



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

02/11/2016

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 PAC2, remarquant les fases completades les que estan en curs i les que s'iniciaran properament, a la PAC3.

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, aquest llistat d'anirà actualitzant en les següent entregues per tal de realitzar-ne també el seguiment.

2. Seguiment del projecte

En aquesta entrega s'ha assolit en gran part el planificat en la primera, això és, tenir un anàlisi de dades completat, i l'entorn de desenvolupament preparat, així mateix s'ha començat la formació en Deep Learning i s'ha iniciat un petit desenvolupament d'alguna part de la solució que serà utilitzada més endavant. A més s'ha començat a fer la selecció de conjunts de dades aprofitant l'experiència amb les eines d'SDSS obtinguda durant la fase d'anàlisi de dades.

A continuació es descriuen les tasques i el seu estat:

Tasca	Estat	Comentaris
Anàlisi inicial i establiment d'objectius	FINALITZADA	Finalitzat a la PAC1
Preses de requisits	FINALITZADA	Finalitzat a la PAC1
Anàlisi de dades	FINALITZADA	S'ha completat l'anàlisi de dades i determinat els datasets concrets que s'utilitzaran
Preparació d'entorn de desenvolupament	FINALITZADA	S'ha instal·lat tots els paquets i llibreries necessaris per tal de dur a terme el desenvolupament
Formació en Deep Learning	EN PROCÉS	S'ha completat la introducció teòrica a DL i està en procés la posada en pràctica dels nous coneixements utilitzant les llibreries necessàries
Construcció de sets de test i entrenament	EN PROCÉS	En acabar l'anàlisi de dades i després de fer proves amb les eines d'SDSS s'han pres certes decisions sobre els conjunts de dades
Desenvolupament	EN PROCÉS	S'ha desenvolupat un algoritme d'arbre de decisió que serà útil en el futur de cara a analitzar els resultats de la FFNN, també s'ha desenvolupat un codi per llegir fitxers fits
Proves amb dades reals	PENDENT	
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
Excesiva 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
No trovar 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	En procés
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	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. Es destaca que les tasques "Construcció de sets de test entrenament" i "Desenvolupament" que tal com s'ha dit, estan iniciades, no es marquen en el Gantt ja que no s'ha assolit el 25% mínim del ser avenç.

