

DEEP LEARNING POUR L'INDEXATION CONCEPTUELLE DE TEXTE

Application à l'extraction d'information

31 janvier 2019

Étienne Simon

Direction: Patrick Gallinari
Encadrement: Benjamin Piwowarski
Vincent Guigue

LIP6 - Sorbonne Université

Compréhension du langage naturel

- Modèle de langage
- Systèmes de questions-réponses
- ...

Extraction d'Information

- Reconnaissance d'entités nommées
- Analyse de sentiment
- Classification de documents
- Extraction de relations
- ...

Transition de modèles basés sur des caractéristiques manuellement définies vers des modèles où l'extraction des caractéristiques est apprise.

Le sol a été la monnaie officielle du Pérou entre 1863 et 1985.



(sol, devise utilisée par, Pérou)

Le sol_{e1} a été la monnaie officielle du Pérou_{e2} entre 1863 et 1985.



(sol, devise utilisée par, Pérou)

Si (e_1, r, e_2) vrai et que e_1, e_2 apparaissent dans une phrase, alors cette phrase exprime r .

Yann LeCun _{e_1} , né le 8 juillet 1960 à Paris _{e_2} , est un chercheur en intelligence artificielle.

(Yann LeCun, lieu de naissance, Paris)



Mintz, Mike, et al. "Distant supervision for relation extraction without labeled data." *ACL 2009*

Si (e_1, r, e_2) vrai et que e_1, e_2 apparaissent dans une phrase, alors cette phrase exprime r .

Yann LeCun _{e_1} , né le 8 juillet 1960 à Paris _{e_2} , est un chercheur en intelligence artificielle.



(Yann LeCun, lieu de naissance, Paris)



Mintz, Mike, et al. "Distant supervision for relation extraction without labeled data." *ACL 2009*

Si (e_1, r, e_2) vrai et que e_1, e_2 apparaissent dans une phrase, alors cette phrase exprime r .

Yann LeCun _{e_1} , né le 8 juillet 1960 à Paris _{e_2} , est un chercheur en intelligence artificielle.



(Yann LeCun, lieu de naissance, Paris)

Le Baron Haussmann _{e_1} a dirigé les transformations de Paris _{e_2} sous le Second Empire.



(Baron Haussmann, lieu de naissance, Paris)




Mintz, Mike, et al. "Distant supervision for relation extraction without labeled data." *ACL 2009*

OpenIE: Utilise le langage comme domaine des relations.

 Banko, Michele, et al. "Open information extraction from the web." *IJCAI 2007*

rel-LDA: Modèle génératif semblable à LDA.

 Yao, Limin, Sebastian Riedel, and Andrew McCallum. "Unsupervised relation discovery with sense disambiguation." *ACL 2012*

VAE: Modèle discriminant à base d'auto encodeur variationnel.
La phrase est la variable observée, la relation la variable cachée.

 Marcheggiani, Diego, and Ivan Titov. "Discrete-state variational autoencoders for joint discovery and factorization of relations." *TACL 2016*

Le sol_{e1} a été la monnaie officielle du ?_{e2} entre 1863 et 1985.

Le sol_{e1} a été la monnaie officielle du ?_{e2} entre 1863 et 1985.

→ Pour compléter la phrase il faut comprendre la relation exprimée.

Le sol_{e₁} a été la monnaie officielle du ?_{e₂} entre 1863 et 1985.

→ Pour compléter la phrase il faut comprendre la relation exprimée.

$$p(e_{-i} \mid s, e_i)$$

Le sol_{e₁} a été la monnaie officielle du ?_{e₂} entre 1863 et 1985.

→ Pour compléter la phrase il faut comprendre la relation exprimée.

$$p(e_{-i} | s, e_i)$$

$$\underbrace{p(e_{-i} | r, e_i)}_{\text{prédicteur de lien}}$$

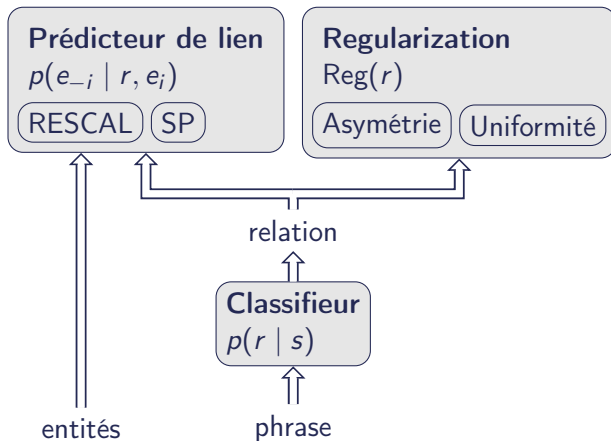
Prédiction de lien: (sol, devise utilisée par, ?)

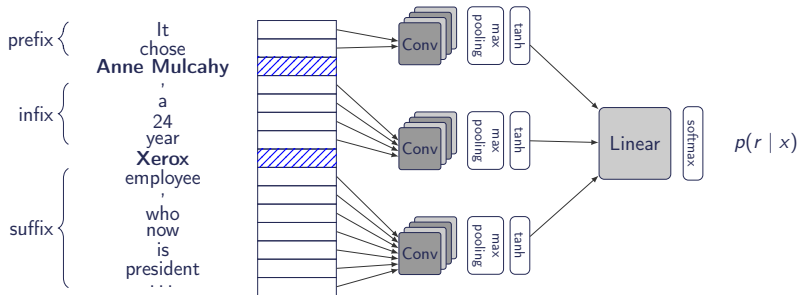
Le sol_{e₁} a été la monnaie officielle du ?_{e₂} entre 1863 et 1985.

→ Pour compléter la phrase il faut comprendre la relation exprimée.

$$p(e_{-i} | s, e_i) = \sum_r \underbrace{p(r | s)}_{\text{classifieur}} \underbrace{p(e_{-i} | r, e_i)}_{\text{prédicteur de lien}}$$

Prédiction de lien: (sol, devise utilisée par, ?)





$$p(e_1 | r, e_2) = \frac{\exp(\psi(e_1, r, e_2))}{\sum_{e' \in E} \exp(\psi(e', r, e_2))}$$

$$p(e_1 | r, e_2) = \frac{\exp(\psi(e_1, r, e_2))}{\sum_{e' \in E} \exp(\psi(e', r, e_2))}$$

Negative Sampling

$$p(D = 1 | e, r) = \sigma(\psi(e_1, r, e_2))$$

$$p(e_1 | r, e_2) = \frac{\exp(\psi(e_1, r, e_2))}{\sum_{e' \in E} \exp(\psi(e', r, e_2))}$$

Negative Sampling

$$p(D = 1 | e, r) = \sigma(\psi(e_1, r, e_2))$$

$$\mathcal{L}_{\text{LP}} = \mathbb{E}_{x \sim \mathcal{X}} \left[-2 \log p(D = 1 | e_1, r, e_2) \right. \\ \left. - \sum_{i=1}^k \mathbb{E}_{e' \sim \mathcal{E}} [\log p(D = 0 | e_1, r, e')] \right. \\ \left. - \sum_{i=1}^k \mathbb{E}_{e' \sim \mathcal{E}} [\log p(D = 0 | e', r, e_2)] \right]$$

$$p(e_1 | r, e_2) = \frac{\exp(\psi(e_1, r, e_2))}{\sum_{e' \in E} \exp(\psi(e', r, e_2))}$$

Negative Sampling

$$p(D = 1 | e, r) = \sigma(\psi(e_1, r, e_2))$$

$$\mathcal{L}_{\text{LP}} = \mathbb{E}_{x \sim \chi} \left[-2 \log \sigma(\psi(e_1, r, e_2)) \right. \\ \left. - \sum_{i=1}^k \mathbb{E}_{e' \sim \mathcal{E}} [\log \sigma(-\psi(e_1, r, e'))] \right. \\ \left. - \sum_{i=1}^k \mathbb{E}_{e' \sim \mathcal{E}} [\log \sigma(-\psi(e', r, e_2))] \right]$$

RESCAL

Encode l'interaction entre les entités deux à deux :

$$\psi^{\text{RS}}(e_1, r, e_2) = \mathbf{u}_{e_1}^T \mathcal{A}_r \mathbf{u}_{e_2}$$

RESCAL

Encode l'interaction entre les entités deux à deux :

$$\psi^{\text{RS}}(e_1, r, e_2) = \mathbf{u}_{e_1}^T \mathcal{A}_r \mathbf{u}_{e_2}$$

Selectional preferences

Encode l'interaction entre la relation et les entités :

$$\psi^{\text{SP}}(e_1, r, e_2) = \mathbf{u}_{e_1}^T B_r + \mathbf{u}_{e_2}^T C_r$$

RESCAL

Encode l'interaction entre les entités deux à deux :

$$\psi^{\text{RS}}(e_1, r, e_2) = \mathbf{u}_{e_1}^T \mathcal{A}_r \mathbf{u}_{e_2}$$

Selectional preferences

Encode l'interaction entre la relation et les entités :

$$\psi^{\text{SP}}(e_1, r, e_2) = \mathbf{u}_{e_1}^T B_r + \mathbf{u}_{e_2}^T C_r$$

Hybrid

$$\psi(e_1, r, e_2) = \psi^{\text{RS}}(e_1, r, e_2) + \psi^{\text{SP}}(e_1, r, e_2)$$

Asymétrie

(0.19 0.20 0.18 0.22 0.21)

Asymétrie

(0.19 0.20 0.18 0.22 0.21)

(0 0 0 1 0)

Asymétrie

$$(0.19 \quad 0.20 \quad 0.18 \quad 0.22 \quad 0.21) \quad \mathcal{L}_S = \mathbb{E}_{x \sim \mathcal{X}} [H(r | x)] \quad (0 \quad 0 \quad 0 \quad 1 \quad 0)$$

Asymétrie

$$(0.19 \quad 0.20 \quad 0.18 \quad 0.22 \quad 0.21) \quad \mathcal{L}_S = \mathbb{E}_{x \sim \mathcal{X}} [H(r | x)] \quad (0 \quad 0 \quad 0 \quad 1 \quad 0)$$

Uniformité

$$\begin{pmatrix} 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

Asymétrie

$$(0.19 \quad 0.20 \quad 0.18 \quad 0.22 \quad 0.21) \quad \mathcal{L}_S = \mathbb{E}_{x \sim \mathcal{X}} [H(r | x)] \quad (0 \quad 0 \quad 0 \quad 1 \quad 0)$$

Uniformité

$$\begin{pmatrix} 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

$$\begin{pmatrix} 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Asymétrie

$$(0.19 \quad 0.20 \quad 0.18 \quad 0.22 \quad 0.21) \quad \mathcal{L}_S = \mathbb{E}_{x \sim \mathcal{X}} [H(r \mid x)] \quad (0 \quad 0 \quad 0 \quad 1 \quad 0)$$

Uniformité

$$\begin{pmatrix} 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix} \quad \mathcal{L}_U = -D_{\text{KL}}(p(r) \parallel U) \quad \begin{pmatrix} 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{pmatrix}$$

NYT+FB: Articles du New York Times alignés avec Freebase

Entités: New York, United States, Manhattan

Relations: containedby, company, place_lived

T-REx: Articles de Wikipédia alignés avec Wikidata

Entités: India, United States, Department, Iran

Relations: country, located in, instance of, borders

Nom	phrases	étiquetés	relations	entités
NYT+FB	2 M	42 k	262	749k
T-REx DS	12 M	12 M	1189	773k
T-REx SPO	763 k	763 k	615	227k

Dataset	Model	B ³			V-measure			ARI
		F ₁	Prec.	Rec.	F ₁	Hom.	Comp.	
NYT+FB	rel-LDA	29.6						
	$\mathcal{L}_{VAE} - \mathcal{L}_S$	35.8	23.4	67.1	27.0	18.6	49.6	18.7
	$\mathcal{L}_{VAE} + \mathcal{L}_S + \mathcal{L}_U$	36.8	24.9	70.2	23.8	15.9	47.3	19.1
	$\mathcal{L}_{LP} - \mathcal{L}_S$	34.8	29.8	41.7	33.1	28.3	39.9	24.5
	$\mathcal{L}_{LP} + \mathcal{L}_S + \mathcal{L}_U$	39.1	32.0	50.4	38.3	32.1	47.3	33.3
T-REx SPO	$\mathcal{L}_{VAE} - \mathcal{L}_S$	21.7	15.4	36.6				
	$\mathcal{L}_{LP} + \mathcal{L}_S + \mathcal{L}_U$	39.9	32.1	52.9	44.1	36.1	56.7	26.0
T-REx DS	$\mathcal{L}_{LP} + \mathcal{L}_S + \mathcal{L}_U$	18.9	13.3	32.6	25.6	20.0	35.5	8.4

rel-LDA: Modèle génératif (Yao, 2012)

\mathcal{L}_{VAE} : Reconstruction VAE (Marcheggiani, 2016)

\mathcal{L}_{LP} : Prédiction de lien

\mathcal{L}_S : Asymétrie

\mathcal{L}_U : Uniformité

Dataset	Model	B ³			V-measure			ARI
		F ₁	Prec.	Rec.	F ₁	Hom.	Comp.	
NYT+FB	rel-LDA	29.6						
	$\mathcal{L}_{VAE} - \mathcal{L}_S$	35.8	23.4	67.1	27.0	18.6	49.6	18.7
	$\mathcal{L}_{VAE} + \mathcal{L}_S + \mathcal{L}_U$	36.8	24.9	70.2	23.8	15.9	47.3	19.1
	$\mathcal{L}_{LP} - \mathcal{L}_S$	34.8	29.8	41.7	33.1	28.3	39.9	24.5
	$\mathcal{L}_{LP} + \mathcal{L}_S + \mathcal{L}_U$	39.1	32.0	50.4	38.3	32.1	47.3	33.3
T-REx SPO	$\mathcal{L}_{VAE} - \mathcal{L}_S$	21.7	15.4	36.6				
	$\mathcal{L}_{LP} + \mathcal{L}_S + \mathcal{L}_U$	39.9	32.1	52.9	44.1	36.1	56.7	26.0
T-REx DS	$\mathcal{L}_{LP} + \mathcal{L}_S + \mathcal{L}_U$	18.9	13.3	32.6	25.6	20.0	35.5	8.4

rel-LDA: Modèle génératif (Yao, 2012)

\mathcal{L}_{VAE} : Reconstruction VAE (Marcheggiani, 2016)

\mathcal{L}_{LP} : Prédiction de lien

\mathcal{L}_S : Asymétrie

\mathcal{L}_U : Uniformité

Dataset	Model	B ³			V-measure			ARI
		F ₁	Prec.	Rec.	F ₁	Hom.	Comp.	
NYT+FB	rel-LDA	29.6						
	$\mathcal{L}_{VAE} - \mathcal{L}_S$	35.8	23.4	67.1	27.0	18.6	49.6	18.7
	$\mathcal{L}_{VAE} + \mathcal{L}_S + \mathcal{L}_U$	36.8	24.9	70.2	23.8	15.9	47.3	19.1
	$\mathcal{L}_{LP} - \mathcal{L}_S$	34.8	29.8	41.7	33.1	28.3	39.9	24.5
	$\mathcal{L}_{LP} + \mathcal{L}_S + \mathcal{L}_U$	39.1	32.0	50.4	38.3	32.1	47.3	33.3
T-REx SPO	$\mathcal{L}_{VAE} - \mathcal{L}_S$	21.7	15.4	36.6				
	$\mathcal{L}_{LP} + \mathcal{L}_S + \mathcal{L}_U$	39.9	32.1	52.9	44.1	36.1	56.7	26.0
T-REx DS	$\mathcal{L}_{LP} + \mathcal{L}_S + \mathcal{L}_U$	18.9	13.3	32.6	25.6	20.0	35.5	8.4

rel-LDA: Modèle génératif (Yao, 2012)

\mathcal{L}_{VAE} : Reconstruction VAE (Marcheggiani, 2016)

\mathcal{L}_{LP} : Prédiction de lien

\mathcal{L}_S : Asymétrie

\mathcal{L}_U : Uniformité

Dataset	Model	B ³			V-measure			ARI
		F ₁	Prec.	Rec.	F ₁	Hom.	Comp.	
NYT+FB	rel-LDA	29.6						
	$\mathcal{L}_{VAE} - \mathcal{L}_S$	35.8	23.4	67.1	27.0	18.6	49.6	18.7
	$\mathcal{L}_{VAE} + \mathcal{L}_S + \mathcal{L}_U$	36.8	24.9	70.2	23.8	15.9	47.3	19.1
	$\mathcal{L}_{LP} - \mathcal{L}_S$	34.8	29.8	41.7	33.1	28.3	39.9	24.5
	$\mathcal{L}_{LP} + \mathcal{L}_S + \mathcal{L}_U$	39.1	32.0	50.4	38.3	32.1	47.3	33.3
T-REx SPO	$\mathcal{L}_{VAE} - \mathcal{L}_S$	21.7	15.4	36.6				
	$\mathcal{L}_{LP} + \mathcal{L}_S + \mathcal{L}_U$	39.9	32.1	52.9	44.1	36.1	56.7	26.0
T-REx DS	$\mathcal{L}_{LP} + \mathcal{L}_S + \mathcal{L}_U$	18.9	13.3	32.6	25.6	20.0	35.5	8.4

rel-LDA: Modèle génératif (Yao, 2012)

\mathcal{L}_{VAE} : Reconstruction VAE (Marcheggiani, 2016)

\mathcal{L}_{LP} : Prédiction de lien

\mathcal{L}_S : Asymétrie

\mathcal{L}_U : Uniformité



- Finir le papier ACL (en particulier l'analyse qualitative)

- Finir le papier ACL (en particulier l'analyse qualitative)
- Nettoyer les équations : lier le negative sampling avec l'équation du modèle global

- Finir le papier ACL (en particulier l'analyse qualitative)
- Nettoyer les équations : lier le negative sampling avec l'équation du modèle global
- Remplacer le classifieur par un modèle plus récent (BERT)

- Finir le papier ACL (en particulier l'analyse qualitative)
- Nettoyer les équations : lier le negative sampling avec l'équation du modèle global
- Remplacer le classifieur par un modèle plus récent (BERT)
- Essayer d'autres fonctions d'énergies

- Finir le papier ACL (en particulier l'analyse qualitative)
- Nettoyer les équations : lier le negative sampling avec l'équation du modèle global
- Remplacer le classifieur par un modèle plus récent (BERT)
- Essayer d'autres fonctions d'énergies
- Utiliser des données semi supervisées pour étiqueter les clusters

- Finir le papier ACL (en particulier l'analyse qualitative)
- Nettoyer les équations : lier le negative sampling avec l'équation du modèle global
- Remplacer le classifieur par un modèle plus récent (BERT)
- Essayer d'autres fonctions d'énergies
- Utiliser des données semi supervisées pour étiqueter les clusters
- Multi-relation

- Finir le papier ACL (en particulier l'analyse qualitative)
- Nettoyer les équations : lier le negative sampling avec l'équation du modèle global
- Remplacer le classifieur par un modèle plus récent (BERT)
- Essayer d'autres fonctions d'énergies
- Utiliser des données semi supervisées pour étiqueter les clusters
- Multi-relation
- Multi-instance

- Finir le papier ACL (en particulier l'analyse qualitative)
- Nettoyer les équations : lier le negative sampling avec l'équation du modèle global
- Remplacer le classifieur par un modèle plus récent (BERT)
- Essayer d'autres fonctions d'énergies
- Utiliser des données semi supervisées pour étiqueter les clusters
- Multi-relation
- Multi-instance
- Gérer une ontologie des relations

- Finir le papier ACL (en particulier l'analyse qualitative)
- Nettoyer les équations : lier le negative sampling avec l'équation du modèle global
- Remplacer le classifieur par un modèle plus récent (BERT)
- Essayer d'autres fonctions d'énergies
- Utiliser des données semi supervisées pour étiqueter les clusters
- Multi-relation
- Multi-instance
- Gérer une ontologie des relations
- Poser d'autres hypothèses permettant d'entraîner des modèles non supervisés

Questions / Suggestions ?