On Algorithmic Statistics for space-bounded algorithms

Alexey Milovanov

National Research University Higher School of Economics, almas239@gmail.com

10 June 2017, Kazan

Algorithmic statistics: motivation

Algorithmic statistics: motivation

- Let x be a binary string (experimental data).

Algorithmic statistics: motivation

- Let x be a binary string (experimental data).
- Our goal is to find $A \ni x$ as a suitable explanation for x.

Algorithmic statistics: motivation

- Let x be a binary string (experimental data).
- Our goal is to find $A \ni x$ as a suitable explanation for x.

Example

Let $x = \overbrace{000\ldots00}^{n \text{ zeros}}$. Then $\{x\}$ is a suitable explanation for x but $\{0,1\}^n$ is not a good explanation.

Algorithmic statistics: motivation

- Let x be a binary string (experimental data).
- Our goal is to find $A \ni x$ as a suitable explanation for x.

Example

Let $x = \overbrace{000\ldots00}^{n \text{ zeros}}$. Then $\{x\}$ is a suitable explanation for x but $\{0,1\}^n$ is not a good explanation.

Example

Let $x = \overbrace{01001011\ldots010}$ be a random string of length n i.e. its *Kolmogorov complexity* $\mathrm{C}(x)$ is equal to n. Then $\{0,1\}^n$ is an reasonable explanation for x, however $\{x\}$ is not adequate.

Randomness deficiency

A set $A \ni x$ is a good explanation for x if

Randomness deficiency

A set $A \ni x$ is a good explanation for x if

- A is *simple*, i.e. $\mathrm{C}(A) \approx 0$;

Randomness deficiency

A set $A \ni x$ is a good explanation for x if

- A is *simple*, i.e. $\mathrm{C}(A) \approx 0$;
- x is *typical* element of A.

Randomness deficiency

A set $A \ni x$ is a good explanation for x if

- A is *simple*, i.e. $\mathrm{C}(A) \approx 0$;
- x is *typical* element of A.

By Kolmogorov x is typical in A if *randomness deficiency* $d(x|A) := \log|A| - \mathrm{C}(x|A)$ is small.

Randomness deficiency

A set $A \ni x$ is a good explanation for x if

- A is *simple*, i.e. $\mathrm{C}(A) \approx 0$;
- x is *typical* element of A.

By Kolmogorov x is typical in A if *randomness deficiency* $d(x|A) := \log|A| - \mathrm{C}(x|A)$ is small. Note that

- $d(x|A) \gtrsim 0$ for every x in A.

Randomness deficiency

A set $A \ni x$ is a good explanation for x if

- A is *simple*, i.e. $\mathrm{C}(A) \approx 0$;
- x is *typical* element of A.

By Kolmogorov x is typical in A if *randomness deficiency* $d(x|A) := \log|A| - \mathrm{C}(x|A)$ is small. Note that

- $d(x|A) \gtrsim 0$ for every x in A.
- The fraction of elements x in A such that $d(x|A) > k$ is less than 2^{-k}.

Optimality deficiency

- Can this theory be used in practice?

Optimality deficiency

- Can this theory be used in practice?
- Kolmogorov complexity is uncountable function.

Optimality deficiency

- Can this theory be used in practice?
- Kolmogorov complexity is uncountable function.
- We can get an upper bound of $\mathrm{C}()$ but we can not prove a lower bound of it.

Optimality deficiency

- Can this theory be used in practice?
- Kolmogorov complexity is uncountable function.
- We can get an upper bound of $\mathrm{C}()$ but we can not prove a lower bound of it.
- So, we can argue that $A \ni x$ is simple, but we can not prove that $d(x|A) = \log|A| - \mathrm{C}(x|A)$ is small.

Optimality deficiency

- Can this theory be used in practice?
- Kolmogorov complexity is uncountable function.
- We can get an upper bound of $\mathrm{C}()$ but we can not prove a lower bound of it.
- So, we can argue that $A \ni x$ is simple, but we can not prove that $d(x|A) = \log|A| - \mathrm{C}(x|A)$ is small.
- Consider another parameter instead of $d(x|A)$: just $\log|A|$.

Optimality deficiency

- Can this theory be used in practice?
- Kolmogorov complexity is uncountable function.
- We can get an upper bound of $\mathrm{C}()$ but we can not prove a lower bound of it.
- So, we can argue that $A \ni x$ is simple, but we can not prove that $d(x|A) = \log|A| - \mathrm{C}(x|A)$ is small.
- Consider another parameter instead of $d(x|A)$: just $\log|A|$.
- What can we say about $\mathrm{C}(A) + \log|A|$ for $A \ni x$?

Optimality deficiency

- Can this theory be used in practice?
- Kolmogorov complexity is uncountable function.
- We can get an upper bound of $\mathrm{C}()$ but we can not prove a lower bound of it.
- So, we can argue that $A \ni x$ is simple, but we can not prove that $d(x|A) = \log|A| - \mathrm{C}(x|A)$ is small.
- Consider another parameter instead of $d(x|A)$: just $\log|A|$.
- What can we say about $\mathrm{C}(A) + \log|A|$ for $A \ni x$?
- $\mathrm{C}(A) + \log|A| \gtrsim \mathrm{C}(x)$.

Optimality deficiency

- Can this theory be used in practice?
- Kolmogorov complexity is uncountable function.
- We can get an upper bound of $\mathrm{C}()$ but we can not prove a lower bound of it.
- So, we can argue that $A \ni x$ is simple, but we can not prove that $d(x|A) = \log|A| - \mathrm{C}(x|A)$ is small.
- Consider another parameter instead of $d(x|A)$: just $\log|A|$.
- What can we say about $\mathrm{C}(A) + \log|A|$ for $A \ni x$?
- $\mathrm{C}(A) + \log|A| \gtrsim \mathrm{C}(x)$.
- The difference $\delta(x, A) := \mathrm{C}(A) + \log|A| - \mathrm{C}(x)$ is called *optimality deficiency*.

The connection between randomness and optimality deficiencies

The connection between randomness and optimality deficiencies

- $d(x|A) := \log|A| - \mathrm{C}(x|A)$, $\delta(x, A) := \mathrm{C}(A) + \log|A| - \mathrm{C}(x)$.

The connection between randomness and optimality deficiencies

- $d(x|A) := \log|A| - \mathrm{C}(x|A)$, $\delta(x,A) := \mathrm{C}(A) + \log|A| - \mathrm{C}(x)$.
- $\delta(x,A) \lesssim d(x|A)$ because $\mathrm{C}(x) \lesssim \mathrm{C}(A) + \mathrm{C}(x|A)$.

The connection between randomness and optimality deficiencies

- $d(x|A) := \log|A| - \mathrm{C}(x|A)$, $\delta(x,A) := \mathrm{C}(A) + \log|A| - \mathrm{C}(x)$.
- $\delta(x,A) \lesssim d(x|A)$ because $\mathrm{C}(x) \lesssim \mathrm{C}(A) + \mathrm{C}(x|A)$.
- The difference can be large.

Example

Let x be random string of length n (i.e. $\mathrm{C}(x) \approx n$). Let y another independent of x random string of length n. Consider $A := \{0,1\}^n \setminus \{y\}$. Then $d(x|A) \approx 0$ however $\delta(x,A) \approx n$.

The connection between randomness and optimality deficiencies

- $d(x|A) := \log|A| - \mathrm{C}(x|A)$, $\delta(x, A) := \mathrm{C}(A) + \log|A| - \mathrm{C}(x)$.
- $\delta(x, A) \lesssim d(x|A)$ because $\mathrm{C}(x) \lesssim \mathrm{C}(A) + \mathrm{C}(x|A)$.
- The difference can be large.

Example

Let x be random string of length n (i.e. $\mathrm{C}(x) \approx n$). Let y another independent of x random string of length n. Consider $A := \{0,1\}^n \setminus \{y\}$. Then $d(x|A) \approx 0$ however $\delta(x, A) \approx n$.

However, the following is true.

The connection between randomness and optimality deficiencies

- $d(x|A) := \log|A| - \mathrm{C}(x|A)$, $\delta(x, A) := \mathrm{C}(A) + \log|A| - \mathrm{C}(x)$.
- $\delta(x, A) \lesssim d(x|A)$ because $\mathrm{C}(x) \lesssim \mathrm{C}(A) + \mathrm{C}(x|A)$.
- The difference can be large.

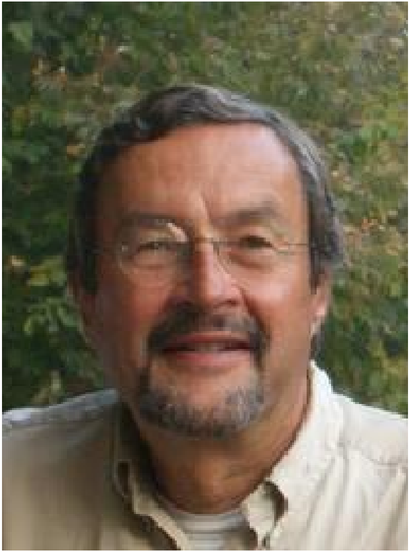
Example

Let x be random string of length n (i.e. $\mathrm{C}(x) \approx n$). Let y another independent of x random string of length n. Consider $A := \{0,1\}^n \setminus \{y\}$. Then $d(x|A) \approx 0$ however $\delta(x, A) \approx n$.

However, the following is true.

Theorem (Vereshchagin, Vitányi)

For every string x and for every $A \ni x$ there exists $B \ni x$ such that $\mathrm{C}(B) \lesssim \mathrm{C}(A)$ and $\delta(x, B) \lesssim d(x|A)$.



Descriptions of Restricted Type

Descriptions of Restricted Type

- So far we considered arbitrary finite sets as models.

Descriptions of Restricted Type

- So far we considered arbitrary finite sets as models.
- However, in practice we usually have some a priori information about the data.

Descriptions of Restricted Type

- So far we considered arbitrary finite sets as models.
- However, in practice we usually have some a priori information about the data.
- Assume that "right" model A belongs to some enumerable family of sets $\mathcal{A}$. (For example, $\mathcal{A}$ is the family of all Hamming balls.)

Descriptions of Restricted Type

- So far we considered arbitrary finite sets as models.
- However, in practice we usually have some a priori information about the data.
- Assume that "right" model A belongs to some enumerable family of sets $\mathcal{A}$. (For example, $\mathcal{A}$ is the family of all Hamming balls.)

It turns out that the previous result holds also for this case.

Descriptions of Restricted Type

- So far we considered arbitrary finite sets as models.
- However, in practice we usually have some a priori information about the data.
- Assume that "right" model A belongs to some enumerable family of sets $\mathcal{A}$. (For example, $\mathcal{A}$ is the family of all Hamming balls.)

It turns out that the previous result holds also for this case.

Theorem (Vereshchagin, Vitányi)

For every string x and for every $A \ni x$ from any enumerable family $\mathcal{A}$ there exists $B \in \mathcal{A}$ containing x such that $\mathrm{C}(B) \lesssim \mathrm{C}(A)$ and $\delta(x, B) \lesssim d(x|A)$.

Space-bounded algorithmic statistics

Space-bounded algorithmic statistics

- The notion of Kolmogorov complexity has the following problem.

Space-bounded algorithmic statistics

- The notion of Kolmogorov complexity has the following problem.
- It ignores time and space needed to produce x from its short description.

- The notion of Kolmogorov complexity has the following problem.
- It ignores time and space needed to produce x from its short description.
- We will consider algorithms whose space (not time) is bounded by a polynomial of the length of a string.

Space-bounded algorithmic statistics

- The notion of Kolmogorov complexity has the following problem.
- It ignores time and space needed to produce x from its short description.
- We will consider algorithms whose space (not time) is bounded by a polynomial of the length of a string.

Definition

The complexity $\mathrm{CD}^m(A)$ of a set A with space bound m is defined as the minimal length of a program p such that

- $p(y) = 1$ if $y \in A$.
- $p(y) = 1$ if $y \notin A$.
- p uses at most m space on every input.

- The notion of Kolmogorov complexity has the following problem.
- It ignores time and space needed to produce x from its short description.
- We will consider algorithms whose space (not time) is bounded by a polynomial of the length of a string.

Definition

The complexity $\mathrm{CD}^m(A)$ of a set A with space bound m is defined as the minimal length of a program p such that

- $p(y) = 1$ if $y \in A$.
- $p(y) = 1$ if $y \notin A$.
- p uses at most m space on every input.

$\mathrm{CD}^m(x)$ is defined as $\mathrm{CD}^m(\{x\})$.

Main result

Main result

We proof an analogue of theorem of Vereshchagin and Vitány for polynomial space bound.

Main result

We proof an analogue of theorem of Vereshchagin and Vitány for polynomial space bound.

Definition

A family of sets $\mathcal{A}$ is called *polynomial-space enumerable* if there is an algorithm that enumerate all subset of $\{0,1\}^n$ from $\mathcal{A}$ in space $\text{poly}(n)$.

Main result

We proof an analogue of theorem of Vereshchagin and Vitány for polynomial space bound.

Definition

A family of sets $\mathcal{A}$ is called *polynomial-space enumerable* if there is an algorithm that enumerate all subset of $\{0,1\}^n$ from $\mathcal{A}$ in space poly(n).

Theorem (Informal)

Let x be a string of length n and let $\mathcal{A}$ be a polynomial-space enumerable family of sets. Then for every set $A \ni x$ from $\mathcal{A}$ there exists a set $B \ni x$ from $\mathcal{A}$ such that $\mathrm{CD}^{poly(n)}(B) \lesssim \mathrm{CD}^{poly(n)}(A)$ and $\delta^{poly(n)}(x, B) \lesssim d^{poly(n)}(x|A)$.

Proof idea

Proof idea

- Define probability distribution $\mathcal{B}$ as follows. Every set from $\mathcal{A}$ of complexity $\mathrm{CD}^{poly(n)}(A)$ belongs to $\mathcal{B}$ with probability $2^{\mathrm{CD}^{\mathsf{poly}(n)}(A|x)-\mathrm{CD}^{\mathsf{poly}(n)}(A)}$.

Proof idea

- Define probability distribution $\mathcal{B}$ as follows. Every set from $\mathcal{A}$ of complexity $\mathrm{CD}^{poly(n)}(A)$ belongs to $\mathcal{B}$ with probability $2^{\mathrm{CD}^{\mathsf{poly}(n)}(A|x)-\mathrm{CD}^{\mathsf{poly}(n)}(A)}$.
- This family $\mathcal{B}$ contains $B \ni x$ with high probability.

Proof idea

- Define probability distribution $\mathcal{B}$ as follows. Every set from $\mathcal{A}$ of complexity $\mathrm{CD}^{poly(n)}(A)$ belongs to $\mathcal{B}$ with probability $2^{\mathrm{CD}^{\mathrm{poly}(n)}(A|x)-\mathrm{CD}^{\mathrm{poly}(n)}(A)}$.
- This family $\mathcal{B}$ contains $B \ni x$ with high probability.
- If $\mathrm{CD}^{poly(n)}(\mathcal{B})$ is small then B satisfies the theorem.

Proof idea

- Define probability distribution $\mathcal{B}$ as follows. Every set from $\mathcal{A}$ of complexity $\mathrm{CD}^{poly(n)}(A)$ belongs to $\mathcal{B}$ with probability $2^{\mathrm{CD}^{\mathrm{poly}(n)}(A|x)-\mathrm{CD}^{\mathrm{poly}(n)}(A)}$.
- This family $\mathcal{B}$ contains $B \ni x$ with high probability.
- If $\mathrm{CD}^{poly(n)}(\mathcal{B})$ is small then B satisfies the theorem.
- We can find $\mathcal{B}$ by brute force.

Proof idea

- Define probability distribution $\mathcal{B}$ as follows. Every set from $\mathcal{A}$ of complexity $\mathrm{CD}^{poly(n)}(A)$ belongs to $\mathcal{B}$ with probability $2^{\mathrm{CD}^{\mathrm{poly}(n)}(A|x)-\mathrm{CD}^{\mathrm{poly}(n)}(A)}$.
- This family $\mathcal{B}$ contains $B \ni x$ with high probability.
- If $\mathrm{CD}^{poly(n)}(\mathcal{B})$ is small then B satisfies the theorem.
- We can find $\mathcal{B}$ by brute force.
- However this requires exponential space.

Proof idea

- Define probability distribution $\mathcal{B}$ as follows. Every set from $\mathcal{A}$ of complexity $\mathrm{CD}^{poly(n)}(A)$ belongs to $\mathcal{B}$ with probability $2^{\mathrm{CD}^{\mathrm{poly}(n)}(A|x)-\mathrm{CD}^{\mathrm{poly}(n)}(A)}$.
- This family $\mathcal{B}$ contains $B \ni x$ with high probability.
- If $\mathrm{CD}^{poly(n)}(\mathcal{B})$ is small then B satisfies the theorem.
- We can find $\mathcal{B}$ by brute force.
- However this requires exponential space.
- Nisan-Wigderson generator helps to reduce it.

Proof idea

- Define probability distribution $\mathcal{B}$ as follows. Every set from $\mathcal{A}$ of complexity $\mathrm{CD}^{poly(n)}(A)$ belongs to $\mathcal{B}$ with probability $2^{\mathrm{CD}^{\mathrm{poly}(n)}(A|x) - \mathrm{CD}^{\mathrm{poly}(n)}(A)}$.
- This family $\mathcal{B}$ contains $B \ni x$ with high probability.
- If $\mathrm{CD}^{poly(n)}(\mathcal{B})$ is small then B satisfies the theorem.
- We can find $\mathcal{B}$ by brute force.
- However this requires exponential space.
- Nisan-Wigderson generator helps to reduce it.
- The same idea was used by Daniil Musatov.

Thank you!