中国科学院软件研究所学术年会’2019
暨计算机科学国家重点实验室开放周

学术论文

基于哈希的模型计数估算算法

葛存菁　马菲菲 刘田 张健 马旭桐

题目：A New Probabilistic Algorithm for Approximate Model Counting

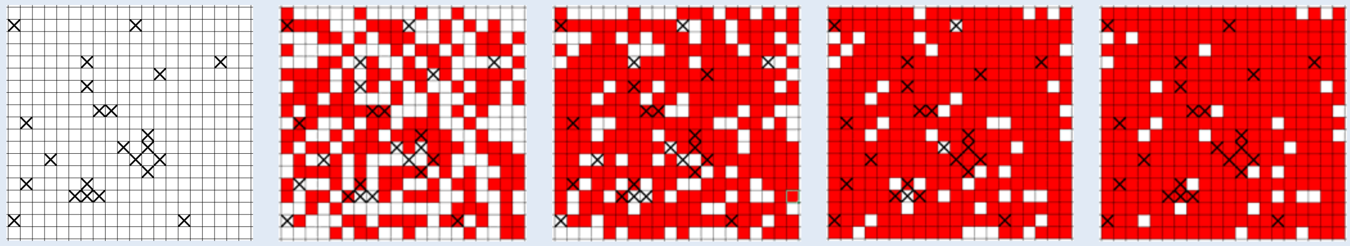
会议：9th International Joint Conference on Automated Reasoning (IJCAR), 312-328, 2018

葛存菁　18611824507　gecj@ios.ac.cn
马菲菲　13699168951　maff@ios.ac.cn

估算方法的直观

□：每个方块格子代表公式的一个赋值；☒：带叉的格子表示公式的解。

在每一个方块格子中掷硬币，若为正面，则涂红。重复若干轮，直到所有带叉的格子都被染红。直观上，一共有约$2^{轮数}$个带叉的格子。



初始 23个☒　　第一轮 13/23　　第二轮 8/23　　第三轮 3/23　　第四轮 0/23

哈希函数 & 哈希过程

定义： 给定公式F及其上变量$x_1, \dots, x_n$。考虑关于F的哈希函数$H(x) \equiv a_0 \oplus_{i=1}^{n} a_i x_i$，其中$a_0, \dots, a_n$为布尔常数。

性质： $Pr_{H \in \mathcal{H}_F}(H(\alpha) = true) = 1/2$，其中$\mathcal{H}_F$表示关于公式$F$的哈希函数的集合。

作用： 给公式F添加哈希函数约束，可以随机屏蔽掉一半的赋值，即实现对每个赋值掷硬币的过程。

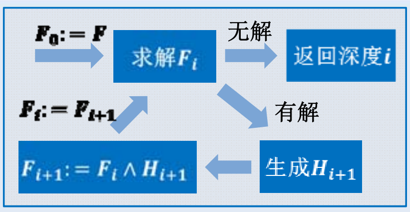
哈希过程： $F_i \equiv F \wedge H_1 \wedge \cdots \wedge H_i$。

$$\begin{cases} Pr_{F_d}(F_d(\alpha) = true) = 2^{-d} \Leftrightarrow Pr_{F_d}(F_d(\alpha) = false) = 1 - 2^{-d} & \text{（哈希函数性质）} \\ Pr_{F_d}(F_d(\alpha) = false, \alpha \in S_F) \approx \prod_{\alpha \in S_F} Pr_{F_d}(F_d(\alpha) = false) & \text{（假设，未证明）} \end{cases}$$

$$\Leftrightarrow Pr_{F_d}\ (F_d\ is\ unsat) \approx \left(1 - 2^{-d}\right)^{\#F} \Leftrightarrow \#F \approx \log_{(1-2^{-d})} Pr_{F_d}\ (F_d\ is\ unsat).$$

算法 & 实现

GetDepth()：如右图，从哈希函数空间中随机生成$F_0, \dots, F_d$，直到F_d无解，返回深度d。



STAC()：重复调用T次**GetDepth()**。统计T次中F_i为无解的次数C_i。选择一个最接近$T/2$的C_k，根据$\#F \approx \log_{(1-2^{-k})} C_k/T$估计解个数。

T的取值：基于前文中的假设，T可以取合适的值使得估计结果满足(ϵ, δ)-界，即估计结果落在区间$[(1+\epsilon)^{-1}\#F, (1+\epsilon)\#F]$中的概率至少为$1-\delta$。

动态终止条件：由于**STAC()**可以看成一个伯努利实验，于是可以计算二项比例置信区间，判断中间结果是否已经满足(ϵ, δ)-界。

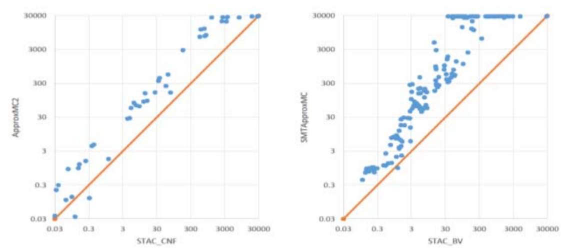
工具实现（https://github.com/bearben/STAC）：

- *STAC_CNF*，命题逻辑公式的模型计数估计；
- *STAC_BV*，SMT(BV)公式的解计数估计。

实验结果

Instance	n	$\#F$	$[1.8^{-1}\#F, 1.8\#F]$	Freq.	$\bar{t}$ (s)	$\bar{T}$	$\bar{Q}$
special-1	20	1.0×10^6	$[5.8 \times 10^5, 1.9 \times 10^6]$	82	0.3	12.2	86.7
special-2	20	1	$[0.6, 1.8]$	86	0.6	12.6	37.6
special-3	25	3.4×10^7	$[1.9 \times 10^7, 6.0 \times 10^7]$	82	11.2	11.8	90.1
5step	177	8.1×10^4	$[4.5 \times 10^4, 1.5 \times 10^5]$	90	0.1	11.9	80.5
blockmap_05_01	1411	6.4×10^2	$[3.6 \times 10^2, 1.2 \times 10^3]$	84	1.1	12.0	73.8
blockmap_05_02	1738	9.4×10^6	$[5.2 \times 10^6, 1.7 \times 10^7]$	89	12.7	11.8	87.7
blockmap_10_01	11328	2.9×10^6	$[1.6 \times 10^6, 5.2 \times 10^6]$	83	80.3	12.0	85.0
fs-01	32	7.7×10^2	$[4.3 \times 10^2, 1.4 \times 10^3]$	80	0.02	12.6	76.2
or-50-10-10-UC-20	100	3.7×10^6	$[2.0 \times 10^6, 6.6 \times 10^6]$	77	7.7	12.0	86.1
or-60-10-10-UC-40	120	3.4×10^6	$[1.9 \times 10^6, 6.1 \times 10^6]$	91	3.5	12.1	86.0

对每个例子100次实验的平均结果（$\epsilon = 0.8$，$\delta = 0.2$）*



与ApproxMC2和SMTApproxMC的比较实验（坐标轴为时间、8小时时限）

***基于假设，我们证明了算法可以返回(ϵ, δ)-界，即估计结果落在区间$[1.8^{-1}\#F, 1.8\#F]$中的概率至少为80%。实验表格中100次实验落在此区间中次数（Freq）也在80左右。**