

The background is a dark blue gradient with a perspective view of a digital landscape. It features a grid of glowing white lines and dots, resembling a network or data flow. Scattered throughout are various sizes of white binary digits (0s and 1s), some appearing to float in the air. The overall aesthetic is futuristic and technological.

ЕЛЕКТРОННІ ІНФОРМАЦІЙНІ РЕСУРСИ: СТВОРЕННЯ, ВИКОРИСТАННЯ, ДОСТУП

ЗБІРНИК МАТЕРІАЛІВ

Міжнародної науково-практичної Інтернет-конференції

9-10 листопада 2020 р.

Міністерство освіти і науки України

Вінницький національний технічний університет

Національна академія Державної прикордонної служби України

ім. Богдана Хмельницького

Вінницький національний медичний університет ім. М.І. Пирогова

Комунальний заклад вищої освіти «Вінницька академія безперервної освіти»

Комунальний заклад «Сумський обласний інститут післядипломної педагогічної освіти»

Люблінська політехніка (Польща)

Новий університет Лісабону (Португалія)

«ЕЛЕКТРОННІ ІНФОРМАЦІЙНІ РЕСУРСИ: СТВОРЕННЯ, ВИКОРИСТАННЯ, ДОСТУП»

ЗБІРНИК МАТЕРІАЛІВ

Міжнародної науково-практичної Інтернет-конференції

9-10 листопада 2020 р.

**Суми/Вінниця
НІКО/ВНТУ
2020**

УДК 004
ББК 32.97
Е50

Рекомендовано до видання Вченою радою Вінницького національного технічного університету Міністерства освіти і науки України (протокол № 5 від 26.11.2020 р.)

Електронні інформаційні ресурси: створення, використання, доступ:
Збірник матеріалів Міжнародної науково-практичної Інтернет конференції 9-10 листопада 2020 р. – Суми/Вінниця : НІКО/ВНТУ, 2020. – 280 с.

ISBN 978-617-7422-13-5

Збірник містить матеріали Міжнародної науково-практичної Інтернет-конференції «Електронні інформаційні ресурси: створення, використання, доступ».

Матеріали збірника подано у авторській редакції. Автори опублікованих матеріалів несуть повну відповідальність за підбір, точність наведених фактів, цитат, статистичних даних, власних імен та інших відомостей, Матеріали відтворюються зі збереженням змісту, орфографії та синтаксису текстів, наданих авторами.

УДК 004
ISBN 978-617-7422-13-5

© Вінницький національний технічний університет, 2020

© Вид-во Суми, НІКО, 2020

Гандрибіда В.О.

МЕТОД ПОШУКУ ГЕОГРАФІЧНИХ АРТЕФАКІВ НА ОСНОВІ ЛОКАЦІЇ КОРИСТУВАЧА..... 70

Гертель І.С., Майданюк В.П., Чернишов К.А.

ФОРМУВАННЯ ПРОСТОРОВИХ ЧАСТОТНИХ КОМПОНЕНТ В ПІРАМІДАЛЬНИХ СХЕМАХ УЩІЛЬНЕННЯ ЗОБРАЖЕНЬ..... 73

Гибало В.В.

РОЗРОБКА МЕТОДІВ І ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ ДЛЯ БАЛАНСУВАННЯ ОБЧИСЛЮВАННОГО НАВАНТАЖЕННЯ НА ОСНОВІ ТЕХНОЛОГІЇ WEBSOCKET..... 76

Горбачова Т.В., Каражекова П.І.

ЕЛЕКТРОННІ РЕСУРСИ GOOGLE ДЛЯ ОРГАНІЗАЦІЇ ДИСТАНЦІЙНОГО ВИВЧЕННЯ МАТЕРІАЛУ..... 78

Денисюк А. В., Озерчук Д.А., Романюк О.В., Романюк О.Н.

ПОРІВНЯННЯ ЗОБРАЖЕНЬ ІЗ ВИКОРИСТАННЯМ ПЕРЦЕПТИВНИХ ХЕШ-МЕТОДІВ..... 80

Зубко А.В., Майданюк В.П.

РОЗРОБКА ПРОГРАМНОГО ЗАСОБУ ДЛЯ ІМІТАЦІЇ ДІЙ ГРАВЦЯ В RTS-ІГРАХ..... 84

Іванов Д.С.

МЕТОДИ РОЗПІЗНАВАННЯ ЧАСТИН ЛЮДСЬКОГО ТІЛА..... 86

Іваха О.А., Романюк О.В., Романюк О.Н.

ГРАФІЧНІ ПРОЦЕСОРИ У ВИРІШЕННІ СУЧАСНИХ ІТ-ЗАДАЧ..... 90

Івацко Т.С.

ПРАКТИЧНЕ ЗАСТОСУВАННЯ КОМУНІКАТИВНОЇ ТА УПРАВЛІНСЬКОЇ ДІЯЛЬНОСТІ У НАВЧАЛЬНО - ВИХОВНОМУ ПРОЦЕСІ ДЛЯ ЗДОБУВАЧІВ ОСВІТИ..... 93

Катєльніков Д.І.

ПРОСТИЙ АЛГОРИТМ КЛАСТЕРИЗАЦІЇ НА ОСНОВІ ТЕОРЕМИ БАЄСА..... 104

**ЕЛЕКТРОННІ ІНФОРМАЦІЙНІ РЕСУРСИ:
СТВОРЕННЯ, ВИКОРИСТАННЯ, ДОСТУП:**

Збірник матеріалів

Міжнародної науково-практичної Інтернет-конференції
9-10 листопада 2020 р.

Редактор С.А.Пойда, Н.А. Ніколаєнко
Комп'ютерне верстання С.А.Пойда, М.С. Ніколаєнко

Підписано до друку 01.11.2020 Гарнітура Times New Roman
Формат 60x84/16 Папір офсетний
Друк цифровий Ум. друк. арк. 16,3
Тираж 300 пр. Зам. № 2/20

Видавництво НІКО
м.Суми, вул.Харківська, 54
Свідоцтво про внесення до Державного реєстру
суб'єктів видавничої справи України
серія СМв № 044
від 15.10.2012
E-mail: ms.niko@i.ua
Телефон для замовлень: +38(066) 270-64-68

7. Національна бібліотека України ім. В.І. Вернадського Національні університети України. [Електронний ресурс]. – Режим доступу: http://www.nbuv.gov.ua/portal/ukr_uni.html - 2013 р. 14

8. Основні показники діяльності вищих навчальних закладів на початок 2011/12 навчального року / Статистичний бюлетень. – Державний статистична служба України, 2012. – 219 с. [Електронний ресурс]. – Режим доступу: <http://www.ukrstat.gov.ua/> - 2013 р.

9. Сухарева, Т. Н. Формирование коммуникативных стратегий словоупотребления у студентов неязыковых вузов : дисс. ... канд. пед. наук / Т. Н. Сухарева. – Тамбов : ТГУ им. Г. Державина, 2002. – 160 с.

*Кательніков Денис Іванович,
кандидат технічних наук,
доцент кафедри програмного забезпечення,
Вінницький національний технічний університет*

ПРОСТИЙ АЛГОРИТМ КЛАСТЕРИЗАЦІЇ НА ОСНОВІ ТЕОРЕМИ БАЄСА

Для випадків, коли потрібно нашвидкоруч оцінити межі кластерів використання громіздких алгоритмів кластеризації є недоцільним. В роботі запропоновано простий але ефективний алгоритм кластеризації який ґрунтується на теоремі Баєса.

Кластеризація даних передбачає розбиття їх на окремі групи за певною мірою подібності. Цей процес є одним з ключовим при автоматичній обробці первинних даних у системах штучного інтелекту та подальшій побудові систем класифікації та розпізнавання [1]. Тому постійно розроблюються все більш нові і досконалі методи: алгоритм кластеризації k-середніх [1], алгоритм мінімального кістякового дерева [2], метод k-найближчих сусідів [3], алгоритм нечіткої кластеризації [4], алгоритм “імітації відпалу” (simulated annealing algorithm) [5], ієрархічні алгоритми кластеризації [6], нейронні мережі для кластеризації [7], генетичні алгоритми для кластеризації [8]. Кожен з цих методів має свої переваги і свою область застосування. Але об’єднує їх те, що всі вони складні і у разі коли необхідне швидке прийняти рішення їх використання є недоцільним. В таких випадках слід звернути увагу на простий,

відомий ще з XIX сторіччя метод Баєса:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}, \quad (1)$$

де $P(A)$ - апіорна ймовірність гіпотези A (Рисунок 1);

$P(B|A)$ - умовна ймовірність спостереження події B за умови істинності A ;

$P(B)$ - повна апіорна ймовірність спостереження події B ;

$P(A|B)$ – апостеріорна ймовірність гіпотези A за умови спостереження події B .

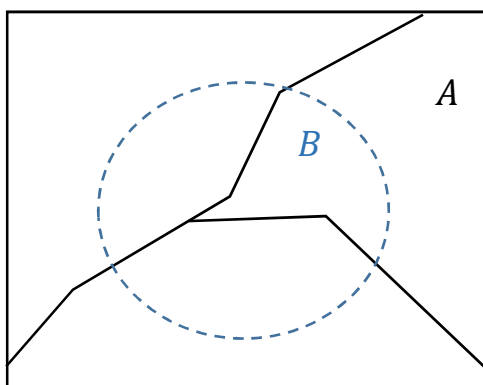


Рисунок 1 – Теорема Баєса

Ця формула дозволяє апостеріорі робити оцінки ймовірності певних подій при умові, що пов'язані з цими подіями умови виконані. Виявляється, що подібний підхід можна застосувати і у випадку коли ймовірності задані у вигляді функції розподілу ймовірностей. Наприклад, для одновимірного випадку такою:

$$p_{A_j}(x_i) = \frac{1}{\sigma_j \cdot \sqrt{2\pi}} \cdot \exp\left(-\frac{(x_i - \mu_j)^2}{2\sigma_j^2}\right) = N(x_i, \mu_j, \sigma_j), \quad (2)$$

де x_i – результат вимірювання ;

μ_j – математичне очікування случайних величин гіпотези A_j ;

σ_j – стандартне відхилення случайних величин гіпотези A_j .

Якщо виміри x_i є не просто числами, а векторами, то буде йти мова про багатовимірний нормальний розподіл (рисунок 2):

$$p_{A_j}(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{n}{2}} |\Sigma|^{\frac{1}{2}}} \cdot \exp\left(-\frac{1}{2}(\mathbf{x} - \mu)^T \Sigma^{-1}(\mathbf{x} - \mu)\right) = N(\mathbf{x}, \mu, \Sigma), \quad (3)$$

де $\mathbf{x} = (x_1, \dots, x_n)$ – результат вимірювання багатовимірної величини;

μ – вектор середніх значень \mathbf{x} ;

Σ - коваріаційна матриця, $|\Sigma|$ - її визначник, Σ^{-1} – обернена до неї матриця;

T – операція транспонування.

На рисунку 2 показано приклад двовимірного нормального розподілу при значеннях $\mu = [0.0 \quad 5.0]$ та $\Sigma = \begin{bmatrix} 5.0 & 2.5 \\ 2.5 & 5.0 \end{bmatrix}$.

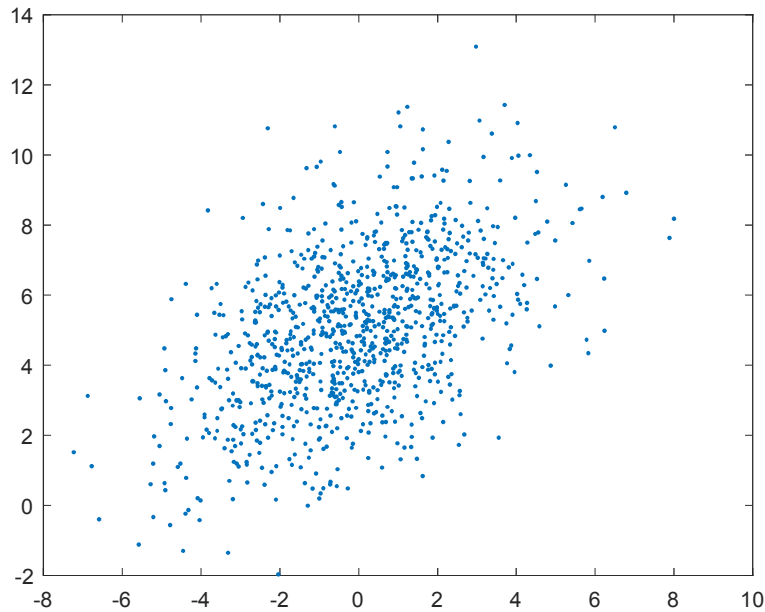


Рисунок 2 - Двовимірний нормальний розподіл

Якщо дано суміш двох подібних випадкових величин (Рисунок 3):

$$p(\mathbf{x}) = w_1 \cdot N(\mathbf{x}, \mu_1, \Sigma_1) + w_2 \cdot N(\mathbf{x}, \mu_2, \Sigma_2), \quad (4)$$

де w_1 – доля вимірів першої випадкової величини, w_2 – доля вимірів другої випадкової величини, $w_1 + w_2 = 1$, то виникає найпростіша задача побудови моделі кластеризації, яка віднесе виміри або до першої, або до другої величини.

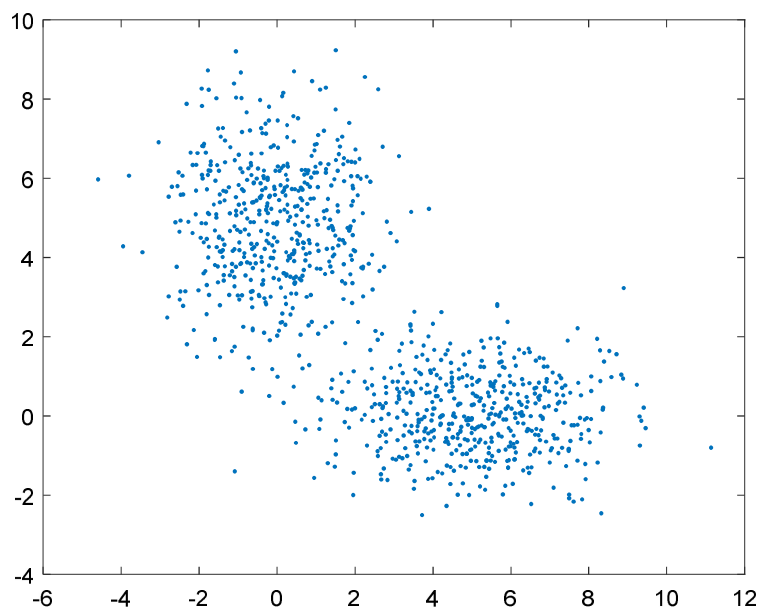


Рисунок 3 - Суміш двох випадкових величин

Використання формули Баєса (1) при використанні багатовимірної функції розподілу ймовірності (3) дозволяє сформулювати наступний простий алгоритм кластеризації.

Введемо позначення:

m - загальна кількість вимірів, m_1 - кількість вимірів, яким було присвоєно мітку першої випадкової величини, m_2 - кількість вимірів, яким було присвоєно мітку другої випадкової величини, $m_1 + m_2 = m$;

\mathbf{x}_j - j -й вимір у суміші, $j \in \{1, 2, \dots, m\}$;

q_j - мітка випадкової величини -го виміру у суміші, $q_j \in \{1, 2\}$.

Крок 1. Присвоюємо початкові оцінки

$$w_1 = 0.5, w_2 = 0.5,$$

$$\mu_1 = [1 \quad 1], \Sigma_1 = \begin{bmatrix} 1.0 & 0.0 \\ 0.0 & 1.0 \end{bmatrix},$$

$$\mu_2 = [10 \quad 10], \Sigma_2 = \begin{bmatrix} 1.0 & 0.0 \\ 0.0 & 1.0 \end{bmatrix}$$

Крок 2. Для кожного j -го виміру у суміші, $j \in \{1, 2, \dots, m\}$,

$$q_j = \begin{cases} w_1 \cdot N(\mathbf{x}, \mu_1, \Sigma_1) \geq w_2 \cdot N(\mathbf{x}, \mu_2, \Sigma_2), & 1 \\ w_1 \cdot N(\mathbf{x}, \mu_1, \Sigma_1) < w_2 \cdot N(\mathbf{x}, \mu_2, \Sigma_2), & 2 \end{cases}$$

Крок 3.

Використовуючи q_j порахуємо m_1 та m_2 - оцінки кількості вимірів першої та другої випадкової величини. Це дозволяє порахувати:

$$w_1 = \frac{m_1}{m} \text{ та } w_2 = \frac{m_2}{m},$$

$$\mu_1 = \text{MEAN}(\{\mathbf{x}_k | q_k = 1\}),$$

$$\mu_2 = \text{MEAN}(\{\mathbf{x}_k | q_k = 2\}),$$

$$\Sigma_1 = \text{STD}(\{\mathbf{x}_k | q_k = 1\}),$$

$$\Sigma_2 = \text{STD}(\{\mathbf{x}_k | q_k = 2\}),$$

де MEAN - функція, яка знаходить вектор середніх значень кожної компоненти;

STD - функція, яка повертає коваріаційну матрицю.

Крок 4. Якщо зміна μ_1 та μ_2 більша за деяке мале $\varepsilon \ll 1$, то зупиняємо виконання алгоритму і поточні значення q_j визначають належність до кластерів. Інакше переходим до Кроку 2.

Результат використання запропонованого алгоритму для суміші з рисунка 3 показано на рисунку 4.

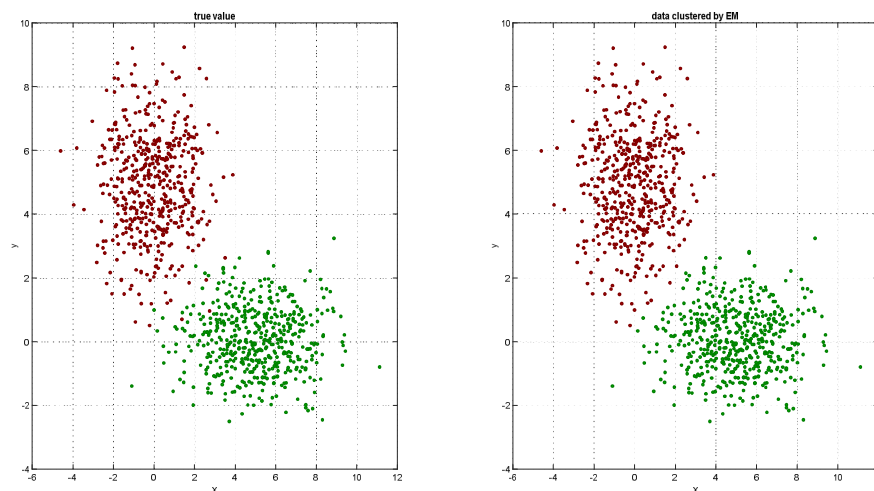


Рисунок 4 - Результат запропонованого алгоритму.
Ліворуч істинні значення, праворуч результат кластеризації.

Висновок. Запропонований в роботі простий але ефективний алгоритм кластеризації, який ґрунтується на використанні теореми Баєса, дозволяє ефективно вирішувати поставлену задачу у випадках, коли використання громіздких повнофункціональних алгоритмів кластеризації є недоцільним.

Список використаної літератури

1. Trevor Hastie, Robert Tibshirani, Jerome Friedman. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Second Edition. Springer series in statistics. New-York:Springer, 2009.
2. Joseph. B. Kruskal. On the Shortest Spanning Subtree of a Graph and the Traveling Salesman Problem. // *Proc. AMS*. 1956. Vol 7, No. 1. C. 48-50
3. Altman, Naomi S. (1992). "An introduction to kernel and nearest-neighbor nonparametric regression". *The American Statistician*. 46 (3): 175–185.
4. Dunn, J. C. (1973-01-01). "A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters". *Journal of Cybernetics*. 3 (3): 32–57.
5. Stuart Russell, Peter Norvig.. Artificial Intelligence: A Modern Approach 3rd ed. London:Pearson, 2009. 1152 p.
6. Frank Nielsen. "Chapter 8: Hierarchical Clustering". Introduction to HPC with MPI for Data Science. New-York:Springer, 2016.
7. K.-L.Du. Clustering: A neural network approach. *Neural Networks*. Volume 23, Issue 1, January 2010, Pages 89-107.
8. M.C.Cowgill, R.J.Harvey, L.T.Watson. A genetic algorithm approach to cluster analysis. *Computers & Mathematics with Applications*. Volume 37, Issue 7, April 1999, Pages 99-108.