Plan du manuscrit

# Introduction

1. Motivation
   1. Motiver l’intérêt des séries temporelles dans les applications aujourd’hui
      1. Données de plus en plus présentes dans de nombreux domaines
      2. Les séries temporelles sont impliquées dans des problèmes de classification, régression et clustering
      3. Pourquoi sont-elles challenging ? (délais, dynamique)
      4. On fait face à la fois, à un problème de small et big data
2. Contribution de la thèse
   1. Définition d’un nouvel espace de représentation : la représentation par paires
   2. Apprentissage d’une métrique multimodale et multi-échelle en vue d’une classification kNN à vaste marge.
   3. Transposition du problème d’apprentissage de métrique dans l’espace des paires
   4. Comparaison de la méthode proposée avec des métriques classiques sur un vaste jeu de données (30 bases) de la littérature dans le cadre de la classification univariée de séries temporelles
   5. Extension du framework d’apprentissage de métrique au problème de régression de séries temporelles univariés
   6. Extension du framework d’apprentissage de métrique au problème de classification/régression de séries temporelles multivariés.
3. Organisation du manuscrit
   1. Présenter les différents chapitres
4. Notations
   1. Donner l’ensemble des notations du manuscrit sur une page séparée

# Chap. 1 : Etat de l’art des algorithmes de Machine Learning

1. Chapeau introductif
   1. Hypothèse : considérer les séries temporelles comme des vecteurs de données statiques.
   2. Appliquer les méthodes classiques de Machine Learning
2. Algorithme
   1. kNN
   2. SVM
      1. Principe général
      2. Forme primale
      3. Forme duale
3. Protocole d’apprentissage
   1. Framework d’apprentissage (train, validation, test)
   2. Cross-validation
4. Evaluation du modèle
   1. Mesure de l’erreur en classification
   2. Mesure de l’erreur en régression
5. Limites des méthodes classiques
   1. Limites : l'aspect ordonné des données temporelles n'est pas pris en compte (dynamique, fréquence, etc.)
   2. Intro à la partie suivante : Or l'ensemble de ces algorithmes (kNN, SVM) reposent sur des notions de calcul de distance ou de similarité. Considérons maintenant l'objet série temporelle et rappelons la notion de distance entre séries temporelles

# Chap. 2 : Etat de l’art des métriques pour les séries temporelles

1. Chapeau introductif
   1. Hypothèse : considérons la série temporelle comme étant un objet ordonné.
   2. On va définir dans la suite la notion d’alignement et de métrique pour des séries temporelles.
2. Alignement de séries temporelles
   1. Les données réelles peuvent présenter des délais. On peut les réaligner avec des techniques existantes.
   2. Définition d’un alignement
   3. Présentation de la DTW
   4. Variante de la DTW
   5. Dans la suite, on considère que les séries temporelles sont alignées
3. Propriétés d’une mesure de distance
   1. Donner les propriétés d’une mesure de dissimilarité (positivité, symétrie, distinguabilité)
4. Métriques basiques pour les séries temporelles
   1. Valeur (dE, Minkowski, Mahalanobis)
   2. Forme (corrélation de Pearson, Cort)
   3. Fréquentiel (dFFT)
5. Kernels pour les séries temporelles (à compléter)
6. Métriques combinées pour les séries temporelles
   1. Dans le cadre des données réelles, plusieurs composantes/modalités peuvent être impliqués (forme, valeur, fréquence).
   2. Certains travaux ont proposé des combinaisons : linéaire, exponentielle, sigmoïde
   3. Limites :
      1. Implique souvent 2 composantes
      2. La combinaison est définie a priori
      3. La combinaison est indépendante de la tâche d’analyse
         * Pour répondre à cela des auteurs ont proposé d’apprendre une métrique en vue de la tâche d’analyse considérée.
7. Metric learning : état de l’art
   1. Donner l’intuition sur la base des travaux de Weinberger
   2. Donner la terminologie (target, imposter, push, pull)
   3. Donner la formalisation du problème
   4. Limite :
      1. On apprend une distance de Mahalanobis
      2. L’apprentissage ne prend pas en compte l’aspect multi-modal
8. Aspect multi-échelles des métriques
   1. Dans le cadre de la classification, on peut avoir des données où l’information qui permet de discriminer une classe d’une autre n’est pas globale mais est localisée sur une partie du signal
   2. Les métriques basiques (valeur, forme, fréquentiel) tel que défini plus haut considère l’intégralité du signal
   3. On propose des métriques locales par dichtomie du signal

# Chap. 3 : Metric Learning multi-modal et multi-scale : formalisation

1. Chapeau introductif
   1. Formaliser le problème mathématiquement
2. Représentation par espace de paires
   1. Expliquer le processus de projection dans l’espace des paires
   2. Donner une interprétation et les warnings de l’espace des paires
      1. Proximité à l’origine
      2. Proximité de 2 points dans l’espace des paires
      3. Norme dans l’espace des paires
3. Les différentes variantes du Metric Learning multi-modal et multi-scale
   1. Résolution LP
   2. Résolution QP
   3. Résolution approximé par un SVM 🡪 Expliquer que dans la suite, pourquoi on va préférer le SVM
      1. Cadre connu
      2. Utilisation de librairies connus de Machine Learning
      3. Extension directe aux métriques non linéaire via le Kernel

# Chap. 4 : Metric Learning multi-modal et multi-scale : Mise en œuvre

1. Chapeau introductif
   1. Donner les étapes principales de l’algorithme avec l’intuition
2. Projection dans l’espace des paires
   1. Expliquer la projection
   2. Expliquer la normalisation log
3. Stratégie M-NN/M-diff
   1. Expliquer les différentes stratégies possibles (kNN/kdiff, MNN/Mdiff, kNN/All)
   2. Expliquer pourquoi on choisit M-NN/M-diff
4. Normalisation par les radius
   1. Expliquer le problème de non-homogénéité des voisinages
   2. Expliquer la normalisation par le radius
5. Apprentissage de la métrique
   1. Apprentissage via un SVM
6. Définition de la mesure de dissimilarité (à faire)
7. Extension à la régression (à faire)
8. Extension au cadre multivarié (à faire)

# Chap. 5 : Application aux données

1. Chapeau introductif
2. Présentation des données
3. Protocole expérimentale
4. Résultats
5. Discussion

# Conclusion et Perspectives

1. Bilan des apports de la thèse
2. Perspectives
   1. Multi-pass Metric Learning
   2. Kernel pour la résolution du problème QP
   3. Utilisation de la distance apprise dans d’autres algorithmes de Machine Learning (Arbre) 🡪 interprétabilité
   4. Utilisation d’autres distances (wavelet)
   5. Apprentissage locale de la métrique

# Annexes

1. Détail des bases de données
2. Présentation des solveurs