

Л.Я. Нич, Р.М. Камінський
Національний університет “Львівська політехніка”,
кафедра інформаційних систем та мереж

ОЦІНЮВАННЯ ЕФЕКТИВНОСТІ ІНДИВІДУАЛЬНОЇ ІНТЕЛЕКТУАЛЬНОЇ ДІЯЛЬНОСТІ ОПЕРАТОРІВ У СИСТЕМАХ ОПРАЦЮВАННЯ ЗОБРАЖЕНЬ НА ОСНОВІ ЕКСПЕРИМЕНТАЛЬНИХ ДАНИХ

© Нич Л.Я., Камінський Р.М., 2011

Теоретично обґрунтована і експериментально підтверджена доцільність використання емпіричної функції розподілу для оцінювання ефективності інтелектуальної діяльності операторського персоналу в системах візуального опрацювання динамічних образів.

Ключові слова: операторська діяльність, оцінювання ефективності роботи, MS Excel 2003, часова послідовність, бета-розподіл.

Legitimacy of the use of empiric function of beta-distribution for the efficiency estimation of intellectual activity of operator personnel in the systems of the visual processing of dynamical objects is theoretically reasonable and experimentally is confirmed.

Key words: operator activity, evaluation of efficiency of work, MS Excel 2003, time sequence, beta-distributing.

Вступ

Сучасний стан інформаційних технологій у своєму розвитку і за рахунок створення відповідного математичного і програмного забезпечення та розширення сфер застосування значно випереджує процеси підготовки висококваліфікованого персоналу.

Виникнення нових завдань, орієнтованих на повне використання обчислювальної техніки та потреба зміни існуючих підходів до розв'язання нових задач в різних предметних областях, вимагають якісної підготовки фахівців і передусім в стислі терміни.

У цьому напрямі працюють численні організації, оснащені не лише тренажерами, психологічними лабораторіями та дослідницькими стендами, але й відповідним сучасним методичним забезпеченням. У підготовці фахівців беруть участь педагоги, психологи, фізіологи, медики; залучаються біологи, фізики, інженери. Важливу роль в обробленні експериментальних даних, розвитку методології побудови математичних моделей, методів, методик, презентації та інтерпретації розроблених моделей, принаймні з точки зору формального пояснення отриманих результатів, відіграють математики.

Проте будь-який навчальний процес відповідає своєму призначенню лише тоді, якщо після його завершення можна встановити об'єктивний рейтинг його учасників – кандидатів на операторські посади. Тому важливе значення має (в кожному окремому випадку) впровадження і використання найадекватнішої та найоб'єктивнішої, тобто на основі власне індивідуальних даних, системи оцінювання основних фахових показників кожного з претендентів.

Підготовка висококваліфікованого персоналу, зокрема операторського, майже у кожному конкретному випадку використовує спеціально адаптовані методики, сценарії та критерії оцінювання результатів, орієнтовані на індивідуальний підхід. Операторська діяльність у багатьох системах опрацювання зображень, наданих операторові на моніторі після попередньої підготовки і обробки, характеризується тим, що вимагає значного інтелектуального ресурсу людини-оператора для підготовки або прийняття рішення і високої відповідальності за наслідки.

У таких системах інформація у формі зображень відображає дійсний стан ситуації (управління технологічними процесами), або минулий (ретроспективний, коли інформація була зафіксована в минулі моменти часу, аерокосмічні знімки, результати дефектоскопії, рентгенівські знімки тощо). Самі зображення у першому випадку відображають у той чи інший спосіб динаміку об'єктів уваги (це зміна показів приладів, зміна значень у відповідних комірках зображуваного табло, подані віртуально на екрані монітора та виділені світловим способом і відповідним кольором на мнемосхемах).

У другому випадку розглядаються статичні зображення (зональні знімки поверхні землі, рентгенограми, а також тексти, графіки, креслення). До таких зображень належать реальні та реалістичні сцени з оточуючого середовища та робочих місць, репродукції картин та історичні пам'ятки, що є відсканованими фотокопіями. Інколи такі зображення також дають змогу виділити рух об'єктів за їхньою динамікою, аналізом змазування окремих фрагментів, або в результаті порівняння двох і більше зображень цієї сцени, знятих у різні моменти часу.

Роль операторського персоналу полягає у тому, що в результаті опрацювання картини деякої ситуації, графічної або текстової інформації, наданих на моніторі, причому не змінюючи їхнього фізичного стану, оператором приймається відповідне рішення. Термін *опрацювання*, як його подає [1], означає розумову діяльність і загалом визначає індивідуальну інтелектуальну діяльність конкретного оператора на його робочому місці за розв'язком поставленої перед ним задачі.

Суть цієї діяльності у цьому дослідженні полягає у тому, що на наданому зображенні треба виявити об'єкти уваги заданого класу та ідентифікувати їх. Такі об'єкти уваги відображають поведінку реальних об'єктів в їхньому середовищі. Якщо ці реальні об'єкти в той чи інший спосіб змінюють свою поведінку так, що порушують визначені вимоги, оператор вибирає і приймає рішення на основі власного досвіду, знань та усіх йому доступних інформаційних джерел, і відновлює параметри цих об'єктів. Рішення можуть бути різними, але в будь-якому випадку вони не мають допустити небажаних наслідків. В одних випадках наслідки таких рішень можуть бути виправлені самим оператором, а в інших – вони є незворотні (катастрофи, нештатні ситуації). Здебільшого операторський персонал є штатним і, як правило, включає висококваліфікованих фахівців.

Підготовка операторів для різних систем здійснюється різними методами та засобами, в основі яких лежать методи професійної орієнтації, відбору, навчання, атестації, переатестації. Останній захід відбувається у випадках тривалої перерви, зміни профілю діяльності та вияснення рівня кваліфікації у зв'язку з підвищенням відповідальності. Усі ці заходи є прерогативою психологів і ергономістів. Проте, незважаючи на достатньо розвинені і ефективні методології, які використовуються в психології та ергономіці, існує певна проблема у виборі і оцінюванні різноманітних показників саме для інтелектуальної діяльності. На практиці як такі показники переважно використовують середні значення та інтервальні оцінки. Такі оцінки дають дуже загальне "середньостатистичне" уявлення про підготовку персоналу і все ж не повною мірою відображають інформацію, яка міститься в отриманих експериментальних даних.

Неповнота висновків за такими оцінками полягає у тому, що в статистичному плані отримані в результаті обробки численні характеристики різносторонньо характеризують вибірку цілим комплексом показників, який називають описовою статистикою, серед яких є лише один усереднений – середнє арифметичне. Проте інші показники переважно ігноруються. Достовірність прийняття правильного рішення можна визначити відношенням кількості правильних рішень до кількості усіх прийнятих оператором рішень, а це означає, що достовірність можна подати як імовірність прийняття правильного рішення. Оцінка оперативності полягає у тому, що, оскільки тривалість часу від моменту появи об'єкта уваги до моменту прийняття рішення є випадковою величиною, то необхідно визначити її закон розподілу, побудувати модель даних і на цій основі розробити відповідний критерій такої оцінки.

У цій роботі обґрунтовується правомірність і доцільність використання для побудови інтегрального кваліфікаційного показника не лише окремих числових характеристик вибірки з описової статистики, але і безпосередньо самого закону розподілу, її варіант, точніше його функції

щільності. Як такий закон розподілу, який також є обґрунтований логікою моделі експериментального дослідження, є бета-розподіл. Крім того, цей розподіл є досить добре вивчений, а його “гнучкість” залежно від параметрів забезпечує великий спектр форм функції щільності.

Об’єктом експериментального дослідження є процес опрацювання зображень на моніторі, який визначає оперативність людини-оператора в задачах розпізнавання рухомих об’єктів. Значення оперативності є фактично неперервною обмеженою величиною у скінченному вибірковому просторі. Саме час опрацювання, зафіксований в експерименті, дає підстави розрізнити операторів, оскільки його латентними складовими є власне когнітивні процеси.

Ефективність операторів в сенсі оцінки їхньої діяльності ґрунтується на двох складових – показниках, які можуть бути кількісно вимірюваними, а саме: достовірності в тому чи іншому сенсі, прийняття у цій ситуації оптимального рішення і оперативності – тривалості самого процесу пошуку, аналізу фрагмента зображення з об’єктом та ідентифікації останнього, і на підставі отриманої інформації – здійснення вибору або створення цього рішення.

Ця робота обґрунтовує доцільність вибору індивідуальних показників ефективності інтелектуальної діяльності операторів на основі форми кривої функції щільності бета-розподілу, параметри якої визначені за результатами експериментальних досліджень.

Постановка проблеми та її актуальність

Опрацювання зображень на моніторі, як правило, включає пошук об’єктів уваги та їхню ідентифікацію з об’єктами заданого класу. Задля цього часто необхідно аналізувати як середовище, яке оточує цей об’єкт уваги, так і поведінку інших об’єктів, вникаючи в суть зображення наданої сцени. У таких ситуаціях оператор, напружуючи пам’ять, створює в уяві різноманітні варіанти можливих наслідків, рішень та способів виходу з критичних ситуацій, покладаючись на своє розуміння ситуації, на свій досвід, уміння і навички. Усе це вимагає великих розумових зусиль, високої професійної підготовки і здебільшого супроводжується великим нервовим напруженням. Якщо врахувати монотонність такої діяльності та гіподинамію положення оператора ще й за значної відповідальності, можна стверджувати, що така діяльність є дуже складною і вимагає ретельного професійного відбору операторського персоналу. На жаль, дати характеристику людині-оператору за допомогою лише одного якогось показника є некоректним, тому у таких ситуаціях використовують цілу систему медико-біологічних та психологічних характеристик. На жаль, дуже мало місця у таких системах відводиться показникам, отриманим на підставі експериментальних досліджень. Річ у тому, що безпосередньо ми можемо вимірювати лише кількість подій та інтервали часу між ними. Проте закладена у цих показниках інформація за належної математичної обробки та фахової інтерпретації може бути дуже ваговою. Використання сучасних засобів та методів для дослідження інтелектуальної діяльності, включаючи лабораторно-дослідні стенди-тренажери, їхнє професійне забезпечення у формі зображень реалістичних сцен та найефективніші методи з арсеналу математичної статистики і теорії ймовірності дає підстави сподіватися отримати об’єктивні показники інтелектуальної діяльності людини-оператора.

У цій роботі на підставі аналізу експериментально отриманих нестационарних часових послідовностей і бета-розподілу їхніх рівнів пропонується метод оцінювання індивідуальних показників оперативності персоналу, які дають об’єктивні підстави для ідентифікації операторів.

Сьогодні психологи використовують великий арсенал готових методик, які не завжди відповідають конкретним цілям. Переважно такі методики стосуються загальноприйнятих підходів щодо обробки даних та значно рідше стосовно побудови математичних моделей. Запропоновані в роботі методи аналізу експериментальних даних є доволі ефективними, а сама методологія є повністю реалізована засобами табличного процесора MS Excel. Враховуючи, що запропоновані методи можуть бути використані для будь-якого типу часових послідовностей, а також в різних предметних областях, ця розробка є актуальною.

Новизна цієї розробки полягає у створенні методу оцінювання ефективності інтелектуальної діяльності операторського персоналу на основі часових послідовностей та апроксимації значень їхніх елементів бета-розподілом за допомогою табличного процесора MS Excel. Робота має як

методологічне, так і прикладне значення, оскільки дає інструктивні рекомендації щодо застосування розробленого методу, включаючи і організацію експериментальних досліджень.

Описання та результати експериментального дослідження

Постановка завдання. Завдання оцінювання ефективності індивідуальної інтелектуальної діяльності операторського персоналу може бути коротко сформульована так: за результатами спостережень за вхідними та вихідними змінними об'єкта дослідження – індивідуальної операторської діяльності побудувати критерій, який в певному сенсі оптимально забезпечує встановлення рейтингу таких об'єктів. Отже, об'єктом експериментального дослідження є конкретна індивідуальна інтелектуальна діяльність кожного з групи операторів.

Вхідними змінними є зображення-кадри деякої реалістичної гіпотетичної ситуації, в процесі перебігу якої у випадкові моменти часу з'являються групи послідовних кадрів, які відображають рух заданого об'єкта уваги. Ця ситуація є статичним зображенням рівномірно розподілених на однорідному тлі плямоподібних (випадкової форми) об'єктів-завад, які імітують хмари. Як об'єкт уваги використано рівнобедрений трикутник 16×8 пікселів темного кольору, який рухається прямолінійно вертикально або горизонтально. Об'єкт з'являється з-під стилізованого зображення хмари і так само зникає під таким самим стилізованим зображенням іншої хмари, що знаходиться на його шляху.

Вихідними змінними є тривалості часу – з моменту появи першого пікселя зображення об'єкта уваги – у цей момент включається таймер, і до моменту прийняття рішення – натискання відповідної клавіші, якою зупиняється таймер. Отже, вихідні змінні визначаються тривалістю часу опрацювання групи кадрів, що відображають рух об'єкта уваги.

Отже, на підставі експериментальних досліджень має бути побудована математична модель індивідуальної оперативності опрацювання зображень сценарію, зображеного на моніторі, а саме: виявлення об'єктів уваги, що рухались з постійною швидкістю і були видимі протягом короткого проміжку часу – від 1 до 3 с, залежно від відстані між об'єктами-завадами цієї сцени, які їх закривали. Моменти появи об'єктів випадкові і розподілені рівномірно в інтервалі 30 – 90 с. Час пошуку і виявлення об'єктів, тобто час опрацювання зображення сцени, фактично повністю визначається функціональним станом оператора, роботою зорового аналізатора та мисленими процесами мозку, і становить основну частину інтелектуальної діяльності оператора. Тривалість експериментів становила 80 – 100 хв. Як оператори брали участь 18 осіб віком 19 – 22 роки.

Обробку отриманих індивідуальних результатів докладно подано нижче у формі таких етапів:

Подання результатів експериментів. Отримані дані є фіксацією подій – часу пошуку та прийняття рішення про виявлення об'єкта уваги – у формі послідовності значень часу з “прив'язкою” до моментів часу проведення експерименту. Оскільки моменти появи подій – зображень з об'єктами уваги, тривалість їхнього пошуку та моменти прийняття рішень про виявлення цих об'єктів є випадковими величинами і загалом відображають одновимірний дискретний випадковий процес, який характеризує оперативність персоналу в процесі діяльності. Такі дані можна подати двома способами: у формі часових рядів та часовими послідовностями або імпульсними потоками. На рис. 1 зображено одну з отриманих індивідуальних часових послідовностей.

Перший підхід в термінологічному плані є дещо некоректним, оскільки в математиці поняття ряду переважно пов'язане з детермінованими, чітко впорядкованими величинами, отриманими в результаті реалізації деякого конкретного алгоритму, наприклад, ряд Тейлора, ряд Маклорена, ряд Фур'є тощо. Крім того, різниця між елементами-доданками у таких рядах знаходиться у певній залежності, наприклад, геометрична прогресія. Зауважимо, що під ознаку ряду підпадає і термін варіаційний ряд – впорядкована за зростанням сукупність випадкових величин значень варіант, але тут головним атрибутом категорії “ряд” є саме порядок варіант.

З практичного погляду використання часових рядів для побудови системи методів та критеріїв оцінювання ефективності операторського персоналу на підставі експериментальних даних

вимагає аналізу випадковості рівнів, виявлення коливних та циклічних складових та встановлення наявності тренду і побудови його математичної моделі у вигляді гладкої неперервної функції.

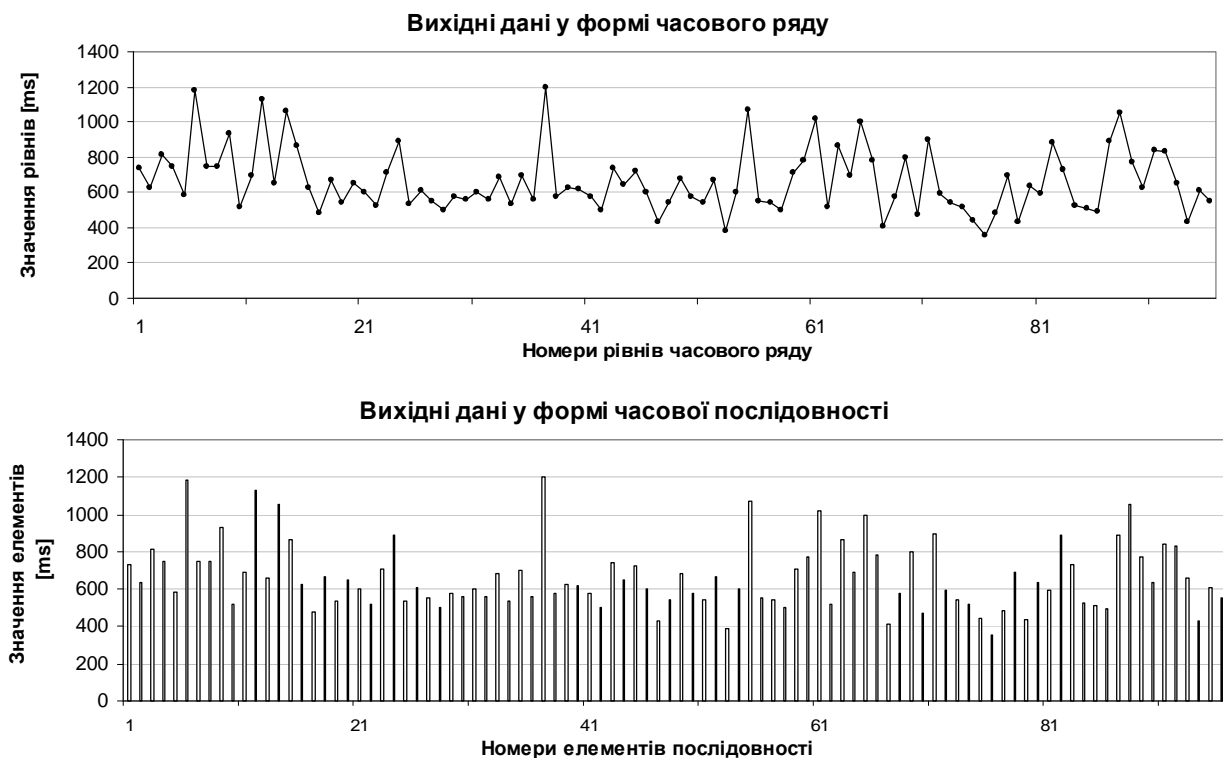


Рис. 1. Подання вихідних даних часовим рядом (вгорі) та часовою послідовністю (внизу)

Для побудови загальної математичної моделі часових рядів використовують такі аналітичні вирази:

$$y(t) = m(t) + j(t) + u(t) + x_t; \quad (1)$$

$$y(t) = m(t) + j(t) + x_t; \quad (2)$$

$$y(t) = m(t) + x_t. \quad (3)$$

У рівнянні (1) $m(t)$ означає тренд, $j(t)$ – коливну, а $u(t)$ – циклічну складові. У рівнянні (2) відсутня циклічна складова. У таких моделях за значної коливальності часто приймають $m = const$. Рівняння (3) описує переважно тенденцію динаміки показника. В усіх трьох рівняннях $x_t = x(t)$ – є випадковою складовою з невідомим законом розподілу, яка включає випадковий характер відхилень від тренду, зумовленого когнітивними процесами самого оператора та помилками і похибками реєстрації. За значної дисперсії значень останніми можна знехтувати і подати x_t як вибірку однаково розподілених випадкових величин. Проте проблема полягає у тому, що невідомо, який розподіл мають ці величини. Найважливішим для побудови системи методів та критеріїв оцінювання операторського персоналу за експериментальними даними, поданими моделями часових рядів, є встановлення наявності тренду та побудова його математичної моделі у вигляді гладкої неперервної функції.

Другий підхід у цій роботі, для означення якого використано термін **часова послідовність**, подає експериментальні дані, як послідовність значень, що характеризують досліджувані нами події, отримані в послідовні випадкові моменти часу. Значення, які власне характеризують ці події, називатимемо елементами часової послідовності. Вони визначають спостережувані значення конкретного показника, а саме: оперативності опрацювання зображень у кожний конкретний момент часу, який є випадковою величиною. Отримані експериментально, такі часові послідовності характеризують індивідуальні значення часу опрацювання наданого зображення і фактично вони є дискретними одновимірними випадковими процесами.

Особливостями часових послідовностей є:

- події, що є однорідними, утворювані одним і тим самим джерелом;
- події, що визначаються кількісними та якісними характеристиками і можуть приймати як детерміновані, так і випадкові;
- події, що можуть бути розподілені в часі спостереження за ними, як рівномірно, так і випадково;
- значення величин, що характеризують події, пов'язані з моментами часу фіксації, проте вони можуть залежати або не залежати від цих моментів;
- якщо члени часової послідовності є створювані одним і тим самим джерелом і за однакових умов, так що $P\{y_i < y\} = P\{y_j < y\}$ для $i \neq j$, тоді їхні значення можна вважати однаково розподіленими і застосовувати до них методи статистичного аналізу випадкової вибірки;
- моменти настання подій можуть бути регулярними, якщо інтервали часу між сусідніми моментами є однаковими і випадковими, підпорядкованими деякому закону розподілу.

Модель часової послідовності у цьому випадку включає дві складові, які можна розглядати незалежно. Перша складова є послідовністю випадкових моментів появи зображень з об'єктами уваги зі своїм власним розподілом, а друга – послідовністю значень сумарного часу: виявлення, ідентифікації та вибору і прийняття рішення оператором, які можна розглядати як вибірку сукупність даних зі своїм власним розподілом.

У цьому плані цей тип експериментального дослідження підпадає під клас послідовності незалежних випробувань, в яких кожна подія – виявлення об'єкта уваги, має імовірність p . Такий процес характеризується біноміальним розподілом ймовірностей, за якого випадкова величина X приймає цілочислові значення $k = 0, 1, 2, \dots, n$ з ймовірностями

$$P(x = k) = C_n^k p^k (1 - p)^{n-k},$$

і характеризується двома параметрами: цілим числом $n > 0$, яке називають числом випробувань, та дійсним числом p – імовірністю успіху, $0 < p < 1$. Фактично, цей розподіл є найпростішою моделлю для першої складової часових послідовностей.

Елементи послідовності як вихідні дані відповідають часу опрацювання наданого зображення на моніторі: з моменту його появи і до моменту прийняття рішення про виявлений об'єкт уваги. У проведених експериментах усі оператори працювали за одним сценарієм, тобто послідовність появи зображень з об'єктами заданого класу була однією і тією самою. За такої умови оцінювання ефективності операторської діяльності для цієї групи спрощується, оскільки використовується лише сукупність значень цих елементів часових послідовностей, а саме – використовується обмежена вибірка даних.

Тривалість опрацювання навіть одних і тих самих зображень є різною і випадковою з невідомим законом розподілу. Встановити адекватний закон для таких даних є дуже складно, оскільки врахувати індивідуальні властивості інтелектуальної діяльності фактично неможливо. На практиці використовують апроксимуючі закони, які вибирають такими, щоб використовувати для різних операторів один закон, щоб індивідуальні відмінності операторів різнилися значеннями його параметрів. Це питання є одночасно і важливим і складним. Одним з чинників тут є те, що для репрезентативної вибірки навіть не дуже відповідний апріорний (у той чи інший спосіб вибраний) розподіл істотно не впливає на апостеріорний. Іншим важливим чинником є складність визначення параметрів вибраного розподілу. На вибір апріорного розподілу також впливає його узгодженість з апостеріорним, тобто і апріорний і апостеріорний розподіли є одним і тим самим розподілом, але з різними значеннями параметрів. У цьому сенсі узгодженість розглядається з точки зору перевірки гіпотез.

Зробивши припущення, що дані операторів підпорядковані одному закону розподілу, і враховуючи зв'язок біноміального розподілу з бета-розподілом як апріорного розподілу, прийнятий бета-розподіл. Тому основною метою обробки даних є визначення параметрів саме бета-розподілу.

Функція щільності бета-розподілу в [2] подана таким виразом:

$$b(x; a, b) = \frac{(a+b+1)!}{a!b!} x^a (1-x)^b, \quad 0 < x < 1; a, b \geq -1,$$

в [3, 4] вона подається через неповну бета-функцію $B(a, b)$:

$$b(x; a, b) = \begin{cases} 0 & \text{за } x \leq 0, \\ \frac{1}{B(a, b)} x^{a-1} (1-x)^{b-1} & \text{за } 0 < x \leq 1, \\ 0 & \text{за } x > 1, \end{cases} \quad (4)$$

де a і b – параметри розподілу, а $B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$, $\Gamma(\cdot)$ – гамма-функція Ейлера.

Визначити значення параметрів a і b можна за значеннями середнього арифметичного та дисперсії, проте це є лише наближені оцінки, за якими можна визначити початкове для оптимізації цих параметрів значення бета-функції. Для знаходження оцінок параметрів a і b використовують такі вирази [5, 6]:

$$\hat{a} = \bar{x} \cdot \left[\left[\frac{\bar{x} \cdot (1-\bar{x})}{s^2} \right] - 1 \right] \quad \text{та} \quad \hat{b} = (1-\bar{x}) \cdot \left[\left[\frac{\bar{x} \cdot (1-\bar{x})}{s^2} \right] - 1 \right], \quad (5)$$

де \bar{x} – середнє арифметичне; s^2 – дисперсія.

У літературі бета-розподіл подається, як визначений на сегменті $[0, 1]$. Приведення індивідуальних розподілів в межі цього сегмента, тобто масштабування значень, дає змогу ідентифікувати операторів шляхом зіставлення графічних форм індивідуальних функцій щільності розподілів, побудованих на підставі знайдених параметрів. Для цього значення x_i , $x_i \in (x_{i\max} - x_{i\min}]$, $i = 1, 2, \mathbf{K}$, тривалості опрацювання зображення шляхом лінійного перетворення, використовуючи вираз [6, 7]:

$$x_i = \frac{x_i - x_{i\min}}{x_{i\max} - x_{i\min}}, \quad (6)$$

приводять до інтервалу $[0, 1]$.

Кожний проведений індивідуальний експеримент був проведений за однакових умов – один і той самий сценарій та ті самі параметри і характеристики сцени, іншими словами, з однією і тією самою моделлю реалістичної ситуації.

Обґрунтування вибору розподілу. Присутність випадкової складової у спостережуваних значеннях характеристик подій вимагає визначення їхнього розподілу, оскільки саме розподіл випадкових величин дає найповніший їхній статистичний опис. Встановити вид розподілу випадкових величин можна теоретичними викладками, як, наприклад, розподіли Больцмана, Гібсона, Максвелла та ін., або ж підібрати подібний серед відомих.

Класичний підхід до аналізу та моделювання експериментально отриманих даних, а відтак і часових послідовностей опирається на гіпотезу про нормальність розподілу. Проте отримані в дослідженні значення подій є випадковими і обмеженими як знизу, так і зверху, крім того, самі послідовності мають різний фіксований обсяг та значну асиметрію розподілу.

Реалістичнішим підходом до розв'язання поставленої задачі є розгляд елементів часової послідовності як множини дискретних однаково розподілених обмежених величин, причому обмеження у кожному окремому випадку є різними як за величиною, так і за положенням на числовій осі. Така ситуація істотно ускладнює ідентифікацію операторів. Стає очевидним, що у цій ситуації загальний показник індивідуальної оперативності має враховувати низку різних вибірових показників, принаймні визначених списком описової статистики табличного процесора MS Excel.

В результаті проведених експериментів для отриманих даних, які в сукупності утворюють вибірку обмежених варіант, визначено показники описової статистики: вибірове середнє,

дисперсію та середнє квадратичне відхилення, асиметрію і ексцес. Усі ці параметри є функціями обсягу вибірки та ФС (функціонального стану) людини-оператора, а отже, відповідними статистичними, точніше емпіричними, оцінками. Оскільки індивідуальні значення оперативності за фіксованих обсягів часових послідовностей лежать у межах сегмента $[x_{\min}, x_{\max}]$, значення асиметрії, що підтверджується відсутністю рівності між модою, медіаною та середнім, а також величини ексцесів, вказують на те, що застосування апарата, який використовується у випадку нормального розподілу, є неправомірним.

Доцільнішим у цьому плані є використання як моделі значень елементів бета-розподілу, який описує неперервні випадкові величини, що мають обмежений розмах варіації. Цей розподіл є доволі відомим, має добре вивчені властивості та функцію щільності, на підставі якої можна створити об'єктивну систему критеріїв ідентифікації операторського персоналу за індивідуальними, експериментально отриманими даними на кшталт системи [8].

Доволі висока гнучкість і загальність форми функції щільності цього розподілу визначається його параметрами. Крім того, цей розподіл є стійким у тому сенсі, що сума випадкових величин, підпорядкованих йому, має також бета-розподіл.

Правомірність переходу до неперервного розподілу для описання дискретних обмежених на сегменті $[x_{\min}, x_{\max}]$ даних, у цьому випадку може бути обґрунтована тим, що значення елементів послідовності належать множині дійсних додатних чисел. Функція щільності бета-розподілу може мати різноманітний вигляд залежно від значень її параметрів, а це уможливорює візуальну відмінність між операторами.

Зауважимо, що значення бета-функції дуже залежить від величини значень параметрів, а побудована функція щільності (5) може істотно відрізнитися від реальної для конкретних даних.

Для уточнення оцінок пропонується така процедура:

Крок 1. Подання даних варіаційним рядом. Традиційно оброблення експериментальних вибірок даних розпочинається з побудови варіаційного ряду, на підставі якого визначаються основні статистичні кількісні характеристики, обмежений список яких подано опцією “Описова статистика” в офісному пакеті MS Excel 2003. Крім цих даних, візуальний аналіз обвідної варіаційного ряду дає ще таку інформацію:

- присутність точки перегину вказує на одномодальність закону розподілу, а порівнювання другої похідної до нуля і розв’язання цього рівняння дає значення координати моди;
- якщо обвідна є прямою, то маємо рівномірний розподіл, аналогічно опуклість вказує на розподіл, пов’язаний з логарифмом, а вгнутість – на експоненціальні функції розподілу: загалом останні дві форми вказують на те, що функції щільності розподілів є нелінійними і не мають моди;
- обвідна варіаційного ряду є оберненою функцією до функції закону розподілу.

Крок 2. Лінійні перетворення приведення даних до одиничного квадрата. Оскільки для апроксимації даних вводиться як апріорний розподіл бета-розподіл, область визначення якого загалом обмежена сегментом $(0, 1]$, використовуючи лінійне перетворення отримані експериментально значення даних можуть бути приведені в межі цього сегмента за допомогою виразу (6).

Для ідентифікації персоналу необхідно забезпечити можливість порівняння індивідуальних варіаційних рядів та індивідуальних графіків функцій щільностей. Іншими словами, обсяги індивідуальних вибірок є різними, що вимагає проведення відповідного масштабування. Це можна реалізувати введенням обернених індексів у такий спосіб. Індекси, а точніше нумерація елементів часового ряду є чітко впорядкована у зростаючому порядку. Здійснивши операцію ділення $j = \frac{i}{n}$,

де n – обсяг вибірки, $i \in \overline{1, n}$ – порядковий номер елемента, можна усі варіаційні ряди подати графічно в межах **одиничного квадрата**. Оскільки і абсциси і ординати є обмежені областю значень від 0 до 1, то подання різних рядів, оскільки їхня форма після таких перетворень

залишається незмінною, дає підстави для візуального аналізу і подання певних висновків стосовно індивідуальних даних. Результати приведення індивідуальних варіаційних рядів та відповідних функцій щільностей в область одиничного квадрата показані на рис. 2 і 3 відповідно.

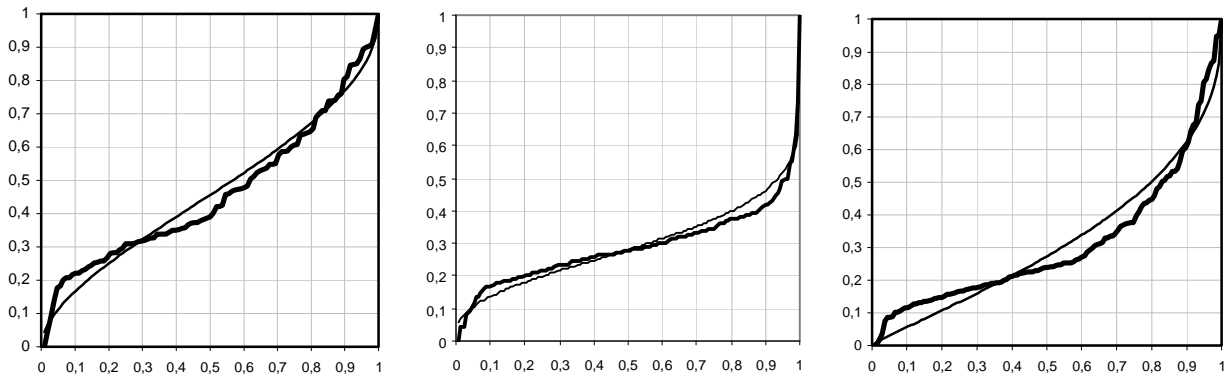


Рис. 2. Варіаційні ряди та їхні апроксимації

З іншого боку, отримані дані є такими і в такій кількості, в якій вони подані, а головне, вони відповідають тому відрізку часу і тим функціональним станам операторів, за яких вони були отримані. Тому закон розподілу цих даних, повністю визначаючи їх як випадкові величини, дає підстави у той чи інший спосіб побудувати певну систему оцінювання ефективності індивідуальної інтелектуальної діяльності.

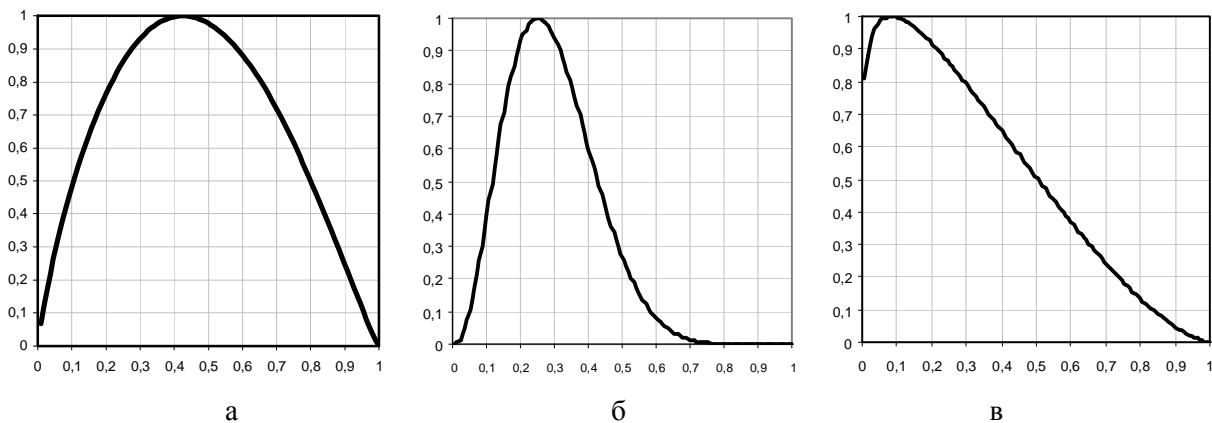


Рис. 3. Індивідуальні функції щільностей бета-розподілів

Крок 3. Описова статистика експериментальних даних. Знаходження характеристик вибірки, визначених показниками описової статистики, повністю забезпечується відповідною опцією табличного процесора MS Excel 2003. У цій роботі використані фактично такі характеристики: середнє арифметичне, дисперсія, ексцес і асиметрія. Перші дві дають змогу розрахувати початкові оцінки параметрів \hat{a} і \hat{b} бета-розподілу за формулами (5). У подальшому аналізі, використовуючи ці оцінки за допомогою функції **БЕТАОБЕР**, будується апроксимуюча обвідна для варіаційного ряду. Використовуючи далі опцію **ПОШУК РІШЕННЯ** і як критерій оптимізації **СУМАКВРІЗН**, уточняємо значення оцінок параметрів a і b .

Крок 4. Побудова графіка функції щільності бета-розподілу. Використовуючи уточнені значення параметрів a і b , будують графік функції щільності бета-розподілу, не враховуючи константу $\frac{1}{B(a, b)}$, яка не впливає на форму графіка функції щільності. Замість неї використовуємо

іншу константу, величина якої робить значення моди Mo для цієї функції таким, що дорівнює $Mo=1$. Побудовані в такий спосіб графіки функцій щільності можна зіставляти між собою та візуально аналізувати їхні особливості. На рис. 3 показані різні за формою графіки індивідуальних функцій щільностей бета-розподілу.

Крок 5. Оцінювання ефективності інтелектуальної діяльності операторів. Основним припущенням в сенсі оцінювання ефективності є те, що кожна індивідуальна функція щільності містить у собі, точніше презентує собою особливості індивідуальної інтелектуальної діяльності – спосіб пошуку і порівняння виявлених об'єктів уваги зі збереженими у пам'яті їхніми еталонами, характер поведінки у цьому середовищі, відношення до поставленої задачі, які, на жаль, не вдається розділити на окремі складові.

У цьому плані показник ефективності операторської діяльності побудований за таких характеристик:

1. **Мода Mo .** За значної асиметрії функцій розподілу роль моди має визначальну роль, оскільки саме в її околі зосереджені найчастіше значення елементів часових послідовностей. У різних довідниках наведені формули для обчислення значення моди, включаючи і MS Excel 2003, дають різні значення цього показника, а тому у цій роботі його визначають на підставі графіка індивідуальної емпіричної функції щільності бета-розподілу.

2. **Ширина функції щільності на половині висоти $d = x_2 - x_1$.** Цей показник і його зміст, відображені на рис. 4, належать до числових характеристик і наведені у [9] як півширина одномодального розподілу, тобто як піврізниця значень функції щільності розподілу у точках x_1 та x_2 , в яких $f(x_1) = f(x_2) = 0.5 \max f(x)$.

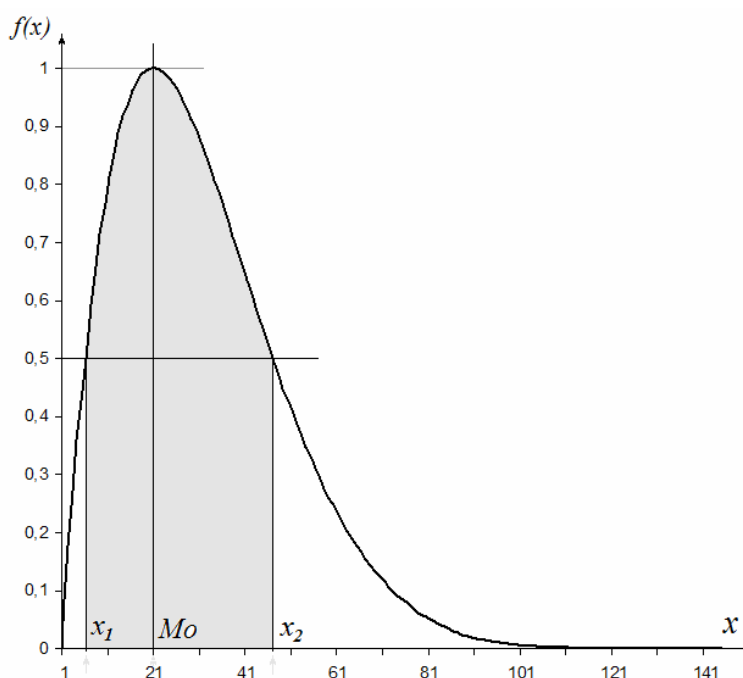


Рис. 4. Ширина функції щільності на половині її максимального значення

Чим більша ця різниця, тим більшою є дисперсія значень, а отже, і слабопрогнозованою, нестабільною є робота оператора.

3. **Асиметрія As .** Цей показник характеризує зміщення форми щільності розподілу вліво або вправо стосовно осі симетрії нормального розподілу. У цій роботі коефіцієнт асиметрії As

визначений за формулою MS Excel 2003, а саме: $As = \frac{n}{(n-1)(n-2)} \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{s} \right)^3$. Чим більша аси-

метрія розподілу тим більше малих значень часу опрацювання, а отже, тим вищою є оперативність опрацювання зображень. Проте, як бачимо з графіків на рис. 3, б і 3, в, асиметрія може бути різною.

4. **Ексцес Ex** . Показник Ex характеризує кривизну кривої щільності розподілу, гостроту або пласкість порівняно з нормальним розподілом. Для визначення цього показника також використано MS Excel 2003.

Показник ефективності операторської діяльності

Подання оброблених даних у цьому випадку функцій щільності до одиничного квадрата дає основу для їхньої об'єктивної ідентифікації, оскільки, з одного боку, вони усі є зображені в одному масштабі – в межах одиничного квадрата, а з іншого, – завдяки особливостям бета-розподілу та індивідуальних параметрів форма їхніх графіків має добре виражений візуально вигляд.

Зображені на рис. 3 різновиди індивідуальних функцій щільностей бета-розподілів відрізняються своєю формою, зокрема: положенням моди, асиметрією, шириною функції щільності на половині висоти та ексцесом. Використання окремо кожного з цих показників не дає об'єктивної картини для оцінювання ефективності діяльності. Наприклад, мода характеризує найчастіші значення цього показника і з цієї точки зору крива щільності на рис. 3, в є кращою, ніж попередні, проте ширина на рівні 0,5 для неї є вдвічі більшою, ніж для зображеної на рис. 3, б, стосовно асиметрії можна вважати, що щільності на рис. 3, а, б є симетричними, в той час, як щільність на рис. 3, в є дуже асиметричною. Значення показників ексцесу для рис. 3, а і б істотно відрізняються, а для щільностей на рис. 3, б і в можна вважати близькими, проте за положенням моди і величини ширини на половині висоти вони усі відрізняються.

Вихід з такої ситуації є у логічному поєднанні цих усіх показників в одному виразі. Зауважимо лише те, що значення моди і ширини на половині висоти мають розмірність абсциси, тобто часу опрацювання зображення, а асиметрія і ексцес є безрозмірними.

Враховуючи логічні міркування, показник ефективності загалом визначається безрозмірною величиною:

$$Q = \frac{Mo}{(x_2 - x_1)} As + Ex.$$

В експериментальних дослідженнях взяли участь дві групи операторів. У першій групі вісім операторів одноразово взяли участь в експериментах і їхні результати подані діаграмою на рис. 4, а. Операторський персонал другої групи налічував 10 операторів, кожен з яких взяв участь у двох експериментах з різницею в один тиждень. Світлі стовпчики діаграми на рис. 4, б належать до першого експерименту, темні – до другого. Оператори **Per**, **Oli**, **Sol** у другому підході показали гірші результати, натомість оператори **Hod**, **Pon**, **Bur**, **Cub**, **Syr**, **Lob** покращали свої результати. Оператор **Lot** і в першому, і в другому експерименті показав однакові результати.

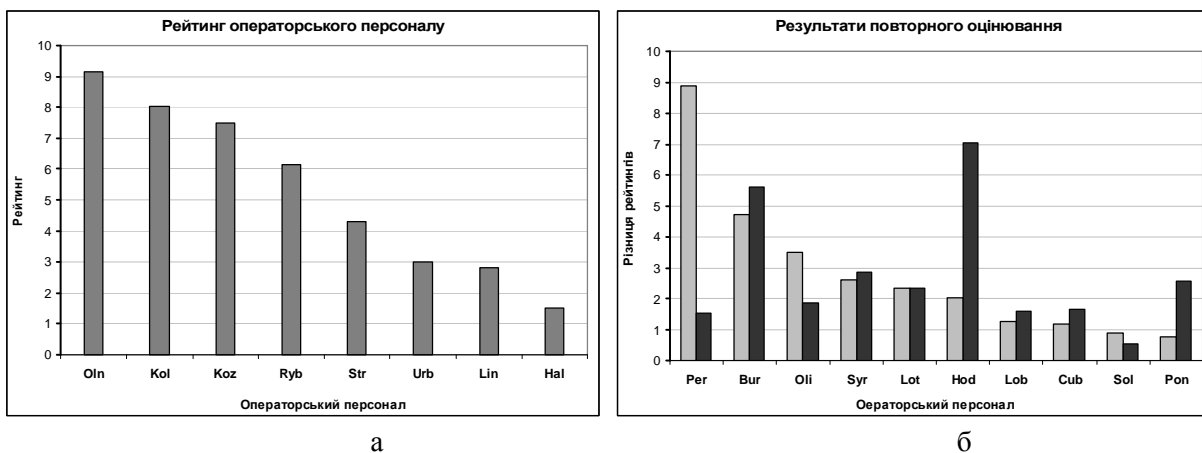


Рис. 4. Результати оцінювання ефективності інтелектуальної діяльності

Висновок

Наведені результати досліджень підтверджують доцільність оцінювання ефективності інтелектуальної діяльності операторського персоналу на підставі емпіричної функції щільності бета-розподілу і можуть бути використані у системах профвідбору, навчання та атестації операторського персоналу у різноманітних системах опрацювання візуальної інформації.

1. Психологічний словник. – К.: Вища шк., 1982. – 216 с. 2. Кобзарь А. И. Прикладная математическая статистика. Для инженеров и научных работников. – М.: Физматлит. – 2006. – 816 с. 3. Айвазян С. А., Мхитарян В. С. Прикладная статистика и основы эконометрики. – М.: Издательское объединение “ЮНИТИ”, 1998. – 1000 с. 4. Вадзинский Р. Н. Справочник по вероятностным распределениям. – СПб.: Наука, 2001. – 295 с. 5. Cooper John C. B. The first class honours grade: an application of the beta distribution // Applied Probability Trust 2007. – P. 73-74. <http://ms.appliedprobability.org/data/files/Articles%2039/39-2-8.pdf>. 6. Ульянов М. В., Петрушин В. Н., Кривенцов А. С. Доверительная трудоемкость – новая оценка качества алгоритмов // Информационные технологии и вычислительные системы. – 2009. – № 2. – С. 23 – 37. 7. Заболотский В. П., Оводенко А. А., Степанов А. Г. Математические модели в управлении: учеб. пособ. – СПб.: СПбГУФП, 2001. – 196 с. 8. Каминский Р. Н. Оценка оперативности действий персонала в системах управления // Проблемы управления и информатики. – 1997. – № 2. – С. 132 – 140. 9. Корн Г., Корн Т. Справочник по математике для научных работников и инженеров. – М.: Наука, 1968. – 720 с. 10. Камінський Р. М. Визначення параметрів розподілу на підставі кумуляти для показника оперативності персоналу автоматизованих систем розпізнавання зображень // Моделювання та інформаційні технології. – К.: ІПМЕ ім. Г. Є. Пухова, 2003. – С. 100 – 108.

УДК 004.942

В.В. Пасічник, Н.М. Іванушак

Національний університет “Львівська політехніка”,
Чернівецький національний університет імені Юрія Федьковича

ЙМОВІРНІСНЕ МОДЕЛЮВАННЯ ПРОЦЕСІВ КЛАСТЕРИЗАЦІЇ КОМП'ЮТЕРНИХ МЕРЕЖ

© Пасічник В. В., Іванушак Н. М., 2011

Здійснено наочне ймовірнісне моделювання процесу зародження та кластеризації локальної комп'ютерної мережі. Розроблено модель, яка дає змогу отримати зображення мережі для різних початкових умов, динамічно візуалізувати процес її структуризації.

Ключові слова: комп'ютерні мережі, дендритний ріст, модель ДОА.

In this work there is given the probabilistic visualization of modeling of nucleation and clustering of local computer network. The developed model allows the network to obtain images of different initial conditions and dynamically visualize the process of structuring.

Key words: computer networks, dendritic growth, DLA model.

Вступ

У географічному просторі, де спостерігається тенденція до утворення локальної комп'ютерної мережі, існує певна кількість потенційних користувачів послуг. Вони розподілені нерівномірно, тому є сенс говорити про різну концентрацію споживачів. У тих місцях, де концентрація є найбільшою, виникає ймовірність утворення серверів мережі, до яких починається