

отримання бази знань (множини значущих правил) використовується в реалізованій системі ефективний метод дерев прийняття рішень. Отримані на основі нього правила мають структуру, що подібна до синтаксичних конструкцій природної мови, тому ці правила легко осмислити.

1. Щавльов Л.В. Способи аналітичного опрацювання даних для підтримання прийняття рішень. СУБД, 4–5/98. 2. http://www.olap.ru/basic/olap_intro6.asp, 10.09.2003 р. 3. Кравець Р.Б. Багатовимірні моделі даних у системах аналітичної обробки інформації. – Вісник Національного університету “Львівська політехніка”. – 1998. – №330. – С. 147–153. 4. Codd E.F., Codd S.B., Salley C.T. *Providing OLAP (on-line analytical processing) to user-analysts: an IT mandate*. – E.F. Codd & Associates, 1993. 5. Han J., Kamber M. *Data mining: methods and technique*. – Morgan Kaufman, 2000. 6. Нікольський Ю., Щербина Ю., Якимечко Р. Древа прийняття рішень та їхнє застосування для прогнозування діагнозу у медицині // Вісник Львівського університету імені Івана Франка. – 2003. – Вип. 6. – С. 195–196.

УДК 004.896

Р.Б. Кравець, Н.Б. Шаховська, А.Р. Продан
Національний університет “Львівська політехніка”,
кафедра інформаційних систем та мереж

СИСТЕМА ВИБОРУ ОПТИМАЛЬНИХ ДАНИХ ДО РЕГУЛЮВАННЯ ПЕРЕТОКУ ЕЛЕКТРОЕНЕРГІЇ НА ЕКСПОРТ

© Кравець Р.Б., Шаховська Н.Б., Продан А.Р., 2004

Розглянуто проблеми вибору оптимального телевіміру ТЕС на основі методу МГУА.

Described main problems of optimal teletype choice based on MGCA.

Вступ

Людство невідворотно вступає в інформаційну епоху, де вага інформації постійно зростає. За підрахунками науковців, з початку нашої ери для подвоєння знань треба було 1750 років, друге подвоєння відбулося в 1900 році, а третє – до 1950 року, тобто вже за 50 років, при зростанні обсягу інформації за ці півстоліття у 8–10 разів. Причому ця тенденція усе підсилюється, тому що обсяг знань у світі до кінця ХХ століття зріс удвічі, а обсяг інформації збільшиться більше ніж у 30 разів. Це явище одержало назву “інформаційний вибух” та вказується серед ознак початку століття інформації.

Відсутність, неповнота або спотворення інформації призводять до викривлення дійсності, втрати управління, призупинення суспільного розвитку. Збирання, передавання, перетворення, обробка інформації та інші інформаційні процедури передують прийняттю будь-яких рішень, тому на сучасному етапі розвитку цивілізації велику увагу приділяють розробці прогресивних інформаційних технологій, підвищенню інформаційної культури суспільства.

Сьогодні є велика потреба саме в інформації для прийняття рішень при великій кількості даних. Коли кількість змінних є більшою ніж кількість випадків, найбільш відомі алгоритми обробки даних зіштовхуються з деякими проблемами. Навіть якщо дані чіткі, велика кількість змінних створює проблему розмірності. Тому прийняття рішення, базоване на аналізі даних – діалоговий та ітераційний процес різних частин робочих задач і рішень – називається отриманням знань з даних.

Є багато різних інструментів видобування даних, описаних у [3, 4]. Важливо обмежити залучення користувача до процесу отримання даних для складних систем при включенні відомого апріорного знання. Це зробить процес більш автоматизованим і об'єктивним. Більшість користувачів насамперед цікавить можливість отримання корисних та зразкових кінцевих результатів без застосування складних математичних, кібернетичних і статистичних методів чи без витрат часу на

складний діалог – керування інструментами моделювання. М'яке обчислення, тобто нечітке моделювання, нейронні мережі (НМ), генетичні алгоритми та ін. є методами автоматичного створення математичних моделей на емпіричних даних.

Використання НМ стикається з проблемою опису отриманих результатів (чому саме такі?) Це означає, що моделі, отримані НМ, усе ще сховані і розподілені по мережі. Немає жодного систематичного підходу для проектування і розвитку НМ [3, 4]. Це – емпіричний процес. Якщо у зразку даних значний шум, то в отриманих моделях систематично спостерігається тенденція до помилок.

У статті розглянуто процедуру та основні труднощі підходів до вибору оптимальної моделі даних, тобто моделі, найбільш подібної до еталонної. Пропонується вибір оптимальних телевимірів для роботи регулятора перетоку потужності експортними лініями електропередач.

Постановка проблеми у загальному вигляді

Електроенергетика України – потужний, складний та багатогранний технологічний комплекс, метою якого є виробництво, передача і розподіл електроенергії між окремими її споживачами. Основою електроенергетики України є Об'єднана енергетична система (ОЕС), яка здійснює централізоване постачання електричної енергії, взаємодіє з енергосистемами суміжних держав, забезпечує експорт та імпорт електроенергії. Важливою технологічною ланкою ОЕС України є магістральні електричні мережі, якими передається електрична енергія від її виробників: атомних, теплових та гідровисхідних електростанцій до енергопостачальних компаній.

Функції оперативного-технологічного управління ОЕС України та експлуатації магістральних і міждержавних електромереж здійснює Державне підприємство Національна енергетична компанія “Укренерго” (НЕК “Укренерго”) та її 8 структурних одиниць в регіонах, в тому числі Західна електроенергетична система (ЗЕС). Частина Західної ЕС під назвою “Острів Бурштинської ТЕС” працює паралельно з об'єднаною енергосистемою Західної Європи (УСТЕ). Технічно це стало можливим після встановлення в диспетчерському центрі ЗЕС і пуску в експлуатацію центрального регулятора (ЦР) перетоку потужності і частоти лініями електропередач, які з'єднують “острів” з УСТЕ, а саме:

ПЛ 750 Західноукраїнська–Альбертірша (Угорщина).

ПЛ 400 Мукачево–Кішварда (Угорщина).

ПЛ 220 Мукачево–Шайосегед (Угорщина).

ПЛ 220 Мукачево–Тісалець (Угорщина).

ПЛ 400 Мукачево–Капушани (Словаччина).

За вимогами УСТЕ, сформульованими в “Каталозі заходів”, які передували приєднанню “острова” до УСТЕ, допускається відхилення сальдо перетоку потужності від планового не більше 50 Мвт. протягом 3-х сек. і відхилення сальдо обміну енергією протягом години не більше 20 Мвт. Досягнути в динаміці таких вимог важко, враховуючи, що епізодично аварійно від'єднуються генеруючі блоки (що призводить до значних збурень в енергосистемі), відмовляють давачі вимірів та каналів передачі даних тощо. Величина відхилення залежить і від точності вимірів, які надходять на регулятор. Однак регулятор працює за вимірами потужності, а оцінка (як було наведено вище) робиться також за відхиленням переданої за кожну годину енергії від планової (менше 20 Мвт. за годину). Звідси випливає вимога до точності вимірів. Кожен вимір має свою реальну похибку, одні виміри завищують, інші занижують покази. Відповідно регулятор то недодає, то передає електроенергію на експорт.

Цілі статті

Ставиться задача мінімізації відхилення експорту енергії від запланованого. Досягнути цього можна підбираючи виміри так, щоб відхилення протягом кожної години були мінімальними. Якщо проінтегрувати кожен вимір потужності протягом години і порівняти його з показами лічильника про передану за цю ж годину енергію кожною з ліній електропередач, то можна визначити відхилення вимірної системи потужності (за якою працює регулятор) від даних лічильників енергії

(які є розрахунковими). Якщо вибрати для роботи регулятора виміри, які дають мінімальний баланс похибок, то досягнемо найкращих результатів регулювання (за точністю). Маніпулюючи пріоритетами вимірів на регуляторі (регулятор завжди опрацьовує вимір з вищим пріоритетом), досягнемо кінцевої мети.

Аналіз останніх досліджень

Сьогодні не існує системи, яка була б здатна оцінювати покази телевимірів та перевіряти їх на точність в реальному часі. Тому робота, що описана у цій пояснювальній записці, є актуальною.

Система, що описується, є інтелектуальною, оскільки дозволяє організувати підтримку прийняття рішення з визначення оптимальних для роботи регулятора перетоку електроенергії вимірів. Очевидним є і економічний ефект, оскільки передбачається зменшення затрат передавання електроенергії партнерам.

Основний матеріал

Зв'язок висвітленої проблеми із важливими практичними завданнями

На вхід регулятора надходять виміри перетоку потужності кожною лінією зв'язку – як із закордонних підстанцій, так і з українських. Основними вважаються виміри із закордонних підстанцій, саме за ними регулюється перетікання потужності. На випадок відмови основних вимірів організовано резервні виміри через диспетчерські центри Варшави, Будапешта, Жиліна і Львова. Крім вимірів потужності на вхід регулятора надходять з Бурштинської ТЕС вимір частоти. Всі виміри надходять з циклічністю 1–2 сек.

Регулятор виробляє регулюючий сигнал залежно від величини відхилення фактичного перетоку потужності від планового та величини відхилення фактичної частоти від заданої за формулою

$$ACE = (P_{\text{план}} - P_{\text{факт}}) + K(F_{\text{задане}} - F_{\text{фактичне}}),$$

де ACE – відхилення регулятора; $P_{\text{план}}$ – планове значення перетоку потужності по лініях зв'язку; $P_{\text{факт}}$ – фактичне (сальдове) значення перетоку по лініях зв'язку; $F_{\text{задане}}$ – задане значення частоти; $F_{\text{фактичне}}$ – фактичне значення частоти; K – частотний коефіцієнт. Його значення залежить від сумарної генеруючої потужності енергосистеми, точніше “острова”.

Вихідний (регулюючий) сигнал регулятора передається на Бурштинську станцію, де він розподіляється між працюючими енергоблоками, збільшуючи або зменшуючи генеровану станцією потужність.

Основні функції Центрального регулятора “острова Бурштинської ТЕС”

Основною задачею Центрального регулятора є утримання заданого планом-графіком перетоку потужності за експортними лініями електропередач і частоти системи. Ця задача реалізується за допомогою основних функцій регулятора, а саме:

- прийому телеграм телемеханіки, що містять інформацію про вимірювання перетікання потужності в граничних пунктах;
- вимірювання частоти системи (з точністю до 1 мГц);
- перевірки і валідації (тобто вибору дійсного значення параметра при інформації з декількох вимірювальних пунктів) вимірювальних даних;
- створення баз даних – картини процесу регулювання;
- створення архівних баз даних;
- розрахунку похибки регулювання відповідно до рівняння:

$$E = \Delta P \pm K(\Delta f)$$

де E – відхилення регулювальної потужності; ΔP – відхилення планованого вихідного сальдо; K – коефіцієнт потужності-частоти; Δf – відхилення частоти від номінального значення (від базової частоти f_0);

- генерації регулюючого сигналу Y ;
- пересилання до системи регулювання на об'єкті (електростанції) заданої базової частоти f_0 , а також регулюючого сигналу Y ;
- блокування пересилання сигналів у разі виявлення неправильної роботи системи;
- повної візуалізації регулювального процесу на X -терміналах;
- самоконтролю правильності роботи системи загалом;
- синхронізації часу всіх комп'ютерів за допомогою супутникового годинника UTC;
- автоматичного вибору шляху для сигналів уведення-виведення за допомогою процесора, що перемикається;
- обміну інформацією в режимі реального часу з розрахунково-регулювальним Центром у Варшаві;
- можливості введення оператором вручну:
- заданої потужності обміну (P_0);
- базової частоти (f_0);
- статичних і динамічних параметрів регулятора;
- можливості інтервенційної зміни регулюючого сигналу Y незалежно від зміни, розрахованої алгоритмом регулятора.

Описання реалізації завдання

Для вирішення задачі можуть використовуватися такі методи: регресійний аналіз, дисперсійний аналіз, МГУА з критерієм вибору без розділення спостережень та МГУА з критерієм вибору з розділенням спостережень.

Оскільки дисперсійний аналіз найчастіше використовується для оцінки середніх значень вибірки, то нам не доцільно використовувати його для аналізу замірів телевимірів.

МГУА з критерієм відбору без розділення на вибірки також оперує середнім значенням вибірки і тому не може використовуватись у даному випадку.

У ході регресійного аналізу ми отримуємо функцію залежності між показами телевиміра та часом заміру. Порівнюючи функції зовнішніх телевимірів та лічильника USTE, можна обрати оптимальний телевимір. Але тут виникає проблема вибору характеру регресії (квадратична, поліміальна, логорифмічна тощо). Для усунення цієї проблеми потрібно вручну досліджувати характер регресії, а програмно розраховувати її коефіцієнти.

МГУА з критерієм відбору з розділенням на вибірки дозволяє не тільки вибрати оптимальну модель (телевимір), але й дослідити її на "довговічність". Одним з недоліків цього методу є оперування середніми значеннями, через які при великих вибірках (як і є у цій задачі) можна втратити точність розрахунків, тобто обрана модель може бути неоптимальною протягом певного періоду часу.

Як і регресійна модель, МГУА з критерієм відбору з розділенням на вибірки базується на функції залежностей між змінними. Отже, для аналізу вибірок показів телевимірів та визначення оптимального телевиміру обрано регресійний аналіз з попереднім дослідженням характеру регресії та МГУА з критерієм відбору з розділенням на вибірки. Користувач буде приймати рішення щодо оптимальності телевимірів на основі результатів застосування цих двох методів.

Концептуальна модель

Потрібно розробити підсистему вибору телевимірів для роботи регулятора електроенергії, яка повинна визначати пріоритети телевимірів шляхом аналізу їхніх замірів та їх порівняння з розрахунковим лічильником USTE (еталоном).

На контекстній діаграмі (рис. 1) зображено основні сутності проектованої системи: бази даних значень телевимірів (*База даних значень зовнішніх телевимірів* та *База даних показів лічильника USTE*). Сутність *Регулятор* призначена для визначення перепаду енергії у певний часовий інтервал та регулювання енергетичного потоку згідно з коефіцієнтами телевимірів, отриманих у процесі моделювання.

Фактично прогнозується вибір найоптимальнішого телевиміру шляхом порівняння моделей.

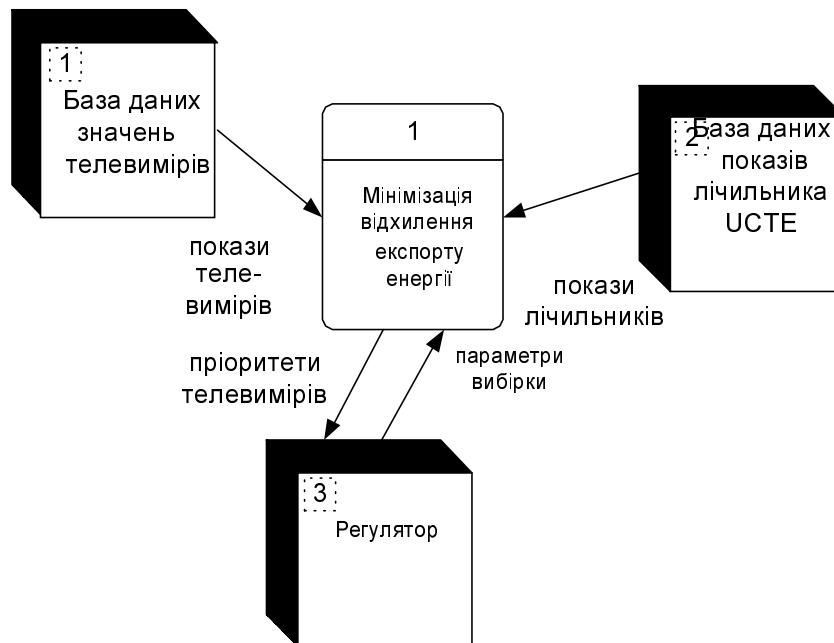


Рис. 1. Контекстна діаграма

Деталізація основного процесу зображена на рис. 2.

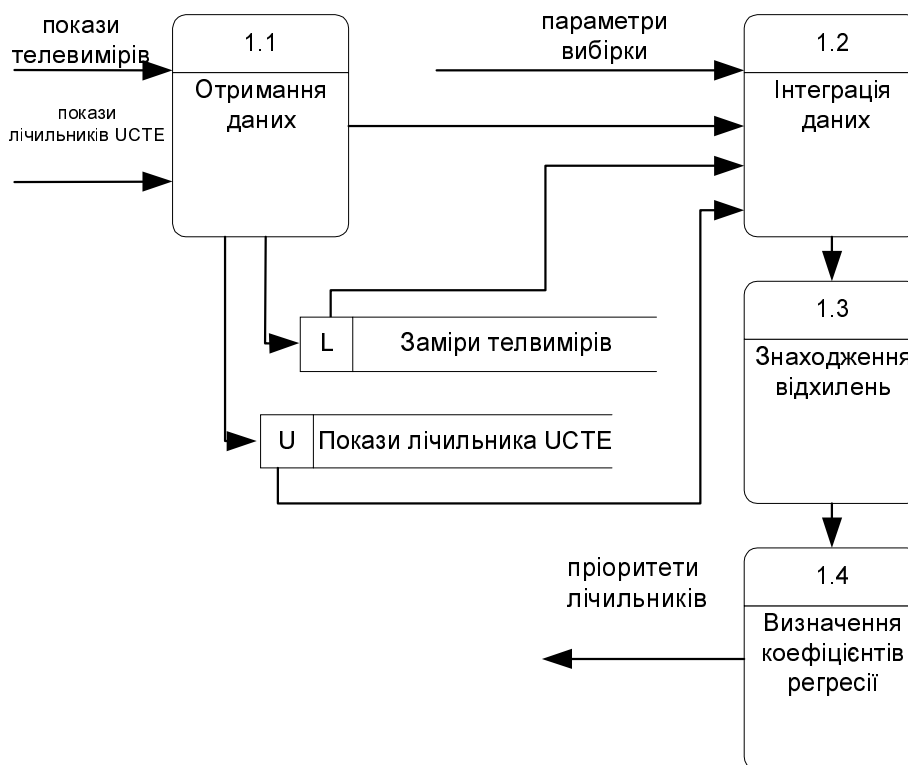


Рис. 2. Деталізація процесу
“Мінімізація відхилення експорту енергії”

Тут процес 1 складається з 4 підпроцесів. Інформація із зовнішніх сутностей заноситься у внутрішні сховища даних. На основі параметрів вибірки вибираються покази зовнішніх телевимірів з вказаного часового діапазону та інтегруються із показами лічильника УСТЕ. Розраховується значення відхилень та будується рівняння регресії, з якого і визначаються пріоритети телевимірів.

Деталізація процесу визначення коефіцієнтів регресії зображена на рис. 3.

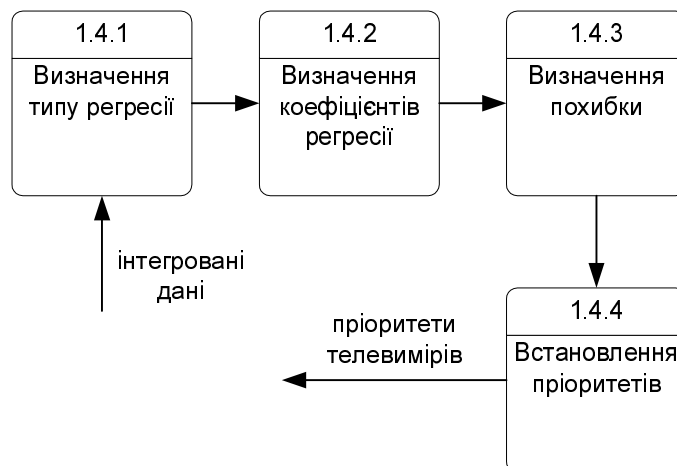


Рис. 3. Деталізація процесу “Визначення коефіцієнтів регресії”

Опис схеми бази даних

Перейдемо до опису реалізації завдання.

Перш за все спроектуємо схему бази даних.

Як вже зазначалось, база даних реалізована у середовищі InterBase.

Схеми відношень, які належать до реалізації завдання, наведені на рис. 4.

Тут відношення з префіксом rf_ призначені для збереження довідкової інформації. Відношення rf_tele призначене для збереження назв телеліній

Відношення RENERGY призначене для збереження величини перетоку активної енергії (в кВт за годину)

В таблиці є три поля типу TdateTime:

GRDate – дата та час за Гринвічем (+0 GMT);

UADate – Київська дата та час (+2 GMT);

PLDate – Варшавська дата та час (+1 GMT);

Type_date – тип часу (літній, зимовий);

Date – дата надходження даних;

Time – час надходження даних;

Name_file – назва файла даних.

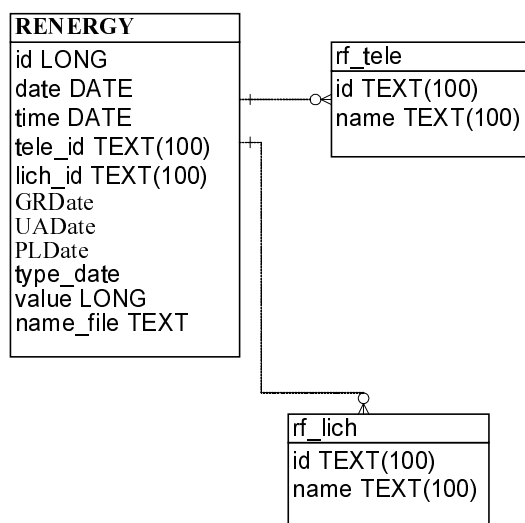


Рис. 4. Схема бази даних

Опис алгоритмів

Алгоритм аналізу даних показів телевимірів полягає у виконанні таких укрупнених кроків:

Крок 1 З БД погодинних значень переданої на експорт електроенергії вибрати значення за потрібний проміжок часу.

Крок 2 З посекундного архіву БД телевимірів вибрати дані за той же самий проміжок часу. Це відпрацьовані ЦР дані.

Крок 3 Дані, отримані на попередньому кроці, інтегруються.

Крок 4 Проінтегровані дані порівнюються з фактичними показами лічильника і визначається відхилення від них.

Крок 5 Використовуючи балансування відхилень, знаходять оптимальні для роботи регулятора телевиміри. За знайденими результатами видають відповідні рекомендації на наступний проміжок часу.

Як вже зазначалося, вибір оптимального телевиміру буде проходити за результатами застосування двох методів: регресійного аналізу та МГУА з критерієм відбору з розділенням на вибірки.

Наведемо блок-схеми відповідних методів (рис. 5, 6).

Помилка наближення або критерію дисперсії прогнозу рекомендується як критерій відбору моделей:

$$\delta_i^2 = \frac{\sum_1^N (y_i - \hat{y}_i)^2}{\sum_1^N (y_i - \bar{y})^2} \rightarrow \min, \quad (1)$$

де y_i – фактичне значення вихідної змінної; \hat{y}_i – значення вихідної змінної, розраховане за оцінюваною моделлю; \bar{y} – її середнє значення.

Критерій регулярності визначає середнє квадратичне відхилення моделі на наборі тестованих даних:

$$AB = \frac{1}{N_B} \sum_{i=1}^{N_B} [y_i - \hat{y}_i(A)]^2 \rightarrow \min,$$
$$\Delta^2(B) = \frac{\sum_{i=1}^{N_B} [y_i - \hat{y}_i(A)]^2}{\sum_{i=1}^{N_B} y_i^2},$$

де $\hat{y}_i(A)$ – зразковий вихід, оцінений на N_A .

Коефіцієнти регресійної поліміальної функції другого порядку та похибка розраховуються за такими формулами:

$$b_0 = \frac{\sum_{i=1}^N y_i \sum_{i=1}^N x_i^3 - \sum_{i=1}^N x_i \sum_{i=1}^N x_i^2 y_i}{N \sum_{i=1}^N x_i^2 - \left(\sum_{i=1}^N x_i^3 \right)^2}, \quad b_1 = \frac{\sum_{i=1}^N y_i \sum_{i=1}^N x_i^2 - \sum_{i=1}^N x_i \sum_{i=1}^N x_i y_i}{N \sum_{i=1}^N x_i^2 - \left(\sum_{i=1}^N x_i^2 \right)^2},$$
$$b_2 = \frac{N \sum_{i=1}^N x_i y_i - \sum_{i=1}^N x_i \sum_{i=1}^N y_i}{N \sum_{i=1}^N x_i^2 - \left(\sum_{i=1}^N x_i \right)^2}, \quad \sigma_e = \frac{\sum_{i=1}^N (b_0 - b_1 x_i - b_2 x_i^2 - y_i)^2}{(N-2)^{1/2}}.$$

Далі похибки впорядковуються за зростанням. Впорядкування похибок означає впорядкування телевимірів за спаданням пріоритету.

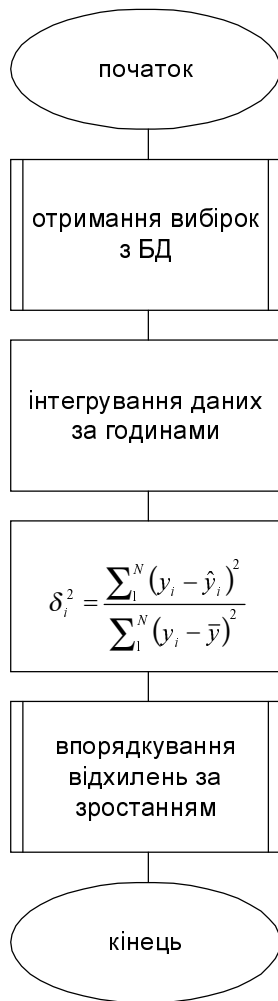


Рис. 5. Блок-схема алгоритму МГВА з критерієм відбору з розділенням на вибірки

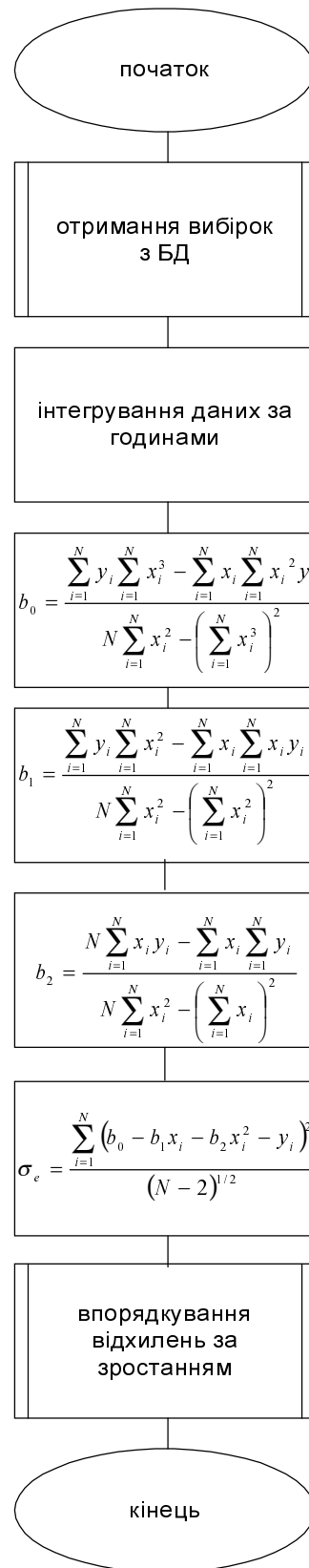


Рис. 6. Блок-схема алгоритму регресійного аналізу

Висновки

Розглянуто проблему визначення оптимального телевиміру на основі порівняння середньо-квадратичних відхилень. Ставилася задача мінімізації відхилення експорту енергії від заплано-

ваного. Досягнути цього можна, підбираючи виміри так, щоб відхилення протягом кожної години були мінімальними. Якщо проінтегрувати кожен вимір протягом години і порівняти його з показами телевимірів за цю ж годину, то можна визначити відхилення кожного виміру від показів лічильника. Якщо вибрати для роботи регулятора виміри, які дають мінімальний баланс похибок, то досягнемо найкращих результатів регулювання (за точністю). Маніпулюючи пріоритетами вимірів на регуляторі (регулятор завжди використовує вимір з вищим пріоритетом), досягнемо кінцевої мети.

1. Ивахненко О.Г.: Метод группового учета аргументов – конкурент метода стохастической аппроксимации // Автоматика. – 1968. – №3. 2. Аксенова Т.Т., Юрачковський Ю.П. Характеризация несмещенной структуры и условия ее J-оптимальности // Автоматика. –1988. – №4. 3. Ивахненко А.Г. Долгосрочное прогнозирование и управление сложными системами. – К.: Техника, 1975. 4. Ивахненко А.Г.: Моделирование сложных систем. Информационный подход. – К.: Высшая школа, 1987. 5. Рекомендации: Предварительные эксплуатационные опыты перед началом прямого подключения новых сетей к объединенной энергосистеме УСРТЕ. – Варшава, 1997. 6. Stepień M., Mrowiec M., Mazurek P. Описание центрального Регулятора для Острова Буриштын. Wrocław, Instytut Automatyki Systemow Energetycznych – December 1997. 7. Cichosz J., Gorski M. Техническая документация центрального Регулятора Острова Буриштын. – Wrocław, Instytut Automatyki Systemow Energetycznych – March 1999.

УДК 004.622

Є.Я. Лещинський

Національний університет “Львівська політехніка”,
кафедра інформаційних систем та мереж

МОДЕЛЮВАННЯ ПРОЦЕСУ ЗБИРАННЯ ТА ПОДАННЯ ІЄРАРХІЧНО ОРГАНІЗОВАНОЇ ІНФОРМАЦІЇ НА ПРИКЛАДІ СПОСТЕРЕЖЕНЬ ЗА ІМУНОЛОГІЧНИМИ ПАЦІЄНТАМИ

© Лещинський Є.Я., 2004

Розглянуто проблематику створення інформаційної інтелектуальної системи підтримки прийняття рішень для предметної галузі з ієрархічною організацією інформації. Задача розглядається на прикладі імунологічних досліджень при скринінговому обстеженні пацієнтів. Значну увагу приділено етапам збирання інформації та її подальшого подання. Проаналізовано реальні дані, отримані з анкет спостереження за пацієнтами. Запропоновано варіант практичної реалізації моделі збирання даних, яка враховує специфіку ієрархічної організації інформації у досліджуваній предметній області.

Problems of creation the informational intellectual system for decision-making support for objective sphere with hierarchical structure of the information are considered. The problem is investigated on immunological observation during examining of patients. Meaningful attention is paid for phase of gathering the information and it's further representation. Analysis of the real data obtained from questionnaire of observation for patients is done. Practical achievements of the model of gathering the data, which take into consideration features of hierarchical structure of the information in concerned sphere, are offered.

Постановка проблеми у загальному вигляді

Створення інформаційної системи, що стосується тієї чи іншої сфери життя людини або її діяльності, починається з етапу опису та моделювання предметної області. Складовою цього процесу є подання об'єктів та понять реального світу у формі, що, з одного боку, відповідає дійсності,