through several correlative experimental simulations. Based on the sequence similarity-matching model, a sequence-matching algorithm is given with seismological similarity. Furthermore, by discovering the history earthquake database (EQDB) in recent 20 years, some experiments are provided to analyze longitudinal thick-granularity sequential similarity and thin-granularity sequential similarity. Finally, the experimental result has found its satisfactory way out by using the proposed algorithm to support earthquake prediction.

Key words: time-sequence; similarity matching; seismological prediction; algorithm; granularity; seismological relevant zones

摘 要: 把时间序列相似性匹配的基本概念和方法引入到地震预报的应用中.在分析现阶段时间序列研究成果的基础上,结合大量地震历史源数据和领域专家经验知识,提出了有关地震地区相关性的地震相似度定义和地震序列相似性匹配模型,并通过大量实验模拟对该模型进行了反复验证,实现了基于地震相似度的时间序列相似性匹配算法.同时,通过分析我国地震活动频繁区域近 20 年来的地震历史数据,应用地震区域序列相似性匹配算法进行了固定时间差的粗粒度和细粒度纵向序列相似性实验分析,取得了可信度较高的实验结果,为地震学预测的应用研究提供了较好的技术支持.

关键词: 时间序列;相似性匹配;地震预报;算法;粒度;地震相关地区

* Supported by the National Joint Foundation of Earthquake of China under Grant No.104090 (国家地震科学联合基金); the Natural Science Foundation of Shanghai Municipality of China under Grant No.03ZR14038 (上海市自然科学基金); the Foundation of Shanghai Municipal Education Commission of China under Grant No.EWL200312 (上海高校网格项目)

Received 2005-03-30; Accepted 2005-07-11

中图法分类号: TP391 文献标识码: A

时间序列相似性匹配(又称为相似性挖掘或相似性查询)是指在时间序列数据库中发现与给定序列模式相似的序列,在统计理论、机器学习以及数据挖掘等方面具有重要的意义.在时间序列数据挖掘中,相似性问题的研究主要与查询匹配以及时序分类紧密结合在一起.在许多信息检索和数据挖掘系统中,定义和度量对象之间的相似性有着重要的地位.相似性度量常用于发现具有相似起伏的股票、确定具有相似销售模式的产品、分类具有相似形状的恒星光谱曲线等等.

在地震预报科学中,专家们经过长期的观测研究和经验积累发现,某些地区大范围的地震活动往往同时趋于平静,即一定区域上的地震活动有同步涨落现象.在这种一定距离的两地区中,某些特定震级以上的显著地震相伴发生的现象称为地震的相关现象,也称为地震的地区相关性.文献[1,2]详尽地研究了北京、新疆及其他地区地震活动的相关性,总结出了如何利用地震活动相关性进行预报的方法.

地震的地区相关性,反映了一定地质构造上地震发生的规律性先后关系.因此,寻找地震相关地区,进而预报相关地震,是人们较为熟悉的地震预报方法之一^[3].然而,传统的经验性手工式方法非常费时、费力.本文把数据挖掘技术引入到地震预报科学中,通过不同地区地震序列的相似性匹配来发现地震相关地区,进而利用地区相关性进行地震预报,这无疑是一项非常有意义的工作.本文首先对原始震例数据进行预处理,将其转化为时间轴上的时间序列,然后结合地震领域的特点,定义序列的相似性度量模型,提出一种基于相似度的序列相似性匹配算法,利用相关地震的相似度快速而全面地发现地震相关的地区.

1 地震序列相似性的定义和度量模型

序列相似性匹配问题的研究核心主要包括 3 个部分,即相似性的定义、相似性度量模型的建立和相似性匹配算法的实现^[4].一般来说,相似性的定义根据应用需求而定,而相似性度量模型则是依据所定义的相似性进行数学抽象而成.在相似性定义方面,Agrawal 等人^[5]提出了一种依据直观意义上时间序列数据的上升、下降趋势定义的相似性;有的相似性定义则比较复杂.相似性度量较早采用 ARMA(auto regression moving average)模型和基于离散傅里叶变换 DFT(discrete Fourier transform)相似性匹配.前者采用随机时间序列分析技术;后者通过 DFT 将时间序列从时域空间映射到频域空间,保留前 k 个傅里叶系数,将序列变为 k -维空间中的点,然后用多维索引方法来存储和检索这些多维空间中的点.这些模型在判断序列相似时,均采用了欧几里德距离作为序列间的相似性评价函数,即当两序列间欧氏距离小于给定阈值时,就认为序列相似.但在实际应用中,由于序列长度不等或采样率不同等问题,使得欧氏距离难以直接应用.

传统的时间序列相似性匹配方法多是针对数值时间序列,而在事件序列的相似性匹配方面所做的研究工作则很少.地震目录数据是记载了历次地震发生的时间、地点(经度、纬度)、震级以及破坏程度的地震事件序列.由于地震事件序列数据往往具有不规则、混沌等非线性特征,因此传统的时间序列相似性度量模型和匹配方法都难以适用于地震数据,必须给出一个合适的地震序列相似性定义和度量模型.

为此,本文依据地震领域知识,通过适当的预处理,将地震目录数据转换为一定时空范围内、一定震级以上的地震事件系列集.这样,不同时空范围内的地震就构成了不同的地震序列.下面给出地震序列及其相似性的定义和地震序列相似性的度量模型.

定义 1(地震事件). 将地震目录数据库中记录的每一个地震看作一次事件.将所有事件按空间属性分离为不同区域的地震事件集,而把地震事件集中在 t 时刻所发生的一定震级以上的地震事件记为 $E(t)$.

定义 2(地震事件序列). 在某地震区域中,发生在时间范围 G 内的事件集 $E_t = \{E(t) | t \in \text{时间范围 } G\}$ 在时间轴上的一个排列,称为地震事件序列.其中每一个单独的事件 $E(t)$ 称为一个事件项.

定义 3(地震序列). 把地震事件项的空间属性略去,依据地震发生的时间和震级,抽象为时间轴上一个时间单位上的地震强度值.这样,一个地震事件序列就被抽象为一个时间序列,我们称其为地震序列,用 $S = \{S(t) | t = 0, 1, 2, \dots, n\}$ 来表示,地震事件序列中,一个事件项对应地震序列中的一个元素.这里要特别说明的是,根据领域特性,

我们将地震事件发生的时间 t 扩展为一个单位时间间隔范围,即 $t \in T_{gap}$.

定义 4(地震序列长度). 对于地震序列 S ,用 $|S|$ 表示其长度,即地震序列 S 的元素个数. $S[i]$ 表示地震序列 S 中的第 i 个元素, $S(t)$ 表示地震序列 S 在 t 时刻的震级取值.

定义 5(地震相似性). 设 x 和 y 是任意两个不同地震序列中的对应元素,用记分函数 $\sigma(x,y)$ 来描述 x 和 y 元素的相似性.记分函数 $\sigma(x,y)$ 定义为

$$\begin{aligned} \sigma(x,y) &= 1, \text{ 如果 } |x(t)-y(t')| \leq M_{threshold}; \\ \sigma(x,y) &= 0, \text{ 如果 } |x(t)-y(t')| > M_{threshold}, \text{ 或者 } x(t')=0, \text{ 或者 } y(t')=0. \end{aligned}$$

这里, $M_{threshold}$ 为震级误差阈值,由用户根据具体需求来确定.震级差阈值越小,表示两个元素所代表的地震强度差别越小.特别值得注意的是,因为允许相关地震发生的时间有一定的间隔,所以 x 和 y 两元素可以对应不同的时间点,用户可以根据需要给定这个时间间隔的范围.

另一方面,地震序列中常常会出现这样的现象:在时间轴的某些单位时间段内,没有地震事件发生.这时,可以在序列中补充所谓的空元素,即在没有地震事件的时间段补充取值为 0 的元素.当非空元素与空元素进行相似性比较时,记分函数取值为 0.

定义 6(稠密地震序列). 在某一时间区域内,对地震序列 S ,在没有发生地震事件的单位时间段补充了空元素以后形成的地震序列,称为稠密地震序列 S' .

定义 7(地震序列相似性度量模型). 设 S 和 R 是两个不同的地震序列,则 S 和 R 的相似性可以用两序列中对应元素的相似性记分函数加权和来度量.两地震序列相似性度量模型定义为

$$Score(S,R) = \sum_{i=1}^{\max(|S'|,|R'|)} \sigma(S'[i],R'[i]) * W(S'[i],R'[i]) \quad (1)$$

这里:(1) S' 和 R' 分别是在指定时间范围内对 S 和 R 补充了空元素以后形成的稠密地震序列,将 S' 和 R' 中的空元素除去后即还原为 S 和 R ;

(2) $|S'|=|R'|$ (即 S',R' 两序列长度相等);

(3) $W(x,y)$ 是权重函数,定义为

$$\begin{aligned} W(x,y) &= \max\{x(t),y(t)\}, \text{ 如果 } x(t) \geq M, \text{ 或者 } y(t) \geq M; \\ W(x,y) &= \min\{x(t),y(t)\}, \text{ 如果 } x(t) < M, \text{ 并且 } y(t) < M. \end{aligned}$$

其中, M 是由用户根据应用情况输入的震级标准.这个 M 值应使权重的定义能够反映地震能量的特性,因为当震级大于一定的震级标准时,震级大的地震其相似远比震级小的地震的相似有意义,所以权重就应该比较大;反之亦然.

定义 8(相关地震支持数及支持度). 相关地震支持数 $Sup(S,R)$ 及支持度 η 定义为

$$\begin{aligned} \forall t \in [T-T_{gap}, T+T_{gap}], \exists E(t) (E(t) \in \{S_t \cap R_t\}) \Rightarrow Sup(E(t)) = 1, \\ \forall t \in [T-T_{gap}, T+T_{gap}], \neg \exists E(t) (E(t) \in \{S_t \cap R_t\}) \Rightarrow Sup(E(t)) = 0, \\ Sup(S,R) = \sum_{t=0}^n Sup(E(t)) \quad (2) \\ \eta = Sup(S,R) / |S'| \quad (3) \end{aligned}$$

其中, $E(t) \in \{S_t \cap R_t\}$ 表示在一个时间间隔范围 $t \in [T-T_{gap}, T+T_{gap}]$ 内,只要序列 S 和序列 R 中均有地震事件发生,支持度总数就增加 1;如果序列 S 和序列 R 有一个没有地震事件,就不记入支持度总数.式(2)中的 n 为时间间隔总数.

2 寻找地震相关地区的序列相似性匹配算法

序列相似性匹配可以分为两类:一类是整体匹配,即查询序列和数据库中的记录序列具有相同的长度;另一类是子序列匹配,查询序列比数据库中的记录序列要短,需要在记录序列中寻找和查询序列相似的子序列.用序列相似性匹配的算法来发现地震相关地区,主要思想就是在各个地区的地震序列中,用整体序列相似性匹配的

方法寻找具有高相似度的地震序列.这种地震序列之间的高相似度,必须反映两地在地震的发生上具有一定规律性的先后关系.为此,我们结合相关地震的特点,在定义了相关地震的支持度和地震序列相似性度量模型的基础上,提出了一种基于支持数的整体序列相似性匹配算法 WSMBSS(whole sequence matching based-on seismo similarity support).下面给出 WSMBSS 算法的形式化描述.

算法. Whole Sequence Matching Based-on Seismo Similarity Support (WSMBSS).

输入:地震目录 M ,时间窗口宽度 T_{gap} ,最小相似度 λ ,最小支持度 S_{min} .

输出:(地震相关地区,支持度).

```

Step1: Transform  $M$  into EQS with Space Label
Step2: For each  $eq$  in EQS catalogue with the same Label  $l$ 
        Add Event ( $eq$ ) into EQ Sequence  $EQS_i$ 
Step3: for ( $i=1; i \leq Num(EQS); i++$ )
        { for ( $j=1; j \leq Num(EQS); j++$ )
          if ( $i < j$  &&  $Score(EQS_i, EQS_j) > \lambda$ )
            { count=0
              for ( $m=1; m \leq |EQS_i|; m++$ )
                for ( $n=1; n \leq |EQS_j|; n++$ )
                  if ( $T(EQS_m(i)) - T(EQS_n(j)) \leq T_{gap}$ ) count ++;
                  if ( $count \geq S_{min}$ ) output ( $EQS_i, EQS_j, count$ )
            }
          }
        }
Step4: Merge the result.

```

WSMBSS 算法说明:

步骤(1):预处理——从地震目录数据中取出数据,根据其包含的经纬度信息,按空间位置划分为不同的区域,并将经纬度和区域的对应信息都一同存放在预处理结果文件中;

步骤(2):构成地震序列——从预处理结果文件中,提取具有相同区域编号的数据归并到一起,分别放入不同的临时文件中,并累计此区域内的地震记录条数,记入(区域号、条数)链表结构中.最后,各区域内的震例记录按时间排序,就形成一组地震时间序列 EQS ;

步骤(3):两两进行相似性分析——具体过程是:取出分别处于不同区域中的地震时间序列,若它们的相似度达到要求,则作相应的记录.然后计算支持度,并根据支持度是否达到阈值输出相关序列;

步骤(4):合并结果——将地震相关地区依据地理位置进行合并,并重新计算支持度.

在此算法中对两个序列进行相似性判定时,需要逐个元素比较.本文在算法的设计过程中借鉴了 KMP 思想,即利用已经得到的“部分匹配”的结果,将比较序列向后“滑动”尽可能远的一段距离后继续比较,这样极大地改善了算法的效率.

3 应用实例及结果分析

3.1 数据准备

数据准备包括数据的选取和预处理.本文所选取的数据资料来源于中国地震局,包括从 1965 年 1 月 1 日~2003 年 5 月 31 日发生在全国境内的地震信息.因为对大于一定震级的地震作出预报才有意义,所以,本文选取了剔除余震后的 4 级以上地震记录,共有记录 8 700 条.进行数据预处理的目的是为了使数据适应时间序列的相似性匹配算法,本文分别对空间窗口、时间窗口和震级进行预处理:

- 1) 设定地震时间序列的空间窗口(经度范围和纬度范围)、震级标准 M (选取一定震级以上的地震);
- 2) 按照空间窗口将整个地球表面划分为若干区域并编号,生成区域和经纬度对应表;

3) 扫描地震目录数据 ML ,对其每一条地震记录 Req ,依照其经纬度将地震划入相应的区域中,设置记录 Req 的区域类标为区域的 ID,同时将地震发生的时间映射为整数;

4) 将各个区域中的地震记录按照时间顺序转换为时间序列,并记录序列长度.

为了便于时间序列的比较,时间轴上的时间间隔应该是一致的.这样,当将地震数据转换为时间序列时,序列中可能出现没有地震记录的空缺值,这时用 0 来填补.

对地震震级进行正规化,即离散并转换为 10 以内的整数,并把正规化所得到的这个整数称为类震级.具体转换关系是:0 对应震级小于 3 级;1 对应震级[3.0,4.0);2 对应震级[4.0,5.0);3 对应震级[5.0,5.5);4 对应震级[5.5,6.0);5 对应震级[6.0,6.5);6 对应震级[6.5,7.0);7 对应震级[7.0,7.5);8 对应震级[7.5,8.0);9 对应震级[8.0,8.5];10 对应震级大于等于 8.5 级.

预处理后的部分地震目录数据格式见表 1.

Table 1 Catalog of seismic data after part pretreatment

表 1 部分预处理后的地震目录数据

Date	Time	Longitude	Latitude	Magnitude	Zone	Sequence time	Pseudo magnitude
19700105	01:00:00	102.35	24.06	7.8	0	17257824	8
19650127	16:56:00	102.18	25.48	4.3	0	17213496	2
19650127	17:00:00	102.24	25.48	5.8	0	17213496	4
19650508	06:52:00	103.24	23.18	3.2	0	17216928	1
19650524	00:00:00	102.35	24.06	5.4	0	17216544	3

3.2 实验设计及结果分析

地震的序列相似性是指某一时空范围内,两个地区发生的地震在时间分布方面具有相似性.以一年时间差为例,如果某区域 A 在 1900 年,1905 年,……,1948 年,……,1976 年...有一定震级以上的地震发生,而在区域 B 相应地于 1901 年,1906 年,……,1949 年,……,1977 年...也有一定震级以上的地震发生,那么就可以认为,区域 A 与区域 B 在一定程度上存在地震相关关系,即 A 发生地震后的第 2 年,B 也发生地震.这种关系称为“固定时间差的纵向序列相似性关系”.这样的相似性是地震地区相关粗粒度的结论,因为在序列相似性匹配中,时间轴上的间隔是 1 年,而且序列元素取值为这一年中发生的地震中震级最大的地震.本文设计了 3 个实验,分别从粗粒度进行固定时间差为 1 年的纵向序列相似性分析、同年的纵向序列相似性分析和细粒度的序列相似性分析.

3.2.1 实验 1:固定时间差为 1 年的粗粒度序列相似性分析

选取自 1980 年 1 月 1 日开始的 5.0 级以上地震序列,用 WSMBSS 算法进行固定时间差为 1 年的纵向序列相似性分析,所得到的实验结果见表 2.实验中具体参数为:地理位置按地理坐标进行 $5^{\circ} \times 5^{\circ}$ 分块, $T_{gap}=1$ year(即两次地震的时间间隔为 1 年), $M_{threshold}=1$ (即震级相差小于 1 级),相关地震支持数阈值 Sup_{min} 为 10.

Table 2 Seismic relevant zones with fixed interval being 1 year and relevant time

表 2 固定时间差为 1 年的地震相关区域及相关次数表

Zone 1	Longitude range	Latitude range	Zone 2	Longitude range	Latitude range	Relevant time
452	100°~105°	25°~30°	492	120°~125°	20°~25°	18
414	90°~95°	30°~35°	492	120°~125°	20°~25°	16
415	95°~100°	30°~35°	492	120°~125°	20°~25°	15
375	75°~80°	35°~40°	492	120°~125°	20°~25°	14
452	100°~105°	25°~30°	414	90°~95°	30°~35°	13
413	85°~90°	30°~35°	492	120°~125°	20°~25°	13
488	100°~105°	20°~25°	492	120°~125°	20°~25°	13
452	100°~105°	25°~30°	375	75°~80°	35°~40°	12
379	95°~100°	35°~40°	492	120°~125°	20°~25°	12
415	95°~100°	30°~35°	452	100°~105°	25°~30°	12
487	95°~100°	20°~25°	492	120°~125°	20°~25°	12

从表 2 中可以看出:我国 4 级以上地震主要发生在台湾、四川以南、云南、青海、新疆以西以及西藏大片区域.从匹配的结果看,这些地区之间存在着较大的关联,尤其是台湾、四川以南地区,从 1980 年~2002 年,仅 22 年里就有 18 年以上发生过近 5 级以上的大地震,而且均存在时间差的关联.

下面对主要的相关区域进行分析.

如图 1 所示,左图表示区域 A(位置在四川-云南境内,其经度范围 $100^{\circ}\sim 105^{\circ}$,纬度范围 $25^{\circ}\sim 30^{\circ}$)和区域 B(台湾省周围,其经度范围 $120^{\circ}\sim 125^{\circ}$ E,纬度范围为 $20^{\circ}\sim 25^{\circ}$ N)相匹配的固定时间差为 1 年的地区相关性分析 M-T 图,其中匹配的次数为 18 次,即在四川-云南境内的每一次 5 级以上地震几乎都可以在台湾省周围找到与之匹配的一次地震记录.右图表示区域 C(位置在四川-云南境内,其经度范围 $100^{\circ}\sim 105^{\circ}$,纬度范围 $25^{\circ}\sim 30^{\circ}$)和区域 D(位置在青海-西藏境内,其经度范 $90^{\circ}\sim 95^{\circ}$,纬度范围 $30^{\circ}\sim 35^{\circ}$)相匹配的固定时间差为 1 年的纵向序列相似性分析 M-T 图,其中匹配的次数为 14 次.

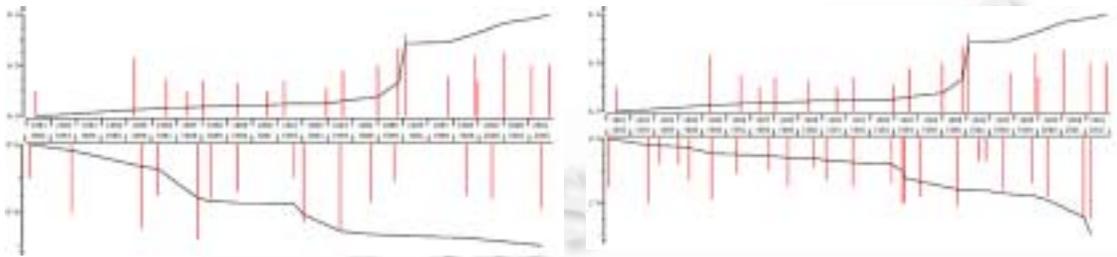


Fig.1 M-T graph interval 1 year

图 1 固定时间差为一年的 M-T 图

从图 2 可以很明显地看出这两对区域间有着很大的相关,特别是台湾和四川-云南的匹配程度很高,即存在固定时间差为一年的纵向序列匹配关系.一方面,这些地区地震发生比较频繁;另一方面,两地可能存在地区相关性.

3.2.2 实验 2:无固定时间差的粗粒度序列相似性分析

采用自 1980 年 1 月 1 日开始的震级在 5.0 以上震例数据转换为时间序列后,用 WSMBS 算法进行同年序列匹配的纵向序列相似性分析.实验中具体参数为:地理位置按地理坐标进行 $5^{\circ}\times 5^{\circ}$ 分块, $T_{gap}=1$ year(即两次地震的时间间隔为 1 年), $M_{threshold}=1$ (即震级相差小于 1 级),相关地震支持数阈值 Sup_{min} 为 10.同年匹配结果与次年相比,匹配次数在 10 次以上的,增加了新疆西边境($70^{\circ}\sim 75^{\circ}$ E, $35^{\circ}\sim 40^{\circ}$ N)和西藏-四川-云南交界($95^{\circ}\sim 100^{\circ}$ E, $25^{\circ}\sim 30^{\circ}$ N)两块区域,这说明它们也可能是地震相关地区.

下面对主要的相关区域进行分析.如图 2 所示.其中左图是台湾和四川-云南两地区 5 级以上相关地震的 M-T 图,图中给出了台湾和四川-云南两地区在 1981 年~1999 年 5 级以上地震构成的地震序列的同年相似性匹配的情况,即 19 年中匹配的次数为 12 次;右图为四川-云南和青海-西藏地区的对应 M-T 图.

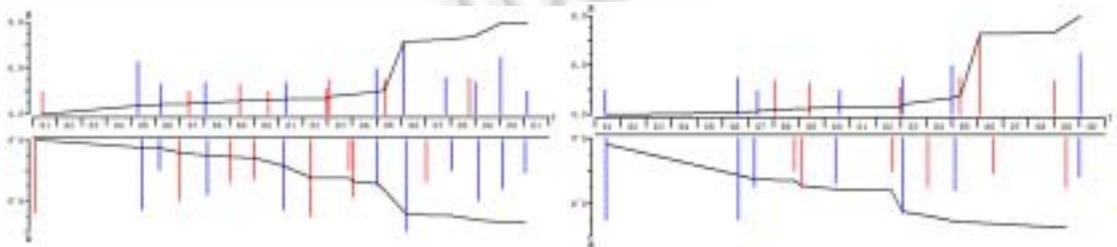


Fig.2 M-T graph in the same year

图 2 无固定时间差的 M-T 图

从图 2 中可以看出,还是在台湾和四川-云南、四川-云南和青海-西藏两组地区,在同一年多次在几乎同一时间两地都发生了 5 级以上的地震,这说明两地有地震相关关系.同样,考虑四川-云南和青海-西藏两地的地理位置关系,也说明这两个地区的地震活动基本是一致的,即一方地震的发生可能牵涉到另一方.

3.2.3 实验 3:较小时空窗口的细粒度序列相似性分析

采用自 1972 年 1 月 1 日开始的震级在 3.0 以上的震例数据,用 WSMBSS 算法进行细粒度的地区相关性分析.实验中具体参数为:地理位置按地理坐标进行 $1^\circ \times 1^\circ$ 分块, $T_{gap}=10$ days(即两次地震的时间间隔为 10 天), $M_{threshold}=1$ (即震级相差小于 1 级),地震相关支持数阈值 Sup_{min} 为 100.实验目的在于发现近 30 年来地震发生的地区相关性,得到的部分实验结果见表 3.

Table 3 Result of comparing zone relativity experiments with interval being 10 days

表 3 时间间隔为 10 天的地区相关性比较实验结果

Zone A	Longitude range	Latitude range	Quick time	Proportion	Zone B	Longitude range	Latitude range	Quick time	Proportion	Relevant time
11821	121°~122°	24°~25°	926	0.139	11260	100°~101°	27°~28°	244	0.529	129
12001	121°~122°	23°~24°	602	0.204	11260	100°~101°	27°~28°	244	0.504	123
9077	77°~78°	39°~40°	493	0.225	11260	100°~101°	27°~28°	244	0.455	111
8897	77°~78°	40°~41°	612	0.167	11260	100°~101°	27°~28°	244	0.418	102
8942	122°~123°	40°~41°	547	0.183	11260	100°~101°	27°~28°	244	0.410	100

以表中第 1 条记录为例进行分析,发现块 11 821 与块 11 260 的地震相关.11 821 块处于我国台湾省台北市的下方,位于我国较大的一个地震频繁发生的地震带上;而 11 260 块处于云南和四川交界处,位于另一个地震带上.结果显示,两地区地震相关次数超过总地震次数的一半.每当 11 821 块内发生一次地震,11 260 块内在 10 天内便会相应发生一次强度较其稍微偏弱的地震.近 30 年来,两地区发生的地震在较小时空窗口中匹配的次数达 129 次.两地发生的地震能如此同步地对应,其偶然巧合的机会是很小的,从本质上说明两地区存在着一定的相关性.

4 结 论

本文的工作首先对原始震例数据进行预处理,将其转化为时间轴上的时间序列.然后结合地震领域的特点,定义了序列的相似性度量模型,提出了一种基于相似度的序列相似性匹配算法.该算法利用相关地震的相似度,可快速、全面地发现地震相关的地区.实验结果表明,该算法发现了一些鲜为人知的地震相关地区.通过分析结果并与地震专家的经验相比较,证实该算法合理、有效.本文的工作为地震学预测的应用研究提供了较好的技术支持.

致谢 在此,我们向对本文的工作给予支持和建议的同行,尤其是上海大学和上海市地震局的所有本课题组的老师和同学们表示感谢.

References:

- [1] Liu WF, Liao QB. Discuss using relevance of earthquake activity to forecast earthquake. In: Xu SX, Lu YZ, Zhu CZ, Liu ZR, *et al.*, eds. Search Corpus of Earthquake Forecast Applied Method (Special of Seismology). Beijing: Academic Press, 1989. 253-267 (in Chinese with English abstract).
- [2] Ao XM, Wang GL, Huang KQ, Yang CR. Research of relevant earthquake forecasting. In: Xu SX, Lu YZ, Zhu CZ, Liu ZR, *et al.*, eds. Search Corpus of Earthquake Forecast Applied Method (Special of Seismology). Beijing: Academic Press, 1989. 268-295 (in Chinese with English abstract).
- [3] Lu YZ, Chen ZL, Wang BQ, Liu PX, Liu WL, Dai WL. Seismology Method of Earthquake Forecasting. Beijing: Earthquake Press, 1985 (in Chinese).
- [4] Bollobas B, Das G, Gunopulos D, Mannila H. Time-Series similarity problems and well-separated geometric sets. Nordic Journal of Computing, 1997,8(4):409-423.
- [5] Agrawal R, Psaila G, Wimmers E, Zait M. Querying shapes of histories. In: Proc. of the 21st Int'l Conf. on Very Large Database (VLDB'95). San Francisco: Morgan Kaufmann Publishers, 1995. 502-514.

附中文参考文献:

- [1] 刘王芬, 廖清波. 利用地震活动的相关性进行地震预报的初步探讨. 见: 许绍燮, 陆远忠, 朱传镇, 刘正荣, 等, 编. 地震预报方法实用化研究文集. 北京: 学术出版社, 1989.
- [2] 敖雪明, 王桂岭, 黄克强, 杨成荣. 相关地震预报方法的研究. 见: 许绍燮, 陆远忠, 朱传镇, 刘正荣, 等, 编. 地震预报方法实用化研究文集. 北京: 学术出版社, 1989. 268-295
- [3] 陆远忠, 陈章立, 王碧泉, 刘蒲雄, 刘文龙, 戴维乐. 地震预报的地震学方法. 北京: 地震出版社, 1985.



吴绍春(1965 -), 女, 江西宜春人, 博士, 副教授, 主要研究领域为数据挖掘, 智能信息处理.



王炜(1947 -), 男, 研究员, 主要研究领域为地震预报, 人工智能应用.



吴耿锋(1945 -), 男, 教授, 博士生导师, 主要研究领域为智能信息处理, 人工智能应用.



蔚赵春(1981 -), 男, 硕士, 主要研究领域为数据挖掘.

全国第 5 次程序设计语言发展与教学学术会议

征文通知

全国第 5 次程序设计语言发展与教学学术会议定于 2006 年 10 月在南京召开。本次会议由东南大学主办, 将就程序设计语言现状与发展、各类程序设计语言及其设计与实现、程序设计语言教学与教材等方面的内容进行广泛的学术交流。会议论文集仍由国家出版社正式出版, 其中部分优秀论文将在会后推荐给国内权威核心杂志发表。

一、征文范围 (包括但不限于)

程序设计语言历史、现状与发展; 面向对象语言及相关技术; 各类建模语言及其设计、实现与应用面向网络应用的程序设计语言; 程序设计语言分析、评价与比较程序设计语言语法、语义与语用; 程序设计语言教学、教材与课件软件开发过程中各类描述语言; 并发、并行与实时程序设计语言; 面向网络与 Web 的各类语言; 第 4 代语言与数据库语言; 程序设计语言形式化描述技术与方法程序设计语言标准化; 其他各种新型程序设计语言 (包括逻辑型语言、函数型语言等); 写作语言与工具。

二、征文要求

1. 本次会议只接受 Email 投稿。
2. 中、英文稿件均可, 一般不超过 6000 字。为了便于出版论文集, 来稿必须附中英文摘要、关键词、资助基金与主要参考文献, 注明作者及主要联系人姓名、工作单位、详细通信地址 (包括 E-mail 地址) 与作者简介。稿件要求采用 Word 或 PDF 格式。

三、重要日期

征文截止日期: 2006 年 4 月 10 日 录用通知发出日期: 2006 年 4 月 25 日 正式论文提交日期: 2006 年 5 月 10 日

四、联系方式

1. 投稿地址: 东南大学 计算机科学与工程系 周晓宇 (zhouxy@seu.edu.cn);
2. 会务情况: 东南大学 计算机科学与工程系 徐宝文 周晓宇 (zhouxy@seu.edu.cn)。