修士論文 要旨

動的区間最頻値問題と頂点発見問題の効率的アルゴリズム

数理情報学専攻 48206208 小畑 哲雅
指導教員 定兼 邦彦 教授

1 動的区間最頻値問題

定義 1 (mode)**.** *A : multiset*
$a \in A$ が A の mode $\Leftrightarrow \forall b \in A$ (a の A での重複度) $\geq$ (b の A での重複度)

mode を扱う問題として、 Range mode problem が存在する。

問題 2 (Range mode problem)**.** アルファベット集合 Σ 上の列 A が与えられたとき，以下のクエリから成るクエリの列を処理せよ．

- mode(l, r)*:* $A[l : r]$ の *mode* を 1 つ出力する．

Range mode problem は Set intersection problem 等への応用が知られており [1]， 多くの先行研究が存在する [1] [2] [3] [4] [5] [6].

Range mode problem を拡張した問題として， Dynamic range mode enumeration problem が考えられる．

問題 3 (Dynamic range mode enumeration problem)**.** アルファベット集合 Σ 上の列 A が与えられたとき，以下のクエリから成るクエリの列を処理せよ．

- insert(c, i)*:* $c\,(\in \Sigma)$ を A の i 番目に挿入する．
- delete(i)*:* A の第 i 要素を削除する．
- modes(l, r)*:* $A[l : r]$ の *mode* を全て出力する．

dynamic range mode enumeration problem の先行研究は存在しない．

定理 4 が本稿の主定理である．

定理 4. *dynamic range mode enumeration problem を $\Omega(\log N + \log \sigma)$ bits wordsize の word RAM モデル上で， insert クエリと delete クエリを $\mathrm{O}\left(N^{\frac{2}{3}} \log \sigma'\right)$ time per query で処理し， modes クエリを $\mathrm{O}\left(N^{\frac{2}{3}} \log \sigma' + |output|\right)$ time per query で処理するデータ構造が存在する．ここで，N はクエリ時点での列の長さを，σ は $|\Sigma|$ を，σ' は列に現れる要素の種類数を現す．空間計算量は，$\mathrm{O}\left(N + N^{\frac{2}{3}} \sigma'\right)$ words である．*

列を $L = \Theta(N^{\alpha})$ 個の長さ $C = \Theta\left(N^{1-\alpha}\right)$ $(0 \leq \alpha \leq 1)$．以下の部分列に分割し、適切に挿入や削除、及び各部分列間での要素の受け渡しを行うことによって、各クエリを高速に処理することが出来る。

insert, delete によって列のサイズが大きく変化した場合、 L の値を変更する必要がある。これに対処する為に、動的データ構造のテクニック [7] を用いる．データ構造を一定数のクエリ毎に再構築することで、各クエリの時間計算量を償却計算量で評価することが出来る。更に、次の再構築後に使用するデータ構造を各クエリのタイミングで少しずつ準備することによって、各クエリの時間計算量を最悪計算量で評価することが出来る。

2 頂点発見問題

以下の形式で表される問題を考える．

問題 5. 辺の情報が未知であるグラフが存在する．以下のクエリを何度でも行うことが出来る．

- *probe* (u, v)*:* 頂点 u と頂点 v の間の辺についての情報を得ることが出来る．

可能な限り少ない回数のクエリを用いて，グラフについての「ある性質」を調べよ．

この問題は，グラフの性質を効率良く特定しようとするアルゴリズムと，グラフの性質を特定されないようにグラフの辺の張り方を決定する戦略の対立構造と見ることが出来る．probe クエリを行うことは、現実のグラフ構造を持つ対象において各ノードの関係性を調べる行為に対応しており、少ない probe でグラフの性質を調べることは、対象の全体像を効率良く調べることに役立つ。クエリ回数の上界や下界について， 多くの研究が行われている． [8] [9] 特に、クエリ回数の下界が $\binom{n}{2}$ 回となる性質は elusive であると定義され、elusiveness については盛んに研究されている。 [8] [10] [11]

本稿では， 以下の問題を考える。

問題 6. 各頂点組間の辺の有無及びその向き付けについての情報が未知である有向グラフが存在する．以下のクエリを何度でも行うことが出来る．

- *probe* (u, v)*:* 頂点 u と頂点 v の間の辺の有無及び

その向き付けについての情報を得ることが出来る.

可能な限り少ない回数のクエリを用いて，出次数が丁度 k である頂点を 1 つ発見せよ．或いは，そのような頂点が存在しないことを報告せよ.

$0 \leq k \leq (n+1)/2$ の場合は， 全ての頂点組について probe する必要があることが知られている [8]．一方， $(n+1)/2 < k$ の場合は，probe 回数についての既知の最良の下界は， $\binom{n}{2} - \binom{k}{2}$ 回である． [8]

本稿は，$(n+1)/2 < k$ の場合の下界を改善した．主要な定理を定理 7 に示す.

定理 7. $(n+1)/2 < k$ のとき，出次数 k の頂点を発見するには， $\binom{n}{2} - \binom{2k-n+1}{2}$ 回のクエリが必要である.

定理 7 の評価は， 元の評価の証明 [8] における辺配置を更に密にすることで達成される.

参考文献

[1] Timothy M. Chan, Stephane Durocher, Matthew Skala, and Bryan T. Wilkinson. Linear-space data structures for range minority query in arrays. In Fedor V. Fomin and Petteri Kaski, editors, *Algorithm Theory – SWAT 2012*, pages 295–306, Berlin, Heidelberg, 2012. Springer Berlin Heidelberg.

[2] Holger Petersen. Improved bounds for range mode and range median queries. In *Proceedings of the 34th Conference on Current Trends in Theory and Practice of Computer Science*, SOFSEM'08, pages 418–423. Springer-Verlag, 2008.

[3] Holger Petersen and Szymon Grabowski. Range mode and range median queries in constant time and sub-quadratic space. *Inf. Process. Lett.*, 109:225–228, 01 2009.

[4] Mark Greve, Allan Grønlund Jørgensen, Kasper Dalgaard Larsen, and Jakob Truelsen. Cell probe lower bounds and approximations for range mode. In Samson Abramsky, Cyril Gavoille, Claude Kirchner, Friedhelm Meyer auf der Heide, and Paul G. Spirakis, editors, *Automata, Languages and Programming*, pages 605–616, Berlin, Heidelberg, 2010. Springer Berlin Heidelberg.

[5] Timothy M. Chan, Stephane Durocher, Kasper Green Larsen, Jason Morrison, and Bryan T. Wilkinson. Linear-space data structures for range mode query in arrays. *Theory of Computing Systems*, 55(4):719–741, November 2014.

[6] Kentaro Sumigawa, Sankardeep Chakraborty, Kunihiko Sadakane, and Srinivasa Rao Satti. Enumerating Range Modes. In Yixin Cao, Siu-Wing Cheng, and Minming Li, editors, *31st International Symposium on Algorithms and Computation (ISAAC 2020)*, volume 181 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 29:1–29:16, Dagstuhl, Germany, 2020. Schloss Dagstuhl–Leibniz-Zentrum für Informatik.

[7] Veil Mäkinen and Gonzalo Navarro. Dynamic entropy-compressed sequences and full-text indexes. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 4009 LNCS:306–317, 2006.

[8] Dishant Goyal, Varunkumar Jayapaul, and Venkatesh Raman. Elusiveness of finding degrees. In Daya Gaur and N.S. Narayanaswamy, editors, *Algorithms and Discrete Applied Mathematics*, pages 242–253, Cham, 2017. Springer International Publishing.

[9] Arindam Biswas, Varunkumar Jayapaul, Venkatesh Raman, and Srinivasa Rao Satti. The complexity of finding (approximate sized) distance-d dominating set in tournaments. In Mingyu Xiao and Frances Rosamond, editors, *Frontiers in Algorithmics*, pages 22–33, Cham, 2017. Springer International Publishing.

[10] Arnold L. Rosenberg. On the time required to recognize properties of graphs: A problem. *SIGACT News*, 5(4):15–16, oct 1973.

[11] H.W. Lenstra, M.R. Best, and Peter van Emde Boas. A sharpened version of the aanderaa-rosenberg conjecture. *Report 30/74, Mathematisch Centrum Amsterdam (1974)*, 01 1974.