

Extraction d'opinions ambiguës dans des corpus d'avis clients

Joseph Lark^{1,2} Emmanuel Morin¹ Sebastián Peña Saldarriaga²

(1) LINA, Université de Nantes, Nantes - France

(2) Dictanova, 2 chemin de la Houssinière, Nantes - France

joseph.lark@univ-nantes.fr, emmanuel.morin@univ-nantes.fr,
sebastian@dictanova.com

RÉSUMÉ

Nous détectons dans des corpus d'avis clients en français des expressions d'opinion ne contenant pas de marqueur d'opinion explicitement positif ou négatif. Nous procédons pour cela en deux étapes en nous appuyant sur des méthodes existantes : nous identifions ces expressions à l'aide de fenêtres de mots puis nous les classifions en polarité. Le processus global présente des résultats satisfaisants pour notre cadre applicatif demandant une haute précision.

ABSTRACT

Ambiguous opinion extraction in user feedbacks

We experiment in French user feedbacks the detection of opinion expressions whose opinion words are not explicitly positive nor negative. Following existing work, we conduct a two-step method : first we identify opinion expressions within word windows, then we infer their polarity. The global process shows satisfactory results for our purpose which requires a high precision.

MOTS-CLÉS : Fouille d'opinion, lexique affectif, avis client, marqueur d'opinion ambigu.

KEYWORDS: Opinion mining, affective lexicon, user feedback, ambiguous opinion word.

1 Introduction

Ce travail s'inscrit dans le domaine de la fouille d'opinion en français, et plus particulièrement dans le cadre de l'analyse des avis clients. Cette analyse est devenue depuis quelques années un point stratégique clé pour de nombreuses entreprises souhaitant mieux comprendre les besoins de leur marché, c'est pourquoi celles-ci font usage d'outils leur permettant de structurer et d'explorer les commentaires écrits spontanément par des consommateurs ou bien lors d'une enquête de satisfaction.

De nombreux travaux en fouille d'opinion ont été développés dans le but de traiter ce type de données (*cf.* section 2), notamment pour la tâche consistant à associer chaque sujet discuté à un ou plusieurs mots portant une polarité (positive ou négative), appelés marqueurs d'opinion. Cette tâche reste cependant complexe lorsque la polarité du marqueur d'opinion est ambiguë car dépendante du contexte (la notion de « *surprise* », par exemple, est positive dans le domaine du divertissement, mais négative dans le domaine bancaire), ou bien lorsque l'opinion ne contient pas de marqueur d'opinion explicite (la proposition « *le conseiller n'a pas répondu à ma demande* » est clairement négative mais ne contient pourtant aucun marqueur explicite de subjectivité).

Dans le cadre de l'amélioration d'un outil professionnel d'analyse des avis clients, nous cherchons ici à identifier ces passages de texte relatifs à une opinion considérés comme ambigus. Les résultats

de l'analyse étant directement visibles par les utilisateurs de l'outil, nous cherchons à maximiser la précision de la détection de ces passages. Nous proposons dans cette optique une méthode en deux étapes reposant sur l'identification de contextes stricts, inspirée de travaux existants (*cf.* section 3). Cette méthode est évaluée sur des corpus d'avis clients (*cf.* section 4), et montre des résultats prometteurs.

2 Travaux connexes

Notre travail se situe à la croisée de deux champs de recherche relativement prolifiques, que sont l'extraction de lexique affectif et la fouille d'opinion à l'échelle des arguments (ou ciblée).

2.1 Extraction de lexique affectif

La majorité des méthodes de fouille d'opinion utilisent un lexique affectif, autrement dit un dictionnaire associant à un marqueur, représenté par un mot simple ou une expression polylexicale, une polarité *a priori*. Parmi les lexiques affectifs les plus utilisés en anglais nous trouvons le lexique de subjectivité MPQA¹ (*Multi-Perspective Question Answering*) ainsi que le lexique SentiWordNet². Si aucun équivalent français n'est encore disponible, nous relevons plusieurs travaux de construction de lexique affectif en français (Mathieu, 1999; Pak & Paroubek, 2010; Vincze & Bestgen, 2011), ainsi que des lexiques tels que Emolex³ disponibles pour les travaux de recherche. Dans la mesure où il n'existe pas de lexique libre adapté spécifiquement à l'analyse des avis clients, nous utilisons ici un lexique affectif propriétaire mis à notre disposition (*cf.* section 3.2).

En ce qui concerne l'extraction de marqueurs d'opinion ambigus et spécifiques à un corpus, qui nous intéresse plus particulièrement ici, nous notons les travaux de Riloff & Wiebe (2003) et Jijkoun *et al.* (2010) dont nous nous sommes inspirés pour la méthode que nous présentons. Le premier de ces travaux présente des résultats en extraction de patrons subjectifs, à l'aide d'une méthode définie précédemment (Thelen & Riloff, 2002) s'appuyant uniquement sur les fenêtres de mots autour d'expressions subjectives du même corpus. Le second montre qu'il est possible d'acquérir un lexique affectif spécifique à un corpus en sélectionnant des expressions syntaxiquement cohérentes contenant un sujet discuté et un marqueur d'opinion.

2.2 Fouille d'opinion ciblée

Les travaux que nous considérons comme précurseurs pour ce type d'analyse (en anglais) sont ceux de Hu & Liu (2004) et de Wilson *et al.* (2005). Dans le premier, les auteurs recherchent les adjectifs cooccurrents des sujets abordés parmi des avis clients, tandis que le second présente une méthode de classification en polarité de segments de phrases. Les travaux plus récents en fouille d'opinion ciblée font usage de l'analyse en dépendances syntaxiques dans le but d'identifier les relations entre un sujet discuté et un marqueur d'opinion. Dans cette voie, Qiu *et al.* (2009) montrent qu'une sélection manuelle de règles d'extraction reposant sur l'analyse en dépendances permet d'atteindre des performances intéressantes à la fois en extraction d'opinion et en inférence de polarité. Jakob & Gurevych (2010) exploitent les dépendances syntaxiques comme traits d'un modèle markovien CRF (*Conditional Random Field*) dans le même but. Enfin, Liu *et al.* (2015) proposent une méthode syntaxique permettant de s'affranchir de la sélection manuelle de règles, et conservant des résultats intéressants. Un état de l'art détaillé des méthodes de fouille d'opinion ciblée peut être observé au

1. http://mpqa.cs.pitt.edu/lexicons/subj_lexicon/

2. <http://sentiwordnet.isti.cnr.it/>

3. <http://emolex.u-grenoble3.fr/emoBase/>

travers des campagnes d'évaluation SEMEVAL 2014 à 2016 (Pontiki *et al.*, 2014, 2015, 2016). Les systèmes montrant les meilleurs résultats en identification des cibles d'opinion (San Vicente *et al.*, 2015; Toh & Su, 2015) conservent cependant un taux d'erreur en précision de l'ordre de 30%, ce qui n'est pas satisfaisant dans le cadre de notre outil professionnel. Par ailleurs, l'analyse en dépendances¹ sur nos corpus se révèle relativement bruitée, aussi nous optons pour une représentation plus robuste des opinions sous forme de règles syntaxiques strictes (des éléments optionnels sont insérables, mais ceux-ci sont restreints à des emplacements et des catégories grammaticales définis)(*cf.* section 3.2). Celles-ci s'approchent des patrons utilisés par Riloff & Wiebe (2003) cependant dans ces derniers le contexte est lexical et non syntaxique. Cette modélisation rend la méthode peu dépendante de la langue, ce qui permet d'envisager l'analyse de langues pour lesquelles l'analyse en dépendances est peu développée.

3 Méthode proposée

À notre sens la détection des expressions d'opinion et l'inférence en polarité de ces expressions sont deux tâches distinctes. La méthode proposée effectue en suivant ce principe une extraction d'opinions en deux étapes : il s'agit premièrement de détecter les segments de textes relatifs à une opinion à partir de fenêtres de mots, puis d'inférer leur polarité par classification.

3.1 Éléments recherchés

Les expressions d'opinion ambiguës recherchées peuvent se présenter sous plusieurs formes, toutefois nous observons trois catégories d'expressions récurrentes, que nous souhaitons pouvoir détecter avec précision : il s'agit des groupes nominaux contenant un adjectif quantitatif (« *file d'attente immense* », « *prix bas* »), ceux contenant un adjectif ambigu car polysémique (« *un repas simple* », « *une soirée étonnante* ») et enfin des propositions négatives ne contenant pas de marqueur explicite d'opinion (« *je n'ai pas reçu de réponse* », « *le compte n'a pas été mis à jour* »).

3.2 Ressources à disposition

Ce travail est mené dans le cadre du développement des outils de l'entreprise Dictanova, qui a mis à notre disposition deux ressources propriétaires constituées semi-automatiquement. La première est un lexique affectif contenant 2 000 marqueurs d'opinions stables de différentes catégories grammaticales, tels que « *aimable* », « *joie* » ou « *détester* ». La deuxième est un ensemble de règles syntaxiques (environ 100) décrivant des séquences d'étiquettes grammaticales propres à l'expression d'une opinion. Pour chaque règle de cette ressource, nous disposons de trois informations cruciales : la place du marqueur d'opinion, la place du sujet, et un indicateur de la présence d'une expression de négation inversant la polarité du marqueur d'opinion. Une règle très simple pourrait être de la forme suivante : <NOM_SUJ> <VERBE> <NEGATION> <ADJECTIF_MO>, où SUJ indique la place du sujet et MO indique la place du marqueur d'opinion. Par exemple, le segment en gras dans la phrase « *Dans ce restaurant les serveurs sont pas aimables.* » respecte cette règle et représente donc une opinion. La présence dans cette règle d'un élément <NEGATION> permet d'attribuer au segment la polarité inverse de celle du marqueur d'opinion « *aimable* ».

3.3 Détection des expressions d'opinions

La première étape consiste à identifier dans le corpus les segments de phrase contenant une opinion.

1. Outils testés : http://alpage.inria.fr/statgram/frdep/fr_stat_dep_parsing.html

Projection des règles syntaxiques À partir des règles syntaxiques fixes à notre disposition, nous recherchons les segments de phrase correspondants. Lorsqu'un segment est reconnu par une règle syntaxique, deux cas de figure sont possibles. Dans le cas où l'emplacement du marqueur d'opinion correspond à un élément du lexique affectif (comme décrit précédemment), le segment de phrase représente une opinion, que nous désignons comme « connue ». Dans le cas contraire, le segment de phrase est désigné comme « segment candidat », et celui-ci est conservé en tant que potentielle expression d'une opinion dans une base de connaissances à part. Nous disposons donc à la suite de cette projection d'une liste d'opinions connues (par la suite O , pour opinions) dont la polarité est issue du lexique affectif, et d'une liste de segments candidats linguistiquement cohérents (par la suite C , pour candidats) mais dont nous ignorons s'il s'agit ou non d'expressions d'opinions.

Sélection des segments candidats Tous les segments candidats que nous avons extraits ne sont pas intéressants du point de vue de la fouille d'opinion. Nous proposons de leur associer un score, en s'inspirant de l'algorithme BASILISK (Thelen & Riloff, 2002), pour retenir les plus pertinents. Nous justifions ce choix par le fait nous avons pu obtenir des résultats encourageants lors d'expériences précédentes grâce à cette méthode. En particulier, le caractère restrictif de l'algorithme vis-à-vis des segments candidats semble adapté à l'analyse de texte bruité.

L'attribution du score se déroule selon les étapes suivantes :

1. Pour chacune des occurrences dans le corpus des opinions de O ou des segments candidats de C , nous retenons les lemmes des 2 mots précédents et suivants, dans la limite de la phrase.
2. Un premier score S_1 est attribué aux fenêtres ainsi extraites, selon la formule (1).
3. Seules les fenêtres dont le score S_1 dépasse un certain seuil sont conservées (par la suite, F). Nous fixons empiriquement ce seuil au quart supérieur du classement.
4. Les segments candidats de C dont une occurrence au moins est entourée par une fenêtre de F sont associés à un second score S_2 , selon la formule (2).
5. Enfin, seuls les segments candidats de C dont le score S_2 dépasse un certain seuil sont considérés comme des expressions d'opinion. Ici encore nous fixons ce seuil au quart supérieur du classement.

$$S_1(f) = \frac{c_f}{t_f} \times \log_2(c_f) \quad (1)$$

$$S_2(o) = \frac{\sum_{i \in N_o} \log_2(1 + c_i)}{N_o} \times \frac{r_c}{r_t} \quad (2)$$

où :

- f est une fenêtre de mots
- c_f est le nombre d'occurrences de f autour d'une opinion connue
- t_f le nombre total d'occurrences de f

où :

- o est une expression d'opinion
- N_o est le nombre de fenêtres contenant o
- c_i est le nombre d'occurrences de la i -ème fenêtre autour d'une opinion connue
- r_c est le nombre de fois où la règle ayant extrait o a extrait une opinion connue
- r_t est le nombre total de fois où la règle ayant extrait o a extrait un segment

3.4 Inférence de polarité

À l'issue de cette première étape, nous obtenons un ensemble de segments candidats sélectionnés, permettant d'identifier les passages de texte où un utilisateur exprime une opinion. Cependant ces segments ne sont que de peu d'utilité tant que nous n'en connaissons pas la polarité.

Traits choisis Les traits que nous avons expérimentés proviennent des méthodes de la littérature présentant des résultats intéressants pour cette tâche et d’une observation de corpus d’avis de domaines spécifiques, définis par des secteurs d’activité. Après sélection, nous utilisons les traits suivants :

- Les trigrammes de mots de la phrase.
- Si parmi les mots compris entre le segment candidat et l’opinion connue précédente (idem suivante) se trouvent des mots préalablement vus entre deux opinions connues. Ces mots sont par conséquent des marqueurs d’une transition de polarité parmi quatre types possibles¹. Le type de transition et la position de l’opinion connue sont précisés pour ce trait.
- Si parmi les mots compris entre le segment candidat et l’opinion connue précédente (idem suivante) se trouve un élément répertorié comme inverseur de polarité, tel que « *mais* », « *en revanche* »... La polarité et la position de l’opinion connue sont alors précisées pour ce trait.

Nous choisissons un modèle de classification SVM (*Support Vector Machine*), en utilisant l’implémentation de la librairie LIBLINEAR².

4 Expériences et résultats

Nous décrivons à présent les expériences que nous avons menées en détection d’opinion puis en inférence de polarité.

4.1 Corpus analysés

Les corpus que nous utilisons pour notre évaluation sont relativement homogènes en taille (de 2 000 à 6 500 documents) et, autant qu’il est possible de le mesurer, homogènes en qualité d’écriture (phrases longues et construites, pas d’écriture abrégée, cependant orthographe très variable, ponctuation parfois non respectée). Ces corpus traitent chacun d’un domaine spécifique différent :

- C1 : Corpus de 2 043 avis spontanés sur plusieurs services de banques en ligne
- C2 : Corpus de 5 217 avis spontanés sur plusieurs restaurants de « hamburgers gourmets »
- C3 : Corpus de 6 536 commentaires suite à l’achat de chaussures sur deux sites d’e-commerce
- C4 : Corpus de 4 994 commentaires suite à la visite de deux parcs d’attractions français

4.2 Constitution des références pour l’évaluation

Afin d’évaluer l’étape de détection des opinions, nous annotons manuellement les segments candidats sélectionnés pour chaque corpus. Un segment est considéré correct si il respecte trois critères :

- le texte délimité par le segment désigne une expression linguistiquement cohérente ;
- le segment fait explicitement référence à un sujet qui est la cible d’une opinion (pas de référence anaphorique) ;
- la polarité de l’opinion exprimée (segment de phrase complet couvrant la cible ainsi que le marqueur d’opinion) est explicite dans le contexte du corpus de l’occurrence.

4.3 Résultats

Notre premier constat est que le nombre de segments candidats sélectionnés lors de la première étape est relativement faible. Les résultats présentés en table 1 indiquent en effet que nous avons sélectionné environ un segment candidat pour cent extraits. Cependant, notre objectif premier étant de détecter

1. Transitions positif → positif, positif → négatif, négatif → positif et négatif → négatif

2. <http://liblinear.bwaldvogel.de/>

des opinions avec précision, nous nous intéressons davantage à cette mesure, qui est très satisfaisante sur chacun des corpus.

Corpus	# Opinions connues	# Segments candidats	# Segments sélectionnés	Précision opinion (%)	Rappel opinion (%)	Précision polarité (%)
C1	2 326	13 197	96	97,91	39,34	81,25
C2	7 481	22 709	269	98,14	43,59	74,34
C3	5 792	12 513	260	98,84	57,77	97,69
C4	12 812	27 908	237	97,04	37,32	72,57

TABLE 1 – Évaluation de la détection des expressions d’opinion. Le rappel indiqué est calculé sur la base des segments extraits lors du premier score (annotés), et non sur tous les segments candidats (non annotés). Le rappel sur l’ensemble du corpus est donc en théorie plus faible.

Parmi les erreurs de détection observées nous constatons deux causes récurrentes, à savoir les expressions descriptives dont la polarité peut être nulle (« *restaurant est plein* »), et l’emploi du conditionnel (« *Le retour est gratuit si le produit acheté ne vous convient pas !* »).

En sus des catégories d’éléments recherchés, nous pouvons observer au travers des exemples présentés en table 2 que cette méthode permet d’extraire des marqueurs d’opinion incorrectement orthographiés (« *décors sont splendides* », « *personel est déplorable* ») ainsi que des marqueurs d’opinion explicites spécifiques au domaine du corpus (« *portions sont copieuses* », « *visuel est soigné* »).

Corpus	Segments extraits sélectionnés
C1	<i>conseillers sont très serviables, conseillers sont peu réactifs, fonctionnalités online sont très complètes, pas eu de retard dans le traitement, pas eu de frais, pas eu de chance, jamais eu de surprise, carte bleue très esthétique, carte bancaire sont gratuites, interface est claire</i>
C2	<i>burgers vraiment pas terribles, créneau horaire est large, rapport qualité prix est évident, offre est variée, produits étaient frais, manager étaient débordés, temps d’attente est faible, personel est déplorable, frites restent moyennes, portions sont copieuses</i>
C3	<i>formalités ont été simples, retours sont assez simples, peintures sont souvent disponibles, qualité des articles est relativement bonne, prix sont très variables, offres promotionnelles sont très fréquentes, prix était plus bas, retour est tellement simple, livraison est aussi gratuite, emballage est très propres</i>
C4	<i>glace artisanale plutôt bonne, peintures sont refaites, queues sont immenses, visuel est soigné, entrée au parc est chère, toilettes sont nombreux, mise en scène est incroyable, personnages sont absents, décors sont splendides, restos sont fermés</i>

TABLE 2 – Exemples de segments extraits sélectionnés (10 premiers selon le score S_2).

Des exemples de marqueurs d’opinion dont la polarité a été inférée sont présentés en table 3. Parmi les erreurs de classification nous retrouvons des marqueurs d’opinion ambigus (« *moyen* », « *géant* ») ou mal orthographiés (« *decevante* », « *sympatique* »). La mauvaise inférence des premiers peut être due aux cas d’ambiguïtés intra-corpus, c’est-à-dire les marqueurs d’opinion pouvant être positifs ou négatifs dans le même corpus (« *un burger géant* », « *une file d’attente géante* »). Les marqueurs d’opinion incorrectement orthographiés sont pénalisés par leur fréquence, dans la mesure où il s’agit pour la plupart d’hapax.

Corpus	Marqueurs d'opinion positifs	Marqueurs d'opinion négatifs
C1	<i>complet, esthétique, fonctionnel, impeccable, inexistant, réactif</i>	<i>frais (nom), long, retard dans le traitement, sympatique</i>
C2	<i>copieux, croustillant, fondant, frais (adj.), original, moyen, sympathique, terrible</i>	<i>approximatif, cher, géant, gras, immangeable, mou, tiède</i>
C3	<i>conforme, dispo, explicite, impressionnant, nickel, simplissime</i>	<i>élevé, long, recommandable</i>
C4	<i>convivable, decevante, enchanteur, impressionnant, incroyable, nouveau, propre</i>	<i>baclé, bluffant, défraîchi, inconfortable, minime, payant</i>

TABLE 3 – Exemples de marqueurs d'opinion classés en polarité (exemples d'erreurs en gras)

5 Conclusion

Nous détectons dans des corpus d'avis clients en français des expressions d'opinion ne contenant pas de marqueur d'opinion explicitement positif ou négatif. Nous procédons pour cela en deux étapes : nous identifions ces expressions à l'aide de fenêtres de mots puis nous les classifions en polarité. Le processus global présente des résultats satisfaisants pour notre cadre applicatif demandant une haute précision. La précision élevée de notre extraction est obtenue au prix d'un rappel très faible. Cependant, l'état actuel de nos travaux laisse penser qu'une expansion significative des expressions détectées est possible en s'appuyant sur les premières extraites. En effet nous avons évalué notre méthode sur les occurrences d'expressions d'opinion et non sur les occurrences des marqueurs détectés, ce qui pourrait permettre d'augmenter le nombre d'expressions ambiguës prises en compte.

Références

- HU M. & LIU B. (2004). Mining and summarizing customer reviews. In *Proceedings of KDD 2004*, p. 168–177, Seattle, WA, USA.
- JAKOB N. & GUREVYCH I. (2010). Extracting Opinion Targets in a Single- and Cross-Domain Setting with Conditional Random Fields. In *Proceedings of EMNLP 2010*, p. 1035–1045, Cambridge, MA, USA.
- JIKKOUN V., DE RIJKE M. & WEERKAMP W. (2010). Generating focused topic-specific sentiment lexicons. In *Proceedings of ACL 2010*, p. 585–594, Uppsala, Sweden.
- LIU Q., GAO Z., LIU B. & ZHANG Y. (2015). Automated Rule Selection for Aspect Extraction in Opinion Mining. In *Proceedings of IJCAI 2015*, p. 1291–1297, Buenos Aires, Argentina.
- MATHIEU Y. Y. (1999). Les prédicats de sentiment. *Langages*, p. 41–52.
- PAK A. & PAROUBEK P. (2010). Construction d'un lexique affectif pour le français à partir de twitter. In *Actes de TALN 2010*, Montréal, Canada.
- PONTIKI M., GALANIS D., PAPAGEORGIOU H., ANDROUTSOPOULOS I., MANANDHAR S., AL-SMADI M., AL-AYYOUB M., ZHAO Y., QIN B., CLERCQ O. D., HOSTE V., APIDIANAKI M., TANNIER X., LOUKACHEVITCH N., KOTELNIKOV E., BEL N., JIMÉNEZ-ZAFRA S. M. & ERYİĞİT G. (2016). SemEval-2016 task 5 : Aspect based sentiment analysis. In *Proceedings of SemEval 2016*, San Diego, CA, USA.
- PONTIKI M., GALANIS D., PAPAGEORGIOU H., MANANDHAR S. & ANDROUTSOPOULOS I. (2015). Semeval-2015 task 12 : Aspect based sentiment analysis. In *Proceedings of SemEval 2015*, p. 486–495, Denver, CO, USA.

- PONTIKI M., GALANIS D., PAVLOPOULOS J., PAPAGEORGIU H., ANDROUTSOPOULOS I. & MANANDHAR S. (2014). Semeval-2014 task 4 : Aspect based sentiment analysis. In *Proceedings of SemEval 2014*, Dublin, Ireland.
- QIU G., LIU B., BU J. & CHEN C. (2009). Expanding domain sentiment lexicon through double propagation. In *Proceedings of IJCAI 2009*, p. 1199–1204, Pasadena, CA, USA.
- RILOFF E. & WIEBE J. (2003). Learning extraction patterns for subjective expressions. In *Proceedings of EMNLP 2003*, p. 105–112, Stroudsburg, PA, USA.
- SAN VICENTE I., SARALEGI X., AGERRI R. & SEBASTIÁN D.-S. (2015). Elixia : A modular and flexible absa platform. In *Proceedings of SemEval 2015*, p. 748, Denver, CO, USA.
- THELEN M. & RILOFF E. (2002). A Bootstrapping Method for Learning Semantic Lexicons using Extraction Pattern Contexts. In *Proceedings of EMNLP 2002*, p. 214–221, Philadelphia, PA, USA.
- TOH Z. & SU J. (2015). Nlangp : Supervised machine learning system for aspect category classification and opinion target extraction. In *Proceedings of SemEval 2015*, Denver, CO, USA.
- VINCZE N. & BESTGEN Y. (2011). Identification de mots germes pour la construction d'un lexique de valence au moyen d'une procédure supervisée. In *Actes de TALN 2011*, p. 223–234, Montpellier, France.
- WILSON T., WIEBE J. & HOFFMANN P. (2005). Recognizing contextual polarity in phrase-level sentiment analysis. In *Proceedings of HLT 2005*, p. 347–354, Vancouver, Canada.