

Introducción

Técnicas no
vistas

Representaciones
distribuidas

Explicabilidad e
Interpretabilidad

DL - GANs

Desarrollo
Histórico

Perspectivas y
Retos

Super-
Inteligencia

Resumen y Perspectivas

Eduardo Morales y Hugo Jair Escalante

INAOE

Contenido

- 1 Introducción
- 2 Técnicas no vistas
 - Representaciones distribuidas
 - Explicabilidad e Interpretabilidad
 - DL - GANs
- 3 Desarrollo Histórico
- 4 Perspectivas y Retos
- 5 Super-Inteligencia

Introducción

Técnicas no vistas

Representaciones distribuidas

Explicabilidad e Interpretabilidad

DL - GANs

Desarrollo Histórico

Perspectivas y Retos

Super-Inteligencia

Resumen

Introducción

Técnicas no vistas

Representaciones distribuidas

Explicabilidad e Interpretabilidad

DL - GANs

Desarrollo Histórico

Perspectivas y Retos

Super-Inteligencia

- En el curso vimos varias de las técnicas recientes o no vistas en los otros cursos de Aprendizaje Computacional
- Existen técnicas que no vimos, por ejemplo, aprendizaje paramétrico basado en máxima verosimilitud, máquinas de Kernel, estimación bayesiana, *stream-learning*, *BigData*, *concept drift*, etc.

Representaciones distribuidas

Introducción

Técnicas no vistas

Representaciones distribuidas

Explicabilidad e Interpretabilidad

DL - GANs

Desarrollo Histórico

Perspectivas y Retos

Super-Inteligencia

- **Representaciones distribuidas:** palabras, frases o documentos son representados en un espacio vectorial, i.e., la semántica se distribuye en un espacio vectorial.
- Los elementos del espacio (las dimensiones) pueden tener un significado o no; pueden capturar información semántica, secuencial, etc.
- **Aplicaciones:** operaciones sobre términos (e.g., agrupamiento), generación de tesauros, etc.

Representaciones distribuidas

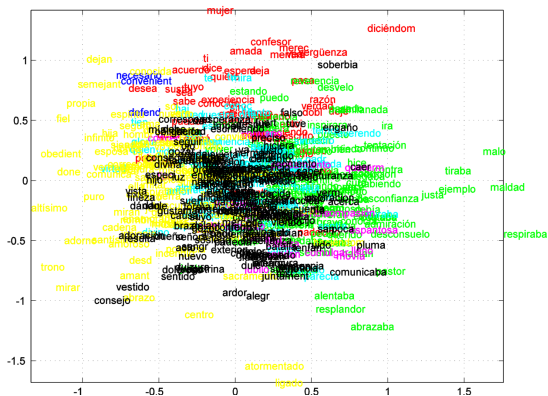


Figure: Representaciones distribuidas de palabras.

Introducción

Técnicas no vistas

Representaciones distribuidas

Explicabilidad e Interpretabilidad

DL - GANs

Desarrollo Histórico

Perspectivas y Retos

Super-Inteligencia

Representaciones distribuidas

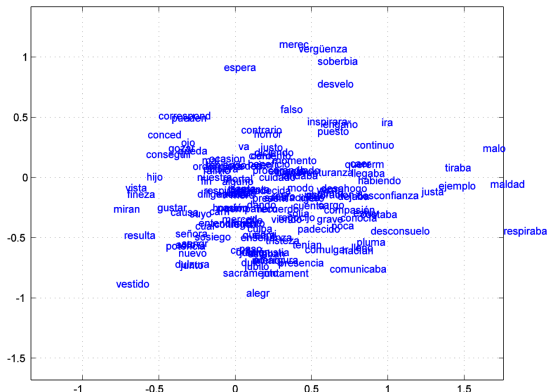


Figure: Representaciones distribuidas de palabras.

Introducción

Técnicas no vistas

Representaciones distribuidas

Explicabilidad e Interpretabilidad

DL - GANs

Desarrollo Histórico

Perspectivas y Retos

Super-Inteligencia

Representaciones distribuidas

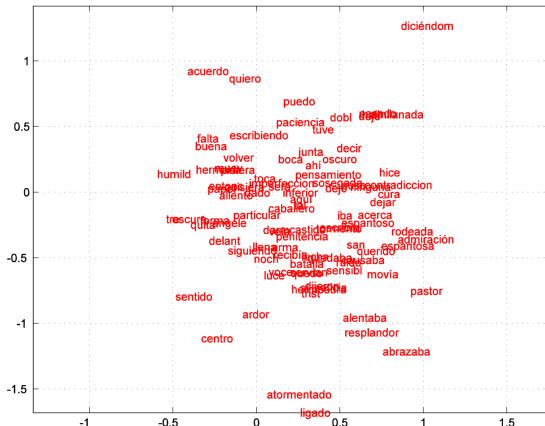


Figure: Representaciones distribuidas de palabras.

Introducción

Técnicas no
vistasRepresentaciones
distribuidasExplicabilidad e
Interpretabilidad

DL - GANs

Desarrollo
HistóricoPerspectivas y
RetosSuper-
Inteligencia

Representaciones distribuidas

Introducción

Técnicas no

vistas

Representaciones
distribuidas

Explicabilidad e
Interpretabilidad

DL - GANs

Desarrollo
Histórico

Perspectivas y
Retos

Super-
Inteligencia

- **Problema:** Cómo aprender estas representaciones?
- **Idea:** Usar modelos lineales y no lineales sobre un conjunto de datos para aprender una representación vectorial para cada palabra
 - El objetivo es aprender un modelo que prediga la palabra faltante a partir de cierto contexto ... (y colateralmente se aprende la representación)

Representaciones distribuidas

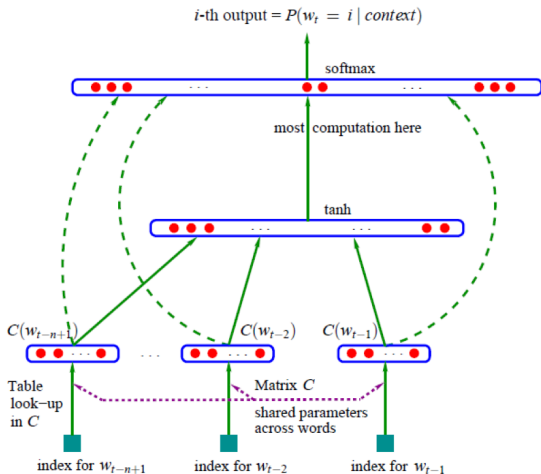


Figure: Lenguaje de modelo neuronal, Bengio et al.

Introducción

Técnicas no vistas

Representaciones distribuidas

Explicabilidad e Interpretabilidad

DL - GANs

Desarrollo Histórico

Perspectivas y Retos

Super-Inteligencia

Representaciones distribuidas

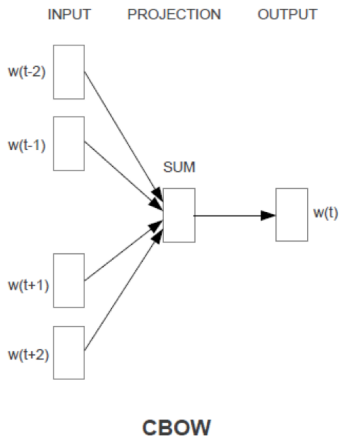


Figure: Modelo CBOW - Word2Vec (Mikolov et al.).

Introducción

Técnicas no vistas

Representaciones distribuidas

Explicabilidad e Interpretabilidad

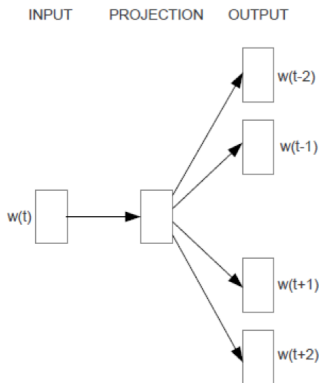
DL - GANs

Desarrollo Histórico

Perspectivas y Retos

Super-Inteligencia

Representaciones distribuidas



Skip-gram

Figure: Modelo Skyp gram - Word2Vec (Mikolov et al.).

Introducción

Técnicas no vistas

Representaciones distribuidas

Explicabilidad e Interpretabilidad

DL - GANs

Desarrollo Histórico

Perspectivas y Retos

Super-Inteligencia

Representaciones distribuidas

- **Extensión:** aprendizaje simultaneo de representaciones para documentos.

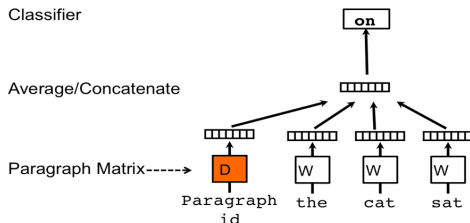


Figure: Vector párrafo - Q. Lee et al..

Explicabilidad e Interpretabilidad

Introducción

Técnicas no vistas

Representaciones distribuidas

Explicabilidad e Interpretabilidad

DL - GANs

Desarrollo Histórico

Perspectivas y Retos

Super-Inteligencia

- Avances en aprendizaje computacional ha permitido obtener modelos que obtiene desempeño comparable al de humanos en diversas tareas (e.g., deepface).
- Sin embargo, dichos modelos están limitados en cuanto a que no se puede decir mucho sobre:
 - Cuál es la razón detrás de una decisión/recomendación? (explicabilidad)
 - Qué explica el funcionamiento interno del modelo? (interpretabilidad)

Explicabilidad e Interpretabilidad

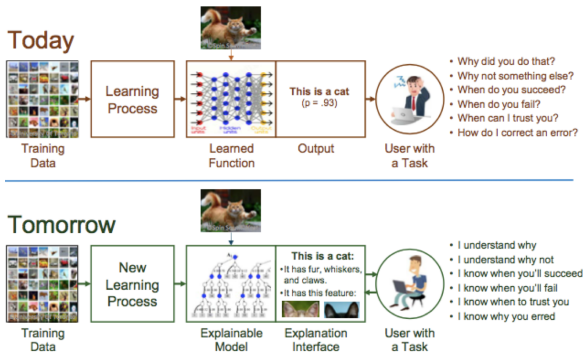


Figure: D. Gunn - DARPA.

Introducción

Técnicas no vistas

Representaciones distribuidas

Explicabilidad e Interpretabilidad

DL - GANs

Desarrollo Histórico

Perspectivas y Retos

Super-Inteligencia

Explicabilidad e Interpretabilidad

Introducción

Técnicas no vistas

Representaciones distribuidas

Explicabilidad e Interpretabilidad

DL - GANs

Desarrollo Histórico

Perspectivas y Retos

Super-Inteligencia

- Esfuerzos actuales:
 - Análisis, interpretación, visualización de pesos de modelos
 - Explicabilidad de modelos de caja negra (e.g., explicar la superficie de decisión de un clasificador)
 - Uso de métodos de *captioning* para explicar las recomendaciones de un modelo
 - Modelos probabilistas, basados en gramáticas para explicar las decisiones de los modelos
 - Identificación de instancias prototípicas para explicar los datos

Redes adversarias - generativas

Introducción

Técnicas no vistas

Representaciones distribuidas

Explicabilidad e Interpretabilidad

DL - GANs

Desarrollo Histórico

Perspectivas y Retos

Super-Inteligencia

- Objetivo: entrenar redes generativas profundas que modelen mejor la distribución de los datos
- Idea: Entrenamiento con adversario, se entrena simultáneamente
 - Un modelo generativo G: intenta capturar la distribución de los datos
 - Un modelo discriminativo D: intenta distinguir ejemplos reales de los generados por G

Redes adversarias - generativas

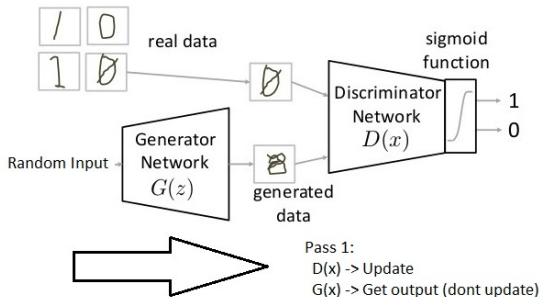


Figure: GANs¹.

¹<https://www.analyticsvidhya.com/blog/2017/06/introductory-generative-adversarial-networks-gans/>

Redes adversarias - generativas

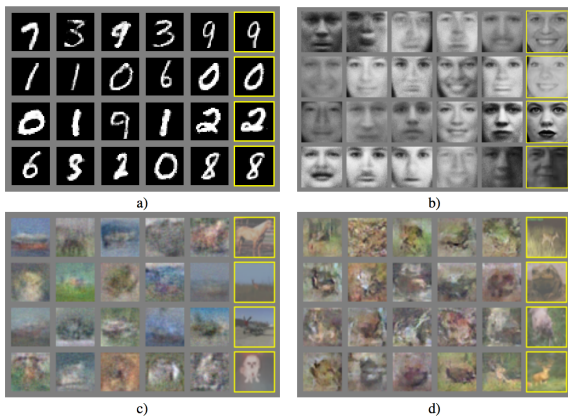


Figure: Generando instancias con GANs.

Introducción

Técnicas no vistas

Representaciones distribuidas

Explicabilidad e Interpretabilidad

DL - GANs

Desarrollo Histórico

Perspectivas y Retos

Super-Inteligencia

Redes adversarias - generativas

Introducción

Técnicas no vistas

Representaciones distribuidas

Explicabilidad e Interpretabilidad

DL - GANs

Desarrollo Histórico

Perspectivas y Retos

Super-Inteligencia

this small bird has a pink breast and crown, and black primaries and secondaries.



the flower has petals that are bright pinkish purple with white stigma



this magnificent fellow is almost all black with a red crest, and white cheek patch.



this white and yellow flower have thin white petals and a round yellow stamen



Figure: Generando instancias multimodales con GANs.

Resumen

Introducción

Técnicas no vistas

Representaciones distribuidas

Explicabilidad e Interpretabilidad

DL - GANs

Desarrollo Histórico

Perspectivas y Retos

Super-Inteligencia

- El área de aprendizaje computacional empezó con las primeras computadoras y con la convicción de que la creación de máquinas inteligentes necesitaba contar con aprendizaje
- Durante su desarrollo ha sufrido diversos cambios y evoluciones

Evolución

Antes	Ahora
<p>Evaluación informal, pocos datos, no generalizable</p>	<p>Evaluaciones formales, medida de desempeños, creación de <i>benchmarks</i></p>
<p>Representaciones simbólicas, cercanía con ciencias cognitivas, restrictivo a pocas áreas</p>	<p>Apertura a otras áreas como PR y NN</p>
<p>Diversos tipos de aprendizaje</p>	<p>Datos y evaluaciones proposicionales, énfasis en clasificación y regresión</p>

Introducción

Técnicas no vistas

Representaciones distribuidas

Explicabilidad e Interpretabilidad

DL - GANs

Desarrollo Histórico

Perspectivas y Retos

Super-Inteligencia

Evolución

Antes	Ahora
Íntimamente ligado a IA	Separación de IA a ML (recientemente se está juntando de nuevo)
I/O: Relacional, razonamiento, planes, experiencias, ...	Atributo-valor (algo en ILP y <i>Statistical Relational Learning</i>), tendencia en que sea DNN quien defina la representación
Sistemas inteligentes (AM, SOAR, ...), comportamientos “ricos” (semánticamente para tareas complejas)	Énfasis en desempeño y estadísticas

Introducción

Técnicas no vistas

Representaciones distribuidas

Explicabilidad e Interpretabilidad

DL - GANs

Desarrollo Histórico

Perspectivas y Retos

Super-Inteligencia

Evolución

Antes	Ahora
<p>Uso de conocimiento del dominio (EBL, ILP, <i>theory revision</i>, ...)</p>	<p>Muchos datos, mejora en desempeño (poco en aprender de pocos datos usando BK)</p>
<p>Reflexiones sobre resultados, a qué se deben, ...</p>	<p>Formalismo matemático (con sesgo) y comparación “ciega” de desempeños</p>

Introducción

Técnicas no vistas

Representaciones distribuidas

Explicabilidad e Interpretabilidad

DL - GANs

Desarrollo Histórico

Perspectivas y Retos

Super-Inteligencia

Persepectivas

Introducción

Técnicas no vistas

Representaciones distribuidas

Explicabilidad e Interpretabilidad

DL - GANs

Desarrollo Histórico

Perspectivas y Retos

Super-Inteligencia

- Desde hace tiempo se tiene un desarrollo acelerado del área (se pueden ver en el número de artículos y foros)
- En parte ha sido impulsado por un interés comercial que le han dado diversos nombres: KDD, *Big Data*, *Deep Learning*, ...
- Se está viviendo un re-surgimiento del área de Inteligencia Artificial (otra vez) con grandes expectativas (antes: Inicios, Sistemas Expertos, NN, KDD; ahora: DL)

Retos

Introducción

Técnicas no vistas

Representaciones distribuidas

Explicabilidad e Interpretabilidad

DL - GANs

Desarrollo Histórico

Perspectivas y Retos

Super-Inteligencia

- Aprendizaje sin (con poca) guía
- Diversos tipos de aprendizaje, concurrentes, y por largo tiempo
- Aprendizaje de “sentido común”
- Selección automática de información relevante
- Aprender con poca información
- Organización del conocimiento aprendido

Super-Inteligencia

- Durante muchos años, los investigadores de IA se dejaron de cuestionar si era posible crear una inteligencia artificial
- A pesar del entusiasmo inicial, el problema resultó ser mucho más difícil de lo anticipado
- Con los avances recientes, tanto en generación y acceso a datos, computadoras más veloces, y algoritmos de DL, la gente ya no se cuestiona si se va a lograr, sino cuándo y cuáles pueden ser sus consecuencias
- Algunos autores predicen que con el desarrollo actual será posible crear una super-inteligencia (entes más inteligentes que el hombre)

Introducción

Técnicas no vistas

Representaciones distribuidas

Explicabilidad e Interpretabilidad

DL - GANs

Desarrollo Histórico

Perspectivas y Retos

Super-Inteligencia

Super-Inteligencia

Existen varios candidatos para crear super-inteligencias:

- IA: algunos creen la más probable y cercana
- Emulación del cerebro: Crear cerebros artificiales con base en los biológicos
- Bio-ingeniería (DNA): Cambiar el DNA para promover inteligencia (*selective breeding*)
- Implantes cerebrales: Combinar el cerebro con implantes que aumenten la capacidad cerebral (e.g., conectarse a Internet - memoria ilimitada, transmisión de datos, cálculos complejos ...)
- Fármacos: Aumentar la capacidad cerebral por medio de fármacos
- Organización Grupal: Crear grupos específicos para resolver problemas difíciles (super-inteligencia colectiva)

Introducción

Técnicas no vistas

Representaciones distribuidas

Explicabilidad e Interpretabilidad

DL - GANs

Desarrollo Histórico

Perspectivas y Retos

Super-Inteligencia

Super-Inteligencia

Introducción

Técnicas no vistas

Representaciones distribuidas

Explicabilidad e Interpretabilidad

DL - GANs

Desarrollo Histórico

Perspectivas y Retos

Super-Inteligencia

Con base en encuestas de cuándo se podría tener una inteligencia a nivel del ser humano (HLMI):

- 2030 - 10%
- 2050 - 50%
- 2100 - 90%

Cuánto tiempo después se va a tener una super-inteligencia:

- 2 años después - 5%-10%
- 30 años después - 50%-75%

Super-Inteligencia

Introducción

Técnicas no vistas

Representaciones distribuidas

Explicabilidad e Interpretabilidad

DL - GANs

Desarrollo Histórico

Perspectivas y Retos

Super-Inteligencia

- Creen que en cuanto nos acerquemos a una inteligencia, se va a dar muy rápido la creación de una super-inteligencia
- Lo que se cree es que al crear un ente inteligente, éste va a poder desarrollar uno más inteligente y así sucesivamente (y algunos creen que se puede dar muy rápido - meses/años)
- Dadas las expectativas que se tiene se está creando una carrera hacia ésto (parecido a la carrera a la luna)

Super-Inteligencia

Introducción

Técnicas no vistas

Representaciones distribuidas

Explicabilidad e Interpretabilidad

DL - GANs

Desarrollo Histórico

Perspectivas y Retos

Super-Inteligencia

- Posiblemente se vaya a crear un “singleton” con ventaja y dominio competitivo sobre los demás
- Si no lo crea un gobierno, que tal un grupo terrorista?
- Se pueden monitorear los avances de los diversos grupos? Se harían colaboraciones internacionales?

Super-Inteligencia

Introducción

Técnicas no vistas

Representaciones distribuidas

Explicabilidad e Interpretabilidad

DL - GANs

Desarrollo Histórico

Perspectivas y Retos

Super-Inteligencia

Miedos:

- Se van a desplazar empleos (los no o poco calificados) y se cree que pronto (aunque no se tenga una super-interligencia)
- Miedo de que termine con la raza humana
- Se podría controlar?
- Se pueden inculcar valores “humanos”?