CFPQ by Matrix Multiplication

Matrix-based algorithm for graph structured data analysis

Rustam Azimov[1], Semyon Grigorev[1],

[1] *Saint Petersburg State University, JetBrains, St. Petersburg, Russia*

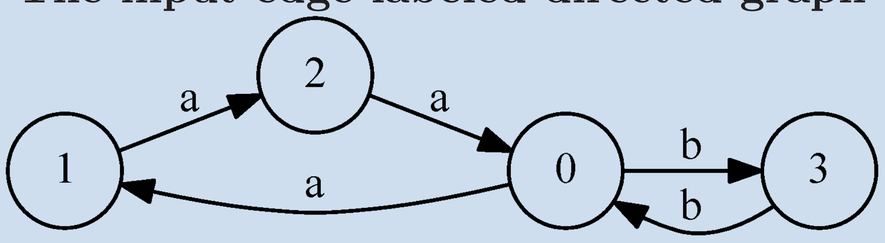
E-mails: *st013567@student.spbu.ru, s.v.grigoriev@spbu.ru*

Motivation

Context-free path querying (CFPQ) is a technique, which recently gains popularity in graph databases, bioinformatics, static analysis, etc. It is often required to query large graphs, and existing algorithms demonstrate a poor performance in this case. We propose the first generalization of matrix-based Valiant's algorithm for context-free path querying. The utilization of matrix operations in the process of context-free path query evaluation makes it possible to efficiently apply a wide class of optimizations and computing techniques for querying large graphs, such as GPGPU, parallel processing, sparse matrix representation, distributed-memory computation, etc.

Context-free path querying

The input edge-labeled directed graph



The input CF grammar for the language $L = \{a^n b^n\}$

$$\begin{array}{llllllll} 0: & S & \to & A\ B & 3: & A & \to & a \\ 1: & S & \to & A\ S_1 & 4: & B & \to & b \\ 2: & S_1 & \to & S\ B & & & & \end{array}$$

The result is a set of node pairs corresponding to paths, whose labeling is in the input language L.

Matrix-based approach

We compute the following matrix transitive closure:

$$T^{(i)} = T^{(i-1)} \cup (T^{(i-1)} \times T^{(i-1)}),\ i \geq 1.$$

The initial adjacency matrix for the input graph

$$T_0 = \begin{pmatrix} \varnothing & \{A\} & \varnothing & \{B\} \\ \varnothing & \varnothing & \{A\} & \varnothing \\ \{A\} & \varnothing & \varnothing & \varnothing \\ \{B\} & \varnothing & \varnothing & \varnothing \end{pmatrix}$$

The first iteration

$$T_1 = T_0 \cup (T_0 \times T_0) = \begin{pmatrix} \varnothing & \{A\} & \varnothing & \{B\} \\ \varnothing & \varnothing & \{A\} & \varnothing \\ \{A\} & \varnothing & \varnothing & \{\boldsymbol{S}\} \\ \{B\} & \varnothing & \varnothing & \varnothing \end{pmatrix}$$

The final matrix

$$T_{13} = \begin{pmatrix} \{S_1, S\} & \{A\} & \varnothing & \{B, S, S_1\} \\ \{S, S_1\} & \varnothing & \{A\} & \{S_1, S\} \\ \{A, S_1, S\} & \varnothing & \varnothing & \{S, S_1\} \\ \{B\} & \varnothing & \varnothing & \varnothing \end{pmatrix}$$

The indices of the elements of T_{13} which contain the nonterminal S are node pairs corresponding to paths, whose labeling is in the given language $\{a^n b^n\}$.

Results

- We propose the matrix-based algorithm for context-free path querying.
- We implemented this algorithm with a number of optimizations (sparse matrix representation, GPGPU) and applied this implementation to the navigation query problem for some popular RDF ontologies, taken from [1].
- We also compared the performance of our implementation with the fastest analog from [2] (based on GLL).
- All materials are available on GitHub: https://github.com/YaccConstructor

Evaluation

Evaluation results for the query $S \to aSb|cSd|ab|cd$ for retrieving the concepts on the same layer.

Ontology	V	E	GLL[2](ms)	sGPU(ms)
biomedical	341	711	261	20
people-pets	337	834	89	32
pizza	671	2604	697	24
g_1	6224	11840	1926	82
g_2	5864	19600	6246	185
g_3	5368	20832	7014	127

Evaluation results for the query $S \to Bb|b$, $B \to aBb|ab$ for retrieving concepts on the adjacent layers.

Ontology	V	E	GLL[2](ms)	sGPU(ms)
funding	778	1480	23	27
wine	733	2450	8	6
pizza	671	2604	29	23
g_1	6224	11840	167	38
g_2	5864	19600	46	21
g_3	5368	20832	393	40

Future Research

Currently, we are working on the matrix-based algorithm for path querying with conjunctive grammars which have more expressive power, than context-free grammars. We want to find new applications for the path querying techniques and implement the required tools.

Acknowledgments

The research was supported by the Russian Science Foundation grant 18-11-00100 and a grant from JetBrains Research.

References

[1] X. Zhang, Z. Feng, X. Wang, G. Rao, and W. Wu. Context-free path queries on rdf graphs. In *International Semantic Web Conference*, pages 632–648. Springer, 2016.

[2] Semyon Grigorev and Anastasiya Ragozina. Context-free path querying with structural representation of result. *arXiv preprint arXiv:1612.08872*, 2016.