

**Вікторія В. Жебка, Володимир І. Виноградов,
Андрій П. Бондарчук, Михайло М. Степанов**

ОПТИМІЗАЦІЯ РОБОТИ АЛГОРИТМУ ГРАДІЄНТНОГО БУСТИНГУ ЗА ДОПОМОГОЮ ПЕРЕХРЕСНОЇ ПЕРЕВІРКИ

У статті розглянуто попередні дослідження перехресної перевірки. Показано як вона послідовно оцінює помилку прогнозування шляхом компромісу до заміщення і дисперсії і показано її застосування в якості критеріїв для зупинки алгоритму градієнтного бустингу.

Використовуючи змодельовані дані представлено збіжність і узгодженість алгоритму градієнтного бустингу. Проведено порівняння результатів для різних правил зупинки.

Розглянуто модель нелінійної логістичної регресії, де метод регуляризації не має рішення. В результаті проведення розрахунків отримано, що рівень похибок дуже швидко зменшується протягом перших кількох ітерацій, потім швидкість спаду стає приблизно лінійною. Коли рівень похибок досягає найнижчого рівня, далі він незначно підвищується. Це показує, що занадто багато ітерації призведе до несприятливого результату. З огляду на високу розмірність і нелінійність цієї проблеми, наш результат є обґрунтованим. Цей приклад показує, що застосування перехресної перевірки для алгоритму градієнтного бустингу є ефективним при розв'язанні задач багатовимірною нелінійною моделювання проблем.

На сьогодні важливі дані виходять з великих наборів неструктурованих даних. Багато закордонних компаній, де збирається велика кількість даних застосовують методи машинного навчання та інтелектуального аналізу даних для їх обробки. Саме у великих неструктурованих даних компанії шукають підказки для рішення проблем. Але дуже часто інструменти, які використовуються компаніями не дозволяють отримати достовірну інформацію. Не менше важливою проблемою є швидкість отримання інформації. З метою поліпшення процесу обробки даних використовуються алгоритми машинного навчання.

Ключові слова: градієнтний бустинг, машинне навчання, критерії зупинки, перехресна перевірка, нелінійна логістична регресія