

ДОСЛІДЖЕННЯ АНСАМБЛЕВИХ МЕТОДІВ ОЦІНКИ СТАНУ КОМП'ЮТЕРНОЇ СИСТЕМИ

Челак В.В., Гавриленко С.Ю.

*Національний технічний університет
«Харківський політехнічний інститут»,
м. Харків*

Стан сучасної комп'ютерної системи може характеризуватися великим об'ємом даних. Статистичні методи, зазвичай працюють з вибірками, що істотно спотворює прогнозованість результатів. Для аналізу величезних масивів інформації широко використовують інтелектуальний аналіз даних (Data mining).

В основі інтелектуального аналізу даних лежить багатоетапний автоматизований ітеративний процес виявлення знань в базах даних, з метою виявлення прихованих моделей. Для сегментації даних і оцінки ймовірності подальших подій використовуються складні математичні алгоритми, що базуються на машинних методах навчання. Найбільш популярними із них є ансамблеві методи або дерева рішень в основі яких лежить набір базових класифікаторів, результати прогнозування яких потім об'єднуються і формують прогноз агрегованого класифікатора. Перевагами ансамблевих методів є найвищий пріоритет точності оцінки об'єктів та можливість обробки неоднорідних даних.

У даній роботі виконано порівняльне дослідження найбільш популярних ансамблевих методів (boosting, bagging та stacking) [1, 2] для оцінки стану комп'ютерної системи з метою захисту даних.

У якості вихідних даних класифікації використано показники функціонування комп'ютерної системи (завантаження центрального процесора, пам'яті, обсяг трафіку, кількість операцій зчитування/запису на диск, статистичні дані аналізу системних подій).

Результати дослідження показали, що бустінг є найбільш ефективним методом з точки зору якості класифікації. Ансамблі схильні давати кращі результати, якщо є суттєва відмінність моделей. На відміну від бустінгу і традиційного бегінга, стекінг потребує використання різних алгоритмів класифікації.

Література:

1. Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani. Tree-Based Methods // An Introduction to Statistical Learning: with Applications in R. – New York: Springer, 2017. – С. 303-336.
2. Leo Breiman, Friedman J.H., Olshen R.A., Stone C.J. Classification and regression trees. – Monterey, CA: Wadsworth & Brooks/Cole Advanced Books & Software, 1984.