

DOSSIER TFG

En aquest document es conforma la llista de canvis i nous avenços al llarg de les diferents entregues. Està organitzat de manera que gairebé secció per secció de cada informe de seguiment es vegi reflexat els canvis i per que s'han fet en alguns casos concrets.

Informe Inicial

Els apartats que van ser desenvolupats pel lliurament de l'informe inicial son els següents:

- Introducció
- State of the Art (SOTA), més concretament sobre Medical Image Captioning and Report Generation
- Explicació amb esquemes del model CLIP d'OpenAI i el seu funcionament.
- Objectius del treball (a gran escala, sense concretar punts, simplement idea del que es volia fer i aconseguir explorar)
- Methodology: Explicar per sobre com es dividira el desenvolupament del projecte i esmentar aquestes parts
- Explicar a alt nivell una mena de vista general sobre una primera proposta de la arquitectura del model.
- Bibliografia dels continguts que es van prendre com a referència en els apartats anteriors.

Adjunto els comentaris que vaig realitzar en aquella entrega:

"He adjuntat al final les referències utilitzades fins al moment. A més, he redactat una proposta d'Abstract (realment s'anirà modificant depenent de com avanci el projecte). S'ha fet una descripció de concepte a alt nivell sobre la idea de l'arquitectura, tot descrivint el flux de dades en cada classe o mètode que es preveu que serà clau. Tot i que és molt per sobre, s'ha intentat descriure que hi entraria d'input a cada part, què i com es processaria i quin seria l'output."

Informe Seguiment I

1. Millores i ampliacions de contingut respecte a l'informe inicial

Part conservada amb millores de redacció o forma:

- La introducció i el context es mantenen iguals, però redactades amb més formalitat i amb format article
- La descripció de CLIP i del seu procés de funcionament bàsic també es manté, ara amb figures afegides (Fig 1 i 2) per visualitzar els components.

Reformulació i desenvolupament ampliat d'objectius i metodologia:

- S'introdueix una metodologia basada en Kanban amb Trello (no apareixia a l'informe inicial).
- Es concreten les fases de desenvolupament, que abans eren genèriques, i es detalla una pipeline completa d'entrenament (text + imatge) amb contrastive learning.

2. Nous continguts afegits

Exploració de dades: nova secció detallada de Data Exploration

- Descripció dels tipus de dades (CT Scans i metadades clíniques).
- Selecció de 5 descriptors clínics (nodule shape, density, infiltration, cdiff, necrosis).
- Estudi estadístic de distribucions de variables i correlacions (χ^2 i Cramér's V).

Desenvolupament del model

- Definició d'un model multimodal anomenat CLIPMedical, no esmentat abans.
- Es detalla l'ús de ClinicalBERT com a encoder de text i l'arquitectura completa del sistema.

- S'explica el flux d'entrenament contrastiu: projecció d'embeddings, càlcul de pèrdues, optimització amb Adam, etc. (Experimental design)

Avaluació i validació

- Secció de test & evaluation on s'introdueix el concepte de "catalog de text embeddings"
- Es fa una recuperació (retrieval) top-1 o s'avalua mitjançant *confusion matrix* i mètriques de classificació.
- Resultats emmagatzemats i visualitzats amb *boxplots* (Figures 10-13)

Disseny experimental

- Es detalla un protocol experimental amb 5-fold-cross-validation
- Es defineix l'experiment de comparació d'espais visuals amb diferents mètriques i s'obtenen i interpreten els primers resultats.

Comentari adjuntat amb l'entrega a l'aplicatiu:

"Hola,

com ja t'he comentat per correu el que he fet en comptes d'entregar un pdf com al lliurament inicial és anar completant ja el treball que vaig fent sobre la plantilla de l'informe final per així anar-lo tenint actualitzat amb cada entrega i tenint ja l'estructura que vull seguir. Sobre això, a la reunió de seguiment, hauriem de comentar que et semblen els apartats i com ho he estructurat. En quant al anàlisis comparatiu d'extractors de la part final, estic acabant d'executar unes proves tant per ResNet152 com per MobileNetV2, així que demà pujaré una nova versió que serà la bona. No està inclòs en aquest informe, però estic començant a fer proves intentar fer un mecanisme d'attention pels embeddings de text i per un altre costat també estic intentant provar l'estrategia de fer servir una loss diferent (TripletMarginLoss en aquest cas). No ho he inclòs per què encara estic fent proves inicials i les primeres execucions, però vull comentar amb tú la lògica de fer servir aquesta funció de pèrdua."

Informe Seguiment II

Millores generals:

- Resum i abstract afegits amb resultats concrets i esmenta experiment realitzats
- Redacció millorada i orientada a resultats
- S'ha donat més èmfasi a la viabilitat clínica i les limitacions pràctiques.

Nous desenvolupaments tècnics:

- Nova forma d'agregar les features de les imatges des dels slices dels CT scans. Es va desenvolupar un mecanisme d'atenció per a obtenir les característiques més rellevants per a la formació dels embeddings (abans la forma estàndard era fer un promig).

Avaluació més profunda:

- S'ha mantingut el pipeline general de test i train, però s'hi ha afegit una comparació entre arquitectures amb anàlisis i interpretació de mètriques i visualitzacions (confusion matrix i boxplots).
- S'identifica problemàtica dels *missing values* i s'explica com s'afronta.
- MobileNetV2: es reporten oscil·lacions en la corba de pèrdua i s'inclou una visualització d'aquesta. A més de un plot explicatiu de com entrena batch a batch el model, indicant que el matching entre text i imatge a cada batch és molt bo, lo qual podria ser causa d'un overfitting clar.
- Anàlisis per descriptor detallat amb plots referenciats.

Primeres conclusions (evidentment no definitives)

- Es reconeix que els resultats no són clínicament viables encara
- Es detecta una forta dependència dels desequilibris de classe
- L'evidència qualitativa no compensa les mancances quantitatives en validació.

Comentari adjuntat amb l'entrega d'aquest informe al aplicatiu:

"Parts de la conclusió son momentànies, pendent d'algunes proves.

En la part d'anàlisis del dataset falten un parell de taules per mostrar la distribució de cada variable.

Falta un esquema en l'explicació de la arquitectura alternativa i anàlisi de resultats."

Proposta Final d'Informe

Millores en la metodologia i la planificació:

- Ús de la metodologia Agile i Kanban amb Trello per la gestió de tasques amb referència i figura al apèndix explicant-ho.
- Planificació més detallada

Exercici d'experimentació més complet:

- Desenvolupament de la comparativa dels extractors que quedaven (ResNet152 i MobileNetV2) amb interpretació de resultats i tria de ResNet152 com a millor extractor pre entrenat

Avaluació i resultats:

- Anàlisis estadístic per descriptor amb mètriques
- Mètodes d'avaluació millorats per a la validació, amb resultats de recuperació de top-1 i matrius de confusió per descriptor.
- S'ha introduït un nou anàlisis per pacient, amb percentatge de descriptors correctes per pacient, ajudant a comprendre millor l'aplicabilitat clínica del model.

Limitacions i conclusions més detallades:

- Reflexió sobre la viabilitat clínica
- S'han esmentat futures millores proposades

Comentari adjuntat amb l'entrega d'aquesta proposta final d'informe a l'aplicatiu:

"T'adjunto la proposta final en format pdf, en principi les imatges i taules haurien d'estar correctes, a mi almenys em surten totes bé. Si no comentem. Al final he adjuntat una imatge del esquema que hi apareix a la figura 7, ja que no es veu massa bé. Com a apèndix he adjuntat una captura de l'eina que he utilitzat per a organitzar les tasques i portar-les al dia (menciono aquesta eina al apartat de metodologia i planificació). Demà comentem a veure que et sembla."

Informe Final

Canvis i millores:

- Per diagnosticar el càncer de pulmó: S'ha afegit una explicació breu sobre els CT scans i la seva estructura, per contextualitzar millor l'ús d'aquestes imatges en el projecte.
- Eliminar l'expressió "contrastive learning": S'ha eliminat aquesta expressió en determinats apartats on no s'utilitzava de manera correcta.
- Canvi en els objectius: S'ha reformulat la secció d'objectius per especificar millor les metes del projecte.
- Clarificació sobre els embeddings: S'ha aclarit que la decisió d'unir els descriptors radiològics en un únic embedding, en comptes de tenir un per descriptor, va ser una decisió pròpia del projecte.
- Reanomenament de "Data Exploration" a "Dataset": S'ha canviat el nom de l'apartat per fer-lo més clar i coherent amb els objectius del projecte.
- Afegir segon paràgraf de "Experimental Design" a "Dataset": S'ha afegit un segon paràgraf de la secció de "Experimental Design" a "Dataset", per clarificar la metodologia utilitzada en la selecció dels descriptors.
- Explicar què indiquen els descriptors en el dataset: S'ha afegit una explicació detallada de què representen i com es defineixen els descriptors utilitzats al dataset.
- Substituir gràfic de "vinfiltration" de la distribució: S'ha substituït el gràfic per un de més clar i actualitzat, que mostra millor la distribució de les dades.
- Explicar el càlcul de la prova del Chi-quadrat: S'ha afegit una explicació sobre com es calcula la prova del Chi-quadrat i la seva aplicació als resultats.
- 6. Development: S'ha explicat exhaustivament la projecció de 1024 dimensions per després passar-ho per una capa lineal + ReLU, i la normalització a 512 dimensions per al processament de les dades.

- Explicar Triplet Margin Loss: S'ha afegit una explicació detallada de com funciona el Triplet Margin Loss i la seva aplicació en el model.
- Eliminar títol "Training": S'ha eliminat el títol "Training" per millorar la coherència estructural i simplificar la presentació.
- Corregir l'ús de la distància Euclidiana: S'ha corregit l'ús de la distància Euclidiana en determinats punts per ajustar-se millor al context tècnic de les operacions realitzades.
- Canviar "top-1 prediction is used": S'ha corregit l'explicació sobre com s'utilitza la predicció top-1, ajustant-la per a una millor comprensió.
- Eliminar figura de "top-5 most similar texts": S'ha eliminat aquesta figura ja que no aportava informació rellevant als resultats obtinguts.
- 7. Experimental design: S'ha detallat millor el propòsit dels experiments, explicant la importància de cada un i les mètriques associades:
 - Experiment 1: Impacte del top-k en l'evaluació.
 - Experiment 2: Anàlisi del Triplet Loss.
 - Experiment 3: Comparativa entre models preentrenats.
 - Experiment 4: Mecanisme d'atenció aplicat en les pèrdues.
- Taula dels hiperparàmetres: S'ha afegit una taula que resumeix tots els hiperparàmetres utilitzats en els experiments.

8. Resultats

- Taula de mètriques: S'ha actualitzat la taula de mètriques a la qual es fa referència, substituint els boxplots per una presentació més clara de les dades amb mitjanes i desviacions estàndard. Les files representen els extractors i les columnes els descriptors, amb les cel·les indicant la mitjana i la desviació estàndard.
- Mètriques de test i train: S'han afegit les mètriques de test i train per cada experiment, per comparar els resultats entre les dues fases.

- Pèrdues de les tres configuracions: S'ha afegit una taula de les pèrdues obtingudes per a les tres configuracions d'entrenament utilitzades.
- Explicar per què es tria ResNet152: S'ha justificat l'elecció de ResNet152 com l'extractor final, ja que ofereix el millor rendiment en top-1, top-3 i top-5, mostrant els resultats corresponents i les matrius de confusió dels entrenaments.
- Matrius de confusió de cada descriptor: S'han afegit totes les matrius de confusió dels descriptors, detallant la capacitat del model per predir correctament en cada cas.

8.2 Experiment d'anàlisi de top-k

- S'ha afegit l'explicació i els resultats de l'experiment relacionat amb l'impacte del valor top-k en l'avaluació del model.

8.3 Experiment d'entrenament de Triplet Loss

- S'ha inclòs el detall de l'experiment dedicat a l'entrenament del Triplet Loss, explicant com afecta al rendiment global del model.

8.4 Experiment amb característiques d'imatge

- S'ha afegit un experiment dedicat a l'anàlisi de característiques d'imatge per estudiar l'impacte de les tècniques d'atenció sobre les representacions visuals.