

Факултет по изчислителна техника и автоматизация

ПРИЛАГАНЕ НА МЕТОДИТЕ НА МАШИННОТО ОБУЧЕНИЕ ПРИ ПАЦИЕНТИ С ОНКОЛОГИЧНИ ЗАБОЛЯВАНИЯ

доц. д-р инж. Недялко Николов
ас. инж. Павлина Линова, докторант, СИТ

Въведение

Технологиите се развиват толкова стремглаво, постиженията на човека са достигнали до такива висоти, има възможност да изследва, да наблюдава и най-отдалечени космически обекти, както и близките ни космически съседи като Марс, но въпреки това има области от науката, където независимо от напредъка в технологиите за наблюдение, скрининг и диагностика в здравословното състояние на човешкото тяло, все още делът на заболяемостта и смъртността на човека е в неприемливо високи стойности, особено за онкологичните заболявания.

Резултати

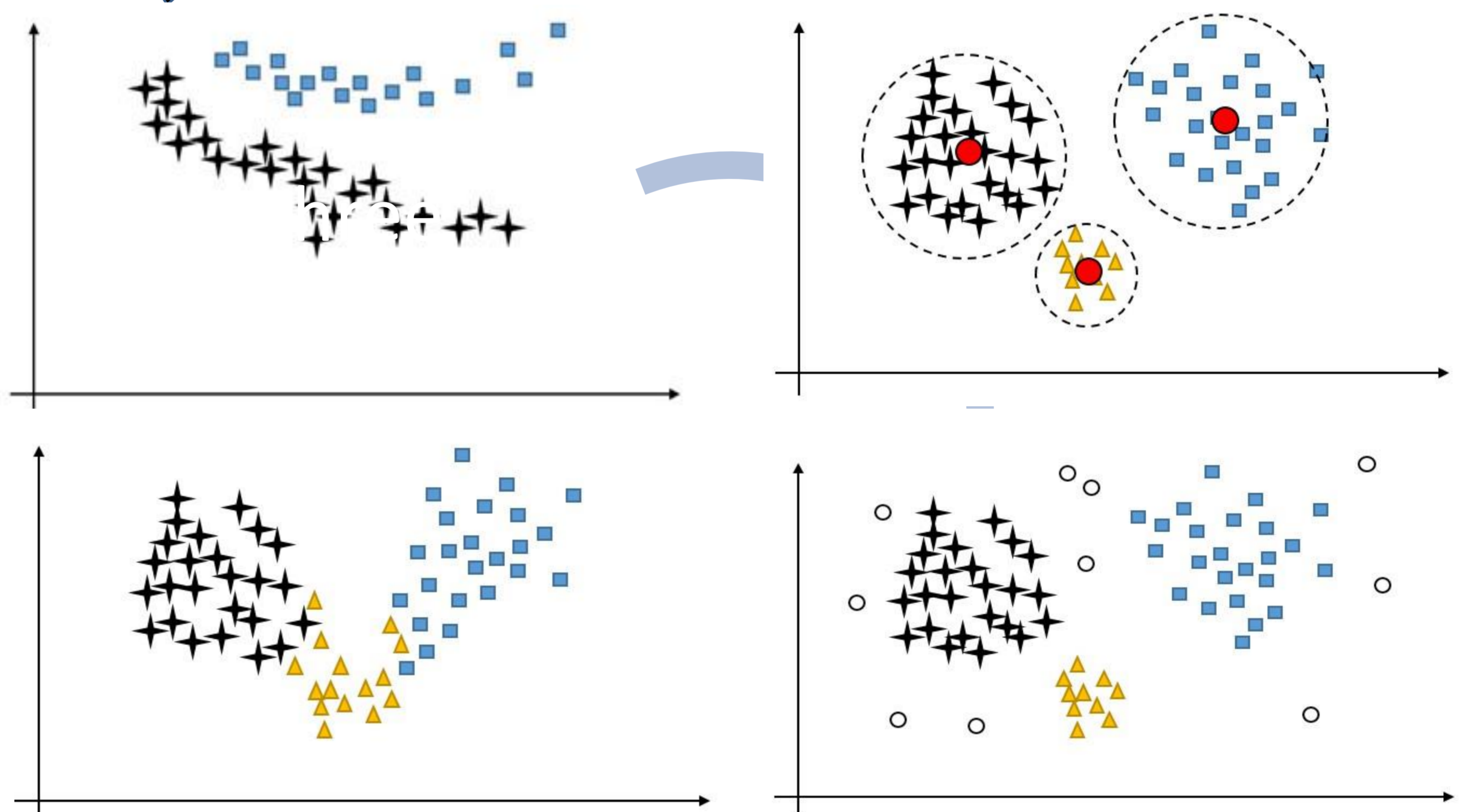
Анализирани са съществуващите алгоритми за клъстеризация, чието множество може да бъде разделено в две основни групи - йерархични и разделящи. Съществуват и още видове алгоритми за клъстеризация, които докторантът разгледа и обобща в разработена статия по този проект. Установи се, че поради големия интерес към този подход и приложимостта му в много направления и сфери на живот, все още се създават нови алгоритми на база на научни изследвания. Бе разгледан един от най-основните алгоритми – алгоритъм на k-средни (k-means), при което множеството от данни се разделя на k клъстера (групи), в които данните от един клъстер имат сходни характеристики, а данните от различни клъстери имат различни характеристики. Има различни метрики за измерване на това сходство / различие, които бяха разгледани от докторанта

Given two p-dimensional data points,
 $x_i = (x_{i1}, x_{i2}, x_{i3}, \dots, x_{ip})$ and
 $x_j = (x_{j1}, x_{j2}, x_{j3}, \dots, x_{jp})$
 The distance between the two data instances can be calculated using the Minkowski metric

$$d(x_i, x_j) = (\sum_{m=1}^p |x_{im} - x_{jm}|^g)^{1/g}, \quad (1)$$

If $g=2$, then the equation is as Eq.2, commonly known as Euclidean distance:

$$d(x_i, x_j) = \sqrt{\sum_{m=1}^p (x_{im} - x_{jm})^2} = \|x_i - x_j\|, \quad (2)$$



Заклучение

Във връзка с поставените задачи са направени опити за анализ на методи на клъстеризация поради все по-навлизашото им прилагане във всички сфери на работа и обработка на данни. Чрез индуктивни подходи в зависимост от типа и структурата на данните, които следва да бъдат клъстеризирани, този алгоритъм на клъстеризация в машинното обучение се обогатява с нови свои алгоритми или тяхното комбиниране. Предстои да бъде разгледано в следващи разработки, които могат да бъдат използвани за по-нататъшни експериментални опити и проверки на хипотези.

Изпълнените дейности по проекта подпомагат научно-изследователската работа на колектива и на научното звено по продължаване на анализа на методите и подходите за обработка на данни чрез машинно обучение, свързани с темата на проекта.

Публикации по проекта

1. Linova P., N. Kalcheva, N. Nikolov, Use of deductive and/or inductive approach in the study of nonlinear programming problems, 2021 XXX International Scientific Conference Electronics, ET 2021, 2021, DOI: 10.1109/ET52713.2021.9579696, 15-17 септември 2021 г., Созопол, Electronic ISBN:978-1-6654-4518-4, Print on Demand(PoD) ISBN:978-1-6654-4519-1 (<https://ieeexplore.ieee.org/document/9579696>) (Scopus)
2. Linova P.S., N. An. Arabadzhieva-Kalcheva, Algorithm of clustering in machine learning, Computer Science and Technologies Journal, TU-Varna, XIX-1/2021, pp.91-95, ISSN 1312-3335
3. Linova P.S., Clustering in machine learning as data science approach, Computer Science and Technologies Journal, TU-Varna, XIX-1/2021, pp.87-90, ISSN 1312-3335
4. P. Linova, Unsupervised learning and clustering approach of medical dataset analysis – summary, Annual Journal, TU-Varna

Благодарности

Научно-изследователски проект в помощ на докторанти ПДЗ/2021 „Прилагане на методите на машинното обучение при пациенти с онкологични заболявания“. Участниците в проекта благодарят на Технически Университет – Варна за финансирането и помощта при реализация на проекта.