

АНАЛІЗ МЕТОДІВ ВІДНОВЛЕННЯ ПРОПУЩЕНИХ ДАНИХ В ЗАДАЧІ ДІАГНОСТИКИ МІТОХОНДРІАЛЬНИХ ЗАХВОРЮВАНЬ

Бойко¹ Д.О., Гречаніна² Ю.Б., Поворознюк¹ А.І., Філатова¹ Г.Є.

¹Національний технічний університет «ХПІ», м. Харків

²Харківський спеціалізований медико-генетичний центр, м. Харків

Встановлення діагнозу з використанням комп'ютеру є сучасною допомогою медицині. Для побудови діагностичних правил повинна бути повна вибірка даних (в якій немає пропусків). За наявності пропусків сучасні методи не дають змоги створити діагностичні правила. Є два способи боротьби з даною проблемою: зменшення навчальної вибірки; відновлення пропущених даних. Зменшення вибірки полягає в видаленні з навчальної вибірки об'єктів, які мають пропуски, але це вносить похибку при побудові діагностичних правил.

Мета роботи – зробити порівняльний аналіз методів відновлення пропущених даних (ВПД). Контрольна вибірка представлена у вигляді таблиці експериментальних даних типа «об'єкт-ознака» $X = [x_{ij}]_{1,1}^{N,m}$, де x_{ij} – j -та ознака i -го об'єкта (пацієнта); N – кількість пацієнтів (27 пацієнтів); m – кількість ознак (8 ознак). Дані були надані Харківським спеціалізованим медико-генетичним центром. Через те що дані представлені в різних одиницях виміру, було проведено нормування контрольної вибірки.

У роботі проведені дослідження ВПД за допомогою наступних методів: відновлення середнім (ВС), k -найближчих сусідів (k -БС), k -найближчих сусідів модернізованим (k -БСМ), медіанний та евристичний. Результати порівняння методів ВПД приведені у табл.

Таблиця – Помилки методів відновлення даних

Похибка відновлення, %	Назва методу						
	ВС	k -бс ($k=1$)	k -бс ($k=2$)	k -бс ($k=3$)	k - бсм	Медіан- ний	Еврис- тичний
Найліпша	7,67	6,98	6,62	6,25	13,20	5,68	4,86
Середня	14,38	19,08	17,26	16,42	27,95	13,73	18,05
Найгірша	23,33	33,69	28,40	26,95	41,86	23,19	34,71

Проаналізувавши методи ВПД, не можливо однозначно вибрати кращий. Як видно з таблиці, найменша помилка отримана за допомогою евристичного метода. Подальше дослідження спрямовані на синтез більш ефективних методів ВПД, які враховують зв'язки між ознаками.