

ResTS : Système de Résumé Automatique des Textes d'Opinions basé sur Twitter et SentiWordNet

Jihene Jmal

LARODEC, ISG, Université de Tunis, 2000, Le Bardo, Tunisie
fer.jmal_jihene@hotmail.fr

RESUME

Comme le E-commerce est devenu de plus en plus populaire, le nombre de commentaires des internautes est en croissance constante. Les opinions sur le Web affectent nos choix et nos décisions. Il s'avère alors indispensable de traiter une quantité importante de critiques des clients afin de présenter à l'utilisateur l'information dont il a besoin dans la forme la plus appropriée. Dans cet article, nous présentons ResTS, un nouveau système de résumé automatique de textes d'opinions basé sur les caractéristiques des produits. Notre approche vise à transformer les critiques des utilisateurs en des scores qui mesurent le degré de satisfaction des clients pour un produit donné et pour chacune de ses caractéristiques. Ces scores sont compris entre 0 et 1 et peuvent être utilisés pour la prise de décision. Nous avons étudié les opinions véhiculées par les noms, les adjectifs, les verbes et les adverbes, contrairement aux recherches précédentes qui utilisent essentiellement les adjectifs. Les résultats expérimentaux préliminaires montrent que notre méthode est comparable aux méthodes classiques de résumé automatique basées sur les caractéristiques des produits.

ABSTRACT

System of Customer Review Summarization using Twitter and SentiWordNet

As E-commerce is becoming more and more popular, the number of customer reviews raises rapidly. Opinions on the Web affect our choices and decisions. Thus, it is more efficient to automatically process a mixture of reviews and prepare to the customer the required information in an appropriate form. In this paper, we present ResTS, a new system of feature-based opinion summarization. Our approach aims to turn the customer reviews into scores that measure the customer satisfaction for a given product and its features. These scores are between 0 and 1 and can be used for decision making and then help users in their choices. We investigated opinions extracted from nouns, adjectives, verbs and adverbs contrary to previous research which use only adjectives. Experimental results show that our method performs comparably to classic feature-based summarization methods.

MOTS-CLES : Fouille d'opinion, Classification, Intensité de l'Opinion, Résumé de texte d'opinion, Popularité.

KEYWORDS: Opinion mining, Sentiment Classification, Opinion Strength, Feature-based Opinion Summarization, Feature Buzz Summary.

1 Introduction

Dans le Web 2.0 (Web social ou participatif), l'utilisateur est un acteur principal qui par-

tage des documents, des informations, des avis. Il interagit, collabore avec autrui, s'exprime et donne son opinion. Il a des services à sa disposition tels que les réseaux sociaux (twitter, facebook, etc.), les blogs, les forums, les wikis, les sites de partages de vidéos, de photos, de musiques, etc. L'utilisation fréquente de ces services fournit un contenu généré par l'utilisateur (UGC : User Generated Content) qui représente de nos jours une quantité de données qui se mesure en yotaoctets (10^{24}). Ce contenu est composé généralement de données textuelles qui sont porteuses d'opinions et de sentiments. L'accès au contenu sémantique des ces données, préalable à la connaissance des opinions qu'elles véhiculent, représente un enjeu pour de nombreux acteurs. Par exemple :

- le consommateur, c'est-à-dire chacun de nous, qui veut s'informer avant toute décision qu'elle soit d'achat ou autre;
- les fournisseurs de biens et de services qui cherchent à se positionner les uns par rapport aux autres dans un univers hautement compétitif et face à une demande de plus en plus complexe à identifier;
- les chercheurs : économistes, sociologues,... ou simplement les responsables publics qui cherchent à comprendre le comportement individuel ou collectif pour anticiper, réguler ou ajuster les rapports entre les différents agents socio-économiques.

C'est dans ce contexte que s'introduit la fouille d'opinion (Opinion Mining, Sentiment Analysis ou Subjectivity Analysis) qui est un sous domaine de la fouille de texte. Son but étant de ressortir les marques d'opinions et de sentiments des documents textuels. Une opinion peut être définie comme l'expression des sentiments d'une personne envers une entité (Liu, 2010). En outre, l'e-commerce devient de plus en plus populaire. Les marchands et les fabricants de produits permettent aux clients de donner leurs avis et opinions sur les produits ou services qu'ils ont vendus (par exemple amazon.com, epinions.com). De plus, les opinions disponibles sur le Web influent sur nos choix et décisions. En effet, d'après une étude menée en 2009 par le CRÉDOC (Centre de Recherche pour l'Étude et l'Observation des Conditions de Vie), 57% des internautes français ont cherché des avis des autres sur le Web et 66% d'entre eux font confiance en ces commentaires (Lehuédé, 2009). La fouille d'opinion peut être divisée en trois sous domaines qui sont la classification de la subjectivité (subjectif/objectif) (Riloff et al, 2003), la classification des sentiments (positif/négatif ou positif/négatif/neutre)(Pang et Lee, 2002), (Wilson et al, 2004) et (Blitzer et al, 2007) et le résumé d'opinions (Hu et Liu, 2004), (Popescu et Etzioni, 2005) et (Gamon et al, 2005).

Nous proposons une nouvelle approche de résumé automatique des textes d'opinions basée sur les commentaires des utilisateurs. Cette approche vise à transformer ces commentaires en des scores qui mesurent l'intensité de l'opinion. Ces scores peuvent être utilisés pour la prise de décision et aident les utilisateurs dans leurs choix. Pour ce faire, nous avons commencé par extraire les caractéristiques des produits à partir des critiques des utilisateurs (exemple batterie, écran, son, image, etc.). Ensuite, nous avons attribué à chaque caractéristique un score calculé à partir de sa fréquence d'apparition dans le corpus pondérée par sa popularité dans le Web 2.0, en particulier sur Twitter¹ ; la plateforme de microblogage la plus populaire. Nous avons par la suite identifié les phrases d'opinion et affecté

¹ www.twitter.com

à chaque verbe et adjectif un score de SentiWordNet (Baccianella et al, 2010). Si la phrase contient un adverbe, ces scores sont pondérés par l'intensité de l'opinion véhiculée par cet adverbe en se référant à la liste de modificateurs (en anglais *intensifier* et *diminisher*) que nous avons préparé. Nous avons enfin calculé le score de tout le produit qui mesure la satisfaction globale des clients. Voici un exemple de résumé généré par notre système pour le produit *iPod* :

Produit : iPod

Satisfaction Client = 60%

Caractéristique 1 : Player : Popularité = 70%

Satisfaction Client = 83%

Caractéristique 2 : Ecran : Popularité = 54%

Satisfaction Client = 62%

....

Les caractéristiques des produits sont classées en fonction de leurs popularités sur le web 2.0. Dans notre conception, un produit n'est pas simplement considéré comme recommandé ou non recommandé, au contraire, nous laissons l'utilisateur libre de faire son choix en se référant aux différents scores que nous mettons à sa disposition traduisant la satisfaction des clients pour l'ensemble du produit et encore pour chacune de ses caractéristiques. Lors du calcul de ces scores, nous avons étudié l'opinion véhiculée par les noms, adjectifs, verbes et adverbes, contrairement aux autres recherches qui utilisent principalement les adjectifs.

2 Etat de l'art

Nous proposons dans cet article deux types de résumés qui sont le résumé d'opinion basé sur les caractéristiques des produits (Feature-based Opinion summarization), et le résumé de leurs popularités qui montre aux entreprises ce qu'intéresse réellement leurs clients (Feature Buzz Summary). Nous avons également fusionné deux axes de recherche à savoir le résumé basé sur les caractéristiques (Hu et Liu, 2004) (Liu et Ding, 2008) (Zhang et Liu, 2011) et l'identification de l'intensité de l'opinion (Wilson et al, 2004). Nous nous sommes basées essentiellement sur l'approche de Hu et Liu (Hu et Liu, 2004). Les deux auteurs utilisent les règles d'association pour extraire les caractéristiques fréquentes des produits. Pour identifier les mots d'opinion (les adjectifs seulement), ils ont eu recours à WordNet² en conjonction avec une liste de mots clés subjectifs (seed words) manuellement préparée. Leur système extrait uniquement les caractéristiques explicites. Une année plus tard, ces auteurs ont mis en œuvre Opinion Observer (Liu et al, 2005), un système offrant une comparaison visuelle entre produits en tenant compte des critiques des utilisateurs sur le Web. Ils identifient les caractéristiques des produits à partir des rubriques *Pros* destinée aux avis positifs et *Cons* celle des avis négatifs.

Plusieurs recherches ont étudié le problème de la détection de mots d'opinion. Il y a des

² <http://wordnet.princeton.edu/>

approches fondées sur le corpus (Corpus-based Approach) (Hatzivassiloglou et McKeown, 1997), (Wiebe, 2000), (Kanayama et Nasukawa, 2006) et (Qiu et al, 2009), d'autres basées sur le dictionnaire (Dictionary-based Approach) (Hu et Liu, 2004), (Kim et Hovy, 2004), (Kamps et al, 2004), (Esuli et Sebastiani, 2005), (Takamura et al, 2005), (Andreevskaia et Bergler, 2006), (Dragut et al, 2010) et (Bouchlaghem et al, 2010). Hu et Liu utilisent seulement les adjectifs pour la détection des opinions. Ils construisent manuellement une liste d'adjectifs qu'ils utilisent pour prédire l'orientation de la phrase et utilisent WordNet pour alimenter la liste par les synonymes et les antonymes des adjectifs dont on connaît la polarité. Ils assignent 1 à chaque adjectif positif et 0 à chaque adjectif négatif. Toutefois, dans notre conception, les adjectifs, les verbes et les adverbes jouent un rôle important dans l'analyse des sentiments. Ils sont tous utilisés pour exprimer une opinion ou une émotion dans le texte, par exemple, le verbe apprécier dans «J'apprécie ce produit» inspire un sentiment positif, même si la phrase ne contient ni adjectif, ni adverbe. Dans (Liu et al, 2005), les auteurs comptent le nombre d'occurrences de chaque entité dans la rubrique *Pros* exprimant un avis positif et *Cons* celle des avis négatifs. Dans (Zhang et Liu, 2011), les auteurs ont montré que les syntagmes nominaux et le substantif peuvent aussi enfermer des opinions. Ils comptent le nombre de phrases positives et négatives pour chaque fonctionnalité du produit en utilisant le lexique d'opinion préparé par (Ding et al, 2008). Leur approche permet d'atteindre une précision moyenne d'environ 0,44. Dans notre conception, nous rejoignons l'avis de ces auteurs. Nous considérons également que les noms peuvent exprimer une opinion. En outre, déceler la polarité de l'opinion n'est toujours pas suffisant. La force (intensité) de l'opinion est également nécessaire. En effet, la subjectivité est exprimée de différentes manières ; «*good battery* » est différent de «*great battery* » et de «*excellent battery* ». (Wilson et al, 2004) et (Pang et Lee, 2005) mettent l'accent sur la détection de la force de l'opinion. (Wilson et al, 2004) utilisent les techniques de boosting, rule learning et support vector regression. (Pang et Lee, 2002) et (Turney, 2002) classent les documents comme « thumbs up » ou « thumbs down », selon l'opinion qu'ils véhiculent. Cependant, (Pang et Lee, 2005) exploitent les techniques d'apprentissage automatique pour donner un score de 1 à 5 aux passages d'opinions.

3 Approche proposée

Notre approche est basée sur les travaux de (Hu et Liu, 2004). La figure 1 présente le modèle proposé. Elle a été mise en œuvre dans notre système ResTS. Nous commençons par recueillir les commentaires des internautes à partir du Web et procédons par l'opération de prétraitement du corpus collecté. Notons que notre système effectue toutes les étapes suivantes d'une manière automatisée et sans aucune intervention humaine. Rappelons que l'opinion est une expression des sentiments d'une personne envers une entité ou un aspect de l'entité (Liu, 2010). Une entité peut être un produit, une personne, un événement, une organisation ou un sujet. Elle est représentée comme une hiérarchie de composants, de sous-composant et ainsi de suite où chaque nœud représente un composant et est associé à un ensemble d'attributs (Liu, 2010). Par conséquent, l'entité elle-même peut également être considérée comme une caractéristique. Une critique de l'entité elle-même est appelée une opinion générale comme dans «*I like this iPod* ». Une critique d'une de ses caractéristiques est appelée une opinion spécifique comme dans «*the battery is really good* ». Comme Hu et Liu, notre tâche est loin d'être un résumé traditionnel de texte. À partir des critiques des utilisateurs, nous proposons un résumé structuré qui donne une vue globale et concise des

opinions des clients. Hu et Liu ne présentent que le nombre de passages jugés positifs et ceux négatifs pour chacune des caractéristique du produit. Notre système offre plus de détails. Nous fournissons un score révélant le degré de satisfaction des clients pour un produit donné et pour chacune de ses caractéristiques. Notre système n'est pas seulement basé sur le corpus puisque nous avons eu recours au Web 2.0 à chaque étape.

3.1 Prétraitement

Selon Liu, les commentaires des utilisateurs sont en trois formats (Liu, 2005):

- Format 1 - Pros et Cons : les consommateurs sont invités à décrire les avantages et les inconvénients séparément dans les rubriques Pros et Cons.
- Format 2 - Pros, Cons et détail : Les consommateurs décrivent les avantages et les inconvénients séparément dans les rubriques Pros et Cons et écrivent de plus des commentaires détaillés.
- Format 3 - Format libre : Les consommateurs écrivent des avis en format libre, sans séparation entre les avantages et les inconvénients.

Dans ce papier nous utilisons les critiques du troisième format. Tous les exemples qui suivent portent sur le produit *iPod* et toutes les critiques sont en anglais. Le tableau 1, ci-dessous, présente quelques exemples de commentaires des internautes.

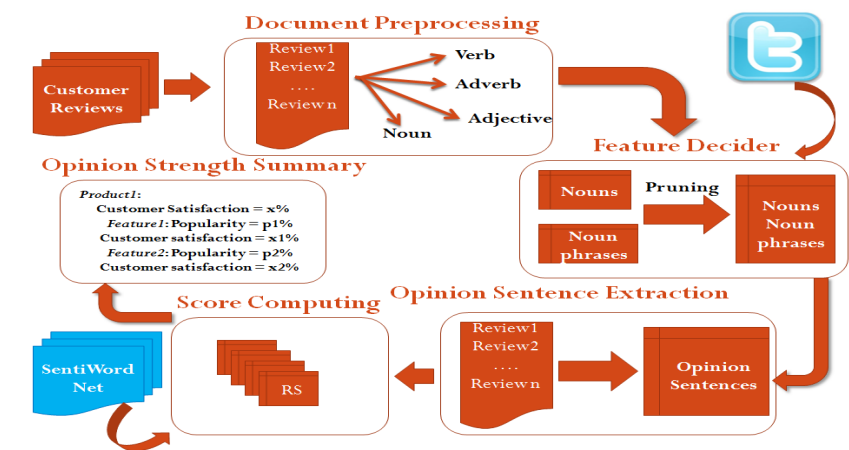


FIGURE 1 – Modèle proposé.

| |
|---|
| ## There isn't much features on the iPod at all, except games. |
| ##The Click Wheel is a great design, something no one else came up with (however, the iRiver has a touchpad). |

TABLE 1 – Exemples de critiques utilisateurs

Nous avons en entrée une base de données d'opinions recueillies à partir de 2 sites marchands (amazon.com et c|net.com) qui constitue notre corpus. Étant donné un nom de

produit, notre système ResTS choisit les documents correspondants dans la base de données et procède à leur segmentation en phrases. Ensuite, il les convertit en minuscule et supprime les caractères non littéraires du début et de la fin de chaque mot (par exemple « ##ipod## » devient « ipod »). Nous mettons également en relief la négation pour l'utiliser plus tard dans la phase de classification (par exemple « don't » ou « dont » devient « do not »). En outre, Hu et Liu (Hu et Liu, 2004) révèlent que les syntagmes nominaux et le substantif dans la phrase sont susceptibles d'être la caractéristique du produit sur laquelle les clients commentent. Par ailleurs, les adjectifs véhiculent l'opinion et le jugement. Nous avons donc effectué l'étiquetage de l'ensemble du corpus en utilisant TreeTagger³ pour identifier les classes grammaticales de chaque mot.

3.2 Extraction des caractéristiques des produits

Nous avons extrait tous les syntagmes nominaux (noms) à partir des critiques des utilisateurs. Ces noms seront considérés comme des caractéristiques des produits. Notons qu'une caractéristique peut être un nom simple ou un terme composé (exemple « picture quality »).

3.2.1 Construction des termes composés

Après avoir collecté les différents noms à partir des critiques des utilisateurs, nous avons procédé à la construction des termes composés qui sont formés de deux noms successifs. Prenons un exemple : « *The Click Wheel is a great design* ». « *Click Wheel* » est considéré comme un terme composé. Nous avons construit de la même manière tous les termes composés mais nous n'avons gardé que ceux qui apparaissent au moins 3 fois dans le corpus.

3.2.2 Caractéristiques fréquentes

Nous avons calculé la fréquence d'apparition des différents noms dans le corpus et nous n'avons gardé que ceux dont la fréquence est supérieure à 0,01. Le Tableau 2 présente quelques résultats.

| Caractéristiques | Nombre d'occurrence | Fréquence |
|------------------|---------------------|------------|
| Click wheel | 9 | 0.07853403 |
| Battery | 30 | 0.2617801 |

TABLE 2 – Exemples de caractéristiques fréquentes

La colonne 1 présente les caractéristiques. La colonne 2 donne le nombre d'occurrences de la fonction et la colonne 3 est la fréquence des occurrences de cette caractéristique dans le corpus.

3.2.3 Popularité dans Twitter

Twitter est le service de microblogage le plus populaire. Les gens peuvent publier et lire de courts messages de 140 caractères maximum appelés tweets⁴. Les textes d'opinion suivent un style particulier (texte libre ou dialecte). On parle de nos jours de Discours Electronique Médié (DEM) qui comporte des fautes d'orthographe, des émoticônes (des smileys), des

³ <http://www.ims.uni-stuttgart.de/projekte/corplex/TreeTagger/>

⁴ Exemple de tweet : « i need an ipod! i have a mill at my house but of course none of them work ☹ ».

acronymes (Exemple : lol), des étirements de mots, etc. Ce type d'écriture est peu étudié par la littérature. Twitter est devenu un domaine attractif pour le traitement automatique de la langue naturelle (NLP). Dans cet article, nous montrons comment les réseaux sociaux, en particulier Twitter, peuvent être utilisés pour détecter la popularité d'un produit donné. Pour ce faire, nous commençons par l'opération de crawling. Nous cherchons seulement les tweets *populaires* parlant d'un produit donné. Notre but étant de déceler les caractéristiques populaires que les gens en montre le plus d'intérêt pour un produit donné en comptant le nombre de personnes qui s'y intéressent. Nous avons utilisé twitter4j⁵, une librairie Java qui permet d'accéder au contenu de Twitter, pour recueillir près de 5000 tweets pour chaque produit posté au cours des derniers jours. Nous avons ensuite calculé le nombre de tweets évoquant chacune des caractéristiques. Le tableau 3 montre une comparaison entre le nombre d'occurrences de certaines caractéristiques dans le corpus et dans Twitter. Après avoir calculé le nombre d'occurrences de chaque caractéristique extraite dans Twitter, nous n'avons gardé que celles dont le nombre d'occurrences est supérieur à 1 ; celles qui sont mentionnées par au moins un tweet.

| Caractéristiques | Occurrences corpus | Occurrences Twitter |
|------------------|--------------------|---------------------|
| Player | 35 | 480 |
| Reputation | 3 | 0 |
| Storage space | 2 | 0 |

TABLE 3 – Nombre d'occurrences dans le corpus Vs nombre d'occurrences dans Twitter

3.3 Extraction des phrases d'opinion

L'un des objectifs de notre système est de détecter les passages subjectifs des commentaires des utilisateurs, de déterminer leur polarité et de mesurer la force de l'opinion exprimée. En utilisant la liste des caractéristiques déjà détectées, notre système ResTS a extrait toutes les phrases qui contiennent au moins une caractéristique. Voici un exemple: « *iPod is brilliant, but service was awful* ». Cette phrase présente deux caractéristiques qui sont « *iPod* » et « *service* ». Les mots d'opinion sont « *brillant* » et « *awful* ».

3.4 Calcul des scores

Dans cette section, nous expliquons comment nous avons procédé pour mesurer l'intensité de l'opinion pour chaque caractéristique, puis pour l'ensemble du produit. Rappelons que nos scores sont compris entre 0 et 1. Le score négatif appartient à l'intervalle [0, 0.5] et le score positif appartient à l'intervalle [0.5, 1]. Pour l'identification de l'intensité de l'opinion nous adoptons l'hypothèse suivante : plus le score est proche de 0, plus le mot est négatif, et vice-versa.

3.4.1 Score d'une caractéristique

Le score d'une caractéristique est sa fréquence d'apparition dans le corpus pondérée par sa popularité sur Twitter. Nous attribuons à chaque caractéristique un score en utilisant la

⁵ <http://twitter4j.org/en/index.html>

formule suivante⁶ :

$$score_f = \alpha freq_f + (1 - \alpha) \frac{nbreTwee\textit{tpf}}{nbreTwee\textit{tp}}$$

Avec : $freq_f$ est la fréquence d'apparition de la caractéristique dans le corpus, $nbreTwee\textit{tpf}$ est le nombre de tweets mentionnant à la fois le produit et la caractéristique, $nbreTwee\textit{tp}$ est le nombre total de tweets collectés pour le produit.

Ce poids mesure l'importance que les gens ont pour une caractéristique d'un produit donné. Il mesure également sa popularité. Prenons l'exemple de la caractéristique « battery », son score est égal à 0.3442 (0.6x0.543 + 0.4x0.046).

3.4.2 Opinion sur Twitter

La contrainte de taille des tweets encourage l'utilisation des émoticônes pour exprimer les opinions et les sentiments. Ces émoticônes résument souvent la polarité de toute la phrase. Nous avons construit notre propre liste d'émoticônes (voir exemples dans le tableau 4) et avons divisé les différents tweets collectés en des tweets positifs et d'autres négatifs selon la polarité de l'émoticône qu'ils contiennent.

| Polarité | Emoticône |
|---------------------|--------------------------|
| Positif | :-) :) :o) :] :3 :c) :^) |
| Extrêmement Positif | <=3 <=8 \o/ |
| Négatif | --!-- :-(:(:{ |
| Extrêmement Négatif | :-9 q(:^;)p |

TABLE 4 – Exemple d'émoticônes avec polarité

Nous avons compté par la suite le nombre de tweets positifs et négatifs pour chaque caractéristique. Notre hypothèse est qu'une caractéristique doit avoir un score élevé si elle appartient plus à des tweets positifs. Donc, si une caractéristique donnée apparaît plus dans des tweets positifs, on doit augmenter son score, sinon on doit le diminuer. Comme nos scores sont entre 0 et 1, nous avons choisi la racine carrée et le carré pour augmenter et diminuer le score des caractéristiques des produits comme le montre l'algorithme suivant.

Algorithm
Feature_Score

Input: $score_f$, nbtweetpos, nbtweetneg //nombre de tweets positifs et négatifs

Begin Feature_Score
If ($score_f \geq 05$) **then**
 If (nbtweetpos > nbtweetneg) **then**
 $score_{ft} = \sqrt{score_f}$

⁶ Pour les expérimentations $\alpha = 0.6$


```

Else


$$score_{ft} = score_f^2$$


EndIf
ElseIf (nbtweetpos < nbtweetneg) then


$$score_{ft} = \sqrt{score_f}$$


Else


$$score_{ft} = score_f^2$$


EndIf
End If
End Feature_Score

```

Output: $score_{ft}$

Prenons l'exemple de la caractéristique « battery », son score est égal à 0.3442. Comme elle apparaît plus dans des tweets négatifs, on doit diminuer son score. Le score devient 0.118.

3.4.3 Score des verbes et des adjectifs

Nous avons utilisé SentiWordNet 3.0 (Baccianella et al, 2006), une ressource lexicale basée sur WordNet 3.0, dans laquelle chaque mot w de WordNet est associé à trois scores numériques $ObjScore(w)$, $PosScore(w)$ et $NegScore(w)$ décrivant à quel point le mot w est objectif, positif ou négatif selon la formule suivante :

$$ObjScore(w) + PosScore(w) + NegScore(w) = 1$$

Par exemple, l'adjectif « great » a six synonymes (synset) et pour chacun un score positif et négatif. Nous ne traitons pas les verbes ou adjectifs objectifs ; ceux dont le score objectif est plus élevé que la somme de leurs scores positifs et négatifs. Étant donné un mot w , et n le nombre de ses synonymes, le score correspondant est calculé en utilisant la formule suivante:

$$score_{w_i} = \frac{\sum_{i=1}^n score_{S_{w_i}}}{n}$$

Avec : $score_{S_{w_i}}$ est le score de SentiwordNet du mot w et donné par l'algorithme suivant.

Algorithm Word_Score_Computing

Input: PosScore, NegScore //les scores de SentiWordNet

Begin Word_Score_Computing

$ObjScore = 1 - (NegScore + PosScore)$

```
If (PosScore + NegScore ≥ ObjScore) then //non abjectif
```

```
  If ( PosScore ≥ NegScore ) then
```

```
    scorewi = PosScore
```

```
  ElseIf ( NegScore ≤ 0.5 ) then
```

```
    scorewi = NegScore
```

```
  Else
```

```
    scorewi = 1-NegScore
```

```
  EndIf
```

```
End If
```

```
End Word_Score_Computing
```

```
Output: scorewi
```

Prenons un exemple : «*The iPod has one of the worst batteries.* ». La phrase d'opinion est «*worst batteries*». Le mot d'opinion est «*bad* ». Il a 14 synonymes dans SentiWordNet et son score calculé en utilisant l'algorithme énoncé ci-dessus est égale à 0.285.

3.4.4 Les adverbes

Les phrases d'opinions peuvent contenir des modificateurs : *intensifier* comme «*Absurdly* », «*Acutely* », «*Alarmingly* » ou *diminisher* comme «*Moderately* », «*Momentarily* », «*Improbably* » qui peut être utilisé de la même manière dans un contexte positif ou négatif comme «*Absolutely great* » ou «*Absolutely bad* ». Nous avons construit notre propre liste d'*intensifier* (192 termes) et *diminisher* (40 termes). Si une phrase contient un modificateur qui précède le verbe ou l'adjectif, nous calculons leurs scores à l'aide de l'algorithme suivant. S'il ya un intensifier précédant un verbe/adjectif positif (score ≥ 0.5), nous devons augmenter son score. Cependant, s'il s'agit d'un *diminisher*, nous devons diminuer le score. Dans le cas d'un verbe/adjectif négatif (score < 0.5), s'il est précédé par un *intensifier*, nous devons réduire son score, sinon, nous devons l'augmenter. Prenons un exemple : «*The battery is extremely bad.* ». Le score de «*bad* » est égal à 0.285. Comme «*Extremely*» est un intensifier et bad est négatif (score < 0.5), le score de «*extremely bad* » devient 0.081(= 0.285x0.285).

Algorithm Word_Score_Computing_Modifier

```
Input: scorew, IntensifierG, DiminisherG
```

Begin Word_Score_Computing_Modifier

```
If ( scorew ≥ 0.5 ) then // the word is positive
```

```
  If ( Modifier ∈ IntensifierG ) then
```

```
    scorew =  $\sqrt{\text{score}_w}$ 
```

```
  ElseIf ( Modifier ∈ DiminisherG ) then
```

| |
|---|
| $score_w = score_w^2$ |
| Endif |
| ElseIf (Modifier ∈ IntensifierG) then |
| $score_w = score_w^2$ |
| ElseIf (Modifier ∈ DiminisherG) then |
| $score_w = \sqrt{score_w}$ |
| Endif |
| End If |
| End Word_Score_Computing_Modifier |
| <hr/> Output: $score_w$ |

3.4.5 Score des phrases d'opinions

Le score des phrases d'opinions dépend en premier lieu des scores des verbes et des adjectifs qu'elles contiennent. Il dépend également du score de la caractéristique qu'elle contient. Si une phrase contient n caractéristiques, son score est donné par la formule suivante⁷ :

$$score_s = \frac{\sum_{i=1}^n \alpha \times score_{f_i} + (1 - \alpha) \times score_{w_i}}{n}$$

Reprenons l'exemple précédent : « *The battery is extremely bad.* ». Le score de « *battery* » est égal à 0.118. Le score de toute la phrase est : $0.3 \times 0.118 + 0.7 \times 0.081 = 0.092$. Prenons maintenant un autre exemple qui montre un score positif : « *The sound is pretty good.* ». Ici, la caractéristique est « *sound* ». Son score est 0.354. Elle apparaît plus dans des tweets positifs, donc son score devient 0.595. La phrase d'opinion est « *pretty good* ». L'adjectif « *good* » a 21 synonymes. Son score est 0.595. Comme « *Pretty* » est un *intensifier*, le score devient 0.771. Le score de la phrase devient 0.718 (= $0.3 \times 0.595 + 0.7 \times 0.771$).

3.4.6 Score du produit

Le score du produit est représenté par le score de tout le corpus relatif à ce produit. Il est donné par la formule suivante:

$$score_r = \frac{1}{n} \sum_{i=1}^n score_{fs_i}$$

Avec : $score_{fs}$ est le score d'une phrases d'opinions et n est le nombre de phrases

⁷ Pour les expérimentations $\alpha = 0.3$

d'opinions dans le corpus.

3.5 Expérimentations

L'approche proposée a été implémentée en langage Java sous l'environnement Eclipse. Nous avons évalué notre système en utilisant plusieurs corpus de critiques des utilisateurs sur les produits suivant : deux appareils photo numériques, un téléphone cellulaire et un iPod. Ces corpus ont été collectées à partir de 2 sites marchands (Amazon.com et C|net.com) et annotées manuellement⁸ par (Hu et Liu, 2004). Le premier objectif de notre système est d'extraire les caractéristiques des produits les plus proches de celles de l'annotation manuelle. Le tableau 4 résume la précision et le rappel de la phase de collecte des caractéristiques des produits. La colonne 1 présente la liste des produits utilisés pour l'évaluation. La colonne 2 donne la précision et le rappel du système de Hu et Liu. La troisième colonne indique la précision et le rappel de notre système. Nous constatons que nos résultats sont très proches de ceux de Hu et Liu ; le F-score moyen du système de Hu et Liu est 0,657, il est de 0,651 pour cette recherche.

| Produit | Hu et Liu | | Collecte | | Collecte (utilisant Twitter) | |
|-------------|-----------|--------|-------------|--------------|------------------------------|--------|
| | Précision | Rappel | Précision | Rappel | Précision | Rappel |
| iPod | -- | -- | 0.702 | 0.697 | 0.754 | 0.518 |
| A Photo1 | 0.634 | 0.658 | 0.617 | 0.679 | 0.743 | 0.55 |
| A Photo 2 | 0.679 | 0.594 | 0.69 | 0.58 | 0.727 | 0.508 |
| Téléphone C | 0.676 | 0.716 | 0.556 | 0.731 | 0.725 | 0.503 |
| Moyenne | 0.663 | 0.656 | 0.641 | 0.671 | 0.737 | 0.519 |

TABLE 4 – Précision et rappel de la méthode proposée Vs Hu et Liu

$$Precision = \frac{NCPR}{NC} \quad Rappel = \frac{NCPR}{NCP} \quad F - score = 2 * \frac{Precision * Rappel}{Precision + Rappel}$$

Avec : NC : Nombre de caractéristiques collectées par le système, NCPR : Nombre de caractéristiques pertinentes collectées par le système (qui correspondent à ceux de l'annotation manuelle), NCP : Nombre de caractéristiques de l'annotation manuelle.

L'utilisation de Twitter au cours de la phase de collecte des caractéristiques du produit a amélioré la précision, mais a causé une baisse du rappel. Ce déclin est dû à la suppression d'un certain nombre de caractéristiques qui ne sont pas populaires, c'est à dire qui n'intéressent pas la majorité des utilisateurs de Twitter.

Le deuxième objectif du système est de résumer l'opinion des utilisateurs envers un produit donné. Pour ce faire, nous avons extrait les phrases d'opinions puis calculé leurs scores. Ces scores sont corrélés à **82%** avec ceux de l'annotation manuelle.

⁸ Exemple d'une phrase annotée par Hu et Liu : "battery[-2]##This is really stupid to me. 18 months for a battery isn't good," "Battery" est la caractéristique et "-2" est le score de la phrase.

3.6 Conclusion

Cet article présente une nouvelle approche de résumé automatique des textes d'opinions des critiques des utilisateurs. Notre approche vise à transformer les critiques des consommateurs en un score qui mesure l'intensité de l'opinion. Ce score est compris entre 0 et 1 et peut être utilisé pour la prise de décision et aide les utilisateurs dans leurs choix. Dans notre conception, un produit n'est pas simplement considéré comme recommandé ou non recommandé, au contraire, nous laissons l'utilisateur libre de faire son choix en fonction de certains scores que nous mettons à sa disposition traduisant la satisfaction des clients pour l'ensemble du produit et encore pour chacune de ses caractéristiques. Lors du calcul de ces scores, nous avons étudié l'opinion véhiculée par les noms, adjectifs, verbes et adverbes, contrairement aux autres recherches qui utilisent principalement les adjectifs. Nous avons de plus montré que les réseaux sociaux tel que Twitter peuvent être exploités pour mettre en évidence les caractéristiques les plus pertinentes pour l'utilisateur et de détecter leurs popularités. Dans les travaux futurs, nous prévoyons améliorer nos résultats (augmenter le rappel), éventuellement en exploitant les passages négatifs et ironiques et d'expérimenter notre méthode à l'aide d'autres entités, non seulement les produits.

Références

- ANDREEVSKAIA, A. AND BERGLER, S. (2006). Mining WordNet for fuzzy sentiment: Sentiment tag extraction from WordNet glosses. *In Proceedings of EACL 2006*.
- BLITZER, J., DREDZE, M., AND PEREIRA, F. (2007). Biographies, Bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification. *In Proceedings of ACL 2007*.
- BACCIANELLA S., ESULI A., SEBASTIANI F. (2010). SentiWordNet 3.0 : An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining. *In Proceedings of LREC'10*.
- BOUCHLEGHEM R., ELKHLIFI A., AND FAIZ R. (2010). Automatic extraction and classification approach of opinions in texts. *ISDA 2010*, IEEE Press, 918-922.
- DING, X., LIU, B., AND YU, P.S. (2008). A Holistic Lexicon-Based Approach to Opinion Mining. *In Proceedings of WSDM*, Stanford University, Stanford, California, USA.
- DRAGUT, E. C., YU, C., SISTLA, P., AND MENG, W. (2010). Construction of a sentimental word dictionary. *In Proceedings of CIKM*.
- ESULI A., AND SEBASTIANI, F. (2005). Determining the Semantic Orientation of Terms through Gloss Classification. *In Proceedings of CIKM*.
- GAMON, M., AUE, A., CORSTON-OLIVER, S., RINGGER, E. (2005). Pulse: Mining Customer Opinions from Free Text. *In Proc. 6th Int. Symp. Advances in intelligent data analysis*, 121-132.
- HU, M., LIU, B. (2004). Mining and Summarizing Customer Reviews. *In Proc. 10th Int. Conf. Knowledge Discovery and Data Mining*, Seattle, WA, 168-177.
- HARRIS, Z. S. (1998). Mathematical structures of language. *Interscience tracts in pure and applied mathematics*, no.21, New York: Interscience Publishers. ix,230 p.
- HATZIVASSILOPOULOS, V., AND MCKEOWN, K. (1997). Predicting the Semantic Orientation of Adjectives. *In Proceedings of ACL 1997*.

- KANAYAMA, K., NASUKAWA, T. (2006). Fully Automatic Lexicon Expansion for Domain-Oriented Sentiment Analysis. *In Proceedings of EMNLP 2006*.
- KAMPS, J., MARX, M., ROBERT J. M., AND RIJKE, M. (2004). Using WordNet to measure semantic orientation of adjectives. *In Proceedings of LREC 2004*.
- KIM, S.M., AND HOVY, E. (2004). Determining the Sentiment of Opinions. *In Proceedings of COLING 2004*.
- LEHUEDE, F. (2009). L'internet participatif redonne confiance aux consommateurs.
- LIU, B., HU, M., AND CHENG, J. (2005). Opinion observer: Analyzing and comparing opinions on the web. *In Proceedings of WWW 2005*.
- LIU, B. (2007). Web Data Mining Exploring Hyperlinks, Contents, and Usage Data, *Springer 2007*, New York.
- LIU, B. (2010). Invited Chapter for the *Handbook of Natural Language Processing*, Second Edition. March, 2010.
- MIHALCEA, R., CORLEY, C., AND STRAPPARAVA, C. (2006). Corpus-based and knowledgebased measures of text semantic similarity. *In Proceedings of the 21st national conference on Artificial intelligence - Volume 1*, pages 775—780, AAAI Press.
- PANG, B., LEE, L., VAITHYANATHAN, S. (2002). Thumbs up? Sentiment Classification Using Machine Learning Techniques. *In Proc. Conf. Empirical Methods in Natural Language Processing*, 79-86.
- PEDERSEN, T., AND PATWARDHAN, S. AND MICHELIZZI, J. (2004). WordNet::Similarity: measuring the relatedness of concepts. *Association for Computational Linguistics*, 2004.
- POPESCU, A. M., ETZIONI, O. (2005). Extracting Product Features and Opinions from Reviews. *In Proc. Conf. Human Language Technology and Empirical Methods in Natural Language Processing*, Vancouver, British Columbia, 339–346.
- QIU, G., LIU, B., BU, J. AND CHEN, C. (2009). Expanding Domain Sentiment Lexicon through Double Propagation. *In Proceedings of IJCAI 2009*.
- RILLOFF, E., JANYCE, W., THERESA, W. (2003). Learning Subjective Nouns Using Extraction Pattern Bootstrapping. *In Proc. 7th Conf. Natural Language Learning*, 25-32.
- TAKAMURA, H., INUI, T., AND OKUMURA, M. (2007). Extracting Semantic Orientations of Phrases from Dictionary. *In Proceedings of HLT-NAACL*.
- TURNER, P. (2001). Mining the Web for Synonyms: PMI-IR versus LSA on TOEFL". *Machine Learning: ECML 2001*, pages 491–502.
- WIEBE, J. (2000). Learning Subjective Adjectives from Corpora. *In Proceedings of AAAI 2000*.
- WILSON, T., WIEBE, J., HWA, R. (2004). Just how mad are you? Finding strong and weak opinion clauses. *In Proceedings of AAAI 2004*.