

ОСОБЛИВОСТІ ЗАСТОСУВАННЯ ГЕНЕТИЧНИХ АЛГОРИТМІВ У ЗАДАЧАХ ПРОГНОЗУВАННЯ

У статті розглядаються особливості алгоритму LGAP та можливість його застосування до вирішення задач прогнозування. Приведено схему алгоритму та розглянуті питання обчислення точності прогнозу.

Ключові слова: Алгоритм LGAP, задачі прогнозування, колективний прогноз.

В статье рассматриваются особенности алгоритма LGAP, и возможность его применения при решении задач прогнозирования. Приведена схема алгоритма и рассмотрены вопросы точности прогноза.

Ключевые слова: Алгоритм LGAP, задачи прогнозирования, коллективный прогноз.

In the article examined the feature of algorithm of LGAP, and possibility of his application at the decision of tasks of prognostication. The chart of algorithm is resulted and the questions of exactness of prognosis are considered..

Key words: Algorithm of LGAP, tasks of prognostication, collective prognosis.

Розвиток методів прогнозування як науки в останні десятиріччя привів до створення багатьох методів, процедур, прийомів побудови прогнозу, нерівноцінних за своїм значенням. За оцінками закордонних і вітчизняних фахівців із прогностики вже налічується понад сто методів прогнозування, у зв'язку із чим, завжди при вирішенні практичних задач виникає проблема вибору методів, які давали б адекватні (точні) прогнози.

Більша частина прогнозів будується на підставі методів експертних оцінок, або ж методів математичної статистики. Застосування експертних методів більш ефективно в короткостроковому прогнозуванні, тому що вони у певній мері спрощують реальні процеси й не виходять за рамки даних сьогодення.

Методи математичної статистики вимагають менше матеріальних витрат, але в основному алгоритми такого роду опираються на минулу ситуацію (або на ситуацію за минулий тиждень, або за вчорашній день – залежно від вибору масштабу) і намагається відобразити динаміку минулого сезону на поточний час. Такий підхід може привести до неспроможності прогнозів, оскільки, наприклад, може зміниться формат та їхня кількість.

Одним із перспективних методів прогнозування є прогнозування за допомогою алгоритмів, що навчаються. Алгоритми, що навчаються аналізують впливи на систему та відповіді системи на ці впливи. Задача може ускладнюватись тим, що вхідні впливи на систему можуть бути невідомими, та можуть проявлятися через вихідні характеристики системи. Подальшим ускладненням може бути наявність прихованих впливів характеристик одних об'єктів на інші. Наприклад, коливання курсів валют та цінних паперів на біржі.

Клас алгоритмів, що навчаються, достатньо широкий. Серед них для вирішення завдань прогнозування найцікавішим є алгоритм LGAP [1]. Генетичний алгоритм LGAP (Learning Genetic Algorithm Prognosis), що навчається, призначений для вилучення закономірностей (знань) з табличних даних і використання цих знань для прогнозування майбутніх подій.

Алгоритм LGAP складається з чотирьох етапів: 1) формування базових елементів; 2) відбір відповідних об'єктів; 3) вироблення первісних варіантів прогнозу; 4) отримання остаточного прогнозу.

1. ФОРМУВАННЯ БАЗОВИХ ОБ'ЄКТІВ

Розглянемо алгоритм на прикладі таблиці «об'єкт-час». Кожен елемент $b_{i,t}$ таблиці (протоколу подій за T минулих днів) відображає значення однієї характеристики i -го об'єкту ($i = 1, 2, \dots, Y$) в t -й момент часу. Моменти часу ($t = 1, 2, \dots, T$) впорядковані в таблиці «за віком»: найсвіжіші дані мають індекс $t = 1$, дані за попередній день $t = 2$ і так далі до дня з індексом $t = T$.

У даному алгоритмі до складу базової множини можуть входити G будь-яких елементів таблиці, узятих з протоколу за τ останніх днів. Такі множини називатимемо базовими об'єктами потужності G . У протоколі за τ днів є N елементів ($N = Y \times \tau$). Потужність базових об'єктів можна міняти у межах від одиниці до N . При кожній заданій потужності G об'єкти можуть мати різну «архітектуру», унаслідок чого кількість варіантів базових об'єктів потужності G рівна числу поєднань з N по G , а загальна кількість різних базових об'єктів виражається рівнянням:

$$M = \sum_{G=1}^N C_N^G.$$

Процедура оцінки якості базових об'єктів включає етап пошуку відповідних об'єктів і етап оцінки очікуваної помилки прогнозу з використанням цих об'єктів.

2. ВІДБІР ВІДПОВІДНИХ ОБ'ЄКТІВ

Елементи базового об'єкта q , що починається з елементів t -й рядка, в масиві початкових даних помічені їх порядковими номерами g в об'єкті: $(1, t), (2, t), \dots, (g, t), \dots, (G, t)$. Набір адрес цих елементів описує структуру (архітектуру) конкретного об'єкта, що складається з G елементів. Якщо індекс t у всіх елементів об'єкта збільшити на задане число p , виходить об'єкт тієї ж структури, що і початковий, але тільки зрушений у часі на p кроків назад. Такий об'єкт назовемо ізоморфним даному базовому об'єкту.

Виділимо серед ізоморфних об'єктів групу, що складається з «потенційно відповідних» об'єктів. У неї включаються ізоморфні об'єкти з найбільшою схожістю на базовий об'єкт q . Міру схожості між об'єктами можна оцінювати різними способами.

Якщо орієнтуватися на абсолютні значення характеристик їх елементів, то можна використовувати евклідову відстань між об'єктами:

$$R_{A,q,p} = \sqrt{\sum_{g=1}^G [b_{g,t} - b_{g,(t+p)}]^2}$$

У групу відповідних об'єктів відбирається об'єкти з найменшими відстанями RA, q, p .

Якщо вважати, що ми маємо справу з даними, вимірними у шкалі відносин, то схожість двох об'єктів можна виявити через відстань між відповідними відносинами:

$$R_{o,q,p} = \sqrt{\sum_{g=1}^G \sum_{s=1}^G [b_{g,t}/b_{s,t} - b_{g,(t+p)}/b_{s,(t+p)}]^2}.$$

3. РОЗРОБКА ЛОКАЛЬНИХ ВАРІАНТІВ ПРОГНОЗУ

Стратегії прогнозування залежать від того, які відносини між елементами різних об'єктів вважаються інваріантними при переході від одного об'єкта до іншого (постійними для різних моментів часу). Якщо вважається, що інформативними є абсолютні значення відповідних елементів об'єктів, то значення прогнозованого елементу можна знайти по абсолютних значеннях відповідних елементів (предикторів), пов'язаних із компетентними об'єктами даної групи. Кількість різних відповідних базових об'єктів потужності G рівна W . Отже, при прогнозуванні може взяти участь $W \times k$ предикторів (W груп по do предикторів). Якщо передбачається лінійна залежність між відповідними елементами різних об'єктів, то можна скористатися прогнозуванням за допомогою

лінійної регресії. Будується лінійна регресія між G елементами базового та p -го компетентного об'єкта.

4. ОТРИМАННЯ ОСТАТОЧНОГО ПРОГНОЗУ

Локальні прогнози отримані за допомогою предикторів, організованих в ієрархічну структуру: окремі предиктори об'єднуються у групи, які об'єднуються в колективи і так далі. Фактично ми маємо справу з колективно-груповим методом прогнозування [2]. При об'єднанні локальних рішень в єдине має сенс звернути увагу на той факт, що компетентність всіх предикторів, породжувачів ці рішення, перевищує деякий поріг, але може бути неоднаковою. В результаті різні групи предикторів виробляють різні набори приватних рішень. Можна розраховувати на те, що компетентніші групи видаватимуть точніші прогнози. Процедура отримання узагальненого рішення полягає в наступному. Спочатку на базі приватних рішень для кожної групи виробляється групове вирішення. При прогнозуванні як групове рішення може бути використане середньоарифметичне значення локальних прогнозів або їх медіана. Обчислюється також характеристика компетентності групи L_j . Ці характеристики знаходяться для всіх до груп предикторів з колективу груп, що брали участь у прогнозуванні.

Узагальнене рішення B на наступному ієрархічному рівні (на рівні колективу) може бути отримано з використанням параметричного сімейства функцій зваженого усереднювання групових рішень:

$$B = \sum_{f=1}^k (B_f \times L_f^a) / \sum L_f^a$$

Тут величина показника ступеня відображає стратегію обліку впливу компетентності. Якщо $a = 0$, то вирішення всіх груп враховуються з рівними вагами. Із зростанням a росте вплив компетентніших груп.

Якщо у прогнозуванні бере участь один колектив групових предикторів, то величина колективного рішення приймається як остаточне значення прогнозованої величини. Якщо визначити дисперсію групових прогнозів, то можна отримати уявлення про компетентність колективу в цілому і про величину очікуваної помилки даного прогнозу.

Метод допускає використання не одного колективу, а множини з W колективів. Нові колективи можуть породжуватися базовими об'єктами, потужність яких відрізняється від G . У цьому випадку за формулою, аналогічній вищевказаній, робиться зважене усереднювання колективних рішень.

На рис. 1 приведена схема метода колективного прогнозування. На першому етапі використовується базовий елемент потужності G_1 і архітектура F_1 . Знаходяться компетентні елементи C_1, C_2, \dots, C_k . Вони утворюють першу групу предикторів. З їх використанням отримують локальні прогнози b_1, b_2, \dots, b_n , обчислюються їх дисперсія D_1 і груповий прогноз b_1' .

На другому етапі ці процедури повторюються для об'єктів тієї ж потужності G_1 , але іншої архітектури: F_2, F_3, \dots, F_m . У результаті виходять групові рішення b_m', b_2' і по ним обчислюється колективне рішення a_1'' і дисперсія D_1'' .

На третьому етапі змінюється потужність базових об'єктів, і для кожної потужності (G_2, G_3, \dots, G_n) повторюються всі попередні процедури. У результаті виходить множина з n колективних рішень: $b_1'', b_2'', \dots, b_n''$ зі своїми значеннями групових дисперсій D_2', D_3', \dots, D_n' . За ними знаходиться остаточний прогноз b_0 і дисперсія колективних прогнозів, D_0 по якій можна судити про очікувану помилку отриманого прогнозу.

Ще одне суттєве питання виникає у процесі аналізу результатів прогнозування, це – точність прогнозу.

Часто береться абсолютне відхилення прогнозу $b_{i,0}'$ від дійсного значення $b_{i,0}$, що ділиться на дійсне значення:

$$d' = |b'_{i,0} - b_{i,0}| / b_{i,0}.$$

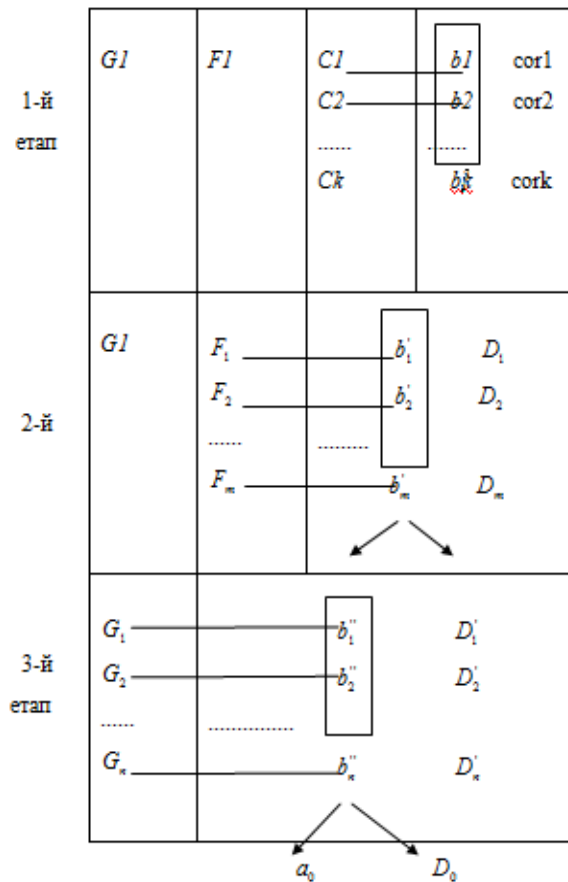


Рис. 1. Схема метода колективного прогнозування

Така відносна величина мало чутлива до помилок прогнозу великих значень і надмірно чутлива до помилок прогнозу величин, близьких до нуля. Крім того, різниця dd_i між мінімальним і максимальним значеннями може бути різною у різних спостережуваних характеристик і однакова відносна помилка d' буде прийнятною для ухвалення рішень в одних випадках і не прийнятною в інших.

У зв'язку з цим пропонується судити про точність прогнозу i -тої характеристики за величиною помилки, що нормується за різницею dd_i :

$$d = |b'_{i,0} - b_{i,0}| / dd_i.$$

Така міра володіє однаковою чутливістю до помилок прогнозу для різних значень прогнозованої характеристики. Її чутливість до помилок тим вище, чим у менших межах коливається прогнозована характеристика, що представляється цілком логічним.

Іноді важливо знати не абсолютну величину $b_{i,0}$ характеристики у майбутньому, а лише те, чи буде вона більше або менше значення $b_{i,t}$ у даний момент часу. У таких випадках застосовна міра точності прогнозу, що враховує лише збіги знаків:

$$d^* = \begin{cases} 0 & \text{якщо } i (b_{i,0} > b_{i,t}) (b'_{i,0} > b_{i,t}) \\ & \text{або } (b_{i,0} < b_{i,t}) i (b'_{i,0} < b_{i,t}); \\ 0.5, & \text{якщо } i (b_{i,0} = b_{i,t}) (b'_{i,0} \neq b_{i,t}); \\ 1 & \text{у інших випадках.} \end{cases}$$

Алгоритм LGAP можна розпаралелювати на число елементів U , що передбачаються. Кожна з V ліній при заданому числі G елементів базової множини розпаралелюється на MG незалежних процесів пошуку компетентних об'єктів і прогнозування за ними. Обміни між процесами робляться на етапі вибору кращих базових і створення нових об'єктів шляхом мутацій або схрещувань, після чого нові базові об'єкти запускаються в незалежні паралельно протікаючі процеси [1, 4].

Алгоритм LGAP найбільш доцільно використовувати для вирішення прикладних задач: прогноз результатів діяльності підприємств, прогноз врожайності сільхозкультур, прогноз стану пацієнтів, які спостерігаються. Значний інтерес представляють нові задачі прогнозування процесів економічного, екологічного і демографічного характеру, проблеми прогнозування стійкого розвитку.

ЛІТЕРАТУРА

1. Загоруйко Н.Г. Прикладные методы анализа данных и знаний. – Новосибирск: НГУ, 2002. – 263 с.
2. Бэстэнс Д.-Э., ван ден Берг В.-М., Вуд Д. Нейронные сети и финансовые рынки: принятие решений в торговых операциях. – М.: ТВП, 1997. – 236 с.
3. Родионов П.Е. Решение задач финансового прогнозирования с использованием нейронных сетей // Современные технологии в задачах управления, автоматизации и обработки информации: Сборник трудов международного научно-технического семинара. – М., 1997. – С. 248-250.
4. Журавлев Ю.И., Загоруйко Н.Г. Класс коллективно-групповых решающих правил, основанных на дисперсионном критерии компетентности предикторов // Анализ данных и сигналов. – Новосибирск. – 1998. – Вып. № 163, Вычислительные системы. – С. 82-90.