

PROJET ANALYSE DE DONNEES SEQUENTIELLES

Objectif :

- Proposer un système de reconnaissance d'écriture.
- Aller plus loin que ce qui a été vu en TP.
- La nouveauté par rapport à ce qui a été vu en TP est plus importante que la performance du système.

Suggestions :

- Exploration des modèles de Markov discriminants (de type CRF)
- Convolution + réseau récurrent
- Combinaison de modèles
- Modèle linguistique

Les données :

Petits mots en français, répartis en 2 sous-ensembles : une base d'apprentissage et une base d'évaluation.

Soutenance :

- **Exécution d'une base de test** : prévoir une fonction d'évaluation pour analyser une base de test qui vous sera transmise le jour de la soutenance. Le chargement de cette base de test sera identique à celui des bases d'apprentissage et d'évaluation. Il vous faudra sauvegarder les paramètres du modèle (il ne faut pas ré-apprendre le modèle le jour de la soutenance!) et les charger pour évaluer les données. On évaluera les 2 métriques suivantes : le *Character Error Rate* et le *Word Error Rate*.
- **Présentation** (7 à 10 min) : Vous présenterez le système choisit, justifierez le choix des hyper-paramètres. Vous montrerez que vous maîtrisez le code de votre modèle (en présentant l'architecture du système et/ou présentant des points clés du code) et présenterez des résultats sur les bases d'apprentissage et de validation.
- **Questions** (3 à 5 min) : le jury vous posera des questions pouvant être théorique et pratique sur votre travail.