Disfluency Detection using Multiple-step Stacked Learning

Xian Qian and Yang Liu

Leandro Matayoshi

Conceptos

Disfluencia
=
Error al hablar

Multiple step stacked learning

Ejemplo

I want a flight to Boston uh I mean to Denver

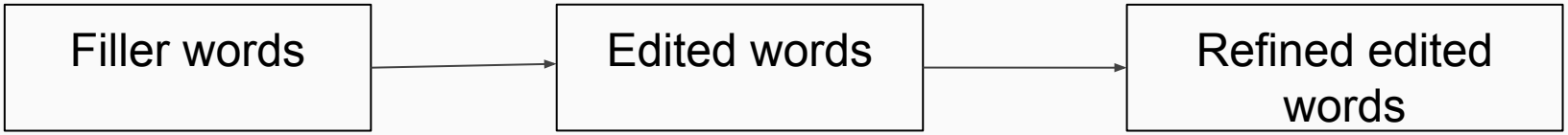
Ejemplo (estructura de disfluencia)

I want a flight *to Boston uh I mean* to Denver

Edited words

Filler words

Overview of the stack



M3N Classifier: Max-Margin Markov's Network.

Filler: Un-weighted hamming loss

Edited: Weighted hamming loss

Etiquetado: Edited words

I/O want/O a/O flight/O to/BE Boston/EE uh/O I/O mean/O to/O Denver/O

BE: Beginning of the multi-word edited region
IE: In the edited region
EE: End of the edited region
SE: Single word edited region
O: Other

Evaluation metric for edited words

$$P = \frac{\#\text{correctly predicted edited words}}{\#\text{predicted edited words}}$$

$$R = \frac{\#\text{correctly predicted edited words}}{\#\text{gold standard edited words}}$$

$$F = \frac{2 \times P \times R}{P + R}$$

N-gram features

Filler word detection

unigrams	$w_0, w_{-1}, w_1, w_{-2}, w_2$ $p_0, p_{-1}, p_1, p_{-2}, p_2, w_0p_0$
bigrams	$w_{-1}w_0, w_0w_1, p_{-1}p_0, p_0p_1$
trigrams	$p_{-2}p_{-1}p_0, p_{-1}p_0p_1, p_0p_1p_2$
logic unigrams	$I(w_i, w_0), I(p_i, p_0), \quad -4 \leq i \leq 4$
logic bigrams	$I(w_{i-1}w_i, w_{-1}, w_0)$ $I(p_{i-1}p_i, p_{-1}p_0)$ $I(w_iw_{i+1}, w_0w_1)$ $I(p_ip_{i+1}, p_0p_1), \quad -4 \leq i \leq 4$
transitions	$y_{-1}y_0$

Edited word detection

All templates in Table 1	
unigrams	w'_1, w'_2, w'_3, w'_4
bigrams	$p_0p'_1, p_0p'_2, p_0p'_3, p_0p'_4$ $w_0p'_1, w_0p'_2, w_0p'_3, w_0p'_4$ $w_0p_1, w_0p_2, w_0p_3, w_0p_4$
logic unigrams	$I(w_0, w'_i), \quad 1 \leq i \leq 4$
transitions	$p_0y_{-1}y_0$

w0, p0 denotan la palabra actual y el POS tag. *I*(a,b) indican si *a* o *b* son idénticas

3ra capa del stack. Refinamiento. Boundary Errors (1)

Correcto: The new type is prettier than what their/SE they used to look like

Posible output: The new type is prettier than what/BE their/EE they used to look like

than they es una expresión improbable y rara (incorrecta)

3ra capa del stack. Refinamiento. Boundary Errors (2)

Coordinates: They can't decide which are the good aspects and which are the bad aspects

Rough copies: It/BE 's/IE a/IE pleasure/IE to/EE it 's good to get outside

All templates in Table 1, Table 2	
word n-grams	$w''_1, w_0 w''_1$
in-between	$L_{AB}, w_0 b_{AB}, b_{AB}$

w'' _*i* = Etiqueta de la palabra *i* posiciones a la derecha
L_*AB* = Longitud de la sub-sequencia en el patrón AB....AB
b_AB = Indica si la palabra actual se encuentra en el medio de 2 bigramas iguales

Weight hamming loss Matrix for Max-Margin Markov's Network

truth \ predict	*BE*	*IE*	*EE*	*SE*	*O*
BE	0	1	1	1	2
IE	1	0	1	1	2
EE	1	1	0	1	2
SE	1	1	1	0	2
O	1	1	1	1	0

Table 4: Weighted hamming loss for M^3Ns.

Comparison and Results

	CRF	AP	PA	M^3N	w. M^3N
Baseline	78.8	79.0	78.9	79.4	80.1
Step 2	81.0	81.1	81.1	81.5	82.3
Step 3	82.9	83.0	82.8	83.3	**84.1**

Table 5: Effect of training strategy and recovered features for stacked learning. F scores are reported. AP = Averaged Perceptron, PA = online Passive Aggresive, M^3N = un-weighted M^3Ns, w. M^3N = weighted M^3Ns.

Preguntas?