

Rapport de projet ISI-10: AdaNet, Adaptive Neural Networks.

Serife Akkoyunlu, Luc Blassel, Romain Gautron

2 mars 2018

Introduction

Le but de ce projet est de reproduire les experiences et les methodes issues du papier suivant : C. Cortes, X. Gonzalvo, V. Kuznetsov, M. Mohri, S. Yang *AdaNet: Adaptive Structural Learning of Artificial Neural Networks* En essayant de reproduire leur methode de construire des reseaux de neurones dont la structure est apprise et optimisee en meme temps que les poids, sur une tache de classification binaire issue du jeu de donnees d'images CIFAR-10.

1 Comment marche AdaNet ?

La structure du reseau est est generee. Le reseau entier qu'on appellera le reseau AdaNet, est augmente a chaque iteration par un sous reseau. Ce sous reseau est choisi en fonction de son effet sur une fonction objective pour pouvoir selectionner le meilleur sous reseau a rajouter au reseau Adanet a chaque etape. Pour bien comprendre le deroulement de ce processus on notera : f_t le reseau AdaNet a l'etape t , h_t et h'_t les sous reseaux candidats (a l'etape t) avec pour chacun de ces sous reseaux $h_{t,k}$ la k^{eme} couche du sous reseau et $h_{t,k,j}$ le j_{eme} neurone de cette couche. l est le nombre de couche maximale du sous reseau.

L'algorithme se deroule alors selon ces etapes :

1. **Initialisation du reseau** : On commence par generer les couches d'entree et de sortie.
2. **Generation de sous-reseaux candidats** : On genere deux sous reseaux candidats (cf. Figure 1) :
 - Un qui a une profondeur identique au sous reseau genere precedement $\rightarrow k$
 - Un qui augmente la profondeur de 1 par rapport au sous reseau precedent $\rightarrow k + 1$Ces deux sous reseaux obeissent aux memes regles de connectivite. C'est a dire que la premiere couche du reseau du sous reseau h_t est reliee a la couche d'entree de f_t , la derniere couche $h_{t,l}$ du sous reseau est forcement connectee a la couche de sortie de f_t . Pour les couches intermediaires, chaque couche $h_{t,k}$ est forcement reliee a $h_{t,k-1}$ et peut etre reliee aux couches de niveau precedent dans tous les sous reseaux precedement generes, c'est a dire $h_{t \in [1, t-1], k-1}$. Lors de la premiere iteration il faut forcement choisir le sous reseau qui augmente la profondeur puisque $k = 0$
3. **Choix du sous-reseau** : Parmi ces deux sous reseaux candidats \mathbf{h} et \mathbf{h}' on choisit celui qui, apres un entrainement, minimise le plus la fonction objective suivante :

$$F_t(\mathbf{w}, \mathbf{h}) = \frac{1}{m} \sum_{i=1}^m \Phi(1 - y_i f_{t-1}(x_i) - y_i \mathbf{w} \cdot \mathbf{h}(x_i)) + \mathcal{R}(\mathbf{w}, \mathbf{h})$$

Avec x_i les exemples d'entrainement, y_i les reponses attendues, m Le nombre d'exemples, \mathbf{w} Les poids associes au sous reseau candidat \mathbf{h} et Φ une fonction (soit la fonction exponentielle soit la fonction logistique). Le terme $\mathcal{R}(\mathbf{w}, \mathbf{h})$ est un terme de regularisation qui est laisse a 0 lors des experimentations.

Si $F(\mathbf{w}, \mathbf{h}) > F(\mathbf{w}, \mathbf{h}')$ alors le sous reseau candidat \mathbf{h}' est choisi et $\mathbf{h}_t \leftarrow \mathbf{h}'$

4. **Condition d'arrêt :** Une fois que le meilleur sous réseau est selectionne on regarde si celui ci a apporte une amelioration au reseau AdaNet. Si la fonction objective de l'entrainement de f_{t-1} (*qui n'est pas le meme que celle utilisee pour le choix du meilleur sous reseau*) est meilleure avec le sous reseau alors que sans alors : $f_t \leftarrow f_{t-1} + \mathbf{w}^*(h)_t$ sinon l'integration du sous reseau n'apporte rien de plus au reseau f_{t-1} et celui est retourne, terminant alors l'execution de l'algorithme. Si cette condition d'arrêt n'est jamais verifiee, l'elrogrithme s'arrete au bout de T iteration et retourne le reseau f_T

On voit donc bien que la generation et l'ajustement des poids du reseau se fait iterativement a chaque etape.

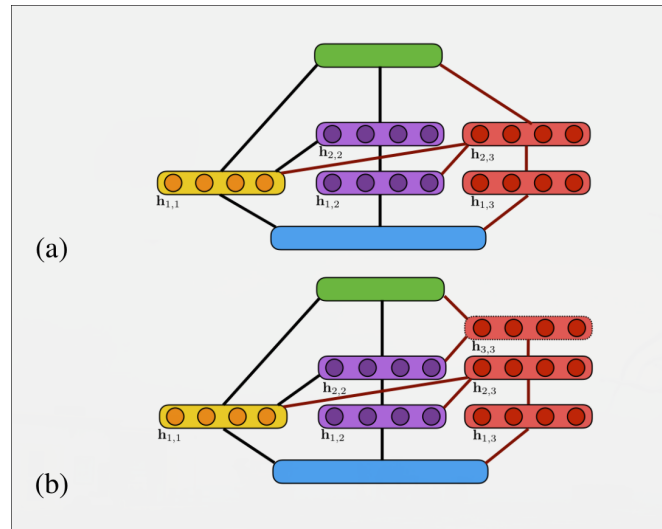


FIGURE 1 – Les deux sous reseaux candidats generes a l'etape 3 (en rouge), un de profondeur $k = 2$ comme le sous reseau choicit a l'etape precedente (en bleu), et un de profondeur $k = 3$ qui augmente la profondeur generale de 1. Ce sont ces deux reseaux AdaNets potentiels qui vont etre evalues

2 Notre implementation

3 Resultats

4 Pistes futures