

ЗАСТОСУВАННЯ ВИПАДКОВИХ ЛІСІВ В ЗАДАЧАХ КЛАСИФІКАЦІЇ ЗОБРАЖЕНЬ.

Філіпчук О.Ю.

Науковий керівник – к.т.н., ст. викл. Маслій Р.В.

Стрімкий розвиток інформаційних технологій, зокрема, прогрес у методах збору, зберігання і обробки даних дозволив багатьом організаціям збирати величезні масиви даних, які необхідно аналізувати. Обсяги цих даних настільки великі, що можливостей експертів вже не вистачає, що породило попит на методи автоматичного дослідження (аналізу) даних, який з кожним роком постійно збільшується. Також він має вбудовану оцінку якостей та показує гарні результати при роботі з великою кількістю ознак (наприклад, в хімії, біології і т.д.). Цей метод, став основою для розробки різних його модифікацій, які є важливими для додатків. Метою даного дослідження є підвищення вірогідності класифікації зашумлених зображень та зменшення часу на навчання класифікатора при використанні випадкових лісів на основі алгоритму дерев рішень.

Метод заснований на побудові великого числа (ансамблю) дерев рішень (це число є параметром методу), кожне з яких будується за вибіркою, яке отримується з вихідної навчальної вибірки за допомогою бутстрепа (тобто вибірки з поверненням). На відміну від класичних алгоритмів побудови дерев рішень в методі випадкових лісів при побудові кожного дерева на стадіях розщеплення вершин використовується тільки фіксоване число випадково відібраних ознак навчальної вибірки (другий параметр методу) і будується повне дерево (без усічення), тобто кожен лист дерева містить спостереження тільки одного класу. Класифікація здійснюється за допомогою голосування класифікаторів, що визначаються окремими деревами, а оцінка регресії – усередненням оцінок регресії всіх дерев. Ймовірність коректної класифікації ансамблів класифікаторів істотно залежить від того, наскільки корельовані їх рішення. А саме, чим різноманітніші класифікатори ансамблю (менша корельованість їх рішень), тим вище ймовірність коректної класифікації

Було проведено експериментальне дослідження розробленого методу класифікації зображення на тестовому наборі зображень AIC (Aerial Image Classification), з використанням різних параметрів та його порівняння з дослідженням класифікації зашумлених даних. Та після того як були розглянуті дослідження на 3 видах шумів: сіль та перець, спекл-шум та гаусівський шум, було встановлено, що найкраща помилка класифікації була отримана при накладанні шуму сіль та перець: для класу field і houses помилка класифікації зовсім не змінилася, і дорівнювала 0 та 0.098 відповідно; для industry на зашумлених зображеннях помилка дорівнювала 0.286 та на оригінальних – 0.333; відповідно для trees 0.097 та 0.064.