

Notas Reunión 15/03/2022

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

Trabajo realizado entre 09/03/2022 y 15/03/2022

- Repetición de experimentos con nuevas etiquetas

- Se discuten los resultados obtenidos tras normalizar las etiquetas.
- Métricas tales como *label density* mejoran considerablemente, más no hay un punto de referencia (no conocemos un nivel "aceptable" o "insuficiente").
- Se comenta la necesidad de tratar el desbalance de casos por etiqueta, así como la posibilidad de estar entendiendo mal el problema (ejemplo de la lotería).
- Cualquier decisión a tomar debe ser justificada por la literatura al respecto.

- Revisión de conceptos en base a clases de NLP (curso del prof. Felipe Bravo)

- El término correcto para el tratamiento realizado a las etiquetas es **normalización**: el proceso de crear clases de equivalencia entre distintos tipos (entradas).
- Se validan los procesos de Stopwords Removal y Lemmatization realizados como enfoques frecuentes en NLP. La librería utilizada, NLTK, es también parte del stack usual para tratar estos problemas.

Trabajo a realizar entre 15/03/2022 y 23/03/2022

- Reunión con equipo el 16/03/2022

- Comentar brevemente el trabajo realizado, donde nos encontramos (post normalización de etiquetas) y hacia dónde vamos (desbalance de clases).

- Tratar el desbalance de clases

- Buscar literatura al respecto para entender el porqué de ciertas cosas. En particular, porqué el Hamming Score aumenta al podar el problema.
- Aplicar un enfoque de construcción "manual" de conjuntos de entrenamiento para asegurar que cada clasificador base reciba un set balanceado entre casos positivos y negativos.
- Repetir experimentos para medir el impacto de dicho tratamiento sobre los resultados previos.
- Incorporar True Positive Rate (TPR) dentro de las métricas. Calcular por etiqueta y average.

- Agendar reunión con prof. Iván Sipirán

- Ya se encuentra en Chile, por lo que debería estar disponible para una reunión. Tratar el tema de las optimizaciones encontradas por él.

Observación: estimo que el tratamiento del desbalance de clases podría extenderse a más de una semana de trabajo.