

Que ressentent les patients ?

Soumia Melzi*, Amine Abdaoui *, Jérôme Azé *,
Sandra Bringay **, Pascal Poncelet *, Florence Galtier***

*LIRMM UM2 CNRS, UMR 5506 - CC 477, 161 rue Ada, 34095 Montpellier

**MIAP UM3, Université Paul-Valéry, Route de Mende, 34199 Montpellier

***CIC, CHU Saint Eloi, 80 avenue Augustin Fliche 34295 Montpellier Cedex 5

Résumé. Les forums de santé en ligne sont des espaces d'échanges où les patients partagent leurs sentiments à propos de leurs maladies, traitements, etc. Sous couvert d'anonymat, ils expriment très librement leurs expériences personnelles. Ces forums sont donc une source d'informations très utile pour les professionnels de santé afin de mieux identifier et comprendre les problèmes, les comportements et les sentiments de leurs patients. Dans cet article, nous proposons d'exploiter les messages des forums via des techniques de fouille de textes pour extraire des traces d'émotions (*e.g. joie, colère, surprise*, etc.).

1 Introduction

Les forums de santé en ligne sont des espaces d'échanges où les patients, sous couvert d'anonymat, relatent très librement leurs expériences personnelles. Ces ressources s'avèrent très riches pour les professionnels de santé qui ont accès à des échanges entre patients, entre patients et professionnels et même entre professionnels. Même si tous les patients ne s'expriment pas dans les forums de santé, ces derniers représentent une base volumineuse et variée des connaissances et des perceptions qu'ont les patients de leur maladie et des soins qui leur sont éventuellement prodigués. Dans le cadre du projet *Parlons de nous*¹, nous cherchons à associer différents marqueurs (émotions, risques, incertitudes, etc.) à des objets médicaux (médicaments, traitements, etc.) pour identifier des cooccurrences fréquentes (*e.g. une association de type Médiateur et peur*). Dans cet article, nous nous focalisons sur l'identification des émotions. Si de nombreuses approches ont été proposées pour l'analyse de la polarité des textes (*positif* et *négatif*), on trouve peu d'approches pour l'analyse des sentiments (*joie, colère, tristesse*, etc.). Nous avons utilisé le lexique des mots d'émotions de (Mohammad et Turney, 2010) pour annoter automatiquement un corpus de messages. Une sous partie de ce corpus a été annotée manuellement. L'étude de l'accord entre ces annotateurs nous a permis de montrer qu'il était difficile, même pour des humains, d'associer une émotion précise à un message. Nous avons donc décidé de donner deux informations aux professionnels de santé : la polarité du texte (*positive* ou *negative*) et les émotions associées. Pour obtenir ces deux informations, nous avons travaillé sur la recherche des meilleurs descripteurs. Des expérimentations sur des jeux de données réelles ont montré l'efficacité de cette approche et des discussions avec les professionnels de santé ont montré l'intérêt médical d'identifier de telles informations.

1. Financé par la MSH-M <http://www.msh-m.fr/programmes/programmes-2013/parlons-de-nous/>

2 État de l'art

Depuis le début des années 2000, l'analyse de sentiments, également appelée fouille d'opinions (*opinion mining*), a connu un intérêt croissant. Beaucoup de communautés se sont intéressées à ce domaine et ont donné des définitions et interprétations variées (e.g. psychologie, sciences sociales, linguistique computationnelle, traitement automatique du langage, fouille de données, etc.). L'analyse de sentiments vise l'extraction des états affectifs exprimés explicitement ou implicitement dans des textes (Liu, 2012). Elle englobe les tâches suivantes : 1) l'analyse de subjectivité porte sur la détection de la présence de sentiments via l'identification d'expressions ou des mots dit subjectifs ; 2) l'analyse de polarité porte sur la détection de la polarité *positive*, *négative* ou *neutre* des textes ; 3) l'analyse des émotions porte sur la catégorie émotionnelle du texte (e.g. *colère*, *dégoût*, *peur*, etc.) ; 4) l'analyse d'intensité porte sur les différents niveaux d'intensité de la polarité et de l'émotion (e.g. *très positif*, *très triste*, etc.). Ces approches offrent une granularité plus précise sur les opinions et les émotions exprimées. Dans ce travail, nous nous focalisons sur la troisième tâche. Comme la plupart des méthodes semi-automatiques de la littérature, nous utiliserons la typologie des émotions de (Ekman, 1992) qui décrit six émotions.

Les méthodes appliquées pour analyser les sentiments sont très nombreuses et généralement, spécifiques aux types des textes : aux tweets (Roberts et al., 2012), aux titres de presse (Strapparava et Mihalcea, 2008), etc. et aux domaines d'application : l'analyse de media sociaux (Balahur, 2013) ou l'identification de mails suicidaires (Pestian et al., 2012).

Quelle que soit la tâche d'analyse de sentiments étudiée (polarité et émotions), la plupart des travaux porte soit sur la création de ressources permettant de décrire les sentiments, soit sur l'utilisation de ces ressources pour classifier des textes selon les sentiments étudiés. Dans la première catégorie de travaux, la plupart des méthodes associe les mots des textes à des termes appartenant à des ressources préalablement annotées par des sentiments. La plupart des ressources ont été construites pour l'anglais et l'analyse de la polarité (e.g. General Inquirer (Stone et al., 1966)). Toutefois des ressources plus spécifiques, comme le lexique de (Mohammad et Turney, 2010) ont été créées pour les mots chargés d'émotions. Pour la classification proprement dite, la plupart des approches utilisent des techniques d'apprentissage basées sur des attributs spécifiques incluant les mots d'émotions (Strapparava et Mihalcea, 2008) pour construire un modèle statistique à partir d'un corpus de textes et l'utiliser pour la détection des sentiments dans d'autres textes. Si beaucoup de ces méthodes s'avèrent efficaces sur des corpus de textes importants, elles se retrouvent limitées dans le cas des textes courts comme les tweets ou spécifiques comme les forums de santé.

3 Corpus d'étude

Nous avons construit un corpus à partir de 17 000 messages issus du forum de santé anglais *Spine-health*. Nous avons annoté automatiquement ce corpus avec le lexique d'émotions de (Mohammad et Turney, 2010). Ce lexique est composé de plus de 14 000 entrées caractérisées par leurs polarités et associées à 8 émotions. Dans ce travail, nous nous intéressons uniquement aux six émotions de (Ekman, 1992) : *colère*, *dégoût*, *peur*, *joie*, *tristesse*, *surprise*. Chaque mot du lexique peut être associé à plusieurs émotions. Cette annotation automatique permet de filtrer les messages objectifs (22% des messages). Afin de ne travailler que

sur des émotions portant sur des objets médicaux, nous avons utilisé le MESH² pour repérer des entités médicales et nous avons filtré 6% des messages n'en contenant pas. Dans un message, plusieurs émotions sont généralement exprimées du fait de sa longueur. Nous avons donc choisi de segmenter les messages en phrases et gardé 3 000 phrases pour constituer un **Corpus Annoté Automatiquement (CAA)**. Toutes les phrases contenant plusieurs émotions ont été étiquetées par l'émotion majoritaire. Un sous-ensemble de ce corpus (600 phrases) a été annoté manuellement par 60 non professionnels de santé³. Nous le notons CAM (**Corpus Annoté Manuellement**).

Pour évaluer l'accord entre les annotateurs, nous avons utilisé la mesure *Kappa*. 150 phrases issues du CAM ont été annotées par deux annotateurs non professionnels (*Kappa* égal à 0,26 soit accord très faible) et par un annotateur professionnel de santé et un non professionnel (*Kappa* de 0,46 soit accord modéré). Cette expérimentation préliminaire souligne la difficulté de la tâche d'annotation manuelle. Par ailleurs, le désaccord entre annotateurs est essentiellement dû à la variabilité entre personnes et non à leur sensibilité au domaine de la santé. Un premier biais consiste à ne considérer que le point de vue de l'annotateur qui est parfois très différent de celui de l'auteur. En effet, les messages des forums de santé traitent de la maladie, des traitements, etc. Ces informations sont par nature négatives et l'annotateur aura, par empathie, tendance à associer une émotion telle que la *tristesse* à une information factuelle comme la description d'un diagnostic. Un deuxième biais réside dans le fait que le corpus est rédigé en langue anglaise, alors que les annotateurs sont des français natifs. Par ailleurs, en étudiant les phrases ayant suscité des écarts d'annotations, nous avons remarqué qu'il est plus facile d'identifier la polarité que de trouver l'émotion elle-même. Il est également plus facile de prédire les émotions *positives* car les émotions *négatives* partagent un vocabulaire très proche. Nous avons noté également que la *surprise* est la plus difficile à identifier. La qualité de notre corpus annoté est finalement assez discutable mais cette étude donne une bonne intuition des difficultés liées à l'obtention d'un corpus de qualité et de la méthodologie à suivre pour améliorer sa qualité. Dans la suite, à partir de ces différents constats, nous avons décidé d'évaluer différentes méthodes permettant de caractériser un texte extrait d'un forum en nous basant sur : 1) une classification bi-classes pour identifier la polarité des émotions (les émotions *positives* correspondant à la *joie*, les émotions *négatives* à la *colère*, la *peur*, la *tristesse* et le *dégoût*). L'émotion *surprise* a été éliminée du fait de sa neutralité. 2) une classification multi-classes pour les 6 émotions : une phrase ne peut être associée qu'à une seule classe d'émotion ; 3) une classification multi-labels nous permettant d'associer une phrase à plusieurs classes d'émotions.

4 Protocole expérimental

Prétraitements : les messages dans les forums contiennent des mots qui ne se retrouvent pas forcément dans les dictionnaires classiques (argot, mise en forme particulière, abréviations, émoticônes, etc.). Il est donc nécessaire de les normaliser en généralisant leur contenu. Pour cela, nous avons appliqué les prétraitements correspondant à la chaîne mise en place par (Balahur, 2013) pour les tweets.

2. Medical Subject Headings, <http://www.ncbi.nlm.nih.gov/mesh>

3. Il s'agit essentiellement d'étudiants de master et d'informaticiens de notre laboratoire.

Classification : nous avons utilisé les attributs ci-dessous afin de trouver les meilleurs descripteurs des émotions : 1) les attributs basés sur les N-Grammes (U,U+B) ; 2) les mots d'émotions (ME) : si une phrase contient deux mots correspondant à l'émotion joie, elle prendra la valeur 2 pour cet attribut ; 3) les smileys (SMI) : l'ensemble des émoticons (:-), :-(...) a été classé selon les six émotions. Si une phrase contient un smiley correspondant à la joie, elle prendra la valeur 1 pour cet attribut ; 4) les intensifieurs (INT) : ponctuations (!!, ??, etc.), lettres répétées (*loool*) et mots en majuscules (*HATE*). Si une phrase contient un intensifieur, elle prendra la valeur 1 pour cet attribut ; 5) le contexte de l'émotion (CONT) : nous utilisons deux attributs que nous appelons émotion voisine et globale. Une phrase prend la valeur vraie pour cet attribut si une phrase qui l'entoure exprime la même émotion et pour l'émotion globale on considère tout le message ; 6) les patrons : en nous inspirant de l'approche de (Béchet et al., 2013), nous enrichissons les attributs en utilisant des patrons obtenus en appliquant un algorithme de recherche de motifs séquentiels. Pour cela, nous utilisons le thésaurus médical MeSH pour identifier des mots d'émotion, le lexique des mots d'émotions pour identifier des traces d'émotions et un lemmatiseur pour obtenir la catégorie grammaticale des mots. Chaque phrase est alors considérée comme une séquence d'itemsets correspondant à une combinaison de ces trois informations. Nous avons ensuite utilisé l'algorithme GSP (Zhang et al., 2002) pour obtenir les sous séquences fréquentes. Nous n'avons conservé que celles contenant au moins une entité médicale et un mot d'émotion. Ces motifs ont ensuite été utilisés comme attributs. Une phrase est étiquetée vraie pour un patron si sa forme syntaxique respecte le patron.

Évaluation : la qualité de la classification bi-classes est évaluée en utilisant les mesures classiques de *précision* P , *rappel* R et la *F-mesure* F . Pour la classification multi-classes, nous calculons à la fois la moyenne au niveau micro F_{mi} et macro F_{ma} . Pour la classification multi-labels, on utilise alors d'autres mesures comme le *Hamming loss* HL , l'*exactitude* A et la *F-mesure* au niveau macro F_{ma} .

Mise en œuvre : nous utilisons les implémentations de Weka⁴ pour la classification bi-classes et multi-classes et Meka⁵ pour la classification multi-labels. Nous utilisons SVM comme classifieur avec sa mise en œuvre SMO dans Weka en utilisant les paramètres par défaut. Nous utilisons le classifieur chain CC mis en œuvre dans Meka pour la classification multi-labels. Nous avons utilisé deux jeux de données : le corpus CAM et le corpus CAA. Nous réalisons une validation croisée (à 10 folds).

5 Résultats et discussions

La table 1 présente les résultats obtenus sur le corpus CAM. Nous ne présentons pas les résultats obtenus sur le corpus CAA qui sont similaires. La classification bi-classes donne les meilleurs résultats, ce qui semble relativement cohérent car cette tâche s'est également avérée plus facile pour les annotateurs humains. La classification multi-labels présente de meilleurs résultats que la classification multi-classes moins efficace lorsqu'un exemple peut être associé à plusieurs classes. Par ailleurs, nous constatons le peu de différence de la F-mesure micro et macro pour la classification multi-classes, qui semble suggérer que toutes les classes sont aussi difficiles à identifier. Ces résultats sont à mettre en parallèle avec l'accord inter-annotateurs (voir section 3). Dans les deux cas, la tâche est difficile mais les méthodes semi-

4. <http://www.cs.waikato.ac.nz/ml/weka/>

5. <http://meka.sourceforge.net/>

automatiques semblent détecter des régularités plus systématiquement sauf cas particuliers comme par exemple l'*ironie*. On peut également conclure que les meilleurs descripteurs sont les unigrammes combinés aux bigrammes et aux mots d'émotions (U+B+ME). La prise en compte des smileys et des intensifieurs n'améliore pas la classification. En effet, ils sont souvent utilisés à des fins d'ironie que nous ne captions pas en regardant uniquement leur présence dans la phrase. De même, le contexte n'est pas un attribut intéressant. En effet, un message est souvent long (7 phrases en moyenne dans notre corpus) et contient de nombreux sentiments (jusqu'à 6 émotions dans 41% des messages). Deux phrases consécutives contiennent donc très souvent des émotions différentes non corrélées. Pour finir, les patrons se sont également avérés peu efficaces car non définis pour chaque classe. Il est important de noter que, quand les différents classifieurs se trompent, ils placent souvent l'exemple dans une classe "proche" appartenant à la même polarité. Ces mauvaises prédictions sont dues au fait que les classes partagent de nombreux mots (comme *colère*, *dégoût* et *tristesse*). Les ressources utilisées étant des dictionnaires et un lemmatiseur, la méthode pourra être reproduite pour d'autres langues en utilisant des ressources similaires.

Attributs utilisés	Bi-classes			Multi-classes		Multi-labels		
	<i>R</i>	<i>P</i>	<i>F</i>	<i>F_{mi}</i>	<i>F_{ma}</i>	<i>HL</i>	<i>A</i>	<i>F_{ma}</i>
U	61,9	57,9	59,8	23,8	21,3	0,21	51,6	57,2
U+B	64,8	61,7	63,2	23,1	22,2	0,13	58,2	58,4
U+B+ME	65,1	62,2	63,6	26,7	25,3	0.11	60,0	62,3
U+B+ME+SMI+INT	50,8	50,3	50,5	22,2	20,8	0,23	54,1	56,9
U+B+ME+CONT	51,6	40,8	45,6	23,7	23,5	0,21	55,3	57,4
U+B+ME+PAT	65,3	61,1	63,1	25,2	25,0	0,04	58,4	63,3

TAB. 1 – Comparaison des résultats obtenus selon les attributs en utilisant le corpus CAM

6 Conclusions et perspectives

Nous avons décrit dans cet article une méthode d'analyse d'émotions dans les messages des forums de santé. La principale difficulté a résidé dans l'acquisition de données annotées et cette étape devra encore être améliorée. Pour l'extraction des émotions, nous avons comparé différents attributs, pour différentes tâches de classification (bi-classes, multi-classes et multi-labels) et montré que les plus efficaces étaient les unigrammes combinés aux bigrammes et aux mots d'émotions pour la classification bi-classes. Toutefois, suggérer une étiquette, malgré la précision obtenue, semble pertinent aux professionnels de santé impliqués dans cette étude. Les perspectives associées à ce travail sont nombreuses. Du point de vue de la tâche d'analyse des émotions, nous allons appliquer notre méthode sur des jeux de données plus importants et non spécifiques à la santé, comme ceux du challenge SEMEVAL⁶. Nous prendrons également en compte les "inverseurs de sens" (*shifters*). Nous validerons la généralité de notre approche sur un corpus en français. En effet, notre méthode repose uniquement sur des lexiques et un outil de lemmatisation. Nous avons également identifié des perspectives liées au domaine d'application. Le forum spine-health est spécialisé dans la thématique "douleur" et la pathologie

6. <http://www.cse.unt.edu/rada/affectivetext/>

discutée est une pathologie de personnes âgées. La nature des textes des messages est fortement liée à ces deux facteurs (peu de smileys ou d'argot, peu d'expressions de joie, etc.). Afin d'étudier d'autres sentiments, nous devons diversifier les thématiques des forums étudiés. Plus généralement, nous étudierons l'influence des médias sur les changements d'émotions des patients comme dans le cas de la pilule de troisième génération.

Références

- Balahur, A. (2013). Sentiment analysis in social media texts. In *4th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, Atlanta, Georgia, pp. 120–128. ACL.
- Béchet, N., P. Cellier, T. Charnois, B. Crémilleux, et S. Quiniou (2013). SDMC : un outil en ligne d'extraction de motifs séquentiels pour la fouille de textes. In *Conférence Francophone sur l'Extraction et la Gestion des Connaissances*, Toulouse, France. session démonstration.
- Ekman, P. (1992). An argument for basic emotions. *Cognition and emotion* 6(3-4), 169–200.
- Liu, B. (2012). *Sentiment Analysis and Opinion Mining*. Morgan & Claypool Publishers.
- Mohammad, S. M. et P. D. Turney (2010). Emotions Evoked by Common Words and Phrases : Using Mechanical Turk to Create an Emotion Lexicon. In *Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*, Stroudsburg, PA, USA, pp. 26–34. ACL.
- Pestian, J. P., P. Matykiewicz, M. Linn-Gust, B. South, O. Uzuner, J. Wiebe, K. B. Cohen, J. Hurdle, C. Brew, et al. (2012). Sentiment analysis of suicide notes : A shared task. *Bio-medical Informatics Insights* 5(Suppl. 1), 3.
- Roberts, K., M. A. Roach, J. Johnson, J. Guthrie, et S. M. Harabagiu (2012). Empatweet : Annotating and detecting emotions on twitter. In *Eight International Conference on Language Resources and Evaluation*, Istanbul, Turkey. ELRA.
- Stone, P. J., D. C. Dunphy, M. S. Smith, et D. M. Ogilvie (1966). *The General Inquirer : A Computer Approach to Content Analysis*. MIT Press.
- Strapparava, C. et R. Mihalcea (2008). Learning to identify emotions in text. In *Symposium on Applied Computing*, New York, NY, USA, pp. 1556–1560. ACM.
- Zhang, M., B. Kao, C. lap Yip, et D. Cheung (2002). A gsp-based efficient algorithm for mining frequent sequences.

Summary

Online health forums are spaces of discussion where patients express their feelings about their diseases, treatments, etc. Because of the anonymity offered by these websites, they express very freely their personal experiences. Health forums are very useful source of information for health professionals in order to better understand patient's feelings, problems and behaviours. In this paper, we use text mining techniques to analyse health forum in order to extract emotions (e.g. joy, anger, surprise) expressed by the patients.