

# 地震数据的模糊信息处理系统\*

伊藤秀美<sup>1)</sup> 若山晶彦<sup>1)</sup> 干场充之<sup>1)</sup> 冯德益<sup>2)</sup> 林命週<sup>3)</sup>

1) 日本茨城县 305 气象研究所

2) 中国天津 300201 天津市地震局

3) 中国上海 200062 上海市地震局

## 摘要

展示了一种带有信息处理系统的工作站。该系统有3种主要功能：从地震数据中检索信息；地震前兆检测；有关结果的图形显示。由于这些功能中充满了模糊性，因而有必要作适当的清晰化处理。该系统能接受从按阶次排列的单词结构中作出成功选择的指令，而各单词的含义则由相应的从属函数来确定。输出处的模糊度可用肉眼，例如通过颜色来识别。这种设计与上述功能中的动态数据变换相结合，能使本系统的操作更为容易。

**关键词** 模糊信息检索；模糊用户接口；地震数据；地震前兆；软件系统

## 1 前言

在地震学的分析工作中，自然语言起着比在其它自然科学（例如物理学）中更为重要的作用，这在地震预报中即可找到一些典型的例子。由于我们还不能从实质上知道详细的孕震物理条件，因此，目前只能依靠地震前兆（地震在语言学上的似然）来代替求解破裂力学方程。

人们历来以为，只有地震学家才能适当处理模糊信息，而至今在所引进的计算机辅助系统中对模糊性却没有给以更多的注意。然而，地震数据的增加使地震学家们在这种状态下的工作更为困难。

1985年在北京召开了“地震研究中的模糊数学方法”国际学术讨论会。在这次会议上，讨论了在模糊集理论的框架内检测地震专家的知识，包括其在检测地震前兆方面知识的问题（Feng and Liu, 1985）。这些成功的研究扩大了模糊集理论的应用范围；不只是把过去的成果经过人的参与重新改造，而且还引进了像模糊分维这样一些新概念（Feng et al., 1990）。我们认为，现在已经进入把这些研究成果汇编为计算机软件系统的阶段，这也就是本文的主要目的。

从地震预报观点来看，前兆检测仅仅是所需的一种功能。但是，只要我们还没有完全建立起一条地震预报途径，那些可以为地震工作者们提供地震分布图以及各种地质参数

\* 1992年6月1日收到本文初稿，1993年1月21日决定采用。

等的信息检索和用户图形接口等功能就不会失去其重要意义。过去的系统在操作上的笨拙可借助于模糊集理论来消除。

我们所说的地震数据不仅仅指那些诸如震中位置与震级那样直接与地震相联系的数据，而且也包括一些相关数据，如灾害描述，活断层数据等等。为更容易展示起见，我们把注意力主要集中在震中位置和震级数据方面，并引进一个自然工作流程系统，以能看到如何处理模糊性。在第 2 节中我们开始介绍地震数据的模糊结构和检索策略；第 3 节讨论模糊用户接口；最后，在第 4 节中我们将给出一种检索地震前兆的最简单方法。

本软件系统建立在 Sun Sparc Station I, AS 4075 GS 之上。程序用 C 语言和 Fortran 语言编出，并使用 Sun PHIGS 作为三维和二维图形显示库。在模糊信息检索中还使用了一个辅助的数据库管理系统 SYBASE。

## 2 地震数据的模糊结构与信息检索

长时期内的地震位置与震级的资料是地震学的一项最基本财富。在地震仪出现之前的若干时期内，这些资料只能根据历史文献中记载的建筑物破坏点的分布或人感地震强度来估计，因而，在这些资料中必然充满模糊性。我们现在就来分析资料的模糊结构。

首先从宇佐美(1987)的“日本破坏性地震总览”当中取出一个例子，该总览覆盖了 5 世纪以来的各个时期。表 1 的第 2 到第 4 列给出了几次地震的发生时间、地点和震级，其模糊性可用几种不同的方式来表达。例：震级大小用 7.9—8.2 这样的范围之内，或者其值用模糊数  $7\frac{1}{4}$ (在地震学中分数值暗示着模糊数)，或者用像“近似于”这样的方式来表示；对于表 1 中的 130 号地震的位置，还有像  $(142\frac{1}{4}, 41.0)$  或  $(144.0, 40.0)$  这样的多种选择的可能性，第 158 号地震的备注表明，还需要对这次地震本身给出可能性程度，即信度的估计。

表 1 破坏性地震的发生时间、位置和震级

| 编 码 | 地 震 时 间<br>(年·月) | 震 中 位 置<br>( $\lambda(^{\circ}E), \phi(^{\circ}N)$ ) | 震 级  | 备 注 |
|-----|------------------|--|--|-----|
| 130 | 1677 · 04        | $(142\frac{1}{4}, 41.0)$ 或 $(144.0, 40.0)$           | $7\frac{1}{4}-7\frac{1}{2}$ 或 $7\frac{3}{4}-8.0$ |     |
| 149 | 1703 · 12        | $(139.8, 34.7)$                                      | 7.9—8.2  |     |
| 158 | 1711 · 12        | $(134.0, 34.3)$                                      | 近似于 6.7  |     |
| 257 | 1854 · 12        | $(137.8, 34.0)$                                      | 8.4  | 可疑  |

上述模糊结构实质上可以通过马野(1981)提出的可能性分布——模糊关系数据库模型来实现(寺野等, 1987)。为了表达可能性分布，我们可以用最简单的三角形函数

$$f(x; \alpha) = O \vee (1 - \frac{|x|}{\alpha}) \quad (1)^*$$

它的高为 1，底长为  $2\alpha$ 。与 7.9—8.2 相对应的可能分布为  $\bigvee_{7.9 \leqslant \alpha \leqslant 8.2} f(x-a; \alpha)$ ，我们将写为  $1/[7.9, 8.2]$ 。类似地我们可以表达各种模糊数。地震位置的描述需要使用二维可能性分布函数。例如，位置  $(144\frac{1}{4}, 41.0)$  的可能性分布可使用四角锥形函数  $f(x-142.25;$

\* 式中的数学符号“ $\vee$ ”表示“并”的运算，即求  $\max$ ；后面用到的符号“ $\wedge$ ”表示“交”的运算，即求  $\min$ 。

$\alpha \wedge f(y-41.0; \alpha)$  来描述, 记为  $1/(142.25, 41.0)_\alpha$ . 参数  $\alpha$  的值对经度和纬度可取为  $0.3^\circ$  左右, 对震级可取  $0.3$ . 如果用“近似于”来描述震级, 则  $\alpha$  值可大到  $0.4-0.5$ . 地震本身的可能性(信度)可用从属度  $\mu$  来表示. 据此解释即可得表 2.

表 2 用可能性分布——模糊关系数据库模型表达的结果

| $\mu$ | 编 码 | 位 置  | 震 级   |
|-------|-----|--|---|
| 1     | 130 | $0.8/(142.25, 41.0)_{0.4} + (144.0, 40.0)_{0.4}$ | $0.8/[7.25, 7.5]_{0.4} + 0.6/[2.75, 8.0]_{0.4}$ |
| 1     | 149 | $1/(139.8, 34.7)_{0.3}$                          | $1/[7.9, 8.2]_{0.4}$                            |
| 0.3   | 158 | $1/(134.0, 34.3)_{0.3}$                          | $1/6.7_{0.4}$                                   |
| 1     | 257 | $1/(137.8, 34.0)_{0.3}$                          | $1/8.4_{0.3}$                                   |

上述模糊数据库可以通过模糊数据库管理软件或通过对常规的相应数据库管理系统附加一个模糊接口来建立. 我们采用后一种途径. 作出这种选择的理由之一是, 新的模糊数据库与到现在为止已建立的常规地震数据库具有相同的管理系统; 第二个理由是. 要表达各数据项之间的内部关系, 在技术上似乎很复杂. 例如, 在表 1 中第 130 号地震的两种备选位置( $142\frac{1}{4}, 41.0$ )和( $144.0, 40.0$ )必需与其备选震级  $7\frac{1}{4}-7\frac{1}{2}$  和  $7\frac{3}{4}-8.0$  相关联.

在我们所用的途径中, 模糊性按以下方式模式化:  $\alpha$ (数据宽度)和  $\mu$ (地震的可能性)的数值储存在新的数据项内. 多重备选结构通过引进地震编码与地震条款之间的一种至数种对应关系来实现. 首先把地震编码重新复制, 使得每一个编码对于一种地震条款只可能有一个备选结果. 这种途径也适用于灾害数据、前兆数据等等, 如图 1 所示.

至此, 检索策略就简单化了. 例如, 当我们有一个关于论域  $U$  上的查

询从属函数  $\mu_q$  时, 我们可以根据阈值取“硬查询”  $q$ , 其指标函数在  $\mu_q$  的支集上取 1, 反之为 0. 对要检索的数据  $d$  可同样进行清晰化处理; 系统具有专门为从属函数支集两端值进行安排的数据项. 我们可以先对  $q$  作一个初步的原始检索, 然后按下式计算符合从属度:

$$\sup_{x \in U} \mu_q(x) \wedge \mu_d(x) \quad (2)$$

最后, 把地震编码与实际地震等同起来.

当我们处理像描述灾害这一类其它地震数据时, 可使用 Zemankova-Leech 和 Kandell (1984), 以及 Miyamoto (1990) 所讨论的更一般的途径.

### 3 模糊用户接口

如第一节中所强调的那样, 引进语言学的表达式可以简化系统的操作. 虽然地震工

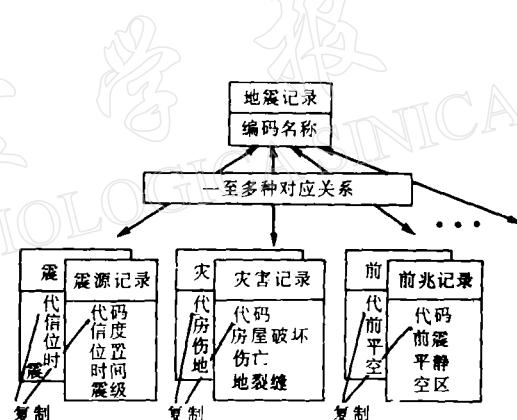


图 1 地震编码与它的条款之间的一种  
至多种对应关系

作者们每天都喜欢用像“深震”和“很大的地震”这一类语言，但他们还是希望能半定量地知道这些词句的含义。这种愿望可以通过使用从属函数来得到满足。

更严密地考察它们在地震学中的用法，可以看出以下途径是可行的。对整个系统的指令是由涉及到几个单项的指令构成的复合体。对每一单项的指令则可通过成功地选择构成从属函数的那些语言学转换词汇来给出。本系统可以在多种指令中考虑接口，并能改变相应的从属函数。如果需要的话，从属函数可以人工调整。

让我们用一个例子来说明上述处理过程。图 2 是一个构成涉及到发震时间、震级、经度、纬度和深度查询的窗口。

假设我们指定一个时期“1980 年中以后”。把它分解为“1980 年的“+”中期“+”以后”（这个顺序在英文中看起来有些笨拙，但在日文和中文中都很自然），我们可从窗口的右边中部显示出按阶次排列的关键词一览表中挑选出这些词来。选出这些词之后，在左边的一个小窗口内就会显示出从属函数，并在其右边注明该从属函数支集端点的值。从属函数的形状可以通过函数图上的两个功能键来改变。

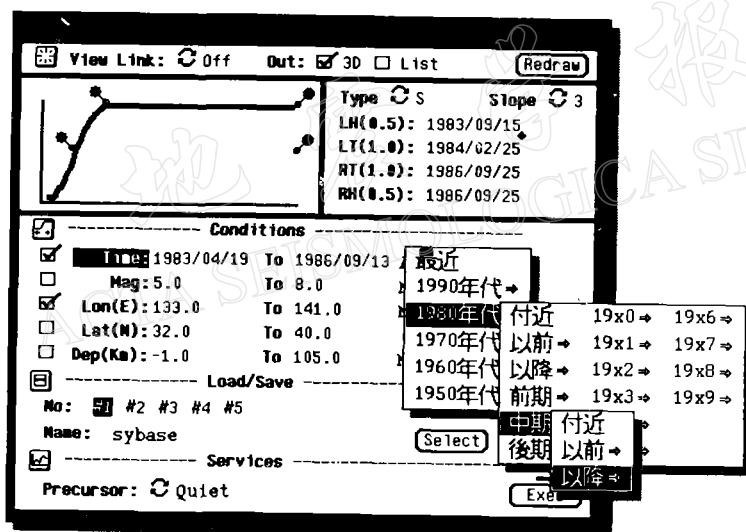


图 2 模糊输入的一个窗口

LH 为左半值, LT 为左全值, RT 为右全值, RH 为右半值, Type 为类型, Slope 为斜率, Conditions 为检索条件, Load/Save 为键入/储存, Services 为服务, Precursor 为前兆, Quiet 为平静(地震活动性)

语言修辞“中期”、“的”和“以后”等应引起从属函数的变形：平移、加宽和类型选择等。我们把这些变形用参量表示为

$$\begin{cases} \mu(x; a, b, \delta_1, \delta_2) = g(x - a; \delta_1) \wedge g(-x + b; \delta_2) \\ a + \delta_1 < b - \delta_2 \end{cases} \quad (3)$$

此处

$$g(x; \delta) = \begin{cases} 0 & x < -\delta \\ (x + \delta)^2 / 2\delta^2 & -\delta \leq x < 0 \\ 1 - (x - \delta)^2 / 2\delta^2 & 0 \leq x \leq \delta \\ 1 & \delta \leq x \end{cases} \quad (4)$$

加宽参数  $\delta_1$  和  $\delta_2$  随着词的组合而改变。例如，对“的”来说，“1980 年中的”词组的  $\delta_1$  和  $\delta_2$  就要比“1980 年的”词组的要小一些。

用类似方法可以定义震级、位置和深度等其它项的从属函数。我们下一步要研究的是各款项中的接口。假定选取“近期的”作为时间项，“关东”作为位置项。“近期的”一词比“1980 年中期”更为含糊，并且要限制在“关东”地区。换句话说，上下文确定的单词含义更明确。“近期的”一词我们可以约定是指最近 10 年左右以来的时期，它包含足够的地震活动序列。本系统通过监测关东地区地震的发生来找出合适的“近期的”一词。

系统的输出用三维或二维图形来显示。通过地震检索与文件之间的对比，很容易看出本系统的图形是先进的。假如像地震位置和震级这样的地震数据的赋值是用关键词来释义的，那么，我们的系统类似于文献检索系统。一个系统若不能显示出所有被检索的文献，它就是笨拙的，因为地震的时间和空间分布都会给我们很重要的信息。

(图 3 为彩色图版，见文末)

图 3 是图形显示的一个实例. 右下部的窗口就是图 2 的窗口, 它给出检索条件. 被检索出的地震示于左边的窗口中. 地震的从属度用不同的颜色标出, 不同颜色代表的标度示于该窗口的左下部. 像旋转、平移、扩大等等图形操作可通过右上部的窗口来进行.

#### 4 地震前兆

地震前兆是主震前的异常变化. 它们是地震预报的决定性因素. 因而, 人们为检测它们已付出很多努力. 典型的地震前兆有前震、地震活动平静异常、空区等等. 前震是正好发生在主震前, 而且在空间上离主震很近的小地震. 地震活动性平静是指在未来主震周围地区地震活动性减弱. 地震空区是一个低地震活动区, 它已积累足以引起主震的能量(藤又護, 1990).

这些概念都是模糊的, 它们只能用自然语言来表达, 而用来定义它们的物理量是不清晰的. 更重要的一点是, 地震前兆是在不同的物理和地质条件下观测到的, 它们的数量也太少, 不易于作常规的统计分析. 持这些观点的研究者们便采用了模糊集理论的途径(Feng et al., 1986).

这一类研究不仅包括了上述常规的地震前兆, 而且也开辟了像模糊分形这样的新领域(Feng et al., 1990). Fractal(分形)一词是由 Mandelbroit 从拉丁文 fractus 移植过来的, 原词义为破碎. 它是借用分维、自相似和幂次定律来分析复杂自然现象的一个典范. 模糊分形是针对带有模糊性的情况所作的一种扩展. 我们把这些结果大都引进了本软件系统当中.

在这一节的剩余部分我们将给出检测地震平静异常的一个最简单的说明书, 它有助于了解所有上述前兆的检测办法.

如果在一个长时期内地震活动性低于其通常的水平, 则可能有大地震. 这里, 我们需要把主震后的几年余震活动时期除开. 这也就是“平静”的模糊性所在(市川政治, 1987).

我们用地震次数来衡量地震活动性的大小. 设  $\lambda$  为地震之间的平均时间间隔. 它可看成一个常数. 把时间分段, 并用  $1, 2, \dots, n$  来编号, 第  $n$  时段代表现在. 下面就来研究第  $m$  时段.

对该时段上的累积地震次数  $N(t)$  与时间  $t$  作回归分析. 设  $\alpha_m$  为回归系数. 如果地震发生的序列是一种泊松过程, 则当  $t \rightarrow \infty$  时有  $N(t)/t \rightarrow \lambda$ . 这样, 我们就可以把地震活动性在第  $m$  时段内为高水平这一事件  $S_m$  的可能性  $s_m$  定义为

$$s_m = g(\alpha_m/\lambda - 1; \delta_1) \quad (5)$$

式中,  $g$  由式(4)给出. 参量  $\delta_1$  具有地震时间间隔起伏的量级, 于是我们取  $\delta_1 = 1/\sqrt{\lambda}$ .

设  $M_l$  为第  $l$  个时段内最大地震的震级, 并有  $l < m$ . 震级 6 可视为大地震的一个阈值, 因而,  $M_l$  是大地震的可能性由下式给出:

$$c_l^1 = g(M_l/6 - 1; \delta_2) \quad (6)$$

在  $M_l$  地震发生后又经过  $m-l$  个时段之后, 第  $m$  个时段属于其余震活动终止期的可能性为

$$c_l^2 = 1 - g[(m-l) - 2L_0]/(L_0 - 1; \delta_3) \quad (7)$$

式中,  $L_0$  为 2.5 年的时间步长, 我们考虑所有从  $t \leq m$  时段来的贡献, 并给出第  $m$  个小时段属于余震终止期的事件  $c_m$  的可能性为

$$c_m = \bigvee_{t \leq m} (c_t^1 \wedge c_t^2) \quad (8)$$

参数  $\delta_2, \delta_3$  可试验性地给定为  $\delta_2 = 0.1, \delta_3 = 0.1$ .

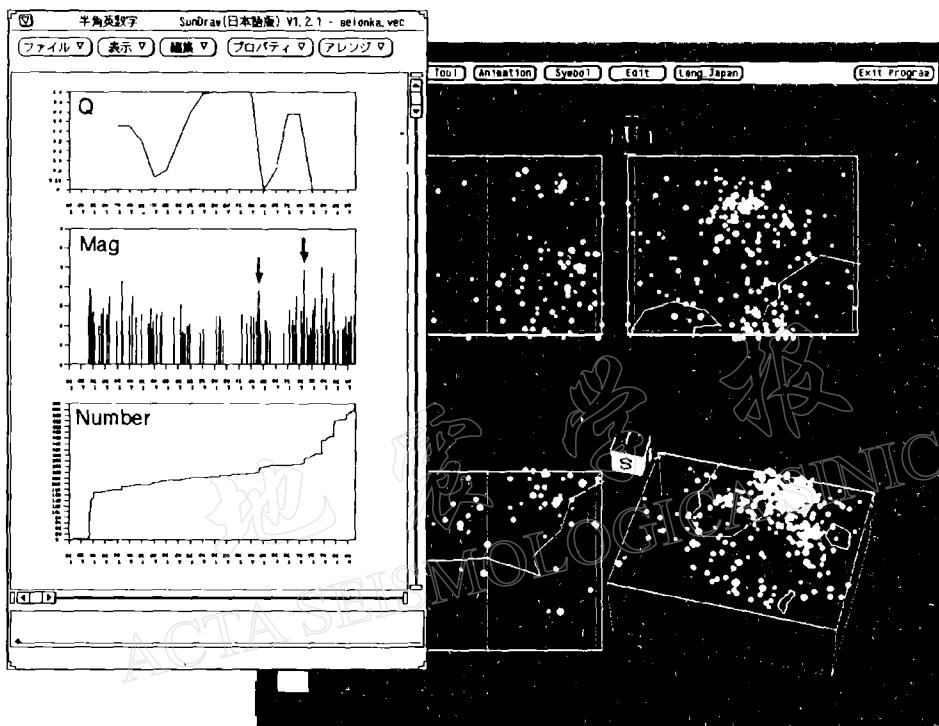


图 4 东海地区观测到的地震活动平静依据

根据我们的定义, 一个平静期持续到大于或等于  $L$  个步长的地震事件  $D_L$  是  $\bigcup_{n-L+1 \leq j \leq n-1} (S_j \cup C_j)$  的补集, 故它的可能性为

$$d_L = 1 - \bigvee_{n-L+1 \leq j \leq n-1} (s_j \vee c_j) \quad (9)$$

当我们有地震事件  $D_L$  时, 就可以估计其平静的条件可能性  $Q^1 = g(L/L_1 + 1; \delta_4)$ , 此处取经验值  $L_1 = 3\delta_2$ ,  $\delta_4 = 0.9$ . 利用此量及  $d_L$ , 我们最后可得出平静的可能性为

$$Q = \bigvee_{L=1} (Q^1 \wedge d_L) \quad (10)$$

图 4 给出利用上述方法对日本东海地区得出的结果. 整个地区可划分为 1, 2, 3, 4 号地区. 右边的窗口示出这 4 个地区内的地震震中分布. 左边给出第 4 号地区(右边下部)地震的  $Q$  值变化图以及相对应的震级时间序列图(中部)和累积地震次数变化图(下部). 由图 4 可以看出, 在大地震(图中箭头所示)发生之前可以观测到平静异常.

## 5 讨 论

本文给出一个拟定的典型工作流程的软件系统的框架. 安排有模糊查询、数据检索、结果的图形窗口显示和地震前兆检测. 同样还讨论了地震数据的模糊数据库结构.

在上述流程中, 按阶次排列的单词的选择指令工作很有效, 它既不构成完整的句子, 但又比简单的单词堆积更精细. 从属函数起着词组与地震现象半定量性质之间的接口的作用. 我们希望许多经验性知识都能沿着这一路线得到处理.

今后, 我们想在结合物理模式的基础上研究地震预报. 这就需要进一步研究, 比方说, 研究模糊环境中的孕震物理模式, 就像物理学家们研究在随机环境中的物理模式一样.

我们感谢气象研究所的冈田先生对地震前兆检测方法给予的启发性讨论. 此项研究部分地得到日本科技厅“模糊系统及其在人文与自然科学中的应用”项目的资助. 伊藤秀美还感谢日本气象厅横田崇博士在本研究工作早期所作的贡献.

## 参 考 文 献

- 马野元秀, 1981. Fuzzy 関系代数によるfuzzyデータベースからの検索. 电子通信学会技術報告, 81, 18, 33—40.
- 勝又謹, 1990. 地震前兆現象データベース. 気象研究所技術報告, 37.
- 市川政治, 1987. 地震前兆現象について2,3問題. 気象研究所研究報告, 38, 3, 203—210.
- 寺野寿郎、浅居喜代治、菅野道夫, 1987. フジイシステム入門. 木々社, 东京 (中译本: 模糊系统理论及应用, 刘维仲、徐萍、姜延宏译, 天津大学出版社, 1991).
- 宇佐美龍夫, 1987. 日本被害地震総覧. 东大出版会, 东京.
- Feng, D. Y. and Liu, X. H. (ed.) 1985. Fuzzy mathematics in earthquake researches. Proc. Int. Symp. on Fuzzy Mathematics in Earthquake Researches, 16—31. Seismological Press, Beijing.
- Feng, D. Y., Lin, M. Z., Jiang, C. and Chen, R. H., 1986. Applications of fuzzy mathematics in earthquake prediction. J. Phys. Earth, 34, Suppl., 213—221.
- Feng, D. Y., Jiang, C., Liu, X. L., Lin, M. Z., Ito, H., Yokota, T. and Yoshida, A., 1990. A study on fuzzy self-similarity and fuzzy fractal dimension of seismicity in China and in Japan. In: Yamakawa, T. (Editor), Proc. Int. Conf. on Fuzzy Logic & Neural Networks, 291—294.
- Miyamoto, S., 1990. Fuzzy Sets in Information Retrieval and Cluster Analysis. Kluwer Academic pub., Dordrecht/Boston/London.
- Zemankova-Leech, M. and Kandel, A., 1984. Fuzzy Relational Data Bases, A Key to Expert Systems, Verlag TÜV Rheinland GmbH, Köln.

