

ОСОБЛИВОСТІ ВИКОРИСТАННЯ МЕТОДІВ ДИСКРЕТНОЇ МАТЕМАТИКИ В ЗАДАЧАХ ОЦІНКИ КНИЖКОВИХ ВИДАНЬ

Параметр якості книжки можна розглядати наступними способами:

- як один інтегральний параметр
- як окремі параметри інтегрального характеру
- як деяка характеристика, що описує зміну певного інтегрального параметру по відношенню до деяких спеціальним чином вибраних змінних.

Основною особливістю інтегральних параметрів є те, що вони не мають, в більшості випадків, безпосередньої фізичної інтерпретації. Відомо, що існують параметри які є функціями ряду інших параметрів і мають свою безпосередню фізичну інтерпретацію, оскільки відомі способи вимірювання його величини, які не залежні від інших параметрів аргументів. Такого типу параметри в рамках даної роботи будемо називати узагальнюючими параметрами хоча можна було би прийняти протилежну термінологію.

Загальноприйнятою інтегральною інтерпретацією до параметру якості є інтерпретація у відповідності з якою параметр якості характеризує поліграфічне видання (PV) в цілому і, тому, являється інтегральним параметром по відношенню до всіх параметрів, що характеризують відповідний продукт. Цьому випадку відповідає ситуація, коли досить важко конструктивно сформулювати спосіб його визначення.

Як що параметри якості розглядати як деякий ряд окремих інтегральних параметрів, то появляється можливість кожний окремий інтегральний параметр більш конструктивно описати у вигляді деякої залежності. Таких розподіл має сенс в тому випадку як що, як що окремі інтегральні параметри можна вважати незалежними. Ситуація з розподілом параметру якості на окремі складові інтегральні параметри, що незалежні між собою є характерною для багатьох типів продукції включаючи і PV. Останнє можна оцінювати по якості дизайну і кольороподілу, окремо оцінювати по механічних параметрах конструкції книги і т.д.

Очевидно, що якість, як і будь який інший параметр, в процесі експлуатації об'єкту, який описується цим параметром може змінюватися. В цьому випадку можна говорити про якість певного виробу як про фактор, яких визначає на яку величину відповідний параметр зменшується. На якісному рівні можна стверджувати що якість виробу, яким в даному випадку є книжка, визначається тим на скільки швидко зменшується значення параметрів, з допомогою яких книжці визначається та, чи інша оцінка. Тоді, параметр якості є тим вищий, чим довше параметри, що вибрані для її оцінки зберігаються, або на скільки вони змінюються. Переважно, при такому підході, говорять про збереження якості, в процесі використання книжки.

Тоді якість книжки є тим більша, чим довше, чи чим краще зберігаються параметри, що оцінюють відповідну книжку.

Приймаючи до уваги приведені вище можливі способи визначення якості, розглянемо теоретичні засоби, з допомогою яких можна було би побудувати моделі, для визначення величини якості: Якість, як деяка характеристика поліграфічного виробу (PV), має наступні особливості, які визначають способи побудови відповідних моделей:

- в більшості випадків, коли мова іде про якість, то в першу чергу, розуміється, що така характеристика, по своїй суті позитивна і повинна тривати перший період часу;

- якість, як деякий параметр PV являється інтегральним, при цьому, міра інтегральності повинна бути настільки висока, що він не повинен мати своєї однозначної, безпосередньої фізичної інтерпретації;

- параметр якості, або характеристика якості повинна допускати досить широкий спектр варіантів її інтерпретації;

- параметр, або характеристика якості, в найбільшій мірі орієнтована на користувача і, в цьому сенсі, він може характеризуватися, як параметр, що залежить від суб'єктів, які визначають параметри якості;

- різні способи визначення якості, або різні інтерпретації параметра якості повинні бути незалежними від інших визначень якості, чи від інших характеристик, які визначають якість.

Перша властивість, або особливість обумовлюється необхідністю використовувати моделі прогнозування, оскільки якість, як моментальна характеристика PV, практично, є мало придатною. Відомий цілий спектр методів побудови моделей прогнозування [1]. Один з методів побудови моделі прогнозування ґрунтується на використанні рівнянь регресії [2]. Регресією прийнято називати умовне математичне очікування $M_x Y$, яке є функцією від x , що формально записується наступним чином:

$$Y = f(x) = M_x Y, \quad x = q(y) = M_y x,$$

де M_x і M_y – математичні очікування по x і y , а Y і X виборки значень відповідних змінних. В найбільш простому випадку, коли обидві функції регресії $f(x) = M_x Y$ та $q(y) = M_y x$ є лінійними, відповідні функції називаються прямими регресіями. Для того, щоб функції регресії можна було успішно використовувати для розв'язку задач прогнозування, останні повинні бути незалежними від початкових умов. У випадку, коли останні залежать від тих, чи інших факторів, відповідні регресійні функції не можуть забезпечити правильного прогнозування. В першу чергу, це стосується функції гусними розподілу ймовірнісних значень змінних, що являються аргументами відповідних моделей. Якщо функцію регресії записати у вигляді співвідношення:

$$y = \Gamma(x, \theta) + \varepsilon,$$

де $\Gamma(-)$ – функція регресії, що має заданий вигляд, а $\theta = (\theta_0, \dots, \theta_r)$ – коефіцієнти регресії, ε – величина похибки, то можна написати вирази для

квадратичної похибки лінійної регресії і, відповідно, похибки прогнозування, що можна представити співвідношенням:

$$R = \lim_{n \rightarrow \infty} E1 [r(x_0, \theta) - r(x_0, \theta_0)]^2$$

При стійкому розв'язку моделі прогнозування величина R повинна мало мінятися, при малих змінах модельних густин розподілу. Оцінка θ_0 максимуму у правдоподібності мінімізує дисперсію регресійного прогнозування R .

В поліграфічній промисловості, якій характерна досить велика різноманітність в способах реалізації окремих фрагментів технологічного прогресу, функція густини розподілу значень параметрів, що використовуються для оцінки, чи характеристики книжкової продукції може представляти собою для окремих екземплярів продукції випадкову величину з забрудненням відповідної вибірки. Очевидно, що відповідно вибірка випадкових значень параметра розглядається в діапазоні значень, що визначаються рядом умов, до яких відносяться:

- характеристики фрагментів технологічного процесу, що відповідальні за формування відповідних параметрів;
- особливості матеріалу з якого передбачаються виготовляти окремі елементи поліграфічної продукції, особливо в частини, що стосується створення конструкції книжки;
- інших факторів що впливають на відповідний фрагмент технологічного процесу та визначаються випадковими подіями.

Для забезпечення стійкості відповідних моделей по відношенню до забруднень, розв'язується задача визначення таких функцій розподілу густини, для яких оцінка μ_0 при зашумненні ε була би стійкою. Наприклад, для функції розподілу густини, що описується наступним співвідношенням:

$$f(x-\mu) = (1-\varepsilon) \varphi(x-\mu) + \varepsilon h(x-\mu),$$

де функція h є симетрична, або $h(y) = h(-y)$ і не відома функція розподілу густини, з врахуванням асимптотики, що описується співвідношенням:

$$0 < \varepsilon = \varepsilon_n = \gamma n^{-1/2}, \quad 0 < \gamma = \text{const}$$

Отримано, що залежність оцінок від невідомої величини ε та інших оцінюваних параметрів означає, якщо $\varepsilon \rightarrow 0$ повільніше ніж $n^{-1/2}$, що можна записати у вигляді:

$$\varepsilon = \varepsilon_n = \gamma n^{-\xi/2}, \quad 0 < \xi < 1, \quad \gamma = \text{const}$$

Для забезпечення стійкості в рамках такого підходу, необхідно розширити модель розглядаючи серію забруднених виборок. В роботі Шуригіна А.Н. [4], розв'язується відповідна задача.

Якість книжки, як певний інтегральний параметр, що її характеризує, визначається для її оцінки не стільки з ціллю визначення її текущего стану, скільки з ціллю визначення придатності книжки до її експлуатації в тих, чи інших умовах, або при дії на них певних ключових факторів. Визначення таких факторів дозволить вибрати параметри, що в найбільшій мірі залежні від цих факторів. Це дозволяє вибрати параметри, по яких слід реалізувати прогнозування, яке оцінює можливість забезпечення відповідних значень визначених параметрів. Оскільки поліграфічних продуктів є багато

параметричним, а забезпечення заданих значень різним параметрам є досить складною технологічною задачею. Більш того, якість книжки в більшості випадків, визначається не як деякий абсолютний, універсальний параметр, а тісно пов'язується з функціональною орієнтацією відповідного продукту. Це означає, що поліграфічний виріб, що орієнтований на використання в якості довідника в тій чи іншій галузі знань, з точки зору параметра якості, може не відповідати якісному показнику виробу, що орієнтований на використання для навчання в школі.

Основною задачею фактичного аналізу є визначення факторів, що в найбільшій мірі впливають на ті чи інші параметри об'єкту, який досліджуються. Очевидно, що фактори, які діють на вибрані параметри досить тісно зв'язані з останніми. Тому, при ітераційній реалізації системи факторного аналізу, можна по визначених факторах та по ефекту їх дії на досліджуваній об'єкт встановити вимоги до значень параметрів, що у відповідності з встановленими факторами, повинні забезпечуватися технологічними процесами виготовлення відповідного продукту.

Систему факторного аналізу будемо розглядати у вигляді певної структури, що складатиметься з ряду компонент, до яких відносяться:

- компонента, що реалізує безпосереднє обчислення або визначення найбільш значимого фактору;
- компонента аналізу результатів одного етапу факторного аналізу з ціллю переходу до формування моделі по інших факторах;
- компонента визначення величини значення параметрів, що визначаються факторами, які виявилися найбільш значущими для даних параметрів;
- компонента, що реалізує ітераційний перехід моделі факторного аналізу;
- компонента, що реалізує управління системного факторного аналізу.

Компонента, що реалізує безпосереднє обчислення, або визначення найбільш значимого фактору серед всіх факторів, що аналізуються в рамках відповідної моделі. В загальному вигляді така модель може бути описана у вигляді наступного співвідношення:

$$E_y(x_j, \dots, x_m) = b_0 + b_1^{(1)} f_1^{(1)}(x_1) + \dots + b_{j_1}^{(j_1-1)} f_{j_1}^{(j_1-1)}(x_{j_1}) + b_m^{(m)} f_m^{(m)}(x_m) + \dots + b_m^{(m-1)} f_m^{(m-1)}(x_m) + \Pi,$$

де $y(x_j, \dots, x_m)$ – спостереження в точці (x_j, \dots, x_m) , Π вміщає члени з добутками $k_{j_1 \dots j_r}^{(j_1-1) \dots (j_r-1)} f_{j_1}^{(j_1-1)}(x_{j_1}) \dots f_{j_r}^{(j_r-1)}(x_{j_r})$, $k_{j_1 \dots j_r}^{(j_1-1) \dots (j_r-1)}$ – константи ($ij \neq \dots \neq ir$), система функцій $f_1^{(1)}(x_{j_1}), \dots, f_{j_1}^{(j_1-1)}(x_{j_1})$ лінійно незалежна в точках x_{j_1}, \dots, x_{j_1} функції $f_{j_1}^{(j_1-1)}(x_{j_1})$ можуть представляти собою поліноми q - тої степені від x_i [5]. Практично для використання факторних моделей використовуються моделі з обмеженою кількістю факторів, наприклад двох факторні моделі. Можливість такого спрощення ґрунтується на тому, що більшість параметрів, що

характеризують виріб зв'язані між собою різною мірою сили такого зв'язку, або зв'язки між різними параметрами мають різну тісність таких зв'язків. Можна прийняти, що ключові параметри найбільш тісно зв'язані з двома чи з трьома параметрами, факторами. Наприклад, у випадку, коли приймаються, як найбільш тісно зв'язаними з деяким параметром, два фактори, то отримуємо двох факторну модель. Факторний аналіз досить часто називають дисперсійним аналізом [6]. В цьому випадку факторну модель можна представити у наступному вигляді:

$$x_{ij} = \mu_0 + d_i + \beta_j + \varepsilon_{ij}, \quad i = \overline{1, I_A}, \quad j = \overline{1, I_B}$$

де x_{ij} – відгук x на i – тому рівні фактора A і j – тому рівні фактора B , $\mu_0 = MX_i$, або $\mu_0 = MX_i$, d_i і β_j – не випадкові величини, що характеризують вклади в x_{ij} , що обумовлені дією відповідних факторів A і B , ε_{ij} – випадкова величина, що характеризує вклад в x_{ij} , які обумовлений дією факторів, що в рамках цієї моделі не враховуються. Якщо прийняти, що ε_{ij} – випадкова величина, що визначається помилками експерименту і мають місце співвідношення $M \varepsilon_{ik} = 0$, випадкові помилки експерименту не корильовані і мають однакову дисперсію, то ε_{ik} має нормальний закон розподілу з нульовим середнім значенням, та невідомою дисперсією, що записується у вигляді $\varepsilon_{ik} \sim N(0, \sigma^2)$. Приймається, що фактори є незалежними. Завдяки цьому перевірка гіпотез про вплив факторів A і B на відгук X проводиться окремо для кожного фактора. Критерії перевірки гіпотез, про вплив факторів A і B на відгук X , ґрунтуються на порівнянні статистики:

$$F = [S_A^2(\bar{X}_i)] / [S_B^2(\bar{X}_j)] \sim F(I_A-1, (I_A-1)(I_B-1))$$

З вибірковою значенням f_B . таке порівняння полягає у визначенні, чи f_B не перевищує $f_{кр} = f_{1-2}(I_A-1, (I_A-1)(I_B-1))$ для заданого рівня значності α . В протилежному випадку, коли $f_B > f_{кр}$, гіпотезу H_0 , про те, що фактор A не впливає на відгук X , відхиляється. В цьому випадку, може виникнути необхідність у перевірці одної з гіпотез $H_0^{(i)}$, згідно з якого вплив га відгук здійснює i -й рівень фактора A . Це означає наступне:

$$H_0^{(i)} : \alpha_1 = \dots = \alpha_{i-1} = \alpha_{i+1} = \dots = \alpha_{I_A} = 0, \alpha_i \neq 0$$

Завдяки використанню факторних моделей стає можливим встановити числові взаємозв'язки між факторами, що впливають на ті, чи інші параметри та визначити міру такого впливу на основі статичних даних. Зрозуміло, що, по крайній мірі, на якісному рівні такі зв'язки фахівцям в поліграфічній галузі є відомими. Але в багатьох випадках дані про ті чи інші можливості виробників поліграфічної продукції відомі з відповідних рекламних матеріалів. Крім того, в рамках поліграфічних виробництв можуть існувати в межах відповідних фірм корпоративні дані про ті чи інші можливості фірми, особливо, коли вони стосуються технологічних характеристик. Це може обмежувати можливості до доступу до даних, які є необхідними для проведення тих, чи інших оцінок можливої поліграфічної продукції і, відповідно, оцінок можливостей окремих поліграфічних

підприємств. Таким чином, проблеми оцінки якості поліграфічної продукції можуть розглядатися в аспекті оцінки окремих поліграфічних виробництв, що являється актуальним для управління виробництвом в умовах ринкових відносин, яким характерна досить висока конкуренція між окремими організаціями.

Процес оцінки книжки, як окрема подія, що підлягає аналізу є досить специфічним і представляє собою процес, який сам по собі має одиничний характер, навіть у тому випадку, коли відповідна книжка розглядається як типовий представник цілого тиражу відповідного видання. Це обумовлено тим, що визначення деякого екземпляру, як типового представника тиражу є суб'єктивним. Такий суб'єктивізм в подальшому аналізі окремих книжок тиражу може посилювати недоліки, або наближення, що характерні для методів, які використовуються для оцінки книжки. Тому в методах, що ґрунтуються на використанні моделей регресії, моделей факторного аналізу, в основу положено аналіз відповідної множини вибірок, які представляють собою статичні дані про певну сукупність об'єктів одного класу, по відношенню до яких реалізуються обчислення вибраних параметрів. Логічно припустити, що поліграфічний продукт певного класу, якими можуть бути клас книжкової продукції з їх різновидностями, що формують підкласи, тип продукції, що відноситься до класу журнальних видань з відповідними різновидностями та інші класи поліграфічної продукції, в рамках одного класу повинен змінюватися від тиражу до тиражу з точки зору зміни інтегрального параметру, яким, в даному випадку є якість книжки. Очевидно, що з точки зору зміни якості поліграфічного продукту остання може не мінятися між двома послідовними тиражами. Але одна з парадигм ринкової конкуренції вимагає постійного підвищення якості продукції, що випускається. Очевидно, що зміна параметру якості чергового продукту може бути фізично не видимою. Ця обставина обумовлюється тим, що, як уже відмічалось, уявлення про якість як деякий параметр є таким, що останній являється, з точки зору його фізичної інтерпретації, таким який не має загальноприйнятої одиниці вимірювань своєї величини в числовій формі. Тому, в цьому сенсі, можна стверджувати, що такий параметр, особливо величина його оцінки, є певною абстракцією. Наприклад, фізично зауважальна зміна параметрів конструкції книжки, що полягає у зміні її розмірів, у використанні більшої гами кольорів, при оформленні книжки, та інші, не обов'язково свідчать про підвищення якості такого виробу. Тим не менше, прийемо, що складовою параметра якості книжки може бути характеристика, яка має фізичну одиницю вимірювання.

Приймаючи до уваги приведену вище парадигму, що обумовлює необхідність підвищення якості книжки, можна говорити про оцінку книжки, яку планується видавати. Прийемо, що оцінка якості проводиться з цілло формування вимог до відповідного видання. У випадку використання рекурентних моделей для оцінки якості поліграфічного продукту, вихідними даними, являються дані про книжку цього ж класу, що уже випускались

відповідним виробником. При цьому, можна формувати рекурентні моделі заданого порядку. В загальному вигляді рекурентну модель можна представити у вигляді лінійного рекурентного рівняння з сталими коефіцієнтами [7].

$$f(x+k) + a_1 f(x+k-1) + a_2 f(x+k-2) + \dots + a_k f(x) = 0, \quad (2.1)$$

де a_1, \dots, a_k – сталі коефіцієнти, $f(x+k-m)$, де $m < k$, функція що описує один рекурентний крок. Очевидно, що функція $f(x)$ повинна описувати нове значення параметра якості книжки. Тому розв’язок цього рівняння дозволяє отримати бажане значення параметра якості на заданому кроці його підвищення, що визначається рангом відповідного рівняння k . Оскільки уявлення про якість книжки, в даному випадку, залежить від ряду вибраних фізичних параметрів, то для використання відповідної моделі необхідно побудувати систему інтерпретації параметра якості, що відповідає змінній X у приведеному рівнянні. Така система інтерпретації параметра якості, що описує його залежність від параметрів, що характеризують книжку і мають власні фізичні міри, разом з рекурентним рівнянням представляє собою рекурентну модель визначення, або оцінки якості поліграфічного продукту.

Використання рекурентних моделей для визначення якості книжки можна було би вважати тривіальним, як би останні, в рамках досліджувальної предметної області не мали певних особливостей.

Перша з особливостей, яку необхідно відмітити полягає у тому, що в класичному випадку, розв’язок рівняння (2.1) шукається у вигляді $f(x) = \lambda^x$, де λ – невідоме число, $\lambda \neq 0$. Тоді з рівняння (2.1) можна отримати рівняння для визначення числа λ :

$$\lambda^k + a_1 \lambda^{k-1} + a_2 \lambda^{k-2} + \dots + a_k = 0. \quad (2.2)$$

Це рівняння прийнято називати характеристичним рівнянням для рекурентного рівняння (2.1). якщо числа $\lambda_1, \dots, \lambda_k$ є коренями рівняння (2.2), то для довільних сталих c_1, c_2, \dots, c_k можна записати:

$$f(x) = c_1 \lambda_1^x + c_2 \lambda_2^x + \dots + c_k \lambda_k^x,$$

де $f(x)$ є функцією, що являється розв’язком рівняння (2.1). Якщо λ має кратність l_k , то можна попереднє співвідношення записати у вигляді:

$$f(x) = (c_k + c_{k-1} x + \dots + c_{k-l_k+1} x^{l_k-1}) \lambda_k^x.$$

Приведені співвідношення ілюструють можливість введення інтерпретації параметрів від яких залежить якість продукту окремих функцій ітераційного процесу рекурентної моделі.

Друга особливість рекурентної моделі полягає у тому, що $f(x)$ з (2.1) може розглядатися як породжуюча функція для послідовності функцій F_0, F_1, \dots, F_n , що в загальному вигляді записується співвідношенням:

$$f(x) = F_0 + F_1 x + F_2 x^2 + \dots + F_k x^k$$

В цьому випадку можна шукати таку інтерпретацію окремих параметрів, що входять в склад параметрів якості, або формують його, яка відповідала би окремим функціям, що породжуються відповідним рекурентним рівнянням.

В цьому випадку, необхідно прийняти до уваги, що окремі параметри, що приймаються складовими для $f(x)$ повинні бути незалежними. Це означає, що в кожній $F_i x^i$, змінні аргументи повинні бути зв'язані відповідними предикатом F_i . Практично, досягнути виконання цієї умови є складно, оскільки в рамках єдиного технологічного процесу виробництва книги, в більшості випадків параметри, що його характеризують і відповідно, визначають параметри продукту, які мають власну фізичну інтерпретацію, зв'язані між собою. Тим не менше, задовольнити сформульовану вимогу можна в процесі формування частини рекурентної моделі, що представляє собою опис інтерпретації елементів предметної області. Інші деталі та особливості використання рекурентних моделей доцільно аналізувати при їх побудові.

1. *Піх І.В.* Параметричне прогнозування якості друкованої продукції . Квалілогія книги. 36. наук. прощ. Львів, 2003, УАД вип. 6. с.200-213
2. *Боровков А.А.* Теория вероятностей. М.: Наука, 1976, -352с.
3. *Шуригин А.М.* Прикладная статистика: робастность, оценивания, прогноз. М.: Финансы и статистика, 2005. – 224 с.
4. *Шуригин А.М.* Робастность и устойчивость статистических оценок. / Статистика, вероятность, экономика. Сб., науч. ст. М.: Наука, с. 90-98
5. *Бродский В.З.* Модели факторных планов. – Тезисы докладов IV Всесоюзной конференции по планированию и автоматизации эксперимента в научных исследованиях, часть 1, М.: 1973, с. 36-38
6. *Бочаров П.П., Печинкин А.В.* Теория вероятностей. Математическая статистика. М.: Гардарика, 1998. – 328 ст.
7. *Ландо С.К.* Лекции о производящих функциях. М.: МЦНМО, 2004. – 144с.

Поступила 16.09.2010р.

УДК 681.322

Ю.В.Стех, к.т.н., доц. каф. САП, НУ «Львівська політехніка»,
Файсал М.Е. Сардіх, аспірант каф. САП НУ «Львівська політехніка»,
М.В.Лобур, д.т.н., професор, зав. каф. САП, НУ «Львівська політехніка»,
М.С. Домброва, аспірант каф. САП НУ «Львівська політехніка»,
В.Є. Арцибасов, аспірант каф. САП НУ «Львівська політехніка»

АДАПТИВНИЙ МЕТОД І АЛГОРИТМ ПОШУКУ ЦЕНТРІВ КЛАСТЕРІВ ЗА ДОПОМОГОЮ НЕЙРОННОЇ МЕРЕЖІ

Developed an adaptive method and algorithm for finding the cluster centers by using neural network. The developed method requires a set threshold. Proposed heuristic computational formula for determining the optimal value of threshold. An