**Лабораторна робота №9**

МОДЕЛЬ ПРОГНОЗУВАННЯ ЧАСОВИХ РЯДІВ ЗА ВИБІРКОЮ МАКСИМАЛЬНОГО ПОДІБНОСТІ

**Мета роботи**: вивчити методи прогнозування за допомогою числових рядів за вибіркою максимальної правдоподібності

### Завдання до роботи

1. Ознайомитися з конспектом лекцій та рекомендованою літературою за темою роботи, а також з додатком І, що містить опис програмного забезпечення для видобування асоціативних правил з великих масивів даних.
2. Сформувати набір даних для обробки та аналізу.
3. На основі обраних даних побудувати модель прогнозування за допомогою числових рядів, як це показано у додатку І.
4. Оцінити коректність побудованої моделі.
5. Оформити звіт з роботи.
6. Відповісти на контрольні питання.

### Зміст звіту

1. Тема та мета роботи.
2. Короткі теоретичні відомості.
3. Набір даних для обробки (якщо він великий – навести фрагмент).
4. Детальний опис процесу використання програмного забезпечення для обробки набору даних, що має бути проілюстрований зображеннями екранних форм з описом кожного їх елементу та коментарями до кожної дії.
5. Текст програми з коментарями.
6. Результати роботи програмного забезпечення (набір отриманих правил, інші характеристики).
7. Висновки, що містять відповіді на контрольні запитання, а також відображують результати виконання роботи та їх критичний аналіз.

### Контрольні питання

1. Що таке часові ряди?
2. Компоненти числових рядів.
3. Для чого призначені часові ряди?
4. Параметри моделі прогнозування EMMSP.
5. Оцінка помилки прогнозування.
6. Алгоритм екстраполяції часового ряду без урахування зовнішніх чинників Визначити вибірку нової історії.
7. Визначити вибірку максимального подібності.
8. Визначити вибірку базової історії.
9. Алгоритм побудови довірчого інтервалу прогнозних значень для розробленої моделі екстраполяції.

**Додаток І**

**ПОБУДОВА МОДЕЛІ ПРОГНОЗУВАННЯ ЗА ВИБІРКОЮ МАКСИМАЛЬНОГО ПОДІБНОСТІ**

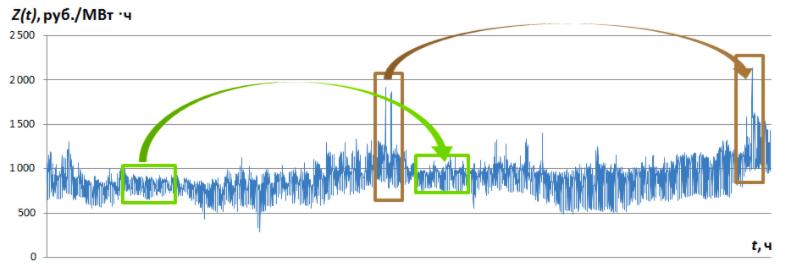
1. Пояснення моделі прогнозування за вибіркою максимального подібності

1.1. Основна ідея та її ілюстрація на вибірках тимчасового ряду

Повний формальний опис моделі прогнозування можна знайти у другому розділі моєї дисертації. Однак якщо говорити просто, то в основу моделі покладено ідею про розвиток історії по спіралі: етапи повторюються, але до мінливих властивостями.

Якщо прикласти зазначену ідею до часових рядах, то можна сказати так: в фактичні значення часового ряду, якщо їх досить багато, напевно, є відрізок, який дуже схожий на той, що відбувається напередодні прогнозу.

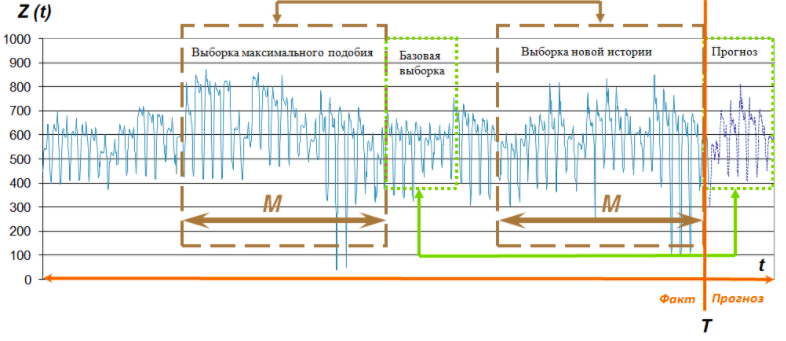
На рис. І.1 представлена частина часового ряду Z (t), для якого без спеціальних обчислень помітні схожі відрізки. Назвемо відрізок тимчасового ряду, що має початок в позначці часу tи довжину (число відліків) M, вибіркою тимчасового ряду (time series pattern) і позначимо ZMt.



Мал. І.1. Часовий ряд Z(t) і деякі його вибірки

Будь-яка модель прогнозування виходить з деякого припущення. У моделі за вибіркою максимального подібності передбачається, що якщо історія повторюється, то для кожної вибірки, що передує прогнозом, є подібна вибірка, що міститься в фактичні значення цього ж часового ряду. Формально це називається гіпотеза подібності.

Співвідношення вибірок і їх назву можна побачити на рис. І.2.



Мал. І.2. Вибірка часового ряду Z (t)

У момент часу T, який називається моментом прогнозу, потрібно визначити P значень часового ряду в майбутньому, тобто обчислити вибірку Прогноз. При цьому значення Вибірки нової історії є доступними. Далі, виходячи з припущення про те, що для кожної вибірки є подібна, потрібно знайти Вибірку максимального подібності для Вибірки нової історії і припустити, що історія повториться, тобто основою для прогнозних значень стане Базова вибірка.

Далі необхідно відповісти на три питання.

1.2. Як визначити подобу вибірок?

У своїй дисертації я пропоную найпростіший варіант визначення подібності - обчислення значення лінійної кореляції. Беремо одну вибірку довжини M, беремо іншу вибірку довжини M, вважаємо значення кореляції, яке і буде відображати подібність двох вибірок.

Шукати Вибірку максимального подібності найпростіше методом перебору серед всіх можливих вибірок. Для часових рядів до 100 000 значень такого сорту перебір займає кілька секунд при реалізації на персональному комп'ютері середньої потужності.

Запропонований підхід визначення подібності не є єдино можливим.

1.3. Як «перенести» зміна властивостей?

У зв'язку з тим, що в основу визначення подібності закладена лінійна кореляція, то найпростішим варіантом «перенесення» властивостей вибірок є лінійна залежність

 (І.1)

Вибірка новітньої історії. Вибірка максимальної подібності

Якщо рівняння (І.1) відображає залежність двох фактичних вибірок за допомогою коефіцієнтів α1 і α0, то на підставі припущення про подібність Прогноз і Базова вибірка співвідносяться наступним чином:

 (І.2)

Прогноз Базова вибірка

2. Приклад реалізації моделі прогнозування за вибіркою максимального подібності в MATLAB

Приклад моделі прогнозування реалізований в MATLAB.

2.1. Вихідні дані, постановка завдання прогнозування і параметри моделі

Вихідними даними є значення цін на електроенергію європейської території РФ оптового ринку електроенергії і потужності в руб. / МВт • год з 01.09.2006 до 13.11.2012 в погодинному дозволі. Архів PRICES\_EUR містить 54 384 значення.

Необхідно обчислити прогноз даного часового ряду на P = 24 значення вперед, тобто, на добу вперед в момент прогнозу T = 01.09.2012 23:00:00.

Завантаження вихідних даних:

*% Ціни ринку на добу вперед оптового ринку електроенергії та потужності*

*% В погодинному дозволі, матриця PRICES\_EUR має три стовпці*

*% <Date> - <Hour> - <Value>*

*load PRICES\_EUR PRICES\_EUR; % Завантаження вихідних даних з файлу PRICES\_EUR.mat*

*TimeSeries = PRICES\_EUR;*

Постановка завдання прогнозування:

*% Постановка завдання прогнозування*

*T = datenum ('01 .09.2012 23:00:00 ',' dd.mm.yyyy hh: MM: ss '); % Момент прогнозу (Origin)*

*P = 24; % Горизонт прогнозування (Forecast horizon)*

Параметри моделі прогнозування EMMSP:

*% Параметри моделі за вибіркою максимального подібності*

*M = 144;*

*Step = 24;*

В даному випадку довжина Вибірки нової історії і Вибірки максимального подібності дорівнює M = 144. Детальніше про те, як визначати M, читайте в третьому розділі дисертації.

Мінлива Step є кроком перебору фактичних значень при визначенні Вибірки максимального подібності.

2.2. Алгоритм моделі прогнозування за вибіркою максимального подібності

2.2.1. Визначити вибірку нової історії

Index = find (TimeSeries (:, 1) == datenum (year (T), month (T), day (T)) & TimeSeries (:, 2) == *hour (T));*

*if length (Index)> 1*

*fprintf ([ 'Помилка часового ряду: відмітка часу T знайдена в тимчасовому ряді більше 1 разу \ n']);*

*elseif isempty (Index)*

*fprintf ([ 'Помилка часового ряду: відмітка часу T в тимчасовому ряді не знайдено \ n']);*

*else*

*HistNewData = TimeSeries ([Index-M + 1: Index], :); % Вибірка нової історії (HistNewData)*

*Index = Index - Step \* 2;*

*end*

2.2.2. Визначити значення подібності

*k = 1;*

*while Index> 2 \* M*

*HistOldData = TimeSeries ([Index-M + 1: Index], :);*

*Likeness (k, 1) = Index;*

*CheckOld = find (HistOldData (:, 3)> 0); % Перевірка на те, нульові вектори*

*CheckNew = find (HistNewData (:, 3)> 0);*

*if isempty (CheckOld) || isempty (CheckNew)*

*Likeness (k, 2) = 0;*

*else*

*Likeness (k, 2) = abs (corr (HistOldData (:, 3), HistNewData (:, 3), 'type', 'Pearson'));*

*end*

*k = k + 1;*

*Index = Index - Step; % Повернення за часом назад на крок Step*

*end*

Крок перебору Step = 24 взятий з того міркування, що часовий ряд має яскраво виражену добову періодичність.

2.2.3. Визначити максимум подібності

*MaxLikeness = max (abs (Likeness (:, 2)));*

*IndexLikeness = find (Likeness (:, 2) == MaxLikeness);*

*MSP = Likeness (IndexLikeness (1), 1);*

Практика показує, що максимумів подібності може бути кілька. Найзручніше взяти перший.

2.2.4. Визначити вибірку максимального подібності

*MSPData = TimeSeries ([MSP-M + 1: MSP], :); % Вибірка максимального подібності (MSPData)*

2.2.5. Визначити базову вибірку

*HistBaseData = TimeSeries ([MSP + 1: MSP + P], :); % Базова вибірка (HistBaseData)*

2.2.6 Пошук коефіцієнтів α1 і α0

*% Робимо апроксимацію HistNewData за допомогою MSPData за методом найменших квадратів.*

*% В даному випадку рішення знаходиться через зворотну матрицю.*

*% У дисертації рішення зроблено іншим способом, проте матричний метод мені більше подобається.*

*X = MSPData (:, 3);*

*X (:, length (X (1,:)) + 1) = 1; % Додаємо стовпець з одиничним вектором I*

*Y = HistNewData (:, 3);*

*E = X (:, 2);*

*Xn = X '\* X;*

*Yn = X '\* Y;*

*invX = inv (Xn);*

*A = invX \* Yn; % Шукані коефіцієнти alpha1 і alpha0 є значеннями матриці A*

Після знаходження значень коефіцієнтів α1 і α0 можна зробити перевірку, з якою помилкою Вибірка нової історії подібна вибірка максимального подібності, т. Е. Яка помилка рівняння (1). У цьому прикладі перевірка для простоти опущена.

2.2.7. прогнозування

*X = HistBaseData (:, 3);*

*X (:, length (X (1,:)) + 1) = 1;*

*Forecast = X \* A; % Вибірка Прогноз, що містить 24 прогнозних значення*

У реалізованому прикладі число прогнозних відліків може бути змінено за допомогою зміни параметра P.

2.2.8. Оцінка помилки прогнозування

*% 1) Визначаємо фактичні значення*

*Index = find (TimeSeries (:, 1) == datenum (year (T), month (T), day (T)) & TimeSeries (:, 2) == hour (T));*

*Fact = TimeSeries ([Index: Index + P-1], 3);*

*% 2) Оцінка помилки MAE*

*MAE = mean (abs (Forecast - Fact));*

*% 3) Оцінка помилки MAPE*

*MAPE = mean (abs (Forecast - Fact) ./ Fact);*

*fprintf ([ 'Момент прогнозу T =', datestr (T, 'dd.mm.yyyy HH: MM'), '\ n', 'Горизонт прогнозу P =', num2str (P), '\ n', 'Помилка прогнозу MAE = ', num2str (MAE),' RUB / MWh, MAPE = ', num2str (MAPE \* 100),' %% \ n ']);*

За підсумками виконання програми в командному рядку MATLAB має з'явитися повідомлення:

Максимальна подібність MaxLikeness = 0.95817

Коефіцієнт alpha1 = 1.0312, коефіцієнт alpha0 = -11.1992

Момент прогнозу T = 01.09.2012 23:00

Горизонт прогнозу P = 24

Помилка прогнозу MAE = 57.6383 RUB / MWh, MAPE = 6.0065%