Journal of Global Optimization (2019) 74:705–735
https://doi.org/10.1007/s10898-018-0680-8

A multi-term, polyhedral relaxation of a 0–1 multilinear function for Boolean logical pattern generation

Kedong Yan[1] · **Hong Seo Ryoo**[2]

Received: 22 July 2017 / Accepted: 18 June 2018 / Published online: 25 June 2018

Abstract
0–1 multilinear program (MP) holds a unifying theory to LAD pattern generation. This paper studies a multi-term relaxation of the objective function of the pattern generation MP for a tight polyhedral relaxation in terms of a small number of stronger 0–1 linear inequalities. Toward this goal, we analyze data in a graph to discover useful neighborhood properties among a set of objective terms around a single constraint term. In brief, they yield a set of facet-defining inequalities for the 0–1 multilinear polytope associated with the McCormick inequalities that they replace. The construction and practical utility of the new inequalities are illustrated on a small example and thoroughly demonstrated through numerical experiments with 12 public machine learning datasets.

Keywords Logical analysis of data · Pattern · 0–1 multilinear programming · Multi-term polyhedral relaxation · Facet-defining inequalities · Graph · Star

1 Introduction

Logical Analysis of Data (LAD) is a combinatorial optimization-based supervised learning methodology that has proven useful across many disciplines (e.g., [1–7,12,18,20–23,27]). A key and bottleneck step in LAD is pattern generation where a set of features and their negations are optimally combined together to form knowledge/rule that distinguishes one

Kedong Yan: Most of the research in this paper was conducted while this author was at Korea University..

This work was supported by Samsung Science and Technology Foundation under Project Number SSTF-BA1501-03..

✉ Hong Seo Ryoo
hsryoo@korea.ac.kr

Kedong Yan
yan@njust.edu.cn

[1] Department of Computer Science and Technology, School of Computer Science and Engineering, Nanjing University of Science and Technology, 200 Xiaolingwei, Xuanwu District, Nanjing 210094, Jiangsu, People's Republic of China

[2] School of Industrial Management Engineering, Korea University, 145 Anam-Ro, Seongbuk-Gu, Seoul 02841, South Korea

type of data/observations from the other(s). Without loss of generality, we consider the analysis of two types of + and − data and denote by $S^\bullet$ the index set of $\bullet$ type of data for $\bullet \in \{+, -\}$. Let $S = S^+ \cup S^-$. We assume S is duplicate and contradiction free (such that $S^+ \cap S^- = \emptyset$) and that the data under analysis are described by n Boolean attributes a_j, $j \in N$:=$\{1, \ldots, n\}$. We let $a_{n+j} = \neg a_j$ for $j \in N$ and let N':=$\{n+1, \ldots, 2n\}$ and $\mathcal{N}$:=$N \cup N'$. Finally, for each data A_i, $i \in S$, we denote by A_{ij} the j-th attribute value of the data; such that $A_{ij} = 1 - A_{i,n+j}$ for $j \in N$ and $A_{ij} = 1 - A_{i,j-n}$ for $j \in N'$. Last, since + and − patterns are symmetric in definition, we present the material below in the context of + pattern generation for convenience, without loss of generality.

To build a mathematical model for pattern generation, we introduce 0–1 indicator variables x_j for $j \in \mathcal{N}$ and let

$$x_j = \begin{cases} 1, & \text{if attribute } a_j \text{ is involved in a pattern; and} \\ 0, & \text{otherwise.} \end{cases}$$

For $i \in S$, we let

$$J_i := \{j \in \mathcal{N} \mid A_{ij} = 0\}.$$

Since the dataset is duplicate and contradiction free, all J_i's are unique. Besides, it is easy to see that $|J_i| = n, \forall i \in S$.

In [31], we showed that the 0–1 MP below holds a unifying theory to LAD pattern generation:

$$\text{(PG)} \;:\; \max \left\{ \varphi^+(\mathbf{x}) + l(\mathbf{x}) \;\middle|\; \varphi^-(\mathbf{x}) = 0,\ \mathbf{x} \in \{0,1\}^{2n} \right\}$$

where $l(\mathbf{x})$ is a linear function and

$$\varphi^\bullet(\mathbf{x}) = \sum_{i \in S^\bullet} \prod_{j \in J_i} (1 - x_j)$$

for $\bullet \in \{+, -\}$.

It is well-known that the constraint of (PG) is equivalent to a set of minimal cover inequalities [17]:

$$\sum_{i \in S^-} \prod_{j \in J_i} (1 - x_j) = 0 \iff \sum_{j \in J_i} x_j \geq 1, \quad i \in S^-$$

The minimal cover inequalities provide a poor linear programming (LP) relaxation bound, however. In [32], we studied a convex underestimation of φ^- whose terms correspond to the − observations of the dataset. For the purpose, we analyzed the − data in a graph and discovered useful neighborhood properties among a set of data (forming a hypercube in a graph) and also a set of sets of data (forming a set of pairwise neighboring hypercubes in a graph) that allow for a 'compact' convexification of φ^- in terms of a smaller number of 0–1 linear inequalities that dominate those obtained for φ^- via methods from the literature [9,17]. [32] also studied the Boolean multilinear polytope that results through the new valid inequalities and identified cases where the new inequalities are facet-defining.

Consider the 0–1 linearization of φ^+ now. In note that φ^+ is in maximization form, one only needs a (piece-wise) concave overestimating function to solve (PG) by techniques for 0–1 mixed integer and linear programming (MIP). For this task, McCormick concave envelope for a 0–1 monomial serves the purpose (e.g., [13,26,28,30]) at the expense of introducing m^+ (where $m^+ = |S^+|$) variables

$$y_i = \prod_{j \in J_i} (1 - x_j), i \in S^+ \tag{1}$$

and $n \times m^+$ inequalities

$$y_i \leq 1 - x_j, \quad j \in J_i, \ i \in S^+ \tag{2}$$

to the formulation of a 0–1 linear relaxation of (PG). Alternatively, one may aggregate the constraints in (2) with respect to j to concavify φ^+ by m^+ valid inequalities (e.g., [29])

$$n y_i + \sum_{j \in J_i} x_j \leq n, \quad i \in S^+,$$

but the resulting is a weaker relaxation of φ^+. In addition, this relaxation requires the m^+ additional y_i's to be 0–1 integers. The constraints in (2) can also be aggregated with respect to i via standard probing techniques and logical implications in integer programming (e.g., [14–16]). This yields

$$\sum_{i \in I_j} y_i \leq |I_j|(1 - x_j), \quad j \in \mathcal{N},$$

where $I_j := \{i \in S^+ \mid j \in J_i\}$ for $j \in \mathcal{N}$. Although a weaker relaxation of φ^+ than the standard, McCormick relaxation above, this method reduces the number of constraints by $m^+/2$ times, thus can be useful in data mining applications (eg, see [32]). Last, a method based on 'mapping' from [8] replaces each y_i in $\sum_{i \in S^+} y_i$ by some $1 - x_j$ for $j \in J_i$. As which j to use is unknown, a naive implementation of this mapping method introduces up to $(m^+)^n$ upper bounding functions, leaving that its most efficient form of implementation may be the standard relaxation.

Extending the line of research in [32], this paper aims to compactly 0–1 linearize φ^+ in terms of a smaller number of stronger linear inequalities based on multi-term relaxation of the terms of φ^+. Toward this goal, we now analyze $+$ and $-$ data simultaneously on a graph for useful properties between terms of φ^+ and those in φ^-; that is, between a set of $+$ data and a set of $-$ data. More specifically, on a graph where individual data are represented as nodes, we introduce an edge between each pair of $+$ and $-$ nodes that are 1 Hamming distance apart in n original attributes. Then, we show that each star in a graph (which, by construction, has a $-$ data as the internal node) simultaneously relaxes d terms of φ^+ to generate $n + d$ valid inequalities for (PG) that dominate $n \times d$ McCormick inequalities from the d leaf node(s) of the star. In addition, we show that a collection of 'neighboring' stars generate a much smaller number of stronger valid inequalities for (PG) based on multi-term relaxation of the $+$ terms/data in the star set that can substitute for the McCormick inequalities from the $+$ nodes of the stars considered for a tighter polyhedral relaxation of (PG). Furthermore, we show that our inequalities are facet-defining of the 0–1 multilinear polytope associated with the McCormick inequalities that they replace. Summarizing in different terms, the main results of this paper provide sufficient conditions and algebraic tools for automatically lifting and tightening McCormick inequalities to a fullest extent, in terms of both dimension and tightness. Mathematically and, also, through an illustrative example, we show that the maximum benefit, in regard to the size as well as the tightness of the relaxation model, is realized when 'a maximal set of maximal stars' are exploited for generating new valid inequalities.

As for the organization of this paper, Sect. 2 first presents the background material and then moves on to discovering the aforementioned main results of this paper. Section 3 illustrates the construction and benefits of the new polyhedral concavification scheme for φ^+ with a small artificial dataset of 10 observations while Sect. 4 demonstrates practical utilities of the main results through experiments with 12 public datasets from [24], in comparison with cutting

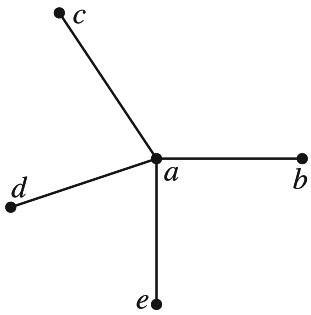


Fig. 1 Degree 4 star with internal node a and 4 leaves b, c, d, e

plane methods implemented in CPLEX [19]. Finally, concluding remarks are provided in Sect. 5.

2 Main results

Similarly as in [32], we engage in polyhedral analysis of feasible region of (PG) via a graph theoretical analysis of data in this paper for a compact and tighter 0–1 linearization of φ^+. The difference is that both $+$ and $-$ data are now analyzed on one graph for the discovery of useful neighborhood properties between a set of $+$ and a set of $-$ data. To obtain a graph representation of a given dataset S, we consider each data as a node in a graph and introduce an edge between a pair of $+$ and $-$ observations if they are 1 Hamming distance apart in n original attributes. This yields an undirected graph G.

Let us recall the definition of a star in a graph (refer Fig. 1) and follow it with the definition of the Hamming distance for measuring similarity between two Boolean data.

Definition 1 A *star* $\mathfrak{S}$ is a tree with one internal/center node adjacent to all other leaf nodes. The degree d of a star refers to the the degree of the internal node; thus, the number of leaves/edges in it.

Definition 2 The *Hamming distance* $H(\mathbf{x}, \mathbf{y})$ between two 0–1 vectors $\mathbf{x}$ and $\mathbf{y}$ is the number of positions in which they differ.

The following provides a means for comparing the relative strength of two valid inequalities for a set of linear inequalities.

Definition 3 If $\pi x \leq \pi_0$ and $\mu x \leq \mu_0$ are two valid inequalities for a polytope in the nonnegative orthant, $\pi x \leq \pi_0$ *dominates* $\mu x \leq \mu_0$ if there exists $u > 0$ such that $u\mu \leq \pi$ and $\pi_0 \leq u\mu_0$, and $(\pi, \pi_0) \neq (u\mu, u\mu_0)$.

Before proceeding, we summarize an important complementary relation between a pair of Boolean attributes/variables that is exploited in obtaining stronger valid inequalities for (PG).

Proposition 1 (Proposition 1 in [32]) *For $j \in \mathcal{N}$, let $j^c = n + j$ if $j \in N$ and $j^c = j - n$ if $j \in N'$. Then, the complementarity cut*

$$x_j + x_{j^c} \leq 1, \tag{3}$$

is valid for (PG).

Let:

$$\mathscr{I}_{PG} := \left\{ \mathbf{x} \in \{0,1\}^{2n},\ \mathbf{y} \in [0,1]^{m^+} \ \middle|\ \varphi^-(\mathbf{x}) = 0,\ (1),\ (3) \right\}$$

The following subsections deal with a tighter polyhedral relaxation of $\mathscr{I}_{PG}$ via a smaller number of stronger valid inequalities and study the multilinear polytope associated with these inequalities.

2.1 Stronger valid inequalities from a star

We begin with a study of a linked pair of nodes in a graph (that is, a pair of + and − data that are 1 Hamming distance away in n original attributes) and then extend the idea to a set of nodes forming a (maximal) star in a graph. For insightful information, we first consider an arbitrary pair of + and − data.

Proposition 2 *For a pair of observations A_i, $i \in S^+$, and A_ℓ, $\ell \in S^-$, the following inequality*

$$y_i \leq \sum_{j \in J_\ell \setminus J_i} x_j \tag{4}$$

is valid for $\mathscr{I}_{PG}$.

Proof As noted earlier, the minimal cover inequality associated with A_ℓ is:

$$\sum_{j \in J_\ell} x_j \geq 1 \tag{5}$$

Note that $|J_\ell \setminus J_i| \in \{1, \ldots, n\}$. Suppose $|J_\ell \setminus J_i| = n$. Then, we have

$$\sum_{j \in J_\ell \setminus J_i} x_j = \sum_{j \in J_\ell} x_j \geq 1 \geq y_i.$$

On the other hand, if $|J_\ell \setminus J_i| \leq n-1$, (2) can be rewritten for A_i as

$$x_j \leq 1 - y_i, \quad j \in J_i$$

and using this in (5) yields:

$$|J_i \cap J_\ell|\,(1 - y_i) + \sum_{j \in J_\ell \setminus J_i} x_j \geq 1$$

Now, with all x_j 0–1 integer variables, y_i is intrinsically binary, too. Finally, this reduces the foregoing inequality to

$$(1 - y_i) + \sum_{j \in J_\ell \setminus J_i} x_j \geq 1$$

and completes the proof. □

An acute reader notes in the proof above that (4) becomes stronger as $H(A_i, A_\ell)$ gets smaller. That is, the strongest inequality of form (4) results when a pair of + and − observations has the Hamming distance equal to 1 in the original attributes such that $|J_\ell \setminus J_i| = 1$. By construction, each such pair has an edge in between in our graph representation of data, and the following is a result that applies to every edge in G.

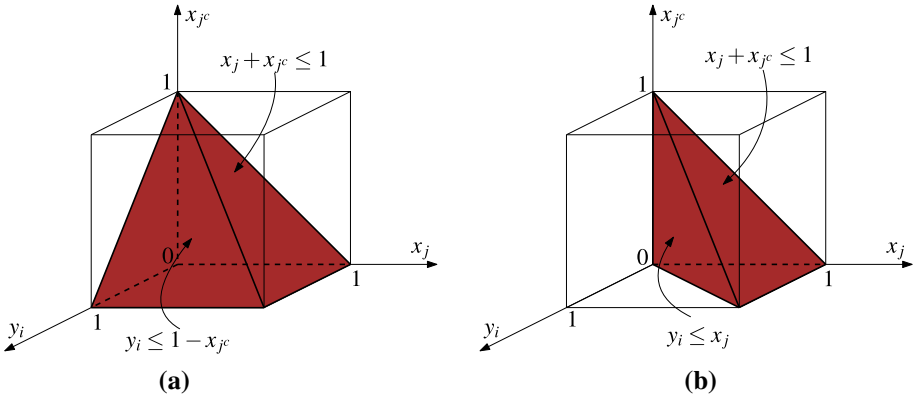


Fig. 2 $y_i \le 1 - x_{j^c}$ (McCormick) versus $y_i \le x_j$ (Lemma 1)

Lemma 1 *For a pair of observations $A_i, i \in S^+$, and $A_\ell, \ell \in S^-$ with $J_\ell \setminus J_i = \{j\}$, the inequality*

$$y_i \le x_j \tag{6}$$

is valid for $\mathscr{I}_{PG}$ and dominates $y_i \le 1 - x_{j^c}$ and $x_j \ge 0$.

Proof $J_\ell \setminus J_i = \{j\}$ indicates that $A_{ij} = 1$ and $A_{\ell j} = 0$ and $J_i \setminus J_\ell = \{j^c\}$, and the validity of (6) is immediate from Proposition 2. Now, (6), along with (3), yields

$$0 \le y_i \le x_j \le 1 - x_{j^c}$$

and completes the proof. □

Remark 1 The strength of (6) primarily lies in that it eliminates the possibilities $x_j = x_{j^c} = 0$ and $y_i = 1$ once for good. This is illustrated in Fig. 2, in comparison with its counterpart McCormick inequality. □

Let us extend the result above for a (maximal) star $\mathfrak{S}$ in G with a $-$ observation A_ℓ as the center node and a set $U (U \subseteq S^+, |U| \ge 2)$ of $+$ observations as its leaves. For notational simplicity, we let

$$\Delta := \bigcup_{i \in U} \{j \in N | A_{ij} \ne A_{\ell j}\}$$

and, for each $j \in \Delta$, denote by i_j the index of $+$ observation in U that is different from A_ℓ in the j-th position; that is, $A_{i_j j} \ne A_{\ell j}, A_{i_j j'} = A_{\ell j'}, \forall j' \in N \setminus \{j\}$.

Theorem 1 *Consider a star $\mathfrak{S}$ formed by one $-$ observation A_ℓ and a set $U (U \subseteq S^+, d = |U| \ge 2)$ of $+$ observations. Then, for each $j \in J^\star := \bigcap_{i \in U} J_i \cap J_\ell (\ne \emptyset)$, the inequality*

$$\sum_{i \in U} y_i \le 1 - x_j \tag{7}$$

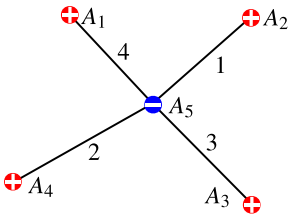
is valid for $\mathscr{I}_{PG}$ and dominates $y_i \le 1 - x_j$ for all $i \in U$.

Proof Consider the McCormick concave envelope for each $y_i, i \in U$:

$$y_i \le 1 - x_j, \quad j \in J_i$$

•	Observation i	A_{ij}	J_i	$J_5 \setminus J_i$
+	A_1	01110	$\{1,5,7,8,9\}$	$\{4\}$
	A_2	11100	$\{4,5,6,7,8\}$	$\{1\}$
	A_3	01000	$\{1,3,4,5,7\}$	$\{8\}$
	A_4	00100	$\{1,2,4,5,8\}$	$\{7\}$
−	A_5	01100	$\{1,4,5,7,8\}$	

(a)



(b)

Fig. 3 Data for Example 1. **a** Observations under analysis **b** A star $\mathfrak{S}$ formed by the five observations

It is easy to see that if $x_j = 1$ for some $j \in J^\star$, then the minimal cover inequality (5) from A_ℓ is satisfied, and $y_i = 0, \forall i \in U$. Hence, we have:

$$\sum_{i \in U} y_i = 0 \leq 1 - x_j, \quad j \in J^\star$$

Now suppose $x_j = 0, \forall j \in J^\star$. Without loss of generality, let $A_{\ell j} = 0, \forall j \in \Delta$. For some $\iota \in \Delta$, by setting $x_\iota = 1$, (5) is satisfied, and $y_i = 0, \forall i \in U \setminus \{i_\iota\}$ since $A_{i\iota} = 0$ (which implies $y_i \leq 1 - x_\iota$), $\forall i \in U \setminus \{i_\iota\}$. In addition, it follows from Lemma 1 that $y_{i_\iota} \leq x_\iota = 1$. Thus,

$$\sum_{i \in U} y_i \leq 1 = 1 - x_j, \quad j \in J^\star.$$

The dominance relation is immediate from $y_i \leq \sum_{i \in U} y_i$ for all $i \in U$. □

We briefly pause for an example to help the reader become acquainted with the notation of this paper and also to illustrate how the results thus far work.

Example 1 [Illustration of Lemma 1 and Theorem 1] Consider a dataset in Fig. 3a that form a star $\mathfrak{S}$ in Fig. 3b, where $J_\ell \setminus J_i$ is indicated for each $i \in \{1, \ldots, 4\}$. Applying the standard McCormick concave relaxation, we obtain 20 inequalities below:

$$\begin{aligned}
&y_1 \leq 1 - x_1, \quad y_1 \leq 1 - x_5, \quad y_1 \leq 1 - x_7, \quad y_1 \leq 1 - x_8, \quad y_1 \leq 1 - x_9 \\
&y_2 \leq 1 - x_4, \quad y_2 \leq 1 - x_5, \quad y_2 \leq 1 - x_6, \quad y_2 \leq 1 - x_7, \quad y_2 \leq 1 - x_8 \\
&y_3 \leq 1 - x_1, \quad y_3 \leq 1 - x_3, \quad y_3 \leq 1 - x_4, \quad y_3 \leq 1 - x_5, \quad y_3 \leq 1 - x_7 \\
&y_4 \leq 1 - x_1, \quad y_4 \leq 1 - x_2, \quad y_4 \leq 1 - x_4, \quad y_4 \leq 1 - x_5, \quad y_4 \leq 1 - x_8
\end{aligned}$$

Applying Lemma 1 for A_1 and A_5, we obtain

$$y_1 \leq x_4$$

and this inequality dominates its counterpart McCormick inequality

$$y_1 \leq 1 - x_9.$$

Likewise, (6) in Lemma 1 yields

$$y_2 \leq x_1, \quad y_3 \leq x_8 \quad \text{and} \quad y_4 \leq x_7$$

which, respectively, dominate

$$y_2 \leq 1 - x_6, \quad y_3 \leq 1 - x_3 \quad \text{and} \quad y_4 \leq 1 - x_2.$$

To apply the result in Theorem 1, we note that $U = \{1, 2, 3, 4\}$ and $J^\star = \{5\}$. Therefore, via (7), we obtain

$$y_1 + y_2 + y_3 + y_4 \leq 1 - x_5.$$

As seen, the above inequality dominates the four inequalities below by the standard linearization method:

$$y_1 \leq 1 - x_5, \quad y_2 \leq 1 - x_5, \quad y_3 \leq 1 - x_5 \text{ and } y_4 \leq 1 - x_5$$

Remark 2 The inequality in (7) resembles the inequality $\sum_{i \in I_j} y_i \leq |I_j|(1 - x_j)$ that is obtained for $j \in \mathcal{N}$ via a simple aggregation based on standard probing techniques and logical implications in integer programming (e.g., [14–16]). The difference is that the former involves less y_i's but has 1 as the coefficient of $1 - x_j$ in the right-hand side of the inequality. An acute reader may note that (7) is an inequality that results when the individual McCormick/standard inequalities from the + leaf nodes of the star are lifted and tightened to the fullest extent. This insight explains why the new inequality from a star is stronger than the individual inequalities used for obtaining it, as illustrated in the preceding example. □

When a leaf node of a star in Theorem 1 is purposefully deleted, we obtain a valid inequality in one less y variable and a different x variable than the one in (7) that is non-dominated by (7) and dominates the set of McCormick inequalities from the remaining leaf nodes (data) of the reduced star. We summarize this result next.

Corollary 1 *Consider a star described in Theorem 1 with $d \geq 3$. For each $\iota \in \Delta$, let $J_\ell \setminus J_{i_\iota} = \{j\}$, the inequality*

$$\sum_{i \in U \setminus \{i_\iota\}} y_i \leq 1 - x_j \tag{8}$$

is valid for $\mathscr{I}_{PG}$ and dominates $y_i \leq 1 - x_j$ for all $i \in U \setminus \{i_\iota\}$.

Proof Consider the star obtained from $\mathfrak{S}$ by removing a leaf A_{i_ι}. The result is immediate from Theorem 1. □

A few remarks are due.

Remark 3 In Corollary 1, $j = \iota$ if $A_{\ell\iota} = 0$; while $j = \iota + n$ if $A_{\ell\iota} = 1$ (refer Fig. 3). □

Remark 4 Note that if $d = 1$ in Theorem 1 or $d = 2$ in Corollary 1, (7) and (8) reduce to the inequalities given by the standard McCormick relaxation method, respectively. The reader can verify that these two inequalities are made the strongest when a maximal star is considered for their construction; the resulting involves the maximum number of y variables in the left-hand side.

Furthermore, note for any star $\mathfrak{S}$ that Δ is fixed. Therefore, removing more than a single observation from U, as done so in Corollary 1, produces a valid inequality that has a less number of y_i's in the left-hand side than (8), thus is dominated by the latter. □

Example 2 (Illustration of Corollary 1) Let us recall the data in Fig. 3. Note that $\Delta = \bigcup_{i \in U} \{j \in N | A_{ij} \neq A_{5j}\} = \{1, 2, 3, 4\}$. For $\iota = 1$, we have $i_1 = \{2\}$, thus $J_5 \setminus J_2 = \{1\}$. So, we obtain via (8)

$$y_1 + y_3 + y_4 \leq 1 - x_1$$

which dominates

$$y_1 \le 1 - x_1, \quad y_3 \le 1 - x_1 \ \text{ and } \ y_4 \le 1 - x_1.$$

In the same fashion, one uses Corollary 1 for other ι's of Δ to obtain

$$y_1 + y_2 + y_3 \le 1 - x_7, \quad y_1 + y_2 + y_4 \le 1 - x_8 \ \text{ and } \ y_2 + y_3 + y_4 \le 1 - x_4$$

which dominate the set of McCormick inequalities involving x_7, x_8 and x_4, respectively.

In summary, for analyzing the dataset in Fig. 3, one can use the 9 valid inequalities in Example 1 and above in place of the 20 McCormick inequalities to obtain a tighter polyhedral overestimation of φ^+. □

For analyzing the strength of the polyhedral inequalities above, we let

$$X := \left\{ x_j \in \{0,1\},\ j \in \mathcal{N},\ \ y_i \in [0,1], i \in U \ \middle|\ (3),\ (5),\ y_i \le 1 - x_j,\ j \in J_i, i \in U \right\}.$$

Then, the following holds (and we defer its proof until the end of the next subsection).

Theorem 2 *For a star $\mathfrak{S}$ described in Theorem 1, (7) and (8) define facets of $conv(X)$.*

2.2 Stronger valid inequalities from a set of stars

Let us consider a set of q (≥ 2) (maximal) stars $\mathfrak{S}_1, \ldots, \mathfrak{S}_q$ and let A_{ℓ_k} denote the internal node of $\mathfrak{S}_k$ and U_k denote the set of $+$ observations (that is, the leaf nodes) of $\mathfrak{S}_k$, $k \in Q := \{1, \ldots, q\}$. Further, we let

$$\Delta_k := \bigcup_{i \in U_k} \{ j \in N \mid A_{ij} \ne A_{\ell_k j} \}, \quad k \in Q$$

for each star $\mathfrak{S}_k$. That is, Δ_k collects the indices where individual leaf nodes of the star $\mathfrak{S}_k$ are different from the center in n original attributes. We also recall that each A_{ℓ_k} corresponds to a term in the constraint function φ^- of (PG), which is 0–1 linearized by the standard method into a minimal cover inequality [17] as follows:

$$\sum_{j \in J_{\ell_k}} x_j \ge 1, \quad k \in Q \tag{9}$$

Finally, for a pair of stars, we let

$$\Delta_{kl} := \{ j \in N \mid A_{\ell_k j} \ne A_{\ell_l j} \}, \quad k, l \in Q$$

to store information on attribute(s) where the two internal nodes of the stars differ (in the original attributes).

With the notation above, consider the following result.

Theorem 3 *Consider a set Q of q (≥ 2) stars. For any $k, l \in Q$ and $j' \in \Delta_k$ and $j'' \in \Delta_l$, suppose that $i_{j'} \ne i_{j''}$ and neither of the two conditions below holds:*

i) $j' \ne j''$ and $j', j'' \in \Delta_{kl}$
ii) $j' = j''$ and $j', j'' \notin \Delta_{kl}$

Then, for each $j \in \mathfrak{J} := \bigcap_{k \in Q} J_k^\star$ ($\ne \emptyset$), where $J_k^\star := \bigcap_{i \in U_k} J_i \cap J_{\ell_k}$, the inequality

$$\sum_{i \in \mathscr{U}} y_i \le 1 - x_j \tag{10}$$

is valid for $\mathscr{I}_{PG}$, where $\mathscr{U} := \bigcup_{k \in Q} U_k$.

Proof If $x_j = 1$ for any $j \in \mathfrak{J}$, then (9) is satisfied and $y_i = 0, \forall i \in \mathscr{U}$. Therefore, $\sum_{i\in\mathscr{U}} y_i = 0$, and (10) is trivially satisfied.

On the other hand, if $x_j = 0, \forall j \in \mathfrak{J}$, we need to show that

$$\sum_{i\in U_k} y_i + \sum_{i\in U_l} y_i \leq 1 \tag{11}$$

holds for any pair $k, l \in Q$ satisfying the conditions of the theorem. Equivalently, it suffices to show that, for every $j' \in \Delta_k$ and $j'' \in \Delta_l$ ($i_{j'} \neq i_{j''}$) satisfying neither of i) and ii), we have:

$$y_{i_{j'}} + y_{i_{j''}} \leq 1 \tag{12}$$

(As for reason, via (7) in Theorem 1, we have: $\sum_{i\in U_k} y_i \leq 1$ and $\sum_{i\in U_l} y_i \leq 1$. Thus, $y_{i_{j'}} = 1$, for instance, indicates that $y_i = 0, \forall i \in U_k \setminus \{i_{j'}\}$ and $y_i = 0, \forall i \in U_l$ via (12).)

For a pair of stars under consideration, the minimal cover inequalities in (9) associated with A_{ℓ_k} and A_{ℓ_l} also need to be satisfied by a 0–1 feasible solution to $\mathscr{I}_{PG}$; that is, at least one x in them needs to be set to 1. Without loss of generality, assume $A_{\ell_k j'} = 0$ and consider the two possible cases for $j' \in \Delta_k$ and $j'' \in \Delta_l$ with $i_{j'} \neq i_{j''}$.

Case 1 ($j' \neq j''$). In this case we have $j' \notin \Delta_{kl}$ or $j'' \notin \Delta_{kl}$. Without loss of generality, assume $j' \notin \Delta_{kl}$. This yields

$$A_{\ell_l j'} = 0, \;\; A_{i_{j'} j'} = 1 \text{ and } A_{i_{j''} j'} = 0.$$

Thus, setting $x_{j'} = 1$ satisfies (9) for ℓ_k and ℓ_l and yields

$$y_{i_{j'}} \leq x_{j'} = 1, \;\; y_{i_{j''}} \leq 1 - x_{j'} = 0,$$

which satisfies (12).

Case 2 ($j' = j''$). In this case, we have $j' (= j'') \in \Delta_{kl}$ such that $A_{\ell_l j'} = 1$, $A_{i_{j'} j'} = 1$ and $A_{i_{j''} j'} = 0$. Therefore, by setting $x_{j'} = 1$, we have (9) satisfied for ℓ_k and obtain

$$y_{i_{j'}} \leq x_{j'} = 1, \;\; y_{i_{j''}} \leq x_{n+j''} = x_{n+j'} = 0.$$

(9) is also satisfied for ℓ_l; depending on which $\mathbf{x}$ is set to 1, we either have $y_{i_{j'}} \leq 1$ or $y_{i_{j'}} \leq 0$. Therefore, we have (12) satisfied for this case as well.

Concluding, with (11) holding for each pair of stars satisfying the conditions of the theorem, at most one $y_i, i \in \mathscr{U}$ can take value 1. Along with $x_j = 0, \forall j \in \mathfrak{J}$, this shows that $\sum_{i\in\mathscr{U}} y_i \leq 1 = 1 - x_j, \; \forall j \in \mathfrak{J}$ and completes the proof. □

Remark 5 Briefly, if either condition of Theorem 3 is satisfied, the right-hand side value of the inequality in (11) becomes larger than 1 for some pair of stars, thereby making (10) invalid for $\mathscr{I}_{PG}$. □

Now, we let

$$\Xi := \left\{ x_j \in \{0, 1\}, j \in \mathscr{N}, y_i \in [0, 1], i \in \mathscr{U} \;\middle|\; (3), (9), \; y_i \leq 1 - x_j, \; j \in J_i, \; i \in \mathscr{U} \right\}$$

and analyze the strength of each of the new inequalities in (10) for $j \in \mathfrak{J}$.

Theorem 4 *For a set of stars in Theorem 3,* (10) *defines a facet of conv*(Ξ).

Proof The validity of (10) was established in the proof of Theorem 3.

Suppose now that (10) is not facet-defining. Then, there exists a facet-defining inequality of Ξ of the form

$$\sum_{j\in\mathscr{N}} \alpha_j x_j + \sum_{i\in\mathscr{U}} \beta_i y_i \le \gamma, \tag{13}$$

where $(\boldsymbol{\alpha}, \boldsymbol{\beta}) \neq (\mathbf{0}, \mathbf{0})$, such that (10) defines a face of the facet of $conv(\Xi)$ defined by (13). That is,

$$F_1 := \left\{ (\mathbf{x}, \mathbf{y}) \in \Xi \;\middle|\; x_j + \sum_{i\in\mathscr{U}} y_i = 1 \right\} \subseteq F := \left\{ (\mathbf{x}, \mathbf{y}) \in \Xi \;\middle|\; \sum_{j\in\mathscr{N}} \alpha_j x_j + \sum_{i\in\mathscr{U}} \beta_i y_i = \gamma \right\}.$$

Consider the following two cases for the solutions in F_1.
Case 1 ($x_j = 1$). Since $j \in \mathfrak{J}$, this solution satisfies (9). The solution with $x_{j'} = 0, \forall j' \in \mathscr{N} \setminus \{j\}$ and $y_i = 0, \ \forall i \in \mathscr{U}$ belongs to F_1 hence F, which yields $\alpha_j = \gamma$. For this solution, suppose $x_{j'} = 1$ for some $j' \in \mathscr{N} \setminus \{j, j^c\}$. This yields $\alpha_j + \alpha_{j'} = \gamma$, thus:

$$\alpha_j = \gamma \ \text{ and } \ \alpha_{j'} = 0, \ \forall j' \in \mathscr{N} \setminus \{j, j^c\}$$

Case 2 ($x_j = 0$). The existence of a pattern is guaranteed for a contradiction-free dataset. This implies that there always exists a 0–1 vector $\mathbf{x}$ that yields $y_i = 1$ for each $i \in \mathscr{U}$. Also, note that the value of x_{j^c} does not affect $\mathbf{y}$; thus, $\beta_i = \beta_i + \alpha_{j^c} = \gamma$ for all $i \in \mathscr{U}$. These yield

$$\alpha_{j^c} = 0 \ \text{ and } \ \beta_i = \gamma, \ \forall i \in \mathscr{U}.$$

Summarizing, the two cases above show that

$$\alpha_{j'} = 0, \ \forall j' \in \mathscr{N} \setminus \{j\}; \ \text{ and } \ \alpha_j = \beta_i = \gamma > 0, \ \forall i \in \mathscr{U},$$

where $\gamma > 0$ is from our supposition that (10) is dominated by (13). This indicates that (13) is a positive multiple of (10) (which is a contradiction) and completes the proof. □

For a set of stars $\mathfrak{S}_k, k \in Q$, let

$$\Delta_\otimes := \bigcup_{k\in Q} \Delta_k \ \text{ and } \ \Delta_\odot := \bigcup_{k,l\in Q,\, k\neq l} \Delta_{kl}$$

and consider the following result.

Theorem 5 *Consider a set of stars in Theorem 3. For each $\iota \in \Delta_\otimes \setminus \Delta_\odot$, assume without loss of generality that $\iota \in \Delta_k$, $k \in Q$, and let $i_\iota \in \mathscr{U}$ be such that $J_{\ell_k} \setminus J_{i_\iota} = \{j\}$. Then, the inequality*

$$\sum_{i\in\mathscr{U}\setminus\{i_\iota\}} y_i \le 1 - x_j \tag{14}$$

is valid for Ξ and defines a facet of $conv(\Xi)$.

Proof The validity of (14) for Ξ is immediate from Theorem 3 for the sub-graph with A_{i_ι} removed from the set of stars for the theorem. As for the facet result, we suppose the contrary that (14) is not facet-defining for $conv(\Xi)$. This suggests

$$F_2 := \left\{ (\mathbf{x}, \mathbf{y}) \in \Xi \;\middle|\; x_j + \sum_{i\in\mathscr{U}\setminus\{i_\iota\}} y_i = 1 \right\} \subseteq F,$$

where F is given by a facet-defining inequality of $conv(\Xi)$ of the form (13), as in the proof of Theorem 3. Now, consider the following two cases for any solution of F_2.

Case 1 ($x_j = 1$). Note that $j \in J^{\cap} := \cap_{k \in Q} J_{\ell_k}$, thus this solution satisfies (9). The solution with $x_{j'} = 0,\ \forall j' \in \mathscr{N} \setminus \{j\}$ and $y_i = 0, \forall i \in \mathscr{U}$ belongs to F_2, hence F. This yields $\alpha_j = \gamma$. For this solution, we can set $x_{j'} = 1$ for each $j' \in \mathscr{N} \setminus \{j, j^c\}$ to obtain $\alpha_j + \alpha_{j'} = \gamma$. We thus have

$$\alpha_j = \gamma \ \text{ and } \ \alpha_{j'} = 0,\ \forall j' \in \mathscr{N} \setminus \{j, j^c\}.$$

Moreover, as noted in the proof of Theorem 4, the existence of a pattern for a contradiction-free dataset guarantees the existence of $\mathbf{x}$ such that $y_{i_\iota} = 1$. This yields:

$$\alpha_j + \beta_{i_\iota} = \gamma \quad \Longrightarrow \quad \beta_{i_\iota} = 0$$

Case 2 ($x_j = 0$). Again, there exists a 0–1 vector $\mathbf{x}$ that yields $y_i = 1$ for each $i \in \mathscr{U} \setminus \{i_\iota\}$. If $x_{j^c} = 0$ (or $x_{j^c} = 1$) in such a solution, we have $\beta_i = \gamma$ (or $\alpha_{j^c} + \beta_i = \gamma$) for $i \in \mathscr{U} \setminus \{i_\iota\}$, hence obtain:

$$\alpha_{j^c} = 0 \ \text{ and } \ \beta_i = \gamma,\ \forall i \in \mathscr{U} \setminus \{i_\iota\}$$

By the two cases above, we have

$$\alpha_{j'} = 0,\ \forall j' \in \mathscr{N} \setminus \{j\};\ \beta_{i_\iota} = 0; \ \text{ and } \ \alpha_j = \beta_i = \gamma > 0,\ \forall i \in \mathscr{U} \setminus \{i_\iota\},$$

where $\gamma > 0$ is from the supposition that (14) is dominated by (13). This shows that (13) is a positive multiple of (14) and completes the proof. □

We have another interesting result below.

Theorem 6 *Consider a set of stars in Theorem 3. For each $\iota \in \Delta_{\odot}$, let:*

$$\mathscr{U}_0^\iota := \{i \in \mathscr{U} \,|\, A_{i\iota} = 0\} \quad \textit{and} \quad \mathscr{U}_1^\iota := \{i \in \mathscr{U} \,|\, A_{i\iota} = 1\}$$

Then, a pair of inequalities

$$\sum_{i \in \mathscr{U}_0^\iota} y_i \le 1 - x_\iota \quad \textit{and} \quad \sum_{i \in \mathscr{U}_1^\iota} y_i \le 1 - x_{\iota+n} \tag{15}$$

are valid for Ξ and define facets of $conv(\Xi)$.

Proof Again, the validity result is immediate; apply Theorem 3 to the two sub-graphs, one with $A_i, i \in \mathscr{U}_0^\iota$ and the other $A_i, i \in \mathscr{U}_1^\iota$, respectively. As for the facet result, since the inequalities in (15) are symmetric, if suffices to show it for the former. For the purpose, we again suppose the contrary to let

$$F_3 := \left\{ (\mathbf{x}, \mathbf{y}) \in \Xi \;\middle|\; x_\iota + \sum_{i \in \mathscr{U}_0^\iota} y_i = 1 \right\} \subseteq F$$

(where F is as defined in Theorem 3) and investigate the two possible cases for the solutions of F_3.

Case 1 ($x_\iota = 1$). The solution with $x_j = 1$ for some $j \in \mathfrak{J}$, $x_{j'} = 0, \forall j' \in \mathscr{N} \setminus \{\iota, j\}$ and $y_i = 0, \forall i \in \mathscr{U}$ satisfies (9), hence belongs to F_3 and F. This yields $\alpha_j + \alpha_\iota = \gamma,\ j \in \mathfrak{J}$. Let $J := J^{\cap} \setminus \{\mathfrak{J}\}$ (where $J^{\cap} = \cap_{k \in Q} J_{\ell_k}$) and note that the solution with $x_{j'} = 1$ for some $j' \in J$, $x_{j''} = 0, \forall j'' \in \mathscr{N} \setminus \{j', \iota\}$ and $y_i = 0, \forall i \in \mathscr{U}$ also satisfies (9), hence belongs

to F_3 and F. This yields $\alpha_{j'} + \alpha_\iota = \gamma,\ j' \in J$. Therefore, the combination of these two solutions yields $\alpha_j + \alpha_{j'} + \alpha_\iota = \gamma,\ j \in \mathfrak{J}, j' \in J$; hence $\alpha_\iota = \gamma,\ \alpha_j = 0,\ j \in J^\cap$. By setting $x_{j''} = 1$ for each $j'' \in \mathscr{N} \setminus (J^\cap \cup \{\iota^c\})$ in the first solution above, we obtain:

$$\alpha_\iota + \alpha_{j''} = \gamma \quad \Longrightarrow \quad \alpha_\iota = \gamma, \ \text{ and } \ \alpha_j = 0, \forall j \in \mathscr{N} \setminus \{\iota, \iota^c\}$$

Note that $x_\iota = 1$ implies $y_i = 0, \forall i \in \mathscr{U}_0^\iota$. Additionally, there is a 0–1 vector $\mathbf{x}$ that yields $y_i = 1$ for each $i \in \mathscr{U}_1^\iota$. Therefore, we have:

$$\alpha_\iota + \beta_i = \gamma,\ i \in \mathscr{U}_1^\iota \quad \Longrightarrow \quad \beta_i = 0, \forall i \in \mathscr{U}_1^\iota$$

Case 2 ($x_\iota = 0$). Again, for a 0–1 vector $\mathbf{x}$ that yields $y_i = 1$ for each $i \in \mathscr{U}_0^\iota$, if $x_{\iota^c} = 0$ (or $x_{\iota^c} = 1$), we have $\beta_i = \gamma$ (or $\alpha_{\iota^c} + \beta_i = \gamma$) for $i \in \mathscr{U}_0^\iota$. Therefore,

$$\alpha_{\iota^c} = 0 \ \text{ and } \ \beta_i = \gamma, \forall i \in \mathscr{U}_0^\iota.$$

The two cases above yield

$$\alpha_j = 0, \forall j \in \mathscr{N} \setminus \{\iota\};\ \ \beta_i = 0, \forall i \in \mathscr{U}_1^\iota;\ \ \text{and} \ \ \alpha_\iota = \beta_i = \gamma > 0, \forall i \in \mathscr{U}_0^\iota$$

where $\gamma > 0$ is from our supposition that (15) is dominated by (13). This shows that (13) is a positive multiple of (15) and completes the proof. □

Remark 6 Denote by d_k the degree of (the internal node of) $\mathfrak{S}_k$ for $k \in Q$. Then, the number of inequalities generated from a set of stars by Theorem 3 is equal to $|\mathfrak{J}| = n - |\Delta_\otimes \cup \Delta_\odot|$. Likewise, the number of inequalities generated via Theorems 5 and 6 are calculated to be $|\Delta_\otimes \setminus \Delta_\odot|$ and $2|\Delta_\odot|$, respectively. Therefore, the total number of inequalities from a set of stars is:

$$\sum_{k \in Q} d_k + n - |\Delta_\otimes \cup \Delta_\odot| + |\Delta_\otimes \setminus \Delta_\odot| + 2\,|\Delta_\odot| = \sum_{k \in Q} d_k + n + |\Delta_\odot|$$

□

Remark 7 As hinted in Remark 2 with a star inequality, the main results of this paper can be understood as providing sufficient conditions and mechanisms for 'algebraic lifting and coefficient tightening' a maximal set of McCormick inequalities to a fullest extent. By fullest, we mean two things. One, the number of y variables in each new inequality is maximized. The other, the coefficient of each and every decision variables $\mathbf{x} \in \{0, 1\}^{2n}$ and $\mathbf{y} \in [0, 1]^{m^+}$ is kept at 1, which attributed to obtaining facet-defining results in Theorems 2, 4, 5, and 6. Furthermore, as understood by now, the maximum benefit is realized when a maximal set of maximal stars are exploited for generating the inequalities of this subsection (refer Remark 4).

As a summary, the dominance relation among the inequalities dealt with and developed in this paper is captured into the hierarchy of their relative strength in Fig. 4. □

Last, we note that Theorem 1 and Corollary 1 are special cases of Theorems 3 and 5, respectively, when Q is a singleton. This yields the following proof for Theorem 2 as an immediate consequence of the results in Theorems 4 and 5.

Proof (of Theorem 2). Since Q is a singleton, we have $\Delta_\otimes = \Delta$, $\Delta_\odot = \emptyset$ and $U = \mathscr{U}$. Use these in Theorem 4 for (7) and Theorem 5 for (8), respectively. □

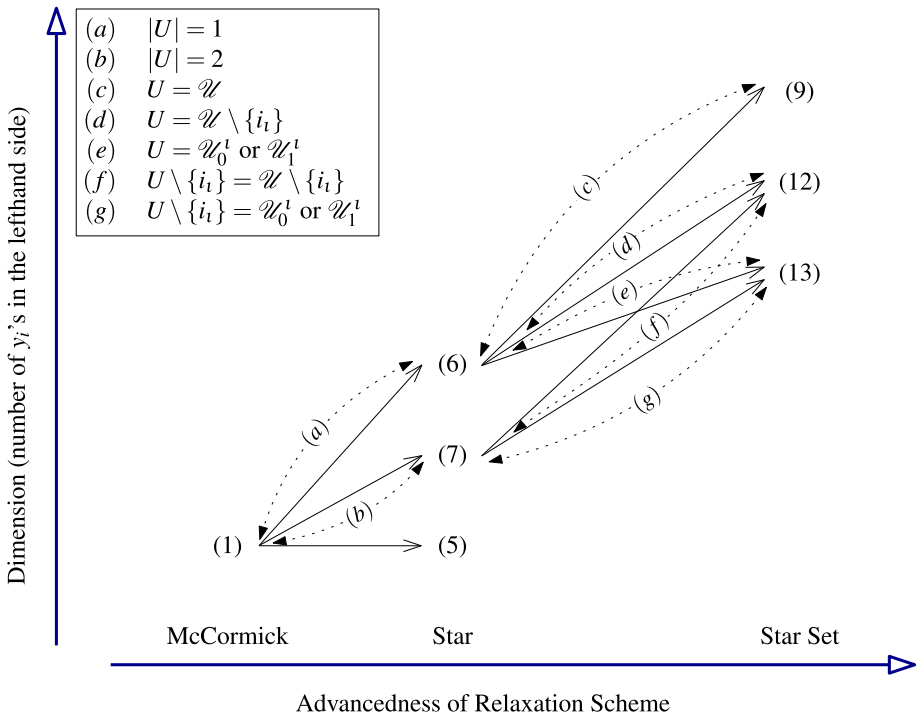


Fig. 4 Dominance relation among valid inequalities for φ^+ of (PG). '$\longrightarrow$' indicates dominance of the head over the tail, while the dotted bidirectional arrow indicates that the two are equivalent; specifically, the one placed northeast reduces to the other under the condition provided as the legend

3 Illustrative example: construction and strength/utility of main results

Consider a set of 6 + observations and 4 − observations in Table 1 that are described in 10 0–1 attributes; that is, $n = 10$, thus $N = \{1 \ldots, 10\}$, $N' = \{11 \ldots, 20\}$ and $\mathscr{N} = \{1, \ldots, 20\}$. For convenience in presentation, we let $S^+ = \{1, \ldots, 6\}$ and $S^- = \{7, \ldots, 10\}$.

When analyzed on a graph, these 10 observations comprise 4 stars, namely, $\mathfrak{S}_1$, $\mathfrak{S}_2$, $\mathfrak{S}_3$ and $\mathfrak{S}_4$, as shown in Fig. 5 and supplemented by Table 1. (We note with A_5 as an example that a + data can belong to multiple stars.)

For easier referencing, we connect each pair of the internal nodes of the 4 stars by a dashed line and provide Δ_{kl} information (that is, the indices where the two internal nodes of the stars differ) along the dashed edge. For each star $\mathfrak{S}_k$, $k \in Q$:={1, 2, 3, 4}, the number on an edge between a leaf node and the center is the index where the pair of + and − observations differ. Therefore, Δ_k in Table 1 for each of the four stars is obtained by simply collecting the indices indicated along the edges of the star.

To illustrate how the inequalities of this paper are generated for this dataset, we first take the 7 edges of the graph and use (6) of Lemma 1 to obtain 7 valid inequalities

$$y_1 \le x_1, \quad y_2 \le x_2, \quad y_3 \le x_{17}, \quad y_4 \le x_3,$$
$$y_5 \le x_{15}, \quad y_5 \le x_{16}, \quad y_6 \le x_4$$

Table 1 Illustrative (PG) Dataset

+ Observation i	$A_{ij}(j \in N)$, J_i		
1	$A_1 = [1000000000]$		
	$J_1 = \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11\}$		
2	$A_2 = [0100000110]$		
	$J_2 = \{1, 3, 4, 5, 6, 7, 10, 12, 18, 19\}$		
3	$A_3 = [0000100100]$		
	$J_3 = \{1, 2, 3, 4, 6, 7, 9, 10, 15, 18\}$		
4	$A_4 = [0010101100]$		
	$J_4 = \{1, 2, 4, 6, 9, 10, 13, 15, 17, 18\}$		
5	$A_5 = [0000001100]$		
	$J_5 = \{1, 2, 3, 4, 5, 6, 9, 10, 17, 18\}$		
6	$A_6 = [0001011100]$		
	$J_6 = \{1, 2, 3, 5, 9, 10, 14, 16, 17, 18\}$		
− Observation i	$A_{ij}(j \in N)$, J_i	Star k	$J_k^\star, U_k, \Delta_k$
7	$A_7 = [0000000000]$	$\mathfrak{S}_1$	$J_1^\star = \{2, 3, 4, 5, 6, 7, 8, 9, 10\}$
	$J_7 = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$		$U_1 = \{1\}$, $\Delta_1 = \{1\}$
8	$A_8 = [0000011100]$	$\mathfrak{S}_2$	$J_2^\star = \{1, 2, 3, 5, 9, 10, 17, 18\}$
	$J_8 = \{1, 2, 3, 4, 5, 9, 10, 16, 17, 18\}$		$U_2 = \{5, 6\}$, $\Delta_2 = \{4, 6\}$
9	$A_9 = [0000000110]$	$\mathfrak{S}_3$	$J_3^\star = \{1, 3, 4, 5, 6, 7, 10, 18, 19\}$
	$J_9 = \{1, 2, 3, 4, 5, 6, 7, 10, 18, 19\}$		$U_3 = \{2\}$, $\Delta_3 = \{2\}$
10	$A_{10} = [0000101100]$	$\mathfrak{S}_4$	$J_4^\star = \{1, 2, 4, 6, 9, 10, 18\}$
	$J_{10} = \{1, 2, 3, 4, 6, 9, 10, 15, 17, 18\}$		$U_4 = \{3, 4, 5\}$, $\Delta_4 = \{3, 5, 7\}$

which, respectively, dominate

$$y_1 \le 1 - x_{11}, \quad y_2 \le 1 - x_{12}, \quad y_3 \le 1 - x_7, \quad y_4 \le 1 - x_{13},$$
$$y_5 \le 1 - x_5, \quad y_5 \le 1 - x_6, \quad y_6 \le 1 - x_{14}.$$

The user can verify that the inequalities generated by Lemma 1 are non-dominated by any other inequality, both new and old.

To illustrate how to generate a set of inequalities from individual stars, we consider $\mathfrak{S}_4$, comprised of the internal node A_{10} and three leaf nodes A_3, A_4 and A_5, hence $U_4 = \{3, 4, 5\}$. First, we collect the information of the edges of the graph to obtain $\Delta_4 = \{3, 5, 7\}$ $(= \bigcup_{i \in U_4} \{j \in N \mid A_{ij} \neq A_{10,j}\})$ and set $J_4^\star = J_3 \cap J_4 \cap J_5 \cap J_{10} = \{1, 2, 4, 6, 9, 10, 18\}$ (refer Table 1 for information on J_i, $i \in U_4$). Thus, (7) of Theorem 1 yields that

$$y_3 + y_4 + y_5 \le 1 - x_j$$

dominates

$$y_i \le 1 - x_j, \quad \forall i \in U_4 = \{3, 4, 5\}$$

for every $j \in J_4^\star = \{1, 2, 4, 6, 9, 10, 18\}$.

Now, applying Corollary 1 with $\Delta_4 = \{3, 5, 7\}$, we obtain additional inequalities involving x_j, $j \notin J_4^\star$. For example, for $\iota = 3 \in \Delta_4$, we have $i_\iota = 4$, as indicated along the edge

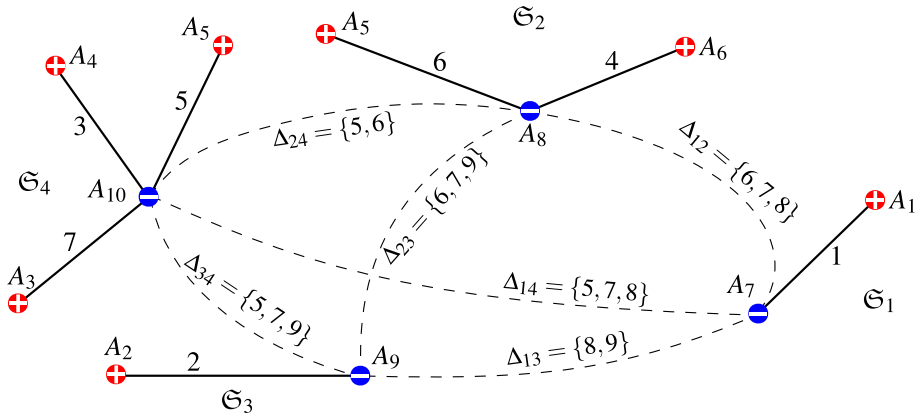


Fig. 5 Graph analysis of data in Table 1

between $A_{\iota(=3)}$ and A_{10} in Fig. 5. Using the information of J_4 in Table 1, one can verify that $J_{10} \setminus J_4 = \{3\}$. Now, using this in (8) of Corollary 1, we obtain

$$y_3 + y_5 \leq 1 - x_3$$

that is stronger than any McCormick inequality that involves x_3, y_3 and/or y_5.

Moving onto more complex results, the reader can confirm that the 4 stars in Fig. 5 satisfy the conditions of Theorem 3. So, we have $\mathscr{U} = U_1 \cup U_2 \cup U_3 \cup U_4 = \{1, 2, 3, 4, 5, 6\}$ and $\mathfrak{J} = J_1^\star \cap J_2^\star \cap J_3^\star \cap J_4^\star = \{10\}$ and use (10) of Theorem 3 to generate

$$y_1 + y_2 + y_3 + y_4 + y_5 + y_6 \leq 1 - x_{10}$$

which not only dominates its McCormick counterparts

$$y_i \leq 1 - x_{10}, \quad \forall i \in \{1, 2, 3, 4, 5, 6\}$$

but also the inequality from $\mathfrak{S}_4$ for $j = 10 \in J_4^\star$ via (7):

$$y_3 + y_4 + y_5 \leq 1 - x_{10}$$

The above serves as an example to show that an inequality from a more complex neighborhood structure (hence, a larger neighborhood) is stronger than the one(s) obtained from its sub-structure(s); with exceptions dealt with in Corollary 1, Theorems 5 and 6 (refer Fig. 4).

To obtain a valid inequality involving a different x_j, $j \notin \mathfrak{J}$, we use Theorem 5 with $\Delta_\otimes = \bigcup_{k\in\{1,\ldots,4\}} \Delta_k = \{1, 2, 3, 4, 5, 6, 7\}$ (refer Table 1 for Δ_k information) and $\Delta_\odot = \bigcup_{k,l\in\{1,\ldots,4\}, k\neq l} \Delta_{kl} = \{5, 6, 7, 8, 9\}$ (refer Fig. 5 for Δ_{kl} information along the dashed lines). For example, for $\iota = 1 \in \Delta_\otimes \setminus \Delta_\odot = \{1, 2, 3, 4\}$, we have $i_\iota = 1$, where A_1 is a leaf node of $\mathfrak{S}_1$ with the internal node A_7. Thus, we obtain $\mathscr{U} \setminus \{i_\iota\} = \{2, 3, 4, 5, 6\}$ and $J_7 \setminus J_1 = \{1\}$ from the center and leaf nodes of A_1. Using these in (14) of Theorem 5, we obtain

$$y_2 + y_3 + y_4 + y_5 + y_6 \leq 1 - x_1$$

that dominates

$$y_i \leq 1 - x_1, \quad \forall i \in \{2, 3, 4, 5, 6\}$$

and, also, the following star inequalities from $\mathfrak{S}_3$, $\mathfrak{S}_4$, and $\mathfrak{S}_2$ respectively:

$$y_2 \leq 1 - x_1, \quad y_3 + y_4 + y_5 \leq 1 - x_1, \quad y_5 + y_6 \leq 1 - x_1$$

Similarly, with $\iota \in \{2, 3, 4\}$, we obtain

$$\begin{array}{rrrrrrrrr} y_1 & & + & y_3 & + & y_4 & + & y_5 & + & y_6 & \leq 1 - x_2, \\ y_1 & + & y_2 & + & y_3 & & + & y_5 & + & y_6 & \leq 1 - x_3, \\ y_1 & + & y_2 & + & y_3 & + & y_4 & + & y_5 & & \leq 1 - x_4, \end{array}$$

that yield a tighter polyhedral relaxation than the one given by the set of their McCormick counterparts

$$\begin{array}{lllll} y_1 \leq 1 - x_2, & y_3 \leq 1 - x_2, & y_4 \leq 1 - x_2, & y_5 \leq 1 - x_2, & y_6 \leq 1 - x_2, \\ y_1 \leq 1 - x_3, & y_2 \leq 1 - x_3, & y_3 \leq 1 - x_3, & y_5 \leq 1 - x_3, & y_6 \leq 1 - x_3, \\ y_1 \leq 1 - x_4, & y_2 \leq 1 - x_4, & y_3 \leq 1 - x_4, & y_4 \leq 1 - x_4, & y_5 \leq 1 - x_4, \end{array}$$

and a set of star inequalities below:

$$\begin{array}{llllll} y_1 \leq 1 - x_2, & y_3 + y_4 + y_5 \leq 1 - x_2, & & y_5 + y_6 \leq 1 - x_2, & & \\ y_1 \leq 1 - x_3, & & y_2 \leq 1 - x_3, & & y_3 + y_5 \leq 1 - x_3, & y_5 + y_6 \leq 1 - x_3, \\ y_1 \leq 1 - x_4, & & y_2 \leq 1 - x_4, & y_3 + y_4 + y_5 \leq 1 - x_4, & & y_5 \leq 1 - x_4. \end{array}$$

Last, for Theorem 6, consider $\iota = 5 \in \Delta_{\odot} = \{5, 6, 7, 8, 9\}$ to obtain $\mathscr{U}_0^5 = \{1, 2, 5, 6\}$ and $\mathscr{U}_1^5 = \{3, 4\}$. These give rise to two inequalities

$$y_1 + y_2 + y_5 + y_6 \leq 1 - x_5 \quad \text{and} \quad y_3 + y_4 \leq 1 - x_{15}$$

that are stronger than 6 McCormick inequalities involving x_5 and x_{15} and, also 3 star inequalities.

Similarly, Theorem 6 yields 4 additional pairs (a total of 8) of valid inequalities below for the illustrative dataset under analysis (Note that 3 are McCormick inequalities and see Fig. 4 for a quick reference for when this happens.):

$$\begin{array}{rrrrrrrrrr} y_1 & + & y_2 & + & y_3 & + & y_4 & + & y_5 & & \leq 1 - x_6 \\ & & & & & & & & y_6 & \leq 1 - x_{16} \\ y_1 & + & y_2 & + & y_3 & & & & & \leq 1 - x_7 \\ & & & & & & y_4 & + & y_5 & + \; y_6 & \leq 1 - x_{17} \\ y_1 & & & & & & & & & \leq 1 - x_8 \\ & & y_2 & + & y_3 & + & y_4 & + & y_5 & + \; y_6 & \leq 1 - x_{18} \\ y_1 & & & + & y_3 & + & y_4 & + & y_5 & + \; y_6 & \leq 1 - x_9 \\ & & y_2 & & & & & & & & \leq 1 - x_{19} \end{array}$$

As a summary, we listed all 60 standard, McCormick inequalities, 45 inequalities from individual stars via Lemma 1, Theorem 1 and Corollary 1 in Sect. 2.1, and 22 inequalities from a set of stars via Lemma 1, Theorems 3, 5 and 6 in Sect. 2.2 in the three columns of Table 2, going from left to right, respectively. For easy comparison, we use '$\longrightarrow$' in this table for dominance relation to state that the inequality in its tail side is dominated by the one on its head. In brief, this table pictorially depicts benefits of the results of this paper in regard to:

- algebraic lifting and strengthening of simple, McCormick envelopes for a 0–1 multilinear objective function φ^+ of (PG);
- a multi-term relaxation of φ^+ (for the example considered, the simultaneous relaxation of all six terms of φ^+) and strengthening; and
- a tighter polyhedral relaxation of (PG) in terms of a much smaller number of stronger 0–1 linear inequalities.

Table 2 Summary of valid inequalities for (PG) instance on dataset in Table 1 and illustration of benefits of new results

McCormick Inequalities	Inequalities from Individual Stars	Inequalities from Star Set
$y_2 \le 1 - x_1$	$y_2 \le 1 - x_1$	$y_2 + y_3 + y_4 + y_5 + y_6 \le 1 - x_1$
$y_3 \le 1 - x_1$	$y_3 + y_4 + y_5 \le 1 - x_1$	
$y_4 \le 1 - x_1$		
$y_5 \le 1 - x_1$	$y_5 + y_6 \le 1 - x_1$	
$y_6 \le 1 - x_1$		
$y_1 \le 1 - x_{11}$	$y_1 \le x_1$	$y_1 \le x_1$
$y_1 \le 1 - x_2$	$y_1 \le 1 - x_2$	$y_1 + y_3 + y_4 + y_5 + y_6 \le 1 - x_2$
$y_3 \le 1 - x_2$	$y_3 + y_4 + y_5 \le 1 - x_2$	
$y_4 \le 1 - x_2$		
$y_5 \le 1 - x_2$	$y_5 + y_6 \le 1 - x_2$	
$y_6 \le 1 - x_2$		
$y_2 \le 1 - x_{12}$	$y_2 \le x_2$	$y_2 \le x_2$
$y_1 \le 1 - x_3$	$y_1 \le 1 - x_3$	$y_1 + y_2 + y_3 + y_5 + y_6 \le 1 - x_3$
$y_2 \le 1 - x_3$	$y_2 \le 1 - x_3$	
$y_3 \le 1 - x_3$	$y_3 + y_5 \le 1 - x_3$	
$y_5 \le 1 - x_3$	$y_5 + y_6 \le 1 - x_3$	
$y_6 \le 1 - x_3$		
$y_4 \le 1 - x_{13}$	$y_4 \le x_3$	$y_4 \le x_3$
$y_1 \le 1 - x_4$	$y_1 \le 1 - x_4$	$y_1 + y_2 + y_3 + y_4 + y_5 \le 1 - x_4$
$y_2 \le 1 - x_4$	$y_2 \le 1 - x_4$	
$y_3 \le 1 - x_4$	$y_3 + y_4 + y_5 \le 1 - x_4$	
$y_4 \le 1 - x_4$		
$y_5 \le 1 - x_4$	$y_5 \le 1 - x_4$	
$y_6 \le 1 - x_{14}$	$y_6 \le x_4$	$y_6 \le x_4$
$y_1 \le 1 - x_5$	$y_1 \le 1 - x_5$	$y_1 + y_2 + y_5 + y_6 \le 1 - x_5$
$y_2 \le 1 - x_5$	$y_2 \le 1 - x_5$	
$y_5 \le 1 - x_5$	$y_5 \le x_{15}$	$y_5 \le x_{15}$
$y_6 \le 1 - x_5$	$y_5 + y_6 \le 1 - x_5$	
$y_3 \le 1 - x_{15}$		
$y_4 \le 1 - x_{15}$	$y_3 + y_4 \le 1 - x_{15}$	$y_3 + y_4 \le 1 - x_{15}$
$y_1 \le 1 - x_6$	$y_1 \le 1 - x_6$	$y_1 + y_2 + y_3 + y_4 + y_5 \le 1 - x_6$
$y_2 \le 1 - x_6$	$y_2 \le 1 - x_6$	
$y_3 \le 1 - x_6$	$y_3 + y_4 + y_5 \le 1 - x_6$	
$y_4 \le 1 - x_6$		
$y_5 \le 1 - x_6$	$y_5 \le x_{16}$	$y_5 \le x_{16}$
$y_6 \le 1 - x_{16}$	$y_6 \le 1 - x_{16}$	$y_6 \le 1 - x_{16}$
$y_1 \le 1 - x_7$	$y_1 \le 1 - x_7$	$y_1 + y_2 + y_3 \le 1 - x_7$
$y_2 \le 1 - x_7$	$y_2 \le 1 - x_7$	
$y_3 \le 1 - x_7$	$y_3 \le x_{17}$	$y_3 \le x_{17}$
$y_4 \le 1 - x_{17}$	$y_4 + y_5 \le 1 - x_{17}$	$y_4 + y_5 + y_6 \le 1 - x_{17}$
$y_5 \le 1 - x_{17}$	$y_5 + y_6 \le 1 - x_{17}$	
$y_6 \le 1 - x_{17}$		
$y_1 \le 1 - x_8$	$y_1 \le 1 - x_8$	$y_1 \le 1 - x_8$
$y_2 \le 1 - x_{18}$	$y_2 \le 1 - x_{18}$	$y_2 + y_3 + y_4 + y_5 + y_6 \le 1 - x_{18}$
$y_3 \le 1 - x_{18}$	$y_3 + y_4 + y_5 \le 1 - x_{18}$	
$y_4 \le 1 - x_{18}$		
$y_5 \le 1 - x_{18}$	$y_5 + y_6 \le 1 - x_{18}$	
$y_6 \le 1 - x_{18}$		
$y_1 \le 1 - x_9$	$y_1 \le 1 - x_9$	$y_1 + y_3 + y_4 + y_5 + y_6 \le 1 - x_9$
$y_3 \le 1 - x_9$	$y_3 + y_4 + y_5 \le 1 - x_9$	
$y_4 \le 1 - x_9$		
$y_5 \le 1 - x_9$	$y_5 + y_6 \le 1 - x_9$	
$y_6 \le 1 - x_9$		
$y_2 \le 1 - x_{19}$	$y_2 \le 1 - x_{19}$	$y_2 \le 1 - x_{19}$
$y_1 \le 1 - x_{10}$	$y_1 \le 1 - x_{10}$	$y_1 + y_2 + y_3 + y_4 + y_5 + y_6 \le 1 - x_{10}$
$y_2 \le 1 - x_{10}$	$y_2 \le 1 - x_{10}$	
$y_3 \le 1 - x_{10}$	$y_3 + y_4 + y_5 \le 1 - x_{10}$	
$y_4 \le 1 - x_{10}$		
$y_5 \le 1 - x_{10}$	$y_5 + y_6 \le 1 - x_{10}$	
$y_6 \le 1 - x_{10}$		

'$\longrightarrow$' indicates that the inequality on its tail is dominated by the one on its head. Star set inequalities are obtained as result of multi-term relaxing 5–6 terms of φ^+ (which has 6 terms) and strengthening

Table 3 CPLEX efforts for solving 0–1 equivalents of (PG) for + pattern generation for data in Table 1

0–1 model	Root relaxation gap (%)	CPU ticks	BB nodes
$(\underline{\text{PG}})_{\texttt{mccormick}}$	81.5	1.85	3
$(\underline{\text{PG}})_{\texttt{star}}$	50.0	0.39	1
$(\underline{\text{PG}})_{\texttt{stars}}$	0[a]	0.05	0[b]

[a] $(\underline{\text{PG}})_{\texttt{stars}}$ is integral; recall that no additional cuts are added
[b] Solved at the root node

Furthermore, the reader confirms the hierarchy of relative strength of the inequalities in Fig. 4 in this table that the more advanced an inequality, the stronger it is; thus, a tightest polyhedral overestimation of φ^+ of (PG) is obtained when a maximum number of maximal stars are exploited for the purpose.

Last, recall (PG) to note that, for + pattern generation experiments, we need to convexify φ^- and concavify φ^+. For a numerical demonstration of the utility of the new polyhedral overestimation scheme for φ^+, we thus obtain 0–1 polyhedral relaxations of the (PG) instance generated on the dataset in Table 1 as

$$(\underline{\text{PG}})_\dagger: \quad \max \left\{ \sum_{i=1}^{6} y_i \;\middle|\; \bar{\varphi}_\dagger^+,\ \sum_{j\in J_i} x_j \geq 1 \text{ for } i = 7,\ldots,10,\ \mathbf{x}\in\{0,1\}^{20},\ \mathbf{y}\in[0,1]^6 \right\}$$

where $\bar{\varphi}_\dagger^+$ is the polyhedral concave overestimation of φ^+ obtained by means of the method specified in $\dagger \in \{\texttt{mccormick}, \texttt{star}, \texttt{stars}\}$, where `mccormick` stands for the standard McCormick relaxation method while `star` and `stars` indicate the improvements made on the McCormick-based model by means of the inequalities from individual stars in Sect. 2.1 and via those from the neighboring stars in Sect. 2.2, respectively. Next, $(\underline{\text{PG}})_{\texttt{mccormick}}$, $(\underline{\text{PG}})_{\texttt{star}}$ and $(\underline{\text{PG}})_{\texttt{stars}}$ were solved by CPLEX 12.8 [19] without utilizing the cutting plane methods of the solver, and information on some important performance indicators is summarized in Table 3. Here, 'Root Relaxation Gap' is the % difference between the root LP relaxation value of the corresponding model and its optimum; 'CPU Ticks' is time in ticks and 'BB Nodes' is the total number of branch-and-bound nodes required by CPLEX for solving these MIP instances.

In brief, the results in Table 3 show the practical side of the mathematically stronger results in this paper well. Particularly, we note that $(\underline{\text{PG}})_{\texttt{stars}}$ is integral, thus solved at the root node.

4 Numerical experiments

For general $\bullet$ pattern generation for $\bullet \in \{+,-\}$, we can use $\bar{\bullet}$ to denote the complementary type of $\bullet$ with respect to $\{+,-\}$ to obtain a 0–1 linear equivalent of $(\text{PG})^\bullet$ as

$$(\underline{\text{PG}})_\dagger^\bullet: \quad \max \left\{ \sum_{i\in S^\bullet} y_i \;\middle|\; \bar{\varphi}_\dagger^\bullet,\ \sum_{j\in J_i} x_j \geq 1,\ \forall i \in S^{\bar{\bullet}},\ \mathbf{x}\in\{0,1\}^{2n},\ \mathbf{y}\in[0,1]^{m^\bullet} \right\}$$

where $\bar{\varphi}_\dagger^\bullet$ is the polyhedral concave overestimation of $\varphi^\bullet$ obtained by any method $\dagger \in \{\texttt{mccormick}, \texttt{stars}\}$ and $m^\bullet$ is the number of $\bullet$ data in the pattern generation task. Recall that $(\underline{\text{PG}})_{\texttt{stars}}^\bullet$ is a strengthened form of $(\underline{\text{PG}})_{\texttt{mccormick}}^\bullet$ by means of the results in Sect. 2.

Table 4 Binary classification datasets from [24]

Dataset (abbreviation)	Number of data		Number of features	
	m^+	m^-	Original	0–1 binarized[b]
BUPA liver disorder[a] (bupa)	145	200	6	269
Cleveland heart disease[a] (clev)	137	160	13	305
Credit card scoring[a] (cred)	357	296	15	773
Pima Indian diabetes[a] (diab)	268	500	8	857
Boston housing[a] (hous)	257	249	13	1209
Wisconsin breast cancer[a] (wisc)	239	444	9	72
Blood transfusion service center (btsc)	125	423	4	144
King Rook vs. King Pawn (krkp)	1668	1527	73	73
Phishing websites (phis)	5945	4753	68	68
QSAR biodegradation (qsar)	356	699	41	4178
Seismic bumps (seis)	170	2414	15	1120
Wilt (wilt)	261	4578	5	2311

[a] 6 well-studied data mining datasets
[b] After 0–1 binarization and before feature selection (e.g., see [10,11])

It is both tighter and smaller than the predecessor, thus is expected to provide better LP relaxation bounds and aid in more efficient LAD pattern generation; that is, a more effective and efficient solution of 0–1 MP in (PG).

To test this, this section runs pattern generation experiments with 12 public data mining datasets from [24], summarized in Table 4. More specifically, we took each dataset in this table and first randomly split it into 3 equal-sized, mutually disjoint partitions of 2/3 of the + and 2/3 of the − data. Next, we combined 2 of them to obtain 3 sub-datasets, each comprised of exactly 2/3 of the original dataset and differ from the other in 50% of the data for generating a pair of + and − patterns via $\underline{(PG)}^{\bullet}_{\texttt{mccormick}}$ and $\underline{(PG)}^{\bullet}_{\texttt{stars}}$. To assess the efficacy of the new relaxation method more objectively, we use CPLEX without utilizing CPLEX cuts when solving these two 0–1 MIP's.

When it comes to solving MIP's, cutting plane methods are a hallmark of efficacy. In note of this, each time $\underline{(PG)}^{\bullet}_{\texttt{mccormick}}$ was solved, we solved the same instance once more with utilizing CPLEX cuts this second time. This 0–1 pattern generation MIP is referred to as $\underline{(PG)}^{\bullet}_{\texttt{mccormick+CPLEX}}$ below. As seen, $\underline{(PG)}^{\bullet}_{\texttt{mccormick+CPLEX}}$ is a strengthened version of $\underline{(PG)}^{\bullet}_{\texttt{mccormick}}$ via cutting plane methods implemented in CPLEX.

Difficulties of mining useful information from real-life data are often caused by the presence of a few 'hard-to-classify' data. Thus, for a fair comparison, we repeated the whole 3-fold validation procedure 10 times to test the three 0–1 MIP models in 30 unique pattern generation tasks with each dataset in Table 4.

The relaxation bound of a hard problem at the root node of the branch-and-bound tree is usually poor, generally speaking. Nevertheless, when an advanced MIP solver like CPLEX is utilized, the root node relaxation values are least influenced by the effect of the arsenal of solution rules and tools as well as branch-and-bound options featured in the solver, hence present itself a valid measure by which the relative strength of relaxation models compared can be assessed. Thus, we compare the average root node relaxation gaps by the three MIP pattern generation models to examine their relative tightness in formulation. Table 5 below summarizes the root node relaxation gaps by the three MIP models and provides the infor-

mation on the %-age of improvements made on $(\underline{\text{PG}})^{\bullet}_{\texttt{mccormick}}$, the basis for all comparison, by means of using CPLEX cuts and our new cuts under their sub-columns labelled by '†' and '‡', respectively. All numbers in this table and those below are in format 'average ± 1 standard deviation' of 30 results, followed by the minimum and the maximum values inside parenthesis.

To assess how the new inequalities enhance the overall efficiency of solution, we compare CPU time (in ticks) required for solving three 0–1 MIP models in Table 6 and summarize the number of CPLEX branch-and-bound nodes explored in Table 7. Between CPU time and the branch-and-bound nodes, we shall note that the former is a more credible measure of efficacy of a method.

Last, Table 8 summarizes information on the number ($n_{\dagger}$) and degree (d) of stars found and the number ($n_{\ddagger}$) of neighboring stars found, along with their size (q) and the number of terms/data ($|\mathscr{U}|$ included in Q's), for each of the 12 datasets. These numbers can be used to infer practical utilities of the mathematical results of this paper, particularly in multi-term relaxing $\varphi^{\bullet}$. For instance, with information on m^+ and $\bar{n}_s$, one notes from last number inside parenthesis for the `seis` dataset that Theorem 3 simultaneously relaxes up to 20 terms of φ^+, which consists of 113(m^+) terms, each a multiplication of 35($\bar{n}_s$) $(1 - x_j)$'s, and so forth.

For comparison, we recall once more that $(\underline{\text{PG}})^{\bullet}_{\texttt{mccormick+CPLEX}}$ and $(\underline{\text{PG}})^{\bullet}_{\texttt{stars}}$ are strengthened forms of $(\underline{\text{PG}})^{\bullet}_{\texttt{mccormick}}$ by means of CPLEX cutting planes and the valid inequalities of this paper, respectively. Thus, one has the following points readily available for comparison and interpretation of numerical results in Tables 5–7:

- The comparative performance between $(\underline{\text{PG}})^{\bullet}_{\texttt{mccormick}}$ and $(\underline{\text{PG}})^{\bullet}_{\texttt{mccormick+CPLEX}}$ is an indicator of the efficacy of cutting plane methods in solving 0–1 pattern generation instances from the 12 datasets;
- Furthermore, the above serves as a measure of numerical difficulties associated with LAD pattern generation and solving (PG);
- The comparative results by $(\underline{\text{PG}})^{\bullet}_{\texttt{mccormick}}$ and $(\underline{\text{PG}})^{\bullet}_{\texttt{stars}}$ illustrate benefits of the new multi-term, polyhedral relaxation scheme of this paper for LAD pattern generation, thus for a class of practical 0–1 MP in (PG); and
- Last, the comparative results by $(\underline{\text{PG}})^{\bullet}_{\texttt{mccormick+CPLEX}}$ and $(\underline{\text{PG}})^{\bullet}_{\texttt{stars}}$ measure a relative superiority of one method over the other in solving the standard 0–1 MIP equivalent of (PG).

In reference to the above, numerical results are self-explanatory, thus we comment only on a few, more important and less obvious issues below.

To begin with, as noted earlier, data mining difficulties are often associated with the presence of (a few) 'difficult' data. As real-life data, the 12 datasets used for comparative experiments in this section contain hard-to-classify data. This is evidenced in Table 6 that time required for solving $(\underline{\text{PG}})^{\bullet}_{\texttt{stars}}$ instances follows a right-skewed distribution for each dataset; namely, the right tail (that is, the difference between the max and average values) is about 2-5 times longer than the left tail regardless of the size of the dataset, despite all 30 $(\underline{\text{PG}})^{\bullet}_{\texttt{stars}}$ instances for a dataset feature the same number of constraints and 0–1 variables. Largely, this explains why n-fold cross validation experiments have been adopted as a standard setting for comparative experiments in the data mining literature as well as in this section.

Now, recall that $(\underline{\text{PG}})^{\bullet}_{\texttt{stars}}$ is built on $(\underline{\text{PG}})^{\bullet}_{\texttt{mccormick}}$ by means of replacing term-wise McCormick envelopes in the latter by a smaller number of stronger inequalities. As practical phenomena can defy mathematical triumphs occasionally, one may not expect $(\underline{\text{PG}})^{\bullet}_{\texttt{stars}}$ to outperform $(\underline{\text{PG}})^{\bullet}_{\texttt{mccormick}}$ on every pattern generation dataset of the 30 fold experiments for

Table 5 Root node relaxation gap

Dataset	•	$(\underline{PG})^{\bullet}_{\texttt{mccormick}}$	$(\underline{PG})^{\bullet}_{\texttt{mccormick+CPLEX}}$	†	$(\underline{PG})^{\bullet}_{\texttt{stars}}$	‡
bupa	+	82.8 ± 2.4(77.1, 86.3)	82.1 ± 2.6(75.6, 85.6)	0.9	65.4 ± 5.5(54.6, 74.6)	21.1
	−	83.7 ± 1.7(79.1, 86.4)	83.3 ± 1.7(78.7, 86.1)	0.5	72.3 ± 3.5(65.0, 78.4)	13.6
clev	+	66.9 ± 4.2(59.1, 74.0)	64.3 ± 4.6(55.9, 72.4)	3.6	40.7 ± 4.3(32.3, 46.0)	39.7
	−	70.6 ± 3.7(64.3, 77.8)	68.4 ± 3.6(62.1, 76.5)	3.2	46.6 ± 4.6(35.7, 54.7)	34.3
cred	+	68.5 ± 6.0(55.9, 79.2)	67.9 ± 6.1(55.5, 78.6)	1.0	57.7 ± 8.0(40.7, 71.5)	16.2
	−	76.8 ± 3.2(70.3, 83.1)	76.2 ± 3.3(69.8, 82.8)	0.8	67.6 ± 3.8(60.6, 75.0)	11.7
diab	+	89.1 ± 1.4(85.9, 91.2)	88.8 ± 1.5(85.7, 91.1)	0.2	82.7 ± 2.6(77.2, 86.1)	7.2
	−	81.1 ± 1.6(77.6, 84.0)	80.9 ± 1.6(77.3, 83.8)	0.3	74.7 ± 2.0(70.7, 78.9)	8.0
hous	+	64.6 ± 5.1(55.4, 72.4)	63.2 ± 5.3(53.4, 71.4)	2.2	41.9 ± 6.0(30.2, 50.1)	35.5
	−	74.1 ± 3.3(67.0, 78.6)	73.1 ± 3.3(66.2, 78.2)	1.3	55.0 ± 6.4(39.0, 65.2)	25.9
wisc	+	59.1 ± 4.4(49.5, 65.4)	56.2 ± 5.3(46.1, 64.0)	4.9	42.1 ± 7.9(23.0, 54.7)	29.0
	−	55.4 ± 4.8(46.5, 63.0)	47.4 ± 6.7(34.2, 58.1)	14.8	20.1 ± 8.9(0.0, 34.2)	64.2
btsc	+	89.4 ± 1.7(85.9, 91.6)	89.0 ± 1.7(85.7, 91.2)	0.4	72.3 ± 3.8(62.5, 79.3)	19.1
	−	83.0 ± 1.9(77.8, 85.9)	82.8 ± 1.9(77.5, 85.8)	0.2	75.9 ± 3.0(68.3, 79.8)	8.7
krkp	+	51.5 ± 1.4(49.1, 53.5)	51.4 ± 1.4(49.0, 53.4)	0.1	34.1 ± 1.9(29.4, 36.7)	33.8
	−	49.1 ± 0.7(47.7, 50.4)	49.0 ± 0.7(47.7, 50.3)	0.2	32.1 ± 1.2(29.8, 34.4)	34.4
phis	+	72.4 ± 0.4(71.8, 73.2)	72.4 ± 0.4(71.8, 73.2)	0.0	70.6 ± 0.4(69.8, 71.3)	2.5
	−	36.9 ± 0.7(35.6, 38.0)	36.9 ± 0.7(35.5, 38.0)	0.1	33.0 ± 0.7(31.7, 34.1)	10.8
qsar	+	76.6 ± 2.0(71.5, 80.5)	76.2 ± 2.0(71.2, 80.2)	0.4	66.5 ± 2.9(60.9, 71.1)	13.4
	−	79.3 ± 2.3(75.2, 83.6)	79.1 ± 2.4(75.0, 83.5)	0.2	74.7 ± 2.9(68.9, 80.4)	5.8
seis	+	93.7 ± 0.6(92.5, 94.5)	93.7 ± 0.6(92.5, 94.5)	0.1	85.3 ± 1.6(81.9, 88.0)	9.0
	−	86.7 ± 1.4(83.7, 88.7)	86.6 ± 1.4(83.7, 88.7)	0.0	85.1 ± 1.6(81.8, 87.6)	1.9
wilt	+	73.6 ± 3.4(65.6, 78.8)	73.3 ± 3.4(65.1, 78.5)	0.4	53.8 ± 6.3(42.4, 64.1)	26.9
	−	67.0 ± 3.0(61.9, 72.0)	66.9 ± 3.0(61.8, 72.0)	0.1	62.9 ± 3.5(56.5, 69.6)	6.1

$(\underline{PG})^{\bullet}_{\texttt{mccormick}}$ and $(\underline{PG})^{\bullet}_{\texttt{stars}}$ are solved by CPLEX without utilizing CPLEX cuts
$(\underline{PG})^{\bullet}_{\texttt{mccormick+CPLEX}}$ is $(\underline{PG})^{\bullet}_{\texttt{mccormick}}$ solved by CPLEX with CPLEX cuts utilized
All results are in format 'average ± 1 standard deviation (min, max)'
†: Average of $\frac{\text{Gap by } (\underline{PG})^{\bullet}_{\texttt{mccormick}} - \text{Gap by } (\underline{PG})^{\bullet}_{\texttt{mccormick+CPLEX}}}{\text{Worse of the 2 results}}$ values: (measures relative efficacy of $(\underline{PG})^{\bullet}_{\texttt{mccormick+CPLEX}}$ over $(\underline{PG})^{\bullet}_{\texttt{mccormick}}$)
‡: Average of $\frac{\text{Gap by } (\underline{PG})^{\bullet}_{\texttt{mccormick}} - \text{Gap by } (\underline{PG})^{\bullet}_{\texttt{stars}}}{\text{Worse of the 2 results}}$ values: (measures relative efficacy of $(\underline{PG})^{\bullet}_{\texttt{stars}}$ over $(\underline{PG})^{\bullet}_{\texttt{mccormick}}$)

each of the test datasets, moderately speaking. Generally speaking, however, mathematical superiority pervades in practice and translates to benefits. Therefore, in light of the mathematical truth that $(\underline{PG})^{\bullet}_{\texttt{stars}}$ is not only smaller in size but also tighter than its predecessor, the former is much easier to solve, regardless of the size of the dataset. Tables 6–7 support this well and clearly demonstrate benefits of the multi-term relaxation method of this paper.

General cutting plane methods have stood the test of time to be effective in MIP solution. However, one sees from comparative results by $(\underline{PG})^{\bullet}_{\texttt{mccormick}}$ and $(\underline{PG})^{\bullet}_{\texttt{mccormick+CPLEX}}$ in the Tables 5–7 that they do not prove to be effective at all for the class of practically important problems dealt with in this paper. Specifically, the use of CPLEX cuts improves the root node bounds for $(\underline{PG})^{\bullet}_{\texttt{mccormick}}$ only by a very small fraction (refer Table 5) but

requires substantially more CPU time (refer Table 6) for solving the 0–1 MIP model. Plausibly, the increased CPU time requirements may arise from adding cuts giving rise to larger LP relaxations to be solved. It may also due to additional constraints getting in the way of the simple structure of $(\underline{\text{PG}})^{\bullet}_{\texttt{mccormick}}$—namely, the minimal cover inequalities and simple upper bounding-type constraints via the McCormick relaxation—from being 'fully' exploited for efficient solution. Simply, one can see from the comparative results by $(\underline{\text{PG}})^{\bullet}_{\texttt{mccormick+CPLEX}}$ that $(\underline{\text{PG}})^{\bullet}_{\texttt{mccormick}}$ is a challenging global optimization problem.

For the same optimization problem, our new cuts reduce the root node gap by a substantial amount and prove to be significantly more effective than CPLEX cuts on average; to see, compare numbers in columns '†' and '‡' of Table 5. They contribute to faster solution of the 0–1 pattern generation instances, too (refer Table 6). As an acute reader notes, this is not a surprise but a direct result of mathematical discoveries of this paper that yield $(\underline{\text{PG}})^{\bullet}_{\texttt{stars}}$ from $(\underline{\text{PG}})^{\bullet}_{\texttt{mccormick}}$ by substituting a smaller number of stronger inequalities for standard McCormick inequalities in the latter, basis model. Furthermore, we (refer the reader back to Table 2 to) note that fixing the value of one y_i in our new inequality to 1 can have a cascading effect on fixing the values of the remaining y_j's in the same inequality to 0 (although we doubt this useful property is fully exploited by CPLEX's default setting). For example, the information for `seis` in Table 8 reveals that our valid inequalities via Theorem 3 involve up to 20 y_i's, obtained from simultaneously relaxing 20 terms of φ^+, that further yield additional inequalities with up to 19 y_i's and a different x_k via Theorems 5 and 6.

5 Conclusion

This paper dealt with a multi-term, polyhedral relaxation of φ^+ of (PG) in terms of a smaller number of stronger linear inequalities. Toward this goal, we analyzed a set of $+$ data and a set of $-$ data on a graph and discovered sufficient conditions and tools for obtaining a tight polyhedral overestimator of φ^+ from individual stars and also from sets of stars. We showed that our inequalities are facet-defining of the 0–1 multilinear polytope associated with the McCormick inequalities that they replace. We further showed that the maximum benefit, in regard to the size as well as the tightness of the relaxation model, is realized when a maximal set of maximal stars are exploited for generating new valid inequalities. With experiments on 12 public data mining datasets, we demonstrated practical utilities of the new results in absolute terms and also in comparison with the cutting plane methods for MIP implemented in CPLEX.

Last, a multi-term, polyhedral convexification/relaxation of highly nonlinear, complex multilinear programs is well-known to be difficult. Numerical studies in Sect. 4 demonstrate benefits of such results with the practical pattern generation MP for LAD. Here, we note that the material of this paper can also be used for other optimization problems/functions, including:

- more general 0–1 multilinear functions of the form $\phi^+(\mathbf{x}) := \sum_{i \in m} c_i \prod_{j=1}^{n_i} x_j$, where $c_i \in I\!R_+$; and
- linear complementary relations/constraints—that is, functions in 0–1 vectors $\mathbf{x}$ and $\mathbf{y}$ with $\mathbf{x} \cdot \mathbf{y} = 0$ (e.g., [25])

Therefore, one may see the numerical results of this paper as calling more attention from the optimization community on the aforementioned, mathematically classic and practically important research subject in nonlinear optimization.

Table 6 CPU time (in ticks) required for solution

Dataset	•	$(\underline{PG})^{\bullet}_{\texttt{mccormick}}$	$(\underline{PG})^{\bullet}_{\texttt{mccormick+CPLEX}}$	†	$(\underline{PG})^{\bullet}_{\texttt{stars}}$	‡
bupa	+	618.7 ± 156.3(379.1, 979.5)	1310.2 ± 282.2(941.4, 1888.6)	− 52.9	419.1 ± 81.8(290.1, 573.4)	31.2
	−	1070.9 ± 237.7(662.6, 1505.8)	1899.9 ± 460.2(1215.2, 2649.7)	− 42.8	757.9 ± 141.3(548.1, 1049.1)	27.6
clev	+	100.3 ± 15.3(81.1, 130.5)	291.4 ± 60.0(196.5, 412.9)	− 63.2	68.5 ± 9.0(55.7, 89.5)	31.1
	−	134.4 ± 24.0(99.2, 177.8)	381.2 ± 85.5(245.0, 548.1)	− 63.5	83.2 ± 13.2(63.3, 104.2)	37.4
cred	+	1593.6 ± 500.7(849.4, 2752.9)	2666.1 ± 733.3(1666.3, 4166.9)	− 40.0	917.0 ± 335.5(450.1, 1513.2)	39.7
	−	1272.5 ± 273.0(881.5, 1948.3)	2185.2 ± 610.3(1381.3, 3430.5)	− 40.2	836.1 ± 224.8(572.7, 1337.6)	32.8
diab	+	5454.7 ± 1628.7(2719.4, 10540.3)	6971.7 ± 1573.6(4509.8, 9762.8)	− 21.1	3462.0 ± 788.8(2503.7, 5905.5)	33.7
	−	7241.0 ± 2363.9(4358.8, 14121.7)	12469.0 ± 4128.7(6739.6, 20659.8)	− 40.0	4704.4 ± 1492.6(2631.0, 8712.5)	33.6
hous	+	255.6 ± 57.3(193.8, 390.6)	625.6 ± 144.2(397.9, 902.2)	− 56.4	167.3 ± 35.3(118.2, 273.9)	32.9
	−	342.6 ± 83.7(201.3, 527.1)	768.4 ± 178.3(519.1, 1095.7)	− 55.4	260.9 ± 83.2(143.6, 453.0)	25.0
wisc	+	74.3 ± 25.9(38.4, 128.0)	237.4 ± 85.5(93.8, 381.7)	− 65.1	58.1 ± 19.3(28.0, 92.0)	21.7
	−	27.1 ± 8.8(14.3, 44.1)	152.8 ± 51.9(70.6, 246.6)	− 80.4	20.0 ± 7.3(0.8, 34.3)	28.2
btsc	+	8662.4 ± 2622.9(3658.0, 15856.2)	15616.3 ± 4748.2(8085.9, 27086.8)	− 41.4	4579.9 ± 1565.4(2507.2, 10097.4)	43.7
	−	33629.6 ± 13732.5(16061.1, 66613.8)	38420.4 ± 14333.2(17175.0, 76132.6)	− 11.1	15165.1 ± 8731.5(6853.0, 35967.8)	53.9
krkp	+	63555.8 ± 17289.8(36621.7, 94731.4)	98159.1 ± 36325.6(57261.9, 184081.6)	− 29.0	42621.6 ± 10541.1(24906.3, 57934.8)	29.5
	−	35619.8 ± 7735.6(22788.0, 53974.6)	51807.2 ± 15428.7(33105.5, 92462.8)	− 28.9	19852.8 ± 6351.2(13097.3, 34824.1)	41.3
phis	+	469689.3 ± 120151.1(263624.4, 748632.2)	923264.7 ± 372797.9(408765.2, 1779841.1)	− 42.1	391650.6 ± 132420.9(219952.7, 647890.0)	14.6
	−	203738.9 ± 75925.8(119254.8, 378760.1)	229619.2 ± 48001.0(165423.6, 339413.2)	− 11.9	143966.7 ± 37728.2(90746.8, 216801.0)	22.5
qsar	+	5537.9 ± 1187.4(3670.9, 8747.5)	8919.7 ± 2521.1(5433.0, 14318.0)	− 35.3	3817.5 ± 1176.1(2483.7, 6583.9)	29.3
	−	16107.6 ± 6089.0(7598.9, 30158.4)	23463.7 ± 6949.0(12601.4, 37517.6)	− 29.4	12027.1 ± 2641.8(6954.1, 16582.5)	20.3

Table 6 continued

Dataset	•	$(\underline{PG})^{\bullet}_{\texttt{mccormick}}$	$(\underline{PG})^{\bullet}_{\texttt{mccormick+CPLEX}}$	†	$(\underline{PG})^{\bullet}_{\texttt{stars}}$	‡
seis	+	93326.2 ± 32215.1(48082.1, 156749.0)	119801.6 ± 35113.2(63782.1, 205955.1)	− 18.5	32268.6 ± 11834.7(20493.5, 64836.5)	63.2
	−	909146.5 ± 329664.2(471838.3, 1497986.9)	1031720.4 ± 328757.8(585722.7, 1718629.5)	− 11.5	591061.7 ± 251797.7(284486.3, 1107060.7)	32.7
wilt	+	4685.5 ± 1387.7(2685.8, 8559.5)	7101.2 ± 2257.2(3778.0, 11227.2)	− 30.4	2910.0 ± 875.2(1588.9, 4864.7)	36.8
	−	59031.9 ± 21781.6(26109.2, 107269.8)	86202.8 ± 25613.5(57374.3, 135757.1)	− 32.0	46752.8 ± 20915.5(20263.4, 116030.4)	17.0

$(\underline{PG})^{\bullet}_{\texttt{mccormick}}$ and $(\underline{PG})^{\bullet}_{\texttt{stars}}$ are solved without utilizing CPLEX cuts

$(\underline{PG})^{\bullet}_{\texttt{mccormick+CPLEX}}$ is $(\underline{PG})^{\bullet}_{\texttt{mccormick}}$ solved with CPLEX cuts utilized

All times are in format 'average ± 1 standard deviation (min, max)'

†: Average of $\frac{\text{Time for } (\underline{PG})^{\bullet}_{\texttt{mccormick}} - \text{Time for } (\underline{PG})^{\bullet}_{\texttt{mccormick+CPLEX}}}{\text{Worse of the 2 results}}$ values: measures relative efficacy of $(\underline{PG})^{\bullet}_{\texttt{mccormick+CPLEX}}$ over $(\underline{PG})^{\bullet}_{\texttt{mccormick}}$

‡: Average of $\frac{\text{Time for } (\underline{PG})^{\bullet}_{\texttt{mccormick}} - \text{Time for } (\underline{PG})^{\bullet}_{\texttt{stars}}}{\text{Worse of the 2 results}}$ values: measures relative efficacy of $(\underline{PG})^{\bullet}_{\texttt{stars}}$ over $(\underline{PG})^{\bullet}_{\texttt{mccormick}}$

Table 7 CPLEX branch and bound nodes required for solution

Dataset	•	$(\underline{PG})^{\bullet}_{\texttt{mccormick}}$	$(\underline{PG})^{\bullet}_{\texttt{mccormick+CPLEX}}$	†	$(\underline{PG})^{\bullet}_{\texttt{stars}}$	‡
bupa	+	494 ± 192(274, 1001)	712 ± 268(322, 1280)	− 28.5	426 ± 124(206, 656)	10.2
	−	686 ± 210(305, 1115)	901 ± 306(460, 1524)	− 19.2	545 ± 125(360, 835)	15.6
clev	+	71 ± 21(32, 108)	114 ± 41(70, 209)	− 36.5	40 ± 17(20, 81)	41.5
	−	116 ± 29(61, 160)	149 ± 38(98, 231)	− 19.0	57 ± 26(26, 111)	47.4
cred	+	412 ± 188(99, 910)	471 ± 201(155, 987)	− 13.4	196 ± 110(49, 402)	49.2
	−	450 ± 135(262, 733)	524 ± 204(268, 1017)	− 8.0	264 ± 85(131, 447)	38.9
diab	+	2896 ± 1099(1285, 6580)	3005 ± 986(1476, 5197)	− 2.8	1653 ± 469(1057, 2888)	37.6
	−	1652 ± 664(1003, 3403)	2768 ± 1254(1087, 5330)	− 33.6	886 ± 341(536, 1982)	42.8
hous	+	102 ± 30(57, 175)	176 ± 71(67, 347)	− 34.8	66 ± 32(23, 137)	36.6
	−	186 ± 56(114, 325)	301 ± 100(170, 497)	− 35.9	199 ± 74(92, 389)	− 3.4
wisc	+	32 ± 16(13, 69)	54 ± 24(17, 104)	− 30.8	23 ± 10(11, 50)	27.0
	−	17 ± 4(12, 25)	11 ± 8(2, 32)	37.5	12 ± 5(0, 22)	31.6
btsc	+	9989 ± 3438(3671, 16827)	17409 ± 6811(5367, 35116)	− 36.4	6464 ± 2866(3100, 17253)	30.4
	−	7128 ± 3498(2705, 15217)	7184 ± 3534(2577, 16849)	2.8	4820 ± 2743(1937, 11800)	30.2
krkp	+	427 ± 156(204, 690)	748 ± 498(225, 2015)	− 25.7	373 ± 100(195, 550)	6.8
	−	255 ± 80(119, 435)	442 ± 247(189, 1063)	− 31.2	163 ± 66(85, 298)	33.7
phis	+	1386 ± 439(709, 2571)	3250 ± 1577(1114, 6902)	− 47.6	1231 ± 466(745, 2084)	8.0
	−	397 ± 147(197, 773)	395 ± 89(197, 572)	− 2.9	246 ± 92(141, 502)	29.9
qsar	+	1420 ± 335(912, 2345)	2120 ± 827(1087, 4235)	− 25.1	921 ± 354(519, 1947)	32.1
	−	2185 ± 860(846, 4215)	2804 ± 954(1347, 5371)	− 17.8	1498 ± 405(747, 2500)	25.6

Table 7 continued

Dataset	•	$(\underline{PG})^{\bullet}_{\texttt{mccormick}}$	$(\underline{PG})^{\bullet}_{\texttt{mccormick+CPLEX}}$	†	$(\underline{PG})^{\bullet}_{\texttt{stars}}$	‡
seis	+	35541 ± 9848(21565, 64075)	45322 ± 14038(22303, 76096)	− 16.4	17147 ± 6173(10478, 32554)	50.7
	−	18786 ± 6963(8758, 33856)	20702 ± 5911(10762, 32685)	− 7.5	11016 ± 4358(5550, 20207)	40.0
wilt	+	919 ± 359(426, 1851)	1275 ± 648(363, 2997)	− 18.2	535 ± 217(220, 1075)	38.5
	−	872 ± 302(421, 1485)	1212 ± 358(632, 1921)	− 25.4	751 ± 297(285, 1337)	11.7

$(\underline{PG})^{\bullet}_{\texttt{mccormick}}$ and $(\underline{PG})^{\bullet}_{\texttt{stars}}$ are solved without utilizing CPLEX cuts

$(\underline{PG})^{\bullet}_{\texttt{mccormick+CPLEX}}$ is $(\underline{PG})^{\bullet}_{\texttt{mccormick}}$ solved with CPLEX cuts utilized

All nodes are in format 'average ± 1 standard deviation (min, max)'

†: Average of $\frac{\text{Nodes by } (\underline{PG})^{\bullet}_{\texttt{mccormick}} - \text{Nodes by } (\underline{PG})^{\bullet}_{\texttt{mccormick+CPLEX}}}{\text{Worse of the 2 results}}$ values: measures relative efficacy of $(\underline{PG})^{\bullet}_{\texttt{mccormick+CPLEX}}$ over $(\underline{PG})^{\bullet}_{\texttt{mccormick}}$

‡: Average of $\frac{\text{Nodes by } (\underline{PG})^{\bullet}_{\texttt{mccormick}} - \text{Nodes by } (\underline{PG})^{\bullet}_{\texttt{stars}}}{\text{Worse of the 2 results}}$ values: measures relative efficacy of $(\underline{PG})^{\bullet}_{\texttt{stars}}$ over $(\underline{PG})^{\bullet}_{\texttt{mccormick}}$

Table 8 Number of stars and neighboring stars utilized

Dataset	$\bar{n}_s$	$\bullet$	$m_t^{\bullet}$	Stars		Neighboring stars		
				$\bar{n}_{\dagger}$	$\bar{d}$	$\bar{n}_{\ddagger}$	$\bar{q}$	$\lvert\bar{\mathscr{W}}\rvert$
bupa	17	+	97	$52 \pm 7(38, 68)$	$1.4 \pm 0.7(5)$	$14 \pm 2(10, 19)$	$3.7 \pm 2.4(13)$	$3.6 \pm 1.6(11)$
		−	133	$46 \pm 6(36, 63)$	$1.6 \pm 0.8(6)$	$14 \pm 3(9, 20)$	$3.3 \pm 1.9(10)$	$3.8 \pm 1.6(11)$
clev	10	+	91	$39 \pm 6(28, 48)$	$1.5 \pm 0.8(7)$	$12 \pm 2(8, 17)$	$3.2 \pm 2.1(13)$	$3.1 \pm 1.1(8)$
		−	107	$35 \pm 4(26, 45)$	$1.7 \pm 0.9(7)$	$13 \pm 2(9, 19)$	$2.6 \pm 1.3(8)$	$3.1 \pm 1.1(8)$
cred	14	+	238	$56 \pm 10(37, 77)$	$1.5 \pm 0.9(7)$	$19 \pm 5(11, 32)$	$3.0 \pm 2.1(12)$	$3.3 \pm 1.4(10)$
		−	197	$58 \pm 10(42, 79)$	$1.4 \pm 0.9(9)$	$20 \pm 5(12, 30)$	$2.9 \pm 2.2(13)$	$3.4 \pm 1.4(9)$
diab	19	+	179	$75 \pm 13(41, 93)$	$1.3 \pm 0.6(7)$	$16 \pm 3(10, 24)$	$4.7 \pm 3.1(15)$	$4.0 \pm 1.9(13)$
		−	333	$66 \pm 11(45, 87)$	$1.5 \pm 0.8(7)$	$16 \pm 3(9, 27)$	$4.0 \pm 2.3(11)$	$4.1 \pm 1.9(13)$
hous	15	+	171	$43 \pm 5(34, 58)$	$1.4 \pm 0.7(5)$	$12 \pm 3(8, 18)$	$3.5 \pm 2.0(12)$	$3.4 \pm 1.3(10)$
		−	166	$38 \pm 5(29, 48)$	$1.6 \pm 0.9(6)$	$13 \pm 2(9, 18)$	$2.9 \pm 1.9(9)$	$3.5 \pm 1.4(11)$
wisc	10	+	159	$16 \pm 3(10, 24)$	$1.9 \pm 1.1(6)$	$8 \pm 2(5, 11)$	$2.1 \pm 1.6(10)$	$3.1 \pm 1.0(8)$
		−	296	$22 \pm 4(14, 30)$	$1.4 \pm 0.7(5)$	$7 \pm 1(4, 10)$	$3.2 \pm 1.6(9)$	$3.0 \pm 1.0(7)$
btsc	39	+	83	$61 \pm 6(48, 77)$	$1.2 \pm 0.5(5)$	$13 \pm 2(9, 16)$	$4.8 \pm 3.5(18)$	$4.4 \pm 2.4(17)$
		−	282	$48 \pm 4(39, 56)$	$1.6 \pm 0.8(5)$	$13 \pm 2(9, 16)$	$3.7 \pm 3.0(15)$	$4.3 \pm 2.3(16)$
krkp	29	+	1113	$280 \pm 10(253, 302)$	$1.2 \pm 0.5(5)$	$158 \pm 9(142, 180)$	$1.8 \pm 1.1(9)$	$2.5 \pm 0.9(9)$
		−	1018	$289 \pm 11(270, 314)$	$1.2 \pm 0.5(5)$	$125 \pm 8(111, 145)$	$2.3 \pm 1.7(16)$	$2.8 \pm 1.4(15)$
phis	26	+	3963	$130 \pm 11(105, 153)$	$1.5 \pm 0.9(8)$	$34 \pm 4(25, 44)$	$3.8 \pm 2.5(15)$	$4.5 \pm 2.5(17)$
		−	3169	$136 \pm 16(115, 169)$	$1.4 \pm 0.8(7)$	$32 \pm 3(26, 39)$	$4.2 \pm 2.8(18)$	$4.5 \pm 2.3(14)$
qsar	20	+	237	$72 \pm 11(52, 91)$	$1.3 \pm 0.6(7)$	$17 \pm 3(12, 25)$	$4.3 \pm 2.6(13)$	$3.8 \pm 1.8(12)$
		−	466	$64 \pm 8(48, 78)$	$1.4 \pm 0.8(7)$	$15 \pm 2(11, 19)$	$4.2 \pm 2.8(14)$	$4.0 \pm 1.9(13)$

Table 8 continued

Dataset	$\bar{n}_s$	•	$m_t^{\bullet}$	Stars		Neighboring stars		
				$\bar{n}_\dagger$	$\bar{d}$	$\bar{n}_\ddagger$	$\bar{q}$	$\lvert\bar{\mathscr{U}}\rvert$
seis	35	+	113	$127 \pm 13(97, 152)$	$1.1 \pm 0.3(4)$	$18 \pm 2(14, 23)$	$7.1 \pm 4.9(27)$	$4.9 \pm 2.8(20)$
		−	1609	$67 \pm 6(57, 80)$	$2.1 \pm 1.3(10)$	$24 \pm 3(17, 31)$	$2.8 \pm 1.9(11)$	$4.6 \pm 2.1(15)$
wilt	27	+	174	$109 \pm 13(88, 154)$	$1.2 \pm 0.6(7)$	$27 \pm 6(18, 42)$	$4.0 \pm 3.0(19)$	$3.7 \pm 1.9(15)$
		−	3052	$69 \pm 7(52, 84)$	$1.9 \pm 1.1(8)$	$30 \pm 5(22, 49)$	$2.3 \pm 1.8(11)$	$3.7 \pm 1.7(14)$

$\bar{n}_s$: Average number of 0–1 attributes after feature selection 1 time (e.g., see [10,11])
$m_t^{\bullet}$: Number of • data for pattern generation (2/3 of those in Table 4)
$\bar{n}_\dagger$: Number of stars $\mathfrak{S}$ found in format 'average ± 1 standard deviation (minimum, maximum)'
$\bar{d}$: Degree (number of leaves) of $\mathfrak{S}$ in format 'average ± 1 standard deviation (maximum)'
$\bar{n}_\ddagger$: Number of star sets Q found in format 'average ± 1 standard deviation (minimum, maximum)'
$\bar{q}$: Cardinality of Q's in format 'average ± 1 standard deviation (maximum)'; trivially, minimum = 1
$\lvert\bar{\mathscr{U}}\rvert$: Number of terms in stars of Q in format 'average ± 1 standard deviation (maximum)'; trivially, minimum = 1

References

1. Abramson, S.D., Alexe, G., Hammer, P.L., Kohn, J.: A computational approach to predicting cell growth on polymeric biomaterials. J. Biomed. Mater. Res. **73A**, 116–124 (2005)
2. Alexe, G., Alexe, S., Axelrod, D., Hammer, P., Weissmann, D.: Logical analysis of diffuse large B-cell lymphomas. Artif. Intell. Med. **34**, 235–267 (2005)
3. Alexe, G., Alexe, S., Axelrod, D.E., Bonates, T., Lozina, I.I., Reiss, M., Hammer, P.L.: Breast cancer prognosis by combinatorial analysis of gene expression data. Breast Cancer Res. **8**, R41 (2006)
4. Alexe, G., Alexe, S., Hammer, P., Vizvari, B.: Pattern-based feature selections in genomics and proteomics. Ann. Oper. Res. **148**(1), 189–201 (2006)
5. Alexe, G., Alexe, S., Hammer, P.L.: Pattern-based clustering and attribute analysis. Soft. Comput. **10**, 442–452 (2006)
6. Alexe, G., Alexe, S., Liotta, L., Petricoin, E., Reiss, M., Hammer, P.: Ovarian cancer detection by logical analysis of data. Proteomics **4**, 766–783 (2004)
7. Alexe, S., Blackstone, E., Hammer, P.L., Ishwaran, H., Lauer, M.S., Snader, C.E.P.: Coronary risk prediction by logical analysis of data. Ann. Oper. Res. **119**, 15–42 (2003)
8. Balas, E., Mazzola, J.B.: Nonlinear 0–1 programming: I. Linearization techniques. Math. Program. **30**, 1–21 (1984)
9. Balas, E., Mazzola, J.B.: Nonlinear 0–1 programming: II. Dominance relations and algorithms. Math. Program. **30**, 22–45 (1984)
10. Boros, E., Hammer, P., Ibaraki, T., Kogan, A.: Logical analysis of numerical data. Math. Program. **79**, 163–190 (1997)
11. Boros, E., Hammer, P., Ibaraki, T., Kogan, A., Mayoraz, E., Muchnik, I.: An implementation of logical analysis of data. IEEE Trans. Knowl. Data Eng. **12**, 292–306 (2000)
12. Brauner, M.W., Brauner, N., Hammer, P.L., Lozina, I., Valeyre, D.: Logical analysis of computed tomography data to differentiate entities of idiopathic interstitial pneumonias. Data Min. Biomed. **7**, 193–208 (2007)
13. Crama, Y.: Concave extensions for nonlinear 0–1 maximization problems. Math. Program. **61**, 53–60 (1993)
14. Fortet, R.: L'algèbre de boole dt ses applications en recherche opérationnelle. Cahiers du Centre d'Études de Recherche Opérationnelle **1**(4), 5–36 (1959)
15. Fortet, R.: Applications de l'algèbre de boole en recherche opérationnelle. Revue Française d'Informatique et de Recherche Opérationnelle **4**(14), 17–25 (1960)
16. Glover, F., Woolsey, E.: Converting the 0–1 polynomial programming problem to a 0–1 linear program. Oper. Res. **12**(1), 180–182 (1974)
17. Granot, F., Hammer, P.: On the use of boolean functions in 0–1 programming. Methods Oper. Res. **12**, 154–184 (1971)
18. Hammer, A., Hammer, P., Muchnik, I.: Logical analysis of Chinese labor productivity patterns. Ann. Oper. Res. **87**, 165–176 (1999)
19. IBM Corp.: IBM ILOG CPLEX Optimization Studio CPLEX User's Manual Version 12 Release 8 (2017). https://www.ibm.com/support/knowledgecenter/SSSA5P_12.8.0/ilog.odms.studio.help/pdf/usrcplex.pdf. Accessed 22 Jun 2018
20. Jocelyn, S., Chinniah, Y., Ouali, M.S., Yacout, S.: Application of logical analysis of data to machinery-related accident prevention based on scarce data. Reliab. Eng. Syst. Saf. **159**, 223–236 (2017)
21. Kim, K., Ryoo, H.: A lad-based method for selecting short oligo probes for genotyping applications. OR Spectr. **30**(2), 249–268 (2008)
22. Kronek, L.P., Reddy, A.: Logical analysis of survival data: prognostic survival models by detecting high-degree interactions in right-censored data. Bioinformatics **24**, i248–i253 (2008)
23. Lauer, M., Alexe, S., Blackstone, E., Hammer, P., Ishwaran, H., Snader, C.P.: Use of the logical analysis of data method for assessing long-term mortality risk after exercise electrocardiography. Circulation **106**, 685–690 (2002)
24. Lichman, M.: UCI machine learning repository (2013). http://archive.ics.uci.edu/ml. Accessed 22 Jun 2018
25. Nguyen, T.T., Tawarmalani, M, Richard, J.-P.P.: Convexification techniques for linear complementarity constraints. In: 15th Conference on Integer Programming and Combinatorial Optimization (2011)
26. McCormick, G.: Computability of global solutions to factorable nonconvex programs: part I—convex underestimating problems. Math. Program. **10**, 147–175 (1976)
27. Mortada, M.A., Yacout, S., Lakis, A.: Fault diagnosis in power transformers using multi-class logical analysis of data. J. Intell. Manuf. **25**, 1429–1439 (2014)
28. Rikun, A.: A convex envelope formula for multilinear functions. J. Global Optim. **10**, 425–437 (1997)

29. Ryoo, H.S., Jang, I.Y.: MILP approach to pattern generation in logical analysis of data. Discrete Appl. Math. **157**, 749–761 (2009)
30. Ryoo, H.S., Sahinidis, N.: Analysis of bounds for multilinear functions. J. Global Optim. **19**(4), 403–424 (2001)
31. Yan, K., Ryoo, H.S.: 0–1 multilinear programming as a unifying theory for LAD pattern generation. Discrete Appl. Math. **218**, 21–39 (2017)
32. Yan, K., Ryoo, H.S.: Strong valid inequalities for Boolean logical pattern generation. J. Global Optim. **69**(1), 183–230 (2017)