

Aprendizaje semisupervisado

UNED

Javier Rodríguez Vidal

**Reading Group Doctorandos NLP&IR@UNED
de noviembre de 2017, Madrid**



Natural Language Processing and
Information Retrieval Group at UNED

nlp.uned.es

Índice de contenidos

- **Introducción**
- **Modelos Generativos**
- **Co-entrenamiento**
- **Métodos basados en grafos**
- **Conclusiones**

Índice de contenidos

- **Introducción**
- Modelos Generativos
- Co-entrenamiento
- Métodos basados en grafos
- Conclusiones

Introducción

- **Aprendizaje supervisado**

- **Desventajas:** coste y dependencia de la anotación de datos, sesgos de los datos de entrenamiento, *overfitting*.

- **Aprendizaje no supervisado**

- **Desventajas:** dependencia del problema, métodos más costosos computacionalmente.

- **Aprendizaje semi-supervisado**

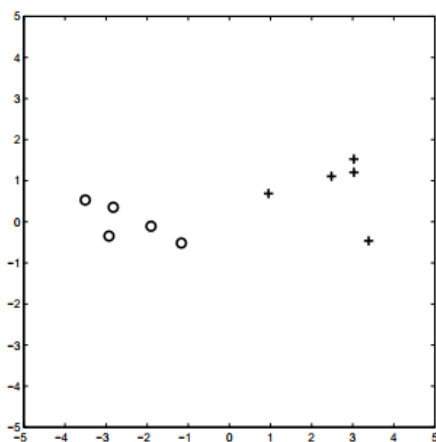
- Uso de datos etiquetados y no etiquetados (con mayor cantidad de no etiquetados).
- **Ventajas:** no es necesario mucho esfuerzo para etiquetar y da mayor precisión que los no supervisados.
- **Desventajas:** escalabilidad (el coste para algunos algoritmos basados en grafos es $O(n^3)$).
- Se usa en el algoritmo de Expectation-Maximization, en co-entrenamiento, etc.

Índice de contenidos

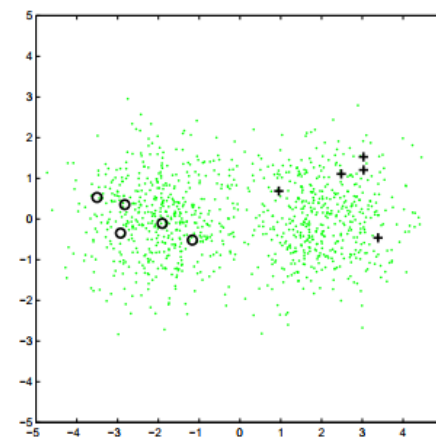
- Introducción
- **Modelos Generativos**
- Co-entrenamiento
- Métodos basados en grafos
- Conclusiones

Modelos Generativos

- Quizás son los métodos
- Tenemos un modelo
- Idealmente se necesita
- Completamente la d

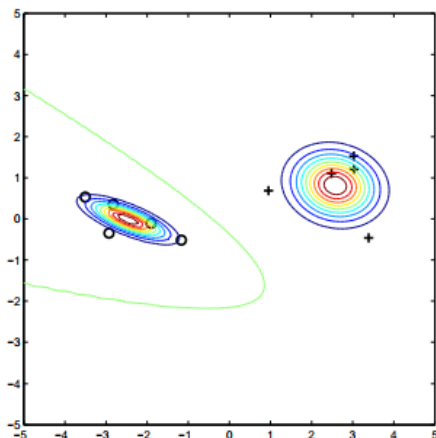


(a) labeled data

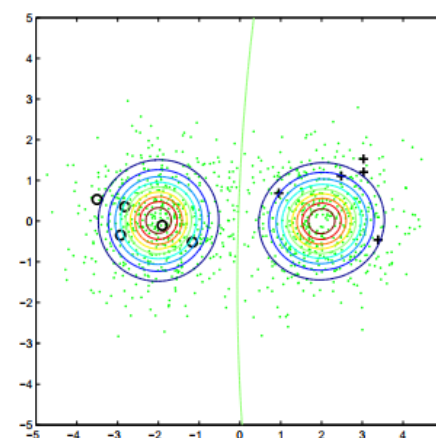


(b) labeled and unlabeled data (small dots)

- Poner atención en lo
- El modelo tiene
- Que el modelo
- EM Local Maxim
- Hacer clústeres
- Kernel de Fisher



(c) model learned from labeled data



(d) model learned from labeled and unlabeled data

nativo identificable,

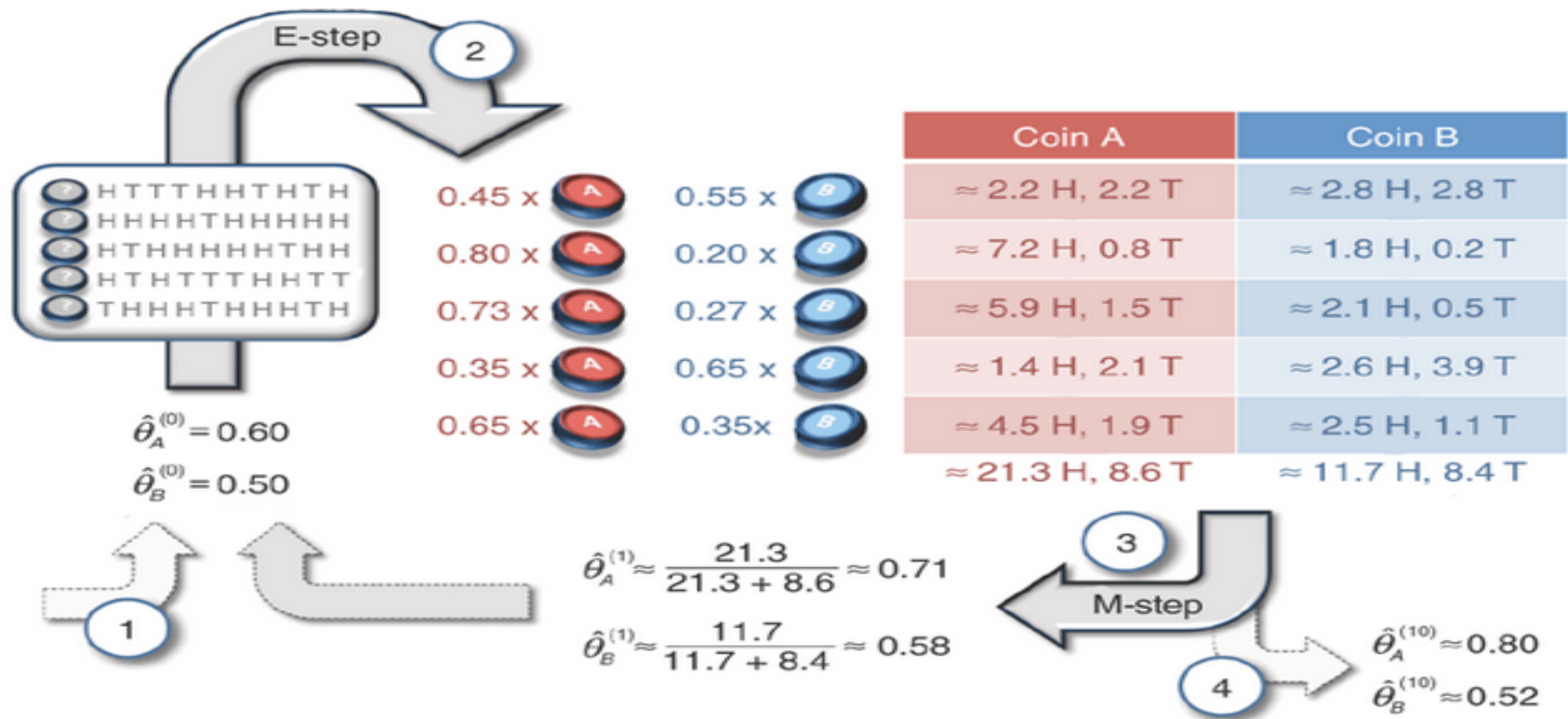
terminar

Modelos Generativos

- **Expectation-Maximization (EM): (Arthur Dempster, Nan Laird & Donald Rubin 1977)**
 - Se intenta encontrar estimadores de máxima verosimilitud de parámetros en modelos probabilísticos que dependen de variables no observables.
 - El paso E: computamos la esperanza de la verosimilitud mediante la inclusión de variables latentes como si fueran observables
 - El paso M: se computan estimadores de máxima verosimilitud de los parámetros mediante la maximización de la verosimilitud esperada en el paso E.
 - Los parámetros que se encuentran en el paso M se usan para comenzar el paso E siguiente (proceso iterativo hasta que converjan)

Modelos Generativos

b Expectation maximization



Índice de contenidos

- Introducción
- Modelos Generativos
- **Co-Entrenamiento**
- Métodos basados en grafos
- Conclusiones

Co-Entrenamiento

- Tiene tres suposiciones:
 - Las señales se pueden dividir en dos subconjuntos.
 - Cada subconjunto de señales es lo suficientemente bueno para entrenar a un buen clasificador.
 - Los dos subconjuntos son condicionalmente independientes.
- Funcionamiento:
 - Inicialmente se entrena dos clasificadores de manera separada con los datos etiquetados y cada subconjunto de señales respectivamente.
 - Cada clasificador clasifica los datos no etiquetados y enseña al otro clasificador con los datos no etiquetados y sus etiquetas predichas (aquellas con las que tenga más confianza).
 - Cada clasificador se re-entrena con los datos proporcionados por su contrario.

Co-Entrenamiento

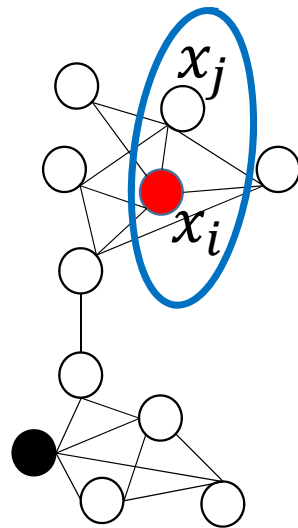
- Los datos no etiquetados sirven para reducir el espacio de la versión (los dos clasificadores deben ponerse de acuerdo en el tamaño de los datos etiquetados y no etiquetados).
- Las señales deben ser lo suficientemente buenas para que se pueda confiar en las etiquetas proporcionadas por cada clasificador.
- Necesitamos que las señales sean condicionalmente independientes para que los puntos de datos con alta confianza de un clasificador sean muestras para el otro.

Índice de contenidos

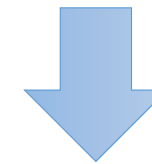
- Introducción
- Modelos Generativos
- Co-Entrenamiento
- **Métodos basados en grafos**
- Conclusiones

Métodos basados en grafos

- Definen un grafo donde los nodos son ejemplos etiquetados y no etiquetados del dataset y los ejes (que pueden tener pesos) reflejan la similitud de los ejemplos.
- Normalmente estos métodos asumen un etiquetado smooth sobre el grafo.



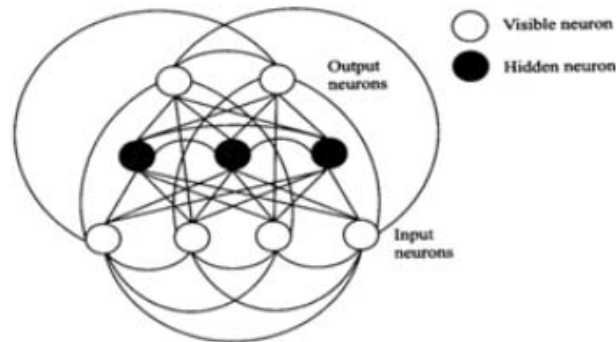
Dos puntos x_i, x_j son cercanos



Sus etiquetas respectivas y_i, y_j probablemente sean las mismas

Métodos basados en grafos

- **Discrete Markov Random Fields: Boltzmann Machines: (Hinton & Sejnowski, 1985)**
 - Las redes de Boltzmann consisten en neuronas conectadas entre sí que pueden estar conectadas bidireccionalmente y que tienen salidas binarias. Las neuronas se distinguen en dos grupos: las visibles y las no visibles.



Métodos basados en grafos

- **Discrete Markov Random Fields: Boltzmann Machines:**

Sea x' el vector de entrada con componentes desconocidos.

1. Asignar los valores conocidos del vector de entrada x' a las neuronas visibles.
2. Imputar todos los valores desconocidos y de las neuronas no visibles con valores aleatorios en el conjunto $\{0,1\}$.
3. Seleccionar una unidad x_k aleatoriamente y calcular su valor de entrada a la red (net_k)
4. Sin importar el valor actual de la unidad seleccionada, asignar el valor $x_k=1$ con probabilidad P_k (definida anteriormente).
5. Repetir los pasos 3 y 4 hasta que todas las unidades tengan una probabilidad de ser seleccionadas para una actualización (ciclo de procesamiento, un ciclo completo no garantiza que todas las unidades hayan sido actualizadas).
6. Repetir el paso 5 hasta llegar al equilibrio térmico.
7. Bajar la temperatura T y repetir pasos 3 a 7.

El algoritmo se detiene cuando T se ha reducido hasta un valor pequeño. Cuando esto sucede la red se ha estabilizado y el resultado final será las salidas de las neuronas visibles.

Índice de contenidos

- Introducción
- Modelos Generativos
- Co-Entrenamiento
- Métodos basados en grafos
- **Conclusiones**

Conclusiones

- Las técnicas de **aprendizaje semisupervisado** evitan la necesidad y el coste de tener todo el conjunto de datos anotado y tienen mayor precisión que los no supervisados al contar con datos anotados.
- **Expectation-Maximization** tiene la capacidad de manejar información faltante y observar variables ocultas por lo que es usado en problemas de clustering.
- En las técnicas de *co-entrenamiento se puede utilizar datos no etiquetados para aumentar los datos etiquetados teniendo dos vistas de un ejemplo que son redundantes pero no están completamente correlacionadas.*
- **Las máquinas de Boltzmann** presentan un grave problema práctico, y es que el aprendizaje parece dejar de producirse correctamente cuando la máquina se amplía a algo más grande que una máquina trivial. Esto se debe a:
 - El tiempo que la máquina necesita para recopilar las estadísticas de equilibrio crece exponencialmente con el tamaño de la máquina, y con la magnitud de la fuerza de las conexiones.
 - Las fuerzas de las conexiones son más flexibles cuando las unidades conectadas tienen probabilidades de activación intermedias entre cero y uno, llevando a la llamada [trampa de varianza](#) (el ruido hace que las fuerzas de las conexiones se vuelvan aleatorias hasta que las actividades se saturan)

Bibliografía

- Zhu X: “Semi-supervised Learning Literature Survey”. Computer Science TR 1530. University of Wisconsin – Madison. July 2008:
<http://pages.cs.wisc.edu/~jerryzhu/research/ssl/semireview.html>
- [Dempster, A.P.](#); [Laird, N.M.](#); [Rubin, D.B.](#) (1977). "Maximum Likelihood from Incomplete Data via the EM Algorithm". [Journal of the Royal Statistical Society, Series B](#). **39** (1): 1–38. [JSTOR 2984875](#). [MR 0501537](#).
- Blum, A., Mitchell, T. [Combining labeled and unlabeled data with co-training](#). *COLT: Proceedings of the Workshop on Computational Learning Theory*, Morgan Kaufmann, 1998, p. 92-100.
- [Hinton, G. E.](#); [Sejnowski, T. J.](#) (1986). «[Learning and Relearning in Boltzmann Machines](#)». En D. E. Rumelhart, J. L. McClelland, and the PDP Research Group. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Volume 1: Foundations* (en inglés) (Cambridge: [MIT Press](#)): 282-317. Archivado desde [el original](#) el 5 de julio de 2010.
- Máquinas de Boltzmann (Scholarpedia):
http://www.scholarpedia.org/article/Boltzmann_machine

¡Gracias!

