

Tarea 2

16 de abril de 2019
Alejandro Quiñones

1. Motivación

Se pueden reconocer distintas motivaciones para esta tarea. En primer lugar, se entrega un ejercicio introductorio al preprocesamiento de cualquier dataset: la necesidad de ordenar la información de manera que sea compatible con librerías y con sus divisiones necesarias para el testeo. Además, se presenta el uso de dos herramientas (LBP y SFS) en un caso de estudio bastante realístico (salvo las buenas condiciones de las imágenes), donde es fácil de pensar en circunstancias de seguridad donde esto sería útil.

2. Solución propuesta

De forma similar a cómo se vio en clases, los pasos a realizar fueron ejecutados en cierto orden lógico que suelen entregar mejores resultados. Se inicia obteniendo las features LBP de las imágenes solicitadas usando ciertas constantes $hdiv = 2$ y $vdiv = 2$. Luego, se divide el dataset según la partición solicitada en el enunciado. Paso siguiente, y solo con el test de entrenamiento, se obtienen los features después del *clean*, y de esos los resultantes después del SFS. Por último, se filtran las columnas eliminadas en el set de training. Una vez obtenidos los filtros, los mismos fueron aplicados en la parte 2. En el informe se entrega adjuntado el código que se utilizó para obtener los resultados, además de un jupyter notebook, que provee lo mismo de una forma más ordenada. Además, se entregan los valores obtenidos de la selección de features a través de ambos *clean* y SFS, junto a una función que permite probar dicha configuración sin la necesidad de realizar la costosa selección.

3. Experimentos realizados

La siguiente tabla resume los experimentos realizados

<i>hdiv</i>	<i>vdiv</i>	Tiempo de ejecución	Accuracy 1	Accuracy 2
2	2	9:37	94 %	88 %
2	3	15:10	86 %	80 %

En particular destaca la ejecución con $hdiv = 2$ y $vdiv = 2$, que provee mejores resultados en un tiempo de ejecución razonable.

4. Conclusión

Es fundamental destacar la importancia del trade-off entre mejores resultados a través de una mejor selección de features, y el tiempo que requiere la selección de features. Además, se concluye que el algoritmo es bastante efectivo en su misión de reconocer imágenes, obteniendo buenos resultados de accuracy incluso en un set completamente nuevo. De esto último se desprende que las características seleccionadas son intrínsecas de caras de las personas, y no parte del sesgo del dataset.