

SAT 问题中隐蔽集求解的改进

李淑霞¹, 龚茜茹¹, 谷文祥²

(1 河南工业职业技术学院 计算机工程系, 河南 南阳 473000; 2 东北师范大学 计算机学院, 吉林 长春 130117)

摘 要: 隐蔽集(backdoor sets)作为隐蔽结构的一种,能有效地提高难求解问题的求解效率,近年来成为人们研究的热点.隐蔽集中变量的赋值能有效减少 SAT 问题求解的搜索分支,从而减少问题求解的时间复杂度和空间复杂度.为提高 SAT 问题的求解效率,提出一种求解 SAT 问题隐蔽集的改进算法,并给出最小隐蔽集的定义.在该算法中加入启发式,使求解出的隐蔽集变量个数较少,最后给出隐蔽集问题的总结和展望.

关键词: SAT 问题;隐蔽集;隐蔽结构;最小隐蔽集;隐蔽集变量

中图分类号: TP18

文献标识码: A

文章编号: 1000-7180(2014)07-0065-04

The Improvment Backdoor Sets for SAT

LI Shu-xia¹, GONG Qian-ru¹, GU Wen-xiang²

(1 Dept. of Computer Engineering, Henan Polytechnic Institute, Nanyang 473000, China;

2 School of Computer Science, Northeast Normal University, Changchun 130117, China)

Abstract: Backdoor is one of these structures, which can effectively improve the efficiency of the SAT problem solving, and which become a focus of study in recent years. The variable assignment for backdoors can reduce the search branch of SAT problem solving process effectively, thereby reducing the time complexity and space complexity of sat problem solver. In order to improve the efficiency of the SAT problem, this paper presents the improved algorithm of backdoor sets for sloving SAT problem, and provides the definition of the smallest backdoor sets. The heuristic is joined in this algorithm, so the smaller backdoor sets can be solved in this way, Finally, this paper proposed summary and outlook.

Key words: Propositional Satisfiability problem; backdoor sets; hidden structure; the smallest backdoor sets; Variables of backdoors

1 引言

人工智能领域的研究存在一些很难求解问题,如命题可满足问题(Propositional Satisfiability problem, 简称 SAT 问题)^[1],该类问题的计算复杂性是 NP 完备的,若 $P \neq NP$ 成立,该类问题是无法在多项式时间内求解的.隐蔽集(backdoor sets)^[2]是隐蔽结构的一种,与某些问题(如 SAT 问题、QBF 问题)的求解难度有很密的关系.如果知道这些难求解问题的隐蔽集,通过该隐蔽集中变量的真值指派,可以使剩下的问题变为多项式时间内可解的问题.

从而提高了 SAT 问题(或 QBF 问题)的求解效率.此时隐蔽集的求解对该类较难求解问题来说尤其重要.

隐蔽集(backdoor sets)是 SAT 问题的重要隐蔽结构,通过对隐蔽集中变量的真值指派能简化 SAT 问题,且简化后的 SAT 问题在多项式时间内可解.2003 年 B. Selman 等人提出了隐蔽集^[3]的概念并用实验证明现实实例中存在一些较小隐蔽集.某一问题的隐蔽集不只一种,各种隐蔽集的变量个数也不同.隐蔽集中变量个数直接影响问题(如 SAT 问题)的求解效率,所以隐蔽集的求解在 SAT

收稿日期: 2013-09-22; 修回日期: 2013-11-10

基金项目: 国家自然科学基金(61070084)

问题求解中具有关键作用。

目前求解较难问题的隐蔽集算法有多种,如文献[4-11],其中 2004 年 S. Szeider 等人提出一种基于参数 k 的求解隐蔽集的 FPT 算法^[12],该算法求解出的隐蔽集变量个数小于等于参数 k ,但不是最小的.本文以此为基础,提出一种改进算法,使求解出的隐蔽集变量个数较少;并给出最小隐蔽集的定义,用实例说明最小隐蔽集的求解能减少求解问题的时间复杂度和空间复杂度.最后用理论的方法证明这一算法的可行性和正确性.

2 相关概念

SAT 问题是确定给定的命题公式是否可满足,即给定一个命题逻辑公式 $F = (a \vee b) \wedge (\bar{a} \vee c)$,其变量集为 $|X| = \text{var}(F) = \{a, b, c\}$,是否存在对其变量集 $|X|$ 的一个真值指派使命题公式 F 可满足,如果可满足则返回这些变量的真值赋值.1971 年 Cook 首次证明了 SAT 问题是 NP-完全的^[13],几乎所有 NP 完全问题都可以在多项式时间内归约到 SAT 问题,其求解更快捷.

隐蔽集是 SAT、约束可满足问题(即 CSP)等问题的一个重要隐藏结构,该隐藏结构与问题难度有很大的关系,隐蔽集的求解能使较难问题的求解变得容易求解.目前求解的隐蔽集大多数是强隐蔽集,强隐蔽集^[12]是指给定一个命题逻辑公式 $F, B \in \text{var}(F)$ 是一个非空变量子集,如果 B 中变量的任一真值指派 τ 都能使得简化公式 $F[\tau]$ 在多项式时间内可解,那么变量集合 B 就是强隐蔽集.删除隐蔽集是指给定一个命题逻辑公式 $F, B \in \text{var}(F)$ 是一个非空变量子集,如果简化公式 $F - B$ 可在多项式时间内求解,那么变量集合 B 是删除隐蔽集.如果简化公式属于多项式时间可解的基本类 C ,且 C 是封闭(closed)的,此时删除隐蔽集就是强隐蔽集^[12].所以比较流行的求解强隐蔽集的方法是用求解删除隐蔽集的方法来求解的.本文同样采用求解删除隐蔽集的方法来求解强隐蔽集.

对一个命题逻辑公式 F 来说,其公式 $F = \bigwedge_{i=1}^k c_i$ 是由 1 个或多个子句 c_i 合取组成,子句 $c_i = \bigvee x_{ij}$ 是由一个或多个字面量(literal)的析取组成.用 $\text{var}(F)$ 表示命题公式中出现的变量集, $F[\tau]$ 表示应用真值指派 τ 后的简化公式, $F - B$ 表示从公式 F 中的子句中删除变量集 B 中出现的变量所得的简化公式,比如命题公式 $F = (a \vee b \vee c) \wedge (a \vee c \vee d)$,其中 $B = \{a\}$, $\text{var}(F) = \{a, b, c, d\}$, $F - B$

$= (b \vee c) \wedge (c \vee d)$,当 $\tau = b$ 时,简化公式 $F[\tau]$ 可以表示为 $F[\tau|_{b=1}] = (a \vee c \vee d) \vee F[\tau|_{b=0}] = (a \vee c) \wedge (a \vee c \vee d)$.用 $\text{num}(B)$ 表示变量集 B 中变量的个数.对一个命题逻辑公式 F 来说,存在多种隐蔽集,即 SAT 问题的隐蔽集是不确定的.而本文中所指的隐蔽集是 SAT 问题的较小隐蔽集.

定义 1 最小隐蔽集 给定一个命题逻辑公式 $F, B \in \text{var}(F)$ 是一个非空变量子集,如果 B 中变量的任一真值指派 τ 都能使得 $F[\tau]$ 在多项式时间内可解,且不存在其他的强隐蔽集 B' 使得 $\text{num}(B') < \text{num}(B)$ 成立,则变量集合 B 就是最小隐蔽集.

比如命题逻辑公式 $F = (a \vee b \vee d) \wedge (b \vee c \vee e)$,存在的 horn 隐蔽集有 $B_1 = \{a, c\}, B_2 = \{b\}$,由于 $\text{num}(B_2) < \text{num}(B_1)$,且没有其他的隐蔽集 B 使得 $\text{num}(B) < \text{num}(B_2)$,故 B_2 是我们求解的最小隐蔽集.如果隐蔽集为 B_1 ,此公式的求解过程可以形象化为图 1 的搜索树,其中 $\square$ 表示叶节点,0 和 1 分别表示变量的真值指派.如果隐蔽集为 B_2 ,此公式的求解过程可以形象化为图 2 的搜索树.从图 1 和图 2 的搜索树中可以看出当隐蔽集为 B_2 时其分支数少于隐蔽集为 B_1 的分支数,且其求解过程的搜索空间也少于隐蔽集为 B_1 时求解过程的搜索空间.故较小隐蔽集能够减少 SAT 问题求解过程的时间复杂度和空间复杂度.

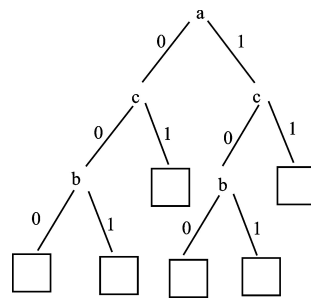


图 1 搜索树

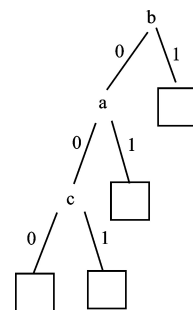


图 2 搜索树

3 改进的求解 SAT 问题的隐蔽集算法

由于较小隐蔽集可以减少 SAT 问题求解过程

的分支数及求解过程的搜索空间. 本文在求解 SAT 问题隐蔽集的算法中加入启发式, 使求解的隐蔽集变量个数减少, 具体算法如下:

3.1 Horn 子句作为子求解器

由于 Horn 子句在多项式时间内可解的^[14], 根据隐蔽集的定义, 可将其作为子求解器, 算法如下:

算法 1 Min-horn-backdoor

输入: SAT 公式 F

输出: 隐蔽集

过程:

```
(1)  $B = \varnothing$ 
(2) while ( $F \notin \text{horn}$ )
(3) { vector<int> count(var_num, 0);
(4)   count = count_formula( $F_{\text{horn}}$ , var_num, count);
(5)   int right_var = right_lits(count);
(6)    $B = \text{BD\_part}(B, \text{right\_var})$ ;
(7)    $F = \text{simplify}(F, \text{right\_var})$ ;
(8) }
(9) return  $B$ 
```

该算法是以 Horn 子句为子求解器, 求解出实例的较小隐蔽集. 该算法中第一行置隐蔽集 B 为空; 第三行定义一个向量变量 count , 并将其中的所有变量初始化为 0, var_num 表示问题实例中的变量个数; 第四行调用子函数 $\text{count_formula}(F_{\text{horn}}, \text{var_num}, \text{count})$, 在其中应用了启发式, 对非 horn 子句中正文字 x 进行计数 $\text{count}[x]$; 第五行是调用子函数 $\text{right_lits}(\text{count})$ 来选择计数文字 $\text{count}[x]$ 最大的文字 x 所对应的变量, 并将该变量记为 right_var ; 第六行调用子函数 $\text{BD_part}(B, \text{right_var})$, 将变量 right_var 放到较小隐蔽集 B 中, 接着调用第 7 行的函数 $\text{simplify}(F, \text{right_var})$, 在此用变量 right_var 的真值指派 τ 来简化公式; 直到简化的问题实例 F 为 Horn 公式. 比如随机生成的 SAT 实例 $F = (2 \vee 4 \vee 3 \vee 5) \wedge (-6 \vee 7 \vee -5 \vee 2) \wedge (-4 \vee -6 \vee 5 \vee 3) \wedge (1 \vee 5 \vee 4 \vee -6) \wedge (2 \vee 5 \vee 1 \vee -6) \wedge (-6 \vee -1 \vee 5 \vee 2) \wedge (-2 \vee 4 \vee -3 \vee -5) \wedge (4 \vee -6 \vee 3 \vee -1) \wedge (-4 \vee 6 \vee 5 \vee -1) \wedge (-1 \vee 4 \vee 2 \vee 3)$, 用模拟算法求得的隐蔽集为 $B = \{2, 4, 3, 1, 6\}$, 用加了启发式的改进算法求解的隐蔽集 $B = \{5, 2, 4\}$, 从中可以看出用改进的求解隐蔽集算法可以求解出较小的隐蔽集, 从而减少了问题求解过程的搜索空间, 提高了问题的求解效率.

3.2 2CNF 子句作为子求解器

由于 2CNF 子句与 HORN 子句一样, 也是多项

时间内可解的类^[14], 故根据隐蔽集的定义, 也可将其作为求解隐蔽集的子求解器, 算法如下:

算法 2: Min-2cnf-backdoor

输入: SAT 公式 F

输出: 隐蔽集

过程:

```
(1)  $B = \varnothing$ 
(2) while ( $F \notin 2\text{cnf}$ )
(3) { vector<int> count(var_num, 0);
(4)   count = count_formula( $F_{2\text{cnf}}$ , var_num, count);
(5)   int  $x = \text{right\_lits}(\text{count})$ ;
(6)    $B = \text{BD\_part}(B, x)$ ;
(7)    $F = \text{simplify}(F, x)$ ;
(8) }
(9) return  $B$ 
```

该算法是以 2CNF 子句为子求解器, 求解出实例的较小隐蔽集. 该算法中第一行置隐蔽集 B 为空; 第三行定义一个向量变量 count , 并将其中的变量初始化为 0, var_num 表示问题实例中的变量个数; 第四行加入了启发式, 即对 F 公式中的非 2cnf 子句中的所有文字 x 或 $-x$ 所对应的变量 x 进行计数 $\text{count}[x]$, 如果简化公式中 2cnf 子句的个数没有发生变化, 则不需对公式中的变量重新计数, 否则需对简化公式重新计数. 第五行是调用子函数 $\text{right_lits}(\text{count})$ 来选择计数 $\text{count}[x]$ 最大变量 x ; 第六行将变量 x 放到较小隐蔽集 B 中, 即 $B = B \cup \{x\}$; 第 7 行调用子函数 $\text{simplify}(F, x)$, 即用变量 $F = F - \{x\}$ 来简化公式; 直到简化的问题实例 F 为 2cnf 公式, 如公式 $F = (1 \vee 5 \vee 4 \vee -6) \wedge (2 \vee 5 \vee 1 \vee -6) \wedge (-6 \vee -1 \vee 5 \vee 2) \wedge (-2 \vee 4 \vee -3 \vee -5) \wedge (4 \vee -6 \vee 3 \vee -1) \wedge (-4 \vee 6 \vee 5 \vee -1) \wedge (-1 \vee 4 \vee 2 \vee 3) \wedge (2 \vee 4 \vee 3 \vee 5) \wedge (-4 \vee -6 \vee 5 \vee 3)$, 用模拟算法求得的隐蔽集为 $B = \{1, 5, 2, 4\}$, 用启发式的改进算法求解的隐蔽集为 $B = \{5, 4, 1\}$

引理 1 应用该启发式求解出的隐蔽集为较小隐蔽集(比较的前提是求解时所应用的子求解器相同)

证明: 假设存在一个 m 个变量 n 个子句的命题公式 F , 其变量数分别为 $x_1, x_2, x_3, \dots, x_m$; 子句分别为 $C_1, C_2, C_3, \dots, C_n$. 当子句 $C_j \notin \text{horn}$ 子句时, 对 C_j 中的正文字 x_i 进行计数, 记为 $\text{count}(x_i)$. 该启发式选择的是计数 $\text{count}(x_i)$ 最大的变量 x_i 作为隐蔽集, 并用该变量来简化命题公式 F . 由于该启发式选择计数 $\text{count}(x_i)$ 最大的变量 x_i 化简公式, 故化简后的 horn 子句个数增加最多, 非 horn 子句个数减少

最多,同理在剩下的非 horn 子句中选择正文字计数 $\text{count}(x_j)$ 最大的变量 x_j ,继续这样的过程直到所有子句都为 horn 子句,即命题公式为 horn 公式. 由于这样的启发式一次能尽可能多地减少非 horn 子句的个数,故用该启发式求解的隐蔽集满足 $\text{num}(B) < \text{num}(B)$,其中 B 为其他任意隐蔽集,所以该方法求解出的隐蔽集是较小隐蔽集.

同理,可以说明该方法求解出的 2cnf 隐蔽集也是较小隐蔽集.

4 结束语

本文在深入研究隐蔽集的基础上,提出一种改进的求解隐蔽集算法,在该算法中加入启发式,使求解出的隐蔽集变量个数比较少,并用实例来说明该算法能求解出的隐蔽集变量数小于之前求解出的隐蔽集变量个数,并给出最小隐蔽集的概念. 目前还没有求解最小隐蔽集的算法,如果能找到较难求解问题的最小隐蔽集,则可以更好地提高该问题的求解效率,所以如何求解问题的最小隐蔽集是一个有待解决的问题.

参考文献:

- [1] Garla P Gomes, Henry Kautz, Ashish Sabharwal, et al. Satisfiability solvers[C]//In Frank van Harmelen, Vladimir Lifschitz, and Bruce Porter, editors, Handbook of knowledge representation. Elsevier, 2008: 89-134.
- [2] Kilby P, Slaney J, Thiebaux S, et al. Backbones and backdoors in satisfiability [C]// AAAI' 05. USA: Pennsylvania, 2005:1368-1373.
- [3] Williams R, Gomes C, Selman B. Backdoors to typical case complexity[C]//Proc of the 18th International Joint Conference on Artificial Intelligence, IJCAI'03. Mexico, 1173-1178.
- [4] Szeider S. Backdoor sets for DLL subsolvers[J]. Journal of Automated Reasoning, 2005, 35(1-3): 73-88.
- [5] Samer M, Szeider S. Backdoor sets of quantified Boolean formulas// SAT' 07. Berlin, LNCS 4501, 2007: 230-243.
- [6] Interian Y. Backdoor sets for random 3-SAT[C]//Sixth International Conference on Theory and Applications of Satisfiability Testing, S. Margherita Ligure, Italy, 2001:231-238.
- [7] Paris L, Ostrowski R, Siegel P, et al. Computing horn strong backdoor sets thanks to local search [C]//Washington DC, ICTAI', 2006:139-143.
- [8] Szeider S, Sakallah K A. Matched formulas and backdoor sets[C]// SAT 2007. Berlin, Springer, LNCS, 2007(4501):94-99.
- [9] Dilkina B, Gomes C P, Sabharwal A. Tradeoffs in the complexity of backdoor detection[C]//CP 2007. USA: Providence LNCS, Springer, 2007(4741):256-270.
- [10] Paris L, Ostrowski R, Siegel Gelbukh P A, et al. From horn strong backdoor sets to ordered strong backdoor sets[C]//MICA 2007. Mexico, LNAI 4827, 2007: 105-117.
- [11] Kottler S, Kaufmann M, Sinz C. Computation of renameable horn backdoors [C]// SAT 2008. China: Guangzhou, LNCS 4996, 2008:154-160.
- [12] Nishimura N, Ragde P, Szeider S. Detecting backdoor sets with respect to horn and binary clauses[C]//Proc of the seventh International Conference of Theory and Applications of Satisfiability Testing. Canada: Vancouver, 2004:96-103.
- [13] Stefan Szeider. Generalizations of matched CNF formulas[J]. Annals of Mathematics and Artificial Intelligence, 2005, 43(1-4):223-238.
- [14] Dowling W F, Gallier J H. Linear-time algorithms for testing the satisfiability of propositional Horn formulas [J]. J. Logic Programming, 1984, 1(3):267-284.

作者简介:

李淑霞 女, (1982-), 硕士研究生, 助教. 研究方向为智能规划与规划识别、嵌入式系统.

龚茜茹 女, (1978-), 硕士研究生, 讲师. 研究方向为图形与多媒体技术.

谷文祥 女, (1947-), 博士生导师. 研究方向为智能规划与规划识别、形式语言与自动机、模糊数学及其应用.