

Analyse de sentiment par réseaux neuronaux récurrents

Bertrand Rondepierre & Thomas Moreau

Télécom paristech - MDI343

13 mai 2014

Overview

- 1 Analyse de sentiment
 - Les objectifs
- 2 L'apprentissage du model
- 3 Les résultats
- 4 Auto encoder

Les Objectifs

- Catégoriser l'opinion général exprimé par une phrase
- Plusieurs niveaux (analyse binaire / fine)
- Représenter la phrase dans un espace propre qui permette de mettre en lumière l'opinion qu'elle contient
- Généré des phrases ?

Le Stanford Tree bank

- 11 855 critiques de film
- Labels et structure de la phrase sous forme d'arbre
⇒ Possibilité d'analyse fine à chaque noeuds

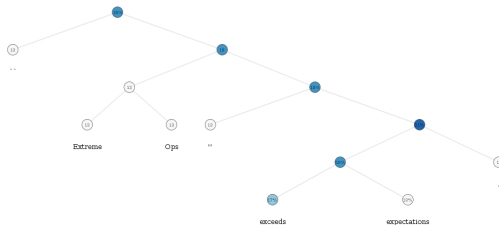


FIGURE: "Extreme Ops" exceeds expectations.

- **Classique** : Bag of words
- N'analyse pas la structure de la phrase (négation, expression)
This movie was actually neither that funny, nor super witty.
- L'idée est d'utiliser la structure d'arbre pour predire le sentiment de la phrase.
- On cherche aussi a optimiser la représentation des mots

- $E = \sum_{phrase} \sum_{nodes} t_i \log y_i.$
- Dans le papier, met les labels en dimension 5
⇒ suppose equi distance entre les labels
- Notre approche : regression et attribution d'un label avec une frontière fixe.

- Calcul du gradient par back propagation

$$\frac{\partial E}{\partial \theta} = \frac{t_i}{y_i} F'(y_i) \left[\frac{\partial W}{\partial \theta} x_{i-1} + \frac{\partial V}{\partial \theta} x_{i-1} x_{i-1}^T + (2Vx_{i-1} + W) \frac{\partial x_{i-1}}{\partial \theta} \right]$$

- Notre model implémente Ada-grad
⇒ Principe est de réduire le learning rate des poids qui ont déjà beaucoup été updaté.
- Autre solution : Rms prop
⇒ On tune le learning rate en fonction de la dynamique du gradient.

- Learning rate, mini batch size, regularisation

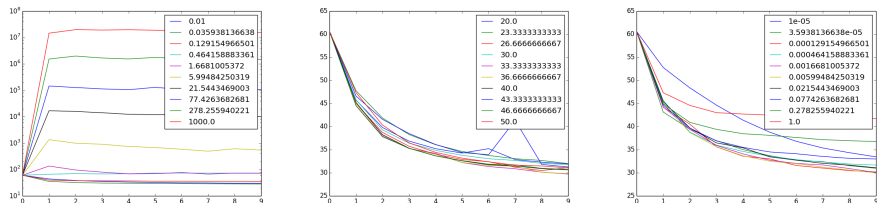


FIGURE: Courbe d'apprentissage en Cross validation pour le learning rate, la taille du mini batch et le facteur de régularisation

Apprentissage

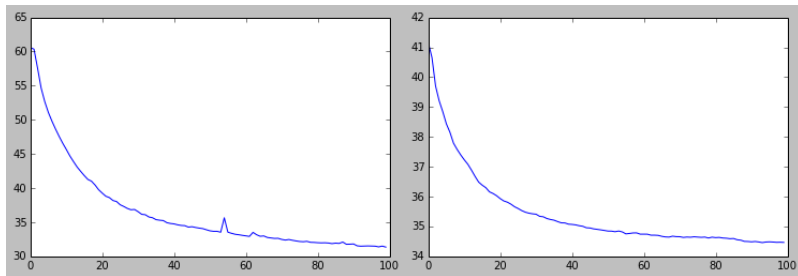


FIGURE: Courbe d'apprentissage pour AdaGrad (*gauche*) et Rprop (*droite*)

Resultat de Classification

	Fine		Binaire	
	All node	Root	All nodes	Root
Socher	80.7	45.7	87.6	85.4
Notre modèle	79.7	42.2	86.7	83.3

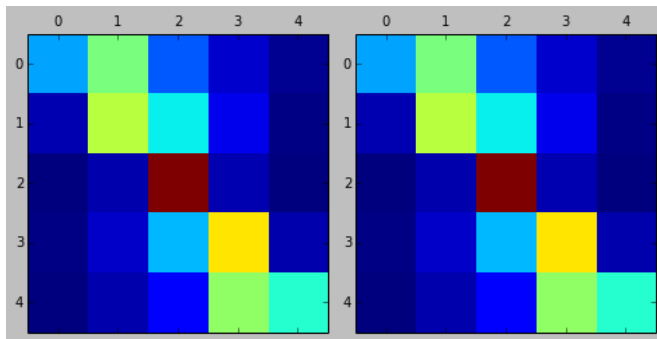


FIGURE: Matrice de confusion (*gauche*) Noeuds (*droite*) Root

Representation apprise - Mots

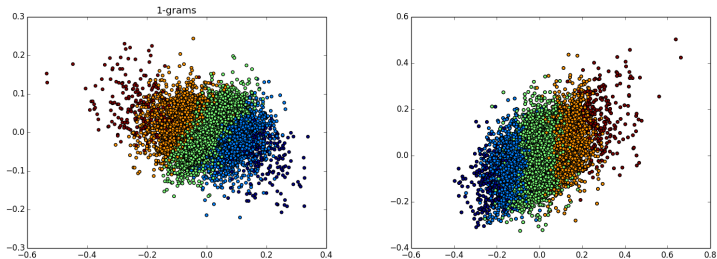


FIGURE: PCA de la representation des mots en dimension 2.
(gauche) PCA sur modèle 30D (droite) modèle 2D

Representation apprise - N-grams

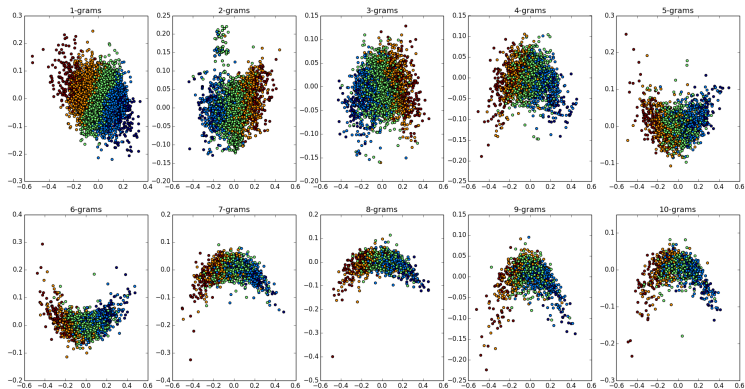
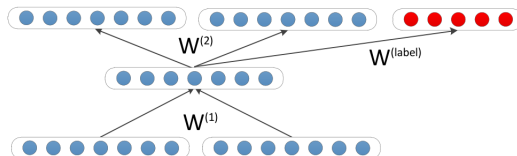


FIGURE: PCA de la representation des mots en dimension 2 pour le modèle 30d

Auto encoder ?

- L'idée est que l'on combine mal les représentations.
- Auto encoder optimize la reconstruction
⇒ Meilleure prise en compte de l'information ?
- L'idée du modèle est



Question ?