УДК 004.822

ПРОБЛЕМА ПОИСКА В ГРАФОВЫХ СТРУКТУРАХ

Лашкевич К.В.
Белорусский государственный университет информатики и радиоэлектроники
E-mail: nos19@mail.ru

Abstract. *PROBLEM OF INFORMATION RETRIEVAL IN GRAPH STRUCTURES. This paper describes the problem of information retrieval in graph structures (particularly, subgraph isomorphism problem) and compares popular algorithms (Ullman`s algorithm for subgraph isomorphism, VF2 algorithm). There are given some features and limitations of mentioned algorithms.*

Название работы – «Проблема поиска в графовых структурах» – охватывает обширное понятие информационного поиска. В контексте предметной области (теория графов) задача информационного поиска может быть сведена к поиску подструктур данных в системе по некоторым указанным критериям. Фактически, данная задача сводится к поиску изоморфных подграфов в графе по заданному шаблону. В данном случае граф – это данные, существующие в системе.

Пусть дан граф G = (V, E). *Подграфом* L = (Y, Z) заданного графа G = (V, E) называется граф, который удовлетворяет условиям [1]: 1. $Y \subset V$; $Z \subset E$; 2. Если $a, b \in Y$ и $(a, b) \in E$, то $(a, b) \in Z$.

Графы G и G' называются *изоморфными*, если между множествами V и V' их вершин можно установить взаимно однозначное соответствие $\leftrightarrow$ [1], сохраняющее отношение смежности вершин, т. е. такое, что для любых $x, y \in V$ и соответствующих им вершин $x', y' \in V'$ ($x \leftrightarrow x'$, $y \leftrightarrow y'$) имеет место $\widetilde{xy} \in E \Leftrightarrow \widetilde{x'y'} \in E'$, где $\widetilde{xy}$ означает смежность вершин x и y. При этом само соответствие называется *изоморфизмом графов.*

Проблема изоморфного подграфа – вычислительная задача, в которой два граф *G* и *H* приведены в качестве входных данных, и нужно определить, содержит ли *G* подграф, изоморфный графу *H*. Задача поиска изоморфного подграфа является обобщением задачи поиска максимальной клики и задачи определения наличия в графе гамильтонова цикла, и, следовательно, относится к NP-полным задачам [2].

Формальный вопрос: Даны два графа: G = (V, E), H = (V', E'). Существует ли подграф $G_0 = (V_0, E_0) : V_0 \subseteq V$, $E_0 = E \cap (V_0 \times V_0)$, такой что $G_0 \simeq H$?

Самым популярным алгоритмом поиска изоморфного подграфа является **алгоритм Ульмана (1976)**. [3]. Хотя время его работы в общем случае экспоненциальное, он находит решение за полиномиальное время для любого фиксированного выбора H (с полиномом, зависящим от выбора H). Когда G является планарным графом и H фиксирован, время работы данного алгоритма можно свести к линейному.

Можно изобразить изоморфизм графов в виде $|V_H|\times|V_G|$ матрицы M, в которой каждая строка содержит ровно одну 1, а каждый столбец содержит не более одной 1. Значение элемента m_{ij} данной матрицы будет равно 1 тогда и только тогда, когда $v_j \in G$ соответствует $v_i \in H$. Тогда $H = M\,(MG)^T$, H и G – матрицы смежности соответствующих графов. Если мы не ищем индуцированный подграф, $P \leq M(MG)^T$ (покомпонентно), т.е. подграф графа G выбранный по матрице M может содержать дополнительные ребра (дуги), отсутствующие в H. Алгоритм работает путем систематического перебора возможных матриц M и проверки, отражают ли изоморфизм подграфа графа G и графа H. Начнем с создания матрицы M_0 размерности $|V_H|\times|V_G|$, элемент m_{ij0} которой будет равняться 1, если возможно соответствие $v_i \leftrightarrow v_j$ (j-я вершина графа G соответствует i-ой вершине H при данном изоморфизме). Будем использовать только степень вершин в качестве критерия, т.е. мы можем отобразить вершину v_i на вершину v_j в том случае, если последняя имеет достаточно достаточное количество смежных вершин: $m_{ij0} = 1 \Leftrightarrow \deg(v_i) \leq \deg(v_j)$.

Теперь все, что нужно сделать, это проверить все матрицы M, которые можно получить из M_0, путем удаления всех единиц в каждой строке, кроме одной при условии наличия максимум одной единицы в каждом столбце. Эти действия производятся рекурсивно.

Алгоритм VF2 был описан специалистами в области информационных технологий на основе алгоритма Ульмана в труде «An Improved Algorithm for Matching Large Graphs» [4]. Суть его состоит в следующем. Даны два графа $G_1 = (N_1, B_1)$ и $G_2 = (N_2, B_2)$, отображение $M \subset N_1 \times N_2$ называется изоморфизмом тогда и только тогда, когда оно представляет собой взаимно однозначную функцию: M отображает каждую ветвь G_1 на ветвь G_2 и наоборот. Отображение M называется граф-подграф изоморфизмом тогда и только тогда, когда M отражает изоморфизм между G_2 и подграфом графа G_1. Будем считать, что графы являются ориентированными, т.е. ветвь (i, j) следует считать отличной от (j, i). Каждое состояние *s* процесса согласования может быть связано с некоторым вариантом частичного отображением M(s), которое содержит только подмножество компонентов функции отображения M. Вариант частичного отображения однозначно идентифицирует два подграфа $G_1(s)$ и $G_2(s)$, полученных путем выбора из G_1 и G_2 узлов, которые были включены в компоненты отображения M(S), и ветвей, соединяющих их. В дальнейшем мы будем обозначать проекция M(S) на N_1 и N_2 как $M_1(s)$ и $M_2(s)$ соответственно, в то время как наборы ветвей $G_1(s)$ и $G_2(s)$ будут обозначаться как $B_1(s)$ и $B_2(s)$ соответственно. Высокоуровневое описание алгоритма на псевдокоде:

```
ПРОЦЕДУРА Match(s)
 ВХОД: промежуточное состояние s; первоначальному состоянию s0 соответствует M(s0)=∅
 ВЫХОД: соответствия между двумя графами
 ЕСЛИ M(s) охватывает все узлы графа G2 ТО
 ВЫВОД M(s)
 ИНАЧЕ
 Рассчитать множество P(s) пар кандидатов на включение в M(s)
 ДЛЯ_КАЖДОГО (n, m) ∈ P(s)
 ЕСЛИ F(s, n, m) ТО
 Рассчитать состояние s', которое получилось при добавлении (n, m) к M(s)
 ВЫЗОВ Match(s')//рекурсивный вызов
 КОНЕЦ_ЕСЛИ
 КОНЕЦ_ДЛЯ_КАЖДОГО
 Восстановить структуры данных
 КОНЕЦ_ЕСЛИ
КОНЕЦ_ПРОЦЕДУРЫ
```

F(s, n, m) – это логическая функция (так называемая функция осуществимости), которая используется, чтобы сократить дерево поиска. Если она возвращает значение «истина», то гарантируется, что состояние s', полученное путем добавления (n, m) к s, представляет собой частичный изоморфизм, если s тоже представляет собой частичный изоморфизм. Поэтому конечное состояние является либо изоморфизмом между G_1 и G_2, либо граф-подграф изоморфизмом между подграфом графа G1 и графом G2. Кроме того, F также будет исключает некоторые состояния, которые хоть и соответствуют изоморфизму между $G_1(s)$ и $G_2(s)$, но не приводят к полному решению.

Список литературы

1. *Зыков А.А.* Основы теории графов / А.А. Зыков. – М.: Вузовская книга, 2004. – 664 с.
2. *Cook, S. A.* "The complexity of theorem-proving procedures", Proc. 3rd ACM Symposium on Theory of Computing, 1971, с. 151–158.
3. *Ullman J. R.* «An Algorithm for Subgraph Isomorphism» / J.R. Ullman. – Journal of the Associations for Computing Machine, 1976. – с. 31 - 42
4. Cordelia L.P., Foggia P., Sansone C., Vento M. An Improved Algorithm for Matching Large Graphs / L.P. Cordelia [et al]. – Proceedings of the 3rd IAPR-TC15 Workshop on Graph-based Representations, 2001. 8 с.