



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

Anàlisis sobre un problema d'Astronomia

Albert Ribé Costa

Màster en enginyeria informàtica
Intel·ligència Artificial

Samir Kanaan Izquierdo

Carles Ventura Royo

30/11/2016

Índex

1. Introducció	4
2. Seguiment del projecte	4
3. Riscos.....	6
4. Diagrama de Gantt actualitzat	7

1. Introducció

En aquest document es descriu l'avenç realitzar al projecte durant el període de la PAC3, remarcant les tasques completades, les que estan en curs i les que s'iniciaran properament, a la PAC4.

Adicionalment 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 entrega s'ha avançat en el projecte en gran part segons el planificat, tot i així s'ha hagut de fer certs ajustos per tal de no incórrer en un retard a l'entrega.

Com es pot veure a la següent taula, les tasques que estaven en procés a l'anterior entrega; Formació en Deep Learning i Construcció de sets de test i entrenament ja han estat finalitzades.

Per altre banda, la tasca de desenvolupament està encara en procés però en un estat molt avançat, es considera que el desenvolupament està pràcticament completat, i només falta acabar de fer algunes modificacions a la part de la FNN.

La major part del temps de la fase s'ha dedicat, precisament, al desenvolupament de la FNN ja que com s'havia previst és la tasca més costosa.

A destacar durant aquesta fase és que el desenvolupament està resultat fins i tot més costós del que es creia, havent de *dedicar més hores de les planificades* inicialment per tal de d'acomplir amb la fita. Durant aquest desenvolupament ha estat molt important l'ajuda del tutor del projecte així com la recerca a través d'internet i sobretot l'assaig i error, ja que ens trobem en un projecte on s'utilitzen tècniques i recursos tècnics força nous i específics que dificulten la possibilitat de trobar fonts d'informació fiables. Sigui com sigui i mirant el diagrama de Gantt del punt 3 sembla que l'avenç del projecte va a bon ritme assolint les fites marcades prèviament.

Pel que fa a la següent fase (PAC4) es començarà l'anàlisi de resultats i alhora es realitzaran les proves amb dades reals, a destacar és que es reservarà part del temps de desenvolupament per solucionar possibles defectes que es puguin descobrir en la fase de procés i anàlisi, aquesta és la principal raó per la qual la tasca encara es marca com a no completada.

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	S'han realitzat proves, estudiat scripts existents i estudiat el temes referents a la programació d'una NN en Python així com els seus conceptes teòrics
Construcció de sets de test i entrenament	FINALITZADA	S'han creat els 3 conjunts de sets explicats al punt 3.2 de la memòria
Desenvolupament	EN PROCÉS	Durant aquesta fase s'ha dedicat la major part del temps a aquesta tasca, en la que s'ha completat el gestor de fitxers que ara s'anomena FileManger i té més funcionalitat, s'ha avançat en la creació de la NN, la qual està quasi completat
Proves amb dades reals	EN PROCÉS	Conforme es va avançant en el desenvolupament també es fan proves amb dades reals corresponents als datasets creats
Anàlisi de resultats	PENDENT	
Preparació de presentació	PENDENT	

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 acomplir els objectius del projecte	Finalitzat (Durant la PAC1)
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
La complexitat del desenvolupament de la NN és major al estimat	Retard i impossibilitat d'acomplir amb els objectius del projecte	Demandar el suport del tutor de projecte en els dubtes trobats Si cal, replanificar algunes tasques i dedicar més temps a la formació i desenvolupament	En procés
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	Demandar 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	Nou

4. Diagrama de Gantt actualitzat

En el diagrama es reflecteix l'evolució del projecte, en negre es pot veure les tasques completades o en procés. L'evolució marcada per la tasca de desenvolupament no correspon el 100% amb la realitat ja que Microsoft Project només permet marcar les tasques amb intervals de 25% de completitud, de manera que aquesta apareix una mica menys avançada del que realment està (apareix al 75%).

