

卢耀健, 刘合香, 王萌. 华南台风灾害的风险熵-极限学习机预测模型研究[J]. 灾害学, 2019, 34(4): 216-221. [LU Yaojian, LIU Hexiang and WANG Meng. Research on extreme learning machine-risk entropy prediction model of typhoon disaster in south China [J]. Journal of Catastrophology, 2019, 34(4): 216-221. doi: 10.3969/j.issn.1000-811X.2019.04.037.]

# 华南台风灾害的风险熵-极限学习机预测模型研究\*

卢耀健<sup>1</sup>, 刘合香<sup>1,2</sup>, 王 萌<sup>1</sup>

(1. 南宁师范大学 数学与统计科学学院, 广西 南宁 530001; 2. 广西北部湾海洋灾害研究重点实验室, 广西 钦州 535000)

**摘 要:** 结合广义模糊熵原理和模糊c均值聚类方法构建华南台风灾害风险熵模型, 对华南台风灾害进行风险分析, 讨论其分布情况; 利用灰色关联分析法, 探讨华南台风灾害的灾情因子、致灾源因子分别和灾害风险熵之间的关系以及二者对风险熵的影响程度; 建立基于极限学习机的非线性回归模型, 以多元线性回归和BP神经网络两种方法作为对照组, 进一步探讨风险熵与灾情因子和致灾源因子关系。结果表明, 华南台风灾害风险熵值呈正态分布, 与灾情因子和致灾源因子的灰色关联度分别为0.716 2和0.794 9, 受灾情因子和致灾源因子的影响较大; 利用构建的极限学习机模型预测的华南台风灾害风险熵值平均绝对误差为0.059, 拟合优度为92.82%, 将预测结果与常规的多元线性回归和BP神经网络方法的预测结果进行对比分析, 结果表明, 用构建的极限学习机模型预测华南台风灾害风险熵值, 其性能比常规多元线性回归和BP神经网络方法有明显的改进。

**关键词:** 华南; 台风灾害; 风险熵; 极限学习机; 灰色关联; 模糊聚类

**中图分类号:** X43; X915.5; P458 **文献标志码:** A **文章编号:** 1000-811X(2019)04-0216-06

doi: 10.3969/j.issn.1000-811X.2019.04.037

华南地区(广东、广西和海南)位于我国南部沿海地带, 靠近西北太平洋台风生源地, 常年遭受台风灾害的侵袭<sup>[1]</sup>。张俊香<sup>[2]</sup>等学者通过研究发现, 随着我国经济的发展, 台风造成的直接经济损失呈上升的趋势。因此, 为了有效减少台风灾害带来的损失, 学者们开始尝试利用不同的方法研究台风灾情、强度及路径<sup>[3-6]</sup>。

熵是人们对于某种事物了解程度的度量, 通常信息熵值越大, 其外延越广, 不确定性就会越大<sup>[7]</sup>。目前, 一些学者已经引用熵理论来对自然灾害进行研究。例如, 金菊良<sup>[8]</sup>等结合了广义分布函数和广义熵理论对洪水灾害系统中的不确定性进行描述, 初步形成了洪水灾害风险管理广义熵智能分析理论框架。魏章进<sup>[9]</sup>等利用改进的熵值法对东南沿海省份的台风灾害进行风险评估, 得到各灾情指标权重, 以此来表示各灾情指标对灾情系统的影响程度。宁岩<sup>[10]</sup>等人利用信息熵原理, 计算了深溪沟流域的41个子流域的8个评价指标的信息熵值, 以此来表示每个指标对泥石流系统的影响程度。谭金凯<sup>[11]</sup>等建立了广西洪涝灾害风险熵模型, 探讨了洪水灾害风险熵与灾情因子和致灾源因子的关系, 并利用支持向量机构建灾害风险熵的非线性回归模型。以上研究成果表明熵理论可以有效地应用在自然灾害研究中。同

时发现, 熵理论应用在华南台风方面的相关研究较少, 关于华南台风的灾害风险程度、致灾因子和灾情因子分别与灾害风险熵的关系以及不确定性等方面的研究尚未出现报道, 因此本文将建立华南台风灾害风险熵模型, 分析其与灾情因子和致灾源因子的关系, 并对灾害风险熵与灾情因子和致灾源因子进行非线性回归拟合。

## 1 数据处理计算与分析方法

### 1.1 构建灾情指数和致灾源指数

根据影响因素的比重, 从灾情因子中选取受灾人口 $X_1$ (人)、死亡人口 $X_2$ (人)、倒塌房屋数 $X_3$ (间)、农作物受灾面积 $X_4$ ( $\text{hm}^2$ )、直接经济损失 $X_5$ (亿元)作为灾情系统中的指标; 从致灾因子中选取最大风速极值 $Y_1$ ( $\text{m/s}$ )、降雨天数 $Y_2$ (d)、降雨极值 $Y_3$ (mm)作为致灾源系统的三项指标。由于各项指标单位不同, 数据离差程度较大, 因此为了达到去除量纲, 缩减数据间的离差程度的目的, 需要对原始数据进行预处理<sup>[11]</sup>:

$$X_{ij} = \begin{cases} \log_2 X_{ij} + 1, & X_{ij} > 1; \\ X_{ij}, & X_{ij} \leq 1; \end{cases} \quad 1 \leq i \leq n, 1 \leq j \leq 5. \quad (1)$$

\* 收稿日期: 2019-03-27 修回日期: 2019-05-31

基金项目: 国家自然科学基金(41665006, 11561009); 广西自然科学基金(2018JJA110052)

第一作者简介: 卢耀健(1996-), 男, 广西玉林人, 硕士研究生, 研究方向为概率统计、数学模型、自然灾害风险分析。

E-mail: 304374001@qq.com

通讯作者: 刘合香(1962-), 女, 山东聊城人, 教授, 硕士生导师, 研究方向为概率统计、数学模型、自然灾害风险分析。

E-mail: hx\_post@163.com

$$Y_{ij} = \frac{10Y_{ij}}{\max(Y_j) + \min(Y_j)}, 1 \leq i \leq n, 1 \leq j \leq 3. \quad (2)$$

式中:  $X_{ij}$  表示灾情系统中第  $i$  个样本的第  $j$  项指标的处理结果,  $Y_{ij}$  表示致灾源系统中的第  $i$  个样本的第  $j$  项指标的处理结果, 每个台风样本的灾情指数和致灾源指数计算如下:

$$X_i = \sum_{j=1}^5 X_{ij}, Y_i = \sum_{j=1}^3 Y_{ij}, 1 \leq i \leq n. \quad (3)$$

式中:  $X_i$  表示第  $i$  个台风灾害的灾情指数,  $Y_i$  表示第  $i$  个台风灾害的致灾源指数。

## 1.2 基于模糊隶属度矩阵的灾害风险熵

熵值作为一种能够度量事物不确定性的工具, 其公式如下。

离散型:

$$H(X) = H(p_1, \dots, p_n) = - \sum_{i=1}^n p_i \log_c \left( \frac{1}{p_i} \right) = - \sum_{i=1}^n p_i \log_c p_i. \quad (4)$$

式中:  $p = (p_1, \dots, p_n)$  是随机事件  $X$  的概率分布,

并且满足  $\sum_{i=1}^n p_i = 1$ 。

连续型:

$$H(X) = - \int_a^b f(x) \log_c f(x) dx. \quad (5)$$

式中:  $f(x)$  是关于随机变量  $X$  的概率密度函数, 并且满足  $\int_a^b f(x) dx = 1$ 。在计算过程中一般取  $c$  值为 2, 信息熵的单位为比特 (bit)。

文献<sup>[11]</sup>指出, 在实际应用中, 随机变量的概率分布  $p = (p_1, \dots, p_n)$  与概率密度函数  $f(x)$  在大多数情况下并不知道, 而且以频率代替概率的方法需要在大样本的前提下才得以实现, 不适用台风灾害过程的小样本。所以, 基于模糊聚类隶属度思想<sup>[12]</sup>, 可以转变为研究以下广义熵:

$$H(x) = - \sum \frac{n_k}{N} u_k \log_c \left( \frac{n_k}{N} u_k \right). \quad (6)$$

式中:  $H(x)$  为样本  $x$  的广义熵值,  $N$  表示样本个数, 假设在模糊  $c$  均值聚类下的将  $N$  个样本共分为  $m$  类, 则  $n_k$  表示第  $k$  类的样本数,  $u_k$  表示在第  $k$  类下的隶属度,  $u_k$  也是  $n_k/N$  的修正系数。通过比较式(4), 实际上就是  $n_k/N$  替代了  $p_i$ 。因此, 可以转化为以下灾害风险熵公式:

$$H(x) = - \sum \left\{ \frac{n_k^{(X)}}{N} u_k^{(X)} \log_c \left( \frac{n_k^{(X)}}{N} u_k^{(X)} \right) \right\} - \sum \left\{ \frac{n_k^{(Y)}}{N} u_k^{(Y)} \log_c \left( \frac{n_k^{(Y)}}{N} u_k^{(Y)} \right) \right\}. \quad (7)$$

式中:  $X$  表示灾情数据,  $Y$  表示致灾源数据,  $n_k^{(X)}$  表示灾情数据经由模糊  $c$  均值聚类下第  $k$  类的样本数,  $u_k^{(X)}$  表示灾情数据在第  $k$  类的隶属度;  $n_k^{(Y)}$  与  $u_k^{(Y)}$  则是把灾情数据换为致灾源数据,  $c$  值同样取 2。参考文献<sup>[13]</sup>, 在模糊  $c$  均值聚类下将样本划分为 5 类, 即将灾害等级划分为 5 个等级。

## 1.3 灰色关联度分析

灰色关联度<sup>[14]</sup> (Grey Relation Analysis, GRA) 是一种多因素统计模型, 由邓聚龙教授首先提出。灰色系统是介于白箱系统和黑箱系统之间的系统, 即只知道部分信息, 其余信息未知。在灰色系统中, 灰色关联度是度量某项指标受其它因素影响

的强弱程度, 关联度会随着时间的变化或者对象的不同而变化, 而相对灰色关联度是通过反映参考序列曲线与比较序列曲线相对于起始点的变化速率的接近程度来反映其关联程度的, 其计算公式如下:

$$\zeta_1 = \min_i \min_k |x_0(k) - x_i(k)|; \quad (8)$$

$$\zeta_2 = \max_i \max_k |x_0(k) - x_i(k)|; \quad (9)$$

$$r(X_0(k), X_i(k)) = \frac{\zeta_1 + \rho \cdot \zeta_2}{|x_0(k) - x_i(k)| + \rho \cdot \zeta_2}. \quad (10)$$

式中: 参考序列为  $X_0 = \{X_0(k), k = 1, 2, \dots, n\}$ , 被比较序列为  $X_i = \{X_i(k), k = 1, 2, \dots, n\}$ ,  $r(X_0, X_i)$  即为  $X_0$  与  $X_i$  的灰色关联度, 通过下列公式得出:

$$r(X_0, X_i) = \frac{1}{n} \sum_{k=1}^n r(X_0(k), X_i(k)). \quad (11)$$

一般认为参考序列与被比较序列的灰色关联度大于 0.5 且越接近 1, 关联程度就越高; 小于 0.5 越接近 0, 关联程度就越低。 $\rho$  是分辨系数, 取值大小在 (0, 1] 范围内, 若  $\rho$  越小, 关联系数间差异越大。根据文献<sup>[14]</sup>的分析结果表明,  $\rho$  值在选定为 0.5 时关联度能更好的反映序列间的影响程度, 因此本文选取  $\rho$  值为 0.5。

## 1.4 极限学习机回归模型

极限学习机<sup>[15]</sup> (Extreme Learning Machine, ELM) 是一种根据 BP 神经网络 (BP Neural Network, BPNN) 改进的, 基于前馈神经网络的机器学习算法, 克服了传统 BPNN 的一些缺陷, 例如它能够进行快速的训练和学习, 在其学习过程中易于在全局极小值收敛, 最大创新点在于输入层和隐含层的连接权值、隐含层的阈值可以任意设置, 无需再做调整; 隐含层与输出层之间的连接权值也不需要迭代调整, 能够通过解方程组的方式一次性解决, 是一种单隐含层神经网络结构模型。具体的算法描述如下。

输入训练集:  $A = \{x_i, t_i \mid x_i \in R^d, t_i \in R^m, i = 1, \dots, N\}$ 。 (12)

隐含层节点输出函数  $G(a_i, b_i, x)$  和隐含层节点个数  $L$ 。

步骤:

Step1: 随机产生隐含层的节点参数  $(a_i, b_i)$ ,  $i = 1, \dots, L$ 。

Step2: 计算出隐含层输出矩阵  $H$  (满秩):

$$H = \begin{bmatrix} G(a_1, b_1, x_1) & \dots & G(a_L, b_L, x_1) \\ \vdots & & \vdots \\ G(a_1, b_1, x_N) & \dots & G(a_L, b_L, x_N) \end{bmatrix}. \quad (13)$$

Step3: 计算最优输出权值:

$$\beta = H^+ T = \begin{cases} H^T (\frac{1}{\lambda} + H^T H)^{-1} T; \\ (\frac{1}{\lambda} + H^T H)^{-1} H^T H. \end{cases} \quad (14)$$

若  $H^T H$  是非奇异矩阵, 则  $H^+ = (H^T H)^{-1} H^T$ ; 若  $HH^T$  是非奇异矩阵, 则  $H^+ = H^T (H^T H)^{-1}$ 。其中  $1/\lambda$  是正则化系数,  $T = [t_1, \dots, t_N]_{M \times N}$ , 一般在  $H^T H$  或者  $HH^T$  对角线上增加一个正则化系数  $1/\lambda$  来获得更好的泛化能力。

$$\text{输出: } f(x) = h(x) H^T (\frac{1}{\lambda} + HH^T)^{-1} T. \quad (15)$$

表 1 1981 – 2016 年登陆华南的 54 个台风灾害灾情指数、致灾源指数和风险熵值

| 编号          | 致灾源指数         | 灾情指数          | 风险熵            | 编号          | 致灾源指数         | 灾情指数          | 风险熵            |
|-------------|---------------|---------------|----------------|-------------|---------------|---------------|----------------|
| 8105        | 10. 22        | 36. 93        | 1. 2175        | 0016        | 9. 57         | 53. 69        | 1. 2495        |
| 8303        | 13. 99        | 38. 66        | 1. 2593        | 0103        | 11. 58        | 59. 50        | 1. 3439        |
| 8309        | 11. 18        | 47. 92        | 1. 5893        | 0107        | 9. 81         | 34. 18        | 1. 1503        |
| 8410        | 11. 35        | 54. 08        | 1. 3580        | 0307        | 12. 38        | 54. 60        | 1. 2516        |
| 8521        | 10. 61        | 35. 25        | 1. 2312        | 0312        | 11. 32        | 42. 06        | 1. 5014        |
| <b>8616</b> | <b>17. 88</b> | <b>53. 73</b> | <b>0. 8523</b> | 0313        | 9. 45         | 31. 83        | 1. 5766        |
| 8710        | 12. 17        | 31. 68        | 1. 0576        | 0320        | 10. 29        | 43. 06        | 1. 1903        |
| 8805        | 11. 87        | 57. 45        | 1. 3907        | <b>0518</b> | <b>17. 31</b> | <b>61. 17</b> | <b>0. 9350</b> |
| 8823        | 10. 84        | 53. 79        | 1. 0253        | 0601        | 14. 4         | 53. 49        | 1. 0228        |
| 8905        | 12. 95        | 47. 96        | 1. 5620        | 0606        | 11. 86        | 61. 79        | 1. 0495        |
| 8908        | 13. 87        | 56. 10        | 1. 2358        | 0814        | 12. 84        | 64. 08        | 1. 4585        |
| 8926        | 11. 99        | 52. 86        | 1. 1095        | 0906        | 8. 49         | 50. 57        | 1. 1977        |
| 9106        | 10. 84        | 49. 59        | 1. 4160        | <b>0915</b> | <b>11. 05</b> | <b>49. 87</b> | <b>1. 7538</b> |
| 9107        | 13. 60        | 53. 35        | 1. 1494        | 1002        | 10. 16        | 50. 53        | 1. 3453        |
| 9108        | 14. 30        | 52. 42        | 1. 0890        | 1003        | 11. 65        | 56. 88        | 1. 4652        |
| 9111        | 14. 02        | 50. 55        | 1. 2755        | 1117        | 12. 08        | 61. 35        | 1. 0468        |
| 9204        | 12. 13        | 39. 41        | 1. 4055        | 1208        | 13. 72        | 49. 14        | 1. 5811        |
| 9205        | 8. 456        | 55. 38        | 0. 8878        | 1213        | 14. 85        | 53. 54        | 1. 1443        |
| 9302        | 12. 47        | 60. 50        | 1. 2559        | 1311        | 19. 50        | 63. 51        | 1. 0937        |
| 9309        | 11. 83        | 64. 37        | 1. 1833        | 1319        | 12. 81        | 49. 11        | 1. 5509        |
| 9315        | 10. 20        | 57. 63        | 1. 3582        | <b>1409</b> | <b>18. 30</b> | <b>64. 31</b> | <b>0. 9605</b> |
| 9316        | 9. 83         | 57. 09        | 1. 4915        | 1415        | 11. 39        | 56. 83        | 1. 6221        |
| <b>9318</b> | <b>17. 87</b> | <b>47. 23</b> | <b>0. 9394</b> | 1510        | 9. 39         | 10. 45        | 1. 5687        |
| 9509        | 10. 61        | 62. 97        | 1. 228         | 1522        | 14. 02        | 59. 99        | 1. 0378        |
| 9612        | 9. 527        | 45. 77        | 1. 1998        | 1604        | 10. 25        | 40. 92        | 1. 3192        |
| <b>9615</b> | <b>8. 81</b>  | <b>72. 24</b> | <b>0. 5405</b> | 1621        | 12. 34        | 46. 44        | 1. 1769        |
| 9903        | 8. 53         | 48. 42        | 1. 1378        | 1622        | 10. 69        | 34. 07        | 1. 2035        |

2 模型的应用与分析

2.1 数据来源

选取 1981 – 2016 年华南地区 54 个台风及其以上灾害样本，其中致灾源数据来源于中国热带气旋年鉴<sup>[16]</sup>和中国台风网(<http://typhoon.weather.com.cn/index.shtml/2017/07/14/>)，灾情数据来源于中国气象灾害大典<sup>[17-19]</sup>以及广东省、海南省和广西壮族自治区气候中心。

2.2 登陆华南台风灾害风险熵的分析

通过式(1)、式(2)、式(3)和式(7)，并结合模糊 *c* 均值聚类得到 54 个台风灾害的灾情指数序列 *X*、致灾源指数序列 *Y* 和灾害风险熵序列 *H*(*x*)，具体结果如表 1 所示。

从表 1 中我们可以发现，9615 号台风灾害风险熵值为 0. 540 5，是所有台风样本中最低，并且该台风灾害过程影响较大，其致灾源系统中最大风速极值为 50 m/s，降雨天数为 3 d，降雨极值为 155 mm，灾情系统中受灾人口为 1 530 万人，死亡人口为 284 人，倒塌房屋数达到 2 491 200 间，农作物受灾面积达到 7 397 hm<sup>2</sup>，造成直接经济损失 218. 63 亿元，是一次极强的台风灾害过程；而 0915 号台风风险熵值为 1. 753 8，是所有台风样本

中最高的。对比实况，可以发现 0915 号台风登陆时最大风速为 35 m/s，造成 24. 2 亿元直接经济损失，是一次强度不大的台风灾害。另外还有编号 8616、9318、0518 和 1409 的台风风险熵值小于 1，它们的灾情指数与致灾源指数都相应的偏高。

为了进一步探讨灾害风险熵的分布情况，我们给出 54 个台风灾害过程的灾害风险熵值的条形分布图和核密度估计曲线图(图 1)。在图 1 中，纵坐标密度 = 频率/组距，从图 1 我们可以发现，密度曲线图最低点的风险熵值约为 0. 540 5，对应的台风灾害过程为 9615 号台风，正是在上述提到的极强台风，属于发生概率极低的事件；最高点的风险熵值为 1. 197 8，对应 0906 号台风，是一次强度不大的台风，造成的直接经济损失为 5. 54 亿元，属于所有台风中发生概率最高的。随着风险熵值的增大，台风发生概率逐渐增大，对应的台风强度会相应降低，在过了最高点后，随着风险熵值增大，台风发生概率逐渐减小，其中低风险熵值对应的多为破坏性强的台风。例如，风险熵值约为 0. 960 5 的 1409 号台风，造成的直接经济损失高达 417. 6 亿元；风险熵值约为 0. 935 的 0518 号台风，降雨极值为 613 mm，受灾人口高达 802. 92 万人。对风险熵进行 Shapiro – wilk 正态性检验，其显著性概率值为 0. 535 8，接受原假设，即说明风险熵服从正态分布。

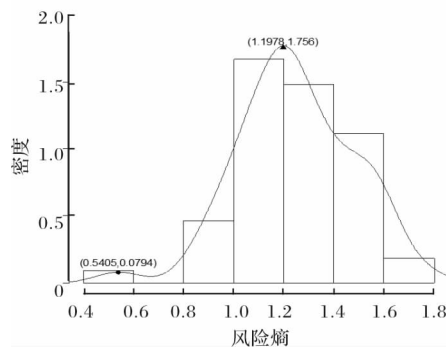


图1 1981-2016年登陆华南的54个台风灾害风险熵值分布图

### 2.3 登陆华南台风灾害风险熵和灾情因子、致灾源因子关联分析

在这里,我们把灾害风险熵  $H(x)$  作为参考序列,把灾情指数序列  $X$  与致灾源指数序列  $Y$  作为被比较序列。首先计算出灾害风险熵分别和致灾源指数与灾情指数的皮尔逊相关系数,得到三者的皮尔逊相关系数矩阵(表2)。从表2可以看出灾情指数与致灾源指数相关系数为0.394,存在正的线性相关性,可认为在台风灾害过程中致灾因子的影响程度越大,造成的灾情影响就越严重;灾害风险熵与灾情指数和致灾源指数的相关系数分别为-0.34和-0.28,存在负的线性相关性,说明当灾害风险熵值越大,对应的台风灾害过程灾情指数和致灾源指数就会相应的越小。但是此相关系数较低,原因在于皮尔逊相关系数只能表示两个指标间的线性关系,无法描述它们的非线性关系,且容易受到样本数量的影响。

表2 灾情指数、致灾源指数和风险熵相关阵

|       | 灾情指数  | 致灾源指数 | 风险熵   |
|-------|-------|-------|-------|
| 灾情指数  | 1     | 0.394 | -0.34 |
| 致灾源指数 | 0.394 | 1     | -0.28 |
| 风险熵   | -0.34 | -0.28 | 1     |

相关研究表明,台风灾害过程是一个灰色系统,受单个因子影响较弱,与众多影响因素存在复杂的非线性关系<sup>[9]</sup>,而灰色关联度能够有效的描述每项指标与参考指标间的非线性关系,刻画它们对参考序列的影响程度。因此我们将利用灰色关联分析法探讨灾情指数和致灾源指数分别与灾害风险熵的关联程度。将数据代入式(8)-式(10),得到每个台风灾害样本的风险熵值与灾情指数和致灾源指数的相对灰色关联度,探讨灾情因子和致灾源因子分别对灾害风险熵的影响程度,得到的结果见图2和图3。

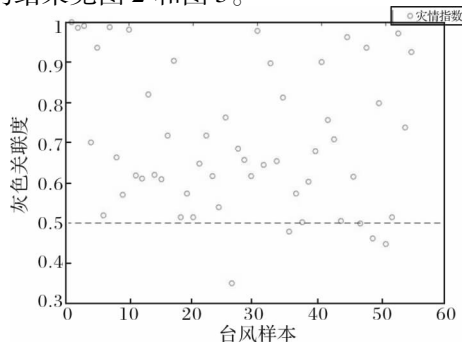


图2 1981-2016年登陆华南的54个台风灾情指数与风险熵灰色关联度散点图

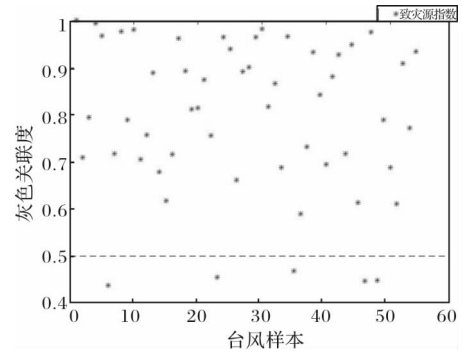


图3 1981-2016年登陆华南的54个台风致灾源指数与风险熵灰色关联度散点图

从图2和图3可以发现,大部分登陆华南的台风灾情指数、致灾源指数与灾害风险熵的灰色关联度大于0.5,只存在少部分台风灾害灰色关联度低于0.5。在图2中,第9615号、0518号、1311号、1409号和1415号台风样本灰色关联度都低于0.5,分别为0.3506、0.4789、0.4985、0.4616和0.4482,表明这些台风灾害风险熵值受灾情因子影响较弱;由图3可知,第8616号、9318号、0518号、1311号和1409号台风样本的灰色关联度分别为0.4376、0.4553、0.4687、0.4407和0.4492,表明这些台风样本风险熵值受致灾源因子影响较弱;另外,比较图2和图3,存在一些台风样本灾情指数与灾害风险熵灰色关联度较低,致灾源指数与灾害风险熵灰色关联度较高,说明在该台风灾害过程中,致灾因子对灾害风险熵的影响大于灾情因子,例如第0814号台风样本灾情指数与风险熵灰色关联度为0.6032,致灾源指数与风险熵灰色关联度为0.933,致灾因子对风险熵的影响程度比灾情因子对风险熵的影响程度更大。接着利用公式(11)计算出灾情指数序列和致灾源指数序列分别与灾害风险熵的灰色关联度,其值分别为0.7162和0.7949。

最终可以得出,在1981-2016年登陆华南的54个台风灾害中,风险熵值受灾情指数序列和致灾源指数序列的影响较大,灾情指数与致灾源指数分别与灾害风险熵值存在负的线性相关性,当风险熵值降低时会导致灾情指数与致灾源指数偏高。在表1中可以发现风险熵值小于1的台风样本致灾能力较强,所造成的灾情损失更为严重。从熵的理论来解释,熵值减小则可认为“负熵”增加,从而导致整个台风系统所携带的信息量增多,不确定性加大,破坏性也就增强。

### 2.4 构建极限学习机回归模型

构建基于极限学习机的华南台风灾害风险熵回归模型,以多元线性回归(Multiple Linear Regression, MLR)模型和BP神经网络回归模型作为对照组,探讨三种不同回归模型的回归与预测效果。在1981-2016年登陆华南的54个台风灾害中选取前48个台风灾害作为训练样本构建回归模型,第49~54个台风灾害作为测试集。在对后6个台风样本进行预测前,对48个台风灾害样本进行回归拟合,初步探讨三种不同模型对风险熵值的回归效果(图4)。

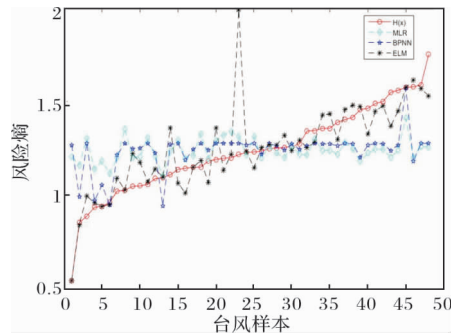


图4 MLR、BPNN 和 ELM 对 48 个台风灾害风险熵回归效果对比

在图 4 中，为了方便比较三种方法的回归效果，将风险熵值从小到大进行排列，从中可以看出 MLR 与 BPNN 模型回归值与真实值差距比较大，而 ELM 回归结果与大多数真实值接近，存在少部分奇异点，其中 MLR 和 BPNN 模型的平均绝对误差分别为 0.176 1 和 0.152 5，而 ELM 的平均绝对误差为 0.078 7，是三者中最低；MLR 和 BPNN 的拟合优度分别为 82.3% 和 83.57%，而 ELM 的拟合优度为 88.7%，是三者中最高。因此，从对 48 个台风样本风险熵值回归效果来看，极限学习机模型要优于多元线性回归模型和 BP 神经网络模型。接下来利用这三种不同的回归模型对后 6 个台风灾害做风险熵值预测，在测试过程中，首先用 48 个训练样本构建的回归模型来预测第 49 个测试样本的灾害风险熵值，将预测结果加入训练样本中，用 49 个训练样本构建的回归模型来预测第 50 个测试样本的灾害风险熵值，如此反复下去，直到预测完第 54 个样本。三种不同方法得到的预测结果及拟合优度如表 3 所示。表 3 是分别用 MLR、BPNN 和 ELM 三种不同方法对后 6 个台风样本灾害风险熵进行预测的结果，从中可以发现 MLR 预测效果最不理想，拟合优度最低，回归方程也没能通过显著性检验；BPNN 在于它本身具有非线性映射能力，对非线性预测的效果会比 MLR 效果要稍好；ELM 与另外两种方法相比预测效果有着显著的提升，其中拟合优度达到 92.82%，平均绝对误差只有 0.059。ELM 算法关键点在于隐含层神经元个数的调整，也仅需要调整隐含层神经元个数，输入层和隐含层的连接权值和隐含层神经元的阈值将由该算法随机产生，训练过程中也无需调整。隐含层神经元个数选取将影响预测的效果，经过多次的实验以及效果对照，最终选取隐含层神经元个数为 29。为了比较三种方法对风险熵值预测的适用性，我们同样给出了 6 个台风灾害样本的预测效果图(图 5)。

表 3 MLR、BPNN 和 ELM 对后 6 个台风灾害的风险熵值预测结果

| 台风编号   | 风险熵    | MLR    | BPNN   | ELM    |
|--------|--------|--------|--------|--------|
| 1415   | 1.6222 | 1.2278 | 1.2328 | 1.4027 |
| 1510   | 1.5687 | 1.4194 | 1.5702 | 1.5835 |
| 1522   | 1.0378 | 1.1841 | 1.2328 | 1.0977 |
| 1604   | 1.3193 | 1.2991 | 1.2328 | 1.3191 |
| 1621   | 1.1769 | 1.2536 | 1.2329 | 1.2351 |
| 1622   | 1.2035 | 1.3183 | 1.2328 | 1.2021 |
| 拟合优度%  | -      | 85.73  | 86.31  | 92.82  |
| 平均绝对误差 | -      | 0.1503 | 0.1262 | 0.059  |

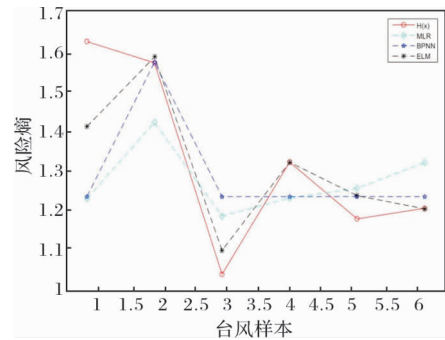


图5 MLR、BPNN 和 ELM 对后 6 个台风风险熵预测效果对照图

从表 3 和图 5 可以看出，ELM 对后 6 个台风灾害风险熵的预测值与计算结果最为接近，在对 1510 号、1604 号和 1622 号台风风险熵值的预测上，这三个台风风险熵值分别为 1.568 7、1.319 3 和 1.203 5，ELM 对它们的预测值分别为 1.583 5、1.319 1 和 1.202 1，与计算结果非常接近；在对 1415 号、1522 号和 1621 号台风风险熵值的预测上，这三个台风风险熵值分别为 1.622 2、1.037 8 和 1.176 9，ELM 对它们的预测值分别为 1.402 7、1.097 7 和 1.235 1，比 MLR 和 BPNN 的预测值更接近计算结果。由此说明，极限学习机模型在对台风灾害风险熵值的预测上比多元线性回归模型和 BP 神经网络模型拥有更好的性能。

### 3 结语

(1)构建华南台风灾害风险熵模型。选取主要的灾情因子和致灾源因子计算灾情指数序列和致灾源指数序列，并结合模糊 c 均值聚类 and 广义模糊熵原理计算出 1981 - 2016 年 54 个台风灾害风险熵值，结果发现风险熵值呈正态分布，低风险熵值对应强度较大的台风。

(2)探讨灾害风险熵与灾情因子和致灾源因子关系。首先计算出风险熵值分别与灾情指数和致灾源指数的皮尔逊相关系数，分别为 -0.34 和 -0.28，呈负相关，但是此相关系数较低，原因在于皮尔逊相关系数只能表示它们之间的线性关系，无法描述它们之间的非线性关系。因此利用式(8)和式(9)分别计算出灾情指数与风险熵、致灾源指数与风险熵的灰色关联度，其结果分别为 0.7162 和 0.794 9。表明灾情指数与致灾源指数对风险熵的影响程度较大，且与风险熵值呈负相关。

(3)进一步探讨风险熵与灾情因子和致灾源因子的非线性关系。利用 MLR、BPNN 和 ELM 三种不同的回归方法，以灾情指数和致灾源指数作为输入变量，灾害风险熵值作为输出变量构建回归模型。通过对后 6 个台风测试样本风险熵的真实值与预测值比较，ELM 方法的平均绝对误差为 0.059，拟合优度为 92.82%，其结果要优于 MLR 和 BPNN。表明极限学习机在风险熵值的预测上具有更好的适用性。

然而在台风灾害过程中各个灾情因子和致灾因子有着不同的影响权重，本文在构造灾情指数序列和致灾源指数序列时没能考虑到这一点；ELM 回归方法对风险熵的回归效果具有普遍一般性，

但是对于某些奇异点的预测偏差依然较大, 极限学习机模型需要进一步优化以解决问题。同时, 本文所研究的台风灾害风险熵-极限学习机预测模型为其它自然灾害的风险分析提供一条新的途径。

## 参考文献:

- [1] 陈联寿, 端义宏, 宋丽莉, 等. 台风预报及其灾害[M]. 北京: 气象出版社, 2009: 178-181.
- [2] 张俊香, 刘旭拢, 宫清华. 近 60 年华南沿海登陆热带气旋特征和热带气旋巨灾灾情分析[J]. 热带地理, 2012, 32(4): 357-363.
- [3] 张广平, 张晨晓, 谢忠. 基于 T-S 模糊神经网络的模型在台风灾情预测中的应用——以海南为例[J]. 灾害学, 2013, 28(2): 86-89.
- [4] 陈燕璇, 刘合香, 谭金凯. 基于等距特征映射降维的台风灾情概率神经网络预评估模型[J]. 灾害学, 2016, 31(3): 20-25, 30.
- [5] 陈仕鸿, 隋广军, 唐丹玲. 一种台风灾情综合评估模型及应用[J]. 灾害学, 2012, 27(2): 87-91.
- [6] 黄小燕, 金龙. 基于主成分分析的人工智能台风路径预报模型[J]. 大气科学, 2013, 37(5): 1154-1164.
- [7] 张东, 张宁. 物理学中的熵理论及其应用研究[J]. 北京联合大学学报(自然科学版), 2007(1): 4-8.
- [8] 金菊良, 王银堂, 魏一鸣, 等. 洪水灾害风险管理广义熵智能分析的理论框架[J]. 水科学进展, 2009, 20(6): 894-900.
- [9] 魏章进, 马华铃, 唐丹玲. 基于改进熵值法的台风灾害风险趋势评估[J]. 灾害学, 2017, 32(3): 7-11.
- [10] 宇岩, 欧国强, 王钧, 等. 信息熵在震后深溪沟流域泥石流危险度评价中的应用[J]. 防灾减灾工程学报, 2017, 37(2): 264-272.
- [11] 谭金凯, 刘合香, 倪增华, 等. 台风致洪灾害的广义模糊熵模型研究[J]. 模糊系统与数学, 2016, 30(1): 182-190.
- [12] 何清. 模糊聚类分析理论与应用研究进展[J]. 模糊系统与数学, 1998(2): 89-94.
- [13] 陈报章, 仲崇庆. 自然灾害风险损失等级评估的初步研究[J]. 灾害学, 2010, 25(3): 1-5.
- [14] 田民, 刘思峰, 卜志坤. 灰色关联度算法模型的研究综述[J]. 统计与决策, 2008(1): 24-27.
- [15] 张海霞. 极限学习机理论与算法研究[D]. 沈阳: 沈阳航空航天大学, 2017: 9-15.
- [16] 中国气象局. 热带气旋年鉴 1981-2014 年[Z]. 北京: 气象出版社, 2015.
- [17] 温克刚. 中国气象灾害大典: 海南卷[Z]. 北京: 气象出版社, 2006.
- [18] 温克刚. 中国气象灾害大典: 广东卷[Z]. 北京: 气象出版社, 2006.
- [19] 温克刚. 中国气象灾害大典: 广西卷[Z]. 北京: 气象出版社, 2006.

# Research on Extreme Learning Machine-Risk Entropy Prediction Model of Typhoon Disaster in South China

LU Yaojian<sup>1</sup>, LIU Hexiang<sup>1,2</sup> and WANG Meng<sup>1</sup>

- (1. Academy of Mathematics and Statistical Sciences, Nanning Normal University, Nanning 530001, China;
2. Guangxi Key Laboratory of Marine Disaster Research in Beibu Gulf, Qingzhou 535000, China)

**Abstract:** Combined with the generalized fuzzy entropy principle and fuzzy c-means clustering method, the risk entropy model of typhoon disaster in South China is constructed. The risk analysis of typhoon disaster in South China is carried out and its distribution is discussed. The relationship between disaster condition factor, disaster source factor and disaster risk entropy of typhoon disaster in South China and the degree of their influence on risk entropy are discussed by means of grey relational analysis. A nonlinear regression model based on extreme learning machine was established. The relationship between risk entropy and disaster condition factor and disaster source factor was further discussed by using multiple linear regression and BP neural network as the compared group. The results show that the entropy value of typhoon disaster risk in South China is normal distribution, which is related to the disaster situation. The grey relation degree of the factor and the disaster-causing source factor is 0.7162 and 0.7949, respectively, and the influence of the disaster factor and the disaster-causing source factor is large; the average absolute error of the risk entropy value of the South China typhoon disaster, which is predicted by the built limit learning machine model, is 0.059, and the fitting goodness is 92.82%. The results of this paper are compared with the predicted results of the conventional multiple linear regression and the BP neural network method. The results show that the risk entropy of the South China typhoon is predicted by the built limit learning machine model, the performance of the method is better than that of the conventional multiple linear regression and the BP neural network.

**Key words:** South China; typhoon disaster; risk entropy; Limit learning machine; grey correlation; fuzzy clusterin