Matyáš Křišťan

kristja6@fit.cvut.cz

presented:

Relationship between supserstring and compression measures: New insights on the greedy conjecture

Bastien Cazaux, Eric Rivals

https://www.sciencedirect.com/science/article/pii/S0166218X1730183X

Definitions

- $||P|| = |w_1| + |w_2| + \cdots + |w_p|$, where $P = \{w_1, w_2, \ldots, w_p\}$ is a set of words.
- $S_A(P)$ is the output of algorithm A with input P.
- $S_{opt}(P)$ is the shortest possible superstring given input P.
- $super(A)$ is the approximation ratio of algorithm A. It is the smallest real value such that for any input P, the following holds:

  $1 \leq \frac{|S_A(P)|}{|S_{opt}(P)|} \leq super(A)$.
- $comp(A)$ is the compression ratio of algorithm A. It is the largest real value, such that for any input P satistying $||P|| \neq |S_{opt}(P)|$, the following holds:

  $0 \leq comp(A) \leq \frac{||P|| - |S_A(P)|}{||P|| - |S_{opt}(P)|}$

Problems

- **Shortest Superstring Problem (SSP)**
  - INPUT: A set of p words $P = \{s_1, s_2, \ldots, s_p\}$ over a finite alphabet Σ.
  - OUTPUT: The shortest string t containing each s_i for $1 \leq i \leq p$ as a substring.
- **r-Shortest Superstring Problem (r-SSP)**
  - INPUT: A set of p words $P = \{s_1, s_2, \ldots, s_p\}$ over a finite alphabet Σ, where $|s_i| = r$ for every $1 \leq i \leq p$.
  - OUTPUT: The shortest string t containing each s_i for $1 \leq i \leq p$ as a substring.

Theorems

- **Theorem 1**: Let P be a set of words satisfying $|S_{opt}(P)| \neq ||P||$. Let γ be a real such that $0 < \gamma \leq \frac{|S_{opt}(P)|}{||P||}$, and let A be an approximation algorithm for SSP. We have: $super(A) \leq \frac{(\gamma-1)comp(A)+1}{\gamma}$.
- **Proposition 1**: Let $P \subseteq \Sigma^r$ and $p = |P|$. Let t be a superstring of P. Then $|t| \geq r + p - 1$.
- **Theorem 2**: Let r be an integer such that $r > 1$ and let P be a subset of Σ^r. For any approximation algorithm A, we have: $\frac{|S_A(P)|}{|S_{opt}(P)|} \leq r - (r-1)comp(A)$.
- **Proposition 2**: GREEDY approximates r-SSP with a ratio of at least $2 - \frac{1}{r}$.
- **Theorem 3**: The superstring approximation ratio of GREEDY for r-SSP is bounded by:

  $2 - \frac{1}{r} \leq super(GREEDY) \leq min(\frac{r+1}{2}, \frac{7}{2})$.