

## **Notas Reunión 22/03/2022**

Presentes en la reunión: Prof. Benjamín Bustos, Prof. Iván Sipirán, Matías Vergara

### **Trabajo realizado entre 15/03/2022 y 22/03/2022**

#### **- Tratamiento del desbalance de clases en Binary Relevance**

- Se entrenan clasificadores para cada etiqueta con un conjunto balanceado de datos, con distintas particiones de train y test.
- Se incorpora la métrica de **Recall** en el estudio, como una medida para ver qué tan bien "capturamos" los casos positivos.
- A diferencia del escenario previo, ahora logramos buena cobertura de casos realmente positivos (Recall macro de 0.84). Sin embargo, se presenta el clásico trade-off precision/recall: incorporamos falsos positivos. Se considera que esto puede deberse a que no mostramos todos los ejemplos negativos en entrenamiento.

#### **- Estudio de literatura al respecto**

- No se encuentran publicaciones en dónde se adopte el enfoque de podar el problema con cierto threshold de ocurrencia sobre las etiquetas.
- Sin embargo, se encuentran dos publicaciones que podrían ser de interés, y se comienzan a estudiar. Estas son: Integration of deep learning model and feature selection for multi-label classification (2022) y Towards Class-Imbalance Aware Multi-Label Learning (2020).

#### **- Reunión con prof. Iván Sipirán para discutir sobre features.**

- En reunión con el prof. Iván, nos percatamos de que los features con los que se trabaja son, efectivamente, de la mejor red encontrada, con las optimizaciones aplicadas.

### **Trabajo a realizar entre 23/03/2022 y 30/03/2022**

#### **- Aplicar Data Augmentation**

- Probar distintos enfoques: transformaciones simples (rotar, cambiar tamaño), teniendo cuidado de respetar las etiquetas como "horizontal" o "vertical", y otras mas agresivas, tales como el recorte de zonas aleatorias de un patrón, asociando cada recorte a las mismas etiquetas del patrón original.

#### **- Estudiar papers adjuntos**

- Continuar la lectura de los papers adjuntos.

#### **- Corroborar si la red se equivoca siempre con los mismos ejemplos.**

- De confirmarse este fenómeno, explicaría porqué la Hamming Loss aumenta a medida que se "simplifica" el problema.

#### **- Evaluar maneras de optimizar la red \***

- Buscar arquitecturas que se adapten de mejor manera al problema actual, considerando que la red que estamos usando no fue entrenada para este problema si no para una clasificación mucho más "gruesa" (y además multiclase, no multilabel).

\*: Puede que postergue esta tarea en función del tiempo para priorizar las otras.