

АЛГОРИТМИ КЛАСИФІКАЦІЇ БІОМЕДИЧНИХ ЗОБРАЖЕНЬ

Дериш Б.Б.¹⁾, Луцик Ю.А.²⁾

Тернопільський національний економічний університет, магістранти

I. Постановка проблеми

Розвиток інформаційних технологій за останнє десятиліття зумовив велику кількість нових методів діагностики. У дослідників з'явилися нові можливості використання візуалізації для якісної діагностики. Автоматизація процесу обробки біомедичних (гістологічних та цитологічних) зображень дозволяє істотно підвищити продуктивність і ефективність при постановці діагнозу в онкології [2]. Кінцевою метою досліджень цих зображень є їх віднесення до певного класу патологій. При цьому використовують різні класифікатори зображень. Вибір і налаштування параметрів конкретного класифікатора є не тривіальною задачею [1]. Тому актуальною проблемою є дослідження сучасних класифікаторів зображень, їх вибір і адаптація до конкретних класів зображень.

II. Мета роботи

Метою роботи є аналіз алгоритмів класифікації біомедичних зображень на основі методів k-means та Random Forest для підвищення достовірності постановки діагнозу.

III. Аналіз методів класифікації

Random forest (англ. випадковий ліс, RF) — алгоритм машинного навчання, запропонований Лео Брейманом і Адель Катлер, що полягає у використанні комітету (ансамблю) вирішальних дерев. Він застосовується для задач класифікації, регресії та кластеризації. Алгоритм поєднує в собі дві основні ідеї: метод баггінга Бреймана і метод випадкових підпросторів, запропонований Tin Kam Ho. У задачах регресії їх відповіді усереднюються, а в задачах класифікації приймається рішення голосуванням за більшістю [5]. Нехай дана навчальна вибірка D розміру n' . Генерується m нових вибірок D_i розміру n' , вибором з D випадково з поверненням. Деякі спостереження можуть потрапити до вибірки кілька разів, деякі можуть не потрапити взагалі. Якщо $n' = n$ і n велике, то частка різних спостережень в D_i буде $(1 - 1/e) \approx 63,2\%$. Далі вивчається m класифікаторів на кожній вибірці D_i . При класифікації нової точки, ці класифікатори голосують і відносять точку до класу, за який проголосувала більшість. У методі випадкових підпросторів (random subspace method, RSM) класифікатори навчаються на різних підмножинах признакового опису, які також виділяються випадковим чином.

На рисунку 1 зображено побудову дерев рішень в глибину і в ширину.

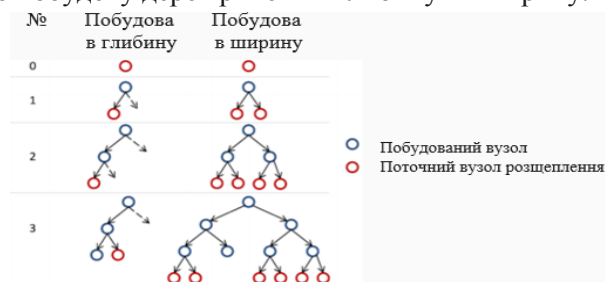


Рисунок 1 – Побудова дерев рішень

Ряд відомих дослідників розробили і дослідили метод k-means [6]. Мета методу — розділити n спостережень на k кластерів, так щоб кожне спостереження належало до кластера з найближчим до нього середнім значенням. Метод базується на мінімізації суми квадратів відстаней між кожним спостереженням та центром його кластера, тобто функції

$$\sum_{i=1}^N d(x_i, m_j(x_i))^2,$$

де d — метрика, x_i — i -ий об'єкт даних, а $m_j(x_i)$ — центр кластера, якому на j -ій ітерації приписаний елемент x_i .

Нехай маємо масив спостережень, що містить певні ознаки. Ці ознаки утворюють багатовимірний простір.

Алгоритм, що реалізує даний метод полягає в пошуку таких k -центрів кластерів, коли сумарне квадратичне відхилення елементів кластерів V від їх центрів буде найменшим:

$$V = \sum_{i=1}^k \sum_{x_j \in S_i} (x_i - \mu_i)^2,$$

де k — число кластерів, S_i — отримані кластери, $i=1,2,\dots,k$, μ_i — центри мас векторів $x_j \in S_i$.

Початкове зображення, отримане від відеокамери, є зашумленим. Для зменшення рівня шумів використовуємо фільтрацію. В роботі використано такі фільтри: медіанний, гаусовий, усереднюючий. Крім цього використано гістограмний аналіз. Для оцінки якості фільтрації застосовано такі показники: MSE – середньоквадратична похибка, PSNR – пікове відношення сигналу до шуму. Приклад фільтрації гістологічного зображення наведено на рисунку 1. Вхідне зашумлене зображення оброблене медіанним фільтром з вікнами $p = 3$ та $p = 7$. В результаті пікове відношення сигналу до шуму склало 42.11 та 32.0 відповідно[3]. На рисунку 2 зображено попередню обробку зображень.

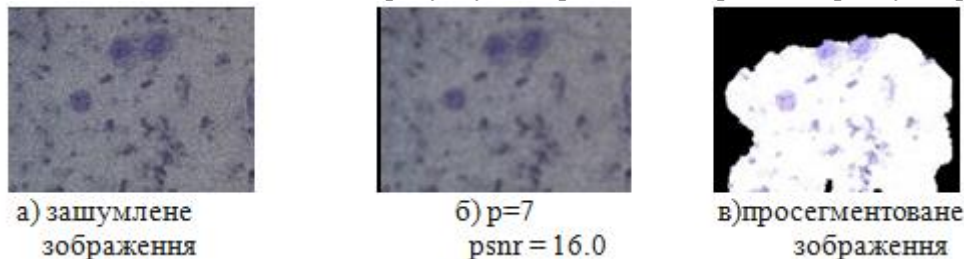


Рисунок 2 – Попередня обробка гістологічного зображення

Після проведених маніпуляцій із зображенням, був отриманий оптимальний варіант[4] зображення. Проводимо експеримент над відфільтрованим зображенням по алгоритму (рисунку 3):



Рисунок 3 – Узагальнений алгоритм роботи класифікатора

Тестування кластеризації наведено на рисунку 4.

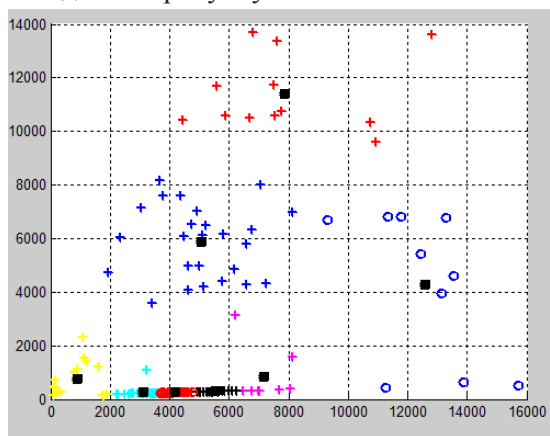


Рисунок 4 – Кластеризація біомедичного зображення

Висновок

У роботі досліджено алгоритми класифікації біомедичних зображень на основі методів k-means та Random forest. Дослідження проведено на навчальній вибірці біомедичних зображень. За допомогою аналітичного підходу досліджено основні властивості та принципи роботи методів класифікації, що дозволило виділити базові компоненти функціонування. Метод k-середніх більш зручний для кластеризації великої кількості спостережень, ніж метод ієрархічного кластерного аналізу. Метод RF зручний для отримання експертних критеріїв.

Список використаних джерел

1. Недзьведь А.М., Абламейко С.В. Обработка медицинских изображений гистологических объектов // Цифровая обработка изображений. Вып. 4. – Мн.: Ин-т техн. кибернетики НАН Беларуси, 2000. – С. 152-164.
2. Бюлетень Національного канцер-реєстру № 14 – «Рак в Україні, 2011-2012»; – 2013. – 123 с.
3. Berezsky O. Automated Processing of Cytological and Histological Images / O. Berezsky, O. Pitsun // Proceedings of XII International Conference Perspective Technologies and methods in mems design - 2016. - pp. 51–53.
4. Breiman L. Random forests / Breiman L. // Machine Learning. – Vol. 45, N 1. –2001. – P. 5–32.
5. E.M. Mirkes, K-means and K-medoids applet. University of Leicester, 2011.