

Apprentissage d'embeddings de phrases

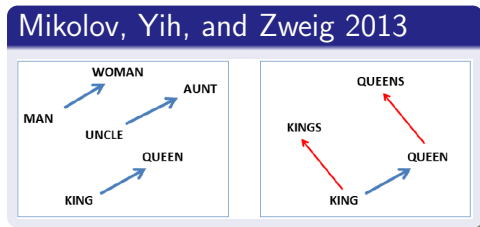
Jeremie Zumer

Université de Montréal

2015/04/16

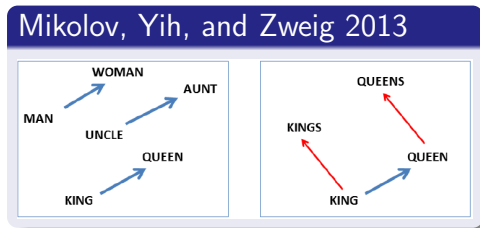
Embeddings: pourquoi

- Information sémantique



Embeddings: pourquoi

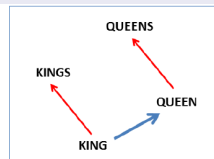
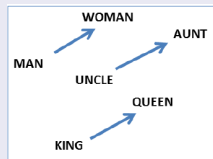
- Information sémantique
- Moins de bruit



Embeddings: pourquoi

- Information sémantique
- Moins de bruit
- Moins de travail manuel

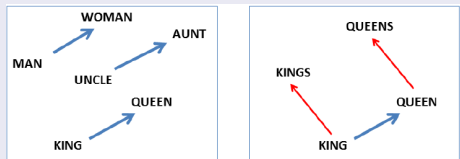
Mikolov, Yih, and Zweig 2013



Embeddings: pourquoi

- Information sémantique
- Moins de bruit
- Moins de travail manuel
- Méthodes génériques

Mikolov, Yih, and Zweig 2013



- Yang, Cui, and Tu 2015

- Yang, Cui, and Tu 2015
- Modélise l'espace des embeddings avec des gaussiennes

- Yang, Cui, and Tu 2015
- Modélise l'espace des embeddings avec des gaussiennes
- EM pour l'apprentissage (lent)

- $p(x|\lambda) = \sum_{i=1}^T \pi_i \mathcal{N}(x|\mu_i, \Sigma_i)$

- $p(x|\lambda) = \sum_{i=1}^T \pi_i \mathcal{N}(x|\mu_i, \Sigma_i)$
- $p(w_i = w | w_{i-m}, \dots, w_{i-1}, d, p) \propto \exp\{c_{document} + c_{phrase} + \sum c_{historique} + b\}$

- $p(x|\lambda) = \sum_{i=1}^T \pi_i \mathcal{N}(x|\mu_i, \Sigma_i)$
- $p(w_i = w | w_{i-m}, \dots, w_{i-1}, d, p) \propto \exp\{c_{document} + c_{phrase} + \sum c_{historique} + b\}$
- 2 phases

- $p(x|\lambda) = \sum_{i=1}^T \pi_i \mathcal{N}(x|\mu_i, \Sigma_i)$
- $p(w_i = w | w_{i-m}, \dots, w_{i-1}, d, p) \propto \exp\{c_{document} + c_{phrase} + \sum c_{historique} + b\}$
- 2 phases
 - Estimer les paramètres

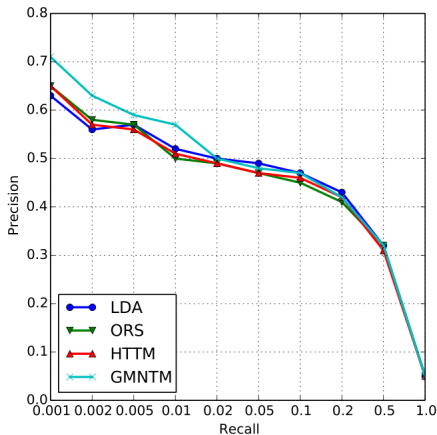
- $p(x|\lambda) = \sum_{i=1}^T \pi_i \mathcal{N}(x|\mu_i, \Sigma_i)$
- $p(w_i = w | w_{i-m}, \dots, w_{i-1}, d, p) \propto \exp\{c_{document} + c_{phrase} + \sum c_{historique} + b\}$
- 2 phases
 - Estimer les paramètres
 - Estimer les facteurs et vecteurs

- $p(x|\lambda) = \sum_{i=1}^T \pi_i \mathcal{N}(x|\mu_i, \Sigma_i)$
- $p(w_i = w | w_{i-m}, \dots, w_{i-1}, d, p) \propto \exp\{c_{document} + c_{phrase} + \sum c_{historique} + b\}$
- 2 phases
 - Estimer les paramètres
 - Estimer les facteurs et vecteurs
- Probabilité d'un sujet $q(z(k) = y) \propto \pi_y \mathcal{N}(\text{vec}(k)|\mu_y, \Sigma_y)$

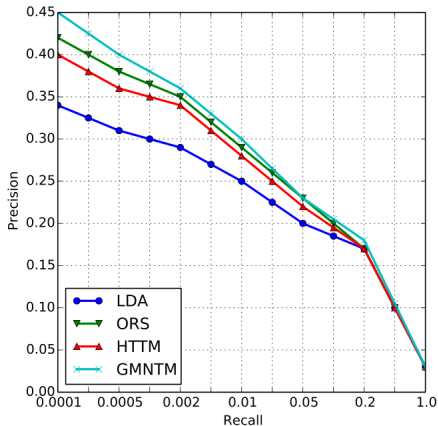
- $p(x|\lambda) = \sum_{i=1}^T \pi_i \mathcal{N}(x|\mu_i, \Sigma_i)$
- $p(w_i = w | w_{i-m}, \dots, w_{i-1}, d, p) \propto \exp\{c_{document} + c_{phrase} + \sum c_{historique} + b\}$
- 2 phases
 - Estimer les paramètres
 - Estimer les facteurs et vecteurs
- Probabilité d'un sujet $q(z(k) = y) \propto \pi_y \mathcal{N}(\text{vec}(k) | \mu_y, \Sigma_y)$
- Pour prédiction: probabilité du sujet $\propto q(z(w) = y), q(z(p) = y), q(z(d) = y)$

Résultats

20 Newsgroups



Reuters RCV1-V2



GMNTM Topic words				LDA Topic Words			
god	space	game	key	god	space	year	key
jesus	orbit	play	public	believe	nasa	hockey	encryption
christ	earth	season	encryption	jesus	research	team	use
believe	solar	team	security	sin	center	division	des
christian	spacecraft	win	escrow	one	shuttle	league	system
bible	surface	hockey	secure	mary	launch	nhl	rsa
lord	planet	hand	data	lord	station	last	public
truth	mission	series	privacy	would	orbit	think	security
sin	satellite	chance	government	christian	april	maria	nsa
faith	shuttle	nhl	nsa	accept	satellite	see	secure

- Backpropagation (rapide)

- Backpropagation (rapide)
- Moins contraint

- Backpropagation (rapide)
- Moins contraint
- Purement génératif

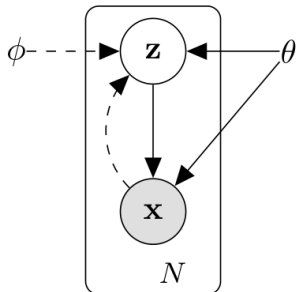
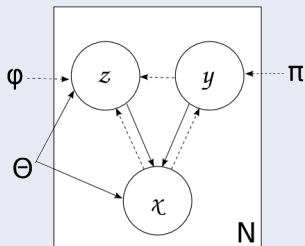


Figure: Kingma and Welling 2013

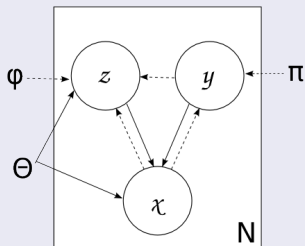
- Ajouter un chemin pour la prédiction de classe

Le modèle



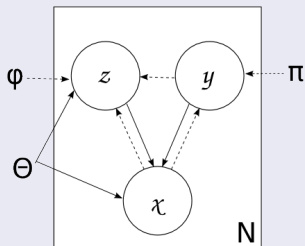
- Ajouter un chemin pour la prédiction de classe
- Utiliser la prédiction de classe durant la reconstruction

Le modèle



- Ajouter un chemin pour la prédiction de classe
- Utiliser la prédiction de classe durant la reconstruction
- Variables latentes encodent le style d'écriture

Le modèle



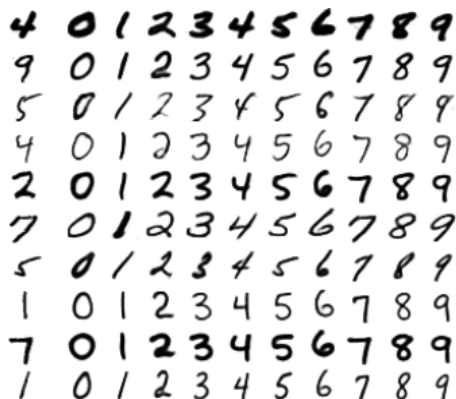


Figure: Kingma et al. 2014

- Utiliser des embeddings est avantageux

Conclusion

- Utiliser des embeddings est avantageux
- Pas (encore) de résultats avec les VAE

- Utiliser des embeddings est avantageux
- Pas (encore) de résultats avec les VAE
- A suivre. . .

- Kingma, D. P and M. Welling (2013). “Auto-Encoding Variational Bayes”. In: *ArXiv e-prints*. arXiv: 1312.6114 [stat.ML].
- Kingma, Diederik P. et al. (2014). “Semi-Supervised Learning with Deep Generative Models”. In: *CoRR* abs/1406.5298. URL: <http://arxiv.org/abs/1406.5298>.
- Mikolov, Tomas, Wen tau Yih, and Geoffrey Zweig (2013). “Linguistic Regularities in Continuous Space Word Representations.” In: *HLT-NAACL*. The Association for Computational Linguistics, pp. 746–751. URL: <http://dblp.uni-trier.de/db/conf/naacl/naacl2013.html#MikolovYZ13>.
- Yang, Min, Tianyi Cui, and Wenting Tu (2015). “Ordering-sensitive and Semantic-aware Topic Modeling”. In: *CoRR* abs/1502.03630. URL: <http://arxiv.org/abs/1502.03630>.