Abedin, Paniz; **Hooshmand, Sahar**; **Ganguly, Arnab**; **Thankachan, Sharma V.**
The heaviest induced ancestors problem revisited. (English) Zbl 07286746
Navarro, Gonzalo (ed.) et al., 29th annual symposium on combinatorial pattern matching, CPM 2018, July 2–4, 2018, Qingdao, China. Wadern: Schloss Dagstuhl – Leibniz Zentrum für Informatik (ISBN 978-3-95977-074-3). LIPIcs – Leibniz International Proceedings in Informatics 105, Article 20, 13 p. (2018).

Summary: We revisit the heaviest induced ancestors problem, which has several interesting applications in string matching. Let $\mathcal{T}_1$ and $\mathcal{T}_2$ be two weighted trees, where the weight $\mathsf{W}(u)$ of a node u in either of the two trees is more than the weight of u's parent. Additionally, the leaves in both trees are labeled and the labeling of the leaves in $\mathcal{T}_2$ is a permutation of those in $\mathcal{T}_1$. A node $x \in \mathcal{T}_1$ and a node $y \in \mathcal{T}_2$ are induced, iff their subtree have at least one common leaf label. A heaviest induced ancestor query $\mathsf{HIA}(u_1, u_2)$ is: given a node $u_1 \in \mathcal{T}_1$ and a node $u_2 \in \mathcal{T}_2$, output the pair (u_1^*, u_2^*) of induced nodes with the highest combined weight $\mathsf{W}(u_1^*) + \mathsf{W}(u_2^*)$, such that u_1^* is an ancestor of u_1 and u_2^* is an ancestor of u_2. Let n be the number of nodes in both trees combined and $\varepsilon > 0$ be an arbitrarily small constant. *T. Gagie* et al. ["Heaviest induced ancestors andlongest common substring", Preprint, arXiv:1305.3164] introduced this problem and proposed three solutions with the following space-time trade-offs:

- an $O(n \log^2 n)$-word data structure with $O(\log n \log\log n)$ query time
- an $O(n \log n)$-word data structure with $O(\log^2 n)$ query time
- an $O(n)$-word data structure with $O(\log^{3+\varepsilon} n)$ query time.

In this paper, we revisit this problem and present new data structures, with improved bounds. Our results are as follows.

- an $O(n \log n)$-word data structure with $O(\log n \log\log n)$ query time
- an $O(n)$-word data structure with $O(\frac{\log^2 n}{\log\log n})$ query time.

As a corollary, we also improve the LZ compressed index of Gagie et al. [loc. cit.] for answering longest common substring (LCS) queries. Additionally, we show that the LCS after one edit problem of size n [*A. Amir* et al., Algorithmica 82, No. 12, 3707–3743 (2020; Zbl 07272778)] can also be reduced to the heaviest induced ancestors problem over two trees of n nodes in total. This yields a straightforward improvement over its current solution of $O(n \log^3 n)$ space and $O(\log^3 n)$ query time.

For the entire collection see [Zbl 1390.68025].

Reviewer: Reviewer (Berlin)

MSC:

68W32 Algorithms on strings

Keywords:

data structure; string algorithms; orthogonal range queries

Full Text: DOI

References:

[1] Amihood Amir, Panagiotis Charalampopoulos, Costas S. Iliopoulos, Solon P. Pissis, and Jakub Radoszewski. Longest common factor after one edit operation. In Gabriele Fici, Marinella Sciortino, and Rossano Venturini, editors, {\it String Processing and Information} {\it Retrieval - 24th International Symposium, SPIRE 2017, Palermo, Italy, September 26-} {\it 29, 2017, Proceedings}, volume 10508 of {\it Lecture Notes in Computer Science}, pages 14-26. Springer, 2017. doi:10.1007/978-3-319-67428-5_2.

[2] Gerth Stølting Brodal and Allan Grønlund Jørgensen. Data structures for range median queries. In Yingfei Dong, Ding-Zhu Du, and Oscar H. Ibarra, editors, {\it Algorithms and Com-} {\it putation, 20th International Symposium, ISAAC 2009, Honolulu, Hawaii, USA, December} {\it 16-18, 2009. Proceedings}, volume 5878 of {\it Lecture Notes in Computer Science}, pages 822- 831. Springer, 2009. doi:10.1007/978-3-642-10631-6_83.

[3] Timothy M. Chan, Kasper Green Larsen, and Mihai Patrascu. Orthogonal range searching on the ram, revisited. In Ferran

Hurtado and Marc J. van Kreveld, editors, {\it Proceedings of} {\it the 27th ACM Symposium on Computational Geometry, Paris, France, June 13-15, 2011}, pages 1-10. ACM, 2011. doi:10.1145/1998196.1998198.

[4] Timothy M. Chan and Bryan T. Wilkinson. Adaptive and approximate orthogonal range counting. In Sanjeev Khanna, editor, {\it Proceedings of the Twenty-Fourth Annual ACM-SIAM} {\it Symposium on Discrete Algorithms, SODA 2013, New Orleans, Louisiana, USA, January} {\it 6-8, 2013}, pages 241-251. SIAM, 2013. doi:10.1137/1.9781611973105.18.

[5] Erik D. Demaine, Gad M. Landau, and Oren Weimann. On cartesian trees and range minimum queries. {\it Algorithmica}, 68(3):610-625, 2014. doi:10.1007/s00453-012-9683-x.

[6] Paolo Ferragina and Giovanni Manzini. Indexing compressed text. {\it J. ACM}, 52(4):552-581, 2005. doi:10.1145/1082036.1082039.

[7] Johannes Fischer and Volker Heun. Space-efficient preprocessing schemes for range minimum queries on static arrays. {\it SIAM J. Comput.}, 40(2):465-492, 2011. doi:10.1137/ 090779759.

[8] Travis Gagie, Pawel Gawrychowski, and Yakov Nekrich. Heaviest induced ancestors and longest common substrings. In {\it Proceedings of the 25th Canadian Conference on Compu-} {\it tational Geometry, CCCG 2013, Waterloo, Ontario, Canada, August 8-10, 2013}. Carleton University, Ottawa, Canada, 2013. URL: http://cccg.ca/proceedings/2013/papers/paper_29.pdf.

[9] Roberto Grossi and Jeffrey Scott Vitter. Compressed suffix arrays and suffix trees with applications to text indexing and string matching. {\it SIAM J. Comput.}, 35(2):378-407, 2005. doi:10.1137/S0097539702402354.

[10] Dov Harel and Robert Endre Tarjan. Fast algorithms for finding nearest common ancestors. {\it SIAM J. Comput.}, 13(2):338-355, 1984. doi:10.1137/0213024.

[11] Joseph JáJá, Christian Worm Mortensen, and Qingmin Shi. Space-efficient and fast algorithms for multidimensional dominance reporting and counting.In Rudolf Fleischer and Gerhard Trippen, editors, {\it Algorithms and Computation, 15th International Sym-} {\it posium, ISAAC 2004, Hong Kong, China, December 20-22, 2004, Proceedings}, volume 3341 of {\it Lecture Notes in Computer Science}, pages 558-568. Springer, 2004. doi:10.1007/ 978-3-540-30551-4_49.

[12] Yakov Nekrich and Gonzalo Navarro. Sorted range reporting. In Fedor V. Fomin and Petteri Kaski, editors, {\it Algorithm Theory - SWAT 2012 - 13th Scandinavian Symposium and} {\it Workshops, Helsinki, Finland, July 4-6, 2012. Proceedings}, volume 7357 of {\it Lecture Notes in} {\it Computer Science}, pages 271-282. Springer, 2012. doi:10.1007/978-3-642-31155-0_24.

[13] Kunihiko Sadakane. Succinct representations of lcp information and improvements in the compressed suffix arrays. In David Eppstein, editor, {\it Proceedings of the Thirteenth Annual} {\it ACM-SIAM Symposium on Discrete Algorithms, January 6-8, 2002, San Francisco, CA,} {\it USA.}, pages 225-232. ACM/SIAM, 2002. URL: http://dl.acm.org/citation.cfm?id=545381.545410.

[14] Kunihiko Sadakane and Gonzalo Navarro.Fully-functional succinct trees.In Moses Charikar, editor, {\it Proceedings of the Twenty-First Annual ACM-SIAM Symposium on Dis-} {\it crete Algorithms, SODA 2010, Austin, Texas, USA, January 17-19, 2010}, pages 134-149. SIAM, 2010. doi:10.1137/1.9781611973075.13.

[15] Daniel Dominic Sleator and Robert Endre Tarjan. A data structure for dynamic trees. In {\it Proceedings of the 13th Annual ACM Symposium on Theory of Computing, May 11-} {\it 13, 1981, Milwaukee, Wisconsin, USA}, pages 114-122. ACM, 1981. doi:10.1145/800076. 802464.

[16] Sharma V. Thankachan, Sriram P. Chockalingam, and Srinivas Aluru. An efficient algorithm for finding all pairs k-mismatch maximal common substrings. In {\it Bioinformatics} {\it Research and Applications - 12th International Symposium, ISBRA 2016, Minsk, Belarus,} {\it June 5-8, 2016, Proceedings}, pages 3-14, 2016.

[17] Peter Weiner. Linear pattern matching algorithms. In {\it 14th Annual Symposium on Switching} {\it and Automata Theory, Iowa City, Iowa, USA, October 15-17, 1973}, pages 1-11. IEEE Computer Society, 1973. doi:10.1109/SWAT.1973.13.

[18] Dan E. Willard. Log-logarithmic worst-case range queries are possible in space theta(n). {\it Inf. Process. Lett.}, 17(2):81-84, 1983. doi:10.1016/0020-0190(83)90075-3.

[19] Gelin Zhou. Two-dimensional range successor in optimal time and almost linear space. {\it Inf.} {\it Process. Lett.}, 116(2):171-174, 2016. doi:10.1016/j.ipl.2015.09.002.

[20] :12

[21] :13

[22] Jacob Ziv and Abraham Lempel. A universal algorithm for sequential data compression. {\it IEEE Trans. Information Theory}, 23(3):337-343, 1977. doi:10.1109/TIT.1977.1055714.