

ANÁLISIS DE DATOS DE ASTRONOMÍA USANDO ALGORITMOS DE DEEP LEARNING Y ESTADÍSTICOS

Vargas Barrera Jorge Luis (1), Jeyakumar Solai (2)

¹ Licenciatura en computación, Universidad de Guanajuato | Dirección de correo electrónico: jorge.vargas@cimat.mx

² Departamento de Astronomía, División de Ciencias Naturales y Exactas, Campus Guanajuato, Universidad de Guanajuato | Dirección de correo electrónico: sjk@astro.ugto.mx

Resumen

Se presentan resultados de la aplicación de técnicas de deep learning al problema de identificación de radio fuentes gigantes (GRSs) presentes en el catálogo NVSS. Se hizo un pre procesamiento de las imágenes para la detección de estas fuentes y posteriormente realizar su clasificación, para esto se implementó una red neuronal convolucional basada en el modelo VGG la cual se entrenó con el catálogo NVSS. El tamaño del conjunto de entrenamiento fue de 1616 imágenes astronómicas distribuidas en 16 categorías. De nuestro análisis, encontramos que nuestro método es efectivo con una precisión de 87%.

Abstract

Results of the application of deep learning techniques to the problem of identifying giant radio sources (GRSs) from the data in the NVSS catalog are presented. A preprocessing of the images was done for the detection of the sources followed by their classification, for this a convolutional neural network was implemented based on the VGG model and then it was trained with the catalog NVSS. The size of the training set was of 1616 astronomical images distributed in 16 categories. In our analysis we found that our method is effective with a precision of 87%.

Palabras Clave

bases de datos astronómicas; galaxias; redes neuronales; deep learning; convolución

INTRODUCCIÓN

Deep Learning

Deep Learning permite modelos computacionales que están compuestos de múltiples capas de procesamiento aprender representaciones de datos con múltiples niveles de abstracción. Estos métodos han mejorado drásticamente el estado del arte en reconocimiento de voz, reconocimiento visual de objetos, detección de objetos y muchos campos más como genética y descubrimiento de drogas.

Building Blocks: Artificial Neural Networks (ANN)

En machine learning y ciencias cognitivas, una red neuronal artificial (ANN) es una red inspirada por las redes neuronales biológicas que son usadas para estimar o aproximar funciones que pueden depender en un gran número de entradas que son generalmente desconocidas. Una (ANN) está constituida de nodos (neuronas) apilados en capas entre el vector de características y el vector objetivo. Un nodo en una red neuronal se construye de pesos y funciones de activación.

El perceptrón es un algoritmo para aprendizaje supervisado de clasificación binaria. Es decir puede decidir si la entrada (representado por un vector de números) pertenece a una clase u otra. Muy similar a regresión logística, los pesos en una red neuronal son multiplicados por el vector de entrada sumados y alimentados al input de la función. Una red perceptrón puede ser diseñada para tener múltiples capas, llevándonos a un perceptrón de multicapa (MLP).

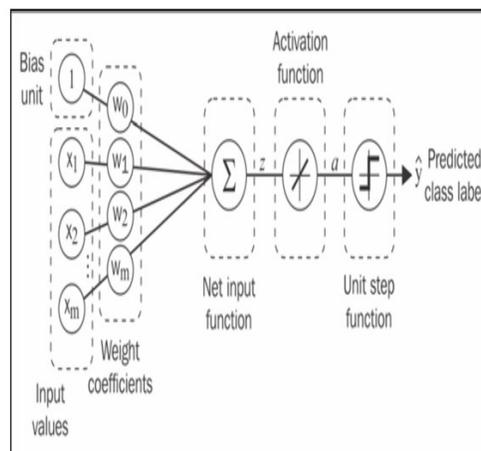


IMAGEN 1: Red neuronal de una sola capa.

Convolutional Neural Network

Una red neuronal convolucional (CNN) es un tipo de ANN del tipo feed-forward en donde el patrón de conectividad entre las neuronas es inspirado por la organización de la corteza visual de los animales. La red consiste de múltiples capas de pequeñas colecciones de neuronas las cuales procesan partes de la imagen, llamados campos receptivos. La salida de estas colecciones son acomodados tal que las regiones de entrada se superpongan, para obtener una mejor representación de la imagen original; esto se repite para cada capa.

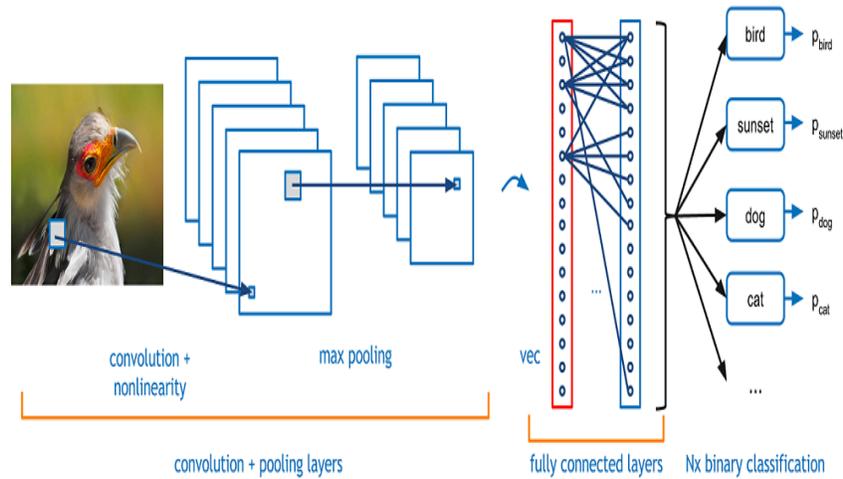


IMAGEN 2: Red neuronal convolucional.

Detección de fuentes en imágenes astronómicas.

Las galaxias pueden ser observadas a diferentes longitudes de onda, cada una de estas longitudes revela diferentes morfologías y características. Las galaxias que contienen un núcleo galáctico activo (AGN) son usualmente poderosas radio fuentes que producen jets y extensas regiones radio emitidas conocidas como lóbulos. Estas regiones son enormes y pueden variar de 100 pc hasta pocos Mpc en extensión lineal.

El primer paso para la clasificación de fuentes es hacer una búsqueda de fuentes. Esto básicamente ajustará Gaussianas a diferentes máximos locales y extraerá varios parámetros de ajuste tales como los parámetros de la gaussiana y su ángulo de posición. Las fuentes serán clasificadas siguiendo la clasificación mostrada en [1] y que se puede observar en la tabla 1. A fin de verificar que las fuentes estén conectadas o no, los contornos son dibujados encima de las imágenes. Con los futuras bases de datos, el volumen de datos/imágenes que se procesara será enorme. La clasificación manual será una tarea tediosa sino imposible. Un acercamiento con Deep Learning (DL) puede ser una herramienta vital para su clasificación. La meta del trabajo presentado aquí muestra que técnicas DL son eficientes y robustas. DL puede ser puesto en uso para analizar grandes cantidades de datos que resultaran de futuros telescopios.

Tabla 1: Clasificación de fuentes

Código	Clasificación o descripción	Código	Clasificación o descripción
1. C	Core	9. L	Lobe
2. CJ	Core jet	10. LCL	Lobe-core-lobe
3. CL	Core lobe	11. LJ	Lobe jet
4. CLU	Cluster	12. LL	Connected double
5. DL	Diffuse lobe	13. RL	Relic lobe
6. EC	Elliptical core	14. RLL	Relic lobe connected
7. JCJ	Jet core jet	15. WLCL	Wide angle lobe-core-lobe connected
8. JL	Jet lobe	16. X	X-shape

MATERIALES Y MÉTODOS

El desarrollo de este proyecto consistió en la aplicación de los siguientes métodos de procesamiento de imágenes, machine learning y deep learning aplicados a una base de datos de 1616 imágenes astronómicas todo esto se programó en Python haciendo uso de Keras y Tensorflow.

Se utilizó el método de threshold para la detección de fuentes en una imagen, donde las fuentes detectadas deben tener un número mínimo de píxeles conectados tal que cada uno es mayor que un valor específico en la imagen. El límite es usualmente definido como un múltiplo de la desviación estándar del fondo. Además la imagen se filtró antes del thresholding para suavizar el ruido y maximizar la detección de objetos con forma similar al kernel del filtro.

Se pre procesaron las imágenes que contienen a las fuentes en particular se utilizó el método de masking como se observa en la imagen 1 que consiste en solo mostrar una fuente por imagen, esto se realiza para cada fuente.

Antes de introducir nuestras imágenes dentro del algoritmo de la red neuronal, se realizó una visualización del espacio de características incrustado en dos dimensiones haciendo uso de la técnica isomap para podernos dar una idea de que como se distribuyen nuestros datos.

Se utilizó una red neuronal convolucional basado en el modelo de redes neuronales convolucionales conocida como VGGNet [2], las arquitecturas basadas en este modelo se caracterizan por: usar solamente capas convolucionales de 3x3 apiladas conforme aumenta la profundidad, reduce volumen haciendo max-pooling y las capas conectadas completamente al final de la red neuronal dan prioridad a un clasificador softmax.

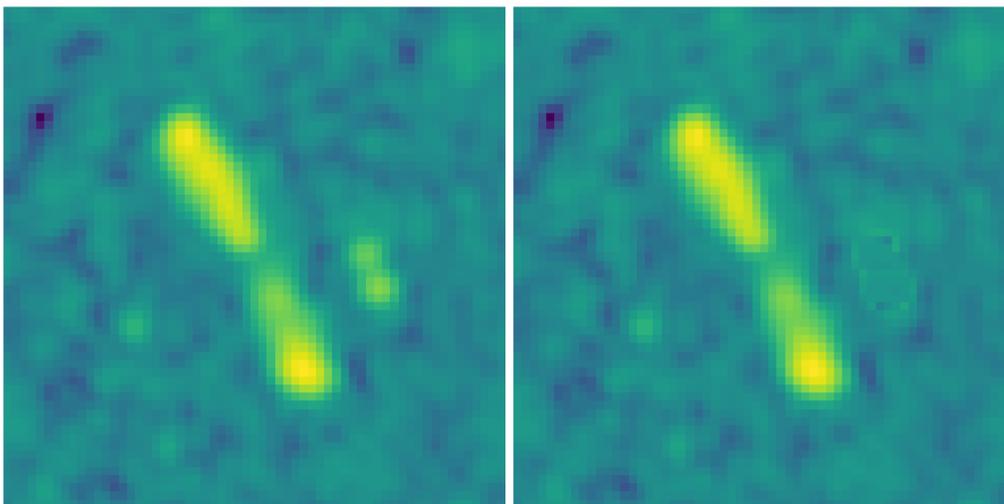


IMAGEN 3: A la izquierda se muestra las diferentes fuentes, a la derecha la misma imagen haciendo masking de una fuente.

RESULTADOS Y DISCUSIÓN

Después de que la red neuronal realizó 80 iteraciones al conjunto de prueba, se obtuvieron los resultados que se muestran en la figura. Como lo muestra la imagen 4 se logró una baja pérdida y un sobreajuste limitado.

Después se procedió a predecir un conjunto de 100 imágenes, lográndolo el resultado correcto en 90 de las imágenes con una certeza promedio de 87% para cada imagen.

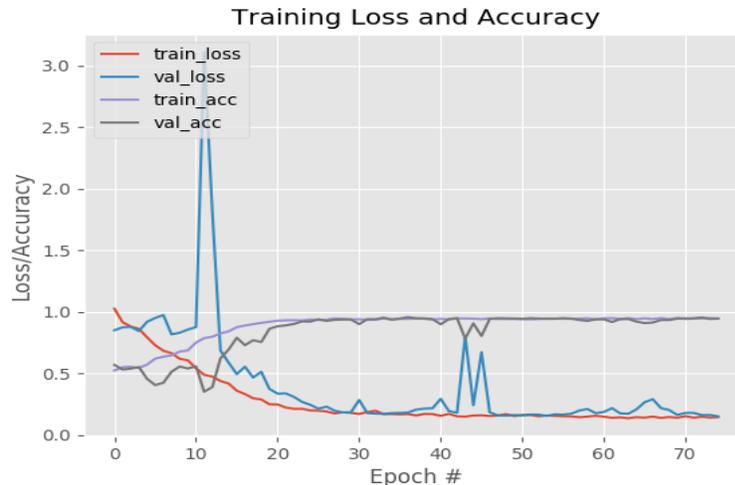


IMAGEN 4: Perdida/precisión del entrenamiento y la validación para el clasificador de fuentes con Keras.

CONCLUSIONES

Los resultados obtenidos fueron satisfactorios, pues se obtienen buenos porcentajes de predicción, sin embargo las diferentes etapas del método pueden mejorarse, en la detección de fuentes se puede explorar otro método, incluso uno basado en una red neuronal, un extensión al proyecto que no se exploró del todo fue la creación de otra red neuronal para el procesamiento de más información relacionada con la fuente y usando la información de ambas redes neuronales formar una mejor predicción. Además con más imágenes de entrenamiento se puede obtener una mejor precisión. Aun cuando no se obtuvieron resultados cercanos a los ya reportados en trabajos similares si se atienden los puntos mencionados anteriormente se puede obtener mayor precisión.

REFERENCIAS

1. Proctor, D. D. (2016). A selection of giant radio sources from NVSS. *The Astrophysical Journal Supplement Series*, 224(18), 13.
2. Simonyan K. & Zisserman A. (2015). Very deep convolutional networks for large-scale image recognition. Published as a conference paper at ICLR 2015, 13.
3. Bastien D. (2017). Classifying bent radio galaxies from a mixture of point-like/extended images with Machine Learning. *Conf. Ser.: Mater. Sci. Eng.* 198 012013
4. Rosebrock A. (2018). Keras and Convolutional Neural Networks. Recuperado en <https://www.pyimagesearch.com/2018/04/16/keras-and-convolutional-neural-networks-cnns/>
5. Maggio V. (2017). Introduction to Deep Neural Networks with Keras and Tensorflow. Recuperado en <https://github.com/leriomaggio/deep-learning-keras-tensorflow>