

Є.О. ГОФМАН, аспірант, ЗНТУ, Запоріжжя,
А.О. ОЛІЙНИК, к.т.н., ЗНТУ, Запоріжжя,
С.О. СУББОТИН, к.т.н., доц. ЗНТУ, Запоріжжя

АГЕНТНИЙ МЕТОД СИНТЕЗУ ДЕРЕВ РІШЕНЬ

Розглянуто завдання побудови моделей у вигляді дерев рішень. Проаналізовано мультиагентний метод з непрямим зв'язком між агентами. Запропоновано мультиагентний метод ідентифікації дерев рішень. Розроблено програмне забезпечення, що реалізує запропонований метод. Бібліогр.: 10 назв.

Ключові слова: агент, дерево рішень, модель, мультиагентний метод.

Постановка проблеми. Розв'язання прикладних завдань у різних областях промисловості, медицини, економіки, веб-технологій пов'язано з необхідністю побудови моделей досліджуваних об'єктів, процесів та систем, для чого запропоновано різні підходи, зокрема методи регресійного аналізу, нейронні мережі, нечітке, нейро-нечітке моделювання й інші [1, 2]. Проте часто необхідно побудувати модель, яка не тільки забезпечую прийнятну точність, а і дозволяє легко зрозуміти, чому приймається відповідне рішення. Досить інтерпретабельними та зрозумілими є моделі, побудовані у вигляді дерев рішень [3 – 7].

Нехай задана навчальна вибірка даних, що складається з N екземплярів, кожний з яких характеризується P атрибутами. При цьому кожний атрибут може відноситися до деякого лінгвістичного терму T . Для кожного i -го екземпляра вказані входження до лінгвістичних термів для кожного атрибута та вказано лінгвістичний терм вихідної змінної. Тоді необхідно побудувати таке дерево рішень, яке дозволяє виконувати віднесення вихідного параметра до лінгвістичного терму із заданою точністю:

$$Q^* \geq Q_{threshold}$$

де Q^* – точність прогнозування за синтезованим деревом рішень;
 $Q_{threshold}$ – прийнятна точність прогнозування.

Аналіз літератури. Відомі різні методи ідентифікації дерев рішень, зокрема ID3, CART, CHAID, QUEST, C5.0 та ін. [3, 4]. Однак більшість із них мають певні недоліки, пов'язані з великою обчислювальною складністю, проблемами формування дерева рішень (ріст дерева, відсікання частини дерева) та ін. [3 – 7].

У зв'язку з цим актуальною є розробка нових методів синтезу дерев рішень, вільних від недоліків існуючих. Одним з нових напрямків

штучного інтелекту є мультиагентні методи з непрямим зв'язком між агентами, що дозволяють вирішувати різні оптимізаційні завдання [2, 8, 9]. Такі методи є особливо ефективними при розв'язанні завдань дискретної оптимізації, тому в цій статті пропонується застосувати мультиагентний підхід з непрямим зв'язком між агентами для розв'язання завдання ідентифікації дерев рішень.

Мета статті – розробка мультиагентного методу з непрямим зв'язком між агентами для синтезу дерев рішень.

Основними завданнями роботи є:

- дослідження основних принципів роботи дерев рішень;
- аналіз мультиагентного методу з непрямим зв'язком між агентами;
- розробка мультиагентного методу ідентифікації дерев рішень;
- розробка програмного забезпечення, що реалізує запропонований мультиагентний метод.

Дерева рішень. Дерева рішень являють собою спадну систему, засновану на підході "розділай і пануй", основною метою якої є розподіл дерева на взаємно непересічні підмножини [3, 5]. Кожна підмножина являє собою підзадачу класифікації.

Дерево рішень описує процедуру прийняття рішення про приналежність певного екземпляра до того або іншого класу.

Дерево рішень є деревоподібною структурою, що складається із внутрішніх і зовнішніх вузлів, зв'язаних ребрами [6]. Внутрішні вузли – модулі, що приймають рішення, – розраховують значення функції розв'язку, на підставі чого визначають дочірній вузол, який буде відвідано далі. Зовнішні вузли (іноді називаються кінцевими вузлами), навпаки, не мають дочірніх вузлів і описують або мітку класу, або значення, що характеризує вхідні дані. У загальному випадку, дерева рішень використовуються в такий спосіб. Спочатку передаються дані (як правило, це вектор значень вхідних змінних) на кореневий вузол дерева рішень. В залежності від отриманого значення функції рішення, використовуваної у внутрішньому вузлі, відбувається перехід до одного з дочірніх вузлів. Такі переходи тривають доти, поки не буде відвідано кінцевий вузол, що описує або мітку класу, або значення, зв'язане з вхідним вектором значень ознак.

Мультиагентний метод з непрямим зв'язком між агентами. Мультиагентний метод з непрямим зв'язком між агентами є багатоагентним евристичним ітеративним методом випадкового пошуку [2, 8]. Поведінка агентів моделюється як процес переміщення й дослідження простору пошуку. Особливістю моделюемого переміщення є моделювання виділення феромонів, які агенти залишають на шляху в

процесі свого переміщення. Феромони в процесі роботи випаровуються. Таким чином, на найкращому шляху залишається більша кількість феромонів, оскільки додавання феромонів відбувається частіше, ніж випаровування. А оскільки вибір шляху для переміщення агентів ґрунтується на інформації про кількість феромонів, то агенти вибирають кращий шлях.

На етапі ініціалізації задаються параметри методу, що впливають на його роботу. Далі відбувається пересування агентів між вузлами графа, у результаті чого, після закінчення пересування кожного агента формуються розв'язки, з яких вибирається кращий на даній ітерації. Далі виконується перевірка критеріїв зупинення. Якщо перевірка була виконана успішно, відбувається закінчення пошуку, у процесі якого вибирається найкращий розв'язок з усіх, що зустрічалися на пройдених ітераціях. Якщо ж перевірка була неуспішною, то проводиться відновлення граней, яке полягає в імітації випаровування феромонів, та перезапуск агентів.

Ґрутуючись на принципах мультиагентного методу, його різновидах і областях застосування [2, 8, 9] можна виділити наступні переваги та недоліки.

До переваг методу можна віднести те, що:

- він може використовуватися в динамічних застосуваннях (агенти адаптуються до змін навколишнього середовища);
- у процесі пошуку метод використовує пам'ять усієї колонії, що досягається за рахунок моделювання виділення феромонів;
- збіжність методу до оптимального розв'язку гарантується;
- стохастичність оптимізаційного процесу, тобто випадковість пошуку, за рахунок чого виключається можливість зациклення в локальному оптимумі;
- мультиагентність методу;
- можливість застосування до розв'язання різних завдань оптимізації.

До недоліків методу варто віднести такі:

- теоретичний аналіз ускладнений, оскільки підсумковий розв'язок формується в результаті послідовності випадкових розв'язків; розподіл ймовірностей змінюється при ітераціях; дослідження є більш експериментальними, ніж теоретичними;
- збіжність гарантується, але час збіжності не визначений;
- висока ітеративність методу;
- результат роботи методу досить сильно залежить від початкових параметрів пошуку, які підбираються експериментально.

Таким чином, можна відзначити, що розглянутий мультиагентний метод з непрямим зв'язком може ефективно вирішувати завдання переважно дискретної оптимізації, які можуть бути узгоджені з наступними вимогами:

- відповідне подання задачі – задача повинна бути описана у вигляді графа з набором вузлів і граней між вузлами;

- евристична придатність елементів графа, на основі яких формується розв'язок, – можливість застосування евристичної міри адекватності окремих елементів у графові пошуку;

- складання альтернативних розв'язків, за допомогою чого можна раціонально визначати припустимі розв'язки;

- правило відновлення феромонів – правило, яке визначає ймовірність переміщення агента з одного вузла графа до іншого.

Мультиагентний метод синтезу дерев рішень. Для виконання ідентифікації дерев рішень з використанням мультиагентного підходу з непрямим зв'язком між агентами слід перетворити основні етапи методу відповідно до особливостей розв'язуваного завдання.

Далі приводяться основні зміни в етапах мультиагентного методу.

1. Ініціалізація. На даному етапі в мультиагентному методі створюється граф пошуку, встановлюються параметри роботи методу, а також розраховуються евристичні міри важливості вузлів графа. Для розв'язання завдання побудови дерев рішень граф пошуку буде складатися з вузлів, що представляють окремі лінгвістичні терми, до яких можуть ставитися лінгвістичні змінні. При цьому для кожного лінгвістичного терма вихідної лінгвістичної змінної створюється окремий граф пошуку, для якого розраховуються окремі матриці евристичних значимостей і феромонів. У зв'язку з цим пошук на кожному графові пошуку виконується окремою множиною агентів. Такий підхід викликаний тим, що при розв'язанні завдання ідентифікації дерев рішень важливість лінгвістичних термів залежить від лінгвістичних термів вихідної змінної, а також має значення порядок відвідування вузлів агентами.

2. Пересування агентів. При пересуванні агенти приймають рішення, у який вузол переміститися, таким чином, формуються окремі дерева рішень для кожного лінгвістичного терма вихідної лінгвістичної змінної. Для такого розв'язку пропонується застосовувати правило випадкового вибору, що базується на евристичних мірах важливості та мірі пріоритетності, заснованої на моделюванні виділення феромонів у процесі пересування. Рішення про завершення переміщення окремого агента слід приймати, виходячи з того, наскільки добре побудоване

дерево рішень виділяє відповідні класи екземплярів вихідної навчальної вибірки.

3. Зміна ступеню значимості вузлів. При розв'язанні завдання синтезу дерев рішень як міру пріоритетності необхідно використовувати якість покриття окремого дерева екземплярів відповідного класу. Крім того, пропонується використовувати елітну стратегію, що дозволить забезпечити більш швидку збіжність до підсумкового розв'язку.

4. Відновлення феромонів. Процедура відновлення феромонів не має істотних особливостей для розв'язуваного завдання, тому її можна застосовувати в традиційному виді.

Виходячи з виділених особливостей, які повинен мати розроблюваний мультиагентний метод ідентифікації дерев рішень, було створено метод синтезу дерев рішень з непрямым зв'язком між агентами, який представлений у вигляді наступної послідовності етапів.

Eman 1. Ініціалізація. Задаються статичні параметри роботи методу: коефіцієнти α, β, ρ . Для кожного з можливих лінгвістичних термів вихідних значень створюється свій граф пошуку, що представляє собою лінгвістичні терми, які можуть бути включені в дерево рішень, і, відповідно, своя окрема множина агентів. Також важливою особливістю розроблювального методу є те, що створюються вузли, які характеризують інверсні лінгвістичні терми, тобто це необхідно для випадку, коли в дереві рішень вибирається варіант, що умова в вузлі не спрацювала. Крім того, для кожного графа пошуку розраховуються евристичні значення значимості окремого терма для відповідного лінгвістичного терма вихідної змінної (1):

$$\eta_p^q = \frac{\sum_{o=1}^N \min(\mu_p(o), \mu_q(o))}{\sum_{o=1}^N \mu_q(o)}, \quad \forall p = \overline{1, T}, \quad q = \overline{1, K}, \quad (1)$$

де η_p^q – значення евристичної значимості лінгвістичного терма p для опису класу q ; o – екземпляр вхідної вибірки, що містить N екземплярів; $\mu_p(o), \mu_q(o)$ – значення функції приналежності об'єкта o терму p й класу q , відповідно; T – кількість лінгвістичних термів для вхідних змінних; K – кількість лінгвістичних термів вихідної змінної.

У кожному просторі пошуку кожному вузлу графа пошуку ставиться у відповідність початкове значення кількості феромонів τ_{init} :

$$\tau_p^q(1) = \tau_{init}, \quad \forall p = \overline{1, T}, \quad q = \overline{1, K}, \quad (2)$$

де $\tau_p^q(1)$ – значення кількості феромонів для p -го терма в просторі пошуку для q -го класу на першій ітерації пошуку.

Етап 2. Вибір терму для додавання в правило j -го агента в графі пошуку i -го лінгвістичного терму вихідної змінної.

Для j -го агента на основі випадкового правила вибору розраховується ймовірність включення k -го лінгвістичного терму в правило, що описує i -ий лінгвістичний терм вихідної змінної (3):

$$P_k^{i,j} = \frac{\eta_k^i \cdot \tau_k^i(t)}{\sum_{p \in R^j} \eta_p^i \cdot \tau_p^i(t)}, \quad (3)$$

де $P_k^{i,j}$ – ймовірність додавання k -го терма в дерево рішень j -го агента в графові пошуку для i -го класу; R^j – множина термів, які можуть бути додані в дерево рішень j -го агента. Оскільки додавання певного лінгвістичного терму означає, що дерево рішень перейшло на наступний рівень і вже аналізується інша вхідна змінна, то крім обраного терму, виключаються також усі терми, що описують дану вхідну змінну.

Після цього перевіряється умова (4)

$$P_k^{i,j} > rand(1), \quad (4)$$

де $rand(1)$ – випадкове число з інтервалу $[0; 1]$.

Якщо умова виконується, тоді лінгвістичний терм k додається в дерево рішень j -го агента, видаляються терми, пов'язані з відповідною вхідною змінною, з множини можливих термів для даного агента й виконується перехід на наступний етап. В іншому випадку продовжується розгляд термів, що залишилися.

Етап 3. Перевірка завершення переміщення j -го агента.

Якщо множина термів, які j -ий агент може додати у формоване правило, є порожньою, то виконується перехід на наступний етап.

В іншому випадку визначається, скільки екземплярів i -го класу покриває дерево рішень j -го агента. Для всіх екземплярів, що відносяться до класу i , розраховується значення вихідної змінної відповідно до дерева рішень j -го агента, і на підставі одержуваних даних збільшується лічильник $cntMatch$, у якому зберігається кількість екземплярів, що покриваються отриманим деревом рішень. Після цього перевіряється умова (5)

$$cntMatch \geq inCntMatchMin_i, \quad (5)$$

де $inCntMatchMin_i$ – гранична мінімальна кількість екземплярів i -го класу, яка повинна визначатися деревом рішень.

Якщо зазначена умова виконується, то вважається, що дерево рішень ідентифікує необхідну кількість екземплярів, та j -ий агент завершив своє переміщення. Якщо усі агенти завершили своє переміщення ($j = cntAgents$) та перевірена приналежність до усіх класів ($i = k$), то виконується перехід на наступний етап. В іншому випадку виконується перехід до етапу 2.

Етап 4. Формування дерев рішень та оцінювання їх якості. Древа рішень формуються випадковим чином шляхом усіляких накладень отриманих агентами розв'язків. При цьому сполучаються такі дерева рішень, які виконують відбір для однакових лінгвістичних термів вихідної змінної та з однаковим коренем (6):

$$p_1^{q,i} = p_1^{q,j}, \quad i \neq j, \quad (6)$$

де $p_1^{q,i}$ та $p_1^{q,j}$ – кореневі вузли дерев рішень i -го та j -го агентів, які використовуються для прогнозування q -го класу екземплярів.

Для оцінювання якості дерев рішень використовується вхідна навчальна вибірка, для кожного екземпляра якої визначається клас за відповідним деревом рішень. Грунтуючись на дані про клас екземплярів, отримані за допомогою дерева рішень, і класі екземплярів, виходячи із заданої навчальної вибірки, розраховують оцінку якості дерева рішень (7)

$$Q = \frac{cntMatch}{N}, \quad (7)$$

де $cntMatch$ – кількість екземплярів, для яких клас був визначений вірно за допомогою заданого дерева рішень; Q – якість прогнозування класу екземплярів на основі відповідної бази правил.

Якщо виконується умова (8)

$$Q_{high} \geq Q_{threshold} \quad (8)$$

де Q_{high} – якість прогнозування дерева рішень, яка характеризується найкращою точністю прогнозування; $Q_{threshold}$ – прийнятна якість прогнозування, то відбувається перехід до етапу формування результуючого дерева рішень.

Етап 5. Додавання феромонів. Додавання феромонів виконується з метою підвищення пріоритетності тих термів, включення яких у дерево рішень сприяє підвищенню якості прогнозування результуючих дерев рішень. У зв'язку з цим кількість додаваного коефіцієнта пріоритетності прямо пропорційна якості прогнозування дерева рішень, у яке входить заданий лінгвістичний терм. При цьому додавання феромонів пропонується виконувати тільки для тих термів, які входять у дерева рішень, для яких виконується умова (9)

$$Q_{DT} \geq \delta \cdot Q_{high}, \quad (9)$$

де δ – коефіцієнт, що визначає, наскільки близько якість прогнозування дерева рішень DT повинно наблизитися до кращої якості прогнозування етапу формування результуючого дерева рішень, щоб можна було застосовувати процедуру додавання феромонів для вузлів, що входять у дане дерево рішень DT .

Таким чином, додавання феромонів виконується для кожного терму, що входить у дерево рішень DT (10):

$$\tau_p^q(t) = \tau_p^q(t) + Q_{DT} \cdot \tau_p^q(t), \quad \forall p \in R, \quad \forall R \subset DT, \quad (10)$$

де $\tau_p^q(t)$ – кількість феромонів для терму p у графові пошуку для класу q , що визначається за допомогою відповідного дерева рішень.

Етап 6. Випаровування феромонів. Для виключення гірших термів, тобто таких, які при включенні їх у дерева рішень, знижують якість прогнозування, застосовують процедуру випаровування феромонів, яка виконується наприкінці кожної ітерації та застосовується для всіх вузлів у всіх графах пошуку. Випаровування феромонів виконується відповідно до формули (11)

$$\tau_p^q(t+1) = \rho \cdot \tau_p^q(t), \quad \forall p = \overline{1, T}, \quad q = \overline{1, K}, \quad (11)$$

де ρ – коефіцієнт випаровування, який задається при ініціалізації.

Якщо $t < t_{\max}$, то виконати перехід до етапу перезапуску агентів, а якщо ні, то вважається, що виконано максимально припустиму кількість ітерацій, і виконується перехід до етапу формування результуючого дерева рішень.

Етап 7. Перезапуск агентів. Усі дані про переміщення агентів у всіх графах пошуку обновляються, агенти розміщуються у випадковій точці графів пошуку, після чого виконується перехід до другого етапу.

Етап 8. Формування результуючого дерева рішень. Краще знайдене дерево рішень модифікується за допомогою традиційного мультиагентного методу з непрямым зв'язком між агентами. При цьому створюється граф пошуку з вузлів, що входять в обране краще дерево рішень. При цьому евристичними мірами пріоритетності вузлів є зважені значення феромонів для кожного терму, обчислені на основі отриманих матриць феромонів для кожного лінгвістичного терму вихідної змінної. Після цього виконується пошук агентами з непрямым зв'язком між ними за традиційною схемою мультиагентного пошуку з непрямым зв'язком. На основі отриманих результатів, з дерева рішень віддаляються ребра з найменшою кількістю феромонів, що дозволяє підвищити інтерпретабельність та логічну прозорість сформованого дерева рішень.

Запропонований мультиагентний метод ідентифікації дерев рішень з непрямим зв'язком між агентами було програмно реалізовано у середовищі пакету Matlab 7.0.

Для експериментів використовувалися тестові дані, які були взяті з загальнодоступних репозиторіїв [10]. Розроблений метод порівнювався з методом CART [5]. Були отримані дерева рішень, які характеризувалися точністю класифікації 81,2% і 86,1% для методу CART та розробленого мультиагентного методу синтезу дерев рішень з непрямим зв'язком між агентами, відповідно.

Виходячи з отриманих результатів, можна відзначити, що запропонований мультиагентний метод ідентифікації дерев рішень забезпечує синтез дерев рішень, які дозволяють виконувати класифікацію з більшою точністю, ніж у випадку синтезу дерев рішень з використанням існуючих методів.

Висновки. У роботі вирішено завдання синтезу дерев рішень.

Наукова новизна роботи полягає в тому, що розроблено мультиагентний метод ідентифікації дерев рішень, який використовує при пошуку декілька графів пошуку для кожного класу об'єктів, що дозволяє синтезувати дерева рішень з високою точністю класифікації.

Практична цінність отриманих результатів полягає в тому, що запропонований мультиагентний метод синтезу дерев рішень був програмно реалізований, і може бути використаний для розв'язання практичних завдань класифікації та прогнозування.

На основі результатів проведених експериментів можна зробити висновок, що запропонований метод ідентифікації дерев рішень дозволяє синтезувати дерева рішень з високими узагальнюючими властивостями, а також дозволяє уникати надлишкового розширення дерева.

Список літератури: 1. *Russel C.* Искусственный интеллект: современный подход / *C. Russel, П. Норвиг.* – М.: Вильямс, 2006. – 1408 с. 2. *Субботін С.О.* Неітеративні, еволюційні та мультиагентні методи синтезу нечіткологічних і нейромережних моделей: монографія / *С.О. Субботін, А.О. Олійник, О.О. Олійник;* під заг. ред. *С.О. Субботіна.* – Запоріжжя: ЗНТУ, 2009. – 375 с. 3. *Rokach L.* Data Mining with Decision Trees. Theory and Applications / *L. Rokach, O. Maimon.* – London: World Scientific Publishing Co, 2008. – 264 p. 4. Classification and regression trees / *L. Breiman, J.H. Friedman, R.A. Olshen, C.J. Stone.* – California: Wadsworth & Brooks, 1984. – 368 p. 5. *Quinlan J.R.* Simplifying decision trees / *J.R. Quinlan* // International Journal of Man-Machine Studies. – 1987. – № 27 (221). – P. 221–234. 6. *Quinlan J.R.* Decision trees and decision making / *J.R. Quinlan* // IEEE Transactions on Systems, Man and Cybernetics. – 1990. – № 2 (20). – P. 339–346. 7. *Quinlan J.R.* Induction of decision trees / *J.R. Quinlan* // Machine Learning. – 1986. – № 1. – P. 81–106. 8. Прогресивные технологии моделирования, оптимизации и интеллектуальной автоматизации этапов жизненного цикла авиадвигателей: монографія / *А.В. Богуслаев, Ал.А. Олейник, Ал.А. Олейник, Д.В. Павленко, С.А. Субботин.* – Запоріжжя: ОАО "Мотор Сич", 2009. – 468 с. 9. *Dorigo M.* Optimization, Learning and Natural Algorithms / *M. Dorigo.* – Milano:

Politecnico di Milano, 1992. – 140 p. **10.** *UCI Machine Learning Repository* [electronic resource] / Center for Machine Learning and Intelligent Systems. – Access mode: <http://archive.ics.uci.edu/ml/datasets.html>.

Статья представлена д.т.н. проф. ЗНУ Гоменюком С.І.

УДК 004.93

Агентный метод синтеза деревьев решений / Гофман Е.А., Олейник А.А., Субботин С.А. // Вестник НТУ "ХПИ". Тематический выпуск: Информатика и моделирование. – Харьков: НТУ "ХПИ". – 2011. – № 17. – С. 16 – 25.

Рассмотрена задача построения моделей в виде деревьев решений. Проанализирован мультиагентный метод с косвенной связью между агентами. Предложен мультиагентный метод идентификации деревьев решений. Разработано программное обеспечение, реализующее предложенный метод. Библиогр.: 10 назв.

Ключевые слова: агент, дерево решений, модель, мультиагентный метод.

UDC 004.93

Agent-based method for the synthesis of decision trees / Gofman E.A., Oleynik A.A., Subbotin S.A. // Herald of the National Technical University "KhPI". Subject issue: Information Science and Modelling. – Kharkov: NTU "KhPI". – 2011. – №. 17. – P. 16 – 25.

The problem of constructing models in the form of decision trees is considered. Multiagent method of indirect communication between agents is analyzed. Multiagent method of decision trees identification is proposed. The software that implements the proposed method is developed. Refs.: 10 titles.

Key words: agent, decision tree, model, multi-agent method.

Поступила в редакцію 20.02.2011