

Resumen: Learning Systems of Concepts with an Infinite Relational Model - Kemp, Tenenbaum, Griffiths

Mauricio Gonzalez Soto

13 de septiembre de 2017

1. Introducción

Si cierto dominio consiste en varios tipos de entidades,

2. Infinite Relational Models

Suponiendo que tenemos una o más relaciones que involucran uno o más tipos. El IRM busca partir cada tipo (type) en clusters, de manera que una buena partición permita que relaciones entre entidades se puedan predecir a partir de la asignación de clusters.

Por ejemplo, si tenemos m relaciones que involucran n tipos, denotamos R^i la i -ésima relación, T^j la j -ésima relación y z^j un vector de asignación de clusters. Lo que se busca es inferir la asignación de clusters, y ultimadamente la posterior $P(z^1, \dots, z^n | R^1, \dots, R^m)$. Para especificar esta distribución, se define un modelo generativo para las relaciones y la asignación de clusters

$$P(R^1, \dots, R^m, z^1, \dots, z^n) = \prod_{i=1}^m P(R^i | z^1, \dots, z^n) \prod_{j=1}^n P(z^j).$$

Estamos asumiendo que las relaciones son condicionalmente independientes dada la asignación de clusters, y que la asignación de clusters es independiente entre tipos.

3. Generación de clusters

Para que el IRM sea capaz de descubrir el número de clusters en el tipo T , se utiliza un prior que asigna probabilidad a todas las posibles particiones de T . Un prior razonable debería descubrir sólo tantos clusters como sugeridos por los datos. A partir de trabajos previos en Bayesiana No paramétrica ([Rasmussen, 2000]) utiliza una distribución sobre particiones inducida por un CRP (Chinese Restaurant Process, [Teh, 2011]).

La idea es ([Teh, 2011]) que en un restaurante chino hay un número infinito de mesas, cada una de las cuales puede sentar un número infinito de clientes. El primer cliente entra y se sienta en la primera mesa, el segundo cliente entra y decide si sentarse en la primera mesa o en otra; en general, el n -ésimo cliente se sienta en una mesa ocupada con una probabilidad proporcional al número de clientes sentados ahí.

Entonces, la manera de construir clusters utilizando el CRP es partir de un cluster con un solo objeto, e ir añadiendo objetos de modo que la probabilidad de caer en cierto cluster sea proporcional a los elementos en él. De esta manera, la distribución de clusters para el objeto i , dada la asignación de los objetos $1, \dots, i-1$ es:

$$P(z_i = a | z_1, \dots, z_{i-1}) = \begin{cases} \frac{n_a}{i-1+\gamma} & \text{if } n_a > 0 \\ \frac{\gamma}{i-1+\gamma} & a \text{ es un nuevo cluster} \end{cases}$$

Donde n_a es el número de objetos asignado al cluster a y γ es un parámetro. La distribución en z inducida por el CRP es intercambiable; es decir, no importa el orden en que los objetos llegan a los clusters. Como nuevos objetos siempre pueden ser asignados, el IRM efectivamente tiene acceso a un número infinito de clusters, de ahí el nombre del Infinite Relational Model.

4. Generación de relaciones a partir de clusters

Se asume que las relaciones son binarias, pero se puede extender a datos continuos y frecuencias. Como ejemplo, veremos primero el caso con un solo tipo T y una relación $R : T \times T \rightarrow \{0, 1\}$. T puede ser, por ejemplo, un conjunto de personas y R una relación de amistad. El modelo generativo para este problema es

$$\begin{aligned} z|\gamma &\sim CRP(\gamma) \\ \eta(a, b)|\beta &\sim Beta(\beta, \beta) \\ R(i, j)|z, \eta &\sim Bernoulli(\eta(z_i, z_j)) \end{aligned}$$

Esto es, estamos suponiendo que la relación R está generada a partir de dos estructuras latentes: una partición z y una matriz de parámetros η . La entrada $R(i, j)$ se genera al lanzar una moneda con sesgo $\eta(z_i, z_j)$ donde z_i, z_j son las asignaciones de cluster de las entidades i, j . El parámetro $\eta(a, b)$ especifica la probabilidad de que exista un link entre las entidades $i \in a, j \in b$. El IRM invierte ese modelo para descubrir la z y la η que mejor explican R .

Estamos asumiendo que la tendencia de una entidad a participar en relaciones está dada por su asignación de cluster.

5. Inferencia

Como en el ejemplo anterior se utilizó una prior conjugada, es sencillo calcular $P(R|z)$

$$P(R|z) = \prod_{a,b} \frac{Beta(m(a, b) + \beta, \bar{m}(a, b) + \beta)}{Beta(\beta, \beta)}$$

donde $m(a, b)$ es el número de pares (i, j) tales que $i \in a, j \in b$ y $R(i, j) = 1$ mientras que $\bar{m}(a, b)$ son las parejas donde $R(i, j) = 0$.

Para muestrear de la posterior $P(z|R) \propto P(R|z)P(z)$ se pueden utilizar técnicas MCMC o al buscar la moda de la distribución.

6. Aplicación

6.1. Objetos y atributos

Aunque el IRM está pensado para datos relacionales, también puede utilizarse para encontrar estructura en datos objeto-atributo. Notemos que cualquier matriz de atributos puede verse como una relación $R : T^1 \times T^2 \rightarrow \{0, 1\}$ entre un set de objetos T^1 y uno de atributos T^2 . También es importante notar que si se decide no clusterizar los atributos, sino asumir que cada uno se genera independientemente sobre las particiones entonces se obtiene el infinite mixture model de Rasmussen.

6.2. Ontologías

Se analizó una red semántica de 135 conceptos y 49 predicados binarios. Los conceptos son, por ejemplo “enfermedad”, “diagnóstico”, etc. Los predicados contienen verbos como “complica”, “causa”, “afecta”. Se aplicó el IRM a la relación ternaria $R : T^1 \times T^1 \times T^2 \rightarrow \{0, 1\}$ donde T^1 es conjunto de conceptos, T^2 conjunto de atributos.

Referencias

- [Rasmussen, 2000] Rasmussen, C. E. (2000). The infinite gaussian mixture model. In *Advances in neural information processing systems*, pages 554–560.
- [Teh, 2011] Teh, Y. W. (2011). Dirichlet process. In *Encyclopedia of machine learning*, pages 280–287. Springer.