Traitement automatique du langage naturel pour la reconnaissance des émotions : vers des assistants virtuels empathiques

Mémoire présenté

dans le cadre du programme de maîtrise en informatique

en vue de l'obtention du grade de maître ès sciences

PAR

© SEYED HAMED NOKTEHDAN ESFAHANI

août 2025

Composition du jury :

Maxime Berger, président du jury, Université du Québec à Rimouski

Mehdi Adda, directeur de recherche, Université du Québec à Rimouski

Hamid Mcheick, membre externe, Université du Québec à Chicoutimi

Dépôt initial le 12 mars 2025 Dépôt final le 27 août 2025

REMERCIEMENTS

Je tiens à remercier le professeur Mehdi Adda pour son aide précieuse dans mes études en informatique. Sans ses conseils, sa supervision, sa tolérance, ses critiques pertinentes, ses vastes connaissances et son appui, je n'aurais pas pu mener à bien ce mémoire. C'est un honneur de poursuivre une maîtrise à l'UQAR sous sa direction.

Je tiens à exprimer ma profonde gratitude à ma famille pour son soutien pendant mes études.

Enfin, je dédie ce mémoire à mes parents.

RÉSUMÉ

L'objectif de cette recherche est d'explorer la détection et l'intégration des états affectifs des personnes lors des interactions avec des assistants virtuels. Le développement d'assistants virtuels empathiques, capables de comprendre et de répondre aux émotions des utilisateurs, présente un potentiel significatif pour améliorer leur expérience et leur satisfaction. Cette recherche se concentre sur l'étape fondamentale de la détection des émotions à partir des entrées textuelles, en s'appuyant sur des techniques avancées d'apprentissage automatique et des grands modèles de langage (GML ou LLM pour l'abréviation en anglais) pour obtenir une reconnaissance précise des émotions.

Dans cette recherche, nous nous sommes concentrés sur le jeu de données de *International Survey on Emotion Antecedents and Reactions* (ISEAR), qui inclut des données textuelles représentant sept émotions distinctes : la joie, la colère, la tristesse, la honte, la culpabilité, le dégoût et la peur. Le jeu de données ISEAR est une collection exhaustive d'expressions textuelles de ces émotions, fournissant une ressource riche pour l'entraînement et l'évaluation des modèles de détection des émotions. Ce jeu de données capture une large gamme d'états émotionnels, ce qui en fait un choix idéal pour le développement et le test des capacités des assistants virtuels empathiques. En utilisant ce jeu de données, nous avons visé à s'assurer que nos modèles puissent détecter et classer avec précision ces sept émotions clés à partir des entrées textuelles, formant ainsi la base des interactions empathiques avec les assistants virtuels.

L'hypothèse de travail de cette recherche est que les GML avancés, en particulier ceux affinés avec des techniques appropriées, peuvent améliorer significativement la précision de la détection des émotions à partir des textes comparés aux modèles d'apprentissage automatique classiques.

Les résultats de la recherche indiquent que les GML avancés, en particulier le modèle Mistral 7B, affiné, surpassent les modèles d'apprentissage automatique classiques dans la tâche de détection des émotions à partir des textes. Le modèle Mistral 7B, affiné en utilisant des techniques d'ajustement fin (en anglais : fine tuning) efficace des paramètres (AFEP), a atteint une précision de 76 %, ce qui est nettement supérieur à la performance d'autres modèles comme Falcon 7B et des algorithmes classiques tels que le MNB et le SVM. Ces résultats valident l'hypothèse de cette recherche, démontrant la capacité supérieure des GML avancés à capturer et à reconnaître des états émotionnels nuancés à partir des entrées textuelles.

Cette recherche démontre avec succès le potentiel de l'utilisation des GML avancés pour la détection des émotions dans le développement d'assistants virtuels empathiques. Les résultats montrent que le modèle Mistral 7B affiné améliore de manière significative la précision de la détection des émotions, fournissant une base solide pour créer des assistants virtuels capables de comprendre et de répondre aux états émotionnels des utilisateurs.

Mots-clés : assistant virtuel empathique, détection des émotions, analyse textuelle, grand modèle de langage, traitement automatique des langues.

ABSTRACT

The aim of this research is to explore the detection and integration of people's affective states during interactions with virtual assistants. The development of empathetic virtual assistants, which can understand and respond to users' emotions, holds significant potential for enhancing user experience and satisfaction. This research focuses on the foundational step of detecting emotions from text inputs, leveraging advanced machine learning techniques and large language models (LLMs) to achieve accurate emotion recognition.

In this research, we focused on the International Survey on Emotion Antecedents and Reactions (ISEAR) dataset, which includes textual data representing seven distinct emotions: joy, anger, sadness, shame, guilt, disgust, and fear. The ISEAR dataset is a comprehensive collection of textual expressions of these emotions, providing a rich resource for training and evaluating emotion detection models. The dataset captures a wide range of emotional states and scenarios, making it an ideal choice for developing and testing the capabilities of empathetic virtual assistants. By utilizing this dataset, we aimed to ensure that our models could accurately detect and classify these seven key emotions from text inputs, forming the foundation for empathetic interactions in virtual assistants.

The working hypothesis of this research is that advanced LLMs, particularly those fine-tuned with appropriate techniques, can significantly improve the accuracy of emotion detection from text compared to classical machine learning models.

The research findings indicate that advanced LLMs, particularly the fine-tuned Mistral 7B model, outperform classical machine learning models in the task of emotion detection from text. The Mistral 7B model, fine-tuned using PEFT techniques, achieved an accuracy of 76%, notably higher than the performance of other models such as Falcon 7B and classical

algorithms like MNB and SVM. These findings confirm our initial hypothesis that advanced LLMs excel at identifying and interpreting subtle emotional nuances in textual data.

This research successfully demonstrates the potential of using advanced LLMs for emotion detection in developing empathetic virtual assistants. The findings show that the fine-tuned Mistral 7B model significantly improves emotion detection accuracy, providing a robust foundation for creating virtual assistants that can understand and respond to users' emotional states.

Keywords: Empathetic Virtual Assistant, Emotion Detection, Text Analysis, Large Language Models, Natural Language Processing.

TABLE DES MATIÈRES

LISTE DES FIGURES

LISTE DES ABRÉVIATIONS

AV	Assistants Virtuels
AI	Artificial Intelligence
NLP	Natural Language Processing
TAL	Traitement Automatique des Langues
ML	Machine Learning
AA	Apprentissage Automatique
LLM	Large Language Model
GML	Grand Modèle de Langage
BOW	Bag of Words
TF-IDF	Term Frequency-Inverse Document Frequency
SVM	Support Vector Machines
MNB	Multinomial Naive Bayes
GPT	Generative Pre-trained Transformer
BERT	Bidirectional Encoder Representations from Transformers
PEFT	Parameter Efficient Fine-Tuning
AFEP	Ajustement Fin Efficace en Paramètres
QLoRA	Quantized Low-Rank Adaptation of Language Models

INTRODUCTION GÉNÉRALE

CONTEXTE

Les assistants virtuels (AV) sont de plus en plus demandés dans le paysage numérique en constante évolution. Leurs applications sont vastes, soutenant les utilisateurs dans leur productivité personnelle, le service client et d'autres domaines. Bien que les AV soient désormais capables d'exécuter des tâches et de fournir des informations, ils manquent encore un élément essentiel dans leurs interactions : le lien émotionnel. Cette lacune limite leur efficacité dans des scénarios où l'intelligence émotionnelle, l'empathie et la compréhension sont essentielles.

L'empathie joue un rôle crucial dans les interactions humaines, facilitant une communication plus efficace, renforçant la confiance et augmentant la satisfaction de l'utilisateur. Des études ont montré que les échanges empathiques enrichissent l'expérience client et stimulent l'engagement ainsi que la fidélité [1]. Dans le domaine de la santé, par exemple, il a été prouvé qu'une communication empreinte de compassion accroît la satisfaction et améliore les résultats des patients [2]. De même, la recherche de Wei et al [3] met en évidence que des réponses empathiques dans le service client peuvent apaiser les tensions et contribuer à une expérience client plus positive. Ces résultats mettent en lumière l'importance de l'intégration de l'empathie aux AV, en particulier pour les fonctions orientées vers l'utilisateur qui nécessitent une compréhension et une prise en compte des besoins émotionnels.

PROBLÉMATIQUE

Malgré leurs capacités avancées, les AV manquent de connexion émotionnelle, essentielle pour des interactions plus efficaces avec l'utilisateur dans des situations émotionnellement nuancées. L'absence d'empathie dans les AV limite leur capacité à répondre aux états émotionnels des utilisateurs, affectant ainsi la qualité et la satisfaction de l'expérience utilisateur. Étant donné ses nombreux avantages dans les interactions humaines, l'intégration de l'empathie dans les AV semble être la prochaine étape logique. Un AV empathique pourrait reconnaître et répondre de manière appropriée aux états émotionnels des utilisateurs, offrant ainsi un niveau de soutien et de compréhension au-delà de la simple assistance fonctionnelle. De telles capacités pourraient transformer les interactions avec l'utilisateur, les rendant plus humaines et satisfaisantes. Par exemple, un AV empathique dans une application de santé mentale pourrait fournir du réconfort et de l'apaisement, ce qui pourrait encourager les utilisateurs à partager leurs préoccupations plus ouvertement. De plus, en comprenant et en répondant aux frustrations des utilisateurs, un AV pourrait améliorer l'expérience du service client, conduisant à une plus grande satisfaction et un meilleur engagement.

En outre, l'intégration de l'empathie dans les AV représente une étape transformatrice dans l'interaction humain-technologie. À mesure que les AV développent la capacité de reconnaître et de répondre aux émotions humaines, ils peuvent fournir un soutien plus personnalisé et significatif, favorisant un sentiment de connexion plus fort. Cette innovation a le potentiel de rendre les AV plus humains, d'offrir une meilleure assistance et de garder les utilisateurs plus engagés, redéfinissant ainsi nos interactions avec la technologie dans un large éventail d'applications.

OBJECTIFS

L'objectif général de cette recherche est d'explorer et de mieux comprendre les mécanismes de reconnaissance des émotions à partir de textes en s'appuyant sur les avancées en traitement automatique des langues (TAL).

Plus spécifiquement, cette étude poursuit deux objectifs :

Le premier objectif est de fournir un aperçu complet des techniques de TAL, en se concentrant sur celles appliquées à la reconnaissance des émotions à partir du texte. Cette étude aborde l'évolution du TAL, examine ses techniques et méthodologies de base, et explore leurs applications dans l'identification des émotions dans les données textuelles. Elle approfondit également diverses méthodes de reconnaissance des émotions, allant des algorithmes classiques d'apprentissage automatique aux approches avancées d'apprentissage profond, dans le but de cartographier le paysage du TAL dans la reconnaissance des émotions.

Le deuxième objectif s'appuie sur cette base en évaluant l'efficacité des grands modèles de langage (GML) par rapport aux techniques classiques d'apprentissage automatique pour la reconnaissance des émotions dans le texte. Cette étude examine les forces et les faiblesses des différents algorithmes d'apprentissage automatique dans ce domaine et examine la performance comparative des GML par rapport aux méthodes traditionnelles. En outre, elle explore comment l'ordre dans lequel les catégories émotionnelles sont présentées à un GML peut influencer sa précision de reconnaissance, fournissant ainsi un aperçu des pratiques optimales pour les tâches de reconnaissance des émotions.

MÉTHODOLOGIE

Cette étude a exploré deux approches majeures pour la classification des émotions à partir des textes : les techniques classiques d'apprentissage automatique (AA) et les GML. La méthodologie suivie, illustrée dans la Figure 1, commence par une revue de la littérature permettant de définir les bases théoriques nécessaires pour aborder cette problématique. À la suite de cette étape, un travail de sélection des données a été effectué afin d'assurer une représentation adéquate des différentes émotions.

Dans l'approche classique d'apprentissage automatique, plusieurs modèles ont été sélectionnés pour évaluer leur performance sur cette tâche. Parmi ces modèles, nous retrouvons des algorithmes bien établis tels que les modèles bayésiens, les SVM, les arbres de décision, les forêts aléatoires, le Gradient Boosting et les réseaux de neurones. L'efficacité de ces modèles dépend largement des représentations des textes, c'est pourquoi différentes techniques d'extraction de caractéristiques ont été appliquées, notamment l'approche Bag-of-Words, la pondération TF-IDF et des représentations vectorielles plus avancées basées sur des modèles pré-entraînés comme BERT [4] et FastText [5]. Afin de garantir la meilleure configuration pour chaque modèle, une optimisation des hyperparamètres a été réalisée à l'aide de la recherche par grille [6]. Une fois les hyperparamètres déterminés, les modèles ont été entraînés puis évalués pour mesurer leurs performances dans la classification des émotions.

Parallèlement à cette approche, les GML tels que Falcon 7B [7] et Mistral 7B [8] ont été évalués. Deux stratégies complémentaires ont été adoptées dans cette partie du travail. La première repose sur l'apprentissage en une seule fois (*one-shot learning*), où un exemple représentatif est fourni au modèle pour chaque émotion afin qu'il classifie un texte donné en se basant sur ces exemples. Cette approche exploite directement les capacités des modèles pré-entraînés sans nécessiter d'adaptation supplémentaire. La seconde stratégie consiste à affiner le modèle Mistral 7B en utilisant la méthode QLoRA [9], une technique d'adaptation efficace permettant d'ajuster partiellement les paramètres d'un modèle pré-entraîné. Ce processus d'affinage a nécessité une phase d'entraînement spécifique suivie d'une évaluation des performances sur les données sélectionnées.

Enfin, une comparaison systématique des résultats obtenus avec les deux approches a été réalisée afin d'évaluer leurs performances respectives. Des indicateurs d'évaluation standard, tels que l'exactitude, la précision et le rappel, ont permis de déterminer l'efficacité de chaque méthode pour la classification des émotions.

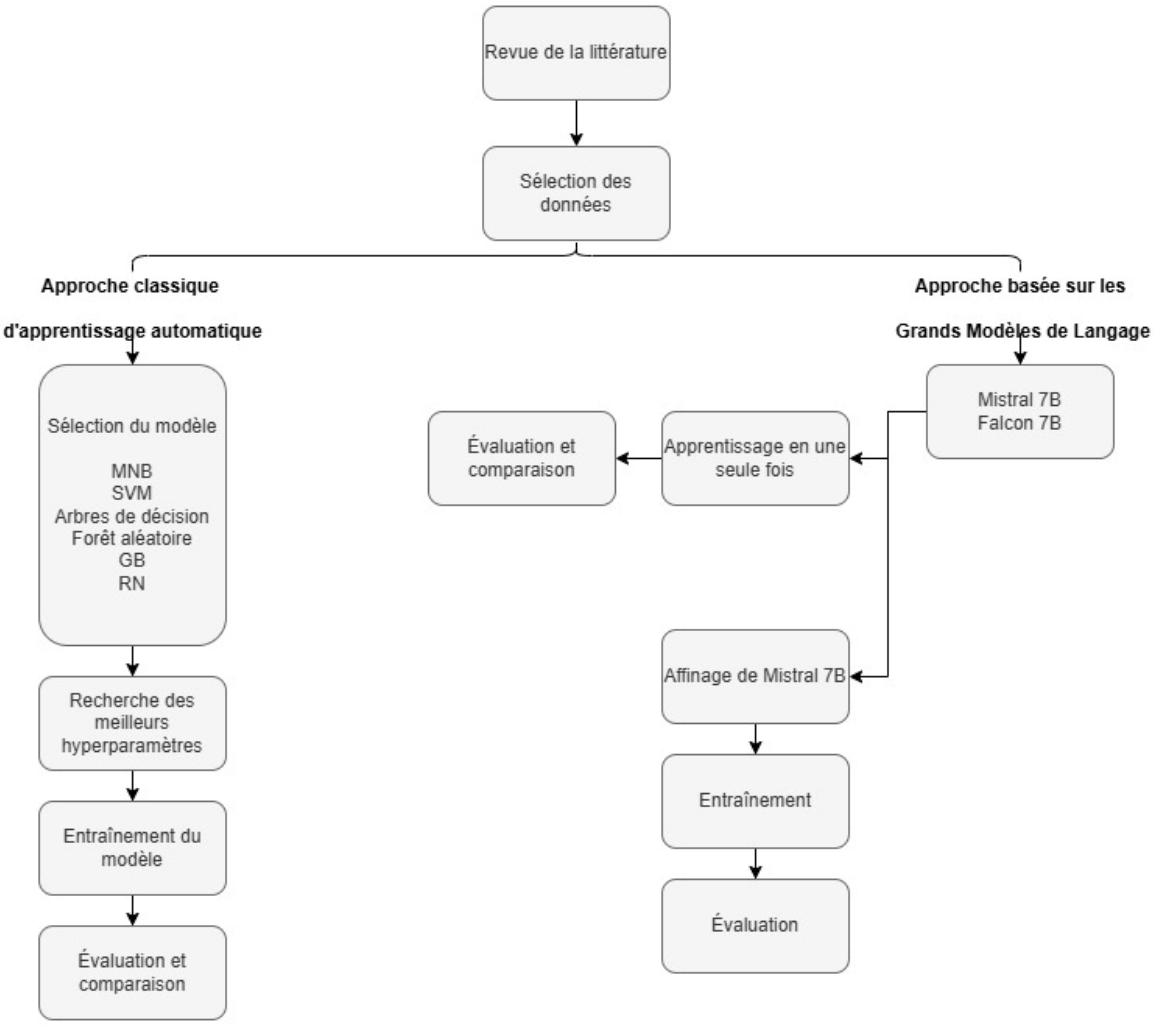

Figure 1. Méthodologie de classification des émotions : approches classiques et Grands Modèles de Langage

CONTRIBUTIONS

Dans cette recherche, nous nous sommes concentrés sur la détection des émotions à partir du texte et avons exploré à la fois les méthodes classiques de AA et les GML pour la reconnaissance des émotions à partir du texte. Nous avons présenté nos résultats lors de la 19^{e} conférence internationale sur les réseaux et communications du futur (FNC 2024),en

comparant différentes approches afin d'identifier les techniques les plus efficaces pour la reconnaissance des émotions dans les AV.. En comparant ces approches, nous avons cherché à identifier les techniques les plus efficaces pour intégrer des capacités de reconnaissance des émotions dans les AV. Cette exploration est essentielle pour développer des AV capables d'interpréter et de répondre avec précision aux nuances émotionnelles des interactions avec l'utilisateur, améliorant ainsi l'expérience utilisateur globale.

ORGANISATION DU MÉMOIRE

Ce mémoire est basé sur deux articles scientifiques :

Le premier est finalisé et s'intitule « Literature review on Natural Language Processing and Emotion Recognition from Text ». Il offre une exploration de l'évolution des techniques de TAL et leur application dans la reconnaissance des émotions à partir de données textuelles. Cette étude discute des défis, des méthodologies impliquées, et met en lumière les techniques de pointe et leurs applications dans divers domaines. Ce premier article constitue le fondement du chapitre 1.

Le second, intitulé « Classical Machine Learning and Large Models for Text-Based Emotion Recognition » [10], a été publié et présenté lors de la 19^{e} conférence internationale sur les réseaux et communications futurs (FNC), tenue du 5 au 7 août 2024, à l'Université Marshall, à Huntington, en Virginie-Occidentale, aux États-Unis. Il analyse et compare l'utilisation de techniques classiques d'apprentissage automatique, comme les SVM, avec des GML avancés tels que BERT, Falcon 7B et Mistral 7B, pour la reconnaissance des émotions dans les textes. Ce travail montre que le modèle Mistral 7B atteint une précision maximale de 76 %, contre 64 % pour SVM. Ce second article forme le socle du chapitre 2.

La conclusion générale résume les principales de l'étude, revisitant ses objectifs principaux et les connaissances acquises grâce à la détection des émotions à partir du texte. Enfin, le chapitre présente des pistes de recherches futures et des applications possibles dans le domaine des assistants virtuels empathiques.

CHAPITRE 1 LITERATURE REVIEW ON NATURAL LANGUAGE PROCESSING AND EMOTION RECOGNITION FROM TEXT

RÉSUMÉ

C'est le premier article, et il est à présent finalisé. Le traitement automatique des langues (TAL) est un sous-domaine essentiel de l'intelligence artificielle qui vise à permettre l'interaction homme-machine par le biais du langage naturel. Cet article fournit une revue détaillé de l'évolution du TAL, des techniques de base et de son rôle dans la reconnaissance des émotions à partir de texte. Nous explorons les principales techniques du TAL telles que la tokenisation, l'étiquetage des parties du discours et la reconnaissance des entités nommées, ainsi que des méthodes avancées comme l'analyse syntaxique et sémantique. L'article se penche également sur les méthodes de représentation de texte, allant des modèles traditionnels comme *Bag-of-Words* (BoW) et *Term Frequency-Inverse Document Frequency* (TF-IDF) aux intégrations modernes comme Word2Vec, GloVe et BERT. Une attention particulière est accordée aux défis et aux méthodologies de la reconnaissance des émotions, notamment l'utilisation d'approches d'apprentissage automatique et d'apprentissage profond. En examinant différents ensembles de données et mesures d'évaluation, cet article met en évidence les techniques de pointe qui façonnent l'avenir de la reconnaissance des émotions à partir de textes, en mettant l'accent sur ses applications dans divers domaines tels que l'analyse des sentiments, la surveillance de la santé mentale et le service client.

Mots-clés : Traitement Automatique des Langues (TAL), Analyse de texte, Analyse des sentiments, Reconnaissance des émotions.

Literature review On Natural Language Processing And Emotion Recognition From Text

Seyed Hamed Noktehdan Esfahani[1] and Mehdi Adda[1*]

[1*]Department of Mathematics, Computer Science and Engineering, UQAR,Quebec, Canada.

*Corresponding author(s). E-mail(s): mehdi_adda@uqar.ca;
Contributing authors: nokh0001@uqar.ca;

Abstract

Natural Language Processing (NLP) is a vital subfield of artificial intelligence that focuses on enabling human-computer interaction through natural language. This article provides a comprehensive overview of NLP's evolution, core techniques, and its role in emotion recognition from text. We explore key NLP techniques such as tokenization, part-of-speech tagging, and named entity recognition, as well as advanced methods like syntactic and semantic parsing. The article also delves into text representation methods, ranging from traditional models like Bag-of-Words (BoW) and Term Frequency-Inverse Document Frequency (TF-IDF) to modern embeddings like Word2Vec, GloVe, and BERT. Special attention is given to the challenges and methodologies in emotion recognition, including the use of machine learning and deep learning approaches. By examining different datasets and evaluation metrics, this article highlights the state-of-the-art techniques that are shaping the future of emotion recognition from text, emphasizing its applications in various domains such as sentiment analysis, mental health monitoring, and customer service.

Keywords: Natural Language Processing (NLP), Text Analysis, Sentiment Analysis , Emotion Recognition

1 Introduction

AI has a subdivision known as NLP, which concentrates on human-computer interaction through natural language. Essentially, the purpose of NLP is to ensure computers

comprehend, process and produce human languages relevantly and effectively. Programs that employ NLP can perform a variety of tasks, from basic ones like spelling correction or word prediction to more complicated ones such as machine translation, mood analysis, and chatbots [1].

The history of NLP has been shaped by important turning points and innovations in technology. The majority of early NLP systems, developed in the 1950s and 1960s, were rule-based and dependent on manually created dictionaries and rules. Large-scale datasets could be learned by systems thanks to statistical techniques brought about by machine learning in the 1980s. Deep learning and neural networks gained popularity in the 2010s, revolutionizing NLP with models like Word2Vec [2] and BERT [3].

This article aims to provide a thorough exploration of NLP techniques, with a focus on emotion recognition from text. The remainder of this paper is organized as follows: Section 2 reviews the foundational techniques in NLP and their concepts. In Section 3 we explain the definitions of emotion and the steps for extracting emotion from text. Section 4 is dedicated to modalities and techniques for emotion recognition. Section 5 presents available datasets with corresponding emotion labels. Finally, we discuss and conclude in Section 6 and Section 7.

2 Core Techniques in NLP

In Natural Language Processing (NLP), several foundational techniques and concepts are essential for understanding and processing human language. Here's an overview of some key concepts :

2.1 Tokenization

Tokenization is the process of splitting text into smaller units called tokens, which can be words, subwords, or characters. It is a fundamental step in text preprocessing for NLP tasks. Word-level tokenization treats each word as a token, while subword tokenization (e.g., Byte Pair Encoding) breaks words into smaller meaningful units. Sentence and document tokenization involve segmenting text into sentences and larger units [1].

For instance, given the sentence "I love natural language processing," the tokenization process would split it into the following components: ["I", "love", "natural", "language", "processing", "."]. Each word and punctuation mark is treated as an individual token. This step is crucial for enabling other NLP tasks, as it simplifies the text by dividing it into manageable parts that machines can understand and analyze.

2.2 Part-of-Speech Tagging

Part-of-speech (POS) tagging involves labeling each word in a sentence with its corresponding part of speech (e.g., noun, verb, adjective). For example, in the sentence "I love natural language processing," POS tagging would yield: [("I", Pronoun), ("love", Verb), ("natural", Adjective), ("language", Noun), ("processing", Noun)]. By identifying these roles, POS tagging assists in deeper linguistic analysis and is essential for tasks like parsing, translation, and information retrieval.

POS tagging is essential for understanding sentence structure and meaning. Techniques for POS tagging include rule-based methods, probabilistic models like Hidden Markov Models (HMMs), and modern machine learning approaches such as Conditional Random Fields (CRFs) [4].

2.3 Named Entity Recognition (NER)

Named Entity Recognition (NER) identifies and classifies entities in text into predefined categories such as names of persons, organizations, locations, dates, and more. For instance, in the sentence "Google was founded by Larry Page and Sergey Brin in 1998," NER would identify ["Google", Organization], ["Larry Page", Person], ["Sergey Brin", Person], ["1998", Date]. This process helps in transforming unstructured text into structured data by detecting and categorizing named entities, making it valuable for applications like information extraction, document classification, and summarization.

Common techniques include rule-based approaches, statistical models, and neural networks, particularly sequence-to-sequence models, which are often used to identify entities such as names, dates, and locations in various contexts.[5].

2.4 Syntactic and Semantic Parsing

Syntactic parsing involves analyzing the grammatical structure of sentences, whereas semantic parsing focuses on understanding the meaning of sentences. Dependency and constituency parsing are two primary syntactic parsing techniques, while semantic role labeling is used for semantic parsing.

For example, in the sentence "The cat sat on the mat," syntactic parsing would reveal that "cat" is the subject of "sat" and "mat" is the object of the preposition "on." This helps in understanding how words are connected within a sentence. Semantic parsing, on the other hand, focuses on interpreting the meaning conveyed by the sentence. For the same sentence, it would derive that the "cat" is the actor performing the action of "sitting" on the "mat." Together, syntactic and semantic parsing provide a deeper understanding of both the structure and the meaning of text, which is essential for more advanced NLP tasks such as translation and text summarization [1].

2.5 Text Representation

Text representation involves converting text into numerical forms that can be processed by machine learning models. Traditional methods include Bag-of-Words (BoW) and Term Frequency-Inverse Document Frequency (TF-IDF). Modern approaches use word embeddings such as Word2Vec [2], GloVe [6], and contextual embeddings like BERT [3] and GPT [7].

For example, the words "king" and "queen" might be represented by vectors like [0.27, -0.32, 0.81, ...] for "king" and [0.21, -0.28, 0.85, ...] for "queen." These vectors encode similarities and differences between words, allowing models to understand relationships such as $king - man + woman \approx queen$. Word embeddings enable machines to grasp subtle nuances in meaning and context, which is crucial for tasks like sentiment analysis, translation, and text generation.

For a clearer understanding and a visual summary of the mentioned techniques, refer to Figure 2.5.

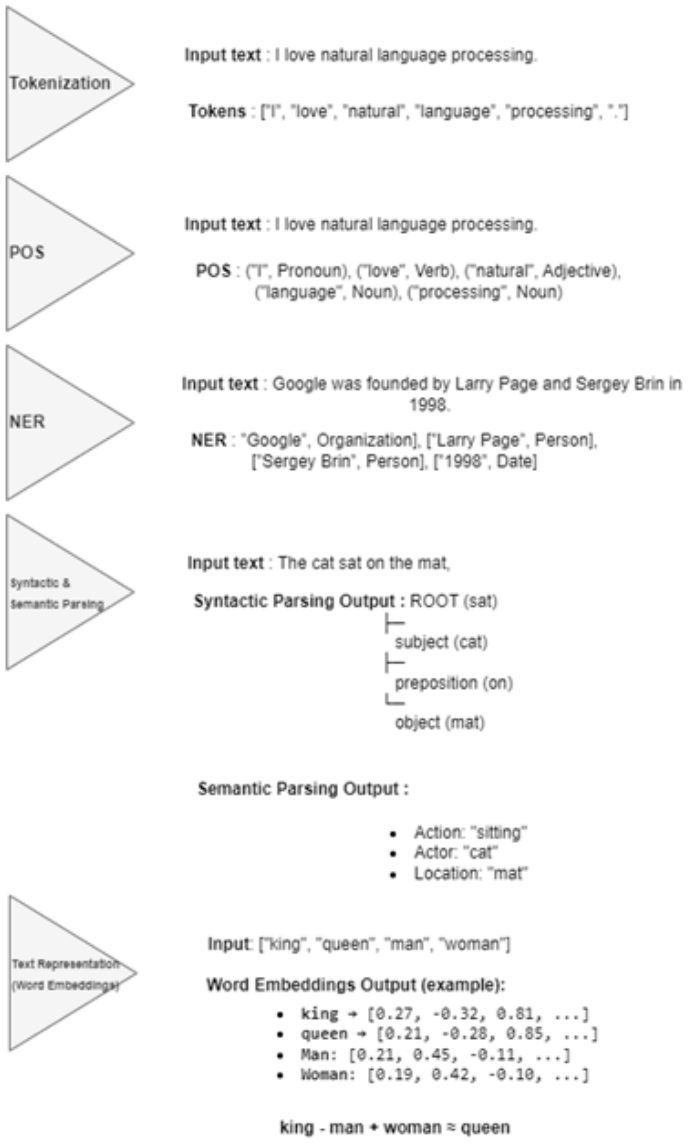

Fig. 1 Core Techniques in NLP

3 Definition of Emotion

Emotion is a complex, multifaceted psychological state, often defined as a combination of subjective experiences, physiological responses, and expressive behaviors. According to Scherer [8], emotions serve adaptive functions by helping individuals respond to events in their environment, processing information to guide survival and social interaction. They are generally considered to have evolved as part of the human brain's mechanisms for decision-making and social bonding [9]. Emotions are triggered by various stimuli, ranging from external events to internal cognitive appraisals, which determine their intensity and nature. While definitions may vary, the consensus in scientific literature is that emotions are central to the human experience, affecting cognition, behavior, and social dynamics [10].

Emotions have been categorized in different ways, depending on the theoretical approach. One prominent model is the Basic Emotion Theory, popularized by Paul Ekman, who identified six primary emotions—happiness, sadness, fear, anger, disgust, and surprise—that are universally recognized across cultures [9]. These basic emotions are believed to be biologically hardwired and accompanied by distinct facial expressions, making them easily recognizable in different societies [11]. This view suggests that certain emotions have evolved to serve key functions, such as survival (fear) or social bonding (happiness).

3.1 Emotion Recognition from Text

The process of identifying and categorizing emotions expressed in textual data is known as emotion recognition from text and is important for applications in sentiment analysis, customer service, mental health monitoring, and more. It improves human-computer interaction by helping to understand user sentiment and emotional states [12]. However, there are a number of obstacles to overcome before emotion recognition from text can be fully utilized, such as ambiguity and context dependence, where a word or phrase can convey multiple emotions depending on the context; sarcasm and irony pose additional challenges because they frequently convey opposite sentiments to what is actually stated; cultural differences also affect the expression and recognition of emotions [13].

Text emotion recognition involves several key steps. First, the raw text data is preprocessed to clean and normalize it, removing noise like punctuation and stop words. Next, feature extraction transforms the text into numerical representations, capturing essential linguistic and semantic information. These features are then fed into a machine learning model, which is trained to recognize patterns associated with different emotions. Once trained, the model can classify new text data into corresponding emotion categories. Finally, the model's performance is evaluated using appropriate metrics to assess its accuracy and effectiveness.

3.2 Preprocessing for Emotion Recognition

Text preprocessing is a critical step in emotion recognition. It involves cleaning and normalizing text, including tasks such as removing stop words, stemming, lemmatization, and handling contractions. Dealing with imbalanced data, where some emotions

are underrepresented, is also essential. Techniques such as resampling and synthetic data generation can help address this issue [14].

3.3 Text Feature Extraction

A crucial stage in NLP is text feature extraction, which converts unprocessed text into numerical vectors that can be processed by machine learning algorithms. This procedure extracts important information from the text and has a big impact on how well NLP models work. Here, we explore various techniques for text feature extraction, from traditional methods to advanced deep learning approaches.

3.4 Traditional Feature Extraction Techniques

A crucial stage in natural language processing NLP is feature extraction, which converts unprocessed text into numerical representations that machine learning algorithms can understand. Traditional feature extraction techniques, while relatively simple, have been foundational in the development of NLP. These methods convert text into structured formats that capture essential information about word occurrences and relationships, enabling models to process and analyze language data. Despite their limitations in capturing context and semantics, these traditional approaches laid the groundwork for more advanced techniques.

3.4.1 Bag-of-Words (BoW)

For text feature extraction, one of the simplest and most popular methods is the BoW model. It illustrates text as an unordered mix of words, preserving diversity while ignoring word order and grammar. With BoW, each document is represented as a vector of word frequencies, and a vocabulary is generated from the text corpus. This technique efficiently converts text into a format that machine learning algorithms can understand. But BoW has some significant drawbacks. Because it ignores word order, context and syntactic relationships are not captured. Furthermore, very high-dimensional vectors can result in a large vocabulary, increasing computational complexity and possibly overfitting.

3.4.2 Term Frequency-Inverse Document Frequency (TF-IDF)

TF-IDF is an enhancement over BoW that aims to reflect the importance of words in a document relative to a corpus. TF-IDF combines term frequency (TF) metric, which measures how often a word appears in a document, and inverse document frequency (IDF), which measures how unique or rare a word is across all documents. This combination helps to downweight common words that are less informative and highlight unique words that are more significant. By adjusting the weight of words based on their overall distribution in the corpus, TF-IDF provides a more informative representation than BoW. Even with these benefits, TF-IDF is still unable to fully capture word order or context beyond individual word importance.

3.4.3 N-Grams

N-grams extend the BoW model by considering sequences of N words together, rather than individual words. This method captures some context and the order of words, providing more information about the structure of the text. For instance, bigrams (2-grams) and trigrams (3-grams) consider pairs and triplets of words, respectively. While N-grams can capture local context and improve the representation of text, they also increase the dimensionality of the feature space, especially for higher values of N. This can lead to computational challenges and may still fall short in capturing long-range dependencies within the text.

3.5 Advanced Feature Extraction Techniques

Traditional techniques often fail to capture the complexities of language, such as context and meaning. To overcome these constraints, sophisticated feature extraction methods, especially those based on neural networks, have been created. These methods focus on creating dense, continuous representations of words that encapsulate both syntactic and semantic information, allowing for more accurate and meaningful analysis of language data.

3.5.1 Word Embeddings

Word embeddings are dense vector representations of words that capture their meanings, contexts, and syntactic relationships. These embeddings are trained using neural networks on large corpora of text data, allowing words with similar meanings to have similar representations.

Word2Vec: Introduced by Mikolov et al. [2], Word2Vec uses two main architectures: the skip-gram model and the continuous bag-of-words (CBOW) model. The skip-gram model predicts the context words given a target word, while the CBOW model predicts the target word based on its context. Word2Vec efficiently captures semantic relationships and analogies between words, providing meaningful vector representations that can be used in various NLP tasks.

GloVe: Developed by Pennington et al. [6], GloVe (Global Vectors for Word Representation) uses a different approach by leveraging global word-word co-occurrence statistics from a corpus. The resulting vectors capture both local and global statistical information, providing rich representations of words. GloVe effectively combines the benefits of matrix factorization methods and local context-based methods like Word2Vec.

Word embeddings offer significant advantages over traditional methods by capturing semantic and syntactic similarities, providing low-dimensional dense vectors that reduce computational complexity. However, they are static and cannot capture word meanings that change with context.

3.5.2 Contextualized Embeddings

Contextualized embeddings generate word vectors that depend on the context in which the word appears, addressing the limitations of static embeddings like Word2Vec and

GloVe. These embeddings are generated by models that consider the entire sentence or paragraph to capture the meaning of words in their specific contexts.

ELMo: Embeddings from Language Models (ELMo), introduced by Peters et al. [15], generate word representations that vary according to their usage in sentences. ELMo uses deep bidirectional language models to produce context-sensitive embeddings, which can significantly improve performance on various NLP tasks by providing richer representations of words based on their surrounding context.

BERT: Bidirectional Encoder Representations from Transformers (BERT), developed by Devlin et al. [3], uses a transformer architecture to create deeply contextualized word embeddings. BERT considers both the left and right context of a word, providing a more comprehensive understanding of word meanings. This bidirectional approach allows BERT to capture nuanced information and dependencies between words, leading to state-of-the-art performance on many NLP benchmarks.

Contextualized embeddings better capture the meaning of polysemous words (words with multiple meanings) and provide improved performance on various NLP tasks. They represent a significant advancement in text feature extraction by incorporating context into the representation of words.

4 Modalities and Techniques for Emotion Recognition

Finding the emotional tone that written language is trying to convey is a difficult task in emotion recognition from text. This task requires sophisticated techniques that can distinguish subtle emotional cues within text. Various approaches have been developed to tackle this challenge, ranging from traditional machine learning models to more recent deep learning architectures. Each approach offers different advantages, depending on the complexity of the data and the specific requirements of the task. Figure 4 illustrates the overview of techniques in emotion recognition in text.

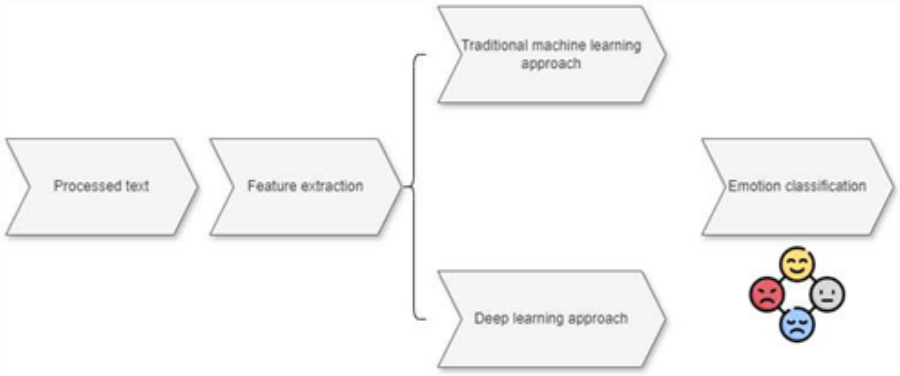

Fig. 2 Methodology for automatic Emotion Detection

4.1 Machine Learning Approaches

Traditional machine learning approaches for emotion recognition from text typically involve training classifiers on labeled datasets, where the goal is to assign an emotional label to a given text based on its content. In this process, algorithms such as Support Vector Machines (SVMs), Naive Bayes, and Decision Trees are frequently employed due to their effectiveness in classification tasks.

Support Vector Machines (SVMs), for instance, are particularly effective in high-dimensional spaces, making them widely used in text classification tasks where the relationships between features can be complex. SVMs work by finding a hyperplane that best separates different classes of data, maximizing the margin between the nearest data points of each class. This allows SVMs to handle sparse and high-dimensional data, which is common in natural language processing tasks [16]. One notable application of SVMs is found in the study by Purver and Battersby [17], who applied the algorithm to classify emotions in Twitter data. They attained an impressive accuracy of 82% when identifying the emotion "Happy" during a 10-fold cross-validation. However, when classifying this emotion across the entire dataset, the accuracy dropped to 67%. Their approach utilized emoticons to label the training data and hashtags for labeling the test data. When testing their trained models on distinguishing specific emotions from one another—rather than just from a general category labeled "Other"—results varied between 13% and 76%, depending on the emotion. Additionally, they created a dataset of 1,000 tweets, manually annotated by human judges, to assess the effectiveness of using hashtags and emoticons as labels. The F-scores across emotions ranged from 0.10 to 0.77, showing that the classifiers performed well for emotions like happiness, sadness, and anger but struggled with others. They concluded that using hashtags and emoticons as labeling methods was a promising alternative to traditional manual annotation.

In a similar method, Roberts et al. [18] gathered tweets on 14 topics likely to provoke emotional reactions and compiled a dataset where all seven emotions (Ekman's six basic emotions plus "Love") were represented. They used seven SVM classifiers to identify emotions, achieving an average F1-score of 0.66. This further demonstrated the utility of SVMs in processing social media text and emotion recognition in general.

Naive Bayes, another widely used approach, is a probabilistic classifier grounded in Bayes' theorem, and it operates under the assumption of feature independence. Although this assumption does not always hold true for real-world data, Naive Bayes is particularly useful for large datasets due to its simplicity and speed [4]. It not only provides predictions but also delivers probability estimates for each classification. An example of Naive Bayes' effectiveness comes from a study that demonstrated the algorithm's 75.2% accuracy in the Twitter Sentiment Analysis dataset [19]. Hasan et al. [20] also used Naive Bayes, along with SVM, Decision Trees, and K-Nearest Neighbors (KNN), to classify emotions in tweets.

Decision Trees offer a more interpretable approach by constructing a tree-like structure where decisions are made based on feature values at each node, ultimately leading to the assignment of an emotional class label. Decision Trees are particularly useful for understanding the decision-making process, as they provide clear visual representations of how specific features contribute to the final classification.

Hasan et al. [21] extended their earlier work by developing an automatic emotion detection system that identified emotions from streams of tweets. This system was designed in two stages: first, it trained an offline model to classify emotions based on their 2014 work; second, it implemented a two-step classification process. This process initially detected whether a tweet contained an emotion, and then assigned a more specific emotional label using soft classification techniques. This two-stage approach enabled a more nuanced identification of emotions within social media streams.

However, these methods rely heavily on extensive feature engineering, which involves the manual selection and crafting of features that can effectively capture emotional nuances in text. N-grams, which are contiguous sequences of words or characters, are commonly used to capture contextual information. For example, bigrams (two-word sequences) or trigrams (three-word sequences) can provide insights into the co-occurrence of words that may signal specific emotions. In addition to n-grams, sentiment lexicons—which are pre-built dictionaries mapping words to their associated sentiments—are often incorporated to enhance the model's understanding of emotionally charged words. Moreover, syntactic features, such as part-of-speech tags, dependency relations, and sentence structure, can offer additional layers of information regarding how words are used in context, further refining the model's ability to recognize emotions.

Despite the strengths of these algorithms, the need for manual feature engineering is one of their significant limitations. Manually designing features is labor-intensive, domain-dependent, and may not generalize well across different datasets or languages. Moreover, these features may not capture the full complexity of human emotions, which are often subtle and context-dependent [22]. Consequently, as the field has evolved, there has been a shift towards more sophisticated techniques, such as deep learning models, which aim to automatically learn and extract features from text data. Nonetheless, traditional machine learning approaches continue to be valuable, particularly in cases where interpretability and computational efficiency are prioritized.

A summary of the key studies discussed in this section is presented in Table 1.

Table 1 Summary of related works on emotion recognition from text with Machine Learning approaches

Authors	Method	Notes
Purver and Battersby [17]	SVM	F-score (varies by emotion): 0.10 to 0.77; manual annotation of 1,000 tweets
Roberts et al. [18]	SVM	Seven SVM classifiers used to detect Ekman's emotions plus "Love"
Hasan et al. [20]	NB, SVM, DT, KNN	Comparison of multiple methods for emotion classification from tweets
Hasan et al. [21]	DT	Two-stage process for emotion detection: presence of emotion, followed by soft classification

4.2 Deep Learning Approaches

Deep learning has transformed emotion recognition with models capable of automatically learning features from raw data.

Recurrent Neural Networks (RNNs) are suited for sequential data and can capture temporal dependencies in text. It learns from past information to predict future outcomes. Essentially, RNNs have a memory that allows them to use previous results to improve their current performance. However, they suffer from vanishing gradient problems with long sequences [23]. For instance, an RNN-based model by Tang et al. [24] achieved an accuracy of 83.1% on the IMDB movie reviews dataset.

Long Short-Term Memory (LSTM) address the limitations of RNNs by introducing memory cells that can maintain information over long sequences, making them effective for capturing context in text [25]. LSTM networks have specialized components called memory cells and gates that enable them to remember information over long periods. Memory cells function like a computer's memory, storing, retrieving, and updating data. The gates, which include input, forget, and output gates, regulate the flow of information into and out of the memory cells, controlling what information is retained or discarded. In a study by Huang et al. [26], an LSTM model achieved an accuracy of 85.7% in detecting emotions in the Semeval dataset.

Convolutional Neural Networks (CNNs), are a type of deep learning model primarily known for their success in image processing. They excel at identifying patterns in data by applying filters to extract features. While originally designed for images, CNNs have also found applications in NLP. They capture local dependencies and hierarchical structures in text. For example, A widely referenced study by Kim et al. [27] demonstrates the effectiveness of CNNs for sentiment analysis. In this study, the author used a CNN architecture to classify movie reviews as either positive or negative. Proposed CNN model achieved an accuracy of 88.1% on the test set, outperforming several traditional models.

Transformer Models, are a type of neural network architecture that has revolutionized NLP. Unlike previous models that relied on recurrent or convolutional structures, transformers employ a self-attention mechanism to weigh the importance of different parts of the input sequence. This allows them to process information in parallel, capturing complex dependencies between words and improving performance on tasks like machine translation, text summarization, and question answering. [28]. A study by Sun et al. [29] fine-tuned BERT for emotion recognition and achieved an accuracy of 88.5% on the EmotionX dataset.

Table 2 summarizes the key studies discussed in this section.

5 Datasets for Emotion Recognition

Several publicly available datasets are used for training and evaluating emotion recognition models:

SemEval-2007 [30]: This dataset is composed of news headlines sourced from outlets like BBC News, CNN, the New York Times, and the Google News search engine. The headlines are suitable for sentence-level emotion annotation. Each one is tagged with one or more emotions, such as anger, disgust, fear, joy, sadness, or surprise.

Table 2 Summary of related works on emotion recognition using deep learning

Authors	Method	Dataset/Context	Accuracy
Tang et al. [24]	RNN	IMDB Movie Reviews	83.1%
Huang et al. [26]	LSTM	Semeval Dataset	85.7%
Kim et al. [27]	CNN	Movie Reviews	88.1%
Sun et al. [29]	BERT	EmotionX Dataset	88.5%

SemEval-2018 [31]: This dataset is made up of tweets, where each tweet is either neutral or conveys one or more emotions from a list of eleven, including anger, disgust, fear, joy, love, optimism, pessimism, sadness, surprise, and trust. The dataset contains tweets in English, Arabic, and Spanish.

SemEval-2019 [32]: This corpus consists of dialogues between two individuals. The first person initiates the conversation, followed by a response from the second person, and the interaction continues. Emotions are labeled based on the third turn, and each dialogue is classified as joy, anger, sadness, or other.

GoEmotions [33]: This dataset contains Reddit comments annotated across 28 emotion categories, offering a diverse range of emotional expressions. It is a key resource for training and evaluating emotion recognition models, supporting tasks like sentiment analysis and empathetic AI.

CBET (Crisis Benchmark Emotion Tagging) [34]: This dataset consists of labeled records, categorized into 9 distinct emotion groups. It is specifically designed to capture emotional responses in crisis situations, making it valuable for analyzing emotional dynamics in critical contexts. This dataset supports the development of emotion recognition models, particularly for applications focused on crisis management and intervention.

ISEAR (International Survey on Emotion Antecedents and Reactions) [35]: This dataset includes 7,666 records categorized into 7 emotion categories.

These datasets are commonly used in research for training and evaluating models in the field of emotion recognition from text. Each dataset varies in size and the number of emotion categories, offering a range of data for different research needs and applications. Table 3 represents details of the mentioned datasets.

6 Discussion

Advancements in NLP, particularly through deep learning and machine learning, have significantly improved computers' ability to process and understand human language. The transition from rule-based systems to neural networks has transformed the field, with models like BERT [3] and ELMo [15] providing better contextual understanding. However, these improvements come with new challenges, particularly in text-based emotion recognition.

One of the core challenges in emotion recognition is the context-dependent nature of language. Words often carry different emotional meanings depending on their context, making it difficult for models to accurately classify emotions. Traditional models

Table 3 Datasets Information

Dataset	Number of Instances	Emotions
GoEmotions	59,331	admiration, amusement, anger, annoyance, approval, caring, confusion, curiosity, desire, disappointment, disapproval, disgust, embarrassment, excitement, fear, gratitude, grief, joy, love, nervousness, optimism, pride, realization, relief, remorse, sadness, surprise, neutral
CBET	81,163	anger, fear, sadness, joy, surprise, disgust, trust, anticipation, guilt
ISEAR	7,666	joy, fear, anger, sadness, disgust, shame, guilt
SemEval-2007	1,250	anger, disgust, fear ,happiness ,sadness ,surprise
SemEval-2018	10,983	neutral, anger, disgust, fear, joy, love, optimism, pessimism, sadness, surprise, trust
SemEval-2019	38,424	Joy, Anger, Sadness, Others

like BoW and TF-IDF were limited in capturing these nuances, and while word embeddings such as Word2Vec [2] , GloVe [6] , and contextualized embeddings like BERT [3] have made strides, they still struggle with linguistic phenomena like sarcasm, irony, and metaphors. These elements can convey emotions opposite to the literal meaning of words, posing a significant barrier to accurate emotion detection.

Most current emotion recognition systems rely on predefined categories (e.g., happiness, sadness, anger), which fail to capture the full range of human emotions. This limits the models' ability to identify subtle emotional states. Recent datasets like GoEmotions [33], which offer 28 fine-grained emotional categories, demonstrate the need for more flexible models that can handle the complexity of real-world emotional expressions.

Emotions are expressed differently across cultures and languages, yet most emotion recognition systems are trained on datasets dominated by Western, English-speaking populations. This raises concerns about the generalizability of these models across diverse cultures and linguistic backgrounds. Addressing this issue requires the development of more culturally inclusive datasets and multilingual models to ensure fair and accurate emotion detection globally.

The imbalance in datasets, where certain emotions are underrepresented, remains a significant challenge. This can result in biased models that underperform in detecting rare emotions. Techniques such as data augmentation, oversampling, and synthetic data generation methods can help mitigate this problem, but more effective solutions are needed. Additionally, adaptive learning approaches could enable models to better handle rare emotions without overfitting on more common ones.

The field of emotion recognition continues to face several research challenges that need to be addressed for the technology to reach its full potential. While significant progress has been made in developing models capable of detecting emotions from text, several open areas of investigation remain, particularly in enhancing the accuracy and adaptability of these systems across diverse contexts and user groups. Future research must focus on overcoming limitations related to data, model architecture, and ethical considerations. Addressing these challenges will not only improve the performance of

emotion recognition systems but also ensure their broader and more responsible use in real-world applications.

One of the key challenges is multimodal emotion recognition. Current research predominantly focuses on text-based systems, but human emotions are expressed through a combination of text, voice, facial expressions, gestures, and more. Integrating data from multiple modalities, such as speech and visual cues, could significantly improve the richness and accuracy of emotion detection.

As emotion recognition technologies become more sophisticated, ethics and privacy concerns also emerge as major challenges. The ability to detect emotions from personal data, particularly in sensitive areas such as healthcare or surveillance, raises serious ethical questions about consent, data security, and the potential misuse of technology. Researchers must prioritize the development of privacy-preserving models that ensure users' autonomy is respected. Techniques like federated learning, which allows models to be trained locally without centralizing personal data, offer promising pathways to address these concerns while still leveraging the benefits of emotion recognition technology.

There is also the challenge of real-time emotion recognition. Many existing emotion recognition models, especially those based on deep learning architectures, are computationally expensive and require significant processing power. This makes real-time emotion detection difficult, particularly in resource-constrained environments such as mobile devices or edge computing systems. Developing more lightweight and efficient models that can operate in real-time while maintaining high accuracy is a key area for future research, especially for applications in interactive systems like virtual assistants and real-time customer service platforms.

To address these challenges, future research should focus on improving the contextual understanding of language, expanding cultural and linguistic inclusivity, integrating multimodal inputs, and building ethical frameworks around emotion recognition technology. By tackling these issues, emotion recognition systems can become more accurate, adaptable, and equitable, unlocking new opportunities in healthcare, customer service, education, and beyond.

7 Conclusion

This review highlights the significant advancements in NLP and emotion recognition from text. Traditional NLP techniques, like Bag-of-Words and TF-IDF, have evolved into sophisticated models such as Word2Vec, GloVe, and BERT, improving machines' ability to understand and process language. Similarly, emotion recognition has benefited from advances in machine learning and deep learning, enabling better identification of emotional cues in text, though challenges like sarcasm and cultural differences persist.

Future developments in emotion recognition will likely involve multimodal approaches that combine text with audio and visual data, further enhancing the accuracy of emotional detection. Ethical concerns around privacy and real-time applications need to be addressed as the field progresses. Despite the challenges, continued

innovation in this area holds great potential for applications ranging from sentiment analysis to mental health monitoring and interactive systems.

References

[1] Jurafsky, D., Martin, J.H.: Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition with language models (2024). Online manuscript released August 20, 2024

[2] Mikolov, T.: Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781 (2013)

[3] Devlin, J.: Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805 (2018)

[4] Manning, C.D., Raghavan, P., Schütze, H.: Introduction to information retrieval (2008)

[5] Lample, G.: Neural architectures for named entity recognition. arXiv preprint arXiv:1603.01360 (2016)

[6] Pennington, J., Socher, R., Manning, C.D.: Glove: Global vectors for word representation. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 1532–1543 (2014)

[7] Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., Sutskever, I., *et al.*: Language models are unsupervised multitask learners. OpenAI blog **1**(8), 9 (2019)

[8] Scherer, K.R.: What are emotions? and how can they be measured? Social science information **44**(4), 695–729 (2005)

[9] Ekman, P.: An argument for basic emotions. Cognition & emotion **6**(3-4), 169–200 (1992)

[10] Izard, C.E.: Basic emotions, natural kinds, emotion schemas, and a new paradigm. Perspectives on psychological science **2**(3), 260–280 (2007)

[11] Ekman, P., Friesen, W.V.: Constants across cultures in the face and emotion. Journal of personality and social psychology **17**(2), 124 (1971)

[12] Poria, S., Cambria, E., Bajpai, R., Hussain, A.: A review of affective computing: From unimodal analysis to multimodal fusion. Information fusion **37**, 98–125 (2017)

[13] Cambria, E., Poria, S., Gelbukh, A., Thelwall, M.: Sentiment analysis is a big suitcase. IEEE Intelligent Systems **32**(6), 74–80 (2017)

[14] Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P.: Smote: synthetic minority over-sampling technique. Journal of artificial intelligence research **16**, 321–357 (2002)

[15] Peters, M.E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., Zettlemoyer, L.: Deep contextualized word representations (2018)

[16] Cortes, C.: Support-vector networks. Machine Learning (1995)

[17] Purver, M., Battersby, S.: Experimenting with distant supervision for emotion classification. In: Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics, pp. 482–491 (2012)

[18] Roberts, K., Roach, M.A., Johnson, J., Guthrie, J., Harabagiu, S.M.: Empatweet: Annotating and detecting emotions on twitter. In: Lrec, vol. 12, pp. 3806–3813 (2012)

[19] Khan, M.T., Durrani, M., Ali, A., Inayat, I., Khalid, S., Khan, K.H.: Sentiment analysis and the complex natural language. Complex Adaptive Systems Modeling **4**, 1–19 (2016)

[20] Hasan, M., Rundensteiner, E., Agu, E.: Emotex: Detecting emotions in twitter messages (2014)

[21] Hasan, M., Rundensteiner, E., Agu, E.: Automatic emotion detection in text streams by analyzing twitter data. International Journal of Data Science and Analytics **7**, 35–51 (2019)

[22] Mohammad, S.M., Turney, P.D.: Crowdsourcing a word–emotion association lexicon. Computational intelligence **29**(3), 436–465 (2013)

[23] Elman, J.L.: Finding structure in time. Cognitive science **14**(2), 179–211 (1990)

[24] Tang, D., Qin, B., Liu, T.: Document modeling with gated recurrent neural network for sentiment classification. In: Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, pp. 1422–1432 (2015)

[25] Schmidhuber, J., Hochreiter, S., *et al.*: Long short-term memory. Neural Comput **9**(8), 1735–1780 (1997)

[26] Huang, M., Zhu, X., Gao, J.: Challenges in building intelligent open-domain dialog systems. ACM Transactions on Information Systems (TOIS) **38**(3), 1–32 (2020)

[27] Kim, Y.: Convolutional neural networks for sentence classification (2014)

[28] Vaswani, A.: Attention is all you need. Advances in Neural Information Processing Systems (2017)

[29] Sun, C., Huang, L., Qiu, X.: Utilizing bert for aspect-based sentiment analysis via constructing auxiliary sentence. arXiv preprint arXiv:1903.09588 (2019)

[30] Strapparava, C., Mihalcea, R.: Semeval-2007 task 14: Affective text. In: Proceedings of the Fourth International Workshop on Semantic Evaluations (SemEval-2007), pp. 70–74 (2007)

[31] Mohammad, S., Bravo-Marquez, F., Salameh, M., Kiritchenko, S.: Semeval-2018 task 1: Affect in tweets. In: Proceedings of the 12th International Workshop on Semantic Evaluation, pp. 1–17 (2018)

[32] Chatterjee, A., Narahari, K.N., Joshi, M., Agrawal, P.: Semeval-2019 task 3: Emocontext contextual emotion detection in text. In: Proceedings of the 13th International Workshop on Semantic Evaluation, pp. 39–48 (2019)

[33] Demszky, D., Movshovitz-Attias, D., Ko, J., Cowen, A., Nemade, G., Ravi, S.: Goemotions: A dataset of fine-grained emotions. arXiv preprint arXiv:2005.00547 (2020)

[34] Alam, F., Sajjad, H., Imran, M., Ofli, F.: Crisisbench: Benchmarking crisis-related social media datasets for humanitarian information processing **15**, 923–932 (2021)

[35] Scherer, K.R., Wallbott, H.G.: Evidence for universality and cultural variation of differential emotion response patterning. Journal of personality and social psychology **66**(2), 310 (1994)

CHAPITRE 2 CLASSICAL MACHINE LEARNING AND LARGE MODELS FOR TEXT-BASED EMOTION RECOGNITION

RÉSUMÉ

Il s'agit du deuxième article, actuellement publié. Cet article [10] traite de l'application des techniques d'apprentissage automatique et des grands modèles de langage (GML) pour la détection des émotions dans les textes. Il compare plusieurs méthodes d'extraction de caractéristiques et examine comment ces approches influencent les performances des modèles. Le modèle Mistral 7B, en particulier, s'est révélé plus efficace que le modèle Falcon 7B, surtout lorsqu'il est affiné pour les tâches spécifiques d'analyse des émotions. Les résultats montrent le potentiel des GML avancés dans le domaine du traitement automatique des langues (TAL), notamment pour l'analyse des sentiments et la reconnaissance des émotions. L'article suggère que les recherches futures devraient se concentrer sur des techniques d'apprentissage en plusieurs exemples (*multi-shot learning*) et sur l'impact du choix des sous-ensembles de données pour l'ajustement, afin d'améliorer la scalabilité et l'adaptabilité des modèles à divers ensembles de données et scénarios.

Mots-clés : TAL, reconnaissance des émotions, apprentissage automatique, grand modèle de langage, GML, apprentissage ponctuel, réglage fin.

Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 241 (2024) 77–84

Procedia Computer Science

www.elsevier.com/locate/procedia

The 19th International Conference on Future Networks and Communications (FNC)
August 5-7, 2024, Marshall University, Huntington, WV, USA

Classical Machine Learning and Large Models for Text-Based Emotion Recognition

Seyed Hamed Noktehdan Esfahani[a], Mehdi Adda[a,*]

[a]*Department of Mathematics and Computer Science, University of Quebec At Rimouski, Quebec, Canada*

Abstract

In the era of digital communication, detecting emotions has become crucial for applications ranging from customer service to mental health assessment. This study examines emotion recognition in text through various machine learning techniques, from traditional machine learning techniques, to advanced large language models (LLMs) such as BERT, Falcon 7B, and Mistral 7B. The results reveal that the fine-tuned Mistral 7B model is the most precise, achieving an accuracy of 76%. Furthermore, Support Vector Machine attained an accuracy of 64%.

Keywords: NLP; Emotion Recognition; Machine Learning; Large Language Models; LLM; One-shot learning; Fine-tuning;

1. Introduction

In the field of natural language processing (NLP), understanding the meaning behind words has significantly improved. Initially, this simply classified opinions as positive, negative, or neutral. Now, the goal is to detect the specific emotions conveyed in the text. These two concepts, sentiment and emotion, are distinct. Sentiment is broad, encompassing positive, negative, and neutral feelings, while emotions are more nuanced categories within the positive-negative spectrum. According to Ekman [10], emotions are categorized into six different types: *JOY*, *SADNESS*, *FEAR*, surprise, *ANGER*, and *DISGUST*. Emotion recognition can be categorized into three main approaches: categorical/discrete, dimensional, and appraisal-based [11]. Various methods exist for recognizing and categorizing emotions through the analysis of diverse types of data, including brain signals, voice/speech, facial movements, and text.

Our research uses machine learning (ML) techniques for emotion recognition through text processing. Traditionally, emotion recognition in ML has been approached through specific methods like training algorithms on labeled

* Corresponding author. Tel.: +1-418-723-9425 ;
E-mail address:mehdi_adda@uqar.ca

1877-0509
10.1016/j.procs.2024.08.013

emotional datasets and employing lexicons of emotional words (lexical approach). However, these techniques are gradually fading in favor of deep learning (DL) models.

The primary objective of this study is to assess the effectiveness of Large Language Models (LLMs) compared to classical ML techniques in emotion recognition.

The remainder of this paper is organized as follows: Section 2 reviews related works. In Section 3, we present our methodology. Sections 4 and 6 outline the experiments conducted and the obtained results, respectively. Section 6 represents a detailed discussion of the results. Finally, we present concluding remarks and discuss future work in Section 7.

2. Related Work

In this section, we categorize the related work on emotion detection from text into two categories. We start by looking at general related work in the field before doing a focused analysis of studies specifically related to the ISEAR dataset. This approach provides a basis for comparison with existing studies.

Santosh Kumar Bharti et al [5] proposed a hybrid model consisting of a deep learning approach and machine learning. They used word2vec to feed the embedding layer to CNN and Bi-GRU. They have removed the last layer of these models to act as an encoder. Furthermore, by concatenating the given latent vector of the models, they fed it to the SVM classifier to predict the emotion of the given text. The performance of the proposed approach was evaluated using a combination of three different types of datasets, namely sentences, tweets, and dialogs, and it reached an accuracy of 80.11%.

According to the findings of Szabóová et al. [17], emotion analysis has proven effective in enhancing human-robot interaction. Conversely, Beridge et al [4]. reported that a majority of respondents (68.7%) doubted the capability of robots to alleviate loneliness, while a significant portion (69.3%) expressed discomfort, ranging from moderate to intense, with the concept of robots serving as companions.

In the study by Chatterjee et al [7]. They introduced a DL method termed sentiment and semantic LSTM (SS-LSTM). This approach involved evaluating various DL techniques such as CNN and LSTM, along with different forms of text representation including Word2Vec, GloVe, FastText, and Sentiment Specific Word Embedding. Additionally, they examined the performance of classical ML algorithms such as SVM, gradient boost decision trees, and Naive Bayes in real text conversations. The evaluation was based on the efficiency of emotion detection, categorized into four classes. The training dataset comprised 17.62 million tweet conversation pairs gathered from Twitter, specifically Q-A tweets. Their method exhibited superior performance compared to most basic ML algorithms.

In Khanpour and Caragea's research [13], emotions were investigated within messages from an online health community. Initially, the authors labeled a dataset sourced from a cancer forum[1] with the six primary emotions outlined by Ekman [9]. They analyzed the prevalence and distribution of these emotions, finding *JOY* and *SADNESS* the most common, followed by *ANGER* and *FEAR*, while *DISGUST* and *surprise* were less frequent. Subsequently, they proposed a computational model that integrates CNN, LSTM, and lexical methods. This model aims to capture underlying semantics in text messages, facilitating a deeper understanding of the emotional content by identifying various emotional expressions in text.

In Kratzwald et al.'s study [14], they introduced a novel neural network method termed the bi-directional LSTM (BiLSTM) network, designed to make predictions based on texts of varying lengths. Their innovation includes bidirectional text processing, layer extraction for regularization purposes, and incorporating a weighted loss function. Additionally, they proposed an extension of transfer learning known as sent2affect. Initially, the network was trained for sentiment analysis. Subsequently, by replacing the output layer, it was reconfigured for emotion detection tasks. Their findings demonstrated performance comparable to advanced research at the time, employing classical ML algorithms such as SVM and random forest decision trees. By nature of these results, Polignano et al [15] developed a model that integrates BiLSTM with Self-Attention and Convolutional Neural Networks (CNN). They focused on extracting word embeddings as a key feature to enhance emotion recognition from text. Consequently, they compared the performance of Google word embeddings, GloVe embeddings, and FastText embeddings using their ensemble of BiLSTM, CNN,

[1] The Cancer Survivors' Network of the American Cancer Society

and Self-Attention model. Evaluation was conducted on multiple datasets including ISEAR, SemEval-2018 Task 1, and SemEval-2019 Task 3, with FastText embeddings showing superior performance across all datasets.

Alotaibi [2] proposed a method employing supervised logistic regression to identify emotions from text. They utilized data from ISEAR, splitting it into training and testing sets. During training, sentences containing emotions were inputted into their logistic regression model and emotion labels. Subsequently, only unseen sentences labeled with emotions were passed through the trained classifier during the testing phase for prediction.

Adoma et al [1] investigated the effectiveness of BERT, RoBERTa, DistillBERT, and XLNet in recognizing emotions from text. Their study concluded that RoBERTa achieved the highest accuracy in emotion recognition on the ISEAR dataset.

Table 1 presents the results of the within-domain classification experiments conducted on the ISEAR dataset.

Table 1. The results of the classification experiments conducted on the ISEAR dataset. All scores represent of F1 scores(in %).

Model	ANGER	DISGUST	FEAR	GUILT	JOY	SADNESS	SHAME
Alotaibi [2] (Logistic Regression)	-	-	64	57	76	73	62
Adoma et al [1] (BERT)	57	67	75	67	88	78	60
Adoma et al [1] (RoBERTa)	62	73	80	68	93	79	65
Adoma et al [1] (DistilBERT)	55	67	73	61	85	78	52
Adoma et al [1] (XLNET)	58	71	78	71	92	79	63
Polignano et al [15] (GoogleEmb)	54	62	74	56	76	65	51
Polignano et al [15] (GloVeEmb)	52	63	72	55	76	66	50
Polignano et al [15] (FastTextEmb)	55	66	71	57	78	65	52

3. Methods

ML techniques for recognizing emotions in text may be categorized into three categories: classical ML, DL, and LLMs. Figure 3 illustrates the general concept of emotion recognition in text and the specific methods we explored in our study. In classical ML and DL approaches, the journey of detecting emotions starts from a preprocessing phase and is fed to the model; finally, the model's output will represent the corresponding class. In the LLMs approach, the preprocessing step is not considered because LLMs are pretrained on extensive datasets and can understand nuanced meanings within text without extensive preprocessing. We employed various classical ML and DL methods to explore emotion detection techniques to ensure comprehensive analysis and comparison. Specifically, we implemented Bag of Words, TFIDF, Fasttext, and BERT techniques for text representation to preprocess and analyze text data to achieve the highest accuracy in recognizing emotions.

Turning our attention to LLMs, we specifically evaluated the Falcon 7B and Mistral 7B models. Initially, we applied these pre-trained LLMs through one-shot learning [6] technique by passing an instance of each emotion sample to the model and asking it to categorize the given text. We did this experience in two different ways, In the first test, we passed the one-shot samples in the order of *JOY*, *FEAR*, *ANGER*, *SADNESS*, *DISGUST*, *SHAME*, *GUILT*, (referred to as Method 1), and in the second experience, we reversed the order by *GUILT*, *SHAME*, *DISGUST*, *SADNESS*, *ANGER*, *FEAR*, *JOY* (referred to as Method 2). We hypothesize that the order of the given text to the LLM models as one-shot learning would affect the output prediction. Further examined a fine-tuned version of the Mistral 7B model.

3.1. Dataset

The ISEAR dataset [16], constructed through cross-cultural questionnaire studies in 37 countries by Scherer et al., consists of 7666 sentences annotated with seven distinct emotion labels: *JOY* (1094), *ANGER* (1096), *SADNESS* (1096), *SHAME* (1096), *GUILT* (1093), *DISGUST* (1096), and *FEAR* (1095). This dataset comes from the answers of 1096 people from a wide range of cultural backgrounds. It has a balanced class distribution across its emotion labels, which can be used for generalized predictive inferences.

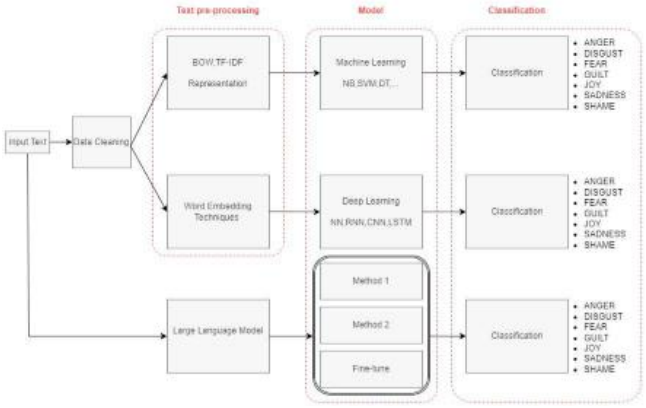

Fig. 1. Methodology for automatic Emotion Detection

4. Experiments

Experiments were carried out using Kaggle[2] GPU hardware accelerator platform. The dataset's initial preprocessing involved removing non-alphabetic characters (excluding '.', '!', '?'), extra spaces, and duplicates, resulting in 7448 unique instances. After excluding 35 entries lacking valuable content, this led to 7413 refined instances for analysis. Subsequent steps for classical machine learning included removing stopwords, tokenizing words, filtering non-alphanumeric characters and stopwords, normalizing text through lowercasing, and lemmatization to consolidate word variations, all aimed at enhancing data analysis.

We divided our experiments into two categories: the classical ML techniques and the LLM techniques. In the classical machine learning techniques, we consistently separated 80% of the data as the training set and tested the model with the remaining data. We determined the hyperparameters for each model using the grid search strategy [18, 3]. In the LLM techniques, we used a fraction of 20% of each emotion category in the dataset to give an equal chance for all the categories. Then the selected instances were given to the LLM models by the one-shot learning approach with Method 1 and 2. We prepared a unique prompt for each instance to ask the LLM to predict its sentiment.

During the experimental evaluation, optimal hyperparameters were determined for the various ML techniques. For Multinomial Naive Bayes using Bag of Words (BOW), the most effective settings included an `ngram_range` of (1, 2) for capturing both unigrams and bigrams, along with an `alpha` value of 1.5 for smoothing. Similarly, Random Forest with BOW performed best with an `ngram_range` of (1, 2) and specific parameters such as `max_depth` set to None, `min_samples_leaf` of 2, `min_samples_split` of 2, and `n_estimators` of 200 for ensemble learning. TF-IDF-based Support Vector Machine (SVM) excelled with an `ngram_range` of (1, 2), `C` value of 10 for regularization, and an `rbf` kernel with `gamma` set to 'scale'. Neural Network with TF-IDF achieved optimal performance with an `alpha` of 0.01 for regularization, a single hidden layer of size 100, and 200 maximum iterations. For models utilizing BERT embeddings, such as SVM, the best settings included a `C` value of 10, an `rbf` kernel, and 'auto' gamma. Lastly, SVM with Fasttext benefited from a `C` value of 10, `rbf` kernel, and 'scale' gamma for effective classification. These identified hyperparameter configurations represent the most successful settings discovered during experimentation, optimizing performance for each machine learning technique.

5. Results

This section analyzes the efficiency of various machine learning algorithms across different feature extraction methods, followed by the LLM approach.

The results obtained through the utilization of Bag of Words, TFIDF, Fasttext, and BERT are provided in Tables 2, 3, 4, and 5, respectively.

[2] https://www.kaggle.com/

Table 2. Results by applying Bag Of Words. All scores represent P: Precision, R: Recall, and F: F1 scores (in %). S represents Support.

Model	ANGER				DISGUST				FEAR				GUILT				JOY				SADNESS				SHAME				Acc %
	P	R	F	S	P	R	F	S	P	R	F	S	P	R	F	S	P	R	F	S	P	R	F	S	P	R	F	S	
MNB	39	**53**	45	206	**70**	54	**61**	226	59	**74**	**66**	184	54	**46**	**50**	224	**72**	72	**72**	230	63	57	60	205	50	46	48	208	**57**
SVM	43	46	44	206	59	57	58	226	62	62	62	184	53	44	48	224	63	74	68	230	**67**	58	**62**	205	46	**49**	48	208	56
DT	33	34	34	206	50	50	50	226	45	64	53	184	43	34	38	224	54	61	57	230	57	49	53	205	44	37	40	208	47
RF	**45**	47	**46**	206	58	**61**	59	226	57	70	63	184	**57**	41	48	224	59	**78**	67	230	64	55	59	205	**56**	44	49	208	**57**
GB	31	50	38	206	68	55	**61**	226	**64**	65	65	184	51	42	46	224	65	71	68	230	60	53	56	205	**56**	43	49	208	54
NN	43	42	43	206	64	56	60	226	57	64	61	184	52	38	44	224	62	77	69	230	54	**60**	57	205	51	**49**	**50**	208	55

Table 3. Results by applying TFIDF. All scores represent Precision, Recall, and F1 scores (in %).

Model	ANGER				DISGUST				FEAR				GUILT				JOY				SADNESS				SHAME				Acc %
	P	R	F	S	P	R	F	S	P	R	F	S	P	R	F	S	P	R	F	S	P	R	F	S	P	R	F	S	
MNB	39	50	44	206	**69**	54	61	226	57	**76**	65	184	54	41	47	224	**69**	73	71	230	62	59	**60**	205	52	47	49	208	57
SVM	42	**53**	**47**	206	64	**60**	62	226	59	72	65	184	**56**	**42**	**48**	224	68	74	71	230	**64**	57	**60**	205	54	48	**51**	208	**58**
DT	35	38	37	206	52	47	49	226	42	60	49	184	40	35	37	224	55	58	56	230	59	49	53	205	42	38	40	208	46
RF	43	42	43	206	62	**60**	61	226	57	70	63	184	**56**	39	46	224	56	76	65	230	56	59	58	205	**59**	46	**51**	208	56
GB	33	46	38	206	62	56	59	226	**64**	66	65	184	50	40	45	224	64	71	67	230	61	53	57	205	52	47	49	208	54
NN	**45**	49	**47**	206	68	59	**63**	226	62	71	**66**	184	**56**	41	47	224	67	**77**	**72**	230	57	**62**	59	205	51	**49**	50	208	**58**

Table 4. Results by applying Fasttext Sentence Embedding. All scores represent Precision, Recall, and F1 scores (in %).

Model	ANGER				DISGUST				FEAR				GUILT				JOY				SADNESS				SHAME				Acc %
	P	R	F	S	P	R	F	S	P	R	F	S	P	R	F	S	P	R	F	S	P	R	F	S	P	R	F	S	
SVM	**31**	**40**	**35**	206	**47**	**48**	**47**	226	**46**	**57**	**50**	184	**39**	31	34	224	**50**	47	**48**	230	**49**	**42**	**45**	205	**40**	**35**	**37**	208	**43**
DT	17	26	20	206	20	21	20	226	21	24	22	184	17	15	16	224	25	17	21	230	23	19	21	205	19	17	18	208	20
RF	24	33	28	206	32	36	34	226	28	39	33	184	32	27	30	224	43	38	40	230	31	24	27	205	26	18	22	208	31
GB	25	32	28	206	36	36	36	226	37	42	39	184	33	29	31	224	46	43	45	230	35	34	34	205	32	28	30	208	35
NN	**31**	38	**35**	206	44	45	44	226	43	**57**	49	184	**39**	**37**	**38**	224	45	**51**	**48**	230	46	38	42	205	38	20	26	208	41

Table 5. Results by applying BERT. All scores represent Precision, Recall, and F1 scores (in %).

Model	ANGER				DISGUST				FEAR				GUILT				JOY				SADNESS				SHAME				Acc %
	P	R	F	S	P	R	F	S	P	R	F	S	P	R	F	S	P	R	F	S	P	R	F	S	P	R	F	S	
SVM	**46**	**51**	**48**	206	**70**	**61**	**65**	226	**67**	**77**	**72**	184	**57**	**51**	**54**	230	**83**	**90**	**86**	230	**68**	**69**	**68**	205	**55**	**51**	**53**	208	**64**
DT	19	30	23	206	38	27	31	226	33	28	30	184	24	29	27	224	57	43	50	230	40	27	33	205	21	25	23	208	30
RF	32	38	35	206	53	51	52	226	62	67	65	184	42	33	37	224	68	90	77	230	62	57	59	205	42	31	36	208	53

5.1. Multinomial Naive Bayes

The Multinomial Naive Bayes algorithm exhibits consistent performance across different feature extraction methods. Whether using Bag of Words or TFIDF techniques, the algorithm achieves an accuracy of 57%. This consistency suggests that Multinomial Naive Bayes is robust and relatively insensitive to the choice of feature extraction method.

5.2. Support Vector Machine

The Support Vector Machine (SVM) algorithm demonstrates varying levels of efficiency depending on the feature extraction method. When applied with Bag of Words, TFIDF, BERT, and Fasttext Sentence Embedding techniques, SVM achieves accuracies of 56%, 64%, 58%, and 43%, respectively. Notably, SVM performs best with BERT, indicating that the BERT technique captures relevant information more effectively for this algorithm compared to the other methods.

5.3. Decision Tree

The Decision Tree algorithm exhibits fluctuating efficiency across different feature extraction methods. With Bag of Words and BERT techniques, the algorithm achieves accuracies of 47% and 30%, respectively. With Fasttext achieves accuracy of 20%. However, its performance improves to 46% when using TFIDF. This suggests that TFIDF captures features more conducive to Decision Tree classification than Bag of Words and BERT techniques.

5.4. Random Forest

Random Forest demonstrates varying levels of efficiency across feature extraction methods. With Bag of Words and TFIDF, the algorithm achieves accuracies of 57% and 56%, respectively. When using BERT and Fasttext sentence embeddings, its accuracy drops slightly to 53% and 31%. This indicates that Bag of Words and TFIDF techniques provide more discriminative features for Random Forest than BERT.

5.5. Gradient Boosting

Gradient Boosting exhibits consistent performance across Bag of Words and TFIDF techniques, achieving accuracies of 54% in both cases. However, its efficiency drops to 35% with Fasttext Sentence Embedding. This suggests that Gradient Boosting performs well with traditional feature extraction methods but may struggle with embedding techniques like Fasttext.

5.6. Neural Network

The Neural Network algorithm shows varying levels of efficiency across different feature extraction methods. With Bag of Words, TFIDF, and Fasttext, the algorithm achieves accuracies of 55%, 58%, and 41%, respectively. Notably, it performs best with TFIDF, indicating that TFIDF captures features more conducive to Neural Network classification than Bag of Words and Fasttext.

Method 1 and Method 2, were employed on Falcon 7B and Mistral 7B models, using identical data sets of 1483 instances to assess their effectiveness. However, the outputs generated by these models were often ambiguous, lacking clarity, and did not align perfectly with the desired emotional categories (*JOY*, *FEAR*, *ANGER*, *SADNESS*, *DISGUST*, *SHAME*, *GUILT*). Consequently, after each trial, the outputs were reviewed and labeled manually whenever feasible. These labeled outputs were then compared against actual labels to gauge accuracy.

5.7. Falcon 7B and Mistral 7B

The comparison of the Falcon 7B and Mistral 7B models reveals distinct differences in their performance on emotion classification tasks. Falcon 7B had 661 and 418 instances labeled as unknown in Methods 1 and 2. In contrast, Mistral 7B had significantly fewer unknown instances, with 267 and 204 in Methods 1 and 2. Mistral 7B also outperformed Falcon 7B in terms of overall accuracy, achieving 52% and 54% accuracy in Methods 1 and 2, compared to Falcon 7B's 23% and 25%. Notably, Mistral 7B excelled in classifying emotions such as *JOY* and *FEAR*, with higher F1-scores: 78% and 79% in Methods 1 and 2 for *JOY*, and 58% and 63% in Methods 1 and 2 for *FEAR*. These results suggest that Mistral 7B is a more effective and reliable model for emotion detection with ISEAR dataset, offering enhanced performance and robustness over Falcon 7B.

5.8. Mistral 7B Fine-tuned

In this study, we utilized QLoRA [8], the latest Parameter Efficient Fine-Tuning (PEFT) [12] technique available at the time of writing, to fine-tune the Mistral 7B model. QLoRA allows for efficient fine-tuning by optimizing only a subset of parameters, reducing computational requirements and improving performance. For this experimentation, 80% of the dataset was allocated for model fine-tuning, while the remaining 20% was reserved for testing purposes. Notably, the test dataset matched the data used to test Method 1 and Method 2 with pre-trained models. Unlike

previous experiences where output revision was required, in this case, there was no necessity for revising the model's output as it was specifically trained to provide responses based on predetermined categories.

Table 6 shows the results after training for 4 epochs and testing the model over the mentioned test data set.

Table 6. Results by applying Mistral 7B Instruct Fine-tune.

	precision	recall	f1-score	support
ANGER	0.69	0.69	0.69	216
DISGUST	0.75	0.75	0.75	214
FEAR	0.78	0.84	0.81	214
GUILT	0.67	0.70	0.69	211
JOY	0.95	0.94	0.95	213
SADNESS	0.81	0.78	0.79	207
SHAME	0.66	0.62	0.64	208
Accuracy				**76%**

6. Discussion

To consider the results obtained using LLMs (Large Language Models), it is important to first consider the performance of classical machine learning techniques across various feature extraction methods. This allows us to assess traditional approaches' relative strengths and weaknesses before discussing the impact of LLMs on the same tasks.

Among the machine learning algorithms analyzed, Multinomial Naive Bayes demonstrated consistent performance irrespective of the feature extraction method used, achieving an accuracy of 57% across Bag of Words and TFIDF techniques. Notably, this algorithm's robustness suggests that it is relatively insensitive to changes in feature representation. Other classical algorithms like Support Vector Machines (SVM) and Decision Trees exhibited varying efficiencies based on the feature extraction method employed, with SVM performing optimally with BERT embeddings (64% accuracy) and Decision Trees showing a preference for TFIDF features (46% accuracy).

Transitioning to the impact of LLMs, Falcon 7B's performance, particularly with Method 1 (23% accuracy), highlighted mixed outcomes across different emotion categories. Notably, it achieved higher precision in identifying *GUILT* instances but struggled with recall, suggesting limitations in capturing guilt-related nuances. Conversely, Mistral 7B demonstrated robust performance across various emotion categories, especially with Method 1 (52% accuracy), showcasing high precision and recall for *JOY* and balanced trade-offs for other emotions. With Method 2, Mistral 7B (54% accuracy) showed similar trends in performance compared to Method 1 (52% accuracy), albeit with slight variations. Notably, it exhibited improvements in recall for *ANGER* and *SADNESS* compared to Method 1, suggesting its enhanced capability to identify instances of these emotions comprehensively. However, the model's performance for *DISGUST* deteriorated slightly with Method 2. Overall, Mistral 7B's performance remained robust across different emotion categories with Method 2, showcasing its consistency. The Mistral 7B Instruct Fine-tune model achieved an accuracy of 76%, which is notably higher than the accuracies reported for Falcon 7B and Mistral 7B. Additionally, the precision, recall, and F1-score values for each emotion category demonstrate robust performance across the board, especially for emotions like *JOY* and *FEAR*. Compared to the earlier techniques described in our study (Table 1), Mistral 7B Instruct Fine-tune performs better in all emotion categories. In particular, our approach outperforms the top F1 results for *JOY* and *FEAR*. For example, our method outperformed the best previous score of 80 by [1], achieving an F1 score of 81 for *FEAR*. Notably, our approach also surpassed the previous best of 93, set by the same study, achieving an impressive F1 score of 95 for *JOY*.

7. Conclusion and Future Work

Our study compared classical machine learning techniques using various feature extraction methods and highlighted the impact of different approaches on performance. Additionally, we focused on the power of LLMs for detecting emotions in texts by one-shot learning technique and examined how altering the sequence of in-context learn-

ing affects perception. Notably, we found that within the Mistral 7B, Method 2 consistently outperformed Method 1 across multiple tests.

When evaluating Mistral 7B against Falcon 7B models, Mistral 7B consistently demonstrated superior performance, particularly when fine-tuned. Notably, Mistral 7B excelled in emotion analysis tasks, showcasing the potential of advanced Large Language Models (LLMs) like Mistral 7B in natural language processing (NLP).

These findings contribute valuable insights to the field of machine learning and NLP. The performance of Mistral 7B in sentiment analysis highlights the effectiveness of fine-tuned LLMs on practical NLP applications, suggesting broader implications for sentiment analysis and emotion recognition tasks.

Future studies should explore alternative sequences of one-shot learning to deepen our understanding of their impact on model performance. Investigating multi-shot learning techniques could provide nuanced insights into model robustness and generalization abilities. Exploring the consequences of selecting different subsets of data for fine-tuning methods will enhance scalability and adaptability to diverse datasets and scenarios. Furthermore, extending fine-tuning processes through increased training epochs holds promise for optimizing model performance and achieving superior results.

Acknowledgement

We acknowledge the support of the Natural Sciences and Engineering Research Council of Canada (NSERC) [funding reference number 06351]).

References

[1] Adoma AF, Henry NM, C.W., 2020. Comparative analyses of bert, roberta, distilbert, and xlnet for textbased emotion recognition. 17th international computer conference on wavelet active media technology and information processing (ICCWAMTIP). IEEE , 117—121.
[2] Alotaibi, F.M., 2019. Classifying text-based emotions using logistic regression .
[3] Bergstra, J., Bardenet, R., Bengio, Y., Kégl, B., 2011. Algorithms for hyper-parameter optimization. Advances in neural information processing systems 24.
[4] Berridge, C., Zhou, Y., Kaye, J., 2023. Companion robots to mitigate loneliness among older adults: Perceptions of benefit and possible deception. Frontiers in Psychology 14, 1106633.
[5] Bharti, S.K., Varadhaganapathy, S., Gupta, R.K., Shukla, P.K., Bouye, M., Hingaa, S.K., Mahmoud, A., et al., 2022. Text-based emotion recognition using deep learning approach. Computational Intelligence and Neuroscience 2022.
[6] Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J.D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., et al., 2020. Language models are few-shot learners. Advances in neural information processing systems 33, 1877–1901.
[7] Chatterjee, A., Gupta, U., Chinnakotla, M.K., Srikanth, R., Galley, M., Agrawal, P., 2019. Understanding emotions in text using deep learning and big data. Computers in Human Behavior 93, 309–317.
[8] Dettmers, T., Pagnoni, A., Holtzman, A., Zettlemoyer, L., 2024. Qlora: Efficient finetuning of quantized llms. Advances in Neural Information Processing Systems 36.
[9] Ekman, P., 2016. What scientists who study emotion agree about. Perspectives on psychological science 11, 31–34.
[10] Ekman, P., et al., 1999. Basic emotions. Handbook of cognition and emotion 98, 16.
[11] Grandjean, D., Sander, D., Scherer, K.R., 2008. Conscious emotional experience emerges as a function of multilevel, appraisal-driven response synchronization. Consciousness and cognition 17, 484–495.
[12] Houlsby, N., Giurgiu, A., Jastrzebski, S., Morrone, B., De Laroussilhe, Q., Gesmundo, A., Attariyan, M., Gelly, S., 2019. Parameter-efficient transfer learning for nlp, in: International Conference on Machine Learning, PMLR. pp. 2790–2799.
[13] Khanpour, H., Caragea, C., 2018. Fine-grained emotion detection in health-related online posts, in: Proceedings of the 2018 conference on empirical methods in natural language processing, pp. 1160–1166.
[14] Kratzwald, B., Ilić, S., Kraus, M., Feuerriegel, S., Prendinger, H., 2018. Deep learning for affective computing: Text-based emotion recognition in decision support. Decision support systems 115, 24–35.
[15] Polignano, M., Basile, P., de Gemmis, M., Semeraro, G., 2019. A comparison of word-embeddings in emotion detection from text using bilstm, cnn and self-attention, in: Adjunct Publication of the 27th Conference on User Modeling, Adaptation and Personalization, pp. 63–68.
[16] Scherer, K.R., Wallbott, H.G., 1994. Evidence for universality and cultural variation of differential emotion response patterning. Journal of personality and social psychology 66, 310.
[17] Szabóová, M., Sarnovskỳ, M., Maslej Krešňáková, V., Machová, K., 2020. Emotion analysis in human–robot interaction. Electronics 9, 1761.
[18] Yang, L., Shami, A., 2020. On hyperparameter optimization of machine learning algorithms: Theory and practice. Neurocomputing 415, 295–316.

CONCLUSION GÉNÉRALE

Le développement d'assistants virtuels empathiques est un domaine d'intérêt croissant, axé sur la compréhension et l'intégration des états affectifs des utilisateurs dans les interactions avec les assistants virtuels. Dans ce travail, nous avons exploré la détection et l'intégration des états émotionnels des personnes lors de conversations avec des assistants virtuels, en mettant particulièrement l'accent sur la détection des émotions à travers le texte, considérée comme une étape fondamentale.

OBJECTIFS ATTEINTS

Tout au long de ce projet, plusieurs objectifs clés ont été identifiés et atteints avec succès. Ces réalisations fournissent une base solide pour le développement d'assistants virtuels empathiques capables de comprendre et de répondre aux états émotionnels des utilisateurs.L'objectif principal de ce projet était de développer un système capable de détecter les émotions à partir d'entrées textuelles. En exploitant diverses techniques d'apprentissage automatique et des GML comme Mistral 7B, nous avons réalisé des progrès significatifs dans l'identification précise des émotions telles que la joie, la peur, la colère, la tristesse, le dégoût, la honte et la culpabilité à partir des entrées textuelles des utilisateurs. Le réglage fin du modèle Mistral 7B, notamment grâce à la technique QLoRA, a permis d'obtenir une précision notable de 76 %, surpassant ainsi d'autres modèles comme Falcon 7B.Un autre objectif clé était d'évaluer les performances des algorithmes d'apprentissage automatique classiques (ex. : MNB, SVM) par rapport aux GML avancés dans le contexte de la reconnaissance des émotions. Nos résultats ont mis en évidence les performances supérieures des GML affinés, en particulier dans les applications pratiques du traitement du langage naturel pour l'analyse des sentiments.

PERSPECTIVES ET TRAVAUX FUTURS

Les résultats de ce travail fournissent une base pour de futures avancées dans le domaine des AV empathiques. Les travaux futurs devraient explorer diverses pistes pour améliorer les capacités et la robustesse du système.

L'étude de l'impact de techniques d'apprentissage alternatives, telles que l'apprentissage multi-coups (en anglais : multishot learning), pourrait offrir des informations plus approfondies sur les performances des modèles et leurs capacités de généralisation. Comprendre comment différentes séquences d'apprentissage en contexte affectent la perception et la précision sera crucial pour développer des systèmes de détection des émotions plus sophistiqués.

L'élargissement du processus de réglage fin à des ensembles de données et à des scénarios plus variés améliorera l'évolutivité et l'adaptabilité du système. En explorant la sélection de différents sous-ensembles de données pour le réglage fin, le modèle peut atteindre de meilleures performances dans diverses applications, le rendant plus polyvalent et efficace dans des environnements réels.

L'augmentation du nombre d'itérations d'entraînement peut optimiser davantage les performances du modèle, ce qui conduit à de meilleurs résultats dans la détection des émotions. Cette approche permettrait d'affiner la capacité du modèle à capturer et à répondre plus efficacement à des états émotionnels nuancés, améliorant ainsi les réponses empathiques des AV.

La détection des émotions par le texte combiné à des entrées multimodales telles que la voix ou les expressions faciales pourrait aussi permettre une compréhension plus complète de l'état affectif de l'utilisateur. Avec une approche plus complète, les assistants virtuels peuvent réagir avec plus d'empathie, ce qui se traduit par des interactions plus intuitives et naturelles.

Le domaine des GML connaît des avancées rapides et de nouveaux modèles émergent régulièrement, souvent même sur une base hebdomadaire [11]. Ces développements transforment le paysage du traitement du langage naturel et de la reconnaissance des émotions à partir du texte. Étant donné la rapidité de ce progrès, il est essentiel que les recherches futures dans ce domaine intègrent et évaluent ces GML émergents. En analysant leurs performances sur des corpus spécialisés en reconnaissance des émotions, les chercheurs pourront identifier les modèles les plus adaptés à l'extraction des signaux émotionnels et comparer leur capacité à reconnaître des émotions subtiles ou complexes.

RÉFÉRENCES BIBLIOGRAPHIQUES

[1] Goleman, Daniel. *Emotional intelligence: Why it can matter more than IQ*. Bloomsbury Publishing, 2020.

[2] Derksen, Frans, Jozien Bensing, and Antoine Lagro-Janssen. "*Effectiveness of empathy in general practice: a systematic review.*" *British journal of general practice* 63.606 (2013): e76-e84.

[3] Wei, Chuang, Maggie Wenjing Liu, and Hean Tat Keh. "The road to consumer forgiveness is paved with money or apology? The roles of empathy and power in service recovery." *Journal of Business Research* 118 (2020): 321-334.

[4] Devlin, Jacob. "Bert: Pre-training of deep bidirectional transformers for language understanding." *arXiv preprint arXiv:1810.04805* (2018).

[5]Bojanowski, Piotr, et al. "Enriching word vectors with subword information." *Transactions of the association for computational linguistics* 5 (2017): 135-146.

[6] Pedregosa, Fabian, et al. "Scikit-learn: Machine learning in Python." *the Journal of machine Learning research* 12 (2011): 2825-2830.

[7] Aridoss, Manimaran, Khushwant Singh Bisht, and Arul Kumar Natarajan. "Comprehensive Analysis of Falcon 7B: A State-of-the-Art Generative Large Language Model." *Generative AI: Current Trends and Applications*. Singapore: Springer Nature Singapore, 2024. 147-164.

[8] Jiang, Albert Q., et al. "Mistral 7B." arXiv preprint arXiv:2310.06825 (2023).

[9] Dettmers, Tim, et al. "Qlora: Efficient finetuning of quantized llms." *Advances in Neural Information Processing Systems* 36 (2024).

[10] Esfahani, Seyed Hamed Noktehdan, and Mehdi Adda. "*Classical Machine Learning and Large Models for Text-Based Emotion Recognition.*" *Procedia Computer Science* 241 (2024): 77-84.

[11] Hugging Face, « Open LLM Leaderboard », Consulté le 19 décembre 2024. [En ligne]. Disponible sur : *https://huggingface.co/spaces/open-llm-leaderboard/open_llm_leaderboard#/*