Algorithm Selection for Maximum Common Subgraph

Paulius Dilkas

22nd October 2017

1 The Problem

Labeled and unlabeled (undireced) maximum common induced subgraph. NB: we are allowing loops. Update formulas accordingly.

2 Algorithms

Clique encoding [7] $k \downarrow$ [3] McSplit [6]

(Somewhere) 1000 s limit, 512 GB limit (clique becomes impossible for some instances), insert CPU specs.

3 Problem Instances

3.1 ARG database

MCS data is from [1][2].

3.2 Benchmarks for the Subgraph Isomorphism Problem

SIP instances are taken from [8] (with the biochemical reaction dataset excluded since we are not dealing with directed graphs).

4 Features

Features are based on the algorithm selection paper for the subgraph isomorphism problem [5]. Simple features:

- number of vertices,
- number of edges,
- density,

- number of loops,
- mean degree,
- maximum degree,
- standard deviation of degrees,
- whether the graph is connected,
- mean distance between all pairs of vertices,
- maximum distance between all pairs of vertices,
- proportion of vertex pairs that are at least 2, 3 and 4 apart (?),
- number of labels,
- number of distinct labels.

The last two were later rethought to be unnecessary and replaced by a boolean feature "labelled" because if labelling is enabled, the number of labels is equal to the number of vertices and the number of distinct labels is equal to 33% of that.

Features that could be computed if we end up using a presolver:

- uniformity of the distribution of edges,
- how many candidate pairs were removed,
- proportion of candidate pairs removed over all pairs,
- min values removed per variable,
- max values removed per variable,
- CPU time taken to compute all this.

Number (distinct or not) of labels includes the label of not having a label (so it's always at least 1).

5 Selection Model

We're using LLAMA [4]. Describe k-folding.

References

[1] M. De Santo et al. "A large database of graphs and its use for benchmarking graph isomorphism algorithms". In: *Pattern Recogn. Lett.* 24.8 (May 2003), 10671079. ISSN: 0167-8655. DOI: 10.1016/S0167-8655(02)00253-2. URL: http://dx.doi.org/10.1016/S0167-8655(02)00253-2.

[2] P. Foggia, C. Sansone and M. Vento. "A Database of Graphs for Isomorphism and Sub-Graph Isomorphism Benchmarking". In: -. 1st Jan. 2001, 176187.

[3] Ruth Hoffmann, Ciaran McCreesh and Craig Reilly. "Between Subgraph Isomorphism and Maximum Common Subgraph". In: *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, February 4-9, 2017, San Francisco, California, USA.* Ed. by Satinder P. Singh and Shaul Markovitch. AAAI Press, 2017, pp. 3907–3914. URL: http://aaai.org/ocs/index.php/AAAI/AAAI17/paper/view/14948.

[4] Lars Kotthoff. *LLAMA: Leveraging Learning to Automatically Manage Algorithms.* Tech. rep. arXiv:1306.1031. arXiv, June 2013. URL: http://arxiv.org/abs/1306.1031.

[5] Lars Kotthoff, Ciaran McCreesh and Christine Solnon. "Portfolios of Subgraph Isomorphism Algorithms". In: *Learning and Intelligent Optimization - 10th International Conference, LION 10, Ischia, Italy, May 29 - June 1, 2016, Revised Selected Papers.* Ed. by Paola Festa, Meinolf Sellmann and Joaquin Vanschoren. Vol. 10079. Lecture Notes in Computer Science. Springer, 2016, pp. 107–122. ISBN: 978-3-319-50348-6. DOI: 10.1007/978-3-319-50349-3_8. URL: https://doi.org/10.1007/978-3-319-50349-3_8.

[6] Ciaran McCreesh, Patrick Prosser and James Trimble. "A Partitioning Algorithm for Maximum Common Subgraph Problems". In: *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI 2017, Melbourne, Australia, August 19-25, 2017.* Ed. by Carles Sierra. ijcai.org, 2017, pp. 712–719. ISBN: 978-0-9992411-0-3. DOI: 10.24963/ijcai.2017/99. URL: https://doi.org/10.24963/ijcai.2017/99.

[7] Ciaran McCreesh et al. "Clique and Constraint Models for Maximum Common (Connected) Subgraph Problems". In: *Principles and Practice of Constraint Programming - 22nd International Conference, CP 2016, Toulouse, France, September 5-9, 2016, Proceedings.* Ed. by Michel Rueher. Vol. 9892. Lecture Notes in Computer Science. Springer, 2016, pp. 350–368. ISBN: 978-3-319-44952-4. DOI: 10.1007/978-3-319-44953-1_23. URL: https://doi.org/10.1007/978-3-319-44953-1_23.

[8] Christine Solnon. *Benchmarks for the Subgraph Isomorphism Problem.* URL: http://liris.cnrs.fr/csolnon/SIP.html.