

---

# Détection d'opinion argumentée à partir de Twitter

Asma<sub>1</sub> Ouertatani<sub>1</sub>\* — Ghada<sub>2</sub> Gasmi<sub>2</sub>\*\* — Chiraz<sub>3</sub> Latiri<sub>2</sub>\*\*\*

\* LIPAH, ENSI, University of Manouba, Tunis, Tunisia

\*\* LISI, INSAT, University of Carthage, Tunis, Tunisia

\*\*\* LIPAH, FST, University of Tunis El Manar, Tunis, Tunisia

---

*RÉSUMÉ. Savoir ce que pensent les gens est fondamental pour la prise de décision. Avec la croissance explosive des réseaux sociaux ces informations sont disponibles à profusion, sous forme d'avis, d'opinions ou d'un jugement formé sur quelque chose ou quelqu'un pour défendre un point de vue.*

*Plusieurs chercheurs ne se sont concentrés que sur l'identification et la définition de l'opinion. Nous proposons dans cet article de définir et de caractériser l'opinion argumentée selon les composantes d'arguments associées. Nous avons mené des expériences en utilisant différents modèles de classification.*

*ABSTRACT. With the explosive growth of Web 2.0 (e.g., blogs, microblogs, Twitter, and postings in social network sites) on the web, individuals and organizations are increasingly using the content to find and monitor opinionated data and distill the information contained in them: opinion supported by arguments or judgment formed about something or someone.*

*Several researchers have focused on the opinion analysis and opinion detection. We attempt in this paper to define and characterize an opinion with the associated prevailing arguments components and we called it argued opinion. we conducted experiments using different classifying models.*

*MOTS-CLÉS : Fouille d'opinion, Fouille d'argument, CRFs.*

*KEYWORDS: opinion detection, Argument identification, CRFs.*

---

## 1. Introduction et motivation

Avec l'avènement du web 2.0 et la croissance explosive des réseaux sociaux (*e.g.* forum, blogs, micro blogs, Twitter) chaque personne a le droit de créer sa propre tribune sur internet. Les organisations utilisent de plus en plus le contenu des réseaux sociaux pour surveiller leurs réputations et leurs marques en extrayant et en analysant les avis positifs ou négatifs de leurs clients concernant un produit ou un service (Pang *et al.*, 2008).

“*Ce que pensent les autres*” est une information importante pour la plupart d'entre nous pendant le processus de prise de décision (Pang *et al.*, 2008). De nos jours, si vous voulez acheter un produit ou bien participer à un événement, vous ne vous limitez plus à demander les avis de vos amis et de vos proches car au lieu de cela, plusieurs avis et opinions sont disponibles dans les forums publics sous forme de commentaires. Pour une organisation, il n'est plus nécessaire de mener des enquêtes et des groupes de discussion pour recueillir l'opinion publique puisqu'il y a une abondance de ces informations accessibles au public (Liu, 2012).

Cependant, la recherche et le suivi de données d'opinions et l'analyse de l'information contenue demeurent une tâche importante en raison de la prolifération de diverses plateformes de réseaux sociaux. Chaque plateforme contient généralement un énorme volume de données d'opinions qui n'est pas toujours facile à déchiffrer surtout dans le cas des blogs et des messages courts. Le lecteur humain aura du mal à identifier les sites pertinents et à extraire et résumer les opinions qu'ils contiennent. Des systèmes de détection d'opinions automatisés sont donc nécessaires (Liu, 2012).

La détection d'opinion a été l'un des domaines de recherche les plus actifs dans le traitement du langage naturel depuis le début de l'année 2000 (Liu, 2010b ; Liu, 2012). C'est un processus qui détermine l'attitude ou la polarité des opinions pour évaluer un produit ou un service, pour comprendre les opinions du public sur des événements sociaux ou sur des mouvements politiques. C'est un domaine qui vise à définir les outils automatiques qui sont capables d'extraire les informations subjectives à partir de textes en langue naturelle afin de créer des connaissances structurées, exploitables et utilisables par un système d'aide à la décision (Pozzi *et al.*, 2016). L'extraction d'opinions peut être appliquée à n'importe quelle forme textuelle d'opinions sur les blogs, les forums, les micros blogs, tels que les tweets (Benamara *et al.*, 2017).

(Pang *et al.*, 2008) ont réalisé une étude approfondie en couvrant des applications, des défis communs comme l'extraction d'opinions, la classification des sentiments, l'analyse de la polarité, et le résumé d'opinion. Ils ont mis en évidence certaines approches basées sur le lexique et d'autres basées sur l'apprentissage supervisé en utilisant la mesure de similarité et plusieurs classifieurs comme le naïve bayes. Ensuite, (Tang *et al.*, 2009) ont discuté des problèmes liés à l'analyse de la subjectivité et l'identification de l'opinion.

Selon le dictionnaire en ligne Larousse <sup>1</sup>, une opinion peut être vue comme : (a) un jugement formé sur quelque chose, un avis, un sentiment qu'un individu ou un groupe émet sur un sujet, des faits, ce qu'il en pense, (b) une déclaration formelle des raisons d'un jugement donné.

De toute évidence, il existe un lien naturel entre l'opinion et l'argument. Dans de nombreux cas l'opinion conduit à mettre en avant des justifications. Lorsqu'une opinion est étendue à une ou plusieurs justifications, on ne parle plus d'une simple opinion, on parle d'une opinion argumentée. Une opinion motivée, appuyée, renforcée par une autre information ou un fait implique la capacité d'une personne à être argumentative. L'argument est considéré comme ayant une énorme influence sur l'opinion d'une personne sur un événement ou un produit spécifique, parce qu'il est basé sur la réflexion de l'argumentateur, sur son expérience et fait appel à la capacité critique du décideur. Un auteur affirme que (Ruffiero, 1988) "*personne ne peut espérer être un bon penseur sans acquérir une compréhension mûre de la nature de l'opinion*". En effet, étudier et comprendre les opinions en prenant en considération les arguments utilisés permet de les classer du moins convaincant au plus convaincant afin de leurs donner de plus en plus de poids, et mettre en évidence leurs aspects qui ont le plus positivement ou négativement impressionné le lecteur.

Les fondements conceptuels de l'étude s'appuient sur la théorie de l'argumentation, qui est une discipline ancienne. Le processus de l'argumentation est capable de fournir un ensemble d'outils fiables et précis pour étudier le raisonnement de l'être humain. Il est couramment utilisé dans plusieurs domaines comme le domaine juridique. La loi a été le domaine d'application pionnier pour la fouille d'argument, et certainement parmi les plus réussis, avec le travail de Mochales Palau et Moens sur la Cour européenne des Droits de l'Homme (CEDH) (Mochales et Moens, 2011a) et les ensembles de données AraucariaDB pour l'extraction des composants d'arguments à partir d'une collection de documents juridiques structurés. Plus récemment, également le projet Vaccin / Vaccin (V / IP) (Ashley et Walker, 2013) a été réalisé, avec l'objectif d'extraire des arguments d'un ensemble de décisions judiciaires impliquant la réglementation des vaccins.

Dans cet article, nous proposons une approche qui fournit une chaîne de traitement pour analyser les flux de données d'opinions argumentées, permettant d'appliquer des techniques de fouille d'opinion et d'argument pour une sortie- de données structurées. Nous mettons en évidence dans cet article les résultats de notre approche sur les données issues de Twitter.

L'article est organisé de la manière suivante : la section 2 présente l'étude préliminaire. Section 3 présente notre approche proposée. Dans la section 4 nous fournissons les résultats de nos différentes expériences, concluons et indiquons nos perspectives.

---

1. <http://www.larousse.fr>

## 2. Étude préliminaire

Nous fournissons dans ce qui suit une étude préliminaire sur la détection d'opinion et sur la théorie d'argument.

### 2.1. Détection d'opinion

La fouille d'opinion peut être décrite comme un processus d'extraction et d'analyse automatiques des opinions sur une cible spécifique. Cette cible peut être un produit, une question comme la politique, l'économie, un événement, *etc.*. (Pang *et al.*, 2002) ont défini l'opinion selon sa polarité positive ou négative. (Liu, 2010a; Eirinaki *et al.*, 2012) ont défini la polarité de l'opinion selon trois classes; positive, négative et neutre. De nombreux chercheurs ont travaillé sur la définition des opinions. La première définition est fournie par (Liu, 2010b) et représentée comme suit :

**Definition 2.1** (*OPINION*) Une opinion est un quadruplet,  $(g, s, h, t)$  : où  $g$  est la cible d'opinion,  $s$  est la polarité de l'opinion sur la cible  $g$ ,  $h$  est le titulaire de l'opinion, et  $t$  est le moment où l'opinion a été exprimée.

Après avoir décomposé la cible d'opinion, (Liu, 2012) ont redéfini l'opinion comme suit :

**Definition 2.2** (*OPINION*) Une opinion est un 5-tuple,  $(e_i, a_{ij}, s_{ijkl}, h_k, t_l)$  : où  $e_i$  est le nom de l'entité,  $a_{ij}$  est l'aspect de l'entité  $e_i$ ,  $s_{ijkl}$  est la polarité de l'opinion sur l'aspect (positif, négatif, ou neutre)  $a_{ij}$  de l'entité  $e_i$ ,  $h_k$  est le détenteur de l'opinion, et  $t_l$  est le moment où l'opinion a été exprimée par  $h_k$ .

### 2.2. Théorie d'argument

Comme nous l'avons expliqué dans la section 1, en discutant et en communiquant, nous échangeons de nombreux points de vue et d'opinions sur un sujet spécifique et discutons les aspects positifs et négatifs de chaque cas, en donnant des raisons et des arguments qui supportent notre opinion.

Nous devons clarifier ce que nous entendons dire par un argument (Mochales et Moens, 2011b).

**Definition 2.3** (*ARGUMENT*) un argument donne une raison de soutenir, une affirmation ou une conclusion qui est susceptible de douter. Un argument est composé

*d'une ou plusieurs prémisses et d'une conclusion :*

**Prémisse :** Une prémisse est une proposition, affirmation entrant dans une démonstration dont on tire une conclusion.

**Conclusion :** Une conclusion est une déclaration supportée par des prémisses.

### 2.3. Travaux connexes

La tâche de détection de l'argument est généralement appelée détection ou identification de l'argument (Walton, 2009). Les sous-tâches les plus spécifiques sont la détection des tweets argumentatifs et la détection de ses composants (prémisses et conclusions) (Levy *et al.*, 2014).

Étant donné que la fouille d'argument s'agit d'un nouveau domaine de recherche, non seulement ses définitions, mais aussi ses approches et ses objectifs varient considérablement. Certains travaux visent à extraire les arguments à partir des données non structurées (Levy *et al.*, 2014). Dans cette section, nous passons en revue les méthodes d'apprentissage supervisée pour extraire automatiquement les arguments du texte

#### 2.3.1. Détection des phrases argumentatives

Plusieurs travaux se sont concentrés sur la détection des phrases argumentatives, des phrases qui contiennent un argument, ou une partie de celui-ci. Cette tâche est définie comme tâche de classification, et généralement résolue avec une méthode d'apprentissage supervisée. Deux questions cruciales dans cette étape impliquent le choix du classifieur et les caractéristiques à utiliser pour décrire l'argument. En ce qui concerne les classifieurs d'apprentissage adoptés, de nombreux travaux dans la littérature ont comparé plusieurs approches, notamment les classifieurs de naïve bayes, les machines à vecteurs de support (SVM), le classifieur d'entropie maximale (Palau et Moens, 2009 ; Park et Cardie, 2014), la régression logistique (Levy *et al.*, 2014), l'arbre de décision et les forêts d'arbres décisionnels (Stab et Gurevych, 2014b).

De nombreux travaux ont employé des méthodes classiques pour la représentation des caractéristiques de classification, y compris la représentation par sac de mots, l'étiquetage morphosyntaxique (POS), bi-grammes et trigrammes (Palau et Moens, 2009 ; Stab et Gurevych, 2014b). Encore plus de caractéristiques sont plus sophistiquées : le lexique d'opinions, le score de la subjectivité et le lexique des mots-clés qui reflètent la présence d'un argument (Levy *et al.*, 2014).

#### 2.3.2. Détection des composants d'argument

D'autres travaux de recherche ont porté sur la détection des composants d'argument en utilisant des fonctions syntaxiques (*e.g.*, règles de production grammaticales) et lexicales (*e.g.*, n-grammes) (Burststein *et al.*, 2003 ; Stab et Gurevych, 2014b). Cependant, les espaces de caractéristiques aussi grands et éparpillés peuvent causer des difficultés pour la sélection des caractéristiques.

Dans (Nguyen et Litman, 2015) les auteurs ont proposé un algorithme qui traite la post-production du modèle LDA (Blei *et al.*, 2003) pour extraire les mots d'argument

(indicateurs d'argument) et les mots de thème qui ont été utilisés comme des nouvelles caractéristiques pour détecter les composants de l'argument.

(Rosenthal et McKeown, 2012) ont déterminé automatiquement si une phrase contient une conclusion en utilisant la régression logistique et un lexique de sentiment. (Goudas *et al.*, 2014) ont identifié les prémisses et les conclusions dans les données issues des réseaux sociaux grecs en utilisant une approche en deux étapes : dans la première étape, ils ont détecté les phrases qui contiennent des arguments (détection de phrase argumentative). Dans la seconde étape, ils ont identifié les fragments contenant les prémisses et les conclusions. Leur approche est basée sur les champs aléatoires conditionnels *CRFs*.

Certains travaux ignorent le problème de détection des composants de l'argument. Plus particulièrement, (Mochales et Moens, 2011a) ont identifié les segments argumentatifs et les composants d'argument en utilisant l'arbre syntaxique, alors que (Stab et Gurevych, 2014b) et (Eckle-Köhler *et al.*, 2015) ont supposé que les phrases ont déjà été segmentées et ils sont concentrés sur leur classification en trois types : prémisses, conclusion et conclusion principale.

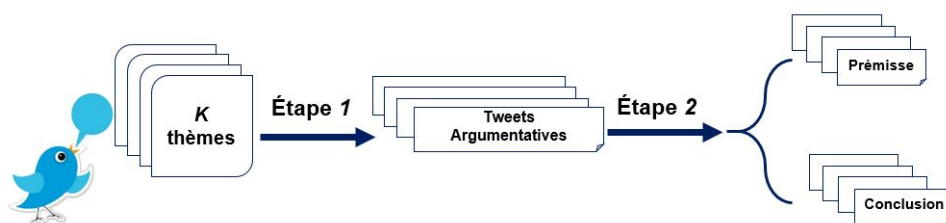
Les champs aléatoires conditionnels, les modèles de Markov cachés et d'autres méthodes similaires ont été appliqués avec succès à une grande variété de problèmes, y compris la reconnaissance d'entités nommées (NER) dans les tweets (Ritter *et al.*, 2011), l'extraction d'informations dans les réseaux sociaux (Imran *et al.*, 2013). Le principal avantage de ce type de méthodes repose sur la possibilité d'effectuer une classification séquentielle sur un ensemble de données, où les instances ne sont pas traitées indépendamment l'une de l'autre, mais l'ordre séquentiel est pris en compte. Une étiquette est assignée à chaque mot de la phrase dans un processus séquentiel unique. Le cadre de classification séquentielle s'est avéré extrêmement puissant non seulement pour les séquences, mais pour tout type de données relationnelles et structurées, où les relations et les dépendances entre les instances de données doivent être prises en compte.

### 3. Approche proposée

L'extraction d'arguments dans les données d'opinions est un nouveau domaine, à notre connaissance, cela constitue une étape innovante dans la détection de l'opinion. L'identification d'opinion basée sur les arguments associés nécessite une compréhension plus approfondie des techniques d'extraction d'arguments dans les réseaux sociaux en premier lieu. Motivés par ce besoin, nous essayons dans cet article de définir et de caractériser l'opinion avec les composants d'arguments associés, que nous appelons *opinion argumentée* et nous la définissons comme suit :

**Definition 3.1** (*OPINION ARGUMENTÉE*) Une *opinion argumentée* est un quadruplet  $(T_i, P_{ik}, Arg, CArg)$  où  $T_i$  est le nom du thème,  $P_{ik}$  est la polarité de l'opinion,  $Arg$  est le type de l'opinion : argumentatif ou non argumentatif,  $CArg$  est les composants d'argument.

Lors de notre participation à Clef 2017 Microblog Cultural Contextualization, nous avons proposé une approche de détection de polarité d'opinion sur des données issues de Twitter, concernant les festivals et les événements culturels, en combinant la fouille de séquences et la modélisation des thèmes *LDA* (Ouertatani et Latiri, 2017). En conséquence, nous avons eu un ensemble de thèmes avec la polarité de chacun, qui sont les deux premiers éléments du quadruplet proposé  $T_i, P_{ik}$ . Dans cet article, nous présentons un pipeline présentée dans la figure 1 pour identifier les éléments restants de l'opinion argumentée ( $T_i, P_{ik}, \mathbf{Arg}, \mathbf{CArg}$ ). L'étape 1 consiste en identification des tweets contenant des arguments. Cette étape



**Figure 1.** *Processus de pipeline*

est nécessaire pour sélectionner uniquement les tweets contenant des arguments, qui constituent l'entrée pour la deuxième étape.

L'étape 2, consiste en détection des composant de l'argument dans le tweet.

Nous employons un modèle de conclusion-prémisse basé sur le travail de Freeman (Freeman, 1991). Nous impliquons l'utilisation des champs aléatoires conditionnels (les *CRFs*) (Lafferty *et al.*, 2001) en suivant l'état d'art de (Goudas *et al.*, 2014; Hou *et al.*, 2012), afin de détecter les composants de l'argument, qui correspondent aux conclusions et prémisses.

L'exemple suivant illustre les composants de l'opinion argumentée :

l'opinion positive de l'auteur sur le festival de cannes est clairement exprimée dans la première phrase par la déclaration (conclusion) ( `#cannesfestival is my favorite moment`) supportée par la prémisse suivante (`movie...and it's an opportunity for fashion creator`):

```
@mccordalex Good morning :), #cannesfestival is my favorite
moment because of movie...and it's an opportunity for fashion
creator. Exciting !!!
```

### 3.1. *Étape 1 : Détection des tweets argumentatifs*

La première étape du processus concerne la tâche d'extraction de tweets qui contient un argument (ou une partie de celui-ci), et qui peut donc être définie comme argumentatifs. En se référant à l'exemple de la section précédente, cela correspond à l'extraction de tous les tweets contenant au moins un ou plusieurs de composants de l'argument. À cet effet nous réduisons le problème à un problème de classification. Nous proposons un classifieur binaire pour distinguer les tweets argumentatifs des tweets non-argumentatifs. Nous avons exploré un ensemble de méthodes d'apprentissage supervisées, tels que la régression logistique (Strano et Colosimo, 2006), les forêts d'arbres décisionnels (Breiman, 2001), les machines à vecteurs de support (*SVM*) (Cortes et Vapnik, 1995), et le classifieur naïve Bayes (Friedman *et al.*, 1997). Ces classifieurs sont entraînés dans un cadre supervisé, donc sur un ensemble de données d'apprentissage, chaque tweet d'apprentissage est présenté sous la forme d'un vecteur de caractéristiques, avec la classe associée (Argumentative / Non-Argumentative). La phase d'apprentissage produit un modèle qui peut ensuite être utilisé pour effectuer des prédictions sur les données de test.

Pour présenter le vecteur de caractéristiques, nous avons utilisé la représentation de mots "*le word embedding*". Nous avons exploité l'outil *Word2vec* (Mikolov *et al.*, 2013b).

*Word2vec* est un modèle, proposé à l'origine dans (Mikolov *et al.*, 2013a), basé sur l'utilisation des réseaux de neurones pour représenter les mots d'un dictionnaire par des vecteurs capables de prendre en compte la notion de contexte, facilitant l'analyse sémantique et syntaxique. (Mikolov *et al.*, 2013b). Cette approche est basée sur l'adoption d'une architecture neuronale simple qui permet des estimations de probabilités significativement meilleures que les modèles n-grammes.

Nous avons utilisé comme caractéristiques la somme des vecteurs de words embeddings pour former le vecteur de caractéristiques (Lilleberg *et al.*, 2015).

### 3.2. *Étape 2 : Détection de composants de l'argument*

Une fois que nous avons identifié les tweets argumentatifs, notre approche procède à l'identification des segments qui expriment une conclusion ou une prémisse. Nous présentons dans cette étape l'intégration de la segmentation et l'étiquetage dans un seul modèle *CRFs* (Lafferty *et al.*, 2001). Nous avons utilisé l'analyse syntaxique de surface (chunking) qui identifie les limites de segments (chunks). Chaque tweet est représenté sous forme des unités lexicales couplées à leurs étiquetage en catégories morphosyntaxiques. En effet, les *CRFs* peut capturer les dépendances de longue distance entre les segments et aide à maintenir les aspects linguistiques tels que l'ordre des mots dans le tweet. Les *CRFs* peut également prendre en compte le contexte local, ce qui est important pour la nature de notre problème.



### 3.2.1. Champs aléatoires conditionnels

Les champs aléatoires conditionnels ou *CRFs* sont des modèles probabilistes discriminants introduits par (Lafferty *et al.*, 2001) pour l'annotation séquentielle. Ils ont été utilisés dans de nombreuses tâches de Traitement des Langues, où ils donnent d'excellents résultats comme dans (Tellier *et al.*, 2010).

Les *CRFs* permettent d'associer à une variable aléatoire qui représente les séquences d'unités  $X$ , à une variable aléatoire représentant les séquences d'étiquettes correspondantes  $Y$ . Tous les composants  $Y_i$  de  $Y$  appartiennent à un alphabet fini.

Les distributions des variables aléatoires  $X$  et  $Y$  sont liées, mais dans une approche discriminante, nous construisons un modèle conditionnel  $p(Y|X)$  qui ne nécessite pas la modélisation de  $p(X)$  :

**Definition 3.2** Soit  $G = (V, E)$  un graphe tel que  $V$  définit l'ensemble des noeuds,  $E$  l'ensemble des arcs.

$Y = (Y_v)_{v \in V} \in V$ , de sorte que  $Y$  soit indexé par les sommets de  $G$ . Alors  $(X, Y)$  est un champ aléatoire conditionnel dans le cas où, conditionnée par  $X$ , les variables aléatoires  $Y_v$  vérifient à la propriété Markov par rapport au graphe :

$$p(Y_v|X, Y_w, w \neq v) = p(Y_v|X, Y_w, w \sim v) \quad [1]$$

Où  $w_v$  signifie que  $w$  et  $v$  sont voisins dans  $G$ . Cette propriété est donc satisfaite si l'état du système (le mot dans lequel on se trouve) ne dépend que des états voisins (la séquence), ainsi que des probabilités de transitions entre états. Le *CRFs* modélise les probabilités conditionnelles suivantes :

$$p(y|x) = \frac{\exp(\theta \cdot f(x, y))}{Z_\theta(x)} \quad [2]$$

Où  $f(x, y)$  sont appelées fonctions caractéristiques (features),  $\theta$  sont les paramètres, et :

$$Z_\theta(x) = \sum_y \exp(\theta \cdot f(x, y)) \quad [3]$$

est une constante de normalisation (Lafferty *et al.*, 2001).

### 3.2.2. Caractéristiques

Pour entraîner nos *CRFs*, nous avons donc transformé le corpus d'apprentissage en isolant les unités composant les prémisses et les conclusions et en les étiquetant. Nous présentons les caractéristiques requises pour les *CRFs* qui comprennent :

**caractéristique lexicale** : nous avons présenté le tweets sous forme des segments de mots contigus comme dans l'exemple 1.1 ci-dessous.

**Étiquetage morphosyntaxique** : pour associer à chaque mot du tweet les informations grammaticales correspondantes comme la partie du discours, le genre, le nombre, etc.

**Analyse syntaxique de surface(chunking)** : pour identifier les unités lexicales élémentaires des tweets et capturer les dépendances syntaxiques entre eux en utilisant le chunker *NLTK-Chunk*. *NLTK-Chunk* identifie les segments non récursifs simples ou chunks (Abney, 1991). Les chunks définissent la structure syntaxique superficielle de chaque segment de tweets ;

– **Balise IOB** : Un chunk présente un mot d’une séquence. Ainsi, la balise *I-XP* signifie que ce mot est dans le bloc *XP* ; le *B-XP* par défaut signifie que le mot est au début du bloc *XP* et *O* signifie que le mot courant est en dehors de n’importe quel groupes syntaxique ;

– **Balise de caractéristique grammaticale** : Si le mot est en tête du groupe, il a la caractéristique grammaticale du groupe entier. Les autres mots du groupe qui ne sont pas la tête ont *O* comme caractéristique. Par exemple, la caractéristique grammaticale *NP* : (Noun Phrase), *VP* :(verbal Phrase).

l’exemple 1.2 présente le vecteur d’entrée pour *CRFs*.

Tweet 1 : @mccordalex Good morning :), #cannesfestival is my favorite moment because of movie...and it’s an opportunity for fashion creator. Exciting !!!

Exemple 1.1 :

Good morning :),  
cannesfestival is my favorite moment because of movie...  
it’s an opportunity for fashion creator.  
Exciting!!!

Exemple 1.2 :

Good <B-NP> morning <I-NP> :) <I-NP> , <O> #cannesfestival <B-NP>  
is <B-VP> my <B-NP> favorite <I-NP> moment <I-NP> because <B-PP>  
of <I-PP> movie <B-NP> ... <O> and <O> it’s <B-VP> an <B-NP>  
opportunity <I-NP> for <B-PP> fashion <B-NP> creator <I-NP>.<O>  
Exciting <B-NP>!!<I-NP>!<O>

#### 4. Étude empirique

Nous présentons les résultats des expérimentations que nous avons menés Pour évaluer notre approche nous suivons deux étapes :

- 1) Étape 1 : détection des tweets argumentatifs en utilisant 4 classifieurs (NB, SVM, forêt d’arbres décisionnels et la régression logistique ),
- 2) Étape 2 : l’identification des composants d’argument en utilisant les *CRFs* et la comparaison les performances de *CRF* aux résultats obtenus en utilisant les 4 classifieurs.

#### 4.1. Mesures

Les mesures d'évaluation utilisées sont : l'Exactitude (*Accuracy*) Précision, Rappel et F-mesure. l'exactitude est la mesure de performance la plus intuitive et il s'agit simplement d'un rapport d'instances correctement prédites des deux classes qui doivent être classées. La précision, le rappel et la mesure F-mesure peuvent nous aider à évaluer notre approche. La mesure de précision <sup>2</sup> indique à quel point le classifieur peut classer correctement les instances, est le rapport entre les observations positives correctement prédites (vrai positif VP) et les observations positives totales prédite (vrai positif VP et faux positif FP). La mesure rappel <sup>3</sup> fournit la fraction des instances pertinentes qui sont correctement récupérées (VP) parmi toutes les instances possibles (vrai positif VP et faux négatif FN). La mesure F-mesure <sup>4</sup> est la moyenne harmonique de la précision et du rappel.

Pour évaluer la performance prédictive des classifieurs et étudier leur performance en dehors de l'échantillon à un nouvel ensemble de données également connu sous le nom de données de test, nous avons utilisé la technique de validation croisée. La motivation de l'utiliser c'est que lorsque nous ajustons un modèle, nous l'adaptions à un ensemble de données d'apprentissage. Sans validation croisée, nous n'avons que des informations sur la façon dont notre modèle fonctionne pour nos données de test. Idéalement, nous aimerions voir comment le modèle fonctionne lorsque nous avons de nouvelles données en termes d'exactitude de ses prédictions.

#### 4.2. Corpus

Les expériences ont été menées sur l'ensemble de données issues de notre participation à la Conférence *CLEF 2017* (Cappellato *et al.*, 2017). Nous avons participé à la sous-tâche de filtrage et d'analyse d'opinion des microblogs traitant les festivals et les événements culturels.

En conséquence, nous avons eu un ensemble de thèmes avec la polarité d'opinions de chacun. Pour l'évaluation, nous avons utilisé le thème Film comme décrit dans le tableau suivant :

	Tweet Positif	Tweet négatif
Thème Film	997	5474

**Tableau 1.** *Thème Film*

$$2. \textit{precision} = \frac{VP}{VP+FP}$$

$$3. \textit{rappel} = \frac{VP}{VP+FN}$$

$$4. \textit{F - mesure} = 2 * \frac{(\textit{Rappel} * \textit{Précision})}{(\textit{Rappel} + \textit{Précision})}$$

### 4.3. Évaluation de la détection des tweets argumentatifs

#### 4.3.1. Protocole

Afin de classer un tweet comme un tweet argumentatif ou non, nous avons utilisé un certain nombre de classifieurs. Chaque tweet est représenté par un vecteur de caractéristiques de taille fixe, en utilisant les words embedding décrites dans la sous-section 3.1.

L'apprentissage et l'évaluation des classifieurs ont été réalisés en utilisant le corpus décrit dans la sous-section 4.3. Nous avons choisi 1000 tweets dont 612 sont positifs et 388 sont négatifs du corpus à annoter sur la plateforme CrowdFlower<sup>5</sup>. Nous avons acquis 264 tweets non argumentatifs et 738 tweets argumentatifs qui ont été utilisés en entrée pour une variété de classifieurs, y compris la régression logistique, les forêts d'arbres décisionnels, les machines à vecteurs de support (*SVM*) et le classifieur naïve Bayes

Nous avons effectué une validation croisée en 10 parties sur l'ensemble de tweets. La performance globale est présentée dans le tableau 2 et dans le tableau 3 les résultats de la classe argumentative :

	Exactitude	Précision	Rappel	F-mesure
SVM	0.806	0.796	0.806	0.800
forêt d'arbres de décision	<b>0.834</b>	<b>0.826</b>	<b>0.835</b>	<b>0.807</b>
Naïve Bayes	0.78	0.812	0.79	0.799
Régression logistique	0.789	0.812	0.790	0.799

**Tableau 2.** Résultats de divers classifieurs pour l'étape 1, évalués avec la validation croisée 10-parties (les deux classes)

	Précision	Rappel	F-mesure
SVM	0.823	0.839	0.829
forêt d'arbres de décision	<b>0.875</b>	<b>0.849</b>	<b>0.861</b>
Naïve Bayes	0.87	0.840	0.808
Régression logistique	0.811	0.827	0.805

**Tableau 3.** Résultats de divers classifieurs pour l'étape 1, évalués avec la validation croisée 10-parties (classe argumentative)

### 4.4. Évaluation de la détection de composants de l'argument

Nous présentons dans cette sous-section la performance de l'étape 2 en utilisant des champs aléatoires conditionnels. Pour l'évaluation, nous devrions utiliser un cor-

5. <https://www.crowdflower.com/>

pus annoté comme un ensemble d'apprentissage. Seuls quelques ensembles de données annotés sont disponibles au public, nous citons ci-dessous quelques corpus :

- L'Araucaria DB <sup>6</sup> un corpus bien connu qui a été utilisé avec succès dans plusieurs travaux sur l'extraction d'arguments, comprend deux types de données : un ensemble de données structurées en anglais collectées dans le cadre de projet à l'Université de Dundee <sup>7</sup>, et un ensemble de données multilingue non structuré d'analyses fournies par les utilisateurs. Les données ont été recueillies à partir de journaux, de dossiers parlementaires (au Royaume-Uni, aux États-Unis et en Inde), de tribunaux (du Royaume-Uni, des États-Unis et du Canada), des magazines (Royaume-Uni, États-Unis et Inde) et de sources supplémentaires, Human Rights Watch et GlobalWarming.org.

- La base de données de la Cour européenne des droits de l'homme (CEDH), spécialement conçue pour l'extraction d'arguments dans les documents juridiques (Mochales et Moens, 2011a).

- Le corpus des essais persuasifs en anglais est choisi parmi les forums en ligne collecté par (Stab et Gurevych, 2014a). Ce corpus contient des annotations de composants d'argument au niveau de la clause ainsi que des relations argumentatives. En particulier, il comprend des annotations de conclusions et de prémisses. Nous avons choisi ce corpus comme jeu de données d'apprentissage parmi d'autres car ce type de données est très proche des données de microblogs.

L'ensemble de données de test a été composé à partir des tweets argumentatifs acquis à partir de l'étape 1. L'objectif de l'étape 2 consiste à examiner si l'un des séquences détectées par le modèle les *CRFs* est soit une conclusion ou une prémisse, et en comparant le résultat obtenu par rapport à la performance des 4 classifieurs. Les performances de la deuxième étape sont indiquées dans le tableau 4. Les résultats de divers classifieurs pour les deux classes Prémisse et Conclusion sont présentés séparément dans les tableaux 5 et 6 respectivement.

	CRFs	SVM	Forêt d'AD	NB	RégressionL
Exactitude	<b>0.76</b>	0.617	0.61	0.595	0.596

**Tableau 4.** Résultats de divers classifieurs et CRF pour l'étape 2, évalués avec une validation croisée 10-parties (les deux classes)

Il est clair que dans l'étape 2 les résultats du modèle les *CRFs* sont au dessus des autres classifieurs. L'application du modèle les *CRFs* pour la tâche d'étiquetage de séquences est bénéfique car elle prend en compte les corrélations entre les séquences (prémisses / conclusions) dans les voisinages et décode conjointement la meilleure chaîne de séquences pour un tweet donné.

6. <http://corpora.aifdb.org/>

7. <http://araucaria.computing.dundee.ac.uk/doku.php>

	Précision	Rappel	F-mesure
CRFs	<b>0.69</b>	<b>0.8</b>	<b>0.74</b>
SVM	0.592	0.617	0.559
forêt d'arbres de décision	0.593	0.7	0.64
Naïve Bayes	0.605	0.62	0.614
Régression logistique	0.623	0.596	0.541

**Tableau 5.** Résultats de divers classifieurs et CRF pour l'étape 2, évalués avec une validation croisée 10-parties ( classe Prémisse)

	Précision	Rappel	F-mesure
CRFs	<b>0.87</b>	<b>0.71</b>	<b>0.78</b>
SVM	0.603	0.57	0.58
forêt d'arbres de décision	0.612	0.612	0.541
Naïve Bayes	0.656	0.596	0.539
Régression logistique	0.603	0.523	0.56

**Tableau 6.** Résultats de divers classifieurs et CRF pour l'étape 2, évalués avec une validation croisée 10-parties (classe Conclusion)

## 5. Conclusion et perspectives

Nous avons présenté dans cet article une définition sur l'opinion argumentée issues de Twitter. Pour extraire l'argument nous avons proposé une approche en deux étapes. Dans la première étape, nous avons détecté les tweets argumentatifs. Nous avons employé une approche basée sur l'apprentissage supervisée. Les meilleurs résultats ont été obtenus en utilisant la forêt d'arbres décisionnels. Dans la deuxième étape, nous avons détecté les composants de l'argument. Nous avons comparé la performance les *CRFs* et les quatre classifieurs utilisé dans la première étape. Les *CRFs* se sont la meilleure méthode par rapport aux approches statistiques.

Comme perspectives, il serait intéressant d'explorer de nouvelles caractéristiques supplémentaires pour la première étape afin d'augmenter la précision encore plus loin comme des mots de repère qui apparaissent dans un argument. De plus, il serait intéressant d'examiner les dépendances entre les différents composants d'un argument donné pour construire la structure de l'argument.

## 6. Bibliographie

- Abney S. P., « Parsing by chunks », *Principle-based parsing*, Springer, p. 257-278, 1991.
- Ashley K. D., Walker V. R., « Toward constructing evidence-based legal arguments using legal decision documents and machine learning », *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Law*, ACM, p. 176-180, 2013.

- Benamara F., Grouin C., Karoui J., Robba V. M. I., « Analyse d'opinion et langage figuratif dans des tweets : présentation et résultats du Défi Fouille de Textes DEFT2017 », *24e Conférence sur le Traitement Automatique des Langues Naturelles (TALN)*, p. 1, 2017.
- Blei D. M., Ng A. Y., Jordan M. I., « Latent dirichlet allocation », *Journal of machine Learning research*, vol. 3, , p. 993-1022, 2003.
- Breiman L., « Random forests », *Machine learning*, vol. 45, n° 1, p. 5-32, 2001.
- Burstein J., Marcu D., Knight K., « Finding the WRITE stuff : Automatic identification of discourse structure in student essays », *IEEE Intelligent Systems*, vol. 18, n° 1, p. 32-39, 2003.
- Cappellato L., Ferro N., Goeuriot L., Mandl T. (eds), *Working Notes of CLEF 2017 - Conference and Labs of the Evaluation Forum, Dublin, Ireland, September 11-14, 2017*, vol. 1866 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2017.
- Cortes C., Vapnik V., « Support-vector networks », *Machine learning*, vol. 20, n° 3, p. 273-297, 1995.
- Eckle-Kohler J., Kluge R., Gurevych I., « On the Role of Discourse Markers for Discriminating Claims and Premises in Argumentative Discourse. », *EMNLP*, p. 2236-2242, 2015.
- Eirinaki M., Pisal S., Singh J., « Feature-based opinion mining and ranking », *Journal of Computer and System Sciences*, vol. 78, n° 4, p. 1175 - 1184, 2012.
- Freeman J. B., *Dialectics and the macrostructure of arguments : A theory of argument structure*, vol. 10, Walter de Gruyter, 1991.
- Friedman N., Geiger D., Goldszmidt M., « Bayesian network classifiers », *Machine learning*, vol. 29, n° 2-3, p. 131-163, 1997.
- Goudas T., Louizos C., Petasis G., Karkaletsis V., « Argument Extraction from News, Blogs, and Social Media. », *SETN*, Springer, p. 287-299, 2014.
- Hou L., Li P., Zhu Q., Cao Y., « Event argument extraction based on crf », *Workshop on Chinese Lexical Semantics*, Springer, p. 32-39, 2012.
- Imran M., Elbassouni S., Castillo C., Diaz F., Meier P., « Practical extraction of disaster-relevant information from social media », *Proceedings of the 22nd International Conference on World Wide Web*, ACM, p. 1021-1024, 2013.
- Lafferty J., McCallum A., Pereira F. C., « Conditional random fields : Probabilistic models for segmenting and labeling sequence data », 2001.
- Levy R., Bilu Y., Hershovich D., Aharoni E., Slonim N., « Context dependent claim detection », 2014.
- Lilleberg J., Zhu Y., Zhang Y., « Support vector machines and word2vec for text classification with semantic features », *Cognitive Informatics & Cognitive Computing (ICCI\* CC), 2015 IEEE 14th International Conference on*, IEEE, p. 136-140, 2015.
- Liu B. *IEEE Intelligent Systems*, vol. 25, n° 3, p. 76-80, 8, 2010a.
- Liu B., « Sentiment Analysis and Subjectivity. », *Handbook of natural language processing*, vol. 2, p. 627-666, 2010b.
- Liu B., « Sentiment analysis and opinion mining », *Synthesis lectures on human language technologies*, vol. 5, n° 1, p. 1-167, 2012.
- Mikolov T., Chen K., Corrado G., Dean J., « Efficient estimation of word representations in vector space », *arXiv preprint arXiv :1301.3781*, 2013a.

- Mikolov T., Yih W.-t., Zweig G., « Linguistic regularities in continuous space word representations. », *hlt-Naacl*, vol. 13, p. 746-751, 2013b.
- Mochales R., Moens M.-F., « Argumentation mining », *Artificial Intelligence and Law*, vol. 19, n<sup>o</sup> 1, p. 1-22, Mar, 2011a.
- Mochales R., Moens M.-F., « Argumentation mining », *Artificial Intelligence and Law*, vol. 19, n<sup>o</sup> 1, p. 1-22, 2011b.
- Nguyen H., Litman D. J., « Extracting Argument and Domain Words for Identifying Argument Components in Texts. », *ArgMining@ HLT-NAACL*, p. 22-28, 2015.
- Ouertatani A., Latiri C., « Opinion Polarity Detection in Twitter Data combining Sequence Mining and Topic Modeling », *Working Notes of CLEF 2017 - Conference and Labs of the Evaluation Forum, Dublin, Ireland, September 11-14, 2017.*, 2017.
- Palau R. M., Moens M.-F., « Argumentation mining : the detection, classification and structure of arguments in text », *Proceedings of the 12th international conference on artificial intelligence and law*, ACM, p. 98-107, 2009.
- Pang B., Lee L. *et al.*, « Opinion mining and sentiment analysis », *Foundations and Trends® in Information Retrieval*, vol. 2, n<sup>o</sup> 1-2, p. 1-135, 2008.
- Pang B., Lee L., Vaithyanathan S., « Thumbs up? : sentiment classification using machine learning techniques », *Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10*, Association for Computational Linguistics, p. 79-86, 2002.
- Park J., Cardie C., « Identifying Appropriate Support for Propositions in Online User Comments. », *ArgMining@ ACL*, p. 29-38, 2014.
- Pozzi F. A., Fersini E., Messina E., Liu B., *Sentiment Analysis in Social Networks*, Morgan Kaufmann, 2016.
- Ritter A., Clark S., Etzioni O. *et al.*, « Named entity recognition in tweets : an experimental study », *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, Association for Computational Linguistics, p. 1524-1534, 2011.
- Rosenthal S., McKeown K., « Detecting opinionated claims in online discussions », *Semantic Computing (ICSC), 2012 IEEE Sixth International Conference on*, IEEE, p. 30-37, 2012.
- Ruffiero V. R., *The art of thinking : A guide to critical and creative thought*, n<sup>o</sup> 153.42 R931a, Harper & Row, 1988.
- Stab C., Gurevych I., « Annotating Argument Components and Relations in Persuasive Essays. », *COLING*, p. 1501-1510, 2014a.
- Stab C., Gurevych I., « Identifying Argumentative Discourse Structures in Persuasive Essays. », *EMNLP*, p. 46-56, 2014b.
- Strano M., Colosimo B. M., « Logistic regression analysis for experimental determination of forming limit diagrams », *International Journal of Machine Tools and Manufacture*, vol. 46, n<sup>o</sup> 6, p. 673-682, 2006.
- Tang H., Tan S., Cheng X., « A survey on sentiment detection of reviews », *Expert Systems with Applications*, vol. 36, n<sup>o</sup> 7, p. 10760-10773, 2009.
- Tellier I., Eshkol I., Taalab S., Prost J.-P., « Pos-tagging for oral texts with crf and category decomposition », *Research in Computing Science*, vol. 46, p. 79-90, 2010.
- Walton D., *Argumentation Theory : A Very Short Introduction*, Springer US, Boston, MA, p. 1-22, 2009.