

Approche immunitaire pour la classification, application à la reconnaissance de locuteur

Eric Charton

Sous la direction de J.F Bonastre



Mémoire de recherche - Master
2007

Classification immunitaire

E. Charton

Introduction

Application à
la reconnais-
sance du
locuteur

Méthode

Résultats

Conclusions

- 1 Introduction
- 2 Application à la reconnaissance du locuteur
- 3 Méthode
- 4 Résultats
- 5 Conclusions

L'approche immunitaire

Classification immunitaire

E. Charton

Introduction

Application à
la reconnais-
sance du
locuteur

Méthode

Résultats

Conclusions

Objectif

Adjoindre à un système de reconnaissance de locuteur un principe immunitaire inspiré du corps humain, capable d'améliorer la détection des imposteurs.

Etat de l'art

L'approche immunitaire

Classification immunitaire

E. Charton

Introduction

Application à la reconnaissance du locuteur

Méthode

Résultats

Conclusions

Objectif

Adjoindre à un système de reconnaissance de locuteur un principe immunitaire inspiré du corps humain, capable d'améliorer la détection des imposteurs.

Etat de l'art

- 1980-90 Systèmes immunitaires génériques appliqués à la protection des réseaux

L'approche immunitaire

Classification immunitaire

E. Charton

Introduction

Application à la reconnaissance du locuteur

Méthode

Résultats

Conclusions

Objectif

Adjoindre à un système de reconnaissance de locuteur un principe immunitaire inspiré du corps humain, capable d'améliorer la détection des imposteurs.

Etat de l'art

- 1980-90 Systèmes immunitaires génériques appliqués à la protection des réseaux
- 1990-00 Systèmes immunitaires spécialisés appliqués à la détection des virus informatiques

L'approche immunitaire

Classification immunitaire

E. Charton

Introduction

Application à la reconnaissance du locuteur

Méthode

Résultats

Conclusions

Objectif

Adjoindre à un système de reconnaissance de locuteur un principe immunitaire inspiré du corps humain, capable d'améliorer la détection des imposteurs.

Etat de l'art

- 1980-90 Systèmes immunitaires génériques appliqués à la protection des réseaux
- 1990-00 Systèmes immunitaires spécialisés appliqués à la détection des virus informatiques
- 2000-07 Algorithmes de détection de modifications inspirés du fonctionnement des lymphocytes

Système immunitaire à base de lymphocytes

Classification immunitaire

E. Charton

Introduction

Application à
la reconnais-
sance du
locuteur

Méthode

Résultats

Conclusions

Lymphocytes : le principe du "non soi"

Système immunitaire à base de lymphocytes

Classification immunitaire

E. Charton

Introduction

Application à
la reconnais-
sance du
locuteur

Méthode

Résultats

Conclusions

Lymphocytes : le principe du "non soi"

- Ce qui est non caractéristique de l'organisme.

Système immunitaire à base de lymphocytes

Classification immunitaire

E. Charton

Introduction

Application à
la reconnais-
sance du
locuteur

Méthode

Résultats

Conclusions

Lymphocytes : le principe du "non soi"

- Ce qui est non caractéristique de l'organisme.
- Correspond à l'opposé du soi représenté par l'organisme.

Système immunitaire à base de lymphocytes

Classification immunitaire

E. Charton

Introduction

Application à
la reconnais-
sance du
locuteur

Méthode

Résultats

Conclusions

Lymphocytes : le principe du "non soi"

- Ce qui est non caractéristique de l'organisme.
- Correspond à l'opposé du soi représenté par l'organisme.
- Postule que le non soi permet de détecter les intrusions sans les connaître préalablement.

Système immunitaire à base de lymphocytes

Classification immunitaire

E. Charton

Introduction

Application à
la reconnais-
sance du
locuteur

Méthode

Résultats

Conclusions

Lymphocytes : le principe du "non soi"

- Ce qui est non caractéristique de l'organisme.
- Correspond à l'opposé du soi représenté par l'organisme.
- Postule que le non soi permet de détecter les intrusions sans les connaître préalablement.
- Reproduit le fonctionnement des lymphocytes T et du thymus

Système immunitaire à base de lymphocytes

Classification immunitaire

E. Charton

Introduction

Application à
la reconnais-
sance du
locuteur

Méthode

Résultats

Conclusions

Lymphocytes : le principe du "non soi"

- Ce qui est non caractéristique de l'organisme.
- Correspond à l'opposé du soi représenté par l'organisme.
- Postule que le non soi permet de détecter les intrusions sans les connaître préalablement.
- Reproduit le fonctionnement des lymphocytes T et du thymus

Principes fonctionnels

Système immunitaire à base de lymphocytes

Classification immunitaire

E. Charton

Introduction

Application à
la reconnais-
sance du
locuteur

Méthode

Résultats

Conclusions

Lymphocytes : le principe du "non soi"

- Ce qui est non caractéristique de l'organisme.
- Correspond à l'opposé du soi représenté par l'organisme.
- Postule que le non soi permet de détecter les intrusions sans les connaître préalablement.
- Reproduit le fonctionnement des lymphocytes T et du thymus

Principes fonctionnels

- Générer une grande quantité de caractéristiques.

Système immunitaire à base de lymphocytes

Classification immunitaire

E. Charton

Introduction

Application à
la reconnais-
sance du
locuteur

Méthode

Résultats

Conclusions

Lymphocytes : le principe du "non soi"

- Ce qui est non caractéristique de l'organisme.
- Correspond à l'opposé du soi représenté par l'organisme.
- Postule que le non soi permet de détecter les intrusions sans les connaître préalablement.
- Reproduit le fonctionnement des lymphocytes T et du thymus

Principes fonctionnels

- Générer une grande quantité de caractéristiques.
- Utilisation de méthodes pseudo aléatoires.

Système immunitaire à base de lymphocytes

Classification immunitaire

E. Charton

Introduction

Application à
la reconnais-
sance du
locuteur

Méthode

Résultats

Conclusions

Lymphocytes : le principe du "non soi"

- Ce qui est non caractéristique de l'organisme.
- Correspond à l'opposé du soi représenté par l'organisme.
- Postule que le non soi permet de détecter les intrusions sans les connaître préalablement.
- Reproduit le fonctionnement des lymphocytes T et du thymus

Principes fonctionnels

- Générer une grande quantité de caractéristiques.
- Utilisation de méthodes pseudo aléatoires.
- La génération aléatoire proposera automatiquement des caractéristiques du non soi.

Le système immunitaire humain

Classification immunitaire

E. Charton

Introduction

Application à la reconnaissance du locuteur

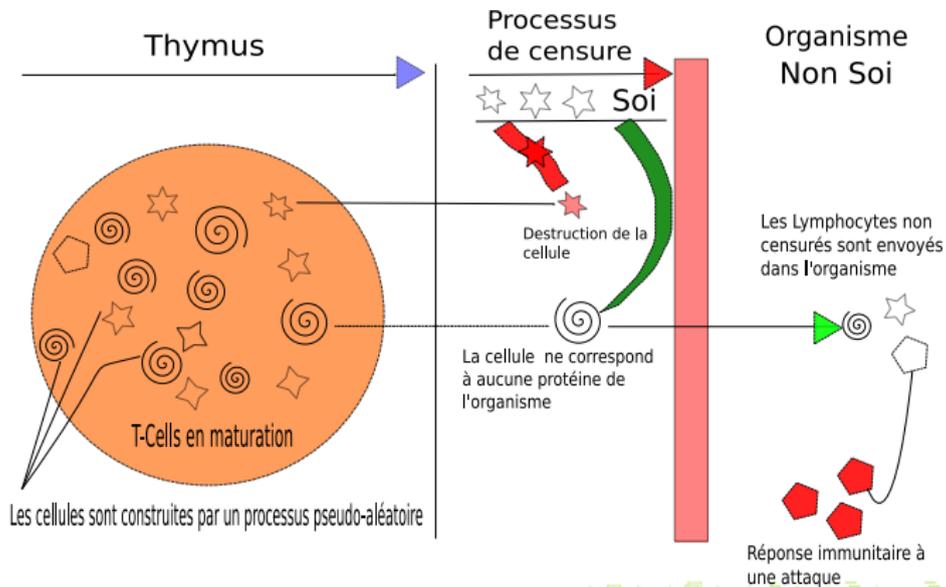
Méthode

Résultats

Conclusions

Prévoir le non soi avec des lymphocytes

Imaginer ce que pourrait être une intrusion sans la connaître



L'algorithme de Forrest

Classification
immunitaire

E. Charton

Introduction

Application à
la reconnaissance du
locuteur

Méthode

Résultats

Conclusions

Repose sur le principe de la sélection négative
Générer le "non Soi" aléatoirement

Etape 1 : Construction des détecteurs

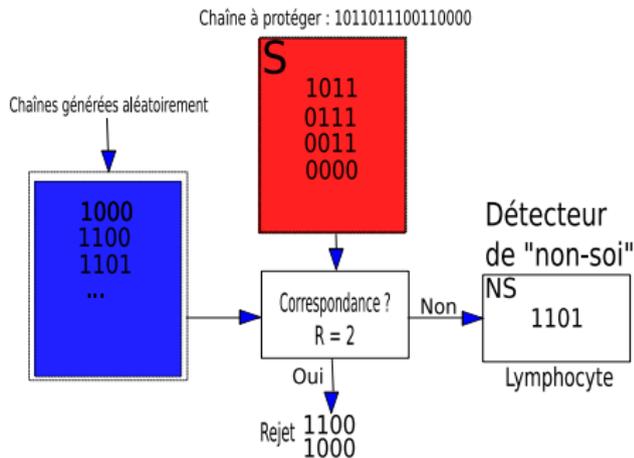


FIG.: Sélection de chaînes de détection du "non Soi"

L'algorithme de Forrest

Classification
immunitaire

E. Charton

Introduction

Application à
la reconnaissance du
locuteur

Méthode

Résultats

Conclusions

Repose sur le principe de la sélection négative

Conserver après apprentissage ce qui ne relève pas du Soi

Etape 2 : système immunitaire

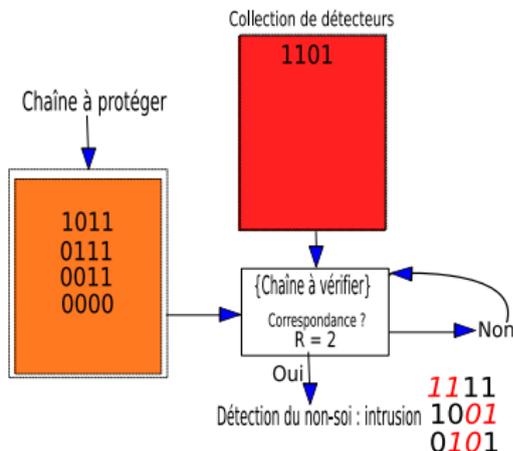


FIG.: Système de détection utilisant les chaînes sélectionnées

Modèle immunitaire et reconnaissance de locuteur

Construire des lymphocytes d'après un système GMM

Classification immunitaire

E. Charton

Introduction

Application à la reconnaissance du locuteur

Méthode

Résultats

Conclusions

Etat de l'art : le modèle GMM UBM

- les classes sont modélisées par des mixtures de gaussiennes multivariées.

Construire des lymphocytes d'après un système GMM

Classification immunitaire

E. Charton

Introduction

Application à
la reconnais-
sance du
locuteur

Méthode

Résultats

Conclusions

Etat de l'art : le modèle GMM UBM

- les classes sont modélisées par des mixtures de gaussiennes multivariées.
- Un GMM "modèle du monde" UBM représentant le modèle universel de locuteurs et un GMM par locuteur.

Construire des lymphocytes d'après un système GMM

Classification immunitaire

E. Charton

Introduction

Application à la reconnaissance du locuteur

Méthode

Résultats

Conclusions

Etat de l'art : le modèle GMM UBM

- les classes sont modélisées par des mixtures de gaussiennes multivariées.
- Un GMM "modèle du monde" UBM représentant le modèle universel de locuteurs et un GMM par locuteur.
- Un test d'hypothèse exprimé dans le domaine logarithmique .

Construire des lymphocytes d'après un système GMM

Classification
immunitaire

E. Charton

Introduction

Application à
la reconnais-
sance du
locuteur

Méthode

Résultats

Conclusions

Etat de l'art : le modèle GMM UBM

- les classes sont modélisées par des mixtures de gaussiennes multivariées.
- Un GMM "modèle du monde" UBM représentant le modèle universel de locuteurs et un GMM par locuteur.
- Un test d'hypothèse exprimé dans le domaine logarithmique .

$$\Lambda(X) = \log p(X|\lambda_{USER}) - \log p(X|\lambda_{WORLD})$$

Partir du modèle GMM UBM

Classification immunitaire

E. Charton

Introduction

Application à la reconnaissance du locuteur

Méthode

Résultats

Conclusions

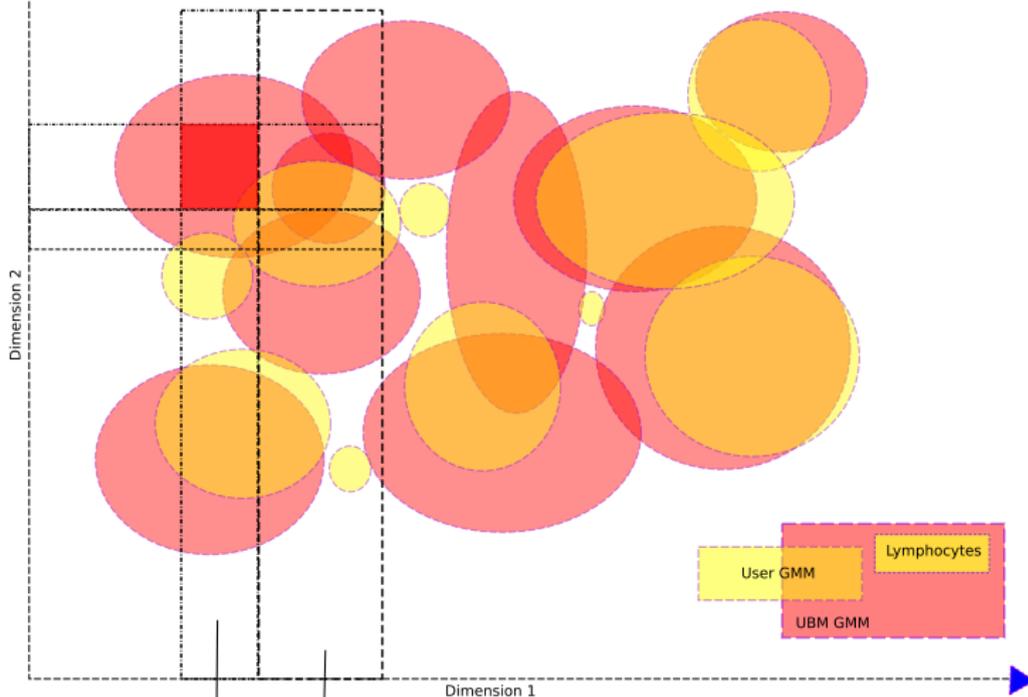
Représentation de deux mixtures de gaussiennes bivariées sans contrainte sphérique
(Vecteur à 2 dimensions - Mixture à 10 gaussiennes - modèle user / world modèle)



Chercher les dissimilarités

Représentation de deux mélanges de gaussiennes bivariées sans contrainte sphérique
(Vecteur à 2 dimensions - Mélange à 10 gaussiennes - modèle user / world modèle)

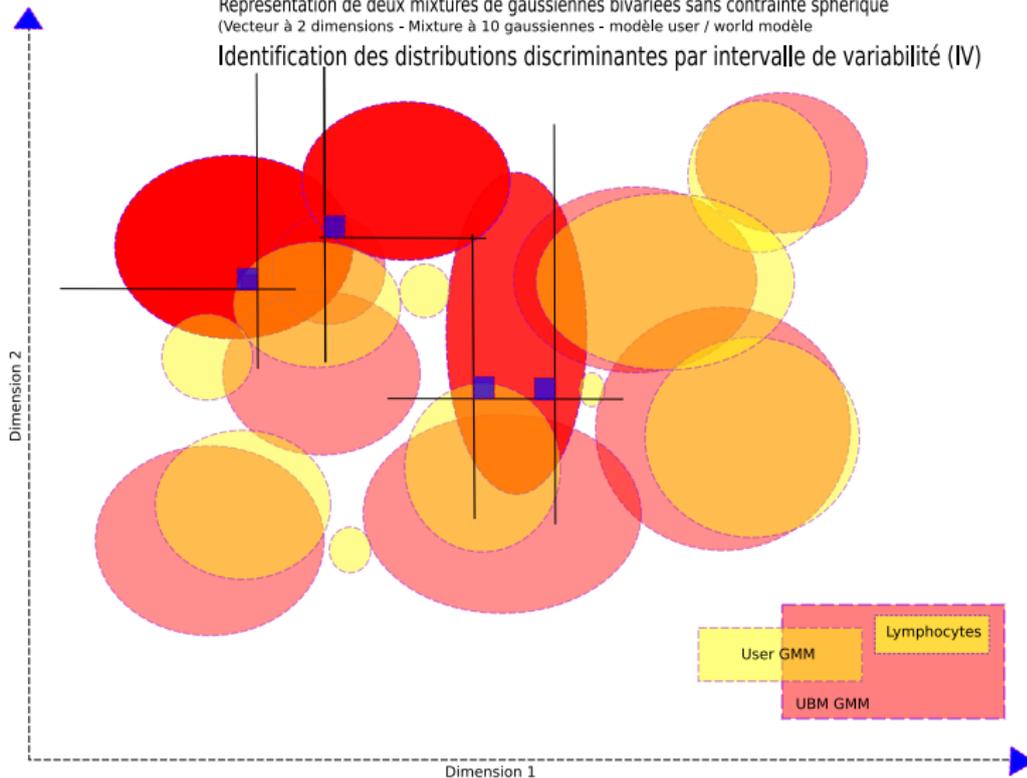
Identification des distributions discriminantes par intervalle de variabilité (IV)



Elaborer les séparateurs

Représentation de deux mixtures de gaussiennes bivariées sans contrainte sphérique
(Vecteur à 2 dimensions - Mixture à 10 gaussiennes - modèle user / world modèle)

Identification des distributions discriminantes par intervalle de variabilité (IV)



Caractère discriminant sur des données

Classification immunitaire

E. Charton

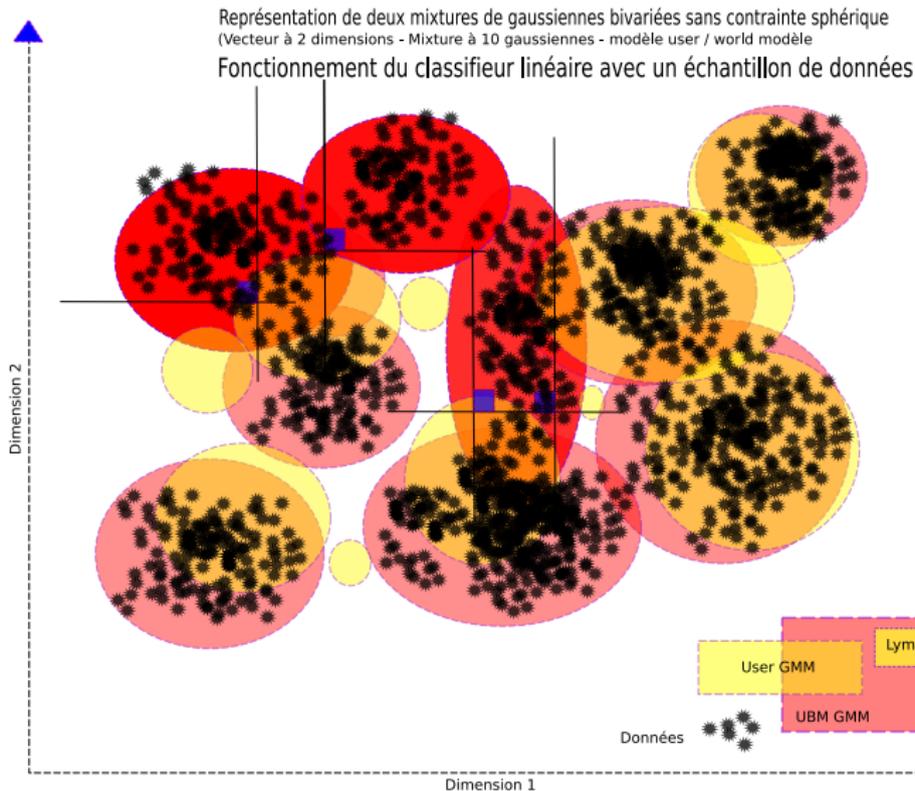
Introduction

Application à la reconnaissance du locuteur

Méthode

Résultats

Conclusions



Test de données

Classification immunitaire

E. Charton

Introduction

Application à la reconnaissance du locuteur

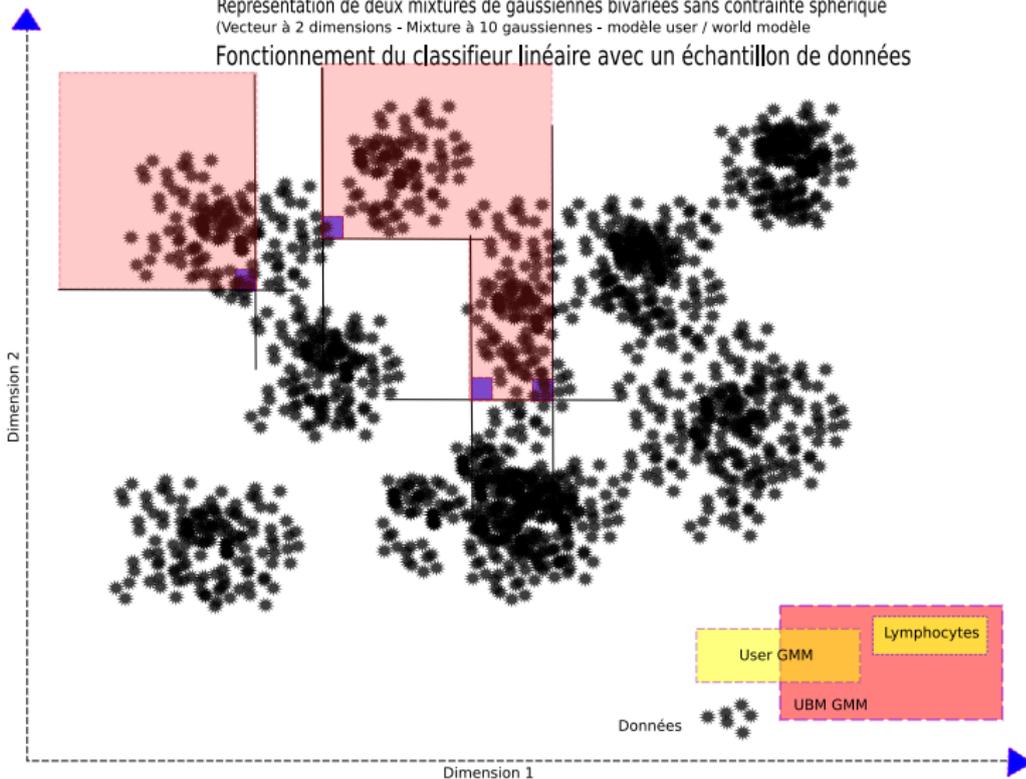
Méthode

Résultats

Conclusions

Représentation de deux mixtures de gaussiennes bivariées sans contrainte sphérique
(Vecteur à 2 dimensions - Mixture à 10 gaussiennes - modèle user / world modèle)

Fonctionnement du classifieur linéaire avec un échantillon de données



Construction d'un système immunitaire d'après un modèle GMM UBM

L'intervalle de variabilité

- Caractériser des distributions "similaires" ou "dissimilaires" en utilisant une mesure dérivée de l'intervalle de confiance
- Rechercher toutes les distributions $\mathcal{N}(\mu_j, \sigma_j)$ d'un GMM_{UBM} ne modélisant aucune des données de $\mathcal{N}(\mu_i, \sigma_i)$, dimension par dimension.

$$\forall i \in \{1, \dots, n\}, \forall j \in \{1, \dots, m\}, IV_{A_i} \cap IV_{B_j} = \emptyset \quad (1)$$

L'intervalle de variabilité

Classification
immunitaire

E. Charton

Introduction

Application à
la reconnaissance
du locuteur

Méthode

Résultats

Conclusions

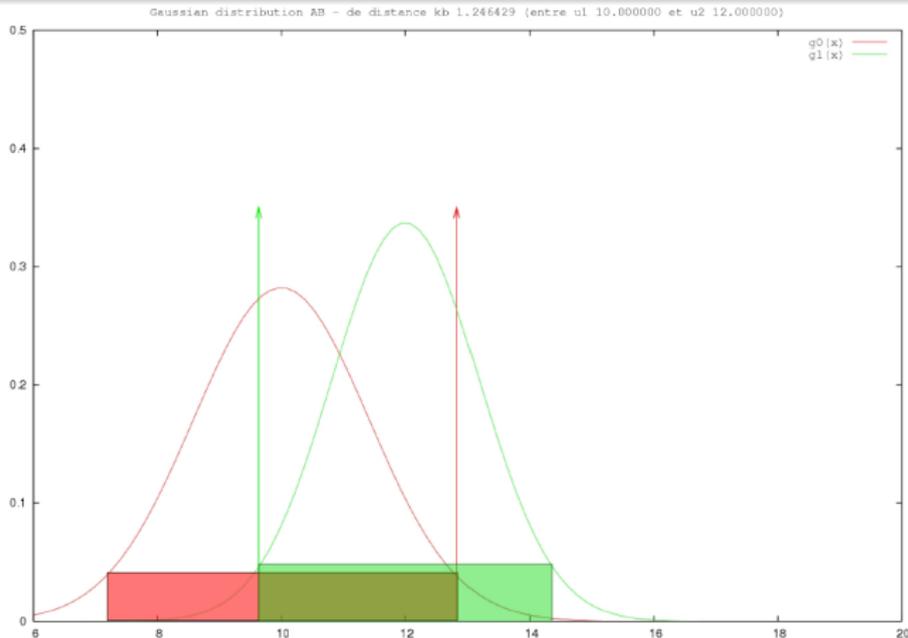


FIG.: Les deux intervalles de variabilité se recouvrent :
 $(\mu_A + (\sigma_A * u_{\alpha/2})) > (\mu_B - (\sigma_B * u_{\alpha/2}))$. La mesure de similarité est
 fixée à 1.

L'intervalle de variabilité

Classification immunitaire

E. Charton

Introduction

Application à la reconnaissance du locuteur

Méthode

Résultats

Conclusions

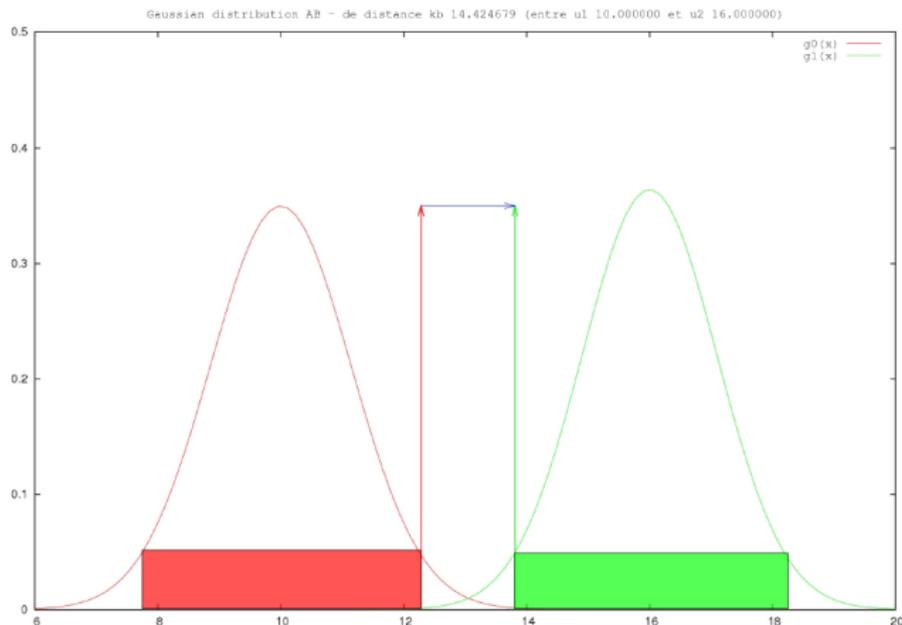


FIG.: Les deux intervalles de variabilité ne se recouvrent pas : $(\mu_A + (\sigma_A * u_{\alpha/2})) < (\mu_B - (\sigma_B * u_{\alpha/2}))$. La mesure de similarité est fixée à 0.

Matrice de dissimilarité

Classification immunitaire

E. Charton

Introduction

Application à la reconnaissance du locuteur

Méthode

Résultats

Conclusions

	GMM	UBM1	UBM2	UBM3	USER vs UBM
USER1		0	1	1	>0
USER2		0	1	1	>0
USER3		0	0	0	>0
USER4		0	1	0	>0
UBM vs USER		0	>0	>0	

TAB.: L'intervalle de confiance de la distribution 1 du GMM_{UBM} n'inclut aucune des valeurs comprises dans les intervalles de confiance des distributions 1 à 4 du GMM_{USER} . La somme de la colonne est donc égale à 0 et la distribution UBM_1 est retenue pour élaborer les séparateurs.

Application au GMM par itérations

Classification
immunitaire

E. Charton

Introduction

Application à
la reconnaissance
du locuteur

Méthode

Résultats

Conclusions

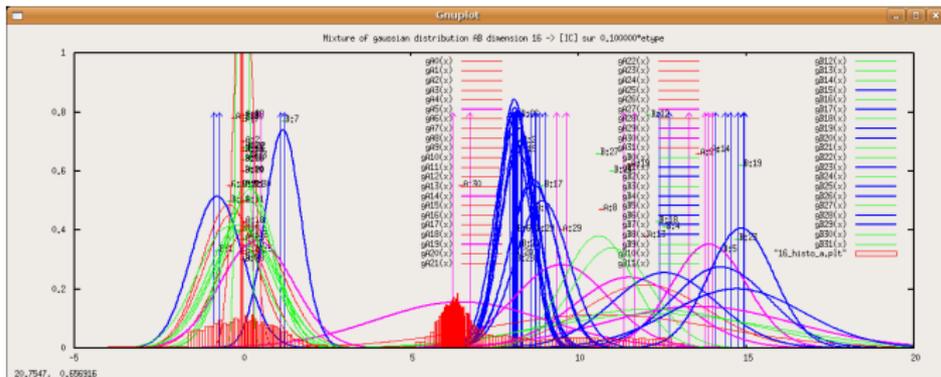


FIG.: Exemple de distributions issues de deux modèles GMM comparés dans une dimension. Les vecteurs verticaux délimitent l'intervalle de variabilité retenu pour comparer les distributions avec leurs voisines.

Combinaisons multidimensionnelles des distributions

Classification immunitaire

E. Charton

Introduction

Application à
la reconnais-
sance du
locuteur

Méthode

Résultats

Conclusions

- Regroupement de distributions dont la taille peut varier de 1 à k' dans la limite des k dimensions de GMM_{UBM} .

Combinaisons multidimensionnelles des distributions

Classification immunitaire

E. Charton

Introduction

Application à
la reconnais-
sance du
locuteur

Méthode

Résultats

Conclusions

- Regroupement de distributions dont la taille peut varier de 1 à k' dans la limite des k dimensions de GMM_{UBM} .
- Permet d'élaborer un hyperplan séparateur dans le sous espace k' de l'espace à k dimensions.

Combinaisons multidimensionnelles des distributions

Classification immunitaire

E. Charton

Introduction

Application à
la reconnaissance du
locuteur

Méthode

Résultats

Conclusions

- Regroupement de distributions dont la taille peut varier de 1 à k' dans la limite des k dimensions de GMM_{UBM} .
- Permet d'élaborer un hyperplan séparateur dans le sous espace k' de l'espace à k dimensions.
- Un hybride entre les systèmes à noyaux et ceux de types gaussiens avec test de vraisemblance.

Evaluation et résultats

Le système immunitaire en complément du GMM

Classification immunitaire

E. Charton

Introduction

Application à la reconnaissance du locuteur

Méthode

Résultats

Conclusions

Une évaluation avec GMM

- Modèle du monde GMM à 2048 gaussiennes : Non X

Le système immunitaire en complément du GMM

Classification immunitaire

E. Charton

Introduction

Application à la reconnaissance du locuteur

Méthode

Résultats

Conclusions

Une évaluation avec GMM

- Modèle du monde GMM à 2048 gaussiennes : Non X
- Modèle des users GMM à 512 gaussiennes : X

Le système immunitaire en complément du GMM

Classification immunitaire

E. Charton

Introduction

Application à la reconnaissance du locuteur

Méthode

Résultats

Conclusions

Une évaluation avec GMM

- Modèle du monde GMM à 2048 gaussiennes : Non X
- Modèle des users GMM à 512 gaussiennes : X
- 37400 tests NIST

Le système immunitaire en complément du GMM

Classification immunitaire

E. Charton

Introduction

Application à la reconnaissance du locuteur

Méthode

Résultats

Conclusions

Une évaluation avec GMM

- Modèle du monde GMM à 2048 gaussiennes : Non X
- Modèle des users GMM à 512 gaussiennes : X
- 37400 tests NIST

Application du système immunitaire

Le système immunitaire en complément du GMM

Classification immunitaire

E. Charton

Introduction

Application à la reconnaissance du locuteur

Méthode

Résultats

Conclusions

Une évaluation avec GMM

- Modèle du monde GMM à 2048 gaussiennes : Non X
- Modèle des users GMM à 512 gaussiennes : X
- 37400 tests NIST

Application du système immunitaire

- Sélection des lymphocytes en utilisant les données d'apprentissage de tous les locuteurs

Le système immunitaire en complément du GMM

Classification immunitaire

E. Charton

Introduction

Application à la reconnaissance du locuteur

Méthode

Résultats

Conclusions

Une évaluation avec GMM

- Modèle du monde GMM à 2048 gaussiennes : Non X
- Modèle des users GMM à 512 gaussiennes : X
- 37400 tests NIST

Application du système immunitaire

- Sélection des lymphocytes en utilisant les données d'apprentissage de tous les locuteurs
- Système immunitaire applicable sur un ensemble de locuteurs clients : 20192 tests de l'évaluation NIST sur les 37400

Le système immunitaire en complément du GMM

Classification immunitaire

E. Charton

Introduction

Application à la reconnaissance du locuteur

Méthode

Résultats

Conclusions

Une évaluation avec GMM

- Modèle du monde GMM à 2048 gaussiennes : Non X
- Modèle des users GMM à 512 gaussiennes : X
- 37400 tests NIST

Application du système immunitaire

- Sélection des lymphocytes en utilisant les données d'apprentissage de tous les locuteurs
- Système immunitaire applicable sur un ensemble de locuteurs clients : 20192 tests de l'évaluation NIST sur les 37400
- Fusion des scores en minorant les valeurs LLR (Log Likelihood Ratio) obtenues par GMM

Le système immunitaire comme un pré-classifieur

Classification immunitaire

E. Charton

Introduction

Application à la reconnaissance du locuteur

Méthode

Résultats

Conclusions

Principe de fusion

Le système immunitaire comme un pré-classifieur

Classification immunitaire

E. Charton

Introduction

Application à la reconnaissance du locuteur

Méthode

Résultats

Conclusions

Principe de fusion

- Si un imposteur est détecté par le système immunitaire, le LLR final doit correspondre à celui d'un "imposteur". Le LLR obtenu par GMM est donc minoré :

Le système immunitaire comme un pré-classifieur

Classification immunitaire

E. Charton

Introduction

Application à la reconnaissance du locuteur

Méthode

Résultats

Conclusions

Principe de fusion

- Si un imposteur est détecté par le système immunitaire, le LLR final doit correspondre à celui d'un "imposteur". Le LLR obtenu par GMM est donc minoré :
 - $LLR - 10$, où 10 correspond à l'amplitude de l'intervalle des valeurs LLR attribuées aux locuteurs dans cette expérience

Le système immunitaire comme un pré-classifieur

Principe de fusion

- Si un imposteur est détecté par le système immunitaire, le LLR final doit correspondre à celui d'un "imposteur". Le LLR obtenu par GMM est donc minoré :
 - $LLR - 10$, où 10 correspond à l'amplitude de l'intervalle des valeurs LLR attribuées aux locuteurs dans cette expérience
- Sinon le score LLR final demeure celui obtenu avec le GMM

Le système immunitaire comme un pré-classifieur

Classification immunitaire

E. Charton

Introduction

Application à la reconnaissance du locuteur

Méthode

Résultats

Conclusions

Principe de fusion

- Si un imposteur est détecté par le système immunitaire, le LLR final doit correspondre à celui d'un "imposteur". Le LLR obtenu par GMM est donc minoré :
 - $LLR - 10$, où 10 correspond à l'amplitude de l'intervalle des valeurs LLR attribuées aux locuteurs dans cette expérience
- Sinon le score LLR final demeure celui obtenu avec le GMM

Résultats

Le système immunitaire comme un pré-classifieur

Classification immunitaire

E. Charton

Introduction

Application à la reconnaissance du locuteur

Méthode

Résultats

Conclusions

Principe de fusion

- Si un imposteur est détecté par le système immunitaire, le LLR final doit correspondre à celui d'un "imposteur". Le LLR obtenu par GMM est donc minoré :
 - $LLR - 10$, où 10 correspond à l'amplitude de l'intervalle des valeurs LLR attribuées aux locuteurs dans cette expérience
- Sinon le score LLR final demeure celui obtenu avec le GMM

Résultats

- Lymphocytes actifs 3600 fois

Le système immunitaire comme un pré-classifieur

Classification immunitaire

E. Charton

Introduction

Application à la reconnaissance du locuteur

Méthode

Résultats

Conclusions

Principe de fusion

- Si un imposteur est détecté par le système immunitaire, le LLR final doit correspondre à celui d'un "imposteur". Le LLR obtenu par GMM est donc minoré :
 - $LLR - 10$, où 10 correspond à l'amplitude de l'intervalle des valeurs LLR attribuées aux locuteurs dans cette expérience
- Sinon le score LLR final demeure celui obtenu avec le GMM

Résultats

- Lymphocytes actifs 3600 fois
- 3598 imposteurs proposés

Le système immunitaire comme un pré-classifieur

Classification immunitaire

E. Charton

Introduction

Application à la reconnaissance du locuteur

Méthode

Résultats

Conclusions

Principe de fusion

- Si un imposteur est détecté par le système immunitaire, le LLR final doit correspondre à celui d'un "imposteur". Le LLR obtenu par GMM est donc minoré :
 - $LLR - 10$, où 10 correspond à l'amplitude de l'intervalle des valeurs LLR attribuées aux locuteurs dans cette expérience
- Sinon le score LLR final demeure celui obtenu avec le GMM

Résultats

- Lymphocytes actifs 3600 fois
- 3598 imposteurs proposés
- 2 faux rejets (tests clients proposés comme imposteurs)

Fusion des scores

Classification immunitaire

E. Charton

Introduction

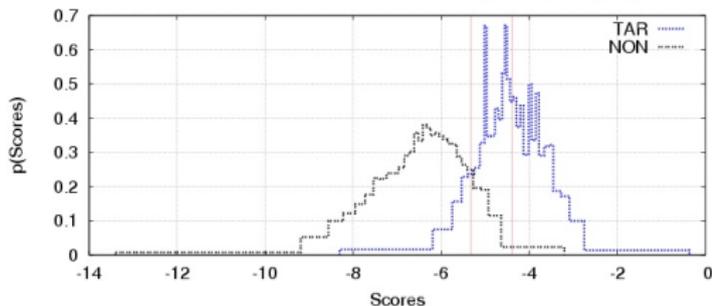
Application à
la reconnaissance
du locuteur

Méthode

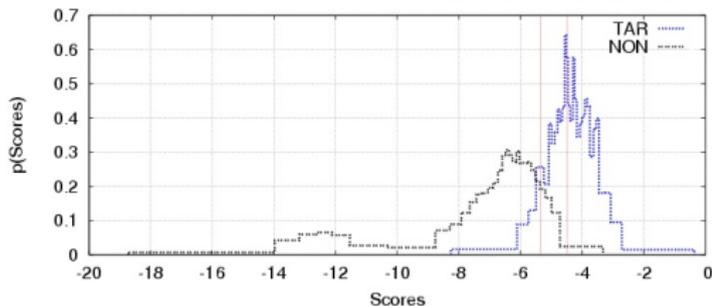
Résultats

Conclusions

LIA: NIST 2006 det3 1conv4w-1conv4w: Histogram target_seg_male₂.nist



LIA: NIST 2006 det3 1conv4w-1conv4w: Histogram lympho, target_seg_male₄.nist



Progression du système

Classification immunitaire

E. Charton

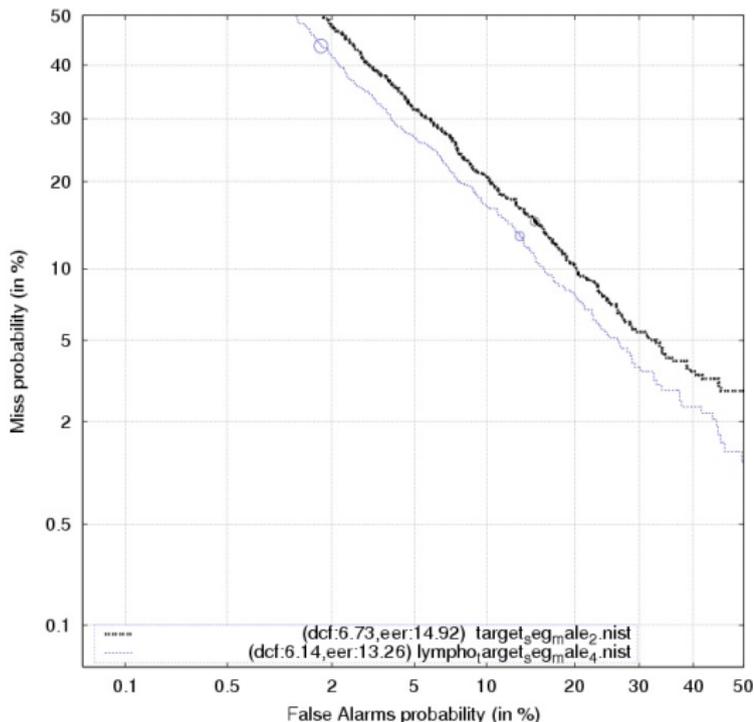
Introduction

Application à la reconnaissance du locuteur

Méthode

Résultats

Conclusions



Conclusions et perspectives

Une méthode prometteuse et originale

Classification immunitaire

E. Charton

Introduction

Application à
la reconnaissance du
locuteur

Méthode

Résultats

Conclusions

En résumé

- Un système complémentaire du modèle GMM original
- Une méthode hybride entre système à noyau et système gaussien
- Une utilisation originale d'un sous ensemble de l'espace de représentation

Une méthode prometteuse et originale

Classification immunitaire

E. Charton

Introduction

Application à
la reconnaissance du
locuteur

Méthode

Résultats

Conclusions

Perspectives

- Passer de séparateurs orthogonaux à des séparateurs linéaires
- Disymétrisation des séparateurs
- Améliorer la sélection des lymphocytes en utilisant les données du modèle du monde
- Appliquer le système immunitaire à une évaluation standard UBM GMM

Classification immunitaire

E. Charton

Introduction

Application à
la reconnais-
sance du
locuteur

Méthode

Résultats

Conclusions

Merci pour votre attention.

Sortie courbe DET avant sélection des lymphocytes

Classification immunitaire

E. Charton

Introduction

Application à la reconnaissance du locuteur

Méthode

Résultats

Conclusions

