

## IA (SYMBOLIQUE ET COMPUTATIONNELLE) – SE

### Objectifs

Ce projet destiné aux élèves ingénieurs en deuxième année a comme objectif d'approfondir les matières enseignées dans le pôle cognitif : Intelligence artificielle, Intelligence artificielle Computationnelle et Systèmes experts.

Le projet est composé de trois parties principales :

1. étude et utilisation des algorithmes d'apprentissage issus de l'intelligence artificielle, avec comme sous-parties :
  - (a) algorithmes de l'intelligence artificielle computationnelle (sous-symbolique), à savoir algorithmes génétiques, réseaux de neurones artificiels et machines à support vectoriel. On utilisera les logiciels JavaNNS (ou SNNS), Weka ainsi que le programme que vous avez réalisé pour les algorithmes génétiques pendant votre projet du 1er semestre.
  - (b) algorithmes de l'intelligence artificielle symbolique et, particulièrement les arbres de décision et les règles d'associations. On utilisera le logiciel Weka.
  - (c) synthèse des résultats de deux sous-parties précédentes;
2. mise en œuvre d'un système expert (en utilisant la synthèse précédente des résultats) pour résoudre le problème choisi. Le système expert doit contenir un ensemble de faits et de règles organisés en modules y compris un module de traitement de données et de connaissances incertaines. Le système doit être implémenté en CLIPS.
3. conclusion générale.

## Outils & Logiciels utilisés

Les élèves utiliseront les logiciels et les programmes suivants : Weka, JavaNNS (ou SNNS) & CLIPS.

N.B.- Pour la partie intelligence artificielle computationnelle, les élèves sont libres d'utiliser d'autres logiciels que ceux proposés, s'ils le souhaitent.

## Étapes de réalisation

Le projet se fera en groupe de quatre personnes. Chaque groupe choisira un ensemble de données décrivant un domaine d'expertise. Nous décrivons dans la suite les trois parties du projet.

### Partie Apprentissage

La **première partie** du projet comportera les étapes suivantes :

1. **Analyse préliminaire du problème et des besoins** : est-ce un problème d'apprentissage *supervisé* ou *non supervisé* ? Dans le cas d'un problème supervisé, on peut envisager de le traiter sans considérer les classes. À l'issue de cette étude, une première idée des méthodes symboliques et sous-symbolique à appliquer est à envisager.
2. **Étude approfondie des données** : Quelle est la nature des données ? S'agit-il des valeurs qualitatives ou quantitatives, discrètes ou continues ? Cette étude permettra d'avoir plus de précision sur le choix des méthodes à appliquer.
3. **Pré-traitement des données** : cette étape est composée de plusieurs phases qui permettent de préparer les données aux différentes méthodes envisagées :
  - (a) Pour pouvoir appliquer les méthodes sous-symboliques il faut *normaliser les données d'entrée* afin d'obtenir des bons résultats ;
  - (b) Pour pouvoir appliquer les méthodes de règles d'association, il faut penser à *discrétiser les attributs continus*. Une façon de faire

et d'y appliquer J48 (via Weka) ce qui permet de découvrir des ensembles finis d'intervalles.

4. *Choisir la méthode de validation* : pour pouvoir comparer les résultats des différents logiciels dont on dispose (Weka, JavaNNS, etc.), il faut se fixer un ensemble d'apprentissage, et un ensemble de test. L'ensemble d'apprentissage permet de construire le modèle (l'arbre de décision ou le réseau neuronal, etc.). L'ensemble de test permet de tester le modèle déjà construit sur des nouvelles données. Nous proposons une méthode expérimentale qui s'approche de la validation croisée. Diviser l'ensemble de données en trois sous ensembles A,B,C ayant la même cardinalité, chaque ensemble doit contenir la même distribution par rapport aux différentes valeurs de la classe. On effectue l'apprentissage et le test trois fois :

- (a) On apprend à partir de A+B et on teste sur C.
- (b) On apprend à partir de A+C et on teste sur B.
- (c) On apprend à partir de B+C et on teste sur A.

On reporte ensuite les trois résultats indiqués en terme de pourcentage d'erreurs ainsi que leur moyenne.

5. *Appliquer les différentes méthodes*<sup>1</sup>, analyser et commenter le modèle. Remarquer bien que le choix judicieux des paramètres pour les méthodes de l'IA computationnelle nécessite un nombre important d'expérimentations pour obtenir les meilleurs résultats en terme de convergence.
6. *Effectuer une étude comparative entre les résultats* obtenus par les différentes méthodes et algorithmes utilisés. Expliquer les résultats.

## Partie Raisonnement

La **deuxième partie** du projet consiste à réaliser un système expert sur le domaine que vous avez choisi. On peut envisager deux possibilités (on peut même les mixer):

---

<sup>1</sup>Il faut expérimenter obligatoirement deux modèles de réseaux de neurones, les arbres de décisions, les règles d'association ainsi que SVM

1. Utiliser les règles ou une partie des règles apprises à partir des arbres de décision et/ou des règles d'association pour fabriquer un système expert. Pour pouvoir tester votre système, les données doivent être transformées en faits. Il est donc nécessaire de prévoir un module de traduction (données - faits).
2. Vous pouvez ajouter des règles que vous jugez nécessaires si vous avez une connaissance sur le domaine du problème.

Le système expert doit être testé et les résultats doivent être commentés et comparés avec les résultats obtenus dans la première partie du projet.

## **Partie Conclusion générale**

La **troisième partie** du projet est consacrée à une discussion qui explique l'intérêt de chaque méthode utilisée. De plus, on doit répondre, entre autre, aux questions suivantes :

- Y-a-t-il de redondance et/ou de complémentarité entre les différentes méthodes que vous avez choisies ?
- Quels sont les perspectives de votre travail ?
- Pourrait-on envisager le travail autrement pour améliorer certains aspects ? Lesquels ?

## **Modalités**

Chaque groupe doit fournir un livrable compsé comme suit :

1. Un fichier avec les données utilisées après prétraitement avec une présentation de la méthode de prétraitement utilisé. Il faut aussi fournir les fichiers des ensembles A, B et C de données utilisées pour l'apprentissage et le test (cf. supra).
2. Le programme utilisé pour l'algorithme génétique et les valeurs des paramètres, avec un manuel d'utilisation.
3. L'architecture des réseaux de neurones et les valeurs des paramètres.

4. Les valeurs des paramètres pour la machine à support vectoriel.
5. Les résultats de l'application de la méthode des arbres de décision.
6. Les résultats de la méthode des règles d'association.
7. La base de faits et des règles du système expert.
8. Le programme en Clips du système expert avec un manuel d'utilisation.
9. Un rapport écrit en  $\text{\LaTeX}$ , qui doit contenir
  - (a) Une description de la réalisation de la première partie du projet ; chaque étape doit être bien expliquée et détaillée. Présentation et discussion des résultats.
  - (b) Une description de la réalisation de la deuxième partie du projet ; chaque étape doit être bien expliquée et détaillée. Présentation et discussion des résultats.
  - (c) Une synthèse des résultats des première et deuxième parties.
  - (d) Le raisonnement suivi pour construire la base de faits et celle de règles du système expert.
  - (e) Présentation et discussion des résultats du système expert.
  - (f) Une conclusion finale détaillée sous forme de discussion qui reprend les éléments de la troisième partie ci-dessus.

N.B. Si une de parties de traitement, à savoir algorithmes génétiques, réseaux des neurones, machines à support vectoriel, arbres de décision, règles d'association, système expert est absente, la note sera 0 (zéro) pour la totalité du projet. Vous êtes donc tenus à travailler toutes les parties du projet de manière exhaustive.

## **Livrables – Délais**

Il ya aura un seul livrable.

Délai : vendredi 15 avril 2011.