

АЛГОРИТМ АНАЛИЗА НЕКОЛИЧЕСТВЕННЫХ ДАННЫХ ЭКОЛОГИЧЕСКОГО МОНИТОРИНГА

Коршунова Ю. Л.

Научный руководитель – д.т.н., профессор Симонов К.В.

Сибирский федеральный университет

Рассмотрим алгоритм анализа неколичественных данных, основанный на экспертных оценках. Количественную оценку качественного признака, характеризующего степень экологической опасности для территории, можно получить на основе сравнения территорий. Получаемые при этом оценки являются относительными, поскольку зависят от того, какие территории сравниваются. Относительные веса по опасности территорий при их небольшом числе ($m \leq 10$) могут быть установлены экспертом. При этом его суждения опираются на профессиональную подготовку и не являются результатом каких-либо рассуждений или вычислений. При большом ($m > 10$) числе территорий определение относительных весов должно опираться на формализованную процедуру.

Для того чтобы оценить степень предпочтения, вводятся количественные оценки предпочтительности территорий h_1, \dots, h_m . Наименее предпочтительной по сравниваемому признаку (m -й) территории присваивается оценка $h_m = 1$. Оценки предпочтительности территорий, выставленные экспертом, считаются согласованными, если имеет место равенство $e_{ij} = e_i / e_j$, где $e_{ij} (i = 1, \dots, m-2; j = 2, \dots, m-1, j > i)$, удовлетворяющие условию $e_{ij} \cdot h_i = h_j$. В противном случае необходимо установить новые значения $e_i (i = 1, \dots, m-1)$. Общее число оценок, которое должен установить эксперт, составляет $m(m-1)/2$. Эксперт сравнивает попарно все территории и дает количественную оценку каждому такому сравнению, после чего возникает необходимость в многочисленных корректировках.

Оценки относительных весов территорий по рассматриваемому признаку вычисляются по формуле $q_i = h_i / \sum_{i=1}^m h_i (i = 1, \dots, m)$. Рассматриваемый подход весьма трудоемок, при этом основные усилия затрачиваются на этапе экспертного оценивания, а не на этапе обработки его результатов.

Далее эксперту поочередно предъявляются все пары территорий, и каждый раз устанавливается, какой из элементов предпочтительнее по рассматриваемому признаку. При этом заполняется матрица попарных сравнений W с элементами:

$$w_{ij} = \begin{cases} 2, & \text{если } \theta_i \succ \theta_j; \\ 1, & \text{если } \theta_i \approx \theta_j; (i, j = 1, \dots, m); \\ 0, & \text{если } \theta_i \prec \theta_j; \end{cases}$$

где θ_i и θ_j – сравниваемые территории.

В качестве относительных весов территорий принимаются компоненты нормированного собственного вектора матрицы W на k -ом шаге итераций:

$$\hat{q}_i^{(k)} = \frac{1}{\lambda^{(k)}} \sum_{j=1}^m w_{ij} \hat{q}_i^{(k-1)}, \quad \lambda^{(k)} = \frac{1}{\lambda^{(k)}} \sum_{i=1}^m \sum_{j=1}^m w_{ij} \hat{q}_i^{(k-1)}.$$

Требуемую точность вычислений ε компонента собственного вектора устанавливают заранее, останавливая расчеты при выполнении m условий

$|\hat{q}_i^{k-1} - \hat{q}_i^k| \leq \varepsilon$. Достоинствами этого варианта алгоритма являются достаточно простая процедура выставления оценок и формализованное вычисление весов территорий.

На следующем этапе анализа данных выбирается нечеткая переменная, которая описывает степень опасности территорий, и затем строится функция принадлежности территорий нечеткому множеству, смысл которого формализован выбранной нечеткой переменной. Для вычисления степеней принадлежности территорий нечеткому множеству \tilde{A} используется метод попарных сравнений по качественному признаку с количественной оценкой предпочтения. Для получения матриц попарных сравнений проводят опрос экспертов относительно того, насколько, по их мнению, территория θ_i более соответствует смыслу нечеткой переменной, в отличие от территории θ_j . Для выставления оценок w_{ij} эксперт с помощью шкалы Т. Саати сравнивает предполагаемые им опасности пары территорий.

Таблица. Шкала попарных сравнений Т. Саати.

Значение w_{ij}	Определение	Пояснение
1	Территории, одинаково опасные (безопасные)	Территории обладают примерно одинаковой опасностью
2	Промежуточное значение	
3	Слабое превосходство	Эксперт считает, что опасность первой территории пары
4	Промежуточное значение	
5	Сильное превосходство	Эксперт считает, что опасность первой территории пары определено выше, чем второй
6	Промежуточное значение	
7	Явное превосходство	Эксперт считает, что опасность первой территории пары явно выше, чем второй и статистика это подтверждает
8	Промежуточное значение	
9	Абсолютное превосходство	У эксперта нет никаких сомнений относительно того, что опасность первой территории пары значительно выше, чем второй

Далее для перевода количественных значений всех показателей в их качественные аналоги используют «статистику интервальных данных».

Количества	Качества
От XI до X 2	качественный уровень – 1
От X 2 до X 3	качественный уровень – 2
От X 3 до X 4	качественный уровень – 3
От X 4 до X 5	качественный уровень – 4
От X 5 до X 6	качественный уровень – 5

Для работы с данными на классах рассматривается числовая функция, обозначенная следующим образом: $\beta(F_1) = 1,0$; $\beta(F_2) = 0,8$, $\beta(F_3) = 0,6$; $\beta(F_4) = 0,4$; $\beta(F_5) = 0,2$. Взвешенные значения опасностей используются как входные данные для расчетов значений суммарной опасности по формуле:

$$H_{\tilde{n}oi} = \alpha_1\beta(F_1) + \alpha_2\beta(F_2) + \alpha_3\beta(F_3) + \dots,$$

где F_i – класс опасности, α – вес рассматриваемой опасности. Значения суммарной опасности рассчитываются как сумма взвешенных опасностей.

На заключительном этапе анализа данных применяется методика рангового анализа:

$$N(i) = \frac{A}{i^b}.$$

Для оценки надежности-согласованности данных применяются несколько подходов:

1. Способ отдельного коррелирования – коэффициент корреляции (как правило, ρ -Спирмена) рассчитывается по соответствующим двум наборам данных – суммарным результатам по каждому набору данных. Для корректировки полученного значения используется формула Спирмена-Брауна:

$$\rho_{SB2} = \frac{2\rho}{1 + \rho},$$

где ρ_{SB2} – скорректированный показатель надежности-согласованности по методу отдельного коррелирования, ρ – коэффициент корреляции Спирмена между двумя наборами показателей.

2. Другой способ корректировки значения коэффициента корреляции между двумя группами показателей (полученных в ходе выполнения 1 этапа) – обобщенная формула Спирмена-Брауна, в которой используется усредненный коэффициент корреляции:

$$\rho_{SB} = \frac{m \cdot \rho}{1 + (m-1)\rho},$$

где ρ_{SB} – усредненный скорректированный показатель надежности-согласованности по методу отдельного коррелирования, m – количество показателей, ρ – усредненный коэффициент корреляции между всеми показателями.

3. Для порядковых шкал с большим количеством позиций (например, шкалы Лайкерта) надежность-согласованность всего инструмента рекомендуется оценивать также при помощи коэффициента альфа Кронбаха α :

$$\alpha = \frac{m}{m-1} * \left(1 - \frac{\sum_{i=1}^m \sigma_i^2}{\sigma_t^2} \right),$$

где σ_i^2 – дисперсия откликов (оценок экспертов) по каждому показателю, $\sum \sigma_i^2$ – дисперсия суммарной шкалы (дисперсия суммы откликов каждого эксперта).

Таким образом, в работе предложена вычислительная технология анализа количественных данных мониторинга для оценки степени экологической опасности для исследуемой территории.