

# INFORMACIÓN MUTUA

Salvador López Mendoza

Mayo de 2018

# MEDIR LA REDUCCIÓN EN LA ENTROPÍA

Se usa la información mutua ( $I(X; Y)$ ) para medir la reducción en la entropía.

¿Qué tanta reducción en la entropía de  $X$  se puede obtener por el hecho de conocer  $Y$ ?

Se define la *información mutua* como:

$$I(X; Y) = H(X) - H(X|Y)$$

# REDUCCIÓN EN LA ENTROPÍA

Propiedades:

- Es un valor no negativo:  $I(X; Y) \geq 0$ .
- Es una relación simétrica:  $I(X; Y) = I(Y; X)$ .
- $I(X; Y) = 0$  sí y sólo sí  $X$  e  $Y$  son independientes.

Si se fija el valor de  $X$  para clasificar distintas  $Y$ , tanto  $I(X; Y)$  como  $H(X|Y)$  resultan en el mismo orden.

Se prefiere  $I(X; Y)$  porque permite comparar distintas parejas  $(X, Y)$ .

# MINANDO RELACIONES SINTAGMÁTICAS

$$I(X; Y) = I(Y; X) = H(X) - H(X|Y) = H(Y) - H(Y|X)$$

Cada vez que aparece la palabra *come*, ¿qué otras palabras pueden aparecer?

¿Qué palabras tienen un valor alto de información mutua con *come*?

$$I(X_{come}; X_{carne}) = I(X_{carne}; X_{come}) > I(X_{come}; X_{el}) = I(X_{el}; X_{come})$$

$$I(X_{come}; X_{come}) = H(X_{come}) \geq I(X_{come}; X_w)$$

# ALGUNAS PROBABILIDADES

Presencia y ausencia de  $w_1$ :  $p(X_{w_1} = 1) + p(X_{w_1} = 0) = 1$

Presencia y ausencia de  $w_2$ :  $p(X_{w_2} = 1) + p(X_{w_2} = 0) = 1$

Co-ocurrencias de  $w_1$  y  $w_2$ :

$$p(X_{w_1} = 1, X_{w_2} = 1) + p(X_{w_1} = 1, X_{w_2} = 0) + \\ p(X_{w_1} = 0, X_{w_2} = 1) + p(X_{w_1} = 0, X_{w_2} = 0) = 1$$

Restricciones:

$$p(X_{w_1} = 1, X_{w_2} = 1) + p(X_{w_1} = 1, X_{w_2} = 0) = p(X_{w_1} = 1)$$

$$p(X_{w_1} = 0, X_{w_2} = 1) + p(X_{w_1} = 0, X_{w_2} = 0) = p(X_{w_1} = 0)$$

$$p(X_{w_1} = 1, X_{w_2} = 1) + p(X_{w_1} = 0, X_{w_2} = 1) = p(X_{w_2} = 1)$$

$$p(X_{w_1} = 1, X_{w_2} = 0) + p(X_{w_1} = 0, X_{w_2} = 0) = p(X_{w_2} = 0)$$

# CÁLCULO DE LA INFORMACIÓN MUTUA

No es necesario calcular todos los términos.

Al calcular la presencia y ausencia de  $w_1$ , basta con conocer  $p(X_{w_1} = 1)$ .

El valor de  $p(X_{w_1} = 0)$  se determina con ese término y despejando en la fórmula.

Sólo es necesario conocer  $p(X_{w_1} = 1)$ ,  $p(X_{w_2} = 1)$  y  $p(X_{w_1} = 1, X_{w_2} = 1)$

# ESTIMACIÓN DE PROBABILIDADES

Depende de los datos.

Se tienen  $N$  segmentos e interesa saber si aparecen las palabras  $w_1$  y  $w_2$ .

$$p(X_{w_1} = 1) = \text{count}(w_1)/N$$

$$p(X_{w_2} = 1) = \text{count}(w_2)/N$$

$$p(X_{w_1} = 1, X_{w_2} = 1) = \text{count}(w_1, w_2)/N$$

$\text{count}(w_1)$  es la cantidad total de segmentos en los que aparece  $w_1$

$\text{count}(w_2)$  es la cantidad total de segmentos en los que aparece  $w_2$

$\text{count}(w_1, w_2)$  es la cantidad total de segmentos en los que aparecen tanto  $w_1$  como  $w_2$ .

# SUAVIZAMIENTO

Agregar seudo datos para que ninguno de los posibles eventos tenga contadores con valor cero.

Se agrega un factor que representa el porcentaje de segmentos en los que aparece esa palabra.

También se agrega una unidad al coeficiente para compensar por el dato añadido.



- La relación sintagmática se puede descubrir al medir las correlaciones entre apariciones de dos palabras.
- Se toman tres conceptos de la Teoría de la Información:
  - Entropía ( $H(X)$ ).
  - Entropía condicional ( $H(X|Y)$ ).
  - Información mutua  $I(X; Y)$ .
- La información mutua es el principal mecanismo para descubrir relaciones sintagmáticas.

# RESUMEN DE ASOCIACIÓN DE PALABRAS

- Asociaciones paradigmáticas y asociaciones sintagmáticas.
- Enfoques totalmente estadísticos para descubrir estas relaciones.
- Las asociaciones que se descubren se pueden usar como la base de otras aplicaciones.