

УДК 519.6 : 519.688 : 004.7

М.Д. ГОДЛЕВСЬКИЙ, О.О. АБАБІЛОВ

Національний технічний університет «ХПІ», Україна

ДОСЛІДЖЕННЯ ЕФЕКТИВНОСТІ ПАРАЛЕЛЬНИХ ГЕНЕТИЧНИХ АЛГОРИТМІВ ДЛЯ ВИРІШЕННЯ ЗАДАЧІ СТВОРЕННЯ РОЗКЛАДУ ЗАНЯТЬ ВУЗУ НА БАЗІ GRID-СИСТЕМИ

Досліджується ефективність застосування паралельних генетичних алгоритмів, що виконуються на Grid-системі, для вирішення задачі створення розкладу занять вузу. Задача, що розглядається, відноситься до класу задач багатокритеріальної оптимізації та є NP-повною. Застосування генетичних алгоритмів спрощує підтримку численних вимог до розкладу та дозволяє легко додавати нові. Розглядається вплив генетичних операторів (відбір, схрещування та ін.) та їхніх налаштувань на якість отриманої популяції. Досліджується ефективність паралельних генетичних алгоритмів двох типів – із використанням острівної моделі та динамічних демів. Розглядається вплив кількості процесів на швидкість генетичного пошуку. Визначаються умови доцільності застосування генетичного алгоритму із динамічними демами для задачі пошуку розкладу.

Ключові слова: NP-повні задачі, багатокритеріальна оптимізація, генетичні алгоритми, паралельні обчислення, Grid-системи, ефективність алгоритмів, складання розкладів.

Вступ

Організація навчального процесу у вузі є основою підготовки висококваліфікованих фахівців. Базисом навчального процесу є розклад занять. У НТУ «ХПІ» складання розкладу дотепер має потребу в автоматизації й тому залишається однією з актуальних проблем.

Задача складання розкладу занять відноситься до класу NP-повних [1]. Це означає, що для її вирішення дотепер не знайдено поліноміального алгоритму й імовірно, що його взагалі не існує. Один з підходів до вирішення задачі складання розкладу, який має задовольняти великій кількості нетривіальних умов, полягає у використанні генетичних алгоритмів [2].

Програма, що реалізує генетичний алгоритм для вирішення задачі складання розкладу занять, має потребу в значних ресурсах: процесорному часі, оперативній й енергонезалежній пам'яті, тому для вирішення задачі доцільно використовувати Grid-систему. У статті [3] описано представлення задачі складання розкладу, структуру особини й три типи паралельних генетичних алгоритмів для вирішення задачі складання розкладу; виділено параметри, які впливають на ефективність описаних алгоритмів.

У даній роботі представлено дослідження ефективності застосування паралельних генетичних алгоритмів. Дослідження ефективності застосування паралельних генетичних алгоритмів проводилося одночасно у двох різних областях. Перша з них – це дослідження різних операторів (мутація, схрещування та ін). Друга мета дослідження – вивчення різних ти-

пів паралельних генетичних алгоритмів: алгоритму, що використовує острівну модель (MultiDeme) і алгоритму з динамічними демами (DynamicDeme).

У статті використовуються найменування відомих стратегій генетичних операторів, які докладно описано в [4].

Постановка задачі дослідження

Необхідно дослідити залежність ефективності паралельних генетичних алгоритмів для вирішення задачі створення розкладу занять вузу від налаштувань. Розклад має задовольняти наступним особливостям організації учебного процесу:

- певні викладачі або групи не можуть мати заняття у певні моменти часу;
- певні аудиторії є недоступними у певні моменти часу;
- для певних занять аудиторія повинна мати певну властивість;
- «вікна» заборонені;
- для певних занять групи об'єднуються в потоки або розділяються на підгрупи;
- заняття бажано починати з ранку;
- розрізняють розклади першого та другого тижня, але ці розклади мають бути схожі;
- групи та викладачі не повинні протягом дня переходити між віддаленими корпусами;
- навантаження викладачів та/або груп має бути розподілене рівномірно по тижню.

Отже, маємо задачу багатокритеріальної оптимізації.

Представлення задачі складання розкладу

Введемо спочатку набір визначень. Ресурс – це кожний з об'єктів, для яких складається розклад: викладач; клас (академічна група, підгрупа чи потік); аудиторія; відрізок часу (тиждень, день та пара). Позначимо i -й тип ресурсу як $T_i = \{R_{ik}\}$, $k = \overline{1, p_i}$, $i = \overline{1, 4}$, де p_i – кількість ресурсів типу T_i , R_{ik} – окремі об'єкти типу ресурсу T_i . Кожне з n занять розкладу представляється кортежем, що містить по одному об'єкту кожного з типів ресурсів. Для кожного заняття E_j , $j = \overline{1, n}$ задано клас та викладача, а час та місце повинні бути визначені в процесі пошуку розв'язку, і саме вони відповідають генам єдиної хромосоми індивіда.

Кожна вимога до задачі складання розкладу – це окрема часткова штрафна функція, яка розраховується для індивіда. Фітнес-функція (зважена сума цих функцій) має бути мінімізована генетичним алгоритмом. Отже, підтримувати наявні та додавати нові критерії порівняно просто.

Вплив обраних стратегій на генетичний пошук

Для досліджень узятя наступна задача складання розкладу. Є 25 викладачів, які мають провести 496 занять (лекційні, лабораторні та практичні) з 15 групами, які об'єднані для лекцій у 5 потоків, та 16 аудиторій, з яких тільки 4 можуть вмістити будь-який потік, і лише 6 підходять для лабораторних занять. Аудиторії розподілено порівну між двома віддаленими корпусами. Студент або викладач не може мати більше 4 занять на день.

У ряді задач, для яких застосовуються генетичні алгоритми, оператор відбору мало впливає на загальну продуктивність і динаміку популяції. Однак, оскільки заздалегідь передбачити ефективність тих або інших операторів для даної задачі в загальному випадку не можна, було досліджено кілька різних схем відбору: roulette wheel, rank, tournament, uniform (рис. 1).

Кращі результати при інших рівних налаштуваннях дає rank, але він робить популяцію занадто однорідною, тому можливо потрапляння до локального оптимуму. Через це рекомендовано застосовувати tournament.

У роботі було досліджено кілька операторів схрещування: even odd, one point, two point, uniform (рис. 2, 3). Частина результатів представлена на рис. 2. Відзначено, що в цілому продуктивність two point краще, ніж в one point, але розходження не-

ликі. Однак у порівнянні із even odd та uniform, two point та one point дають відчутно кращий результат.

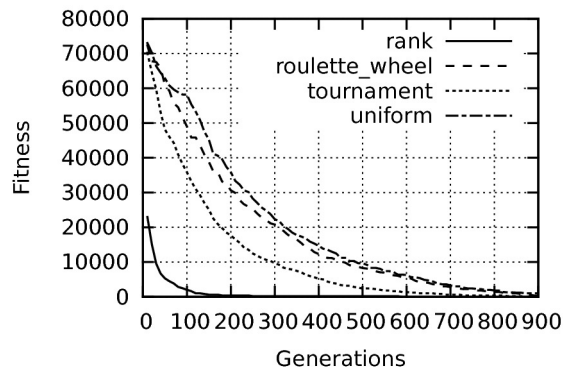


Рис. 1. Вплив оператора відбору на генетичний пошук

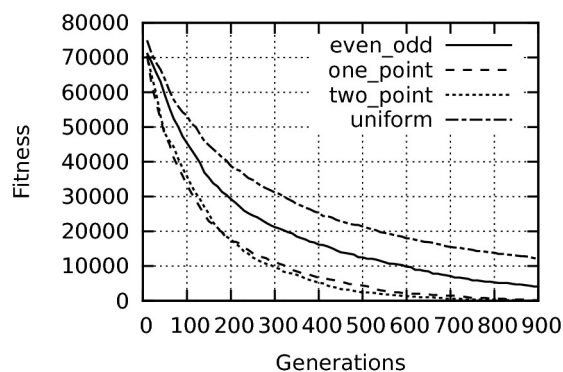


Рис. 2. Вплив оператора схрещування на генетичний пошук

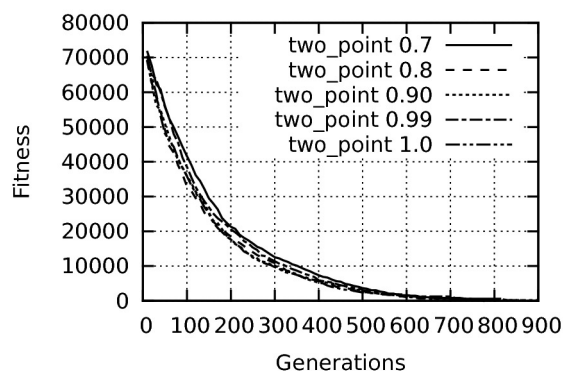


Рис. 3. Вплив частоти схрещувань на генетичний пошук

Всі ці оператори мало змінюють популяцію, коли вона стає одноманітною. З іншого боку, оператори схрещування грають дуже важливу роль на початку пошуку, коли популяція характеризується більшою розмаїтістю. Експерименти показали, що, навіть якщо рівень схрещувань дуже високий, немає майже ніякого ризику в тім, що пошук зупиниться на локальному оптимумі. Як правило, чим більше

високий рівень схрещувань було використано, тим краща продуктивність була досягнута (див. рис. 3, рівень схрещувань змінювався від 0.7 до 1.0).

Мутація, так само як схрещування, звичайно є дуже впливовим оператором. У задачі пошуку розкладу використовується два види мутації (рис. 4, 5).

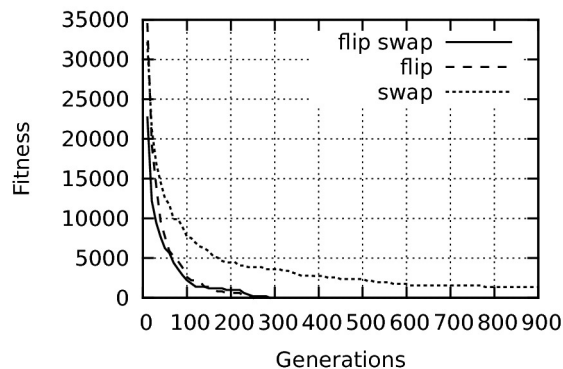


Рис. 4. Вплив операторів мутації на генетичний пошук

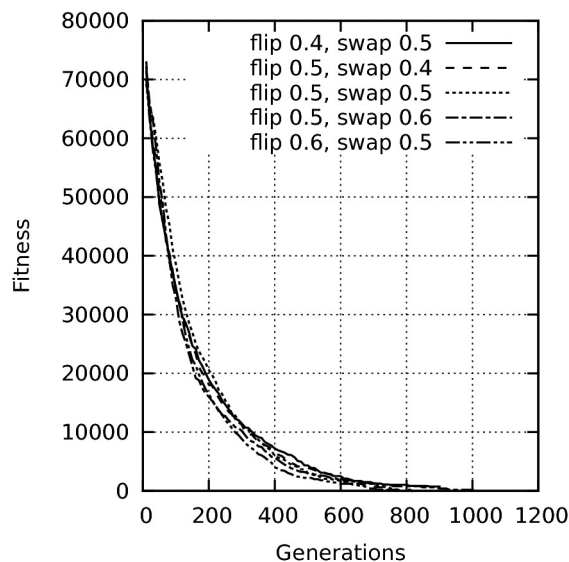


Рис. 5. Вплив частоти мутацій на генетичний пошук

Перший вид – це заміна гена в хромосомі на довільний алель із припустимих (flip). Для роботи такого оператора необхідно почати додаткові дії по визначенню припустимих значень генів у кожній позиції. Тоді для задач пошуку розкладу така мутація має наступний сенс: для даного заняття призначили довільне (із припустимих) місце або час проведення.

Другий вид – це обмін двох генів у хромосомі (swap). Для задач пошуку розкладу це буде мати наступний сенс: у двох заняттях обміняли місце або час проведення.

З рис. 4 видно, що найбільш ефективною є комбінація flip та swap, а найменш – оператор swap самотужки.

Рівень мутацій може сильно вплинути на швидкість пошуку рішення. На рис. 5 представлена статистика для різних рівнів. Мутація проводилася двома описаними вище операторами одночасно. Видно, що найкращими рівнями мутації є 0.6 для swap та 0.5 для flip.

Результати дослідження паралельних генетичних алгоритмів

Основна мета паралелізації генетичного алгоритму полягає в тому, щоб збільшити швидкість і продуктивність пошуку.

Алгоритм, що використовує острівну модель (MultiDeme), вимагає підрозділу популяції на деяке число підпопуляцій (демів). Демі розділяються один від іншого (географічна ізоляція), і індивіди конкурують тільки в межах дема. Час від часу спрацьовує додатковий оператор – міграція: деякі індивіди переміщуються (копіюються) від одного дема до іншого.

Як і передбачалося, збільшення числа запущених процесів призводить до більш швидкого пошуку розв'язку, якщо частота міграції не занадто висока (у проведених дослідженнях – раз у 50 поколінь), що видно з рис. 6 (використано від 2 до 5 процесів). Найбільш надійним способом визначення необхідності міграцій є комбінування умови періоду (раз у кілька поколінь) та умови застою.

Алгоритм із динамічними демками (DynamicDeme) дозволяють уникнути недоліків мікро- та крупнозернистих паралельних генетичних алгоритмів [3, 5]. Цей алгоритм є комбінацією глобального паралелізму із крупнозернистим алгоритмом (модель підпопуляцій, що перекриваються). Немає оператора міграції як такого, тому що популяція під час розвитку обробляється як єдиний набір індивідів.

Основна ідея – скоротити час очікування найбільш повільних індивідів в моделі ведучого-веденого динамічним поділом популяції в демки, які можуть тоді бути оброблені без затримки. Це дає більше ефективності з погляду швидкості обробки. Алгоритм повністю масштабується. Починаючи від глобального паралелізму з розподіленим розрахунком функції придатності, можна збільшити масштаб алгоритму до мікрозернистої версії, з кількома індивідами в межах кожного дема й великою кількістю демів.

Алгоритм DynamicDeme досить складний і використовує чотири типи процесів: Manager, Counter, Master, Slave.

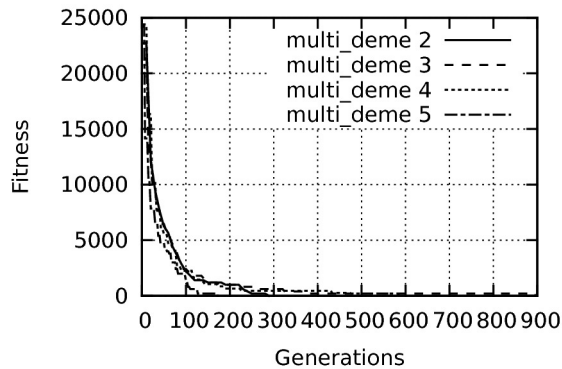


Рис. 6. Вплив кількості процесів MultiDeme на генетичний пошук

З рис. 7 видно, що збільшення кількості процесів з 5 до 6 (при цьому є один процес Master) приводить до уповільнення генетичного пошуку, але це пов'язано з тим, що досліджувалася порівняно проста для обчислення цільова функція.

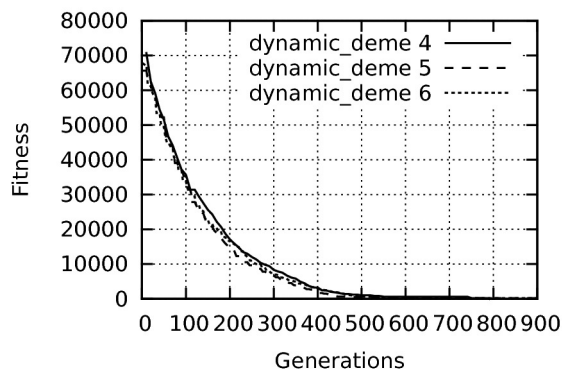


Рис. 7. Вплив кількості процесів DynamicDeme на генетичний пошук

З рис. 8 видно, що для простої цільової функції кількість оброблених за секунду індивідів зменшується при збільшенні кількості процесів як для великої популяції (2000 індивідів), так і для малої (500).

Для порівняння наведено швидкість непаралельного алгоритму (local). Це має місце тому, що час на передачу геномів між процесами порівняний із часом обчислення цільової функції. Напроти, якщо цільова функція складна, то збільшення кількості процесів (наприклад, від 4 до 8 на рис. 9) збільшує і кількість індивідів на секунду, що прискорює генетичний пошук.

Отже, при високо витратній оцінці пристосованості алгоритм DynamicDeme є досить ефективним, щоб використовуватися в задачі складання розкладів.

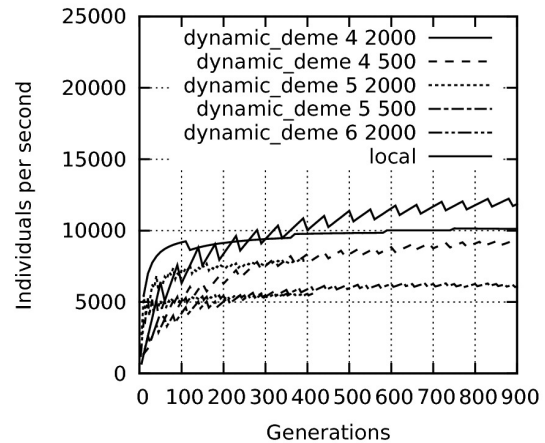


Рис. 8. Вплив кількості процесів на швидкість обробки індивідів для простої функції придатності

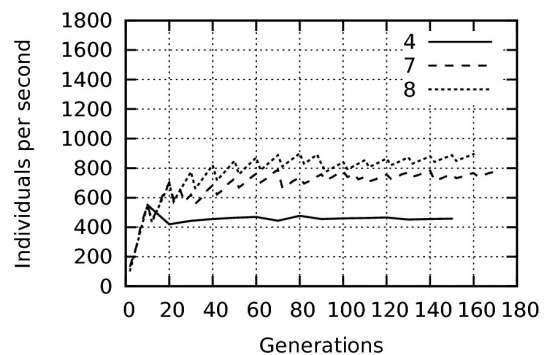


Рис. 9. Вплив кількості процесів на швидкість обробки індивідів для складної функції придатності

Висновки

Проведені дослідження показали, що використання паралелізації генетичного пошуку є ефективним для задачі створення розкладу, приводячи до вищої якості популяції та швидшого пошуку розв'язку. Алгоритм MultiDeme доцільно застосовувати у загальному випадку; для задачі із багатьма одночасними вимогами до розкладу (які роблять оцінку пристосованості високо витратною) також вигідно застосовувати DynamicDeme.

Реалізація алгоритмів дозволяє збільшувати масштаб для даної задачі шляхом збільшення кількості паралельних процесів і виконувати налаштування генетичного алгоритму (число індивідів у межах популяції, вид і рівень мутацій і схрещувань тощо).

Розвитком даної роботи є моделювання роботи паралельних генетичних алгоритмів, виконуваних у Grid-системі.

Література

1. Holland, J.H. *Adaptation in Natural and Artificial Systems [Текст] / J. H. Holland.* – Ann Arbor, MI: University of Michigan Press, 1975. – 228 p.

2. Кормен, Т. *Алгоритмы: построение и анализ [Текст]: пер. с англ. / Т. Кормен, Ч. Лейзерсон, Р. Ривест.* – М.: МЦНМО, 2002. – 960 с.

3. Годлевський, М.Д. *Розробка та налаштування паралельних генетичних алгоритмів для розв'язання задачі створення розкладу занять вузу на основі GRID-системи [Текст] / М.Д. Годлевський, О.О. Абабілов // Вісник Національного технічного*

університету «ХПІ». Збірник наукових праць. Тематичний випуск «Системний аналіз, управління та інформаційні технології». – Х.: НТУ «ХПІ». – 2010. – № 67. – С. 3 – 7.

4. Goldberg, D.E. *Genetic Algorithms in Search, Optimization, and Machine Learning [Текст] / D.E. Goldberg.* – Addison-Wesley Publishing Company, Inc., 1989. – 414 p.

5. Nowostawski, M. *Parallel Genetic Algorithm for Sequencing Optimisation. Mini-project report [Текст] / M. Nowostawski.* – The University of Birmingham, School of Computer Science, 1998. – 33 p.

Надійшла до редакції 2.09.2011

Рецензент: д-р техн. наук, проф., декан факультету інформатики і управління І. П. Гамаюн, Національний технічний університет «Харківський політехнічний інститут», Харків.

ИССЛЕДОВАНИЕ ЭФФЕКТИВНОСТИ ПАРАЛЛЕЛЬНЫХ ГЕНЕТИЧЕСКИХ АЛГОРИТМОВ ДЛЯ РЕШЕНИЯ ЗАДАЧИ СОЗДАНИЯ РАСПИСАНИЯ ЗАНЯТИЙ ВУЗА НА БАЗЕ GRID-СИСТЕМЫ

М.Д. Годлевский, А.А. Абабилов

Исследуется эффективность применения параллельных генетических алгоритмов, выполняемых на Grid-системе, для решения задачи создания расписания занятий вуза. Рассматриваемая задача относится к классу задач многокритериальной оптимизации и является NP-полной. Применение генетических алгоритмов упрощает поддержку многочисленных требований к расписанию и позволяет легко добавлять новые. Рассматривается влияние генетических операторов (отбор, скрещивание и др.) и их настроек на качество полученной популяции. Исследуется эффективность параллельных генетических алгоритмов двух типов – с использованием островной модели и динамических демов. Рассматривается влияние количества процессов на скорость генетического поиска. Определяются условия целесообразности применения генетического алгоритма с динамическими демами для задачи поиска расписания.

Ключевые слова: NP-полные задачи, многокритериальная оптимизация, генетические алгоритмы, параллельные вычисления, Grid-системы, эффективность алгоритмов, составление расписаний.

EFFICIENCY RESEARCH OF PARALLEL GENETIC ALGORITHMS FOR SOLVING UNIVERSITY TIMETABLE SCHEDULING PROBLEM BASING ON A GRID SYSTEM

M.D. Godlevskiy, A.A. Ababilov

The efficiency of parallel genetic algorithms running on Grid system for solving a timetable scheduling problem is investigated. The mentioned problem belongs to the class of multicriteria optimization problems and is NP-complete. Application of genetic algorithms simplifies the support of numerous requirements to schedule and makes it easy to add new ones. Impact of genetic operators (selection, crossover etc.) and their parameters for population quality are reviewed. The efficiency of parallel genetic algorithms of two types – using the island model and dynamic demes – is investigated. The influence of processes on speed of the genetic search is considered. The conditions of appropriate use of dynamic demes genetic algorithm for a timetable scheduling problem are defined.

Key words: NP-complete problem, multicriteria optimization, genetic algorithms, parallel computing, Grid system, efficiency of algorithms, timetable scheduling.

Годлевський Михайло Дмитрович – д-р техн. наук, проф., завідувач кафедри автоматизованих систем управління, Національний технічний університет «Харківський політехнічний інститут», Харків, Україна.

Абабілов Олексій Олександрович – магістр кафедри автоматизованих систем управління, Національний технічний університет «Харківський політехнічний інститут», Харків, Україна, e-mail: ilovegnu-linux@gmail.com.