

Clasificación de razas de perros^{*}

19 de julio de 2018

Descripción del proyecto: La identificación de razas de perros es una tarea difícil. En ocasiones, incluso expertos no son capaces de identificar la raza de un perro simplemente mirándolo [1]. Sumado a esto, la mayoría de los perros son en realidad de raza mixta; es decir, son una combinación de dos o más razas. La determinación de las razas en un perro de raza mixta mediante inspección visual es casi imposible para los humanos. Por ejemplo, la Figura 1 muestra un border collie y un beagle de raza pura junto con una mezcla border collie-beagle. Note que la raza mixta no se parece en nada a un beagle.



(a) Border Collie



(b) Beagle



(c) Mezcla Border Collie - Beagle

Figura 1: Un ejemplo de cómo un perro de raza mixta puede verse muy diferente a sus razas constituyentes.

^{*}Proyecto formulado por el Profesor Pavlos Protopapas

La identificación de las razas de perros no es un problema simplemente académico. De hecho, la identificación correcta de las razas tiene implicaciones en las políticas públicas y las estrategias de adopción de los refugios para perros [2]. Por ejemplo, una familia activa puede querer adoptar un perro que los acompañe en caminatas y otras actividades al aire libre. Otra familia puede, en cambio, querer un perro que se acurruque con ellos en el sofá. Aunque cada perro tiene su propia personalidad, su raza puede dar indicios de sus características. Además, la raza de un perro es también un gran indicador de lo grande que será cuando esté completamente desarrollado. El tamaño de un perro puede ser un factor determinante en la adopción.

Así, el problema de identificar la raza de un perro es relevante y al mismo tiempo, es un desafío interesante para ser abordado con técnicas de machine learning. En este proyecto, se busca clasificar las razas de perros usando un conjunto de datos que consta de más de 20 mil imágenes que involucran a 120 razas diferentes.

Objetivo general: proponer, construir y evaluar modelos para clasificar razas de perros. Por facilidad, no se considerarán razas mixtas en este proyecto.

Bases de datos: Los datos que se usarán en este proyecto provienen de la base de datos *Stanford Dogs Dataset*: <http://vision.stanford.edu/aditya86/ImageNetDogs/main.html>.

Objetivos específicos: .

1. Construir algunos modelos para clasificar las razas de perros.
2. Evaluar la calidad predictiva de los modelos.
3. Comparar los resultados de cada modelo.
4. Usar su modelo para detectar valores atípicos (*outliers*).

Cabe resaltar que este proyecto presenta varios retos:

1. Se debe realizar un pre-procesamiento significativo de las imágenes. No todas las imágenes son del mismo tamaño. La mayoría son de color, pero algunas son en blanco y negro. Las etiquetas y la estructura del directorio deberán limpiarse y modificarse para que su modelo pueda procesar fácilmente los datos.
2. Aunque la base de datos consta de 20.580 imágenes, también incluye 120 razas. Esto significa que en realidad no hay muchos datos para el proceso de aprendizaje puesto que solo hay alrededor de 150-200 imágenes por raza de perro. Es posible que deba considerar técnicas para aumentar los datos. También puede pensar en trabajar con superclases en lugar de razas. Una superclase puede definirse como un conjunto de razas de perros. Por ejemplo, border collies, blue heelers y Australian Shepherds son ejemplos de razas de pastoreo. El American Kennel Club (AKC) ha agrupado razas de perros en siete grupos distintos (ver AKC Dog Breed Groups). Una buena descripción de cada grupo se puede encontrar en el sitio web de National Dog Show: The Seven Groups. Es posible que deba considerar el uso de los siete grupos como superclases.

3. Tendrá que pensar en una buena forma de hacer una detección de valores atípicos. Consideramos que una raza de perro es atípica si es significativamente diferente de todas las demás razas. Por ejemplo, un perro pomerano luce completamente diferente a los border collies y shephards australianos. Por lo tanto, consideramos que un pomerano es un valor atípico. Una posibilidad de evaluar la detección de valores atípicos es “ocultar” algunas razas durante el proceso de aprendizaje y verificar si el modelo las identifica como valores atípicos.

Referencias

- [1] K. R. Olson, J. Levy, B. Norby, M. Crandall, J. Broadhurst, S. Jacks, R. Barton, and M. Zimmerman, “Inconsistent identification of pit bull-type dogs by shelter staff,” *The Veterinary Journal*, vol. 206, no. 2, pp. 197–202, 2015.
 - [2] T. King, L. C. Marston, and P. C. Bennett, “Breeding dogs for beauty and behaviour: Why scientists need to do more to develop valid and reliable behaviour assessments for dogs kept as companions,” *Applied Animal Behaviour Science*, vol. 137, no. 1, pp. 1–12, 2012.
-