

17/004.568

République Algérienne Démocratique et Populaire

Ministère de l'enseignement supérieur et de la recherche scientifique

Université de 8 Mai 1945 - Guelma -

Faculté des Mathématiques, d'Informatique et des Sciences de la matière

Département d'Informatique



Mémoire de fin d'études Master

Filière : Informatique

Option : Système informatique

Thème :

Vers une classification fine de la subjectivité à l'aide d'une modélisation floue

Encadré Par :

Dr. Farek Lazher

Présenté par :

Amirouche Abdel Djalil

Juin 2018

Résumé

De nos jours, le commerce électronique se développe rapidement, de sorte que les critiques de produits ont rapidement augmenté sur le Web. Le grand nombre de revues rend difficile pour les fabricants ou les entreprises de les classer automatiquement dans différentes orientations sémantiques (positives, négatives et neutres). La plupart des méthodes existantes utilisent une liste de mots d'opinion pour la classification des sentiments. Nous proposons un modèle de logique floue pour effectuer des classifications sémantiques sur des critiques introduites par des clients dans les sous-classes suivantes: très faible, faible, modérée, forte et très forte.

La logique floue, contrairement aux techniques d'exploration de données statistiques, permet non seulement d'utiliser des valeurs non numériques mais introduit également la notion de variables linguistiques. Utilisation de termes et de variables linguistiques résultera en un processus d'interrogation plus humain.

Les résultats obtenus nous ont mené à découvrir les facteurs qui ont influencé la performance de notre système, qui seront sujets d'éventuelles améliorations dans nos futurs travaux de recherche.

Table des matières

RESUME	1
INTRODUCTION GENERALE.....	5
CHAPITRE 1. FOUILLE D'OPINION : ETAT DE L'ART	9
1.1 INTRODUCTION.....	10
1.2 OPINION, FAIT ET SENTIMENT:	10
1.2.1 Une opinion :.....	10
1.2.2 Les sentiments	10
1.2.3 Les faits.....	11
1.3 SOUS-DOMAINES DE RECHERCHE :	11
1.4 APPROCHE DE L'ANALYSE DES SENTIMENTS ET LA DETECTION D'OPINIONS :	12
1.4.1 Approches d'apprentissage automatique :	12
1.4.2 Approche Lexique :	13
1.4.3 Approches hybride :	13
1.5 DEFINITION DE FOUILLE D'OPINION :	13
1.6 TRAITEMENT AUTOMATIQUE DES LANGAGES NATURELS (TALN)	14
1.7 RECHERCHE D'INFORMATION (RI) :	15
1.8 DOMAINE D'APPLICATION DE L'ANALYSE DES SENTIMENTS :	15
1.8.1 Les entreprises :	16
1.8.2 Les clients :	16
1.8.3 Politique :	16
1.9 COMPLEXITE DE NOTION D'OPINION	16
1.10 FACTEURS DE DIFFICULTE DE LA FOUILLE D'OPINION :	18
CHAPITRE 2. CONCEPTS FONDAMENTAUX DE LA THEORIE DE LA LOGIQUE FLOUE	21
2.1. INTRODUCTION :	22
2.2. LA DEFERENCE ENTRE LA LOGIQUE FLOUE ET LOGIQUE CLASSIQUE :	22
2.3. ENSEMBLES FLOUS :	23
2.4. HAIES :	26
2.5. OPERATEURS FLOUS :	27
2.6. REGLES FLOUES :	28
2.7. FUZZIFICATION :	29
2.8. DEFUZZIFICATION :	30
3.1 INTRODUCTION :	38
3.2 ÉTUDE CONNEXES :	38
3.3 APPROCHE PROPOSE :	39
3.4 PRINCIPALES CARACTERISTIQUES :	40
3.4.1 Phase de prétraitement :	40
3.4.2 Phase d'extraction de caractéristique :	40
3.5 PHASE DE CLASSIFICATION DE CARACTERISTIQUE :	41
3.5.1 Fuzzification des variables d'entrée :	41
3.5.3 Agrégation:	45
3.6 DEFUZZIFICATION :	46
3.7 CONCLUSION	47
CHAPITRE 4 : IMPLEMENTATION	49
4.1 INTRODUCTION :	50
4.2 PRESENTATION DES OUTILS :	50

4.2.1 Le langage de programmation python 3.6 :	50
4.2.2 L'éditeur spyder :	51
4.2.3 La bibliothèque NLTK :	51
4.2.4 La bibliothèque skfuzzy :	52
4.2.5 Bibliothèque textblob :	52
4.2.6 La bibliothèque numpy :	52
4.2.7 La phase de prétraitement :	52
FIGURE 4.1 : LA SUPPRESSION DE PONCTUATIONS ET LA REPETITION DES LETTRE	53
4.3 L'EXTRACTION DES CARACTERISTIQUES :	53
4.3.1 Tokenisation et catégorisations :	53
4.3.3 extraction score :	53
4.3.4 Dépendance grammatical :	54
4.4 Phase de classification de caractéristique :	54
4.4.1 La Fuzzification des caractéristiques :	54
LISTE DES FIGURES :	62
LISTE DES TABLEAUX	63

Introduction générale

L'analyse des sentiments (ang. Sentiment Analysis), ou opinion mining (OM), est une notion beaucoup évoquée mais souvent mal comprise, Il s'agit d'un processus qui permet de déterminer la tonalité émotionnelle qui se cache derrière une série de mots. Cette analyse est utilisée pour mieux comprendre la perception, les opinions et les émotions exprimées dans une mention en ligne, cette analyse également appelée analyse de subjectivité (ang. Subjectivity Analysis) réunit plusieurs disciplines : Traitement automatique des langages naturels (TALN), recherche d'information (RI), et là dont le but de classifier ces sentiments en deux catégories sémantique : positive ou négative, cette identification des sentiments ou bien des opinions nécessite une analyse lexicale et syntaxique profonde.

La quantité énorme des informations disponibles notamment sur le Web rend aujourd'hui les systèmes actuels chaque jour plus préoccupés par la manière de gérer la surcharge d'information, en s'assurant que l'utilisateur aura l'accès aux meilleures sources avec le moindre effort. Ces dernières années, une attention particulière a été donnée pour résoudre ce problème. Le secteur de e-commerce est l'un des plus influencé par la quantité de données produites par les clients sous forme de commentaires et critiques sur les produits commercialisés, cette augmentation a été remarquée surtout après l'apparence du Web 2.0.

Actuellement il y a une grande quantité d'information sur le web qui rend les système actuels plus préoccupés par la gérance de la surcharge de ces information dont le but est de faciliter de trouver la meilleur source demandée par l'utilisateur, par exemple le secteur de e-commerce a un énorme surcharge de données produites par les client sous forme de commentaires et des critiques sur les produits commercialisés.

Les opinions des clients forment une source d'informations qu'il ne faut pas être maltraitée ou ignorée par la communauté de recherche, donc le développement des approches qui sont capables de profiter le maximum de ces données est nécessaire pour exploiter ces données en vue de faciliter la prise de décision rapidement en moindre coût.

Ce travail souligne la nécessité de développer une approche capable de profiter pleinement de ces données, dont l'objectif est classifier les sources subjectives en passant par un niveau de

granularité plus fine. Au lieu classer une source subjective (phrase ou document) en positive ou négative (classification binaire), nous raffinons les deux catégories en plusieurs sous-catégories, par exemple : positive (négative) faible, positive (négative) moyen, positive (négative) fort, ou encore : très faible, faible, moyen, fort, très fort. Cette catégorisation fine permet de voir en détail et évaluer adéquatement les orientations positives et négatives de chaque source.

Dans ce contexte, nous proposons d'employer les capacités de logique floue (eng ; fuzzylogic) qui nous servira à calculer l'orientation globale d'une source subjective, et la classer selon son intensité à plusieurs niveaux de granularité.

A) Motivation

Du point de vue des clients avant d'acheter un produit, donner une considération aux opinions des autres acheteurs est un comportement commun même avant l'existence de l'internet, dans l'époque numérique le client a l'accès d'un milliers d'opinions ce qui lui aide pour prendre une décision précise pour ce produit, les clients veulent toujours de trouver le meilleur produit avec un prix raisonnable.

L'importance de souligner ces opinions c'est qu'elles sont des opinions des personnes neutres qui représentent la voix des consommateurs ordinaires et ils ne sont pas lié à une organisation ou une entreprise, ces opinions sont généralement déférentes aux annonces publicitaires qui favorise le produit.

Du point de vue du commerce électronique (*ang. e-commerce*), les opinion des clients peuvent être utilisées de manière où elle peuvent améliorer le produit selon la satisfaction des client afin d'augmenter les bénéfices du secteur.

Parfois, nous trouvons des produits qui ont des milliers d'opinions, donc il pourrait être une tâche difficile au client de prendre une décision précise ou même il peut être ennuyant de lire toutes ces opinions sur quelques caractéristiques particulières dans le cas d'une exigence d'un client expérimenté.

Les techniques de classification actuelles ne sont pas assez efficaces pour décrire des informations qui représentent des opinions ou des faits, les informations qui représentent des faits sont généralement les mêmes, cependant les opinions sont des jugements ou pensées personnels sur un sujet ou un objet, qui sont généralement différentes aux autres. Dans ce cas, un mécanisme de résumé est indispensable pour rendre les opinions facilement exploitables.

B) Organisation du mémoire

Dans notre travail nous proposons de transformer la classification binaire classique (positive et négative) des sentiments en classification multi-classes en raffinant les deux catégories positive et négative en plusieurs sous-catégories, et cela dans le but de donner une vision plus détaillée sur les orientations positives et négatives exprimées par les clients.

Notre travail est divisé en quatre (4) chapitres organisés comme suit :

– Chapitre 1.

Ce chapitre est un état de l'art sur les techniques utilisées en fouille d'opinion : Nous commençons par la définition des différents concepts tels qu'opinion, sentiment, émotion, fait, et les liens entre eux et certains d'autres concepts comme la polarité et l'intensité de l'opinion. Nous nous concentrons sur les techniques de classification de la subjectivité utilisées dans le domaine.

– Chapitre 2.

Dans le deuxième chapitre nous introduisons le principe de fonctionnement des systèmes à base de la logique floue, en premier lieu, une définition formelle de la logique floue et la logique classique sont données, ainsi que la différence entre eux, nous présentons aussi l'ensemble flou, les fonctions d'appartenance, les règles floues, et enfin, les systèmes de fuzzification et défuzzification.

– Chapitre 3.

Le chapitre 3 décrit les étapes de notre approche proposée qui se focalise autour de la logique floue. Nous allons décrire comment exploiter la puissance de la logique floue pour créer un

ystème capable de détecter l'orientation globale d'une source subjective. Les variables linguistiques, les fonctions d'appartenance et les ensembles flous de notre système sont détaillés. La conception des règles floues ainsi que du choix système d'inférence et la technique de défuzzification sont motivés.

– **Chapitre 4.**

Ce dernier chapitre illustre étape par étape, l'implémentation de notre système. Dans la première section, nous présentons l'environnement de développement : le langage de programmation choisi, les bibliothèques nécessaires, etc. Dans la deuxième section, nous illustrons et argumentons les résultats obtenus.

Chapitre 1. Fouille d'opinion : état de l'art

1.1 Introduction

Ce chapitre donne un aperçu général sur le problème de fouille d'opinions, en présentant les différents concepts utilisés, les axes de recherche, et le problème de classification d'opinions. Des concepts assez compliqués sont exposés dans ce chapitre, comme les opinions implicites et explicites exprimées sur des caractéristiques (*ang. Features*) implicites et explicites, illustrés par des exemples.

1.2 Opinion, fait et sentiment:

Les messages que nous recevons, de proches, d'adultes, de la télévision, des livres, des journaux sont composés d'un ensemble de faits, d'opinions et de sentiments. Il n'est pas toujours facile de savoir quels sont les faits, les opinions et les sentiments.

Cette confusion entraîne bien souvent des disputes et peut conduire à des situations de blocage.

Toute discussion devrait avoir pour point de départ l'identification des faits, chacun pourrait ensuite exposer ses opinions ou ses sentiments. La confusion entre les 3 catégories est source de conflits et d'incompréhensions. Certains confondent et ne savent plus distinguer leurs opinions des faits.

1.2.1 Une opinion :

Est un jugement, un avis portant sur un sujet, sur des personnes, des objets, des situations. C'est aussi une croyance et une conviction. Il peut exister autant d'opinions différentes sur un même sujet qu'il y a de personnes présentes !

Bien souvent nos discussions reposent sur des opinions pour lesquelles nous sommes prêts nous " battre " pour les défendre et préserver ce en quoi nous croyions.

Nos opinions peuvent venir des autres : sans nous en rendre compte, nous reprenons des idées que nous avons entendues autour de nous (famille, amis, lectures)

Nos opinions peuvent aussi venir de nos propres observations et idées.

1.2.2 Les sentiments :

Le terme sentiment trouve sa racine dans le verbe " sentir ". C'est une connaissance plus ou moins claire, donnée d'une manière immédiate, c'est une sensation, une impression liée à ses émotions. Les sentiments sont l'expression de ce que l'on ressent au plus profond de soi. Les sentiments les plus connus sont l'amour, la haine, l'amitié.

1.2.3 Les faits :

C'est ce qui existe, c'est la réalité ou ce qui est démontré, c'est un acte, un évènement.

On va citer quelques phrases et pour chacune, on va écrire un F si elle est un fait, S pour un sentiment ou un O pour une opinion.

1. Je t'aime **S**
2. Il pleut **F**
3. Je suis heureuse **S**
4. Tu es énervant **O**
5. Tu m'énerves **S**
6. La soupe c'est bon **O**
7. LE corps a besoin de vitamines **F**

1.3 Sous-domaines de recherche :

Les informations textuelles sont soit des faits qui sont des expressions subjectives ou neutres qui n'expriment pas un sentiment, soit des opinions (sentiments), une opinion est subjective qui décrit des sentiments, des jugements ou bien l'évaluation d'une personne à propos d'un sujet, un objet d'intérêt ou de ses propriétés. Dans [17], l'auteur décrit la fouille d'opinion comme un sous-domaine qui analyse les textes dans le but d'extraire les informations pertinentes aux sentiments, il a une tâche qui est appelée classification d'opinions et qui classe les textes suivant les opinions qu'ils expriment, elle peut souvent se faire sur deux classes **positive** ou **négative**, ou trois classes **positive**, **négative** ou **neutre**.

Dans [7], et d'après les auteurs, la fouille d'opinion peut être divisée en trois sous-domaines:

- *L'identification des textes d'opinion :*

Il consiste à localiser et classer les textes ou des parties des textes selon ce qu'elles soient objectives ou subjectives porteuses des sentiments ou d'opinions.

- *Le résumé d'opinions :*

Ce résumé peut être textuel (contenant les parties du texte qui expriment des opinions), chiffré (pourcentage, notes), graphique (histogramme) ou encore image (thermomètre, étoiles, pouce levé ou baissé...), il consiste à mettre les opinions exprimées en avant et les cibles de ces opinions qui existent dans le texte.[2]

1.4 Approche de l'analyse des sentiments et la détection d'opinions :

Dans littératures, nous pouvons distinguer trois types d'approches pour la détection d'opinions et l'analyse des sentiments :

1.4.1 Approches d'apprentissage automatique :

Appelé aussi approche statistique, cette approche se basée sur l'apprentissage automatique, elle utilise la technique de classification pour classer le texte en des classes déférentes. Il existe principalement deux types de techniques d'apprentissage.[1] [3]

1.4.1.1 Apprentissage supervisé :

Il est basé sur les données libellées et par conséquent, les étiquettes sont fournies au modèle au cours du processus d'apprentissage. Ces donnés libellées sont utilisées par l'algorithme d'apprentissage pour donnée un modèle qui sera utilisée lors de la prise de décision.

Certains modèles d'apprentissage automatique ont été formulés pour classer les tweets en classes. Les techniques d'apprentissage automatique comme Naive Bayes (NB) , L'entropie maximale (ME). Et les machines à vecteurs de support (SVM) ont donné un grand succès dans l'analyse des sentiments.

L'apprentissage automatique commence par la collection des données d'apprentissage. Ensuite, on entraîne un classificateur sur ces données. Une fois une technique de classification supervisée est sélectionnée, une décision importante à faire est la sélection des descripteurs.

Ces derniers nous disent comment les documents sont caractérisés. Les descripteurs les plus couramment utilisés dans la classification de sentiment sont :

- Présence de termes et leur fréquence
- Information de la partie de discours (POS)
- Négations.
- Mots et des phrases d'opinion.

1.4.1.2 Apprentissage Non-supervisé :

Il ne consiste pas d'une classification précise, donc il se base sur le regroupement.

Le succès de ces deux méthodes d'apprentissage dépend principalement de la sélection et l'extraction de l'ensemble des descripteurs utilisés pour détecter le sentiment (la classe), Les

algorithmes d'apprentissage non supervisés classification hiérarchique ascendante, centres mobiles, règles d'association... etc.

1.4.2 Approche Lexique :

La principale tâche dans cette approche est la conception de lexiques (ou dictionnaires) d'opinions. L'objectif de ces lexiques est de répertorier le plus de mots possible qui sont porteurs d'opinions. Ces mots permettent ensuite de classer les textes en deux (positif et négatif) ou trois catégories (positif, négatif et neutre)

Les approches fondées sur le lexique reposent principalement sur un lexique de sentiments, à savoir, une collection de termes de sentiments connue et précompilée, des phrases et même des expressions idiomatiques, développés pour les genres traditionnels de communication. Tels que le lexique Opinion Finder.[1]

1.4.3 Approches hybride :

Cette approche est appelée aussi classification semi-supervisées. Elles combinent les points forts des deux approches précédentes. Elles prennent en compte tout le traitement linguistique des approches statistiques.[2]

1.5 Définition de fouille d'opinion :

Dans la littérature. sentiment analysis, opinion mining, opinion extraction, sentiment mining, subjectivity analysis, affect analysis, emotion analysis, review mining, appraisal extraction, sont des termes utilisés pour désigner des technologies d'analyse automatique des discours, écrits ou parlés, afin d'en extraire des informations subjectives comme des jugements, des évaluations ou des émotions.

L'origine de la discipline l'analyse des sentiments se réfère aux des sciences des psychologies, la sociologie et l'anthropologie [4]. Le terme Analyse sentimentale se réfère à l'extraction automatique de texte évaluative. Qui aide à produire des résultats prédictifs. Le terme d'analyse de sentiment est apparu en Nasukawa et Yi en 2003 [5]. Et le terme extraction de l'opinion terme est apparu dans Dave. Laurent et Pennock en 2003 [6]. Cependant la recherche sur des sentiments et des opinions est apparue plus tôt dans [7] [8] Bing Liu [9] a présenté une définition de l'analyse des sentiments comportant les domaines d'application ainsi que sa relation avec le TALN : l'analyse des sentiments est le domaine de l'étude qui analyse les opinions ; les sentiments, les évaluations, les attitudes et les émotions des gens vers des entités telles que des produits, des services, des organisations, des

particuliers, des problèmes. Des événements, des sujets et leurs attributs. Il représente un grand espace de recherche. L'analyse des sentiments est un domaine de recherche extrême active en traitement automatique des langues.

Pour mettre en valeur l'intérêt de l'échange d'opinions dans l'analyse des sentiments, pang et Lee en 2008 [10] considère que l'opinion des autres a toujours été une pièce d'information très précieuse au moments de se faire une opinion ou de prendre une décision. En effet avant l'apparition du Web et l'internet. Les gens avaient intérêt à connaître les opinions de leurs amis ou de leur famille. Il leur était demandé de faire savoir quel parti politique recevrait leur voix lors des prochaines élections. Grâce à l'essor considérable qu'ont connu le Web et l'internet à partir des années quatre-vingt-dix, il est devenu possible pour tous de consulter l'opinion d'un groupe de personne à travers le Web. Donc l'échange d'opinion est la phase principale qui permet d'effectuer une analyse de sentiment sur un sujet donné.

Selon H tang et S tan X la plupart des recherches existantes se sont portées sur la fouille et l'extraction de faits. Par exemple, la recherche d'information. La recherche sur le Web et beaucoup d'autres. On assiste, ces dernières années, à une prise de conscience de l'importance de l'opinion sur le web, ce qui explique les nombreux et récents travaux dans ce domaine [9]. Ils montrent l'importance de l'analyse de sentiment dans le temps actuel.

1.6 Traitement automatique des langages naturels (TALN)

Le traitement automatique du langage naturel (TALN), est une discipline s'appliquant au domaine de l'informatique et du langage. Il est utilisé par exemple pour la traduction automatique, la reconnaissance vocale ou encore les réponses automatiques aux questions.[11]

Ces domaines représentent des défis majeurs, car les mots du langage sont souvent traités un à un par l'ordinateur. Or, de nombreux mots sont polysémiques et recouvrent différentes réalités, en s'inscrivant notamment dans des contextes ou expressions qui peuvent changer complètement leur sens d'origine. A titre d'exemple, le verbe « rimer » ne va pas avoir la même signification entre « ne pas rimer » et « ne rimer à rien », tout comme le mot « film » ne va pas se référer à la pellicule photographique dans l'expression « ne rien comprendre au film ». Les différents domaines d'application de la linguistique sont pris en compte par le TALN[12] :

- La **morphologie**, qui s'intéresse à la composition des mots et leur corrélation avec d'autres mots.
- La **syntaxe**, qui définit comment les mots sont agencés dans une phrase.
- La **sémantique**, qui correspond à la signification des mots et des groupes de mots
- La **pragmatique**, grâce à laquelle le contexte est pris en compte.
- Enfin la **phonologie**, qui s'occupe des sonorités de la langue orale, ce qui est important pour la reconnaissance vocale.

1.7 Recherche d'information (RI) :

Un système de recherche d'information (RI) est un système qui permet de retrouver les documents liés à une requête d'utilisateur, à partir d'une base de documents volumineuse.

Il y a trois notions clés: documents, requête, pertinence [13]:

Document: Un document peut être un texte, un morceau de texte, une page Web, une image, une bande vidéo, etc. On appelle document toute unité qui peut constituer une réponse à une requête.

Requête: Une requête exprime le besoin d'information d'un utilisateur. Elle est en général de la forme suivante: « Trouvez les documents qui ... ».

Pertinence : Le but de la recherche informatique est de trouver seulement les documents pertinents. La notion de pertinence est très complexe. Généralement, dans un document pertinent, l'utilisateur doit pouvoir trouver les informations dont il a besoin. C'est sur cette notion de pertinence que le système doit juger si un document doit être donné à l'utilisateur comme réponse.

1.8 Domaine d'application de l'analyse des sentiments :

La détection des opinions joue un rôle très important dans plusieurs domaines ainsi plusieurs applications ont vu le jour dans ce contexte on cite brièvement quelques applications ci-dessous[14] :

1.8.1 Les entreprises :

Grace a la détection des sentiments, les entreprises peuvent connaitre les opinions des clients sur leur produit ou leur service, dans le but d'améliorer leur produit a travers ces opinions pour augmenter leur chiffre d'affaire

1.8.2 Les clients :

De nos jours, avant que le client prenne une décision d'achat tel qu'un produit technologique, Hôtel, etc. il cherche des conseils à travers les avis des autres acheteurs qui ont une expérience avec le produit souhaitable

1.8.3 Politique :

Avant de diffuser une nouvelle loi, les politiciens essayent de récolter l'avis des internautes sur cette loi. Il est intéressant de connaître aussi l'avis des internautes sur tel homme politique pour une élection présidentielle. [20]

1.9 complexité de notion d'opinion

Selon la définition citée dans [15], les opinions sont généralement des expressions de jugement que l'on porte sur un individu, un être vivant, un phénomène ou bien un objet, et elle peut aussi être des expressions décrivant des sentiments.

On doit connaitre la différence entre un « fait » et une « opinion ». Un *fait* est quelque chose qui est empiriquement vrai et peut-être étayé par des preuves, par contre une opinion est une croyance ou un jugement personnel que l'on porte sur un sujet ou un ensemble de sujet qui n'implique pas qu'elle soit obligatoirement vraie.

Pour démontrer la complexité de la notion d'opinion, nous allons nous baser sur un ensemble de critiques¹ exprimées par des clients ont déjà visité un restaurant à la ville de Montréal(Canada). Les exemples sont les suivants, pris sous forme de captures écran :

¹<http://www.restomontreal.ca/restaurant-reviews/>

Nombre de convives : 2 | Date : 2011-12-31

J'y vais avec ma famille le 31 décembre chaque année. La fondue est excellente. Le prix est raisonnable mais le service est un peu lent

Nombre de convives : 2 | Date : 2011-10-31

Bon resto pas cher. à éviter les dimanches(soirée des enfants) si vous voulez être tranquille.

Nombre de convives : 9 | Date : 2011-11-24

Moi et ma famille adorons ce restaurant italien! Tous les plats sont bons, et fait avec passion. Le service est bien, pas trop d'attente. Il y a une belle ambiance, et une belle déco. C'est un petit restaurant, donc ça vaut la peine de réserver si vous y aller en groupe.

Figure 1.1 : Ensemble de commentaires collectés du site web *restomontreal.ca*

Dans la première critique, il y a trois phrases, la première phrase « *J'y vais avec ma famille chaque année* » est objective, car elle ne porte aucun mot porteur d'opinion, tandis que les deux dernières phrases « *La fondue est excellente* » et « *Le prix est raisonnable mais le service est un peu lent* » sont subjectives à cause de la présence des mots et d'expressions sorteurs d'opinions : *excellente*, *raisonnable* et *un peu lent*. La présence de deux polarités opposées dans la dernière phrase « *le prix est raisonnable mais le service est un peu lent* » qui rend difficile à extraire l'orientation globale dans cette phrase.

Dans la deuxième critique, dans la première phrase « *Bon resto pas cher* », il est facile de déduire que l'orientation sémantique globale est positive parce que l'auteur a utilisé que des mots et des expressions positives « *bon* » et « *pas cher* ». Dans la deuxième phrase « *à éviter les dimanches (soirée des enfants) si vous voulez être tranquille* » : le client donne un conseil aux autres clients d'éviter ce restaurant les dimanches pour qu'ils soient tranquilles. Bien que le verbe « *éviter* » ait une polarité négative, la phrase porte une opinion positive, parce que l'auteur a spécifié que le bruit causé par la soirée des enfants est seulement les dimanches. Ce genre de phrases nécessite une analyse profonde pour extraire correctement leurs orientations globales.

Dans la troisième critique, « *Moi et ma famille adorons ce restaurant italien ! Tous les plats sont bons, et fait avec passion. Le service n'est pas bien, pas trop d'attente. Il y a une belle ambiance, et une belle déco. C'est un petit restaurant, donc ça vaut la peine de réserver si vous y aller en groupe.* ». L'orientation sémantique globale est claire, toutes les expressions et les mots porteurs d'opinions sont positifs : *adorer*, *bon*, *pas trop d'attente*, *belle ambiance*,

belle déco, sont des porteurs d'opinions positives, mais, comment nous pouvons savoir que la phrase « *C'est un petit restaurant* » porte une opinion positive, sachant que seule porteur d'opinion dans cette phrase est l'adjectif « *petit* » ? Comme l'auteur du commentaire n'a pas utilisé des modificateurs qui changent la direction de l'opinion comme *mais, pourtant, tandis que*, etc., nous pouvons facilement déduire que « *petit* » porte une polarité positive en appliquant la pseudo-règle linguistique des conjonctions intra-phrase (nous détaillons les règles de conjonction dans le chapitre 5 du présent manuscrit).

Généralement, les sentiments exprimés dépendent fortement du domaine, il se peut que la même expression indique deux polarité négatif et positif dans deux domaines différents. L'exemple suivant, cité dans [16], illustre la Dépendance de domaine : Dans le domaine électronique, « long » peut avoir une polarité positive, comme dans le commentaire « *the battery life of Camera X is long* ». Tandis que dans le domaine informatique, il peut avoir une polarité négative, comme dans « *Program X takes a long time to complete* ». Dans [16], les auteurs ont mentionné aussi qu'il n'est pas facile de construire des lexiques de sentiments pour tous les domaines d'intérêt.

De plus sur Internet, chacun utilise son propre vocabulaire, ce qui rend la tâche plus difficile même s'il s'agit du même domaine. En plus, il est très difficile d'affecter correctement le poids pour des phrases de la critique. Très souvent nous avons une description très positive d'un restaurant, avec la meilleure nourriture, le bon accueil, la propreté, etc., mais une phrase comme « *Malgré tout ça j'ai quitté le restaurant avant de terminer mon repas* » peut changer toute l'opinion en négative bien que toutes les autres phrases soient positives.

1.10 Facteurs de difficulté de la fouille d'opinion :

L'opinion-mining est un domaine jeune il a toujours des nombreux problèmes, malgré l'amélioration depuis les débuts de la recherche, ils n'ont pas été résolu de manière convaincante.

Malgré l'amélioration de ce jeune domaine <<Opinion-mining>>, on trouve toujours qu'il a des nombreux problèmes vue qu'il n'ont pas été résolu de manière convaincante Plusieurs facteurs qui sont cités ci-dessous ont été trouvés qui rendent la fouille d'opinion difficile ;

L'opinion et le sentiment sont le plus souvent décrits par la polarité. Cette dernière est en générale, soit positive (opinion favorable), soit négative (opinion défavorable), soit neutre. Nous montrons ci-dessous quelques difficultés de cette fouille d'opinions [14][2] :

–**l'ambiguïté des mots** : Par exemple le mot «grand» est un fait dans la phrase suivante «il est grand». Par contre il exprime une opinion (par exemple, pour designer la taille d'un mobile).

–**la structuration de la phrase** : Par exemple on oppose deux parties d'une phrase avec la conjonction «mais», par exemple «le restaurant est agréable mais les serveurs sont lent. Dans ce cas la polarité de la deuxième partie est opposée à la première.

–**Le contexte** : par exemple dans la phrase «je trouve que le restaurant est magnifique mais ma sœur le trouve mauvais», les deux opinions sont données par des personnes différentes.

–**le vocabulaire** : qu'on utilise pour exprimer une opinion. Il diffère d'une personne à une autre, comme par exemple un anglo-saxon lorsqu'il exprime ses sentiments utilise des mots bien représentatifs de ce qu'il ressent contrairement aux personnes qui ne connaissent pas ou peu sa langue.

–**Dépendance du contexte** : Une même thématique peut être utilisée dans différentes classes et peut exprimer une toute autre signification, comme Par exemple, le mot « longue » peut indiquer une opinion positive ou une opinion négative sur une caractéristique d'un produit dépendant de la caractéristique elle-même, comme indiqué dans l'exemple suivant :

« La durée de vie de la batterie de mon ordinateur portable est longue, c'est magnifique ».

« Mon ordinateur portable prend une longue durée pour démarrer, c'est un problème vraiment gênant ».

– **Hétérogénéité des données** : la difficulté due au langage libre qui est utilisé par les internautes pour exprimer leur avis. Les ponctuations ne sont pas forcément utilisées pour marquer les fins de phrases, des mots spécifiques sont utilisés tel que : «ha ha ha», «Good», «pffffff».

– difficulté de déterminer un lexique adapté à l'analyse de l'ensemble des textes d'opinion.

– difficulté à trouver une association entre l'opinion et la requête. En effet une opinion dans un document ne porte pas forcément sur la requête considérée.

- ***La classification d'opinions*** :

A pour but de classer les textes qui sont identifiés sur des classes selon l'opinion qu'ils expriment, généralement sur deux classes positives et négatives ou bien sur trois classes **positives négatives ou neutres**.

1.11 Conclusion :

On a présenté un état de l'art de la classification des opinions, on a commencé par un brève historique sur la l'analyse des sentiment ainsi qu' une définition des concept opinion, sentiment ou un fait, et les déférentes approche utilisés , on a aussi cité quelque facteurs de difficulté qui ont un mauvais impact sur la perfection de cette classification.

Dans le chapitre suivant, nous allons présenter les notions de base de la logique floue

Chapitre 2. Concepts Fondamentaux de la Théorie de la Logique Floue

2.1. Introduction :

La logique floue est une extension de la logique booléenne créée par Lotfi Zadeh en 1965 en se basant sur sa théorie mathématique des ensembles flous [18], qui est une généralisation de la théorie des ensembles classiques. En introduisant la notion de degré dans la vérification d'une condition, permettant ainsi à une condition d'être dans un autre état que vrai ou faux, la logique floue confère une flexibilité très appréciable aux raisonnements qui l'utilisent, ce qui rend possible la prise en compte des imprécisions et des incertitudes.

Le chapitre commence par une brève comparaison de la logique floue et de la logique classique. Une définition plus formelle de la logique floue s'en suit. Des modificateurs linguistiques aux règles de contrôle, des "haies", sont ensuite présentées. Pour appliquer les règles de contrôle floues et développer un système de contrôle réel, deux méthodes sont présentées : la "Défuzzification" par les méthodes de Mamdani et de Sugeno. On terminera ensuite par une étude de cas.

2.2. La déférence entre la logique floue et logique classique :

La logique booléenne classique ne permet que deux états : VRAI ou FAUX. La logique floue fut proposée par Zadeh en 1965[18] ; elle permet d'exprimer différents niveaux, plutôt que seulement 1 ou 0. Par exemple : le moteur est chaud, le moteur est très chaud. Quelle est la différence entre « chaud » et « très chaud » ? Ou encore, un homme est haut s'il mesure 170cm. Un homme est très haut s'il mesure 190cm. Où est la ligne de démarcation ? Un homme de 180cm est-il haut ou très haut ? 180.5cm ? 179.5cm ?

La logique floue est une branche des mathématiques qui permet à un ordinateur de modéliser le monde réel de la même façon que les personnes. Elle est préoccupée par la quantification et le raisonnement en utilisant un langage qui permet des définitions ambiguës, comme beaucoup, peu, petit, haut, dangereux. Elle s'occupe de situations où la question qui est posée et la réponse obtenue contiennent des concepts vagues.

Selon la logique floue, le raisonnement exact est un cas limité du raisonnement approximatif tout n'est qu'un degré. Tout système logique peut être rendu flou. Les connaissances sont interprétées comme une collection de contraintes élastiques ou floues d'un ensemble de variables. L'inférence est un processus de propagation de contraintes élastiques. La logique booléenne est un sous-ensemble de la logique floue.

La logique floue permet d'accommoder le concept de vérité partielle : des valeurs entre complètement vrai et complètement faux sont admises. On supporte des modes de raisonnement approximatifs plutôt qu'exacts. Son importance provient du fait que le raisonnement humain est approximatif. [18][20]

La définition de Zadeh [18] :

Fuzzy Logic is determined as a set of mathematical principles for knowledge Representation based on degrees of membership rather than on crisp membership of

2.3. Ensembles Flous :

La logique floue repose sur la théorie des ensembles flous, qui est une généralisation de la théorie des ensembles classiques [Zadeh, 1965]. Par abus de langage, suivant les us de la littérature, nous utiliserons indifféremment les termes sous-ensembles flous et ensembles flous. Les ensembles classiques sont également appelés ensembles nets, par opposition à flou, et de même la logique classique est également appelée logique booléenne ou binaire.

En logique floue, un ensemble flou contient plusieurs valeurs. L'ensemble flou est concerné par un degré d'appartenance (ou degré de vérité). On utilise un continuum de valeurs logiques entre 0 (complètement faux) et 1 (complètement vrai). Une fonction d'appartenance est utilisée pour mapper un item X dans le domaine des nombres réels à un intervalle de 0 à 1, ce qui permet un degré de vérité. L'appartenance à un ensemble représente une valeur entre 0 et 1. Un ensemble flou peut être défini comme un ensemble ayant des frontières floues. Un ensemble flou est défini comme suit : soit S un ensemble et x un membre de cet ensemble. Un sous-ensemble flou F de S est défini par une fonction d'appartenance $\mu_F(x)$ qui mesure le degré auquel x appartient à F. [19]

Un exemple : Soit S un ensemble des entiers positifs et F un sous-ensemble flou de petits entiers. Des entiers peuvent avoir une distribution de probabilité qui indiquent leur appartenance au sous-ensemble flou F : $\mu_F(1) = 1.0$, $\mu_F(2) = 1.0$, $\mu_F(3) = 0.9$, ... $\mu_F(30) = 0.01$. La figure 2.1 montre cette fonction d'appartenance.

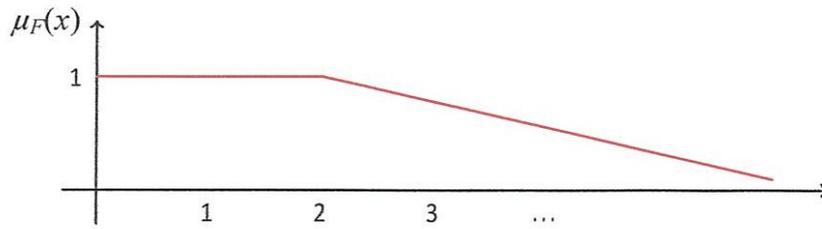


Figure 1.1 : Représentation du sous-ensemble flou F des petits entiers.

$\mu_F(x)$ est défini comme une fonction :

$$\mu_F(x) : X \rightarrow 0,1$$

Où $\mu_F(x) = 1$ si x est totalement dans A, $\mu_F(x) = 0$ si x n'est pas dans F et $0 < \mu_F(x) < 1$ si x est partiellement dans F. La fonction d'appartenance est une mesure :

- Du degré auquel un élément est membre d'un ensemble
- Du degré d'appartenance
- De la valeur de l'appartenance
- Du degré de confiance

La logique floue permet de transformer plusieurs valeurs réelles en quelques variables floues avec différentes appartenances, ce qui permet de réduire le nombre de règles. On utilise ces règles pour faire la commande d'un système.

Le tableau 2.1 compare des règles de commande classiques à la logique floue pour un système de chauffage, où v est la vitesse du ventilateur

Logique Classique	Logique Floue
Si temps = 21 ALORS v = 30%	Si temps = élevé ALORS v = lent
Si temps = 22 ALORS v = 40%	Si temps = assez élevé ALORS v = rapide
Si temps = 23 ALORS v = 50%	Si temps = très élevé ALORS v = très rapide
Si temps = 24 ALORS v = 60%	
Si temps = 25 ALORS v = 75%	
Si temps > 26 ALORS v = 100%	

Tableau 2.1 – Comparaison entre la logique classique et la logique floue pour un système de chauffage

Pour représenter la logique floue sur ordinateur, il faut déterminer la fonction d'appartenance. On utilise le plus souvent une représentation graphique. Les bornes de ces graphiques proviennent d'experts dans le domaine. La figure 2.2 montre deux exemples de représentation de la température, une en logique classique, et l'autre en logique floue. Selon la figure 2.2, en logique classique, une température de 22.5° est considérée comme élevée. En logique floue, une température de 22.5° appartient au groupe "moyenne" avec un degré d'appartenance de 0.167, et appartient au groupe "élevée" avec un degré d'appartenance de 0.75.

Voici une Comparaison entre l'appartenance de la température en logique classique vs la logique floue

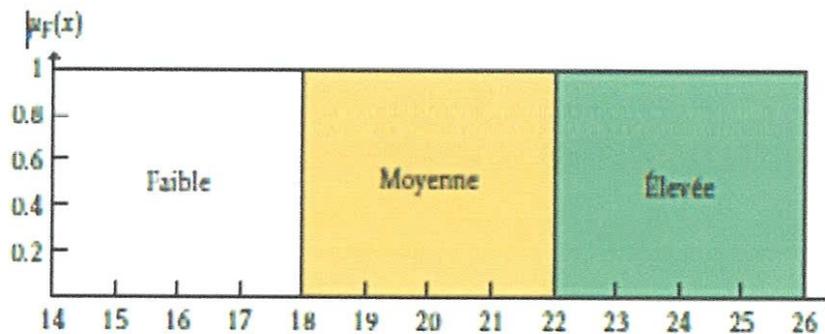


Figure 2.2 : (a) Représentation classique

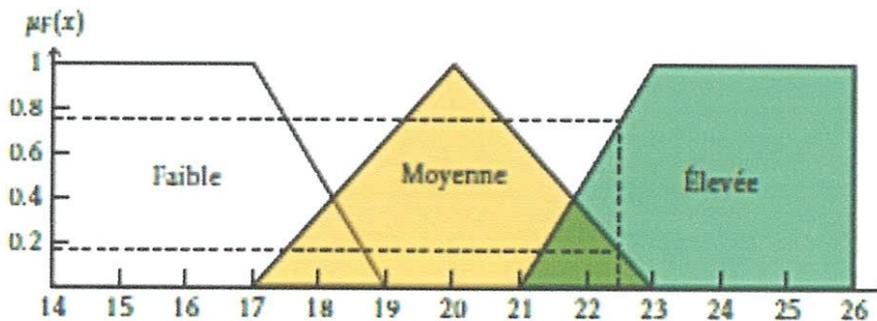


Figure 2.3 : (b) Représentation floue

Les variables floues faibles, moyennes et élevées sont représentées par des fonctions linéaires. D'autres fonctions auraient pu être utilisées, comme des trapézoïdes, des paraboles, etc.

Cependant, les fonctions linéaires sont beaucoup plus faciles à implémenter de façon pratique, et donnent de bons résultats. On utilise souvent une notation vectorielle pour représenter les fonctions. Pour les fonctions d'appartenance de la figure 2.2, on peut utiliser la notation suivante :

- Température faible : (1/17, 0/19)
- Température moyenne : (0/17, 1/20, 0/23)
- Température élevée : (0/21, 1/23)

2.4. Haies :

Les haies sont des modificateurs de valeurs floues et permettent la génération de déclaration floue à l'aide de calculs mathématiques. Elles modifient la forme des ensembles flous. Elles ont le même rôle que des adverbes et adjectifs. Selon leur impact sur la fonction d'appartenance, on les classifie selon leur effet [21]:

- **Concentration** : aide à intensifier un ensemble. Exemple : “très” crée une concentration et crée un nouveau sous ensemble.
- **Dilatation** : agrandi l'ensemble. Exemple : “plus” ou “moins” rend l'ensemble plus grand que l'original.
- **Contraste** : change la nature de l'ensemble en l'intensifiant ou en l'agrandissant .Exemple : “généralement”.

Les haies sont utilisées comme modificateurs tout-usage. Ils aident à refléter la pensée humaine. Un exemple de haies est montré à la figure 2.3, ou on montre les ensembles très faible et très élevée.

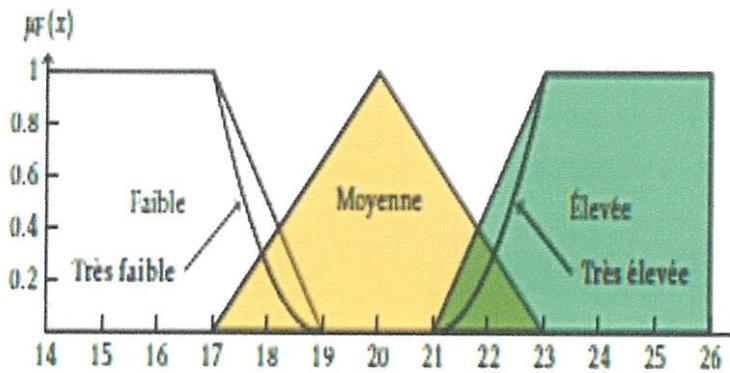


Figure 2.4 : Exemple de haies

Le tableau 2.2 montre quelques exemples de haies.

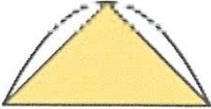
Haie	Représentation mathématique	Représentation graphique
Très	$[\mu_A(x)]^2$	
Plus ou moins	$\sqrt{\mu_A(x)}$	

Tableau 2.2 : Exemples de haies et représentation mathématique

2.5. Opérateurs Flous :

Les opérateurs flous décrivent comment des ensembles flous interagissent ensembles. On regardera certaines opérations communes, comme le complément, l'intersection et l'union. Le complément permet de vérifier de combien un élément n'appartient pas à un ensemble. Comme exemple, si on a l'ensemble des températures élevées, le complément est l'ensemble des températures qui ne sont pas élevées. Si A est l'ensemble flou, son complément $\neg A$ est :

$$\mu_{\neg A}(x) = 1 - \mu_A(x)$$

L'intersection de deux ensembles, en logique floue, est un peu différente des méthodes classiques. On cherche à savoir de combien un élément est-il dans les deux ensembles. On utilise alors la valeur minimale d'appartenance pour calculer l'intersection.

$$\mu_{A \cap B}(x) = \min [\mu_A(x), \mu_B(x)]$$

En logique floue, l'union est le contraire de l'intersection. On cherche à savoir de combien est-il dans l'un des deux ensembles. On utilise la valeur maximale d'appartenance.

$$\mu_{A \cup B}(x) = \max [\mu_A(x), \mu_B(x)]$$

La figure 2.4 résume ces opérations, de façon graphique.

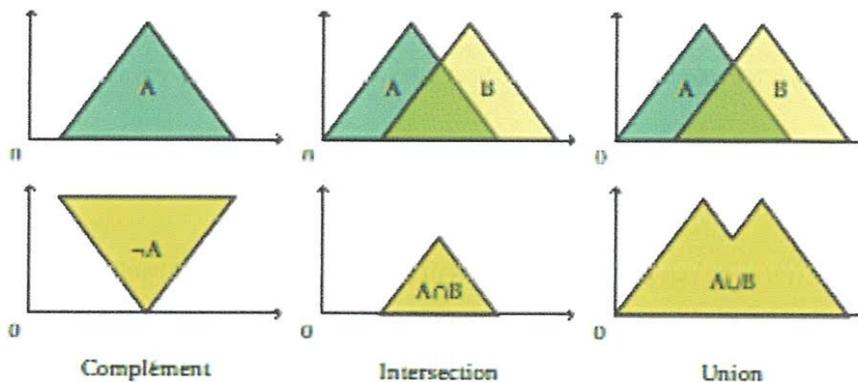


Figure 2.5 : Exemple d'opérations sur des ensembles flous

2.6. Règles floues :

Une règle floue est une déclaration de la forme suivante:

SI x est A ALORS y est B

Où x et y sont des variables linguistiques, et A et B sont des valeurs linguistiques, déterminées par les ensembles flous sur les ensembles X et Y. Une variable linguistique est une variable floue. Par exemple : La tension est haute. La variable linguistique tension prend

la valeur linguistique élevée. La plage de valeurs linguistiques possibles d'une règle représente l'univers de cette variable. Un exemple de règle floue est :

SI vitesse est lente ALORS arrêt est court

La variable vitesse peut avoir une plage de valeurs entre 0 et 220 km/h. On peut inclure des sous-ensembles flous (très lent, lent, moyenne, rapide, très rapide) pour modifier cette règle. Chaque sous-ensemble flou représente une valeur linguistique pour la variable.

La logique classique (SI – ALORS) utilise la logique binaire. La logique floue permet d'associer une plage de valeurs (un ensemble flou) à des variables linguistiques[18].

2.7. Fuzzification :

La Fuzzification est l'opération de rendre une entrée classique en valeur linguistique. Des valeurs d'entrée sont traduites en concepts linguistiques représentés comme des ensembles flous. Les fonctions d'appartenance sont appliquées aux mesures et des degrés de vérité sont établis pour chaque proposition. Les règles d'inférence permettent de calculer les valeurs d'appartenance de règles qui ont plusieurs antécédents. Si la conjonction qui unit deux antécédents est ET (AND) [20], on prend le minimum des deux. Exemple :

SI voiture a de l'essence ET voiture a un moteur ALORS voiture peut fonctionner

Supposons que la voiture a le plein d'essence (appartenance 1.0), mais qu'elle n'a pas de moteur (appartenance 0). La règle doit être déclenchée ? Bien sûr que non ; une voiture sans moteur ne fonctionne pas.

Si la conjonction qui unit deux antécédents est OU (OR), on prend le maximum des deux. Exemple :

Si le chien jappe OU il fait très froid dehors ALORS ouvrir la porte.

Supposons que le chien ne jappe pas (appartenance 0.0), mais qu'il fait très froid dehors (appartenance 1.0). Doit-on ouvrir la porte au chien ? Bien sûr que oui, on ne va pas laisser le chien dehors en temps très froid.

Les conclusions atteintes par les systèmes flous sont des faits flous ayant des degrés d'appartenance. Exemple : risque est faible avec une appartenance de 0.5. Cependant, le

déroulement final doit être une décision concrète ; exemple : louer de l'argent. Le processus de transformer un fait flou en un fait net est la Défuzzification.

2.8. Défuzzification :

La Défuzzification est le processus de convertir une valeur floue en valeur nette. Quelques méthodes existent, comme l'appartenance maximale, la méthode du centroïde, et la méthode des moyennes pondérées.[18]

2.9 Inférence de Mamdani

L'inférence de Mamdani procède selon quatre étapes [18]:

2.9.1. Fuzzification des variables d'entrée

On prend l'exemple d'un système de contrôle d'un ventilateur de maison, ayant 2entrées (température et humidité) et une sortie (vitesse du ventilatcur). Des règles possibles sont :

SI x est A1 OU y est B1 ALORS z est C1

SI x est A2 ET y est B2 ALORS z est C2

SI x est A3 ALORS z est C3

Où x est la température, y est l'humidité et z est la vitesse du ventilateur. Les ensembles flous sont :

- A1 = faible, A2 = moyenne, A3 = élevée
- B1 = sec, B2 = humide
- C1 = lent, C2 = moyenne, C3 = rapide

Les ensembles flous sont donnés dans les figures ci-dessous :

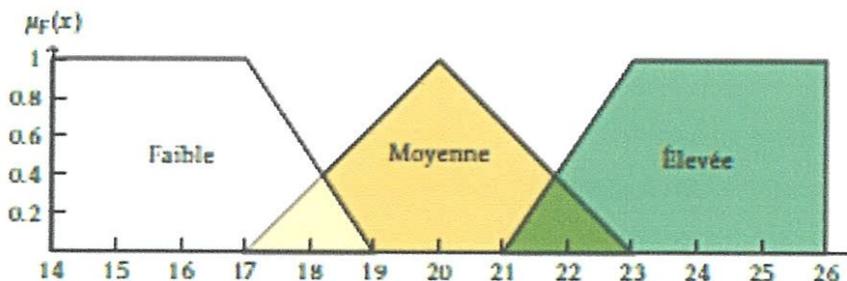


Figure 2.6 : Entrée x : température (°)

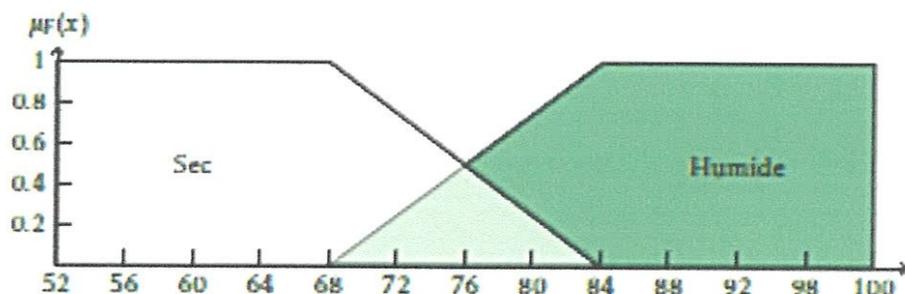


Figure 2.7 : Entrée y : humidité (%)

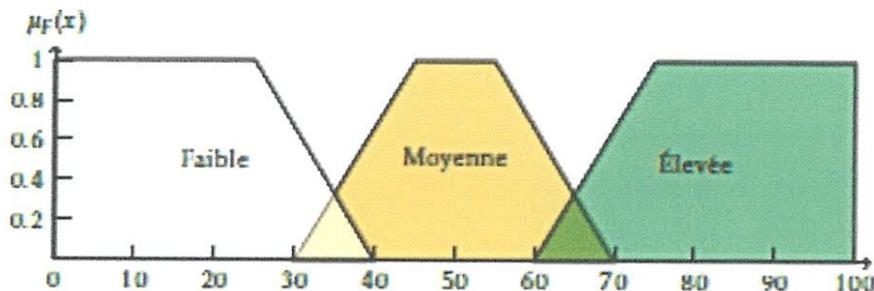


Figure 2.8 : Sortie z : vitesse du ventilateur (%)

Supposons qu'il fait actuellement 18°C, et que l'humidité est de 80%. On applique ces entrées sur les fonctions d'appartenance pour déterminer l'appartenance à chaque variable. Une température de 18°C correspond à une appartenance de 0.5 à l'ensemble faible ($\mu_{A1} = 0.5$) et une appartenance de 0.33 à l'ensemble moyenne ($\mu_{A2} = 0.33$). Une humidité de 80% correspond à une appartenance de 0.25 à l'ensemble sec ($\mu_{B1} = 0.25$) et une appartenance de 0.75 à l'ensemble humide ($\mu_{B2} = 0.75$).

2.9.2 Evaluation des règles

Il faut maintenant évaluer les règles en fonction des entrées obtenues. On applique les opérateurs flous correspondants pour combiner les règles. Si une règle a plusieurs antécédents, un opérateur flou est utilisé pour obtenir un seul chiffre qui représente le résultat. Ce résultat est ensuite appliqué à la fonction d'appartenance de la conséquence. Le résultat peut être produit par deux méthodes : coupure ou mise à l'échelle.

Si on reprend l'exemple précédent, on applique les valeurs d'appartenance à la règle 1 :

$$\text{SI } x \text{ est } A_1(0.5) \text{ OU } y \text{ est } B_1(0.25) \quad \text{ALORS } z \text{ est } C_1(?)$$

On peut utiliser deux méthodes pour combiner ces antécédents :

1. Méthode 1 : maximum

$$\mu_{C1} = \max [\mu_{A1}, \mu_{B1}] = 0.5$$

2. Méthode 2: probor

$$\begin{aligned}\mu_{C1} &= \text{probor} [\mu_{A1}, \mu_{B1}] = (\mu_{A1} + \mu_{B1}) - (\mu_{A1} \cdot \mu_{B1}) \\ &= (0.5 + 0.25) - (0.5 \cdot 0.25) = 0.625\end{aligned}$$

Si on utilise la méthode 1, l'appartenance est 0.5.

Pour la règle 2 :

SI x est $A_2(0.33)$ ET y est $B_2(0.75)$ ALORS z est C_2 (?)

On peut utiliser deux méthodes pour combiner ces antécédents :

1. Méthode 1 : minimum

$$\mu_{C1} = \min [\mu_{A2}, \mu_{B2}] = 0.33$$

2. Méthode 2 : prod

$$\mu_{C1} = \text{prod} [\mu_{A2}, \mu_{B2}] = \mu_{A1} \cdot \mu_{B1} = (0.33)(0.75) = 0.25$$

Si on utilise la méthode 1, l'appartenance est 0.33.

Au total, on obtient :

SI x est $A1$ (0.5) OU y est $B1$ (0.25) ALORS z est $C1$ (0.5)

SI x est $A2$ (0.33) ET y est $B2$ (0.75) ALORS z est $C2$ (0.33)

SI x est $A3$ (0.0) ALORS z est $C3$ (0.0)

2.9.3. Agrégation des sorties des règles

Si on utilise la méthode de coupure pour combiner les règles, il faut créer un nouveau

Polygone à partir des trois fonctions d'appartenance de la conséquence C. La hauteur du polygone est déterminée à partir de la valeur d'appartenance calculée plus haut.

L'agrégation pour cet exemple est montrée à la figure 2.9.

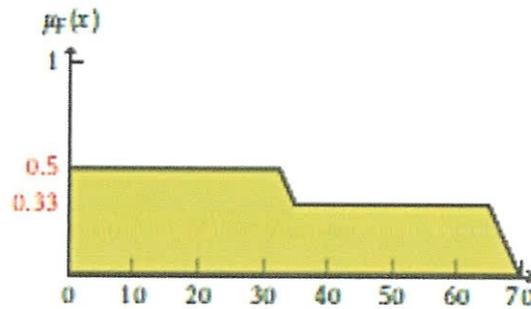
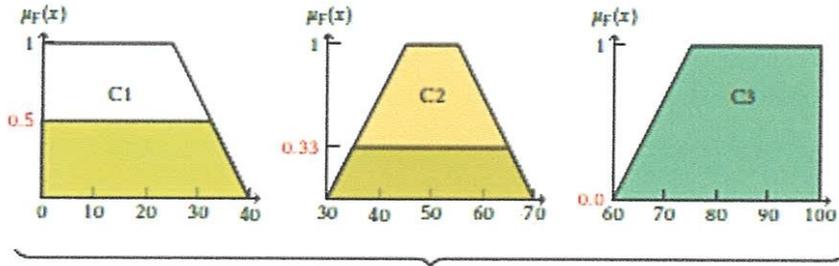


Figure 2.9 : Agrégation des règles de sortie par coupure

2.9.4 Défuzzification

Après avoir combiné les règles, il faut maintenant produire un chiffre net comme sortie. Dans ce cas-ci, la sortie doit être la vitesse du ventilateur. La technique la plus populaire est la méthode du centroïde : on cherche le centre de gravité du polygone obtenu :

$$CG = \frac{\sum_{x=a}^b \mu A(x) \cdot x}{\sum_{x=a}^b \mu A(x)}$$

Le centre de gravité n'a pas besoin d'être calculé des façons très précises. On peut approximer, en calculant à tous les 10, par ex. Pour l'exemple précédent :

$$CG = \frac{(0 + 10 + 20 + 30)(0.5) + (40 + 50 + 60)(0.33)}{0.5 + 0.5 + 0.5 + 0.5 + 0.33 + 0.33 + 0.33} = 26,67$$

Le ventilateur doit donc être à 26.67% de sa vitesse maximale.

2.9.4 Mise à l'échelle

La mise à l'échelle est une autre méthode pour générer la figure de sortie. Dans ce cas, au lieu de simplement couper la figure, on conserve la forme générale. Cette approche permet de mieux préserver l'intention de la règle, mais elle est un peu plus complexe à implanter mathématiquement. La figure 2.7 montre la différence entre la coupure et la mise à l'échelle pour la fonction d'appartenance C2 de l'exemple précédent.

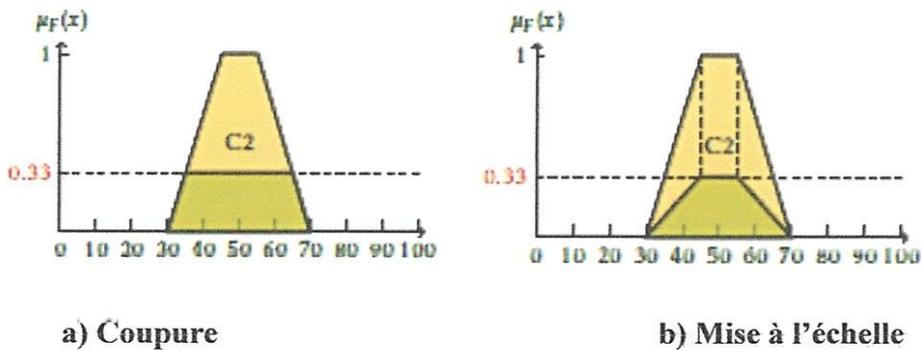


Figure 2.10 : Comparaison entre la coupure et la mise à l'échelle

2.10. Inférence de Sugeno

La méthode de Sugeno permet de simplifier le calcul de l'agrégation, afin d'obtenir plus rapidement une solution nette. Cette méthode est souvent utilisée pour des applications à temps réel, où le temps de calcul est important. La méthode de Sugeno utilise un seul pic comme fonction d'appartenance, plutôt qu'un polygone. Ce pic est un ensemble flou ayant une fonction d'appartenance 1 à un point particulier de l'espace, et 0 ailleurs. Le résultat de l'évaluation des règles devient l'amplitude du pic. La figure 2.8 montre un exemple de fonction de style Sugeno pour le même exemple.

Les conséquences sont évaluées selon le résultat de l'évaluation des règles. Dans l'exemple précédent, on a obtenu une appartenance de 0.5 pour la règle 1, de 0.33 pour la règle 2,

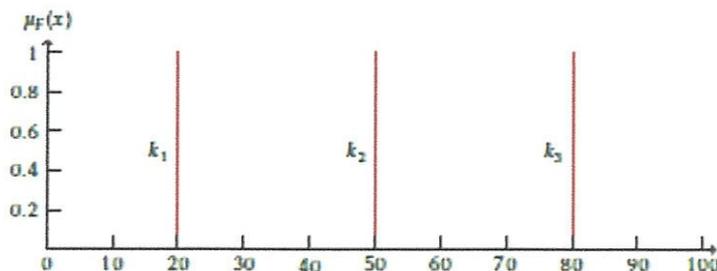


Figure 2.11 : Conséquences de style Sugeno

Et 0 pour la règle 3. Ces trois valeurs deviennent alors l'amplitude des pics, comme a la figure 2.9.

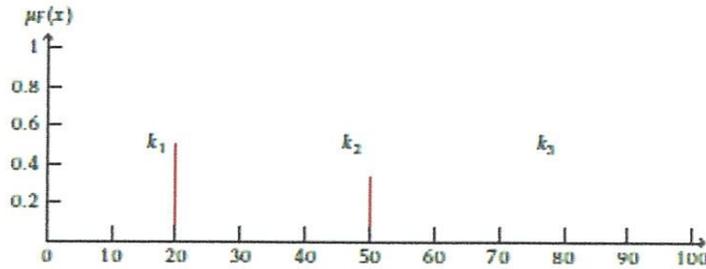


Figure 2.12 : Evaluation des conséquences de style Sugeno

Pour faire l'agrégation des règles, on utilise une moyenne pondérée :

$$CG_{\text{Sugeno}} = \frac{\sum_i \mu A(k_i) \cdot k_i}{\sum_i \mu A(K_i)}$$

Selon l'exemple précédent, on obtient :

$$CG_{\text{Sugeno}} = \frac{(0.5)(20) + (0.33)(50)}{0.5 + 0.33} = 32\%$$

On obtient une vitesse de ventilateur de 32%, ce qui est quand même près de la valeur obtenue par l'autre méthode.

2.11. Conclusion

Ce chapitre a pour objectif de présenter les notions fondamentales et les principes de base de la logique floue. Dans cette présentation, nous nous sommes limités au point indispensable à la compréhension du principe et étape de mise en œuvre d'un mécanisme d'inférence floue que nous utiliserons dans notre application pour raffiner la classification des sentiments.

On a commencé par une brève comparaison de la logique floue et de la logique classique, ainsi que les modificateurs Haies, à la fin deux méthodes on été présentées : la Défuzzification développé par Mamdani et de Sugeno, dans notre travaille on va utiliser la méthode de Mamdani.

Chapitre 3 : Conception

3.1 Introduction :

L'analyse du sentiment vise à identifier la polarité d'un document par le traitement du langage naturel, l'analyse du texte et linguistique computationnelle. Au cours de ces dernières années, il y a eu beaucoup d'attention sur l'analyse des sentiments parce que les contenus textuels n'ont pas cessé de progresser en termes de nombre et les données disponibles en ligne ont augmenté de façon exponentielle qui inclut de nombreux sentiments dans des documents (avis, commentaires, articles).

Nombreuses approches envisagent des techniques d'apprentissage automatique ou d'analyse statistique, mais il y a eu peu d'utilisation des classificateurs flous (le but de notre étude) dans ce domaine qui contient l'ambiguïté de la langue et de la pertinence des approches floues pour faire face à celle-ci (l'ambiguïté).

Notre travail propose un système basé sur des règles floues pour l'analyse des sentiments, qui peut offrir des résultats plus raffinés grâce à l'utilisation de la logique floue avec le degré d'appartenance. On compare ainsi les performances de notre approche proposée avec les classificateurs des sentiments couramment utilisés (par exemple les arbres de décision, Naïve Bayes) qui sont connus pour réaliser cet objectif.

Les résultats expérimentaux indiquent que notre approche basée sur le flou fonctionne légèrement mieux que l'autre algorithme. De plus, l'approche floue permet la définition de différents degrés de sentiment sans avoir besoin d'utiliser un grand nombre de classes.

3.2 Étude connexes :

Dans cette section, on donne un aperçu des travaux antérieurs liés à l'analyse des sentiments comprenant les types de classification des tâches, types de données, techniques de prétraitement et algorithmes d'apprentissage machine. Nous donnons également un aperçu complet des approches floues pour le traitement du texte. La plupart des approches dans le domaine de l'analyse des sentiments se concentrent sur la polarité ou la classification d'opinion en positif ou négatif [8], [9], [10]; certains chercheurs comprennent également une classe neutre [11], [12].

D'autres tâches de classification se focalisent sur la subjectivité contre l'objectivité [13], sur l'anticipation des catégories d'émotions (ex. la colère, la joie la peur la déception la tristesse) [14], [15] ou la force du sentiment [16]. La recherche dans ce domaine a été appliquée à une variété de sources de données, telles que les critiques de films [10], les critiques sur les films [17], [18], les données sur Facebook [11] et données de micro-blog [12], [19].

Pour préparer les données des tâches de classification, plusieurs techniques de prétraitements sont généralement utilisées, telles que la correction d'orthographe, tokenisation (diviser le texte en jetons tels que les mots), et la suppression des nombres, la ponctuation et la répétition des lettres. [6], [11], [12].

Les approches d'apprentissage automatique ont été utilisées avec succès pour l'analyse des sentiments et plusieurs algorithmes ont démontré leur efficacité pour les tâches d'analyse des sentiments: Bayes naïves, machines vectorielles de support, entropie maximale et arbres de décision [20], [21]. Au cours des dernières années, des approches floues ont commencé à émerger pour le traitement du texte. En 2012, un examen des approches floues pour le traitement du langage naturel [22] a été mis en évidence via la régression du pourcentage d'articles relatifs aux approches floues dans tous les articles de la littérature sur le traitement du langage naturel. . Depuis lors, un certain nombre d'approches floues ont été proposées pour une variété d'applications, comme indiqué ci-dessous.

3.3 Approche proposée :

La méthode proposée est basée sur les combinaisons de mots d'opinion autour de chaque caractéristique du produit dans une phrase. Cette méthodologie détermine la force de l'orientation de l'opinion (très faible, faible, modérée, très forte et forte) sur la caractéristique du produit en utilisant la technique de la logique floue. Par exemple, considérons ces phrases en anglais

- 1) *We extremely enjoy this camera.*
- 2) *We like this camera.*
- 3) *The picture quality is very good.*
- 4) *The picture quality is good*

Par interprétation humaine, il est évident que l'intensité des phrases 1 et 3 est plus que les phrases 2 et 4.

L'approche basée sur le lexique holistique classe ces phrases dans les classes positives, négatives et neutre tandis que la méthode proposée vise à classer ces phrases dans les niveaux de granularité (très faible, faible, modérée, forte et très forte en combinant les mots qui portent un opinion (Adverbe, adjectif et verbe on peut aussi considérer les noms comme des mots porteuse d'opinion).

Notre proposition est l'utilisation des systèmes à base des règles floues pour la classification des sentiments. Cette section présente les caractéristiques principales de l'approche floue et justifie son importance dans les applications pratiques.

3.4 Principales caractéristiques :

3.4.1 Phase de prétraitement :

Les textes en ligne générés par l'utilisateur nécessitent un prétraitement pour supprimer le bruit avant que le processus d'exploration d'opinion puisse être effectué. Ces revues contiennent généralement des fautes d'orthographe, de grammaire, l'utilisation de mots non-dictionnaire tels que des abréviations ou des acronymes de termes communs, des erreurs de ponctuation, des majuscules incorrectes etc. Ces erreurs dans l'ensemble de données rendent l'étape de prétraitement essentielle. Dans cette étape, des tâches telles que la correction des erreurs d'orthographe à l'aide d'un traitement de texte standard, la détection des limites de phrases et le regroupement répétitif de ponctuations sont effectuées. Les étapes de prétraitement génèrent des phrases qui peuvent être analysées automatiquement par l'analyseur linguistique. Si un mot non-dictionnaire se produit dans l'ensemble de données de révision, ses synonymes sont vérifiés dans le dictionnaire en utilisant nltk.Synset et les nouveaux mots sont ajoutés dans le dictionnaire.

3.4.2 Phase d'extraction de caractéristique :

Dans l'exemple passé les mots qui expriment l'orientation positive et négative sont généralement adjectifs, adverbe, verbe et nom peuvent être utilisés pour exprimer une opinion. Considérez cette phrase: "very" comme adverbe, "good" comme adjectif, "enjoy" et "like" comme verbe. Les chercheurs ont compilé un ensemble de mots et de phrases pour l'adjectif, l'adverbe et le verbe et le nom respectivement. Ces listes sont collectivement appelées le lexique d'opinion.

Afin d'utiliser différentes listes, nous devons effectuer un marquage sur les mots car de nombreux mots peuvent avoir plusieurs tags en fonction de leurs usages. La partie de parole d'un mot est une catégorie linguistique définie par son comportement syntaxique ou morphologique. Les catégories de marquage communes en anglais sont: nom, verbe, adjectif, adverbe, pronom, proposition, conjonction et interjection. Nous utilisons l'analyseur linguistique du processeur NLP [27] pour le marquage

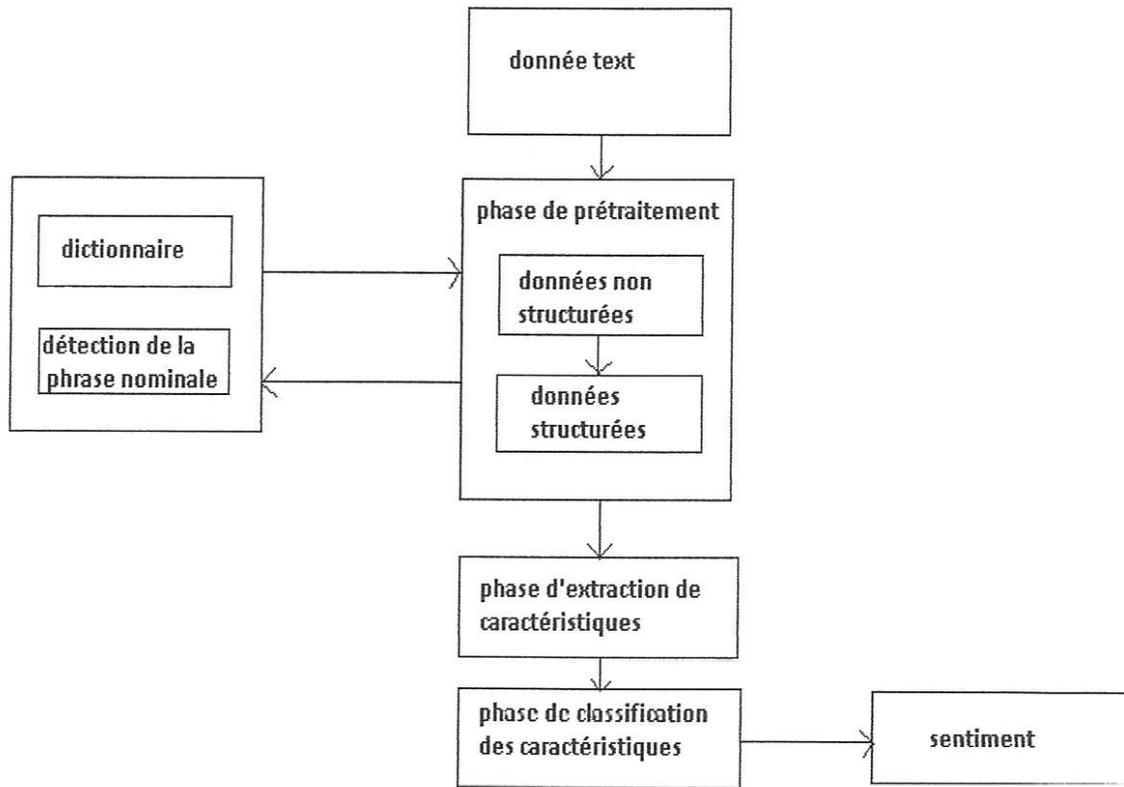


Figure 3.1 : Architecture d'approche proposé

3.5 Phase de classification de caractéristique :

3.5.1 Fuzzification des variables d'entrée :

la fuzzification consiste à définir les ensembles flous pour des variable d'entrée qui sent des mot dans notre message ou texte, dans notre approche on va faire des fonction d'appartenance floue triangulaire pour les caractéristique porteuse d'opinion de la phrase (verbe ,noms , adjectif) , Ces caractéristiques peut avoir des valeurs floues comme bonnes, excellentes ou satisfaisantes (low positif, medium positif, high positif) qui ont une polarité positive; ou il peut avoir des valeurs floues comme très mauvaises ,mauvaise ou bien moins mauvaise(low negatif, medium negatif, high negatif) qui ont une polarité négative

Un exemple est présenté par la figure 3.2 et 3.3

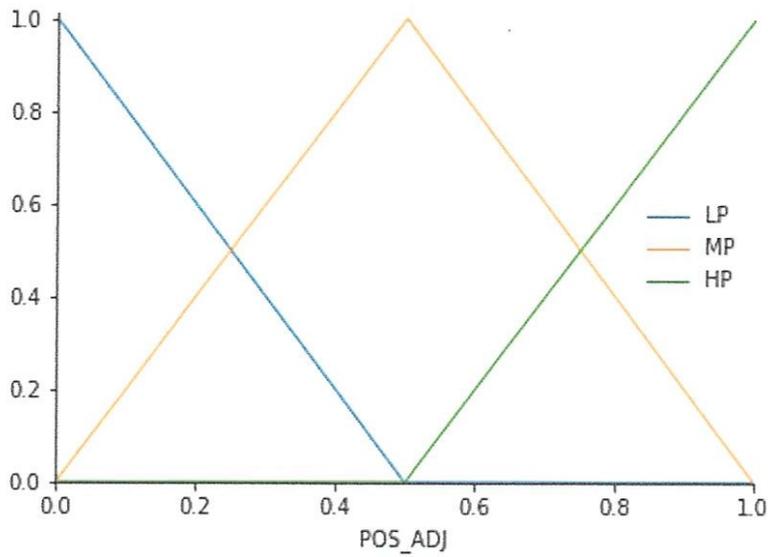


Figure 3.2 : représentation de la fonction d'appartenance triangulaire positive

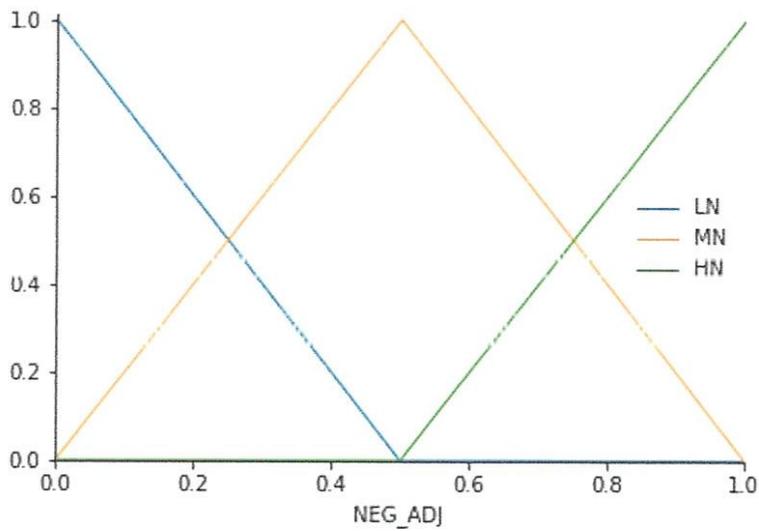


Figure 3.3 : représentation de la fonction d'appartenance triangulaire négative

Pour les adverbes on va utiliser une fonction floue trapézoïdale, ces adverbes prennent une valeur floue <<increaser>> et <<decreaser>> qui influence le degré de sentiment la figure 3.4 ci-dessous cette fonction

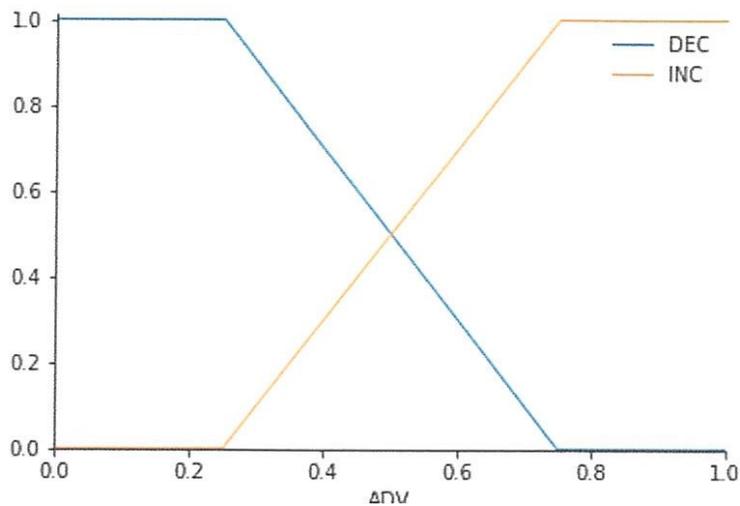


Figure 3.4 : représentation de la fonction d'appartenance trapézoïdale floue

3.5.2 Evaluation des règles:

Cette étape où l'on établit les règles floues qui permettent d'aboutir à la sortie en fonction des valeurs des variables d'entrée.

Ces règles floues lient les variables d'entrée afin de tirer des conclusions ou des déductions, après l'édition des règles, il reste à calculer les degrés d'appartenance de la variable de sortie à tous les ensembles flous qui lui sont associés.

Chaque règle est composée d'une prémisse liée par des opérateurs ET, OU et donne lieu à une implication. Par l'opérateur alors (prédicat).

Nous définissons quelques règles «IF - THEN» qui peuvent supporter toutes les possibilités. La forme générale de la règle floue peut être exprimée comme suit:

IF (X1 est A) AND (X2 est B or X3 est C) THEN Orientation est D.

A, B, C ∈ {faible, modérée, élevée}

D ∈ {très fort, fort, modéré, très faible et faible}

Voilà un exemple :

SI l'adverbe est «increaser» et l'adjectif ou verbe est «VIP» alors l'orientation est VIIP.

Le tableau 3.1 montre quelques règles IF-THEN. Pour calculer l'orientation des phrases 1 et 2, basé sur le tableau 1, nous définissons ces règles:

<p>Règle 1 : if adverb (dec) AND [pos_adj (LP) OR pos_verbe (LP) OR pos_noms(LP) then orientation est VLP]</p>
<p>Règle 2 : if adverbe (incresear) AND [pos_adj (LP) OR pos_verbe (LP) OR pos_noms(LP) then orientation est MP]</p>
<p>Règle 3 : if adverbe (decresear) AND [pos_adj (MP) OR pos_verbe (MP) OR pos_noms(MP) then orientation est (LP)]</p>
<p>Règle 4 : if adverbe (incresear) AND [pos_adjectif (MP) OR pos_verbe (MP) OR pos_noms(MP) then orientation est (H) positif]</p>
<p>Règle 5 : if adverbe (decresear) AND [pos_adjectif (HP) OR pos_verbe (HP) OR pos_noms(HP) then orientation est (M) positif]</p>
<p>Règle 6 . if adverbe (incresear) AND [neg_adjectif (HP) OR neg_verbe (HP) OR neg_noms(HP) then orientation est(VHP)]</p>
<p>Règle 7 : if adverbe (decresear) AND [neg_adjectif (LN) OR neg_verbe (LN) OR neg_noms(LN) then orientation est (VLN)]</p>
<p>Règle 8 : if adverbe (incresear) AND [neg_adjectif (LN) OR neg_verbe (LN) OR neg_noms(LN) then orientation est (MN)]</p>
<p>Règle 9 : if adverbe (decresear) AND [neg_adjectif (MN) OR neg_verbe (MN) OR neg_noms(MN) then orientation est (LP)]</p>
<p>Règle 10 : if adverbe (incresear) AND [neg_adjectif (MN) OR neg_verbe (MN) OR neg_noms(MN) then orientation est (MN)]</p>
<p>Règle 11 : if adverbe (decresear) AND [neg_adjectif (HN) OR neg_verbe (HN) OR neg_noms(HN) then orientation est (MN)]</p>
<p>Règle 12 : if adverbe (incresear) AND [neg_adjectif (HN) OR neg_verbe (HN) OR neg_noms(high HN) then orientation est (VHN)]</p>

Règle 13 : if [pos_adjectif (LP) OR pos_verbe (LP) OR pos_noms(LP) then orientation est (LP)]
Règle 14 : if [pos_adjectif (MP) OR pos_verbe (MP) OR pos_noms(MP) then orientation est (MP)]
Règle 15 : if [pos_adjectif (HP) OR pos_verbe (HP) OR pos_noms(HP) then orientation est (HP)]
Règle 16 : if [neg_adjectif (LN) OR neg_verbe (LN) OR neg_noms(LN) then orientation est (LN)]
Règle 17 : if [neg_adjectif (MN) OR neg_verbe (MN) OR neg_noms(MN) then orientation est (MN)]
Règle 18 : if [neg_adjectif (HN) OR neg_verbe (HN) OR neg_noms(HN) then orientation est (HN)]

3.1 Tableau : les règles << if >> et << then >>

3.5.3 Agrégation:

Il faut maintenant évaluer les règles en fonction des entrées obtenues. On applique les opérateurs flous correspondants pour combiner les règles. Si une règle a plusieurs antécédents, un opérateur flou est utilisé pour obtenir un seul chiffre qui représente le résultat. L'ensemble flou global de sortie est construit par agrégation des ensembles flous obtenus par chacune des règles concernant cette sortie.

Ce résultat est ensuite appliqué à la fonction d'appartenance de la conséquence. Le résultat peut être produit par deux méthodes : coupure ou mise à l'échelle.

Dans les règles proposées on trouve des intersections(ET) et des unions(OU), pour l'intersection nous devons sélectionner le degré minimum de la fonction d'appartenance des deux variables antécédentes. Cela donnera le degré d'appartenance aux nombres flous conséquents

$$\mu_{A \cap B}(x) = \min [\mu_A(x), \mu_B(x)]$$

Contrairement dans le cas de calculer l'union on utilise la valeur maximale de la fonction d'appartenance des variables antécédentes est ça donne les degrés d'appartenance flous conséquents est-il dans l'un des deux ensembles :

$$\mu_{A \cup B}(x) = \max [\mu_A(x), \mu_B(x)]$$

Pour la négation on utilise l'opérateur flou complément $\neg A$, donc le degré d'appartenance flou conséquent est donné par le complément des variables antécédentes :

$$\mu_{\neg A}(x) = 1 - \mu_A(x)$$

Si on utilise la méthode de coupure pour combiner les règles, il faut créer un nouveau Polygone à partir des fonctions d'appartenance de la conséquence. La hauteur du polygone est déterminée à partir de la valeur d'appartenance calculée plus haut.

3.6 Défuzzification :

Pour calculer la sortie finale, la fonction de défuzzification doit être utilisée pour les convertir en valeur net et trouver l'orientation finale du sentiment. Dans cette étude, on utilise le défuzzificateur de Mamdani l'opérateur de défuzzification le plus connu dans la méthode de défuzzification est le centre de gravité, qui calcule le centre de gravité de la zone sous la fonction d'appartenance:

$$\frac{\int x \mu(x) dx}{\int \mu(x) dx}$$

Où x est la valeur non-floue et $\mu(x)$ est la fonction d'appartenance, Nous définissons une valeur nette pour chacun des résultats, par exemple:

Neutre: 0, VL(positif\negatif): 0,2, L(positif\negatif): 0,4 M (positif\negatif): 0,6,

H (positif\negatif): 0,8, VH(positif\negatif): 1.

Sur la base de ces chiffres, nous pouvons trouver l'orientation de la phrase 1 et 2 appartiennent à tres Fort Positif et Modéré Positif.

3.7 Conclusion :

Dans ce chapitre on a proposé un modèle de logique floue pour identifier l'orientation sémantique des opinions exprimées par les critiques sur les caractéristiques du produit. La méthode est capable de classer les revues en cinq catégories: (VH) négatif (ou positif), (H) négatif (ou positif), (M) négatif (ou positif), (L) négatif (ou positif) et VL négatif (ou positif). Nous nous attendons à ce que l'exactitude de la classification soit augmentée en combinant des mots d'opinion (c'est-à-dire adverbe, adjectif, verbe, nom).

Chapitre 4 : Implémentation

4.1 Introduction :

Dans ce chapitre, nous présentons l'implémentation de notre approche présentée dans le chapitre 3, puis l'évaluation et la discussion des résultats obtenus.

Ce chapitre décrit notre application sur la Classification fine d'opinions sur le commentaire et les critique en utilisant la fusion de la logique floue avec la classification du sentiment pour caractériser une opinion positive ou négative. On va donner une brève explication sur les outils qu'on a utilisés dans notre application.

4.2 Présentation des outils :

En effet le choix des outils n'est pas un caractère de base pour notre projet essentiel pour nous et d'obtenir de bonne résultat .ces résultat ne dépend réellement que de la notre approche qu'on a expliqué dans le chapitre précédent, on a exploité les outilles ci-dessous pour atteindre notre application

4.2.1 Le langage de programmation python 3.6 :

Python est un langage de programmation, dont la première version est sortie en 1991. Créé par **Guido van Rossum** jusqu'à se voir associer une organisation à but non lucratif particulièrement dévouée, Lors de la création de la Python Software Fondation en 2001, et durant les années qui ont suivi, le langage Python est passé par une suite de versions que l'on a englobées dans l'appellation Python 2.x (2.3, 2.5, 2.6...). Depuis le 13 février 2009, la version 3.0.1 est disponible. Cette version casse la compatibilité ascendante qui prévalait lors des dernières versions dans notre application on a utilisé la version 3.6.3.

Python n'a pas besoin d'un compilateur, donc il n'a pas besoin d'attendre un temps avec des projets de 10'000 lignes que l'on doit tester souvent, et Python n'a pas besoin d'un IDE car il a déjà un interpréteur intégré.[pyton]



Figure 4.1 : langage de programmation python.

4.2.2 L'éditeur spyder :

Spyder est un environnement de développement intégré multiplateforme (IDE) open source pour la programmation scientifique en langage Python. Spyder intègre NumPy, SciPy, Matplotlib et IPython, ainsi que d'autres logiciels open source. IL est publié sous licence MIT. Spyder est un puissant environnement de développement interactif pour le langage Python, avec des fonctionnalités avancées d'édition, de test interactif, de débogage, d'introspection et d'informatique numérique. Spyder fournit également un inspecteur d'objets qui s'exécute dans le contexte de la console. Spyder nous permet de travailler facilement avec les meilleurs outils de la pile scientifique Python dans un environnement simple mais puissant. Ces fonctionnalités interagissent facilement avec notre code au fur et à mesure que nous progressons, parcourez et analysez tous les résultats générés par votre code, explorez de nouvelles bibliothèques, parcourez les commandes enregistrées sur toutes les consoles par ordre chronologique et gérez des tâches de programmation importantes.



Figure 4.2 :l'éditeur spyder.

4.2.3 La bibliothèque NLTK :

Le Natural Language Toolkit (NLTK) est une plate-forme utilisée pour construire des programmes Python qui fonctionnent avec des données de langage humain pour l'application dans le traitement statistique du langage naturel (NLP).

Il contient des bibliothèques de traitement de texte pour la tokenisation, l'analyse, la classification, le stemming, le marquage et le raisonnement sémantique. Il comprend un livre expliquant les principes sous-jacents aux tâches sous-jacentes de traitement de la langue prises en charge par NLTK.

The Natural Language Toolkit est une bibliothèque open source pour le langage de programmation Python écrit à l'origine par Steven Bird, Edward Loper et Ewan Klein pour le développement et l'éducation.

NLTK comprend plus de 50 corpus et sources lexicales tels que Penn Treebank Corpus Open Multilingual Wordnet, Problem Report Corpus, et Lin's Dependency Thesaurus.

4.2.4 La bibliothèque skfuzzy :

Scikit-Fuzzy est une collection d'algorithmes de logique floue destinés à être utilisés dans SciPy Stack, écrits en Python.

4.2.5 Bibliothèque textblob :

TextBlob est une bibliothèque Python pour le traitement des données textuelles. Il fournit une API simple permettant de plonger dans des tâches courantes de traitement du langage naturel (TAL), telles que l'étiquetage de parties du discours, l'extraction de syntagmes nominaux, l'analyse des sentiments, la classification, la traduction, etc.

4.2.6 La bibliothèque numpy :

Numpy est une bibliothèque de base pour l'informatique scientifique en Python. Il fournit un objet tableau multidimensionnel haute performance et des outils pour travailler avec ces tableaux.

Outre ses utilisations scientifiques évidentes, NumPy peut également être utilisé comme un conteneur multidimensionnel efficace de données génériques. Des types de données arbitraires peuvent être définis. Cela permet à NumPy de s'intégrer facilement et rapidement à une grande variété de bases de données.

4.2.7 La phase de prétraitement :

Dans notre travail nous allons nous focaliser sur la langue anglaise par ce que la bibliothèque NLTK que nous allons utiliser par la suite support bien cette langue. Avant d'entamer notre analyse nous devons nous pencher sur la préparation du message qu'on veut analyser afin de fragmenter celui-ci et éradiquer tous bruit, avant que le processus d'exploration d'opinion puisse être effectué convenablement, ce message pourrait contenir des fautes d'orthographe, de grammaire et l'utilisation des mots non-dictionnaire tels que des abréviations, des acronymes de termes communs, des erreurs de ponctuation, et les fautes d'orthographe.

Exemple dans la figure 4.1 :

```
entrez votre commentaire : tthis,,boook ?is veeery reccomended?!
```

```
this,,book ?is very recommended?!
```

```
this book is very recommended
```

Figure 4.1 : la suppression de ponctuations et la répétition des lettre

4.3 L'extraction des caractéristiques :

4.3.1 Tokenisation et catégorisations :

Dés qu'on termine la phase du prétraitement on s'empresse à une opération d'extraction des caractéristiques en élaborant une liste des différents jetons (Token) qui constituent le message, et nous allons affecter des étiquètes sur la liste des mots dégagés, la figure 4.2 montre cette étape :

```
['this', 'book', 'is', 'very', 'recommended']
```

```
[('this', 'DT'), ('book', 'NN'), ('is', 'VBZ'), ('very', 'RB'), ('recommended', 'JJ')]  
... ]
```

figure 4.2 : Toknisation de phrase et l'affectation des étiquètes

4.3.3 extraction score :

Dans cette étape on va extraire le score de chaque mot positif ou négatif vis-à-vis leur sentiment ce score pourrait varier de 0 à 1 ainsi que un jetons ayant 0 comme score est un mot neutre qui n'a pas de sentiment (un fait) par contre le score 1 c'est le parfait négatif ou positif cette exemple dans la figure 4.3 :

```
book -- POS et NEG scores  
0.0  
0.0  
be -- POS et NEG scores  
0.25  
0.125  
very -- POS et NEG scores  
0.25  
0.25  
recommended -- POS et NEG scores  
0.5  
0.0
```

Figure4.3 :Ectraction du score

4.3.4 Dépendance grammatical :

Cette étape sert à extraire la dépendance grammaticale dans la phrase pour faire sortir les clauses exprimées et la relation entre les adverbes (increase) et (decrease) que nous avons déjà expliqués dans le chapitre trois et leur modification qui influence le score de sentiment exprimé par un nom, un verbe ou un adjectif dans notre exemple 'this book is very recommended', on s'intéresse sur la clause 'very recommended', 'recommended' est un adjectif modifié par l'adverbe 'very'.

4.4 Phase de classification de caractéristique :

4.4.1 La Fuzzification des caractéristiques :

Cette étape sert à traduire les valeurs d'entrée en concepts linguistiques représentés comme des fonctions d'appartenance floue trapézoïdale pour les adverbes et triangulaire pour les verbes, les adjectifs et les noms porteurs d'opinion comme dans la figure ci-dessous:

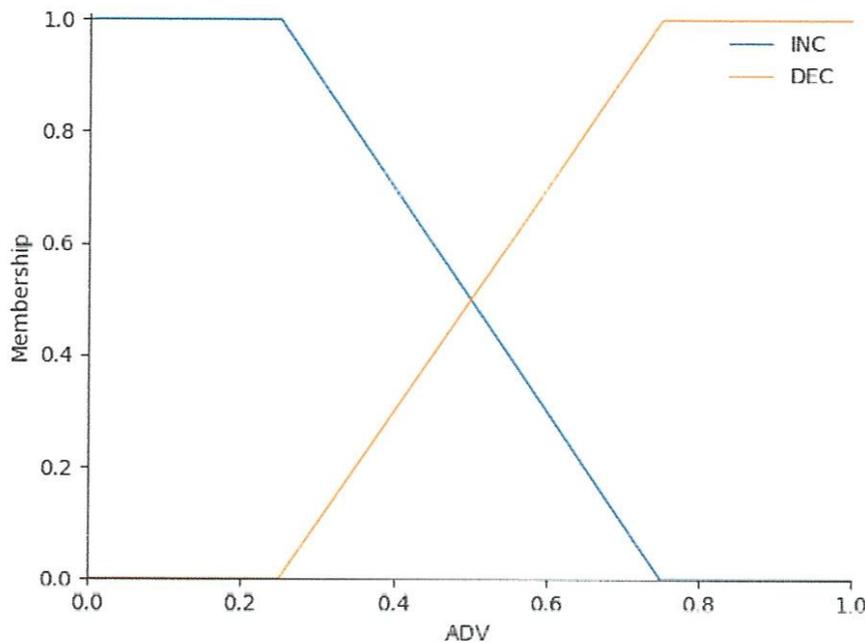


Figure 4.4 : la fonction trapézoïdale des adverbes

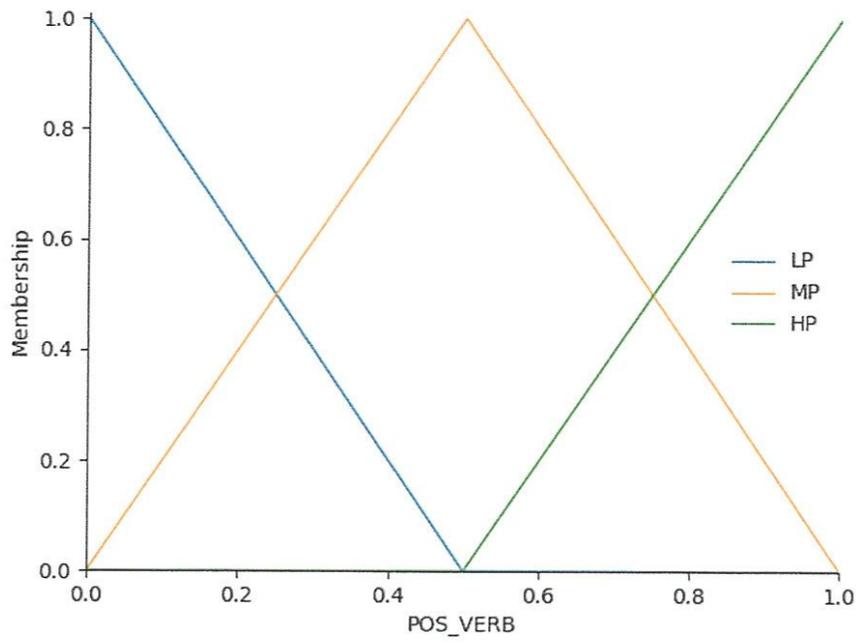


Figure 4.5 : la fonction d'appartenance triangulaire pour les verbe positif

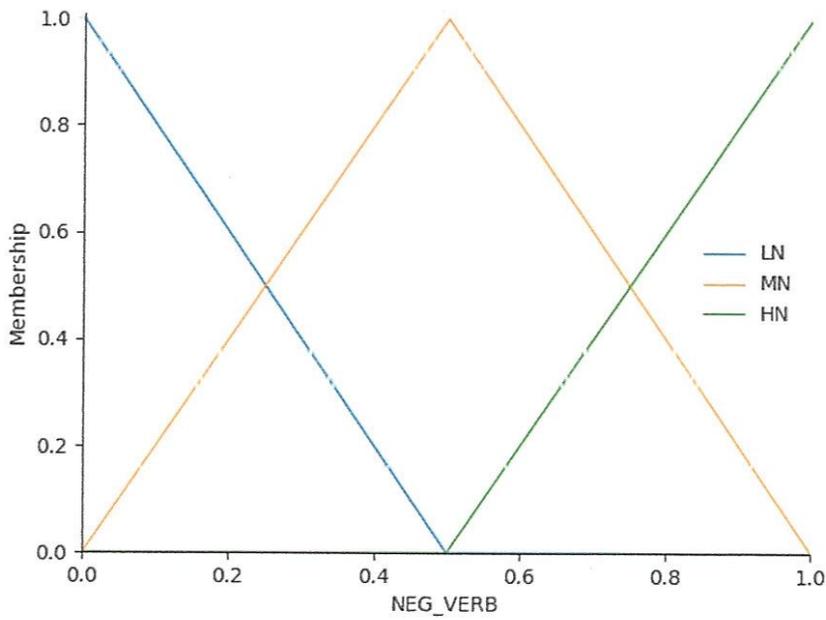


Figure 4.6 : la fonction d'appartenance triangulaire pour les verbe negatif

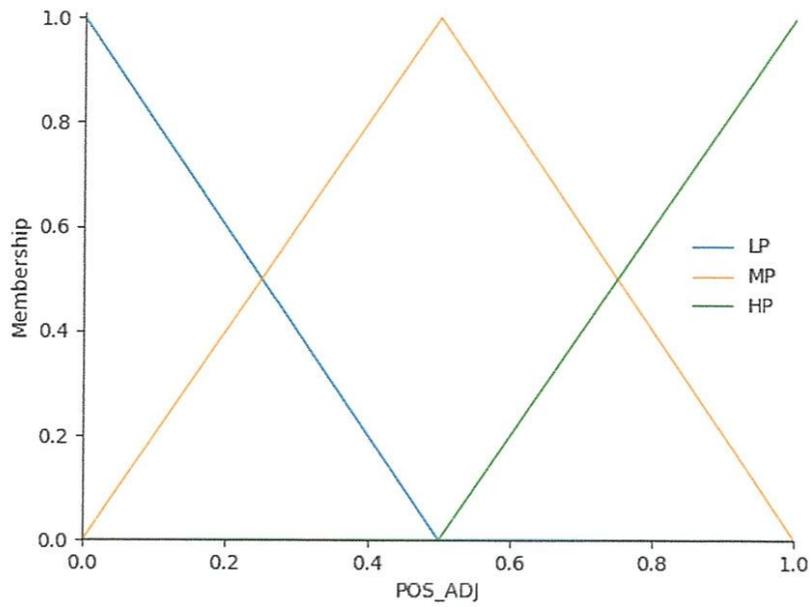


Figure 4.7 : la fonction d'appartenance triangulaire pour les adjectif positif

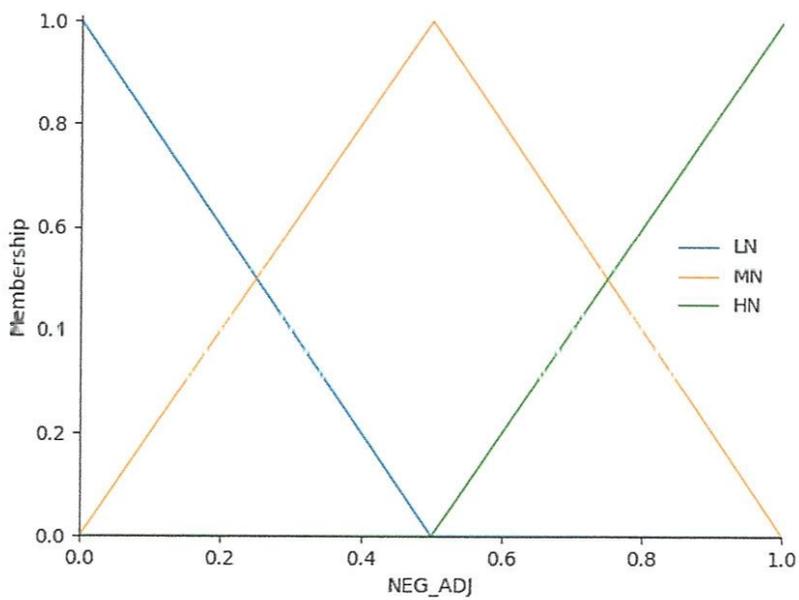


Figure 4.8 : la fonction d'appartenance triangulaire pour les adjectif positif

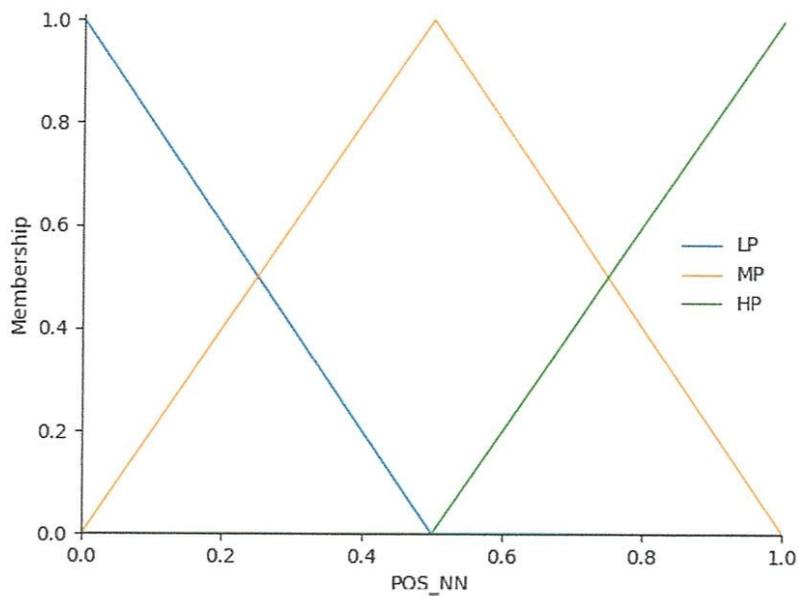


Figure 4.9 : la fonction d'appartenance triangulaire pour les noms positifs

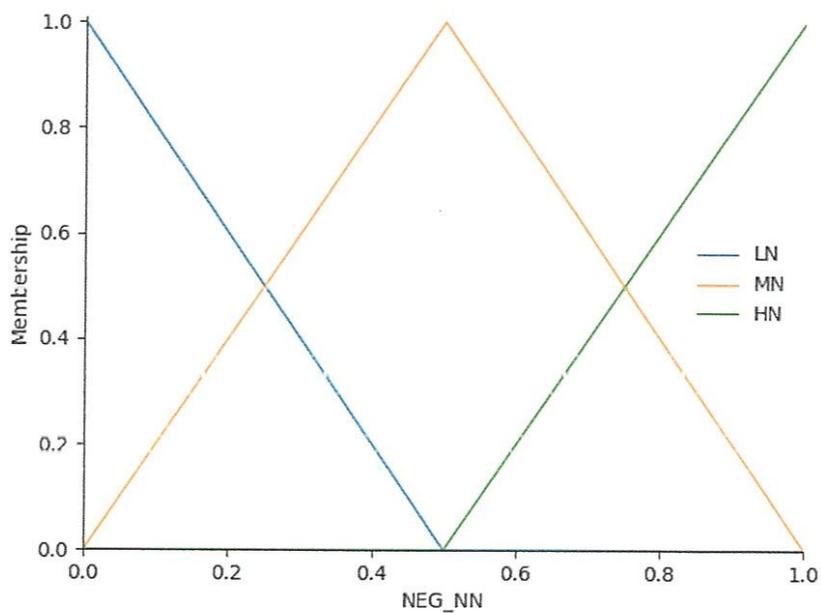


Figure 4.10 : la fonction d'appartenance triangulaire pour les noms négatifs

4.4.2L'agregation et la defuzification :

Après avoir évalué les règles qu'on a parlé dans le chapitre 3 L'ensemble flou global de sortie est construit par agrégation des ensembles flous obtenus par chacune des règles, et pour le résultat final, il est calculé avec la méthode de centre de gravité.

La figure 4.11 montre l'ensemble flou global avant l'agregation :

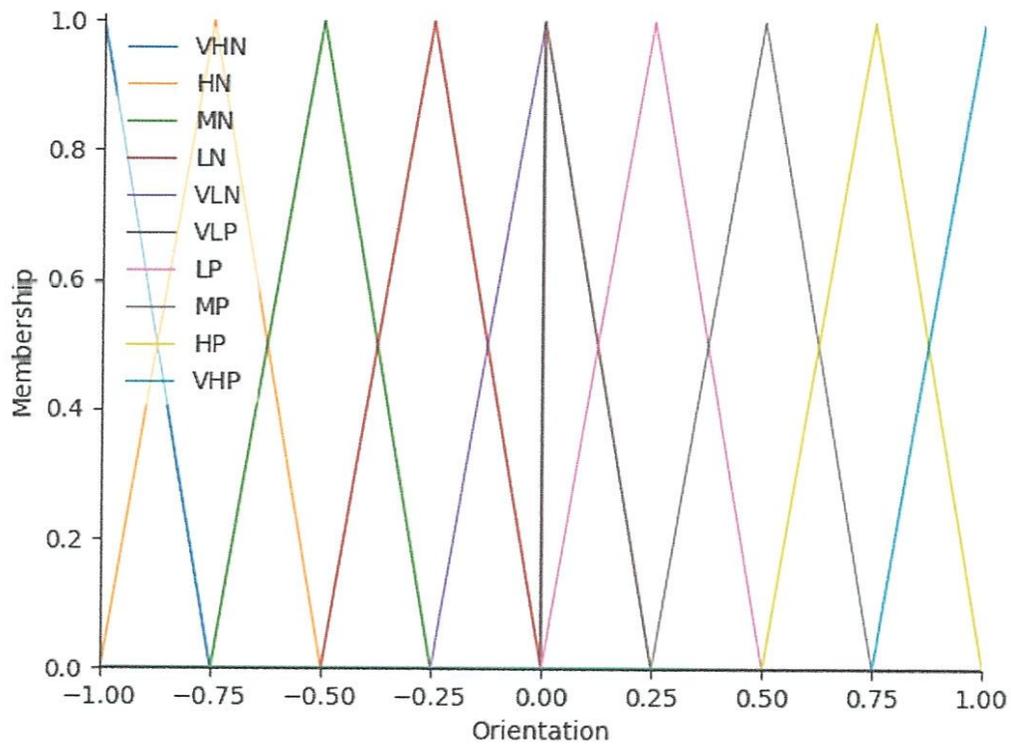


Figure 4.11 : l'ensemble flou global

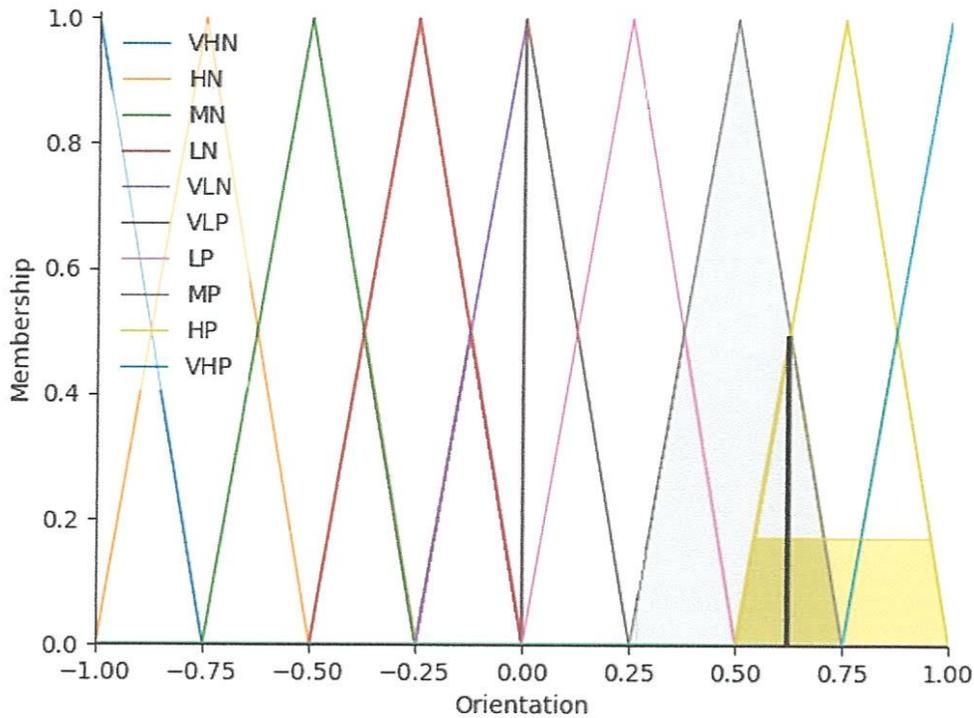


Figure 4.11 :l'agrégation et la defuzification

Donc le résultat obtenue c'est 0.67 high positif

4.5Conclusion

L'expérimentation que nous avons réalisée a donnée des très bonne résultat par apport au résultat des autres approche, la classification des sentiment avec l'amélioration de logique floue les résultats deviennent plus précises.

A l'avenir nous espérons introduire d'autres caractéristiques qui peuvent améliorer notre approche, et réalisé une approche qui peut travailler sur des sites qui ont une grande quantité des critique et confirmer la performance de notre technique.

Conclusion générale

Dans notre mémoire, nous avons présenté une approche d'identification d'opinions dans des commentaires introduit par les clients sur un produit, en utilisant la logique floue pour donner une classification fine et présenter son influence et performance sur le processus d'identification d'opinions.

Les techniques de classification actuelle sont pas assez efficaces pour décrire des informations qui représentent des opinions ou des faits, les informations qui représentent des faits sont généralement les mêmes, cependant les opinions sont des jugements ou pensées personnels sur un sujet ou un objet, qui sont généralement différentes aux autres. Dans ce cas, un mécanisme de résumé est indispensable pour rendre les opinions facilement exploitables.

On a proposé une approche dans le but de raffiner le résultat en divisant la polarité d'opinion en plusieurs classes. Dans le but d'obtenir un processus d'interrogation plus humain.

LISTE DES FIGURES :

<i>Figure 1.1 : Ensemble de commentaires collectés du site web restomontreal.ca.....</i>	<i>09</i>
<i>Figure 2.2 : Représentation classique.....</i>	<i>15</i>
<i>Figure2.3 : Représentation floue.....</i>	<i>16</i>
<i>Figure 2.4 : Exemple de haies.....</i>	<i>18</i>
<i>Figure 2.5 : Exemple d'opérations sur des ensembles flous.....</i>	<i>19</i>
<i>Figure 2.6 :Entrée x : température (°).....</i>	<i>21</i>
<i>Figure 2.7 : Entrée y : humidité (%).....</i>	<i>22</i>
<i>Figure2.8 : Sortie z : vitesse du ventilateur (%).....</i>	<i>24</i>
<i>Figure 2.9 – Agrégation des règles de sortie par coupure.....</i>	<i>25</i>
<i>Figure 2.10 : Comparaison entre la coupure et la mise à l'échelle.....</i>	<i>25</i>
<i>Figure 2.11 : Conséquences de style Sugeno.....</i>	<i>25</i>
<i>Figure 2.12 : Evaluation des conséquences de style Sugeno.....</i>	<i>26</i>
<i>Figure 3.1 : L'architecture d'approche proposée.....</i>	<i>30</i>
<i>Figure 3.2 représentation de la fonction d'appartenance triangulaire positif.....</i>	<i>31</i>
<i>Figure 3.3 représentation de la fonction d'appartenance triangulaire negatif.....</i>	<i>31</i>
<i>Figure 3.4 représentation de la fonction d'appartenance trapzoidale floue.....</i>	<i>32</i>
<i>Figure 4.1 : langage de programmation python.....</i>	<i>36</i>
<i>Figure 4.2 :l'éditeur spyder.....</i>	<i>37</i>
<i>Figure 4.1 : la suppression de punctuations et la répétition des lettre.....</i>	<i>39</i>
<i>figure 4.2 : Toknisation de phrase et l'affectation des étiquètes.....</i>	<i>39</i>
<i>Figure4.3 :Ectraction du score.....</i>	<i>39</i>
<i>Figure 4.4 :la fonction trapézoïdal des adverbe.....</i>	<i>40</i>

<i>Figure4.5 :la fonction d'appartenance triangulaire pour les verbe positif</i>	<i>41</i>
<i>Figure4.6 :la fonction d'appartenance triangulaire pour les verbe negatif.....</i>	<i>41</i>
<i>Figure4.7 :la fonction d'appartenance triangulaire pour les adjectif positif.....</i>	<i>42</i>
<i>Figure4.8 : la fonction d'appartenance triangulaire pour les adjectif positif</i>	<i>42</i>
<i>Figure4.9 : la fonction d'appartenance triangulaire pour les noms positifs.....</i>	<i>43</i>
<i>Figure4.10 : la fonction d'appartenance triangulaire pour les noms négatifs.....</i>	<i>54</i>
<i>Figure 4.11 :l'ensemble floue global.....</i>	<i>55</i>
<i>Figure 4.11 :l'agrégation et la defuzification avec le centre de gravité.....</i>	<i>56</i>

Liste des tableaux

Tableau 2.1 : Comparaison entre la logique classique et la logique floue pour un système de chauffage	15
Tableau 2.2 : Exemples de haies et représentation mathématique.....	18
Tableau 3.1 : des règles << if >> et << then >>	34