不完全例題に対するブール的解析

Endre Boros§, 茨木俊秀†, 牧野和久†‡
§ラトガース大学ラトコー研究所
† 606-01 京都市左京区吉田本町 京都大学 工学部 数理工学教室
‡日本学術振興会特別研究員

あらまし　本論文では, データからの知識獲得の形態として, 不完全なビットをもつ部分定義論理関数 $(\tilde{T},\tilde{F})$ が拡張を持つかどうかを決定する問題を考察する. ただし, $\tilde{T}\subseteq\{0,1,*\}^n$ は正例の集合, $\tilde{F}\subseteq\{0,1,*\}^n$ は負例の集合を表し, "*" はデータにおける不完全ビットを表している. 論理関数 $f:\{0,1\}^n\mapsto\{0,1\}$ が正例の集合 $\tilde{T}$ と負例の集合 $\tilde{F}$ に対応する論理ベクトルをそれぞれ真, 偽にするとき, f を $(\tilde{T},\tilde{F})$ の拡張という. より正確には, 不完全ビットをどう扱うかによって, コンシステント, ロバスト, 最大ロバストと呼ばれる 3 種類の拡張を定義する. 本論文では, 様々な制限の下でそれらの拡張に関する問題に対して, 多項式時間アルゴリズムの存在, あるいは, NP-困難であるを示す.

和文キーワード: 不完全ビットをもつ部分定義論理関数, コンシステント拡張, ロバスト拡張, 最大ロバスト拡張, 知識獲得.

Boolean Analysis of Incomplete Examples

Endre Boros§, Toshihide IBARAKI† and Kazuhisa MAKINO†‡
§RUTCOR, Rutgers University, P.O. Box 5062, New Brunswick, NJ 08903, USA.
†Department of Applied Mathematics and Physics, Graduate School of Engineering, Kyoto University, Kyoto 606-01, Japan.
‡Research Fellow of the Japan Society for the Promotion of Science

abstract As a form of knowledge acquisition from data, we consider the problem of deciding whether there exists an extension of a partially defined Boolean function with missing bits $(\tilde{T},\tilde{F})$, where $\tilde{T}$ (resp., $\tilde{F}$) is a set of positive (resp., negative) examples. Here, "*" denotes a missing bit in the data, and it is assumed that $\tilde{T}\subseteq\{0,1,*\}^n$ and $\tilde{F}\subseteq\{0,1,*\}^n$ hold. A Boolean function $f:\{0,1\}^n\mapsto\{0,1\}$ is an extension of $(\tilde{T},\tilde{F})$ if it is true (resp., false) for the Boolean vectors corresponding to positive (resp., negative) examples; more precisely, we define three types of extensions called consistent, robust and most robust, depending upon how to deal with missing bits. We then provide polynomial time algorithms or prove their NP-hardness for the problems under various restrictions.

英文 key words: partially defined Boolean function with missing bits, consistent extension, robust extension, most robust extension, knowledge acquisition.

1 Introduction

The knowledge acquisition in the form of Boolean logic has been intensively studied in the recent research (e.g., [3, 6, 9]): given a set of data, represented as a set T of binary "true n-vectors" (or "positive examples") and a set F of "false n-vectors" (or "negative examples"), establish a Boolean function (extension) f, such that f is true (resp., false) in every given true (resp., false) vector; i.e., $T\subseteq T(f)$ and $F\subseteq F(f)$, where $T(f)$ (resp., $F(f)$) denotes the set of true vectors (resp., the set of false vectors) of f. A pair of sets (T,F) is called a *partially defined Boolean function* (*pdBf*).

For instance, data x represent the symptoms to diagnose a disease, e.g., x_1 denotes whether temperature is high ($x_1=1$) or not ($x_1=0$), and x_2 denotes whether blood pressure is high ($x_2=1$) or not ($x_2=0$), etc. Establishing an extension f, which is consistent with the given data, amounts to finding a logical diagnostic explanation of the given data. Therefore, this may be considered as a form of knowledge acquisition from given examples. It is evident that the problem is closely related to learning theory [1, 10], in which a systematic improvement of the obtained extensions is also taken into account.

Unfortunately, the real-world data might not be complete. As for the above examples, for some data x, temperature might not be measured, that is, it is not known whether $x_1 = 0$ or 1, which is represented as $x_1 = *$. For another instance, we have a battery of 45 biochemical tests for carcinogenicity. However, we do not usually apply all tests, since all tests cannot be checked in a laboratory or some tests are very expensive. When a test is not applied, we say that the test result is *missing*. A set of data $(\tilde{T}, \tilde{F})$, which includes the missing results, is called a *partially defined Boolean function with missing bits (pBmb)*, where $\tilde{T}$ (resp., $\tilde{F}$) $\subseteq \{0, 1, *\}^n$ denotes the set of "positive examples" (resp., "negative examples") of such vectors. To cope with such situations, we introduce in this paper three types of complete Boolean functions called *consistent*, *robust* and *most robust* extensions, respectively. More precisely, given a pBmb $(\tilde{T}, \tilde{F})$, (i) a consistent extension (CE) is a Boolean function f such that, for every $\tilde{a} \in \tilde{T}$ (resp., $\tilde{F}$), there is a 0-1 vector a obtained from $\tilde{a}$ by fixing missing bits appropriately, for which $f(a) = 1$ (resp., $f(a) = 0$) holds, (ii) a robust extension (RE) is a Boolean function f such that, for every $\tilde{a} \in \tilde{T}$ (resp., $\tilde{F}$), any 0-1 vector a obtained from $\tilde{a}$ by fixing missing bits arbitrarily satisfies $f(a) = 1$ (resp., $f(a) = 0$), and (iii) a most robust extension (MRE) is a Boolean function f which is a robust extension of a pBmb (T', F'), where (T', F') is obtained from $(\tilde{T}, \tilde{F})$ by fixing a smallest set of missing bits appropriately (the remaining missing bits in $T' \cup F'$ are assumed to take arbitrary values). All of these extensions provide logical explanations of a given pBmb $(\tilde{T}, \tilde{F})$ with varied freedom given to the missing bits in $\tilde{T}$ and $\tilde{F}$. By definition, if $(\tilde{T}, \tilde{F})$ has a robust extension f, it is also a most robust extension and is a consistent extension, and if $(\tilde{T}, \tilde{F})$ has a most robust extension f, it is a consistent extension. In case of most robust and consistent extensions, they also provide information such that some missing data must take certain values if $(\tilde{T}, \tilde{F})$ can have a consistent extension.

Table 1: Summary of complexity results.

Restrictions	RE	CE	MRE
$\lvert AS(a)\rvert \le 1$ for all $a \in \tilde{T} \cup \tilde{F}$	P	P	P
$\lvert AS(a)\rvert \le 2$ for all $a \in \tilde{T} \cup \tilde{F}$	P	NPC	NPH
General case	P	NPC	NPH

P: Polynomial, NPC: NP-complete, NPH: NP-hard

In this paper, we study the problems of deciding the existence of these extensions for a given pBmb $(\tilde{T}, \tilde{F})$, under various restrictions, mainly from the view point of their computational complexity. We obtain computationally efficient algorithms in some cases, and prove NP-completeness in some other cases, as summarized in Table 1, where $AS(a)$ denotes the set of missing bits in a vector a.

2 Partially Defined Boolean Functions with Missing Bits

A *Boolean function*, or a *function* in short, is a mapping $f : \mathbf{B}^n \mapsto \mathbf{B}$, where $\mathbf{B} = \{0, 1\}$, and $V \in \mathbf{B}^n$ is called a *Boolean vector* (or a *vector* in short). If $f(V) = 1$ (resp., 0), then V is called a *true* (resp., *false*) vector of f. The set of all true vectors (resp., false vectors) is denoted by $T(f)$ (resp., $F(f)$).

A *partially defined Boolean function (pdBf)* is defined by a pair of sets (T, F) such that $T, F \subseteq \mathbf{B}^n$. A function f is called an *extension* (or a *theory*) of the pdBf (T, F) if $T \subseteq T(f)$ and $F \subseteq F(f)$. Obviously, a pdBf (T, F) has an extension if and only if $T \cap F = \emptyset$.

To handle missing components, we introduce set $\mathbf{M} = \{0, 1, *\}$, and interpret the asterisk components $*$ of $v \in \mathbf{M}^n$ as missing bits. For a vector $v \in \mathbf{M}^n$, let $ON(v) = \{j \mid v_j = 1, j = 1, 2, \ldots, n\}$ and $OFF(v) = \{j \mid v_j = 0, j = 1, 2, \ldots, n\}$. For a subset $\tilde{S} \subseteq \mathbf{M}^n$, let $AS(\tilde{S}) = \{(v, j) \mid v \in \tilde{S}, j \in V \setminus (ON(v) \cup OFF(v))\}$ be the collection of all missing bits of the vectors in $\tilde{S}$. If $\tilde{S}$ is a singleton $\{v\}$, we shall write simply

$AS(v)$ instead of $AS(\{v\})$. Clearly, $\mathbf{B}^n \subseteq \mathbf{M}^n$, and $v \in \mathbf{B}^n$ if and only if $AS(v) = \emptyset$. Let us consider binary assignments $\alpha \in \mathbf{B}^Q$ to subsets $Q \subseteq AS(\tilde{S})$ of the missing bits. For a vector $v \in \tilde{S}$ and an assignment $\alpha \in \mathbf{B}^Q$, let v^α denote the vector obtained from v by replacing the $*$ components belonging to Q by the binary values assigned to them by α, i.e.,

$$v_j^\alpha = \begin{cases} v_j & \text{if } (v,j) \notin Q \\ \alpha(v,j) & \text{if } (v,j) \in Q. \end{cases}$$

For vectors $v, w \in \mathbf{M}^n$, we shall write $v \approx w$ if there exists an assignment $\alpha \in \mathbf{B}^{AS(\{v,w\})}$ for which $v^\alpha = w^\alpha$ holds, and we say that v is *potentially identical* with w.

A *pdBf with missing bits* (or in short *pBmb*) is a pair $(\tilde{T}, \tilde{F})$, where $\tilde{T}, \tilde{F} \subseteq \mathbf{M}^n$. To a pBmb $(\tilde{T}, \tilde{F})$ we always associate the set $AS = AS(\tilde{T} \cup \tilde{F})$ of its missing bits. A function f is called a *robust extension* of the pBmb $(\tilde{T}, \tilde{F})$ if

$$f(a^\alpha) = 1 \text{ and } f(b^\alpha) = 0 \quad \text{for all } a \in \tilde{T},\ b \in \tilde{F} \text{ and for all } \alpha \in \mathbf{B}^{AS}.$$

We first consider the problem of deciding the existence of a robust extension of a given pBmb $(\tilde{T}, \tilde{F})$.

Robust Extension (RE)
Input: A pBmb $(\tilde{T}, \tilde{F})$, where $\tilde{T}, \tilde{F} \subseteq \mathbf{M}^n$.
Question: Does $(\tilde{T}, \tilde{F})$ have a robust extension ?

It may happen that a pBmb $(\tilde{T}, \tilde{F})$ has no robust extension, but it has an extension if we change some (or all) $*$ bits to appropriate binary values. A function f is called a *consistent extension* of pBmb $(\tilde{T}, \tilde{F})$, if there exists an assignment $\alpha \in \mathbf{B}^{AS}$ for which $f(a^\alpha) = 1$ and $f(b^\alpha) = 0$ for all $a \in \tilde{T}$ and $b \in \tilde{F}$. In other words, a pBmb $(\tilde{T}, \tilde{F})$ is said to have a consistent extension if, for some assignment $\alpha \in \mathbf{B}^{AS}$, the pdBf $(\tilde{T}^\alpha, \tilde{F}^\alpha)$ defined by $\tilde{T}^\alpha = \{a^\alpha \mid a \in \tilde{T}\}$ and $\tilde{F}^\alpha = \{b^\alpha \mid b \in \tilde{F}\}$ has an extension. This leads us to the following decision problem.

Consistent Extension (CE)
Input: A pBmb $(\tilde{T}, \tilde{F})$, where $\tilde{T}, \tilde{F} \subseteq \mathbf{M}^n$.
Question: Does $(\tilde{T}, \tilde{F})$ have a consistent extension ?

It may also happen that not all missing bits are necessary to be specified in order to have a robust extension. An assignment $\alpha \in \mathbf{B}^Q$ for a subset $Q \subseteq AS$ is called a *robust assignment* if the resulting pBmb $(\tilde{T}^\alpha, \tilde{F}^\alpha)$ has a robust extension. We are interested in finding a robust assignment with the smallest size $|Q|$.

Most Robust Extension (MRE)
Input: A pBmb $(\tilde{T}, \tilde{F})$, where $\tilde{T}, \tilde{F} \subseteq \mathbf{M}^n$.
Output: NO if $(\tilde{T}, \tilde{F})$ does not have a consistent extension; otherwise a robust assignment $\alpha \in \mathbf{B}^Q$ for a subset $Q \subseteq AS$, which minimizes $|Q|$.

It follows from definition that if RE or CE are NP-complete, then MRE is NP-hard, and conversely, if MRE is solvable in polynomial time, then both RE and CE are polynomially solvable.

Let us add that we shall also consider various restricted variants of the above problems, in which the input pBmb $(\tilde{T}, \tilde{F})$ is restricted to satisfy the following condition for a given k:

$$|AS(a)| \leq k, \text{ for every } a \in \tilde{T} \cup \tilde{F}.$$

3 Robust and Consistent Extensions

Theorem 1 *Problem* RE *can be solved in polynomial time.*

Proof. It is easy to see that a pBmb $(\tilde{T}, \tilde{F})$ has a robust extension if and only if there exists an index j such that $a_j \neq b_j$ and $\{a_j, b_j\} = \{0, 1\}$ (i.e., either $a_j = 0$ and $b_j = 1$, or $a_j = 1$ and $b_j = 0$) for every pair of $a \in \tilde{T}$ and $b \in \tilde{F}$. Obviously, this can be checked in $O(n|T||F|)$ time. □

Let us note next that CE can be trivially solved if $|AS(a)| > 0$ holds for all $a \in \tilde{T} \cup \tilde{F}$, since in this case $(\tilde{T}, \tilde{F})$ always has a consistent extension f. Indeed, let us consider an assignment $\alpha \in \mathbf{B}^{AS}$ such that $|ON(a^\alpha)|$ is odd for all $a \in \tilde{T}$, and $|ON(b^\alpha)|$ is even for all $b \in \tilde{F}$, and let f be the parity function for which $f(v) = 1$ if and only if $|ON(v)|$ is odd. Problem CE becomes more complicated when not all input vectors have missing bits, although it remains polynomially solvable if each input vector contains at most one missing bit.

Theorem 2 *Problem* CE *can be solved in polynomial time for a pBmb* $(\tilde{T}, \tilde{F})$ *for which every* $a \in \tilde{T} \cup \tilde{F}$ *satisfies* $|AS(a)| \leq 1$.

Proof. Let j_a be the index of the $*$ in each vector $a \in \tilde{T} \cup \tilde{F}$ (i.e., $AS(a) = \{(a, j_a)\}$), if any. Then $(\tilde{T}, \tilde{F})$ has a consistent extension if and only if (i) there is no pair of $a \in \tilde{T}$ and $b \in \tilde{F}$ such that $a, b \in \mathbf{B}^n$ and $a = b$, and (ii) there is an assignment $\alpha \in \mathbf{B}^{AS}$ satisfying the conditions

$$\alpha(a, j_a) \neq b_{j_a} \quad \text{if } a \notin \mathbf{B}^n \text{ and } b \in \mathbf{B}^n \qquad (1)$$

$$\alpha(b, j_b) \neq a_{j_b} \quad \text{if } a \in \mathbf{B}^n \text{ and } b \notin \mathbf{B}^n \qquad (2)$$

$$\alpha(a, j_a) \neq b_{j_a} \text{ or } \alpha(b, j_b) \neq a_{j_b} \quad \text{if } a, b \notin \mathbf{B}^n \text{ and } j_a \neq j_b \qquad (3)$$

$$\alpha(a, j_a) \neq \alpha(b, j_b) \quad \text{if } a, b \notin \mathbf{B}^n \text{ and } j_a = j_b \qquad (4)$$

for every pair of $a \in \tilde{T}$ and $b \in \tilde{F}$ with $a \approx b$. Obviously, condition (i) can be checked in $O(n|\tilde{T}||\tilde{F}|)$ time. To check (ii), let us observe that each of the conditions (1)–(4) can equivalently be represented as clauses in the variables $\alpha(v, j)$ for $(v, j) \in AS$. Namely, (1) and (2) can be represented by linear clauses, (3) by a clause containing two variables, and (4) by the conjunction of two clauses, each of which contains two variables. E.g. (4) is equivalent with the condition

$$1 = (\alpha(a, j_a) \vee \alpha(b, j_b))(\overline{\alpha(a, j_a)} \vee \overline{\alpha(b, j_b)}).$$

In total, we have a 2-SAT problem containing at most $2|\tilde{T}||\tilde{F}|$ clauses, which is solvable in time linear in its input size (see e.g., [2]). This shows that problem CE can be solved in $O(n|\tilde{T}||\tilde{F}|)$ time. □

Example 1 *Let us define* $\tilde{T}, \tilde{F} \subseteq \{0,1\}^3$ *by*

$$\tilde{T} = \left\{ \begin{array}{l} a^{(1)} = (1,1,*) \\ a^{(2)} = (0,0,1) \\ a^{(3)} = (0,1,*) \\ a^{(4)} = (*,0,0) \end{array} \right\}, \tilde{F} = \left\{ \begin{array}{l} b^{(1)} = (1,1,1) \\ b^{(2)} = (0,*,1) \\ b^{(3)} = (*,0,0) \end{array} \right\}.$$

Then we have the following 2-SAT:

$$\overline{\alpha(a^{(1)},3)}\,\alpha(b^{(2)},2)\left(\overline{\alpha(a^{(3)},3)} \vee \overline{\alpha(b^{(2)},2)}\right)$$
$$\left(\alpha(a^{(4)},1) \vee \alpha(b^{(3)},1)\right)\left(\overline{\alpha(a^{(4)},1)} \vee \overline{\alpha(b^{(3)},1)}\right) = 1.$$

For this, the assignment $\alpha \in \mathbf{B}^{AS}$ *given by* $\alpha(a^{(1)},3) = \alpha(a^{(3)},3) = \alpha(a^{(4)},1) = 0$ *and* $\alpha(b^{(2)},2) = \alpha(b^{(3)},1) = 1$, *is a satisfying solution.* □

In general, however, we have the following negative result (see [4]).

Theorem 3 *Problem* CE *is NP-complete, even if* $|AS(a)| \leq 2$ *holds for all* $a \in \tilde{T} \cup \tilde{F}$.

4 Most Robust Extensions

As Theorem 3 implies that MRE is NP-hard even if $|AS(a)| \leq 2$ for all $\tilde{T} \cup \tilde{F}$. Therefore, we only consider the case in which

$$|AS(a)| \leq 1 \text{ for all } a \in \tilde{T} \cup \tilde{F},$$

and show that it can be solved in polynomial time.

Let us remark first that any assignment $\alpha \in \mathbf{B}^{AS}$ for which $(\tilde{T}^\alpha, \tilde{F}^\alpha)$ has an extension must satisfy the conditions (i) and (ii) in the proof of Theorem 2. Hence, some components of such an α may be forced to take a unique binary value by conditions (1) and (2). Let us assume therefore that we fix all such asterisks $*$ in advance, and let us consider only conditions (3) and (4) in the sequel.

Let us define next a bipartite graph

$$\begin{array}{l} G_{AS} = (V, E), \\ V = AS(\tilde{T}) \cup AS(\tilde{F}), \text{ and} \\ E = \{(q, r; \alpha) \mid q = (a, i) \in AS(\tilde{T}), r = (b, j) \in \\ \quad AS(\tilde{F}), \text{ there exists an assignment } \alpha \in \\ \quad \mathbf{B}^{\{q,r\}} \text{ such that } a^\alpha = b^\alpha\}, \end{array} \qquad (5)$$

where the label $c(e)$ of each edge $e = (q, r; c(e))$, as defined in (5), is called the *configuration* of e. If there are more than one assignments $\alpha \in \mathbf{B}^{\{q,r\}}$ for some $q \in AS(\tilde{T})$ and $r \in AS(\tilde{F})$, for which $a^\alpha = b^\alpha$ (this occurs if $q = (a, i)$ and $r = (b, j)$ satisfy $i = j$), then the graph G_{AS} has parallel edges corresponding to such different configurations. Let us note that, since $|AS(a)| \leq 1$ holds for all $a \in \tilde{T} \cup \tilde{F}$, every pair of $q = (a, i) \in AS(\tilde{T})$ and $r = (b, j) \in AS(\tilde{F})$ has at most two assignments $\alpha \in \mathbf{B}^{\{q,r\}}$ such that $a^\alpha = b^\alpha$.

Example 2 *Let us define* $\tilde{T}, \tilde{F} \subseteq \{0,1\}^6$ *by*

$$\tilde{T} = \left\{ \begin{array}{l} a^{(1)} = (*,1,1,1,1) \\ a^{(2)} = (1,1,1,1,*) \\ a^{(3)} = (1,1,1,*,1) \\ a^{(4)} = (1,1,*,1,1) \\ a^{(5)} = (1,*,0,1,0) \end{array} \right\}, \tilde{F} = \left\{ \begin{array}{l} b^{(1)} = (1,*,1,1,1) \\ b^{(2)} = (1,1,1,1,*) \\ b^{(3)} = (1,1,*,1,0) \\ b^{(4)} = (1,1,0,1,*) \end{array} \right\}.$$

Then the graph G_{AS} is given in Figure 1. Although the configurations of edges are not indicated, they are easy to find out. For example, the edge $e = (a^{(1)}, b^{(1)})$ has $c(e) = (a_1^{(1)} = 1, b_2^{(1)} = 1)$, and the parallel edges $e' = (a^{(2)}, b^{(2)})$ and $e'' = (a^{(2)}, b^{(2)})$ have $c(e') = (a_5^{(2)} = 0, b_5^{(2)} = 0)$ and $c(e'') = (a_5^{(2)} = 1, b_5^{(2)} = 1)$, respectively. □

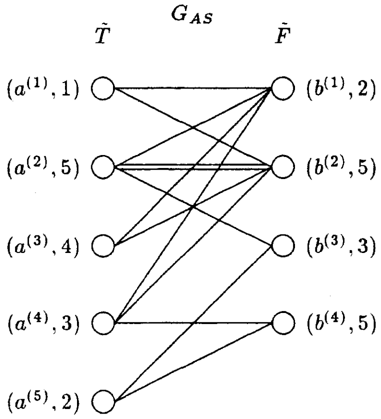

Figure 1: The graph G_{AS} of the pBmb $(\tilde{T}, \tilde{F})$ in Example 2.

Lemma 1 *Given a pBmb $(\tilde{T}, \tilde{F})$, an assignment $\beta \in \mathbf{B}^Q$ for a subset $Q \subseteq AS$ is a robust assignment of $(\tilde{T}, \tilde{F})$ (i.e., $(\tilde{T}^\beta, \tilde{F}^\beta)$ has a robust extension) if and only if, for every edge $e = (q, r; \alpha)$ of G_{AS}, we have either $q = (a, i) \in Q$ and $a^\beta \neq a^\alpha$, or $r = (b, j) \in Q$ and $b^\beta \neq b^\alpha$, or both.*

Proof. Let us first show the only-if-part. Let f be a robust extension of $(\tilde{T}^\beta, \tilde{F}^\beta)$, and let $e = (q, r; \alpha)$ be an edge of G_{AS}. We can assume, without loss of generality that $q = (a, i) \in AS(\tilde{T})$.

Let us assume that either $q \notin Q$ or $a^\beta = a^\alpha$. Let us show first that $f(a^\alpha) = 1$. Indeed, if $q = (a, i) \notin Q$, then $(a^\beta)^\alpha = a^\alpha$, and since $\beta \in \mathbf{B}^Q$ is a robust assignment, $f(a^\alpha) = 1$ must hold. On the other hand, if $a^\beta = a^\alpha$, then obviously $f(a^\alpha) = f(a^\beta) = 1$ must hold, since $a \in \tilde{T}$.

We then show that $f(a^\alpha) = 1$ implies $r = (b, j) \in Q$ and $b^\beta \neq b^\alpha$. If $r \notin Q$, then $(b^\beta)^\alpha = b^\alpha = a^\alpha$, and hence $f(a^\alpha) = f(b^\alpha) = 0$ by $b \in \tilde{F}$, which is a contradiction. Similarly $b^\beta = b^\alpha$ leads to the same contradiction. Hence $r \in Q$ and $b^\beta \neq b^\alpha$ must hold.

To prove the if-part, assume that $\beta \in \mathbf{B}^Q$ for a subset $Q \subseteq AS$ is not a robust assignment of $(\tilde{T}, \tilde{F})$. Then, by the definition of robustness, we have a pair of vectors $a \in \tilde{T}$ and $b \in \tilde{F}$ such that $a^\beta \approx b^\beta$. Then the edge $e = (q, r; \alpha)$ with $q = (a, i)$ and $r = (b, j)$ does not satisfy the statement of the lemma. □

For a vector $d \in \mathbf{B}^n$, let $E(d)$ denote the set of edges $e = (q, r; \alpha) \in E$ with $a^\alpha = b^\alpha = d$, where $q = (a, i)$ and $r = (b, j)$ Then $E = \cup_d E(d)$. Let us define a *coherent domain* $D(d)$ as the set of vertices incident to some edges of $E(d)$, and let D_0 denote the set of isolated vertices (i.e., incident to no edge $e \in E$). (Vertices in D_0 do not belong to any coherent domain.) In the following discussion, we only consider nonempty coherent domains. Figure 2 shows all nonempty coherent domains of the graph G_{AS} of $(\tilde{T}, \tilde{F})$ in Example 2.

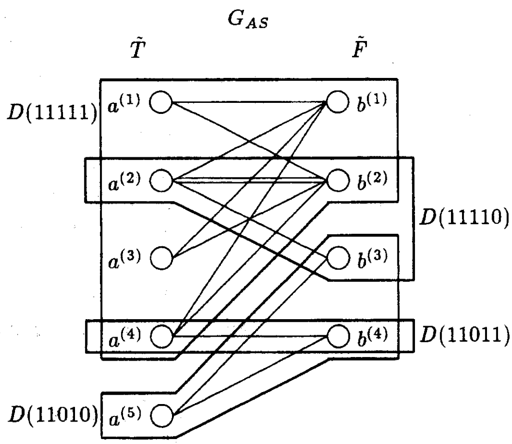

Figure 2: Coherent domains of the graph G_{AS} of $(\tilde{T}, \tilde{F})$ in Example 2.

Lemma 2 *Every coherent domain $D(d) \subseteq V$ of G_{AS} induces a complete bipartite subgraph of G_{AS}.*

Proof. Take any pair $q = (a, i) \in AS(\tilde{T})$ and $r = (b, j) \in AS(\tilde{F})$ that satisfy $q, r \in D(d)$. Then there exist assignments $\alpha \in \mathbf{B}^{\{q\}}$ and $\beta \in$

$\mathbf{B}^{\{r\}}$ such that $d = a^\alpha = b^\beta$. We concatenate these assignments to have an assignment $\gamma = (\alpha, \beta) \in \mathbf{B}^{\{q,r\}}$ for which $a^\gamma = b^\gamma = d$, implying that there is an edge $(q, r) \in E(d)$. □

Lemma 3 *Let $D(d)$ and $D(d')$ be two coherent domains of G_{AS}, where $d, d' \in \mathbf{B}^n$ and $d \neq d'$. If $D(d) \cap D(d') \neq \emptyset$, then $\| d - d' \| = 1$ holds, where $\| x \| = \sum_{i=1}^n |x_i|$.*

Proof. Let $q = (a, i) \in D(d) \cap D(d')$. Then there exist two assignments $\alpha, \beta \in \mathbf{B}^{\{q\}}$ ($= \{0, 1\}$) such that $a^\alpha = d$ and $a^\beta = d'$. Since $|AS(a)| \leq 1$ is assumed, $\| d - d' \| = 1$ is implied. □

Lemma 4 *Let $D(d)$ and $D(d')$ be two coherent domains of G_{AS}, where $d, d' \in \mathbf{B}^n$ and $d \neq d'$. Then $|D(d) \cap D(d')| \leq 2$ holds. Furthermore, if $D(d) \cap D(d') = \{q, r\}$, then the graph G_{AS} has two parallel edges between q and r.*

Proof. If $q = (a, i), r = (b, j) \in D(d) \cap D(d')$, then by assigning 0 and 1 to q and r, each of a and b can become both d and d'. Since $\| d - d' \| = 1$ by Lemma 3, this can only happen if the vectors a and b are identical, missing the same component $i = j$. Therefore $|D(d) \cap D(d') \cap AS(\tilde{T})| \leq 1$ and $|D(d) \cap D(d') \cap AS(\tilde{F})| \leq 1$, and hence $|D(d) \cap D(d')| \leq 2$. Finally, if $D(d) \cap D(d') = \{q, r\}$, where $q = (a, i) \in AS(\tilde{T})$ and $r = (b, j) \in AS(\tilde{F})$, then $q = r$ implies that there are two assignments $\alpha, \beta \in \mathbf{B}^{\{q,r\}}$ such that $a^\alpha = b^\alpha = d$ and $a^\beta = b^\beta = d'$, i.e. the graph G_{AS} has two parallel edges between q and r. □

Let us now color the edges of G_{AS} by "yellow" and "blue", so that all edges of a set $E(d)$ have the same color, and every pair of sets $E(d)$ and $E(d')$ with $D(d) \cap D(d') \neq \emptyset$ has different colors. We call such a two coloring *alternating.* The following lemma shows that an alternating coloring is always possible. Furthermore, it can be uniquely completed after fixing a color of a set $E(d)$ in each connected component of G_{AS}.

Lemma 5 *Let $D(d^{(0)}), D(d^{(1)}), \ldots, D(d^{(l)})$ denote a cycle of coherent domains such that $d^{(i-1)} \neq d^{(i)}$ and $D(d^{(i-1)}) \cap D(d^{(i)}) \neq \emptyset$ hold for all $i = 1, 2, \ldots, l-1$, and $D(d^{(l)}) = D(d^{(0)})$. Then l is even.*

Proof. Lemma 3 tells that $\| d^{(i-1)} - d^{(i)} \| = 1$ holds for all $i = 1, 2, \ldots, l-1$. Since $\| d^{(0)} - d^{(l)} \| = 0$ is even, l must be even. □

Finally, let us orient the edges of G_{AS} according to a given alternating coloring, as follows. Every yellow edge (q, r) is oriented from $q \in AS(\tilde{T})$ to $r \in AS(\tilde{F})$, and every blue edge (q, r) is oriented from $r \in AS(\tilde{F})$ to $q \in AS(\tilde{T})$. Let G'_{AS} denote the resulting directed graph. E.g. Figure 3 shows the directed graph G'_{AS} corresponding to the pBmb $(\tilde{T}, \tilde{F})$ of Example 2. Let us observe that every directed path of this graph is alternating in colors, and every alternating undirected path is either forward directed or backward directed. The next lemma

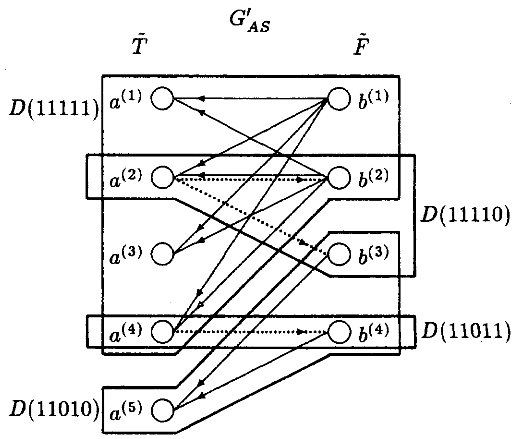

Figure 3: The directed graph G'_{AS} of $(\tilde{T}, \tilde{F})$ in Example 2.

characterizes a robust assignment by a directed path of G'_{AS}.

Lemma 6 *Let $(\tilde{T}, \tilde{F})$ be a pBmb, and let $q^{(0)} \xrightarrow[e_1]{} q^{(1)} \xrightarrow[e_2]{} q^{(2)} \ldots q^{(l-1)} \xrightarrow[e_l]{} q^{(l)}$ be a directed path in G'_{AS}. Then $\beta \in \mathbf{B}^Q$ for $Q \subseteq AS$ is a robust assignment if and only if the following properties hold, where $q^{(i)} = (a^{(i)}, j_i)$ and $\alpha_i = c(e_i)$ for all i.*

(i) If $q^{(0)} \notin Q$ or $(a^{(0)})^\beta = (a^{(0)})^{\alpha_1}$, then $q^{(i)} \in Q$ and $(a^{(i)})^\beta \neq (a^{(i)})^{\alpha_i}$ hold for all $i = 1, 2, \ldots, l$.

(ii) *If* $q^{(l)} \notin Q$ *or* $(a^{(l)})^\beta = (a^{(l)})^{\alpha_l}$ *for some* $l > 0$, *then* $q^{(i)} \in Q$ *and* $(a^{(i)})^\beta \neq (a^{(i)})^{\alpha_{i+1}}$ *hold for all* $i = 0, 1, \ldots, l-1$.

Proof. We first prove the only-if-part. For condition (i), we first consider $e_1 = (q^{(0)}, q^{(1)})$. By Lemma 1, $q^{(0)} \notin Q$ or $(a^{(0)})^\beta = (a^{(0)})^{\alpha_1}$ implies that $q^{(1)} \in Q$ and $(a^{(1)})^\beta \neq (a^{(1)})^{\alpha_1}$. Now, since $e_1 = (q^{(0)}, q^{(1)}) \in E(d)$ and $e_2 = (q^{(1)}, q^{(2)}) \in E(d')$ have different colors, we must have $d \neq d'$ and $q^{(1)} \in D(d) \cap D(d')$, and hence $\| d - d' \| = 1$ by Lemma 3. Therefore, $(a^{(1)})^\beta \neq (a^{(1)})^{\alpha_1}$ $(= d)$ implies $(a^{(1)})^\beta = (a^{(2)})^{\alpha_2}$ $(= d')$, and hence $q^{(2)}$ satisfies $(a^{(2)})^\beta \neq (a^{(2)})^{\alpha_2}$ by Lemma 1. This assignment can proceed in a similar manner to $q^{(i)}$, $i = 2, 3, \ldots, l$. Case (ii) is similar to (i).

Conversely, if conditions (i) and (ii) hold, then, by Lemma 1, $\beta \in \mathbf{B}^Q$ is a robust assignment. □

Let C_i, $i = 1, 2, \ldots, s$, denote all the strongly connected components of this directed graph G'_{AS}. Furthermore, let G^*_{AS} denote the transitive closure of G'_{AS} (i.e., (s,t) is an arc in G^*_{AS} if there is an s-t directed path in G'_{AS}), and let G_0 denote the directed subgraph of G^*_{AS} induced by

$$W = \cup_{i \text{ s.t. } |C_i|=1} C_i. \tag{6}$$

It is easy to see that the set of isolated vertices D_0 in G_{AS} satisfies $D_0 \subseteq W$. Figure 4 contains the graph G_0 of $(\tilde{T}, \tilde{F})$ in Example 2, where, for simplicity, arcs (u,v), for which there is a directed path of length at least 2 from u to v, are not indicated.

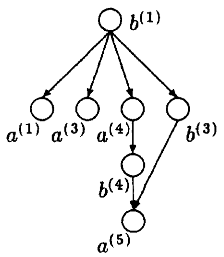

Figure 4: The graph G_0 corresponding to G'_{AS} of $(\tilde{T}, \tilde{F})$ in Example 2.

Lemma 7 *Let* $(\tilde{T}, \tilde{F})$ *be a pBmb, let* $\alpha \in \mathbf{B}^Q$ *for some* $Q \subseteq AS$ *be a robust assignment, and let* C_i *and* W *be defined as above. Then the following two conditions hold:*

(i) $C_i \subseteq Q$ *for all* C_i *with* $|C_i| > 1$, *and*

(ii) $W \setminus Q$ *is an antichain in* G_0 *(i.e., for any pair of* $q, r \in W \setminus Q$, *there is no directed path from* q *and* r *in* G_0, *and vice versa).*

Proof. Consider a robust assignment $\alpha \in \mathbf{B}^Q$. Assume $q \in C_i \setminus Q$ for some C_i with $|C_i| > 1$. Then there is a directed cycle $q^{(0)} (= q)$, $q^{(1)}, q^{(2)}, \ldots, q^{(l)} (= q)$ of length $l > 1$ in G'_{AS}, and $q \notin Q$ implies $q \in Q$ by Lemma 6, which is a contradiction. Hence condition (i) holds. To prove condition (ii), let us assume that for some pair of $q, r \in W \setminus Q$, there exists a directed path from q and r in G'_{AS}. This is again a contradiction since $q \notin Q$ implies $r \in Q$ by Lemma 6. □

Lemma 8 *Let* $(\tilde{T}, \tilde{F})$ *be a pBmb, and let* $S \subseteq W$ *be any maximal antichain in* G_0. *Then for* $Q = AS \setminus S$, *there is a robust assignment* $\alpha \in \mathbf{B}^Q$ *of* $(\tilde{T}, \tilde{F})$.

Proof. For $Q = AS \setminus S$, we shall construct a robust assignment $\alpha \in \mathbf{B}^Q$. In the following, we shall consider the directed graph G'_{AS}, and let us note that, by definition, S is also an antichain in G'_{AS}. Lemma 6 tells that, starting from a vertex $q \in S$ (i.e., $q \notin Q$), a robust assignment β for all vertices t which are either reachable from q or reachable to q is uniquely determined, unless the following cases of conflicts are encountered.

(i) For $q, r \in S$, there is a vertex t for which there are two directed paths $P_1 = q^{(0)} (= q) \to q^{(1)} \to \ldots \to q^{(k)} (= t)$ and $P_2 = r^{(0)} (= r) \to r^{(1)} \to \ldots \to r^{(l)} (= t)$ such that $t^\alpha \neq t^{\alpha'}$, where $\alpha = c(q^{(k-1)}, t)$ and $\alpha' = c(r^{(l-1)}, t)$.

(ii) For $q, r \in S$, there is a vertex t for which there are two directed paths $P_1 = q^{(0)} (= t) \to q^{(1)} \to \ldots \to q^{(k)} (= q)$ and $P_2 = r^{(0)} (= t) \to r^{(1)} \to \ldots \to r^{(l)} (= r)$ such that $t^\alpha \neq t^{\alpha'}$, where $\alpha = c(t, q^{(1)})$ and $\alpha' = c(t, r^{(1)})$.

If one of these conflicts occurs, Lemma 6 tells that t must be assigned in different ways, and hence we cannot construct an appropriate robust assignment β.

However, we now show that none of these conflicts can occur. Let us consider case (i) only, since case (ii) can be analogously treated. Now $t^{\alpha} \neq t^{\alpha'}$ implies $(q^{(k-1)}, s) \in E(d)$ and $(r^{(l-1)}, s) \in E(d')$ for some $d \neq d'$. Thus $(q^{(k-1)}, t)$ and $(r^{(l-1)}, t)$ have different colors, since $D(d) \cap D(d') \neq \emptyset$. By the rule of orienting edges (yellow edges are oriented from $AS(\tilde{T})$ to $AS(\tilde{F})$, and blue edges are oriented from $AS(\tilde{F})$ to $AS(\tilde{T})$), this means that one of $(q^{(k-1)}, t)$ and $(r^{(l-1)}, t)$ is oriented towards t, and the other is away from t, a contradiction to the assumption in (i).

Let us denote by R the set of all vertices $t \notin S$ such that either t is reachable from some $q \in S$ or some $q \in S$ is reachable from t. The above argument shows that a robust assignment β for R is uniquely determined by Lemma 6. Finally, we consider an assignment $\gamma \in \mathbf{B}^{AS \setminus (S \cup R)}$. By the maximality of S, every vertex $t \in AS \setminus (S \cup R)$ has an incoming arc $e = (r, t) \in E(d)$. Therefore, determine the robust assignment β of this t so that $t^{\beta} = d$ holds. This is well-defined because all incoming arcs to t belong to the same $E(d)$ by the definition of G'_{AS}. It is easy to see that the resulting β over AS is in fact a robust assignment. □

Lemmas 7 and 8 tell that problem MRE is equivalent to the problem of finding a maximum antichain of G_0. Since G_0 is acyclic, we can find such an antichain in polynomial time by Dilworth's theorem (see e.g. [8]). Hence, we have shown the following theorem.

Theorem 4 *Problem* MRE *can be solved in polynomial time for a pBmb* $(\tilde{T}, \tilde{F})$ *in which all* $a \in \tilde{T} \cup \tilde{F}$ *satisfy* $|AS(a)| \leq 1$.

5 Discussion

From the view point of knowledge acquisition, it is interesting to consider extensions for restricted classes of Boolean functions, because some structural information that justifies such restrictions might be available beforehand. In this case, however, even the problem of deciding whether there is an extension or not for a given pdBf (T, F) may not be trivial, depending on the class of functions at hand. This problem and the problem of finding an extension with a minimum number of errors are extensively discussed in [3] for such classes as positive (i.e., monotone) functions, Horn functions, functions with k-DNF and/or h-term-DNF, threshold functions, regular functions, read-once functions, self-dual functions, decomposable functions and so on. This direction is further pursued in [4] to consider the problems RE, CE and MRE of this paper, assuming that a pBmb $(\tilde{T}, \tilde{F})$ is given.

References

[1] D. Angluin, Queries and concept learning, Machine Learning 2 (1988) 319-342.

[2] B. Apswall, M. F. Plass, R. E. Tarjan, A linear-time algorithm for testing the truth of certain quantified Boolean formulas. Information Processing Letters 8 (1979) 121-123

[3] E. Boros, T. Ibaraki, K. Makino, Error-free and best-fit extensions of a partially defined Boolean function, RUTCOR Research Report RRR 14-95, Rutgers University 1995.

[4] E. Boros, T. Ibaraki, K. Makino, Extensions of partially defined Boolean functions with missing data, RUTCOR Research Report RRR 06-96, Rutgers University 1996.

[5] E. Boros, T. Ibaraki, K. Makino, Boolean Analysis of Incomplete Examples, SWAT'96, LNCS 1097 (1996) 440-451.

[6] Y. Crama, P. L. Hammer, T. Ibaraki, Cause-effect relationships and partially defined boolean functions, Annals of Operations Research 16 (1988) 299-326.

[7] R. Dechter, J. Pearl, Structure identification in relational data, Artificial Intelligence 58 (1992) 237-270.

[8] L. R. Ford, D. R. Fulkerson, Flows in Networks, Princeton University Press 1962.

[9] J. R. Quinlan, Induction of decision trees, Machine Learning 1 (1986) 81-106.

[10] L. G. Valiant, A theory of the learnable, Communications of the ACM 27 (1984) 1134-1142.