

基于模糊预测对象的证据权方法及其在土地沙漠化评价中的应用

张生元^{1,2,3}, 武强¹, 成秋明^{3,4}, 葛咏^{3,5}

1. 中国矿业大学资源与安全工程学院, 北京 100083
2. 石家庄经济学院经贸学院, 河北石家庄 050031
3. 中国地质大学地质过程与矿产资源国家重点实验室, 湖北武汉 430074
4. 加拿大约克大学地球与空间科学工程系, 多伦多 M3J1P3
5. 中国科学院地理科学与资源研究所资源与环境信息系统国家重点实验室, 北京 100101

摘要: 为了使在地理信息系统中被广泛用于点事件预测的证据权方法能对面事件进行评价和预测, 提出了一种新的基于模糊训练层的证据权方法. 它是一种更广泛的证据权方法, 与普通证据权方法所不同的是, 它的训练层是模糊集合, 其取值是它的隶属度. 通过适当的变换也可以把点训练层转换为模糊集合. 因此, 该方法可以对面事件、点事件和线事件进行评价和预测. 该方法可以处理训练层和证据层均为模糊集合的情况, 被称为双重模糊证据权方法. 作为该方法的一个应用实例, 本文介绍毛乌素沙漠边缘的晋陕蒙地区土地沙漠化评价的应用实例.

关键词: 证据权方法; 面事件; 点事件; 模糊预测对象; 沙漠化评价; 训练层.

中图分类号: P628

文章编号: 1000-2383(2006)03-0389-05

收稿日期: 2005-08-12

Weights of Evidence Method Based on Fuzzy Training Layer and Its Application in Desertification Assessment

ZHANG Sheng-yuan^{1,2,3}, WU Qiang¹, CHENG Qiu-ming^{3,4}, GE Yong^{3,5}

1. School of Resource and Safety Engineering, China University of Mining & Technology, Beijing 100083, China
2. School of Economics and Commerce, Shijiazhuang University of Economics, Shijiazhuang 050031, China
3. State Key Laboratory of Geological Processes and Mineral Resources, China University of Geosciences, Wuhan 430074, China
4. Department of Earth and Space Science and Engineering, York University, Toronto M3J1P3, Canada
5. State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China

Abstract: This paper proposes an extended version of the fuzzy weights of evidence method, based on the fuzzy training layer for assessing and predicting the occurrence of area-events. It can be considered as a generalization of the ordinary weights of evidence method, which is used to predict the occurrence of point-events with known evidences and training layers. Unlike the ordinary weights of evidence using point training sets, the new method involves training data as a fuzzy set. Point training data can be converted to a fuzzy set, therefore the new method can predict both point-events and area-events. When both evidence and training data are fuzzy sets, the new method acts as a dual fuzzy weights of evidence method. A case study has been used to demonstrate the application of the method in assessing desertification in the area in the contact of Shanxi and Shaanxi Provinces and Inner Mongolia Autonomous Region.

Key words: weights of evidence method; area-event; point-event; fuzzy set; fuzzy object; desertification; training layer.

0 引言

证据权方法是一种贝叶斯方法,20 世纪 80 年代末以来被广泛用于矿产资源评价(Bonham-Carter *et al.*, 1988, 1990; Agterberg, 1989; Agterberg *et al.*, 1990; Bonham-Carter, 1994). 随着该方法应用的不断深入,特别是与 GIS 的结合,如 Arc-WofE(Kemp *et al.*, 1999),使得该方法得到了进一步的发展(Cheng and Agterberg, 1999; Raines *et al.*, 2000; 成秋明,2003; Cheng, 2004). 但到目前为止,证据权方法仅被用于点事件的预测.可是在实际问题中不仅仅是点事件的预测,经常会涉及到面事件的预测,例如区域环境污染预测、沙漠化预测等,有时也会遇到线预测问题.在应用证据权方法进行点事件预测时,为计算先验概率和后验概率,必须对研究区进行单元网格划分,使得训练点落入某些单元网格中.当把研究区进行单元网格划分并且将训练层与其进行叠加的时候,研究区中的单元网格按照是否包含训练点分成 2 类,即包含训练点和不包含训练点.在这种意义上,这些单元网格构成了一个面图层,其取值为 1 或 0,1 表示单元格中包含训练点,而对应单元格的面积可以认为是训练点代表的面积.这个面图层和训练层是等价的,用以代替训练层.当把这个面图层进一步推广到模糊集合时,针对这种模糊训练层,本文提出了基于模糊预测对象的证据权方法.

1 基于模糊预测对象的证据权方法

1.1 基本方法

证据权方法是一种贝叶斯概率准则下的对数线性模型,基本假设是各个证据层关于预测事件是相互独立的.用 D 表示所研究的事件, S 是研究区域.为了表示方便仍用 D 表示在研究区 S 内已知的事件 D 出现的训练层.事件 D 的类型决定训练层的类型,对于普通证据权方法,由于其预测的是点事件,其训练层由已知的点组成.对于本文所研究的模糊预测对象,其训练层的类型是面图层或栅格数据图.设 $\mu_D(x)$ 表示训练层 D 在位置 x 的值, $x \in S, 0 \leq \mu_D(x) \leq 1, \mu_D(x)$ 值越接近 1,表示事件 D 在位置 x 已经出现的程度越高,如果 $\mu_D(x) = 1$,表示事件 D 在位置 x 已经出现, $\mu_D(x) = 0$,表示事件 D 在位置

x 没有出现.因此,可以把训练层看作是关于事件 D 的模糊集合, $\mu_D(x)$ 是其隶属函数.设 X_1, X_2, \dots, X_n 是与 D 相关的证据层.证据层通常是分类数据或有序数据,每个层有 $k=1, 2, \dots, m_i$ 个离散的分类值或状态. X_{ik} 表示第 i 个证据层的第 k 个分类值.

由模糊概率的定义(王新洲等,2003),可导出 D 在 S 中出现的模糊先验概率为:

$$P(D) = A(D)/B(S).$$

其中, $A(D) = \int_{x \in S} \mu_D(x) dx$, $B(S)$ 是研究区面积.

在 X_{ik} 的条件下, D 出现的模糊条件概率为:

$$P(D | X_{ik}) = \frac{A_{ik}}{B_{ik}}.$$

其中, B_{ik} 表示第 i 个证据层取值为 X_{ik} 时的面积,而

$$A_{ik} = \iint_{x \in S \cap (X_i = X_{ik})} \mu_D(x) dx.$$

同理可导出下列几个模糊条件概率:

$$P(\bar{D} | X_{ik}), P(D | \bar{X}_{ik}) \text{ 和 } P(\bar{D} | \bar{X}_{ik}).$$

由加权定义(Bonham-Carter *et al.*, 1990),可导出第 i 个证据层取值为 X_{ik} 时的模糊加权值:

$$W_{ik}^+ = \ln \frac{P(X_{ik} | D)}{P(X_{ik} | \bar{D})} = \ln \left(\frac{A_{ik}}{B_{ik} - A_{ik}} \cdot \frac{B(S) - A(D)}{A(D)} \right),$$

$$W_{ik}^- = \ln \frac{P(\bar{X}_{ik} | D)}{P(\bar{X}_{ik} | \bar{D})} = \ln \left(\frac{A(D) - A_{ik}}{(B_i - B_{ik}) - (A(D) - A_{ik})} \cdot \frac{B(S) - A(D)}{A(D)} \right).$$

在上述定义中,如果分子或者分母之一为 0,则无法计算加权值.为了克服这一问题,对其进行如下的修正:

$$W_{ik}^+ = \ln \left(\frac{A_{ik} + 0.5}{B_{ik} - A_{ik} + 0.5} \cdot \frac{B(S) - A(D)}{A(D)} \right),$$

$$W_{ik}^- = \ln \left(\frac{A(D) - A_{ik} + 0.5}{(B_i - B_{ik}) - (A(D) - A_{ik}) + 0.5} \cdot \frac{B(S) - A(D)}{A(D)} \right).$$

可以证明如果各证据层关于 D 是条件独立性的,事件 D 的后验 logit 值为:

$$\ln O(D | X_{i_1 k_1} X_{i_2 k_2} \dots X_{i_n k_n}) =$$

$$\sum_{j=1}^n W_{ij k_j}^+ + \ln O(D).$$

当证据层有多个分类值时将会导致更大的噪声

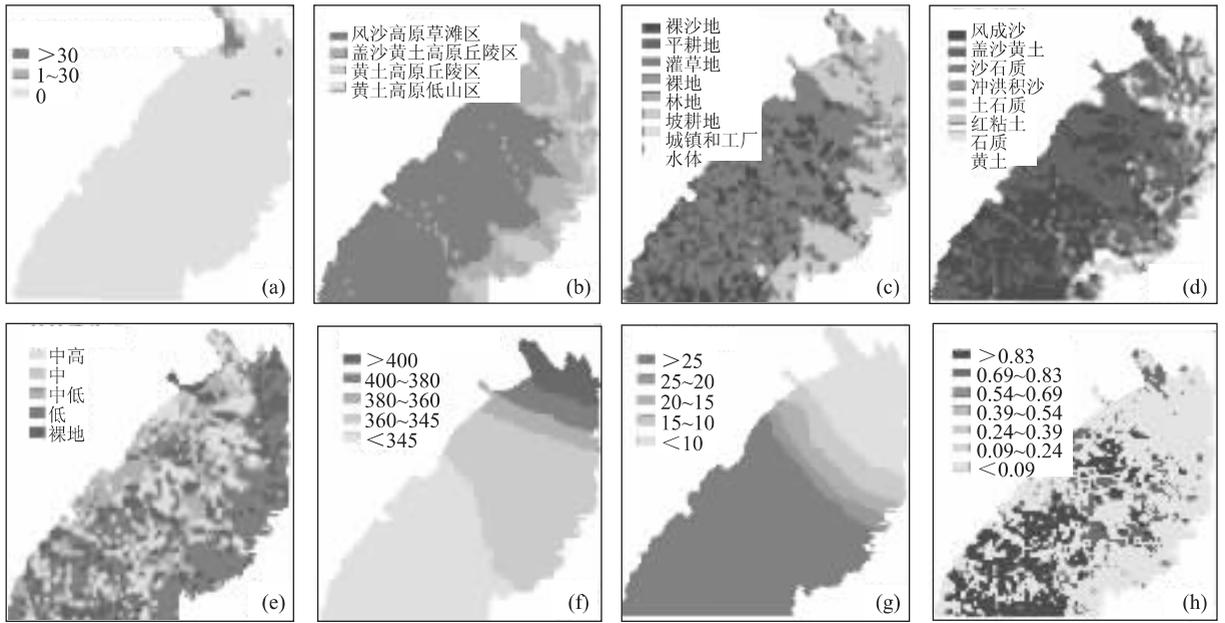


图 1 7 个证据层和作为训练层的土地沙漠化隶属度

Fig. 1 Seven evidence layers and the fuzzy membership function of training layer

a. 矿山利用面积(km²); b. 地貌类型; c. 土地利用类型; d. 母质类型; e. 植被覆盖度; f. 年平均降水量(mm/a); g. 年风暴次数; h. 训练层的沙漠化隶属度

干扰,影响加权值估计的可靠性(Bonham-Carter *et al.*, 1990). 为此,在证据权方法中通常将证据层重新分为 2 类或 3 类,以便获得更稳健的证据权估计. 但对于连续的证据层,这种重新分类一般会产生部分信息损失. Cheng and Agterberg(1999)提出了模糊证据权方法,该方法不需要对证据层进行重新分类,而是把证据层看作与事件 D 有关的模糊集合,定义了模糊概率、条件模糊概率以及模糊证据权的计算模型. 当训练层和证据层都是模糊集合时,作者将模糊证据权方法扩展为双重模糊证据权方法.

1.2 双重模糊证据权方法

设 $\mu_X(x)$ 是某个证据层 X 的隶属函数, $A_1 = \{x; \mu_X(x) = 1\}$, $A_2 = \{x; \mu_X(x) = 0\}$, 模糊证据层的加权公式(Cheng and Agterberg, 1999)为

$$W_{\mu_X(x)} = \log \frac{P(\mu_X(x) | D)}{P(\mu_X(x) | \bar{D})}$$

其中:

$$P(\mu_X(x) | D) = (1 - \mu_X(x))P(A_2 | D) + \mu_X(x)P(A_1 | D),$$

$$P(\mu_X(x) | \bar{D}) = (1 - \mu_X(x))P(A_2 | \bar{D}) + \mu_X(x)P(A_1 | \bar{D}).$$

与模糊证据权方法不同的是,条件概率 $P(A_1 |$

$D)$, $P(A_2 | D)$, $P(A_1 | \bar{D})$, $P(A_2 | \bar{D})$ 是基于模糊训练层的模糊条件概率.

1.3 将点事件转换为模糊事件

用基于模糊训练层的证据权方法也可以对点事件、线事件进行预测,但不需要对研究区域进行单元网格划分,只需将训练层按照一定的方法转换为具有面积属性的模糊集合. 以矿产资源评价为例,可以考虑如下的转换方法:

(1)等面积转换. 如果仅知道矿点的位置,对矿点的其他属性知道得很少,可以通过空间分析的缓冲区方法使每个点生成具有相同面积的面积单元. 也可进行多层缓冲,使得不同缓冲层之间具有不同的赋值,其值在 0~1 之间.

(2)按矿床规模. 如果知道矿床的规模,首先将其按照规模大小进行分类,分为特大型、大型、中型和小型矿床,相同类型的矿床赋予相同的面积,不同类型的矿床按照其规模大小赋予不同的面积,具体可通过空间分析的缓冲区方法实现.

(3)按矿床已探明或预测的面积. 如果已经知道矿床的已探明或预测面积、规模以及产状,可按照空间分析的加权缓冲区方法为每一个矿床点生成具有相应面积的区域.

表 1 用 GeoDAS GIS 得到的植被覆盖度的统计结果

Table 1 Statistics of the degree of vegetation coverage calculated by means of GeoDAS GIS

植被覆盖程度	分类	$B(i)$	$A(i)$	W^+	$S(W^+)$	W^-	$S(W^-)$	C	STDEV	T-Value
较好	1	21 163	5 172.50	-0.35	0.04	0.04	0.01	-0.40	0.04	-9.40
好	2	22 525	4 730.00	-0.55	0.04	0.07	0.01	-0.62	0.04	-14.35
一般	3	42 613	11 023.75	-0.28	0.03	0.08	0.01	-0.35	0.03	-11.44
差	4	68 625	23 482.50	0.12	0.02	-0.07	0.02	0.19	0.03	7.58
裸地	5	28 088	13 897.50	0.75	0.03	-0.15	0.01	0.91	0.03	27.51

$B(i)$. 各个分类的面积; $A(i)$. 隶属函数在各个分类的积分; W^+ . 正加权值; W^- . 负加权值; $S(W^+)$. W^+ 的标准差; $S(W^-)$. W^- 的标准差; $C=W^+-W^-$; STDEV. C 的标准差; T-Value. 显著性水平.

2 应用实例——晋陕蒙地区土地沙漠化评价

研究区位于鄂尔多斯高原东南部,毛乌素沙漠与陕北高原接壤地带,地貌单元以长城为界,基本分为两大类:南及东南部为黄土梁峁丘陵区,区内沟壑纵横,地面起伏不平;西及西北部为沙漠滩地,沿陕蒙边界展布,沙丘累累,地形平坦,是正在发展的沙漠化地区.

2.1 数据准备

通过收集资料、野外踏勘和征求有关方面专家的意见,将研究区的土地沙漠化灾害因子分为自然因子和人为因子.自然因子包括:地貌类型(地形)、土地利用类型、降雨量大小(不同年度)、土壤类型(表土成分)、风力(起风沙次数/年).人为因子包括:矿山开发面积、植被覆盖.

植被覆盖度、土地利用类型、地貌类型和母质类型 4 个评价层的基础数据以及沙漠化现状基础数据,是通过 1999 年 TM543 波段假彩色卫片合成图像解译并最后通过野外踏勘验证获得;降雨强度数据是以研究区 1999 年七、八、九 3 个月降雨量为主而进行全年统计计算获得;大风强度评价层数据是通过研究区 1999 年的平均风速计算得到;矿山开发面积评价层数据是由研究矿区的总体开发设计规划获得;训练层是由 1999 沙漠化现状基础数据经过土地沙漠化危险度评估得到.7 个证据层和模糊训练层如图 1 所示.

2.2 用证据权方法进行评价和预测

为了用基于模糊预测对象的证据权方法对研究区的土地沙漠化进行评价,首先需要计算各个证据层的每个分类值的统计结果(Cheng and Agterberg, 1999; Cheng and Zhang, 2002),包括 W^+ 、 $S(W^+)$ 、 W^- 、 $S(W^-)$ 、 C 、 $S(C)$ 以及 T-value. 表 1 是

表 2 各个证据层的隶属度及其模糊证据权

Table 2 Fuzzy membership function, fuzzy weights and standard deviation for each evidence

证据层	分类	隶属度	模糊权	标准差
植被覆盖度	较好	0	-0.442	0.029
	一般	0	-0.442	0.029
	比较差	0.21	-0.235	0.024
	差	0.53	0.106	0.021
	裸地	1	0.731	0.030
土地利用类型	裸沙地	1	1.992	0.052
	平耕地	0.8	0.531	0.043
	灌草地	0.6	-0.199	0.037
	裸地	0.4	-0.775	0.035
矿山开采(面积)	林地、主要城镇及工矿用地、水体	0	-1.937	0.047
	未开采	0	0.020	0.010
	开采面积 $\leq 30 \text{ km}^2$	0.5	0.020	0.070
母质类型	开采面积 $> 30 \text{ km}^2$	1	0.130	0.140
	黄土、石质、红黏土、土石质	0	-3.440	0.120
	冲洪积沙	0.3	0.050	0.090
	沙石质	0.4	0.390	0.070
	沙土质	0.5	0.660	0.060
降雨强度(mm/a)	盖沙黄土、风成沙	1	1.700	0.020
	> 380	0	-1.369	0.058
	380~360	0.3	0.062	0.041
	360~345	0.6	0.427	0.026
大风强度(次/年)	< 345	1	0.652	0.018
	< 10	0	-1.620	0.050
	10~15	0.17	-0.550	0.040
	16~20	0.39	-0.060	0.030
	21~25	0.66	0.230	0.020
	> 25	1	0.440	0.020

经计算得到植被覆盖度的各个分类值统计结果.

由于证据层的分类值都比较多,所以将证据层作为模糊集合,使用双重模糊证据权的方法对研究区域的土地沙漠化进行评价.在该方法中,证据层隶属函数的确定对评价结果影响较大,每个分类值隶属度可以根据表 1 所示的统计结果,运用一定的数学方法,如拟合方法并结合实际问题和相关专业知识确

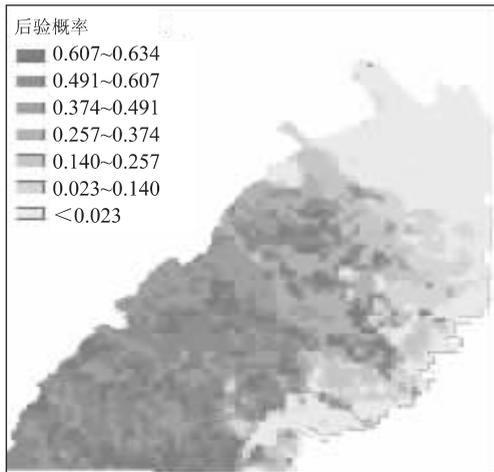


图 2 应用基于模糊预测对象的证据权方法得到的沙漠化后验概率

Fig. 2 Posterior probability obtained by using fuzzy weights of evidence model based on fuzzy training layer

定。以植被覆盖度为例,根据植被的分类值及其与土地沙漠化的关系,首先确定当植被是裸地时其隶属度为 1,当植被的覆盖较好或好时其隶属度为 0,在植被比较差和差时其隶属度为 0~1,将其对比差 C (contrast) 用修正的指数曲线进行拟合,得到其隶属度分别为 0.21 和 0.53(表 2)。

表 2 是各个证据层的隶属度和模糊加权值及其标准差。图 2 是土地沙漠化后验概率分布图,其均值为 0.315,标准差为 0.234。可以根据后验概率进行土地沙漠化危险性分区:当后验概率大于 0.409 时,该区域是沙漠化强度危险区;0.374~0.409 时是沙漠化中度危险区;0.257~0.374 时是轻度危险区;0.140~0.257 是潜在危险区;小于 0.140 是未沙化区。把后验概率分区结果与研究区土地沙漠化现状进行对比,具有较好的一致性。

References

Agterberg, F. P., 1989. Computer programs for mineral exploration. *Science*, 245: 76—81.

Agterberg, F. P., Bonham-Carter, G. F., Wright, D. F., 1990. Statistical pattern integration for mineral exploration. Computer applications in resource estimation prediction and assessment of metals and petroleum. Pergamon Press, New York, 1—12.

Bonham-Carter, G. F., 1994. Geographic information systems for geosciences: Modelling with GIS. Pergamon, Oxford, 1—398.

Bonham-Carter, G. F., Agterberg, F. P., Wright, D. F., 1988. Integration of geological datasets for gold exploration in Nova Scotia. *Photogrammetric Engineering and Remote Sensing*, 54(11): 1585—1592.

Bonham-Carter, G. F., Agterberg, F. P., Wright, D. F., 1990. Weights of evidence modelling: A new approach to mapping mineral potential. Geological Survey of Canada Paper 89-9, 171—183.

Cheng, Q., Agterberg, F. P., 1999. Fuzzy weights of evidence method and its applications in mineral potential mapping. *Natural Resource Research*, 8(1): 27—35.

Cheng, Q., 2004. Application of weights of evidence method for assessment of flowing wells in the Greater Toronto Area, Canada. *Natural Resource Research*, 13(2): 77—86.

Cheng, Q., Zhang, S., 2002. Fuzzy weights of evidence method implemented in GeoDAS GIS for information extraction and integration for prediction of point events. In: Proceedings (CD) of 2002 IEEE Int. Geoscience and Remote Sensing Symposium, Toronto, 1—3.

Cheng, Q. M., 2003. Non-linear mineralization model and information processing methods for prediction of unconventional mineral resources. *Earth Science—Journal of China University of Geosciences*, 28(4): 445—454 (in Chinese with English abstract).

Kemp, L. D., Bonham-Carter, G. F., Raines, G. L., 1999. ArcWofE: Arcview extension for weights of evidence mapping. <http://gis.nrcan.gc.ca/software/arcview/wofe>.

Raines, G. L., Bonham-Carter, G. F., Kemp, L., 2000. Predictive probabilistic modelling using ArcView GIS. *ArcUser*, 3(2): 45—48.

Wang, X. Z., Shi, W. Z., Wang, S. L., 2003. Information process in fuzzy space. Wuhan University Press, Wuhan, 1—280 (in Chinese).

附中文参考文献

成秋明, 2003. 非线性矿床模型与非常规矿产资源评价. *地球科学——中国地质大学学报*, 28(4): 445—454.

王新洲, 史文中, 王树良, 2003. *模糊空间信息处理*. 武汉: 武汉大学出版社, 1—280.