Représentations vectorielles de mots : application à la ponctuation à l'aide de réseaux de neurones récurrents

Thomas Pellegrini

Université de Toulouse ; IRIT

June 21, 2015

Outline

Word representations

d1: Toulouse est chef-lieu de la région Midi-Pyrénées

d2: Bordeaux est chef-lieu de la région Aquitaine

Word representations

$$\text{'Toulouse'} \rightarrow \begin{bmatrix} 1 \\ 0 \\ . \\ . \\ . \\ 0 \end{bmatrix} \rightarrow [0] \qquad \text{'Bordeaux'} \rightarrow \begin{bmatrix} 0 \\ 0 \\ . \\ . \\ 1 \\ 0 \end{bmatrix} \rightarrow [7]$$

Chaque mot $\in \{0, 1\}^9$

Word representations

d1: Toulouse est chef-lieu de la région Midi-Pyrénées

d2: Bordeaux est chef-lieu de la région Aquitaine

d1: [0, 1, 2, 3, 4, 5, 6]

d2: [7, 1, 2, 3, 4, 5, 8]

→ [0] et [7] très similaires

→ Comment mesurer leur similarité ?

Word representations

Comment mesurer leur similarité ? $\rightarrow$ Cosine similarity

$$\text{sim}(\mathbf{u}, \mathbf{v}) = cos(\theta) = \frac{\mathbf{u} \cdot \mathbf{v}}{\left|\mathbf{u}\right| \left|\mathbf{v}\right|}$$

Word representations

$$\text{sim('Toulouse', 'Bordeaux')} = \text{sim}(\begin{bmatrix} 1 \\ 0 \\ . \\ . \\ . \\ 0 \end{bmatrix} , \begin{bmatrix} 0 \\ 0 \\ . \\ . \\ 1 \\ 0 \end{bmatrix}) = 0$$

Un autre problème : sparsity (clairsemé, dispersé ?)...

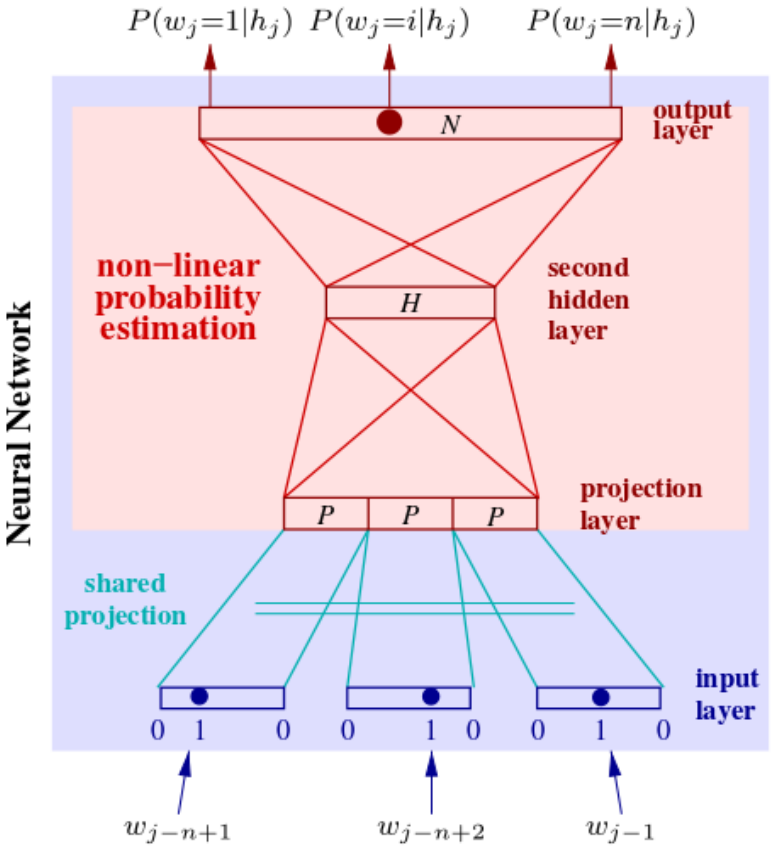
The curse of dimensionality

Word representations

Idea: to learn a distributed representation for words (Bengio, 2003)

1. associate with each word in the vocabulary a distributed *word feature vector* (a real-valued vector in $\mathbb{R}^m$),
2. express the joint *probability function* of word sequences in terms of the feature vectors of these words in the sequence, and
3. learn simultaneously the *word feature vectors* and the parameters of that *probability function*.

Word representations



from (Schwenk, 2007)

Word representations

$\text{sim}(\text{'Toulouse'}, \text{'Bordeaux'}) = 0,87$

	sim('France', .)
'Grande-Bretagne'	0,83
'RFA'	0,75
'Grèce'	0,75
'Suède'	0,74
'Suisse'	0,73

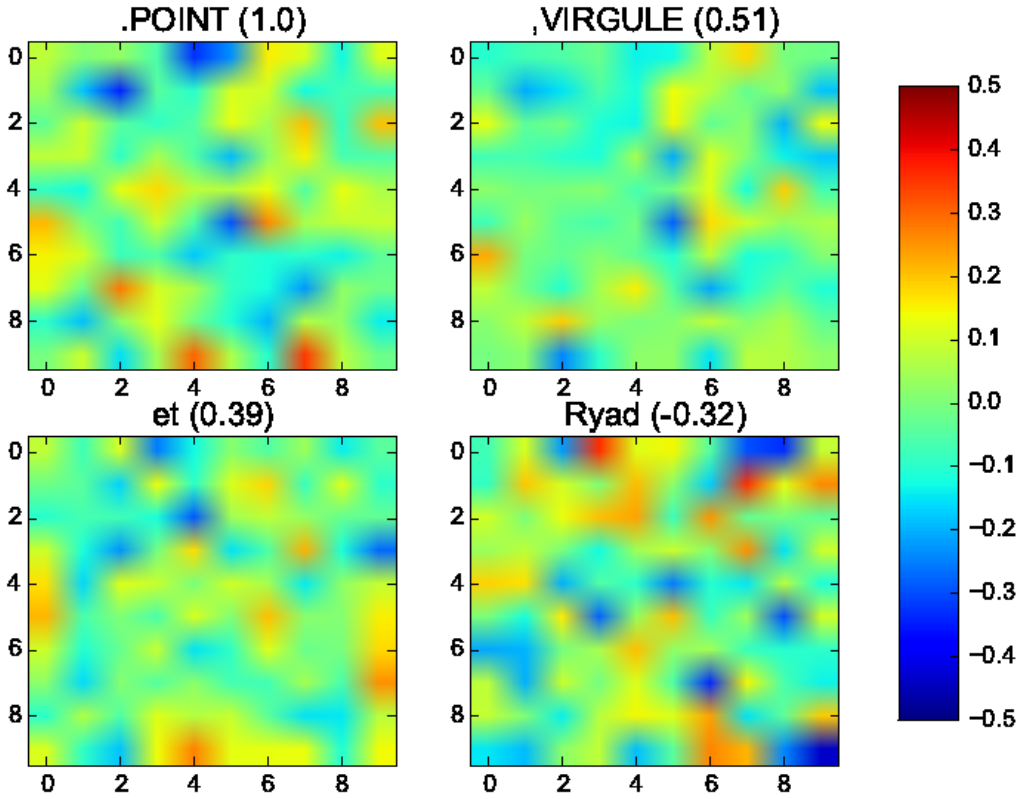
Word representations

Extrait du corpus Le Monde (janvier 1988)

```
Martin a une fille ,VIRGULE Peggy .POINT
et le fantôme de Murdoch ,VIRGULE arrivé en Amérique avec le château ,VIRGULE n' est évidemment pas insensible à ses charmes .POINT
il courtise Peggy ,VIRGULE qui le prend pour Donald ,VIRGULE amoureux timide de la belle .POINT
on voit le quiproquo ,VIRGULE mais ce n' est pas le seul ressort humoristique de cette comédie où ,VIRGULE par l' entremise de René Cl
Joe Martin ,VIRGULE qui connaît ,VIRGULE vite ,VIRGULE la présence du fantôme ,VIRGULE s' en sert pour la publicité de ses produits al
son rival  en affaires ,VIRGULE  Bigelow ,VIRGULE ne croit pas aux fantômes et sème d'autant plus le doute que ,VIRGULE au cours d' un
fantôme à vendre ,VIRGULE avec ses fausses pistes ,VIRGULE ses poursuites ,VIRGULE sa poésie burlesque ,VIRGULE rappelle les premiers
les histoires de fantômes appartiennent surtout à la tradition anglo-saxonne .POINT
mais celui de René Clair ne vient pas d' un conte de terreur .POINT
le cinéaste a cultivé l' opposition des sourires et des rires ,VIRGULE du merveilleux à la réalité prosaique .POINT
Murdoch bouscule toutes les conventions et amène un changement chez Donald ,VIRGULE Anglais très fin de race .POINT
Robert Donat  interprète d' Hitchcock ,VIRGULE la même année ,VIRGULE pour les Trente-Neuf Marches  tient les deux rôles avec esprit ,
et René Clair a fait des acteurs et des actrices les personnages de son propre univers .POINT
on oublie trop souvent à quel point ,VIRGULE au-delà des procédés techniques ,VIRGULE il se préoccupait du langage visuel .POINT
le critique Alexandre Arnoux ne manqua pas de le signaler et définit ,VIRGULE ainsi ,VIRGULE le film .POINT
 mélange aisé de légende anglaise ,VIRGULE de netteté française ,VIRGULE de farce américaine ,VIRGULE cet ouvrage ,VIRGULE où abondent
René Clair reviendra au merveilleux pendant son exil de guerre aux Etats-Unis avec Ma femme est une sorcière et C' est arrivé demain .
plus tard ,VIRGULE en France ,VIRGULE il y aura aussi les Belles de nuit .POINT
Denis Laget ,VIRGULE Albert Merz ,VIRGULE
```

$\text{sim}(\text{'.POINT'}, \text{',VIRGULE'}) = 0,51$

Word representations



Maximum entropy approach

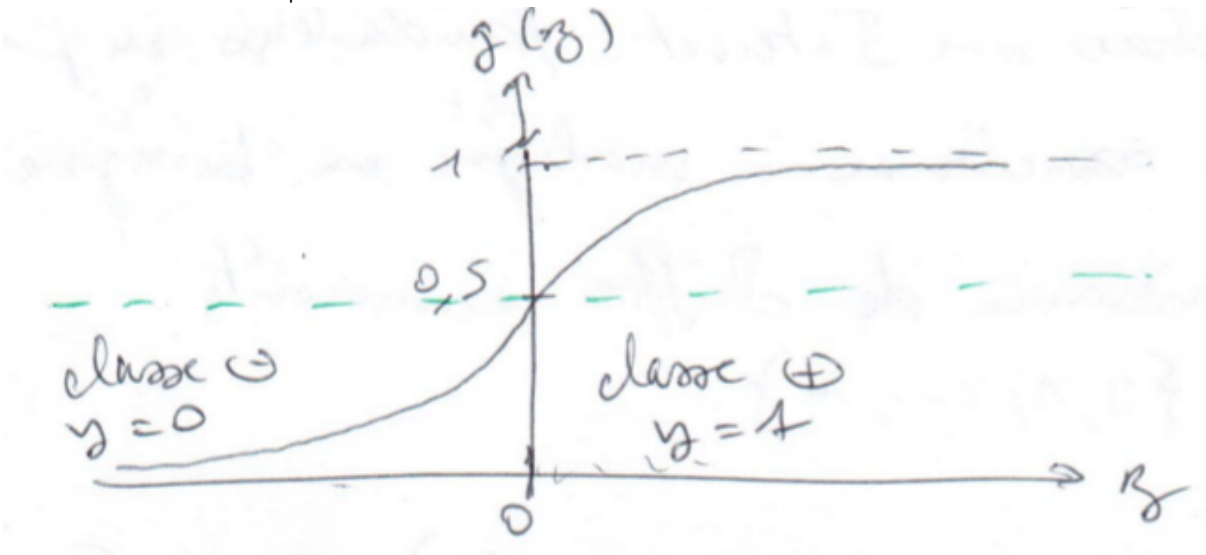
- Algorithme de classification
- Exemples d'utilisation :
  - emails : spam / no-spam ?
  - tumeur : cancéreuse / bégnine ?
  - cas multinomial : reconnaissance de chiffres manuscrits
- En entrée, des paramètres :
  - continus : âge, poids, valeurs de pixels
  - catégoriels : groupe sanguin, présence de tel ou tel mot
- En sortie :
  - cas binaire : échec/succès, positif/négatif $\leftarrow$ fonction sigmoide ou logisitque
  - cas multinomial (multi-classes) : $\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\} \leftarrow$ fonction softmax

Maximum entropy approach: binary case

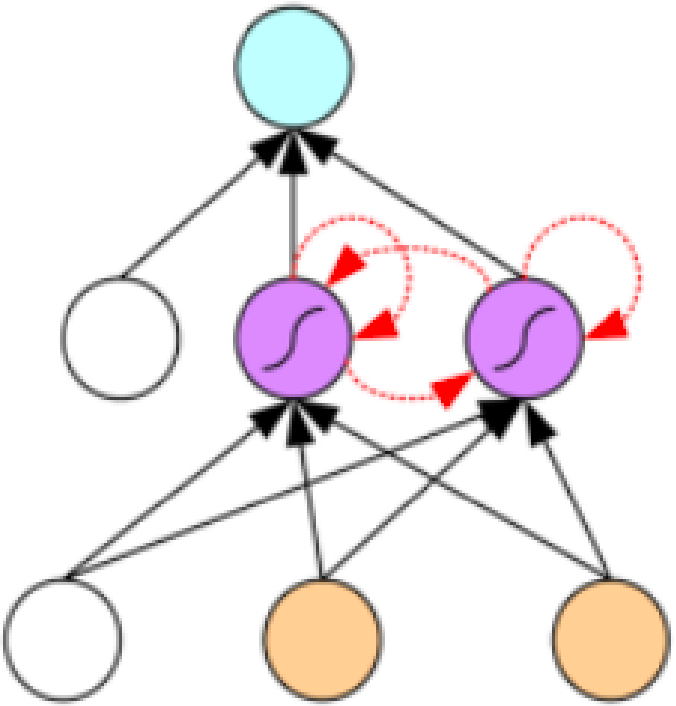
On considère $h_\theta(x) = g(\theta_0 + \theta_1 x_1 + \ldots + \theta_n x_n) = g(\theta^t x)$

On veut $h_\theta(x) = P(y = 1|x; \theta)$ telle que $\begin{cases} h_\theta(x) < 0,5 \Rightarrow y = 0 \\ h_\theta(x) \geq 0,5 \Rightarrow y = 1 \end{cases}$

On prend $g(z) = \dfrac{1}{1 + e^{-z}}$



Recurrent Neural Networks' approach



$$h(t) = f(Ux(t) + Vh(t-1))$$

Experimental evaluation

	Train	Test
# Tokens	17.3M	1.7M
# Types	200.0K	66.2K
# Full Stops	757.2K	75.4K
# Commas	1.4M	128.6K

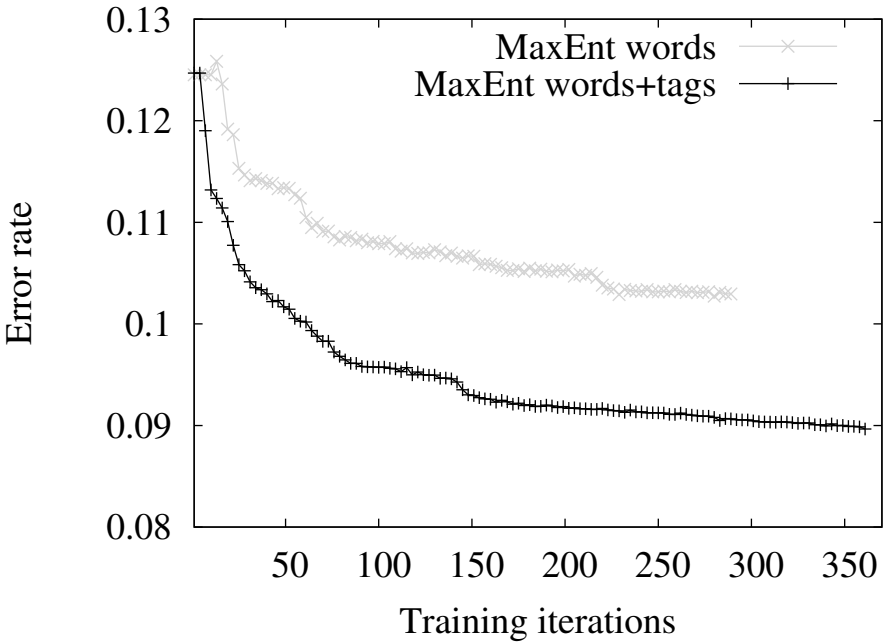
Table 1: *Le Monde* corpus description.

Experimental evaluation

- ME_W and RNN: word context window of size 7 words
- ME_W+POS: unigrams, bigrams, trigrams of words and POS tags

Experimental evaluation

Comparaison apprentissage des ME

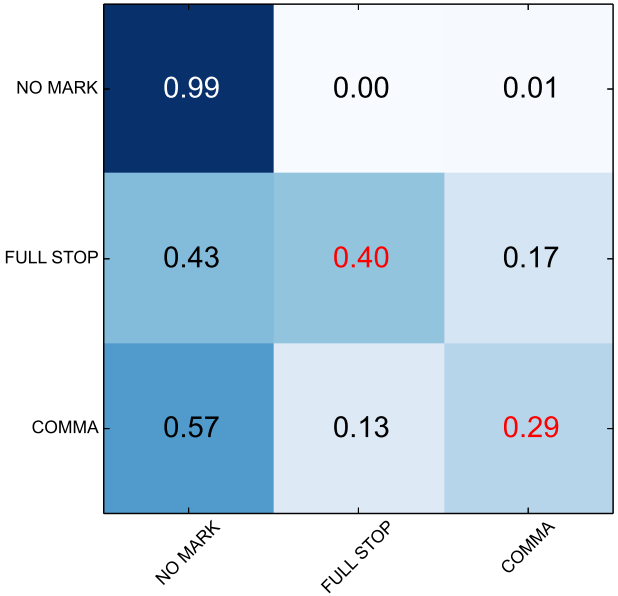


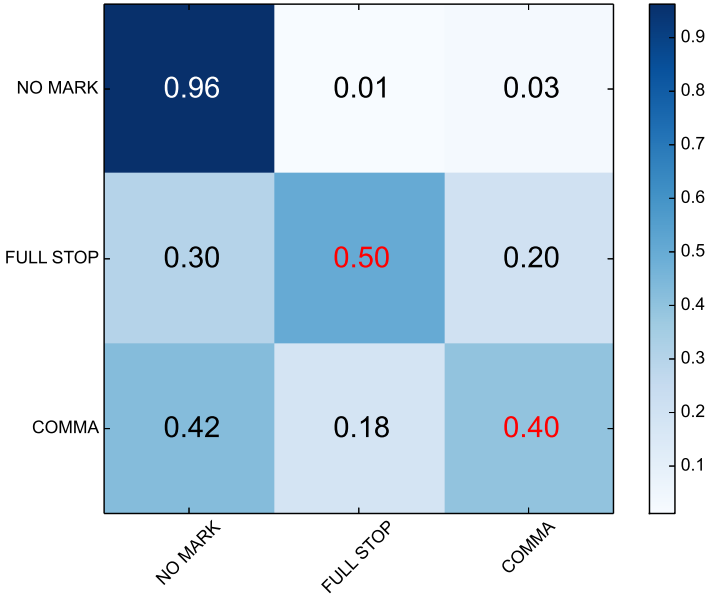
Experimental evaluation

Model	Train error (%)	Test error (%)	full stops			commas		
			Prec	Rec	F	Prec	Rec	F
ME_W	10.29	10.31	55.0	24.4	33.8	55.3	15.8	24.5
ME_W+POS	8.96	9.11	55.8	39.6	46.3	57.8	29.2	38.8
RNN	N/A	9.97	48.1	49.9	49.0	48.2	39.6	43.5

Table 2: Punctuation recovery results with the MaxEnt and the RNN approaches.

Experimental evaluation





Future work

- Valider ces résultats avec plus de données
- Étudier plus en détail les RNN : word-level vs minibatch learning, variantes de modèles
- Comment ajouter les POS tags et d'autres paramètres dans le modèle RNN ?