Complexité de Kolmogorov

Jacques-Henri Jourdan

20 janvier 2009

Introduction

Introduction

- But : trouver une notion d'information contenue dans une chaîne (binaire)
- 0000000000000000 contient moins d'information que 1011000101100111
- Idée : utiliser la longueur d'un programme informatique générant cette chaîne.

Définition

Définition : complexité de Kolmogorov relativement à une machine de Turing

Soit $\mathcal{M}$ une machine de Turing, et w un mot binaire.

- $\mathcal{K}_{\mathcal{M}}(w)$: longueur du plus petit mot binaire u tel que $\mathcal{M}(u) = w$
- si un tel mot u n'existe pas, $\mathcal{K}_{\mathcal{M}}(w) = +\infty$

Exemples

- Si $\mathcal{M}$ est un interpréteur de programme OCaml codé en ASCII, la complexité de la chaîne "Hello world !" est plus faible que la longueur du code ASCII du programme :
`let _ = print_string "Hello world !"`.
- Si $\mathcal{M}$ est la machine de Turing qui écrit "Hello world !", la complexité de cette chaîne devient nulle mais toutes les autres chaînes ont une complexité infinie.

Théorème d'invariance

Théorème d'invariance

Il existe $\mathcal{M}_0$ une machine de Turing, appelée *machine de Turing de référence*, telle que pour toute machine de Turing $\mathcal{M}$, il existe une constante $c_{\mathcal{M}}$ telle que :

$$\forall w\ \mathcal{K}_{\mathcal{M}}(w) \leq \mathcal{K}_{\mathcal{M}_0}(w) + c_{\mathcal{M}}$$

Preuve : si on choisit pour $\mathcal{M}_0$ un interpréteur OCaml, on peut passer à $\mathcal{M}_0$ le code de $\mathcal{M}$ en Cal, suivi du mot qu'il faut donner à $\mathcal{M}$ pour obtenir w.

Définition invariante

Le théorème permet de définir la notion de complexité de Kolmogorov de façon intrinsèque :

Définition : complexité de Kolmogorov

On se fixe $\mathcal{M}_0$ une machine de Turing de référence et on définit $\mathcal{K}(w) = \mathcal{K}_{\mathcal{M}_0}(w)$ la complexité de Kolmogorov de w.

Propriétés élémentaires

Propriétés élémentaires

1. $\mathcal{K}(u) \leqslant |u| + O(1)$
2. $\mathcal{K}(uu) \leqslant \mathcal{K}(u) + O(1)$
3. $\mathcal{K}(uv) \leqslant \mathcal{K}(u) + \mathcal{K}(v) + O(\log \mathcal{K}(u))$

Preuve : avec le théorème d'invariance.

Calculer la complexité de Kolmogorov

Incalculabilité

- La complexité de Kolmogorov n'est pas calculable
- Idée de la preuve : sinon, on pourrait formaliser l'idée de la phrase «Soit n le plus petit entier naturel impossible à décrire en moins de 42 mots français».

Incalculabilité des minorant

- Facile de majorer la complexité de Kolmogorov.
- Difficile, en revanche, de la minorer.
- Pourtant, elle tend vers $+\infty$ lorsque $|w|$ tend vers $+\infty$.
- On ne peut pas minorer la complexité de Kolmogorov par une fonction calculable non bornée.
- Idée de la preuve : identique à l'incalculabilité.

Passage à la limite

Il existe une machine de Turing $\mathcal{M}'$ qui prend un mot et un entier t telle que, pour tout mot u, on a :

$$\lim_{t \to +\infty} \mathcal{M}'(u, t) = \mathcal{K}(u)$$

Application aux langages rationnels

Définition préliminaire

Définition : mot caractéristique d'un langage

Soit L un langage. Soit ϕ une énumération calculable de Σ^*. Le *mot caractéristique* de L est le mot infini $\chi_L = \chi_L^1 \chi_L^2 \cdots$ défini par :

$$\chi_L^i = \begin{cases} 1 & \text{si } \phi(i) \in L \\ 0 & \text{sinon} \end{cases}$$

On note $\overline{n}$ le codage de n en binaire.

KC-caractérisation des langages rationnels

Soit L un langage. Les propositions suivantes sont équivalentes :

1. L est rationnel
2. Il existe une constante c telle que :

$$\forall x \in \Sigma^* \quad \forall n \in \mathbb{N} \quad \mathcal{K}(\chi_{x^{-1}L}[1..n]) \leq \mathcal{K}(\overline{n}) + c$$

3. Il existe une constante c telle que :

$$\forall x \in \Sigma^* \quad \forall n \in \mathbb{N} \quad \mathcal{K}(\chi_{x^{-1}L}[1..n]) \leq \log_2 n + c$$

C'est une sorte de lemme de pompage !

Conclusion

- Notion de complexité intrinsèque d'une chaîne
- Notion abstraite, car difficile (et inutile) à calculer en pratique
- Pas inutile pour autant (le théorème de KC-caractérisation des langages rationnels est effectivement plus puissant qu'un simple lemme de pompage)