



Són les Xarxes Neuronals més eficients que els arbres de decisió?

Anàlisi sobre un problema d'Astronomia

Albert Ribé Costa

Màster en enginyeria informàtica
Intel·ligència Artificial

Samir Kanaan Izquierdo

Carles Ventura Royo

28/12/2016

Índex

1. Introducció.....	4
2. Seguiment del projecte	4
3. Riscos	5
4. Diagrama de Gantt actualitzat.....	6

1. Introducció

En aquest document es descriu l'avenç realitzar al projecte durant el període de la PAC4, remarcant les tasques completades, i la finalització del projecte.

Addicionalment també s'inclou un apartat de riscos, identificant els que s'han detectat durant aquest període així com les mesures que s'han decidit prendre per tal de mitigar-los.

2. Seguiment del projecte

En aquesta darrera entrega s'ha finalitzat totes les tasques del projecte. Un cop el gruix del desenvolupament va estar acomplert durant la PAC3, en aquesta només ha calgut fer alguns ajustos, millores i solució d'incidències.

A continuació s'ha procedit a l'anàlisi dels dos algoritmes realitzant-ne tantes execucions amb diferents configuracions com calgués, cal destacar que aquesta ha estat la tasca que més temps a suposat en aquesta fita, ja que moltes de les execucions trigaven un temps considerable, en alguns casos superant varis dies de duració, per aquesta raó i tal com es va fer en el desenvolupament s'ha augmentat el nombre d'hores diàries dedicades per tal d'arribar a la fita.

Addicionalment durant aquesta fase s'ha acabat la redacció de la documentació i preparat el paquet d'entrega que conté tant els documents com els software desenvolupat.

A continuació es descriuen les tasques i el seu estat:

Tasca	Estat	Comentaris
Anàlisi inicial i establiment d'objectius	FINALITZADA	Finalitzat a la PAC1
Presa de requisits	FINALITZADA	Finalitzat a la PAC1
Anàlisi de dades	FINALITZADA	Finalitzat a la PAC2
Preparació d'entorn de desenvolupament	FINALITZADA	Finalitzat a la PAC2
Formació en Deep Learning	FINALITZADA	Finalitzat a la PAC3
Construcció de sets de test i entrenament	FINALITZADA	Finalitzat a la PAC3
Desenvolupament	FINALITZADA	S'han realitzat algunes millores i solucionat incidències
Proves amb dades reals	FINALITZADA	
Anàlisi de resultats	FINALITZADA	S'han realitzat els anàlisis dels algoritmes segons la planificació prèvia
Preparació de presentació	FINALITZADA	

3. Riscos

En quan als riscos, a continuació en podem veure les conseqüències i les tasques de mitigació empreses per reduir-ne l'efecte:

Risc	Conseqüència	Mitigació	Estat
Excessiva complexitat de les dades d'origen	Augment dels temps de desenvolupament, formació i anàlisi de dades	S'ha escollit un origen de dades concret que sigui més assequible per aconseguir els objectius del projecte	Finalitzat (Durant la PAC1) (EVITAT)
No trobar informació concreta d'exemples de FFNN fets amb Lasagne	Retard en la formació en DL i en l'inici del desenvolupament	S'ha demanat informació al respecte al tutor de projecte, amb els exemples proveïts s'ha iniciat la formació i proves més concretes	Finalitzat (Durant la PAC2) (MITIGAT)
La complexitat del desenvolupament de la NN és major al estimat	Retard i impossibilitat d'acomplir amb els objectius del projecte	Demander el suport del tutor de projecte en els dubtes trobats Si cal, replanificar algunes tasques i dedicar més temps a la formació i desenvolupament	Finalitzat (Durant la PAC3) (MITIGAT)
Problemes per fer funcionar Theano amb CUDA	Es possible que si no es poden solucionar els problemes d'incompatibilitat no es pugui executar l'algoritme en una GPU	Demander suport al tutor, i en cas de no poder solucionar els problemes realitzar les proves utilitzant la CPU tot tenint en compte a l'hora de fer l'anàlisi els canvis que una GPU suposaria en l'execució de l'algoritme. Per tal de poder completar el projecte es decidirà si cal deixar l'execució en GPU fora de l'abast	Finalitzat (ACCEPTAT)

3.1 Síntesi i lliçons apreses

Es pot afirmar que en quasi tots els riscos identificats s'ha pogut actuar i mitigar o fins i tot evitar. El cas més interessant de revisar, però, és l'últim, "Problemes per fer funcionar Theano amb CUDA" que finalment s'ha decidit acceptar i finalment no fer res per evitar-lo.

Per problemes amb l'entorn de desenvolupament i proves no era possible configurar-lo per utilitzar la GPU amb les llibreries de Nvidia CUDA, a primer cop d'ull era un greu problema ja que el no incloure-ho afecta de forma decisiva al funcionament de l'algoritme FNN, tot i així trobar el problema i intentar solucionar-lo va permetre descobrir modes de funcionament de Theano que a posteriori han ajudat en altres parts de l'algoritme.

Pel que fa a efectes comparatius amb l'arbre de decisió no afectava de forma decisiva, la velocitat d'execució, punt en que més ajuda utilitzar un GPU, ja quedava àmpliament superat utilitzant la FNN sense GPU (revisar l'anàlisi i conclusions a la memòria). Així mateix a la secció de conclusions de la memòria s'analitzen treballs futurs que podrien millorar la FNN, entre ells l'activació de l'execució en GPU.

4. Diagrama de Gantt actualitzat

En el diagrama es reflecteix l'evolució del projecte, on totes les tasques queden marcades com a finalitzades.

