

## **КОМПЛЕКСУВАННЯ ДЕКІЛЬКОХ АЛГОРИТМІВ ПРИ ВИРІШЕННІ ЗАДАЧІ ПРОГНОЗУВАННЯ**

Підхід комплексування декількох алгоритмів успішно застосовується для вирішення великого спектру задач машинного навчання на протязі багатьох років. Дійсно, поєднання одразу декількох методів зазвичай дозволяє досягнути кращого результату, ніж при використанні кожного методу окремо.

### **Постановка задачі прогнозування.**

Нехай задано  $n$  значень часового ряду:  $\vec{x} = [x_1, \dots, x_n]$ . Тоді задача прогнозування, що розглядається у даній роботі, полягає у побудові деякої моделі прогнозованого об'єкту, що б залежала від  $k$  послідовних значень часового ряду  $[x_i, \dots, x_{i+k}]$  та видавала значення прогнозу для значення  $x_{i+k+t}$ :  $F(x_i, \dots, x_{i+k}) = \hat{x}_{i+k+t}$ .

### **Огляд методів.**

Штучні нейронні мережі (ШНМ). ШНМ є систему з'єднаних і взаємодіючих між собою штучних нейронів. ШНМ не програмується в звичному сенсі цього слова, вони навчаються. У процесі навчання нейронна мережа здатна виявляти складні залежності між вхідними та вихідними даними, а також виконувати узагальнення. Здібності нейронної мережі до прогнозування безпосередньо впливають з її здатності до узагальнення і виділення прихованих залежностей між вхідними і вихідними даними. Після навчання мережа здатна прогнозувати майбутні значення деякої послідовності на основі декількох попередніх значень і / або якихось існуючих зараз чинників.

Метод групового урахування аргументів. Метод групового урахування аргументів (МГУА) [2] - це набір алгоритмів прогнозування (а точніше математичного моделювання), який ґрунтується на розбитті вихідних даних на дві вибірки: навчальну і перевірочну і використанні опорних функцій деякого вигляду, параметри яких знаходяться на навчальній вибірці, а перевірка того, наскільки добре вони моделюють заданий ряд, виконується на перевірочній вибірці.

Лінійна регресія з динамічними вагами (ЛРДВ). Даний метод є «вдосконаленням» звичайної лінійної регресії, що дозволяє знаходити декілька наборів ваг лінійної моделі для певних підвбірок (при чому ці ваги знаходяться з урахуванням не тільки прикладів відповідної підвбірки, але і усіх прикладів вибірки) і подальшого прогнозування шляхом усереднення усіх ваг, де коефіцієнти усереднення залежать від близькості приклади до відповідної підвбірки (кластеру).

### **Комплексування декількох алгоритмів.**

У даній роботі для вирішення задачі прогнозування пропонується використання підходу bagging для комплексування моделей, отриманих за допомогою 3 методів: ШНМ, МГУА та ЛРДВ. Для оцінки усереднюючих коефіцієнтів моделей використаємо зовнішній критерій – а саме помилку моделі на окремій, валідаційній вибірці.

Таким чином, алгоритм побудови прогнозуючої моделі має наступний вигляд:

1. Попередня обробка даних.
2. Генерація підвбірок. Згідно з підходом bagging, генеруються 3 підвбірки розміру  $n' \leq n$  шляхом семплювання з заміною.
3. На кожній підвбірці навчається модель з використанням певного методу: ШНМ, МГУА чи ЛРДВ.
4. Знаходяться усереднюючі ваги. Для цього спочатку розраховуються середньоквадратичні помилки кожної моделі на валідаційній вибірці:

$$\vec{e} = [e_1, e_2, e_3]$$

$$e_i = \sum_{j=1}^m (F_i(\vec{x}_j^{(v)}) - y_j^v)^2$$

Після цього, усереднюючі коефіцієнти моделей розраховуються як:

$$\alpha_i = \frac{1}{2} * (1 - \frac{e_i}{\sum_i e_i}), i = 1, 2, 3.$$

5. Отже, «фінальна» прогнозуюча модель буде мати вигляд:

$$F(\vec{x}) = \sum_{i=1}^3 \alpha_i * F_i(\vec{x}).$$

### **Висновки.**

У даній роботі запропоновано метод вирішення задач прогнозування на основі комплексування наступних алгоритмів: ШНМ, МГУА та ЛРДВ, де усереднюючі коефіцієнти моделей знаходяться на валідаційній вибірці. Поєднання цих підходів дозволяє на практиці знайти більш адекватну модель прогнозованого об'єкту.