

КОМБИНАТОРНЫЙ ПОИСК ПЕРВИЧНЫХ УЗЛОВ В ОРИЕНТИРОВАННЫХ АЦИКЛИЧЕСКИХ ГРАФАХ

Представлен алгоритм поиска первичных узлов в разреженном ориентированном ациклическом графе при наличии произвольных дуг на основе обратного обхода графа и решении задачи о сумме подмножеств с оптимизацией кодами Грея

Введение

В задачах финансового мониторинга (AML) [1] часто возникает необходимость определения первичных узлов при наличии избыточных связей. Работа посвящена комбинаторному алгоритму поиска множества уникальных путей в ориентированных ациклических графах, основанному на решении задачи о сумме подмножеств

1. Модель и алгоритм поиска

Пусть дан взвешенный ориентированный ациклический граф (DAG) $G = (V, E)$ не имеющий кратных дуг, $\deg^-(v) \leq 16$. Каждой дуге $e \in E$ сопоставлено значение веса $A_e > 0$. Для целевого узла v_e задана сумма $A_{out}(v_e)$. Узел v допустим, если

$$\sum_{e \in E'_{in}(v)} A_e = A_{out}(v), \quad E'_{in} \subseteq E_{in}(v)$$

где $A_{out}(v)$ – сумма весов дуг, ведущих к уже включенным в путь узлам. Алгоритм возвращает мультимножество из K первичных узлов. Кратность узла в наборе равна числу построенных от него уникальных путей к v_e . Поиск ограничен первыми найденными решениями.

Для оптимизации обработки графов объемом $|V| \approx 2 \cdot 10^6$ применены следующие решения:

- Использование CSR-представления графа
- Решение задачи о сумме подмножеств с предварительным расчётом словаря сумм-масок кодами Грея [2] за $O(2^k)$
- Предварительный фильтр (обратный BFS), отбрасывающий недостижимые в обратном графе узлы

II. Верификация и анализ производительности

Для оценки алгоритма использованы синтетические DAG до $2 \cdot 10^6$ узлов. Структура генерировалась по нормальному распределению с последующим добавлением произвольных узлов и межуровневых связей.

Таблица 1 – Результаты тестирования алгоритма на различных масштабах данных

Параметры графа (N/E)	Ср. время, ms	Ср. потребление памяти, Mb
$2 \cdot 10^6 / 2.5 \cdot 10^6$	2311	1117
$2 \cdot 10^5 / 2.5 \cdot 10^5$	236.6	140.6
$2 \cdot 10^4 / 2.5 \cdot 10^4$	20.6	11.6

Заключение

Разработанный алгоритм позволяет восстанавливать мультимножества первичных узлов в разреженных DAG до $2 \cdot 10^6$ узлов за 2.3 с. Полученные характеристики производительности и потребления памяти (1.1 Гб) подтверждают применимость выбранного комбинаторного подхода для графов большой размерности.

Список литературы

1. Li, X. FlowScope: Spotting Money Laundering Based on Graphs // AAAI-20: proc. of the 34th AAAI Conf. on Artif. Intell. – 2020. – P. 4733.
2. Кнут, Д. Э. Искусство программирования. Т. 4, вып. 3: Генерация всех сочетаний и разбиений. – М.: Вильямс, 2007. – С. 17–19.

Синькевич Ярослав Игоревич, студент 1 курса факультета информационной безопасности Белорусского государственного университета информатики и радиоэлектроники, sinkevich.yraslau@gmail.com

Научный руководитель: Кукин Дмитрий Петрович, Научный руководитель: Кукин Дмитрий Петрович, доцент кафедры вычислительных методов и программирования БГУИР, kudin@bsuir.by.