

INFORMATIONAL SUPPLY OF THE CONTROL SYSTEM AT THE BEETPROCESSING DEPARTMENT OF THE SUGAR FACTORY

A. Bezuhlov, A. Ladanyuk

National University of Food Technologies

Key words:

*Intellectual methods
Beet processing
Corelation
Silhouette method
MARGINAL distribution
Copula*

Article history:

Received 07.11.2018
Received in revised form
22.11.2018
Accepted 10.12.2018

Corresponding author:

A. Bezuhlov
E-mail:
bezuhlov2andrii@
gmail.com

ABSTRACT

The paper deals with an approach of better representation of technological objects with the help of the latest methods of data analysis. The need to improve models of technological processes raises question of finding approaches that will simplify finding the regularities of complex technological processes and systems. The technological process such as the beet-processing department(BPD) of a sugar plant has a number of problems that have not been solved to date in terms of the process efficiency at the level of loss reduction and productivity increase. For these purposes, the researchers use methods of analysis of input data and the probabilistic method of their presentation.

The use of the silhouette method makes it possible to find the number of groups by their interaction with each other without being tied to the time coordinates. This provides a qualitatively new result based on the possibility of using intellectual and situational approaches. The method of silhouettes is used to analyze technological situations in which the copulas determine the interdependence between the technological variables of the BPD due to the fact that copulas are built on the basis of marginal distribution and have different variants of their structure. Unlike the correlation analysis, copulas provide the opportunity to simultaneously analyze and find dependencies of an area of variables at once, and for each pair of variables the copula may be different.

In the paper, the authors used the time series data of the technological process of BPD, carried out the analysis of literary sources and considered methods that were previously not used for the technological processes of sugar production.

As a result, the silhouette of the real data of the technological process of BPD are shown, which determined the technological regimes which correspond to the normal course of the process and to individual regimes that do not satisfy it. The analysis by the copula method showed the possibility of finding the regularities between variables if they are nonlinear and are not tied to time. Analysis by the copula can be also a prototype for creating larger models for integration into decision support systems or intellectual management systems. The beginning of such implementation is laid out in the article.

DOI: 10.24263/2225-2924-2018-24-6-5

ІНФОРМАЦІЙНЕ ЗАБЕЗПЕЧЕННЯ СИСТЕМИ КЕРУВАННЯ БУРЯКОПЕРЕРОБНИМ ВІДДІЛЕННЯМ ЦУКРОВОГО ЗАВОДУ

А.О. Безуглов, А.П. Ладанюк

Національний університет харчових технологій

У статті розглядається забезпечення кращого представлення технологічних об'єктів за допомогою новітніх методів аналізу даних. З огляду на вимогу покращення моделей технологічних процесів зроблено спробу знайти підходи, які спростять знаходження закономірностей складних технологічних процесів і систем. Технологічний процес, зокрема бурякопереробне відділення (БПВ) цукрового заводу, має ряд проблем, які не вирішені на сьогодні з точки зору ефективності на рівні зменшення втрат і збільшення продуктивності. Для цих цілей використовують методи аналізу вхідних даних і ймовірнісний метод їх представлення.

Використання методу силуетів дає можливість знайти кількість груп за їх взаємодією між собою без прив'язки до часових координат, що забезпечує якісно новий результат, зважаючи на можливість використання інтелектуальних і ситуаційних підходів. Метод силуетів застосовується для аналізу технологічних ситуацій. Копули визначають взаємозалежності між технологічними змінними БПВ завдяки тому, що побудовані на основі маргінального розподілу і мають різні варіанти своєї структури. На відміну від кореляційного аналізу, копули дають можливість одночасно аналізувати та знаходити залежності відразу масиву змінних, причому для кожної пари змінних копула може бути різною.

У статті використано дані часових рядів технологічного процесу БПВ, проаналізовано літературні джерела та розглянуто методи, які раніше для технологічних процесів цукрового виробництва не використовувались.

Показано силуети реальних даних технологічного процесу БПВ, які визначили технологічні режими, що відповідають нормальному перебігу процесу та таким процесам, що не задовольняють його. Аналіз методом копул показав можливість знаходження закономірностей між змінними, якщо вони мають нелінійний характер і не прив'язані до часу. Також аналіз копулами може служити прототипом створення більших моделей для інтеграції в системи підтримки прийняття рішень або системи інтелектуального управління в цілому. Початок такої реалізації покладено у поданій статті.

Ключові слова: інтелектуальні методи, обробка буряків, кореляція, метод силуетів, маргінальний розподіл, копула.

Постановка проблеми. Бурякопереробне відділення (БПВ) цукрового заводу практично забезпечує темп виробництва щодо кількості перероблених буряків та зміни матеріальних потоків на системах очистки дифузійного соку та випарювання, формуючи таким чином основні техніко-економічні показ-

ники виробництва і здійснюючи забезпечення необхідної якості готової продукції. В технічній літературі є значна кількість методів, в яких розглядаються різні аспекти функціонування та управління БПВ, основна частина присвячена керуванню дифузійними установками різних типів. Однак практично відсутні роботи комплексного характеру з огляду на те, що БПВ включає, крім дифузійних апаратів, різні механізми та прилади для транспортування буряків та дифузійного соку, насоси, збірники тощо, а значне місце займають бурякорізки, які забезпечують необхідну кількість та якість бурякової стружки (її довжину).

Аналіз останніх досліджень і публікацій. У Національному університеті харчових технологій (НУХТ) проведені дослідження, значна частина яких присвячена задачам керування БПВ (в основному дифузійними апаратами) [1—3]. На сучасному етапі у НУХТ використовуються сучасні підходи для розвитку напряму управління БПВ і цукровим заводом у цілому.

Мета статті: дослідити можливості підвищення ефективності систем автоматичного управління (САУ) шляхом використання новітніх методів аналізу даних, таких як метод ліквітів і метод копул, для розробки моделей і в подальшому систем управління бурякопереробними відділеннями цукрового заводу, провести аналіз технічної літератури для оцінки можливості використання цих методів у контексті автоматизації бурякопереробного відділення.

Викладення основних результатів дослідження. Розробка та практичне використання методів сучасної теорії керування та системного підходу потребують новітніх рішень, насамперед інтелектуальних засобів, способів, підсистем і систем, які базуються на комплексуванні форм об'єднання ефективних керувальних дій [4]. У результаті формуються системи, які засновані на робастно-оптимальних і нейронечітких методах, способах ситуаційного та прецедентного керування (у тому числі адаптивного), можливостях компенсації збурень і запізнювань. Ці методи та способи описані в [5; 6].

Для БПВ з урахуванням його особливостей актуальними є задачі керування, які базуються на сучасному інформаційному забезпеченні з використанням методів:

- аналізу часових рядів (кластеризація, класифікація, обробка з виявленням трендів та аномальних даних тощо);
- побудови нейронечітких регуляторів;
- формування підсистем(систем) підтримки прийняття рішень;
- обґрунтуванні ситуаційних і прецедентних методів керування;

На вимогу оптимізації виробничого процесу цукрового заводу, зокрема БПВ, класичних методів аналізу даних і створення математичних моделей недостатньо. Для того, щоб моделі відтворювали об'єкт у більш широкому, ніж до того вигляді (з притаманним їм сучасним вимогам точністю), вибір та аналіз даних, і подальша з ними робота доповнюється ймовірнісними підходами, що у синтезі з класичними дають збільшення інформативної бази на одному з основних етапів створення моделей.

Серед таких підходів останнім часом поширення набули варіанти з багатовим або маргінальним розподілом величин. До таких методів належать також і методи копул. Для БПВ вперше використано копули для аналізу даних як

ідентифікатора та аналізатора роботи похилого дифузійного апарату. Для того, щоб з'ясувати, що собою являє метод копул для задачі аналізу даних у БПВ, приведемо аналіз самого методу.

Відмінності між копула-функціями Клейтона, Гумбеля та Франка продемонстровані графічно на рис. 6. Видно, що зображені копула-функції мають різний характер поведінки в околицях точок (0,0) і (1,1). Наприклад, копула-функція Гумбеля в межах (1,1) демонструє різкий ріст, тоді як ріст в окремій точці (0,0) менш крутий. Як буде показано далі, вимірюванням характеру цих копул-функцій є коефіцієнт «хвостової залежності». Копула-функція Гумбеля має «верхню граничну відстань», копула-функція Клейтона — «нижні гаусові відстані», а копула-функція Франка не має хвостової залежності. Для двох розподілів, двох умовно незалежних величин для БПВ створюють копули: Стюдента, Гаусової копули, Клейтона Гумбеля, Франка та оцінюють параметри отриманих функцій. Відповідно до існуючих методів, копули аналогічно до кореляцій (Пірсона і рангових кореляцій Спірмена та Кендала) мають критерії, які перевіряють значення на адекватність. Для копул такими параметрами є: основні умови існування, властивості копул і теорема Скляра, характеристики яких відомі з літератури [8].

Існує загальноприйнята схема, за якою отримують необхідні вибірки даних вільного розподілу змінних, залежності яких описуються одним із видів копул. За теоремою Скляра, генерування вибірки випадкових величин $X_1 \dots X_n$ з маргінальними функціями розподілу $F_1 \dots F_n$ здійснюється за схемою, наведеною у [9]: спершу здійснюють генерацію випадкових величин $u_1 \dots u_n$ з рівними на проміжку $[0, 1]$ одновимірними розподілами; одержують значення x_i :

$$x_i = F_{i-1}(u_i), \quad (1)$$

де $F_{i-1}(u_i)$ обернена функція $F(i)$, а з вибірки формують навчальні та робочі множини.

Функція спільного розподілу може бути виражена як функція копули, яка застосовується до окремих розподілів. Розраховується монотонна асоціація (тільки строго збільшена або зменшена, але не змішана) між двома змінними, що залежить від рангу і порядку значень. Використовуючи копули, можна створювати багатоваріантні розподіли специфічних маргінальних одиничних розподілів і обирати тип копул для забезпечення кореляційного зв'язку між змінними.

Згідно з літературою [8; 9], коефіцієнт кореляції Кендалла оцінює ступінь подібності між двома наборами даних, що належать одній і тій же сукупності об'єктів або явищ. Цей коефіцієнт залежить від кількості інверсій пар об'єктів, які будуть необхідні для перетворення одного порядку в інший. Для цього кожен порядок представлений набором усіх пар об'єктів (наприклад, $[a, b]$ і $[b, a]$ — це дві пари, що представляють об'єкти a та b), а значення 1 або 0 призначається цій парі, якщо її порядок відповідає або не відповідає способу об'єднання цих двох об'єктів. Ця схема забезпечує набір бінарних значень, які потім використовуються для обчислення коефіцієнта кореляції Пірсона.

Для порівняння двох упорядкованих «випадкових» множин (одного й того ж об'єкта) підхід Кендалла полягає в тому, щоб підрахувати кількість різних

пар між цими двома упорядкованими множинами. Це число дає відстань між множинами, що називається симетричною різницею відстаней (симетрична різниця — це задана операція, яка асоціює два набори з набором елементів, що належать тільки одному набору). Симетрична різницева відстань між двома множинами упорядкованих пар $P1$ і $P2$ позначається $d \Delta (P1, P2)$.

Коефіцієнт кореляції Кендалла залежить лише від порядку пар, і його завжди можна обчислити, припускаючи, що один з елементів рангової зупинки служить як контрольна точка (наприклад, з $N = 4$ елементів ми вважаємо до-вільно, що перший порядок дорівнює 1234).

Для порівняння двох упорядкованих «випадкових» множин (одного й того ж об'єкта) підхід Кендалла полягає в тому, щоб підрахувати кількість різних пар між цими двома упорядкованими множинами. Це число дає відстань між множинами, що називається симетричною різницею відстаней (симетрична різниця — це задана операція, яка асоціює два набори з набором елементів, що належать тільки одному набору).

Коефіцієнт кореляції Кендала також може інтерпретуватися як стандартний коефіцієнт кореляції, який обчислюється між двома множинами на $N*(N - 1)$ бінарних значень. Після визначення порядку копул, визначають, які копули використовувати для визначення зв'язків (Гаусса, Стьюдента, Гумбеля тощо). Вченими не визначено точного наукового методу для цієї задачі, тому існує декілька основних методологій: аналізують парні графіки розподілів величин і визначають копули графічно; використовують досвід проектування та експертів.

У статті обрано для дослідження змінні витрат бурякової стружки на вході в дифузійний апарат та витрата кількості соку на виході з дифузійного апарата, тому що вони часто мають змінюваний характер у незвичайних технологічних ситуаціях. Графічний еквівалент подання залежності двох змінних стосовно ймовірного розподілу подано нижче на рис. 1а та 1б.

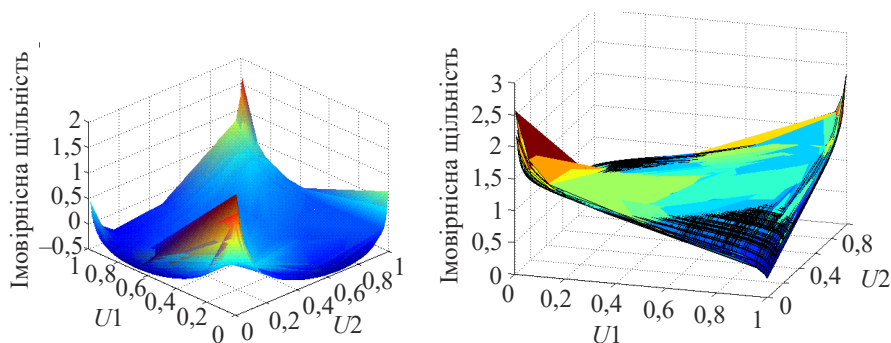


Рис. 1. Розподіли копул: а — розподіл копули Гаусса; б — розподіл Т-копули

Коефіцієнти, отримані при розрахунках (розмах копули та співвідношення викривлень), використовують для створення моделей і надбудов у системи підтримки прийняття рішень, адже їхня структура при навчанні лише доповнюється новою інформацією, що розширює можливі варіанти аналізу технологічних режимів, де змін зазнаватимуть тільки коефіцієнти $\rho_{11} = 9,4391$ для нашого випадку та матриця $Rho = [1,0000 \ 0,1479; 0,1479 \ 1,0000]$.

Цей підхід доповнюється адаптацією методу класифікації за схемою силуетів, який виконує задачу оптимізації кластеризації часових рядів з урахуванням специфіки цукрового виробництва з його інформаційними потоками, які викликають труднощі, однак існують популярні методи для того, щоб задати базові значення класів: метод ліктів і метод силуетів

При навчанні з учителем кількість класів у певному наборі даних відома явно, оскільки кожен екземпляр даних записується як член одного з існуючих класів. Це не можна зробити при навчанні без учителя, де ідея класів та атрибутів як така не створюється, наприклад, в одному з поширених видів навчання без учителя — кластеризації даних, що дає змогу проводити апроксимацію приналежності класів, мінімізуючи міжкласову подібність. Головним недоліком, як відомо з [10], у кластеризації може бути необхідність задавання загальної кількості класів, які можуть існувати у непозначених вхідних і вихідних наборах даних у тому чи іншому вигляді, коли при відомому машинному навчанні використовують так звані базові значення класів (ground truth). Однак не завжди можливо застосувати їх з багатьох причин. Так, у даних взагалі можуть бути невизначені класи (і, відповідно, кластери), тож завдання навчання без учителя полягає в тому, щоб забезпечити кількість класів, які існують у непозначеному (без розмітки) наборі даних на початку аналізу даних, зробити спробу нав'язати певну значиму структуру оптимальних кластерів і класів.

Метод силуету оцінює подібність об'єкта до власного, створеного в процесі кластера, що називається об'єднанням. Цей процес здійснюється у контексті порівняння з іншими кластерами відповідно до ймовірнісного розподілу, зокрема маргінального. Інтерпретуючі знання методу, тобто значення силуетів, є засобом порівняння, яке має діапазон $[-1, 1]$; значення, близьке до 1, вказує на тісний зв'язок з об'єктами у власному кластері, тоді як значення, близьке до -1 , вказує на протилежність [11].

Значення силуету для кожної точки є мірою того, наскільки подібна ця точка до точок у своєму кластері, порівняно з точками в інших кластерах. Значення силуету для i -ї точки, S_i , визначається таким чином:

$$S_i = (b_i - a_i) / \max(a_i, b_i), \quad (2)$$

де a_i — середня відстань від i -ї точки до інших точок того ж кластера, що a_i , і b_i — мінімальна середня відстань від i -ї точки до точок іншого кластера, мінімізована над кластерами. Кластеризовані дані в X використовують підхід із пакета Matlab `kmeans. cidx = kmeans(X, 2)`; після чого виконують створення форми силуету з кластеризованих даних: `silhouette(X, cidx)`.

За замовчуванням силует використовує квадратичну евклідову відстань між точками в X , яка, зважаючи на дані експериментів, дає хибні значення для технологічних змінних (рис. 2).

За допомогою стандартної функції з евклідовою відстанню описати достатньо точно кластери неможливо через великі розкиди початкових даних, які часто виникають у виробничих процесах через свою нестабільність і нестаціонарність або невірно задані початкові розподіли. Натомість можна використати інший метод з тією самою кількістю кластерів.

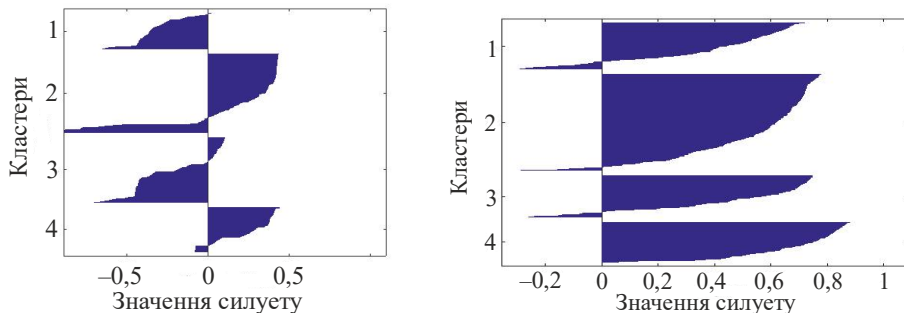


Рис. 2. Кластери: а — нечіткі кластери за евклідовою відстанню;
б — тенденція кластерів силуетами за методом суми абсолютних різниць

З вигляду силуету можна побачити, що більшість точок у ньому мають велике значення силуету — більше 0,6, тож цей кластер має найбільше зі значень, кластер з якого дещо віддалений від сусідніх кластерів. У кожному кластері є певна кількість від'ємно розміщених точок, що показує їхню незадовільну роздільність. Це не дає змоги точно їх призначити класу, але велика кількість точок має значення силуетів більше 0,7—0,8. На основі отриманих даних методом силуетів підставляються отримані кластери з відповідними розподілами для використання у методі копул, що забезпечує більш точні результати аналізу, які неможливо досягти класичними методами для великих об'ємів вибірок.

На основі теоретичних досліджень та експериментальних даних отримано ряд результатів, які дають можливість сформулювати вимоги та практичну реалізацію до систем автоматизації БПВ з урахуванням сучасних технологій теорії керування та їх реалізації в мікропроцесорних системах [12].

Розроблено ряд математичних моделей, які описують динамічні процеси в різних технологічних агрегатах і допоміжних пристроях, наприклад: моделі різки, транспортери, збірники, дифузійні апарати. Це дає можливість оцінити процес функціонування БПВ на різних періодах роботи протягом сезону. В комплексі математичних моделей наявні диференціальні рівняння, статичні характеристики та ланки запізнення. Для використання цих моделей на першому етапі необхідно мати оцінки часових рядів, їх класифікацію та визначення аномальних даних, що відповідає різним виробничим ситуаціям.

Показано, що на відміну від традиційного статистичного аналізу, додаткові технологічні дані може дати підхід, заснований на використанні копул. Копульний аналіз дає змогу аналізу розподілу більше, ніж двох змінних. А сам алгоритм працює над частковим розподілом між змінними. Також слід зазначити, що гауссівські копули і копули Стьюдента найточніше передають залежність між даними для більшості ознак і класів у складних об'єктах, на відміну від інших типів копул.

Висновки

1. Для ефективного керування БПВ як складної системи необхідно використовувати ряд методів, зокрема сучасні підходи аналізу часових рядів, які дають можливість аналізу процесу функціонування.

2. На відміну від існуючих методів, показано, що використання методів копул та кластеризації методом силуетів дає можливість отримати нову інформацію щодо виробничих ситуацій, зокрема на основі аналізу даних.

3. Наведені в статті методи можуть використовуватися для створення бази даних і знань, а також формування підсистеми підтримки прийняття рішень.

Література

1. Способ автоматического управления процессом экстрагирования сахара из свеклы А.С. 1458391 СССР, МПК С 13 D 3/10. 4208860/31-13; заявл. 16.03.1987. опубл. 15.02.1989. Бюл № 6. 3 с.

2. Способ автоматического управления процессом экстракции сахара из свекловичной стружки в непрерывнодействующем диффузионном аппарате А.С. 1129236 СССР, М.Кл. С 13 D 3/10. 3495531/28-13; заявл. 17.09.1982. опубл. 15.12.1984. Бюл № 46. 3 с.

3. Система автоматизації процесів координації підсистем технологічного комплексу цукрового заводу з використанням ситуаційного управління: пат. №73051 Україна: МПК G05B 13/04 №и 201202161; заявл. 24.02.2012 р. опубл. 10.09.2012. Бюл № 17. 4 с.

4. Ладанюк А.П. Синтез АСР методом компенсування керувальних дій. *Автоматизоване управління багатовимірними об'єктами на засадах обчислювального інтелекту: матеріали наук.-практ. конф.*, (Івано-Франківськ 17—19 жовтня 2018р.) Івано-Франківськ ІФНТУНГ. С. 59.

5. Ladaniuk A., Bezuhlov A., Boiko R. Mathematical models for the functioning of the beet-processing department in the sugar production factory. Kriviy Rih: CSITA, 2018. С. 3—12.

6. Ладанюк А.П., Власенко Л.О., Кишенько В.Д., Методологія наукових досліджень: Навч. Посіб. Київ : Видавництво Ліра-К, 2018. 352 с.

7. Berkes P, Wood F, Pillow J (2009) Characterizing neural dependencies with copula models. In: Koller D, Schuurmans D, Bengio Y, Bottou L., editors. *Advances in Neural Information Processing Systems* 21. P. 129—136.

8. de Winter, Joost & Gosling, Samuel & Potter, Jeff. Comparing the Pearson and Spearman Correlation Coefficients Across Distributions and Sample Sizes: A Tutorial Using Simulations and Empirical Data. — *Psychological Methods*. 21. 273—290. 2016. 290 p.

9. Nelsen R. *An Introduction to Copulas*, 2nd ed. — Berlin: Springer-Verlag, 2006. 269 p.

10. János Abonyi, Balázs Feil. *Cluster Analysis for Data Mining and System Identification*. Springer, 2007. P. 303.

11. Rousseeuw P.J. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*. 1987. 20. P. 53—65.

12. Ладанюк А.П., Луцька Н.М., Кишенько В.Д., Власенко Л.О., Іващук В.В. *Методи сучасної теорії управління: підручник*. Київ: Видавництво Ліра-К, 2018. 368 с.