

Conclusión

Lección 29

Dr. Pablo Alvarado Moya

CE5506 Introducción al reconocimiento de patrones
Área de Ingeniería en Computadores
Tecnológico de Costa Rica

II Semestre, 2019

Contenido

- 1 Redes neuronales recurrentes
- 2 Ensamblés
- 3 Detección de anomalías
- 4 Sistemas de recomendación
- 5 Aprendizaje Profundo
 - Redes neuronales convolucionales
 - Auto-codificadores
 - Redes adversarias generativas

Conclusión

- El área de reconocimiento de patrones y aprendizaje automático crece rápidamente.
- Hemos revisado en el curso las bases que hacen posible comprender las propuestas que aparecen día a día.
- Los temas son sin embargo abundantes y de evolución rápida
- En esta conclusión del curso presentaremos algunos tópicos “calientes” en el área, solo como invitación a su revisión

Redes neuronales recurrentes

- Las **redes neuronales recurrentes** son una alternativa a los HMM, para la clasificación de sucesiones.
- Son redes neuronales con realimentación, es decir, salidas de una capa pueden ingresar como entrada a la misma capa o a capas inferiores.
- Particularmente las llamadas redes LSTM (*long short temporal memory*) (Hochreiter&Schmidhuber, 1997) han recibido mucha atención, y se usan actualmente para el reconocimiento de habla.

Combinación de clasificadores débiles

(1)

- Una pregunta natural: ¿es posible mejorar el desempeño de clasificadores o regresores combinando varios de ellos?
- La respuesta es **sí**, y de distintas formas y a distintos niveles:
 - **Bagging**: corrige alta varianza
Entrena varios clasificadores similares con muestras aleatorias del conjunto de datos original.
 - **Boosting**: corrige alto sesgo
Usa una sucesión de clasificadores, donde cada nueva etapa recibe los datos que no clasifican bien las etapas anteriores
 - **Stacking**: mejora capacidad predictiva
Similar a boosting. Se usan distintos clasificadores y sus salidas se usan como entrada a otros clasificadores, por lo que clasificadores se “apilan”.
- En principio, todos los métodos:

Combinación de clasificadores débiles

(2)

- 1 Producen una distribución de modelos usando subconjuntos de los datos originales.
- 2 Combinan los modelos en un modelo **agregado**
- Bosques aleatorios (Junglas, Helechos, etc.) combinan varios árboles de decisión al estilo de bagging (cada árbol es de alta varianza), aunque cada nodo del árbol está haciendo una clasificación muy simple, lo que se puede ver como boosting.

Detección de anomalías

(1)

- La tarea consiste en detectar si un nuevo dato es anómalo o sigue la distribución que caracteriza la operación normal de algún sistema.
 - 1 Detección de operación correcta de motores, máquinas, sistemas, etc.
 - 2 Detección de fraude
- En principio, aplicamos aquí técnicas de estimación de PDF, similares a las que revisamos en el curso:
 - Mezcla de gaussianas
 - Histogramas
 - Kernels
 - ...

Detección de anomalías

(2)

- Por medio de PCA, ICA u otros podemos decorrelar/independizar componentes y analizarlos independientemente (reducir dimensiones)

Detección de anomalías

(3)

- Dificultades:
 - Usualmente para entrenamiento hay muchos más datos disponibles para operación normal que para operación defectuosa.
 - Hay una alta probabilidad de que anomalías futuras no tengan ninguna similitud con anomalías disponibles en entrenamiento
 - Otro concepto importante aquí es distancias de un dato a una distribución: por ejemplo: distancia de Mahalanobis

Sistemas de recomendación

(1)

- Esta es una de las aplicaciones frecuentes actualmente
- Es lo que usa Amazon, Netflix, Spotify, etc. para recomendar
- Es un tema activo en industria aunque no en academia.
- El problema es el siguiente:
dado un conjunto de u usuarios y m posibles productos, un usuario califica un subconjunto pequeño de los m productos, el sistema debe predecir qué calificación daría ese usuario a productos no evaluados, con base a la información total disponible.

Sistemas de recomendación

(2)

Película	María	Pedro	Ana	Juan
Amor eterno	5	4	0	0
Romance	5	?	0	0
Mascotas	?	5	0	?
Autos 43	0	?	5	4
Karate vs. Kung-Fu	?	0	5	?

- Las técnica “**basada en contenido**” usualmente crean un vector de descriptores asociados a cada producto, y luego por regresión sobre dichos descriptores se busca predecir para cada usuario cuál calificación le asignaría a los productos no evaluados.

Sistemas de recomendación

(3)

- Otra técnica de “**filtrado colaborativo**” supone que cada usuario indica sus gustos (por ejemplo, qué tipos de producto gusta), y entonces se infieren los vectores de entrada automáticamente a partir de los gustos indicados por todos los usuarios y las calificaciones de los productos.

Aprendizaje profundo

Deep learning

(1)

- Con la disponibilidad actual de grandes cantidades de datos, es posible aplicar modelos de alta varianza en el aprendizaje, pues los datos masivos evitan el sobreajuste.
- Como aprendizaje profundo se conocen técnicas, particularmente basadas en redes neuronales, en las que múltiples capas producen la alta variabilidad.
- Los modelos usualmente tienen decenas de miles, centenas de miles o millones de parámetros.
- **Aprendizaje profundo** engloba gran cantidad de técnicas para regresión, clasificación, aprendizaje reforzado, recomendación, etc.

Aprendizaje profundo

Deep learning

(2)

- Las técnicas están diseñadas para aprovechar las grandes cantidades de datos, para lo cual también usan nuevos métodos para regularizar los modelos de alta varianza y acelerar el entrenamiento.
- Las demandas computacionales de los métodos son altas, tanto para entrenamiento como para reconocimiento, por lo que el desarrollo del área incluye investigación en nuevas arquitecturas de hardware especializadas:
 - Neuro-procesadores
 - TPU/GPU
 - Procesadores de visión

Aprendizaje profundo

Deep learning

(3)

- El aprendizaje profundo ha permitido incursionar en nuevas áreas teóricas, particularmente el llamado “aprendizaje de representación”, en donde el mismo modelo aprende qué características son útiles para la tarea en cuestión.
- Estas áreas a su vez se usan en el llamado “aprendizaje transferido”, que permite usar lo aprendido en un contexto en otro distinto.

Aprendizaje profundo: CNN

Convolutional Neural Networks

(1)

- Las CNN son uno de los métodos de aprendizaje profundo más populares
- Se usan fuertemente en visión por computador y otras áreas donde los datos están en una rejilla 1D, 2D o 3D.
- La idea es reemplazar capas iniciales totalmente conectadas por capas de filtrado, donde los filtros se aprenden de los datos.
- Las capas convolucionales reducen fuertemente el número de parámetros requeridos.

Aprendizaje profundo: CNN

Convolutional Neural Networks

(2)

- Entre capas de filtrado se introducen capas de submuestreo no lineal (*pooling*).
(Esta operación está en discusión académica → redes capsulares, G. Hinton)
- Usualmente las últimas capas son completamente conectadas (como en redes neuronales convencionales)

Aprendizaje profundo: Auto-codificadores

Autoencoders

- Los auto-codificadores son redes neuronales, que buscan copiar la entrada en la salida.
- Internamente tienen un cuello de botella en el camino que obliga a encontrar una representación eficiente comprimida de la información a la entrada, tal que de allí la entrada pueda ser reconstruida.
- La red se conforma de un **codificador** desde la entrada a la capa escondida con el código, y un **decodificador** desde el código a la salida.
- Las aplicaciones son abundantes: compresión, eliminación de ruido, generación automática de características, modelos con variables latentes, etc.

Aprendizaje profundo: GAN

Generative Adversarial Networks

(1)

- Las GAN son un modelo generativo.
- Aprenden la distribución probabilística que subyace los datos de entrenamiento.
- Si generamos datos de la distribución probabilística aprendida, querríamos obtener datos indistinguibles de aquellos en el conjunto de entrenamiento original.
- Las GAN consisten en dos redes que compiten entre sí:
 - la red **discriminante** recibe datos reales o generados y devuelve la probabilidad de que la imagen sea o no real
 - la red **generativa** recibe un ruido aleatorio a la entrada y con él genera un dato a partir de la distribución aprendida.

Aprendizaje profundo: GAN

Generative Adversarial Networks

(2)

- La segunda red es similar al decodificador de un auto-codificador
- La idea en GAN es entrenar la red generativa para engañar a la red discriminante, pero la red discriminante se entrena de forma supervisada para detectar las imágenes falsas.
- Las GAN son la base de los procesos “creativos” artificiales: arte sintético, texto a imagen, texto a sonido, etc.

FIN

Resumen

- 1 Redes neuronales recurrentes
- 2 Ensamblés
- 3 Detección de anomalías
- 4 Sistemas de recomendación
- 5 Aprendizaje Profundo
 - Redes neuronales convolucionales
 - Auto-codificadores
 - Redes adversarias generativas

Este documento ha sido elaborado con software libre incluyendo L^AT_EX, Beamer, GNUPlot, GNU/Octave, XFig, Inkscape, GNU-Make y Subversion en GNU/Linux



Este trabajo se encuentra bajo una Licencia Creative Commons Atribución-NoComercial-LicenciarIgual 3.0 Unported. Para ver una copia de esta Licencia, visite <http://creativecommons.org/licenses/by-nc-sa/3.0/> o envíe una carta a Creative Commons, 444 Castro Street, Suite 900, Mountain View, California, 94041, USA.

© 2017–2019 Pablo Alvarado-Moya Área de Ingeniería en Computadores Instituto Tecnológico de Costa Rica