

---

# Réseau de neurones profond et SVM pour la classification de sentiments

**Abdelhalim Rafrafi — Vincent Guigue — Patrick Gallinari**

Laboratoire d'Informatique de Paris 6 (LIP6)  
Université Pierre et Marie Curie, Paris 6  
{abdelhalim.rafrafi, vincent.guigue, patrick.gallinari}@lip6.fr

---

*RÉSUMÉ. Le développement des forums, des blogs et de la vente en ligne pousse les utilisateurs à laisser de plus en plus d'informations en libre accès sur le web. Une partie de ces informations décrit des sentiments: elles permettent de développer des modèles d'analyse d'opinions et de faire des sondages dans divers domaines en récupérant simplement ces données textuelles. Nous proposons d'utiliser des réseaux de neurones pour apprendre des modèles de classification d'opinions efficaces. L'architecture retenue mêle des couches de neurones classiques et des couches de convolution, elle permet de projeter les mots dans un espace sémantique continu. Nous avons comparé notre travail avec un modèle SVM sur le corpus multi-thématiques Amazon. Nos expériences aboutissent à des performances identiques aux SVM et à de l'état de l'art. Cependant les modèles se comportent différemment et les erreurs ne concernent pas les mêmes documents: il est possible d'améliorer les performances en combinant les modèles.*

*ABSTRACT. As web 2.0 is spreading, users get used to give there opinions on forums, blogs and e-commerce websites. This is a valuable piece of information for many applications such as consumer modeling, sales prediction or opinion survey. According to the literature, the efficiency of opinion mining tools will mainly relies on the ability of discriminating texts that express positive sentiments from texts that express negative ones. Previous experiments show that this task is difficult.*

*Our approach relies on deep neural network (NN), mixing standard layers and convolutionnal layers. We demonstrate the efficiency of our approach on classical Amazon reviews data-set: raw performance equals Support Vector Machine (SVM) and the state of the art but our predictions differ significantly from SVM and combining SVM and NN enables us to overcome SVM.*

*MOTS-CLÉS: Classification de sentiments, réseaux de neurones profonds*

*KEYWORDS: Sentiment classification, deep neural network*

---

## 1. Introduction

Les tâches de recherche d'informations sont longtemps restées cantonnées à des approches thématiques. L'essor du Web 2.0 pousse les internautes à donner leur avis en ligne sur les forums, les blogs et les sites de commerce électronique notamment. Ces données présentent un fort potentiel, elles vont constituer la base de nouvelles applications en détection de sentiments. La modélisation des utilisateurs bénéficie de ces nouveaux modèles, les sites de e-commerce améliorent leurs outils de recommandation personnalisée et plusieurs entreprises sont d'ores et déjà capables de fournir des sondages d'opinions basés sur les sources ouvertes du web. Plusieurs bases de données sont disponibles sur différents sujets : (Ding *et al.*, 2008) proposent une base de revues d'utilisateurs, (Blitzer *et al.*, 2007) ont rassemblé les contributions des clients Amazon, (Pang *et al.*, 2002) ont publié plusieurs jeux de données sur les critiques de films des spectateurs, (Whitehead *et al.*, 2009) ont récupéré des données allant de la critique de colonies de vacances à la notation des avocats. Dans le domaine du jeu vidéo, (Marcoux *et al.*, 2009, Zhu *et al.*, 2006) ont utilisé avec succès la détection de sentiments dans les forums afin de prédire les ventes de jeux vidéos avant même leurs sorties en magasin.

Le domaine de la détection de sentiments est vaste et présente de gros enjeux économiques. Cela explique l'engouement actuel des chercheurs et le nombre important de publications sur le sujet. Un état de l'art relativement exhaustif a été dressé en 2008 par (Pang *et al.*, 2008). La détection de sentiments est également utile en recherche d'informations classique : elle permet de distinguer différentes sortes de contributions (revue d'expert, article de journaux, blog, forum...) et de construire des résumés automatiques triés par thèmes et opinions.

Dans leur étude sur la détection de sentiments, (Pang *et al.*, 2008) différencient plusieurs tâches allant de la construction d'une taxonomie des émotions à la quantification de ces émotions mais il apparaît que la tâche clé consiste à classer les textes positifs par rapport aux textes négatifs. Comme le montre l'ensemble des articles précédemment cités, apprendre un tel classifieur est difficile. Les approches thématiques se basent d'une manière ou d'une autre sur l'étude des distributions de mots : une telle approche n'est pas directement transposable à la détection de sentiments car les auteurs utilisent souvent les mêmes mots pour exprimer une opinion positive ou négative. Par exemple, la négation (bien/pas bien, pas vraiment bien/vraiment pas bien) ou la comparaison (A est mieux que B/B est mieux que A) ont des représentations très proches voire similaires lorsque nous utilisons des sacs de mots. De plus les mêmes mots peuvent avoir des connotations positives ou négatives en fonction du contexte (rabais en français ou *cheap* en anglais en sont des exemples).

Pour aborder ce problème, trois grandes stratégies sont généralement utilisées : (1) modéliser conjointement le thème et le sentiment en utilisant des méthodes à variables latentes comme Probabilistic Latent Semantic Analysis (PLSA) ou Latent Dirichlet Allocation (LDA) pour lever les ambiguïtés des expressions spécifiques à un thème donné (Mei *et al.*, 2007, Titov *et al.*, 2008, Lin *et al.*, 2009).

(2) Ajouter des connaissances extérieures comme des lexiques afin de traiter n'importe quelles données étiquetées ou non. Certains lexiques sont fournis par les linguistes (Esuli *et al.*, 2006, Wilson *et al.*, 2005), d'autres sont issus des techniques d'apprentissage comme (Blitzer *et al.*, 2007) qui proposent d'extraire des mots pivots à partir de corpus multi-domaines.

(3) Travailler au niveau des phrases permet d'extraire des motifs thème/sentiment caractéristiques locaux. C'est une manière implicite de lever les ambiguïtés liées à la négation. (Zhao *et al.*, 2008) utilisent des Conditional Random Fields (CRF), (Li *et al.*, 2010) utilisent des chaînes de Markov tandis que (Titov *et al.*, 2008) ont développé une approche LDA multi-échelles : ces références montrent l'intérêt des approches locales pour l'extraction de sentiments.

Etant donné le nombre croissant de bases de données étiquetées en sentiments, nous proposons d'utiliser une approche supervisée basée sur une architecture de réseaux de neurones proche de (Collobert *et al.*, 2008). La représentation jointe des thèmes et sentiments (1) est obtenue en projetant les mots dans un espace continu. L'apprentissage du réseau de neurones entraîne un rapprochement des mots qui apparaissent dans le même contexte : il s'agit donc d'une alternative aux méthodes à variables latentes. La problématique (2), consistant à enrichir les connaissances *a priori* pour améliorer les performances, prend la forme de pré-traitements ciblés dans notre approche : nous utilisons des *tags part-of-speech* et des filtres sur les mots rares afin de nous concentrer sur l'information discriminante lors de l'apprentissage. Enfin, l'architecture du réseau de neurones retenue traite les documents en concaténant les phrases (3). Une couche de convolution permet de traiter tous les mots du document séquentiellement quelle que soit sa longueur.

Les résultats obtenus sont intéressants à plusieurs titres : les performances du réseau de neurones sont au niveau des *Support Vector Machines* (SVM) et de l'état de l'art mais avec des prédictions significativement différentes. Seules environ 85% des décisions prises sont communes entre les SVM et le réseau de neurones : il est possible de dépasser l'état de l'art en combinant les sorties des deux modèles. L'autre intérêt de cette approche concerne l'interprétation des résultats et la construction d'un espace sémantique : une fois le réseau appris, il est possible de mesurer des distances entre les mots et de rechercher les voisins d'un mot en particulier.

Après avoir décrit l'état de l'art en section 2, nous décrivons le modèle et les pré-traitements que nous avons utilisés en section 3. La section 4 présente les résultats que nous avons obtenus.

## 2. Etat de l'art

### 2.1. Approches en Traitement de la Langue Naturelle (TALN)

Une grande partie de la littérature en classification de sentiments repose sur le TALN et les classifieurs à base de règles. De tels systèmes requièrent de nombreuses connaissances *a priori* et mêlent des lexiques orientés sentiments et des motifs gram-

A. Rafrafi, V. Guigue et P. Gallinari

maticaux spécifiques. B. Liu et son équipe ont exploré de nombreuses pistes dans ce domaine. Par exemple, (Mukherjee *et al.*, 2010) proposent d'étiqueter un ensemble de structures grammaticales en se basant sur un étiquetage *Part-Of-Speech* (POS). Ils sélectionnent ensuite les motifs et les mots les plus discriminants pour améliorer les performances. (Ding *et al.*, 2007) analysent les co-occurrences de mots à l'intérieur d'une phrase puis les co-occurrences entre les phrases. Ils combinent des règles issues des deux échelles pour obtenir un meilleur taux de bonne classification en classification de sentiments.

## **2.2. Approches supervisées & Support Vector Machines (SVM)**

Les approches supervisées sont également largement utilisées en classification de sentiments. Leur usage croît régulièrement avec le nombre de bases de données étiquetées disponibles. (Pang *et al.*, 2002, Pang *et al.*, 2008) ont comparé les approches classiques (SVM, *Naive Bayes*,...) et les représentations classiques (unigrams, N-grams, filtres *Part-Of-Speech*, tf-idf, codage présence/absence) sur plusieurs jeux de données (revues de films, Amazon).

Ils montrent l'existence d'un palier de performance autour de 80% de taux de bonne classification. Ce palier semble difficile à franchir, les exemples étant ensuite particulièrement complexes à analyser. Le codage unigrams+présence/absence semble le plus efficace : les N-grams introduisent un problème de dimensionnalité et les codages fréquentiels génèrent du bruit qui pénalise la généralisation du modèle. Afin de franchir le palier mentionné ci-dessus, (Blitzer *et al.*, 2007) combinent des heuristiques et utilisent des données non-étiquetées : ils atteignent 85% de taux de bonne classification sur certaines données Amazon. Une voie classique pour améliorer les performances consiste à enrichir la description en sacs de mots pour lever certaines ambiguïtés : (Kennedy *et al.*, 2006) prennent en compte les négations et les adverbes qui modifient la valence des sentiments, (Read, 2005) étudie l'impact des *smileys* sur la classification de sentiments.

## **2.3. Approches statistiques non supervisées**

Plusieurs articles récents utilisent des algorithmes dérivés de *Latent Dirichlet Allocation* (LDA) (Blei *et al.*, 2003) et *Probabilistic Latent Semantic Analysis* (PLSA) (Hofmann, 1999) pour extraire simultanément les thèmes et les sentiments présents dans les documents. (Mei *et al.*, 2007) proposent par exemple une version de PLSA où chaque mot est généré selon une mixture de modèles : une distribution de mots modélise l'arrière plan du document, chaque thème est représenté par une distribution dite neutre tandis que deux autres distributions modélisent respectivement les probabilités d'apparition des mots dans les documents positifs et négatifs. La plupart de ces approches nécessitent une initialisation spécifique afin d'extraire efficacement les opinions : (Mei *et al.*, 2007) proposent d'utiliser 10 mots positifs et 10 mots négatifs comme *a priori*. D'autres approches utilisent des données étiquetées ou des lexiques.

L'expression des sentiments est très liée aux thèmes abordés. L'usage d'un lexique général comme SentiWordNet (Esuli *et al.*, 2006) ne permet pas d'obtenir des performances brutes satisfaisantes : il est nécessaire de prendre en compte la spécificité du thème abordé en utilisant par exemple des documents non étiquetés du thème en question (Blitzer *et al.*, 2007). (Titov *et al.*, 2008, Lu *et al.*, 2008, Ghose *et al.*, 2007) n'attaquent pas non plus directement la tâche de classification de sentiments et modélisent conjointement les thèmes et sentiments présents dans les documents. (Turney, 2002) utilise l'algorithme non supervisé PMI (Pointwise Mutual Information) pour mesurer la distance entre les phrases de ses documents et les mots de référence *poor* et *excellent*.

### 3. Modèles et pré-traitements des données

Nous utiliserons les notations suivantes dans la suite. Soient  $d_i = \{m_j\}$  le  $i^{eme}$  document et  $y_i \in \mathcal{Y}$  l'étiquette qui lui est associée. Le document  $d_i$  est composé de mots indicés en  $j$ . Nous travaillerons exclusivement en apprentissage supervisé sur des problèmes bi-classes. Les étiquettes prennent pour valeurs  $\mathcal{Y} = \{+1, -1\}$ .

#### 3.1. Modèle de référence : SVM

Nous utilisons les SVM comme référence pour montrer l'intérêt de notre approche. La solution SVM est obtenue en résolvant le problème suivant :

$$W^* = \arg \min_W \left( \|W\|^2 + C \sum_i (1 - y_i \langle d_i, W \rangle)_+ \right)$$

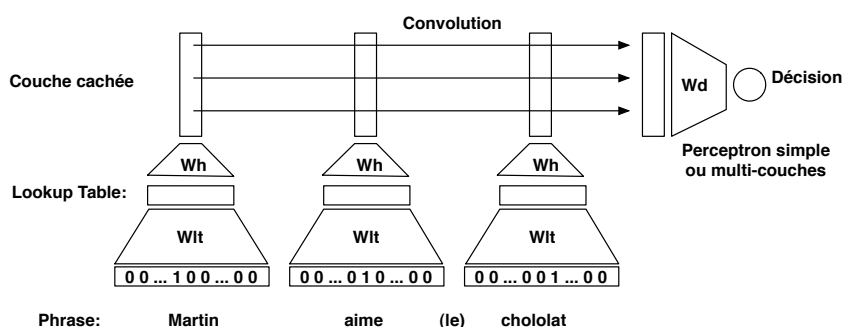
Dans cette équation,  $d_i$  désigne la représentation en sac de mots du  $i^{eme}$  document et  $W$  est un vecteur de poids de la taille du dictionnaire. La résolution est effectuée par descente de gradient stochastique.

#### 3.2. Architecture du réseau de neurones

Un document est vu comme une séquence de mots en entrée du réseau de neurones. Toutes les phrases sont prises les unes à la suite des autres.

##### 3.2.1. Couche 1 : Espace sémantique

Le réseau de neurones que nous avons utilisé est présenté en figure 1. Il se compose d'une première couche, notée  $W_{lt}$ , qui permet de projeter un mot dans dans un espace sémantique. L'ensemble des poids issus du mot  $j$  ( $W_{lt_j}$ ) forme la représentation du mot. La dimension de la représentation est un hyper-paramètre noté  $N_{lt}$ . Cette représentation est appelée *Lookup Table* dans (Collobert *et al.*, 2008). Il n'y a pas de fonction d'activation sur cette couche.



**Figure 1.** Réseau de neurones pour la détection de sentiment et la construction d'un espace de représentation sémantique des mots.

### 3.2.2. Couche 2 : Couche cachée et convolution

Chaque mot est ensuite projeté sur une couche cachée de taille  $N_h$ . Pour les neurones de la couche cachée, la fonction d'activation est la fonction *log soft max* :  $a_i = \log \frac{\exp(z_i)}{\sum_{j=1}^n \exp(z_j)}$

La phase de convolution consiste à sommer terme à terme toutes les couches cachées : à la fin de la convolution, la représentation du document est donc de taille  $N_h$ . La couche de convolution permet de passer d'un problème de taille variable (en fonction de la taille du document) à une représentation de taille fixe.

Dans l'article original, (Collobert *et al.*, 2008) proposent de construire chaque couche cachée en utilisant plusieurs mots avec une logique de fenêtre glissante : ils ont donc deux phases de convolution, une pour construire la couche cachée et une pour agréger les différentes couches cachées. Nous n'avons pas obtenu de résultats intéressants avec cette architecture et nous l'avons donc simplifiée.

### 3.2.3. Couche 3 : Décision

Le dernier problème consiste à exploiter la représentation construite pour prendre une décision. Il est possible d'utiliser un perceptron multi-couches ou un perceptron simple. Après expériences, nous avons opté pour un perceptron simple en utilisant une couche cachée assez grande (composée de plus de 100 neurones).

La couche de sortie compte deux neurones (un par classe) et est normalisée afin d'avoir directement une interprétation probabiliste de l'appartenance du document à chaque classe.

### 3.2.4. Optimisation du réseau de neurones

L'optimisation est effectuée par rétro-propagation du gradient sur l'ensemble du réseau en utilisant la librairie Torch (Collobert, 2008). Le pas de gradient est mis à jour après chaque passage sur l'ensemble des données selon la formule proposée par (Plagianakos *et al.*, 2001) :  $learningRate = learningRate(1 - \delta)$  où  $\delta$  désigne la différence entre les taux de bonne classification aux itérations  $t - 1$  et  $t$ . Le pas de mise à jour initial est réglé à  $10^{-2}$ .

### 3.2.5. Hyper-paramètres

Les hyper-paramètres à régler sont donc les suivants pour le modèle :  $N_{it}$  modélise la taille de l'espace de projection des mots, il sera pris entre 5 et 50,  $N_h$  modélise la taille de la couche cachée, il sera pris entre 100 et 500. En dessous de certains seuils, le réseau de neurones manque d'expressivité et n'apprend plus correctement, au delà d'une certaine limite les temps de calcul sont rédhibitoires et la variabilité des résultats augmente, certains *runs* étant totalement inexploitable.

Les valeurs précises que nous avons utilisées sont données dans la partie expériences.

## 4. Expériences

Nous présentons ici les résultats que nous avons obtenus avec les SVM et les réseaux de neurones présentés précédemment (section 3).

### 4.1. Bases de données

Toutes les bases de données sont bi-classes : elles contiennent des documents positifs et négatifs étiquetés.

Nous avons utilisé le corpus Amazon qui est divisé en 14 sous-corpus thématiques (Blitzer *et al.*, 2007) : *software, kitchen, apparel, books, magazines, health, games, sports, video games, camera, music, DVD, baby, electronics*. La plupart des bases contiennent 1000 documents de chaque classe. Dans *software*, il y a seulement 915 documents négatifs et 950 dans *magazines*.

### 4.2. Pré-traitements

Afin d'améliorer les performances de notre système, nous avons utilisé plusieurs pré-traitements :

– Nous utilisons un filtre *Part-Of-Speech* avec les options : JJ JJP JJR JJS RB RBR RBS NN NNS VB VBD VBG VBN VBP VBZ MD afin de sélectionner les noms, verbes, adjectifs et adverbes.

A. Rafrafi, V. Guigue et P. Gallinari

- Les mots rares (apparaissant une seule fois) ne sont pas éliminés : ils sont remplacés par leurs étiquettes grammaticales.

- les mots qui apparaissent au moins dans deux documents et dont leurs étiquettes grammaticales (*Part-Of-Speech*) n'appartiennent pas à la liste des étiquettes définie par le filtre vu au-dessus ne sont pas éliminés : ils sont remplacés par un mot générique.

### 4.3. Evaluation des performances et réglages

Les performances sont mesurées en taux de bonne classification. Toutes les expériences sont effectuées 10 fois et les résultats présentés sont des moyennes sur les 10 *runs*. Nous avons appris nos modèles sur 80% des données et testé les modèles sur les 20% restantes. Chaque *run* correspond à un tirage aléatoire différent (*bagging*).

Les hyper-paramètres sont les suivants :  $N_{it} = 10$ ,  $N_h = 100$  (cf section 3.2). Pour l'optimisation, nous avons opté pour un pas de gradient de  $10^{-2}$  avec une politique adaptative (cf section 3.2.4).

Les SVM sont utilisés avec un noyau linéaire régularisé. La valeur du paramètre de régularisation  $C$  est égale à -3.5.  $C$  a été fixé en étudiant différentes valeurs (grille) sur 3 sous-bases de données en validation croisée. Nous avons volontairement fixé un unique compromis pour l'ensemble des bases afin de ne pas biaiser les résultats en réglant *trop* finement les hyper-paramètres.

Nous avons comparé les SVM bruts et avec pré-traitements (SVM-SGD et SVM-SGD PP). Nous avons calculé les performances des réseaux de neurones (RNA) et nous les avons combinés avec les résultats des SVM selon deux stratégies : (MAX) règle du max entre les sorties probabilistes des SVM et des RNA, (CL) combinaison linéaire entre les sorties probabilistes avec un facteur de combinaison 0.5.

Nous avons enfin calculé les performances théoriques obtenues avec un oracle donnant le classifieur à utiliser sur chaque document (SVM+RNA Oracle).

### 4.4. Performances & analyses

Le tableau 1 présente les résultats que nous avons obtenus sur les jeux de données Amazon avec les algorithmes et les réglages détaillés ci-dessus. Plusieurs analyses s'imposent suite à ces expériences :

- Les pré-traitements apportent un gain compris entre 1 et 4% en taux de bonne classification. Les autres pré-traitements que nous avons essayés (lemmatisation/*stemming*, codage des négation) n'apportent pas de gain significatif.

- Les SVM et les RNA offrent globalement le même niveau de performance.

- Les résultats sont très différents entre SVM et RNA : les premiers apprennent des poids pour les mots du dictionnaire alors que les seconds apprennent des espaces sémantiques pour les mots. La dernière colonne du tableau permet de se rendre compte



de la différence entre les résultats des deux modèles : si nous étions capable de choisir toujours le modèle optimal pour un document, les performances seraient encore nettement meilleures.

– Nous avons donc décidé de fusionner les sorties des modèles et cela apporte un gain significatif en taux de bonne classification (entre 0.5 et 2.5% d'amélioration). Seule la base de données *appareil* se comporte différemment. La fusion n'apporte pas d'amélioration des performances sur la base de données *sport*.

– Les performances varient beaucoup d'un thème à l'autre et les modèles appris sont très différents : l'analyse des poids des SVM montrent que les mots participant le plus à la décision changent beaucoup d'une base à l'autre. Par exemple, sur les articles matériels, le mot *return* est toujours classé parmi les plus négatifs, dans le domaine des jeux vidéos, le mot *wait* est très négatif.

– Le temps d'apprentissage est nettement à l'avantage des SVM : sur le thème *video* le temps moyen d'apprentissage d'un RNA est supérieur à 45 minutes alors que le SVM met 6 secondes. En prédiction cette différence devient négligeable : les deux systèmes mettent de l'ordre d'une seconde pour traiter une base de 200 documents. NB : les deux implémentations (SVM et RNA) sont écrites en C et optimisées. Il s'agit respectivement de LIBSVM et Torch5.

Pour faire face à la tâche difficile qu'est la classification de sentiments, la fusion tardive entre différents modèles capables d'extraire des informations très variées à plusieurs niveaux dans les documents semble être une bonne solution.

#### 4.5. Analyse de l'espace sémantique des mots (*lookup table*)

Dans le modèle que nous avons créé, tous les mots ont une représentation continue dans un espace de dimension 10. Nous avons fait une projection en 2D afin de naviguer plus facilement dans les mots mais la taille du dictionnaire et la perte d'information liée à la réduction de dimension rend la projection 2D inutilisable. Nous avons finalement opté pour une analyse sur requête en calculant, pour quelques mots, les plus proches voisins dans la *lookup table*. Des extraits sont proposés dans le tableau 2. L'espace de projection propose réellement des rapprochements sémantiques en rapport avec les étiquettes des documents. C'est un point très intéressant par rapport aux techniques à variables latentes (PLSA, LDA) qui ne proposent que des probabilités d'apparition de mots par thème.

## 5. Conclusion

Cet article propose d'utiliser un réseau de neurones profond avec une couche de convolution pour détecter les sentiments présents dans les documents textuels. Ce réseau de neurones construit implicitement un espace sémantique continu où sont projetés les mots du dictionnaire.

Bases de données	Algorithmes					
	SVM-SGD	SVM-SGD PP	RNA	SVM + RNA MAX	SVM + RNA CL	SVM + RNA Oracle
Apparel	79.3	<b>85.76</b>	81.95	83.83	85.34	92.1
Baby	78.1	83.49	84.19	85.02	<b>85.76</b>	91.69
Book	78.3	81.34	80.51	81.86	<b>83.01</b>	89.95
Camera	83.5	88.39	87.96	89.27	<b>89.87</b>	94.65
DVD	77.8	82.582	80.556	82.46	<b>83.44</b>	89.68
Electronics	78.7	83.7	82.1	83.56	<b>84.56</b>	91.42
Games	81.2	85.32	81.1	85.14	<b>86.19</b>	91.92
Health	80.6	84.51	82.86	84.48	<b>85.97</b>	91.94
Kitchen	80.5	83.84	81.86	83.26	<b>85.39</b>	91.89
Magazines	84.3	88.59	86.15	87.61	<b>89.561</b>	94.75
Music	78.3	81.7	80.27	82.12	<b>83.07</b>	90
Software	83.9	86.16	86.22	87.78	<b>88.65</b>	92.84
Sports	81.1	<b>85.72</b>	82.8	84.37	85.7	92
Video G.	79.6	82.33	81.23	82.83	<b>83.14</b>	89.74

**Tableau 1.** Résultats obtenus sur les données Amazon. Les performances sont exprimées en taux de bonne classification moyen sur 10 runs en utilisant 80% des données en apprentissage et 20% en test. Les deux premières colonnes montrent les performances des SVM avec et sans pré-traitements, l’algorithme RNA est présenté seul puis combiné avec les SVM. La dernière colonne montre les performances théoriques obtenues en utilisant un oracle choisissant entre SVM et RNA pour chaque document.

Camera		Kitchen	
easy	difficult	bad	cook
flawlessly	kiddie	hear	meals
transport	cables	vendor	hate
documentation	upgrade	evolution	cleaner
immediate	worst	years	coffeemakers
exaggerated	robust	packaging	batter
soooo	screwdriver	amazed	accommodates
amateurs	lady	begin	cooking
controls	wrestle	salad	tasks

**Tableau 2.** Présentation ordonnée de 8 proches voisins (parmi les 24 plus proches) de certains mots dans la lookup table après apprentissage du réseau de neurones. Les bases de données sont données en première ligne, les mots requêtes sont sur la deuxième ligne du tableau.

Nous avons montré que les performances obtenues sont au niveau des meilleurs algorithmes actuels. De plus, les erreurs ne surviennent pas aux mêmes endroits qu'avec les SVM, ce qui offre la possibilité de combiner efficacement les des deux algorithmes pour obtenir des performances de premier plan.

Les perspectives autour de ce travail sont doubles. D'une part, il s'agit d'exploiter les documents non étiquetés et d'attaquer plusieurs tâches à la fois. Ce type de réseau de neurones a été conçu initialement pour traiter plusieurs problèmes à la fois : une première partie de l'apprentissage est alors non supervisée, modélisant par exemple une bonne phrase (issue des documents) par rapport à une mauvaise (générée aléatoirement). Chaque tâche est ensuite traitée en optimisant légèrement le réseau de neurones (*fine tuning*). Sur la détection de sentiments, cela revient à apprendre un modèle de langage avant de spécialiser notre réseau sur les sentiments. Cette approche permet de bénéficier des documents non étiquetés sans modifier profondément le modèle. Ce type d'approche est aussi prometteur pour les tâches de détection de sentiments multi-domaines.

L'autre perspective concerne les approches multi-échelles : le but serait de construire un réseau de neurones exploitant à la fois les informations locales au niveau de la phrase et les informations globales au niveau du document. Après avoir testé cette architecture au niveau du document, notre idée consiste à construire un réseau de neurones qui apprendra un espace sémantique au niveau de la phrase et qui l'utilisera pour la prise de décision au niveau du document. La complexification de l'architecture sera également un moyen d'intégrer de plus nombreuses informations *a priori* comme les étiquettes *Semantic Role Labeling* (SRL) et de retravailler sur la séparation entre les phrases pour la prise de décision.

## Remerciements

Les auteurs tiennent à remercier le pôle de compétitivité cap digital et la DGCIS (Direction Générale de la Compétitivité, de l'Industrie et des Services) pour le financement du projet DOXA qui a contribué à la réalisation des travaux présentés dans cet article.

## 6. Bibliographie

- Blei D. M., Ng A. Y., Jordan M. I., « Latent dirichlet allocation », *The Journal of Machine Learning Research*, vol. 3, p. 993-1022, 2003.
- Blitzer J., Dredze M., Pereira F., « Biographies, Bollywood, Boom-boxes and Blenders : Domain Adaptation for Sentiment Classification », *Proceedings of the Association for Computational Linguistics (ACL)*, 2007.
- Collobert R., « Torch », presented at the Workshop on Machine Learning Open Source Software, NIPS, 2008.

A. Rafrafi, V. Guigue et P. Gallinari

- Collobert R., Weston J., « A Unified Architecture for Natural Language Processing : Deep Neural Networks with Multitask Learning », *International Conference on Machine Learning, ICML*, 2008.
- Ding X., Liu B., « The utility of linguistic rules in opinion mining », *SIGIR '07 : Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval*, ACM, New York, NY, USA, p. 811-812, 2007.
- Ding X., Liu B., Yu P. S., « A holistic lexicon-based approach to opinion mining », *WSDM '08 : Proceedings of the international conference on Web search and web data mining*, ACM, New York, NY, USA, p. 231-240, 2008.
- Esuli A., Sebastiani F., « SENTIWORDNET : A Publicly Available Lexical Resource for Opinion Mining », *Proceedings of LREC-06, 5th Conference on Language Resources and Evaluation*, Genova, IT, p. 417-422, 2006.
- Ghose A., Ipeirotis P., Sundararajan A., « Opinion mining using econometrics : A case study on reputation systems », *Computational Linguistics*, vol. 45, p. 416, 2007.
- Hofmann T., « Probabilistic latent semantic indexing », *SIGIR*, p. 50-57, 1999.
- Kennedy A., Inkpen D., « Sentiment Classification of Movie Reviews Using Contextual Valence Shifters », *Computational Intelligence*, vol. 22, p. 2006, 2006.
- Li F., Huang M., Zhu X., « Sentiment Analysis with Global Topics and Local Dependency », 2010.
- Lin C., He Y., « Joint sentiment/topic model for sentiment analysis », *CIKM '09 : Proceeding of the 18th ACM conference on Information and knowledge management*, ACM, New York, NY, USA, p. 375-384, 2009.
- Lu Y., Zhai C., « Opinion integration through semi-supervised topic modeling », *17th International Conference on World Wide Web*, p. 121-130, 2008.
- Marcoux J., Selouani S.-A., « A Hybrid Subspace-Connectionist Data Mining Approach for Sales Forecasting in the Video Game Industry », *Computer Science and Information Engineering, World Congress on*, vol. 5, p. 666-670, 2009.
- Mei Q., Ling X., Wondra M., Su H., Zhai C., « Topic sentiment mixture : modeling facets and opinions in weblogs », *WWW '07 : Proceedings of the 16th international conference on World Wide Web*, ACM, New York, NY, USA, p. 171-180, 2007.
- Mukherjee A., Liu B., « Improving gender classification of blog authors », *EMNLP '10 : Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, Association for Computational Linguistics, Morristown, NJ, USA, p. 207-217, 2010.
- Pang B., Lee L., « Opinion mining and sentiment analysis », *Foundations and Trends in Information Retrieval*, vol. 2, p. 1-135, 2008.
- Pang B., Lee L., Vaithyanathan S., « Thumbs up ? : sentiment classification using machine learning techniques », *ACL Conference on Empirical Methods in NLP*, vol. 10, p. 79-86, 2002.
- Plagianakos V., Magoulas G., Vrahatis M., « Learning rate adaptation in stochastic gradient descent », *Advances in convex analysis and global optimization*, vol. 54, p. 433-444, 2001.
- Read J., « Using emoticons to reduce dependency in machine learning techniques for sentiment classification », *Proceedings of the ACL Student Research Workshop*, ACL '05, Association for Computational Linguistics, Morristown, NJ, USA, p. 43-48, 2005.

- Titov I., McDonald R., « Modeling online reviews with multi-grain topic models », *Proceeding of the 17th international conference on World Wide Web*, p. 111-120, 2008.
- Turney P. D., « Thumbs up or thumbs down ? : semantic orientation applied to unsupervised classification of reviews », *40th Annual Meeting on Association for Computational Linguistics*, p. 417-424, 2002.
- Whitehead M., Yaeger L., « Building a General Purpose Cross-Domain Sentiment Mining Model », *CSIE '09 : Proceedings of the 2009 WRI World Congress on Computer Science and Information Engineering*, IEEE Computer Society, Washington, DC, USA, p. 472-476, 2009.
- Wilson T., Wiebe J., Hoffmann P., « Recognizing contextual polarity in phrase-level sentiment analysis », *HLT '05 : Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*, Association for Computational Linguistics, Morristown, NJ, USA, p. 347-354, 2005.
- Zhao J., Liu K., Wang G., « Adding redundant features for CRFs-based sentence sentiment classification », *EMNLP '08 : Proceedings of the Conference on Empirical Methods in Natural Language Processing*, Association for Computational Linguistics, Morristown, NJ, USA, p. 117-126, 2008.
- Zhu F., Zhang X., « The influence of online consumer reviews on the demand for experience goods : The case of video games », *International Conference on Information Systems*, p. 367-382, 2006.

