Exploring the Loss Landscape in Neural Architecture Search

Colin White
Abacus.AI

Sam Nolen
Abacus.AI

Yash Savani
Abacus.AI, Carnegie Mellon University

Slides (with hyperlinks): https://crwhite.ml
Neural architecture search

Architectures are getting increasingly more specialized and complex

Source: https://towardsdatascience.com/understanding-and-visualizing-densenets-7f688092391a
NAS Algorithms

Algorithms are getting complex too

[Bender et al. 2018]

[Mellor et al. 2020]

[White et al. 2021]
Noise in the loss landscape

NAS: given a set $A$, find $\max_{a \in A} f(a)$
Is NAS only hard because of noise?

Some recent NAS techniques have been designed specifically to deal with noisy architecture evaluations.

Image source: https://www.cs.umd.edu/~tomg/projects/landscapes/
Outline

● Motivation

● Local Search
  ○ Strong baseline for NAS
  ○ SotA on de-noised search spaces

● Theoretical analysis
  ○ Characterization of local search
  ○ Characterization of number of local minima
  ○ Simulation results
Local Search

- Five lines of code

**Algorithm 1** Local search

**Input:** Search space $A$, objective function $\ell$, neighborhood function $N$

1. Pick an architecture $v_1 \in A$ uniformly at random
2. Evaluate $\ell(v_1)$; denote a dummy variable $\ell(v_0) = \infty$; set $i = 1$
3. While $\ell(v_i) < \ell(v_{i-1})$ :
   i. Evaluate $\ell(u)$ for all $u \in N(v_i)$
   ii. Set $v_{i+1} = \text{argmin}_{u \in N(v_i)} \ell(u)$; set $i = i + 1$

**Output:** Architecture $v_i$

Evaluate all architectures in the neighborhood
Performance vs. Noise on NAS-Bench-301

The graphs illustrate the relationship between noise factor added (standard deviation of ensemble) and validation error, as well as the number of iterations to convergence for different methods. The methods compared include Random, Reg. Evolution, BANANAS, Local Search, and BayesOpt with GP.
Performance vs. Noise on NASBench-101/201

**NASBench-101, standard**

- Random
- Reg. Evolution
- BANANAS
- BayesOpt with GP
- Local Search

**NASBench-101, reduced noise**

- Random
- Reg. Evolution
- BANANAS
- BayesOpt with GP
- Local Search

**NASBench-201, CIFAR-10, standard**

- Random
- Reg. Evolution
- BANANAS
- Local++
- Local Search
- BayesOpt with GP

**NASBench-201, CIFAR-10, reduced noise**

- Random
- Reg. Evolution
- BANANAS
- Local++
- Local Search
- BayesOpt with GP
## NASBench-201: Computing #minima

| Version   | # iters | # local min. | % reached global min. |
|-----------|---------|--------------|-----------------------|
| Denoised  | 5.36    | 21           | 47.4                  |
| Standard  | 4.97    | 55           | 6.71                  |
| Random    | 2.56    | 616          | 0.717                 |

NAS-Bench-201 cifar10, 15,625 total architectures
Theoretical results

- Given the distribution of accuracies and neighborhood graph

**Theorem 5.1.** Given $|A| = n$, $\ell$, $s$, $\epsilon$, $\text{pdf}_n$, and $\text{pdf}_e$, we have

$$
\mathbb{E}[[\{v \in A \mid LS^*(v) = v\}]] = n \int_{\ell(v^*)}^{\infty} \text{pdf}_n(x) \left( \int_x^{\infty} \text{pdf}_e(x, y) dy \right)^s dx, \text{ and}
$$

$$
\mathbb{E}[[\{v \in A \mid \ell(\ell S^*(v)) - \ell(v^*) \leq \epsilon\}]] = n \int_{\ell(v^*)}^{\ell(v^*)+\epsilon} \text{pdf}_n(x) \left( \int_x^{\infty} \text{pdf}_e(x, y) dy \right)^s \mathbb{E}[\|\ell S^*(x)\|] dx.
$$

**Lemma 5.2.** Given $A$, $\ell$, $s$, $\text{pdf}_n$, and $\text{pdf}_e$, then for all $v \in A$, we have the following equations.

$$
\mathbb{E}[[LS^{-1}(v)]] = s \int_{\ell(v)}^{\infty} \text{pdf}_e(\ell(v), y) \left( \int_{\ell(v)}^{\infty} \text{pdf}_e(y, z) dz \right)^{s-1} dy, \text{ and}
$$

$$
\mathbb{E}[[LS^{-k}(v)]] = b_{k-1} \cdot \mathbb{E}[[LS^{-1}(v)]] \left( \frac{\int_{\ell(v)}^{\infty} \text{pdf}_e(\ell(v), y) \mathbb{E}[[LS^{-(k-1)}(v)]] dy}{\int_{\ell(v)}^{\infty} \text{pdf}_e(\ell(v), y) dy} \right).
$$

Example:
For uniform dist, #local minima = |search space| / |neighbors|

$$
15625/24 = 651.04
$$
Neighborhood distributions

CIFAR-10  CIFAR-100  ImageNet16-120
Simulation Results

- Unif. Random
- ImageNet16-120
- CIFAR-10
- CIFAR-100

- Thm 5.1 sim. - Unif. Random
- Thm 5.1 sim. - ImageNet16-120
- Thm 5.1 sim. - CIFAR-10
- Thm 5.1 sim. - CIFAR-100

- Lemma 5.2
- NASBench-201 with losses in $U([0,1])$
Conclusion

- Local search is a strong baseline for NAS
- Local search is SotA on de-noised search spaces
- Theoretical characterization of local search

https://github.com/naszilla/naszilla

Thanks!