DOI : 10.33274/2079-4827-2020 -41-2-97-86
УДК 621.325.5

***Цвіркун Л. О.**, канд. пед. наук* [1]
***Цвіркун С. Л.**, канд. техн. наук* [2]
***Бондаренко О. О.**, канд. екон. наук* [3]

[1] Донецький національний університет економіки і торгівлі імені Михайла Туган-Барановського (м. Кривий Ріг, Україна), e-mail: cvirkun@donnuet.edu.ua

[2] Криворізький національний університет (м. Кривий Ріг, Україна), e-mail: tserg30@ukr.net

[3] Донецький національний університет економіки і торгівлі імені Михайла Туган-Барановського (м. Кривий Ріг, Україна), e-mail: bondarenko_oo@donnuet.edu.ua

ДОСЛІДЖЕННЯ ТЕХНОЛОГІЇ ВИЛУЧЕННЯ ЯБЛУК ПЕВНОГО РІЗНОВИДУ В УМОВАХ ХАРЧОВОЇ ПРОМИСЛОВОСТІ ІЗ ЗАСТОСУВАННЯМ МЕТОДІВ КЛАСТЕРИЗАЦІЇ

UDC 621.325.5

***Tsvirkun L. A.**, PhD in Pedagogical sciences* [1]
***Tsvirkun S. L.**, PhD in Technical sciences* [2]
***Bondarenko O. O.**, PhD in Economic sciences* [3]

[1] Donetsk National University of Economics and Trade named after Mykhailo Tugan-Baranovsky, Kryvyi Rih, Ukraine, e-mail: cvirkun@donnuet.edu.ua

[2] Kryvyi Rih National University, Kryvyi Rih, Ukraine, e-mail: tserg30@ukr.net

[3] Donetsk National University of Economics and Trade named after Mykhailo Tugan-Baranovsky, Kryvyi Rig, Ukraine, e-mail: bondarenko_oo@donnuet.edu.ua

THE RESEARCH OF TECHNOLOGY OF APPLE EXTRACTION OF A CERTAIN FORM IN THE CONDITIONS OF THE FOOD INDUSTRY WITH APPLICATION OF CLUSTERIZATION METHODS

***Мета**. Метою статті є дослідження технології вилучення яблук певного різновиду в умовах харчової промисловості.*

***Методи**. У роботі для розбиття яблук на певні різновиди, наведені декількома видами, застосовані методи чіткої і нечіткої кластеризації.*

***Результати**. Зазначено, що на попередньому етапі сортування відбувається вилучення маленьких яблук, битих, неякісних, а також із дефектами; за розміром, різновидом, кольором яблука сортують вже згодом на виробництві. Акцентовано увагу на тому, що належність яблук до певного різновиду може бути визначена за кількома параметрами, а саме розмір (d), вага (m), колір (g), тому для виконання цієї операції доцільно використовувати операцію кластеризації. Запропоновано для розбиття яблук на певні різновиди, наведені декількома сортами, застосовувати методи чіткої і нечіткої кластеризації. Аналіз досліджень показав, що чітка кластеризація характеристик яблук X означає розбиття даних на певне число взаємовиключних підмножин з подібними характеристиками, які притаманні певному різновиду. При цьому вважають, що кількість різновидів відома. За допомогою класичної теорії множин чітка кластеризація визначається як сімейство підмножин, які відповідають умовам: всі об'єкти розподілені за кластерами, кожен об'єкт належить тільки одному кластеру, жоден з кластерів не порожній. Було встановлено, що з-поміж методів чіткої кластеризації для вирішення задачі розпізнавання яблук за сортами найбільш ефективними є методи K-means та K-medoid, які визначають належність кожного набору характеристик яблук для одного із кластерів. Інтерпретовано результати нечіткої кластеризації яблук за алгоритмами (Fuzzy C-means, Густафсона-Кесселя, Гаса-Гева). Наведено результати нечіткого розбиття у вигляді тривимірного графіка. Запропоновано для розбиття яблук на певні різновиди, наведені*

Надійшла до редакції 09.09.2020 р.

декількома сортами, схожість показників яких дозволить вилучити певний різновид яблук із загального потоку, метод, що дозволяє здійснювати розпізнавання певних сортів яблук шляхом нечіткої кластеризації його характеристик з використанням алгоритму Густафсона-Кесселя і подальшої апроксимації гауссовими функціями належності проекцій результатів кластеризації на безліч вхідних даних.

***Ключові слова**: технологія, сортування, характеристики яблок, різновиди яблук, харчова промисловість, яблука.*

Постановка проблеми. Яблука — невід'ємна частина українського експорту. За кордоном вони користуються величезним попитом і щороку завойовують нові ринки. Щоб отримати яблука вищого ґатунку, необхідно дотримуватися правильної технології зберігання, сортування та пакування. На попередньому етапі сортування відбувається вилучення маленьких яблук, битих, неякісних, а також із дефектами. Однак за розміром, різновидом, кольором яблука сортують вже згодом на виробництві. Відповідно від того, наскільки якісно яблука будуть відсортовані, залежить подальший експорт продукції.

Аналіз останніх досліджень і публікацій. Оскільки належність яблук до певного різновиду може бути визначена за кількома параметрами, а саме розмір (d), вага (m), колір (g), то для виконання цієї операції доцільно використовувати операцію кластеризації (рис. 1). Більшість алгоритмів кластеризації не спираються на традиційні для статистичних методів допущення, вони можуть використовуватися в умовах майже повної відсутності інформації щодо законів розподілу даних [1; 10].

Рисунок 1 — Різновиди яблук

Вихідною інформацією для кластеризації є матриця вимірювання непрямих ознак певних різновидів яблук, що складається з M рядків, кожна з яких містить значення ознак:

$$\mathbf{x} = \begin{vmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{21} & \dots & x_{2n} \\ \dots & \dots & \dots & \dots \\ x_{M1} & x_{M2} & \dots & x_{Mn} \end{vmatrix}, \quad (1)$$

де n — кількість ознак; M — кількість різновидів яблук.

У нашому випадку завданням є розбиття яблук за різновидом, наведеним декількома сортами, на кілька кластерів, схожість показників яких дозволить вилучити певний різновид яблук із загального потоку.

Для підвищення ефективності процесу кластеризації була виконана нормалізація вхідних даних [2; 3; 10]:

$$\overline{x} = \frac{x - x_{\min}}{x_{\max} - x_{\min}}, \quad (2)$$

де x — поточне значення характеристики яблук; $[x_{\min}, x_{\max}]$ — діапазон значень характеристики яблук.

Результати нормалізації пари характеристик яблук: розміру d і колірного тону g (у відтінках сірого) після виключення з вибірки наведено на рис. 2.

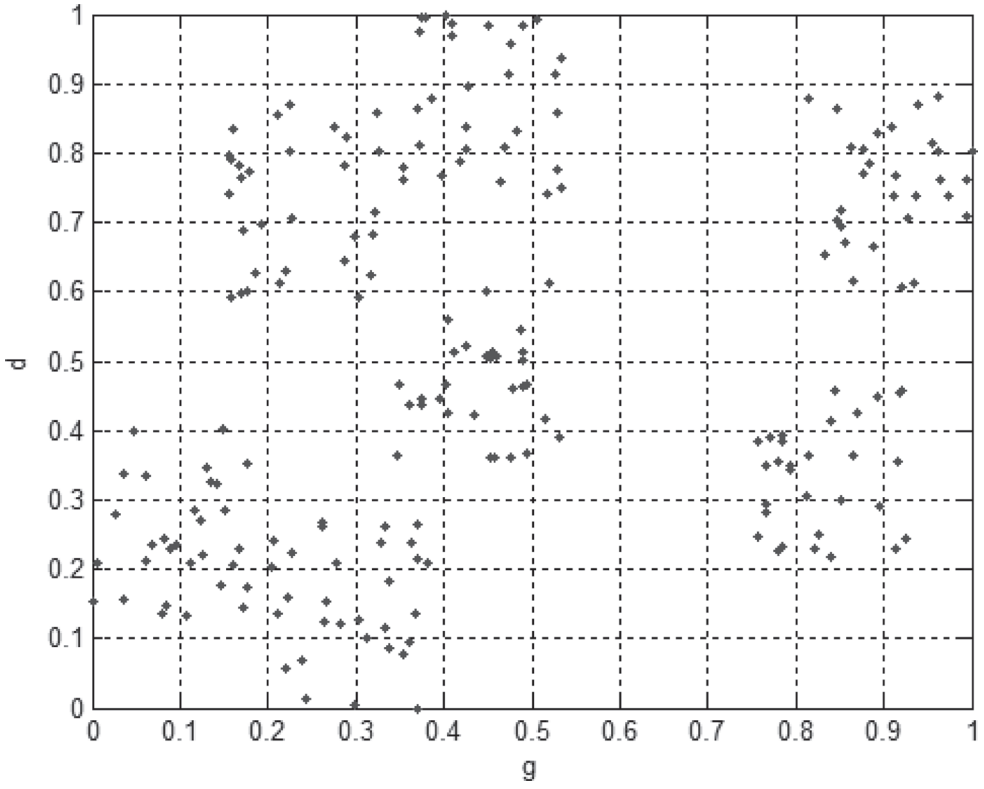


Рисунок 2 — Результати нормалізації характеристик яблук

Для здійснення кластеризації характеристик яблук було розглянуто методи чіткої і нечіткої кластеризації. Чітка кластеризація характеристик яблук X означає розбиття даних на певне число взаємовиключних підмножин з подібними характеристиками, які притаманні певному різновиду. При цьому вважають, що кількість різновидів відома. За допомогою класичної теорії множин чітка кластеризація визначається як сімейство підмножин $\{A_i | 1 \le i \le c \subset P(X)\}$, які відповідають умовам: всі об'єкти розподілені за кластерами, кожен об'єкт належить тільки одному кластеру, жоден з кластерів не порожній [1; 4].

Серед методів чіткої кластеризації для вирішення задачі розпізнавання яблук за сортами найбільш ефективними є методи K-means та K-medoid, які визначають належність кожного набору характеристик яблук для одного із кластерів, щоб мінімізувати в межах кластера суму квадратів [1, 4].

$$\sum_{i=1}^{c} \sum_{k \in A_i} \left\| x_k - v_i \right\|^2, \tag{3}$$

де A_i — набір об'єктів (опорних точок) в i-й кластер; v_i — середнє, на що вказують точки кластера i.

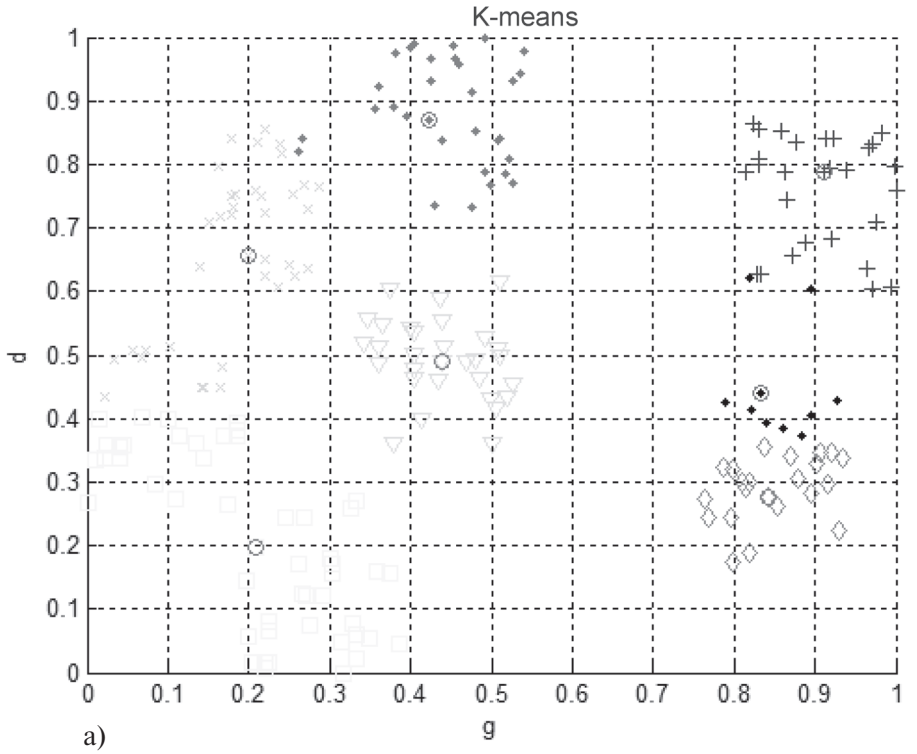
Відповідно до алгоритму K-means кластеризації v_i називається центрами кластерів:

$$v_i = \frac{\sum_{k=1}^{N_i} x_k}{N_i}, \quad x_k \in A_i, \tag{4}$$

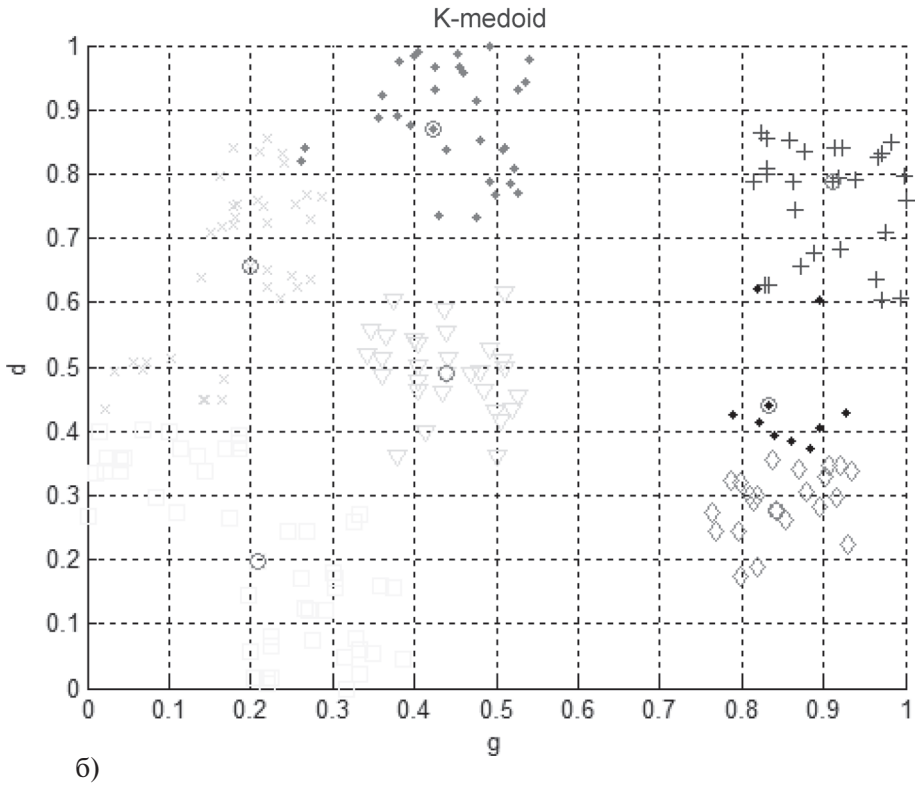
де N_i — кількість об'єктів у A_i.

Мета статті — дослідження технології вилучення яблук певного різновиду в умовах харчової промисловості

Виклад основного матеріалу дослідження. Результат кластеризації яблук, виконаної за алгоритмом K-means, наведено на рис. 3(а). Згідно з методом K-medoids кластеризація центрами кластерів — найближчі об'єкти середніх даних в одному кластері $V=\{V_i \in X 1 \le i \le c\}$ [1; 4]. Результат кластеризації яблук, виконаної за алгоритмом K-medoids, наведено на рис. 3(б).



а)



б)

Рисунок 3 — Результат кластеризації яблук: а) алгоритм K-means; б) алгоритм K-medoids

У реальній ситуації поділ яблук на різновиди недоцільно представляти двома ступенями приналежності 0 або 1, як це відбувається при чіткій кластеризації. Більш природним є використання часткової належності в діапазоні від 0 до 1, що дозволить певним яблукам, характеристики яких знаходяться в межах між декількома кластерами, належати їм з різним ступенем.

Нечіткі кластери описують матрицею нечіткого розбиття [1]:

$$F = \left[\mu_{ki}\right], \quad \mu_{ki} \in [0,1], \quad k = \overline{1,M}, \quad i = \overline{1,c}, \tag{5}$$

де рядок з номером k містить ступені приналежності об'єкта $(x_{k1}, x_{k1}, ..., x_{kn})$ до відповідних кластерів A_1, A_2, ..., A_c.

Алгоритм кластеризації Fuzzy C-means, що використовувався для кластеризації характеристик яблук, заснований на мінімізації функціонала C-means [1; 4]

$$J(X,U,V) = \sum_{i=1}^{c}\sum_{k=1}^{N} \mu_{ik}{}^{m} \left\| x_k - v_i \right\|_A^2, \tag{6}$$

де

$$V = \left[v_1, v_2, ..., v_c \right], \quad v_i \in R^n, \tag{7}$$

вектор центрів кластерів

$$D_{ikA}^2 = \left\| x_k - v_i \right\|_A^2 = (x_k - v_i)^T A_i (x_k - v_i). \tag{8}$$

Результат кластеризації яблук, виконаної за алгоритмом Fuzzy C-means ((рис. 4(а)), показав правильне, порівняно з еталоном визначення центрів кластерів, що дозволяє зробити висновок про перспективність використання цього методу.

Також для класифікації характеристик яблук був використаний алгоритм Густафсона-Кесселя (Gustafson-Kessel), який удосконалює алгоритм Fuzzy C-means, використовуючи адаптивну норму відстані для кожного кластера [5; 6]

$$D_{ikA}^2 = (x_k - v_i)^T A_i (x_k - v_i), \quad 1 \le i \le c, \quad 1 \le k \le N. \tag{9}$$

При цьому матриці Ai використовують в методі C-means для оптимізаційних змінних. Цільова функція алгоритму визначається формулою:

$$J(X,U,V) = \sum_{i=1}^{c}\sum_{k=1}^{N} \mu_{ik}{}^{2} D_{ikA_i}^2. \tag{10}$$

Результат кластеризації яблук, виконаної за алгоритмом Густафсона-Кесселя, наведено на рис. 4(б). Результат кластеризації яблук, виконаний за алгоритмом Гаса-Ґева наведений на рис. 3(в).

Якість алгоритмів кластеризації як ітераційних процесів оцінювалась за швидкістю досягнення оптимуму із заданою точністю за кінцеве число кроків. Збіжність алгоритмів кластеризації характеристик яблук перевірялася при розбитті на близьку до оптимальної кількість кластерів: 5–9, що дозволило оцінити не тільки власне показники збіжності, а й додатково перевірити доцільність обраного раніше вибору оптимальної кількості кластерів.

Найкращі в середньому показники мають методи Густафсона-Кесселя і Гаса-Ґева. Однак недоліком алгоритму Гаса-Ґева є необхідність попередньої обробки вихідних даних шляхом кластеризації з використанням, наприклад, методу нечіткої кластеризації FCM. Таким чином, найбільш доцільним видається використання методу нечіткої кластеризації Густафсона-Кесселя [7; 8; 9].

Наведемо результати нечіткого розбиття у вигляді тривимірного графіка. Для кожного зразка яблук відкладемо по осях абсцис і ординат значення характеристик, а по осі аплікат — ступінь приналежності непевному кластеру.

Тривимірні зображення нечітких кластерів відповідних зразків досліджуваних різновидів яблук наведені на рис. 5.

Таким чином, розроблений метод дозволяє здійснювати розпізнавання певних різновидів яблук шляхом нечіткої кластеризації його характеристик з використанням алгоритму Густафсона-Кесселя і подальшої апроксимації гауссовими функціями належності проекцій результатів кластеризації на безліч вхідних даних.

Висновки. Отже, задля розбиття яблук на певні різновиди, наведені декількома сортами, схожість показників яких дозволить вилучити певний різновид яблук із загального потоку, було розроблено метод, що дозволяє здійснювати розпізнавання певних сортів

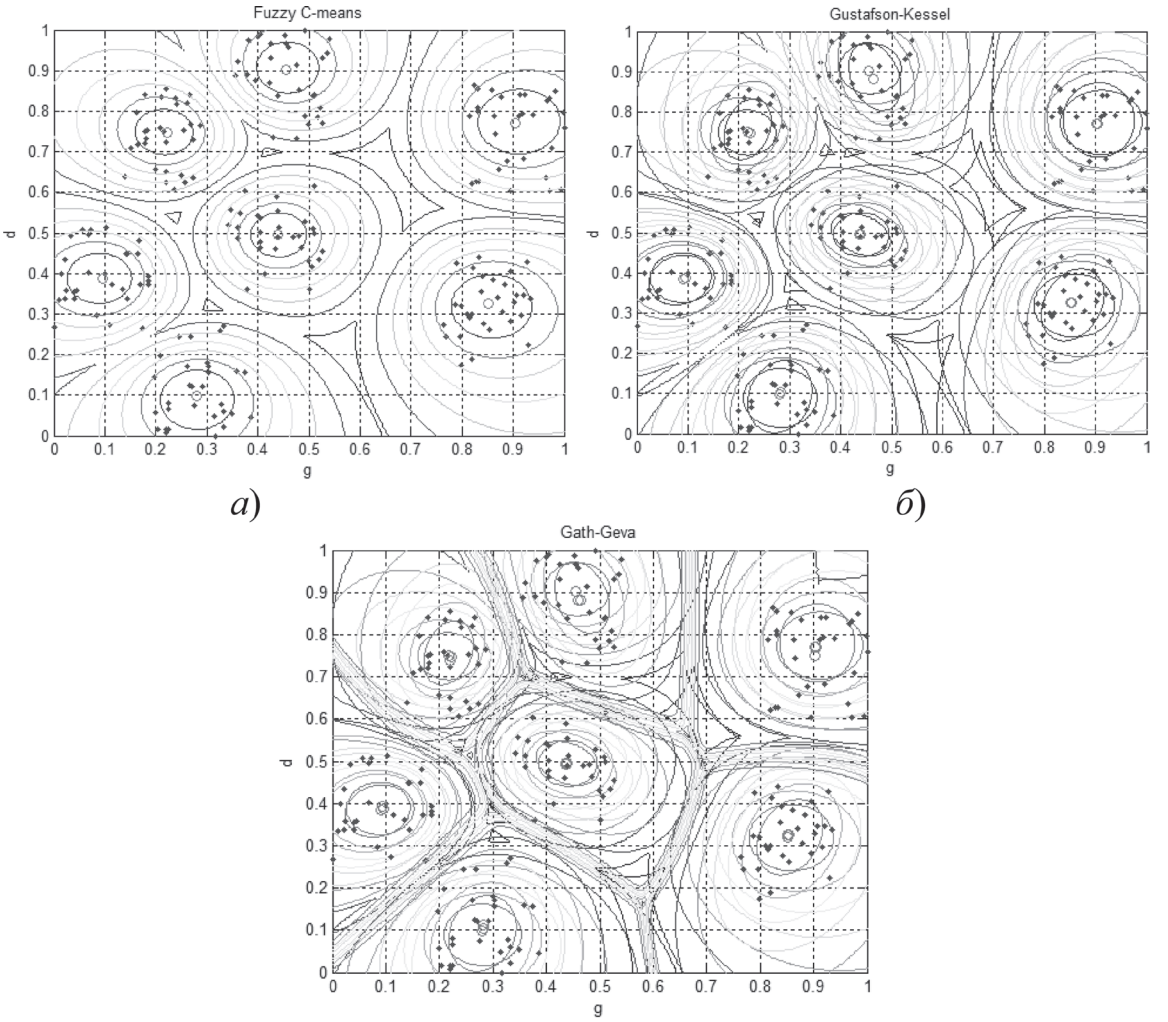


Рисунок 4 — Результат нечіткої кластеризації яблук
а) Fuzzy C-means; б) Густафсона-Кесселя; в) Гаса-Гева

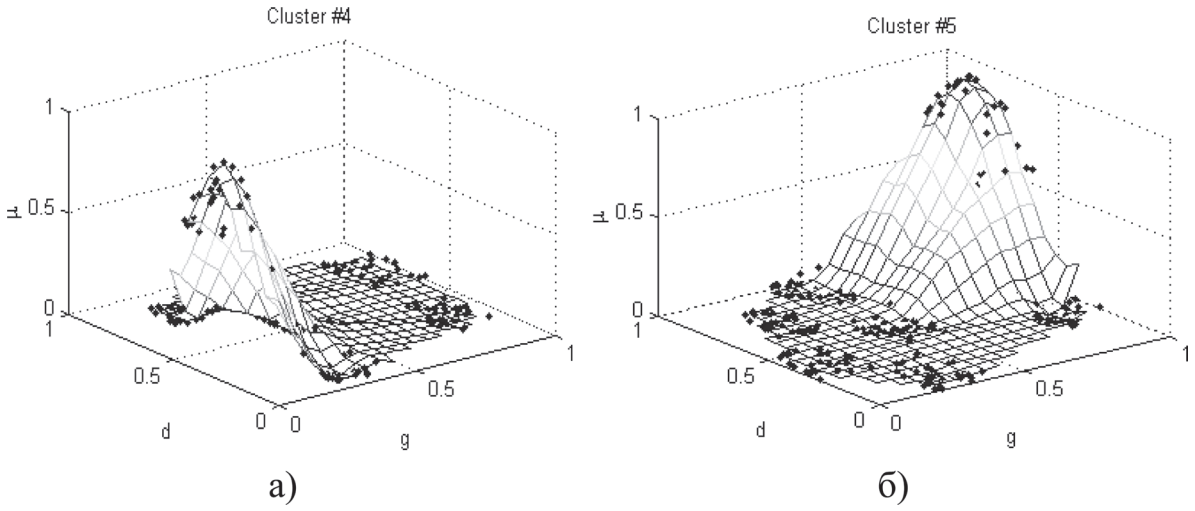


Рисунок 5 — Функції приналежності характеристик: а — кластер № 4, б — кластер №5

яблук шляхом нечіткої кластеризації його характеристик з використанням алгоритму Густафсона-Кесселя і подальшої апроксимації гауссовими функціями належності проекцій результатів кластеризації на безліч вхідних даних. Подальшим кроком є апробація запропонованого методу.

Список літератури

1. Штовба С. Д. Введение в теорию нечетких множеств и нечеткую логику. URL : //http: matlab/exponenta.ru/fuzzylogic/book1.

2. Balasko B., Abonyi J., Feil B. Fuzzy clustering and data analysis toolbox. *Veszprem: Department of Process Engineering University of Veszprem*, 2006. 74 p.

3. Нечеткие множества в моделях управления и искусственного интеллекта / под. ред. Д. А. Поспелова. М. : Наука, 2006. 312 с.

4. Bezdek J. C. Pattern recognition with fuzzy objective function algorithms. Plenum Press. 2016. Vol. 4. P. 1–14.

5. Gustafson D. E., Kessel W. C. Fuzzy clustering with a fuzzy covariance matrix. IEEE Conference on Decision and Control including the 17th Symposium on Adaptive Processes. San Diego, CA, USA. 2008. Vol. 7. P. 773–781.

6. Gath I., Geva A. Unsupervised optimal fuzzy clastering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2005. Vol. 7. P. 773–781.

7. Lim Jae S. Two-Dimensional Signal and Image Processing. Englewood Cliffs. New Jersey: Prentice Hall, 2007.p. 694.

8. Rosten E., Drummond T. Fusing points and lines for high performance tracking. *10th IEEE International Conference on Computer Vision. Beijing, China*. 2010. Vol. 2. P. 1508–1515.

9. Bay H., Tuytelaars T., Van Gool L. SURF: Speeded up robust features. *Proceedings of the European Conference on Computer Vision*. San Diego, CA, USA. 2006. Vol. 4. P. 404–407.

10. Balasko B., Abonyi J., Feil B. Fuzzy Clustering and Data Analysis Toolbox. *University of Pannonia*, no. 12, pp. 274–286.

References

1. Shtovba, S. D. (2010). *Vvedeniye v teoriyu nechetkikh mnozhestv i nechetkuyu logiku* [Introduction into the theory of fuzzy sets and fuzzy logic]. Access mode : http://matlab/exponenta.ru/fuzzylogic/book1.

2. Balasko, B., Abonyi, J., Feil, B. (2006). *Nabor instrumentov dlya nechetkoy klasterizatsii i analiza dannykh* [Fuzzy clustering and data analysis toolbox]. *Veszprem: Department of Process Engineering University of Veszprem*, 74 p.

3. Pospelova, D. A. (2006). *Nechetkiye mnozhestva v modelyakh upravleniya i iskusstvennogo intellekta [*Fuzzy sets in management models and artificial intelligence]. Moscow, Science, 312 p.

4. Bezdek, J. C. (2016). *Raspoznavaniye obrazov s pomoshch'yu algoritmov nechetkoy tselevoy funktsii* [Pattern recognition with fuzzy objective function algorithms]. *Plenum Press* [Plenum Press], vol. 4., pp.1–14.

5. Gustafson, D. E., Kessel, W. C. (2008). *Nechetkaya klasterizatsiya s nechetkoy kovariatsionnoy matritsey* [Fuzzy clustering with a fuzzy covariance]. *Konferentsiya IEEE po prinyatiyu resheniy i kontrolyu, vklyuchaya 17-y simpozium po adaptivnym protsessam* [IEEE decision making and control conference, including 17th adaptive processes symposium]. San Diego, CA, USA, vol. 7.pp. 773–781.

6. Gath, I., Geva, A. (2005). *Optimal'naya nechetkaya klasterizatsiya bez uchitelya* [Unsupervised optimal fuzzy clastering]. *IEEE Transactions po analizu shablonov i mashinnomu analizu* [IEEE Transactions on Pattern Analysis and Machine Intelligence], vol. 7, pp. 773–781.

7. Lim Jae, S. (2007). *Dvumernaya obrabotka signalov i izobrazheniy* [Two-Dimensional Signal and Image Processing. Englewood Cliffs]. New Jersey, Prentice Hall, 2007. 694 p.

8. Rosten, E., Drummond, T. (2010). *Tochki soyedineniya i linii dlya vysokoproizvoditel'nogo otslezhivaniya* [Fusing points and lines for high performance tracking]. *Mezhdunarodnaya konferentsiya IEEE po komp'yuternomu zreniyu* [IEEE International Conference on Computer Vision]. Beijing, China, vol. 2, pp. 1508–1515.

9. Bay, H., Tuytelaars, T., Van Gool, L. (2006). *SURF: Uskorennyye nadezhnyye funktsii* [SURF: Speeded up robust features]. *Trudy Yevropeyskoy konferentsii po komp'yuternomu zreniyu* [Proceedings of the European Conference on Computer Vision]. San Diego, CA, USA, vol. 4, pp. 404–407.

10. Balasko, B., Abonyi, J., Feil, B. (2005). Fuzzy Clustering and Data Analysis Toolbox. University of Pannonia, no. 12, pp. 274–286.

Objective. *The aim of the article is to study the technology of extracting apples of a certain species in the food industry.*

Methods. *In the work for the division of apples into certain varieties, represented by several varieties, the methods of clear and fuzzy clustering are used.*

Results. *It is noted that at the preliminary stage of sorting there is a removal of small apples, broken, low-quality, as well as with defects; by size, variety, color apples are sorted later in production. Emphasis is placed on the fact that the affiliation of apples of a certain species can be determined by several parameters, namely size (d), weight (m), color (g), so it is advisable to use the clustering operation to perform this operation. It is proposed to use clear and fuzzy clustering methods to divide apples into certain varieties represented by several varieties. Analysis of research has shown that a clear clustering of the characteristics of apples X means the division of data into a certain number of mutually exclusive subsets with similar characteristics that are inherent in a particular variety. In this case, it is believed that the number of species is known. Using classical set theory, a clear clustering is defined as a family of subsets corresponding to the conditions: all objects are distributed in clusters, each object belongs to only one cluster none of the clusters is empty. It was found that among the methods of clear clustering to solve the problem of recognizing apples by variety, the most effective methods are K-means and K-medoid, which determine the affiliation of each set of characteristics of apples for one of the clusters. The results of fuzzy clustering of apples by algorithms (Fuzzy C-means, Gustafson-Kessel, Gus Giva) are interpreted. The results of fuzzy partitioning in the form of a three-dimensional graph are presented. It is proposed to divide apples into certain varieties represented by several varieties, the similarity of which will exclude a certain type of apples from the general flow, a method that allows the recognition of certain varieties of apples by fuzzy clustering of its characteristics using Gustafson-Kessel algorithm and subsequent approximation clustering to multiple inputs.*

Key words: *technology, sorting, characteristics of apples, varieties of apples, food industry, apples.*