

La vérité sur la tonalité & l'analyse automatique du langage

Réalisé par Synthesio

Sommaire

Introduction	.2
Pas d'intelligence artificielle pour l'analyse des sentiments	.3
L'analyse humaine, un passage obligatoire de l'analyse des contenus	.5
Les avancées des technologies actuelles	.6
Le futur des technologies sémantiques	.8
Conclusion	.10

Introduction

Depuis près de 30 ans de nombreuses entreprises et chercheurs en linguistique s'intéressent à l'analyse automatique de contenus. Cette discipline communément appelée le traitement automatisé de la langue naturelle (TALN) se retrouve au cœur des débats avec l'avènement des médias sociaux. Les professionnels du marketing digital, du service client et de la communication cherchent aujourd'hui sur l'outil idéal leur permettant d'analyser automatiquement la teneur et la tonalité des conversations publiées sur leurs marques sur le web. Quel est l'état actuel des technologies d'analyse du langage ? Dans quelle mesure peut-on se fier à leurs résultats ? Quel est l'impact des médias sociaux sur les futurs développements de cette discipline controversée ?

Pas d'intelligence artificielle pour l'analyse des sentiments

De la difficulté d'attribuer le bon sentiment à un message

La classification automatique des textes par grandes thématiques peut s'avérer être relativement simple. Si le mot « voiture » ainsi qu'un ensemble de termes associés au champ lexical de l'automobile apparaissent à plusieurs reprises dans un texte, un outil peut en déduire rapidement que le texte traite de l'automobile. La classification par sentiment, elle, s'avère plus difficile même si elle est bien souvent binaire (positif/négatif). En effet, il arrive qu'un message relate les paroles d'un tiers en contradiction avec sa propre opinion, ou bien qu'il énumère des qualités qui ne sont pas parvenues à convaincre le locuteur. Dans ces deux cas, l'ordinateur relèvera seulement le sentiment majoritaire, gommant les nuances du texte conduisant souvent à des contresens.

Intéressons-nous à l'avis suivant : « Ce film a tout pour être *excellent* : une intrigue *palpitante*, un casting de rêve, de réelles performances d'acteurs et pourtant le film ne tient pas la route. »

Ici, la machine relèvera les mots en italiques et en déduira que la critique est positive. Tandis qu'un lecteur saura immédiatement, en lisant la dernière phrase notamment, que cet avis est négatif.

La tâche est complexe et les dictionnaires utilisés pour l'analyse automatique de sentiments doivent pouvoir tenir compte des éléments suivants :

- La polarité du sentiment et les degrés de positivité
- La détection de la subjectivité et l'identification d'une opinion
- L'analyse conjointe sujet-tonalité
- Les points de vue
- Les informations non-factuelles (liées aux six émotions universelles : la colère, le dégoût, la peur, la joie, la tristesse et la surprise)
- La syntaxe
- La négation

Classer les sentiments à l'aide d'un lexique

Dans la quête de l'automatisation de l'analyse de tonalités, certains chercheurs ont voulu créer une typologie des sentiments, attribuant aux mots une tonalité positive ou négative. Avec cette méthode, les outils n'analysent pas la tonalité du message dans son ensemble mais chacun des mots qui le compose.

Cette démarche soulève plusieurs problèmes et a jusqu'à présent obtenu des résultats peu satisfaisants. En effet, l'absence de prise en compte du contexte ne permet pas d'interpréter le sentiment général véhiculé par un message. Or, il s'agit bien là du but de l'analyse de sentiment. Par ailleurs, chaque chercheur impliqué dans la mise au point du logiciel peut avoir une perception différente du sentiment lié à un mot et fausser les réglages.

Diversité des médias et diversité des formes d'expression

La diversité des types de médias présents sur le web implique un premier niveau de tri dans l'analyse de sentiments. Tandis que les médias traditionnels sont généralement factuels, les médias sociaux présentent des opinions souvent tranchées. Mais il arrive également qu'un message sur un réseau social transmette une information neutre ou bien qu'un article de magazine cite

une opinion. L'algorithme doit donc être capable de détecter la présence d'une opinion dans un texte, en faisant par exemple la différence entre les phrases « la batterie dure deux heures » et « la batterie dure seulement deux heures », ce qui n'est pas encore le cas des technologies actuelles.

Par ailleurs, les médias sociaux ont induit une nouvelle forme de communication semblable au langage SMS avec des émoticônes, de l'argot fréquent, des fautes d'orthographe ou une syntaxe simplifiée. Ces nouvelles manières de s'exprimer ne suivent plus des règles et des codes automatisables.

L'ironie, trop subtil pour une machine

Tout comme c'est le cas à l'oral, l'expression d'une opinion peut se faire de manière indirecte, grâce à l'humour, au sarcasme ou à l'ironie. Relever l'ironie à l'écrit n'est pas toujours une tâche évidente pour un être humain, mais pour une machine cela devient quasiment impossible puisqu'elle se fonde sur la tonalité des mots employés et leur fréquence pour déduire un sentiment. Ainsi, l'ordinateur ne pourra pas détecter l'emploi d'antiphrases et les interprètera au premier degré faussant ainsi son analyse.

La machine ne sait pas différencier les sujets au sein d'un article

Au sein d'un article de blog ou d'un avis, un internaute peut choisir de comparer différents produits ou marques en variant les tonalités. Or, l'ordinateur analyse la retombée comme un bloc monolithique et ne sait pas attribuer les sentiments aux objets décrits. Cette situation devient d'autant plus complexe qu'elle apparaît au sein de conversations, sur des forums par exemple.

Les limites de l'analyse multilingue

La majorité des recherches consacrées à l'analyse de tonalité s'est concentrée sur des textes écrits en anglais. Par conséquent, la plupart des outils d'analyse de sentiment ont été développés en langue anglaise.

Seth Grimes, expert en linguistique



Certaines entreprises traitent la tonalité en une seule langue (l'anglais généralement) tandis que d'autres proposent des analyses en 10 langues différentes. Les entreprises faisant de l'analyse de sentiments dans peu de langues utilisent des outils linguistiques (lexiques, dictionnaires, règles de grammaire, etc.) mais l'analyse multilingue est encore quasiment inexploitable. On ne peut pas se contenter de traduire du contenu linguistique anglais en français ou en chinois et s'attendre à des résultats raisonnables.

L'analyse humaine un passage obligatoire de l'analyse des contenus.

L'analyse automatique, moyen le plus rapide de décrypter un volume important de données

Il est possible de rationaliser le traitement d'un flux important d'informations en effectuant d'abord une analyse automatique de tonalités qui sera suivie d'un ajustement de la part des utilisateurs. Cette tâche automatique peut également permettre d'obtenir un premier aperçu des tendances d'opinions (positives/négatives) à une échelle très globale et de leur évolution dans le temps. En faisant cela, la marque peut commencer avec un tableau clair de la situation avant de s'intéresser aux détails explicatifs.

Préférer les retours précis aux moyennes statistiques

Les notations en ligne peuvent donner une vision déformée de la réalité, car les internautes ont tendance à exagérer leur jugement lorsqu'on leur demande d'attribuer une note, aussi bien positivement que négativement. Par ailleurs, elles ne fournissent pas de données vraiment actionnables. Il est important pour une marque de connaître le sentiment général des internautes tout en comprenant les raisonnements justifiant leurs avis. Certains commentaires méritent plus d'intérêt que d'autres et peuvent s'avérer déterminant dans la mise en place d'une nouvelle stratégie.

« Laissons l'expertise aux êtres humains » -KD Paine

Les ordinateurs ne sont pas encore assez évolués pour gérer les subtilités du langage humain, en particulier sur le web, malgré leurs avantages nombreux. L'analyse humaine reste le moyen le plus sûr de déchiffrer correctement les informations qui circulent en ligne.

La spécialiste des Relations Publiques, KD Paine, explique d'ailleurs :



« Les ordinateurs sont capables de beaucoup de choses, mais la différenciation entre commentaires positifs et négatifs sur les médias sociaux ne fait pas partie de leur champ de compétence. Le problème de ces médias est qu'ils sont parsemés d'ironie, de sarcasme et d'expression non traditionnelle des sentiments. C'est pourquoi nous recommandons une solution hybride. Laissons les tâches grossières aux ordinateurs et l'expertise aux êtres humains. »

Les avancées des technologies actuelles

Des ordinateurs en progrès permanents

Bien que la technologie n'ait pas encore atteint un stade suffisant pour fonctionner en autonomie, la recherche informatique ne cesse d'avancer et présente des résultats de plus en plus surprenants, laissant présager une victoire de la machine sur l'homme à long terme.

Le projet Mao and Lebanon [201] s'intéresse à la détection des sentiments locaux, à savoir expression par expression, afin de résoudre les difficultés évoquées précédemment. Cet algorithme est fondé sur l'utilisation de contraintes mathématiques reflétant l'idée qu'un mot à connotation positive (ou négative) influe positivement (ou négativement, respectivement) sur la tonalité locale du texte.

Le Watson d'IBM, champion de Jeopardy



Avec Watson, la société informatique IBM n'en était pas à son coup d'essai. En 1997, déjà, son ordinateur Deep Blue avait vaincu le champion Gary Kasparov aux échecs. Mais tandis que Deep Blue s'appuyait sur la puissance de ses calculs, Watson utilise les technologies d'analyse sémantique du langage naturel. Capable de gérer 128 tâches simultanées et disposant de plusieurs téraoctets de données, Watson a participé au célèbre jeu télévisé américain, Jeopardy, remportant plusieurs victoires contre les meilleurs joueurs. Le principe de Jeopardy est le suivant : le présentateur donne une réponse et il s'agit de trouver la question correspondante. Malgré le fort taux de bonnes réponses chez les

candidats humains opposés à Watson, le superordinateur a su l'emporter grâce à sa rapidité, donnant 13 bonnes réponses avant même que l'un des champions n'ait pu débiter.

Mais si la nouvelle machine d'IBM est douée aux jeux télévisés, c'est, à terme, une utilisation médicale que les ingénieurs d'IBM avaient en tête lors de sa conception. En effet, Watson pourrait être capable d'accélérer le processus de détermination des diagnostics.

L'ordinateur est capable d'attribuer un sens à un corpus homogène

Plus le domaine d'analyse est limité, plus l'ordinateur sera efficace. Sur un corpus homogène de textes, avec une langue unique et un champ lexical limité, l'ordinateur est capable d'analyser de manière très précise les tonalités des messages, et ce, peu importe la quantité d'information à traiter.

Un ordinateur peut donc s'avérer extrêmement précieux dans un cadre d'analyse bien défini, puisqu'il accélère le processus de traitement de l'information avec une faible marge d'erreur.

Les intentions détectées grâce à des modèles syntaxiques complexes¹

Certains experts ont essayé d'incorporer les relations syntaxiques aux logiciels d'analyse textuelle. Une fois mises au point, ces fonctionnalités sont testées sur des corpus de textes dans le but de former le logiciel à la détection d'expressions de la subjectivité. Ces tests se déroulent de la manière suivante : des modèles syntaxiques sont construits avant d'être utilisés sur un corpus d'entraînement, déclenchant ainsi un modèle d'extraction spécifique à chaque apparition d'une nouvelle forme syntaxique différent. Par exemple, « <x> m'a fait plaisir » apparaîtrait à chaque occurrence du mot « plaisir ». Toutefois, ces techniques sont limitées puisque le logiciel se réfère à des expressions syntaxiques préconstruites plutôt qu'aux mots exacts

¹ [Learning Extraction Patterns for Subjective Expression](#)

utilisés. Quand on analyse le sentiment d'un extrait, il s'agit ici uniquement de la première étape du processus visant à découvrir s'il y a effectivement un sentiment exprimé.

Des êtres humains trop subjectifs pour être fiables

Ces nouveaux ordinateurs apparaissent d'autant plus puissants que l'être humain a encore ses faiblesses, principalement dues à sa subjectivité.

Celle-ci se retrouve bien sûr chez les analystes qui ont chacun leur propre système d'interprétation : certains ont tendance à surévaluer la négativité des messages qu'ils observent, d'autres à les voir d'un œil plus positif. Une analyse humaine n'est donc pas toujours fiable, surtout si elle est réalisée par plusieurs personnes.

Un ordinateur au contraire fait preuve de constance, reproduisant à la chaîne le même algorithme et il est impossible, à moins d'une faille dans la conception, qu'il analyse de deux manières différentes la même information. Et il est également beaucoup plus rapide, comme le montre l'exemple de l'ordinateur Watson d'IBM.

Le futur de l'analyse sémantique

Interview de Seth Grimes, un "visionnaire de l'analytique"



"Watson", l'ordinateur d'IBM a gagné au jeu télévisé Jeopardy, créant ainsi un buzz autour de sa technologie. Pourquoi tant de bruit à votre avis ?

Faire participer un ordinateur à Jeopardy a été un joli numéro. IBM est parvenu à accomplir un exploit compréhensible par tous. Si j'appelle cet accomplissement un « numéro », c'est parce qu'avoir un ordinateur capable de gagner à Jeopardy ne fait pas partie des priorités du monde du commerce, ou de la société de manière plus générale. Néanmoins, le fait de voir Watson participer à Jeopardy a permis au public non-scientifique de se rendre compte du potentiel et de la réalité de la technologie.

Les systèmes de questions-réponses existent déjà, automatisant les réponses à des questions d'ordre commercial sans que personne n'ait besoin d'être à l'autre bout du fil (ex : support client, requêtes, e-commerce). Pour le moment, Watson se contente d'extraire des informations factuelles, mais cette technologie pourrait tout aussi bien travailler sur les sentiments, grâce à un « annotateur » de sentiments. Alors, nous ne devrions plus nous limiter aux questions factuelles : nous serions capables de l'interroger sur les opinions et les émotions.

Quel degré de précision cette technologie peut-elle atteindre ?

Les objectifs de précision et la quantité de travail consacrés à leur réalisation devraient être fixés en vue d'une commercialisation. Certaines difficultés peuvent être levées avec un faible niveau de précision (ex : la classification en sentiments positifs/négatifs) tandis que d'autres applications nécessitent une précision plus élevée.

D'après ce que j'ai pu constater, la plupart des outils extracteurs de tonalité disposent d'une précision de 40 à 50% à l'état brut, mais ils peuvent atteindre un niveau supérieur à 80% s'ils sont formés par des humains se chargeant de créer des échantillons annotés ou des règles de langage et de corriger l'outil. J'ai vu quelqu'un revendiquer une précision de 98%, ce qui est risible et ridicule. Le seul moyen d'atteindre un tel niveau, c'est de restreindre les limites du problème au maximum, de créer un logiciel sur mesure pour le cas donné et d'être plus indulgent face aux erreurs.

Le plus important, c'est de commencer par détecter la présence de tonalité avant même de déterminer si elle est positive ou négative, puis de soumettre le texte à une classification humaine ou automatique. En utilisant la machine pour filtrer et les capacités humaines pour l'analyse, on peut, dans certains cas, engendrer des niveaux très élevés de précision. Si l'on souhaite que la machine s'occupe du processus entier, la quantité de travail à fournir pour obtenir un niveau de précision élevé est beaucoup plus importante, mais, encore une fois, ces décisions doivent être prises en fonction des besoins commerciaux et de la nature des textes à traiter.

Les outils effectuant uniquement des analyses globales (au niveau du message ou du document) fournissent très souvent des résultats beaucoup moins utiles malgré un meilleur taux de réussite. Réfléchissez un peu. Il peut être utile pour une personne à la tête d'un groupe hôtelier, composé de 4 200 établissements, de savoir que 77% des critiques sont positives, 17% neutres et 6% négatives (nombres inventés). Mais ne serait-ce pas encore plus utile de connaître le détail des opinions pour chaque hôtel ? Cette personne a besoin de savoir si un client, qui a trouvé sa chambre propre et le personnel exemplaire, a en revanche été gêné par le bruit. Au sein d'un avis globalement positif, tous les détails ne seront pas également positifs, et seule une analyse détaillée des sentiments peut permettre d'améliorer son produit ou son service.

Néanmoins, essayons de ne pas surestimer l'exactitude de l'analyse humaine des sentiments. La meilleure étude que j'ai lue sur la précision a été réalisée par l'université de Pittsburgh en 2005. Ils ont découvert que les analyses de sentiment réalisées à la main concordaient à hauteur de 82%. Ces résultats passaient à 90% lorsqu'on ne tenait pas compte des cas incertains, à savoir les cas pour lesquels certains disaient ne pas être sûrs de leur analyse.

Qu'en est-il de l'analyse multilingue ?

Certaines entreprises traitent la tonalité en une seule langue (l'anglais généralement) tandis que d'autres proposent des analyses en 10 langues différentes. Les entreprises faisant de l'analyse de sentiments dans peu de langues utilisent des outils linguistiques (lexiques, dictionnaires, règles de grammaire, etc.) mais l'analyse multilingue est encore quasiment inexploitable. On ne peut pas se contenter de traduire du contenu linguistique anglais en français ou en chinois et s'attendre à des résultats raisonnables.

D'autres outils choisissent, d'appliquer des méthodes statistiques pour extrapoler les tonalités à partir d'un large volume de données.

Globalement, je pense que les meilleurs résultats multilingues sont obtenus grâce à un savant mélange d'approches linguistique et statistique, renforcé par un apprentissage automatisé de la machine.

Pour revenir sur l'exactitude, la question à se poser est la suivante : Est-il préférable de passer des mois à bâtir un algorithme d'analyse pour un sujet donné, là où un être humain mettrait certainement moins de temps et apporterait un meilleur degré de précision ?

Certains canaux de communication (parmi les forums, les blogs, Twitter, etc.) sont-ils plus faciles à analyser que d'autres grâce à la fouille de textes ?

Un message sur un forum peut décrire la propreté d'une chambre d'hôtel, la sympathie du personnel ou d'autres qualités mais l'utilisation d'un algorithme ne permettra pas de retenir beaucoup de ces informations utiles. Pour obtenir des résultats corrects, il faut les analyser de manière plus détaillée, sur des segments de phrase ou des mots.

Twitter présente beaucoup d'intérêt car il est difficile d'exprimer plus d'une idée en un tweet. La plupart des tweets tournent autour d'une idée unique, ce qui, en théorie, devrait les rendre plus simples à analyser. Le problème, est que, contrairement aux blogs ou aux articles de presse, les tweets sont remplis de mots d'argot et d'abréviations, ce qui ne facilite pas l'analyse. En outre, un tweet fait souvent partie d'une conversation. Très peu de tweets font sens tout seuls : beaucoup incluent un lien vers un article ou répondent à d'autres, par exemple. Certains tweets font partie d'une conversation à plusieurs et la compréhension du contexte du tweet passe par la compréhension de l'ensemble de la conversation. La majorité des outils existant ne prend pas en compte cette composante, elle ne va pas au-delà du tweet pour intégrer la nature conversationnelle de Twitter.

De quelles avancées récentes dans la technologie d'analyse de sentiments êtes-vous témoin ?

Nous devons aller plus loin que la simple analyse de tonalité. Par exemple, on peut vouloir s'intéresser au sentiment par catégories émotionnelles (colère, tristesse, joie) dans le cadre d'un service hôtelier. Je suis sûr que l'on peut imaginer, dans un contexte commercial, de nombreuses applications utiles de la compréhension automatisée de la tonalité émotionnelle.

Ensuite, il y a les signaux d'intention dont je viens de parler : des sentiments qui indiquent des projets ou des actions.

Afin d'optimiser la flexibilité de son entreprise, il faut savoir créer des catégories adaptées à son commerce grâce à une approche statistique. Ainsi, l'analyste crée des catégories pertinentes et il place chaque document dans la bonne catégorie afin d'entraîner l'outil. La machine utilise alors des mesures statistiques de similarité pour découvrir les points communs des objets présents dans chaque catégorie, dans le but d'automatiser la classification.

Conclusion

Aucun fournisseur de solution de veille en ligne n'ose prétendre que la technologie est suffisamment avancée pour déterminer les sentiments autour d'un sujet donné. Néanmoins, certains affirment que le Traitement Automatique du Langage Naturel peut au moins permettre d'identifier des macro-tendances comme les sujets de conversation les plus populaires ou les évolutions des sentiments dans le temps. A ce jour encore très limitées, ces technologies sont utilisées quasi exclusivement en anglais et se heurtent à la complexité du langage sur les médias sociaux. Encore de nombreuses années de recherche seront nécessaires avant que des outils universels d'analyse de sentiment ne voient le jour.

A propos de Synthesio

Synthesio est une société innovante spécialisée dans l'analyse de réputation des marques sur le web. Nous proposons tableaux de bord et études pour accompagner au quotidien nos clients, annonceurs et agences, dans la définition et le suivi de leur stratégie sur les médias sociaux (Facebook, Twitter, Blogs, Forums, etc.).

De nombreuses marques internationales telles que Sanofi, Orange ou Accor font appel aux services de Synthesio pour comprendre la perception de leurs marques, anticiper les situations de crise ou encore préparer et analyser le lancement de nouveaux produits ou de campagnes de communication.