KOLMOGOROV COMPLEXITY

1.4.2019

MARTIN RAŠKA

INTRODUCTION

HOW TO MEASURE COMPLEXITY OF OBJECTS ?

OBJECTS $\rightsquigarrow$ STRINGS $\rightsquigarrow$ $\{0,1\}^*$

S: 010101010101010101010101

T: 110100010111010010011

INTRODUCTION

HOW TO MEASURE COMPLEXITY ?

- BY THE LENGTH OF THE SHORTEST DESCRIPTION.

S = "01" 10-TIMES

T = ?

COMPLEXITY ≈ COMPRESSIBILITY

INTRODUCTION

DOES THERE EXIST AN OPTIMAL DESCRIPTION LANGUAGE?

CONSIDER DESCRIPTION LANGUAGE U WHERE STRINGS ARE DESCRIBED BY (SUITABLE CODING OF) TURING MACHINES WHICH COMPUTES THEM.

→ DECODING IS COMPUTED BY UNIVERSAL TURING MACHINE

INTRODUCTION

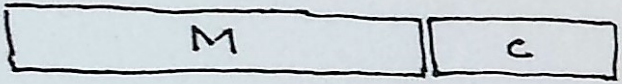
OPTIMALITY OF U:

CONSIDER ANY OTHER LANGUAGE D

DECODING FUNCTION FOR D IS COMPUTABLE BY TURING MACHINE M

FOR STRING S, LET C BE THE SHORTEST DESCRIPTION OF S IN D.

THEN WE OBTAIN DESCRIPTION OF S IN U



INTRODUCTION

DENOTING RESPECTIVE KOLMOGOROV COMPLEXITIES $K_U(s)$, $K_D(s)$ FOR U, D

WE OBTAIN

$$K_U(s) \leq |M| + |c| = |M| + K_D(s)$$

INTRODUCTION

$\rightsquigarrow$ INVARIANT THEOREM:

TWO TURING COMPLETE DESCRIPTION LANGUAGES U, V ARE EQUALLY EFFICIENT

I.E. $\exists\, m \in \mathbb{N}$:

$$\forall\, s \in \{0,1\}^*: \; |K_U(s) - K_V(s)| \leq m$$

INTRODUCTION

CALL STRING s

- COMPRESSIBLE, IF

$$K(s) \ll |s|$$

- INCOMPRESSIBLE, OTHERWISE

INTRODUCTION

HOW MANY (IN)COMPRESSIBLE STRINGS THERE ARE ?

MORE CONCRETELY: HOW MANY OF STRINGS OF LENGTH n ARE COMPRESSIBLE BY 4 ?

- AT MOST AS MANY AS PROGRAMS OF LENGTH $\leq n-4$

$$\leq 1 + 2 + 2^2 + \ldots + 2^{n-4} = 2^{n-3} - 1$$

INTRODUCTION

AT MOST 2^{m-3} OUT OF 2^m STRINGS OF LENGTH m ARE COMPRESSIBLE BY 4

→ PROBABILITY THAT RANDOMLY CHOOSEN STRING IS COMPRESSIBLE BY 4 IS

$$\leq \frac{2^{m-3}}{2^m} = 2^{-3} = \frac{1}{8}$$

→ MOST OF THE STRINGS ARE COMPLEX

HOW TO DETERMINE KOLMOGOROV COMPLEXITY K(s) FOR PARTICULAR STRING?

INTRODUCTION

UPPER BOUNDS FOR $K(s)$:

SUFFICES TO FIND A PROGRAM P
THAT OUTPUTS s

$$\rightsquigarrow K(s) \leq |P|$$

TRIVIAL APPROACH:

PRINT "LITERAL NOTATION OF s"

$$\rightsquigarrow K(s) \leq |s| + 8$$

LOWER BOUNDS FOR $K(s)$:

NAIVE APPROACH:

GO THROUGH ALL PROGRAMS
OF LENGTH $< m$
IF NO ONE OUTPUTS s, THEN

$$K(s) \geq m$$

$\rightsquigarrow$ HALTING PROBLEM

IT TURNS OUT

THAT WE ARE UNABLE TO DETERMINE $K(s)$

FOR ALMOST ALL STRINGS s

CHAITIN'S INCOMPLETENESS THEOREM

THERE EXISTS A CONSTANT L SUCH THAT FOR NO STRING OF BITS S WE CAN PROVE

$$K(S) > L .$$

CHAITIN'S INCOMPLETENESS THEOREM

WHAT DOES IT MEAN "TO PROVE $K(s) > L$"?

CHOOSE YOUR FAVOURITE FORMAL THEORY T STRONG ENOUGH TO TALK ABOUT KOLMOGOROV COMPLEXITY

→ ZFC, PEANO ARITHMETICS, ...

CHAITIN'S INCOMPLETENESS THEOREM

WHAT IS THEORY ?

- LANGUAGE = FUNCTION AND RELATION SYMBOLS
- SET OF AXIOMS (= FORMULAS) IN THAT LANGUAGE

PROOF OF FORMULA A IN THEORY T

= SEQUENCE OF FORMULAS

$$B_0, B_1, \ldots, B_m = A$$

SUCH THAT ...

CHAITIN'S INCOMPLETENESS THEOREM

IN ADITION, WE REQUIRE THAT WE CAN VERIFY IN SOME MECHANICAL WAY, WETHER GIVEN FORMULA IS AXIOM OF THE THEORY

CHAITIN'S INCOMPLETENESS THEOREM

WHAT DOES IT MEAN THAT

"THEORY T CAN TALK ABOUT KOLMOGOROV COMPLEXITY" ?

- FOR GIVEN STRING s AND $m \in \mathbb{N}$, WE CAN EXPRESS THE CONDITION "$K(s) > m$" BY FORMULA $A_{s,m}$ SO THAT

  IF T PROVES $A_{s,m}$, THEN $K(s) > m$

- WE ALSO REQUIRE FORMULAS $A_{s,m}$ TO BE RECOGNIZABLE IN MECHANICAL WAY

CHAITIN'S INCOMPLETENESS THEOREM

BACK TO THE THEOREM ...

THERE EXISTS A CONSTANT L SUCH THAT FOR NO STRING S WE CAN PROVE

$$K(s) > L .$$

- WE FIXED THEORY T
- WE HAVE FORMULAS $A_{s,m}$ SAYING "$K(s) > m$"

CHAITIN'S INCOMPLETENESS THEOREM

IDEA OF THE PROOF:

FIRST NOTE THE FOLLOWING...

IF P IS A PROGRAM WHICH FOR INPUT $m \in \mathbb{N}$ OUTPUTS STRING s_m, THAN

$$K(s_m) \leq |P| + \log_2 m$$

CHAITIN'S INCOMPLETENESS THEOREM

IDEA OF THE PROOF :

CONSIDER A PROGRAM P WHICH FOR INPUT $m \in \mathbb{N}$ WILL GO THROUGH ALL PROOFS IN THE THEORY T LOOKING FOR A PROOF OF $A_{s,m}$ FOR SOME s.

IF SUCH PROOF IS FOUND, PROGRAM OUTPUTS THE STRING s.

CHAITIN'S INCOMPLETENESS THEOREM

RUNNING THE PROGRAM P ON INPUT m,
THERE ARE TWO CASES:

- THERE IS NO PROOF OF ANY FORMULA $A_{s,m}$, SO THE PROGRAM WILL RUN FOREVER
- OR, THERE IS A STRING s FOR WHICH THE PROOF OF $A_{s,m}$ EXISTS; THEN P OUTPUTS SUCH STRING s

IN THE SECOND CASE, WE HAVE $K(s) \leq |P| + \log_2 m$,
BUT ALSO $K(s) > m$, BECAUSE T PROVES IT.

CHAITIN'S INCOMPLETENESS THEOREM

PUTING THESE INEQUALITIES TOGETHER, WE OBTAIN

$$n < K(s) \leq |P| + \log_2 n$$

$$n < |P| + \log_2 n$$

IF WE CHOOSE CONSTANT L LARGE ENOUGH SO THAT

$$L \geq |P| + \log_2 L,$$

THE SECOND CASE CAN'T OCCUR,

CHAITIN'S INCOMPLETENESS THEOREM

THAT IS, FOR SUCH L PROGRAM WILL NEVER STOP, BECAUSE THERE IS NO PROOF OF $A_{s,L}$ FOR ANY STRING s

IN OTHER WORDS, WE CAN'T PROVE

$$K(s) > L$$

FOR ANY STRING s.

CHAITIN'S INCOMPLETENESS THEOREM

REMAINS TO SHOW THAT SUCH PROGRAM P REALLY EXISTS ...

```
PROGRAM P (m : INTEGER);
  FOR i := 1 TO INFINITY DO
    FOR EACH STRING α OF LENGTH i DO
      IF α IS PROOF OF FORMULA A_{s,m} THEN
        RETURN s;
```

WE ARE ABLE TO VERIFY BY COMPUTER WHETHER GIVEN STRING IS FORMULA, AXIOM OF T, PROOF IN T, FORMULA $A_{s,m}$ AND FINALLY PROOF OF $A_{s,m}$ IN T

CHAITIN'S INCOMPLETENESS THEOREM

(COUNTER)EXAMPLE :

CONSIDER THE THEORY

$$T' := ZFC + \{ A_{s,m} ; K(s) > m \}$$

FOR ALL L, THERE EXISTS STRING s SUCH THAT $K(s) > L$, THEREFORE $A_{s,L} \in T'$, IN PARTICULAR $A_{s,L}$ IS PROVABLE IN T'

IS IT AGAINST THE STATEMENT OF THE THEOREM?