УДК 004.02:004.67:51.7

Л.М. Божуха

ПРО СИСТЕМУ ПЕРЕВАГ ЛІНГВІСТИЧНОГО КЛАСИФІКАТОРА

Анотація. У роботі представлені результати побудови найпростішої комплексної оцінки на основі трапецієвидного числа лінгвістичного класифікатора з використанням показника Херста для виявлення трендовості показників. Використання запропонованої модифікації методу може розширити область якісних критеріїв оцінки з лінгвістичним класифікатором при аналізі кількісних характеристик.
Ключові слова: лінгвістичний класифікатор, стандартний 01-класифікатор, функція належності, показник Херста

Постановка проблеми

Для аналізу роботи будь-якої системи можна використовувати математичний апарат на основі вибірки статистичних даних системи. На систему впливає цілий ряд взаємопов'язаних між собою причин, які можна об'єднати за групами. Зв'язок між цими групами може бути дуже суттєвим і тому завжди виникає складність у побудові комплексної математичної моделі системи.

Аналіз кількісних характеристик кожного параметра у причинних групах може бути показаний на побудованій інтерполяційній функції або при використанні математичного апарату планування експерименту. Отримані таким чином критерії є локальними для кожної з груп та не є корисними для систем з іншим набором кількісних характеристик. У більшій мірі математична модель, яка побудована тільки на основі статистичних даних, показує необхідні результати тільки при використанні великого обсягу цих даних. Як правило такі моделі можуть бути багатовимірними, але кількість параметрів, які аналізуються, тільки ускладнюють математичні перетворення і іноді не враховуються зв'язки між факторами.

Багато причинних груп можуть не мати кількісних характеристик, або мати дуже велику їх кількість з невизначеними пороговими значеннями. При оцінюванні цих слабо вимірюваних факторів використовують штучні прийоми. Проблемою цього підходу є створення для аналітика або експерта методики виставлення відповідних балів по критерію з термінологєю нечітких виражених степенів: «дуже низька», «низька», «середня», «висока», «дуже висока». Проблемою побудови таких критеріїв по якісним характеристикам є існування набору показників невпорядкованих факторів одного рівня ієрархії.

Постановка задачі

Проблематикою роботи є співставлення лінгвістичного опису з кількісними характеристиками факторів. При побудові комплексної оцінки [1] проаналізувати кількісні характеристики факторів, відношення між факторами у структурі ієрархії цих факторів.

Метою даної роботи є представлення модифікації схеми агрегування даних на одному рівні ієрархії при аналізі довільної системи з зафіксованим набором її показників.

Основна частина

Нехай моделлю математична модель:

$$\mathrm{FSM} = < \mathrm{G}, \mathrm{L}, \mathrm{S} >. \tag{1}$$

де G - деревоподібна ієрархія факторів.

Набір якісних оцінок рівнів кожного фактору в ієрархії G має вигляд:

$$\mathrm{L} = \left\{ \begin{matrix} \textit{дуже низький рівень } (ДН) \\ \textit{низький рівень } (Н) \\ \textit{середній рівень } (С) \\ \textit{високий рівень } (В) \\ \textit{дуже високий рівень } (ДВ) \end{matrix} \right\}. \tag{2}$$

Система переваг одних факторів іншим для одного рівня ієрархії факторів:

$$S = \left\{ F_{k,i} \Re\, F_{k,j} \mid \Re = \left\{ \begin{matrix} > \textit{відношення переваги}, \\ \approx \textit{відношення рівноваги} \end{matrix} \right\} \right\} \tag{3}$$

Деревоподібна ієрархія факторів G може бути описана орієнтованим графом без циклів, петель, горизонтальних ребер у межах одного рівня ранжування, який містить одну кореневу вершину:

$$\mathrm{G} = < F, V >, \tag{4}$$

де $F = \{ F_i \}$ - множина вершин факторів, $V = \{ V_i \}$ - множина дуг.

Вершину, яка відповідає фактору у цілому позначимо F.

Необхідно вибрати ряд окремих показників, про які можна сказати, що вони якнайкраще характеризують окремі сторони і при

 ISSN 1562-9945

цьому утворюють якусь закінчену сукупність, що дає вичерпне уявлення в цілому. Вибір системи показників для аналізу може бути індивідуальним для кожної системи. Значимість тих чи інших показників для оцінки системи різна і тому, перед експертом постає важке завдання відбору і ранжирування чинників аналізу. Показники, які класифіковані за групами, можуть утворювати ієрархію, але в найпростішому випадку вони просто складають невпорядкований набір.

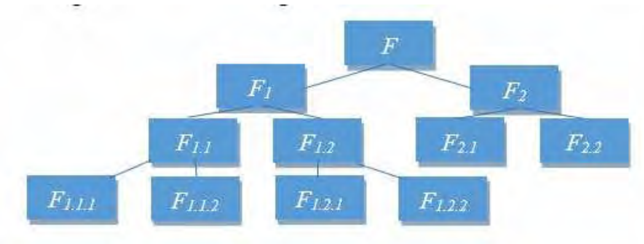

Рисунок 1 - Деревовидна ієрархія

Розглянемо деревовидну ієрархію F, яка залежить від двох основних груп існування критеріїв аналізу F_1, F_2. Дерево можна розширити, якщо додавати до графу G нові вузли. На ієрархії критеріїв (*рис. 1*) можна обрати систему відношень переваг на основі експертної оцінки груп факторів. На нашій схемі ця система є початковою і на даному етапі моделювання не є експертною. Припустимо, що всі фактори рівнів ієрархії (*рис. 1*) знаходяться у відношенні рівноваги:

$$S = \begin{Bmatrix} F_1 \approx F_2,\, F_{1.1} \approx F_{1.2},\, F_{2.1} \approx F_{2.2}, \\ F_{1.1.1} \approx F_{1.1.2},\, F_{1.2.1} \approx F_{1.2.2} \end{Bmatrix}. \quad (5)$$

Сформуємо лінгвістичну змінну «Рівень фактору» з терм-множиною значень L. Алгоритм побудови змінної агрегування даних проходить за напрямом дуг графа ієрархії при застосуванні OWA-оператора Ягера [2] з вагами у згортці у вигляді коефіцієнтів Фішберна і кожному рівню відповідає функція належності.

Припустимо, що лінгвістичні оцінки рівнів факторів не є експертними, не відомі їх характеристики і вони відповідають набору якісних оцінок L:

$$\{F_{1.1.1}(C), F_{1.1.2}(C), F_{1.2.1}(C), F_{1.2.2}(C), F_{2.1}(C), F_{2.2}(C)\}. \quad (6)$$

Для запису якісних характеристик параметру рівня надалі використовується стандартний кількісний вигляд у вигляді відповідної функції належності. Функція належності $\mu^*(x)$ представляється стандартним п'ятирівневим 01-класифікатором. Функції належності

$\mu^*(x)-$ трапецієвидні трикутні числа $\left(a_1^*, a_2^*, a_3^*, a_4^* \right)$, де a_i^*- абсциси вершин трапеції.

Отже, для подальшої побудови алгоритму у якості функції належності обираємо стандартну функцію у вигляді:

$$\mu^*(x) = \begin{cases} ДН: & \mu_1(x) = (\ -0.05, 0.05, 0.15, 0.25\ ) \\ Н: & \mu_2(x) = (\ 0.15, 0.25, 0.35, 0.45\ ) \\ С: & \mu_3(x) = (\ 0.35, 0.45, 0.55, 0.65\ ) \\ В: & \mu_4(x) = (\ 0.55, 0.65, 0.75, 0.85\ ) \\ ДВ: & \mu_5(x) = (\ 0.75, 0.85, 0.95, 1.05\ ) \end{cases} \quad (7)$$

За кожним показником $F^* = \left(F_1^*, F_2^*, \ldots, F_{\text{i}}^*\right)$ на обраному підрівні графу G відомі лінгвістичні оцінки $L^* = \left(L_1^*, L_2^*, \ldots, L_{\text{i}}^*\right)$ та визначена вагова система Фішберна $P^* = \left(p_1^*, p_2^*, \ldots, p_{\text{i}}^*\right)$. Показник підрівня F^* характеризується своєю лінгвістичною оцінкою, яка визначається функцією належності на 01-носії $x \in [0;1]$ за допомогою OWA-оператора Ягера:

$$\mu^*(x) = \sum_{k=1}^{i} p_k^* * \mu_k^*(x) = \sum_{k=1}^{i} p_k^* * \left(\ a_{\text{k}1}^*, a_{k2}^*, a_{k3}^*, a_{k4}^*\ \right) = \\ = \left(\sum_{k=1}^{i} p_k^* * a_{\text{k}1}^*, \sum_{k=1}^{i} p_k^* * a_{\text{k}2}^*, \sum_{k=1}^{i} p_k^* * a_{\text{k}3}^*, \sum_{k=1}^{i} p_k^* * a_{\text{k}4}^* \right), \quad (8)$$

де $\mu_k^*(x)$ визначені формулами (7).

Проводячи обчислення послідовно знизу догори по усім рівням ієрархії G, застосовуючи співвідношення (7), (8) і доведені у роботі [1] результати лінгвістичного розпізнавання показника верхнього рівня ієрархії, отримуємо функцію належності фактору F та його інтерпретацію: $\mu_1(x)$-ДН-Дуже низький; $\mu_2(x)$-Н-Низький; $\mu_3(x)$-С-Середній; $\mu_4(x)$-В-Високий; $\mu_5(x)$-ДВ-Дуже високий.

Критерій для побудованої моделі, яка представлена графом G (*рис. 1*) з системою переваг (5) та лінгвістичними оцінками факторів (6) на рівні F має вигляд $F_1 \approx F_2$ і $p_1 = 1/2$, $p_2 = 1/2$. Функція належності цього рівня може бути записана у вигляді: $\mu^*(x) = \left(0.35 \;\; 0.45 \;\; 0.55 \;\; 0.65\ \right)$.

 ISSN 1562-9945

Для того, щоб визначити рівень кожного з отриманих факторів, необхідно визначити ступінь схожості трапецієвидного числа $\left(a_1^*, a_2^*, a_3^*, a_4^* \right)$ відповідного рівня і стандартного трапецієвидного числа $\left(b_1^*, b_2^*, b_3^*, b_4^* \right)$ вигляду (7) за допомогою міри розпізнавання рівня (різновидність міри Хемінга):

$0 \le \nu = 1\text{-}\max\left\{ \left|a_1^* - b_1^*\right|, \left|a_2^* - b_2^*\right|, \left|a_3^* - b_3^*\right|, \left|a_4^* - b_4^*\right| \right\} \le 1$. Рівень фактору F відповідає лінгвістичній змінній «Середній (С)» при $\nu = 1\text{-}\max\left\{ 0, 0, 0, 0 \right\} = 1 \le 1$.

Отже, при порівнянні отриманої функції належності і ступені оцінки критерія маємо, що при побудованій залежності критеріїв ступінь критерія є «Середнім».

Одним з інструментів дослідження показників системи є математичний апарат теорії ймовірностей, яка працює з випадковими явищами та рядами. Багато стохастичних процесів в природі та техніці володіють довгостроковою залежністю та фрактальною структурою. Найбільщ адекватним математичним апаратом для дослідження динаміки та структури рядів є фрактальний аналіз, який враховує поведінку системи в період вимірювань та її попередню історію.

Параметр Херста H представляє собою міру самоподібності або міру тривалої залежності стохастичного процесу. Значення $H = 0.5$ показує на випадковий ряд та на відсутність довготривалої залежності. Чим ближче H до 1, тим вище степінь стійкості довготривалої залежності. Для оцінки показника Херста існує багато методів, які володіють декотрими недоліками. Порівняльний аналіз статистичних властивостей оцінок показника Херста [3] показав, що збільшення точності оцінки H необхідно використовувати середнє арифметичне виправлених незміщених оцінок, які отримані декількома методами, один з яких бажано обрати метод вейвлет-перетворення. Для визначеності надалі використовуємо R/S-аналіз (метод нормованого розмаху):

$$H = \frac{\ln\left(R / S\right)}{\ln\left(a\tau\right)}, \tag{9}$$

де $R = \max\limits_{1 \le t \le \tau}\left(x^{cum}(t,\tau) \right) - \min\limits_{1 \le t \le \tau}\left(x^{cum}(t,\tau) \right)$- розмах кумулятивного ряду

$$x^{cum}(t,\tau)=\sum_{i=1}^{t}\left(x(t)-\overline{x}(\tau)\right),\ \overline{x}(\tau)=\frac{1}{\tau}\sum_{i=1}^{\tau}x(t);$$

$S=\sqrt{\frac{1}{\tau-1}\sum_{i=1}^{\tau}\left(x(t)-\overline{x}(\tau)\right)^2}$ - середньоквадратичне відхилення ряду;

a - декотре додатне стале число;

τ - кількість періодів спостережень.

При значеннях $0,5<H<1$ у часового ряду трендостійка поведінка і ряд буде зберігати тенденцію росту(спаду) такий же час у майбутньому. Діапазон $0<H<0,5$ відповідає антиперсистентним рядам. Модифікована система переваг одних факторів іншим для одного рівня ієрархії факторів базується на порівнянні міри самоподібності стохастичного процесу:

$$S=\left\{F_{k,i}\Re\, F_{k,j}\mid \Re=\left\{\begin{array}{l}>\textit{відношення переваги},\ H_{k,i}>H_{k,j}\\ \approx \textit{відношення рівноваги},\ H_{k,i}=H_{k,j}\end{array}\right\}\right\}. \quad (10)$$

Отже, всі фактори рівнів ієрархії (*рис. 1*) можуть знаходитися у відношенні:

$$S=\left\{\begin{array}{l}F_1<F_2,\, F_{1.1}>F_{1.2},\, F_{2.1}<F_{2.2},\\ F_{1.1.1}<F_{1.1.2},\, F_{1.2.1}>F_{1.2.2}\end{array}\right\}. \quad (11)$$

На рівні F побудови критерію $F_1<F_2$ і $p_1=1/3$, $p_2=2/3$. Функція належності цього рівня не зміниться в порівнянні з функцією належності в умовах (5) і (6): $\mu^*(x)=\left(0.35\ \ 0.45\ \ 0.55\ \ 0.65\right)$. При побудованій залежності критеріїв ступінь критерія F є «Середнім».

Отже, в лінгвістичному класифікаторі показник Херста для аналізу відношень факторів одного рівня ієрархії (10) з визначеною ваговою системою Фішберна $P^*=\left(p_1^*\ ,\, p_2^*,\,\ldots,\, p_i^*\right)$. не впливає на побудову функції належності критерія найвищого рівня.

Існує декілька варіацій фрактальної розмірності [4]. Для визначення рівнів класифікатора використовується розмірність $A=1/H$ простору ймовірностей для оцінки товщини хвостів в функції щільності ймовірності (по Мандельброту). Якісні оцінки кожного фактору в ієрархії отримані на заданій множині значень показника Херста $H\in\left\{\ 1,\, 2/3,\, 1/2,\, 2/5,\, 1/3\ \right\}$. Співвідношення (2), яке характеризує якісні оцінки кожного фактору в ієрархії G, визначимо так:

 ISSN 1562-9945

$$\mathrm{L} = \left\{ \begin{array}{c} \textit{дуже низький рівень}\,(ДН), \quad 0 < H < 1/3 \\ \textit{низький рівень}\,(Н), \quad 1/3 < H < 2/5 \\ \textit{середній рівень}\,(С), \quad 2/5 < H < 1/2 \\ \textit{високий рівень}\,(В), \quad 1/2 < H < 2/3 \\ \textit{дуже високий рівень}\,(ДВ), \quad 2/3 < H \end{array} \right\}. \qquad (12)$$

Співвідношення (6) для розрахункової моделі виглядає так:

$$\left\{F_{1.1.1}(C), F_{1.1.2}(B), F_{1.2.1}(C), F_{1.2.2}(C), F_{2.1}(C), F_{2.2}(B)\right\}. \qquad (13)$$

На рівні F побудови критерію при системі переваг (11) функція належності може бути записана у вигляді: $\mu^*(x) = (0.47 \;\; 0.57 \;\; 0.67 \;\; 0.77)$. Рівень фактору F відповідає лінгвістичній змінній «Високий (В)» при

$\nu = 1\text{-max}\{\, 0.08, 0.08, 0.08, 0.08 \,\} = 0.92 \le 1$.

Отже, при порівнянні отриманої функції належності і ступені оцінки критерія маємо, що при побудованій залежності критеріїв ступінь критерія є «Високим».

Висновки

Побудована найпростіша комплексна оцінка на основі трапецієвидного числа лінгвістичного класифікатора з використанням показника Херста для виявлення трендовості показників.

Проаналізований алгоритм побудови критерію з врахуванням кількісних характеристик системи для формування початкових умов на кожному з етапів постановки задачі. Алгоритм не дає відповіді на питання про коректність експертної оцінки факторів побудованої ієрархії, про зв’язки між рівнями.

Використання кількісних характеристик ряду з маленькою вибіркою для побудови класифікатору може бути якісною заміною при переході від відношень рівноваги до відношень переваг. Показник Херста є кількісною характеристикою при побудові лінгвістичного класифікатора для аналізу відношень факторів одного рівня ієрархії. Досліджено, що визначена в роботі вагова система Фішберна $P^* = \left(p_1^*, p_2^*, \ldots, p_i^*\right)$ не впливає на побудову функції належності критерія найвищого рівня.

Використання запропонованої модифікації методу може розширити область існування критеріїв комплексної оцінки з лінгвістичним класифікатором при використанні кількісних характеристик показників.

ЛІТЕРАТУРА

1. Недосекин А. О. Комплексная оценка риска банкротства корпорации на основе нечетких описаний. — На сайте:
http://sedok.narod.ru.
2. Yager R. Families of OWA operators // Fuzzy Sets and Systems, 59. 1993.
3. Кириченко Л.О. Сравнительный анализ статических свойств оценки показателя Херста. - Вісник Національного технічного університету "Харківський політехнічний інститут". Збірник наукових праць. Тематичний випуск: Інформатика і моделювання. – Харків: НТУ "ХПІ", 2010. – № 21. – с. 88-95
4. Эрик Найман. Расчет показателя Херста с целью выявления трендовости (персистентности) финансовых рынков [Электронный ресурс]: (Статья). // Э. Найман. 2010. – Режим доступа:
http://www.capital-times.com.ua/index.php?
option=com_content&task=view&id=11623& Itemid=88888963

 ISSN 1562-9945