Metodología de entrenamiento de un Transformer visual

La arquitectura del Transformer ha representado una revolución en las tareas de procesamiento de lenguaje natural. El éxito de los mecanismos de atención y de los grandes modelos pre-entrenables permitieron su expansión a otros dominios. Uno de estos dominios es la visión por computador, en donde el concepto de atención y la habilidad del modelo para enfocarse en características clave ha sido de gran importancia. Esta revolución dio origen a los Transformers visuales, los cuales continúan revolucionando el campo del análisis y procesamiento de imágenes y se convirtieron en el estado del arte para las principales tareas de la visión por computador: detección, clasificación y segmentación. En esta lectura, recorreremos la metodología de entrenamiento del Transformer visual.

Entrenar un Transformer visual

Recordemos que una de las ventajas principales de los Transformers visuales es que pueden ser pre-entrenados con una cantidad significativa de datos. Por ejemplo, una de las bases de datos en las que ViT fue entrenado para la tarea de clasificación es ImageNet-21k, la cual contiene 14 millones de imágenes. En caso de que deseemos utilizar esta arquitectura para clasificar razas de perros, un gran punto de inicio es partir de los pesos pre-entrenados de ViT en alguna base de datos, por ejemplo, ImageNet-21k. Posteriormente, se realiza un entrenamiento refinado sobre las imágenes de diferentes razas de perros. A este proceso se le conoce como **fine tuning** o **refinamiento**.

Durante el fine tuning, el modelo se entrena en los datos de interés utilizando una tasa de aprendizaje más baja y cargando los pesos pre-entrenados. Los otros aspectos de la metodología de entrenamiento son similares a los mencionados previamente en otras arquitecturas:

1) Se deben dividir los datos en segmentos de entrenamiento, validación y evaluación.

Universidad de los Andes | Vigilada Mineducación Reconocimiento como
Universidad: Decreto 1297 del 30 de mayo de 1964.
Reconocimiento personería jurídica: Resolución 28 del 23 de febrero de 1949 Minjusticia.

Universidad de los Andes Colombia

2) Se entrena el modelo utilizando una tasa de aprendizaje más baja y cargando los pesos pre-entrenados en otra base de datos.
3) Se evalúa el desempeño en el segmento de validación y se realizan ajustes al entrenamiento y a la arquitectura para mejorar la métrica de evaluación.
4) Se evalúa en el set de evaluación para la divulgación de sus resultados o la implementación en un escenario específico.

Conclusiones

En conclusión, la metodología de entrenamiento de un Transformer visual sigue los mismos principios de otras redes neuronales. La diferencia principal se encuentra en la capacidad de los Transformers para ser entrenados en grandes bases de datos y posteriormente ser refinados para tareas específicas.

Bibliografía

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., ... & Houlsby, N. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929.

Universidad de los Andes | Vigilada Mineducación Reconocimiento como
Universidad: Decreto 1297 del 30 de mayo de 1964.
Reconocimiento personería jurídica: Resolución 28 del 23 de febrero de 1949 Minjusticia.

Universidad de los Andes
Colombia

© - Derechos Reservados: la presente obra, y en general todos sus contenidos, se encuentran protegidos por las normas internacionales y nacionales vigentes sobre propiedad Intelectual, por lo tanto su utilización parcial o total, reproducción, comunicación pública, transformación, distribución, alquiler, préstamo público e importación, total o parcial, en todo o en parte, en formato impreso o digital y en cualquier formato conocido o por conocer, se encuentran prohibidos, y solo serán lícitos en la medida en que se cuente con la autorización previa y expresa por escrito de la Universidad de los Andes.

De igual manera, la utilización de la imagen de las personas, docentes o estudiantes, sin su previa autorización está expresamente prohibida. En caso de incumplirse con lo mencionado, se procederá de conformidad con los reglamentos y políticas de la universidad, sin perjuicio de las demás acciones legales aplicables.

Universidad de los Andes | Vigilada Mineducación Reconocimiento como Universidad: Decreto 1297 del 30 de mayo de 1964.
Reconocimiento personería jurídica: Resolución 28 del 23 de febrero de 1949 Minjusticia.