Artificial Intelligence - Assignment 2

Theme

This work intends to use computer learning to perform one of the following tasks:

1. Teach an agent to solve a puzzle based on reinforcement learning;
2. Teaching an agent to play a board game, based on reinforcement learning;
3. Build classification models, based on supervised learning algorithms, to a specific dataset;
4. Build regression models, based on supervised learning algorithms, to a specific dataset;
5. Apply Natural Language Processing techniques and classification models to address a classification task in natural language.

In works 1 and 2, the game developed in work 1 may be used or a new game implemented among those suggested. In works 3, 4 and 5, datasets chosen from a suggested set.

Themes 3 and 4: Supervised Learning

For supervised learning jobs, it is intended to learn to classify or predict values for the concept under analysis. Initially, an exploratory analysis of the data (distribution classes, distribution of values in each attribute, etc.).

Different methods must be implemented / applied learning algorithms, comparing their performance using appropriate metrics (errors obtained throughout the learning process, analysis of the confusion matrix, quality measures of the classification / forecast obtained such as precision, recall, accuracy, F-measure) and the time taken to perform the training / test.

Supervised learning includes performing the following procedures: analyzing the set of data in order to verify the possible need for pre-processing and data transformation, identification of the concept (s) to learn (dependent variable (s)) and definition of the training set and set of test, selection and parameterization of the learning algorithm, and learning assessment obtained by measuring the results in the training and test sets.

At least **3 supervised learning algorithms**

- Decision Trees
- Neural Networks
- K-Nearest Neighbor
- Support Vector Machines, etc

based on a data set made available for this purpose.

You should **compare the results** obtained with the different algorithms of learning and its parameterization, using appropriate tables or graphic elements, through the use of a tool for

this purpose such as Seaborn or **Matplotlib**. In both themes, classification and regression problems can be solved, using supervised learning methodologies.

However, the work on theme 3 should be focused essentially in solving classification problems and those of theme 4 in regression problems. In both cases, the suggested datasets should be the basis of the work although they can be complemented with additional information / datasets

Programming language

Any programming language and development system can be used, including of languages, **Python**, C ++, Java, C #, among others. Libraries or tools can be used specific for the implementation of supervised learning experiences such as **Scikit Learn**, Rapid Miner, Weka or R, among others.Reinforcement learning systems can be used such as ML-Agents, Open AI Gym, etc.
However, the following systems are suggested for the various works:

- Themes 1 and 2: Unity and ML-Agents;
- **Themes 3 and 4: IPython, Jupyter Notebook and Scikit-learn**.
- Theme 5: IPython, Jupyter Notebook, Scikit-learn, NLTK, StanfordNLP.

Constitution of Groups

The groups must be composed of 2 or 3 students. Individual or compound groups are not accepted by 4 students. Groups can be made up of students from different classes (maximum of two different classes) but all students must be present at the follow-up sessions and presentation / demonstration of the work. The constitution of groups composed of students from different classes is not advised, given the logistical difficulties of carrying out work that this can cause.

Checkpoint

Each group must submit a short presentation in Moodle (max. 5 slides), in PDF format, which will be used in class to analyze, together with the teacher, the progress of the work.

The presentation must contain:

1. specification of the work to be carried out (definition of the learning problem (s) computational to be solved);
2. work related to references to works found in the research (articles, web pages and / or source code);
3. description of the tools and algorithms to be used at work;
4. implementation work already done.

Final Delivery

Each group must submit three files in Moodle:

1. a **presentation**, in PDF format
2. an **article**, in PDF format, following the model to be made available in Moodle of the discipline,

3. and **the code** implemented, properly commented, including a “readme.txt” file with instructions on how to compile it, run and use.

Based on the submission submitted, students **must conduct a demonstration** (about 15 minutes) of the work, in the practical class, or in another period to be designated by the teachers of the discipline.
The presentation should focus:

- in addition to the elements already described for the interim delivery (checkpoint)
- all relevant aspects related to data processing
- developed models
- their evaluation and comparison
- using appropriate graphic elements.

Dataset

Below you can find the dataset:

1. Kaggle
2. Dataset Search

Dates

1. Checkpoint: **11/05/2020**
2. Final Delivery: **27/05/2020**
3. Presentation: **28-29/05/2020**