

Détection automatique de la subjectivité dans les textes - État de l'art ^{*}

Jo Oleiwan^[22005314], Adrien Delmote^[22001108], Mortimer Meunier^[22304652],
Martin Mouysset^[22013781], and Jules Bienassis^[21801606]

Université Toulouse III - Paul Sabatier

Abstract. Cet état de l'art présente la détection automatique de la subjectivité dans les textes, une dimension essentielle pour différencier opinions et faits dans le traitement du langage naturel. L'analyse couvre une gamme de techniques, des méthodes traditionnelles d'apprentissage automatique aux modèles neuronaux avancés, soulignant l'importance des représentations textuelles et les stratégies d'augmentation des données face aux déséquilibres des ensembles de données. Le document évoque également le raffinement des grands modèles linguistiques, illustrant leur efficacité accrue en matière de détection de subjectivité. Les limites des approches existantes sont examinées, offrant un aperçu des directions futures pour renforcer la précision et l'utilité de ces technologies.

Keywords: Détection de subjectivité · Traitement du langage naturel · Modèles d'apprentissage automatique · Représentations textuelles · Augmentation de données · Analyse de sentiments · Techniques de classification textuelle.

1 Introduction

L'ère numérique a donné naissance à un flux massif d'informations, rendant la distinction entre faits objectifs et opinions subjectives un enjeu important pour diverses applications, de la modération des contenus sur les réseaux sociaux à l'analyse des sentiments dans les critiques de produits dans le but de pouvoir créer par exemple des systèmes de recommandation [5, 6]. Cet enjeu est d'autant plus important dans le domaine de la presse, où l'intégrité du contenu peut influencer largement l'opinion publique. Par conséquent des travaux visant à détecter la désinformation ont été réalisés [7, 8]. Dans ce but, la recherche scientifique s'est orientée vers la création de modèles avancés capables de détecter la subjectivité dans les textes, une tâche qui s'inscrit dans le domaine plus large de l'analyse des sentiments et de l'intelligence artificielle appliquée au langage naturel [7].

Il est essentiel de souligner les définitions convergentes de la subjectivité selon différentes équipes de recherche, révélant une compréhension unifiée de ce concept traditionnellement ambigu. Selon l'article intitulé "On the Definition of Prescriptive Annotation Guidelines for Language Agnostic Subjectivity

^{*} L'équipe IAFA-tigables

Detection” [1], la subjectivité est vue comme une empreinte personnelle que le locuteur laisse dans son énoncé, reflétant ses positions, attitude et sentiments. Cette étude spécifie des critères précis d’annotation qui définissent la subjectivité notamment par l’usage d’intensificateurs ou de suppositions qui tendent à exprimer une opinion.

Un autre travail, discuté dans SubjectivITA [2] définit la détection de la subjectivité comme la capacité à discerner si un texte est influencé par les préjugés de son auteur, soulignant la complexité de la tâche en raison de la variabilité des niveaux d’expertise et des biais personnels.

Enfin, dans ”A Corpus for Sentence-level Subjectivity Detection on English News Articles” [3], les auteurs perçoivent la subjectivité comme l’interprétation individuelle du langage et des expériences personnelles. Dans cette optique, une phrase est subjective si elle révèle des opinions, des goûts ou des préférences personnels, notamment à travers des moyens linguistiques tels que le sarcasme, les figures de rhétorique ou l’expression d’émotions personnelles.

Comme expliqué par Pang et al [4], La subjectivité est étroitement liée à l’analyse de sentiments. En effet, l’analyse de sentiments a besoin de phrases qui évoquent le sentiment de l’auteur, donc de phrases subjectives. En détectant les phrases qui sont importantes pour le sentiment, on détecte aussi les phrases subjectives. Cette connexion est fondamentale pour comprendre comment les opinions et émotions exprimées influencent la polarité contextuelle positive, négative ou neutre d’un texte.

Pour réaliser cette tâche, l’idée est d’utiliser les méthodes d’apprentissage automatique pour arriver à classifier des données textuelles, que ce soit des documents ou des phrases, en tant que subjectives ou objectives.

Ce document tente de présenter un panorama des méthodes existantes et des avancées récentes dans ce domaine, tout en mettant en évidence l’importance de la représentation textuelle et du raffinement des modèles linguistiques pré-entraînés. Nous tentons également d’aborder le défi posé par les déséquilibres dans les jeux de données et les stratégies d’augmentation de données qui sont au cœur des solutions proposées.

Notre état de l’art est structuré de la façon suivante. La section 2 résume les collections de test existantes pour la détection de subjectivité, et répond ensuite à l’enjeu du déséquilibre des données à travers leur augmentation. La section 3 explique l’importance des représentations de texte dans le contexte de leur utilisation dans les modèles d’apprentissage à travers les plongements de mots ou de grands modèles de langues qui permettent de générer directement du texte avec du sens. La section 4 regroupe tout d’abord les différents modèles non neuronaux utilisés dans le cadre de la détection de subjectivité et l’analyse de sentiments, puis sont décrites en section 5 les approches neuronales pour plusieurs domaines d’applications en lien avec la détection de subjectivité. Enfin, nous discutons des limites de cet ensemble d’approches et des perspectives de recherches à l’avenir dans la section 6, avant de conclure en section 7.

2 Données et prétraitement

2.1 Collection de test

Afin d'évaluer leurs modèles, les groupes de recherches, les ingénieurs et les chercheurs ont besoin de collections de test[6] pour se situer en termes de performances. Les collections de test dans le contexte de l'apprentissage automatique sont des ensembles de données utilisés pour évaluer et comparer les modèles dans le cadre de tâches spécifiques. Dans cette section nous présentons un résumé sur des collections de test utilisés pour la détection de la subjectivité.

Chaturvedi et al [6] présentent de nombreuses collections de test, telles que :

| Nom du Dataset | Langue | Nombre d'Éléments | Détails des Annotations | Usage Principal |
|---------------------------------|----------|-------------------|---|---|
| MPQA | Anglais | 535 articles | 9 700 phrases (55% subjectives, 45% objectives) | Analyse de subjectivité et opinion |
| MPQA Gold | Espagnol | 504 phrases | 273 phrases subjectives, 231 phrases objectives | Validation d'annotations en subjectivité |
| Critique de Film (Movie Review) | Anglais | 10 000 phrases | 5 000 critiques de films (subjectives), 5 000 résumés d'intrigues (objectives) | Analyse de sentiment, subjectivité vs objectivité |
| TASS (SEPLN) | Espagnol | 7 219 tweets | Annotés comme positif, neutre, négatif, sans opinion (positif et négatif considérés subjectifs) | Analyse des sentiments, subjectivité dans les réseaux sociaux |
| Twitter (Collection Test) | Anglais | 498 tweets | Annotés comme positifs, négatifs ou neutres (positif et négatif considérés subjectifs) | Analyse des sentiments, subjectivité sur les réseaux sociaux |

Table 1. Aperçu de collections de tests utilisés en traitement du langage naturel

Dans le cadre de la tâche 2 sur la subjectivité de la recherche sur la détection de fake news, l'atelier CheckThat! utilise des collections de test pour valider les soumissions des équipes participantes. Le fichier Test Gold est construit par les organisateurs de la recherche, il utilisent des règles d'annotations spécifiques traduisant ce qu'est la subjectivité [1].

SubjectivITA est une autre collection de test, elle a été créée en rassemblant manuellement des articles de journaux en ligne italiens. Les articles sont choisis sur des médias nationaux, généralement considérés comme politiquement impartiaux, des chroniques et des blogs. Les sujets des articles sont nombreux afin de couvrir un large éventail de styles. Les articles ont été collectés entre le 20 janvier et le 1er février 2021 et ont été choisis au hasard. La collection contient 1 841 phrases et 103 documents. Les articles sont annotés entre subjectif et objectif [2].

2.2 Augmentation des données

Les différents modèles d'apprentissage automatique demandent tous une quantité de données plus ou moins importante. Il arrive souvent que le lot de données disponible soit insuffisant, ou que les auteurs souhaitent en augmenter la taille pour augmenter la précision du modèle. Différentes méthodes d'augmentation de données ont été observées [9, 10].

Traduction Si le jeu de données utilisé possède des données en plusieurs langues, la traduction peut être une technique efficace pour augmenter le nombre de données disponibles. En effet, dans le cas d'une langue cible, il est possible de traduire les autres textes de l'ensemble de données dans cette langue. De plus, même si le modèle à construire doit être multilingue, "l'inter-traduction" (traduire chaque texte dans chaque langue disponible) permet de significativement augmenter le nombre de données disponibles [9].

Si l'on souhaite éviter tout biais qu'une traduction pourrait apporter, il est possible de réunir toutes les langues en un seul ensemble de données multilingue "data aggregation". C'est ce qu'a fait l'équipe Thesis Titan dans le cadre de la tâche 2 de CheckThat! 2023 [10]. Cela leur a permis d'obtenir un ensemble de données multilingue plus important que celui fourni à la base.



Fig. 1. La back translation cf. [11]

Traduction aller-retour (back translation) Lorsque le jeu de données n'est constitué que d'une seule langue, la traduction aller-retour "back-translation" peut se montrer efficace. Elle se découpe en deux étapes : traduire une partie des données dans une langue tampon (aller) puis retraduire les phrases obtenues dans la langue d'origine (retour). Cette technique permet, en théorie, d'obtenir des variations des données d'entrée. L'équipe Accenture, dans le cadre du projet CheckThat! 2023, a expérimenté les limites de cette méthode [12]. Lorsque une classe est plus présente que d'autres dans un jeu de données, cela crée un déséquilibre qui se répercute sur le modèle entraîné. Pour pallier à ce déséquilibre, ils ont augmenté la quantité de données de la classe minoritaire. Ils ont essayé différents niveaux de traduction aller-retour: arabe - anglais - arabe ou arabe - anglais - espagnol - anglais - arabe. Ces tentatives ont permis d'augmenter la précision de leur modèle (de 0.816 à 0.823) mais ils ont constaté une baisse de

la précision et donc de la qualité des données d'entrée en effectuant plusieurs niveaux de traduction aller-retour. En effet, le but de cette pratique est de modifier le contenu des phrases récupérées comparé aux phrases données en entrée. Mais certains mots clefs sont porteurs de subjectivité dans une phrase : "Are there any resolutions that the Security Council may issue to ensure tat Egypt's water share in the Nile River will not be affected?" a fini par se modifier en "Are there resolutions that the Security Council could adopt to ensure that Egypt's share of water in the Nile is not affected?" . Le ton de la phrase est devenu légèrement plus objectif alors que, d'après la méthode, elle est catégorisée en tant que subjective. BLEU est une méthode pour l'évaluation automatique de machines de traduction, comme expliqué par Papineni et al [13]. Il faut que ce BLEU score soit assez élevé pour ne pas perdre d'informations, mais assez faible pour ne pas obtenir exactement la phrase de départ. C'est d'ailleurs le point abordé par l'équipe Accenture [12] dans la conclusion de leurs expérimentations: la recherche d'un point d'équilibre vis-à-vis de la qualité des traductions.

Reformulation + Génération Il est possible d'utiliser des outils de génération de texte pour reformuler des phrases existantes, ou bien en générer de nouvelles. Pour cela, GPT semble être un modèle de génération de texte efficace. Cependant l'utilisation de ces méthodes semble générer quelques problèmes, comme soulevé par [14]. En effet, le style d'écriture de GPT4 étant assez différent de celui habituellement trouvé, la différence de style présent dans le jeu de données après modification a été pointé comme une raison probable de la faible efficacité de leur modèle . Ce n'est cependant pas un problème systématiquement présent avec l'utilisation de GPT4. Comme décrit dans [15] par Shushkevich et al, des améliorations de performances ont été notées après augmentation de la taille du jeu de données avec de la génération de phrases ressemblantes en utilisant chatGPT4, en augmentant la taille des jeux de données de 830 à 1546 phrases, et de 1613 à 2118 phrases.

Afin d'améliorer la qualité des résultats obtenus, l'équipe DWReCo [16] a interviewé des rédacteurs de la Deutsche Welle (la radio internationale allemande) afin d'obtenir plus d'informations sur les styles de subjectivité journalistique. De leurs discussions ils ont extrait qu'un texte subjectif peut : être émotionnel, impliquer de la propagande, inclure des préjugés, être partisan, être désobligeant ou exagéré. Ils ont utilisé text-davinci-003 afin d'augmenter le nombre de données subjectives en créant des variations de certaines phrases dans chaque style. Ils ont alors pu créer plusieurs ensembles de données, un pour chaque style de subjectivité utilisé lors de l'augmentation de données. Grâce à cela, ils ont pu comparer les scores des modèles entraînés sur chaque ensemble, sur l'ensemble augmenté en combinant tous les styles et enfin sur l'ensemble non augmenté. Leurs résultats indiquent que le style de subjectivité permettant d'obtenir le meilleur modèle diffère pour chaque langue testée, à savoir l'anglais, le turc et l'allemand. L'anglais et le turc ont de meilleures performances en utilisant des exagérations; la propagande est également efficace pour l'anglais là où ce sont des phrases partisans et contenant des préjugés qui sont au même niveau pour

le turc (tous les styles combinés aussi). Cependant, les modèles allemands basés sur des ensembles de données avec styles de subjectivité ne semblent pas avoir obtenu de meilleurs résultats que ceux entraînés sur l'ensemble sans style. Après avoir analysé la qualité des échantillons créés, il leur est apparu que les modifications apportées aux données allemandes étaient plus subtiles, ce qui expliquerait la similarité du score avec le modèle sans style.

3 Modèle de représentation de texte et langues

3.1 Plongement de mots

Afin d'améliorer l'apprentissage des modèles, les chercheurs et les ingénieurs utilisent du plongement de mots pour obtenir une représentation vectorielle des mots. Le plongement de mots, ou "word embeddings" est une technique fondamentale dans le traitement du langage naturel (NLP), fournissant une représentation vectorielle dense et significative des mots; qui capture leur sémantique et leurs relations contextuelles [8]. Dans le contexte de la détection de la subjectivité, ces représentations vectorielles offrent une base riche pour distinguer les nuances et les caractéristiques spécifiques des textes subjectifs par rapport aux objectifs de la tâche. En convertissant les mots en vecteurs, les modèles d'apprentissage automatique peuvent efficacement traiter et analyser de grandes quantités de données textuelles pour identifier des schémas qui indiquent la présence de subjectivité, tels que l'expression d'opinions, d'émotions ou de jugements personnels [1]. Dans cette section nous présentons un résumé sur les plongements de mots souvent utilisés.

Glove GloVe (Global Vectors for Word Representation) est un modèle de plongement de mots qui combine les avantages de deux approches principales de la modélisation sémantique : la factorisation de matrice et les modèles contextuels locaux. Conçu pour capturer à la fois les statistiques globales d'un corpus et les informations contextuelles locales, GloVe est entraîné sur des matrices de cooccurrence des mots, reflétant la fréquence à laquelle chaque paire de mots apparaît ensemble dans un contexte donné. Cette méthode permet à GloVe de saisir des nuances de signification et de relation entre les mots, offrant une représentation dense efficace pour diverses tâches de traitement du langage naturel. De plus, contrairement à d'autres modèles de plongement, GloVe ne repose pas uniquement sur les contextes locaux immédiats mais intègre des informations sur l'ensemble du corpus. [8, 17].

Word2Vec Word2Vec se base sur l'hypothèse distributionnelle, selon laquelle les mots qui apparaissent dans des contextes similaires tendent à avoir des significations similaires. L'algorithme parcourt le corpus de texte et pour chaque mot, il définit un contexte autour de ce mot. Ce contexte est défini par une fenêtre de mots voisins.

Word2Vec utilise deux architectures principales : Skip-gram et CBOW (Continuous Bag of Words). Dans l'architecture Skip-gram, le modèle prend un mot en entrée et tente de prédire les mots environnants (contexte), tandis que dans CBOW, le modèle prend le contexte comme entrée et tente de prédire le mot cible. Pendant l'apprentissage, Word2Vec ajuste les poids du réseau neuronal pour maximiser la probabilité de prédiction des mots dans le contexte donné, comme décrit dans la figure ci-dessous.

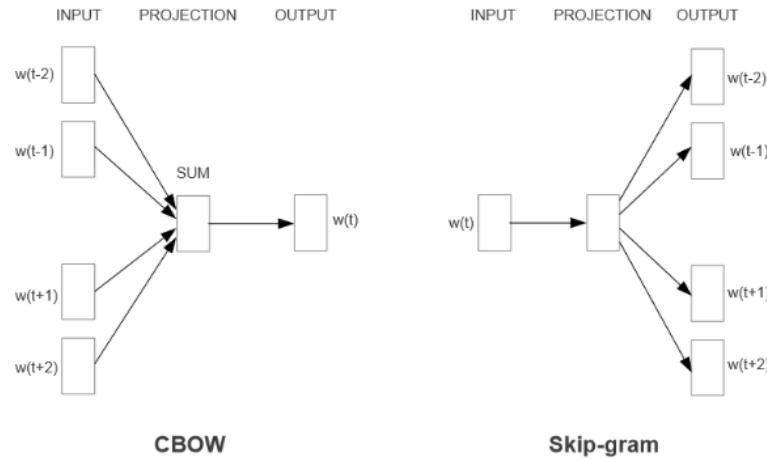


Fig. 2. Composition du Word2Vec cf. [18]

Une fois l'apprentissage terminé, chaque mot du vocabulaire est associé à un vecteur de nombres réels de taille fixe, appelé vecteur de mots. Les valeurs de ces vecteurs sont apprises pendant l'entraînement et capturent les relations sémantiques et syntaxiques entre les mots. Les mots similaires seront plus proches dans l'espace vectoriel, mesuré généralement par la similarité cosinus entre les vecteurs.

Transformers Un transformers est une architecture introduite fin 2018 par Vaswani et al. (2017). Le transformers peut être utilisé pour le pré-entraînement des modèles de langue. Le mécanisme d'auto-attention multi-têtes nouvellement introduit du Transformers [32] permet à chaque mot d'observer tous les mots précédents ou tous les mots à l'exception de la cible, ce qui permet au modèle de capturer efficacement les dépendances à longues distance dans le coût computationnel élevé des calculs récurrents dans les LSTM. [19]. Plusieurs modèles ont émergé du transformer comme GPT, BERT...

BERT BERT (Bidirectional Encoder Representations from Transformers) est une avancée significative dans le domaine des plongements de mots et de phrases. De plus, Bert a révolutionné la compréhension du langage naturel (NLP) depuis son introduction en 2018 par Devlin et al [20]. Contrairement à des modèles précédents, BERT est conçu pour fonctionner de manière bidirectionnelle, permettant ainsi au modèle d'examiner le contexte d'un mot à la fois à gauche et à droite dans une phrase. Cela est essentiel pour saisir les nuances et les sens contextuels des mots qui sont fondamentaux dans la détection de la subjectivité, où la signification peut changer radicalement en fonction du contexte environnant. Ce modèle est pré-entraîné à des tâches en NLP, nous détaillerons ce pré-entraînement dans la section suivante.

3.2 BERT, LLM et variantes multilingues

BERT Comme expliqué dans une partie précédente, on utilise BERT encodeur pour le plongement de mot. Cependant on peut utiliser le BERT pré-entraîné pour la représentation de texte. BERT (Bidirectional Encoder Representations from Transformers) [20] s'appuie sur l'architecture des transformers pour offrir une représentation contextuelle profonde des mots et des phrases. Contrairement aux approches antérieures qui traitent le texte de manière unidirectionnelle, BERT utilise un modèle de langage masqué (MLM) pour apprendre à prédire l'identifiant d'un mot masqué en se basant sur le contexte fourni par les mots environnants, à gauche comme à droite. Cette méthode permet à BERT de comprendre le contexte d'un mot dans toutes les directions, ce qui se traduit par une compréhension plus nuancée du langage.

BERT est disponible en deux variantes principales, adaptées à différents besoins et capacités de calcul :

- BERT Base, qui est plus compact et conçu pour des opérations moins gourmandes en ressources
- BERT Large, qui offre une capacité de traitement plus élevée pour les tâches de traitement de texte complexes, nécessitant plus de puissance de calcul mais fournissant des performances supérieures. (REF)

Ces modèles peuvent être utilisés pour encoder des phrases entières, extrayant des plongements qui intègrent une compréhension profonde du contexte, ce qui est particulièrement bénéfique dans des domaines tels que la compréhension de texte, la traduction automatique, et d'autres applications avancées en NLP.

LLM Également, les grands modèles de langue ou LLM sont souvent utilisés pour effectuer des représentations de texte pour réaliser leur tâche. Les LLM sont basés sur les réseaux neuronaux [21].

Une bonne majorité des LLM actuels sont issues des transformers. On peut en distinguer 3 types d'utilisation du transformer [19]. Les modèles autorégressifs de langage, qui prédisent le prochain mot en fonction des mots précédents, utilisent la partie décodeur du transformer. Les modèles de langage masqué, qui vont

prédire le mot masqué en fonction de tous les mots de la séquences (avant ou après), utilisent eux la partie encodeur du transformer. Les modèles BERT sont de ce type. Enfin, nous avons les modèles de langue encodeur-décodeur. Ce type de modèle utilise un ensemble des deux modèles précédents pour rendre les modèles plus flexibles. Ce type de modèle prend un texte en entrée et renvoie un texte en sortie. La figure ci-dessous explique les modèles sous forme de schéma.

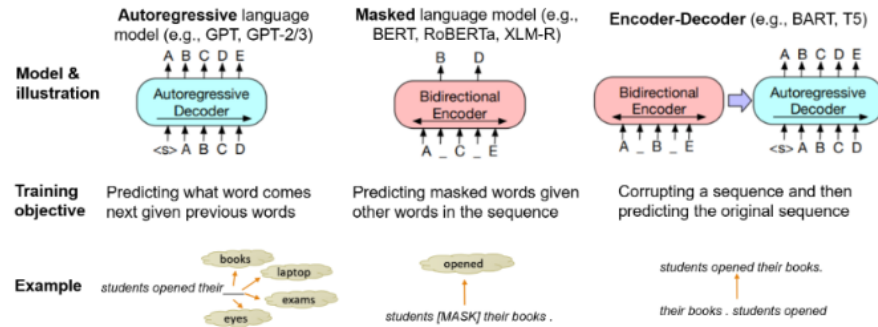


Fig. 3. Les 3 types de larges langages modèles cf. [19]

Il existe plusieurs autres LLMs issue de l'architecture transformer :

- Les modèles GPT distribués par OpenAI, utilisent le mécanisme d'auto-régressions en changeant l'ordonnancement du transformer en réalisant en premier l'étape de normalisation plutôt qu'à la fin. [22].
- Introduit par le groupe Meta, les modèles LLaMa et LLaMa 2 [23, 24] sont basés sur la même architecture, s'inspirant de nouveautés introduites par différents modèles comme GPT ou PaLM, mais se différenciant par la quantité de fractions de mots, mots ou suite de mots, (usuellement tokens en anglais), utilisés pour leur entraînement. Ces modèles utilisent aussi la prénormatisation, la fonction d'activation ReLU est aussi modifiée pour SwiGLU, plus performant et où les positions des plongements de mots sont modifiées, au lieu d'être utilisé directement, elles sont encodées avec les tokens au travers d'une matrice de rotation. Ce qui permet de garder une meilleure attention des tokens distants en incorporant les positions dans les tokens [25].
- Développés par MistralAI, les modèles Mistral 7B et Mixtral 8x7B ont modifié la façon dont ils lisent des phrases en créant une fenêtre glissante d'attention où chaque token d'une couche ne peut accéder qu'à une fenêtre, un nombre W de jeton de la couche précédente, et donc glissante, car à chaque couche la fenêtre se décale ce qui permet d'avoir de bonnes performances avec des modèles plus rapides et compacts. [26] Le modèle Mixtral remplace la couche feed-forward du transformers par une couche qui va servir à router les tokens vers un modèle parmi un ensemble disponible

[27]. Chaque modèle étant entraîné sur différents tokens, ces modèles en plus de celui réalisant le routage sont entraînés ensemble ce qui permet de faire en sorte que chaque modèle produise une sortie différente et que le routeur amène l'entrée vers les modèles qui produisent la meilleure sortie [28].

Il n'y a encore aucun travail sur la détection de subjectivité avec les LLMs à notre connaissance. Cependant, il existe des travaux sur l'analyse de texte avec ces modèles. On peut donc facilement en conclure qu'on peut utiliser les LLMs dans la détection de subjectivité.

Prompt Pour faire exécuter un LLM, nous avons besoin de lui envoyer des invites de commande ou prompt. Une invite de commande est une formulation d'invites aux LLMs pour effectuer des tâches, L'invite dit aux LLMs comment analyser le texte. C'est très important et obligatoire pour guider le LLM pour la réponse. Il y a plusieurs aspects qui doivent être pris en compte pour la formulation d'une demande [21]:

- Nous devons définir clairement la tâche. Déterminez les informations spécifiques que nous souhaitons extraire.
- Déterminer le résultat souhaité
- Prendre en compte la longueur du prompt et sa spécificité
- Inclure des invites ou un contexte afin de guider le LLM
- Répéter et tester, afin de trouver le meilleur prompt avec les meilleurs résultats

4 Approches non neuronales pour la détection de la subjectivité

4.1 Apprentissage automatique

Depuis les années 1960, les méthodes traditionnelles de classification de texte ont joué un rôle clé dans le domaine de la détection de la subjectivité. Parmi ces modèles, le Bayésien Naïf est un modèle simple mais puissant, qui est particulièrement apprécié pour sa capacité à fonctionner sous l'hypothèse d'indépendance entre les attributs, facilitant ainsi la modélisation malgré la corrélation potentielle entre les caractéristiques. Les K-plus proches voisins (KNN) offrent une approche basée sur la proximité où la classification d'un élément est déterminée par la majorité des classes de ses voisins les plus proches, permettant une certaine flexibilité et adaptabilité. Les séparateurs à vaste marge (SVM) sont réputés pour leur robustesse dans des espaces de grande dimension, identifiant l'hyperplan qui maximise la marge entre les classes de données. Les arbres de décision, avec leur structure hiérarchique simplifiant la prise de décision en séparant l'espace des attributs en régions homogènes, fournissent une méthode intuitive et facilement interprétable pour la classification. Enfin, les forêts aléatoires améliorent la précision et contrôlent le surapprentissage en combinant plusieurs arbres de décision lors du processus de prise de décision. Ces méthodes ont

établi les fondements sur lesquels la classification de texte a évolué, permettant des avancées significatives dans la détection de la subjectivité avant l'arrivée des approches basées sur les réseaux de neurones. [29].

Feature extraction Souvent la détection de subjectivité inclut l'analyse de sentiments, l'extraction de caractéristiques d'une langue lié à la polarité de l'opinion d'un texte est ainsi pertinente pour détecter la subjectivité.[5]. Les caractéristiques extraites permettent de transformer des documents en vecteurs, lesquels sont utilisés par différentes stratégies d'apprentissage automatique pour en ressortir l'opinion du contenu.

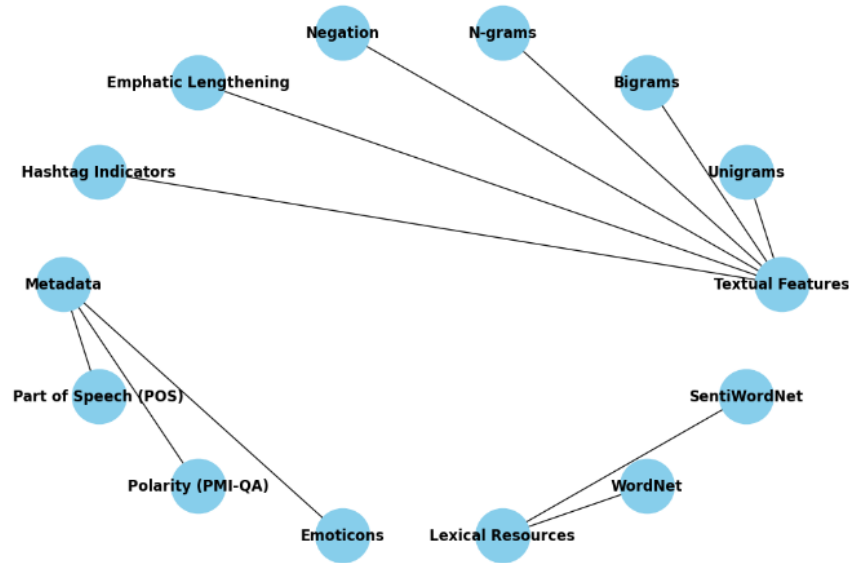


Fig. 4. Familles de caractéristiques exploitables en NLP

Parmi ces caractéristiques [5], on trouve des éléments de métadonnées comme la partie du discours (POS), la polarité via des métriques spécifiques (par exemple, PMI-QA), la présence d'émoticônes, ainsi que l'utilisation de ressources lexicales telles que WordNet et SentiWordNet[5]. Les caractéristiques textuelles incluent des unigrammes, bigrammes, et d'autres n-grammes, qui capturent la fréquence et la co-occurrence de termes dans les textes. D'autres études intègrent des aspects tels que la négation, l'allongement emphatique de mots, et des indicateurs de polarité issus de hashtags, soulignant ainsi l'importance de contextes sociaux et émotionnels spécifiques à des plateformes comme Twitter. Il peut être efficace de détecter certains mots-clés pertinent et de classer les textes en fonction de la présence de ces mots. Chaque mot qui peut évoquer une

émotion sera pris en compte pour déterminer l’objectivité du texte. La principale faiblesse de la détection de mots-clés réside dans l’obtention de ces mots-clés, souvent donnés par des avis experts. [6].

Une autre caractéristique intéressante à extraire peut être l’affinité lexicale. Elle attribue à des mots une affinité probabiliste pour une catégorie particulière. Par exemple, la proposition “je suis allé me promener dans le parc” contient les concepts “aller promenade” et “aller parc”. De tels concepts peuvent être plus facilement classifiés comme positifs ou négatifs. Cependant, il ne permet pas de détecter les sentiments neutre (objectif). Pour résoudre cela, ils ont remarqué que les mots neutres ont une plus proche affinité que les mots à émotion [6]. La détection de subjectivité au niveau des phrases a été intégrée dans la détection de la subjectivité au niveau des documents en utilisant les coupes minimales de graphe sur les phrases. En utilisant les contraintes contextuelles entre les phrases dans un graphe, on pourrait obtenir de meilleurs résultats sur la polarité résultante.

4.2 Graphe

En 2004, Bo Pang et Lillian Lee [4] ont développés une technique de classification de subjectivité afin de synthétiser des revues de film (qu’ils ont ensuite catégorisé en positive ou négative). Leur méthode suppose qu’ils ont accès à deux informations essentielles: $indj(x)$, l’estimation que la phrase x soit dans la classe j et $assoc(xi, xk)$, l’estimation que xi et xk soient dans la même classe. La manière utilisée pour calculer ces deux informations, ainsi que d’autres propositions sont consultables dans leur travaux de recherche. Afin d’éviter de rencontrer un problème insoluble à cause du nombre de possibilités, ils ont modélisé ce problème par un graphe.

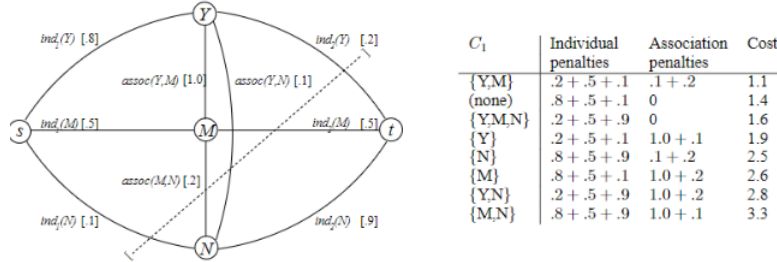


Fig. 5. Exemple de graphe modélisé avec 3 phrases cf. [4]

La source et le puits représentent respectivement chaque label, chaque phrase y sont reliées par leur score $indj$. Et enfin, chaque donnée est reliée aux autres par son $assoc(xi, xk)$, avec xi E classe 1 et xk E classe 2. L’objectif est de pouvoir classer les phrases dont l’appartenance à une classe est trop floue; cette

modélisation à donc pour but de pénaliser le fait que deux phrases proches l'une de l'autre soient classées différemment. Avec ce modèle, il est possible d'appliquer les algorithmes de gestion de flot maximum qui s'exécutent avec des temps d'exécution asymptotiques polynomiaux. Ce modèle s'inscrivant dans un objectif d'analyse de sentiments, les scores de ce modèle ne sont pas détaillés mais simplement leur impact positif sur la qualité du modèle final.

5 Approches neuronales pour la détection de la subjectivité

5.1 CNN

En 2011, les chercheurs ont commencé à utiliser les réseaux de neurones convolutionnels (CNN) car ils étaient considérés comme plus efficace que les anciennes méthodes. Les filtres dans les couches supérieures du CNN peuvent capturer les relations syntaxiques entre les phrases éloignées dans la phrase en paramètre d'entrée.

Certains auteurs ont créé un modèle CNN permettant de prendre en compte le sentiment du texte, souvent ignoré dans les techniques précédentes. Et pour finir d'autre chercheurs présentent un réseau neuronal convolutionnel hiérarchique (HCNN) qui permet d'appliquer les convolutions à chaque caractéristiques des vecteurs de mots dans la phrase [6].

5.2 GRU

Comme montré par [2] et [30], les GRU (Gated Recurrent Unit), que ce soit de manière uni ou bi directionnelle, sont parfois utilisées dans la tâche de détection de subjectivité. Il s'agit d'une architecture basée sur les réseaux de neurones récurrents (RNN), initialement créé pour la traduction automatique de texte [31]. Ces modèles nécessitent une représentation vectorielle d'un mot ou d'une phrase pour pouvoir en faire la classification. Pour cette dite représentation, le modèle utilisé par ces auteurs a été GloVe entraîné préalablement dans la langue voulue. Cependant, même si les résultats obtenus dans ces deux études semblent prometteurs comme on peut le voir graphiquement, ils sont systématiquement dépassés par des modèles basés uniquement sur l'attention tels que BERT.

Les transformers, modèle basé sur l'attention [32], semblent permettre une représentation plus précise du sens des phrases représentées. Cependant, la tâche de représentation ne semble pas suffisante quant à la classification d'une phrase telle que subjective ou objective. À cet effet, une seconde couche est nécessaire en sortie du transformer pour effectuer la classification. Usuellement, une simple couche dense est utilisée.

Table 2
Comparison of F1 scores between the proposed model, the baseline, and other related works on T2.

| Model | F1(%) | Model | P-SUBJ | R-SUBJ | F1-SUBJ | Accuracy | F1-macro |
|------------------------------|--------------|------------------|-------------|-------------|-------------|-------------|-------------|
| Baseline | 30.97 | GRU | 0.46 | 0.73 | 0.56 | 0.63 | 0.62 |
| Bi-LSTM-CRF (word2vec) | 66.32 | MultilingualBERT | 0.62 | 0.67 | 0.64 | 0.76 | 0.73 |
| Bi-LSTM-CRF (fastText) | 69.98 | AIBERTv | 0.62 | 0.65 | 0.63 | 0.75 | 0.72 |
| CNN-BGRU-CRF (word2vec) | 69.44 | | | | | | |
| CNN-BGRU-CRF (fastText) | 70.67 | | | | | | |
| Attention-Based Neural Model | 72.83 | | | | | | |

Fig. 6. Comparaison du GRU avec d'autres modèles (cf.[30] pour le tableau de gauche et [2] pour le tableau de de droite)

5.3 BERT+LSTM et CNN+LSTM

Cependant de nombreux travaux montrent que d'autres approches peuvent être utilisées. Kaliyar et al [8] ont montré qu'il est possible d'utiliser différentes couches de convolution entre un modèle de représentation d'une phrase, de type BERT ou GloVE appelé "Embedding", et une couche dense devant effectuer la classification, comme visible sur la figure ci-contre.

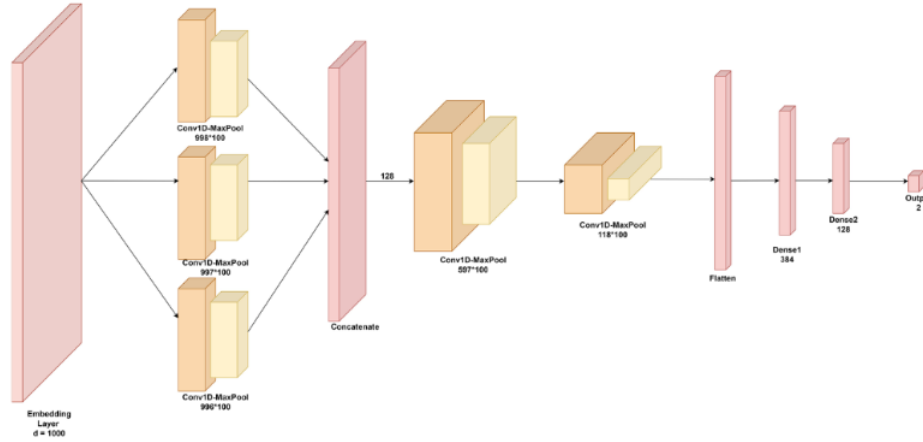


Fig. 7. Architecture du modèle FakeBERT cf.[8]

Le FakeBERT se compose de 5 couches:

- Couche de convolution chargée d'améliorer la représentation sémantique des mots de longueur différente.
- Couche de mise en commun maximale chargée de réduire la taille de la représentation des mots afin d'optimiser le nombre de calculs.
- Couche d'aplanissement chargée de modifier le format des données.
- Deux couches denses chargées d'effectuer la classification finale.

Ce modèle ayant une précision de 0,989 est très performant pour une tâche de classification. Il est un bon candidat pour l'adapter à tâche mais il faudrait

Table 21 Classification results with BERT and GloVe

| Word embedding model | Classification model | Accuracy (%) |
|-------------------------------------|--------------------------------------|--------------|
| TF-IDF (using unigrams and bigrams) | Neural Network | 94.31 |
| BOW (Bag of words) | Neural Network | 89.23 |
| Word2Vec | Neural Network | 75.67 |
| GloVe | MNB | 89.97 |
| GloVe | DT | 73.65 |
| GloVe | RF | 71.34 |
| GloVe | KNN | 53.75 |
| BERT | MNB | 91.20 |
| BERT | DT | 79.25 |
| BERT | RF | 76.40 |
| BERT | KNN | 59.10 |
| GloVe | CNN | 91.50 |
| GloVe | LSTM | 97.25 |
| BERT | CNN | 92.70 |
| BERT | LSTM | 97.55 |
| BERT | Our Proposed model (FakeBERT) | 98.90 |

Fig. 8. Comparaison du modèle FakeBERT avec d'autres modèles cf.[8]

réexaminer et potentiellement reconfigurer le modèle pour qu'il puisse distinguer correctement entre subjectivité et objectivité. Les caractéristiques linguistiques qui déterminent la subjectivité peuvent être subtiles et différer de celles utilisées pour juger de la véracité. Par exemple, le modèle devra être affiné pour reconnaître les opinions, l'ironie et le sarcasme, et les expressions discriminantes d'après la définition de la subjectivité de CheckThat! [1] plutôt que la présence de faits vérifiables.

L'utilisation de couches dites "Long Short Term Memory" (LSTM) pour effectuer la classification est ici intéressante. Ce type d'architecture est une variante des RNN, modifiée afin pallier le problème de disparition du gradient [33]. Il s'agit là d'une piste explorée ailleurs, notamment par [34]. Même si la tâche du modèle décrit ici est la détection de sarcasme et non de subjectivité, la proximité de ces deux concepts nous amène à penser que ces travaux puissent être utilisables pour la tâche décrite dans ce papier.

5.4 Ajustement fin de LLM

L'ajustement fin (fine-tuning) d'un LLM implique le réentraînement d'un modèle préexistant, et pré-entraîné sur vaste ensemble de données, en utilisant un ensemble de données plus petit et spécifique à un domaine. Cette approche est significative car l'entraînement d'un grand modèle de langue à partir de zéro est extrêmement exigeant en termes de puissance de calcul et de temps. Utiliser les connaissances existantes intégrées dans le modèle pré-entraîné permet d'atteindre des performances élevées sur des tâches spécifiques avec des besoins considérablement réduits en données et en calcul [19]. Ce processus applicable à la détection de subjectivité est décrit dans la section suivante par rapport aux recherches courantes. A ce jour, il n'existe que des travaux sur BERT et GPT sur la détection de subjectivité. Cependant, on peut facilement en conclure qu'on peut utiliser les LLMs

(comme Llama et Mistral) dans la détection de subjectivité car ils sont utilisés dans l’analyse de texte en général (par exemple la détection de sentiment).

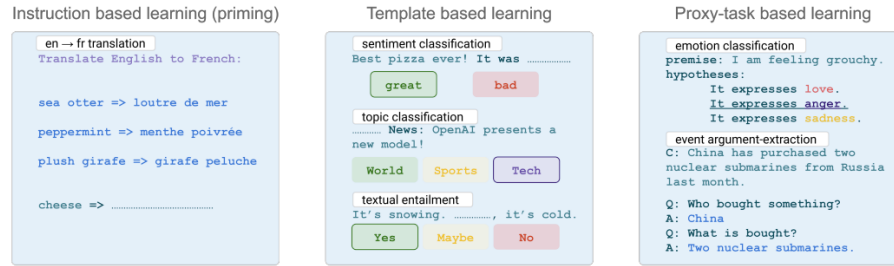


Fig. 9. 3 types d’apprentissage cf.[19]

Comme montrer graphiquement, il existe 3 type d’apprentissage [19]:

- L’apprentissage basé sur les invites (Instruction based learning) : Cette méthode consiste à enseigner au modèle des tâches variées en utilisant des invites directes comme "traduire X en Y". Par exemple, pour une tâche de résumé, on pourrait donner l’invite "TL;DR" qui veut dire "Too long didn’t read" pour que le modèle produise un résumé du texte suivant l’invite. Ces approches nécessitent souvent de grands modèles de langage et une quantité importante de données d’entraînement. L’invite fournit un contexte explicite pour le modèle sur ce qu’il doit faire.
- L’apprentissage basé sur les patrons (Template-based learning) : Cette méthode réformule les tâches de PNL pour les rapprocher des tâches de pré-entraînement des modèles en utilisant des patrons avec des emplacements à compléter. Par exemple, pour la classification de sentiment, on pourrait avoir un patron où le modèle complète les blancs pour exprimer un sentiment. Cette méthode exploite les connaissances acquises lors du pré-entraînement et peut réduire le nombre d’exemples spécifiques nécessaires pour atteindre de bonnes performances.
- L’apprentissage basé sur les tâches proxy (Learning from proxy tasks): Ici, des tâches de compréhension du langage naturel supervisées sont utilisées comme intermédiaires. Par exemple, la classification des émotions ou l’extraction d’événements pourrait être reformulée en tâches de questions-réponses ou d’inférence textuelle où le modèle doit remplir un patron basé sur un prompt. Cela permet d’obtenir de meilleures performances en zero-shot ou en few-shot learning, même avec des modèles de langage de taille plus réduite.

Dans le contexte de la détection de la subjectivité, ces méthodes peuvent être utilisées pour entraîner des modèles à reconnaître des éléments subjectifs dans le texte, tels que les opinions, sentiments ou biais, souvent en formulant la tâche

comme une classification, une inférence ou un remplissage de patron basé sur des invites ou des données exemples.

5.5 Les travaux sur la détection de subjectivité

Comme expliqué précédemment, BERT encodeur est utilisé pour le plongement de mot. Cependant, BERT fine-tuned peut être utilisé pour réaliser la tâche de la détection de subjectivité. Comme il a été montré par les équipes de CheckThat! [12] [16] [35], BERT fine-tuned est très efficace avec des très bon score en anglais.

Ce modèle a aussi permis de créer de nouveaux modèles comme RoBERTa [36]. Ce modèle est entraîné sur plus de données et plus longtemps que BERT. De plus, RoBERTa supprime l'objectif de BERT sur la prédiction de la phrase suivante. Et pour finir, ils appliquent des changements dynamiques du motif de masquage appliqué aux données d'entraînement. Ce qui permet de rendre ce modèle plus efficace que son prédécesseur, comme on peut le voir avec des équipes CheckThat! [12], [16] qui ont un plus grand score que ceux qui utilisent BERT simple.

Le problème est que RoBERTa se concentre que sur une langue, pour remédier à cela des chercheurs ont créé le modèle XML-RoBERTa. Ce modèle est la version multiLangue de RoBERTa. Selon une équipe CheckThat! [37], elle a eu de très bon résultats dans les langues autre que l'anglais.

Durant ces dernières années, beaucoup de modèles avec des réseaux neuronaux ont émergé notamment GPT, ou cette version 3 a été utilisée pour la détection de subjectivité. Plusieurs équipes CheckThat! [16], [14] ont utilisé ce modèle. Dont l'un deux a été premier dans la langue anglais mais ce modèle dépend des invites de commandes qu'on lui envoie.

6 Limites et perspectives

6.1 Limites

Traduction pour l'augmentation de données La traduction peut introduire des biais et des inexactitudes. L'équipe Accenture a noté que cela pouvait altérer le ton de la phrase, la rendant plus objective ou subjective que l'intention originale. [12] De plus, l'utilisation de modèles de génération de texte comme GPT peut entraîner les mêmes problèmes. En effet, l'utilisation de GPT crée des problèmes de styles, ou les phrases générées diffèrent des données d'origine, ce qui affecte la qualité de l'entraînement [16].

Apprentissage automatique Les modèles d'apprentissage automatique qui ne sont pas des réseaux de neurone requièrent essentiellement un processus de sélection et d'extraction de caractéristiques qui est généralement coûteux et long si fait manuellement sans outil pré-faits, cela peut impliquer des hypothèses qui ne sont pas toujours valides dans différents contextes ou langues. Ce processus peut nécessiter également un ajustement continu à l'arrivée de nouvelles données, ce qui limite la généralisation du modèle [29].

Approches neuronales classiques Les réseaux de neurones convolutionnels (CNN) sont en général peu efficaces pour gérer des dépendances à longue distance dans un texte. Ils sont reconnus pour leur capacité à reconnaître des motifs locaux mais pas dans la compréhension globale de séquences de données, ce qui est impératif pour déterminer l’objectivité ou la subjectivité globale d’une phrase.

Les Unités Récurrentes Fermées permettent généralement une évaluation et représentation efficace des phrases, notamment lorsque elles sont couplées à d’autres couches. Moins coûteux en paramètre que les LSTM, ses performances sont souvent inférieures. Il s’agit tout de même d’une piste intéressante lorsque le lot de données est faible.

Les réseaux LSTM ont été conçus pour éviter le problème de disparition du gradient inhérent aux RNN classiques mais leur performance peut être surpassée par les architectures basées sur l’attention comme les Transformers, notamment dans des contextes où la compréhension du contexte global est la base pour la détection de la subjectivité. De plus, ils ont un nombre de paramètres bien plus élevé qu’un RNN classique, faisant de leur apprentissage une tâche plus lourde nécessitant un lot de données plus important.

6.2 Perspectives

Au vu des résultats obtenus par certaines équipes de recherches qui ont appliqué une architecture utilisant BERT et des couches neuronales supplémentaires comme CNN ou LSTM, il serait probablement pertinent, à l’avenir, d’avancer la recherche vers une architecture similaire adaptée pour détecter la subjectivité.

En outre, l’ajustement fin de modèles de langage pré-entraînés open-sources comme BERT, Mistral, LLaMA, BLOOM ou d’autres devient une pratique de plus en plus courante dans la communauté de recherche comme en témoignent de nombreuses discussions et articles diffusés sur des plateformes de blog telle que Medium. Ces modèles ré-entraînés pour des tâches spécifiques, permettent une accélération de la recherche en NLP, et une collaboration fluide à travers la communauté scientifique grâce aux re-publications de modèles. En poursuivant dans cette voie, la détection de la subjectivité peut devenir plus accessible, précise et efficace, alimentant le développement de systèmes intelligents capables de comprendre la subjectivité de l’humain.

7 Conclusion

Pour conclure cet état de l’art sur la détection de la subjectivité dans le traitement du langage naturel, il est important de rappeler l’importance croissante de cette tâche dans un monde numérique en rapide évolution. Comme souligné dans notre introduction, la distinction entre faits objectifs et opinions subjectives est requise pour diverses applications, allant de la modération de contenus sur les réseaux sociaux à l’analyse des sentiments dans les critiques de produits. Cette nécessité se reflète également dans les efforts pour lutter contre la désinformation

et promouvoir l'intégrité du contenu dans la presse, influençant ainsi l'opinion publique de manière significative.

Dans le contexte du traitement de langage naturel, beaucoup de méthodes ont été réalisées pour détecter la subjectivité.

Dans cet état de l'art, nous avons examiné des approches et des avancées significatives dans la détection de subjectivité, allant de méthodes traditionnelles comme l'apprentissage automatique à des méthodes plus avancées comme les architectures neuronales. Nous avons expliqué l'importance de la représentation textuelle pour pouvoir utiliser ces modèles et nous avons discuté de techniques d'augmentation de données pour remédier aux problèmes liés aux déséquilibres dans les jeux de données ainsi que l'ajustement des grands modèles de langue à notre tâche brièvement.

L'intérêt de ce travail repose sur la compilation et l'analyse approfondie des différentes stratégies employées pour la détection de la subjectivité, fournissant ainsi un aperçu détaillé des forces et des limites de chaque approche. Enfin, l'application des dernières avancées en traitement automatique des langues a été mise en évidence à travers les grands modèles de langue, montrant l'évolution et l'adaptabilité des techniques face aux enjeux de classification textuelle actuels.

References

1. Ruggeri, F., Antici, F., Galassi, A., Korre, K., Mutti, A. et Barron-Cedeno, A. : On the Definition of Prescriptive Annotation Guidelines for Language-Agnostic Subjectivity Detection, 103–111 (2023)
2. Antici, F., Bolognini, L., Inajetovic, M.A., Ivasiuk, B. : SubjectivITA: An Italian Corpus for Subjectivity Detection in Newspapers. Springer International Publishing, International Conference of the Cross-Language Evaluation Forum for European Languages, (2021), 40–52
3. Aker, A., Gravenkamp, H., Mayer, S.J. et Hamacher, M. : A Corpus for Sentence-level Subjectivity Detection on English News Articles, 1–6 (2019)
4. Pang, B. et Lee, L. : A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts. Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics (ACL-04), 271–278 (2004)
5. Sekharan, S.C. : Subjectivity Detection for sentiment analysis on Twitter Data. Artificial Intelligence Techniques for Advanced Computing Applications, 467–476 (2020)
6. Chaturvedi, I., Cambria, E., Welsch, R.E. et Herrera, F. : Distinguishing between facts and opinions for sentiment analysis: Survey and challenges, 65–77 (2018)
7. Galassi, A., Ruggeri, F., Barron-Cedeno, A., Alam, F., Caselli, T., Kutlu, M., Strub, J.M., Antici, F., Hasanain et M. Kohler, J. : Overview of the CLEF-2023 CheckThat! Lab: 24th Working Notes of the Conference and Labs of the Evaluation Forum, CLEF-WN 2023. In: Working Notes of the Conference and Labs of the Evaluation Forum (CLEF 2023), pp 236–249. Aliannejadi, M., Faggioli, G., Ferro, N. et Vlachos, M. Thessaloniki, Greece (2023)
8. Kaliyar, R.K., Goswami, A. et Narang, P. : FakeBERT: Fake news detection in social media with a BERT-based deep learning approach, 11765–11788 (2021)
9. Pachov, G., Dimitrov, D., Kyochev, I. et Nakov, P. : Gpachov at CheckThat! 2023: A Diverse Multi-Approach Ensemble for Subjectivity Detection in News Articles, 1–9 (2023)

10. Leistra, F. et Caselli, T. : Thesis Titan at CheckThat! 2023: 24th Working Notes of the Conference and Labs of the Evaluation Forum, CLEF-WN 2023, 351–359 (2023)
11. Li, B., Hou, Y. et Che, W.: Data augmentation approaches in natural language processing: A survey. *AI open*; 71–90 (2022)
12. Sieu, T., Rodrigues, P., Strauss, B. et Williams, E.M : Accenture at CheckThat! 2023: Impacts of Back-translation on Subjectivity Detection, 1–11 (2023)
13. Papineni, K., Roukos, S., Ward, T. et Zhu, W.J. : BLEU: a method for automatic evaluation of machine translation, 311–318 (2022)
14. Frick, R.A. : Fraunhofer SIT at CheckThat! 2023: Can LLMs Be Used for Data Augmentation & Few-Shot Classification? Detecting Subjectivity in Text Using Chat-GPT, 1–7 (2023)
15. Shushkevich, E. et Cardif, J. : TUDublin at CheckThat! 2023: ChatGPT for Data Augmentation, 1–9 (2023)
16. Schlicht, I.B., Khellaf, L. et Altiok, D. : DWReCO at CheckThat! 2023: Enhancing Subjectivity Detection through Style-based Data Sampling, 1–12 (2023)
17. Pennington, J., Socher, R. et Manning, C.D. : GloVe: Global Vectors for Word Representation, 1532–1543 (2014)
18. Mikolov, T., Chen, K., Corrado, G.S. et Dean : Efficient Estimation of Word Representations in Vector Space, 1–13 (2013)
19. Min, B., Ross, H., Sulem, E., Veyseh, A.P.B., Nguyen, T.H., Sainz, O., Agirre, E., Heintz, I., Roth, D. : Recent Advances in Natural Language Processing via Large Pre-Trained Language Models: A Survey, 1–40 (2023)
20. Devlin, J., Chang, M.W., Lee, K. et Toutanova, K. : BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pp. 4171–4186. Burstein, J., Doran, C. et Solorio, T., BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding (Devlin et al., NAACL 2019) (2019)
21. Tornberg, P. : How to use Large Language Models for Text Analysis, 1–6 (2023)
22. Brown, T.B, Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D.M., Wu, J., Winter, C., Hesse, C., Chen, M., Sigler, E., Litwin, M., Gray, S., Chess, B., CLark, J., Berner, C., McCandlish, S., Radford, A., Sutskever, I. et Arnadei, D.: Language Models are Few-Shot Learners, 1–75 (2020)
23. Touvron, H., Lavril, T., Izacard, G., Martinet, X., Lachaux, M.A., Lacroix, T., Rozière, B., Goyal, N., Hambro, E., Azhar, F., Rodriguez, A., Joulin, A., Grave, E. et Lample, G.: LLaMA: Open and Efficient Foundation Language Models, 1–27 (2023)
24. Touvron, H., Martin, L., Stone, K., Albert, P., Almahairi, A., Babaei, Y., Bashlykov, N., Batra, S., Bhargava, P., Bhosale, S., Bikel, D., Blecher, L., Ferrer, C.C, Chen, M., Cucurull, G., Esiobu, D., Fernandes, J., Fu, J., Fu, W., Fuller, B., Gao, C., Goswami, V., Goyal, N., Hartshorn, A., Hosseini, S., Hou, R., Inan, H., Kardas, M., Kerkez, V., Khabsa, M., Kloumann, I., Korenev, A., Koura, P.S., Lachaux, M.A., Lavril, T., Lee, J., Liskovich, D., Lu, Y., Mao, Y., Martinet, X., Mihaylov, T., Mishra, P., Molybog, I., Nie, Y., Poulton, A., Reizenstein, J., Rungta, R., Saladi, K., Schelten, A., Silva, R., Smith, E.M, Subramanian, R., Tan, X.E., Tang, B., Taylor, R., Williams, A., Kuan, J.X., Xu, P., Yan, Z., Zarov, I., Zhang, Y., Fan, A., Kambadur, M., Narang, S., Rodriguez, A., Stojnic, R., Edunov, S. et Scialom, T.: Llama 2: Open Foundation and Fine-Tuned Chat Models, 1–77 (2023)

25. Su, J., Ahmed, M., Lu, Y., Pan, S., Bo, W., et Liu, Y.: RoFormer : enhanced transformer with rotary position embedding, 1–14 (2023)
26. Jiang, A.Q., Sablayrolles, A., Mensch, A., Bamford, C., Chaplot, D.S., de las Casas, D., Bressand, F., Lengyel, G., Lample, G., Saulnier, L., Lavaud, L.R., Lachaux, M.A., Stock, P., Le Scao, T., Lavril, T., Wang, T., Lacroix, T. et El Sayed, W. : Mistral 7B, 1–9 (2023)
27. Jiang, A.Q., Sablayrolles, A., Roux, A., Mensch, A., Savary, B., Bamford, C., Chaplot, D.S., de las Casas, D., Hanna, E.B., Bressand, F., Lengyel, G., Bour, G., Lample, G., Saulnier, L., Lavaud, L.R., Lachaux, M.A., Stock, P., Subramanian, S., Yang, S., Antoniak, S., Le Scao, T., Gervet, T., Lavril, T., Wang, T., Lacroix, T. et El Sayed, W. :, F.: Mixtral of Experts, 1–13 (2024)
28. Fedus, W., Dean, J. et Zoph, B. : A review of sparse expert models in deep learning, 1–23 (2022)
29. Li, Q., Peng, H., Li, J. et Xia, C. : A Survey on Text Classification : From Traditional to Deep Learning, 1–41 (2022)
30. Abdelgwad, M.M, A Soliman, T.H, Taloba, A.I et Farghaly, M.F : Arabic aspect based sentiment analysis using bidirectional GRU based models. Journal of King Saud University - Computer and Information Science, 6652–6662 (2022)
31. Cho, K., van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwank, H. et Bengio, Y.: Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp 1724–1734. Moschitti, A., Pang, B. et Daelmans, W., Doha, Qatar (2014)
32. Vaswani, A., Shazeer, N., Parmar, N., Jakob, U., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I. : Attention is all you need, 1–15 (2023)
33. Hochreiter, S.: Long Short Term Memory. Neural Computation, 1735–1780 (1997)
34. Panday, R. et Singh, J.P. : BERT-LSTM model for sarcasm detection in code-mixed social media post. Journal of intelligent information systems, 235–254 (2023)
35. Agirrezabal, M.: KUCST at CheckThat 2023: How good can we be with a generic model?, 99–110 (2023)
36. Liu, Y., Ott, M., Goyal, N. et Du, J. : RoBERTa: A Robustly Optimized BERT Pretraining Approach, 1–13 (2019)
37. Dey, K., Tarannum, P., Hassan, M.A. et Noori, S.R.H.: NN at CheckThat! 2023: Subjectivity in News Articles Classification with Transformer Based Models, 1–11 (2023)