Extraction des caractéristiques lexico-grammaticales et couplage des unités CRF (Conditional Random Field) au réseau de neurones profond pour l'extraction des aspects

Saint Germes Bienvenu Bengono Obiang, Norbert Tsopze

Analyse des sentiments

Extraction des caractéristiques lexico-grammaticales et couplage des unités CRF (Conditional Random Field) au réseau de neurones profond pour l'extraction des aspects

Saint Germes Bienvenu Bengono Obiang * — Tsopze Norbert**

* Département d'informatique
Université de Yaoundé 1
CAMEROUN
saintgermes1996@gmail.com

** Département d'informatique
Université de Yaoundé 1
Sorbonne Université, IRD, UMMISCO, F-93143, Bondy, France
CAMEROUN
tsopze.norbert@gmail.com

RÉSUMÉ. L'analyse des opinions consiste à extraire des connaissances à partir des commentaires laissés par les utilisateurs à propos d'un produit, service, texte,... L'analyse des opinions basée sur les aspects consiste alors à décomposer le commentaire afin d'extraire les aspects que cet utilisateur a évalué. Le modèle proposé par Jebbara et Cimiano, vainqueur de la compétition SemEval2016 n'extrait pas correctement des aspects composés et ne prend pas en compte les caractéristiques lexico-grammaticales des textes en entrée, ce qui limite aussi ses performances dans la détection des aspects. Nous proposons une amélioration du modèle de Jebbara et Cimiano. en y introduisant des unités CRF afin de prendre en compte les dépendances entre les étiquettes et ajoutant aux entrées du modèle des caractéristiques lexico-grammaticales. Les expérimentations faites sur les deux jeux de données de SemEval2016 ont permis de tester cette approche et montrer une amélioration de la mesure F-score d'environ 3.5%.

ABSTRACT. The Internet contains a wealth of information in the form of unstructured texts such as customer comments on products, events and more. By extracting and analyzing the opinions expressed in customer comments in detail, it is possible to obtain valuable opportunities and information for customers and companies. The model proposed by Jebbara and Cimiano. for the extraction of aspects, winner of the SemEval2016 competition, suffers from the absence of lexico-grammatic input characteristics and poor performance in the detection of compound aspects. We propose the model based on a recurrent neural network for the task of extracting aspects of an entity for sentiment analysis. The proposed model is an improvement of the Jebbara and Cimiano model. The modification consists in adding a CRF to take into account the dependencies between labels and we have extended the characteristics space by adding grammatical level characteristics and lexical level characteristics. Experiments on the two SemEval2016 data sets tested our approach and showed an improvement in the F-score measurement of about 3.5%.

MOTS-CLÉS : Analyse des sentiments, ABSA , Apprentissage profond , Unité récurrente à portes

KEYWORDS : Sentiments analysis , ABSA , Deep learning , Gated Recurrent Unit

1. Introduction

L'avènement du web 2.0 a offert aux utilisateurs la possibilité de donner leurs avis sur des sujets, articles,... à travers des commentaires postés en ligne. L'exploitation de ces commentaires dans le but d'aide à la décision est appelée fouille (analyse) d'opinions (ou de sentiments). Dans ce processus de fouille, un commentaire peut être traité globalement ou décomposé en aspects décrivant le sujet. L'analyse des sentiments basée sur les aspects [2] se fait généralement en deux étapes : l'extraction des aspects et la polarisation de l'aspect (positif ou négatif). Par exemple, à partir du texte "le repas était de bonne qualité" dans une revue de restaurant, l'aspect "repas" est extrait avec une polarité positive.

Ce travail se limite à l'extraction des aspects. Récemment, des modèles d'apprentissage profond supervisés ont connus des très bonnes performances notamment le modèle de Jebbara et Cimiano ([1]) vainqueur de la compétition SemEval 2016 basé sur des unités neuronales GRU bidirectionnelles (Bidirectional Gated recurrent unit). Bien que cette approche permet d'obtenir de meilleures performances que les travaux antérieurs, elle présente des limites : (1) performances relativement faibles pour le cas d'extraction des aspects composés (mots composés), (2) la non-exploitation des caractéristiques du niveau grammatical. La première est particulièrement importante car une solution permettra au modèle d'extraire par exemple d'un corpus du domaine de l'informatique où les caractéristiques sont souvent des mots composés, les aspects "Core i7", "Window 7". Une solution à la deuxième limite permet d'améliorer la représentation du texte en entrée.

Pour apporter une solution à la première limite, nous proposons un mécanisme d'analyse des dépendances entre les étiquettes de sortie des diffèrent mots, ceci est fait en ajoutant une couche CRF (Conditional Random Field) au modèle de Jebbara et Cimiano. Les unités CRF permettent de prendre en compte les relations en les classes associées aux différents mots. Pour la seconde limite, nous étendons l'espace des caractéristiques en y ajoutant des caractéristiques grammaticales prises sur le graphe de dépendances grammaticales fourni par stanford dependencies [3] et des caractéristiques lexicales. Nous appelons le modèle proposé Grammatical BiGRU-CRF (GRAM-BiGRU-CRF). L'avantage de ce modèle par rapport aux autres est sa capacité à étiquetter la séquence en se basant sur un BiGRU tout exploitant des caractéristiques grammaticales et en utilisant la spécificité des CRF pour la prise en compte des dépendances entre les étiquettes de sortie.

Dans la prochaine section, nous présenterons un état de l'art sur l'extraction des aspects. La section suivante aura pour objet la présentation de notre proposition. Les expérimentations ont été réalisées et les résultats feront l'objet de la section 4.

2. Etat de l'art

Hu et Liu présentent dans le livre [2] les différentes approches pour la tâche d'extraction des aspects. Ces approches sont notamment basées sur les noms et groupes de noms fréquents, sur l'apprentissage supervisé et celle basée sur la modélisation des sujets. La méthode basée sur les noms et groupes de noms fréquents trouve des aspects explicites qui sont des noms et des groupes de noms qui apparaissent dans un grand nombre de revues dans un domaine donné, cette méthode a été améliorée par Popescu et Etzioni [4] en supposant que la classe du produit est connue à l'avance. Leur algorithme détecte le nom ou groupe de noms caractérisant un produit en calculant l'information mutuelle ponctuelle entre le nom ou groupe de noms et la classe du produit.

Scaffidi et al. [5] ont présenté une méthode qui utilise un modèle linguistique pour identifier les caractéristiques du produit. Ils ont supposé que les caractéristiques des produits sont plus fréquentes dans les revues de produits que dans un texte en langage naturel général. Cependant, leur méthode semble avoir une faible précision puisque les aspects extraits sont affectés par le bruit. Certains méthodes ont traité l'extraction du terme d'aspect comme un étiquettage de séquences et ont utilisé un CRF pour cela. De telles méthodes ont donné de très bons résultats sur les ensembles de données, même dans les expériences inter domaines [6, 16].

La modélisation des sujets a été largement utilisée comme base pour effectuer l'extraction et le regroupement des aspects [7]. Le principe ici est le suivant : sachant que les opinions ont des cibles, elles sont évidemment liées. Leurs relations peuvent être exploitées pour extraire des aspects qui sont des cibles d'opinions parce que les mots de sentiment sont souvent connus. Deux modèles ont été considérés : pLSA (Probabilistic latent semantic analysis) et LDA (Latent Dirichlet allocation) [8]. Les deux modèles introduisent une variable latente " sujet " entre les variables observables " document " et " mot " pour analyser la distribution sémantique des sujets dans les documents. Dans les modèles thématiques, chaque document est représenté comme un mélange aléatoire des sujets latents, où chaque sujet est caractérisé par une distribution sur les mots. De telles méthodes ont gagné en popularité dans l'analyse des médias sociaux comme la détection de sujets politiques émergents sur Twitter[**?**]. Le modèle LDA définit un processus de génération probabiliste de Dirichlet pour la distribution sujet-document ; dans chaque document, un aspect latent est choisi selon une distribution multinomiale, contrôlée par un Dirichlet antérieur α. Puis, étant donné un aspect, un mot est extrait selon une autre distribution multinomiale, contrôlée par un autre Dirichlet antérieur β. Parmi les travaux existants utilisant ces modèles figurent l'extraction d'aspects globaux (marque du produit) et d'aspects locaux (propriété d'un produit), l'extraction de phrases clés [9] et la notation de multi-aspects.

L'apprentissage supervisé est aujourd'hui l'approche la plus utilisée et aussi celle présentant des meilleures performances. IHS-R&D [15] meilleur système lors du semEval 2014 pour la sous-tache d'extraction des aspects sur les données des restaurants attaque la sous-tache comme un problème d'étiquetage de séquences en utilisant un classificateur de champs aléatoires conditionnel (CRF). Les mots, les groupes de mots et le rôle sémantique sont les caractéristiques utilisées pour prédire les aspects et leurs catégories. Le modèle proposé obtient une amélioration de près de 10% par rapport au baseline (fourni lors de la compétition semEval). Les approches récentes utilisant les réseaux de neurones profonds ont montré une amélioration significative des performances, c'est notamment le cas du modèle de Al-Smadi et al [10] qui améliore le modèle IHS-R&D en ajoutant à la structure du modèle une BiLSTM (Bidirectional Long Short-Term Memory) [11] pour résoudre le problème d'extraction des aspects dans les revues d'hôtels arabes. Le modèle présenté lors du SemEval 2016 par Jebbara et Cimiano. [1], le BiGRU (Bidirectional gated recurrent unit) [11] présente des résultats encore meilleurs que le IHS-R&D sur le même jeu de données. Jebbara et Cimiano pour ce modèle d'extraction, utilisent comme caractéristiques le mot, l'étiquette morphosyntaxique du mot et la sémantique du mot fourni par SenticNet. Contrairement aux autres modèles, ils introduisent une nouvelle unité de réseaux de neurones récurrent : le GRU (Gatted recurrent unit) [17].

Li et al. [18] ont proposé HAST, un schéma basé sur deux réseaux de neurones récurrents. Un premier niveau pour capturer l'historique et un deuxième pour le résumé de l'opinion ; les unités neuronales dans les deux réseaux sont de type LSTM. Les sorties de ces deux réseaux sont concatenés pour constituer les caractéristiques à présenter en

entrée à un réseau de nerones feedfrward complètement connecté. Dans [19] Poria et al. proposent une approche combinant un réseau de neurones convolutionnels à sept couches (une couche d'entrée, deux couches de convolution, deux couches de pooling, et couches complètement connectées) et les caractéristiques lingustiques pour étiquetter les mots du corpus en "aspect" et "non aspect". Dans [20], Xu et al. proposent un modèle appelé DE-CNN (Dual Embeddings CNN) basé sur deux niveaux d'extraction de caractéristiques : un premier niveau pour l'extraction des caractéristiques du domaine et une deuxième pour les caractéristiques générales. Les caractéristiques obtenues sont alors soumises en entrée à un réseau de neurones convolutionnels de quatre couches, qui enverra aussi se sorties en entrée à un réseau de neurones feedforward complètement connecté.

3. Solution proposée

Elle consiste à modifier le modèle de Jebbara et Cimiano [1] en ajoutant à ses entrées les caractéristiques lexico-grammaticales et à ses sorties une couche formée des unités CRF.

3.1. Graphe de dépendances grammaticales

C'est un graphe orienté mettant en évidence les relations grammaticales entre les mots dans une phrase ainsi que l'étiquette morpho-syntaxiquement associée a chaque mot.

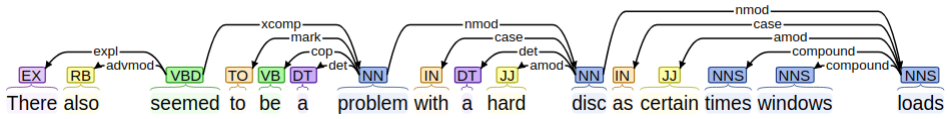

Figure 1. *Exemple de graphe de dépendances grammaticales généré pas la library StanforLP (http ://corenlp.run/)*

Le graphe de dependance grammaticale obtenu a partir de la phase :There alsoseemed to be a problem with the hard disc as certain times windows loads est présenté à la figure 1. Les aspects dans cette phrase sont : hard disc, windows, drivers. Les aspects dans cette phrase sont : hard disc, windows. Chacun des aspects (mots) possede une dependance avec un autre mot (nœud) dans l'arbre : amod(disk,hard) - ce qui signifie que "hard" est un modificateur adjectival de "disk".
Ce graphe possède une propriété importante : il existe au plus une liaison (relation avec un autre mot) incidente sur un mot. Le tableau 1 présente un ensemble de relation entre mots où *NN* est un nom, *VB* le verbe .

Modèle de relation de dépendance	Mot caractéristique	Mot d'opinion
NN - amod - JJ	NN	JJ
NN - nsubj - JJ	NN	JJ
NN - nsubj - VB - dobj - NN	Le premier NN	Le dernier NN
VB - advmod - RB	VB	RB

Tableau 1. *Exemple de relations impliquant les mots*

3.2. Les caractéristiques

Avant d'extraire les caractéristiques, les opérations de prétraitements effectuées sur le corpus sont la mise en minuscule de chaque mot et la tokenisation. Les caractéristiques extraites sont les suivantes :

Word Embeddings (WE) : qui est la représentation numérique sous forme de vecteur d'un mot. Le modèle utilisé ici est celui de Google contenant trois millions de mots pour le calcul des WE de dimension 300, contrairement au modèle de Jebbara et Cimiano qui utilise un WE de dimension 100. La séquence de WE pour une phrase constituée d'une séquence de N mots est le vecteur :

$$[W]_1^N = \{W_1, W_2, ..., W_N\} \text{ avec } W_i \in R^{300} \text{ et } R \text{ l'ensemble des nombres réels} \quad (1)$$

Étiquetage morphosyntaxique (POS) : Treebank Tagger [12] est l'outil d'étiquetage utilisé en raison de son utilisation dans le graphe de dépendance grammaticale. Il possède 36 étiquettes. Chaque étiquette est ensuite encodée en un vecteur de dimension 36.

$$[P]_1^N = \{P_1, P_2, ..., P_N\} \text{ avec } P_i \in R^{36} \quad (2)$$

Pour un mot d présenté en entrée, son étiquette (POS_d) dans le graphe et l'étiquette de sa cible (mot cible) POS_t lui sont associées. Le vecteur correspondant au mot d est le triplet (d, POS_d, POS_t). Le mot t est le mot tel qu'il existe au plus un arc (d, t) avec d comme noeud incident dans le graphe de dépendance grammaticale.

Sémantique (Sn) : c'est une ressource niveau concept basée sur un graphe fournissant des informations sémantiques et effectives. Pour chacun des 30.000 concepts faisant partie du graphe de connaissance, Sn fournit des scores pour 5 sensations : Le plaisir, l'attention, la sensibilité, l'aptitude, la polarité [13]. Pour l'ensemble des mots en entré nous obtenons donc la séquence :

$$[S]_1^N = \{S_1, S_2, ..., S_N\} \text{ avec } S_i \in R^5 \quad (3)$$

Sémantique des groupes nominaux (Sng) : l'une des remarques faites sur le modèle de Jebbara et Cimiano. En ce qui concerne les entrées est le fait que celui-ci ne prend pas en compte la sémantique des mots dans leurs contextes. Or la sémantique d'un mot s'il se trouve dans un groupe nominal dépend de celui-ci. La sémantique de groupes dans le modèle est obtenue par le biais d'une fonction ζ définie comme suit :

$$\zeta = \Gamma(f(G,X)) \quad (4)$$

Avec :

– G et X respectivement la grammaire et la séquence de mots en entrée, G est définie par l'expression régulière $G =< (NN|NNP|NNPS|NNS)+ >$ où NN est le noun, NNP le nom propre, $NNPS$ le nom propre au pluriel et NNS le nom pluriel

– $f(G,X)$ est la fonction qui permet d'extraire l'ensemble des groupes nominaux contenus dans la sequence d'entrée en se basant sur la séquence spécifiée.

$$f(G,X) = \{T_1, T_2, ..., T_M\} \text{ Avec } [T]_i \text{ un groupe nominal} \quad (5)$$

– $\Gamma(f)$ est la fonction qui étant donné un ensemble de groupes nominaux permet de déterminer la sémantique de chaque groupe.

$$\Gamma(f) = \{Sng_1, Sng_2, ..., Sng_N\} \text{ avec } Sng_i \in R^5 \quad (6)$$

Les éléments du vecteur représentent respectivement le plaisir, l'attention, la sensibilité, l'aptitude et la polarité.

Le rôle (Rl) : cette caractéristique est obtenue à partir d'un graphe de dépendance grammaticale [3]. Elle représente l'étiquette de la relation ayant pour nœud (mot) d'arrivé le mot courant. **Stanford typed dependencies** possède 56 types de dépendances, chaque type sera donc représenté par un vecteur de taille 56, chaque vecteur représentant une dépendance. Pour l'ensemble des mots en entrés nous obtenons donc la séquence :

$$[Rl]_1^N = \{Rl_1, Rl_2, ..., Rl_N\} \text{ avec } Rl_i \in R^{56} \tag{7}$$

Les éléments du vecteur représentent respectivement l'une des 56 dépendances, un seul composant du vecteur peut avoir la valeur 1 et les autre 0 permettant ainsi d'identifier une dépendance de façon unique.

L'appartenance à un groupe (InG) : Qui est un vecteur de dimension 2 permettant de représenter l'appartenance et la position de l'entrée dans un groupe de mots g_i. La grammaire permettant d'extraire les g_i est $GG =< (NN|NNP|NNPS|NNS)+ >$ où le premier élément du vecteur représente la position du mot dans le groupe, le second élément représente l'appartenance à un groupe.

3.3. Encodage des sorties

Le problème d'extraction est abordé comme un problème d'étiquetage de séquence. Pour cela, les termes d'aspects exprimés sont codés en utilisant le formalisme IOB2 [14]. Selon ce schéma, chaque mot de notre texte reçoit l'une des 3 balises, à savoir I, O ou B, qui indiquent si le mot est au début, à l'intérieur ou à l'extérieur. La représentation binaire de ces étiquettes est la suivante : $I = (1,0,0)$, $B = (0,1,0)$ et $O = (0,0,1)$.

3.4. Modèle proposé

Le réseau neuronal proposé lit une séquence de mots et prédit une séquence de balises IOB2 correspondante. La Figure 2 présente l'architecture du modèle que nous proposons. La première partie (couches d'entrée) représente l'ensemble des caractéristiques extraites du corpus. En gras sont les caractéristiques que nous avons ajoutées à celles proposées par Jebarra et Cimiano. Les couches suivantes (Bi-GRU et Dense) sont composées des unités neuronales. La dernière couche (CRF) est aussi une couche que nous avons ajoutée au modèle de Jebbara et Cimiano Elle a pour but la détection des dépendances entre les mots voisins. Un aspect est donc tout mot de la grammaire BI^*.

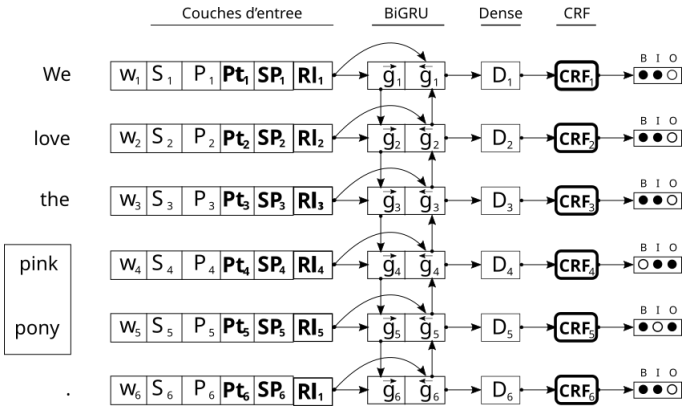

Figure 2. *Modèle GRAM-BiGRU-CRF proposé*

4. Expérimentations

Les données utilisées et les résultats obtenus en expérimentant notre proposition sont présentés dans cette section, ainsi qu'une comparaison avec le modèle de Jebbara et Cimiano.

4.1. Données

Les données de SemEval2016 ont été utilisées à cet effet. Elles consistent en des évaluations des clients sur les restaurants et les ordinateurs portables. Les étiquettes **"NULL"** sont conservées, de même que les étiquettes conflictuelles. Le modèle de Jebbara et Cimiano [1] a été ré implémenté afin de comparer à notre proposition. Le Tableau 2 résume la composition du jeu de données. Celui concernant les restaurants est constitué de 2000 revues et 2507 aspects pour l'entrainement et 676 pour les tests avec 859 aspects et le jeu de données concernant les ordinateurs portables contient 3048 revues avec 2373 aspects pour l'entrainement et 800 revues avec 654 aspects pour les tests.

	Restaurants		*Laptops*	
Entrainement	**Test**	**Entrainement**	**Test**	
Nombre de revues	2000	676	3048	800
Nombre d'aspects	2507	859	2373	654
revues contenants un aspect	1151	410	931	266
revues contenants deux aspects	381	128	355	105
revues contenants trois aspects	126	33	141	34
revues ne contenants pas aspects	292	89	1556	378
Nombre d'aspects NULL	627	209	0	0

Tableau 2. *Description des jeux de données utilisés pour les expérimentations*

La majeure partie des phrases du jeu de données contient plusieurs aspects. (avec des sentiments différents par aspects) par phrase. Contrairement au modèle de Wei et al.[21] qui duplique les phrases contenant plusieurs aspects afin que chaque phrase dans son jeu d'entrainement ne contienne que un aspect, Nous avons conservé les aspects multiples dans les phrases afin d'entrainer le modèle avec données ayant une structure similaire à celle du test.

4.2. Résultats

BiGRU-CRF	*Laptops*			*Restaurants*		
	Précision	**Rappel**	**F-mesure**	**Précision**	**Rappel**	**F-mesure**
$WE + POS_d$	70.58%	66.22%	67.99%	68.03%	68.00%	67.94%
$WE + POS_d + Sn$	69.10%	66.26%	67.59%	68.07%	67.64%	67.84%
$WE + POS_d + Rl$	68.67%	**70.25%**	69.44%	69.86%	67.67%	68.50%
$WE + POS_d + Rl + POS_t$	67.85%	69.93%	68.81%	69.83%	69.30%	69.56%
$WE + POS_d + Rl + POS_t + NNi$	70.04%	69.47%	69.71%	67.23%	69.34%	68.18%
$WE + POS_d + Rl + POS_t + NNi + Sn + GSn$	**70.70%**	69.17%	**69.83%**	**70.01%**	**69.69%**	**69.75%**

Tableau 3. *Evaluation du modèle GRAM-BiGRU-CRF sur les données d'entrainement*

Nous évaluons le modèle proposé en utilisant différentes combinaisons de caractéristiques en entrée. La première combinaison contient uniquement les caractéristiques

de base [1]. Nous ajoutons ensuite des caractéristiques lexicales et des caractéristiques du niveau grammaticale. La dernière ligne de Tableau 3 montre qu'en combinant les 7 caractéristiques, nous obtenons un équilibre entre le rappel et la précision.

BiGRU	*Laptops*			*Restaurants*		
	Précision	**Rappel**	**F-mesure**	**Précision**	**Rappel**	**F-mesure**
$WE + POS_d$	69.75%	68.55%	68.97%	69.24%	69.00%	68.89%
$WE + POS_d + Sn$	69.90%	68.24%	69.01%	67.95%	68.84%	67.94%
$WE + POS_d + Rl$	70.03%	67.33%	68.61%	67.73%	70.60%	69.13%
$WE + POS_d + Rl + POS_t$	**70.34%**	66.95%	68.60%	**70.74%**	68.12%	69.23%
$WE + POS_d + Rl + POS_t + NNi$	68.93%	**71.60%**	**70.17%**	66.46%	**73.81%**	**69.89%**
$WE + POS_d + Rl + POS_t + NNi + Sn + GSn$	69.55%	69.55%	69.79%	67.78%	71.77%	69.58%

Tableau 4. *Evaluation du modèle de Jebbara et Cimiano. sur les données d'entrainement*

Le Tableau 4 présente Les résultats obtenus en utilisant le modèle proposé par Jebbara et Cimiano en ajoutant en entrée les caractéristiques que nous avons proposées. Les résultats sont moins bons en utilisant les caractéristiques de base. Une meilleure précision est obtenue en ajoutant à l'espace de caractéristiques de base le Role (Rl) et l'étiquette morphosyntaxique de la cible (POS_t). En précisant l'appartenance à un groupe (NN_i), nous obtenons un meilleur rappel et une meilleure f-mesure.

Évaluation sur les données de test : en plus de l'évaluation sur le jeu de données d'entrainement, nous avons aussi testé les différents modèles sur le jeu de données de test en utilisant les combinaisons de caractéristiques ayant obtenues les meilleurs résultats lors de la phase d'entrainement. Le tableau 5 présente les résultats obtenus.

Laptops						
	GRAM − BiGRU − CRF			*BiGRU*		
	Précision	**Rappel**	**F-mesure**	**Précision**	**Rappel**	**F-mesure**
$WE + POS_d$	**68.67%**	**70.25%**	**69.44%**	67.85%	69.93%	68.81%
$WE + POS_d + Rl$	**74.85%**	**68.27%**	**71.41%**	70.76%	65.51%	68.03%
$WE + POS_d + Sn$	73.76%	**66.89%**	**70.16%**	**74.05%**	61.03%	66.91%
$WE + POS_d + Rl + POS_t$	**68.05%**	67.58%	67.82%	67.62%	**68.79%**	**68.20%**
$WE + POS_d + Rl + POS_t + NNi$	**73.28%**	64.31%	68.50%	71.32%	**68.62%**	**69.94%**
$WE + POS_d + Rl + POS_t + NNi + Sn + GSn$	67.07%	66.72%	66.89%	**70.58%**	**68.27%**	**69.41%**
Restaurants						
	GRAM − BiGRU − CRF			*BiGRU*		
	Précision	**Rappel**	**F-mesure**	**Précision**	**Rappel**	**F-mesure**
$W + POS_d$	**71.17%**	70.19%	**70.68%**	67.34%	**71.17%**	69.20%
$W + POS_d + Sn$	68.90%	**70.39%**	69.64%	**73.36%**	68.03%	**70.60%**
$W + POS_d + Rl$	69.33%	**69.60%**	69.47%	**73.68%**	68.62%	**71.06%**
$W + POS_d + Rl + POS_t$	**69.69%**	**70.78%**	**70.23%**	69.32%	69.99%	69.65%
$W + POS_d + Rl + POS_t + NNi$	68.03%	68.43%	68.23%	**72.26%**	**69.99%**	**71.11%**
$W + POS_d + Rl + POS_t + NNi + Sn + GSn$	**70.66%**	**72.74%**	**71.69%**	70.50%	70.78%	70.61%

Tableau 5. *Comparaison des modèles BiGRU et GRAM-BiGRU-CRF sur les données de test.*

Nous évaluons les deux modèles (BiGRU et GRAM-BiGRU-CRF) sur les données de test en utilisant deux combinaisons de caractéristiques. Les résultats présentés dans le Tableau 5 sur les données concernant les laptops montrent qu'en ajoutant aux caractéristiques de base le rôle (Rl) tiré du graphe de dépendances grammaticales la précision augmente significativement. En ce qui concerne les données des restaurants, la combinai-

1. Caractéristiques utilisées par Jebbara

son des 7 caractéristiques permet d'observer une augmentation considérable de la précision dans le cas où elle était très faible, mais aussi une augmentation du rappel et de la F-mesure dans le cas du modèle proposé.

Le tableau 6 compare les modèles GRAM-BiGRU-CRF et BiGRU sur la détection des aspects composés. Les ajouts apportés au modèle BiGRU permettent d'améliorer les résultats du modèle BiGRU sur ces données.

Restaurants		
	Aspects non composés	**Aspects composés**
BiGRU	77.3%	53.6%
$GRAM - BiGRU - CRF$	**78.6%**	**55.8%**

Tableau 6. *Comparaison des modèles BiGRU et GRAM-BiGRU-CRF sur la détection des apsects composés.*

Nous combinons ensuite les données des restaurants de SemEval 2014 ce qui nous permet de passer de 2000 évaluations à 3886 évaluations. Les expériences réalisées sur cet ensemble de données nous permettent d'obtenir les résultats du tableau 6 en donnant le pourcentage d'aspects extraits en fonction du nombre de mots qui composent l'aspect. Ces résultats montrent l'efficacité du modèle sur les aspects non composés et composés avec une amélioration de **2.9%**.

Le tableau 7 présente une comparaison des modèles GRAM-BiGRU-CRF, BIGRU, HAST [18] et DE-CNN [20] sur les données de SemEval 2014 où P, R et F désignent respectivement la précision, le rappel et la mésure F. Les résultats des approches HAST et DE-CNN ont été obtenus en utilisant leur code source disponible sur Internet. Le code de DE-CNN ne fournit pas en résultat le rappel et la précision. On constate la supériorité des modèles HAST et DN-CNN sur les modèles à base du BiGRU. Cela peut être dû à une bonne représentation des entrées car pour les modèles HAST et DE-CNN, les caractéristqiues sont extraites à l'aide des couches neuronales.

SemEval2014											
$GRAM - BiGRU - CRF$			*BiGRU*			*HAST*			$DE - CNN$		
P	R	F_1	P	R	F_1	P	R	F_1	P	R	F_1
75.30 %	73.40 %	74.26 %	72.22 %	70.33 %	70.77 %	77.33 %	77.80%	77.56%	–	–	81.59 %

Tableau 7. *Comparaison du GRAM - BiGRU - CRF et d'autres modèles de la littérature*

5. Conclusion

Il était question d'améliorer le modèle de Jebbara et Cimiano afin qu'il détecte les aspects composés. Pour le faire, ce papier propose l'ajout des caractéristiques lexico-grammaticales en entrée et une couche CRF en sortie du modèle. Les expérimentations faites sur les mêmes données de la compétition SEMVAL2016 ont montré que ces ajouts améliorent sensiblement les résultats du modèle proposé par Jebbara et Cimiano.

La suite de ce travail consistera à étudier l'influence des caractéristiques lexico-grammaticales et des unités CRF dans ce modèles. Nous envisagerons aussi d'étudier la possibilité de tester ce modèle avec d'autres unités neuronales que les GRU et le LSTM.

6. Bibliographie

[1] SOUFIAN JEBBARA AND PHILIPP CIMIANO, « Aspect-Based Sentiment Analysis Using a Two-Step Neural Network Architecture », *In : Semantic Web Challenges. Third SemWebEval Challenge at ESWC 2016. Revised Selected Papers. Sack H, Dietze S, Tordai A, Lange C (Eds) ; Communications in Computer and Information Science, 641. Cham : Springer*, 153-170 2017.

[2] LIU, BING, « Sentiment Analysis and Opinion Mining », *Synthesis Lectures on Human Language Technologies*, vol. 5/ 1-167, 2012.

[3] DE MARNEFFE, MARIE-CATHERINE AND MANNING, CHRISTOPHER D., « Stanford typed dependencies manual », 2008.

[4] A.-M. POPESCU, O. ETZIONI, « Extracting product features and opinions from reviews », *Proc. of EMNLP-2005*, 2005, pp. 3–28.

[5] C. SCAFFIDI, K. BIERHOFF, E. CHANG, M. FELKER, H. NG, C. JIN, RED OPAL, « Product-feature scoring from reviews », *Proc. of the 8th ACM Conference on Electronic Commerce* , vol. ACM, 2007, pp. 182–191.

[6] T. ZHIQIANG, W. WENTING, « DLIREC : Aspect term extraction and term polarity classification system », *Proc. of the 8th Int. Workshop on Semantic Evaluation (SemEval 2014)*,pp. 235–240, 2014.

[7] Y. HU, J. BOYD-GRABER, B. SATINOFF, A. SMITH, « Interactive topic modeling », *Mach.Learn.*, vol. . 95 (3) (2014) 423–469.

[8] T. HOFMANN,« , Probabilistic latent semantic indexing », *Proc. of 22nd ACM SIGIR Conference on Research and Development in Information Retrieval*,ACM, 1999, pp. 50–57.

[9] S. BRANAVAN, H. CHEN, J. EISENSTEIN, R. BARZILAY, « , Learning document-level semantic properties from free-text annotations, *J. Artif. Intell Research* », vol. Res. 34 (2) (2009) 569-603.

[10] MOHAMMAD, A.-S., AL-AYYOUB, M., AL-SARHAN, H., AND JARARWEH, Y, « , Using aspect-based sentiment analysis to evaluate arabic news affect on readers », *Int. Journal of Machine Learning and Cybernetics.*, 2015.

[11] GRAVES, A. AND SCHMIDHUBER, J, « , Framewise phoneme classification with bidirectional lstm and other neural network architectures. », *Neural Networks*, vol. 18(5-6) :602–610. 2005.

[12] SANTORINI, BEATRICE, « Part-Of-Speech Tagging Guidelines for the Penn Treebank Project 2nd printing », *Department of Linguistics, University of Pennsylvania*, 1995.

[13] CAMBRIA, ERIK AND OLSHER, DANIEL AND RAJAGOPAL, DHEERAJ, « SenticNet 3 : A Common and Common-Sense Knowledge Base for Cognition-Driven Sentiment Analysis », *Proc. of AAAI* , 2014.

[14] TJONG KIM SANG, E.F., VEENSTRA, J., « , Representing text chunks, » *Proc. of European Chapter of the ACL (EACL).*, pp. 173–179. Bergen, Norway (1999).

[15] CHERNYSHEVICH M., « cross-domain extraction of product features using conditional random fields », *Proc. of the 8th int. workshop semantic evaluation (SemEval)*, pp 309–313, 2014.

[16] JAKOB, NIKLAS AND GUREVYCH, IRYNA, « Extracting Opinion Targets in a Single- and Cross-domain Setting with Conditional Random Fields », *Association for Computational Linguistics*, pp 1035–1045, 2010.

[17] "CHO, KYUNGHYUN AND VAN MERRIENBOER, BART AND BAHDANAU, DZMITRY AND BENGIO, YOSHUA, « On the Properties of Neural Machine Translation : Encoder–Decoder Approaches », *Association for Computational Linguistics*, pp 103–111, 2014.

[18] XIN LI AND LIDONG BING AND PIJI LI AND WAI LAM, « Aspect term extraction with history attention and selective transformation », *IJCAI'18 : Proceedings of the 27th International*

Joint Conference on Artificial Intelligence pp 4194–4200, 2018.

[19] SOUJANYA PORIA AND ERIK CAMBRIA AND ALEXANDER GELBUKH, « , Aspect extraction for opinion mining with a deep convolutional neural network », *Knowledge-Based Systems*, vol. 108 pp. 42–49, 2016.

[20] XU, HU AND LIU, BING AND SHU, LEI AND YU, PHILIP S.« , Double Embeddings and CNN-based Sequence Labeling for Aspect Extraction », *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2 : Short Papers)*, vol. pp. 592–598, 2018.

[21] XUE, WEI AND LI, TAO, « , Aspect Based Sentiment Analysis with Gated Convolutional Networks », *Association for Computational Linguistics*, vol. pp. 2514–2523. Melbourne, Australia (2018).