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
norbert.tsopze@facsciences-uy1.cm

RÉSUMÉ. L'analyse des opinions consiste à extraire des connaissances à partir des commentaires laissés par les utilisateurs à propos d'un produit, service, actualité,... L'analyse des opinions basée sur les aspects consiste à décomposer le commentaire afin d'en extraire les aspects évalués par l'utilisateur. Ces aspects contribuent à la prise de décision en fournissant plus de détails sur les avis des utilisateurs. Le modèle proposé par Jebbara et Cimiano, vainqueur de la compétition *SemEval2016* ne prend pas en compte les caractéristiques lexico-grammaticales des textes en entrée, ce qui limite les performances tant dans la détection des aspects simples que celle des aspects composés. Nous proposons une amélioration du modèle de Jebbara et Cimiano en y introduisant des unités CRF (Conditional Random Field) afin de prendre en compte les dépendances entre les étiquettes et en 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 de montrer une amélioration de la F-mesure d'environ 3.5%.

ABSTRACT. Sentiment analysis consists of extracting knowledge off customers' comments about a service, a product, news...etc. When an analysed comment is decomposed to extract targeted aspects of the assessed object, we speak of aspect based sentiment analysis. These aspects are considered to offer a better decision making by providing more details on the user's judgement. Winner of the *Semeval2016* competition, Jebbara and Cimiano model does not take into account lexico-grammatical characteristics of the input texts. Thus, limiting its performances in detecting both simple and composed aspects. We propose a model meant to improve Jebbara and Cimiano's by introducing CRF (Conditional Random Field) units, with the purpose of including dependencies between labels and adding to the model entries, lexico-grammatical characteristics. Experiments were conducted on *SemEval2016* datasets to test this approach and have shown a raise of about 3.5% on the F-score measure.

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. Beaucoup de travaux se sont intéressés à l'analyse des sentiments. Ces travaux distinguent trois niveaux : documents, phrases et aspects. 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). Ce travail se limite à l'extraction des aspects. 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.

Plusieurs approches ont été proposées pour l'extraction des aspects. L'approche basée sur l'extraction des mots fréquents consiste à extraire les mots et groupe de mots fréquents [4, 5, 6, 16], dont les limites sont la détermination du seuil de fréquence et la non extraction des mots peu fréquents. La modélisation de sujets est une approche qui a pour idée générale de derterminer la cible d'un mot d'opinion dans une séquence [7, 8, 22, 9]. Ces approches ne tiennent pas compte des informations sur la sémantique des éléments de la phrase.

Récemment, des modèles d'apprentissage profond supervisé ont connus des très bonnes performances [10, 18, 20, 19] notamment le modèle de Jebbara et Cimiano ([1]) vainqueur de la compétition *SemEval2016* basé sur des unités neuronales GRU [17] bidirectionnelles ou BiGRU (*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. En effet, 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"*, *"Windows 7"*. Une solution à la deuxième limite contribue à améliorer la représentation du texte en entrée du modèle. Comme présenté dans le Tableau 1 cette proportion est non négligeable dans les données de la compétition.

Pour apporter une solution à la première limite, nous proposons un mécanisme d'analyse des dépendances entre les étiquettes de sortie des diffèrents mots (Section 3.3), 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 entre 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 Section 2, nous présenterons un état de l'art sur l'extraction des aspects. La Section 3 aura pour objet la présentation de notre proposition. Les expérimentations ont été réalisées et les résultats feront l'objet de la Section 4.

SemEval2016		*SemEval2014*		*Alc-14-short-data*
Laptops	***Restaurants***	***Laptops***	***Restaurants***	
36.62 %	25.35 %	36.64 %	24.52 %	69.8 %

Tableau 1. *Proportions des mots composés dans quelques jeux de données*

2. Etat de l'art

Hu et Liu [2] présentent 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 [2] 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. Certaines méthodes ont traité l'extraction des aspects comme un étiquettage de séquences et ont utilisé un CRF pour cela. De telles méthodes ont donné de bons résultats (soit une F1 de 75.21% sur les données Laptop du SemEval2014) sur les ensembles de données, même dans les expériences inter-domaines [6, 16]. Mais ces modèles ne sont pas adéquats dans le cas où le domaine etudié possède une multitude d'aspects dont certains apparaissent trés rarement dans le corpus. Ainsi, certains auteurs [7, 8, 22, 9] ont plutôt adopté des méthodes basées sur la modélisation des sujets.

La modélisation des sujets a été largement utilisée comme base pour effectuer l'extraction et le regroupement des aspects [7]. Le principe est le suivant : sachant que les opinions ont des cibles dans la relation grammaticale entre un nom et un adjectif, ces 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 pour cette analyse : 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 [22]. 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 figure l'extraction d'aspects globaux (marque du produit) et d'aspects locaux (propriété d'un produit), l'ex-

traction de phrases clés [9] et la notation de multi-aspects. Cependant les méthodes basées sur la modélisation des sujets ne permettaient pas de prendre en compte des informations telles que : la syntaxe, le sens des mots et même le dictionaire du domaine. Pour pallier cette limite, des méthodes basées sur l'apprentissage supervisé sont proposées [15, 10, 1].

L'apprentissage supervisé est aujourd'hui l'approche la plus utilisée et aussi celle présentant des meilleures performances. L'équipe IHS-R&D [15] présente le 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 conditionnels (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 modèle de référence (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 fournie 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 (History attention and selective transformation), un modèle 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 concatenées pour constituer les caractéristiques à présenter en entrée à un réseau de nerones feedforward 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 un 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. Le Tableau 2 présente un récapitulatif des méthodes de la littérature. Les auteurs de ces méthodes utilisent différents ensembles de données SemEval2014 et SemEval2016. Les évaluations se focalisent principalement sur les métriques F1 et rappel.

Le Tableau 2 récapitule des méthodes de l'état de l'art. La colonne "Limites" présente les limites principales de chaque méthode.

Auteurs, Année	Réf	Modèles	Spécificités	Limites
Popescu, 2005	[4]	OPINE	Utilisation d'un modèle non supervisé	Perfomances limitées dans la detection des aspects peu fréquents
Chernyshevich, 2014	[15]	IHS-R&D	Introduction des attributs : FA, FH et FPA	Contexte limité lié au traitement de la séquence dans un seul sens
Xin Li et al. 2018	[18]	HAST	Utilisation d'un modèle basé sur l'attention	Utilisation de peu de caractéristiques pour representer un mot
Soujanya et al. 2016	[19]	CNN+LP	Utilisation d'un modèle basé sur un CNN	Utilisation de peu de caractéristiques pour representer un mot
Jebbara et al. 2016	[1]	BiGRU	Introduction d'un modèle récurent bidirectionel	Mauvaise performance dans l'extraction des aspects composés

Tableau 2. *Récapitulatif des modèles de l'état de l'art.*

3. Solution proposée

Nous proposons le modèle GRAM-BiGRU-CRF qui 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

Le graphe de dépendances grammaticales 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.

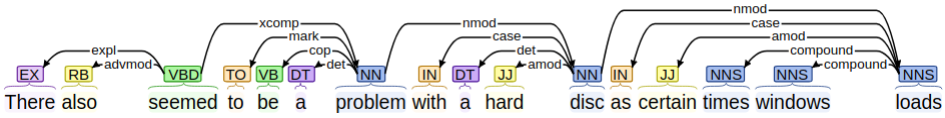

Figure 1. *Exemple de graphe de dépendances grammaticales généré pas la library StanforLP (http ://corenlp.run/)*

Le graphe de dépendances grammaticales est un graphe orienté $G = (S, A)$ tel que :

– S **est l'ensemble des nœuds :** un nœud du graphe de dépendance grammaticale est représenté par un mot de la sequence d'entrée et a pour attribut l'étiquette morpho-syntaxique du mots. Pour une sequence en entrée composée de n mots, le graphe obtenu a $n+1$ nœuds, Le nœud supplémentaire est représenté par le mot clé *ROOT*.

$$Card(S) = Card(t_1, t_2, t_3, ...t_n) + 1 = n + 1 \text{ avec } t_1, t_2, t_3, ...t_n \text{ les mots de sequence}$$

– A **est l'ensemble des relations** : une relation dans le graphe est représentée par trois composantes (le nœud de départ d, Le nœud d'arrivée t et l'étiquette de la relation).

Du fait de son orientation, ce graphe peut être exploité pour la tache d'extraction des aspects. Ce graphe possède une propriété importante : il existe au plus une liaison incidente sur un noeud.

Le Tableau 3 présente la signification des étiquettes sur les liaisons de la Figure 1 . Les étiquettes morphosyntaxique utilisées dans cette Figure 1 proviennent de *Penn Treebank tagset*[1].

Dépendances	Significations	Descriptions
expl	*expletive*	Cette relation capture un "there" existentiel
advmod	*adverb modifie*	Adverbe (non clausal) qui sert à modifier le sens du mot.
xcomp	*open clausal complement*	Complément prédicatif ou clausal sans sujet propre.
mark	*marker*	Mot qui introduit une clause finie subordonnée à une autre clause.
cop	*copula*	Relation entre le complément d'un verbe copulaire et le verbe copulaire.
det	*determiner*	Relation entre la tête d'un NP et son déterminant.
amod	*adjectival modifier*	Toute phrase adjectivale qui sert à modifier le sens du NP.
nmod	*nominal modifier*	Nom fonctionnant comme un argument ou un adjuvant non essentiel (oblique).
compound	*noun compound modifier*	Tout nom qui sert à modifier le nom de tête.
case	*case marking*	La relation de cas est utilisée pour toute préposition en anglais.

Tableau 3. *Relations grammaticales Stanford décrite dans "The Stanford typed dependencies manual" [26]*

3.2. Caractéristiques extraites

Avant d'extraire les caractéristiques, les opérations de prétraitements effectuées sur le corpus sont la mise en minuscule de chaque mot et la tokenisation. Le but de la mise en minuscule de chaque mot et la tokenisation est d'utiliser la même casse pour écrire tous les mots et de représenter les commentaires en unités linguistiques. Les caractéristiques extraites sont les suivantes (le Tableau 4 présente un résumé de celle-ci) :

Plongement de mots (Word Embeddings noté WE) [2] : qui est la représentation numérique sous forme de vecteur d'un mot qui a été utilisé avec succès dans de nombreuses tâches de TAL [23, 24, 25, 10, 1]. Le modèle utilisé ici est celui de *Google* contenant trois millions de mots pour le calcul des WE de dimension 300 (la taille 300 a été choisi expérimentalement après avoir testé glove-100d[3]), 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

1. https ://www.sketchengine.eu/tagsets/penn-treebank-tagset/
2. gensim : https ://radimrehurek.com/gensim/index.html
3. https ://nlp.stanford.edu/projects/glove/

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 entrée nous obtenons donc la séquence :

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

Sémantique des groupes nominaux (Sng)[4] **:** l'une des remarques faite sur le modèle de Jebbara et Cimiano en ce qui concerne les entrées est le fait que ce-dernier 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 :

$$\zeta = \Gamma(f(G,X)) \tag{4}$$

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 nom, *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.

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

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

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

Les éléments du vecteur Sng_i représentent respectivement le plaisir, l'attention, la sensibilité, l'aptitude et la polarité [1].

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'arrivée 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 entrée 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 autres la valeur 0 permettant ainsi d'identifier une dépendance de façon unique.

L'appartenance à un groupe (InG) [5] **:** C'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.

4. SenticNet : https ://pypi.org/project/senticnet
5. spaCy : https ://spacy.io/

Caractéristiques	Composantes	Descriptions
WE	300	Descripteur vectoriel de type "Word embeddings"
POS_d	36	Étiquetage morphosyntaxique do mot
POS_t	36	Étiquetage morphosyntaxique du mot cible de la relation
Sn	5	Informations sémantiques sur les mots
Sng	5	Information sémantique d'un groupe nominal
Rl	56	Relation gramaticale entre le mot et sa cible dans le graphe gramatical
InG	2	Représentation de l'appartenance du mot a un groupe nominal et du début du groupe nominal

Tableau 4. *Résumé des caractéristiques*

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 $O = (0,0,1)$. La notation IBO (pour Inside, Begin, Outside) respectivement permet d'étiquetter les mots comme étant le début d'un aspect (pour B), comme étant à l'intérieur de l'aspect (pour I) et pas n'appartenant pas à l'aspect (pour O). Un aspect composé commence par le mot de début (étiquetté B) et s'enchaine par les autres mots (étiquettés I) ; tandis qu'un aspect simple est tout simplement étiquetté I ou B.

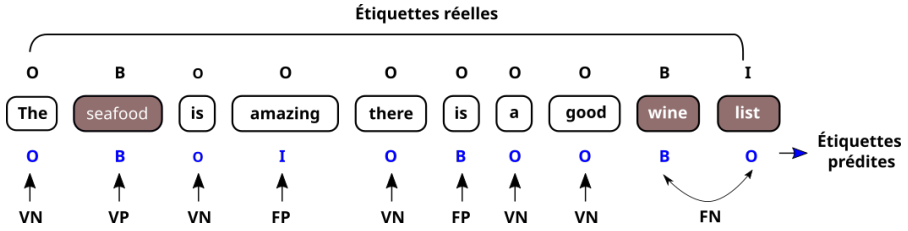

Figure 2. *Etiquettage d'une phrase au format IOB pour l'extraction des aspects et définition des différentes composantes pour le calcul des mesures.*

3.4. Modèle proposé

Le réseau neuronal proposé lit une séquence de mots et prédit une séquence de balises IOB2 correspondante. La Figure 3 présente l'architecture du modèle que nous proposons. La première partie (couches d'entrée) représente l'ensemble des caractéristiques extraites du corpus. Les caractéristiques que nous avons ajoutées à celles proposées par Jebarra et Cimiano sont en gras. 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^*.

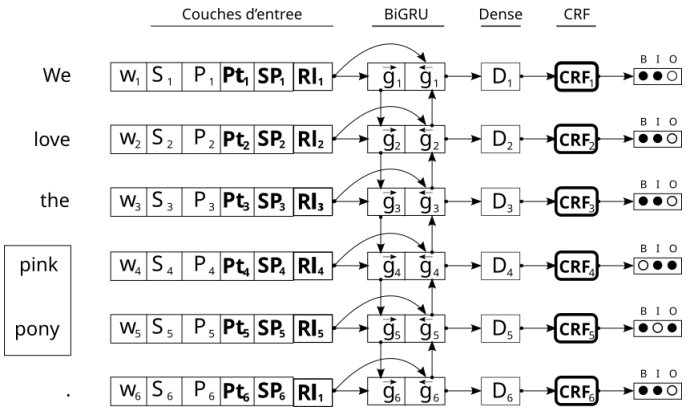

Figure 3. *Modèle GRAM-BiGRU-CRF proposé*

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. Les mesures évaluées au cours de ces expérimentations sont la précision, le rappel et la F-mesure (F1).

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 réimplémenté afin de comparer à notre proposition. Le Tableau 5 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 5. *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 aspect) 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 qu'un aspect. Nous avons conservé les aspects multiples dans les phrases afin d'entrainer le modèle avec des données ayant une structure similaire à celle du test. Les expériences ont été réalisées en appliquant une validation croisée d'ordre 5.

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 6. *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 [6]. Nous ajoutons ensuite des caractéristiques lexicales et des caractéristiques du niveau grammatical. La dernière ligne de Tableau 6 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 7. *Evaluation du modèle BiGRU sur les données d'entrainement*

Le Tableau 7 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 différentes combinaisons de caractéristiques. Le Tableau 8 présente les résultats obtenus.

6. Caractéristiques utilisées par Jebbara

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 8. *Evaluation 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 plusieurs combinaisons de caractéristiques. Les résultats présentés dans le Tableau 8 sur les données concernant les laptops montrent qu'en ajoutant aux caractéristiques de base celles tirées du graphe de dépendances grammaticales la précision augmente significativement (4.61%). En ce qui concerne les données des restaurants, la combinaison des 7 caractéristiques permet d'observer une augmentation considérable de la précision (4.92%), mais aussi une augmentation de la F-mesure (1.91%).

Le Tableau 9 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 9. *Comparaison des modèles BiGRU et GRAM-BiGRU-CRF sur la détection des apsects composés.*

Nous combinons ensuite les données des restaurants de *SemEval2014* 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 9 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 **2.2%** pour les aspects composés.

Les caractéristiques presentées dans ce document ont aussi été utilisées pour la tache de polarisation d'un aspect specifique. Comme Jebbara and Cimiano [1], nous avons ajouté une couche entrainée pour la polarité des aspects. Les resultats obtenus sont consignés dans le Tableau 10. La caractéristique *Dist* représente la distance relative du mot par rapport à l'aspect sélectionné [1]. Nous pouvons déduire du Tableau 10 que l'ajout des caractéristiques issues du graphe de dépendances gramaticales ($Rl - POS_t$) améliore

également la polarisation des aspects extraits de ce jeu de données **0.47%** ; toutefois les meilleurs résultats sont obtenus en ajoutant la sémantique du groupe nominal (*GSn*) soit **0.51%**.

Caractéristiques	**Taux de succès**
$WE + POS_d + Dist + Sn$	87.21%
$WE + POS_d + Dist + Sn + Rl$	87.68%
$WE + POS_d + Dist + Sn + GSn$	**87.74%**
$WE + POS_d + Dist + Sn + Rl + POS_t$	87.28%

Tableau 10. *Taux succès obtenu sur la polarité des aspects extraits*

Le Tableau 11 présente une comparaison des modèles GRAM-BiGRU-CRF, BIGRU, HAST [18] et DE-CNN [20] sur les données de SemEval 2014 pour la tâche d'extraction des aspects où P, R et F désignent respectivement la précision, le rappel et la F-mesure. Les résultats des approches HAST et DE-CNN ont été obtenus en utilisant leur code source disponible sur Internet[7][8]. 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 11. *Comparaison du GRAM - BiGRU - CRF et d'autres modèles de la littérature*

5. Conclusion

Cet article propose une amélioration du modèle de Jebbara et Cimiano en améliorant la représentation du texte en entrée et la détection les aspects composés. Les caractéristiques lexico-grammaticales ont été extraites des textes en entrée et combinées aux caractéristiques du modèle de base. Une couche CRF a également a été ajoutée en sortie du modèle pour Les prendre en compte les aspects composés. Les expérimentations faites sur les mêmes données de la compétition *SemEval2016* ont montré que ces ajouts améliorent sensiblement les résultats du modèle proposé par Jebbara et Cimiano.

En perspectives, nous allons étudier l'influence des caractéristiques lexico-grammaticales et des unités CRF dans ce modèle. Nous envisagerons aussi d'étudier la possibilité de tester ce modèle avec d'autres unités neuronales que les GRU et le LSTM.

7. https ://github.com/lixin4ever/HAST
8. https ://github.com/howardhsu/DE-CNN

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).

[22] S. RILL, D. REINEL, J. SCHEIDT, R. ZICARI, POLITWI, « , Early detection of emerging political topics on twitter and the impact on concept-level sentiment analysis,Knowl.-Based Syst. »,vol. 69 (2014) 14–23.

[23] COLLOBERT, R., WESTON, J., BOTTOU, L., KARLEN, M., KAVUKCUOGLU, K., KUKSA, P« , Double Embeddings and CNN-based Sequence Labeling for Aspect Extraction », *Natural language processing (almost) from scratch. Journal of Machine Learning (Research 12)*, vol. pp. 2493–2537, 2011.

[24] DOS SANTOS, C., ZADROZNY, B.« , Learning character-level representations for part-of-speech tagging », *Proceedings of the 31st International Conference on Machine Learning.*, vol. pp. 1818–1826, 2014.

[25] LE, Q., MIKOLOV, T.« , Distributed Representations of Sentences and Documents. », *ICML 32*, vol. pp. 1188–1196, 2014.

[26] DE MARNEFFE, MARIE-CATHERINE MANNING, CHRISTOPHER D« , The Stanford typed dependencies representation. », , vol. pp. 1-8, 208.