

Notas

Alcantar Gómez Alan Arnoldo

27 de mayo de 2019

1. Introducción

El reconocimiento óptico de caracteres (ROC) es una rama de las ciencias computacionales enfocada a la obtención de información alfanumérica a partir de imágenes escaneadas o fotográficas. En la actualidad la mayoría de los artículos de investigación utilizan la base de datos MNIST la cual es un conjunto de referencia para probar el rendimiento de la extracción de características o modificaciones a los modelos. La forma habitual de trabajo consiste en separar el conjunto de datos disponibles en dos subconjuntos llamados prueba y entrenamiento, y optimizan algunos parámetros del modelo tomando pequeños conjuntos de muestreo. El mayor hueco de conocimiento que se observa es la falta de un análisis de las curvaturas de aprendizaje, las cuales consiste en ajustar un modelo a un conjunto de datos de tamaño diferente, con el objetivo de obtener el comportamiento del modelo ante la cantidad de datos disponibles. El presente trabajo hace un diseño de experimentos para algunos algoritmos de clasificación de la librería SK-Learn y hace un análisis estadístico a las curvas de aprendizaje de cada algoritmo.

2. Antecedentes y literatura

[Lecun et al., 1998] Este trabajo presenta el entrenamiento de varias versiones de la reconocida red neuronal convolucional LeNet5, con el fin de observar la influencia del tamaño del conjunto de entrenamiento se establecen tres tamaños; 15 000, 30 000 y 60 000, y la representación de los datos se hace usando los 784 píxeles que forman la imagen. Los resultados de las curvas de aprendizaje muestran que es posible incrementar la base de datos MNIST para mejorar el rendimiento de reconocimiento. En base a lo anterior se aplicaron distorsiones aleatorias al MNIST hasta conseguir 540 000 datos nuevos, con ello fue posible mejorar el rendimiento de reconocimiento de 99.05 % (60 000 datos) a 99.2 % (600 000 originales más distorsionados). Por último, el artículo presenta una extensa comparación entre clasificadores entre los cuales están los lineales, vecinos mas cercanos, polinomiales, redes RBF y algunas redes neuronales, y las comparaciones se hace utilizando el porcentaje de error sobre el conjunto de prueba, número de operaciones para reconocer un dígito y número de variables

para cada modelo.

Para 2007 se publica el artículo [Lauer et al., 2007], el cual se enfoca en desarrollar un extractor de características basado en la arquitectura de la LeNet5 el cual toma el vector original y cuya salida es un vector de 120 elementos, el cual representa el vector las características extraídas y este es conectado a una multitud de SVM, bautizando a este algoritmo como TFE-SVM. Para verificar la relevancia del TFE-SVM se entreno el modelo con los tamaños de 15 000, 30 000 y 60 000 de la base MNIST, y se presenta el error porcentual sobre el conjunto de entrenamiento contra el tamaño del conjunto de entrenamiento.

El artículo mas reciente encontrando sobre curvas de aprendizaje sobre MNIST fue publicado en 2009 [Levi and Ullman, 2009] bajo la idea que el uso de grandes bases de datos para entrenar un modelo de clasificación no solo permite hacer mejoras en la precisión, sino que además permite utilizar modelos más complejos los cuales pueden tener una mejora substancial en el rendimiento, este trabajo propone la implementación de una máquina de soporte vectorial con kernel en lineal (OKSVM), en el cual la complejidad del modelo es aprendida a medida que más datos están disponibles. La representación de los datos es la de origen (vector de 784 elementos). Durante la fase entrenamiento los datos son presentados de forma aleatoria y el rendimiento es representado por el porcentaje de error sobre el conjunto de prueba .

3. Bibliografía

Referencias

- [Lauer et al., 2007] Lauer, F., Suen, C. Y., and Bloch, G. (2007). A trainable feature extractor for handwritten digit recognition. *Pattern Recognition*, 40(6):1816 – 1824.
- [Lecun et al., 1998] Lecun, Y., Bottou, L., Bengio, Y., and Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324.
- [Levi and Ullman, 2009] Levi, D. and Ullman, S. (2009). Learning model complexity in an online environment. In *2009 Canadian Conference on Computer and Robot Vision*, pages 260–267.