

LAMA-WeST Lab

Artificial Intelligence
Natural language processing
Semantic web



Création de négatifs dans le TP2

INF8460 - Traitement automatique de la langue naturelle

Edouard Albert-Roulhac
Polytechnique Montréal

16 septembre 2024

L'objectif du TP2 est d'**entraîner un modèle de plongements** qui puisse distinguer les synonymes des antonymes. Pour ce faire, il nous allons **rapprocher les mots similaires** et **éloigner les mots non similaires**.



Pendant l'entraînement, nous utilisons des lots (batches) de paires de synonymes B_S et d'antonymes B_A .

- ▶ **Négatif d'un synonyme** : Le mot du lot le plus proche qui n'est pas dans la même paire.
- ▶ **Négatif d'un antonyme** : Le mot du lot le plus éloigné qui n'est pas dans la même paire.



Considérons un lot B_S de synonymes de taille 3. Nous voulons construire le lot de négatifs T_S :

- ▶ **Lot de synonymes B_S :**
 - ▶ (arbre, plante)
 - ▶ (voiture, véhicule)
 - ▶ (arbre, buisson)
- ▶ **On en déduit les voisins**, tout mot apparaissant dans au moins une paire en commun (et un mot est son propre voisin) :
 - ▶ **Voisins de arbre** : arbre, plante, buisson
 - ▶ **Voisins de plante** : plante, arbre
 - ▶ **Voisins de voiture** : voiture, véhicule
 - ▶ **Voisins de véhicule** : véhicule, voiture
 - ▶ **Voisins de buisson** : buisson, arbre



	arbre	plante	voiture	véhicule	buisson
arbre	1.0	0.8	0.1	0.2	0.9
plante	0.8	1.0	0.3	0.4	0.7
voiture	0.1	0.3	1.0	0.9	0.2
véhicule	0.2	0.4	0.9	1.0	0.3
buisson	0.9	0.7	0.2	0.3	1.0



	arbre	plante	voiture	véhicule	buisson
arbre	$-\infty$	$-\infty$	0.1	0.2	$-\infty$
plante	$-\infty$	$-\infty$	0.3	0.4	0.7
voiture	0.1	0.3	$-\infty$	$-\infty$	0.2
véhicule	0.2	0.4	$-\infty$	$-\infty$	0.3
buisson	$-\infty$	0.7	0.2	0.3	$-\infty$

Les valeurs $-\infty$ indiquent que les mots correspondants sont des voisins et ne peuvent pas être sélectionnés comme négatifs.



Pour chaque mot, nous sélectionnons le négatif comme étant le mot avec la plus grande similarité qui n'est pas un voisin, c'est-à-dire avec le maximum dans une ligne.

- ▶ **Négatif de arbre** : véhicule (similarité 0.2)
- ▶ **Négatif de plante** : buisson (similarité 0.7)
- ▶ **Négatif de voiture** : plante (similarité 0.3)
- ▶ **Négatif de véhicule** : plante (similarité 0.4)
- ▶ **Négatif de buisson** : plante (similarité 0.7)



Nous construisons maintenant le lot de négatifs T_S en utilisant les mots négatifs trouvés pour chaque mot du lot de synonymes B_S .

- ▶ (arbre, plante) \rightarrow (véhicule, buisson)
- ▶ (voiture, véhicule) \rightarrow (plante, plante)
- ▶ (arbre, buisson) \rightarrow (véhicule, plante)

Ainsi, le lot de négatifs T_S sera :

$$\{(\text{véhicule, buisson}), (\text{plante, plante}), (\text{véhicule, plante})\}$$

