

Emotion Estimation in Speech Using a 3D Emotion Space Concept

Par : Michael Grimm and Kristian Kroschel, Universität Karlsruhe (TH), Germany

Présenté par :

D. Lahbib¹ A. Mohamed¹

¹Université Paris Sud XI

Proposé par : Laurence Devillers

Enseignant : Emmanuel Waller

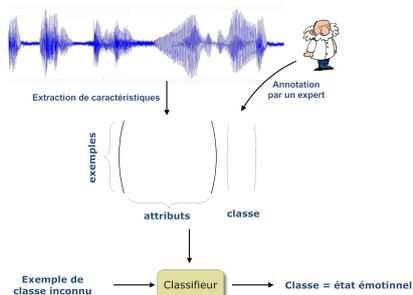
Contexte Général

La reconnaissance automatique d'émotions contenues dans la parole...

- Objectif
 - Reconnaître l'état émotionnel d'un locuteur étant donné son discours vocal
- Applications
 - l'amélioration des interfaces homme-machine
 - robots assistants
 - ordinateur tuteurs
 - reconnaissance automatique de la parole

Contexte Général

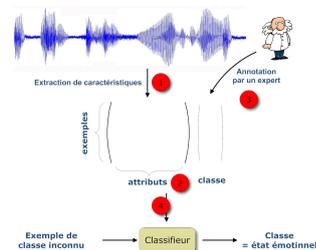
Apprentissage supervisé



Contexte Général

Problématiques

1. quels attributs ?
2. Combien d'attributs ?
3. Evaluer l'émotion ?
4. comment l'estimer ?



Plan

1. Contexte Général
2. Pré-traitement des Données
3. Estimation des Emotions
4. Résultats
5. Conclusion et Perspectives

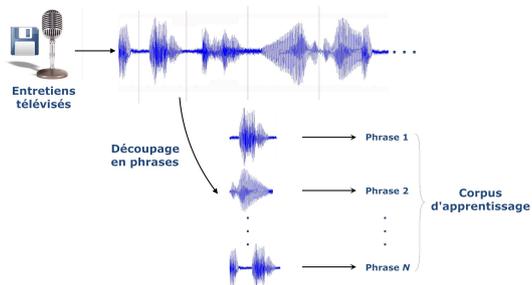
Construction du corpus d'apprentissage
Réduction de dimensionnalité

Plan

1. Contexte Général
2. Pré-traitement des Données
 - Construction du corpus d'apprentissage
 - Réduction de dimensionnalité
3. Estimation des Emotions
 - SVM
 - k-Plus Proches Voisins Flous
 - Système d'Inférence Floue
4. Résultats
5. Conclusion et Perspectives

Corpus d'apprentissage

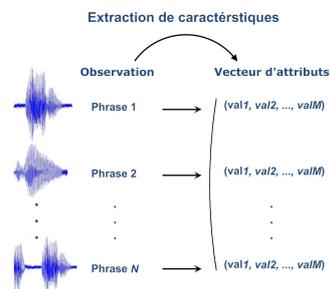
Découpage en phrase du signal vocal



Construction du corpus d'apprentissage
Réduction de dimensionnalité

Corpus d'apprentissage

Extraction des caractéristiques



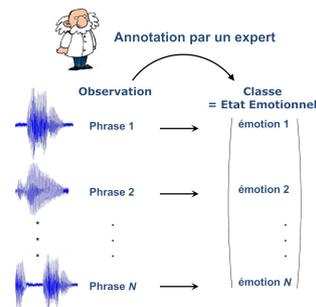
Corpus d'apprentissage

Extraction des caractéristiques

- 137 attributs acoustiques normalisés dans l'intervalle [0, 1] :
 - 53 dérivant du ton et de l'énergie
 - 6 caractéristiques temporelles
 - 78 attributs décrivant la variabilité spectrale
- par exemple : la valeur moyenne, l'écart type, durée moyenne du discours ...

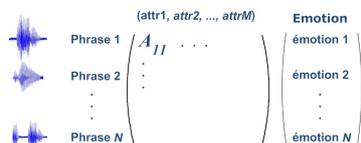
Corpus d'apprentissage

Annotation des Observations



Corpus d'apprentissage

Représentation classique de la base d'apprentissage

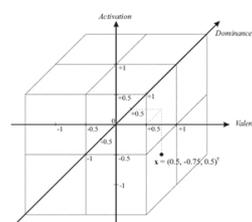


Format exploitable pour l'apprentissage ...

Corpus d'apprentissage

Comment décrire les états émotionnels

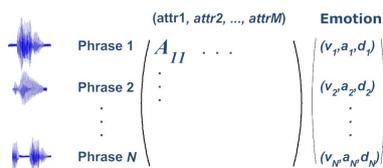
L'émotion est un triplet à valeurs continues (valence, activation, dominance)
Un dans point dans un espace 3D.
Exemple la colère :



- une valence négative,
- un niveau élevé d'excitation
- et une forte dominance.

Corpus d'apprentissage

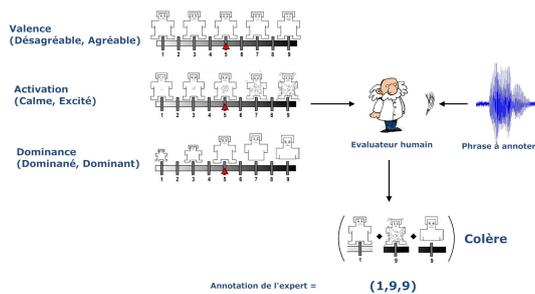
Base d'apprentissage



Problème :
Comment attribuer les valeurs de valence, de l'activation et de la dominance ?

Annotation des émotions

par Self-Assessment Manikins



Plan

- Contexte Général
- Pré-traitement des Données
 - Construction du corpus d'apprentissage
 - Réduction de dimensionnalité
- Estimation des Emotions
 - SVMR
 - k-Plus Proches Voisins Floues
 - Système d'Inférence Floue
- Résultats
- Conclusion et Perspectives

Réduction de dimensionnalité

Approches

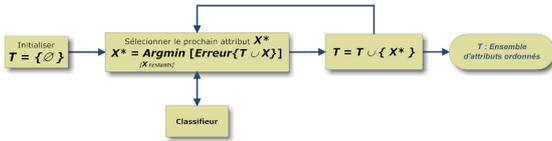


2 approches proposées :

- Sélection Séquentielle Ascendante (Sequential Forward Selection)
- Analyse en Composante Principale ACP

Sélection Séquentielle Ascendante

Etape 1 : Ordonner les attributs

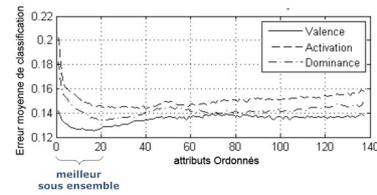


- **Principe** : Ordonner une à une les variables tant que l'ajout minimise l'erreur.
- On fait appel alors à l'algorithme d'apprentissage
- Classifieur utilisé : SVM pour la régression

Sélection Séquentielle Ascendante

Etape 2 : choisir les premiers meilleurs attributs

Choisir les premiers meilleurs attributs qui minimisent l'erreur



Analyse en Composante Principale

ACP

ACP : Principe

Les meilleurs attributs correspondent aux vecteurs propres de la matrice de covariance dont les valeurs propre associées représentent 90% de la somme de toutes les valeurs propres.

L'ACP n'utilise pas les résultats de classification comme feedback : elle se base seulement sur le nombre d'observations et le nombre d'attributs

Réduction de dimensionnalité

Résultat

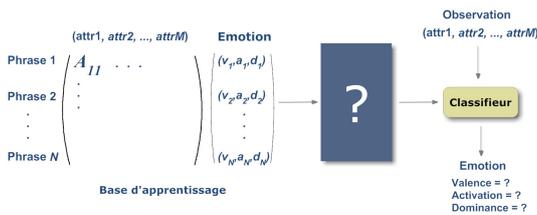
- ACP : pour inclure 90% de la somme totale des valeurs propres 61 composantes doivent être incluses.
- 20 attributs sont suffisants pour avoir un taux d'erreur acceptable.

Meilleure approche !

Sélection Séquentielle Ascendante.

Estimation des Emotions

Objectif



Estimation des Emotions

Approches

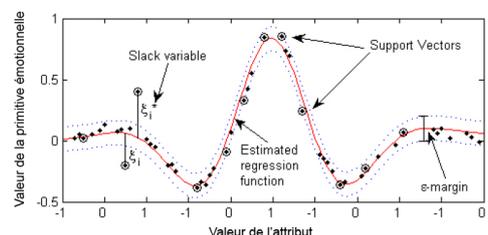
- La sortie désirée n'est pas une classification dans un ensemble fini de catégories mais une estimation à valeurs continues des primitives émotionnelles.
- Plusieurs classifieurs ont été analysés :
 - 1 Régression pas Séparateur à Vaste Marge,
 - 2 Classifieur par l'algorithme ds k-Plus Proche Voisin Flou
 - 3 Une méthode d'Inférence à base de règles Floues

Plan

- 1 Contexte Général
- 2 Pré-traitement des Données
 - Construction du corpus d'apprentissage
 - Réduction de dimensionnalité
- 3 Estimation des Emotions
 - SVMR
 - k-Plus Proches Voisins Flous
 - Système d'Inférence Floue
- 4 Résultats
- 5 Conclusion et Perspectives

Régression par Séparateur à Vaste Marge

SVMR : Description



Plan

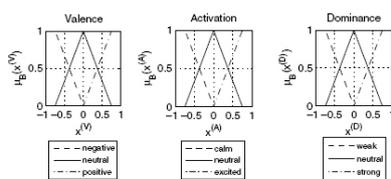
- 1 Contexte Général
- 2 Pré-traitement des Données
 - Construction du corpus d'apprentissage
 - Réduction de dimensionnalité
- 3 Estimation des Emotions
 - SVMR
 - k-Plus Proches Voisins Floues
 - Système d'Inférence Floue
- 4 Résultats
- 5 Conclusion et Perspectives

Plan

- 1 Contexte Général
- 2 Pré-traitement des Données
 - Construction du corpus d'apprentissage
 - Réduction de dimensionnalité
- 3 Estimation des Emotions
 - SVMR
 - k-Plus Proches Voisins Floues
 - Système d'Inférence Floue
- 4 Résultats
- 5 Conclusion et Perspectives

Système d'Inférence Floue

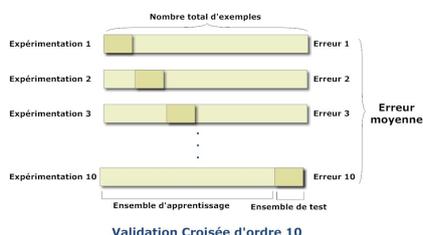
Fonctions d'appartenance



Résultats

Validation croisée d'ordre K

Toutes les expérimentations ont été effectuées en utilisant une validation croisée d'ordre 10



k-Plus Proches Voisins Floues

k-PPVF : Description

- C'est une approche de classification basée sur la définition d'une distance.
 - Elle détermine les k plus proche voisins d'une observation en question dans l'espace des attributs
 - et affecte les propriétés de ces voisins à l'observation.
- Dans notre cas, les propriétés des voisins sont les valeurs des primitives émotionnelles.
- Ces valeurs sont agrégées pour avoir une estimation de l'émotion pour une d'observation, qui correspond à la moyenne des valeurs des voisins.
- À cause de cette moyenne, tous les voisins ont une influence sur l'estimation, d'où le nom k-Plus Proche Voisins Floues

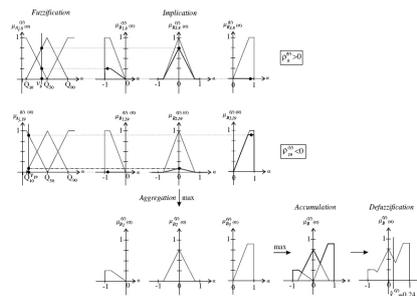
Système d'Inférence à base de Règles Floues

SIF : Description

- La logique floue reproduit bien la nature des émotions qui sont en général floues
- Cette description floue se reflète dans un nombre de terme linguistiques utilisé pour les sentiments, l'humeur et les états affectifs
- Un système d'Inférence Floue consiste en trois éléments principaux :
 - Fuzzification
 - Inférence
 - Défuzzification

Estimation des Emotions

Algorithme



Résultats

Deux mesures pour évaluer les résultats d'apprentissage :

- L'erreur moyenne entre l'émotion estimée et la référence annotée par les évaluateurs humains.
- Le coefficient de corrélation empirique entre l'émotion estimée (automatiquement) et la référence (annotée par l'expert)

Résultats

	(a) Mean Error			(b) Correlation Coefficient		
	Valence	Activation	Dominance	Valence	Activation	Dominance
SVR-RBF	0.13	0.15	0.14	0.46	0.82	0.79
SVR-Pol	0.14	0.16	0.15	0.39	0.80	0.77
SIF	0.27	0.17	0.18	0.28	0.75	0.72
KNN	0.13	0.16	0.14	0.46	0.80	0.78

Merci pour votre attention !

Conclusion et Perspectives

- De bons résultats pour l'estimation de l'émotion
 - faibles taux d'erreurs de 0.13 à 0.15
 - corrélation significative entre les références annotées manuellement et les estimations calculés : modéré (0.46, valence) voire élevée (0.82/0.79, activation/dominance).
- Perspectives :
 - La fusion de l'estimation des primitives émotionnelles avec la reconnaissance automatique de la parole
 - La conception d'un système temps réel utilisant l'algorithme reporté ici.
 - Conception d'interfaces ou des humanoïdes capables de s'adapter aux personnalités et aux humeurs à long terme