第20卷 第6期 大 气 科 学 Vol. 20. No. 6
1996 年 11 月 SCIENTIA ATMOSPHERICA SINICA Nov. 1996

多元聚合分类方案

吴红光

(南昌气象学校，南昌 330043)

摘 要 取得对样品（或逆分析中的要素）的客观分类（型），对于提高天气预报的效果有着十分重要的意义，遗憾的是目前尚未见到能够获得合理分类结果的可行客观分类方法。本方案是以聚合为策略，以计算机为工具，通过逐次全面比较所给样本资料各种情况下的分类，试图在这方面进行的一项尝试，方案目前已取得下列成果：①提供出一种新的分类思路和方法；②分类客观且有较好的效果；③对小样本可确保取得合理分类。

关键词 合理分类 划分 聚合

1 引言

随着科学和探测技术的飞跃发展，目前获取的资料大大增加。在预报中，要全面有效地应用这样多的信息，就要求对资料进行适当的分类，而目前还没有一种可为大家所接受的完善的客观分类方法，这就影响了预报的效果。

本方案以逐次全面比较所给样本资料各种情况下的分类为思路，以使分类结果类内的样品达到最相似（同质），类与类间的样品达到最相异（异质）为原则，以计算机为工具，对取得样品的客观合理分类进行了尝试。

2 方案设计

给定N个样品，对每个样品观测M个要素（特征），问：用什么方法可以将它们聚合成若干个可以定义的类，并使分类的结果客观合理?

阳含熙和卢泽愚在合著的《植物生态学的数量分类方法》中[1]，介绍了一种全面比较的多元划分法，即对于给出的样品逐次划分成两个子组，每次划分均全面比较分成两个子组的所有可能分法，从中找出使子组间相异最大的划分来。例如，最先一次划分，有N个样本，将它分成两个子组的所有可能方式数为$(C_N^1+C_N^2+\cdots+C_N^{N-1})/2=[(C_N^0+\cdots+C_N^{N-1}+C_N^N)-C_N^0-C_N^N]/2=(2^N-2)/2=2^{N-1}-1$，一旦分成了有$N_1$和$N_2$个样本的两个子组后，对它们还要进行同样的分法，又需计算$2^{N_1-1}+2^{N_2-1}-2$次，如此进行下去……。该方案由于是对样品逐次分成两个子组的所有分法进行全面比较，可保证找出类间相异性最大的划分，分类结果客观合理。但它的计算量太大，一般只宜处理 16 个以下的样本，样本再大，则在计算上就较难实现。例如，对于 40 个样本，其初次分类数就有 $2^{39}-1$ 种之多。故该方案没有什么实用价值。

1994-12-07 收到，1995-08-28 收到再改稿

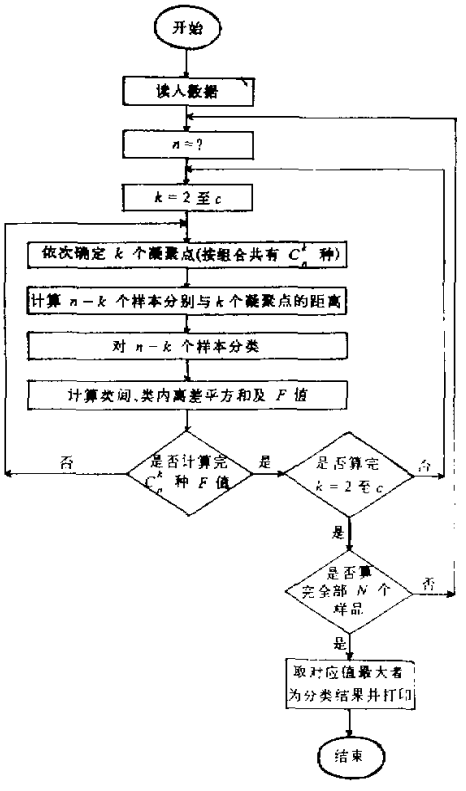

图1 计算机操作程序

那么我们能否解决该方案计算量的问题，使之能在现实条件下得以实施呢?

上面方案是从总体样本开始，进行全面二级划分，然后再对划分出来的两个子类依次逐次进行二级划分的。由于该方案是属于数量分类中的划分方案，是从总体样本N开始的，首次划分要比较$2^{N-1}-1$种分法，所以，一遇到大样本，立刻就进行不下去了。但如果改成聚合方案，从总体样本中先取出实际可能实施的适当数量样品，例如取n个，对这n个样品的所有可能分类进行全面比较，得出相对于这些样品的合理分类，例如为三类，将这三类的类中心分别作为新的样品替代原样品，则n个样品就缩减为三个，再将这三个样品放到下步中与后面取出的适量样品一道参加聚合分类，按上作法，逐次缩减，就可以实现对大样本资料的分类。操作思路见图1。图1中C一般取为$n/2$，亦可根据实际情况取定。F可按综合反映类内同质、类间异质来设定。

3 应用举例

为了说明其应用，我们举一个用该方案进行天气分类的例子。

资料为江西省气象台1980年2月29天的实测资料，取每日14时的地面气压、气温、湿球温度、水汽压、温度露点差和风向作为要素。对地面气压和风向作技术处理，气压取为实测值-1000；风向从0至16取17个值，具体见表1。

另外4个要素均取观测原值。共取得6个要素29个样品，按本方案计算，得最后分类结果为下述五类（此时$F=9.85$[1]为最大）。

各类与天气的对照关系如下：

1) 此处设$F=\frac{Q_{IT}/f_{IT}}{Q_{IN}/f_{IN}}$。式中$Q_{IN}=\sum_{i=1}^{k}\sum_{j=1}^{m}\sum_{l=1}^{n_{ij}}(x_{il}-\bar{x}_{ij})^2$，$Q_{IT}=\sum_{i=1}^{k}\sum_{j=1}^{m}\sum_{l=1}^{n_{ij}}(\bar{x}_{ij}-\bar{x}_j)^2$分别为类内、类间离差平方和，$f_{IN}=n-b$，$f_{IT}=b-1$分别为类内、类间自由度，$b$分类数，$m$因子数，$n_{ij}$系分类号为$i$时的样本数

表1　风向的取值

静风	NNE	NE	ENE	E	ESE	SE	SSE	S	SSW	SW	WSW	W	WNW	NW	NNW	N
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16

第1类，有11个样品，对应为2月1～11日，属冰雪天气，月报表的W天气现象栏上10天有结冰记录，含7天微弱降雪记录。

第2类也是11个样品，对应为2月12～22日，属弱降雨天气，有9天降雨记录，最大日降雨量为14.7 mm。

第3类有5个样品，对应为23、24、26～28日，属雨雾天气，有4个降雨、4个雾天记录，其中3天同日出现了雨和雾。

第4类只有1个样品，对应为25日，出现了该月最大的降雨，日降雨量为24.8 mm，远远大于该月次大雨日14.7 mm。

第5类也只有1个样本，对应为29日，为雷雨天气，日降雨量为6.6 mm。

由上对照可见，各类与天气之间的对应关系明显，反映了类内同质、类间异质，取得了较理想的分类效果。

4　讨论和结论

本方案的主旨实际在于寻找一种可以实现文献[1]方案的算法。它是通过改变分类策略，以聚合代替划分，使之可以从部分样本开始，逐步完成对全体样本的分类。该方案由于在样本缩减过程中，用类中心替代原类样品会产生一定的中心位移，目前尚不能保证对大样本（N）的分类结果一定达到类间相异性最大。但是这种位移作用并非一定不能减小甚至消除，这就为最终取得对大样本的客观合理分类指出了一条途径。该方案就目前来说还取得了下列成果：①提供出一种新的进行客观分类的思路和方法。②克服了现行一些分类方法的弊病，保证分类结果的客观性，有比较满意的分类效果。③对于一次就可实现计算的小样本（n），可确保取得客观合理的分类结果。

关于n的取法，在计算条件可以实现的情况下取最大值。因为在逐次用类中心替代原类样品的中心有位移影响，循环次数过多对最后分类结果会有不良作用。

本方案编成程序后，在计算机上易于实行。该方案对于气候的分型及生物学、地质学等的数据分类同样适用。对于要求分类数很多的情况，亦可采用该方案分步进行，即先采用该方案获得较少的类，对获得的各类分别采用该方法继续往下再分。

参 考 文 献

1　[1]含熙、卢泽愚，1983，植物生态学的数量分类方法，北京：科学出版社，1～84.
2　[英]M. Kendall，1983，多元分析，北京：科学出版社，38～58.
3　国家气象局，1989，气象站天气分析和预报，北京：农业出版社，332～340.
4　[英]P. 史尼斯、[美]R. 索卡尔，1984，数值分类学，北京：科学出版社.

Multiple Aggregation Classification Method

Wu Hongguang

(*Nanchang School of Meteorology, Nanchang* 330043)

Abstract Classifying samples into types is very important to the correctness of weather forecasts. It is the pity that there is probably no classificational methods which can be used to obtain a satisfactory result. The present paper attempts to have a try in this field by aggregating, computing and comparing various classifications with the given data. The results show that (i) it is a new type of classificational methods and (ii) it is objective and can be used to obtain fairly good results especially with small sample size.

Key words rational classification dividing aggregation

《大气科学》获1996年度中国科学院优秀科技期刊二等奖

根据《中国科学院优秀自然科学期刊奖暂行条例》的有关规定，近日进行了1996年度中国科学院优秀科技期刊评比活动。中国科学院优秀期刊评选委员会对申请参评的期刊进行了认真评比，共评选出1996年度中国科学院优秀科技期刊77种，其中一等奖14种，二等奖26种，三等奖37种。《大气科学》获二等奖。

为参加1996年度中国科学院优秀科技期刊评比活动，广大读者、作者为编辑部提供了有价值的信息和意见，对本刊的参评起了重要作用，对此我们表示深深的谢意。

通过这次评比使我们对《大气科学》存在的问题有进一步的了解和认识，我们要认真向兄弟刊物学习，努力解决办刊中出现的各种问题，在广大读者、作者的支持下把《大气科学》办得更好，完成时代赋予我们的任务。

（《大气科学》编辑部）