

УДК 621.391

Є.С.Проскурка, В.Д.Кишенько

Національний університет харчових технологій

## ПРЕЦЕДЕНТНИЙ АНАЛІЗ ТЕХНОЛОГІЧНИХ СИСТЕМ ЦУКРОВОГО ВИРОБНИЦТВА З ВИКОРИСТАННЯМ ТОПОЛОГІЧНОЇ ГРАМАТИКИ

*В даній статті досліджуються часові ряди технологічних параметрів на цукровому заводі при використанні топологічної граматики. Часові ряди кодують, потім розбивають на фігури – гексограми, які мають свої топологічні коди. Серед цих гексограм шукають однакові за топологічними кодами. Відслідковується історія повторення паттернів для подальшого прогнозування стану системи.*

Ключові слова: *топологічна граматика, цукрове виробництво, прецедентний аналіз, технологічні системи.*

Технологічні процеси харчових виробництв мають прояви складної поведінки, що характеризуються утворенням і дисипацією просторово-часових структур, визваною суттєвою нелінійністю характеристик об'єкта керування, механізмами виникнення нових режимів, які докорінно відрізняються від функціонування у врівноважених системах. Разом з тим, в поведінці складної динамічної системи, як технологічного комплексу спостерігається певна регулярність, яка може визначена за допомогою прецедентного аналізу.

У більшості енциклопедичних джерел прецедент визначається як випадок, що мав місце раніше й слугує прикладом або виправданням для наступних випадків подібного роду. Вивід на основі прецедентів (CBR - Case-Based Reasoning) є підходом, що дозволяє вирішити нове завдання, використовуючи або адаптуючи рішення вже відомого завдання. Як правило, такі методи розмірковування містять у собі чотири основних етапи, що утворюють так званий цикл розмірковування на основі прецедентів або CBR-цикл [1].

До переваг розмірковування на основі прецедентів можна віднести наступні аспекти:

- можливість прямо використовувати досвід, накопичений системою без інтенсивного залучення експерта в тій або іншій предметній області;
- можливість скорочення часу пошуку рішення поставленого завдання за рахунок використання вже наявного рішення для подібного завдання;
- існує можливість виключити повторне одержання помилкового рішення;
- відсутня необхідність повного і поглибленого розгляду знань про конкретну предметну область;
- можливе застосування евристик, що підвищують ефективність рішення завдань.

До недоліків міркувань на основі прецедентів можна віднести наступне:

- при описі прецедентів звичайно обмежуються поверхневими знаннями про предметну область;
- велика кількість прецедентів може привести до зниження продуктивності системи;
- проблематичним є визначення критеріїв для індексації і порівняння прецедентів;
- проблеми з налагодженням алгоритмів визначення подібних (аналогічних) прецедентів;
- неможливість одержання рішення завдань, для яких немає прецедентів або ступінь їхньої подібності менше заданого граничного значення.

Основна мета використання апарату прецедентів у рамках інтелектуальних систем підтримки і прийняття рішень (ІСППР) полягає у видачі готового рішення оператору для поточної ситуації на основі прецедентів, які вже мали місце в минулому при керуванні даним об'єктом або процесом.

На першому етапі CBR-циклу виконується визначення ступеня подібності поточної ситуації з прецедентами із бази даних (БД) системи і наступний її аналіз з метою вирішення нової проблемної ситуації, що склалася на об'єкті.

Пошук прецедентів в часових рядах відбувається за різними методами. Застосуємо пошук на основі топологічних фігур.

Методи паттерного аналізу визначають тенденцію розвитку поведінки динамічної системи, причому розглядаються тренди різного масштабу (по осі часу). Можливо представити, що поведінка системи описується послідовністю трендів, фрагмент такої послідовності називається фігурою. Тренд є елементарною фігурою. Виникає питання, як записати наступне твердження: поведінка системи описується трендом, цей тренд складається з послідовності трендів і т.д.

Такий рекурсивний опис може бути задано спеціальним механізмом – формальною топологічною граматику.

Побудова тренду, який складається з послідовності трендів, задається формальним правилом згортки такого вигляду [2]:

$$T_{\lambda}^i \rightarrow T_{\lambda_1}^{i-1} * \dots * T_{\lambda_i}^{i-1} * \dots * T_{\lambda_n}^{i-1} \quad (1)$$

Це правило визначає одноразове перетворення і читається так; “тренд і-го рівня складається з послідовності n трендів (і-1)-го рівня”. Символ  $\lambda_j \in \{0,1\}$  позначає спадний і висхідний тренд. Права частина правила фактично визначає фігуру, складену з елементів, "\*" – визначає операцію приписування і утворення ланцюгів трендів.

Більш складне правило задає багатокрокове перетворення:

$$T_{\lambda}^i \leftarrow T_{\lambda_1}^{i-x_1} * \dots * T_{\lambda_j}^{i-x_j} * \dots * T_{\lambda_n}^{i-x_n} \quad (2)$$

де  $x_j < i$ ,  $T_{\lambda_j}^{i-x_j}$  – тренд довільного рівня, але нижче тренду і-го рівня.

Можна собі представити ще більш складне правило:

$$T_{\lambda_1}^{y_1} * \dots * T_{\lambda_j}^{y_j} * \dots * T_{\lambda_k}^{y_k} \leftarrow T_{\lambda_1}^{x_1} * \dots * T_{\lambda_j}^{x_j} * \dots * T_{\lambda_n}^{x_n} \quad (3)$$

коли деяка послідовність трендів нижнього рівня (фігура нижнього рівня) перетвориться в послідовність високого рівня (фігуру більше високого рівня), де  $k < n$  і для всіх  $i, j, y_j < x_i$ .

Існує, ще один механізм конструювання фігур за допомогою рекурсивного уточнення тренду через тренди, що складають його. Перепишемо правило згортки, змінивши напрям згортки і збільшення індексу трендів Tr:

$$Tr_{\lambda_i}^i \rightarrow Tr_{\lambda_{i+1,1}}^{i+1} Tr_{\lambda_{i+1,2}}^{i+1} \dots Tr_{\lambda_{i+1,n}}^{i+1} \quad (4)$$

Тоді правило можна розуміти, як механізм, що породжує тренди. Набір правил, що породжує тренди, визначає породжувальну граматику:

$$\begin{aligned} Tr^0 &\rightarrow Tr_{1,1}^1 \dots Tr_{1,m_1}^1 \\ Tr_1^1 &\rightarrow Tr_{2,1}^2 \dots Tr_{2,m_2}^2 \\ &\dots \dots \dots \\ Tr_n^1 &\rightarrow Tr_{n,1}^2 \dots Tr_{n,m_n}^2 \dots \end{aligned} \quad (5)$$

Зміст цієї граматики – послідовне уточнення (розкриття тренду) від простого до складного (зверху-вниз). Для правил породження так само, як і для правил складання, вводиться процедура наскрізної нумерації точок, котрі «звужують» тренд.

Топологічна граматика задається чотирма компонентами:

$$GT = \langle Tr^0(1,0), P, Q, \delta \rangle \quad (6)$$

1.  $Tr^0$  – аксіома ( $Tr^0(01)$  або  $Tr^0(1,0)$ );
2. P – правила структуризації, уточнення трендів зі стандартним способом наскрізної нумерації;
3. Q – топологічні коди фігур, що відповідають правилам;
4.  $\delta$  – посилання термінальних трендів на аксіому.

Метод виділення топологічних фігур (паттернів) та пошук серед них однакових в історії часового ряду для прогнозування майбутньої тенденції бере свій початок в технічному аналізі. Використовують наступні вимоги для виділення фігур:

1. Обов'язково для графічної побудови фігури виділяється кінцева множина точок.
2. Фігура складається з елементів, для цього виділяються пари точок. Факт виділення позначається на малюнку лінією, що покриває початкову і кінцеву точку пари.
3. Фігури представляються у вигляді ланцюга із цих ліній, що замикає початкову і кінцеву точки цієї фігури;
4. Кожній лінії (парі точок) ставиться у відповідність деяке відношення (ознака): висхідний, спадний або незмінний характер положення лінії в ланцюзі, що описує фігуру.

5. Розпізнавання фігури (віднесення її до якого-небудь класу) відбувається візуально.

Надалі для формального опису фігур буде використаний «візуальний» підхід, що склався в технічному аналізі. Ідея формалізації полягає в навішенні на фігуру «ярлик» з кодом, по якому можна розпізнати фігуру.

Введемо наступний спосіб кодування фігур:

1. Проводиться нумерація точок. Початкова точка фігури нумерується «1». Далі нумерація йде по порядку в історії появи точок. Кінцева точка фігури нумерується «7», вона є точкою «1» для наступного паттерну.

2. Точки, що становлять фігуру, проєктують на вісь (площину) характеристик. Утворюються рівні, що проходять через кожну окрему точку.

3. Код визначається як послідовність номерів точок, починаючи із самої нижньої по осі характеристик і закінчуючи самої верхньою.

На рис.1 показаний спосіб кодування деякої фігури і її код (TU=3214756). Спосіб кодування називається топологічним, а отриманий код – топологічним кодом фігури (TU). Назва «топологічний» підкреслює той факт, що при описі фігури не враховуються її метричні властивості. TU-коди «записують» відносини порядку (вище/нижче) між рівнями, на яких перебувають точки, і відношення порядку в часовому відліку. Тому TU-коди називаються топологічними рівневими кодами.

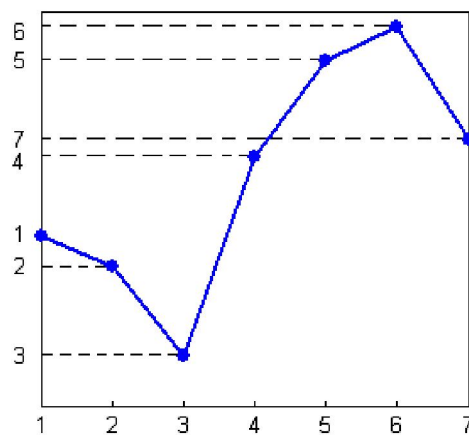


Рис. 1. Топологічна фігура з кодом TU=3214756

Кількість точок було вибрано сім. Сім точок утворюють шість відрізків. Фігура, що складається з шести відрізків називається гексограмою.

Особлива цінність гексограми полягає в тім, що це мінімальна фігура, що дозволяє виділити перелом або збереження поведінки динамічної системи [2].

На рис. 2 зображений часовий ряд витрати дифузійного соку на цукровому заводі. В цьому часовому ряду спробуємо знайти однакові паттерни за допомогою програми, яка була розроблена за допомогою математичного пакету MATLAB. Програма розбиває часовий ряд на N-ну кількість паттернів, де N визначається за формулою (1), і нумерує їх від 1 до N.

$$N = \text{floor}(n/6) - 1, \quad (7)$$

де  $n$  – кількість відрізків часового ряду;  $\text{floor}(n/6)$  – операція, що округлює результат ділення до меншого цілого числа.

Програма присвоює кожному паттерну свій топологічний код. Потім шукає серед цих паттернів однакові за топологічним кодом і видає їх номери. Маючи номери паттернів за формулами (2) і (3) можна дізнатися, початкову і кінцеву точку паттерну на осі часу:

$$t_p = (N - 1) \cdot 6 + 1, \quad (8)$$

$$t_k = t_p + 6, \quad (9)$$

де  $t_p$  – початкова точка паттерну;  $t_k$  – кінцева точка паттерну.

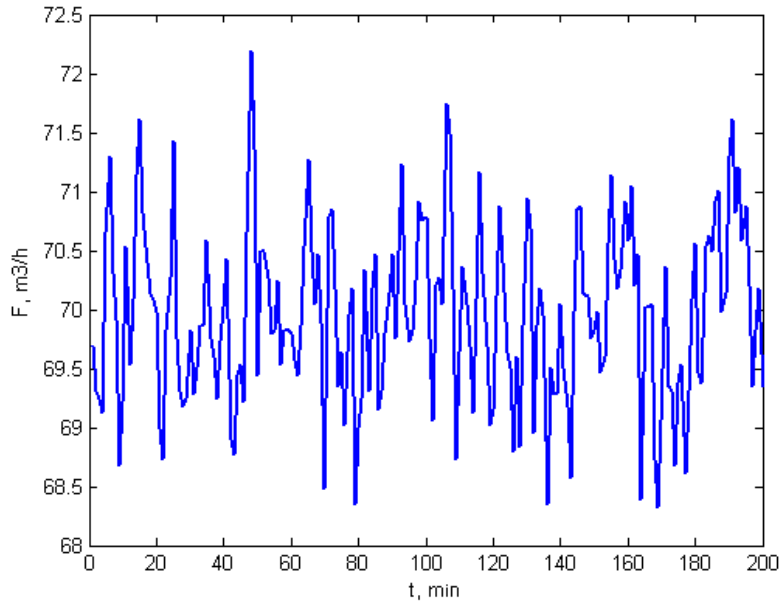
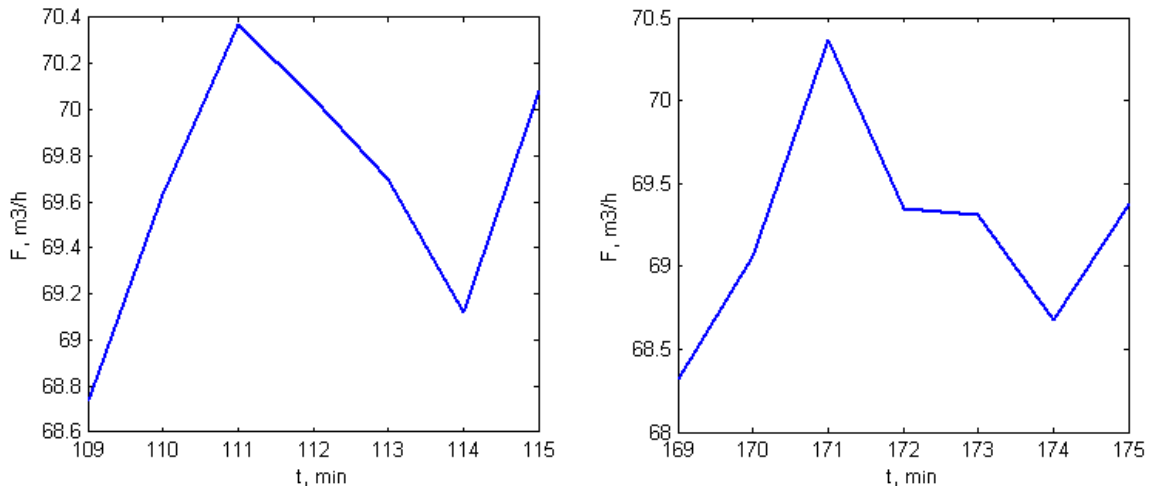


Рис. 2. Часовий ряд витрати дифузійного соку



а) виділений паттерн №19, TU=1625473

б) виділений паттерн №29, TU=1625473

Рис. 3. Виділені паттерни з часового ряду

Отримані паттерни із часового ряду, представлено на рис. 2, та їх номери зображені на рис. 3. Як видно з рисунку програма знайшла два однакові паттерни (одну пару), під номерами №19 та №29 і топологічним кодом TU=1625473.

Ці дві фігури не мають однакового виду. Зрозуміло, що більшість реальних фігур, зображених у метричному просторі, може мати той самий TU-код, але різний вид. Тоді ця множина називається множиною еквівалентності. TU-код фігури є образом (паттерном) деякої множини фігур, що існують у метричному просторі. Уводячи в топологічний простір яку-небудь метрику, ми стискаємо множину еквівалентності. Загалом кажучи, можна звести множину еквівалентності до єдиного екземпляра. Проблема розширення або звуження паттернів лежить поза математичними методами і повинна вирішуватися при аналізі характеристик динамічної системи [2].

З рис. 2 видно, що сигнал має шумові компоненти. Спробуємо відфільтрувати сигнал та видалити шумові компоненти та знайти в очищеному сигналі однакові паттерни.

Сигнал будемо фільтрувати за допомогою дискретного вейвлет-перетворення використовуючи при цьому вейвлет Добеші другого порядку (db2), та розкладаючи сигнал до другого рівня деталізації [3].

На рис. 4 представлений сигнал витрати дифузійного соку на цукровому заводі, який ми очистили за допомогою дискретного вейвлет-перетворення. За допомогою програми MATLAB © Є.С.Проскурка, В.Д.Кишенько

знайдемо однакові паттерни по топологічними кодами в часовому ряду. Результати, які були отримані програмою наведенні в таблиці 1.

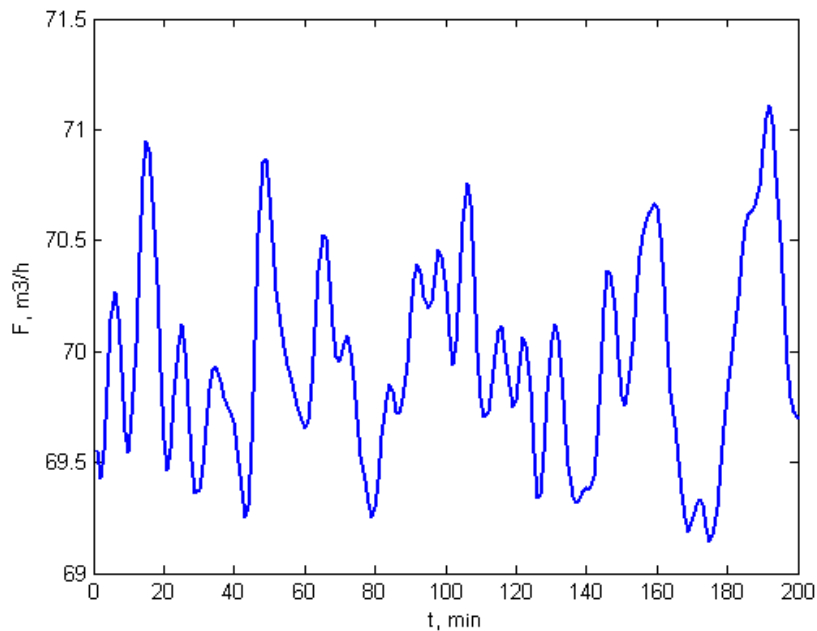


Рис. 4. Сигнал витрати дифузійного соку, очищений від шумів

Таблиця 1

Результати дослідження на виявлення паттернів в часовому ряду витрати дифузійного соку

№ з/п	Номер паттерну 1	Номер паттерну 2	№ з/п	Номер паттерну 1	Номер паттерну 2
1	2	3	4	5	6
1	3	49	18	13	33
2	7	9	19	13	39
3	7	13	20	15	35
4	7	28	21	17	21
5	7	33	22	22	37
6	7	39	23	26	30
7	8	26	24	26	31
8	8	30	25	26	42
9	8	31	26	28	33
10	8	42	27	28	39
11	9	13	28	30	31
12	9	28	29	30	42
13	9	33	30	31	42
14	9	39	31	33	39
15	11	22	32	36	43
16	11	37	33	40	44
17	13	28	34	45	47

З таблиці 1 видно, що в часовому ряду було знайдено 34 пари паттернів. Деякі номери паттернів повторюються один та більше разів.

З проведеного досліді видно, що чим менше шумових компонент в часовому ряду тим більша кількість знайдених однакових паттернів за топологічними кодами.

---

**Висновки**

Проведені дослідження показують, що кількість знайдених однакових паттернів за топологічними кодами зростає при зменшенні шумових компонент в сигналі, який досліджується. Знайдені паттерни однакові за топологічними кодами, але різні за формою.

Надалі ведеться робота по створенню бази прецедентів, в якій можна було знайти поточний паттерн та паттерни, що йдуть за ним, для прогнозування майбутнього стану системи. Для точного прогнозування потрібно вводити метрику, яка зменшить кількість однакових паттернів за топологічним кодом, але різних за формою та коефіцієнти адаптації, що будуть розтягувати чи звужувати паттерни, не змінюючи їх топологічні коди, але робити їх однаковими за метричними показниками.

1. Aamodt A., Plaza E. Case-based reasoning: foundational issues, methodological variations, and system approaches // AI Communications. IOS Press. Vol. 7: 1. 1994. – P. 39-59.
2. Столяров Л.Н. Введение в теорию дискретного прецедентного анализа динамических систем // Труды научной конференции студентов и аспирантов в секции «Финансовая аналитика». – М.: МФТИ. 1996. – с. 13-45.
3. Кишенько В.Д., Проскурка Є.С. Фільтрація вхідної інформації в підсистемах технологічного моніторингу систем керування цукровим виробництвом. – Х.: “Восточно-Европейский журнал передовых технологий”, (Математика и кибернетика - фундаментальные и прикладные аспекты), 4/8 (40) 2009.- С. 16-20.