

ІНФОРМАТИКА ТА МЕТОДИКА ЇЇ ВИКЛАДАННЯ

УДК 519.712

Сенченко А.С., Бобырь А.В.

¹ кандидат физико-математических наук, доцент кафедры алгебры, ГВУЗ «ДГПУ»

² студент 5 курса физико-математического факультета, ГВУЗ «ДГПУ»

e-mail: senchenko@pisem.net

ИСПОЛЬЗОВАНИЕ ГЕНЕТИЧЕСКИХ АЛГОРИТМОВ ДЛЯ РЕШЕНИЯ ЗАДАЧИ О РАСЩЕПЛЕНИИ МНОЖЕСТВА

В работе предложен генетический алгоритм расщепления заданного множества. Этот алгоритм разбивает множество на два подмножества так, чтобы в каждом из них не содержалось целиком ни одно из подмножеств заданного семейства. Предложена программная реализация алгоритма в среде C++.

Ключевые слова: *расщепление, генетический алгоритм, подмножества.*

Вступ

Термин расщепление используется во многих областях науки: в физике существует понятие расщепления ядра и атома; в генетике – понятие расщепления по фенотипу; в психологии – расщепление сознания; в органической химии – расщепление белков, жиров, углеводов. В этих научных направлениях науки под расщеплением понимают выделение каких-то особенностей какого-либо единого объекта. Понятие расщепление существует и в теории алгоритмов. Многие практические задачи эффективно решаются с помощью построения дерева расщепления. Внутренние узлы этого дерева являются атрибутами некоторой базы данных, а каждая его ветвь, идущая от внутреннего узла, отмечена предикатом расщепления, причем каждая запись использует уникальный путь к каждому узлу-решению.

В математике известна задача о расщеплении множества: задан набор подмножеств C множества S ; необходимо определить существует ли такое разбиение множества S на два подмножества S_1 и S_2 , что ни одно подмножество из C не содержится целиком ни в S_1 , ни в S_2 . Эта задача имеет много различных практических применений, рассмотрим некоторые из них:

© Сенченко А.С., Бобырь А.В., 2013

1. *Задача выбора переводчиков.* N человек являются переводчиками с различных иностранных языков, причем для каждого из них задано владение определенными иностранными языками. Некоторая организация нанимает этих переводчиков для обслуживания двух групп туристов; необходимо выделить переводчиков для каждой из этих групп так, чтобы в каждой из групп были переводчики со всех необходимых языков.

2. *Задача распределения парков вагонов.* Некоторая крупная организация владеет парками вагонов, причем каждый вагон может быть использован только для перевозки определенных видов продукции. Необходимо так распределить парк вагонов между двумя заводами этой организации, чтобы из каждого завода можно было осуществить вывоз продукции.

3. *Задача раскрашиваемости графа в два цвета.* [1, 2] В дискретной математике эта проблема известна под названием *задача Эрдеша-Хайнала*. Эта задача решена для частных случаев, а в общем случае остается открытой.

Доказано что задача о расщеплении множества является NP -сложной [3, 5]. Поэтому все алгоритмы являются эвристическими и дают приближенный к оптимальному результат. Большой вклад в решение задачи внесли ученые Р. Эрдеш, Л. Ловас, Л. Лу, Дж. Бек, А.В. Косточка, Д.А. Шабанов. Исследования в данном направлении продолжаются, поскольку используемые методы не всегда позволяют найти оптимальное решение из-за неудовлетворительных показателей быстродействия. Наиболее перспективными с нашей точки зрения являются алгоритмы с использованием методов теории искусственного интеллекта, в первую очередь генетические алгоритмы, поэтому именно их мы и использовали для решения нашей задачи.

Постановка задачи. Задачей работы является разработка генетического алгоритма построения расщепления заданного множества и его программная реализация в среде C++. Исходное множество и семейство подмножеств $C_1 \dots C_k$ в нем задается таблицей. Необходимо разбить заданное множество на два подмножества S_1 и S_2 таким образом, чтобы ни одно из подмножеств семейства $C_1 \dots C_k$ не входило полностью ни в S_1 ни в S_2 . Мощности подмножеств S_1 и S_2 должны быть равными или близкими к равенству. Задаются ограничения на количество элементов множества и количество подмножеств в семействе.

Основная часть

Задача может быть решена с помощью полного перебора вариантов. В этом случае сложность решения задачи равна 2^n , где n – количество элементов множества. Известно что функция $f(n) = 2^n$ при возрастании n

растет очень быстро. Поэтому для множества, состоящего из 60 элементов, решение перебором вариантов найти за реальное время невозможно.

Одним из подходов к решению нашей задачи может являться эволюционное моделирование. Этот подход использует принцип природной эволюции процесса. Он и послужил основой для генетических алгоритмов.

Рассмотрим основные понятия, связанные с генетическими алгоритмами. Хромосому – определенную последовательность символов – будем считать отдельным вариантом решения задачи. Каждый входящий в хромосому символ будем называть геном.

Генетический алгоритм определенным или случайным образом генерирует начальную популяцию хромосом. Будем считать, что вся популяция состоит из q хромосом. Затем с помощью фитнес-функции проверяем каждую хромосому насколько хорошо она подходит для решения задачи. Из начальной популяции выбираем $\frac{q}{2}$ хромосомы по наилучшим значениям фитнес-функции. Эти хромосомы переходят в следующее поколение. Из выбранных $\frac{q}{2}$ хромосом получаем еще $\frac{q}{2}$ хромосом путем применения операций кроссинговера, сдвига, мутации.

Мутация – такое преобразование хромосомы, в ходе которого случайным образом изменяется один или несколько генов. Наиболее распространенный вид мутации – случайное изменение одного из генов хромосомы.

Кроссинговер – операция в ходе которой из двух хромосом порождаются две новые хромосомы. Одноточечный кроссинговер работает таким образом. Сначала, случайным образом, выбирается одна из точек разрыва. Обе родительские структуры разбиваются на два сегмента по этой точке. Затем соответствующие сегменты разных родителей склеиваются по точке разрыва и получаем две хромосомы наследников.

Сдвиг состоит в циклическом изменении позиции всех генов на определенное количество единиц. Работа генетического алгоритма это последовательное приближение и проверка условия достижения искомого результата.

Алгоритм выполняется до тех пор, пока не пройдет определенное число поколений или не будет получен любой другой критерий остановки, в противном случае работа алгоритма будет продолжаться.

Определение операций генетического алгоритма для поставленной задачи

В нашей задаче рассматривается множество из n элементов, каждый из которых входит хотя бы в одно из подмножеств элементов $C_1 \dots C_k$. Информация о вхождении или не вхождении каждого элемента множества в подмножество $C_1 \dots C_k$ задаем таблицей.

На рисунку 1 зображено зовнішній вигляд головного вікна програми.

Всі хромосоми є послідовностями довжини n , що складаються з нулів і одиниць, кожен ген відповідає за розміщення відповідного елемента в то чи інше підмножество: значення i -го гена 0 означає, що елемент i розміщується в S_1 ; значення 1 – що елемент i розміщується в S_2 . Таким чином кількість нулів в хромосомі відповідає кількості елементів в підмножині S_1 .

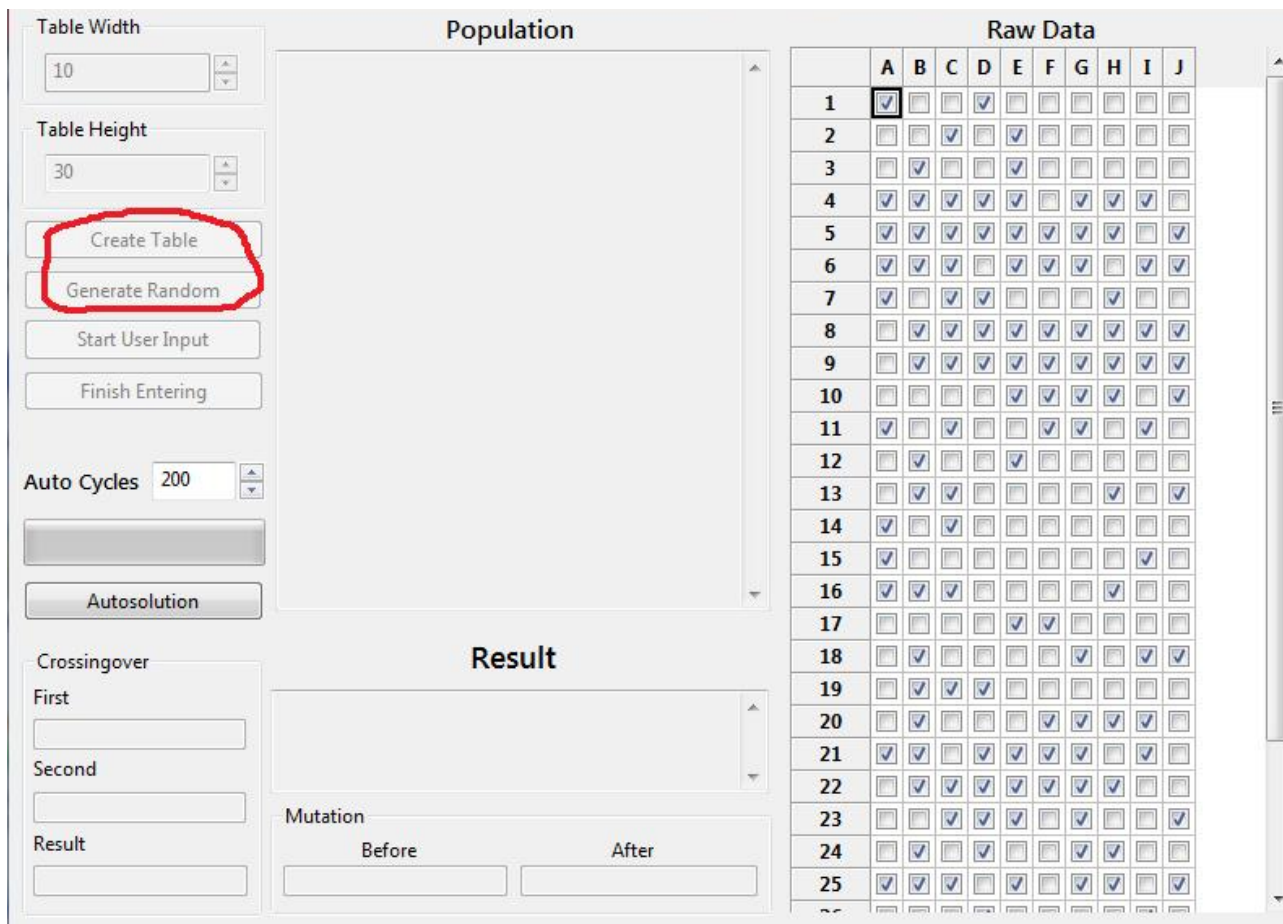


Рис. 1: Інтерфейс вид головного вікна програми

Фітнес-функція задається як кількість підмножин $S_1 \dots S_k$ цілком входять в S_1 або S_2 . Розв'язком задачі буде хромосома, для якої значення фітнес-функції дорівнює нулю.

В наступне покоління переходять хромосоми з найменшим значенням фітнес-функції. Для отримання хромосом наступного покоління застосовуються операції зсуву, мутації та одноточкового кроссинговера.

Вручну встановлюємо максимальну кількість поколінь (до 500 хромосом). При цьому, якщо в одному поколінні буде декілька однакових хромосом, то всі повторення хромосом ліквідируються за допомогою примусового

применения к ним мутации.

Разработана программная реализация алгоритма в среде программирования C++. По результатам тестирования программы найдены оптимальные параметры генетических операций: мутация с вероятностью 0,18 – 0,25, сдвиг 0,17 – 0,2, кроссинговер 0,55 – 0,65.

Наша программа с этими параметрами находит расщепление множества с вероятностью 0,68, причем наличие решение для каждого случая изначально неизвестно, поэтому неблагоприятным считается также случай, при котором расщепление невозможно.

Выводы

В работе предложен генетический алгоритм решения задачи расщепления множества. Рассмотрен вариант кодирования исходных данных, оценка фитнес-функции и способы выполнения генетических операций. Разработана программная реализация алгоритма, выбраны оптимальные, на наш взгляд, вероятности генетических операций.

Дальнейшие исследования должны рассматривать особенности выполнения генетических операций и их вероятность для отдельных случаев вида семейства подмножества, в частности для случая когда подмножества в семействе отличаются друг от друга очень слабо.

Литература

- [1] Райгордский А.М. Задача Эрдеша-Хайнала о раскраске гиперграфов, ее обобщение и смежные проблемы / А.М. Райгордский, Д.А. Шабанов // Успехи математических наук. — 2011. — Т. 66, № 3. — С. 109 – 182.
- [2] Шабанов Д.А. Раскраски гиперграфов / Д.А. Шабанов // Труды МФТИ. — 2012. — Т. 4, № 1. — С. 131 – 140.
- [3] Гери М. Вычислительные машины и трудно решаемые задачи / М. Гери, Д. Джексон. — М.: Мир, 1982. — 416 с.
- [4] Воронский Г.К. Генетические алгоритмы, искусственные нейронные сети и проблемы виртуальной реальности / Г.К. Воронский. — Х.: Основа, 1997. — 112 с.
- [5] Алгоритмы: построение и анализ / [Т. Кормен, Ч. Лейзерсон, Р. Ривест, К. Штайн]. — М.: Вильямс, 2005. — 1296 с.
- [6] Рутковская Д. Нейронные сети, генетические алгоритмы и нечеткие системы / Д. Рутковская, М. Пилинский, Л. Рутковский. — М.: Горячая линия, 2006. — 452 с.