Discrete Applied Mathematics 160 (2012) 1872–1874

Contents lists available at SciVerse ScienceDirect

Discrete Applied Mathematics

journal homepage: www.elsevier.com/locate/dam

Note

Greedy colorings of words

Dieter Rautenbach [a,*], Zoltán Szigeti [b]

[a] Institute of Optimization and Operations Research, University of Ulm, Ulm, Germany

[b] Laboratoire G-SCOP, CNRS, INP Grenoble, UJF, Grenoble, France

ARTICLE INFO

Article history:
Received 15 November 2011
Received in revised form 28 March 2012
Accepted 30 March 2012
Available online 25 April 2012

Keywords:
Binary paint shop problem
Greedy algorithm
Evenly laminar hypergraph
Odd transversal

ABSTRACT

In the Binary Paint Shop Problem proposed by Epping et al. (2004) [4] one has to find a 0/1-coloring of the letters of a word in which every letter from some alphabet appears twice, such that the two occurrences of each letter are colored differently and the total number of color changes is minimized. Meunier and Sebő (2009) [5] and Amini et al. (2010) [1] gave sufficient conditions for the optimality of a natural greedy algorithm for this problem. Our result is a best possible generalization of their results. We prove that the greedy algorithm optimally colors every suitable subword of a given instance word w if and only if w contains none of the three words (a, b, a, c, c, b), (a, d, d, b, c, c, a, b), and (a, d, d, c, b, c, a, b) as a subword. Furthermore, we relate this to the fact that every member of a family of hypergraphs associated with w is evenly laminar.

1. Introduction

The following so-called *Binary Paint Shop Problem* proposed by Epping et al. [4], has recently received considerable attention:

> Given a word $w = (w_1, \ldots, w_n)$ over some alphabet Σ such that each letter from Σ appears twice in w, find a 0/1-coloring $f = (f_1, \ldots, f_n) \in \{0, 1\}^n$ of w in which the two occurrences of each letter are colored differently, that is, $w_i = w_j$ for some $i < j$ implies $f_i \neq f_j$, such that the number of color changes between consecutive letters is minimized.

A word in which each letter from some alphabet Σ appears either twice or never is called an *admissible word*. A coloring f of a word w with 0 and 1 is called a 0/1-*coloring*. A 0/1-coloring f of an admissible word w is called *good* if the two occurrences of each letter of w are colored differently. If $f_i \neq f_{i+1}$, then we say that there is a *color change* at position $i + \frac{1}{2}$, that is, there is a color change *between* i *and* $i + 1$.

The Binary Paint Shop Problem is APX-hard [3,4] but there is a natural *greedy algorithm*, which works for an admissible word $w = (w_1, \ldots, w_n)$ as follows:

> Set $g_1 := 0$ and for j from 2 up to n, set $g_j := g_{j-1}$ unless there is some $i < j$ with $w_i = w_j$ and $g_i = g_{j-1}$. In the latter case, set $g_j := 1 - g_{j-1}$.

The coloring produced by the greedy algorithm is called the *greedy coloring*. The following claim is evident.

Claim 1. *For every admissible word, its greedy coloring is good.* □

The performance of the greedy algorithm on random instances was studied in [1,2].

* Corresponding author. Fax: +49 3677 69 3272.
E-mail addresses: dieter.rautenbach@uni-ulm.de, dieter.rautenbach@gmx.de (D. Rautenbach), zoltan.szigeti@g-scop.inpg.fr (Z. Szigeti).

0166-218X/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.dam.2012.03.038

A word w' is a *subword* of a word $w = (w_1, \dots, w_n)$ if there are indices $1 \le i_1 < i_2 < \cdots < i_k \le n$ with $w' = (w_{i_1}, w_{i_2}, \dots, w_{i_k})$. Meunier and Sebő [5] proved the optimality of the greedy algorithm for instances that do not contain (a, b, b, a) as a subword. Their result was extended by Amini et al. [1] who proved the optimality of the greedy algorithm for instances that do not contain (a, b, a, c, c, b) or (a, b, b, c, a, c) as a subword. Our aim in the present note is to extend the last two results optimally in a certain sense. We need some terminology.

For two integers $i < j$, let

$$I(i,j) = \left\{ i + \frac{1}{2}, (i+1) + \frac{1}{2}, \dots, (j-1) + \frac{1}{2} \right\}.$$

For an admissible word $w = (w_1, \dots, w_n)$ and its greedy coloring $g = (g_1, \dots, g_n)$, let

$$\mathcal{I}(w) = \{I(i,j) \mid 1 \le i < j \le n, w_i = w_j\} \quad \text{and}$$
$$\mathcal{G}(w) = \{I(i,j) \in \mathcal{I}(w) \mid g_{j-1} \ne g_j\}.$$

By the definition of a good coloring, the following is obvious.

Claim 2. *For every good coloring $f = (f_1, \dots, f_n)$ of an admissible word $w = (w_1, \dots, w_n)$, if w_i and w_j are the two occurrences of some letter in w, then there is an odd number of color changes between i and j.* □

A hypergraph $\mathcal{H}$ is called *evenly laminar* if it is laminar, that is, every two hyperedges of $\mathcal{H}$ are disjoint or one contains the other, and each of its hyperedges properly contains an even number of hyperedges of $\mathcal{H}$. An *odd transversal* of a hypergraph $\mathcal{H}$ is a subset of its vertex set that intersects each hyperedge of $\mathcal{H}$ in an odd number of elements.

We can now state and prove our result.

Theorem 1. *For every admissible word w, the following three statements are equivalent.*

(a) *The greedy coloring is optimal for every admissible subword w' of w.*
(b) *$\mathcal{G}(w')$ is an evenly laminar hypergraph for every admissible subword w' of w.*
(c) *w contains none of the three words (a, b, a, c, c, b), (a, d, d, b, c, c, a, b), and (a, d, d, c, b, c, a, b) as a subword.*

Proof. We prove the three implications (a) ⇒ (c), (c) ⇒ (b), and (b) ⇒ (a).

(a) ⇒ **(c)**: Let $w_1 = (a, b, a, c, c, b)$, $w_2 = (a, d, d, b, c, c, a, b)$, and $w_3 = (a, d, d, c, b, c, a, b)$. The greedy coloring g_i and an optimal coloring f_i of the word w_i are as follows

$$g_1 = (0, 0, 1, 1, 0, 1), \quad g_2 = (0, 0, 1, 1, 1, 0, 1, 0), \quad \text{and} \quad g_3 = (0, 0, 1, 1, 1, 0, 1, 0)$$
$$f_1 = (0, 1, 1, 1, 0, 0), \quad f_2 = (0, 1, 0, 0, 0, 1, 1, 1), \quad \text{and} \quad f_3 = (0, 1, 0, 0, 0, 1, 1, 1).$$

Counting the color changes, the desired implication follows.

(c) ⇒ **(b)**: Let $w = (w_1, \dots, w_n)$ be an admissible word as in (c) and let $g = (g_1, \dots, g_n)$ be its greedy coloring. To show (b), clearly, it suffices to prove that $\mathcal{G}(w)$ is evenly laminar. If $I(i,j) \in \mathcal{G}(w)$, then $w_i = w_j$ and $g_i \ne g_j \ne g_{j-1}$. By Claims 1 and 2, there is an odd number of color changes between i and j, hence an even number between i and $j-1$, which, by the construction of $\mathcal{G}(w)$, implies that $\mathcal{G}(w)$ contains an even number of sets $I(i',j')$ with $(j'-1) + \frac{1}{2} \in I(i,j)$ that are distinct from $I(i,j)$. Therefore, if $\mathcal{G}(w)$ is laminar, then it is evenly laminar and we may assume, for a contradiction, that $\mathcal{G}(w)$ contains two sets $I(a_1, a_2)$ and $I(b_1, b_2)$ and neither $I(a_1, a_2) \subseteq I(b_1, b_2)$ nor $I(b_1, b_2) \subseteq I(a_1, a_2)$ nor $I(a_1, a_2) \cap I(b_1, b_2) = \emptyset$. We may suppose that

$$a_1 < b_1 < a_2 < b_2. \tag{1}$$

Let us choose the two sets such that a_1 is minimal and, subject to this, b_2 is minimal.

Claim 3. *There is a unique color change between a_2 and b_2, which is at position $(b_2 - 1) + \frac{1}{2}$.*

Proof. Suppose, for a contradiction, that there is a set $I(c_1, c_2)$ in $\mathcal{G}(w)$ with $a_2 < c_2 < b_2$. If $a_2 < c_1$, then, by (1), $a_1 < b_1 < a_2 < c_1 < c_2 < b_2$ and w contains the subword (a, b, a, c, c, b), which is a contradiction. If $a_1 < c_1 < a_2$, then $a_1 < c_1 < a_2 < c_2$, and $c_2 < b_2$ implies a contradiction to the minimality of b_2. If $c_1 < a_1$, then, by (1), $c_1 < b_1 < c_2 < b_2$, and $c_1 < a_1$ implies a contradiction to the minimality of a_1. Hence the claim follows. □

By Claims 1–3, there is a color change between b_1 and $a_2 - 1$, that is, $\mathcal{G}(w)$ contains a set $I(c_1, c_2)$ with $b_1 < c_2 < a_2$. We assume that this set is chosen such that c_2 is maximal. By this choice, we obtain the following.

Claim 4. *There is a unique color change between c_2 and a_2, which is at position $(a_2 - 1) + \frac{1}{2}$.* □

If $c_1 < a_1$, then $c_1 < a_1 < c_2 < a_2$, and $c_1 < a_1$ implies a contradiction to the minimality of a_1. Hence $a_1 < c_1$.

Claim 5. *If I_1 and I_2 are sets in $\mathcal{G}(w)$, then the number of color changes in $(I_1 \setminus I_2) \cup (I_2 \setminus I_1)$ is even.*

Proof. The number of color changes in $(I_1 \setminus I_2) \cup (I_2 \setminus I_1)$ is equal to the number of color changes in I_1 plus the number of color changes in I_2 minus twice the number of color changes in $I_1 \cap I_2$, that is, by Claims 1 and 2, its parity is odd plus odd minus even, which is even. □

Suppose that $b_1 < c_1$. Then, by Claims 3 and 5, there is a set $I(d_1, d_2)$ in $\mathcal{G}(w)$ with $a_1 < d_2 < b_1$. If $d_1 < a_1$, then $d_1 < a_1 < d_2 < a_2$, and $d_1 < a_1$ implies a contradiction to the minimality of a_1. Hence, by (1), $a_1 < d_1 < d_2 < b_1 < c_1 < c_2 < a_2 < b_2$ and w contains the subword (a, d, d, b, c, c, a, b), which is a contradiction.

Finally, suppose that $c_1 < b_1$. Then, by Claims 4 and 5, there is a set $I(d_1, d_2)$ in $\mathcal{G}(w)$ with $a_1 < d_2 < c_1$. If $d_1 < a_1$, then $d_1 < a_1 < d_2 < a_2$, and $d_1 < a_1$ implies a contradiction to the minimality of a_1. Hence, by (1), $a_1 < d_1 < d_2 < c_1 < b_1 < c_2 < a_2 < b_2$ and w contains the subword (a, d, d, c, b, c, a, b), which is a contradiction and completes the proof of this implication.

(b) $\Rightarrow$ **(a)**: This implication follows from Lemma 2 in [1], which states that every odd transversal T of an evenly laminar hypergraph $\mathcal{H}$ contains at least as many elements as there are hyperedges in $\mathcal{H}$, that is, $|T| \geq |E(\mathcal{H})|$. For the sake of completeness, we give a short proof. Possibly by adding a new vertex, we may assume that every hyperedge of $\mathcal{H}$ is a proper subset of the vertex set of $\mathcal{H}$.

Claim 6. *If e is a set of vertices of $\mathcal{H}$ that properly contains k hyperedges of $\mathcal{H}$, then $|e \cap T| \geq k$.*

Proof. We prove this claim by induction on k. For $k = 0$, it is trivial. For $k > 0$, let $e_1, \ldots, e_\ell$ denote the maximal hyperedges properly contained in e. Let k_i for $1 \leq i \leq \ell$ denote the number of hyperedges properly contained in e_i. Clearly, $k = (k_1 + 1) + \cdots + (k_\ell + 1)$. Since $\mathcal{H}$ is evenly laminar, k_i is even for every $1 \leq i \leq \ell$. Since T is an odd transversal, $|e_i \cap T|$ is odd for every $1 \leq i \leq \ell$. Hence, by induction, $|e_i \cap T| \geq k_i + 1$ and we obtain $|T \cap e| \geq |T \cap (e_1 \cup \cdots \cup e_\ell)| = |T \cap e_1| + \cdots + |T \cap e_\ell| \geq (k_1 + 1) + \cdots + (k_\ell + 1) = k$. □

Applying Claim 6 to the vertex set of $\mathcal{H}$ implies $|T| \geq |E(\mathcal{H})|$.

Let $w = (w_1, \ldots, w_n)$ be an admissible word as in (b). It suffices to argue that the greedy coloring $g = (g_1, \ldots, g_n)$ of w has at most as many color changes as any other good coloring $f = (f_1, \ldots, f_n)$ of w. Let

$$G = \left\{ i + \frac{1}{2} \mid 1 \leq i \leq n-1, g_i \neq g_{i+1} \right\} \quad \text{and} \quad F = \left\{ i + \frac{1}{2} \mid 1 \leq i \leq n-1, f_i \neq f_{i+1} \right\}.$$

By definition, we have $|G| = |\mathcal{G}(w)|$. By Claim 2, F is an odd transversal of the evenly laminar hypergraph $\mathcal{G}(w)$. Hence $|F| \geq |\mathcal{G}(w)|$ and thus $|F| \geq |G|$, which completes the proof. □

Clearly, the optimality statements concerning the greedy algorithm proved in [1,5] follow from Theorem 1.

If $w' = (w_1, \ldots, w_n)$ is an admissible word and $g = (g_1, \ldots, g_n)$ is its greedy coloring, then inserting (x_i, x_i) between every two consecutive letters w_i and w_{i+1} of w' with $g_i \neq g_{i+1}$ where the x_i's are new and distinct letters from the alphabet, results in an admissible word w that contains w' as a subword and is optimally colored by the greedy algorithm. This construction implies our final claim.

Claim 7. *It is impossible to characterize the admissible words that are optimally colored by the greedy algorithm using forbidden subwords.* □

References

[1] H. Amini, F. Meunier, H. Michel, A. Mohajeri, Greedy colorings for the binary paintshop problem, Journal of Discrete Algorithms 8 (2010) 8–14.
[2] S.D. Andres, W. Hochstättler, Some heuristics for the binary paint shop problem and their expected number of colour changes, Journal of Discrete Algorithms 9 (2011) 203–211.
[3] P.S. Bonsma, T. Epping, W. Hochstättler, Complexity results on restricted instances of a paint shop problem for words, Discrete Applied Mathematics 154 (2006) 1335–1343.
[4] T. Epping, W. Hochstättler, P. Oertel, Complexity results on a paint shop problem, Discrete Applied Mathematics 136 (2004) 217–226.
[5] F. Meunier, A. Sebő, Paintshop, odd cycles and necklace splitting, Discrete Applied Mathematics 157 (2009) 780–793.