

Miniproyecto 3: Clasificación de Animales

Silvia Estefanía Jiménez Pilonieta
Universidad de los Andes
Bogotá, Colombia
se.jimenezp@uniandes.edu.co

Andrés Serrano Conti
Universidad de los Andes
Bogotá, Colombia
a.serranoc@uniandes.edu.co

Jacobo Valderrama Rovira
Universidad de los Andes
Bogotá, Colombia
j.valderramar@uniandes.edu.co

1. Introducción

Para esta entrega se va a trabajar sobre la clasificación de animales a partir del aprendizaje supervisado. A diferencia del aprendizaje no supervisado en el que únicamente se tienen bases de datos, en el supervisado estos datos también vienen acompañados de unas etiquetas hechas por seres humanos las cuales ayudan a identificar cada objeto y que hacen que el algoritmo haya analizado dichos datos y respectivas etiquetas para poder generalizar los datos[1]. Como se mencionó anteriormente, la tarea de clasificación no consiste únicamente en categorizar imágenes de acuerdo a una determinada etiqueta sino que también pretende que se puedan utilizar características de cada imagen(recortes) de modo que se puedan diferenciar. La clasificación de animales es uno de los muchos problemas que permite abordar este tipo de problemas lo que puede llevar a identificar diferentes especies de estos animales, sin embargo, esto se puede extrapolar a tareas en otros campos.

En la actualidad existen múltiples métodos de clasificación manual y automática. Entre los mas utilizados están los siguientes métodos:regresión logística, Naive Bayes, K-vecino, arboles de decisión, bosques aleatorios, redes neuronales artificiales y SVM(maquinas de soporte vectorial). Ahora bien, es importante mencionar que todos los métodos tienen sus ventajas y desventajas por lo que es difícil determinar cual es el mejor. En el caso de los algoritmos automáticos,K-vecinos y las maquinas de soporte vectorial se caracterizan por ser buenos y no tan complicados de aplicar por lo que son ampliamente utilizados, mientras que las redes neuronales artificiales son altamente eficientes pero requieren de muchos recursos para funcionar[3][5]. En el caso de los métodos mencionados anteriormente, se encontró que los arboles de decisión son utilizados de forma manual mas frecuentemente debido a que no requieren de

una previa preparación de los datos lo que lo hace efectivo para realizarse sin la necesidad de utilizar algoritmos complejos[3][4].

Actualmente hay muchas implementaciones del problema de clasificación supervisada. Hay muchas patologías médicas en las que si no se hace un diagnóstico temprano esto puede llevar a que la salud del paciente se vea afectada hasta el punto de causarle la muerte.En el año 2022 un grupo de investigadores utilizó un algoritmo de clasificación supervisada con el propósito de identificar si una persona tenía hígado graso o no[2]. En el caso anterior se utilizó un problema de clasificación binaria en el que se pretendía ver si un hígado tenía la enfermedad o no mas no en que etapa o que tan avanzado estaba el problema como seria en otro tipo de clasificación.

Por último, es importante resaltar que es necesario usar análisis de imágenes en esta tarea, puesto que para poder clasificar elementos en distintas clases se deben generar etiquetas y correlaciones de las imágenes con el predictor, así entrenar el modelo correctamente y detectar el elemento con la mejor precisión posible, en este caso detectar al animal correctamente.

2. Base de datos

La base de datos de este problema se compone de 11.236 imágenes repartidas en las carpetas de entrenamiento, validación y prueba de la siguiente forma: 2.759 imágenes de prueba, 5.610 de entrenamiento y 2.867 de validación. Dentro de cada una de las tres carpetas se encuentran otras 5 carpetas correspondientes a las clases caballo, oveja, elefante, gato y perro. Cada una de las imágenes se encuentra en formato .jpeg y presentan variación en los tamaños por lo que al inicio del proceso se corrige el tamaño para que cada una quede con tamaño 300 x 300 píxeles, la distribución de estas imágenes por clases en entrenamiento es de 9.46%

para elefantes, 23.56% para caballos, 11.42% para gatos, 12.55% para perros y 43% para ovejas.

3. Abordaje

3.1. Descriptores

Para solucionar este problema de clasificación se utilizarán 3 tipos de descriptores, textura, color y forma. La textura se basa en patrones que se repiten espacialmente en una ventana, brindándonos información acerca del material de los objetos y permitiéndonos diferenciar objetos con una representación de color o forma similares. El elemento fundamental de la textura es el textón, los cuales, al contenerlos desde una ventana nos permiten obtener descriptores de textura, ya sea un histograma de textones donde se representa la frecuencia o probabilidad de cada textón en la ventana, o un mapa de textones donde se asigna un color a cada textón del grupo de textones dado y esto se presenta dentro de la imagen. Para generar estos descriptores se realiza un diccionario de textones, donde primero se crea un banco de filtros con patrones sencillos que busque estructuras como puntos, bordes o barras.

Luego se halla la respuesta de la imagen a cada uno de los filtros para hallar diferentes patrones dentro de la imagen, al haber usado todos los filtros se pueden representar cada uno de los píxeles como una respuesta a todos los filtros del banco, por lo que cada píxel estará representado por un vector tamaño N que al juntarlos todos genera una matriz $H \times W$ vectores de tamaño N . Seguido a esto agrupamos las representaciones en clusters, por lo que se hace uso de Kmeans, con un K arbitrario, en el espacio de representación, para finalmente hallar los centroides de cada cluster, obteniendo un textón por cada cluster, es decir obtendremos la misma cantidad de textones que el valor de K . En este caso para nuestro descriptor de textura realizamos el diccionario de textones para generar el mapa de textones y a partir de este mapa crear el histograma de textones que será nuestro descriptor final y nos permitirá diferenciar las imágenes dado la diferencia en la textura del pelaje de las distintas clases.

En segunda instancia tenemos los descriptores de color, en este caso usaremos como descriptor los histogramas de color, en los cuales se modela la distribución de intensidades o color de una imagen o región de interés. Por lo tanto, si dos imágenes tienen histogramas similares, estas son similares o tienen distribución de colores similares. Los histogramas de intensidades pueden ser conjuntos o concatenados. Los histogramas conjuntos se generan asignando un canal de color a cada eje y luego hallando las intensidad de cada canal por cada píxel de la imagen para finalmente señalar la distribución de intensidades entre los 3 ejes. Por otro lado, en el histograma concatenado se concatenan los histogramas marginales de cada canal, para

obtener un solo histograma que se puede comparar entre imágenes. Este descriptor es de utilidad para la clasificación de las distintas clases pues en algunas de estas los animales tienen colores de pelaje muy específicos, aún así si existen dos imágenes con distribución de color similar en diferentes clases podría clasificarse erróneamente.

Por último, el descriptor de forma se utiliza para analizar la forma de los objetos y sus contornos, considerando los gradientes orientados de la ventana de una imagen, este descriptor es especialmente importante porque es el único que toma en cuenta la posición espacial, en esta caso utilizaremos el histograma de gradiente orientado (HOG), por lo que primero se comprime la imagen para luego hacer cross-correlación con los filtros y así calcular la magnitud y dirección del gradiente, tomando para cada píxel el canal de color con mayor magnitud como el gradiente final. Seguido a esto, dividimos la imagen en celdas de 8×8 ajustando la dirección de cada píxel a una de las orientaciones del gradiente y a partir de esto construimos el histograma. Se debe tener en cuenta que el histograma se normaliza en bloques, para esto, se toma un número definido de celdas y normalizar los histogramas en bloques de las celdas que se sobrelapan. Finalmente se concatenan los histogramas normalizados en el descriptor, generando un mapa de gradientes de la imagen. Este descriptor, es especialmente importante en este problema de clasificación ya que nos permite evaluar la forma específica de cada clase, permitiendo diferenciar con mayor facilidad a que clase pertenece cada imagen.

3.2. Baseline

El método de K vecinos mas cercano o también denominado KNN es un método del aprendizaje supervisado utilizado para la clasificación de puntos basado en la cercanía a partir de un punto de referencia. Este tipo de método se basa en el uso de diferentes métricas de distancia y de este modo la similitud entre puntos con características parecidas[10]. En este caso utilizamos el método de vecino más cercano en una clase que contiene 3 funciones que en orden inicializan, entrenan y predicen el clasificador. En la inicialización, como su nombre lo indica inicializa los vectores del descriptor y las etiquetas, en la parte de entrenamiento se cargan las imágenes y etiquetas destinadas al entrenamiento del modelo y por último, en la predicción se genera el recorrido donde se evalúa la distancia que hay entre imagen con todas las imágenes de entrenamiento. Posteriormente a la imagen que entra se le asignará como etiqueta la etiqueta de la imagen de entrenamiento con la que tenga la menor distancia posible.

References

- [1] "Supervised vs. Unsupervised Learning: What's the Difference?," IBM.com, Mar. 12, 2021.

<https://www.ibm.com/cloud/blog/supervised-vs-unsupervised-learning> (accessed Nov. 02, 2022).

Realizado en L^AT_EX

- [2] A. Gaber, H. A. Youness, A. Hamdy, H. M. Abdelaal, and A. M. Hassan, "Automatic Classification of Fatty Liver Disease Based on Supervised Learning and Genetic Algorithm," *Applied Sciences*, vol. 12, no. 1, p. 521, Jan. 2022, doi: 10.3390/app12010521.
- [3] M. Waseem, "Python Machine Learning Certification Training," *Edureka*, Dec. 04, 2019. <https://www.edureka.co/blog/classification-in-machine-learning/#tree> (accessed Nov. 03, 2022).
- [4] R. Schapire, "Machine Learning Algorithms for Classification Machine Learning Algorithms for Classification Machine Learning Algorithms for Classification Machine Learning Algorithms for Classification." [Online]. Available: http://protocols.netlab.uky.edu/liuj/teaching/CS685_s17/CS485-2016/CS485g/notes/Reference/picasso-minicourse-classification.pdf
- [5] Z. Keita, "Classification in Machine Learning: An Introduction," *Datacamp.com*, Sep. 21, 2022. <https://www.datacamp.com/blog/classification-machine-learning> (accessed Nov. 03, 2022).
- [6] Kumar N, Nachamai M.Noise Removal and Filtering Techniques Used in Medical Images. *Orient.J. Comp. Sci. and Technol*;10(1). Available from: <http://www.computerscijournal.org/?p=4800>
- [7] Ortiz, N. (2020, 3 marzo). Clasificación con datos desbalanceados. *Aprende Machine Learning*. Recuperado 20 de septiembre de 2022, de <https://www.aprendemachinelarning.com/clasificacion-con-datos-desbalanceados/>
- [8] El concepto y principio de segmentación semántica de imágenes y métodos comunes. - programador clic. (s. f.). Recuperado 27 de septiembre de 2022, de <https://programmerclick.com/article/71061394999/>
- [9] El concepto y principio de segmentación semántica de imágenes y métodos comunes. - programador clic. (s. f.). Recuperado 27 de septiembre de 2022, de <https://programmerclick.com/article/71061394999/>
- [10] "¿Qué es el algoritmo de k vecinos más cercanos? — IBM," *Ibm.com*, 2022. <https://www.ibm.com/cookies/topics/knn#:~:text=El%20algoritmo%20de%20k%20vecinos%20m%C3%A1s%20ceranos%2C%20tambi%C3%A9n%20conocido%20como,un%20> (accessed Nov. 14, 2022).
- [11] la, "Descriptores," *Genial.ly*, 2022. <https://view.genial.ly/634ecef76bd6300119fd316/interactive-content-descriptores> (accessed Nov. 18, 2022).