

УДК 621.391

Варфоломєєва О. Г., к.т.н.; (Державний університет телекомунікацій, м. Київ. o\_g\_var@i.ua)

Колченко Т. В., аспірантка (Український НДІ зв'язку, м. Київ. +380 (44) 248 86 67. post1@undiz.org.ua)

## ЗАСТОСУВАННЯ МЕХАНІЗМІВ КОРЕЛЯЦІЙНОГО АНАЛІЗУ ІНФОРМАЦІЇ В СИСТЕМАХ УПРАВЛІННЯ

**Варфоломєєва О. Г., Колченко Т. В. Застосування механізмів кореляційного аналізу інформації в системах управління.** Розглянуто проблему необхідності проведення досліджень з метою окреслення напрямків і шляхів подальшого розвитку систем управління телекомунікаційними мережами, визначення певних механізмів, що дозволить оцінювати коефіцієнти, які характеризують кореляцію між випадковими величинами. Параметри, що характеризують функціонування системи управління, прямо або побічно впливають на критерії її оптимальності, можуть варіюватися з метою підвищення показників ефективності роботи системи управління. Застосування цього механізму забезпечить життєздатність системи управління телекомунікаційними мережами, а також самої телекомунікаційної мережі, та її цілісність при подальшому розвитку мереж в умовах ринкової багатоператорської діяльності.

**Ключові слова:** телекомунікаційна мережа, система управління, кореляційний аналіз, критерії оптимальності

**Варфоломєєва О. Г., Колченко Т. В. Применение механизмов корреляционного анализа информации в системах управления.** Рассмотрена проблема необходимости проведения исследований с целью очерчивания направлений и путей последующего развития систем управления телекоммуникационными сетями, определения соответствующих механизмов, позволяющих оценивать коэффициенты, которые характеризуют корреляцию между случайными величинами. Параметры, характеризующие функционирование системы управления, прямо или косвенно влияют на критерии ее оптимальности, могут варьироваться с целью повышения показателей эффективности работы системы управления. Применение этого механизма обеспечит жизнеспособность системы управления телекоммуникационными сетями, а также самой телекоммуникационной сети, и ее целостность при последующем развитии сетей в условиях рыночной многооператорской деятельности.

**Ключевые слова:** телекоммуникационная сеть, система управления, корреляционный анализ, критерии оптимальности

**Varfolomeyeva O. H., Kolchenko T. V. Application of mechanisms of correlation analysis of information in the management systems.** The problem of necessity of conducting of researches is considered with the purpose of outlining of directions and ways of subsequent development of the management systems of telecommunications networks, determination of certain mechanisms, which will allow estimating coefficients which characterize correlation between casual sizes. Parameters which characterize functioning of the management system straight or side influence on the criteria of its optimum can be varied with the purpose of increase of indexes of efficiency of work of the management system. Application of this mechanism will provide viability of the management systems of telecommunications networks, and also telecommunication network, and its integrity at subsequent development of networks in the conditions of market multioperator activity.

**Keywords:** telecommunication network, management system, correlation analysis, optimum criteria

**Загальні положення.** Критеріями оптимальної роботи системи управління телекомунікаційними мережами є її продуктивність і надійність, які визначаються конкретними показниками оцінки, наприклад, часом реакції, коефіцієнтом помилок, вартістю системи тощо. Ці параметри, що прямо або побічно впливають на критерії оптимальності, можуть варіюватися з метою підвищення показників ефективності роботи системи управління.

**Застосування кореляції.** Кореляційний аналіз – сукупність методів математичної статистики, які дозволяють оцінювати коефіцієнти, які характеризують кореляцію між випадковими величинами, і перевіряти гіпотези про їх значення на підставі розрахунку їх вибірових аналогів [1, 2].

Припустимо, що є  $N$  об'єктів, які характеризуються двома групами показників:

1) факторні ознаки (показники  $x_1, \dots, x_n$ ) – показники, що характеризують стан об'єкту з погляду його основних характеристик;

2) результативні ознаки (показники  $y_1, \dots, y_m$ ) – результуючі показники, що характеризують поведінку об'єктів з погляду проблеми, яка досліджується.

Отже, кожен об'єкт можна уявити у вигляді точки в  $n$ -мірному просторі за першою групою показників і в  $t$ -мірному – за другою групою показників. При цьому, як правило, і в тому, і в іншому середовищі, одержані точки не утворюють однорідної множини, а мають деякі сфери згущування. Множина об'єктів, наданих точками, які належать одній сфері згущування, є одноріднішим в порівнянні зі всією сукупністю досліджуваних об'єктів. Тобто, вся множина об'єктів, що вивчаються, розбивається на класи двома способами – відповідно до двох груп показників. Рішення задачі визначення взаємозв'язків між двома одержаними класифікаціями досліджуваної множини об'єктів допоможе побудувати оптимальну систему управління і розрахувати для неї необхідну кількість інформації. Інше кажучи, потрібно встановити діапазони спільної зміни показників, що визначають поведінку об'єктів, які досліджуються.

При розробці оптимальної системи управління необхідно задаватися пороговими значеннями критерію оптимальності. Якщо продуктивність системи оцінювати логічними значеннями "Працює"/"Не працює", то тоді оптимізація зводиться до діагностики несправностей і приведення мережі в будь-який працездатний стан. Іншим випадком є тонка настройка мережі, за якої параметри працюючої системи можуть варіюватися за продуктивністю навіть на декілька відсотків.

Побудова раціонального варіанту системи пов'язана з вибором великої і досить різномірної множини параметрів – типів і моделей обладнання, що використовується, модифікацій цього обладнання, кількості об'єктів управління, типів операційних систем, стеків протоколів, їх параметрів тощо [3].

Під час обробки інформації управління виникають ситуації, коли дані типу "об'єкт-ознака" містять достатньо велику кількість ознак, а кількість об'єктів може ще у декілька разів перевищувати кількість ознак.

Класифікація, створення нової структури з меншою розмірністю ознакового простору, зберігання, передача по каналах зв'язку, обробка і наочне уявлення та інтерпретація таких даних викликає труднощі. Виникає проблема скорочення розмірності ознакового простору. Таке скорочення можливе, оскільки в більшості випадків ознаки досить взаємозв'язані (корельовані) і, таким чином, дані є надмірними з погляду інформації і ця надмірність повністю визначається кореляційною матрицею вихідних змінних.

Для зменшення надмірності дані потрібно стиснути. Стиснення зводиться до перетворення вихідного простору  $X$  в інший простір  $Y$ , в якому можна вибрати підмножину, як правило, не спостережуваних (латентних) змінних меншої розмірності  $L < P$ , що не викликає значної втрати інформації.

Вибір виду перетворення  $Y = f(X)$  і кількості латентних змінних, що пояснюють спостережувані змінні, залежить від конкретної специфіки задачі, що вирішується, повинен спиратися на критерій, який забезпечує збереження інформації про  $X$  в стислому образі  $Y$ . Для здійснення такого переходу до нових змінних можна використовувати статистичні властивості матриці  $X$ . Якщо дані мають багатовимірний нормальний розподіл, то ці властивості визначаються кореляційною матрицею "ознака-ознака".

Кореляційний аналіз вирішує два основних завдання:

- визначення форми зв'язку, тобто встановлення математичної форми, якою виражається даний зв'язок;
- вимірювання тісноти, тобто ступеня зв'язку між ознаками з метою встановити ступінь впливу даного чинника на результат.

Застосування методів кореляційного аналізу дає можливість уявляти зв'язок між ознаками аналітично – у вигляді рівняння – і додавати йому кількісний вираз.

Розглянемо застосування прийомів кореляційного аналізу на конкретному прикладі. Із зростанням факторної ознаки тут, як правило, росте і результативна, тому залежність між ними може бути виражена рівнянням, яке називається *лінійним рівнянням регресії*. Допустимо, що між пропускною спроможністю каналу зв'язку і величиною затримки серіалізації в пакетній мережі існує прямолінійний зв'язок, який виражається рівнянням

прямої  $Y=a+b*x$ . Параметр  $b$  називається *коефіцієнтом регресії* і показує наскільки в середньому відхиляється величина результативної ознаки  $y$  при відхиленні величини факторної ознаки на одну одиницю.

Необхідно знайти параметри  $a$  і  $b$ , що дозволить визначити теоретичні значення  $Y$  для різних значень  $X$ . Причому,  $a$  і  $b$  повинні бути такими, щоб було досягнуто максимальне наближення до первинних (емпіричних) значень отриманих теоретичних значень  $Y$ . Ця задача розв'язується за допомогою способу найменших квадратів, основна умова якого зводиться до визначення параметрів  $a$  і  $b$ , так, щоб  $\sum (y_i - Y)^2 = \min$ .

Відомо, що умова мінімуму забезпечується, якщо параметри  $a$  і  $b$ , визначаються за допомогою системи двох нормальних рівнянь, що відповідають вимозі методу якнайменших квадратів:

$$\sum y = na + b \sum x.$$

Перше рівняння є сумою всіх первинних рівнянь. Друге є добутком обох частин рівняння прямої на один і той же множник.

Математично доведено, що умова  $\sum xy = a \sum x^2$  дотримується, якщо такого множника прийняти як значення факторної ознаки, тобто якщо рівняння прямої помножити на  $x$ .

Якщо співвідношення між величинами  $X$  і  $Y$  має вид степеневі функції  $Y=ax^b$ , то параметр  $b$  рівняння називається показником еластичності і вказує наскільки відсотків зміниться  $y$  при зростанні  $x$  на 1 % (при  $x = 1$ ,  $a = Y$ ).

Для визначення параметрів степеневі функції приводимо її до лінійного вигляду шляхом логарифмування:  $lgy = lga + b * lgx$ , а потім будемо систему нормальних рівнянь:

$$\begin{cases} \sum lgy = n lga + b \sum lgx \\ \sum lgy lgx = lga \sum lgx + b \sum lgx^2 \end{cases}$$

Розв'язавши систему двох нормальних рівнянь, можна знайти логарифми параметрів логарифмічної функції  $a$  і  $b$ , а потім і самі параметри  $a$  і  $b$ .

Іноді зворотний зв'язок між факторною і результативною ознаками може бути представлений рівнянням гіперболи:

$$y = a + \frac{b}{x}$$

Тоді задачею є знаходження параметрів  $a$  і  $b$  за допомогою системи двох нормальних рівнянь:

$$\begin{cases} \sum y = na + b \sum \frac{1}{x} \\ \sum y \frac{1}{x} = a \sum \frac{1}{x} + b \sum \frac{1}{x^2} \end{cases}$$

Для вимірювання тісноти прямолінійного зв'язку між двома ознаками використовується парний коефіцієнт кореляції  $r$ .

Визначення тісноти зв'язку, за суттю, зводиться до вивчення цієї зв'язаності, тобто визначення того, якою мірою відхилення від середнього рівня однієї ознаки зв'язане з відхиленням іншої. За наявності повного прямого зв'язку всі значення  $(x - \bar{X})$  і  $(y - \bar{Y})$  повинні мати однакові знаки, при повному зворотному – різні, при частковому зв'язку знаки частіше всього будуть співпадати, а за відсутності зв'язку – співпадатиме приблизно в рівній кількості випадків.

Для оцінки істотності коефіцієнта кореляції користуються спеціально розробленою таблицею критичних значень  $r$ . Коефіцієнт кореляції  $r$  застосовується тільки тоді, коли між явищами існує прямолінійний зв'язок. Якщо зв'язок криволінійний, то користуються індексом кореляції, який розраховується за формулою:

$$R = \sqrt{1 - \frac{\sum (y - Y)^2}{\sum (y - \bar{y})^2}},$$

де  $y$  – первинні значення;  $\bar{y}$  – середнє значення;  $Y$  – теоретичні значення змінної величини.

Показник випадкової дисперсії визначається за формулою:

$$\sigma_Y^2 = \frac{\sum (y - Y)^2}{n}.$$

Загальна дисперсія

$$\sigma_o^2 = \frac{\sum (y - \bar{y})^2}{n}$$

характеризує розмір відхилень емпіричних значень результативної ознаки  $y$  від  $\bar{y}$ , тобто загальну варіацію.

Відношення випадкової дисперсії до загальної характеризує частку випадкової варіації в загальній варіації, а  $\sigma_o^2 = \frac{\sum (y - \bar{y})^2}{n}$  – це частка факторної варіації в загальній, тому що за правилом складання дисперсій загальна дисперсія дорівнює сумі факторної і випадкової дисперсій:

$$\sigma^2 = \sigma_Y^2 + \sigma_o^2.$$

Підставимо до формули індексу кореляції відповідні позначення випадкової, загальної і факторної дисперсій та одержимо:

$$R = \sqrt{1 - \frac{\sum (y - Y)^2}{\sum (y - \bar{y})^2}} = \sqrt{1 - \frac{\sigma_o^2}{\sigma^2}} = \sqrt{\frac{\sigma_Y^2}{\sigma^2}}$$

Отже, *індекс кореляції* характеризує частку факторної варіації в загальній:  $R = \sqrt{\frac{\sigma_Y^2}{\sigma^2}}$ ,

проте з тією лише різницею, що замість групових середніх беруться теоретичні значення  $Y$ . Індекс кореляції за своїм абсолютним значенням коливається в межах від 0 до 1.

При функціональній залежності випадкова варіація  $\sum (y - \bar{Y})^2 = 0$ , індекс кореляції дорівнює 1. За відсутності зв'язку  $R = 0$ , тому що  $Y=y$ . Коефіцієнт кореляції є мірою тісноти зв'язку тільки для лінійної форми зв'язку, а індекс кореляції – і для лінійної, і для криволінійної. При прямолінійному зв'язку коефіцієнт кореляції за своєю абсолютною величиною дорівнює індексу кореляції:

$$|r|=R.$$

Якщо індекс кореляції звести в квадрат, то одержимо коефіцієнт детермінації:

$$R^2 = \sigma_Y^2 / \sigma^2.$$

Коефіцієнт детермінації є найконкретнішим показником, оскільки він відповідає на питання про те, яка частка в загальному результаті залежить від чинника, встановленого в основу угруповання. При прямолінійному парному зв'язку факторну дисперсію можна визначити без обчислення теоретичних значень  $Y$  за формулою:

$$\sigma^2 = \frac{1}{n} (a \sum y + b \sum xy) - y^{-2}.$$

Отже, теоретичні дослідження сучасного стану розвитку телекомунікаційних мереж та системи управління ними, дозволяє визначити низку параметрів і характеристик, визначення вимог до яких сприяє підвищенню надійності систем управління телекомунікаційними мережами.

**Висновки.**

1. Для успішної діяльності різних установ, підприємств та організацій (державного управління, промислових, фінансових, транспортних тощо) необхідні сучасні телекомунікаційні мережі та системи управління, засновані на спільному використанні технічних засобів телекомунікацій та обчислювальної техніки. Відмови в роботі таких систем призводять до величезних збитків. Для уникнення таких збитків необхідно підвищити надійність телекомунікаційних мереж. Одним із напрямків забезпечення підвищення надійності є ефективне управління об'єктами (ресурсами) телекомунікаційних мереж:

2. Зміна стану деякого об'єкту або параметра означає виникнення певної події в телекомунікаційній мережі, тому взаємний вплив елементів мережі може розглядатися як кореляція подій і може бути представлено як процес перегляду великої кількості подій, який призводить до відкидання деяких з них, і аналіз відносин типу "причина-ефект" між подіями, що залишилися, з виділенням тривожних сигналів з набору взаємозв'язаних подій або ідентифікацію події, що є причиною виникнення зміни стану елемента мережі.

3. При побудові оптимального варіанту системи управління необхідно провести оцінку кореляційних характеристик об'єктів і параметрів управління та визначити обсяги управляючої інформації в системі.

**Література.**

1. Варфоломеєва О. Г. Методика розрахунку показників ефективності системи управління мережами телекомунікацій із застосуванням методу експертних оцінок / О. Г. Варфоломеєва // Зв'язок. – 2005. – №7(59). – С. 22-25.

2. Вентцель Е. С. Прикладные задачи теории вероятностей / Е. С. Вентцель, Л. А. Овчаров. – Москва : Радио и связь, 1983. – 415 с.

3. Стеклов В. К. Проектирование телекоммуникационных сетей // В. К. Стеклов, Л. Н. Беркман. – К.: Техніка, 2002. – 791 с.

4. Шварц М. Сети связи: протоколы, моделирование и анализ, ч. 2 / М. Шварц. – Москва : Наука, 1992. – 272 с.

5. Беркман Л. Н. Информационно-энтропийные подходы к расчету параметров системы управления интеллектуальной сети / Л. Н. Беркман // Зв'язок. – 2001. – №4. – С. 14-16.

6. Слюсарь В. А. К вопросу об оценке некоторых критериев функционирования систем управления телекоммуникационными сетями / В. А. Слюсарь, О. Г. Варфоломеєва, Г. Ф. Колченко // Наукові записки Українського науково-дослідного інституту зв'язку. – 2007. – № 1. – С. 20-24.

7. Методи підвищення показників якості системи управління телекомунікаційними мережами : монографія / [В. В. Хиленко, Л. Н. Беркман, Г. Ф. Колченко, О. Г. Варфоломеєва]. – К.: Норіта-плюс, 2007. – 236 с.

8. Стеклов В. К. Визначення параметрів оптимальної системи управління за багатокритеріальним методом : навчальний посібник для дипломного проектування // В. К. Стеклов, Л. Н. Беркман, Г. Ф. Колченко. К.: ДП «УНЦІЗ», 2001. – 105 с.

9. Гребенніков В. О. Оптимізація загальнодоступності інфокомунікаційних послуг в Україні / В. О. Гребенніков, Г. Ф. Колченко // Наукові записки Українського науково-дослідного інституту зв'язку. – 2013. – № 2(26). – С. 5-13

10. Система управління якістю. Основні положення та словник термінів // ДСТУ ISO 9000-2007 (ISO 9000:2005, IDT).

11. Principles for a Telecommunications Management Network (Принципи управління телекомунікаційними мережами) // ITU-T Recommendation M.3010. – 2010.