Resumen: Jointly Identifying

Mauricio Gonzalez Soto

6 de septiembre de 2017

1. Intro

En este artículo se estudia el problema de reconocimiento de entidades y extracción de relaciones.

- El documento básico es un artículo, el cual suponemos que trata sobre una entidad, llamada la entidad principal
- El documento menciona entidades secundarias que están relacionadas a la principal, pero no entre sí
- Se trabaja con entidades pre-definidas así como relaciones

2. Formulación del problema

- Sea $\mathbf{x} = \{x_1, ..., x_N\}$ una sucesión observada de *tokens*.
- Sea s_p la entidad principal, que suponemos es claramente identificable
- Sea $\mathbf{s} = \{s_1, ..., s_L\}$ una segmentación, donde cada s_i es una tripleta $\{\alpha_i, \beta_i, y_i\}$ dode α_i es una posición inicial, β_i posición final y y_i una etiqueta asignada
- Debe cumplirse que $0 \leq \alpha_i \leq \beta_i \leq |x|$ y $\alpha_{i+1} = \beta_i + 1$
- Sea r_{pn} la asignación de relaciones entre la principal s_p y un candidato a entidad secundaria s_n y sea $\mathbf{r}$ el conjunto de asignación de relaciones para la $\mathbf{x}$
- Sea $\mathbf{y} = \{\mathbf{r}, \mathbf{s}\}$
- Dada una observación $\mathbf{x}$, queremos encontrar $\mathbf{y}^*$ tal que

$$\mathbf{y}^* = \text{ arg max }_y p(\mathbf{y}|\mathbf{x})$$

3. Modelo propuesto

- Sean L y M el número de segmentos y el número de relaciones para una secuencia observada $\mathbf{x}$
- Se define una distribución condicional conjunta para $\mathbf{s}, \mathbf{x}, \mathbf{r}$ en un modelo grafico no dirigido $\mathcal{G}$
- Se pueden dividir los factores de $\mathcal{G}$ en tres grupos: ϕ^S, ϕ^R, ϕ^∇
- Respectivamente, el potencial de segmentación, el potencial de relaciones, y el potencial conjunto de segmentación-relación.
- La función potencial $\phi^S(i, \mathbf{s}, \mathbf{x})$ modela la segmentación
- El potencial $\phi^R(r_{pm}, r_{pn}, \mathbf{r})$ representa dependencias entre relaciones en $\mathbf{r}$
- El potencial conjunto $\phi^\nabla(s_p, s_j, \mathbf{r})$ captura interacciones entre segmentaciones para entidades secundarias

- Por el Teorema de Hammersley-Clifford, la condicional conjunta $P(\mathbf{y}|\mathbf{x})$ se factoriza como

$$P(\mathbf{y}|\mathbf{x}) = \frac{1}{Z(\mathbf{x})} \prod_{C_S} \phi^S(i, \mathbf{s}, \mathbf{x}) \prod_{C_R} \phi^R(r_{pm}, r_{pn}, \mathbf{r}) \prod_{C_\nabla} \phi^\nabla(s_p, s_j, \mathbf{r})$$

- Tenemos que la función Z es el normalizante

$$Z(x) = \sum_{\mathbf{y}} P(\mathbf{y}|\mathbf{x}).$$

- Se hace el supuesto de que las funciones potencial se factorizan según un set de features y correspondientes pesos
- Con más detalle,

$$\phi^S(i, \mathbf{s}, \mathbf{x}) = \exp(\sum_{i=1}^{|s|} \sum_{k=1}^{K} \lambda_k g_k(i, \mathbf{s}, \mathbf{x}))$$

- Además, haremos que la función g_k dependa del segmento s_i, del anterior s_{i-1} y de toda la observación $\curvearrowleft$ de modo que

$$g_k(i, \mathbf{s}, \mathbf{x}) = g_k(s_{i-1}, s_i, \mathbf{x})$$

- De la misma manera para ϕ^R y ϕ^∇ tenemos que

$$\phi^R(r_{pm}, r_{pn}, \mathbf{r}) = \exp(\sum_{m,n}^{M} \sum_{w=1}^{W} \mu_w q_w(r_{pm}, r_{pn}, \mathbf{r}))$$

-

$$\phi^\nabla(s_p, s_j, \mathbf{r}) = \exp(\sum_{j=1}^{L} \sum_{t=1}^{T} \nu_t h_t(s_p, s_j, \mathbf{r}))$$

- Entonces, tenemos que

$$P(\mathbf{y}|\mathbf{x}) = \frac{1}{Z(\mathbf{x})} \exp\left(\sum_{i=1}^{|s|} \sum_{k=1}^{K} \lambda_k g_k(i, s, x) + \sum_{m,n}^{M} \sum_{w=1}^{W} \mu_w q_w(r_{pm}, r_{pn}, r) + \sum_{j=1}^{L} \sum_{t=1}^{T} \nu_t h_t(s_p, s_j, r) \right)$$

4. Aprender los parámetros

- Dada una muestra $\mathcal{D} = (\mathbf{x}_i, \mathbf{y}_i)_{i=1}^N$ queremos estimar los parámetros $(\lambda_k, \mu_w, \nu_t)$
- La log-verosimilitud regularizada es

$$\mathcal{L} = \log[\Phi(r, s, x)] - \log[Z(x)] - \sum_{k=1}^{K} \frac{\lambda_k^2}{2\sigma_\lambda^2} - \sum_{w=1}^{W} \frac{\mu_w^2}{2\sigma_w^2} - \sum_{t=1}^{T} \frac{\nu_t^2}{2\sigma_\nu^2}$$

- Donde,

$$\Phi(r, s, x) = \exp(\sum_{i=1}^{|s|} \sum_{k=1}^{K} \lambda_k g_k(i, s, x) + \sum_{m,n}^{M} \sum_{w=1}^{W} \mu_w q_w(r_{pm}, r_{pn}, r) + \sum_{j=1}^{L} \sum_{t=1}^{T} \nu_t h_t(s_p, s_j, r))$$

- Tenemos que $\mathcal{L}$ es concava por lo que es fácil maximizarla

5. Asignación más probable

- El objetivo es encontrar
$$\mathbf{y}^* = \text{ arg max }_y p(\mathbf{y}|\mathbf{x})$$
- No se puede resolver de manera exacta, se necesitan métodos aproximados
- Se propone decodificar las variables ocultas objetivo basados en asignar etiquetas a las variables muestreadas.
- Se hace en dos pasos: Primero, se predice una etiqueta inicial para una x_i dado el modelo ya entrenado
- En segundo lugar, se re-estima la asignación a x_i