



Neural Architecture Search

With Distributed
Reinforcement Learning
George Kyriakides

Σύνοψη



Κίνητρα - Στόχοι

- Κίνητρα
 - Ο χρόνος που χρειάζεται για το χειροκίνητο σχεδιασμό και έλεγχο των αρχιτεκτονικών των δικτύων.
 - Ο ρόλος της αρχιτεκτονικής στις επιδόσεις του δικτύου.
- Στόχοι
 - Μπορούν να χρησιμοποιηθούν κατανεμημένοι αλγόριθμοι τύπου actor-critic για την εκπαίδευση ενός πράκτορα στο σχεδιασμό αρχιτεκτονικών?
 - Μπορούν να εκπαιδευθούν μερικώς οι αρχιτεκτονικές προκειμένου να αξιολογηθούν?

Νευρωνικά Δίκτυα

- Παρέχουν πάρα πολλές δυνατότητες
 - Universal Function Approximators
 - Επιβλεπόμενη μάθηση (Image – speech recognition)
 - Μη-επιβλεπόμενη μάθηση (Image style transfer)
- Περίπλοκες αρχιτεκτονικές
 - Fully-connected, dropout, dropconnect, convolutions, pooling Layers
 - Tanh, arctan, sigmoid, relu, leaky relu, exponential relu, softplus, softmax, softsign
 - Residual nets, Recursive nets
- Μεγάλο υπολογιστικό κόστος για την εκπαίδευση μιας σύγχρονης αρχιτεκτονικής

HPC

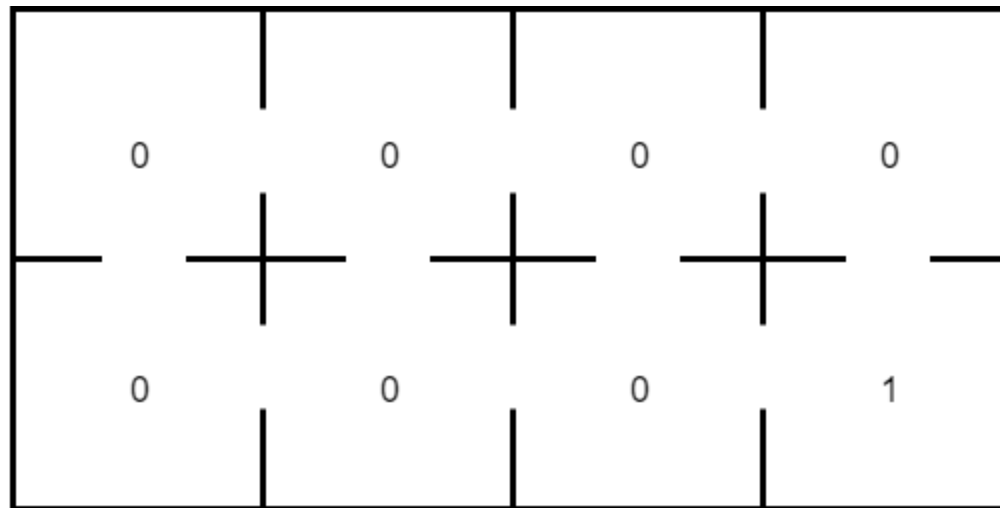
- Παραλληλισμός εργασιών ή/και δεδομένων
 - Πολλοί πυρήνες ή
 - Πάρα πολλοί απλοί πυρήνες
- Shared Memory
 - Multi-core (π.χ. Xeon® Platinum 8176 – 28core)
 - Multi-processor (π.χ. 2 Xeon® Platinum 8176)
- Distributed Memory
 - MPI
 - Hadoop
- GPGPU
 - CUDA (Βασικός παράγοντας Deep Learning)
 - OpenCL
- FPGAs
- Distributed GPUs/FPGAs

Reinforcement Learning

- Τρόπος για την επίλυση προβλημάτων όπου ένας πράκτορας πρέπει να αποφασίσει το πως θα δράσει προκειμένου να μεγιστοποιήσει (αθροιστικά) μία αμοιβή.
- Προβλήματα που επίσης μελετώνται σε άλλους κλάδους όπως:
 - Βελτιστοποίηση
 - Θεωρία ελέγχου
 - Θεωρία παιγνίων
 - Πολυπρακτορικά συστήματα κλπ.
- Συνήθως το πρόβλημα μοντελοποιείται ως Διαδικασία Αποφάσεων Μαρκόβ (MDP).
- Χρησιμοποιείται όταν δεν υπάρχει πρότερη γνώση σχετικά με τα χαρακτηριστικά της MDP (πιθανότητες μετάβασης, αμοιβές κλπ.)
- Τα χαρακτηριστικά μαθαίνονται από τον πράκτορα, καθώς αλληλεπιδρά με το περιβάλλον του (ενεργεί και λαμβάνει ανάδραση)

Παράδειγμα MDP

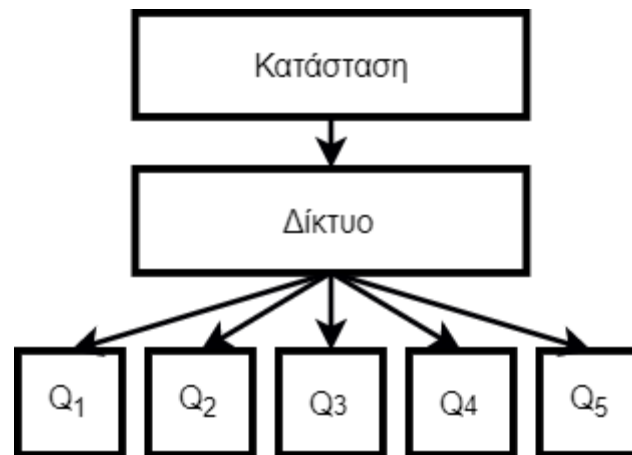
- Καταστάσεις $S = \{1, 2, 3, 4, 5, 6, 7, 8\}$
- Δράσεις $A = \{\Pi, K, A, \Delta, M\}$
- Πιθανότητα Μετάβασης $P_a(s, s')$
- Αμοιβή $R = \{0, 0, 0, 0, 0, 0, 0, 1\}$
- Παράγοντας έκπτωσης μελλοντικών αμοιβών $0 \leq \gamma \leq 1$
- Πολιτική $\pi =$ πως δρα ο πράκτορας σε κάθε κατάσταση
- Αξία της κατάστασης $V_\pi(s) = E(\sum_{t=0}^T \gamma^t R_{t+1} | s, \pi)$
- Αξία της ενέργειας $Q_\pi(s, a) = E(\sum_{t=0}^T \gamma^t R_{t+1} | s, \pi, a)$



Deep Q-Learning

- Προσέγγιση του Q με νευρωνικό δίκτυο
- Επιλογή ενέργειας με βάση τη μέγιστη απολαβή, $\max_a Q(s_{t+1}, a)$
- Εκπαίδευση $Q(s_t, a) = R_a(s_t, s_{t+1}) + \gamma \max_a Q(s_{t+1}, a)$
- Σύνολο εκπαίδευσης

| Κατάσταση | Ενέργεια | Αθρ.Αμοιβή | Πρόβλεψη | X | Y |
|-----------|----------|------------|-----------------------|----------------------------|--------------------------------|
| | 7Δ | 1 | [0.5,0.1,0.8,0.3,0.6] | [0,0,0,0,0,0, 1 ,0] | [0.5,0.1,0.8, 1.0 ,0.6] |
| | 8Α | 0 | [0.7,0.8,0.9,0.8,0.8] | [0,0,0,0,0,0,0, 1] | [0.7,0.8, 0.0 ,0.8,0.8] |



Double Deep Q-Learning

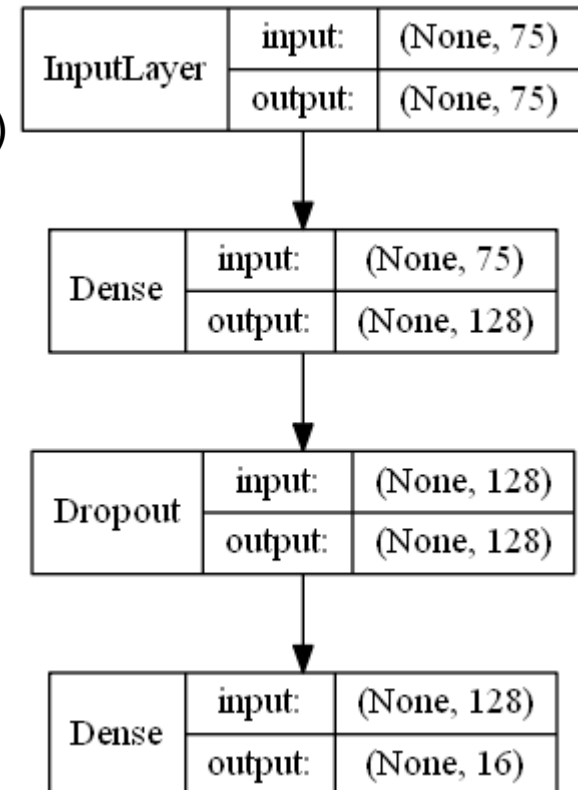
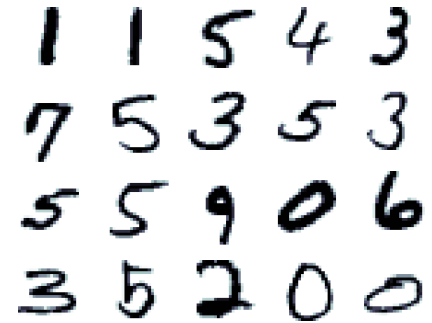
- 2 νευρωνικά δίκτυα προσεγγίζουν το Q (αξία της ενέργειας)
- Το ένα παράγει το $\max_a Q_B(s_{t+1}, a)$
- Το δεύτερο προσεγγίζει το $Q_A(s_t, a)$

$$Q_A(s_t, a) = R_a(s_t, s_{t+1}) + \gamma Q_A(s_{t+1}, \max_a Q_B(s_{t+1}, a))$$

- Μετά από ένα batch, τα βάρη του A αντιγράφονται στο B (απλούστερη έκδοση)
- Αποτελέσματα
 - Μειώνεται η υπερεκτίμηση του Q και σταθεροποιείται η εκπαίδευση.
 - Δεν παράγονται καλύτερες λύσεις.

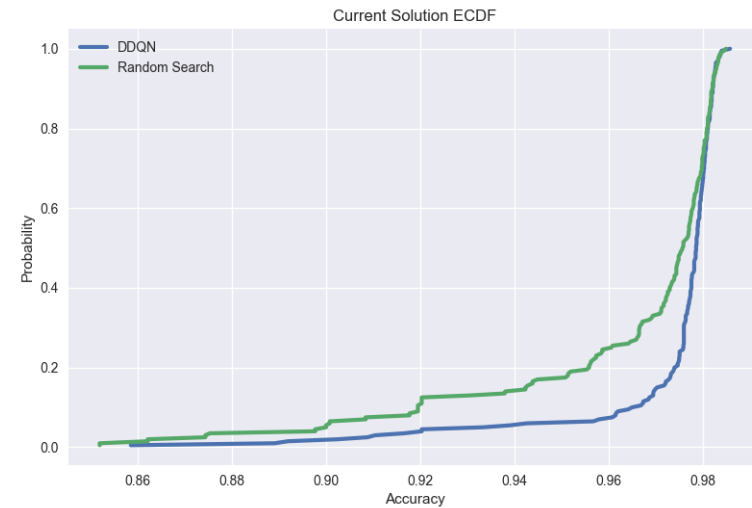
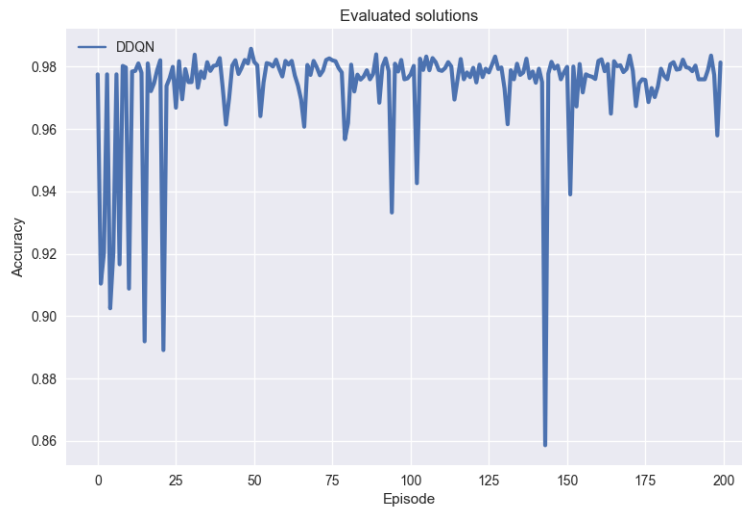
DDQN NAS

- Σύνολο δεδομένων MNIST
- Δίκτυα 4 επιπέδων
- Τύποι νευρώνων
 - Fully Connected(128, 256, 384, 512, 640, 768, 896, 1024)
 - Dropout(0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7)
- 50.625 δυνατοί συνδυασμοί – καταστάσεις
- 200 επεισόδια εκπαίδευσης
- Κατάσταση: 4 one-hot vectors
- Αμοιβή: βελτίωση της ακρίβειας
- (3 εποχές εκπαίδευσης)
- Υλοποίηση με Keras

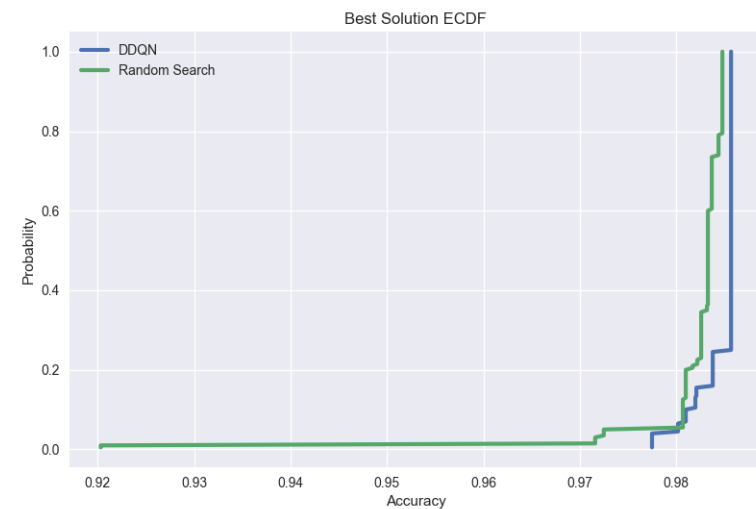


DDQN NAS

Αποτελέσματα



| Method | Best Mean | Current Mean | Total Time (s) |
|---------------|-----------|--------------|----------------|
| DDQN | 98.5% | 97.4% | 25598 |
| Random Search | 98.2% | 96.4% | 29629 |



DDQN NAS

Αποτελέσματα

Πιθανότητα επαναδειγματοληψίας μίας αρχιτεκτονικής

$$P(S = x) = \binom{r}{x} \times \left(\frac{1}{n}\right)^x \times \left(\frac{n-1}{n}\right)^{r-x}$$

$$x = 2, r = 200, n = 50625$$

$$P(S = x) = 7.85e-06$$

Advantage Actor-Critics

- Advantage Function

$$A_{\pi}(s_t, a_t) = Q_{\pi}(s_t, a_t) - V_{\pi}(s_t) \approx R - V(s_t)$$

- Actor (Αποφασίζει για την ενέργεια/πολιτική)

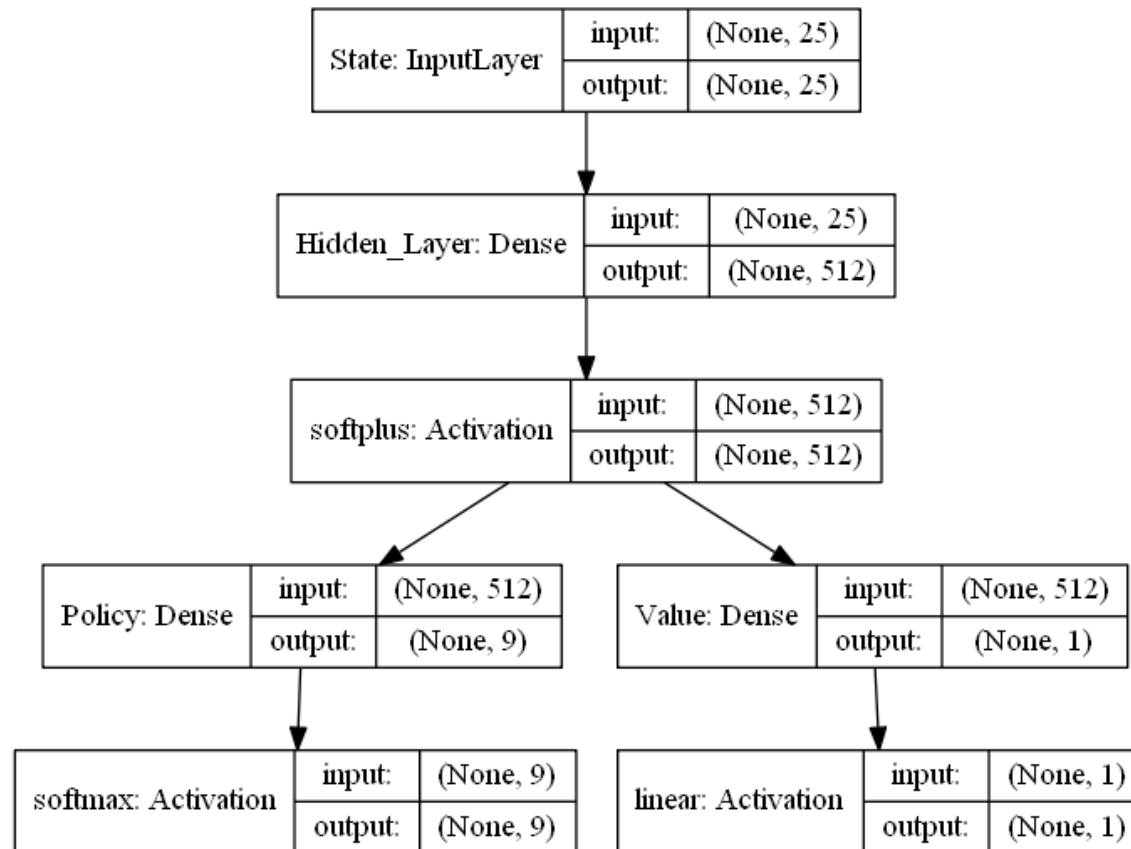
$$d\theta_a = d\theta_a + \nabla_{\theta_a} \log \pi(a_i | s_i)(R - V(s_i)) + \beta \nabla_{\theta_a} H(\pi(s_i))$$

- Critic (Αποφασίζει-κρίνει την ποιότητα της ενέργειας/πολιτικής)

$$d\theta_v = d\theta_v + \frac{\partial (R - V(s_i))^2}{\partial \theta_v}$$

- Υλοποίηση ως ένα νευρωνικό δίκτυο με δύο output layers

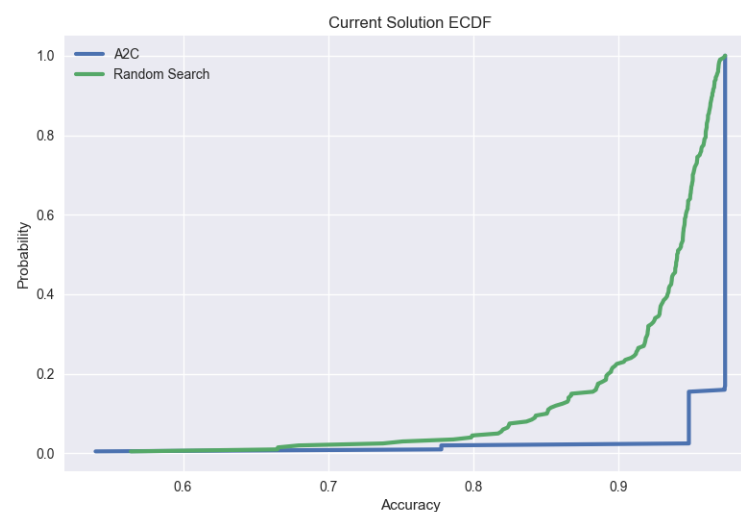
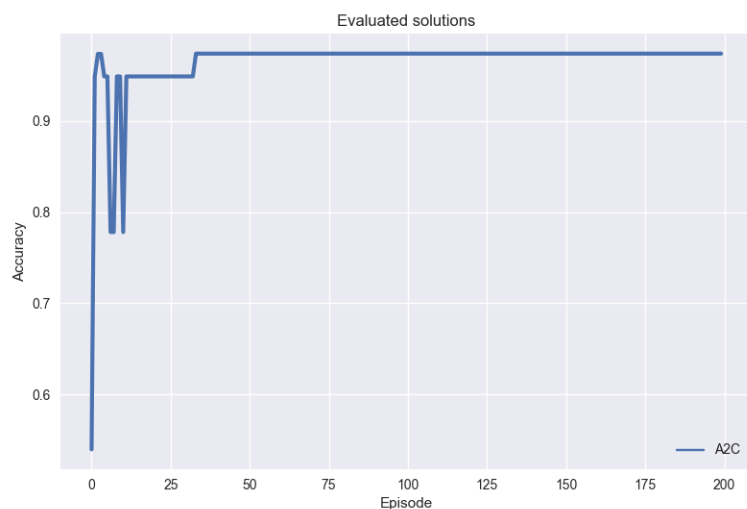
Advantage Actor-Critic Controller



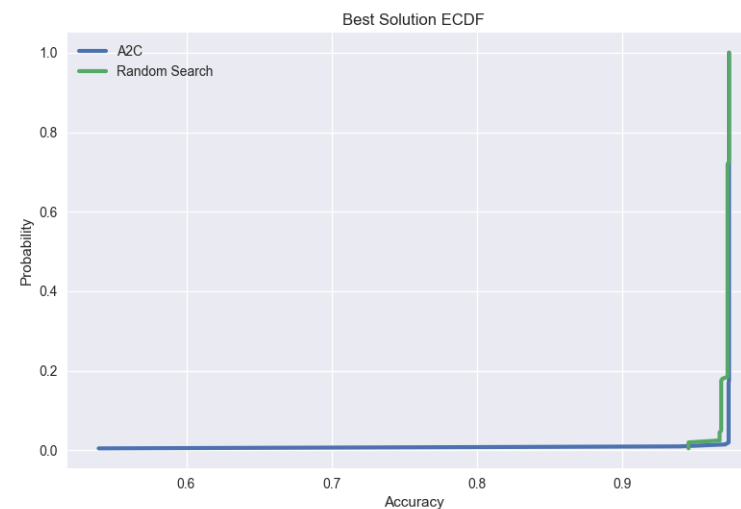
A2C Discrete NAS

- MNIST
- Τύποι νευρώνων
 - Fully Connected (64, 128, 256, 512, 1024)
 - Dropout (0.1, 0.3, 0.5)
- 4 Επίπεδα
- 200 Επεισόδια
- Tensorflow + MPI
- 4 Workers, 1 Master

A2C Discrete NAS

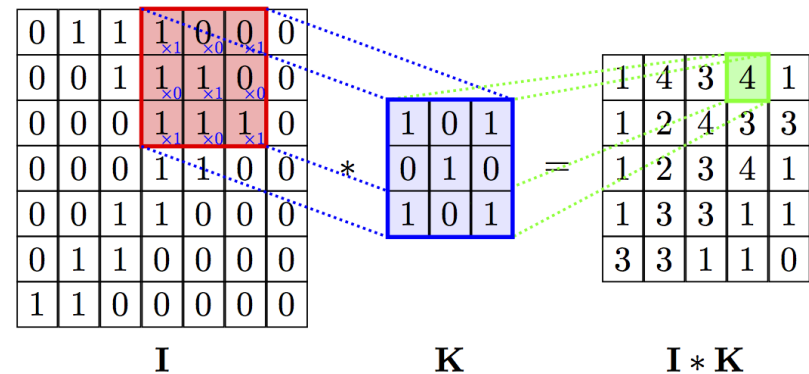


| Method | Best Mean | Current Mean | Total Time (s) |
|---------------|-----------|--------------|----------------|
| A2C | 97.4% | 96.5% | 1212 |
| Random Search | 97.2% | 92.2% | 295 |

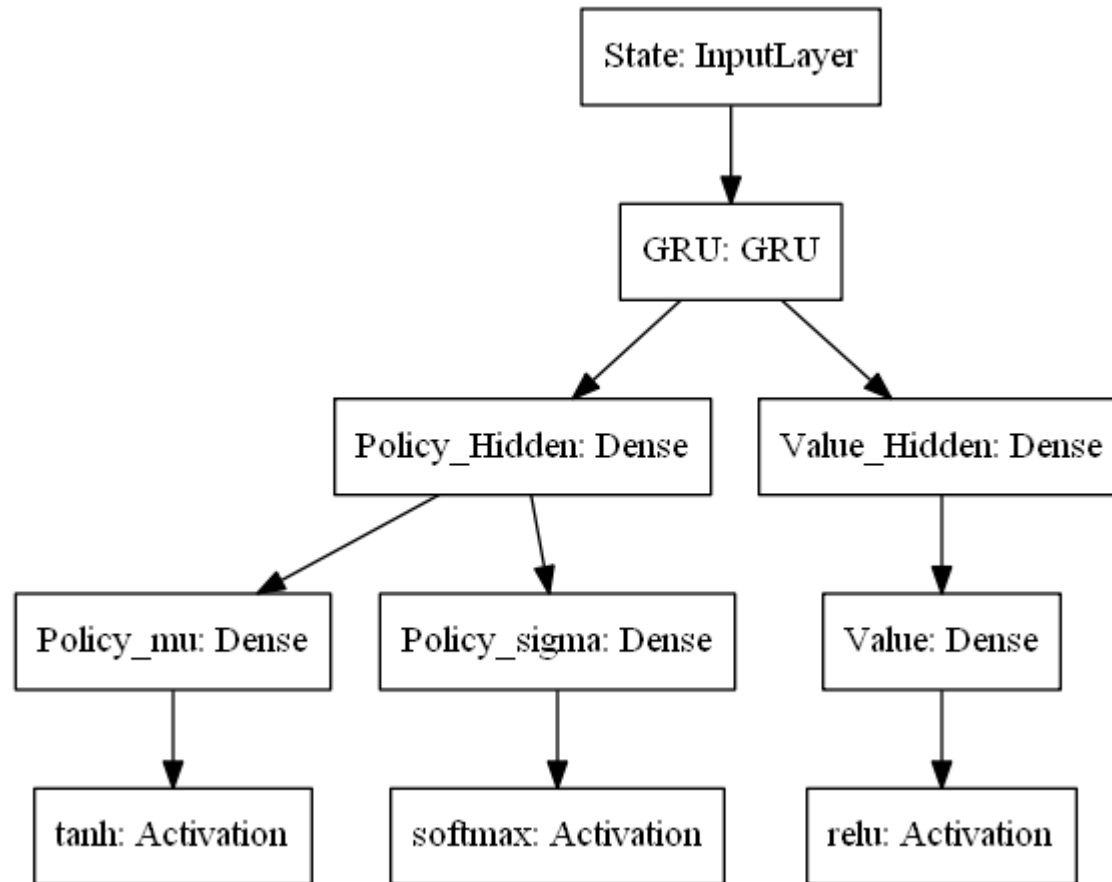


A2C Continuous NAS

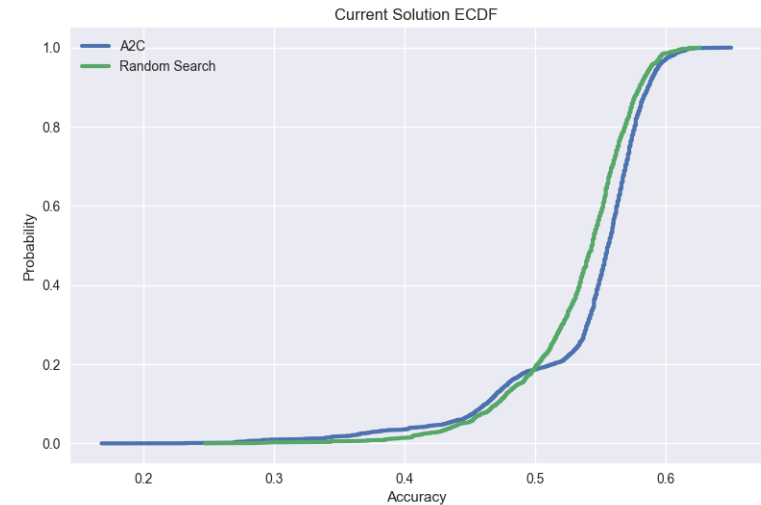
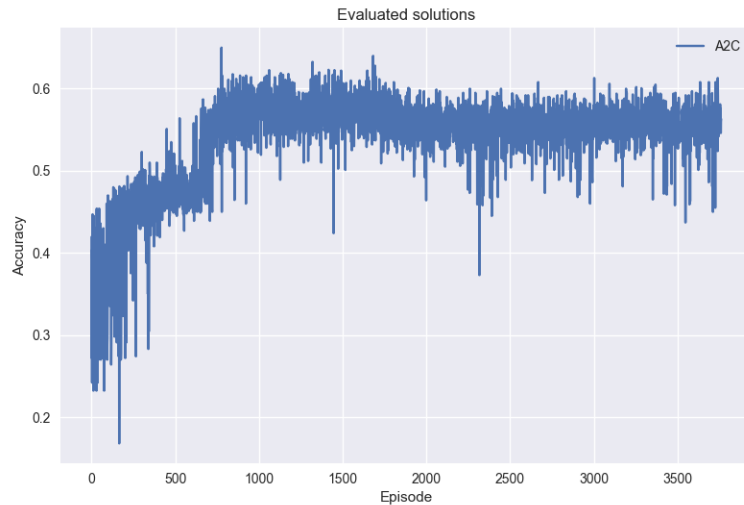
- CIFAR10 Dataset
- Fully convolutional networks 2 επιπέδων
- Παράμετροι: Μέγεθος φίλτρου και αριθμός
- 4000 επισόδεια
- Κατάσταση: (s_1, n_1, s_2, n_2)
- Tensorflow+MPI



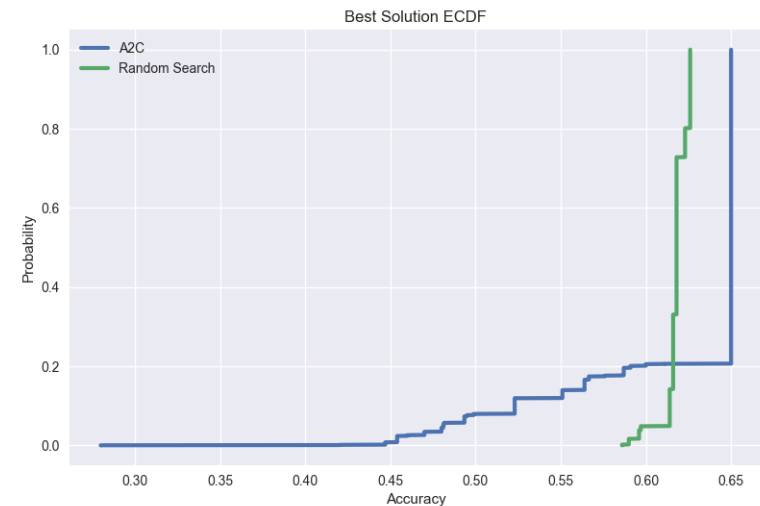
Continuous Controller



A2C Continuous NAS

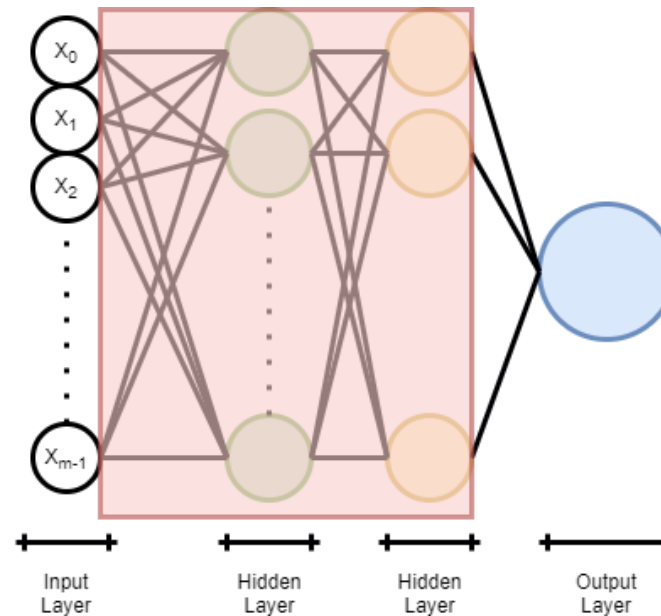


| Method | Best Mean | Current Mean | Total Time (s) |
|---------------|-----------|--------------|----------------|
| A2C | 62.4% | 53.9% | 88679 |
| Random Search | 61.8% | 53.4% | 106909 |

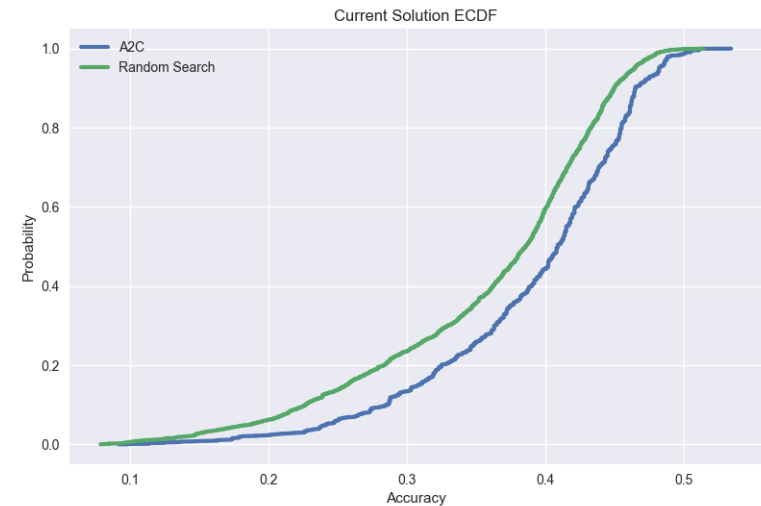
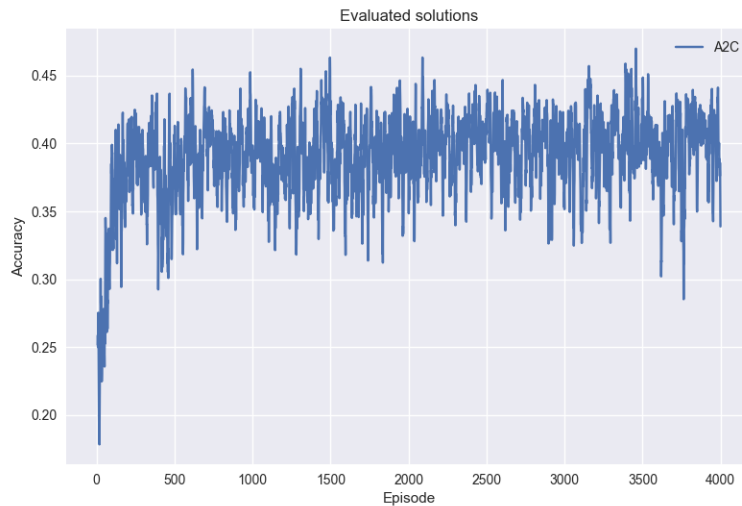


Μερική εκπαίδευση

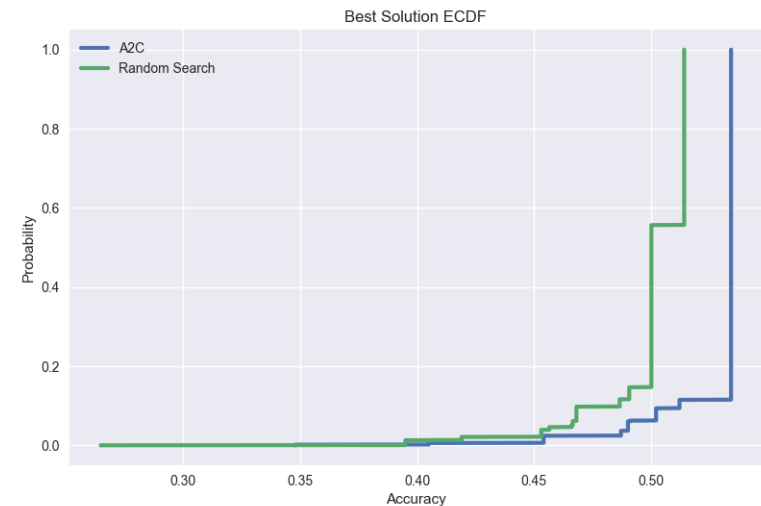
- Εκπαιδεύεται μόνο το επίπεδο εξόδου του δικτύου
- Στόχοι
 - Λιγότεροι παράμετροι εκπαίδευσης
 - Ταχύτερη αξιολόγηση



Μερική εκπαίδευση



| Method | Best Mean | Current Mean | Total Time (s) |
|---------------|-----------|--------------|----------------|
| A2C | 52.8% | 39.0% | 21934 |
| Random Search | 50.2% | 36.2% | 56463 |



Ευχαριστώ πολύ!

