Coláiste na hOllscoile Corcaigh, Éire
University College Cork, Ireland

CS4618 Artificial Intelligence I

Today: Analysing and Designing Mutation

Thomas Jansen

November 21st

Plans for Today

Thoughts about Applications

Input some nodes $V = \{v_1, v_2, \ldots, v_n\}$
costs for connecting two nodes $w_{i,j} \in \mathbb{R}^+$
$\widehat{=}$weighted graph (V, E, w)

Output connections such that
1. for any two nodes $\exists$path of connections between them
2. a cheapest set of such connections

$\widehat{=}$minimal spanning tree (MST)

Observation for MST efficient algorithm known
thus don't use RSH

Add additional restriction 'maximal degree k'
$\rightsquigarrow$ problem NP-hard
$\rightsquigarrow$ application of RSH is reasonable

RSH for the Degree-Restricted MST

What can we do to make better RSH for the problem?

Known $\binom{n}{2} = \frac{n(n-1)}{2} = \Theta(n^2)$ potential connections
optimal solutions contain $n-1 = \Theta(n)$ connections

Mutations standard bit mutations maintain the uniform distribution
tend towards the uniform distribution
$\rightsquigarrow$ tend towards solutions with $\approx \frac{1}{2} \cdot \frac{n(n-1)}{2}$ connections
$\widehat{=}$ looking 'in the wrong part of the search space'

Idea bias search towards promising region of the search space

Caution Know what you are doing!

Asymmetric Mutations

Standard Bit Mutations

1. For $i \in \{1, 2, \ldots, n\}$
2.   With probability $1/n$
3.   set $y[i] := 1 - x[i]$
4.   else set $y[i] := x[i]$

Asymmetric Mutations

1. For $i \in \{1, 2, \ldots, n\}$
2.   If $x[i] = 1$
3.   then $p_m = 1/(2\,|x|)$
4.   else $p_m = 1/(2(n - |x|))$.
5.   With probability p_m
6.   set $y[i] := 1 - x[i]$
7.   else set $y[i] := x[i]$

Observations

Standard Bit Mutations:

- mutation probability equal for all bits (in particular, independent of bit value)
- $p_m = 1/n$

Asymmetric Mutations:

- mutation probability depends on bit value, equal for all bits of the same value
- $1/(2n) \leq p_m \leq 1/2$

Properties of Asymmetric Mutations

Standard Bit Mutations

- $p_m = 1/n$
- $\mathrm{E}\left(|\mathsf{mutation}(x)| \mid x\right)$
  $= |x| \cdot \left(1 - \frac{1}{n}\right) + (n - |x|) \cdot \frac{1}{n}$
  $= |x| + 1 - \frac{2|x|}{n}$
- $\mathsf{H}\left(x, y\right) = 1 \Rightarrow$
  $\mathrm{Prob}\left(y = \mathsf{mutation}(x)\right)$
  $= \frac{1}{n}\left(1 - \frac{1}{n}\right)^{n-1} = \Theta\left(\frac{1}{n}\right)$
- $\mathrm{Prob}\left(\mathsf{H}\left(\mathsf{mutation}(x), x\right) = k\right)$
  $= \binom{n}{k}\left(\frac{1}{n}\right)^{k}\left(1 - \frac{1}{n}\right)^{n-k}$
  $\leq \frac{1}{k!}$
  $= \Theta\left(1/\left(\sqrt{k} \cdot (k/e)^{k}\right)\right)$

Asymmetric Mutations

- $1/(2n) \leq p_m \leq 1/2$
- $\mathrm{E}\left(|\mathsf{mutation}(x)| \mid x\right)$
  $= |x| \cdot \left(1 - \frac{1}{2|x|}\right) +$
  $(n - |x|) \cdot \frac{1}{2(n-|x|)} = |x|$
- $\mathsf{H}\left(x, y\right) = 1 \Rightarrow$
  $\mathrm{Prob}\left(y = \mathsf{mutation}(x)\right) =$
  $\frac{1}{2l}\left(1 - \frac{1}{2l}\right)^{l-1}\left(1 - \frac{1}{2(n-l)}\right)^{n-l}$
  $= \Theta\left(\frac{1}{l}\right)$
- $\mathrm{Prob}\left(\mathsf{H}\left(\mathsf{mutation}(x), x\right) \geq k\right)$
  $\leq \left(\frac{e^{k-1}}{k^k}\right) < (e/k)^k$

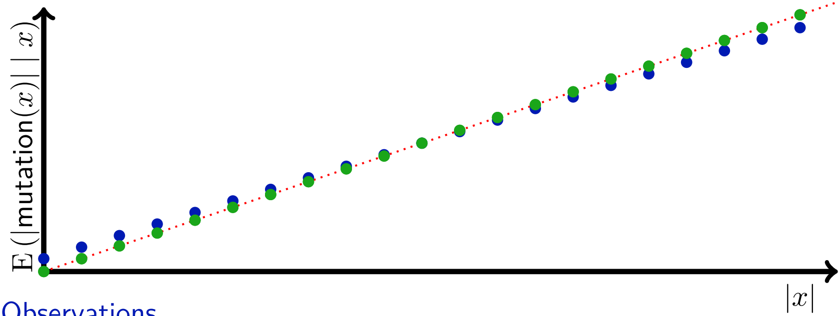
Expected Number of 1-Bits

Standard Bit Mutations

$\mathrm{E}\left(|\text{mutation}(x)| \mid x\right)$
$= |x| + 1 - \frac{2|x|}{n}$

Asymmetric Mutations

$\mathrm{E}\left(|\text{mutation}(x)| \mid x\right)$
$= |x|$

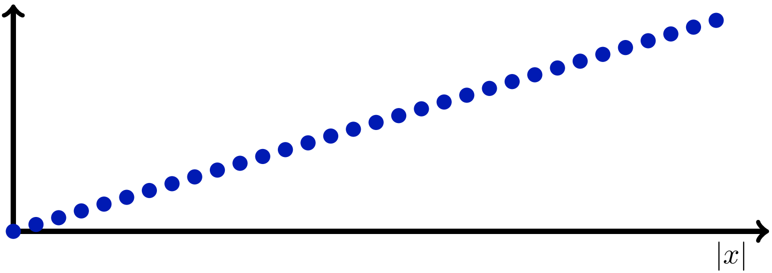


Observations

- standard bit mutations steer towards the middle of the search space
- asymmetric mutations tend not to leave the current level
- asymmetric mutations tend to spend more time at levels with few 0-bits or few 1-bits

OneMax

Remember $\textsc{OneMax}(x) = \sum_{i=1}^{n} x[i]$



Remember $\mathrm{E}\left(T_{\text{(1+1) EA},\textsc{OneMax}}\right) = \Theta(n \log n)$

Theorem

$\mathrm{E}\left(T_{\text{asym. (1+1) EA},\textsc{OneMax}}\right) = \Theta(n)$

Asymmetric Mutations on OneMax: Lower Bound

Claim $\mathrm{E}\left(T_{\text{asym. (1+1) EA},\textsc{OneMax}}\right) = \Omega(n)$

Technique drift analysis
distance measure $d\colon \{0,1\}^n \to \mathbb{R}_0^+$
expected initial distance $\mathrm{E}\left(d(x_0)\right)$
maximal expected drift $\Delta = \max \mathrm{E}\left(d(x_{t-1}) - d(x_t) \mid x_{t-1}\right)$
$\mathrm{E}\left(T\right) \geq \mathrm{E}\left(d(x_0)\right)/\Delta$

Observe $\mathrm{E}\left(d(x_0)\right) = n/2$

Observe $\mathrm{E}\left(\mathrm{H}\left(\text{mutation}(x), x\right)\right) = |x| \cdot \frac{1}{2|x|} + (n - |x|) \cdot \frac{1}{2(n-|x|)} = 1$

Thus $\Delta \leq 1$

Thus $\mathrm{E}\left(T_{\text{asym. (1+1) EA},\textsc{OneMax}}\right) \geq \frac{n/2}{1} = \frac{n}{2} = \Omega(n)$ ✓

Asymmetric Mutations on OneMax: Upper Bound

Claim $\mathrm{E}\left(T_{\text{asym. (1+1) EA},\textsc{OneMax}}\right) = O(n)$

Observation
$$\begin{aligned}
&\mathrm{E}\left(T_{\text{asym. (1+1) EA},\textsc{OneMax}}\right)\\
&\leq \sum_{h=1}^{n} \mathrm{E}\left(\text{time for } \mathrm{H}\left(x, 1^n\right) \text{ `} h \text{ to } h-1\text{'}\right)\\
&\leq \sum_{h=1}^{n} \left(\binom{h}{1}\left(\frac{1}{2h}\right)\left(1-\frac{1}{2h}\right)^{h-1}\left(1-\frac{1}{2(n-h)}\right)^{n-h}\right)^{-1}\\
&= 2\sum_{h=1}^{n} \left(\left(1-\frac{1}{2h}\right)^{(2h-2)/2}\left(1-\frac{1}{2(n-h)}\right)^{2(n-h)/2}\right)^{-1}\\
&\leq 2\sum_{h=1}^{n} \left(e^{-1/2}\cdot e^{-1/2}\right)^{-1} = 2en = O(n) ✓
\end{aligned}$$

□

Discussing the Result on OneMax

We have $\mathrm{E}\left(T_{\text{asym. (1+1) EA},\text{OneMax}}\right) = \Theta(n)$
speed-up $\Theta(\log n)$ in comparison to standard bit mutations

Observation global optimum of OneMaxis 1^n
in a region of the search space with few 0-bits
thus speed-up could be expected

Do we become much slower if optimum is somewhere else?

Remember $f^* = \{f_a \mid a \in \{0,1\}^n\}$
with $f_a(x) = f(x \oplus a)$ for $f\colon \{0,1\}^n \to \mathbb{R}$

What is $\mathrm{E}\left(T_{\text{asym. (1+1) EA},\text{OneMax}_a}\right)$ (depending on a)?

What is $\mathrm{E}\left(T_{\text{asym. (1+1) EA},\textsc{OneMax}_a}\right)$?

Remember $\forall f, a\colon \mathrm{E}\left(T_{\text{(1+1) EA},f_a}\right) = \mathrm{E}\left(T_{\text{(1+1) EA},f}\right)$

Observe $\forall f, a\colon \mathrm{E}\left(T_{\text{asym. (1+1) EA},f_a}\right) = \mathrm{E}\left(T_{\text{asym. (1+1) EA},f_{\overline{a}}}\right)$

Theorem

$\mathrm{E}\left(T_{\text{asym. (1+1) EA},\textsc{OneMax}_a}\right) = O(n \log(2 + |a|))$

Proof

Remember $\overline{a}$ is unique global opt. of $\textsc{OneMax}_a$

Idea partition run into two phases

1. from start until $\leq 2(n - |\overline{a}|)$ 0-bits, length T_1
2. from end of 1 until optimum found, length T_2

First Phase: $n - |x| > 2\,(n - |\overline{a}|)$

Idea '$n - |x| > 2\,(n - |\overline{a}|)$' $\widehat{=}$ far away
probability for an improvement large
rough estimate suffices

Observe $\mathsf{H}\,(x, \overline{a}) \geq (n - |\overline{x}|) - (n - |\overline{a}|)$
$\Rightarrow \mathsf{H}\,(x, \overline{a}) \geq n - |\overline{a}|$

Thus $\mathrm{E}\,(\text{`}\mathsf{H}\,(x, \overline{a})\text{'} \rightsquigarrow \text{`}\mathsf{H}\,(x, \overline{a}) - 1) \leq \left(\binom{n-\overline{a}}{1} \frac{1}{2(n-\overline{a})} e^{-1}\right)^{-1}$
$\leq 2e$

Thus $\mathrm{E}\,(T_1) \leq (n - 2\,(n - |\overline{a}|))\,2e \leq 2en = O(n)$ ✔

Second Phase: $n - |x| \leq 2\,(n - |\overline{a}|)$

We have

- $n - |x| \leq 2\,(n - |\overline{a}|)$
- $\mathsf{H}\,(x, 1^n) = n - |x| \;\leq 2\,(n - |\overline{a}|)$
- $\mathsf{H}\,(\overline{a}, 1^n) = n - |a|$
- thus $\mathsf{H}\,(x, \overline{a}) \leq 3\,(n - |\overline{a}|)$

Observe $\leq 3\,(n - |\overline{a}|)$ improving mutations sufficient
each with probability $\geq \binom{h}{1}\frac{1}{2en}$
for Hamming distance h

Thus $$\mathrm{E}\,(T_2) \leq \sum_{h=1}^{3(n-|\overline{a}|)} \frac{2en}{h} = \sum_{h=1}^{3|a|} \frac{2en}{h} \leq 2en\,(\log(3\,|a|) + 1)$$
$$= 2en\,(\log(|a|) + \log(3) + 1) \;= O(n \log(2 + |a|))$$ ✔

Thus $$\mathrm{E}\left(T_{\text{asym. (1+1) EA},\textsc{OneMax}_a}\right) = \mathrm{E}\,(T_1) + \mathrm{E}\,(T_2)$$
$$= O(n) + O(n \log(2 + |a|)) = O(n \log(2 + |a|))$$

□

Looking Back on $\textsc{OneMax}_a$

We have
$$\forall a\colon \mathrm{E}\left(T_{\text{asym. (1+1) EA},\textsc{OneMax}_a}\right) = O(n \log(2 + |a|))$$
$$\forall a\colon \mathrm{E}\left(T_{\text{(1+1) EA},\textsc{OneMax}_a}\right) = \Theta(n \log n)$$

Thus

- faster by a factor of up to $\Theta(\log n)$ if optimum has few 1-bits or few 0-bits
- equal speed (asymptotically) if optimum is 'in the middle of the search space'
- never (much) slower, sometimes clearly faster

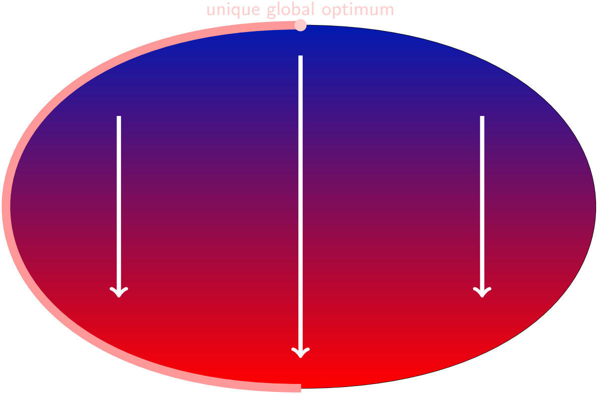
Caution These are results on $\textsc{OneMax}_a$.

Can it get worse?

An Example Function

Def. $\textsc{Plateau}\colon \{0,1\}^n \to \{0,1,\ldots,n+1\}$

$$\textsc{Plateau}(x) = \begin{cases} n+1 & \text{if } x = 1^n \\ n & \text{if } x = 1^i 0^{n-i},\ i \in \{0,1,\ldots,n-1\} \\ n-|x| & \text{otherwise} \end{cases}$$



(1+1) EA on Plateau

Fact $\mathrm{E}\left(T_{(1+1)\ \mathrm{EA},\mathrm{PLATEAU}}\right) = \Theta(n^3)$
(remember simulated annealing on similar example function)

Proof Ideas fair random walk on $\{0, 1, \ldots, n\}$ takes time $\Theta(n^2)$
(1+1) EA on Plateau is similar
probability for moving $\Theta(1/n)$
$\rightsquigarrow$ additional factor $\Theta(n)$

Essential unbiased random walk
(or only very, very, very slightly biased)

Is the bias of asymmetric mutations harmful?

Asymmetric Mutations on Plateau

Theorem

$$\text{Prob}\left(T_{\text{asym. (1+1) EA,PLATEAU}} = n^{o\left(n^{1/6}\right)}\right) = 2^{-\Omega\left(n^{1/6}\right)}$$

Proof

Observation hit plateau with $\leq 3n^{2/3}$ 1-bits
with probability very close to 1
since ① initially $> (2/3)n$ 1-bits,
② number decreases until on plateau and
③ this happens quickly (time $O(n)$)

Observation 'big jumps' play no role
since k-bit mutations have probability $< (e/k)^k$

What happens in small mutations with $|x| = \Theta\left(n^{2/3}\right)$?

Asymmetric Mutations with $|x| = \Theta\left(n^{2/3}\right)$ on the Plateau

We have $x = 1^i 0^{n-i}$ with $i = \Theta\left(n^{2/3}\right)$

Observe $\text{Prob}\,(\text{increase number of 1-bits})$

$$\leq \sum_{j=1}^{n-i} \left(\frac{1}{2(n-i)}\right)^j < \sum_{j=1}^{n-i} \left(\frac{1}{n}\right)^j = O\left(\frac{1}{n}\right)$$

Observe $\text{Prob}\,(\text{decrease number of 1-bits})$

$$\geq \sum_{j=1}^{i} \left(\frac{1}{2i}\right)^j \cdot e^{-1} = \Omega\left(\frac{1}{n^{2/3}}\right)$$

Thus $\text{Prob}\,(\text{increase number of 1-bits} \mid \text{change number of 1-bits})$

$$= \frac{O(1/n)}{\Omega\left(1/n^{2/3}\right)} = O\left(\frac{1}{n^{1/3}}\right)$$

Thus $\text{Prob}\,(\text{decrease number of 1-bits} \mid \text{change number of 1-bits})$

$$= 1 - \text{Prob}\,(\text{increase number of 1-bits} \mid \text{change number of 1-bits})$$

$$= 1 - O\left(\frac{1}{n^{1/3}}\right)$$

Asym. Mutations with $|x| = \Theta\left(n^{2/3}\right)$ on Plateau (cont.)

We have $\text{Prob}\left(\text{increase number of 1-bits} \mid \text{change number of 1-bits}\right) = O\left(\frac{1}{n^{1/3}}\right)$

$\text{Prob}\left(\text{decrease number of 1-bits} \mid \text{change number of 1-bits}\right) = 1 - O\left(\frac{1}{n^{1/3}}\right)$

We have unfair random walk
biased away from the global optimum

We know reaching global optimum in spite of this bias
very unlikely (weakly exponential)
due to result on gambler's ruin problem

Observation still weakly exponential
in a weakly exponential number of attempts
(number of attempts sufficiently small)

□

Summary & Take Home Message

Things to remember

- asymmetric mutations
- comparison with standard bit mutations
  - small and large mutations
  - speed-up on OneMax
- carefully biasing variation

Take Home Message

- Incorporating domain knowledge is usually beneficial.
- Biasing variation can speed search considerably.
- Think and check what you've done.