

République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieure et de la Recherche Scientifique



Université Saad Dahlab de Blida (U.S.D.B)

Faculté des Sciences

Département d'Informatique



Mémoire Présenté par :

AISSAOUI Abdenour

BENKRID Mahrez

En vue d'obtenir le diplôme de master

Domaine : Mathématique et informatique

Filière : Informatique

Spécialité : Informatique

Option : Ingénierie de logiciel

Thème : Apprentissage non supervisé pour le résumé vidéo multi-sources

Soutenu le :

M. *CHERIF ZAHAR AMING* Président

M. *BENYAHIA* Examineur

Mr. Kamech hichem abdellah Promoteur

Promotion 2015/2016

Remerciements

Tout d'abords, nous remercions dieu de nous avoir donné le courage et la volonté pour mener à bien ce modeste travaille au sein de notre université Saad Dahleb de Blida.

Nous souhaitons adresser notre sincères remerciements à notre promoteur monsieur kamech abdellah Hicham de nous avoir proposé le thème de ce mémoire ainsi que pour sa patience, ces aides et ces précieuse conseil et de nous avoir dirigés tout au long de notre travail avec son expérience et sa compétence.

Nous tenons également à remercier « TOUS » les Messieurs et dames, nos professeurs qui nous ont enseigné durant deux ans de formation master en Informatique, pour leurs précieux conseils et leurs orientations,

Nous remercions le membre de jury de nous avoir fait l'honneur d'évaluer notre travail tout en apportant leurs remarques et leurs contributions à l'enrichissement de ce mémoire.

Nous remercions également nos amis et collègues et tous ceux qui ont contribué de près ou de loin a la conclusion de ce mémoire.

En fin Nous remercions également nos familles respectives pour leur soutien et leurs encouragements tout au long de la réalisation de ce travail.

Mahrez & Abdenour

Dédicace

Je dédie ce modeste travail

A mes très chers parents qui ont tout fait pour que je réussisse dans ma vie et mes études, je leur souhaite le bonheur et la bonne santé et j'espère que vous seriez toujours fiers de moi

Je dédie ce travail à mes chers sœur Sabrina et Céline

A mon cher frère Belaid

A mon grand père mes grandes mères, à tous mes oncles et tantes, cousin et cousine

A toute ma famille sans exception

A tous mes amis en particulier mohamed, walid, abdou...

A mon chère frère et binôme mahrez

A tous ceux qui me sont chères

Abdenour

Je dédie ce modeste travail

A mes très chers parents qui ont tout fait pour que je réussisse dans ma vie et mes études, je leur souhaite le bonheur et la bonne santé et j'espère que vous seriez toujours fiers de moi

A ma chère grande sœur Hassiba, son épouse Ahmed et leur enfant Yassine

A mon chère frère Youcef, sa femme Nawel et leur enfant Rania

A mon grand père, ma grande mère, a tous mes oncles et mes tantes, cousines et cousins

A toute ma famille sans exception

A tous mes amis en particulier Mohamed, Ahmed, Abdou, Walid....

A mon cher binôme Abdenour

A tous ceux qui me sont chères

Mahrez

Résumé:

De nombreuses tâches de surveillance visuelle, comme le résumé vidéo, sont accomplies en analysant les caractéristiques basées sur les images. En s'appuyant uniquement sur des repères visuels pour la vidéo surveillance publique, cela s'avère peu fiable, les événements d'intérêt peuvent être subtiles. Nous pensons que les sources de données non visuelles telles que la météo et la circulation peuvent être exploitées afin de compléter les données visuelles à l'analyse de contenu vidéo et le résumé.

En particulier, nous étudions des moyens de composer avec les écarts de dimension et de représentation tout en associant ces sources de données hétérogènes, et dérivons un mécanisme efficace pour les données manquantes et incomplètes de différentes sources.

En outre, une étude de l'utilisateur complète est effectuée afin de valider la qualité du résumé vidéo généré à l'aide du modèle d'apprentissage multi-sources proposées.

Mots clé : données multi-sources, données hétérogènes, surveillance visuelle, clustering, la reconnaissance de l'événement, résumé vidéo

Abstract:

Many visual surveillance tasks, e.g. video summarization, is conventionally accomplished through analyzing imagery based features. Relying solely on visual cues for public surveillance video understanding is unreliable; since visual observations obtained from public space CCTV video data are often not sufficiently trustworthy and events of interest can be subtle. We believe that non-visual data sources such as weather reports and traffic sensory signals can be exploited to complement visual data for video content analysis and summarization. In this paper, we present a novel unsupervised framework to learn jointly from both visual and independently-drawn non-visual data sources for discovering meaningful latent structure of surveillance video data.

In particular, we investigate ways to cope with discrepant dimension and representation whilst associating these heterogeneous data sources.

In addition, a comprehensive user study is conducted to validate the quality of video summarisation generated using the proposed multi-source model.

Keywords: Multi-source data, heterogeneous data, visual surveillance, clustering, event recognition, video summarisation

Sommaire

Introduction générale.....	11
----------------------------	----

Chapitre 1: Etat de l'art

1. Introduction :	14
2. Définition du résumé vidéo.....	14
2.1. Vidéo :	15
2.2. Scènes :	15
2.3. Shots :	15
2.4. Frame « image »:	15
3. Objectif de résumé vidéo :	16
4. Les phases de résumé vidéo :	16
4.1 Segmentations de la vidéo :	16
4.2 Calcul des paramètres de résumé pour chaque image :	16
Mouvement « Motion » :	16
Extraction des images clés :	17
niveau audio :	17
4.3 calcul de l'importance des shots :	17
4.4 Choix des shots :	17
4.5 Créations du résumé vidéo :	18
5. Les techniques du résumé vidéo :	18
5.1 Résumé vidéo basés sur les fonctionnalités (feature based) :	18
5.2 Méthode basé sur motion :	18
5.3 Résumé vidéo basés sur les clusters :	18
5.3.1 Les méthodes du spectral clustering :	19
5.3.2 Les méthodes basées sur K-means :	19
5.4 Résumé Vidéo basée sur Clustering et utilise la Distance euclidienne :	20
5.5 Les méthodes basées sur les images-clés « key-frames »:	20
6. Domaine d'application :	20
6.1 Le domaine de la médecine :	21

6.2 Le sport :	22
6.3 Les vidéo surveillances :	22
7. Conclusion :	22

Chapitre 2: Généralité sur la classification

1. Introduction :	24
2. Définitions :	24
3. Domaines d'application et points de vocabulaire :	25
4. Exemples de problèmes de classification :	26
4.1. Prédiction e-mail / Spam :	26
4.2 Reconnaissance de formes :	27
4.3. Des Tableaux :	28
5. Fondements :	29
6. Les étapes d'une classification :	29
7. Approche Paramétrique versus non-paramétrique :	30
7. 1. Non paramétrique :	30
7.2. Paramétrique : « Probabilistes » :	30
8. Les types des méthodes de Classifications :	30
8.1. Méthode supervisée «Classement» ou « Discrimination» :	31
8.1.1 Définition :	31
8.1.2 Principe :	31
8.1.3 Les k plus proches voisins (K-PPV) :	32
8.1.4 Arbre de Décision :	34
8.1.5. Support Vector Machine :	35
8.1.6. Réseaux de neurone :	36
8.2 Méthode non supervisé :	37
8.2.1 Définition :	37
8.2.2. Principe général	38
8.2.3. Les exigences de Clustering :	39
8.2.4. Les types de Clustering :	39
8.2.5. Les algorithmes de Clustering :	42
8.2.6. Mesure de similarité	49

8.2.7. Les limites de Clustering :	53
9. Conclusion :	54

Chapitre 3: Solution proposé et conception

1. Introduction :	56
2. Problématique :	56
3. Solution proposée :	57
4. Apprentissage par Forêts aléatoires :	58
5. Données visuelles :	59
7. Les forêts de classification:	60
8. Les Forêts de clustering :	61
9. Multi-Source Clustering Forest:	62
9.1 Les termes visuels :	62
9.2 Les termes non visuels :	62
10. Comment faire face avec les données non-visuelles manquant :	63
11 .Découpage en frames :	63
12. Extraction des données visuelles :	63
13. Choix du descripteur dans nos travaux :	64
13.1. Histogramme de couleur RGB :	64
13.2 Histogramme de gradient orienté (HOG) :	66
13.2.1 Distance euclidienne	67
13.2.2 Distance de Mahalanobis	67
14. Conclusion :	68

Chapitre 4: Teste et résultats

1. Introduction :	70
2. Outils de développement :	70
2.1. Environnement matériel :	70
2.2. Environnement logiciel :	70
3. Présentation des benchmark utilisé :	71

4. Résultats et interprétations :	73
4.1 F-mesure :	73
4.2 L'entropie :	73
6. Conclusion :	75
 Conclusion Générale:	 76

Liste des figures:

Figure.1.1. Structure d'une vidéo	15
Figure 2.1 – E-mail valide / SPAM	27
Figure.2.2 – reconnaissance des caractères manuscrits	28
Figure 2.3. – Reconnaissance d'empreintes digitales.....	28
Figure 2.4. – Reconnaissance vocale.....	28
Figure 2.5. – données bancaires.....	29
Figure. 2.6– illustration des k-ppv d'un point	29
Figure. 2.7– . Le calcul des paramètres.....	32
Figure. 2.8–. Illustration de réseaux de neurone.....	33
Figure. 2.9– Illustration de regroupement en clusters.....	34
Figure. 2.10 – les deux types de clustering non-hiérarchique/hiérarchique.....	40
Figure. 2.11– Exemple d'un problème de discrimination à deux classes, avec une séparatrice linéaire : la droite d'équation $y=x$. Le problème est linéairement séparable.....	41
Figure. 2.12 – Fonction d'appartenance dans kmeans/Fuzzy C-means.....	45
Figure. 2.13 – le dendrogramme de la hiérarchie H de la suite de partitions d'un ensemble {a, b, c, d, e}.....	47
Figure. 2.14– différents écaillages peuvent conduire à différents clustering	50

Figure3.1 : illustration du modèle multi-source proposée.....	58
Figure 3.2:déférentes étapes du multi-source clustering avec le model multi-source clustering forest.....	60
Figure 3.3: Une illustration du clusturing des données avec les forêts de clusturing	61
Figure3.4 : image et son histogramme de couleur	65
Figure 3.5: Exemple d'image et son histogramme RGB	65
Figure 4.1: exemple sur l'ensemble des donnée TISI et ERCe	72
Figure4.2. Les valeurs de F-mesures pour l'approche proposée	74

Liste des tableaux :

Tableau. 2.1 : les noms attributs à la classification en Français/Anglais.....	25
Tableau 2.2 : Base d'exemples "jouer un match" pour la classification.....	32
Tableau 4.1 : Les mesures d'évaluation de l'approche proposée	73
Tableau 4.2 : comparaison des résultats obtenue avec les travaux antérieure.....	74

Introduction générale

De nos jours, le monde a connu beaucoup de développement dans les multimédia, et les contenus numériques. Nous observons aujourd'hui une augmentation dans le domaine de la vidéo. Il est entrain de devenir une des principales sources d'informations. Ce qui exige des outils pour manipuler et présenter les données de manière efficace. Toutefois, l'utilisateur peut ne pas avoir suffisamment de temps pour regarder la vidéo complète ou l'ensemble du contenu vidéo peut ne pas être intéressant pour l'utilisateur. Dans de tels cas, l'utilisateur peut vouloir juste voir les informations importantes de la vidéo originale au lieu de regarder toute la vidéo. Le résumé doit transmettre autant d'informations sur l'apparition de divers incidents dans la vidéo. En outre, la méthode doit être générale afin qu'elle puisse travailler avec les vidéos d'une variété de genres et de sélectionner ce qui doit rester dans le résumé et la façon dont cette information peut être extraites automatiquement.

Dans ce cas il existe plusieurs algorithmes qui nous permettent de réaliser ces résumés vidéo selon les besoin d'application.

La tâche cruciale dans la surveillance visuelle est d'effectuer automatiquement le résumé vidéo en prenant en compte l'énorme quantité de données. Dans la plupart des techniques, les données visuelles sont généralement exploitées comme la seule source de renseignements pour le résumé vidéo, d'autre part les événements ne peuvent pas être visuellement bien définis et facilement détectables. En particulier, la surveillance des données visuelles de l'espace public est souvent inexacte et incomplète due à la variation de sources incontrôlables telles que les variations de la lumière, de l'occlusion.

Dans cette étude, nous voulons exploiter l'information secondaire non visuelle pour compléter la perspective unilatérale d'après des observations visuelles. Des exemples de sources non visuelles comprennent les renseignements météorologiques, les données de trafic basés sur GPS, les données textuelles des réseaux sociaux, et les calendriers d'événements.

Les sources de données auxiliaires sont bénéfiques pour la modélisation de données visuelles car malgré que les données visuelles et non visuelles peuvent avoir des caractéristiques très différentes, elles dépeignent le phénomène physique commun dans une scène. Ces données hétérogènes sont liées, même si ce lien s'avère indirect.

Pour réaliser notre objectif et arriver à concevoir un apprentissage multi-sources pour un résumé vidéo, nous avons passé par plusieurs étapes. Ces étapes sont présentées dans les chapitres ci-dessous :

Chapitre 1 : Etat de l'art qui permet de donner une description de la structure et les différents composants et caractéristiques d'une vidéo, et décrit quelques méthodes de génération de résumés vidéo.

Chapitre 2 : Généralité sur la classification dans ce chapitre nous présentons quelques définitions et les différentes méthodes de classification ainsi que leurs domaines d'application.

Chapitre 3 : Solution proposée et conception ce chapitre comporte notre solution proposée et une explication sur la méthode utilisée ainsi que la conception de notre travail.

Chapitre 4 : Teste et résultats ce chapitre vise à décrire les différents outils utilisés pour la réalisation de notre projet et démontrer que notre système répond a l'objectif souhaité.

1. Introduction :

Récemment, le résumé vidéo a suscité beaucoup d'intérêt parmi les chercheurs et en conséquence, divers algorithmes et techniques ont été proposés. Dans ce travail, nous avons effectué un examen de la recherche sur le résumé vidéo : nous avons identifié des éléments importants et décrit la manière dont elles sont abordées dans les travaux spécifiques. Nous avons défini les techniques proposées pour produire un résumé vidéo d'une haute qualité visuelle et certaines approches sont adaptées pour le traitement vidéo en temps réel de divers types de vidéos compressées. Toutefois, une méthode d'évaluation valide peut prendre en charge le domaine à l'avance et la meilleure technique pour abstraire une séquence vidéo doit être identifiée. Bien que, la vidéo est toujours en grande partie en abstraction de la phase de recherche, les applications pratiques sont encore limitées dans la complexité de la méthode et l'ampleur du déploiement.

2. Définitions de résumé vidéo :

Un résumé vidéo est un résumé qui représente une vue abstraite de la séquence vidéo originale et peut être utilisé comme navigation vidéo et les systèmes de récupération. Il peut être aussi un moment fort de la séquence originale qui est la concaténation d'un nombre de segments vidéo sélectionnés ou peut être une collection d'images clés. [1]

Un résumé d'une vidéo doit satisfaire les trois principes suivants énoncés par Fayzullin et Al :

- Le résumé vidéo doit contenir des entités et des événements de haute priorité de la vidéo (les événements importants).
- le résumé doit être libre de la répétition.
- le résumé devrait présenter un degré raisonnable de continuité.
- Il peut être soit une image (images-clés) ou des images animées (vidéo) [1], d'où la vidéo se compose de scènes, shots et les frames (figure 1.1)

Chapitre 1 : état de l'art

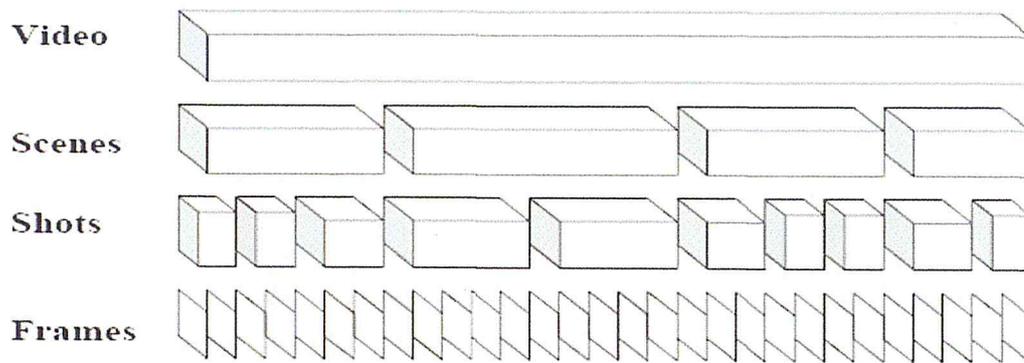


Figure.1.1. Structure d'une vidéo

2.1. Vidéo :

La vidéo regroupe l'ensemble des techniques permettant l'enregistrement ainsi que la restitution d'images animées, accompagnées ou non de son, sur un support électronique et non de type photochimique.[2]

2.2. Scènes :

La scène est le regroupement autour d'une même action de plusieurs séquences qui se déroulent en un ou plusieurs lieux, et dans la même temporalité [2].

2.3. Shots :

Le Shot est la définition la plus fondamentale dans la vidéo, c'est une unité de la vidéo enregistrée sur les caméras. En d'autres termes, la vidéo est composée de différents shots, qui comporte différents volets et types [3].

2.4. Frame « image »:

Une image est une représentation visuelle de quelque chose (objet, être vivant et /ou concept), Elle peut être naturelle (ombre, reflet) ou artificielle (peinture, photographie), visuelle ou non, tangible ou conceptuelle (métaphore), elle peut entretenir un rapport de ressemblance directe avec son modèle ou au contraire y être liée par un rapport plus symbolique[2].

Chapitre 1 : état de l'art

3. Objectif de résumé vidéo :

Le résumé vidéo est un mécanisme pour produire une courte vidéo pour donner à l'utilisateur une vue abstraite, synthétique et utile d'une séquence vidéo et pour faciliter le processus de recherche des contenus importants dans la vidéo en construisant une vidéo qui possède que les scènes importantes.

Nous considérons que le résumé vidéo est une partie importante du développement technologique, pour garder une collection de toutes les images des caméras. [1]

4. Les phases de résumé vidéo :

4.1 Segmentations de la vidéo :

La première étape de résumer vidéo est de segmenté la vidéo en plusieurs photos, la segmentation de l'image vise à regroupé les pixels similaires en régions d'où la segmentation de la vidéo généralise ce concept au regroupement des pixels en régions spatio-temporelles qui présentent une cohérence entre l'apparence et le mouvement. Une telle segmentation est utile pour plusieurs tâches de vision du niveau supérieur, tels que l'activité reconnaissance, suivi d'objet, la recherche basée sur le contenu, et l'amélioration visuelle [1].

On distingue deux types de segmentation :

La segmentation temporelle et la segmentation spatio-temporelles.

4.2 Calcul des paramètres de résumé pour chaque image :

Pour chaque image, les paramètres utilisé pour le résumer sont le mouvement, extraction des images-clés et le niveau du audio [1].

➤ **Mouvement « Motion » :**

Le mouvement dans une image vidéo c'est le critère le plus important pour faire le résumer vidéo car les images qui ont plus de mouvement dans la vidéo portes des informations très importantes sur ce qui se passe, pour cela ces informations devront être incluses dans le résumer vidéo.

Il existe plusieurs approches pour calculer le mouvement dans une image, l'une des approches la plus utilisé c'est le calculer de flux optique en utilisant l'algorithme de flux optique Horn et Schunck [1].

➤ Extraction des images clés :

C'est un autre critère très important pour décider quel 'shots' sont importants et quel sont les images clés qu'il contient, l'images-clés c'est une image qui représente le mieux le contenu de la vidéo de manière abstraite, Les moments sont calculés en utilisant la formule [1].

$$\mu_i = \frac{1}{N} \sum_{j=1}^N p_{ij},$$
$$\sigma_i = \left(\frac{1}{N} \sum_{j=1}^N (p_{ij} - \mu_i)^2 \right)^{1/2},$$
$$s_i = \left(\frac{1}{N} \sum_{j=1}^N (p_{ij} - \mu_i)^3 \right)^{1/3},$$

➤ Niveau audio :

Le son réalise également des informations très importantes pour le résumer vidéo par exemple dans une explosion dans un film ou un but dans un match de football les niveaux de l'audio sont très élevé, une telle scène devraient avoir grande chance d'être une partie de la vidéo résumé [1]

4.3 calcul de l'importance des shots :

Après que le calcul de tout les paramètres est fait, les coefficients de pondération pour chaque shot sont calculés aussi. Tous les paramètres (mouvement moyen, nombre d'images clés, entropie moyenne et audio moyenne) pour un shot sont normalisés dans l'intervalle [0, 1] en divisant chaque mesure par la valeur maximale de cette métrique pour tous les shots. Le poids affecté à chaque shot est la somme de chacune des métriques normalisées. Ainsi, les poids attribués aux shots se situent dans l'intervalle [0, 4].

4.4 Choix des shots :

La longueur de la vidéo résumée est déterminée en multipliant la valeur du pourcentage fourni par la longueur totale de la vidéo originale. Après avoir attribué des poids à chaque shot, il est évident que les shots qui ont les poids les plus élevés devraient être une partie de la vidéo résumée.

Chapitre 1 : état de l'art

C'est possible que la longueur du shot le plus important puisse dépasser la longueur de la vidéo résumée (lorsque la longueur du résumé doit être très petite), à chaque shot est attribué un «bit d'écriture» qui est mis à 0 au départ pour tous les shots. Si le shot est choisi pour être une partie de la vidéo résumée, ce bit est mis à 1.

4.5 Créations du résumé vidéo :

Les données de prise de vue sont analysées dans un ordre chronologique et le bit d'écriture de chaque shot est coché. Si le bit est 1, le shot est écrit dans le fichier de résumé. Les données audio correspondant sont écrites dans le fichier audio du résumé. Après le shot est terminer, nous procédons au shot suivant. Nous terminons le processus après tous les shot sont écrits dans la vidéo.

5. Les techniques du résumé vidéo :

5.1 Résumé vidéo basés sur les fonctionnalités (feature based) :

La vidéo numérique offre de nombreuses fonctions comme la couleur, la motion, et l'audio, ces techniques fonctionnent bien si l'utilisateur souhaite se concentrer sur les caractéristiques de la vidéo. Par exemple, si l'utilisateur veut voir la fonctionnalité couleur puis il faut choisir les couleurs de la vidéo basée sur les techniques de synthétisation. Les techniques « feature based» du résumé vidéo sont classés sur la base de la motion, couleur, contenus dynamiques, geste, audio-visuels, Discours Transcription, Object. [4]

5.2 Méthode basé sur motion :

C'est une méthode qui consiste à extraire une intensité de l'activité de mouvement des shots dans une vidéo compressée. La méthode utilise ensuite l'intensité de l'activité de mouvement pour segmenter la vidéo en segments faciles et difficiles à résumer. Les segments facile à résumer sont représentés par n'importe quelle frames sélectionnés à partir des segments facile à résumer tant que le processus d'extraction de résumé ce base sur la couleur génère les séquences des frames difficiles à résumer. Les frames sélectionné et généré de chaque segment à chaque shot sont combinés pour former le résumé de la vidéo compressée [5].

5.3 Résumé vidéo basés sur les clusters :

Dans cette technique on distingue plusieurs méthodes :

5.3.1 Les méthodes du spectral clustering :

Les méthodes de spectrale clustering se sont avérées efficaces dans la détection de structure de bloc d'une matrice de similarité pour certains dataset, Les algorithmes spectrale utilise le contenu de l'information dans les vecteurs propres de la matrice d'affinité des données pour détecter les structures.

La première étape dans cette méthode est de créer la matrice de similarité, à l'aide d'un ensemble prédéfinis de trames afin de stimuler la structure de bloc de la matrice de similarité. Après le calcul de la matrice de similarité, l'étape suivante consiste à déterminer le nombre de clusters dans la vidéo, Ce nombre est ensuite passé à l'algorithme K-means pour la détermination automatique du nombre de Cluster qui ne sont pas des tâche importantes.

La dernière étape de l'algorithme consiste à trouver les segments qui représentent la durée spécifiée de chaque cluster, cette méthode est basée sur l'estimation du mouvement.[6]

5.3.2 Les méthodes basées sur K-means :

C'est des méthodes basées sur l'algorithme de classification k-means pour séparer les séquences de la vidéo, le principe est de calculer la distance euclidienne des histogrammes des segments correspondant à la distance métrique.

La première étape est de créer les segments de la vidéo puis sélectionné l'histogramme, après on applique l'algorithme k-means pour regroupé les scènes reliaer. Le principe de l'algorithme k-means est de : [7]

1. Sélectionnez un k aléatoire.
2. Calculer les histogrammes sur les centroïdes des clusters.
3. Chaque histogramme est affecté à un cluster qui minimise la fonction d'erreur.
4. calculer centroïdes cluster.
5. A chaque itération, vérifiée pour voir si les centroïdes convergent. Sinon, nous allons à l'étape 2.
6. Après l'application de l'algorithme k-means l'étape prochaine est de sélectionné les segments pour construire la vidéo résumé.
7. L'avantage de k-means est une méthode simple et efficace pour le regroupement (clustering) des images similaire. [7]

Chapitre 1 : état de l'art

5.4 Résumé Vidéo basée sur Clustering et utilise la Distance euclidienne :

Cette méthode est basée sur la suppression des images vidéo redondantes qui ont les mêmes contenus. Comme beaucoup d'autres approches, le contenu vidéo entier est tout d'abord regroupé en nœuds, chacun contenant des images du même contenu visuel, en représentant chaque groupe avec l'image la plus représentative, un ensemble d'images clés est obtenu qui résume ensuite la séquence donnée, les étapes de cette méthode sont: [8]

1. acquisition de la vidéo et faire sortir les images de la vidéo.
2. calcul de la distance euclidienne.
3. Clustering « regroupement ».
4. Réduction des images.
5. Tri des images sélectionnés.
6. Composition de la vidéo [8]

5.5 Les méthodes basées sur les images-clés « key-frames »:

La plupart des travaux antérieurs dans le résumé vidéo se base sur le choix de des images-clés par hasard ou l'échantillonnage de l'image de la séquence originale de la vidéo à certains intervalles de temps, qui a été appliqué dans la vidéo de grossissement, système MiniVideo. Bien que c'est probablement la façon la plus simple d'extraire des images clés, l'inconvénient est que ce type d'arrangement peut entraîner des courts segments qui ne sont pas importants et encore représente des segments plus longue qui pourrait avoir plusieurs images avec le même contenu, ou bien des images manquantes ainsi de modifier et de représenter de la vidéo [9].

6. Domaine d'application :

De nombreuses applications professionnelles et éducatives qui impliquent la génération ou utilisant de gros volumes de données vidéo et multimédia sont les premiers candidats pour tirer parti des techniques d'analyse de contenu vidéo. Les techniques développées dans différents domaines la synthèse vidéo, nous trouvons dans trois catégories présentées : les applications vidéo grand public, gestion de bases de données et de surveillance Image-Vidéo. Pour chaque catégorie, certaines applications de l'exemplaire de référence sont énumérées.

Avec l'augmentation du stockage et de la capacité computationnelle de la consommation des appareils électroniques tels que les enregistreurs vidéo personnels, les applications vidéo

Chapitre 1 : état de l'art

grand public permet à l'utilisateur final de parcourir le contenu enregistré dans des moyens efficaces et visualiser les parties intéressantes rapidement. D'autre part, la gestion de bases de données d'images et de vidéos comprend différents domaines d'application comme le moteur de recherche vidéo, bibliothèque vidéo numérique, l'objet d'indexation et d'extraction, l'étiquetage et d'objet automatique.

Par conséquent, les organisations de médias et entreprises de radiodiffusion de télévision ont manifesté beaucoup d'intérêt dans ces applications, en particulier dans l'organisation et de l'indexation de gros volumes de données vidéo pour faciliter l'utilisation efficace et efficace de ces ressources à l'usage interne. Ces grandes libraires crée vidéo une occasion unique pour l'utilisation de techniques d'analyse des médias intelligente créer les techniques de recherche et de navigation avancée pour trouver l'information pertinente rapidement et à peu de frais. La segmentation intelligente de la vidéo.

6.1 Le domaine de la médecine :

Nous pouvons imaginer deux types de résumés pour l'échocardiographie vidéo. Dans le résumé statique toutes les vues de l'écho, la vidéo est représentée par une trame-clé. Ce hautement la navigation de facilite échocardiogramme vidéo. En sélectionnant l'une des images de l'aperçu, l'utilisateur devrait pouvoir visualiser le segment vidéo de ce point de vue seulement. Par conséquent, au lieu de regarder la totalité de la vidéo et à l'aide de l'outil de navigation Série d'un magnétoscope analogique, le clinicien a accès aléatoire aux différents points de vue.

Pour construire le résumé dynamique, la séquence d'images entre deux pics de l'onde R consécutifs sont extraites de chaque vue puis concaténés pour former le résumé. On peut créer des résumés dynamiques avec différents durée ou même différents contenus. Par exemple, depuis le 2D Echo vues sont vues les plus importants sur le plan clinique, on ne peut que créer un résumé qui est composé des vues 2D uniquement. Ou même basée sur la préférence de l'utilisateur, le résumé peut avoir différentes longueurs et au lieu d'un seul cycle de R-R pour chaque vue, on ne peut inclure plusieurs cycles de R-R dans le résumé. La dynamique résumé est également nommé résumé clinique parce que chaque cycle de R-R d'une vue du cœur dispose de suffisamment d'informations sur cette vue [10].

Chapitre 1 : état de l'art

6.2 Le sport :

Le résumé vidéo est appliqué dans différents domaines sur plusieurs genres de sport, y compris le football, le rugby et le hockey [11].

On peut utiliser les résumé vidéo dans la vie quotidienne aussi pour résumé des vidéo prise a l'aides des caméscopes et des Smartphones a fin de réduire la taille des vidéo pour gagné de l'espace.

6.3 Les vidéo surveillances :

Des millions de caméras de surveillance enregistrent de la vidéo, produisant d'énormes archives vidéo. Même quand un archive de vidéo est connu pour inclure des activités critiques, les trouver est difficile, pour cela il est important d'utiliser le résumé vidéo pour récupérer les activités suspect [12].

7. Conclusion :

Dans ce chapitre, nous avons élaboré un ensemble de définitions qui nous permettent de mieux comprendre les approches, aspects et composants d'une vidéo ainsi que déférentes méthodes du résumé vidéo et ces domaine d'application afin de mieux cerné le sujet et de pouvoir bien placé notre problématique .

Chapitre2 :

Généralités sur la classification

Chapitre2 : Généralités sur la classification

1. Introduction :

Les applications se sont multipliées, chacune apportant son lot de problèmes, plus complexe les uns que les autres. Après les biologistes, ce sont les historiens, les médecins, les sociologues ou les didacticiens qui ont manifesté le besoin de classer leurs reliques, leurs patients ou leurs élèves. Aujourd'hui, en nous intéressant aux problèmes issus des technologies de l'information et de la communication, nous sommes amenés à traiter des données très complexes, par leur nature, leur taille, leur provenance, leur diversités. Ces données peuvent être, par exemple, du texte, des images, des séquences vidéo, des bandes sonores et tout support combinant ces différents types de données.

L'envie d'organiser pour simplifier a progressivement évolué vers l'ambition de classer (ou classifier) pour comprendre et, pourquoi pas, pour prédire. Cette évolution a conduit à dégager deux stratégies de classification : supervisé et non-supervisé.

La notion de prédiction fait référence à une stratégie particulière d'Apprentissage Automatique (AA), appelée apprentissage supervisé. Ce domaine d'étude à part entière, peut être considéré comme une sous-thématique de l'Intelligence Artificielle (IA). De façon synthétique, l'apprentissage supervisé consiste à faire émerger d'un ensemble de données d'entraînement pré-classifié, les caractéristiques nécessaires et suffisantes pour permettre de classer correctement une nouvelle donnée. Dans ce type d'approche, les classes sont connues à l'avance, cette connaissance est utilisée dans le processus d'apprentissage.

La classification non-supervisée correspond davantage au problème initial d'organisation, et puise ses fondements théoriques en Analyse de Données (AD).

2. Définitions :

La classification est une discipline reliée de près ou de loin à plusieurs domaines, elle est connue aussi sous noms variés (classification, clustering, segmentation, . . .) selon les objets qu'elle traite et les objectifs qu'elle vise à atteindre.

Pour attribuer une définition au terme « classification », il faudrait d'abord définir ses racines, ça vient du verbe “classer” qui désigne plus une action qu'un domaine, ou plutôt une série de méthodes qu'une théorie unifiée.

En mathématique, On appelle classification, la catégorisation algorithmique d'objets. Elle consiste à attribuer une classe ou catégorie à chaque objet (ou individu) à classer, en se basant

Chapitre2 : Généralités sur la classification

sur des données statistiques. Elle fait couramment appel aux méthodes d'apprentissage et est largement utilisée en reconnaissance de formes.

Il est important de noter qu'il ne faut pas confondre entre ces deux termes : «classification» et « classement », au fait le mot classification en anglais signifie une chose, alors que le même mot en français ait une autre signification (utilité).

Dans un classement on affecte les objets à des groupes préétablis, c'est le but de l'analyse discriminante que de fixer des règles pour déterminer la classe des objets. La classification est donc, en quelque sorte, le travail préliminaire au classement, savoir la recherche des classes "naturelles" dans le domaine étudié, en anglais « Cluster Analysis ».

Cette collision entre les termes peut se résumer comme suite :

Français	Anglais
Classification	Clustering
Classement	Classification

Tableau. 2.1 – les noms attributs à la classification en Français/Anglais

D'une manière général en vertu de ces définitions, la classification se définit alors comme une méthode mathématique d'analyse de données, pour faciliter l'étude d'une population d'effectif important, généralement des bases d'observations caractérisent un domain particulier (animaux, plantes, malades, gènes,. . . etc.), où on les regroupe en plusieurs classes[13].

3. Domaines d'application et points de vocabulaire :

La classification comme dit préalablement joue un rôle dans presque toutes les sciences et techniques qui font appel à la statistique multidimensionnelle. A titre d'exemple les sciences biologiques : botanique, zoologie, écologie, ... qui utilisent le terme "taxinomie" pour désigner l'art de la classification. Ainsi que les sciences de la terre et des eaux : géologie, pédologie, géographie, étude des pollutions, font grand usage de classifications.

Une autre forte utilité des techniques de classification dans les sciences de l'homme : psychologie, sociologie, linguistique, archéologie, histoire, etc. ... et sans oublier les techniques dérivées comme les enquêtes d'opinion, le marketing, etc... Ces dernières emploient parfois les mots de "typologie" et "segmentation" pour désigner la classification, Citons encore la médecine [13], l'économie, l'agronomie . . . etc! Dans toutes ces disciplines

Chapitre2 : Généralités sur la classification

la classification peut être employée comme un domaine particulier ; mais elle l'est souvent vue comme une méthode complémentaire à d'autres méthodes statistiques. Elle est très largement utilisée à l'interprétation des graphiques d'analyse factorielle, ou bien déterminer des groupes d'objets homogènes, préalablement à une régression linéaire multiple.

Voilà les quelques exemples de ses utilités :

4. Exemples de problèmes de classification :

Ce sont les domaines que la classification pourrait viser et en même temps ils représentent les différents types de données d'entrées des techniques de classification, ce qui est nécessaire d'en présenter Avant d'aborder les méthodes classificatoires.

4.1. Prédiction e-mail / Spam :

Comme le fait de différencier un E-mail valide d'un SPAM, d'ailleurs la classification est fort utile en ce qui concerne la catégorisation des documents sur internet quel que soit la nature du document (image , fichier , son . . .etc) [14] et elle peut même être utilisée pour classifier les document selon leur sens (le web sémantique et les moteurs de recherche où on associe des sens pour les termes et pour les classifier il faut développer un langage de traitement/classification sémantique par exemple à base d'ontologie.) ou tout simplement pour classifier les ouvrages dans le monde des bibliothèques et des archives(le système de Classification de la Bibliothèque du Congrès LCC [1.1] (Figure 2.1)

Chapitre2 : Généralités sur la classification

e-mail valide	SPAM
<p>From : KOUDRI <KOUDRI_MOHAMMED@LIVE.fr> To : Mr BENAZOUZ <BENAZOUZ@UNIV_TLEMCEN.fr> Subject : SALAM ! Au fait, j'ai oublié de vous dire [...] Koudri</p>	<p>From : Pugh F Trina <LarsenSRTQG@power.ufscar.br> To : hue@ensmp.fr, hulin@ens.fr, item@ens.fr,jacq@ens.fr Subject: ENTIRE AMERICANHOSPITAL RESOURCE, NEW! NEW! INTRODUCTORY OFFER! JUST RELEASED! In a rapidly-changing industry, current healthcare information is an invaluable resource to businesses and organizations. The United States Healthcare Email Database includes [...]</p>

Figure 2.1 – E-mail valide / SPAM

4.2 Reconnaissance de formes :

Généralement c'est une question qui vise à reconnaître ou identifier certains motifs à partir de données brutes afin de prendre une décision dépendant de la catégorie attribuée à ce motif [15], ces motifs (formes) peuvent s'agir d'une image (visage, empreinte digitale, rayon X, EEG,...) ou sonore (reconnaissance de parole), et bien d'autres. Comme la reconnaissance des caractères manuscrits (Figure.2.2) :

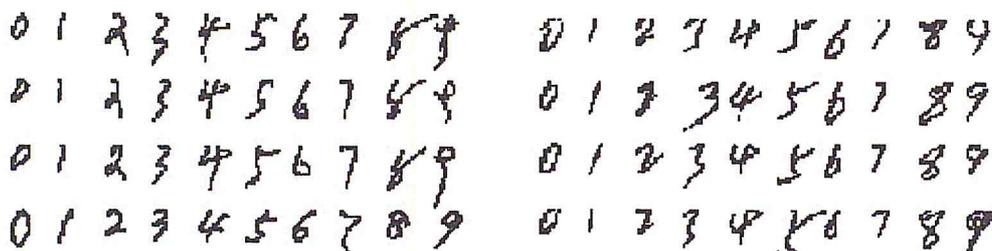


Figure.2.2 – reconnaissance des caractères manuscrits

Chapitre2 : Généralités sur la classification

Ou d'empreintes digitales : C'est l'exemple où on cherche à identifier une personne grâce à son empreinte (Figure 2.3),



Figure 2.3 – Reconnaissance d'empreintes digitales

Ou de Reconnaissance vocale : Dans le cas des signaux il ya des méthodes de traitement des langues pour fixer la classe d'appartenance d'un signale, ie : $\text{signal}(j)$ appartient à {français, anglais, ...} (Figure 2.4).

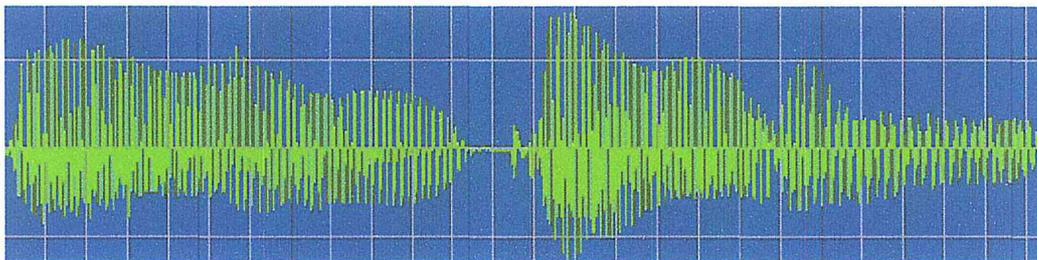


Figure 2.4. – Reconnaissance vocale

4.3. Des Tableaux :

Comme les données bancaires : ce type de données est représenté sous forme des tableaux $N \times M$ où N le nombre d'exemples (individus, objets) et M l'ensemble des descripteurs ou qualité de ce qui a été décrit (attributs) la case (i, j) de ce tableau contient l'information relative de l'élément j sur i ; le rôle de la classification dans ce cas, consiste à déterminer le comportement d'un individu par rapport a ce qu'on a appris des autres (base d'apprentissage) (Figure 2.5)

Chapitre2 : Généralités sur la classification

Transactions nationales	Taux de fraude (Montant de la fraude, en millions d'euros)			
	2004	2005	2006	2007
Palements	0,036 % (81,2)	0,033 % (82,8)	0,035 % (92,3)	0,032 % (95,6)
- dont paiements de proximité et sur automate	0,029 % (63,5)	0,025 % (59,2)	0,024 % (59,1)	0,017 % (45,4)
- dont paiements à distance	0,177 % (17,7)	0,195 % (23,6)	0,199 % (33,2)	0,236 % (50,1)
- dont par courrier / téléphone	nd	nd	0,194 % (18,8)	0,201 % (23,8)
- dont sur Internet	nd	nd	0,208 % (13,4)	0,281 % (26,4)
Retraits	0,027 % (22,7)	0,017 % (15,0)	0,019 % (17,4)	0,020 % (19,0)
Total	0,033 % (103,9)	0,029 % (97,8)	0,031 % (109,6)	0,029 % (114,5)

Source : Observatoire de la sécurité des cartes de paiement

Figure 2.5. – données bancaires

5. Fondements :

La philosophie des techniques classificatoires sur des objets qu'on souhaite traiter en catégories, se déroule généralement comme suite :

La classification vise à créer ces catégories à partir de traitements ne faisant intervenir que les données et pas la subjectivité de l'utilisateur (expert, simple utilisateur. . .) mais il est parfois nécessaire d'apprendre à partir d'exemples pour cela on aura besoin de la subjectivité de l'expérimentateur par le choix des descripteurs qu'il utilise, le classement selon un seul critère nous donnera pas le même classement selon un autre, la classification prends en considération tous les critères disponibles pour les classer en espace multi- dimensionnel.

Bien que les premières bases de l'approche algorithmique de la classification soient relativement anciennes, ce n'est qu'avec le développement de l'informatique qu'il est devenu possible de les mettre en œuvre sur de grands échantillons de données. Le résultat d'une méthode de classification peut être soit une partition mathématique soit une hiérarchie (mathématiques).

6. Les étapes d'une classification :

1. Choix des données.
2. Calcul des similarités entre les n individus à partir des données initiales.
3. Choix d'un algorithme de classification et exécution.

Chapitre2 : Généralités sur la classification

4. L'interprétation des résultats :

- évaluation de la qualité de la classification,
- description des classes obtenues.

7. Approche Paramétrique versus non-paramétrique :

7. 1. Non paramétrique :

Les approches dites non paramétriques (classification hiérarchique, méthode des centres mobiles) basée sur l'hypothèse : plus deux individus sont proches, plus ils ont de chances de faire partie de la même classe, en plus ce que distingue cette approche est qu'on ne fait aucune hypothèses sur le modèle que suivent les données, C'est le cas des plus proches voisins (k-PPV), donc il suffit de trouver les propriétés de convergence quand le nombre de données est grand.

7.2. Paramétrique : « Probabilistes » :

La seconde grande famille des méthodes de classification, ce sont les approches probabilistes, utilisent une hypothèse sur la distribution des individus à classer, c'est-à-dire, on suppose que l'on connaît la forme du modèle qui a généré les données. Par exemple, on peut considérer que les individus de chacune des classes suivent une loi normale. Le problème qui se pose, est de savoir déterminer ou estimer les paramètres des lois (moyenne, variance) et à quelle classe les individus ont le plus de chances d'appartenir à partir de l'ensemble d'apprentissage.

Les paramètres d'une loi peuvent être déterminés de maintes façons, C'est le cas par exemple des classifications bayésiennes ou encore l'algorithme espérance-maximisation. (Ce qu'on va l'attribuer tout un chapitre).

8. Les types des méthodes de Classifications :

On peut grouper les méthodes classificatoires en deux grandes familles, cette fois-ci , on prends en considération l'intervention ou non d'un « attribut classe » au fur et à mesure du processus de la classification, ces deux types sont : « supervisée (Classement)» et « non supervisée (Classification, Clustering) ».

1. supervisé (classement) : groupes fixés, exemples d'objets de chaque groupe.

2. non supervisé (classification) : on ne connaît pas de groupe.

Chapitre2 : Généralités sur la classification

Cependant, Il existe d'autres types de classification qui s'appuient sur d'autres types de méthodes d'apprentissages comme « l'apprentissage semi-supervisé » et « l'apprentissage par renforcement ». En effet, l'apprentissage semi-supervisé est un bon compromis entre les deux types d'apprentissage « supervisé » et « non-supervisé », car il permet de traiter un grand nombre de données sans avoir besoin de toutes les étiqueter, et il profite des avantages des deux types mentionnés. Alors que L'apprentissage par renforcement est fort utilisé dans le cas d'apprentissage interactif, Dans le reste de ce chapitre nous ne présenterons que le premier type « supervisé ».

8.1. Méthode supervisée «Classement» ou « Discrimination» :

8.1.1 Définition :

Le «classement» est une méthode supervisée qui consiste à définir une fonction qui attribue une ou plusieurs classes à chaque donnée. Dans cette approche on suppose qu'un expert fournit auparavant les étiquettes pour chaque donnée, les étiquettes sont des classes d'appartenance.

Selon [16] : « (la classification supervisée (appelée aussi classement ou classification inductive) a pour objectif « d'apprendre » par l'exemple. Elle cherche à expliquer et a prédire l'appartenance de documents à des classes connues a priori. Ainsi c'est l'ensemble des techniques qui visent à deviner l'appartenance d'un individu à une classe en s'aidant uniquement des valeurs qu'il prend)»

8.1.2 Principe :

La conception supervisée d'un classifieur à C classe (ensemble fini de classe c_i) est le fait de classifier N objets (x_i) de même nature (des phonèmes, caractères manuscrits, . . .) sachant que ces N objets sont supposés avoir été préalablement « étiquetés » par un « superviseur » en C ensembles qui forme un ensemble d'apprentissage.

Le superviseur n'est qu'un Classifieur en lequel on a confiance (expert humain, caractère répétitif, le système visuel humain . . .), donc notre système de la classification supervisée va être conçu en basant sur les exemples du superviseur (l'ensemble d'apprentissage où pour tout exemple on connaît à priori sa classe.) c'est-à-dire, on cherche à prédire si un objet (élément) « x_i » de la base de données, décrit par un ensemble de descripteurs «d», appartient ou non à une classe « c_j » parmi N Classes, pour le faire, on a un ensemble d'apprentissage décrit par :

Chapitre2 : Généralités sur la classification

$$A = (x_1, c_2), (x_2, c_4), (x_3, c_2) \dots (x_i, c_j) / x_i \in R^d, c_j \in C$$

Donc pour chaque objet x_i de l'ensemble de données, on peut connaître sa classe a priori c_j . La classification supervisée tente de chercher, à partir des données de A , une fonction de décision Γ qui va associer à tout nouveau élément x_i de test une classe c_j , puis on compare ce que nous a donné cette fonction avec la classe connue a priori de cet élément, de sorte à minimiser les mauvais classements $\Gamma(x_i \neq C_j)$.

Donc l'objectif est de chercher à prédire la classe de toute nouvelle donnée.

Exemple :

On pourrait donner l'exemple le plus connu : problèmes d'aide ou diagnostic médical, où les superviseurs sont généralement les médecins afin de noter la classe des objets de l'ensemble d'apprentissage à partir des remarques constatées. Ou bien l'exemple d'un tableau où le dernier descripteur (Jouer) représente la classe des exemples (Tableau.2.2)

Numéro	Ensoleillement	Température (°F)	Humidité (%)	Vent	Jouer
1	soleil	75	70	oui	oui
2	soleil	80	90	oui	non
3	soleil	85	85	non	non
4	soleil	72	95	non	non
5	soleil	69	70	non	oui
6	couvert	72	90	oui	oui

Tableau.2.2 – Base d'exemples "jouer un match" pour la classification

Ce tableau contient les données de l'apprentissage pour la classification dont chaque instance est labélisée par "Oui" et "Non". Ce qui permet de construire un modèle de classification permettant de prédire si ça marche pour jouer un match ou non ?

La déduction se fait par rapport à l'apprentissage sur le jeu de données.

Dans cet exemple on parle de « classification binaire », car on classe les données en deux classes ($|C| = 2$), et idem pour « n-aire » si on classe les données en n classes.

Nous allons présenter quelques techniques classiques de la classification supervisée :

8.1.3 Les k plus proches voisins (K-PPV) :

La méthode des plus proches voisins (noté parfois k-PPV ou k-NN pour -Nearest-Neighbor) consiste à déterminer pour chaque nouvel individu que l'on veut classer, la liste des plus proches voisins parmi les individus déjà classés. L'individu est affecté à la classe qui contient le plus d'individus parmi ces plus proches voisins. Cette méthode nécessite de choisir

Chapitre2 : Généralités sur la classification

une distance, la plus classique est la distance euclidienne (voir chapitre3.section6), et le nombre de voisins à prendre en compte.

Cette méthode supervisée et non-paramétrique est souvent performante. De plus, son apprentissage est assez simple, car il est de type apprentissage par coeur (on garde tous les exemples d'apprentissage). Cependant, le temps de prédiction est très long, car il nécessite le calcul de la distance avec tous les exemples, mais il existe des heuristiques pour réduire le nombre d'exemples à prendre en compte [17] (Figure. 2.6).

8.1.3.1 Algorithme :

1. initialisation, choix de :

- Nombre de classes, Valeur de k , exemples initiaux, mesure de similarité.

2. pour chaque vecteur d'objet à classer :

- mesurer la distance du vecteur avec tous les autres déjà classés
- déterminer la liste des k vecteurs les plus proches de lui (k -ppv)
- déterminer la classe la plus représentée dans la liste des k -ppv et affecter notre vecteur à cette classe.

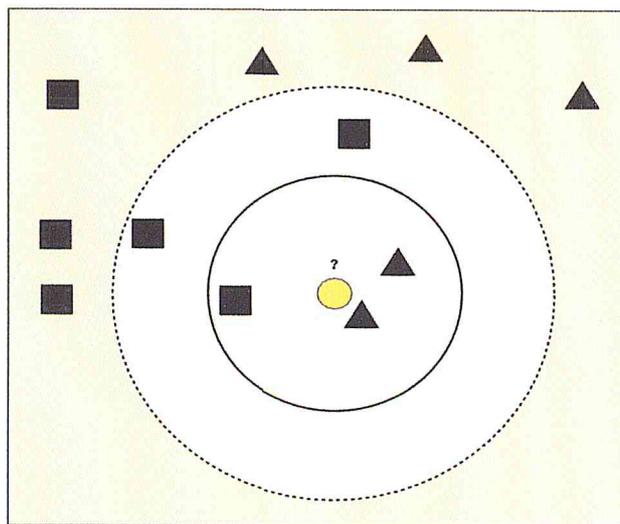


Figure. 2.6 – illustration des k -ppv d'un point

(La décision sera pour l'affecter à la classe majoritairement présente dans les k -ppv).

Chapitre2 : Généralités sur la classification

8.1.3.2 Discussion :

Ce qu'on peut remarquer sur cette méthode, c'est le cout de calcul qu'elle impose au fur et a mesure de ce processus de classification, car ce cout augmente avec chaque vecteur qu'on vient de classifié, plus on ajoute des nouveaux vecteurs déjà classés , plus que ce coute augmente ce qui explique le temps d'exécution qu'elle prend pour classifier. en plus de la sensibilité de cet algorithme à l'initialisation des paramètres d'entrées (le choix de k , la distance utilisée ..) alors il faut que lors de la sélection des paramètres d'entrées que ces derniers respectent certaines contraintes (comme que k ne soit pas un multiple du nombre de classes pour éviter une surreprésentation d'une classe par apport à une autre .malgré ces points , k-ppv reste une des méthodes les plus utilisées grâce à sa simplicité et robustesse et son caractère de généralisation a partir d'un nombre éminent de données d'apprentissage .

8.1.4 Arbre de Décision :

L'apprentissage se fait par partitionnement récursif selon des règles sur les variables explicatives. Suivant les critères de partitionnement et les données, on dispose de différentes méthodes, dont CART, CHAID ... Ces méthodes peuvent s'appliquer à une variable à expliquer qualitative ou quantitative. Deux types d'arbres de décision sont ainsi définis [14]:

8.1.4.1 Arbres de classification:

La variable expliquée est de type nominal (facteur). A chaque étape du partitionnement, on cherche à réduire l'impureté totale des deux nœuds fils par rapport au nœud père.

8.1.4.2 Arbres de régression:

La variable expliquée est de type numérique et il s'agit de prédire une valeur la plus proche possible de la vraie valeur.

Construire un tel arbre consiste à définir une suite de nœud, chaque nœud permettant de faire une partition des objets en 2 groupes sur la base d'une des variables explicatives. Il convient donc :

- De définir un critère permettant de sélectionner le meilleur nœud possible à une étape donnée,
- De définir quand s'arrête le découpage, en définissant un nœud terminal (feuille),
- D'attribuer au nœud terminal la classe ou la valeur la plus probable,

Chapitre 2 : Généralités sur la classification

- D'élaguer l'arbre quand le nombre de nœuds devient trop important en sélectionnant un sous arbre optimal à partir de l'arbre maximal,
- Valider l'arbre à partir d'une validation croisée ou d'autres techniques

8.1.4.3 Avantage :

- Méthode est non paramétrique et insensible aux valeurs extrêmes
- Elle permet de traiter de variables de natures différentes
- Elle comporte une sélection des variables
- Elle détermine des sous-populations définies par des règles facilement interprétables.
- On peut isoler certains nœuds et définir des classes de risque

8.1.5. Support Vector Machine :

8.1.5.1. Principe :

Les entrées X sont transformées en un vecteur dans un espace de Hilbert F . Dans le cas d'un classement en 2 classes, on détermine un hyperplan dans cet espace F . La solution optimale repose sur la propriété que les objets sont les plus éloignés possibles de l'hyperplan, on maximise ainsi les marges [11].

Soit x le vecteur associé. On définit $f(x) = \omega x + \beta$ et l'hyperplan a pour équation $\omega x + \beta = 0$.

La distance d'un point au plan est donnée par

$$d(x) = \frac{|\omega x + \beta|}{\|\omega\|}$$

Le classement est correct si $yf(x) > 0$ ou à un coefficient près $yf(x) \geq 1$.

8.1.5.2. Ajustement :

Maximiser la marge revient à minimiser $\|\omega\|$ ou $\|\omega\|^2/2$ sous les contraintes $y_i f(x_i) \geq 1$.

On utilise la méthode des multiplicateurs de Lagrange en ne conservant que les vecteurs x_i les plus proches de l'hyperplan (vecteurs supports).

Lorsque tous les cas ne sont pas séparables, on introduit un terme d'erreur : $y_i f(x_i) \geq 1 - \xi_i$.

Chapitre2 : Généralités sur la classification

La transformation en vecteur ne fait intervenir que l'expression du produit scalaire dans F . On recherche en fait directement l'expression du produit scalaire à partir des coordonnées initiales à l'aide d'une fonction k appelée noyau. On distingue les noyaux linéaire, polynômiaux, gaussien ...s'adaptant aux différentes problématiques rencontrées [11].

8.1.6. Réseaux de neurone :

8.1.6.1. Principe :

Cette méthode repose sur la notion de neurone formel. Un neurone formel est un modèle caractérisé par des signaux d'entrée (les variables explicatives par exemple), une fonction d'activation f , $f(\alpha_0 + \sum a_i x_i)$. f peut être linéaire, à seuil, stochastique et le plus souvent sigmoïde. Le calcul des paramètres se fait par apprentissage (Figure. 2.7).

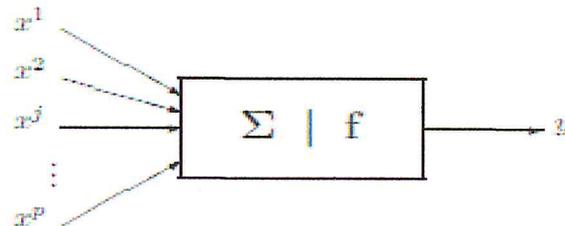


Figure. 2.7- . Le calcul des paramètres

Les neurones sont ensuite associés en couche. Une couche d'entrée lit les signaux entrant, un neurone par entrée x_j , une couche en sortie fournit la réponse du système. Une ou plusieurs couches cachées participent au transfert. Un neurone d'une couche cachée est connecté en entrée à chacun des neurones de la couche précédente et en sortie à chaque neurone de la couche suivante.

De façon usuelle et en régression (Y quantitative), la dernière couche est constituée d'un seul neurone muni de la fonction d'activation identité tandis que les autres neurones (couche cachée) sont munis de la fonction sigmoïde.

En classification binaire, le neurone de sortie est muni également de la fonction sigmoïde tandis que dans le cas d'une discrimination à m classes (Y qualitative), ce sont m neurones avec fonction sigmoïde, un par classe, qui sont considérés en sortie [11].

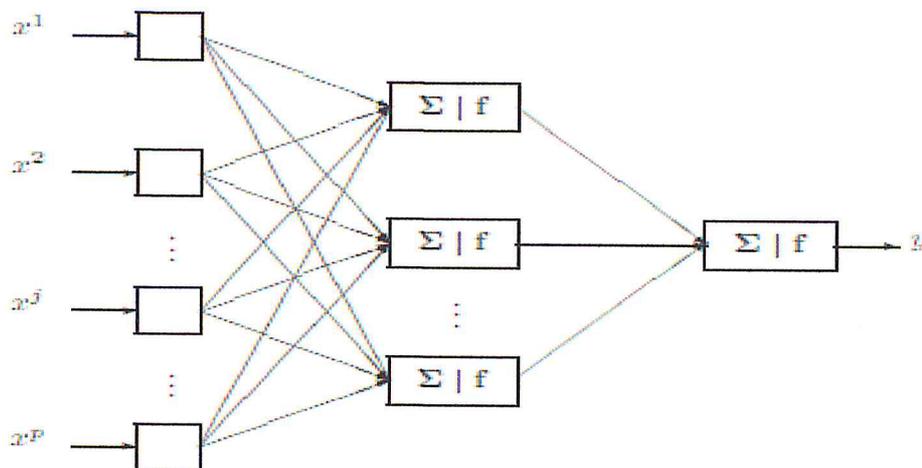


Figure. 2.8–. Illustration de réseaux de neurone

8.1.6.2. Paramètres de complexité :

Le modèle dépend de plusieurs paramètres :

- L'architecture du réseau : nombre de couches cachées (une ou deux en général) et le nombre de neurones par couche,
- Le nombre d'itération, l'erreur maximale tolérée et un terme de régularisation (decay).

Les paramètres de réglage sont difficiles à définir correctement. On peut utiliser library (e1071) par exemple pour rechercher les valeurs optimales

8.1.6.3. Apprentissage (ajustement) :

On minimise une fonction objective $Q(\alpha)$ (perte quadratique si Y est quantitative ou une fonction entropie en classification). A partir des gradients de cette fonction, on utilise un algorithme d'optimisation.

8.2 Méthode non supervisé :

8.2.1 Définition :

Le Clustering aussi connu sous nom (Segmentation) est un regroupement en classes homogènes consistant à représenter un nuage des points d'un espace quelconque en un ensemble de groupes appelé Cluster.

Chapitre2 : Généralités sur la classification

C'est un traitement sur un ensemble d'objets qui n'ont pas été étiquetés par un superviseur. Généralement lié au domaine de l'analyse des données comme ACP (analyse linéaire en composantes principales) [18] [19], ce type de méthodes vise à répondre au problème de : diminution de la dimension de l'espace d'entrée, ou pour le groupement des objets en plusieurs catégories (clusters) non définies à l'avance. Parmi les méthodes qu'on peut trouver dans ce type de classification : les cartes auto-organisatrices de kohonen [20], GMM . . . etc.

Un «Cluster» est donc une collection d'objets qui sont «similaires» entre eux et qui sont «dissemblables » par rapport aux objets appartenant à d'autres groupes. On peut voir cette définition clairement graphiquement dans l'exemple suivant (Figure. 2.9) :

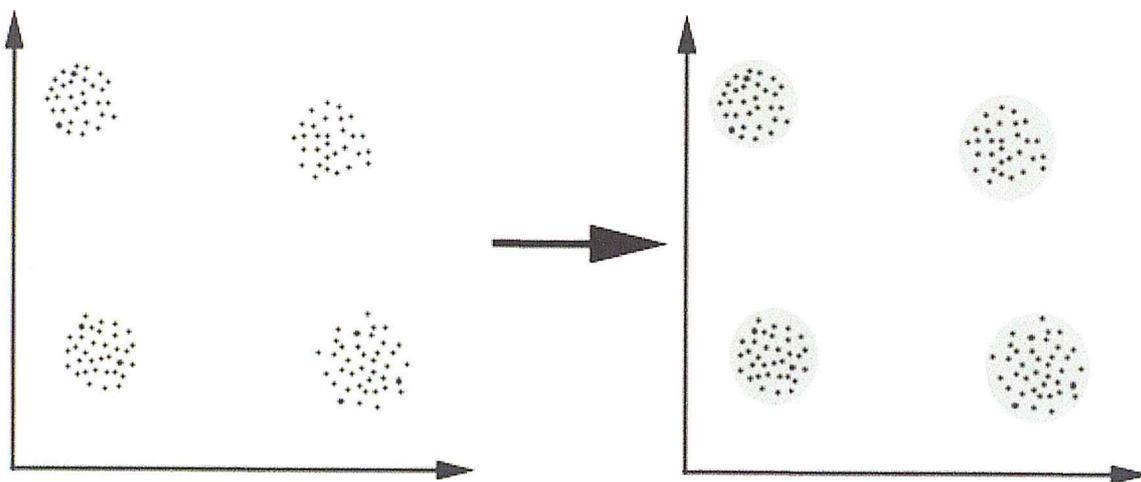


Figure. 2.9 – Illustration de regroupement en clusters

Dans ce cas, il est très facile pour une personne d'identifier 4 Clusters dans lesquels les données (nuage des points) peuvent être divisées, le critère de similarité est la distance : deux ou plusieurs objets appartiennent au même cluster s'ils sont «proches», bien sûr cela dépend d'une distance donnée (dans ce cas la distance géométrique).

Un autre type de regroupement est le clustering conceptuel : deux ou plusieurs objets appartiennent au même cluster si celui-ci définit un concept commun à tous les objets.

En d'autres termes, les objets sont regroupés en fonction de leur adéquation aux concepts descriptifs, et non pas en fonction de mesures de similarité simple.

8.2.2. Principe général :

Contrairement à la classification (méthodes supervisées), on ne possède pas des connaissances a priori sur les classes prédéfinies des éléments. Donc La division des objets

Chapitre2 : Généralités sur la classification

dans les différents groupes (clusters) se procède en se basant sur le calcul de similarité entre les éléments.

Alors que l'objectif des méthodes du Clustering est de grouper des éléments proches dans un même groupe de manière à ce que deux données d'un même groupe soient le plus similaires possible et que deux éléments de deux groupes différents soient le plus dissemblables possible

[21].

Mathématiquement, on a un ensemble X de N données décrites chacune par leurs P attributs.

Donc Le Clustering consiste à créer une partition ou une décomposition de cet ensemble en sous parties (clusters) telle que :

- Les données appartenant au même groupe se ressemblent.
- Les données appartenant à deux groupes différents soient peu ressemblantes

Exemple :

On utilise souvent ce type de classification en traitement d'images pour fixer les divers objets qu'elles contiennent (segmentation) : routes, villes, rues , des organes humaines (pour les images médicales) . . .

8.2.3. Les exigences de Clustering :

Les principales exigences qu'un algorithme de clustering doit répondre sont les suivantes :

- Evolutivité des clusters
- traiter les différents types d'attributs
- découvrir les clusters de forme arbitraire
- exigences minimales pour la connaissance du domaine afin de déterminer les paramètres d'entrée.
- capacité de composer avec le bruit et les valeurs manquantes traiter les dimensionnalités élevées. L'intelligibilité et la convivialité.

8.2.4. Les types de Clustering :

Il existe deux grands types du clustering (Figure. 2.10):

A/ le clustering hiérarchique : d'agglomération («bottom-up»)

B/ le clustering non-hiérarchique : de division («top-down»)

Dans le premier cas, on décompose l'ensemble d'individus en une arborescence de groupes.

Chapitre2 : Généralités sur la classification

Dans le 2ème, on décompose l'ensemble d'individus en K groupes, les algorithmes de ce type peuvent aussi être utilisés comme algorithmes de division dans le clustering hiérarchique.

TYPES OF CLUSTERING

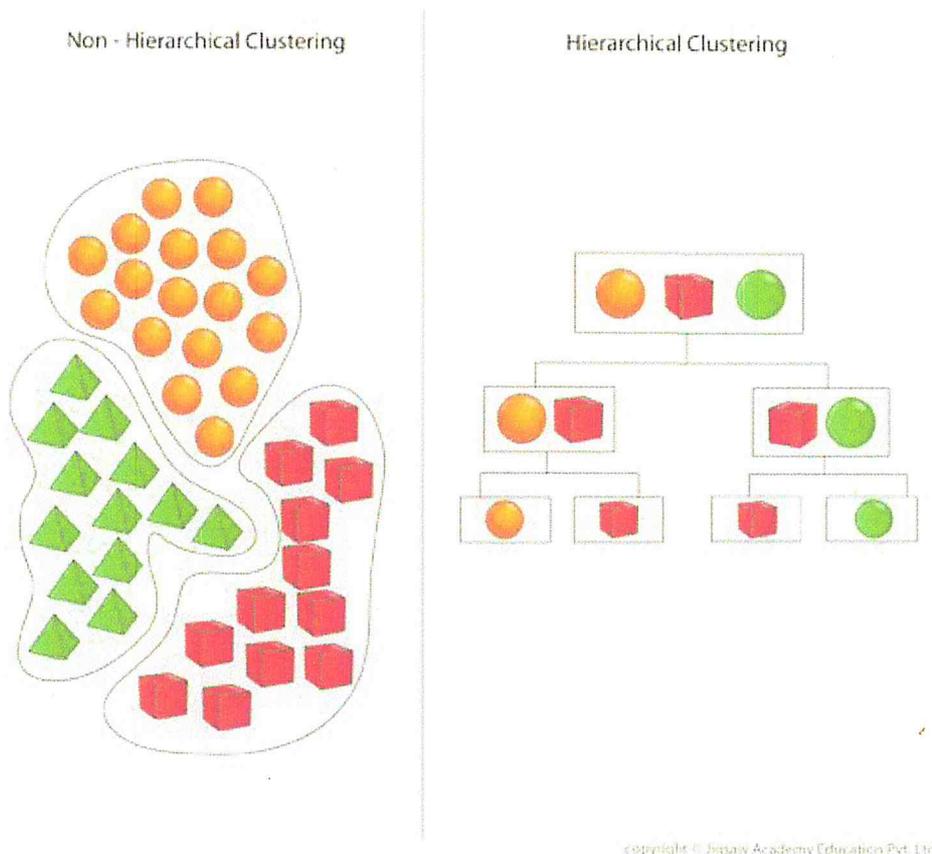


Figure. 2.10 – les deux types de clustering non-hiérarchique/hiérarchique

Cependant dans certains ouvrages on classifie les types des Algorithmes de clustering en 4 groupes à cause des méthodes qui ne respectent plus les normes du premier classement comme le cas de la règle « Chaque objet doit appartenir à un seul groupe. » alors que les versions floues la tempèrent et permettent à un objet d'appartenir à plusieurs classes selon un certain degré.

Les 4 types sont:

1. Clustering exclusif
2. Overlapping Clustering (fuzzy clustering)

Chapitre2 : Généralités sur la classification

3. Clustering Hiérarchique

4. Clustering probabiliste

Dans le premier cas, les données sont regroupées d'une manière exclusive, de sorte que si une donnée certaine appartient à un amas définie alors il ne pourrait pas être inclus dans un autre cluster. Un simple exemple de cela est montré dans la figure ci-dessous (Figure. 2.11), où la séparation des points est définie par une ligne droite sur un plan bidimensionnel.

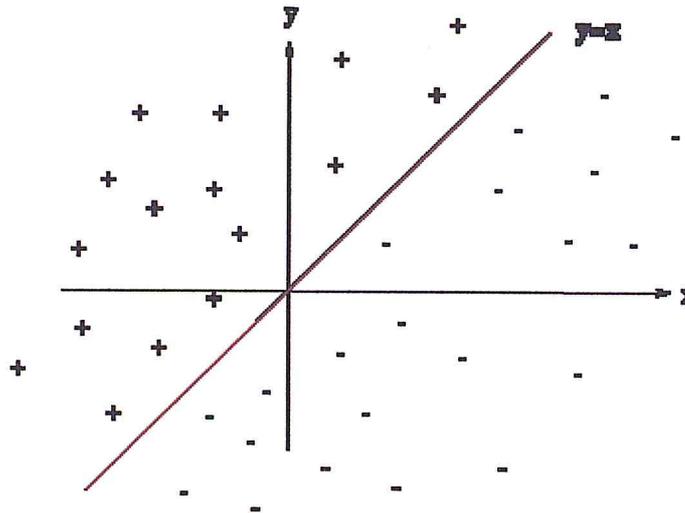


Figure. 2.11– Exemple d'un problème de discrimination à deux classes, avec une séparatrice linéaire : la droite d'équation $y=x$. Le problème est linéairement séparable.

Au contraire le second type, le regroupement overlapping, utilise des ensembles flous aux données de cluster, de sorte que chaque point peut appartenir à deux ou plusieurs groupes avec différents degrés d'appartenance. Dans ce cas, les données seront associées à une valeur d'une composition appropriée.

Comme nous l'avons dit, un algorithme de clustering hiérarchique est fondé sur l'union entre les deux plus proches clusters cad : consiste à trouver des clusters successifs utilisant des clusters précédemment établis. La première condition est de mettre, au début, chaque objet dans un cluster distinct et les fusionner en clusters successivement plus grand. Après quelques itérations on atteint le final Cluster voulu qui regroupe tous les sous-clusters (sous-partitions).

Enfin, le dernier type de regroupement utilise une approche complètement probabiliste basant sur la probabilité d'appartenance aux clusters.

Chapitre2 : Généralités sur la classification

8.2.5. Les algorithmes de Clustering :

Dans ce qu'il suit nous présentons quelques algorithmes de Clustering, voilà quelques exemples:

1. K-means
2. Fuzzy C-means
3. Hierarchical clustering
4. Forêt d'arbres décisionnels

Chacun de ces algorithmes appartient à l'un des types de clustering énumérés ci-dessus. Par exemple, K-means est un algorithme de clustering exclusif, pendant que Fuzzy C-means est un algorithme de Overlapping Clustering, alors que clustering hiérarchique il est claire qu'il s'agit de troisième type de clustering,. Nous allons discuter et définir les principes de ces méthode de clustering dans quelques lignes.

8.2.5.1. K-means :

L'algorithme k-means mis au point par McQueen en 1967[22], un des plus simples algorithmes d'apprentissage non supervisé, appelée algorithme des centres mobiles [23] [24], il attribue chaque point dans un cluster dont le centre (centroïde) est le plus proche. Le centre est la moyenne de tous les points dans le cluster, ses coordonnées sont la moyenne arithmétique pour chaque dimension séparément de tous les points dans le cluster c.à.d chaque cluster est représentée par son centre de gravité.

8.2.5.1.1 Principe :

L'idée principale est de définir les k centroïdes arbitraires c_1, c_2, \dots, c_k (k le nombre de clusters fixé a priori, chaque c_i représente le centre d'une classe), Ces centroïdes doivent être placés dans des emplacements différents. Donc, le meilleur choix est de les placer le plus possible éloignés les uns des autres. La prochaine étape est de prendre chaque point appartenant à l'ensemble de données et l'associer au plus proche centroïde. C à d Chaque classe Si sera représentée par un ensemble d'individus les plus proches de sont c_i , Les nuées dynamiques sont une généralisation de ce principe, où chaque cluster est représenté par un noyau mais plus complexe qu'une moyenne.

Lorsqu'aucun point n'est en attente, la première étape est terminée et un groupage précoce est fait. À ce point nous avons besoin de recalculer les k nouveaux centroïdes mi des groupes issus de l'étape précédente qui vont remplacer le c_i (m_j est le centre de gravité de la classe

Chapitre2 : Généralités sur la classification

S_j , calculé en utilisant les nouvelles classes obtenues). Après, on réitère Le processus jusqu'à atteindre un état de stabilité où aucune amélioration n'est possible, nous pouvons constater que les k centroïdes changent leur localisation par étape jusqu'à plus de changements sont effectués. En d'autres termes les centroïdes ne bougent plus.

8.2.5.1.2. Algorithme :

Choisir k moyennes c_1, c_2, \dots, c_k initiales (par exp au hasard)

1. Répéter :

Affectation de chaque point à son cluster le plus proche :

$$S_i^{(t)} = \{x_i : \|x_i - m_i^{(t)}\| \leq \|x_i - m_{i^*}^{(t)}\| \text{ for all } i^* = 1, \dots, k\}$$

Mettre à jour la moyenne de chaque cluster

$$m_i^{(t+1)} = \frac{1}{|S_i^{(t)}|} \sum_{x_j \in S_i^{(t)}} x_j$$

2. Jusqu'à : atteindre la convergence quand il n'y a plus de changement.

Fin.

8.2.5.1.3. Discussion :

Cette méthode est la plus populaire des méthodes de clustering, malgré ça, un de ses problèmes majeurs est qu'il tend à trouver des classes sphériques de même taille. En plus K-means est connu par sa complexité de « NP-difficile ». Il est donc fréquemment faire appeler une heuristique en pratique, ce qui explique qu'elle est convergente et surtout avantageuse du point de vue calcul mais elle dépend essentiellement de la partition initiale (Des initialisations différentes peuvent mener à des clusters différents « problèmes de minima locaux ») cela risque d'obtenir une partition qui ne soit pas optimale pourtant qu'elle donne sûrement une partition meilleure que la partition initiale. De plus, la définition de la classe se fait à partir de son centre, qui pourrait ne pas être un individu de l'ensemble à classer, d'où le risque d'obtenir des classes vides.

Chapitre2 : Généralités sur la classification

8.2.5.2 Méthode Fuzzy C-means :

8.2.5.2.1. Principe :

Fuzzy C-means (FCM) est une méthode de clustering qui permet à un objet de données d'appartenir à deux ou plusieurs clusters. Cette méthode dérivée de l'algorithme c-means [25], identique à l'algorithme k-means décrit précédemment, elle a été développée par Dunn [26] en 1973 et améliorée par Bezdek [27] en 1981, est fréquemment utilisée dans la reconnaissance des formes. Il est basé sur la minimisation de la fonction objective suivante :

$$J_m = \sum_{i=1}^N \sum_{j=1}^C U_{ij}^m \|x_i - c_j\|^2 \quad 1 \leq m < \infty$$

Où m est un nombre réel (> 1), U_{ij} est le degré d'appartenance de x_i dans le j ème Cluster, x_i est le i ème élément des données mesurées, c_j est le centre d'un cluster et $\|*\|$ est toute norme exprimant la similarité entre les données mesurées et le centre. Ce Partitionnement logique flou (fuzzy) est réalisé grâce à une optimisation itérative de la fonction objectif indiqué ci dessus, avec la mise à jour de l'appartenance u_{ij} et les centres des clusters c_j

On peut résumer la différence entre fuzzy C-means et k-means dans la fonction d'appartenance d'un nuage de points dans deux clusters dans l'exemple suivant (Figure. 2.12):

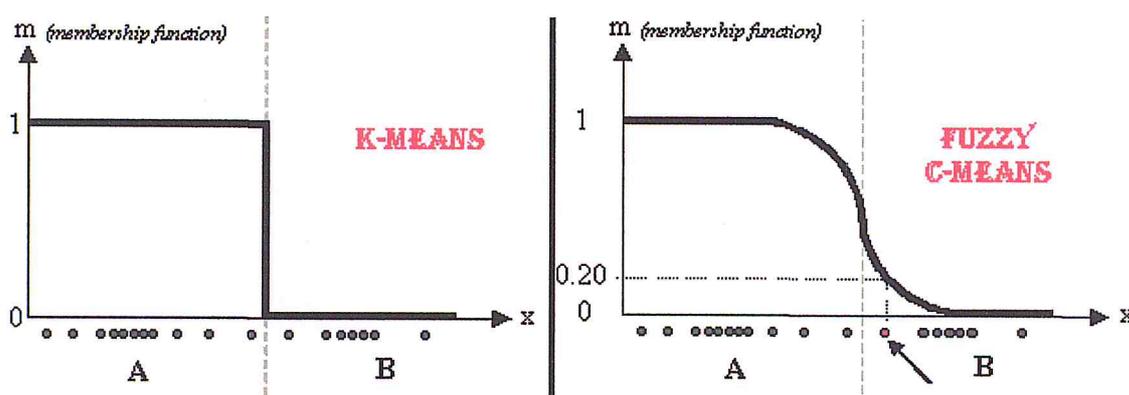


Figure. 2.12 – Fonction d'appartenance dans kmeans/Fuzzy C-means

Dans le cas de k-means un objet ne peut pas appartenir dans deux clusters Simultanément, ce qui explique la Discrimination binaire entre les clusters mais en FCM il est possible qu'un objet appartienne à deux ou plusieurs clusters selon différents pourcentages cad que les données

Chapitre2 : Généralités sur la classification

sont liés à chaque groupe par le biais d'une fonction d'appartenance, ce qui représente le comportement flou de cet algorithme. Pour le faire, nous devons simplement construire une matrice appropriée nommée U dont les facteurs sont des nombres entre 0 et 1, et représentent le degré d'appartenance entre les centres de données et des clusters.

$$U_{\text{MFC}} = \begin{bmatrix} 0.8 & 0.2 \\ 0.3 & 0.7 \\ 0.6 & 0.4 \\ \dots & \dots \\ 0.9 & 0.1 \end{bmatrix}$$

Il est également important de noter que les initialisations différentes causent différentes évolutions de l'algorithme. En fait, il pourrait converger vers le même résultat, mais probablement avec un nombre différent d'itérations.

8.2.5.2.2. Algorithme

Il ya des parties des equations cathé il faut les récriture

1. Initialiser $U = [u_{ij}]$ matrice $U_{(0)}$.
2. A la k-étape : calculer les centres $C_{(k)} = [c_j]$ avec $U_{(k)}$

$$c_j = \frac{\sum_{i=1}^N u_{ij}^m \cdot x_i}{\sum_{i=1}^N u_{ij}^m}$$

3. Mise à jour de $U_k, U_{(k+1)}$

$$u_i^{j=1} C_{k=1} \|x_i - c_j\| \|x_i - c_k\|^{2m} - 1$$

4. Si $\|U_{(k+1)} - U_k\| < \varepsilon$ ($0 < \varepsilon < 1$), alors STOP, sinon le retour à l'étape 2.

8.2.5.2.3. Discussion :

Une méthode que son caractère hybride (la notion de centre de gravité et la notion Floue) le rend simple, rapide. La FCM exige des paramètres d'entrées, et que la matrice de partition floue, doit être initialisée d'une manière appropriée. Ces paramètres sont choisis d'une façon

Chapitre2 : Généralités sur la classification

arbitraire, ces paramètres ont une grande influence sur le résultat attendu. Ce qu'il nous oblige de faire une étude appropriée sur les données en entrée et le regroupement que l'on souhaite obtenir.

Ce type d'algorithme est fort utilisé en traitement d'images [28] [29] [30] afin d'identifier des zones similaires (contours, coins, région homogènes. . .).

8.2.5.3. Méthodes hiérarchiques :

Le processus basique des méthodes hiérarchiques a été donné par Johnson et Lance Et Williams, [31] [32], Ce type de clustering consiste à effectuer une suite de regroupements en Clusters de moins en moins fines en agrégeant à chaque étape les objets (simple élément) ou les groupes d'objets (un Cluster-partition-) les plus proches. Ce qui nous donne une arborescence de clusters. Cette approche utilise la mesure de similarité pour refléter l'homogénéité ou l'hétérogénéité des classes.

8.2.5.3.1. Principe :

Son principe est simple, initialement chaque individu forme une classe, soit n classes , donc on cherche à réduire ce nombre de classe $n_{\text{newnbrclss}} < n$ itérativement de sorte que dans chaque étape on fusionne deux classes ensemble (Les deux classes choisies pour être fusionnées sont celles qui sont les plus "proches" en fonction de leur dissimilarité) ou ajouter un nouveau élément à une classe (un élément appartient à une classe s'il est plus proche de cette classe que de toutes les autres) La valeur de dissimilarité est appelée indice d'agrégation. Qui commence dans la première itération faible, et croît d'itération en itération.

Parmi les algorithmes plus connus de ce type : La classification ascendante hiérarchique (CHA) où le mot ascendante est utilisé pour désigner qu'elle part d'une situation dont tous les individus représentent des clusters à part entière, puis on cherche les rassembler en classes de plus en plus grandes. Ainsi Le qualificatif "hiérarchique" désigne le fait qu'elle produit une hiérarchie, (une amélioration a été proposée en 2002 par P. Bertrand , appelée Classification Ascendante Hiérarchique).

8.2.5.3.2. Algorithme de CHA :

1. Initialisation :

Chaque individu est placé dans son propre cluster, Calcul de la matrice de ressemblance M entre chaque couple de clusters (ici les points)

2. Répéter :

- Sélection dans M des deux clusters les plus proches C_i et C_j

Chapitre2 : Généralités sur la classification

- Fusion de C_i et C_j par un cluster C_G plus général
 - Mise à jour de M en calculant la ressemblance entre C_G et les clusters existants
- Jusqu'à fusionner les 2 derniers clusters.

Dans la figure suivante, on représente une illustration du principe de CHA et la hiérarchie finale obtenue où Les liens hiérarchiques apparaissent clairement (Figure. 2.13).

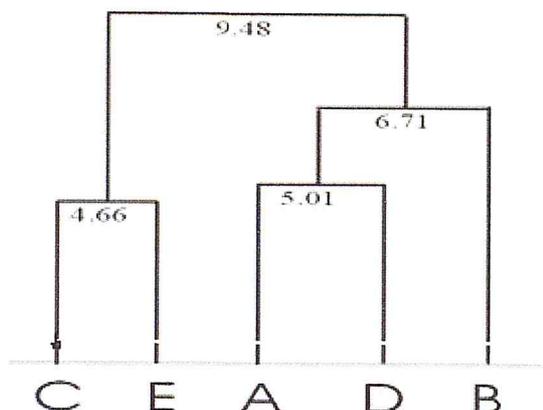


Figure. 2.13 – le dendrogramme de la hiérarchie H de la suite de partitions d'un ensemble {a, b, c, d, e}

note : Un dendrogramme = la représentation graphique d'une classification ascendante hiérarchique sous forme d'un arbre binaire

8.2.5.3.3. Avantages de CHA :

La classification ascendante hiérarchique (CAH) est une méthode de classification qui présente les avantages suivants [14] :

- On travaille à partir des dissimilarités entre les objets que l'on veut regrouper. On peut donc choisir un type de dissimilarité adapté au sujet étudié et à la nature des données.
- L'un des résultats est le dendrogramme, qui permet de visualiser le regroupement progressif des données. On peut alors se faire une idée d'un nombre adéquat de classes dans lesquelles les données peuvent être regroupées.

Chapitre2 : Généralités sur la classification

8.2.5.3.4. Discussion :

la CAH ne nécessite pas de connaître le nombre de clusters a priori. De plus, il n'y a pas de fonction d'initialisation, ainsi une seule construction d'un cluster (équivalent à une itération pour les méthodes de partitionnement).

En ce qui concerne généralement les méthodes hiérarchiques le problème qu'on peut rencontrer réside dans la sélection d'une ultra-métrique (distance pour calculer la similarité entre clusters) soit la plus proche de la métrique utilisée pour les individus, car ces méthodes sont heuristiques, pour cela ya plusieurs techniques permet de le faire : Saut minimal (single linkage) ; Saut maximal (complete linkage) ; Saut moyen ; Barycentre. . . une autre faiblesse est : la complexité de temps d'au moins $O(n^2)$, où n est le nombre d'objets au total, ainsi qu'on pourrait jamais défaire ce qui a été fait précédemment.

Il est difficile parfois d'apporter une justification aux méthodes hiérarchique (CAH, CDH.), Cependant, dans [33], une interprétation probabiliste de la CAH, basée sur une estimation par maximum de vraisemblance des modèles de mélange, est proposée comme solution pour mieux interpréter les résultats.

Un autre inconvénient de ce type de méthodes est qu'une action effectuée (fusion ou décomposition), elle ne peut être annulée. Cela permet de réduire le champ d'exploration, mais une telle astuce ne peut corriger une décision erronée. Afin améliorer la qualité d'une classification hiérarchique, on peut profiter de deux techniques:

- analyser attentivement les liens entre objets à chaque étape [34] et [35].
- améliorer la partition obtenue avec une méthode de deuxième type de clustering (partitionnement) [36].

8.2.5.4. Forêt d'arbres décisionnels :

Les forêts d'arbres décisionnels [14] (ou forêts aléatoires de l'anglais « *Random decision forest* ») ont été formellement proposées en 2001 par Leo Breiman et Adèle Cutler. Elles font partie des techniques d'apprentissage automatique. Cet algorithme combine les concepts de sous-espaces aléatoires et de « *bagging* ». L'algorithme des forêts d'arbres décisionnels effectue un apprentissage sur de multiples arbres de décision entraînés sur des sous-ensembles de données légèrement différents

8.2.5.4.1. Algorithme :

La base du calcul repose sur l'apprentissage par arbre de décision. La proposition de Breiman vise à corriger plusieurs inconvénients connus de la méthode initiale, comme la sensibilité des

Chapitre2 : Généralités sur la classification

arbres uniques à l'ordre des prédicteurs, en calculant un ensemble de N arbres partiellement indépendants.

Une présentation rapide de la proposition peut s'exprimer comme suit :

1. Créer N nouveaux ensembles d'apprentissage par un double processus d'échantillonnage :
 - 1.1 sur les observations, en utilisant un tirage avec remise d'un nombre K d'observations identique à celui des données d'origine (technique connue sous le nom de bootstrap),
 - 1.2 et sur les P prédicteurs, en n'en retenant qu'un échantillon de cardinal $m < \sqrt{P}$ (La limite n'est qu'indicative).
- 2 Sur chaque échantillon, on entraîne un arbre de décision selon une des techniques connues, en limitant sa croissance par validation croisée.
- 3 On stocke les N prédictions de la variable d'intérêt pour chaque observation d'origine.
- 4 La prédiction de la forêt aléatoire est alors un simple vote majoritaire (*Ensemble learning*).

Le principal revers de cette méthode est que l'on perd l'aspect visuel des arbres de décision uniques.

8.2.6. Mesure de similarité

Pour comparer homogénéité ou le ressemblance, la similarité entre deux objets (points, images, classes, phonème ..), il faut pouvoir mesurer la similarité (ou la dissimilarité) entre eux.

Nous allons décrire maintenant des mesures de similarité pour prouver la similarité entre les objets, selon [37], «tout système ayant pour but d'analyser ou d'organiser automatiquement un ensemble de données ou de connaissances doit utiliser, sous une forme ou une autre, un opérateur de similarité dont le but est d'établir les ressemblances ou les relations qui existent entre les informations manipulées».

Donc la similarité est une partie importante de la définition d'une méthode de clustering, elle consiste en effet à définir et formaliser une mesure de similarité adaptée aux caractéristiques des données. Si les composantes des vecteurs de données d'instance sont toutes dans les mêmes unités physiques alors il est possible que la distance euclidienne soit suffisante pour

Chapitre2 : Généralités sur la classification

réussir à grouper les données similaires. Cependant, même dans ce cas, la distance euclidienne peut parfois être trompeuse. La Figure ci-dessous illustre ceci avec un exemple vu selon la largeur et la hauteur d'un objet. Malgré que les deux mesures aient été prises dans les mêmes unités physiques (Figure. 2.14).

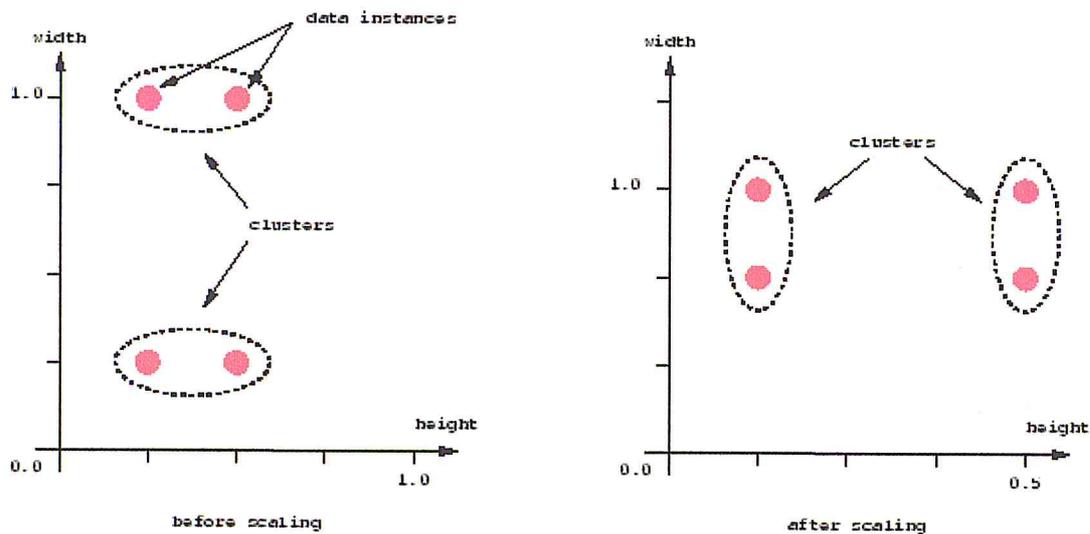


Figure. 2.14 – différents écaillages peuvent conduire à différents clustering

Donc une décision éclairée doit être faite quant à la mise à l'échelle relative. Comme le montre la figure, différents écaillages peuvent conduire à différents clustering.

8.2.6.1. Vocabulaire :

Il est à noter qu'il ya deux concepts pour exprimer la notion de proximité entre les objets à classifier :

1. Mesure de dissimilarité DM : plus la mesure est faible plus les points sont similaire (distance).
2. Mesure de similarité SM : plus la mesure est grande, plus les points sont similaires.
3. On parle souvent de « distances » en désignant une mesure de similarité, lorsque ces mesures ont les propriétés de non-négativité, réflexivité, symétrie (la distance entre l'objet A à B est la même que la distance de B à A) et qui respectent l'inégalité triangulaire.

Il existe un grand nombre de mesures de similarité, dans ce qui suit, nous présentons quelques unes des fonctions entre deux objets $d(x_1, x_2)$.

8.2.6.2. Fonctions de similarité :

- **La distance euclidienne** : (aussi appelée la distance à vol d'oiseau) Un rapport de clusters analysis en psychologie de la santé a conclu que la mesure de la distance la

Chapitre2 : Généralités sur la classification

plus courante dans les études publiées dans ce domaine de recherche est la distance euclidienne ou la distance au carré euclidienne.

$$d^2(x_1, x_2) = \sum_i (x_{1i} - x_{2i})^2 = (x_1 - x_2)'(x_1 - x_2)$$

- **La distance de Manhattan** : (appelée aussi taxi-distance)

$$d^2(x_1, x_2) = \sum_i |x_{1i} - x_{2i}|$$

- **La distance de Mahalanobis** : corrige les données pour les différentes échelles et des corrélations dans les variables, L'angle entre deux vecteurs peuvent être utilisés comme mesure de distance quand le regroupement des données de haute dimension. Voir l'espace produit scalaire.

$$d^2(x_1, x_2) = (x_1 - x_2)'C^{-1}(x_1 - x_2)$$

($C = \text{covariance}$)

- **La distance de Sebestyen** :

$$d^2(x_1, x_2) = (x_1 - x_2)'W(x_1 - x_2)$$

$W = \text{matrice diagonale de pondération}$

- **La distance de Hamming** : mesure le nombre minimum de substitutions nécessaire pour changer un membre dans un autre. Elle permet ainsi, de quantifier la différence entre deux séquences de symboles, généralement utilisée dans le cas des valeurs discrètes (vecteurs)

$$d(a, b) = \sum_{i=0}^{n-1} (a_i \oplus b_i)$$

Où d est la dimensionnalité des données. La distance euclidienne est un cas particulier où

Chapitre2 : Généralités sur la classification

$p = 2$, alors que Manhattan $p = 1$. Néanmoins, il n'existe pas de directives générales théoriques pour la sélection d'une mesure à une application donnée. Une autre question, est de savoir comment mesurer la distance entre 2 classes $D (C_1; C_2)$? Pour cela il ya certaines fonctions permettent de mesurer cette distance comme :

Plus proche voisin :

$$\min(d(i, j), i \in C_1, j \in C_2)$$

Diamètre maximum :

$$\max(d(i, j), i \in C_1, j \in C_2)$$

Distance moyenne :

$$\frac{\sum_{i,j} d(i, j)}{n_1 n_2}$$

Distance des centres de gravité :

$$d(\mu_1, \mu_2)$$

Distance de Ward :

$$\sqrt{\frac{n_1 n_2}{(n_1 + n_2)}} d(\mu_1, \mu_2)$$

8.2.6.3. Discussion

Une note importante est de savoir si le clustering utilise une distance symétrique ou asymétrique. Bon nombre des fonctions énumérées ci-dessus ont la propriété que les distances sont symétriques. Dans d'autres applications (par exemple, la séquence-alignement).

Certaines mesures sont spécifiques aux domaines particuliers comme histogrammes ou aux distributions. Dans, on trouvera une étude comparative de quelques de ces fonctions .

Chapitre2 : Généralités sur la classification

En plus, ces mesures rencontrent certaines difficultés lorsqu'on change le jeu de données comme le fait de travailler sur des espaces de couleurs où quelque distance ne sont pas recommandées. L'inconvénient major de la plupart de ces fonctions, c'est qu'elles sont coûteuses en temps de calcul et sont de plus sensibles à la dimension des données. Pour remédier le problème de dimensions, il ya des techniques ont été proposées pour la réduction de dimensions, qui permettent d'appréhender cette difficulté [38].

8.2.7. Les limites de Clustering :

Il ya un certain nombre de problèmes avec le clustering. Parmi eux :

- les techniques de clustering actuelles ne traitent pas tous les besoins de façon adéquate (et simultanément), comme le fait que si nous n'avons pas des variables continuées (la longueur), mais les catégories nominales, comme les jours de la semaine. Dans ces cas encore, la connaissance du domaine doit être faite pour formuler le clustering appropriée.
- traitement d'un grand nombre de dimensions et de grand nombre de données, question peut être problématique en raison de la complexité du temps de calcule.
- l'efficacité de la méthode dépend de la définition de «distance» utilisée.
- si la mesure de la distance n'existe pas, nous devons la «définir», ce qui n'est pas toujours facile, surtout dans des espaces multidimensionnels.
- le résultat de l'algorithme de clustering peut être interprété de différentes manières.
- Beaucoup d'algorithmes de clustering exigent la spécification du nombre de clusters à produire en entrée de l'ensemble de données, avant l'exécution de l'algorithme. ie : connaissance de la valeur correcte à l'avance, la valeur appropriée doit être déterminée un problème pour lequel un certain nombre de techniques ont été développées.

Chapitre2 : Généralités sur la classification

9. Conclusion :

Nous avons vu une généralité sur les méthodes de classification. La Discrimination (ou les méthodes supervisées) peut être basée sur des hypothèses probabilistes (Classifieur naïf de Bayes, méthodes paramétriques) ou sur des notions de proximité (plus proches voisins) ou bien encore sur des recherches dans des espaces d'hypothèses (arbres de décision, réseaux de neurones).

Certes l'approche supervisée est très utilisée pour les raisons et les avantages qu'on a mentionnés pour chaque méthode, néanmoins il reste qu'il y a un manque de stratégies pour les exemples d'auto-apprentissage (c'est-à-dire, d'apprendre à partir d'une base sans aucune connaissance préalable) que les méthodes supervisées ne peuvent pas traiter, dans ce cadre vient la deuxième approche des méthodes de classification, qui est : l'approche non-supervisée (ou spécifiquement « la classification automatique »).

Les méthodes de clustering comme toutes les autres méthodes de classification, ont leurs avantages, faiblesses (voir section : discussion), cependant, il n'y a pas que le type statistique, il y'en a d'autre type qui s'appuient sur la théorie de probabilité. Dans le chapitre suivant nous nous intéresserons à une nouvelle méthode qui les le clustering multi source en forêts en démontrant ces avantages et sa robustesse en qualité de classification.

Chapitre 3 :

Solution proposé et conception

Chapitre 3 : solution proposé et conception

1. Introduction :

Après avoir élaboré les chapitres précédents respectivement l'état de l'art sur les résumé vidéo et les différentes notions sur la classification utilisé dans les résumé vidéo, nous passons dans ce chapitre à la conception et la modélisation du travail, commençant par la spécification des besoins, puis la conception du système en proposant de faire une classification des données visuelle tout en exploitant l'information secondaire non visuelle en utilisant la méthode des random forest et un ensemble de descripteurs et d'algorithmes pour une meilleure sémantique.

2. Problématique :

Il n'est pas facile de formuler un cadre qui exploite à la fois des données visuels et non visuelles à l'analyse de contenu et de résumé vidéo, tant en termes d'algorithmes et dans la pratique.

En termes d'algorithmes, l'interaction entre des sources de données hétérogènes est confrontée à un certain nombre de défis :

1/ la disparition des sources qui sont différentes dans la représentation (continues ou catégoriques), et varient largement en ampleur et covariance. En outre, la dimension des sources visuelles dépasse souvent celle des informations non-visuelle dans une grande mesure. En raison de ce problème de divergence dans la dimensionnalité, une simple concaténation de caractéristiques se traduira par une représentation défavorable des données visuelles.

2/ Les deux données visuelles et non-visuelles dans l'isolement peuvent être inexactes et incomplètes.

Dans la pratique, les sources de données secondaires, par exemple la météo, les rapports de circulation et les tableaux d'événements, peuvent être peu fiables en disponibilité. En outre, les salles de commande vidéo existantes ne possèdent pas forcément un accès direct à ces sources. Cela rend les modèles qui attendent l'information visuelle et non visuelle complète au cours du déploiement impossible.

Chapitre 3 : solution proposé et conception

3. Solution proposée :

Dans cette étude, nous abordons le problème d'apprentissage multi-source dans le contexte de résumé vidéo, conventionnellement basé sur l'analyse de fonction visuelle et la détection d'objets/segmentation. En particulier, nous formulons un nouveau cadre qui est capable d'effectuer l'apprentissage Multi-sources sur des données hétérogènes (Fig.3.1). Nous considérons les données visuelles comme source principale et les données non visuelles comme sources secondaire, puisque nous croyons que l'information visuelle joue toujours le rôle principal dans l'analyse de contenu vidéo. Au cours de la formation, le modèle effectue le clustering de données multi-source et découvre un ensemble de clusters de données visuelles marquées avec la distribution de données non visuelles, par exemple, le climat et la vitesse de circulation.

Par la suite, les clips clés sont automatiquement choisis parmi les clusters découverts. Le résumé vidéo peut être produit par la composition chronologique de ces clips clés.

Nous proposons un cadre unifié d'apprentissage multi-sources capable de découvrir les structures sémantiques de contenu vidéo collectivement à partir des données hétérogènes visuelle et non visuelles. Ceci est rendu possible par la formulation d'une nouvelle forêt de clustering. Bien que les deux données visuelles et non visuelles puissent être inexactes et incomplètes, notre modèle est capable de découvrir et ensuite exploiter le partage pour une meilleure corrélation des données. La figure suivante illustre la solution proposée (Figure3.1)

Chapitre 3 : solution proposé et conception

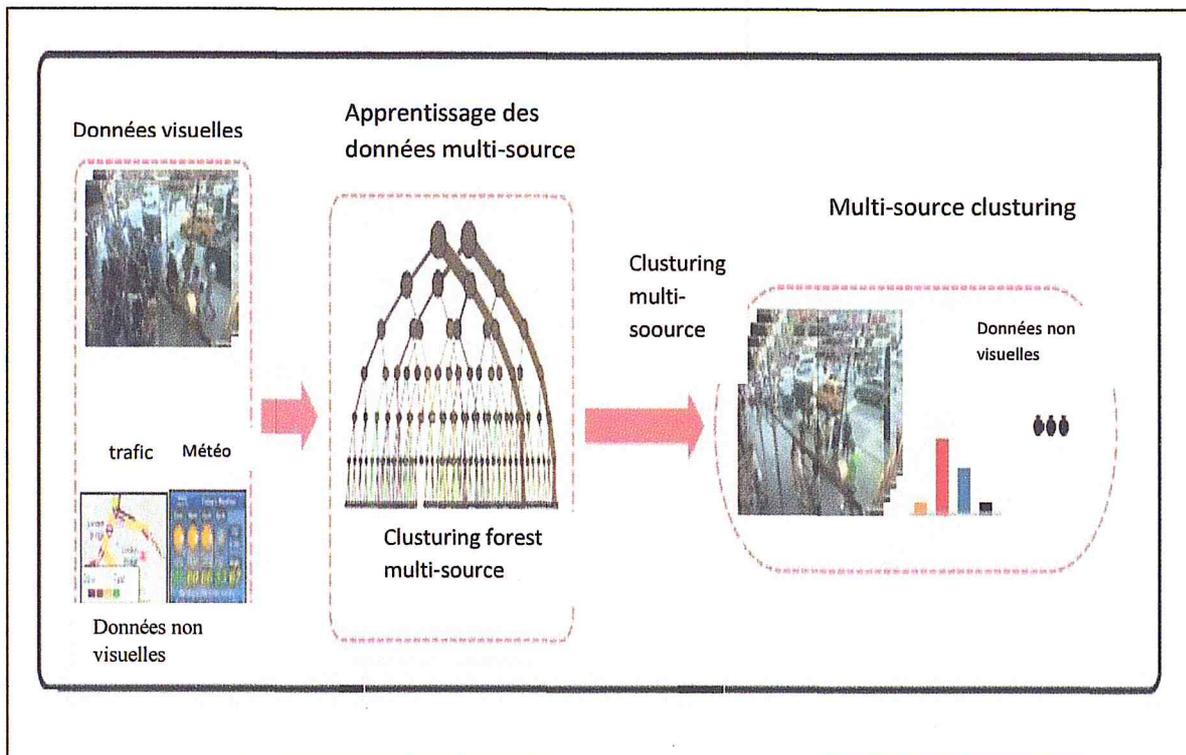


Figure 3.1 : illustration du modèle multi-source proposée

4. Apprentissage par Forêts aléatoires :

Différentes variantes de forêts aléatoires ont été élaborées, soit supervisées, ou non supervisées [39], les modèles supervisées ne conviennent pas à notre problème puisque nous ne prenons pas la disponibilité des étiquettes pendant la formation du modèle. Les modèles de forêts de clustering existant, d'autre part, supposent que les sources de données soient homogènes. Aucun principe de combiner plusieurs sources de données indépendantes et hétérogènes dans les modèles de forêt n'est disponible.

Une façon commune d'établir un résumé vidéo est en extrayant puis en combinant une série d'images clés. Ces matières sont généralement détectées et sélectionné à partir de clusters d'images vidéo ou clips [40].

Dans cette étude, nous avons suivi l'approche susmentionnée mais nous avons envisagé non seulement le contenu visuel de la vidéo, mais aussi un grand corpus de données non visuelles recueillies auprès de sources indépendantes hétérogènes (Fig.3.2(a)). Plus précisément, via l'apprentissage de structures latentes de données multi-sources (Fig. 3.2 (b-c)). Nous souhaitons faire référence à ça et imposer la sémantique non visuelles directement dans le clustering sans aucune annotation manuelle humaine des données vidéo (Fig.3.2(d)).

Chapitre 3 : solution proposé et conception

5. Données visuelles :

Une vidéo peut être vue comme un ensemble de N clips d'une durée de T clips chacun. Chaque clip peut être défini par un descripteur visuel de dimension d , on note alors les descripteurs visuels ainsi :

$$X_i = (x_{i,1}, \dots, x_{i,d}) \in \mathbb{R}^d, i = 1, \dots, N$$

6. Données non visuelles :

Les données non visuelles sont recueillies auprès de sources hétérogènes indépendantes. Nous représentons collectivement m Types de données non visuelles associés au i ème clip vidéo comme suite :

$$y_i = (y_{i,1}, \dots, y_{i,m}) \in \mathbb{R}^m, i = 1, \dots, N$$

Notez qu'une ou plusieurs informations liées à un clip peuvent être manquantes.

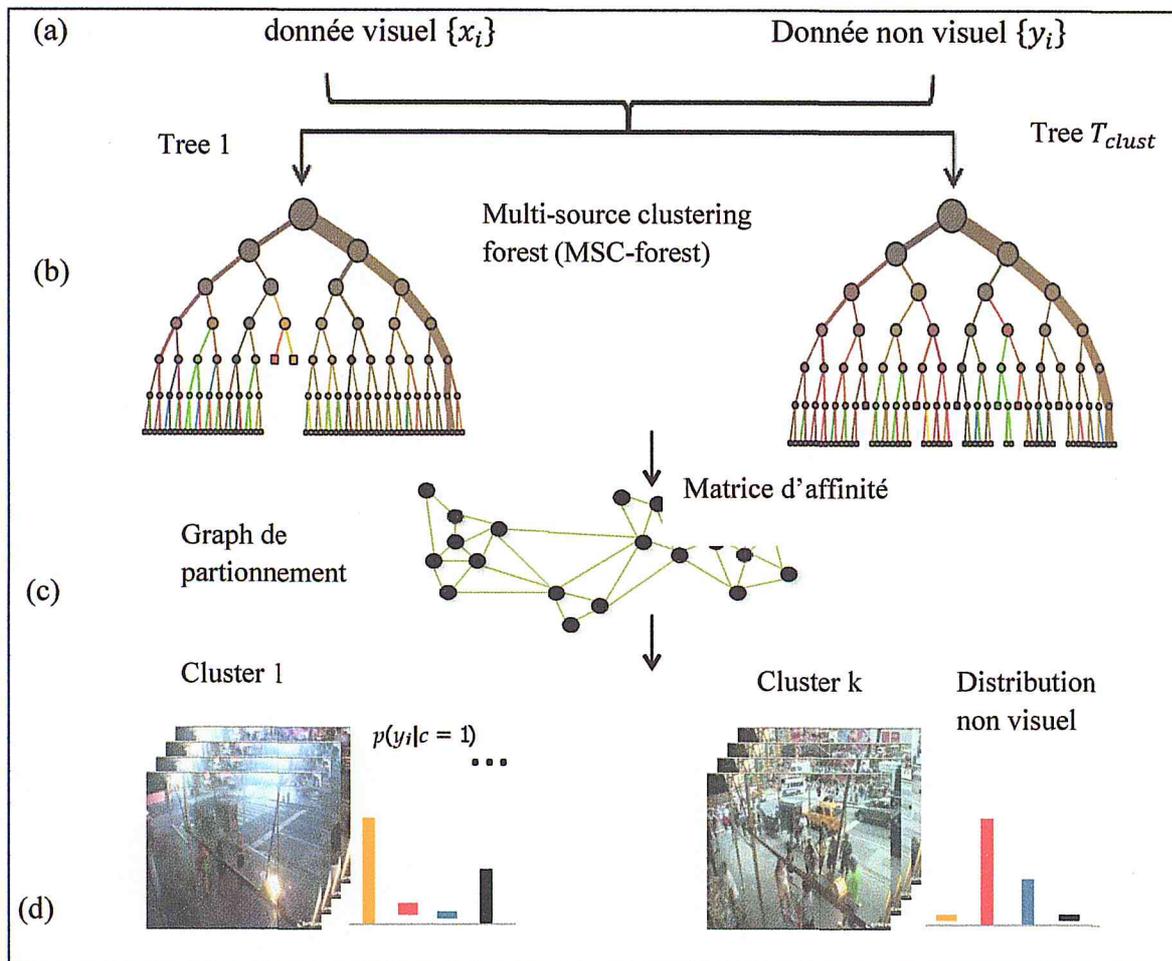


Figure 3.2: différentes étapes de la multi-source clustering avec le model multi-source clustering forest

Chapitre 3 : solution proposé et conception

Nous proposons ici un algorithme de clustering par forêts aléatoires, le Multi-Source Forêt Clustering (MSC-Forest). On y introduira une fonction objective permettant l'optimisation conjointe de gains d'information de différentes sources. Mais d'abord, nous décrivons les forêts classiques avant de détailler les MSCForest.

7. Les forêts de classification:

. Une forêt de classification, est un ensemble de T_{class} d'arbres de décision binaire $T(x) \rightarrow R^k$ Avec X est une donnée de dimension d , et $R^k = [0,1]^k$ dénotant l'espace de distribution de probabilité de classe sur l'étiquette $L = \{1, \dots, K\}$

Les arbres de décision sont tirés indépendamment les uns des autres, chacune avec un sous-ensemble aléatoire X_t de la formation d'échantillons $X = \{x_i\}$ [41].

La base du calcul repose sur l'apprentissage par arbre de décision. La proposition de Breiman[42] vise à corriger plusieurs inconvénients connus de la méthode initiale, comme la sensibilité des arbres uniques à l'ordre des prédicteurs, en calculant un ensemble de B arbres partiellement indépendants.

Une présentation rapide de la proposition [43] peut s'exprimer comme suit :

1. Créer B nouveaux ensembles d'apprentissage par un double processus d'échantillonnage :
 - (a) sur les observations, en utilisant un tirage avec remise d'un nombre N d'observations identique à celui des données d'origine (technique connue sous le nom de bootstrap),
 - (b) et sur les p prédicteurs, en n'en retenant qu'un échantillon de cardinal $m <$

p

p (la limite n'est qu'indicative).

2. Sur chaque échantillon, on entraîne un arbre de décision selon une des techniques connues, en limitant sa croissance par validation croisée.

3. On stocke les B prédictions de la variable d'intérêt pour chaque observation d'origine.

4. La prédiction de la forêt aléatoire est alors un simple vote majoritaire (*Ensemble learning*).

Le principal revers de cette méthode est que l'on perd l'aspect visuel des arbres de décision uniques.

Chapitre 3 : solution proposé et conception

8. Les Forêts de clustering :

Contrairement à la classification des forêts, les forêts de clustering ne nécessitent aucun motif d'information sur l'étiquette au cours de la phase de formation. Un clustering en forêt se compose d'arbres de décision binaire T_{clust} . Les nœuds feuilles de chaque arbre définit une répartition spatiale des données d'apprentissage. Le fait intéressant, c'est que la formation d'un clustering en forêt peut être effectuée à l'aide de l'approche d'optimisation des forêts de classification en adoptant le pseudo algorithme de classe. [44].

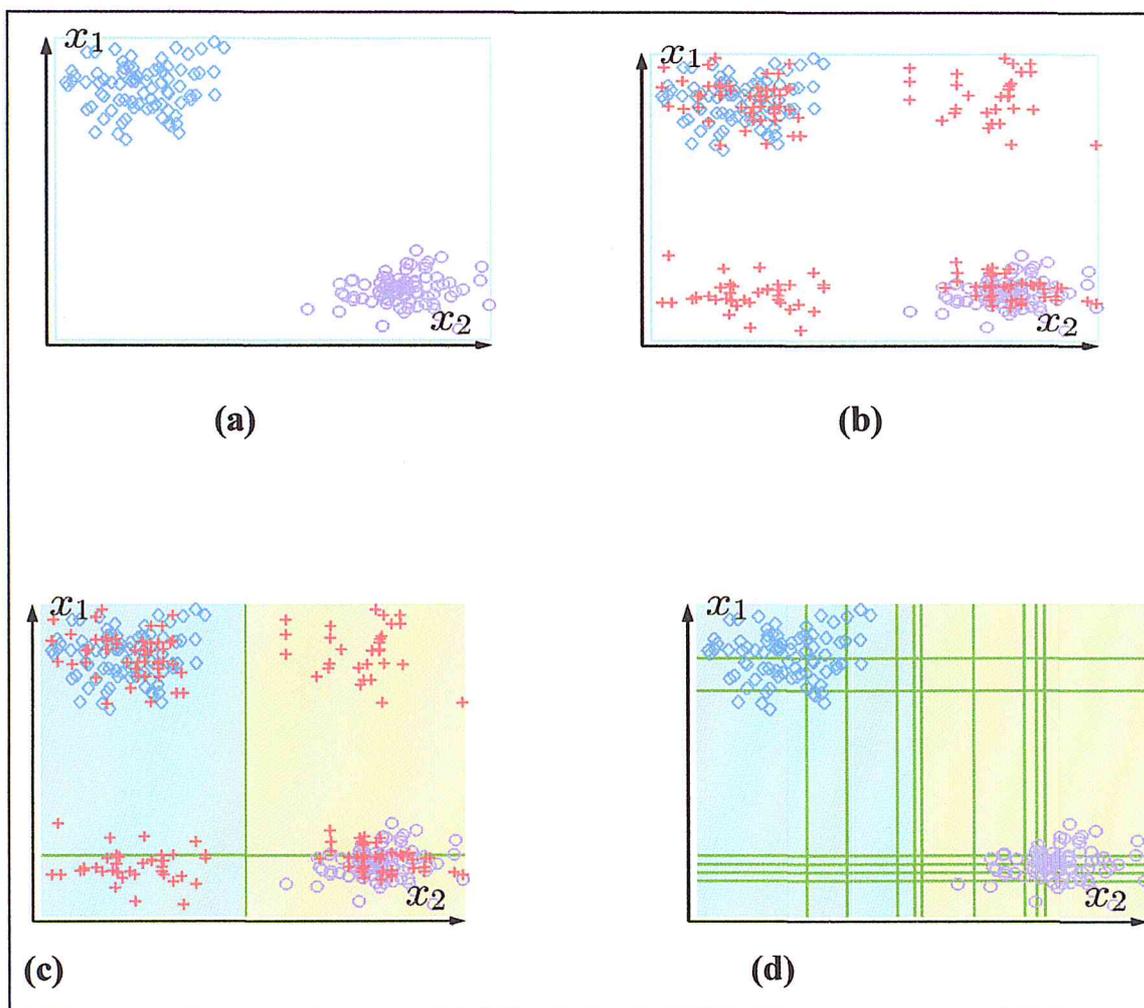


Figure 3.3: Une illustration du clustering des données avec les forêts de clustering

Chapitre 3 : solution proposé et conception

9. Multi-Source Clustering Forest:

Les forêts de clustering classiques supposent seulement les sources de données homogènes. Le modèle proposé utilise les caractéristiques visuelles comme variables de fractionnement et exploite les informations non visuelles comme des données supplémentaires pour aider la détermination :

Formellement, nous définissons une nouvelle fonction de gain d'information conjointe pour le fractionnement du nœud pendant la formation des arbres comme suite :

$$\Delta I = \underbrace{\alpha_v \frac{\Delta I_v}{I_{v0}}}_{\text{Visuel}} + \underbrace{\sum_{j=1}^m \alpha_j \frac{\Delta I_j}{I_{j0}}}_{\text{Non visuels}}$$

Cette formulation définit les meilleures données répartis sur les espaces communs de données multi-sources. Les termes de cette équation sont interprétés comme ci-dessous.

9.1 Les termes visuels :

$$\Delta I_v = \Delta I_{\text{class}}$$

Le gain d'informations visuelles est défini par ΔI_{class} , Il est estimé par le coefficient de Gini G [44] qui est calculé par $g = \sum_{i \neq j} P_i P_j$ avec P_i et P_j étant la proportion d'échantillons appartenant au i ème et j ème catégorie dans un nœud de répartition.

9.2 Les termes non visuels :

Ils sont présentés comme information secondaire pour les termes visuelles. Le ΔI_j désigne le gain d'information dans la J ème donnée non visuelle. Une source non visuelle peut être soit discrète ou continue. Pour une source non visuelle discrète, nous utilisons le coefficient de Gini G [44] comme critère de répartition des données. Dans le cas de source non visuelle avec des valeurs continues, nous adoptons la régression des moindres carrés pour faire respecter la continuité dans l'espace de clustering:

$$R = \frac{1}{|S|} \sum_{l=1}^{|S|} \left(y_{i,j} - \frac{1}{|S|} \sum_{i=1}^{|S|} y_{i,j} \right)^2$$

Chapitre 3 : solution proposé et conception

Avec Y_{ij} représente la valeur dans l'espace non visuel j ème associé à l' i ème échantillon

$X_i \in S$ Et S est l'ensemble des échantillons atteignant nœud s , c'est $\Delta J=R$

10. Comment faire face avec les données non-visuelles manquant :

Nous introduisons un mécanisme de pondération adaptative pour traiter dynamiquement avec le manque de données non visuelles. Plus précisément, lorsque certaines données non visuelles sont disparues et supposons que la proportion manquante du i ème type non visuel dans l'ensemble de la formation X_t pour MSC-tree t est δ_i , nous avons réduis son poids de a_i à $a_i - \delta_i a_i$. Le poids réduit total $\sum_i \delta_i a_i$ Est ensuite répartie uniformément au poids de toutes les sources afin de s'assurer $\alpha_v + \sum_{i=1}^m a_i + a_t = 1$. Cette méthode de pondération adaptative linéaire qui produit des résultats satisfaisants dans nos expériences.

Après avoir vu les différentes explications sur la méthode proposée, on va expliquer les différentes étapes pour extraire les informations visuelles.

11 .Découpage en frames :

La première étape de notre procédure est le découpage de la vidéo en frames, comme on l'a défini avant, la vidéo se compose de plusieurs frames qui sont les images. Plus précisément, 25 images par seconde dont on aura besoin dans notre classification. A chaque image, une donnée non visuelle lui est associée. Qui comprend la météo, l'humidité, rapport du temps et rapport de circulation qu'on récupère grâce à des objets connectés.

12. Extraction des données visuelles :

Elle consiste en des transformations mathématiques calculées sur les pixels d'une image numérique. Les caractéristiques visuelles permettent généralement de mieux rendre compte de certaines propriétés visuelles de l'image, utilisées pour des traitements ultérieurs entrant dans le cadre d'applications telles que la détection d'objets ou la recherche d'images par le contenu. Dans cette partie nous allons extraire les données des images grâce à des descripteurs globaux, qui représentent l'image dans son ensemble, nous nous intéressons à deux descripteurs qui nous permettent d'avoir les caractéristiques visuelles de toute l'image sous forme d'un vecteur de taille fixe.

Chapitre 3 : solution proposé et conception

13. Choix du descripteur dans nos travaux :

Pour nos travaux, nous avons choisi d'employer les descripteurs HOG et histogramme de couleurs RGB. En effet, ces derniers fournissent de bons résultats de classification tout en étant compatibles avec le temps réel et plus rapides à calculer que les matrices de covariances.

13.1. Histogramme de couleur RGB :

La couleur est un des attributs primaires d'images. Rappelons qu'un pixel d'une image couleur est caractérisé par un vecteur à trois composantes, quel que soit le système colorimétrique utilisé. Les couleurs présentes dans une image sont décomptées dans l'histogramme des couleurs. Cet histogramme renferme une masse énorme d'informations.

L'histogramme d'une image couleur de résolution $M \times N$ est la représentation d'une fonction discrète $v(C)$ qui à chaque couleur C présente dans l'image, associe le nombre $v(C)$ de pixels ayant cette couleur, encore appelé population de couleur C . cette fonction est donc une application de l'ensemble R^3 des couleurs dans l'ensemble N des entier naturels [45]:

$$v : \begin{cases} R^3 \rightarrow N \\ C \rightarrow v(C) \end{cases}$$

Le quotient de $v(C)$ par le nombre total MN de pixels de l'image peut être vu comme la densité de probabilité $k(C)$ d'une variable aléatoire dont l'image constitue un ensemble de réalisation :

$$k(C) = \frac{v(C)}{MN} \quad \text{avec} \quad \sum_c k(C) = 1$$

La couleur moyenne de l'image est donnée par :

$$\bar{C} = \sum_c C \cdot k(C)$$

En règle générale, chaque composante colorimétrique est codée sur 8bits. L'histogramme d'une image de résolution $M \times N$ est alors constitué par un tableau tridimensionnel comportant 2^{24} cellules. Chaque cellule de ce tableau doit pouvoir contenir un nombre entier positif ou nul au maximum égal à MN , donc codé sur un nombre n_b de bits tel que $n_b > \log_2(MN)$ En effet, n bits ne permettent de compter que jusqu'à $2^n - 1$, et un nombre de bits est toujours entier, de plus, le codage des nombres en informatique fait en général appel à des multiples de 8 bits. L'histogramme occupe alors $2^{24} \cdot n_b$ bits. Pour une image dont la

Chapitre 3 : solution proposé et conception

résolution est de 512×512 pixels, $n_b = 24$ bits et l'histogramme occupe un espace mémoire de $E = 3 \times 2^{24}$ octets, soit environ 50 Mo.

Réaliser un système de recherche d'images fondé sur l'analyse des couleurs nécessite donc une réduction drastique de l'espace mémoire utilisé pour modéliser l'histogramme. Trois procédures peuvent être envisagées :

- la réduction du nombre de bits utilisé pour quantifier l'espace colorimétrique ;
- l'utilisation d'un histogramme compact ;
- la modélisation de l'histogramme par les premiers moments de la distribution des couleurs.

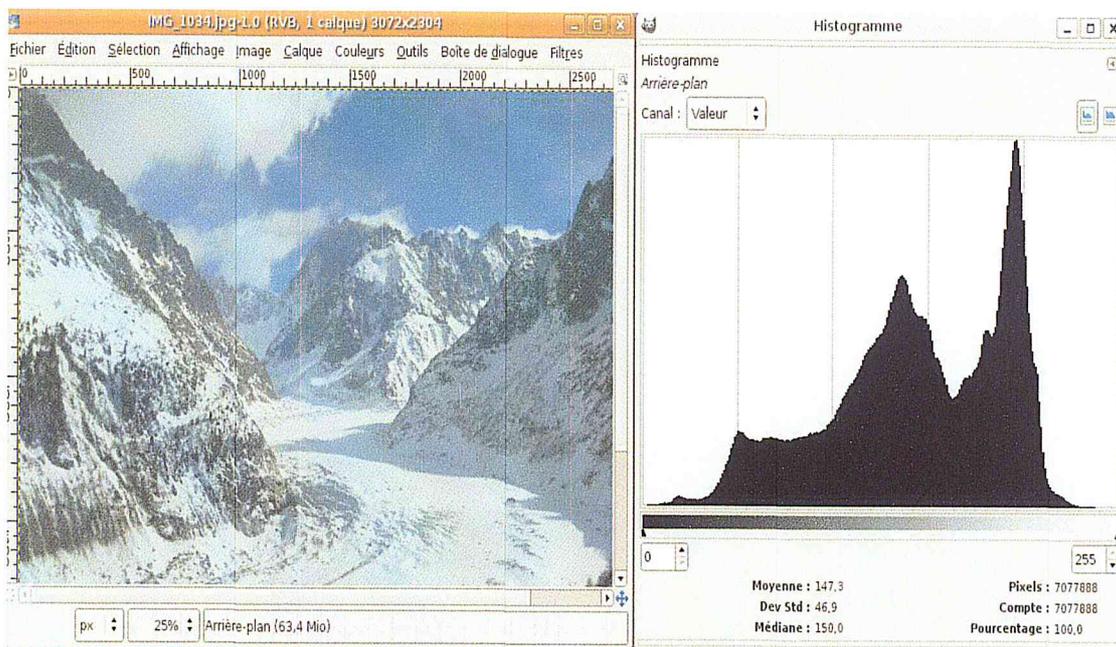


Figure 3.4 : image et son histogramme de couleur

Chapitre 3 : solution proposé et conception

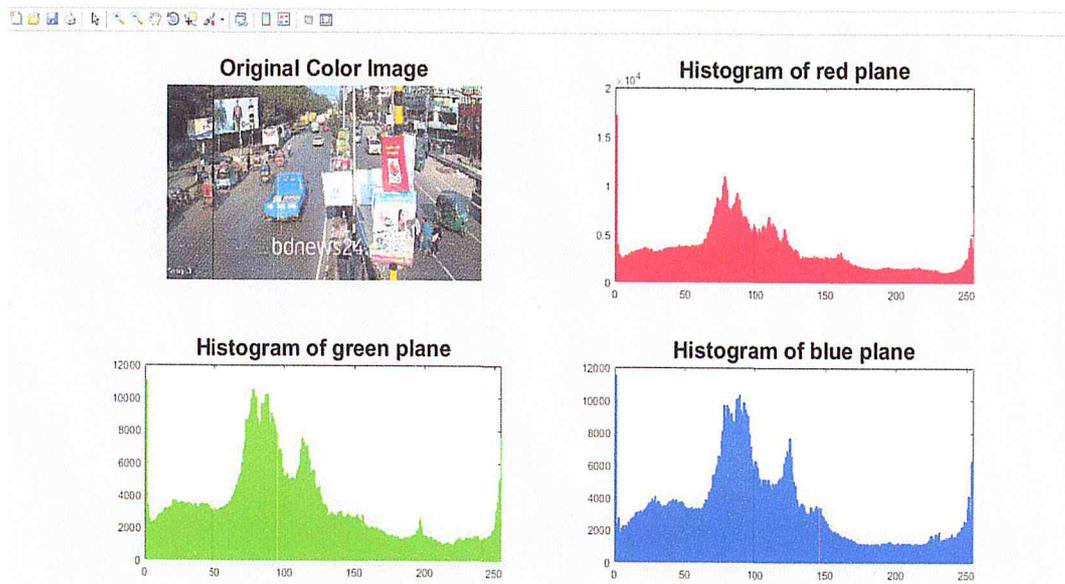


Figure 3.5: Exemple d'image et son histogramme RGB

13.2 Histogramme de gradient orienté (HOG) :

Les histogrammes de gradients orientés (*Histograms of Oriented Gradients* ou HOG) et leurs variantes sont sans conteste les descripteurs les plus populaires dans le domaine de la détection d'objet et extraction des caractéristiques des images.

Ils ont été introduits par Dalal et Triggs en 2005 et s'inspirent fortement des travaux précédents de Lowe, 2004 sur le descripteur SIFT (*Scale-invariant feature transform*).

L'idée mise en œuvre ici est de discrétiser l'orientation du gradient dans l'image à l'aide d'histogrammes. Ainsi les contours, donc la forme des objets, sont codés ce qui permet de les reconnaître [46].

D'un point de vue pratique, le calcul d'un HOG commence par l'évaluation des gradients horizontaux G^x et verticaux G^y pour tous les pixels de l'image. Ce filtrage est réalisé à l'aide de deux convolutions effectuées séparément. La première utilise le masque $[1, -1]$ pour les gradients horizontaux et la seconde emploie le masque $[-1, 1]^T$ pour les gradients verticaux. Ainsi à chaque point (i, j) de l'image est associé un couple $(G_{i,j}^x, G_{i,j}^y)$

Après le calcul des gradients l'étape suivante est la construction de l'histogramme de l'orientation des gradients. L'image est ensuite décomposée en blocs, généralement de 8 pixels de côté, et un histogramme est construit pour chacun d'entre eux. Les classes des histogrammes représentent l'angle du gradient. Leur nombre est variable suivant les

Chapitre 3 : solution proposé et conception

implémentations. Un pixel d'un bloc vote donc pour une classe en fonction de l'angle du gradient en ce point et son importance dans l'histogramme est la norme du gradient.

La dernière étape est la normalisation des histogrammes entre les blocs. Les histogrammes de blocs adjacents (par exemple des regroupements 4x4) sont concaténés pour former un vecteur qui est ensuite normalisé pour former une partie du descripteur. Le descripteur complet est obtenu en répétant l'opération sur tous les groupements de blocs possibles.

Après l'extraction des caractéristiques des images à l'aide des descripteurs qui sont représentés par des matrices, on considère ces images comme la réunion de n classes à un seul élément et on regroupe progressivement les classes deux à deux selon l'algorithme suivant :

La première étape est de calculer la matrice des distances $D = (d(M_i, M_j)_{0 \leq i \leq n, 0 \leq j \leq n})$, dans notre travail on a utilisé deux méthodes de calcul de mesure de similarité entre les images qui sont :

13.2.1 Distance euclidienne :

La distance la plus connue est la distance Euclidienne, qui définit l'espace cartésien.

$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$ (Pour des vecteurs de dimension n). La distance Euclidienne est souvent utilisée au carré [47].

13.2.2 Distance de Mahalanobis :

La distance de Mahalanobis permet de calculer la distance entre deux points dans un espace à p dimensions, en tenant compte de la structure de variance-covariance sur ces p dimensions. Le carré de cette distance se définit par :

$$dM^2 = (x1 - x2) \Sigma^{-1} (x1 - x2)$$

Si les éléments de x et de y sont indépendants, la matrice de Covariance sera l'identité et la distance de Mahalanobis serait égale à la distance Euclidienne [48].

La seconde étape est d'itérer les deux étapes suivantes jusqu'à l'agrégation en une seule classe :

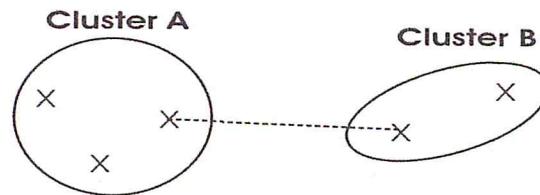
- Regrouper les deux éléments (classe) les plus proches au sens de la distance entre groupes choisis.

Chapitre 3 : solution proposé et conception

- Mettre à jour le tableau de distance en remplaçant les deux classes regroupé par la nouvelle et en calculent sa distance avec chacune des autre classe.

La distance entre la classe est calculée par la stratégie du saut minimum (single linkage) qui est la plus petite distance entre les éléments des deux parties :

$$D(I_1, I_2) = \min_{i \in I_1, j \in I_2} d(i, j)$$



Après la classification des données visuelles et les donnée non visuelles on doit rassembler les résultats obtenus, pour cela on à généré un algorithme qui permet de rassembler les résultats des donnés visuelles qui sont les classes de chaque image avec ces donnés non visuelles pour qu'on puisse utiliser la classification Forest sous java.

14. Conclusion :

La conception de la solution est une étape primordiale est importante, elle prépare a la phase de réalisation. Dans ce chapitre on a montré notre solution proposée avec une description textuelle bien détaillée .le chapitre suivant porte sur la phase d'implémentation, celle qui donne lieu au système conçu précédemment.

Chapitre 4 :

Teste et résultats

Chapitre 4 : Teste et résultats

1. Introduction :

Dans les chapitres précédent on à donné quelque notion de classification, de résumé vidéo et la solution proposé pour exploiter les données visuelles et les données non visuelles dans le contexte de résumé vidéo. Dans ce chapitre consacré a l'implémentation, nous allons présenter les différentes étapes utilisées ainsi que les outils qui nous ont servis pour ce travail et les résultats obtenus de notre approche.

2. Outils de développement :

2.1. Environnement matériel :

L'implémentation à été réalisé sur deux ordinateurs portable:

- DELL inspiron 3520, windows 10 32bits, Processeur Intel (R) Core™ i3-3110M CPU @ 2.40 GHZ, 4 Go de Ram.
- TOSHIBA Satellite C660, windows 10 32bits, Processeur Intel (R) Core™ i3 CPU M380 @ 2.53 GHZ, 4 Go de Ram.

2.2. Environnement logiciel :

Les algorithmes ont été développée à l'aide de l'outil de programmation Matlab R2015b de l'éditeur Mathworks , il est utilisé à des fins de calcul numérique, MATLAB permet de manipuler des matrices, d'afficher des courbes et des données, de mettre en œuvre des algorithmes, de créer des interfaces utilisateurs, et peut s'interfacer avec d'autres langages comme le C, C++, Java, et Fortran. Les utilisateurs de MATLAB sont de milieux très différents comme l'ingénierie, les sciences et l'économie dans un contexte aussi bien industriel que pour la recherche. La compilation sous *Matlab* se fait en appelant la fonction « Deploytool » qui utilise le fichier *.m* (code source) et le fichier *.fig* (spécification de l'interface utilisateur GUI) pour ainsi générer le fichier exécutable.

L'algorithme des forêts aléatoire est implémenté sous java grâce a l'outil weka, ce dernier est un logiciel de data mining en Java.

Weka est une collection d'algorithmes d'apprentissage machine pour les tâches d'extraction de données. Les algorithmes peuvent être appliquées directement à un dataset ou appelés à partir un code Java. Weka contient des outils pour le prétraitement des données, classification,

Chapitre 4 : Teste et résultats

régression, clustering, règles d'association, et de visualisation. Il est également bien adapté pour l'élaboration de nouveaux régimes d'apprentissage machine.

3. Présentation des benchmark utilisé :

Nous avons mené des expériences sur deux ensembles de données collectées à partir de webcams accessibles au public qui comportent un en plein air et une scène intérieure, respectivement:

1/ les Times Square Intersection (TISI) dataset, et 2/ l'éducation Centre de ressource (ERCE) dataset. Il y a un total de 7324 clips vidéo couvrant plus de 14 jours dans le jeu de données(TISI), tandis qu'un total de 13817 extraits ont été recueillis sur une période de deux mois dans l'ensemble de données (ERCE). Chaque clip a une durée de 20 secondes.

L'ensemble de données TISI est difficile en raison des comportements complexes, et une grande variation d'illumination due à la fois par les sources d'éclairage naturel et artificiel à différents temps de la journée. L'ensemble de données ERCE n'est pas facile aussi en raison d'événements physiques qui sont impliqués et caractérisé par de grands changements dans la configuration de l'environnement, les participants, et les modèles d'activité complexe

Les sources visuelles et non visuelles que nous avons extrait de ce qui suit l'ensemble de caractéristiques visuelles pour représenter le contenu visuel dans chaque clip: (a) les caractéristiques de couleur, y compris RGB et HSV; (b) flux optique; (c) des caractéristiques globales de la scène sur la base de GIST, et (d) la personne et détection des véhicules [49].

Chapitre 4 : Teste et résultats

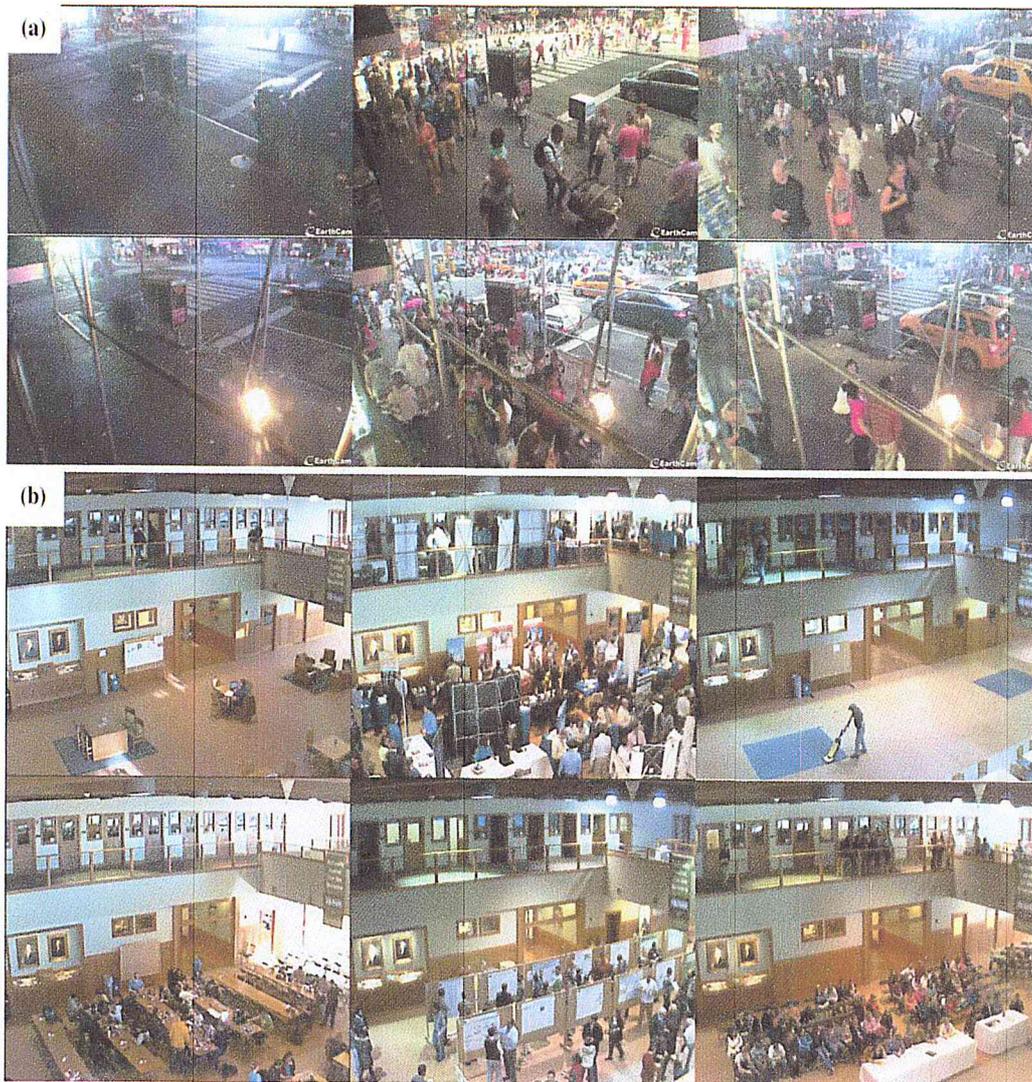


Figure 4.1: exemple sur l'ensemble des données TISI et ERCE

Nous avons recueilli 10 types de sources non-visuelles pour l'ensemble de données TISI: (a) les données météorologiques extraite du World-Weather Online avec 9 éléments: température, type de temps, la vitesse du vent, direction du vent, les précipitations, l'humidité, la visibilité, la pression, et la couverture nuageuse; (B) les données de vitesse de circulation avec 4 niveaux de vitesse de la circulation: très lent, lente, modérée et rapide. Pour l'ensemble de données ERCE, nous avons recueilli les données provenant de multiples sources en ligne indépendantes. Calendrier des événements sur le campus, y compris: aucun événement (NoEvt), Nettoyage (CLN), Salon de l'emploi, Groupe étudiant (GrStd), Concours de bourses d'études (SchlCpt), service accommodantes(AcmSvc), Orientation des élèves (StdOrt).

Chapitre 4 : Teste et résultats

4. Résultats et interprétations :

Afin d'évaluer les performances de la méthode proposée, nous l'avons implémenté puis nous l'avons testé sur une sélection de vidéos de la base de données. Nous avons utilisé deux mesures distinctes pour évaluer cette technique qui sont : la F-mesure et l'entropie. Les résultats obtenus sont présentés dans le tableau *Table4.1* :

4.1 F-mesure :

C'est la Moyenne harmonique de la précision et du rappel. Elle mesure la capacité de la méthode à donner toutes les solutions pertinentes et à refuser les autres. Elle se calcule comme suit :

$$F - mesure = \frac{2Précision \times Rocall}{Précision + Rocall}$$

Elle est également connue sous le nom de mesure, car la précision et le rappel sont Pondérés de façon égale et sa valeur varie entre 0 et 1, plus sa valeur se rapproche de 1 et mieux.

4.2 L'entropie :

C'est une Grandeur caractérisant le désordre d'un système, on calculant l'entropie on peut s'avoir si les données sont bien classer dans le cluster ou non.

L'entropie peut être soit positive ou nulle et tant que sa valeur est petite les données sont bien classer.

Chapitre 4 : Teste et résultats

Vidéo	Nombre d'image	F-mesure	L'entropie
V 1	1680	0.887	0.456
V2	2325	0.755	0.754
V3	1959	0.804	0.575
V4	1879	0.725	0.653
V5	2322	0.684	0.971
V6	1856	0.603	1.123
V7	1845	0.728	0.847
V8	2078	0.847	0.604
V9	1984	0.708	0.745
V10	2271	0.643	1.032

Table4.1. Les mesures d'évaluation de l'approche proposée

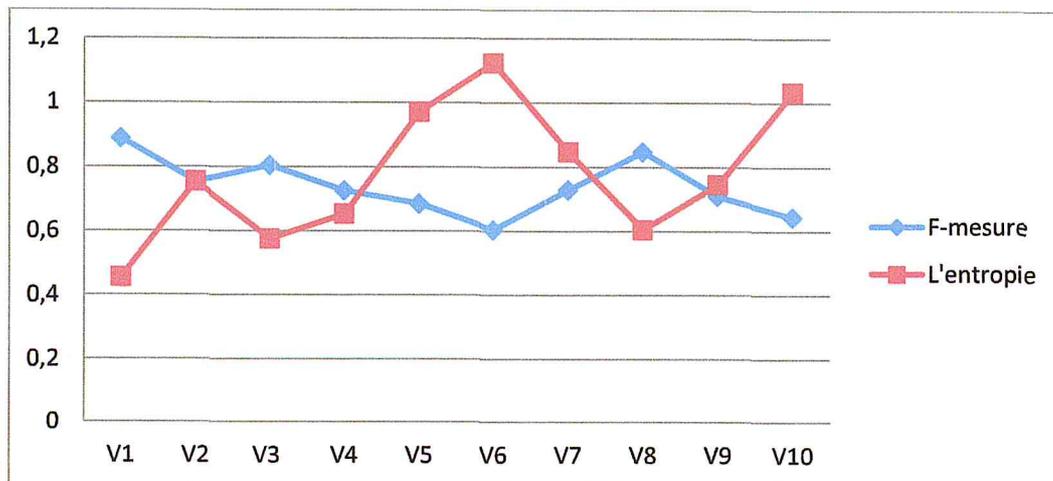


Figure4.2. Les valeurs de F-mesures et l'entropie pour l'approche proposée

Les résultats obtenus dans le tableau et la figure montrent que la valeur maximale de F-mesure est égale à 0.887 et sa valeur moyenne est de 0.603 et aussi l'entropie ou sa valeur maximale est égale à 1.123 et sa valeur moyenne est égale à 0.456. Ces moyennes sont acceptables mais comme nous ne pouvons pas nous contenter de ces résultats de notre projet

Chapitre 4 : Teste et résultats

de fin d'étude, nous avons essayé de les comparés avec les mesures obtenues dans des travaux précédents.

5. Etude comparative :

Après avoir testé l'approche proposé on à fait une étude comparative par rapport à des travaux antérieures dans le même domaine les résultats obtenu sont représenté dans le tableau suivant :

dataset	TISI		ERCe
	Vitesse de circulation	Le climat	Evénement
VO-Forest [50]	0.8675	1.0676	0.0616
VNV-Kmeans [51]	0.9197	1.4994	1.2519
VNV-AASC [52]	0.7217	0.7039	0.0691
MSC-Forest	0.702	0.6071	0.0024

Table 4.2: comparaison des résultats obtenue avec les travaux antérieure

Les résultats obtenus dans le tableau montrent que par rapport aux 3 travaux antérieurs les valeurs de l'entropie de notre approche qui est le MSC-Forest sont les plus petit et le reste des valeurs sont tous supérieur, nous remarquons qu'il ya une nette amélioration par rapport aux autres méthodes d'apprentissage, on déduit alors que notre approche est plus performante en matière d'apprentissage.

6. Conclusion :

Dans ce chapitre nous avons présenté les outils nécessaires pour développer notre approche d'apprentissage multi-source qui permet d'exploiter les données visuelles et les données non visuelles dans le résumé vidéo, et les différents tests utilisés pour vérifier les résultats obtenus. On peut conclure que la robustesse de notre approche à été démontrée par des mesures statistiques et que notre approche répond aux besoins spécifiés dans le chapitre précédent.

Conclusion Générale

Durant ce mémoire, Nous avons présenté une nouvelle approche d'apprentissage multi-source dans le résumé vidéo qui permet d'exploiter à la fois les données visuelles et les données non visuelles à l'analyse de contenu et de résumé vidéo.

L'achèvement de notre mémoire nous a permis d'avoir des notions avancées sur les approches d'apprentissages et approfondir nos connaissances dans le domaine des traitements d'images ainsi sur les outils de développement. Au fait ce projet nous a permis de s'intégrer dans la recherche scientifique, de plonger dans la pratique et d'activer nos connaissances acquises durant nos cinq années universitaires.

Le modèle que nous avons présenté effectue le clustering de données multi-source sur des données hétérogènes et découvre un ensemble de clusters de données visuelles marqués avec la distribution de données non visuelles. Ceci est rendu possible par la formulation d'une forêt de clustering (MSCForest Multi-Source).

La solution que nous avons présentée et le projet que nous avons réalisé dans ce mémoire n'est en réalité qu'une ouverture vers d'autres travaux car notre approche peut encore évoluer et se voir améliorer, dans un court terme l'amélioration va concerner l'extraction des images clés pour la génération de résumé vidéo finale et aussi réduire le temps d'exécution qui n'est pas négligeable à l'heure actuelle.

Par la suite on envisage d'élargir notre champ de source de données vidéo en faisant la relation et la combinaison des données vidéo de plusieurs angles pour arriver à un meilleur résumé et une meilleure sémantique.

Référence bibliographique :

- [1] Alan Hanjalic , HongJiang Zhang. An Integrated Scheme for Automated Video Abstraction based on Unsupervised Cluster-Validity Analysis, 2-30
- [2]Dynamic video summarization of home video, Proc. SPIE 3972, Storage and Retrieval for Media Databases 2000, 378 (December 23, 1999)
- [3] Digital learning center, 25 février 2016, shot definition, http://www.homevideomaking.com/Lessons/beginner_lessonpages/shotdefinition.html
- [4]Muhammad Ajmal, Muhammad Husnain Ashraf, Muhammad Shakir, Yasir Abba, Faiz Ali Shah. Computer Vision and Graphics. Video Summarization: Techniques and Classification.vol 7594. p 1-13
- [5] Ajay Divakaran, Kadir A. Peker, Huifang Sun . Method for summarizing a video using motion and color descriptors. 24 févr. 2004
- [6] Uros Damjanovic, Tomas Piatrik, Divna Djordjevic, and Ebroul Izquierdo . Video Summarisation for Surveillance and News Domain. University of London.[1-14]
- [7] Tommy Chheng . Video Summarization Using Clustering. University of California, Irvine.p1-7
- [8] RaviKansagara, DarshakThakore, MahaswetaJoshi. A Study on Video Summarization Techniques. International Journal of Innovative Research in Computer and Communication Engineering. Vol. 2, février 2014, p1-8
- [9] RaviKansagara¹, DarshakThakore², MahaswetaJoshi³. A Study on Video Summarization Techniques. International Journal of Innovative Research in Computer and Communication Engineering. Vol. 2, février 2014, p2964-2965.
- [10] H. Feigenbaum, Echocardiography, LEA & FEBIGER, 1993
- [11] D. A. Sadlier , N. E. O'Connor , Event detection in field sports video using audio-visual features and a support vector Machine , IEEE, 26 septembre 2005, 1225 – 1233

[12] Yael Pritch , Sarit Ratovitch , Avishai Hendel , Shmuel Peleg, Clustered Synopsis of Surveillance Video, IEEE, 195 – 200, 2-4 Sept. 2009

[13] Marc Jamouille, Michel Roland , Jacques Humbert, Jean-François Brûlet. Traitement de l'information médicale par la Classification internationale des soins primaires, deuxième version : CISP-2. Care Edition, Bruxelles, 2000

[14] W.B. Michael and Malu Castellanos, survey of text mining clustering, classification and retrieval Survey of Text Mining : Clustering, Classification, and Retrieval , Second Edition, Springer, pp 3-22, 2007.

[15], Richard O. Duda, Peter E. Hart, David G. Stork, Pattern classification, Wiley-interscience, 2001.

[16] Gérard. Govaert, Analyse des données. IC2(série Traitement du signal et de l'image), Lavoisier.2003

[17]. Berrani, S.-A., Amsaleg, L., & Gros, P. Recherche par similarités dans les bases de données multidimensionnelles : panorama des techniques d'indexation. Ingénierie des systèmes d'information (RSTI série ISI-NIS), 7(5-6), pp 65-90.2002.

[18] Saporta G. , probabilités, analyse des données et statistiques. Technip , paris, 1990.

[19], Lebart L., Morineau A. & piron M.statistique exploratoire multidimensionnelle. Dunod, 3ème édition, paris, 2000.

[20] Kohonen T. self-organized formation of topologically correct feature maps. Biological cybernetics no 43, pp59-69, reprinted in Anderson & Rosenfeld , Eds, Neurocomuting : foundations of research, MIT press, Cambridge Ma, 1988.

[21]. J. Hartigans. clustering algorithms. John Wiley and Sons, Inc., 1975.

[22], J. B. MacQueen (1967) : "Some Methods for classification and Analysis of Multivariate Observations, Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability", Berkeley, University of California Press,no 1, pp281-297.1967.

[23] :Benzécri J.P. L'analyse des données. Dunod, Paris, 197.

[24], Celeux, G., Diday, E., Govaert, G., Lechevallier, Y., and Ralambondrainy, H. Classification automatique des données, environnement statistique et informatique. DUNOD informatique. 1989.

[25]: Ball, G. H. et Hall, D. J. ISODATA, an Iterative Method of Multivariate Analysis and Pattern Recognition. Behavior Science, 153, 1967.

[26]: J. C. Dunn (1973) : "A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters", Journal of Cybernetics ,no 3, pp 32-57. 1973.

[27]: J. C. Bezdek (1981) : "Pattern Recognition with Fuzzy Objective Function Algorithms", Plenum Press, New York. 1981.

[28] Mathematical Morphology and Image Analysis : A Fuzzy Approach ».Workshop on Knowledge-Based Systems and Models of Logical Reasoning, Reasoning, 1988.

[29] Fairouz Hadi , Khier Benmahammed , Etude comparative entre la morphologie mathématique floue et le regroupement flou , Faculté des Sciences de l'Ingénieur, Université Ferhat Abbas-Sétif, Algérie., 3rd International Conference : SETIT 2005

[30] F. H'oppner, F. Klawonn, R. Kruse, T. Runkler. Fuzzy Cluster Analysis, Methods for classification, data analysis and image recognition. Wiley, 2000.

[31] S. C. Johnson : "Hierarchical Clustering Schemes" Psychometrika, no 2, pp 241-254, 1967.

[32], Lance, G.N., & Williams, W.T. : A general theory of classificatory sorting strategies : I. Hierarchical systems. Computer Journal, no 9,pp 373-380, 1967.

[33] Kamvar, S. D., Klein, D., & Manning, C. D., Interpreting and Extending Classical Agglomerative Clustering Algorithms using a Model-Based approach. Pp 283-290 of : International Conference on Machine Learning (ICML).2002.

[34] et Guha, S., Rastogi, R., et Shim, K. CURE : an e_icient clustering algorithm for large databases. Dans Proceedings of ACM SIGMOD International Conference on Management of Data, pp 73-84, 1998.

- [35]. Karypis, G., Eui-Hong, H., et Kumar, V. Chameleon : Hierarchical Clustering Using Dynamic Modeling. *Computer*, no 32(8) :68-75, 1999.
- [36]. Zhang, T., Ramakrishnan, R., et Livny, M. BIRCH : an efficient data clustering method for very large databases. Dans *Proceedings of the 1996 ACM SIGMOD International Conference on Management of Data*, pp 103-114, 1996.
- [37], Bisson, G, La similarité : une notion symbolique/numérique. Chap. XX of : *Apprentissage symbolique-numérique (tome 2)*. Editions CEPADUES.2002.
- [38]. Berrani, S.-A., Amsaleg, L., & Gros, P. Recherche par similarités dans les bases de données multidimensionnelles : panorama des techniques d'indexation. *Ingénierie des systèmes d'information (RSTI série ISI-NIS)*, 7(5-6), pp 65-90.2002.
- [39] X. Zhu, C. C. Loy, and S. Gong, "Constructing robust affinity graphs for spectral clustering," in *Proceedings of the 27th IEEE Conference on Computer Vision and Pattern Recognition*, Jun. 2014, pp. 1450–1457.
- [40] B. T. Truong and S. Venkatesh, "Video abstraction: A systematic review and classification," *ACM Transactions on Multimedia Computing, Communications, and Applications*, 2007.
- [41] L. Breiman, "Random forests," *Machine Learning*, 2001.
- [42] (en) Leo Breiman, « Random Forests », *Machine Learning*, vol. 45, 2001, p. 5-32.
- [43] (fr) Pirmin Lemberger, Marc Batty, Médéric Morel et Jean-Luc Raffaëlli, *Big Data et Machine Learning*, Dunod, 2015, pp 130-131.
- [44] T. Shi and S. Horvath, "Unsupervised learning with random forest predictors," *Journal of Computational and Graphical Statistics*, pp. 118–138, 2006
- [45] Dawei LIU, *Systèmes Dynamiques et Signaux*, Rapport de stage du master 2 recherche, Institut des Sciences et Techniques de l'Ingénieur d'Angers, février au 30 juin 2007
- [46] Thierry Chesnais , *Contextualisation d'un détecteur de piétons , Application à la surveillance d'espaces publics*, Université Blaise Pascal - Clermont II, 24 juin 2013

[47] Barile, Margherita. "Euclidean." From *MathWorld*--A Wolfram Web Resource, created by Eric W. Weisstein. <http://mathworld.wolfram.com/Euclidean.html>

[48] <https://www.xlstat.com/fr/solutions/fonctionnalites/tests-multidimensionnels>.

[49] www.eecs.qmul.ac.uk/%7Exz303/download.html.

[50] Breiman, L. (2001). Random forests. *Machine Learning*, 45, 5.

[51] Jain, A. K. (2010). Data clustering: 50 years beyond k-means. *Pattern Recognition Letters*, 31(8), 651–666.

[52] Huang, H. C., Chuang, Y. Y., & Chen, C. S. (2012). Affinity aggregation for spectral clustering. In *IEEE conference on computer vision and pattern recognition*.