CSE 352 Artificial Intelligence
Fall 2018
Professor Anita Wasilewska

http://www.cs.sunysb.edu/ ~ cse352

Meets Monday, Wednesday 2:30 - 3:50pm

Place JAVITS 109

Professor Anita Wasilewska e-mail
anita@cs.sunysb.edu,
Office phone number: 632 8458
Office location: New Computer Science Department building, office 208

Office Hours Monday, Wednesday 12:30 pm - 2:00 pm
Wednesday 7:10 pm - 8:00 pm and by appointment.

Place New Comp. Science Building, Room 208, telephone: 2-8458

Textbook

The Essence of ARTIFICIAL INTELLIGENCE, Alison Cawsey, Prentice Hall, 1998.

This is a short condensed book (not expensive!) and not very technical

We will cover in detail first 3 chapters (plus my lecture notes for technical details) and chapter 7.

It is your responsibility to read all of the book and all other materials published on the course website

Lecture Notes posted on the course webpage EXTEND the material from the book providing TECHNICAL details and are the **major source** for the course.

Additional Textbook

DATA MINING - Concepts and Techniques

Jiawei Han and Michelle Kamber, Morgan, Kauffman Publishers, 2006, 2010, 2013

Here is the author webpage: https://hanj.cs.illinois.edu/

You can download text and slides for CHAPTER 6: Classification and Prediction at

http://web.engr.illinois.edu/ hanj/bk2/slidesindex.htm

Our Lecture Notes cover almost all of this chapter.

Class attendance is the most important, as the Lectures and **Lecture Notes** serve as an **extra textbook** for the course and are integral and as important part of the course design

I will **check class attendance** by giving and collecting answers (almost each class) to small questions connected with the lecture; you will get **1-2 extra credit points** for your answers.

Grading

During the semester you have to complete the following.

1. Quizzes - (30pts)

There will be 2 Quizzes (20 minutes), 15 points each

Each quiz will consist of **1 - 2** questions only:

NO make-up for quizzes.

I might give some **additional quizzes** for extra credit

2. Midtem - (65 pts)

3. Project (40pts);

4 Final] (65pts)

5. Extra Credit I will give during the class small questions for extra credit

You can earn up to **20 extra points** during the semester

Quizzes and Tests are **closed book** (and cell phones) examinations

None of the grades will be curved.

Homeworks I posted 4 Homework Assignments AND Homework Solutions. I encourage students to SOLVE homework problems first- and then to compare their solutions with those posted. Quizzes and TESTS will contain problems very similar to the Homework Problems.

Final grade computation During the semester you can earn **200pts** plus **extra credit**.

The grade will be determined in the following way: # of earned points divided by 2 = % grade.

The % grade is translated into a letter grade in a standard way i.e.

$100 - 95$ % is A, $94 - 90$ is $A-$,
$89 - 86$% is $B+$, $85 - 83$ % is B, $82 - 80$ % is $B-$,
$79 - 76$ % is $C+$, $75 - 73$ % is C, $72 - 70$ % is $C-$,
$69 - 60$ % is D range and F is below 60%.

Tentative Quizzes and Tests schedule

Fall Break October 8 -9

Q1 Wednesday, October 10

MIDTERM Wednesday, November 7 in class

PROJECT DATA = Project Homework - **10 extra points** due Wednesday, November 20 - or any date before November 20

Thanksgiving Break November 21 - 25

PROJECT due Wednesday, November 28

Q2 Wednesday, December 5

Last Class (Review for Final) Monday , December 10

FINAL time and place as scheduled by University

PROJECT DESCRIPTION

Project **goal** is to use Internet based Classification Tools to build two type classifiers: **descriptive** and **non-descriptive**. Discuss the results. **Compare** these two approaches on the basis of obtained results.

1. Descriptive Classifier

Use a **Decision Tree** tool to generate sets of **discriminant rules** describing the content of the data.

Use WEKA:
http://www.cs.waikato.ac.nz/~ ml/weka/index.html)

2. Non-Decsriptive Classifier

Use **Neural Networks** tool to build your Classifier

Use WEKA or a tool of your choice. Describe specifics of your tool in a way that makes your report comprehensible for others.

Here are some tools suggestions:

http://www.mathworks.com/products/neural-network/?requestedDomain=www.mathworks.com

http://www.simbrain.net/

PROJECT DATA is provided on the course web page.

This is a real life classification data with TYPE DE ROCHE (Rock Type) as a CLASS attribute. There are 98 records with 48 attributes and 6 classes.

Classes are:

C1 : R. Carbonatees AND R. Carbonatees impures

C2 : Pyrate

C3 : Charcopyrite

C4 : Galene

C5 : Spahlerite

C6 : Sediments terrigenes

Most important attributes (as determined by the expert) are: **S, Zn, Pb, Cu, CaO+MgO, CaO, MgO, Fe2O3**

This is a real life experimental data and it contains a lot of missing data (no value).

The project has to follow the following steps of **DM Process** to build different classifiers defined to three experiments.

S1: Data Preparation that includes attributes selection, cleaning the data, filling the missing values, etc... to build Project DATA - **PD**.

S2: Data preprocessing

1. For the Decision Trees **Descriptive Classifier** you use 2 methods of data discretization to the Project Data **PD** creating two data sets: **PD1** and **PD2**. Describe which methods you used.

2. For the Neural Network **Non -descriptive Classifier** use the Project DATA - **PD** and your tool method of normalization of your choice. Specify which.

Building Classifiers

For each sets of data **PD1**, **PD2** (for Decision Trees), and **PD** (for Neural Networks) perform the following **Experiments 1- 3**.

For each Experiment **compare** the resulting **Descriptive Classifiers** with each other and compare each **Descriptive Classifier** with the resulting **Non-Decsriptive Classifier**.

Experiments 1- 3

Experiment 1 : use the preprocessed data to perform a **full classification** (learning), i.e. build a classifier for all classes **C1- C6** simultaneously.

Experiment 2 : use all records to perform a **contrast classification** (contrast learning), i.e. a classification contrasting class **C1** with a class **notC1** that contains other classes.

Experiment 3 : repeat Experiments 1, 2 for all records with the **most important attributes** as defined by the expert.

Write a detailed Project Description with methods, motivations, results and submit via Blackboard and e-mail to TA

It is a TEAM PROJECT - we will help you to form TEAMS

Here are some suggestions for **Neural Network** tools:

http://www.mathworks.com/products/neural-network/index.html

https://pypi.python.org/pypi/neurolab

RAPID MINER https://rapidminer.com/

Course Content

The book is very thin. It is a short overview of major areas of AI. I supplement it with LECTURE NOTES for detailed information. In particular we will cover all or majority of the following book chapters and LECTURE NOTES subjects, **not always in the order they are listed**

The course will to cover in depth the majority of the following subjects.

Chapter 1 AI history and applications. Lecture Notes

Chapter 2 Knowledge Representation and Inference.

Propositional and Predicate Calculus is presented in detail in Lecture Notes are on the douse webpage

Chapter 2- Logic Details for Chapter 2: Overview of Propositional and Predicate Logic; Predicate languages and basic Laws of Quantifiers; Predicate Logic Arguments in **Lecture Notes**

Automated theorem proving DETAILS for Chapter 2: Propositional Resolution. EXTRA HANDOUTS and Lecture notes.

Automated theorem proving 2 DETAILS for Chapter 2: Predicate Resolution. Lecture notes distributed in class.

Chapter 3 Expert Systems. Overview of EXPERT SYSTEMS Design and Technology.

Lecture notes and EXTRA HANDOUT distributed in class.

Chapter 5 Natural Language Processing- reading assignment.

Chapter 7 Machine Learning - concentration on INDUCTIVE, or CLASSIFICATION Learning.

The book is not technical- we will use the extensive Lecture Notes

Lecture Notes : In particular we cover the following techniques.

Decision Trees - detailed algorithm on lecture slides posted on the web and intuitive introduction is in the book.

Neural Networks - detailed algorithm on lecture slides on the web and intuitive introduction in the book.

Genetic Algorithm - detailed algorithm on the lecture slides on the web and intuitive introduction in the book.

Association Analysis - Apriori Algorithm

Classification by Association

Chapter 8 Agent and Robots - readings

New Advances Published Lectures Slides for Reading

Deep Learning

Web Mining: an overview of methods and problems

Text Mining: an overview of methods and problems

Visualization and DM techniques

Natural Language Processing and DM techniques

ACADEMIC INTEGRITY STATEMENT (Adopted by the Undergraduate Council September 12, 2006)

Each student must pursue his or her academic goals honestly and be personally accountable for all submitted work. Representing another person's work as your own is always wrong. Any suspected instance of academic dishonesty will be reported to the Academic Judiciary. For more comprehensive information on academic integrity, including categories of academic dishonesty, please refer to the academic judiciary website at http://www.stonybrook.edu/uaa/academicjudiciary/

Stony Brook University Syllabus Statement If you have a physical, psychological, medical, or learning disability that may impact your course work, please contact Disability Support Services at (631) 632-6748 or http://studentaffairs.stonybrook.edu/dss/. They will determine with you what accommodations are necessary and appropriate. All information and documentation is confidential.

Students who require assistance during emergency evacuation are encouraged to discuss their needs with their professors and Disability Support Services. For procedures and information go to the following website: http://www.sunysb.edu/ehs/fire/disabilities.shtml