

## МУЛЬТИАГЕНТНИЙ МЕТОД ПОШУКУ ЧИСЛОВИХ ДАНИХ В ГЛОБАЛЬНІЙ МЕРЕЖІ ІНТЕРНЕТ

Піговський Ю.Р.<sup>1)</sup>, Парій Р.П.<sup>2)</sup>

*Тернопільський національний економічний університет*  
<sup>1)</sup> к.т.н., доцент; <sup>2)</sup> магістр

### I. Постановка проблеми

Пошук сторінок в мережі Інтернет за певними ознаками є важливою задачею, що має практичну цінність. Проблема пошуку сторінок за змістом представленої у документах числової інформації є недостатньо вивчена, а тому розробка методів її розв'язання є актуальною.

Розглянемо постановку задачі пошуку числових даних на множині сторінок-документів як задачу максимізації деякої функції корисності  $f_U(P_1, P_2 \dots P_N)$  шляхом варіювання набору параметрів її шляху  $P_1, P_2 \dots P_N$ :

$$f_U(P_1, P_2 \dots P_N) \xrightarrow{P_1, P_2 \dots P_N} \max, \quad (1)$$

де кожен параметр  $P_i$  у стрічці параметрів  $P_1, P_2 \dots P_N$  відповідає одній з компонент URL-адреси і може приймати значення з деякої наперед-визначеної множини можливих варіантів:

$$P_i \in \Omega_i, \quad (2)$$

де  $\Omega_i$  – множина варіантів реалізації  $i$ - того компонента URL-адреси. Наприклад, множини субдоменів, доменів (у випадку пошуку в межах одного сайту може складатися з єдиного елемента) та сторінок можуть бути описані так:

$$P_1 \in \Omega_1 = P_1 \in \left\{ \begin{array}{l} \text{"olatheroofing"}, \\ \text{"raytownddoors"}, \\ \text{"olathewindows"}, \\ \dots \\ \text{"www"} \end{array} \right\}, \quad P_2 = \text{"site.com"}, \quad P_3 \in \Omega_3 = P_3 \in \left\{ \begin{array}{l} \text{"",} \\ \text{"design.php"}, \\ \text{"doors.php"}, \\ \dots \\ \text{"windows.php"} \end{array} \right\}. \quad (3)$$

Вибираючи елементи по-одному з кожної множини і компонуючи їх отримуємо URL-адреси такого вигляду:

<http://olatheroofing.totalrenovationgroup.com/>

...

<http://www.totalrenovationgroup.com/windows.php>

Не усі з наведених URL-адрес присутні, завантаження деяких з них може призводити до помилки відсутності сторінки 404, що відповідатиме нульовому значенню функції корисності  $f_U(P_1, P_2 \dots P_N) = 0$ .

### II. Мета роботи

Метою дослідження є розробка мультиагентного методу пошуку Інтернет-сторінок, що базується на алгоритмах поведінкової оптимізації, зокрема, алгоритмі бджолиного рою (Bees Algorithm).

### III. Метод мультиагентного пошуку

Задача пошуку (1)–(3) є оптимізаційною задачею на максимум, що характеризується присутністю великої кількості локальних екстремумів. Функція корисності може мати вигляд алгебричного запису або імітаційного алгоритму. Другий випадок на практиці спостерігається найчастіше, що унеможливає знаходження похідної від (1). Наслідком цього є непридатність методів градієнтного пошуку для її розв'язання.

З літератури [1, 2] відомо, що при розв'язанні оптимізаційних задач з мультимодальною і недиференційованою функцією мети ефективними є алгоритми еволюційної і поведінкової оптимізації. В даній роботі адаптовано алгоритм бджолиного рою (Bees Algorithm) [3] для розв'язання задачі (1)–(3). Розглянемо задачу глобальної умовної оптимізації

$$\max F(X) = F(X^*) = F^*, X \in D \subset R^k, \quad (4)$$

де  $X = (x_1, x_2, \dots, x_k)$  - вектор варійованих параметрів,

$$D = \{X \mid x_i^- \leq x_i \leq x_i^+, i \in [1:k]\} - \text{множина допустимих значень цього вектору.}$$

Позначимо бджолиний рій  $B = \{B_a, a \in [1:Z]\}$ , де  $B_a$  - бджола (агент),  $Z$  - число агентів в рою. Положення бджоли  $B_a$  в момент часу  $t = 0, 1, \dots$  визначається вектором її координат  $X_{a,t} = (x_{a,t,1}, x_{a,t,2}, \dots, x_{a,t,k})$ . Нехай  $B_\beta^S \notin B, \beta \in [1:S]$  - бджола-розвідник;  $S < Z$ .

Схема використовуваного варіанту методу рою бджіл має наступний вигляд.

На першому кроці методу в точці з випадковими координатами  $B_{\beta,0} \in D$ , відправляються бджоли-розвідники. В залежності від значень цільової функції  $F(X)$  в цих точках, в області  $D$  виділяються два типи ділянок (підобластей)  $d_\beta$ :  $n$  кращих ділянок, які відповідають найбільшим значенням цільової функції;  $m$  перспективних ділянок, які відповідають значенням цільової функції, найбільш близькими до найкращих значень.

Підобласть  $d_\beta$ ,  $\beta \in [1:s]$  називається підобластю локального пошуку і представляє собою гіперкуб в просторі  $R^k$  з центром в точці  $X_{\beta,0}$  і довжинами сторін, рівними  $2\Delta$ . Тут  $\Delta$  - параметр, який позначає розмір області локального пошуку.

Якщо виявилось, що евклідова відстань  $\|X_{\beta,0} - X_{\gamma,0}\|$  між двома агентами-розвідниками  $B_\beta^S, B_\gamma^S \in B, \beta \neq \gamma$  менше деякої фіксованої довжини, то можливі два наступних варіанти методу, із яких в роботі використовується другий варіант: поставити в відповідність цим агентам дві різних ділянки які перетинаються  $d_\beta, d_\gamma$  (найкращих і/або перспективних); поставити в відповідність тим же агентам одну ділянку, центр якого знаходиться в точці, яка відповідає агенту з великим значенням цільової функції.

В кожній із найкращих і перспективних ділянок відправляється по  $N$  і по  $M$  агентів, відповідно. Координати цих агентів в вказаних ділянках визначаються випадковим чином.

Відмітимо, що варіантом розглянутого рішення є відправка в вказані під області не фіксовану кількість агентів, а кількість, пропорційну відповідним значенням цільової функції. Розміри під областей, в які відправляються агенти, можуть зменшуватися з ростом числа ітерацій з тим, що в кожній підобласті рішення зводилось до локального максимуму цільової функції в цій підобласті.

На основі аналізу значень функції, яка відповідає всім агентам рою, після деякого числа ітерацій знаходяться  $n$  нових найкращих і  $m$  нових перспективних ділянок. В якості критерію закінчення ітерацій можна використовувати досягнення заданої кількості ітерацій  $T$ . Можна також закінчувати ітерації, якщо на протязі  $\tau \leq T$  ітерацій не вдалося збільшити максимальне значення цільової функції. Тут  $\tau$  - параметр зупинки.

### Висновки

У роботі поставлено задачу пошуку числових даних в мережі Інтернет і описано основні положення мультиагентного методу її розв'язання. Метод базується на алгоритмах поведінкової оптимізації бджолиного рою (Bees Algorithm). Реалізація цього методу дасть змогу покращити відповідність знайдених документів критеріям пошуку та скоротити тривалість його виконання.

### Список використаних джерел

2. Рутковская Д. Нейронные сети, генетические алгоритмы и нечеткие системы / Д. Рутковская, М. Пилиньский, Л. Рутковский ; [пер. с польск. И. Д. Рудинского]. – Москва : Горячая линия – Телеком, 2006. – 452 с.
3. Dorigo M. Ant system: optimization by a colony of cooperating agents / M. Dorigo, V. Maniezzo, A. Colomi // IEEE Trans. on Systems, Man, and Cybernetics. – 1996. – Part B: Cybernetics. – № 26 (1). – P. 29–41.
4. Гришин А. А. Исследование эффективности метода пчелиного роя в задаче глобальной оптимизации / А.А.Гришин, А.П.Карпенко // Наука и Образование №8, 2010. – 28 с. [Електронний ресурс]. — Режим доступу <http://technomag.edu.ru/doc/154050.html>