Reconnaissance du genre par analyse de dynamique de frappe au clavier sur texte libre

SARSSI 2012

R. Giot[1], C. Rosenberger[1], B. Dorizzi[2]

[1] Université de Caen Basse-Normandie, UMR 6072 GREYC
ENSICAEN, UMR 6072 GREYC
CNRS, UMR 6072 GREYC

[2] Institut Télécom; Télécom SudParis
UMR 5157 SAMOVAR

25 Mai 2012

Sommaire

- Objectif
- Contexte
  - Dynamique de frappe au clavier
  - Biométrie douce
- Méthode développée
  - Données extraites
  - Classifieur utilisé
  - Fonctionnement général
- Protocole expérimental
  - Base de données
  - Scénarios et résultats
- Résultats et discussion
  - Reconnaissance sur phrase connue
  - Reconnaissance sur phrase inconnue
  - Utilisation de plusieurs phrases
- Conclusion

Sommaire

Objectif

Contexte

Méthode développée

Protocole expérimental

Résultats et discussion

Conclusion

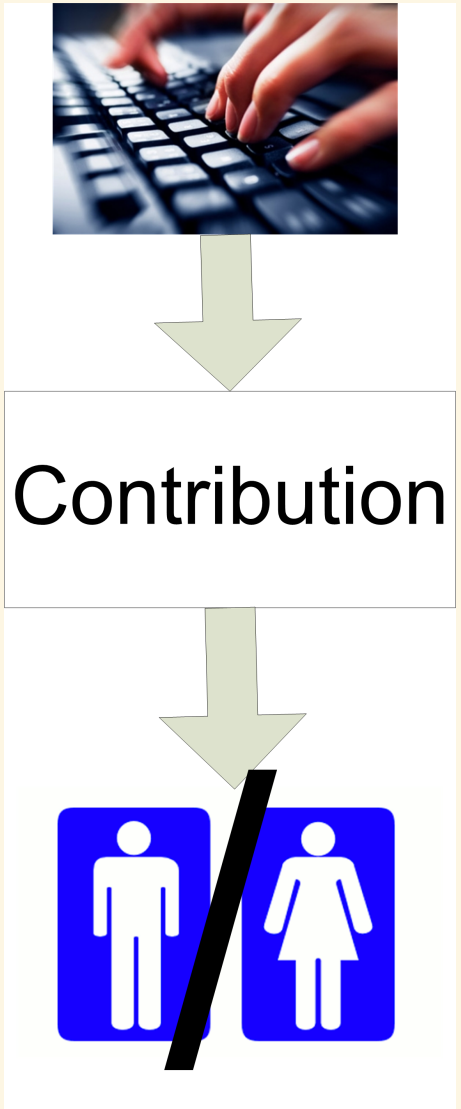
Objectif = Reconnaissance du genre sur texte libre

Objectif

- Reconnaissance du genre sur texte libre

Application

Sécurité Lutte contre la pédophilie
Amélioration dynamique de frappe
Usage Auto-configuration applications



Objectif = Reconnaissance du genre sur texte libre

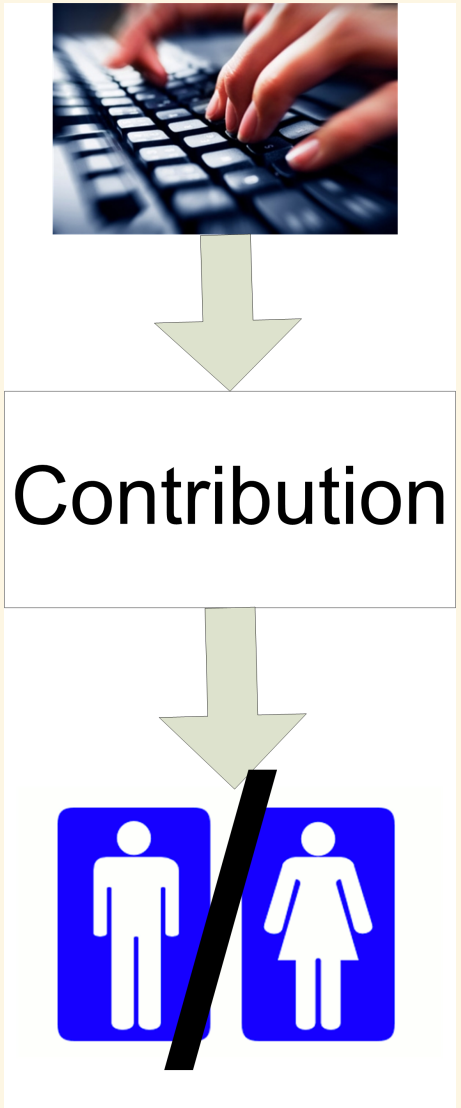
Objectif

- Reconnaissance du genre sur texte libre

Application

Sécurité Lutte contre la pédophilie
Amélioration dynamique de frappe

Usage Auto-configuration applications



Sommaire

Objectif

Contexte
- Dynamique de frappe au clavier
- Biométrie douce

Méthode développée

Protocole expérimental

Résultats et discussion

Conclusion

La dynamique de frappe est une modalité comportementale

- Vérification de la façon de taper au clavier
- Ne nécessite pas de capteur spécifique
  - Uniquement le clavier d'un ordinateur
  - $\Rightarrow$ Faible cout
- Bonne acceptation des utilisateurs
  - Non intrusif (*El-Abed et al. 2010*)
- Mais, sujette à une forte variabilité intra-classe (*Giot et al. 2011*)

La dynamique de frappe est une modalité comportementale

- Vérification de la façon de taper au clavier
- Ne nécessite pas de capteur spécifique
  - Uniquement le clavier d'un ordinateur
  - $\Rightarrow$ Faible cout
- Bonne acceptation des utilisateurs
  - Non intrusif (*El-Abed et al. 2010*)
- Mais, sujette à une forte variabilité intra-classe (*Giot et al. 2011*)

La dynamique de frappe est une modalité comportementale

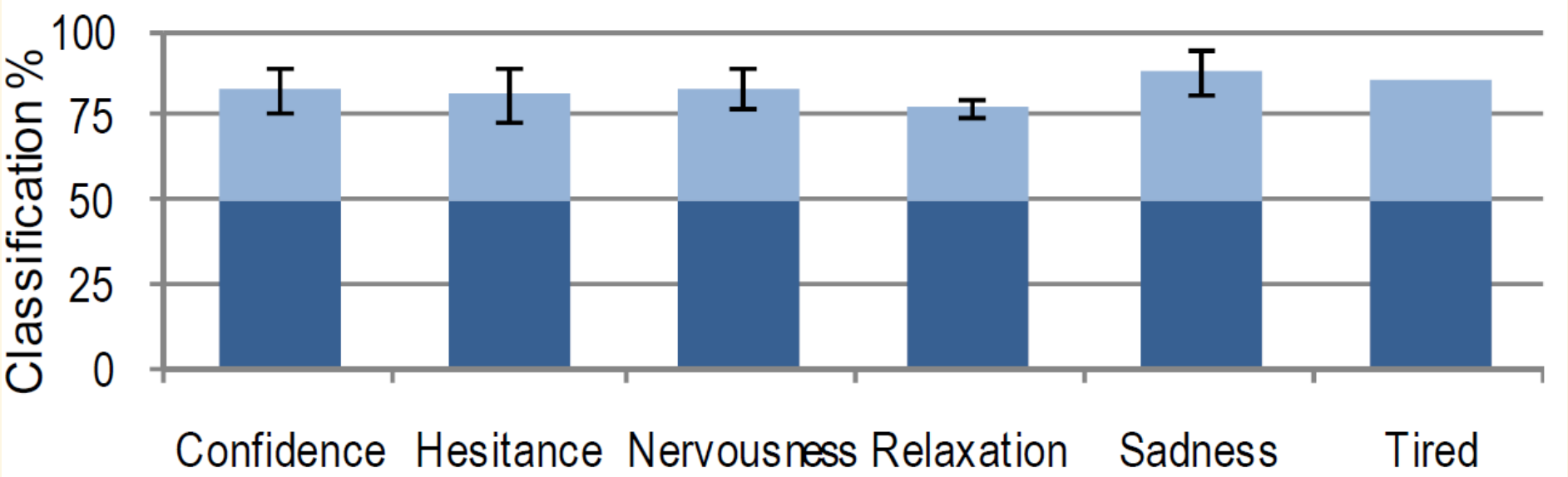
- Vérification de la façon de taper au clavier
- Ne nécessite pas de capteur spécifique
  - Uniquement le clavier d'un ordinateur
  - $\Rightarrow$ Faible cout
- Bonne acceptation des utilisateurs
  - Non intrusif (*El-Abed et al. 2010*)
- Mais, sujette à une forte variabilité intra-classe (*Giot et al. 2011*)

Le taux de reconnaissance du genre à partir d'une photo est de 90%

Fusion	FERET	UND
Intensity	95.33	85.46
Shape	96.26	86.78
Texture	93.46	84.58
20 × 20	85.98	73.57
36 × 36	91.59	80.62
128 × 128	95.33	79.30
All	99.07	91.19

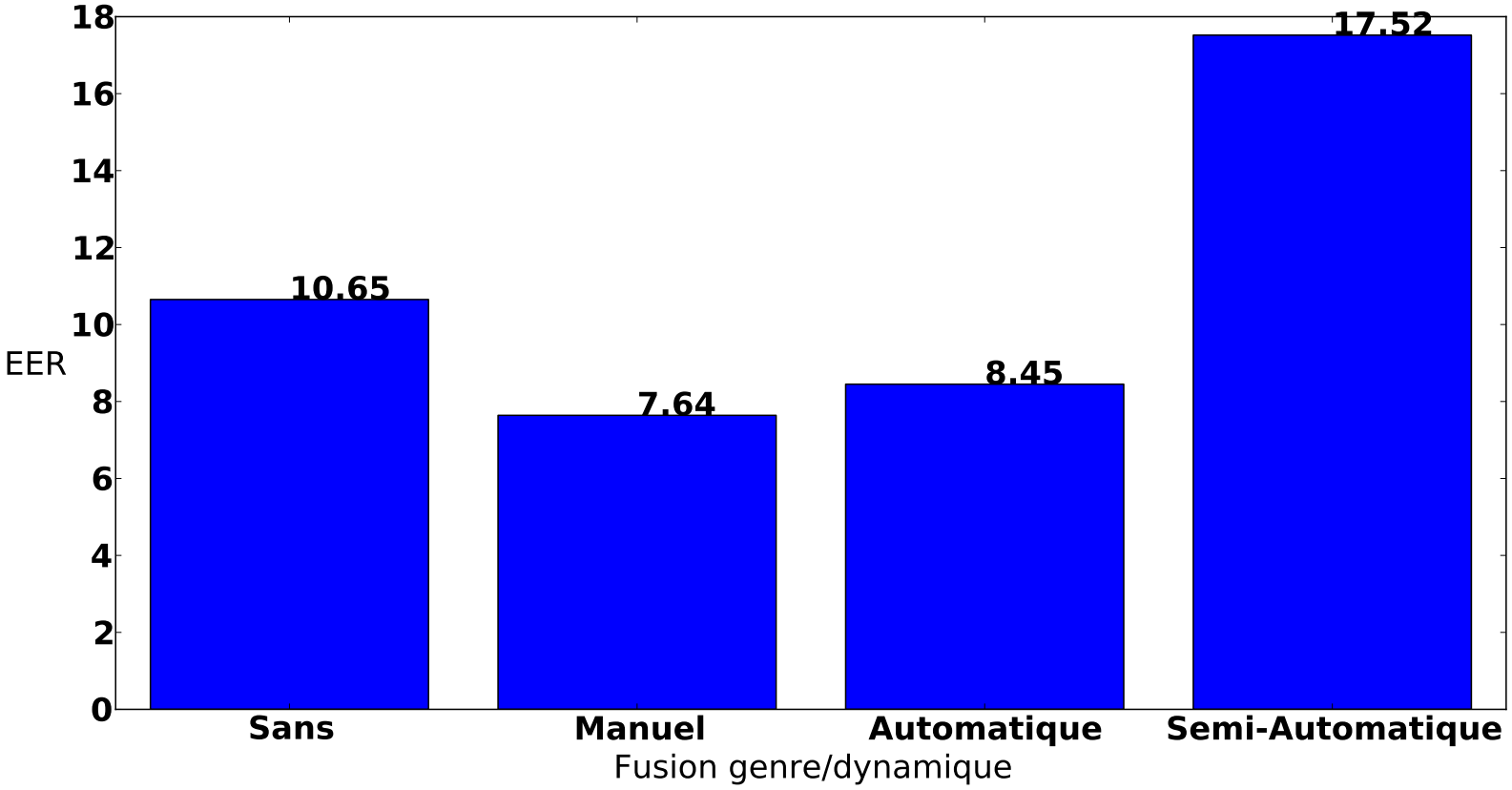
L. A. Alexandre, “Gender recognition : a multiscale decision fusion approach,” Pattern Recognition Letters, "February" 2010.

Il est possible de reconnaitre l'état émotionnel avec la dynamique de frappe



C. Epp, M. Lippold, and R. Mandryk, "Identifying emotional states using keystroke dynamics," in Proceedings of the 2011 annual conference on Human factors in computing systems, 2011, pp. 715–724.

Il est possible de reconnaitre le genre sur texte fixe à 90%



R. Giot and C. Rosenberger, "A new soft biometric approach for keystroke dynamics based on gender recognition," International Journal of Information Technology and Management (IJITM). Special Issue on : "Advances and Trends in Biometrics by Dr Lidong Wang, vol. 11, no. 1/2, pp. 35–49, 2012.

Sommaire

Objectif

Contexte

Méthode développée
- Données extraites
- Classifieur utilisé
- Fonctionnement général

Protocole expérimental

Résultats et discussion

Conclusion

Données extraites pour chaque phrase

Type de données

- Temps de pression (PR)
- Intervalles de temps (RP, PP, RR)
- Pas de filtrage de données aberrantes

Digraphes sélectionnés

- Digraphes communs dans la langue
- Limitation à ceux du corpus

Descripteur

- Vecteur de 23 éléments

Données extraites pour chaque phrase

Type de données

- Temps de pression (PR)
- Intervalles de temps (RP, PP, RR)
- Pas de filtrage de données aberrantes

Digraphes sélectionnés

- Digraphes communs dans la langue
- Limitation à ceux du corpus

Descripteur

- Vecteur de 23 éléments

Données extraites pour chaque phrase

Type de données

- Temps de pression (PR)
- Intervalles de temps (RP, PP, RR)
- Pas de filtrage de données aberrantes

Digraphes sélectionnés

- Digraphes communs dans la langue
- Limitation à ceux du corpus

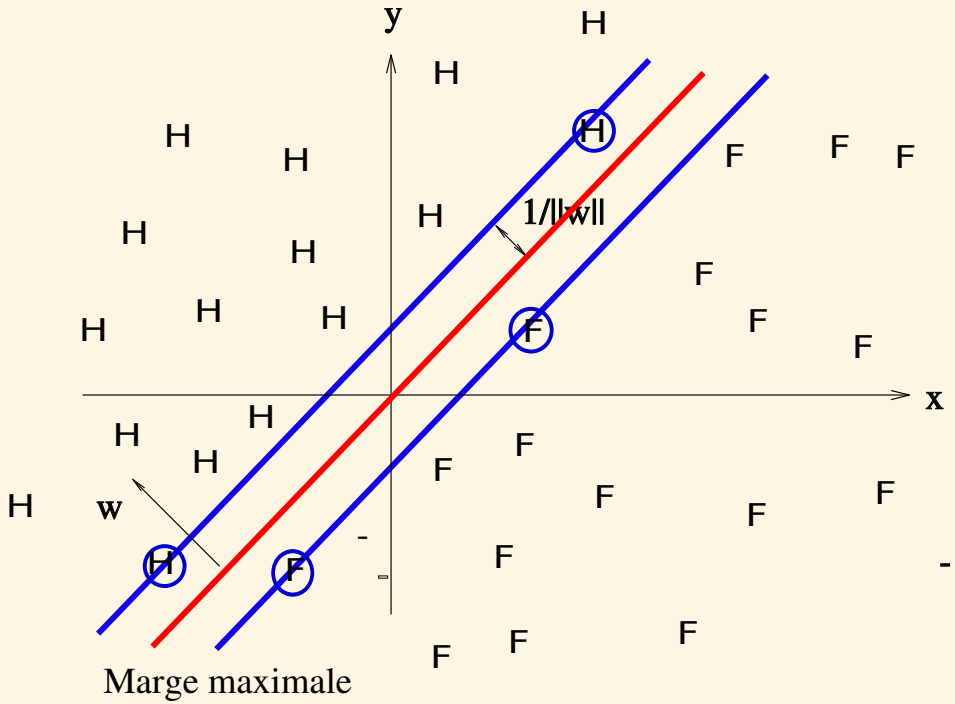
Descripteur

- Vecteur de 23 éléments

Utilisation d'un Séparateur à Vastes Marge pour classer hommes/femmes

Principes du SVM

- Classifieur à 2 classes
- Principes principaux :
  - Sélection de la marge maximale
  - Utilisation de fonction noyau pour séparations non linéaires



Utilisation du SVM

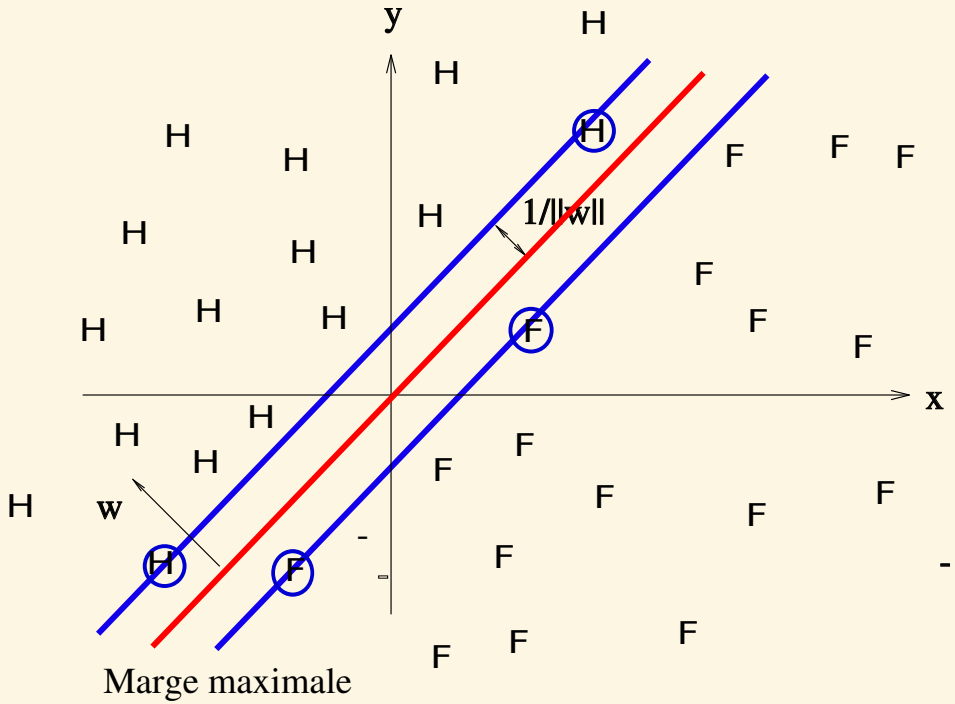
- Normalisation des données avant utilisation
- Obtention du % d'appartenance à la classe (plutôt que de l'étiquette)

V. Vapnik, Statistical learning theory. Wiley New York, 1998.

Utilisation d'un Séparateur à Vastes Marge pour classer hommes/femmes

Principes du SVM

- Classifieur à 2 classes
- Principes principaux :
  - Sélection de la marge maximale
  - Utilisation de fonction noyau pour séparations non linéaires

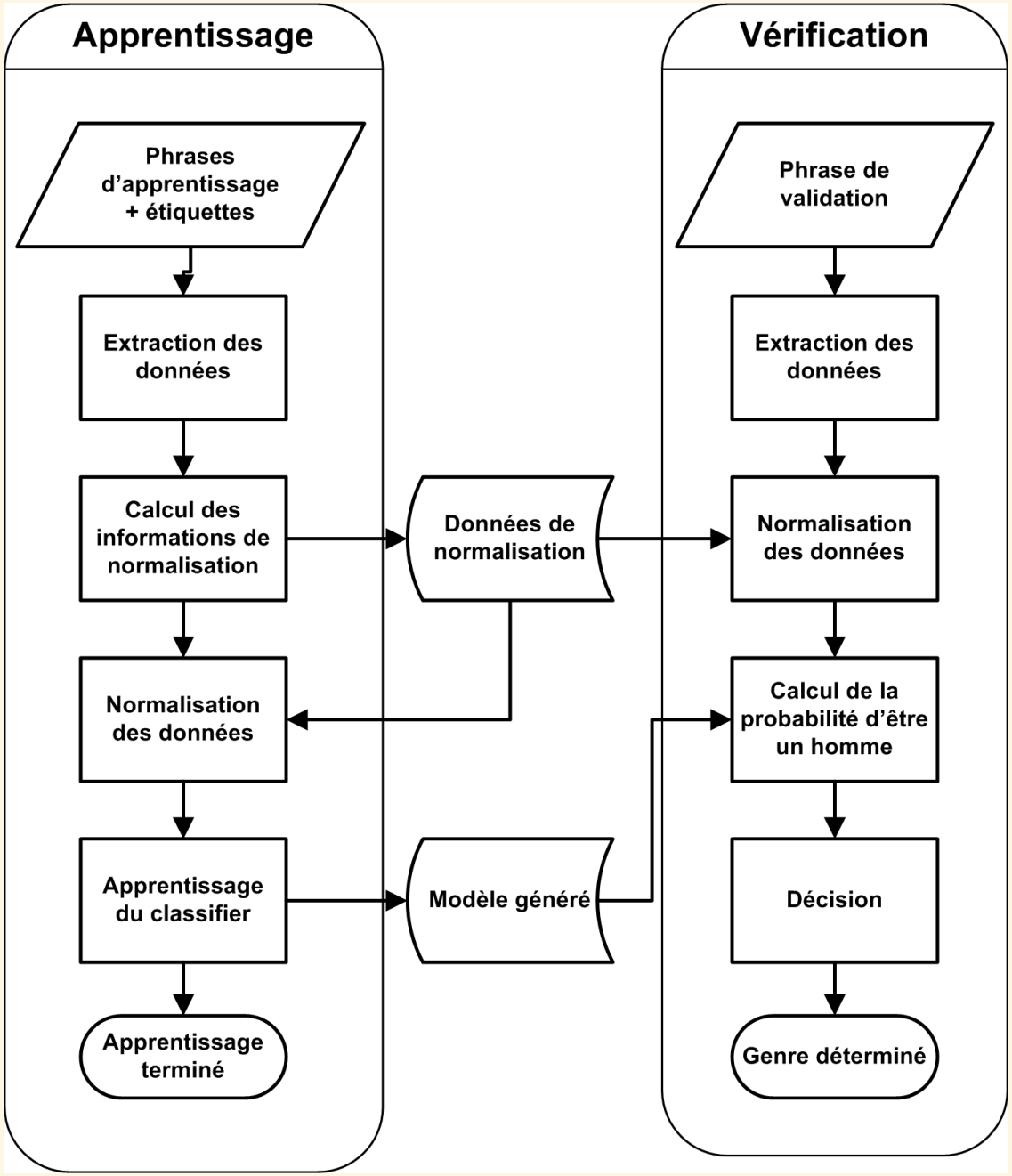


Utilisation du SVM

- Normalisation des données avant utilisation
- Obtention du % d'appartenance à la classe (plutôt que de l'étiquette)

V. Vapnik, Statistical learning theory. Wiley New York, 1998.

Fonctionnement général



Sommaire

Objectif

Contexte

Méthode développée

Protocole expérimental
- Base de données
- Scénarios et résultats

Résultats et discussion

Conclusion

Base de données utilisée

- Base publique
- 54 volontaires
  - 39 hommes
  - 15 femmes
- 15 phrases en espagnol
- phrases potentiellement bruitées
  - entre 90% et 130% du texte attendu
- Seuls 6 digraphes sont retenus

L. Bello, M. Bertacchini, C. Benitez, J. Carlos, Pizzoni, and M. Cipriano, "Collection and publication of a fixed text keystroke dynamics dataset," in XVI Congreso Argentino de Ciencias de la Computacion (CACIC 2010), 2010.

Scénarios de reconnaissance

- 3 Scénarios de reconnaissance testés
- 1 scénario = 1 combinaison différente des phrases
  - d'apprentissage
  - de reconnaissance
- Exemples de validation toujours différents des exemples d'apprentissage
- Peu de données $\Rightarrow$ Validation croisée

Sommaire

Objectif

Contexte

Méthode développée

Protocole expérimental

Résultats et discussion
- Reconnaissance sur phrase connue
- Reconnaissance sur phrase inconnue
- Utilisation de plusieurs phrases

Conclusion

Scénario 1 : reconnaissance sur phrase connue

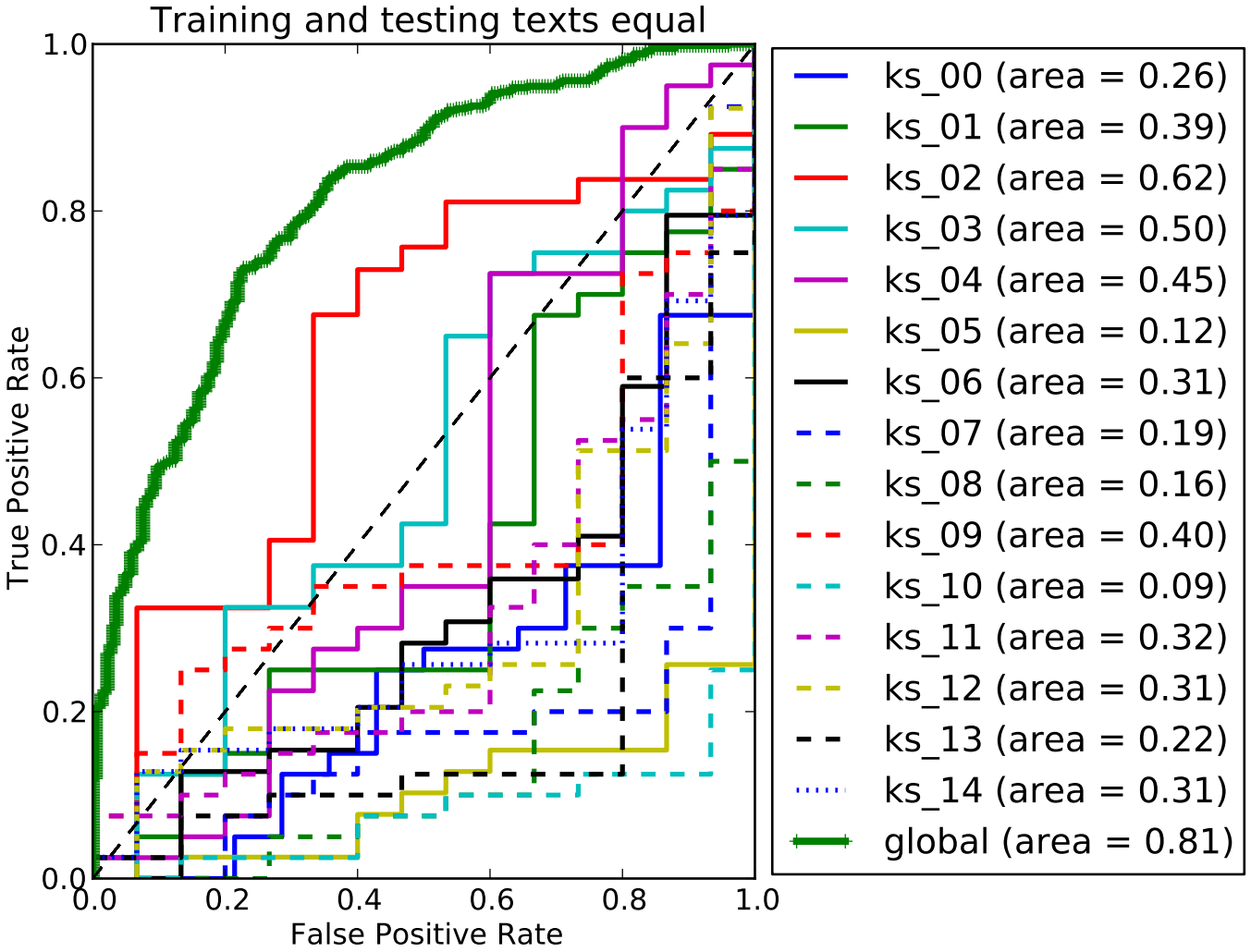
Scénario

1. Apprentissage avec les exemples d'une phrase
2. Reconnaissance avec un exemple de la même phrase

Chiffres

- 15+1 scénarios
- 54 exemples pour les 15 premiers sous-scénarios
- 54*15 = 810 exemples pour le 16e

Résultats 1 : reconnaissance sur phrase connue



Scénario 2 : reconnaissance sur phrase inconnue

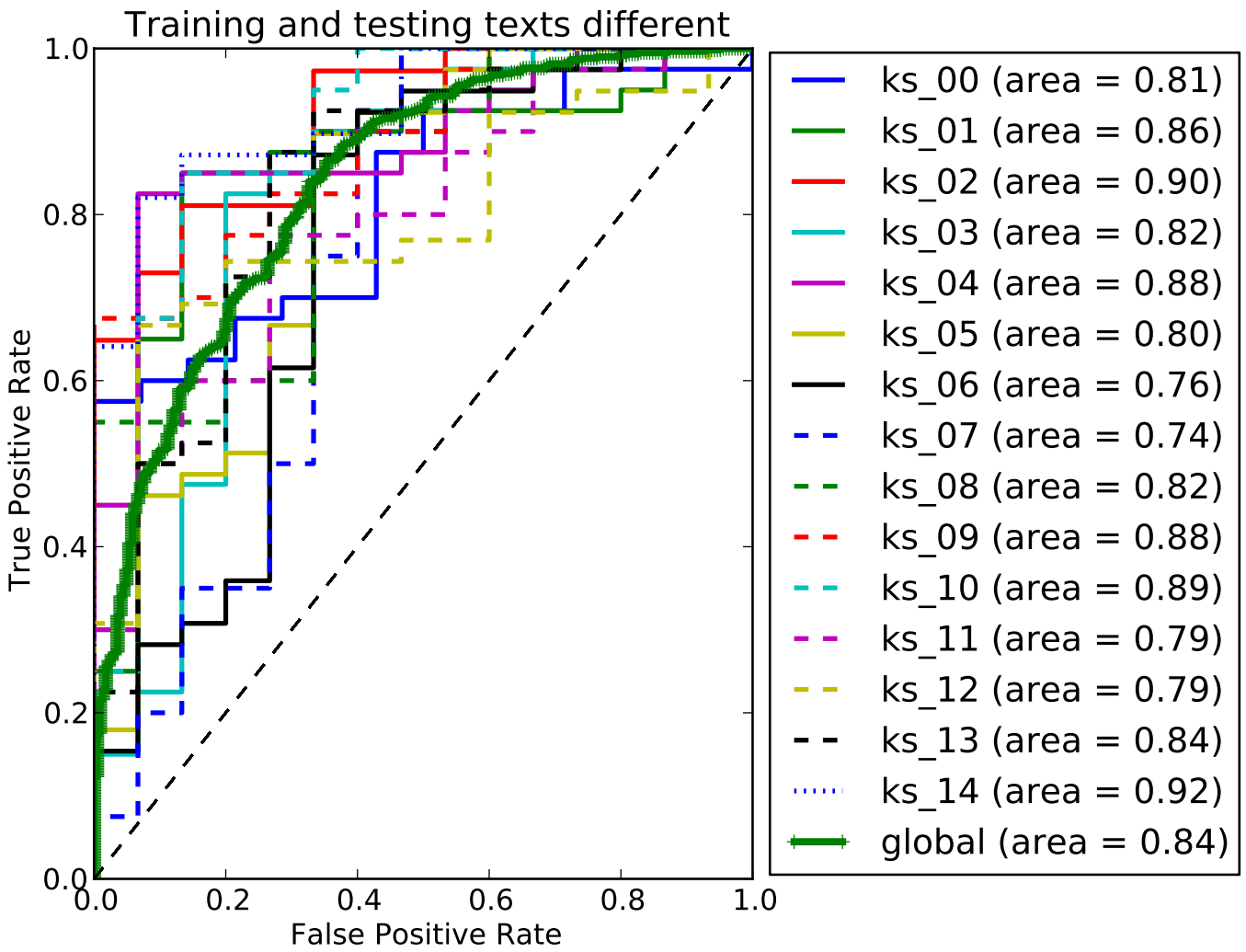
Scénario

1. Apprentissage avec tous les exemples sauf ceux d'une phrase
2. Reconnaissance avec un exemple de la phrase exclue

Chiffres

- 15+1 sous-scénarios
- 14*54=756 exemples pour les 15 premiers sous-scénarios

Résultats 2 : reconnaissance sur phrase inconnue



Scénario 3 : utilisation de plusieurs phrases inconnues

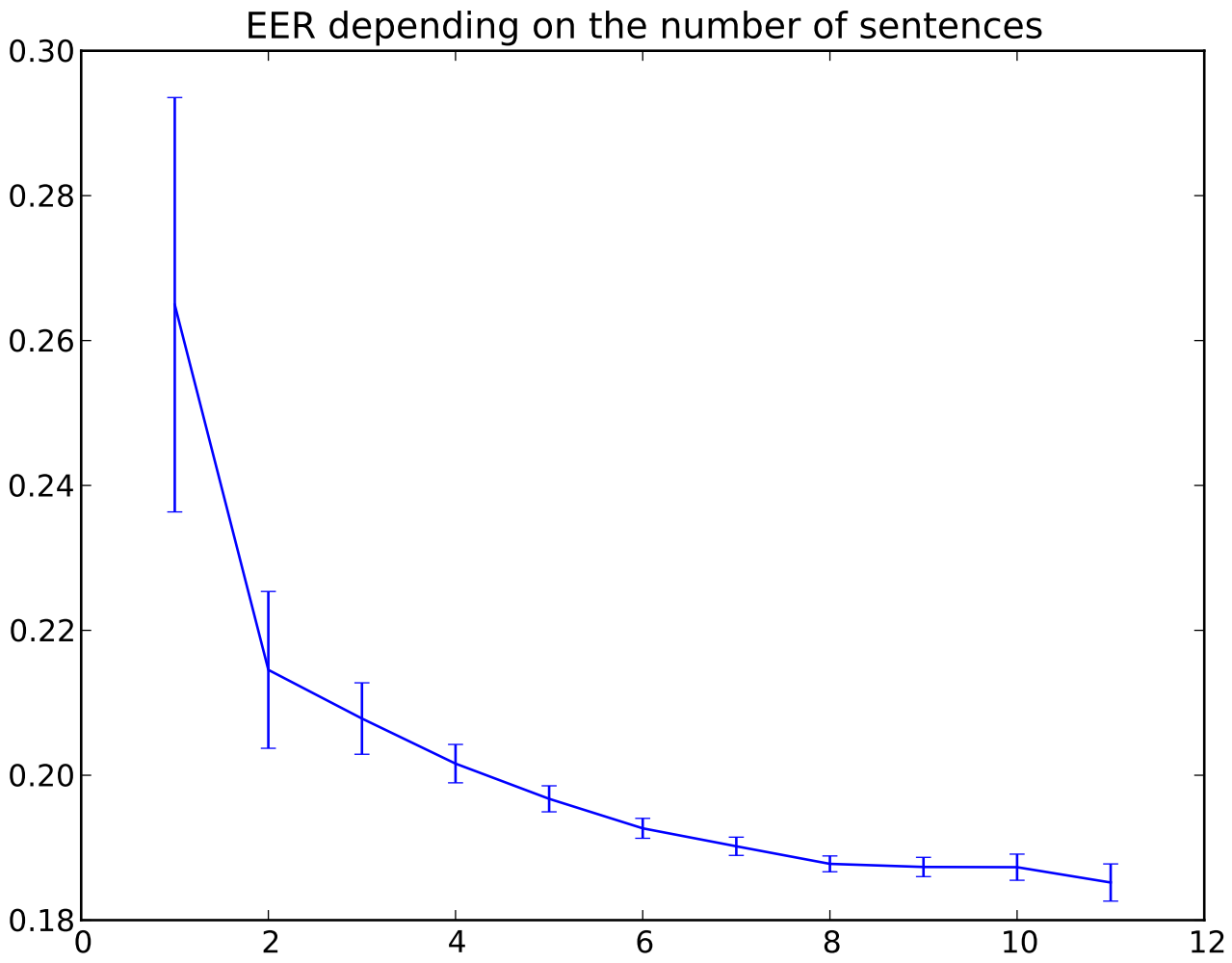
Scénario

- Reconnaissance après la saisie de plusieurs phrases
- Utilisation des scores du scénario 2
- Calcul de toutes les combinaisons
  - Nombre de phrases disponibles
  - Nombre de phrases utilisées
- Fusion = moyenne arithmétique des scores

Chiffres

- Fusion des scores pour chaque utilisateur
- C_N^{15} scores par utilisateur pour N phrases

Résultats 3 : utilisation de plusieurs phrases



Sommaire

Objectif

Contexte

Méthode développée

Protocole expérimental

Résultats et discussion

Conclusion

Conclusion

Avant

- Dynamique de frappe capable de
  - Vérifier l'identité
  - Détecter le genre par texte fixe
  - Déterminer les émotions

Maintenant

- Possibilité de détecter le genre par texte libre
- Performances de reconnaissance de 75% à 80%

Futur

- Amélioration des performances d'authentification
- Détermination d'autres informations (age ?)

Conclusion

Avant

- Dynamique de frappe capable de
  - Vérifier l'identité
  - Détecter le genre par texte fixe
  - Déterminer les émotions

Maintenant

- Possibilité de détecter le genre par texte libre
- Performances de reconnaissance de 75% à 80%

Futur

- Amélioration des performances d'authentification
- Détermination d'autres informations (age ?)

Conclusion

Avant

- Dynamique de frappe capable de
  - Vérifier l'identité
  - Détecter le genre par texte fixe
  - Déterminer les émotions

Maintenant

- Possibilité de détecter le genre par texte libre
- Performances de reconnaissance de 75% à 80%

Futur

- Amélioration des performances d'authentification
- Détermination d'autres informations (age ?)

Merci pour votre attention

http://www.epaymentbiometrics.ensicaen.fr/