

云推理模型在贵州省粮食产量预测中的应用

张和喜^{1,2}, 迟道才¹, 王群², 房军²

(1. 沈阳农业大学 水利学院, 沈阳 110016; 2. 贵州省水利科学研究院, 贵阳 550002)

摘要:粮食生产是国民经济重要的组成部分,粮食生产的波动必然会引发整个国民经济的波动。基于贵州省1949—2008年粮食总产量数据,采用云推理预测模型动态模拟该省粮食产量变化态势,并以1999—2008年10 a数据与预测数据进行验证。结果表明:云推理模型在粮食产量预测中的应用效果较好,尤其是对趋势的预测效果令人满意,具有较高的实用价值和推广应用的意义。

关键词:云推理; 粮食产量; 预测

中图分类号:F307.11;O29

文献标识码:A

文章编号:1005-3409(2011)02-0184-04

Application of Cloud Reasoning Model to the Yield Prediction of Guizhou Province

ZHANG He-xi^{1,2}, CHI Dao-cai¹, WANG Qun², FANG Jun²

(1. College of Water Resources, Shenyang Agricultural University, Shenyang 110161, China;

2. Guizhou Academy of Hydraulic Sciences, Guiyang 550002, China)

Abstract: Grain production is the important part of national economy development, whose fluctuation will affect national economy security. Based on the grain yield data from 1949 to 2008 of Guizhou Province, cloud reasoning model was used to simulate the change situation of grain yield. And confirmation and forecast of the model was carried out according to the data of 1999—2008. The results indicated that the cloud reasoning model is good in grain yield forecast, especially in the tendency forecast effect, demonstrates the high practical value and can be applied.

Key words: cloud reasoning model; grain yield; prediction

粮食安全始终是关系贵州国民经济发展、社会稳定的全局性重大战略问题,而为了保障贵州中长期粮食安全,其中的任务之一就是要对粮食未来产量做出准确的预测。当前预测粮食产量的方法主要有生产函数模型、多元回归模型和指数平滑模型,但这些预测方法考虑的参数比较单一,忽略一些对预测有用的信息,从而对未来粮食产量不能做出较为准确的预测。粮食产量的形成是一个十分复杂的生物学和生态学过程,粮食产量的高低取决于生产投入、科学技术发展等社会因素的影响,同时又受到气候、环境等自然条件和国家政策调整等多种因素的影响,是多因素综合作用的结果^[1]。

云推理模型是既含有已知信息又含有未知信息考虑参数较为全面的灰色动态数学方法,可以较为准确的预测未来信息。本文通过建立贵州粮食产量预测的云推理模型,探索粮食产量发展的内在规律,从

而对粮食产量做出较为科学准确的预测,为指导粮食生产提供科学的理论依据。

1 云模型

1.1 云和云滴

根据概念的外延特性可分为精确概念和不精确概念。不精确概念的外延是不确定的、定性的,它涉及到两种不确定性:一是随机性,二是模糊性,人们通常用不确定语言值描述不精确概念。模糊数学的引入为处理这些不确定概念中广泛存在的模糊现象提供了研究工具,但是,隶属函数一旦成为精确数值表达后,就不再有丝毫模糊性了。针对当前存在的这些问题,在传统模糊数学和概率统计的基础上建立起来的云模型提出了不确定定量互换模型,它把不确定概念的模糊性和随机性有机地结合在一起,实现了不确定语言值与定量数值之间的自然转化^[2-3]。

定义 1: 设 U 是一个用精确数值表示的定量论域, C 是 U 上的定性概念, 若定量值 $x \in U$, 且 x 是定性概念 C 的一次随机实现, x 对 C 的确定度 $u(x) \in [0, 1]$ 是有稳定倾向的随机数。

$$u: U \rightarrow [0, 1] \quad \forall x \in U \quad x \rightarrow u(x)$$

则 x 在论域 U 上的分布称为云, 每一个 x 称为一个云滴^[4-5]。云具有以下性质。

- (1) 论域 U 可以是一维的, 也可以是多维的。
- (2) 定义中提及的随机实现, 是概率意义下的实现; 定义中提及的确定度, 是模糊集意义下的隶属度, 同时又具有概率意义的分布。所有这些都体现了模糊性和随机性的关联性。
- (3) 对于任意一个 $x \in U$, x 到区间 $[0, 1]$ 上的映射是一对多的变换, x 对 C 的确定度是一个概率分布, 而不是一个固定的数值。
- (4) 云由云滴组成, 云滴之间无次序性, 一个云滴是定性概念在数量上的一次实现, 云滴越多, 越能反映这个定性概念的整体特征。
- (5) 云滴出现的概率大, 云滴的确定度大, 则云滴对概念的贡献大。

1.2 云的数学特征

云用期望 Ex , 熵 En 和超熵 He 三个数字特征来整体表征一个概念。

期望 Ex : 云滴在论域空间分布的期望。通俗地说, 就是最能够代表定性概念的点, 或者说是这个概念量化的最典型样本。

熵 En : 定性概念的不确定性度量, 由概念的随机性和模糊性共同决定。

超熵 He : 是熵的不确定性度量, 即熵的熵。由熵的随机性和模糊性共同决定。

云模型是云的具体实现方法, 也是云运算、云聚类、云推理、云控制等方法的基础。

云可以用图形表示, 简称云图。正态云图及数字特征的意义如图 1 所示^[6]。

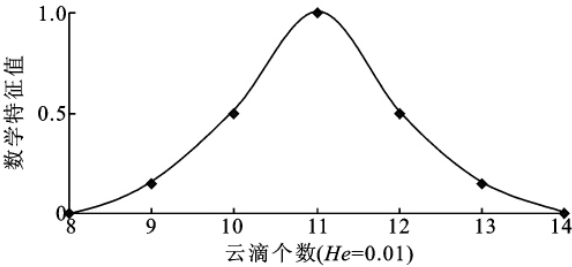


图 1 一维正态云图

2 正态云发生器

2.1 正向云发生器

正向正态云发生器是从定性到定量的映射, 它根

据云的数字特征 (Ex, En, He) 产生云滴。正向正态云的定义如下。

定义 2: 设 U 是一个用精确数值表示的定量论域, C 是 U 上的定性概念。若定量值 $x \in U$, 且 x 是定性概念 C 的一次随机实现, 若 x 满足: $x \sim N(Ex, En'^2)$, 其中, $En' = N(En, He^2)$, 且 x 对 C 的确定度满足

$$U = \exp\left[-\frac{(x - Ex)^2}{2(En)^2}\right]$$

则 x 在论域 U 上的分布称为正态云。

正向正态云发生器的具体算法为:

- (1) 生成以 En 为期望值, He^2 为方差的一个正态随机数 $En_i = \text{NORM}(En, He^2)$;
 - (2) 生成以 Ex 为期望值, En_i^2 为方差的一个正态随机数 $x_i = \text{NORM}(Ex, En_i^2)$;
 - (3) 计算 $u_i = \exp\left[-\frac{(x_i - Ex)^2}{2(En)^2}\right]$
 - (4) 具有确定度 u_i 的 x_i 成为数域中的一个云滴;
 - (5) 重复步骤(1)到(4) n 次, 产生要求的 n 个云滴。
- 其中 NORM 为产生服从正态分布随机数的函数。

2.2 逆向云发生器

逆向云发生器是实现从定量值到定性概念的转换模型^[7]。它可以将一定数量的精确数据转换为以数字特征 (Ex, En, He) 表示的定性概念, 具体算法如下:

- (1) 根据 x_i 计算这组数据的样本均值 $\bar{X} = \frac{1}{n} \sum_{i=1}^n x_i$, 一阶样本绝对中心矩 $\frac{1}{n-1} \sum_{i=1}^n |x_i - \bar{X}|$, 样本方差 $S^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{X})^2$;
- (2) $En = \bar{X}$
- (3) $En = \sqrt{\frac{\pi}{2}} \times \frac{1}{n} \sum_{i=1}^n |x_i - Ex|$
- (4) $He = \sqrt{S^2 - En^2}$

3 云推理

3.1 前件云和后件云发生器

定义 3 如果给定论域 U 中的一个特定点 α , 通过云发生器可以生成这个特定点 α 属于概念 C_1 的确定度分布, 这时的云发生器称为前件云发生器; 给定一确定度 $u, u \in [0, 1]$, 通过云发生器可以生成论域 U_2 中的概念 C_2 上满足这个确定度的云滴的分布, 这时的云发生器称为后件云发生器。

一维前件云发生器如下:

- (1) 根据定性概念的数字特征熵 En 和 He 生成

正态分布的随机数。 $En_i = \text{NORM}(En, He^2)$;

(2) 根据期望值 Ex 和特定输入值 a 计算确定度

$$u = \exp\left[-\frac{(a - Ex)^2}{2(En')^2}\right]$$

一维后件云发生器如下:

(1) 根据定性概念的数字特征熵 En 和 He 生成正态分布的随机数。

$$En_i = \text{NORM}(En, He^2)$$

(2) 根据期望值 Ex 和确定度 u 计算其相应的云滴。 $b = Ex \pm En' \sqrt{-2\ln u}$

3.2 规则发生器

把一个前件云发生器和一个后件云发生器,按照图 2 所示进行连接,可以构造出单条件规则,称为单条件单规则发生器。

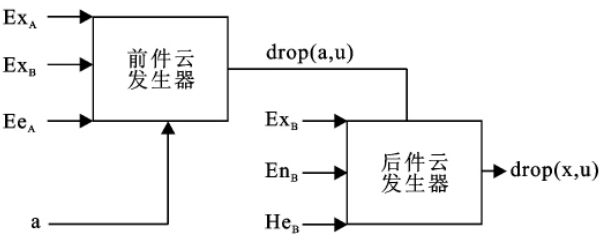


图 2 单条件单规则发生器

规则发生器具体的算法如下:

(1) 根据规则前件 A 的定性概念 $C_1(Ex_A, En_A, He_A)$ 生成正态分布随机数 $En'_A = \text{NORM}(En_A, He_A)$;

(2) 计算特定值 a 的确定度

$$u = \exp\left[-\frac{(a - Ex_A)^2}{2(En'_A)^2}\right]$$

(3) 根据规则后件 B 的定性概念 $C_2(Ex_B, En_B, He_B)$ 生成正态分布随机数 $En'_B = \text{NORM}(En_B, He_B)$;

(4) 如果输入值是规则前件的上升沿,即 $a \leq Ex$, 则规则的后件也选择上升沿,即 $b = Ex_B - \sqrt{-2\ln(u)}En'_B$, 反之, $b = Ex_B + \sqrt{-2\ln(u)}En'_B$

3.3 规则的合并

当两个规则的前件云或后件云相近时,可以进行云合并运算。给定两个相邻的云模型 $C_1(Ex_1, En_1, He_1), C_2(Ex_2, En_2, He_2)$, 令合并后的云模型为 $C(Ex, En, He)$, 则有式(1)、式(2)、式(3)。

$$Ex = \frac{Ex_1 En'_1 + Ex_2 En'_2}{En'_1 + En'_2} \tag{1}$$

$$En = En'_1 + En'_2 \tag{2}$$

$$He = \frac{He_1 En'_1 + He_2 En'_2}{En'_1 + En'_2} \tag{3}$$

3.4 基于云推理的中长期粮食产量预测步骤

步骤 1: 根据 n 年粮食产量序列 $E = (e_1, e_2, e_3, \dots, e_n)$, 得出输入、输出向量 X, Y , 其中, 输入向量的分量 $x_i = e_i, i = 1, 2, \dots, n-1$, 输出向量分量 $y_i = e_{i+1}, i = 1, 2, \dots, n-1$ 。

步骤 2: 将 X 中的数据从大到小排列, Y 中的数据位置做相应的改变。使用模糊聚类方法对 X 中的数据进行聚类。

步骤 3: Y 中数据进行相应的分类, 并根据 X, Y 的聚类情况, 利用逆向云发生器生成 c 组预测云。

步骤 4: 根据当前 t 年粮食产量数据, 利用逆向云发生器, 生成当前趋势云, 并与各个预测云合并, 形成最终的规则的后件云。

步骤 5 利用生成的规则云进行预测。

4 算例

以贵州省 1949—2008 年的 60 a 粮食产量为例, 以 1949—1998 年 50 a 的数据作为模型计算的样本, 用 1999—2008 年 10 a 的数据作为验证模型精度的样本。在云推理时, 考虑相邻年份对未来年份的影响以及当前趋势对未来的影响。

令 1949—2008 年的数据归一为 $E = \{e_1, e_2, \dots, e_{60}\}$ 。将 1949—1998 的粮食产量数据作为输入样本 X , 1950—1999 年的粮食产量数据作为输出样本 Y 。即

$$\begin{aligned} X &= \{x_1, x_2, \dots, x_{50}\} = \{e_1, e_2, \dots, e_{50}\} \\ &= \{0.256, 0.259, \dots, 0.950\} \end{aligned}$$

$$\begin{aligned} Y &= \{y_1, y_2, \dots, y_{50}\} = \{e_2, e_3, \dots, e_{51}\} \\ &= \{0.259, 0.276, \dots, 0.972\} \end{aligned}$$

利用 (x_i, y_i) 推导相邻年份之间的不确定性规则。

将 X 中的数据从大到小排列, Y 中的数据位置做相应的改变。使用某种聚类方法对 X 中的数据进行聚类。由于属于单指标聚类, 许多方法均能取得很好的效果, 使用模糊聚类, 并将数据分为 3 类, Y 中的数据做相应于 X 的分类。这样我们得到了 3 组挖掘规则数据。

$$\begin{cases} (0.255, 0.258, 0.272, 0.275, 0.280, 0.287) \\ (0.258, 0.275, 0.280, 0.297, 0.315, 0.327) \\ (0.315, 0.321, 0.327, 0.354, \dots, 0.610, 0.621, 0.653, 0.649) \\ (0.321, 0.392, 0.354, 0.364, \dots, 0.621, 0.763, 0.512, 0.749) \\ (0.749, 0.763, 0.808, 0.817, 0.863, 0.883, 0.947) \\ (0.808, 0.679, 0.817, 0.863, 0.883, 0.947, 0.969) \end{cases}$$

其中,第一行数据作为输入,利用逆向云发生器生成规则的前件云。第二行数据作为输出,首先利用逆向云发生器生成预测云,同时考虑预测年份前 5 a 的数据,利用逆向云发生器生成当前趋势云,利用公式(1)–(3)进行预测云和趋势云的合并运算,将生成的云作为规则的后件云。

生成的规则用云表示为

$$C_{A1}(Ex=0.273,En=0.014,He=0)$$
$$C_{B1}(Ex=0.292,En=0.026,He=0)$$
$$C_{A2}(Ex=0.481,En=0.105,He=0)$$

表 1 贵州省多年粮食产量数据

年份	粮食/万 t	年份	粮食/万 t	年份	粮食/万 t	年份	粮食/万 t
1949	296.55	1964	455.65	1979	623.05	1994	938.70
1950	299.80	1965	489.60	1980	648.35	1995	948.85
1951	319.05	1966	474.00	1981	567.36	1996	1002.60
1952	344.65	1967	516.30	1982	654.00	1997	1025.90
1953	379.40	1968	485.05	1983	703.00	1998	1100.00
1954	411.05	1969	464.85	1984	757.80	1999	1125.21
1955	426.35	1970	516.45	1985	594.96	2000	1161.30
1956	486.50	1971	557.95	1986	672.34	2001	1100.30
1957	535.60	1972	428.25	1987	673.17	2002	1034.20
1958	524.90	1973	598.95	1988	635.68	2003	1104.30
1959	423.20	1974	543.55	1989	708.25	2004	1149.58
1960	316.15	1975	538.91	1990	721.00	2005	1152.06
1961	324.95	1976	569.70	1991	885.50	2006	1038.00
1962	365.50	1977	666.10	1992	788.90	2007	1100.86
1963	372.30	1978	643.36	1993	869.50	2008	1158.00

表 2 基于云推理的 1999—2008 年粮食产量预测

年份	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
期望	0.926	0.813	0.778	0.830	0.786	0.809	0.833	0.797	0.891	0.951
最大	0.998	0.988	0.999	0.995	0.966	0.964	0.994	0.967	0.937	0.964
最小	0.553	0.508	0.502	0.487	0.478	0.483	0.490	0.598	0.484	0.522
真实	0.968	0.981	0.947	0.890	0.950	0.980	0.972	0.893	0.947	0.927
误差/%	4.33	17.13	17.84	6.74	17.26	17.44	14.30	10.75	5.91	2.58

由表 2 可知,模型预测值比较接近于观测值,误差都在一定范围内,这说明云推理模型可以较好地模拟与预测粮食产量的变化趋势。根据云推理模型处理后数据进行拟合,其趋势一致,预测值见图 3,通过图 3 可知,预测值接近真实值,最大相对误差 17.84%,最小相对误差 2.58%。

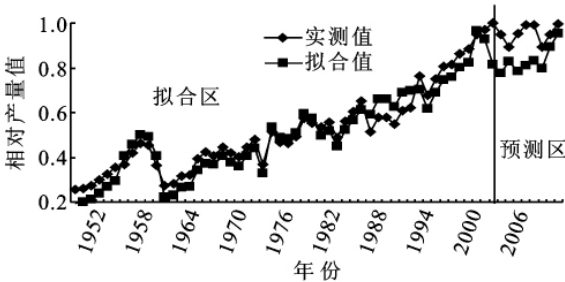


图 3 粮食产量拟合及预测结果

$C_{B2}(Ex=0.491,En=0.113,He=0.015)$
 $C_{A3}(Ex=0.833,En=0.0696,He=0)$
 $C_{B3}(Ex=0.853,En=0.0903,He=0.036)$
式中: C_{Ai} ——前件云; C_{Bk} ——后件云。

根据隶属度最大原则,确定出预测应选用的规则,并根据规则发生器的方法,即可推理出未来年的粮食产量值。由于云推理是一种不确定性的推理,每次计算所得的数值不尽相同,但必然在某一范围之内。因此本文运行程序 10 次,将预测结果的期望值、最大值、最小值列于表 2 中。

5 小 结

本文利用定性定量不确定性之间转换的云模型研究了粮食产量预测的问题,本文建立的判别模型充分考虑了粮食产量和相邻年份的匹配程度,克服了传统判别方法在多年粮食产量预测上的不足。这为粮食产量的预测提供了一种新的方法与途径。通过逆向和正向云发生器算法建立的云推理预测模型,能够充分挖掘数据中不确定的信息。通过预测值与真实数据的比较,预测效果较好。但粮食产量预测仍然还有很多因素需要考虑,因此,将云模型优势与粮食产量特点广泛地结合,尚需在今后做进一步的研究。

析法作为一种纯数理分析方法,数据本身的准确性影响到分区结果的合理性,且该方法具有一定的机械性。故进一步对以上6个类型区进行分析,综合经验判断,从工程模式、工程措施和排灌工程设计来比较,滨海涝渍治理工程类型区和滨海干旱治理工程类型区两者特点不明显,干旱区较干旱,但是工程措施只增加12.7%~17.2%,可是干旱区即旱又涝,涝的严重性却大于涝渍区,排水措施增加11.54%~80.56%。所以,以不分干旱和涝渍,将两个区合并为一个区,滨海旱涝治理类型区为宜。故二级类型区最终调整为五个:滨海平原旱涝治理工程类型区、台地治石工程类型区、低丘干旱治理工程类型区、山间阶地治理工程类型区、河谷平原治涝工程类型区(见图3)。

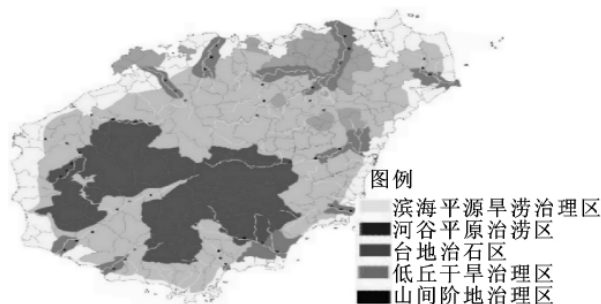


图3 海南省土地开发整理工程类型区划图

2 结论

(1)海南省不同区域土地开发整理的制约性因素不同,土地开发整理影响要素存在较大区域分异性,有必要进行类型区划分研究;(2)不同类型区划分方法均存在一定局限性,类型区不应机械确定;(3)采用图形叠置法,快速聚类法与经验判断法综合分析,海南省土地开发整理一级类型区划分为4个:即滨海平原类型区、河谷平原类型区、环山丘陵类型区与南部山地类型区。二级工程类型区划分为5个:即滨海平原旱涝治理工程类型区、台地治石工程类型区、低丘干旱治理工程类型区、山间阶地治理工程类型区、河谷平原治涝工程类型区。

(上接第187页)

参考文献:

- [1] 杨友孝,蔡运龙. 中国粮食可持续生产条件和结构变化趋势及其影响[J]. 农业现代化研究,2000,21(5):257-260.
- [2] 李德毅,邸凯昌,李德仁,等. 用语言云模型挖掘关联规则[J]. 软件学报,2000,11(2):143-158.
- [3] 宋远俊,杨孝宗,李德毅. 多机多任务实时系统云调度策略[J]. 计算机学报,2000,23(10):1107-1113.

3 讨论

(1)关于类型区成图问题。类型区划分是针对某一土地开发整理工程模式而言。影响类型区划分的各类要素在空间上很难完全叠置,往往在空间上相互交错。由综合要素构成的分类单元,一般没有明晰的空间分布范围与边界,而在划定类型区时往往具有一定面积,成片划定。此外,土地开发整理是以项目为单元进行,故类型区划图在一定程度上仅仅是示意图。

(2)关于类型区边界的确定问题。土地开发整理类型区划区别于区域区划与一般的类型区划,是多因素综合作用的结果,与现存行政边界不可能完全吻合。海南省类型区划分以乡镇为单元,类型区边界的确定在遵循工程模式一致性的基础上,兼顾行政区划的完整性原则。故类型区边界具有一定程度的示意性。即每个类型区有一定的地理分布范围,但不是一个由某个特定边界划分的封闭区域,只是某个特定类型地理位置的界定。

参考文献:

- [1] 蔡海生,陈美球,赵建宁,等. 土地开发整理工程类型区划分的概念与方法探讨[J]. 农业工程学报,2009,25(10),290-295.
- [2] 国家测绘局海南测绘资料信息中心编制. 海南省地图集[M]. 广州:广东省地图出版社,2006:1-5.
- [3] 海南省农业厅土肥站. 海南土壤[M]. 海南:三环出版社,1994:10-25.
- [4] 高素华,黄增明,张统钦,等. 海南岛气候[M]. 北京:气象出版社,1988:41-118.
- [5] 海南省水务局,中国水利水电科学研究院. 海南省水资源综合规划报告[R]. 海口:海南省水务局,2005:12-45.
- [6] 江泽林. 海南省优势农产品区域布局研究[M]. 北京:中国农业出版社,2005:45-68.
- [7] 海南省统计局. 海南统计年鉴(2006)[M]. 北京:中国统计出版社,2007:1-39.
- [4] Li Deyi, Han Jiawei, Shi Xuemei. Knowledge Representation and Discovery Based on Linguistic Atoms[J]. Knowledge-based Systems,1998,15(10):431-440.
- [5] Li Deyi. Knowledge Representation in KDD Based on linguistic Atoms[J]. Journal of Computer Science and Technology,1997,12(6):481-496.
- [6] 茆诗松,王静龙,濮晓龙. 高等数理统计[M]. 北京:高等教育出版社,1998.
- [7] 吕辉军,王晔,李德毅. 逆向云在定性评价中的应用[J]. 计算机学报,2003,26(8):30-33.