

On Information Size of Combinatorial Problems

Maciej Drozdowski

Institute of Computing Science, Poznań University of Technology, Poland

This talk is about:

- combinatorial problem is a object, a thing, that can be measured
- exponential amount of information in SAT
- constant Kolmogorov complexity of SAT
- the consequences

SAT

Definition

INPUT I : sums $k_j, j = 1, \dots, m$, of binary variables, or their negations, chosen over a set of n binary variables $x_1, \dots, x_n$.

REQUEST: find an assignment of 0/1 values to $x_1, \dots, x_n$, i.e. vector $\bar{x}$, such that the conjunction $F(I, \bar{x}) = \prod_{j=1}^m k_j$ is 1.
If such a vector does not exist then signal $\emptyset$.

$|I|$ – instance I size, i.e., length of the string encoding I

"yes" instance – if for instance I : $\exists \bar{x} : F(I, \bar{x}) = 1$

"no" instance – otherwise

Search problem, e.g. SAT, as a string relation

Let:

Σ – an alphabet

e – some reasonable encoding scheme over Σ ,

Σ^+ – a set of strings encoding instances of SAT using scheme e over alphabet Σ .

SAT-search is an example of a string relation:

Definition

Search problem Π is a string relation

$$R[\Pi, e] = \left\{ (a, b) : \begin{array}{l} a \in \Sigma^+ \text{ is the encoding of } I \text{ and} \\ b \in \Sigma^+ \text{ is the encoding of a solution} \\ \text{under coding scheme } e \end{array} \right.$$

Information size and Kolmogorov complexity

Definition

Problem information size is the number of bits necessary to write it as a string relation.

Definition

Problem Kolmogorov complexity is the number of bits necessary to write the shortest program to write the problem as a string relation.

Truly random bit sequence

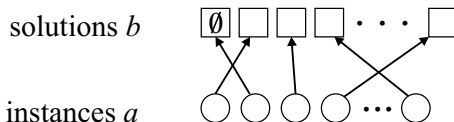
Definition

Truly random bit sequence (TRBS) is a sequence of bits, that has no shorter representation.

Thus:

- the only way to represent it is to store it in its whole entirety,
- an N -bit TRBS has information size N bits.
- an N -bit TRBS Kolmogorov information complexity is also N .

How much information in SAT?



SAT can be thought of as string relation $R[SAT, e]$:

- a mapping from strings a – instances, to strings b – solutions,
- set of arcs from instances a to solutions b ,

How much information does such an object comprises?

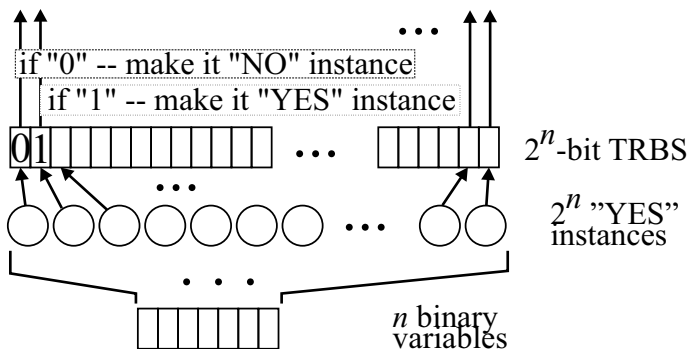
Exponential Information Size of SAT

Theorem

The amount of information in SAT grows at least exponentially with instance size.

Exponential Information Size of SAT

Proof: the idea



Exponential Information Size of SAT

Proof: 2^n of "Yes" instances in n variables

- n – number of variables,
- Let there be 4 clauses for $i = 1, \dots, n$:
 $k_{i1} = x_a + x_b + \tilde{x}_i$, $k_{i2} = \overline{x_a} + x_b + \tilde{x}_i$,
 $k_{i3} = x_a + \overline{x_b} + \tilde{x}_i$, $k_{i4} = \overline{x_a} + \overline{x_b} + \tilde{x}_i$.
- $\tilde{x}_i$ – a variable x_i with or without negation
- No valuing of x_a, x_b alone makes the four clauses simultaneously equal 1.
- The four clauses may simultaneously become equal 1 only if $\tilde{x}_i = 1$ (i.e. either $x_i = 1$ or $\overline{x_i} = 1$).
- Satisfying formula $F = k_{11}k_{12}k_{13}k_{14} \dots k_{n4}$ depends on valuing of variables $\tilde{x}_i$ for $i = 1, \dots, n$.
- There are 2^n different ways of constructing formula F , leading to 2^n different "yes" instances with 2^n different solutions.

Exponential Information Size of SAT

Proof: Injecting 2^n -bit-long TRBS into SAT

- Consider a truly random bit sequence (TRBS) of length 2^n .
- Consider $j = 0, \dots, 2^n - 1$ instances with clauses $k_{1j}, \dots, k_{4j}$ and variables $\tilde{x}_i$ as constructed above.
- If TRBS bit $j = 0, \dots, 2^n - 1$ **is 1**, then the j th instance is constructed as "yes" instance by setting $\tilde{x}_i$ in $k_{1j}, \dots, k_{4j}$ consistently with number j binary encoding, for $i = 1, \dots, n$.
- If TRBS bit $j = 0, \dots, 2^n - 1$ **is 0**, then the j th instance is **spoiled** to a "no" instance by setting some $\tilde{x}_i$ variable(s) in $k_{1j}, \dots, k_{4j}$ inconsistently (we have both x_i and $\bar{x}_i$).

Exponential Information Size of SAT

Proof: From the number of variables n to instance size $|I|$

- each number is limited from above by some constant K ,
- the length of the encoding of the instance data is
$$|I| = 4n \times 3 \log K + \log K = 12n \log K + \log K$$
- $\Rightarrow n = (|I| - \log K) / (12 \log K)$
- $\Rightarrow$ Number of incompressible bits of TRBS which is hidden in SAT is $\Omega(2^n) = \Omega(2^{d_1 |I|})$, where $d_1 = 1/(12 \log K) > 0$ is constant.
- $\Rightarrow$ the amount of information in SAT grows at least exponentially with instance size. □

SAT information size upper bound

Observation

SAT information size is $O(|I|2^{|I|})$.

Proof:

- there are $2^{|I|}$ instances of size $|I|$,
- $O(|I|)$ bits are necessary to record an instance and a solution in string relation pair (a, b) .

Kolmogorov complexity of SAT

SAT, as a string relation $R[\text{SAT}, e]$, can be recreated by the fixed length algorithm:

```
def G :  
  for  $n$  in  $1, 2, \dots$  :  
    for  $s$  in Binary_Strings_Long( $2n + 2n2^{2n}$ ) :  
       $l = n_2 | \sim | s$            #  $n_2$  binary encoded  $n$ ,  $| \sim |$  concatenation  
      SolutionExists=False  
      for  $\bar{x}$  in Binary_Strings_Long( $n$ ) :  
        if  $V(l, \bar{x})$  :         # solution verifier  $V$  exists since  $\text{SAT} \in \mathbf{NP}$   
          record( $l, \bar{x}$ )  
          SolutionExists=True  
          break  
      if not SolutionExists :  
        record( $l, \emptyset$ )
```

Kolmogorov complexity of SAT

Observation

Kolmogorov complexity of SAT, $K(R[\text{SAT}, e])$, is $O(1)$.

SAT Computational Complexity with Exponential Information

Observation

SAT is solvable in polynomial time using $O(|I|2^{|I|})$ amount of information by referring to precomputed solutions.

Proof:

- `solution ← TableOfSolutions[I];`
- in a data structure of size Z , a piece of information can be found in time $O(p(\log(Z)))$ by use of B/B+ trees, p is a polynomial,
- `TableOfSolutions[I]` has size $O(|I|2^{|I|})$,
- Hence computational complexity is $O(p(|I| + \log |I|))$. □

Implications of SAT exponential information size

- ① Randomness (represented by TRBS) and computational complexity (represented by **NP**-complete SAT) are connected,
- ② SAT ultimately is not learnable for machine learning problems because it comprises pattern-less TRBSs,
- ③ this can be extended to **NP**-complete/hard problems which reduce to SAT,

Implications of SAT exponential information size

- ④ Kolmogorov information complexity $K(x)$ for strings x is not a measure in mathematical sense because it does not satisfy countable additivity/ σ -additivity principle:
should be:

$$K(\cup(SAT_subsets)) = \sum_{SAT_subsets} K(S)$$

is:

$$K(\cup(SAT_subsets)) = O(1) < \sum_{SAT_subsets} K(S) \approx \Omega(2^{|I|})$$

Implications of SAT exponential information size

- ⑤ Implications for problem solvability and requirements on algorithm information and time complexity:
 - information conservation principle (cf. unitarity in physics)
 - how much information must algorithms have to solve combinatorial problems (e.g. SAT)?
 - how much time must fixed-code algorithms have to solve combinatorial problems (e.g. SAT)?
 - where is this information coming from?
 - can ML algorithms have fair advantage due to their learning stage?

Conclusions

- SAT as a string relation has $\Omega(2^{|I|})$ and $O(|I|2^{|I|})$ information size
- SAT can be solved in polynomial time given exponential amount of information
- Kolmogorov complexity is not a measure
- consequences for ML and problem solvability

Thank you for listening

more details on SAT information size at <https://arxiv.org/abs/2401.00947>