'INTELLIGENT SYSTEMS' –THE COURSE

Ana Madevska – Bogdanova Nevena Ackovska
Faculty of NaturalSciences and Mathematics - Institute of Informatics
Skopje, Macedonia,

ABSTRACT

The elective course 'Intelligent Systems' thought at Institute of Informatics in Skopje offers interesting and modern topics that include modeling the real world, Data Mining, Mechanical Robotics, Bioinformatics, i.e. pattern recognition, prediction. A new methodology of teaching and training students is presented in building modern Intelligent Systems.

I. INTRODUCTION

The elective course 'Intelligent Systems' represents totally different point of view on technology development, compared to any other course taught in the computer science curriculum.
The students become interested to understand the connection of the essential sciences (statistic, mathematics, physics, biology etc) with the Information technology. They are thought to understand the theoretical background of building different 'Intelligent Systems' (pattern recognition, prediction) and developing simulations and actual physical models (movement of intelligent robots). They also learn how to deal with the real problems that arise from the implementation of real-time systems.
This paper covers the methodology used in teaching the 'Intelligent Systems' course. Some examples of student work on this field are also presented. Furthermore, these systems can be extended to other useful applications.

II. KNOWLEDGE BACKROUND

The teaching material in this course mainly is divided in four parts – Machine Learning techniques, Bioinformatics, Robotics and Data mining. The first part – ML techniques are used in problem solving and building intelligent systems within the other three parts.

2.1 Machine Learning

The first part of the course is about the Machine Learning techniques. Machine learning usually refers to the changes in systems that perform tasks associated with artificial intelligence. Such tasks involve recognition, diagnosis, planning, robot control, prediction. Computer Science students have solid mathematical background and for the first time, in this module they understand the potential of the mathematical disciplines such as numeric optimization, statistic, vector spaces when used in solving practical problems.
We teach ML techniques mainly for dealing with the pattern recognition problems. We used Artificial Neural Networks [7] and Support Vector Machines. It was revealing for the students to understand the other, non-standard way of dealing with the pattern recognition issues. They were confronted with problems from science, medicine, linguistics, trading. They were taught to see the similarities in this kind of problems as well as the differences that characterize the way of dealing with them (adjusting the parameters in the ML techniques). The performances of the given tasks in the before mentioned areas using ML techniques were much higher comparing to the standard statistical procedures used over the same problems.

2.2 Bioinformatics

The second part of the Intelligent System course is Bioinformatics. Our students who posses' sound mathematical background can find a great challenge in using mathematics' and informatics' tools in discovering meaningful sequences in the genetic material [2], [3]. There are many ways to approach the problem of understanding the processes in the biological cell. For the purposes of 'Intelligent Systems' course, we are mainly concerned with the basic string processing aspects of DNA and RNA, and the modeling of the genetics processes is done using linguistic approach [1].
Defining the terms bioinformatics and computational biology is not necessarily an easy task. Computational biology and bioinformatics are multidisciplinary fields, involving researchers from different areas of specialty. In order to be a good bioinformatitionist, it is important for the students to understand the terminology and basic processes behind the biological problems. Many interesting problems arise out of sequence analysis. There are two different types of biological sequences studied in this class: DNA/RNA and amino acids (proteins) [8], [9].
Deoxyribonucleic Acid (DNA) is the basis for the building blocks encoding the information of life. A single stranded DNA molecule, called a ***polynucleotide*** or ***oligomer***, is a chain of small molecules called nucleotides.

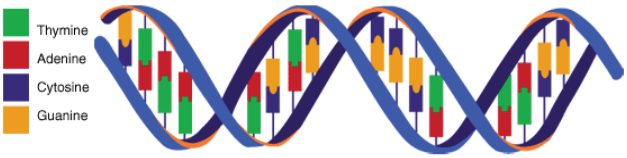

Figure 1: There are four different nucleotides, or bases: adenine (A), cytosine (C), guanine (G) and thymine (T)

It was important for the students to understand that by stringing together a simple alphabet of four characters together we can get enough information to create a complex organism!

2.3 Robotics

Robotics for educational purposes is becoming very popular in the previous years. Almost every curricula of Computer science has an advanced robotics course [13]. It is becoming so popular, that many student robotics competitions are being held worldwide [15, 16].The robotics part of the 'Intelligent Systems' course teaches students to deal with the constraints of the real world. They had several tasks to deal with the direct and the inverse kinematics in order to move real robots. For the needs of this course we used the robot Lynx 5 (Fig. 2) designed by Lynxmotion Inc [14]. The comparison of the use of theoretical algorithms for movement and the real movement proved crucial for the students to understand the problems that arise when dealing with concepts like matter, gravitation, part assembling etc. Since the curricula on the Institute of Informatics primarily is based on software techniques, the knowledge gained when dealing with hardware is a unique opportunity for our students to gain capabilities of creating embedded software for real life applications.

Figure 2: Lynx 5 robot for educational use

The basic idea behind these projects was to embed the software based on sensor inputs (voice and video) in movement of the robot arm Lynx 5. The projects created included chess playing (Fig.3), video motion detection, sound forced movement, robot dancing etc

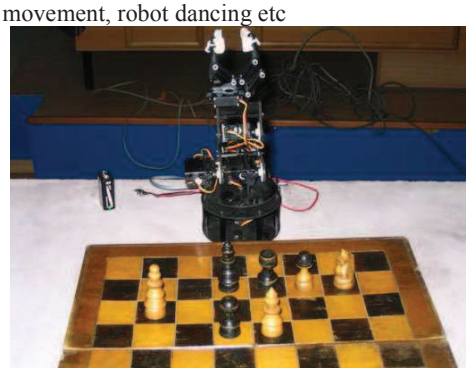

Figure 3: Chess playing using the robot arm Lynx 5

2.4 Data Mining

Another part that is taught in the course, in a smaller scale, is Data Mining.The students were very interested to understand the main principles of the knowledge discovering in data bases process- selection of data, preprocessing, transformation, data mining, Interpretation and evaluation and finally, discovering new knowledge. It was intriguing for them to see the differences between the data base queries and searching data bases for new knowledge. We have considered real world problems from big mobile companies (customer churn) and Customer Relationship Management. Also, more scientific problems were considered, as exploring the web logs – practical issue is predicting the next cash address, or personalization in offering wed addresses in the persons cache.

III. METHODOLOGY

Regardless of the topic (Bioinformatics, Data Mining, Robotics), the main methodology is to confront the students with a problem (pattern recognition, prediction), try to solve it in the traditional way (developing algorithm, using statistical procedures) and then comparing with the outcomes when state-of the-art machine learning techniques were used. Usually traditional approach is not efficient to find common characteristics in the given set of data – as in pattern recognition problems. The students were presented with papers from the main results in the field of ML, and then they had several tasks to solve by themselves. The practical approach, not only theoretical view of the filed, has helped them to develop new dimension of using informatics tools and understand that hard problems from real world – fraud detection, cancer detection, presented by manipulation of large amount of data (modeling, prediction, pattern recognition) can be solved by using ML techniques.
As practical part of Data Mining is using the open source package Weka [11]. An excellent diploma work was done with this software, where different ML techniques were used in solving several toy problems. The main goal of this diploma work was achieved – to understand the power of the non-standard ML techniques in discovering common fitchers in large amount of data.

IV. TEACHING MATERIAL

4.1 LECTURES

The teaching material is recruited from different sources – text books, journal papers, conference papers, web pages with high level of credibility. Text books were used in understanding the theoretical background of the offered topics (mathematics, statistics, Bayesian models, numerical optimization). Journal and conference materials were considered when explaining the latest achievements in pattern recognition in Bioinformatics (gene, promotors recognition), Data Mining (discovering new knowledge in given data set), Robotics. As for the web sources, they were used to download software for using Machine Learning techniques, such as SVM light [10], Weka for Data Mining [11], and also for gathering material from real life pattern recognition problems, different benchmark data sets.

4.2 EXERCISES

The course 'Intelligent Systems' has 3 classes for exercises per week. The exercises part included basic algorithms for the thought technique, as well as user manuals for the used tools (Weka, Lynx 5 etc). Part of the exercises time is used for the explanation of the tool, and the rest of the time is used for a problem solving. The bioinformatics part of the problems included information search, finding similarities, finding exceptions in the DNA/RNA files, building protein files, RNA folding problem etc. The robotics part of the exercises included direct and inverse kinematics of the robot Lynx 5, signal input, signal transformation etc. Data Mining part included shop surveys, redesign of a shop in order to group the products by the way customers tend to select them etc.

V. PREVIOUS EXPERIENCE

Each year, (4 generations by 2007) the best students in the generation have attended the course 'Intelligent Systems'. They develop interest in different parts covered by the course. Each year, more students decide to do their BSc thesis in this area (statistical pattern recognition by Neural Networks, expert system in Entomology, audio signal processing, Pattern Recognition in genetics with Support Vector Machines, Data Mining with Weka). Usually, these theses are extensions of the individual work in the course – essays or their individual practical projects.

VI. CONCLUSION

Instead of conclusion, we can say that the last four years, since the 'Intelligent Systems' course is developed, there is an expansion in interest in the fields of Bioinformatics, Signal processing, data mining among the computer science students. That was not the case before establishing the Intelligent System course. Our students successfully continue their education in some topic covered by this course, by their own choice. They are pursuing or finished their MSc thesis in Bioinformatics, Machine learning in the well known University centers across Europe.

REFERENCES

[1] S. Bozinovski, G. Jovancevski, N. Bozinovska, "Minimal Cognition Systems: Current Metaphors" *Proceedings of the Sixth International Conference on Cognitive and Neural Systems,* Boston, 2002

[2] A. Madevska, D. Nikolic "Automatic Classification With Support Vector Machines In Molecular Biology", *Third Annual Conference on Cognitive and Neural Systems*, Boston, USA (1999)

[3] A. Madevska-Bogdanova, D. Nikolic, L. Curfs, "Classification of DNA sequences using modified SVM outputs", *Proceedings of 13- th International Conference of Artificial Neural Networks, ICANN /IJCNN, June, Istanbul, Turkey, 2003;*

[4] J. Berg, J. Tymoczko, L. Stryer, *Biochemistry*, Freeman and Company, 2002;

[5] S. Bozinovski. "Flexible manufacturing systems: A biocybernetic approach" in M. Vukobratovic, E. Popov (Ed.) *Proc. Symp on Robotics and Flexible Manufacturing, Moskow, 1986, p. 192-197*

[6] S. Bozinovski, B. Mueller, F. diPrimio " Biomimetic Autonomous Factories: Autonomous Manufacturing Systems and Systems Software" *GMD Report 115*, German National Center for Information Technology, Sankt Augustin, 2000;

[7] Bishop, C. M. Pattern Recognition and Machine Learning. Springer, 2006;

[8] B. Lewin, *Genes VII*, Oxford University Press, 1999;

[9] H. Lodish, A. Berk, L. Zipursky, P. Matsudaira, D. Baltimore, J.Darnell, *Molecular Cell Biology*, 4th edition W. H. Freeman and Company, New York, 2000;

[10] SVM light, http://svmlight.joachims.org/

[11] Weka, open source Machine Learning Software, http://www.cs.waikato.ac.nz/~ml/weka/

[12] N. Ackovska, A. Madevska-Bogdanova, "Teaching Bioinformatics to Computer Science Students", Proceedings of IEEE EUROCON 2005, Belgrade, Serbia and Montenegro, 2005, p 811-814

[13] S. Bozinovski, G. Jovancevski, N. Ackovska, "Distributed Interactive Robotics Classroom", Periodical in Optoelectronic Information – Power Technologies, No. 2(4), 2002, p. 5-10.

[14] Lynx5, http://www.lynxmotion.com/Category.aspx?CategoryID=2

[15] Botball http://www.botball.org/?gclid=CJnjyaeI9pACFSTaXgodPVJesQ

[16] IEEE SoutheastCon 2008 http://ewh.ieee.org/reg/3/secon/08/competition.html