Relating description complexity to entropy

Reijo Jaakkola, Antti Kuusisto, Miikka Vilander
University of Tampere, Finland

Description complexity

Description complexity of a property: the **minimum length** of a formula defining the property.

Description complexity

Description complexity of a property: the **minimum length** of a formula defining the property.

Depends on the logic studied.

Logics

Modal logic with the universal modality:

$$p \mid \neg\varphi \mid \varphi \wedge \psi \mid \blacklozenge\varphi$$

Logics

Modal logic with the universal modality:

$$p \mid \neg\varphi \mid \varphi \wedge \psi \;\mid \blacklozenge\varphi$$

$$\mathfrak{M}, w \models \blacklozenge\varphi \quad \Leftrightarrow \quad \mathfrak{M}, u \models \varphi \text{ for some } u \in \mathit{domain}(\mathfrak{M})$$

Logics

Let $\tau = \{p_1, \ldots, p_k\}$ be a finite vocabulary consisting of proposition symbols.

A τ-model $\mathfrak{M}$ is a structure (W, V) where

1. W is a finite, non-empty **domain**,
2. $V : \tau \to \mathcal{P}(W)$ is a **valuation function**.

A pointed model is a pair $(\mathfrak{M}, w)$ with $w \in W$.

We study the setting with a finite "universe" $\mathcal{U}$ consisting of all τ-models with the finite domain $\{1, \ldots, n\}$ of size n.

Logics

$$
\begin{array}{lcl}
\mathfrak{M}, w \models p & \Leftrightarrow & w \in V(p) \\
\mathfrak{M}, w \models \neg\varphi & \Leftrightarrow & \mathfrak{M}, w \not\models \varphi \\
\mathfrak{M}, w \models \varphi \wedge \psi & \Leftrightarrow & \mathfrak{M}, w \models \varphi \text{ and } \mathfrak{M}, w \models \psi \\
\mathfrak{M}, w \models \blacklozenge\varphi & \Leftrightarrow & \mathfrak{M}, v \models \varphi \text{ for some } v \in W
\end{array}
$$

Logics

$$
\begin{array}{lcl}
\mathfrak{M}, w \models p & \Leftrightarrow & w \in V(p) \\
\mathfrak{M}, w \models \neg\varphi & \Leftrightarrow & \mathfrak{M}, w \not\models \varphi \\
\mathfrak{M}, w \models \varphi \wedge \psi & \Leftrightarrow & \mathfrak{M}, w \models \varphi \text{ and } \mathfrak{M}, w \models \psi \\
\mathfrak{M}, w \models \blacklozenge\varphi & \Leftrightarrow & \mathfrak{M}, v \models \varphi \text{ for some } v \in W
\end{array}
$$

This logic is **expressively complete** for defining **sets of propositional assignments**.

Logics

Graded modal logic with universal modality:

$$p \mid \neg\varphi \mid \varphi \wedge \psi \ \mid \blacklozenge^{\geq d}\varphi$$

Logics

Graded modal logic with universal modality:

$$p \mid \neg\varphi \mid \varphi \wedge \psi \;\mid \blacklozenge^{\geq d}\varphi$$

$$\mathfrak{M}, w \models \blacklozenge^{\geq d}\varphi \quad \Leftrightarrow \quad \begin{array}{l} \mathfrak{M}, u \models \varphi \text{ for at least } d \text{ elements} \\ u \in \mathit{domain}(\mathfrak{M}) \end{array}$$

This logic is **expressively complete** for defining **multisets of propositional assignments**.

GMLU: graded modal logic with universal modality
MLU: modal logic with universal modality (so $d = 1$)

Formulas in negation normal form.

Formula size:

- $\mathit{size}(\alpha) = 1$ for a literal α,
- $\mathit{size}(\varphi \vee \psi) = \mathit{size}(\varphi \wedge \psi) = \mathit{size}(\varphi) + \mathit{size}(\psi) + 1$,
- $\mathit{size}(\blacklozenge^{\geq d}\varphi) = \mathit{size}(\blacksquare^{< d}\varphi) = \mathit{size}(\varphi) + d$.

$\blacksquare^{< d}$ is the dual of $\blacklozenge^{\geq d}$ equivalent to $\neg \blacklozenge^{\geq d} \neg$

Entropy

Entropy:

- A family of notions relating to randomness.
- The notions come from thermodynamics, statistical mechanics and information theory.

Shannon entropy for a distribution $p : X \to [0,1]$ is

$$-\sum_{y \in X} p(y) \log p(y)$$

Entropy

Shannon entropy: $H_S(\equiv) := -\sum_{i \in I} p(M_i) \log p(M_i)$

- $\equiv$ is logic-based equivalence relation over the model class $\mathcal{U}$.
- $M_i \subseteq \mathcal{U}$ with $i \in I$ are the equivalence classes.
- $p(M_i) = \dfrac{|M_i|}{|\mathcal{U}|}$

Entropy

Boltzmann entropy: $k_B \ln \Omega$

- Ω is a set of microstates.
- k_B is the Boltzmann constant.

Entropy

Boltzmann entropy: $H_B(M_i) := \log |M_i|$

where M_i is an equivalence class

Entropy

Boltzmann entropy: $H_B(M_i) := \log |M_i|$

where M_i is an equivalence class

Suppose φ defines M_i w.r.t. $\mathcal{U}$.

Intuitively, each $\mathfrak{M} \in M_i$ is a microstate realizing the macrostate φ

Entropy

Proposition. $H_S(\equiv) + \langle H_B \rangle \; = \; \log(|\mathcal{U}|)$

MLU

Theorem. In MLU, among the equivalence classes of $\equiv$, the largest equivalence class M_i has maximum description complexity (i.e., requires a formula of maximum length).

Corollary. In MLU, the equivalence class with maximum Boltzmann entropy has maximum description complexity.

MLU

Theorem. In MLU, among the equivalence classes of $\equiv$, the largest equivalence class M_i has maximum description complexity (i.e., requires a formula of maximum length).

Corollary. In MLU, the equivalence class with maximum Boltzmann entropy has maximum description complexity.

- Holds for sufficiently large models.
- The largest class is the one realizing all types. (Recall types are maximally consistent conjunctions of literals.)

MLU

The proof uses **formula length games**.

- Game position: $(\mathcal{A}, \mathcal{B}, r)$ where
  - $\mathcal{A}$ and $\mathcal{B}$ classes of pointed models $(\mathfrak{M}, w)$.
  - $r \in \mathbb{N}$
- Disjunction move
  1. Samson chooses $\mathcal{A}_0, \mathcal{A}_1$ such that $\mathcal{A}_0 \cup \mathcal{A}_1 = \mathcal{A}$, and Samson also chooses s, t such that $s + t = r$.
  2. Delilah chooses the next position which is either
     - $(\mathcal{A}_0, \mathcal{B}, s)$ or
     - $(\mathcal{A}_1, \mathcal{B}, t)$
- Diamond move
  1. Samson modifies each $(\mathfrak{M}, w) \in \mathcal{A}$ to some $(\mathfrak{M}, w')$
  2. The game continues from $(\mathcal{A}', \mathcal{B}', r - 1)$ where
     - $\mathcal{A}'$ contains the models $(\mathfrak{M}, w')$
     - $\mathfrak{B}'$ contains all models $(\mathfrak{N}, v')$ obtainable by modifying models $(\mathfrak{N}, v) \in \mathcal{B}$.

MLU

- Literal move:
  - Samson chooses a literal α, and the game ends.
  - Samson wins if $\mathcal{A} \models \alpha$ and $\mathcal{B} \models \neg\alpha$.

MLU

- Literal move:
  - Samson chooses a literal α, and the game ends.
  - Samson wins if $\mathcal{A} \models \alpha$ and $\mathcal{B} \models \neg\alpha$.

Theorem: Samson has a winning strategy for $(\mathcal{A}, \mathcal{B}, r)$ iff there is a formula φ of size at most r such that $\mathcal{A} \models \varphi$ and $\mathcal{B} \models \neg\varphi$.

MLU

To show the class with models realizing all types requires a maximum length formula, we let

- $\mathcal{A} = \{(\mathfrak{M}, w)\}$ where $\mathfrak{M}$ realizes all types, and
- $\mathcal{B}$ is a class with all the models omitting precisely one type.

GMLU

Theorem: For GMLU, we have $\langle H_B \rangle \sim |\tau| \langle C \rangle$

- $\langle H_B \rangle$ is expected Boltzmann entropy.
- τ is the propositional vocabulary considered.
- $\langle C \rangle$ is expected description complexity.

GMLU

Ingredients of the proof:

- We prove $\langle H_B \rangle \sim |\tau| n$ by a calculation utilizing, inter alia, the Stirling approximation, the weak law or large numbers and further sophisticated estimates.
- We show $\langle C \rangle \sim n$ by a formula-size game for GMLU.

First-order logic over a general relational vocabulary

We show that the expected description complexity of $\equiv_{\mathrm{FO}}$ grows asymptotically faster than its expected Boltzmann entropy.

- Show description complexity of an isomorphism class is $\Omega(\frac{n^m}{\log(n)})$ with high probability.
- Use this to get an estimate for expected description complexity, and compare this to an estimate for expected Boltzmann entropy.

Thanks