224072722

Fileppus S Moses

Trends in artificial intelligence and Machine Learning

Chen, Xianrun, Dachuan Xu, Yicheng Xu, and Yong Zhang. 2024. “Parameterized Approximation Algorithms for Sum of Radii Clustering and Variants.” In *Proceedings of the Thirty-Eighth AAAI Conference on Artificial Intelligence and Thirty-Sixth Conference on Innovative Applications of Artificial Intelligence and Fourteenth Symposium on Educational Advances in Artificial Intelligence*. AAAI’24/IAAI’24/EAAI’24. AAAI Press. https://doi.org/10.1609/aaai.v38i18.30053.

1. Summary of the Paper

The computationally difficult Sum of Radii (SoR) clustering issue and its constrained variations, Fair SoR and Matroid SoR, are the subject of the paper "Parameterized Approximation Algorithms for Sum of Radii Clustering and Variants." In the SoR problem, k clusters are chosen in a metric space so that the sum of the radii of these clusters is as small as possible while covering all data points. This problem provides a compromise between minimizing the greatest radius and total distance, and it falls between the well-known k-center and k-median clustering objectives.

In order to obtain effective approximations, the authors suggest a broad and cohesive algorithmic framework that makes use of fixed-parameter tractable (FPT) techniques. Among their most significant contributions are the following:

- A randomized FPT algorithm that approximates the classical SoR problem with a (2+ε)(2 + \varepsilon)(2+ε)-approximation.

- Extensions of this method to derive algorithms for (3+ε)(3 + \varepsilon)(3+ε)-approximation for MatSoR and FairSoR challenges.

- Theoretical assurances that every algorithm executes on schedule O(1) = 2O(klog(k/ε))· 2^{O(k \log (k/\varepsilon))} \cdot n^{O(1)}2O(klog(k/ε))·nO(1), where n is the number of data points and k is the number of clusters.

Thorough probabilistic analysis, radius-profile enumeration, iterative greedy coverings, and transformations that manage fairness and matroid constraints using bipartite matching and matroid intersection approaches are used by the authors to develop their theoretical results. Notably, even if the algorithms are probabilistic, repetition or derandomized search might increase the likelihood of finding a valid answer at the expense of a thorough search.

Nevertheless, there are no actual trials in the text to back up practical performance. Furthermore, FPT algorithms could not scale effectively in situations when k is large, even though they are useful for small k.

2. Discussion of Related Literature

The paper is situated within an extensive body of research in approximation algorithms and clustering theory. The authors reference several foundational and contemporary works to contextualize their contributions. The following ten sources are especially pertinent:

Charikar & Panigrahy (2004). presented the first constant-factor approximation algorithm for SoR, employing a primal-dual approach to obtain a 3.504-approximation. For almost twenty years, this was the standard.

The most well-known polynomial-time solution before the current study was obtained by Friggstad & Jamshidian (2022), who used sophisticated LP rounding techniques to enhance the approximation ratio for SoR to 3.389.

The SoR problem was officially introduced and its NP-hardness was confirmed by Montag and Suri (1989), providing the theoretical underpinnings for subsequent approximation techniques.

To show that FPT techniques are feasible in limited environments, Inamdar & Varadarajan (2020) suggested an FPT (28-approximation) algorithm for uniform capacitated SoR.

Bandyapadhyay et al. (2023) made notable progress in capacitated clustering by increasing the approximation guarantees to 4 for the uniform case and to 15 for the non-uniform capacitated SoR.

By presenting the idea of fairlets and balance fairness, Chiarichetti et al. (2017) established the foundation for fair clustering. FairSoR and other fairness-aware clustering algorithm enhancements were influenced by their work.

For data summarization, Kleindessner et al. (2019) suggested fair k-center clustering, which produced linear-time algorithms with exponential approximation factors.

Using combinatorial structures to manage matroid restrictions, Chen et al. (2016) presented a 3-approximation approach for the matroid center problem. Using LP-rounding, Swamy (2016) presented an 8-approximation for the matroid median problem, which influenced MatSoR techniques.

To demonstrate the usefulness and strength of FPT techniques in confined settings, Goyal & Jaiswal (2023) provided tight FPT approximation solutions for constrained k-center and k-supplier clustering problems.

Together, these studies enhance approximation ratios and expand application to complex clustering constraints, forming the theoretical foundation of the current paper.

3. Critical Evaluation

By presenting a generalized FPT approximation framework for SoR and its constrained versions, the reviewed study significantly advances the theoretical understanding of clustering optimization problems. A major breakthrough in the field is represented by the unification of methods under a single framework and the ensuing improvements in approximation guarantees (from over 3.3 to below 3 for SoR, and similarly reduced bounds for FairSoR and MatSoR).

The paper's rigorous algorithmic design, which is backed by approximation theory and probabilistic analysis, is one of its strongest points. The authors' thorough proofs, formal lemmas, and algorithmic procedures improve the approaches' reproducibility and transparency. Additionally, their discovery of structural parallels between matroid independence and fairness constraints demonstrates a thorough comprehension of constraint modeling in clustering.

The paper does, however, have several shortcomings. Initially, the suggested algorithms are all theoretical; no actual tests are carried out to evaluate their scalability, runtime performance, or practicality. The FPT technique is less practical in high-dimensional or large-cluster issues, even if it works well when k is small, which is a typical situation in practice. Even while the algorithms are theoretically effective, they also depend on frequent guessing and randomization, which can be computationally costly.

The study is also limited to metric spaces. Practical relevance could be greatly increased by extensions to non-metric domains, dynamic datasets, or streaming systems. The authors explicitly admit that it is still difficult to establish constant approximations for constrained forms of SoR in polynomial time, indicating areas that could use more investigation.

4. Conclusion

In conclusion, this research offers a sophisticated and practical theoretical approach for dealing with the constrained variations of the Sum of Radii clustering problem. The authors make a substantial contribution to the literature on clustering algorithms by using fixed-parameter tractable strategies to obtain state-of-the-art approximation ratios under matroid and fairness constraints.

Although the paper's theoretical innovation and technical precision are praiseworthy, its immediate usefulness is limited by its lack of experimental validation and scaling considerations. Through actual research, parallelized implementations, and more extensive generalizations outside of metric situations, future work should try to close this gap.

5. Reference

- Charikar, M., & Panigrahy, R. (2004). Clustering to minimize the sum of cluster diameters. *Journal of Computer and System Sciences*, 68(2), 417–441. https://doi.org/10.1016/j.jcss.2003.06.004
- Friggstad, Z., & Jamshidian, M. (2022). Improved polynomial-time approximations for clustering with minimum sum of radii or diameters. In *Proceedings of the 30th Annual European Symposium on Algorithms (ESA)* (LIPIcs, Vol. 244, pp. 56:1–56:14). Schloss Dagstuhl–Leibniz-Zentrum für Informatik. https://doi.org/10.4230/LIPIcs.ESA.2022.56
- Monma, C., & Suri, S. (1989). Partitioning points and graphs to minimize the maximum or the sum of diameters. In *Graph Theory, Combinatorics and Applications* (Vol. 2, pp. 899–912). Wiley.
- Inamdar, T., & Varadarajan, K. R. (2020). Capacitated sum-of-radii clustering: An FPT approximation. In *Proceedings of the 28th Annual European Symposium on Algorithms (ESA)* (LIPIcs, Vol. 173, pp. 62:1–62:17). Schloss Dagstuhl–Leibniz-Zentrum für Informatik. https://doi.org/10.4230/LIPIcs.ESA.2020.62
- Bandyapadhyay, S., Lochet, W., & Saurabh, S. (2023). FPT constant-approximations for capacitated clustering to minimize the sum of cluster radii. In *Proceedings of the 39th International Symposium on Computational Geometry (SoCG)* (LIPIcs, Vol. 258, pp. 12:1–12:14). Schloss Dagstuhl–Leibniz-Zentrum für Informatik. https://doi.org/10.4230/LIPIcs.SoCG.2023.12
- Chierichetti, F., Kumar, R., Lattanzi, S., & Vassilvitskii, S. (2017). Fair clustering through fairlets. In *Advances in Neural Information Processing Systems* (Vol. 30, pp. 5029–5037).
- Kleindessner, M., Awasthi, P., & Morgenstern, J. (2019). Fair k-center clustering for data summarization. In *Proceedings of the 36th International Conference on Machine Learning (ICML)* (PMLR, Vol. 97, pp. 3448–3457).

- Chen, D. Z., Li, J., Liang, H., & Wang, H. (2016). Matroid and knapsack center problems. *Algorithmica*, 75(1), 27–52. https://doi.org/10.1007/s00453-015-0016-0
- Swamy, C. (2016). Improved approximation algorithms for matroid and knapsack median problems and applications. *ACM Transactions on Algorithms*, 12(4), 49:1–49:22. https://doi.org/10.1145/2846092
- Goyal, D., & Jaiswal, R. (2023). Tight FPT approximation for constrained k-center and k-supplier. *Theoretical Computer Science*, 940(Part A), 190–208. https://doi.org/10.1016/j.tcs.2022.11.018