Image captioning con redes neuronales convolucionales y recurrentes para reconocimiento de arácnidos y su importancia médica

Eddson Sierra
Instituto de Investigación de Operaciones
Universidad Galileo
eddson.sierra@galileo.edu

Abstract

Existen más de 102,000 especies de arácnidos de las cuales solo un pequeño porcentaje corresponde a especies de importancia médica y/o epidemiológica. La gran diversidad de especies y su particular distribución geográfica son factores que dificultan su reconocimiento e identificación, trabajo que recae en profesionales aracnólogos expertos en el campo. A través de una red neuronal convolucional basada en la arquitectura AlexNet, se desarrolló un modelo capaz de catalogar arácnidos a partir de fotografías, alcanzando hasta un 82% de exactitud en datos de prueba. Este modelo se presenta como parte del proyecto final del curso de Statistical Learning II, de Universidad Galileo; proyecto que también incluye una red neuronal tradicional para detección de churn y un modelo que combina una red convolucional y una red recurrente LSTM para generar una descripción a partir de imágenes de arañas, indicando la especie y la importancia médica.

1 Problema y objetivo

El objetivo de este proyecto es desarrollar tres de los modelos de deep learning más utilizados en la actualidad: una red neuronal feed forward (o multilayer perceptron) sobre datos estructurados, una red neuronal convolucional sobre datos espaciales y una red neuronal recurrente sobre datos secuenciales. La implementación va desde la preparación de los datasets hasta el entrenamiento, validación, evaluación y selección de los modelos.

El primer problema identificado se trata de modelar el churn de clientes, o el ritmo al que los clientes abandonan una empresa. Este es sin duda alguna, uno de los grandes problemas que las empresas enfrentan y que no tienen del todo resuelto. Una de las principales razones de esto, es que los negocios deben ser hábiles en reconocer las señales que indican que un cliente podría retirarse y luego emprender acciones para impedir que eso suceda. Por lo tanto, este problema puede tener dos causas: la poca habilidad de las empresas para detectar los patrones del churn o la falta de acción para impedir que se pierdan los clientes. En el primer caso, un modelo de churn ayuda a identificar las características de los clientes que tienen alta probabilidad de abandonar (por ejemplo en un banco, estas características pueden ser el límite de crédito, la edad, el género, el balance de saldos y el salario del cliente). En este caso, el objetivo es implementar una red neuronal que modele el churn de clientes y permita predecir qué clientes están por retirarse.

Por otro lado, en el mundo de la identificación de especies, existe una oportunidad (tanto interesante académicamente como de provecho para los expertos) de mejorar el reconocimiento de especies de arañas potencialmente peligrosas para los seres humanos. En octubre de 2018, el biólogo mexicano Diego Barrales Alcalá con estudios especializados en arácnidos, creó una cuenta en Twitter llamada Arachno_Cosas con el fin de difundir información sobre los arácnidos y resolver las preguntas que los usuarios le hacen (normalmente enviando una fotografía del arácnido). Por lo tanto, se presenta la

oportunidad de utilizar las imágenes de las arañas enviadas por los usuarios de Twitter y las respuestas del biólogo para entrenar una red neuronal capaz de catalogar una fotografía de un arácnido de acuerdo con su importancia médica.

Adicionalmente, se propone también utilizar el mismo conjunto de datos (imágenes y las respuestas de Twitter) para entrenar una red neuronal recurrente para hacer *image captioning* sobre las imágenes, es decir, generar una descripción de las imágenes que emula las respuestas del experto en la plataforma.

2 Metodología

A continuación, se describe el conjunto de datos y la arquitectura empleada en cada una de las partes del proyecto. En general, cada parte sigue la misma metodología: preparación de datos, definición de arquitectura, parámetros de optimización y métricas de evaluación y finalmente, entrenamiento, validación y selección del mejor modelo. La selección del mejor modelo se ha realizado en los tres casos, mediante comparación de las métricas en el data set de pruebas correspondiente. Las métricas utilizadas fueron: exactitud, precisión, recall y f1 Score.

2.1 Red neuronal feed forward con datos estructurados

Se utilizó el dataset estructurado "Churn Modelling" de Siddharth Dixit. La preparación de los datos consistió principalmente en el tratamiento de variables categóricas resuelto mediante one hot encoding y el escalamiento de los datos. El dataset incluye las siguientes características para 10,000 clientes de un banco: la puntuación de crédito, geografía, género, edad, la tenencia, el balance de saldos, número de productos, el salario estimado, un indicador de si el cliente cuenta con tarjeta de crédito, un indicador de miembro activo y un indicador de si el cliente se retiró o no.

Para la separación de datos se utilizó sckikit learn y la arquitectura de la red neuronal se definió en keras. Se utilizaron dos capas ocultas: una de 512 neuronas y la segunda de 256 neuronas con batch normalization entre las capas y activación ReLU. La capa de salida consiste en una única neurona por ser un problema de clasificación binaria con activación sigmoide. Para el inicializador de los kernels se utilizó un inicializador Xavier normal también llamado inicializador Glorot.

Para los parámetros de optimización se utilizó un optimizador Adam y la pérdida del modelo corresponde a la función de entropía cruzada binaria.

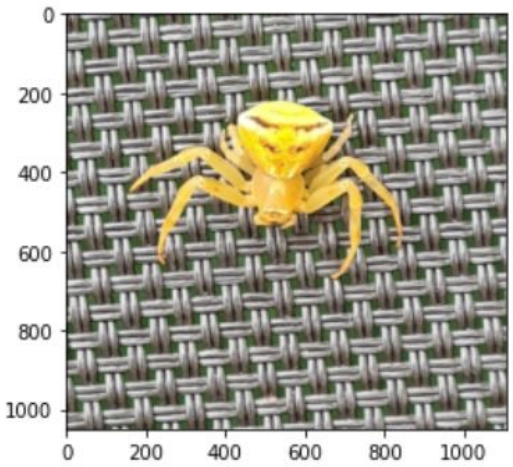

Figure 1: Ejemplo de una imagen en el dataset.

2.2 Red neuronal convolucional con datos espaciales

Los datos espaciales consisten en 5,574 imágenes a color de arácnidos y sus corrrespondientes etiquetas de acuerdo a 2 categorías: especie de importancia médica (IM) o especie que no supone importancia médica (NIM). El dataset se extrajo de la cuenta de Twitter @Arachno_Cosas mediante

web scraping con la librería snscrape. Los datos pasaron por una fase de pre-procesamiento en la que las imágenes se transformaron a un tensor de 4 dimensiones de forma (5574 x 128 x 128 x3) y las etiquetas se codificaron en un vector binario.

Este modelo se inspiró en una arquitectura AlexNet al utilizar tres capas consecutivas de convolución y max pooling pero añadiendo batch normalization luego de la convolución. En la estructura final de la red se utilizaron tres capas densas de 4096, 4096 y 1000 unidades con activación ReLU respectivamente y una neurona en la salida con activación sigmoide. En la siguiente figura, se presenta un diagrama de la arquitectura utilizada.

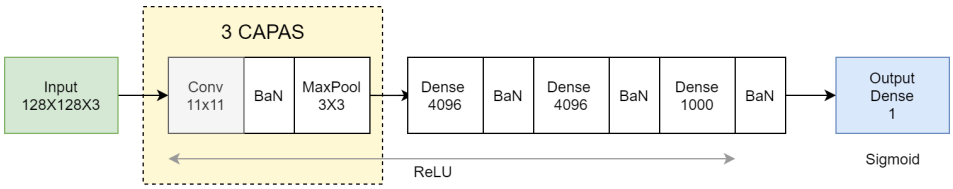

Figure 2: Arquitectura de la CNN inspirada en AlexNet.

2.3 Red neuronal recurrente con datos secuenciales

Para este modelo los datos consisten en las mismas imágenes usadas en la CNN junto con el contenido (como texto) de las respuestas a los tweets sobre las fotografías de los arácnidos. Aunque el origen de los datos es el mismo que para el modelo anterior, el pre-procesamiento es diferente. En este caso, se requiere limpiar los datos de texto, removiendo signos de puntuación, emojis, enlaces de internet y otros caracteres especiales. Para codificar la sequencia de texto se propuso mapear cada palabra a un vector multi dimensional por medio de un modelo pre-entrenado de Vectores Globales para Representación de de Palabras o *GloVe* como es conocido en inglés. Para la generación de las descripciones (tweets) se propuso también el uso de dos métodos: *Greedy Search* y *Beam Search*.

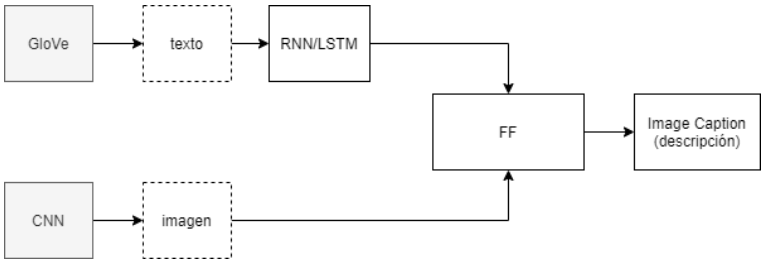

Figure 3: Modelo encoder-decoder.

Respecto a la arquitectura se propuso utilizar un modelo encoder-decoder, combinando la forma codificada de la imagen y la forma codificada del texto (en este caso el contenido del tweet) para alimentarlos a un decodificador. En la práctica, esto significa que el modelo se compone de una red neuronal convolucional que provee la imagen codificada, el modelo *Glove* que provee el texto codificado en secuencias, y una red neuronal recurrente LSTM.

3 Resultados

Los resultados del proyecto corresponden a dos modelos de redes neuronales entrenadas (una feed forward para predicción de churn y otra convolucional para clasificaciónd e imágenes de arañas según si importancia médica). En el caso de image captioning se implementó el modelo de lenguaje inicial para la preparación de los datos de entrada. Estos modelos se encuentran disponibles en el repositorio (ver apéndice A). A continuación se enumeran los resultados alcanzados en la fase de evaluación en los datasets de pruebas.

- Para el modelo de predicción de churn de clientes se alcanzó una exactitud de 85.55% en un entrenamiento de solo 100 epochs, utilizando un batch size de tamaño 16.

- Para el modelo de clasificación de imágenes de arácnidos según su importancia médica se alcanzó un 82.24% de exactitud en datos de pruebas. Para ello, se realizó el entrenamiento en 100 epochs, con 30 pasos por epoch y batch size de tamaño 16.

4 Conclusiones y trabajo futuro

Se presentó un modelo para predicción de churn de los clientes de un banco utilizando una red neuronal feed forward. También se presentó un modelo de deep learning, para clasificación de imágenes de arácnidos de acuerdo con su importancia médica, utilizando una red neuronal convolucional basada en Arquitectura AlexNet con batch normalization.

Debido a los prolongados tiempos de entrenamiento y la limitación de tiempo, no fue posible capturar más experimentos con variaciones en los hiper-parámetros, pero se considera que es posible mejorar las métricas de desempeño del modelo al someterlo a más entrenamiento, así como regular el sobre ajuste potencial con capas de dropout.

Referente al modelo propuesto RNN/LSTM se prevee seguirlo desarrollando para obtener un modelo final funcional capaz de emular las respuestas del experto sobre la clasificación y nivel de importancia médica de los arácnidos a partir de las imágenes. En este caso, el desarrollo alcanzó hasta la fase de pre-procesamiento de los datos de texto.

References

[1] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. Imagenet classification with deep convolutional neural networks. In *Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1*, NIPS'12, page 1097–1105, Red Hook, NY, USA, 2012. Curran Associates Inc.

[2] F. Méndez, D. Mendoza. *El aracnólogo más famoso de las redes sociales* UNAM Global. (2020, September 1). https://unamglobal.unam.mx/el-aracnologo-mas-famoso-de-las-redes-sociales/

A Código y detalles de implementación

Repositorio de GitHub con el código de la implementación y pre-procesamiento de datos de cada parte del proyecto disponible en: `https://github.com/eddson90/tareas_data_science/tree/main/statistical-learning-2/proyecto-final`