

Merchant Venturers Building
Woodland Road,
Bristol, BS8 1UD
United Kingdom

ML-NET

ACAI-99

René MacKinney Romero

University of Bristol

·
·
·
·
·
·
·
·

ACAI-99

René MacKinney Romero

Introduction

ACAI summer school was a very instructive event. It allowed the participants to learn about the different approaches that exists in Machine Learning (ML), as well as recent progress and the shortcomings that have to be addressed in the near future. This report is divided into Lectures and Workshops. The focus of the Lectures was on theoretical aspects of ML while the Workshops showcased more applications of ML..

Lectures

Prof. Tom Mitchell set the scene with general overview of the Machine Learning Field. He discussed in some depth learning trees and neural networks. Bayesian learning was viewed in greater depth and included an example on learning to classify text found on the Web. This is an area of great interest for research and applications and despite the fact that the example was a simple one it allowed him to show several interesting concepts. Prof. Mitchell finally talked about CES a programming language that is oriented to make learning simpler.

Dr. Maarten van Someren lecture was on applications of machine learning. This was a lecture devoted to talk about some successes and shortcomings of applications of machine learning. It included several projects as well as a survey of machine learning applications made in Holland some years ago. The survey gave two results. One positive, that the companies are in general satisfied with the results they obtain applying machine learning and a second negative, that the use of machine learning tools is still rather amateurly. A further point was made on the fact that most machine learning conferences focus on the technique or tool used rather on the application. Therefore making hard for people to share their experiences on a given field.

Prof. Ryszard S. Michalsky talked about concept learning and induction. This lecture was more theoretical and presented a framework for induction and formation of concepts. It presented a formal definition of concept learning and several issues surrounding this problem such as complexity, noise and others. Prof. Michalsky also presented AQ learning which is a learning approach on which he has been working for several years. He presented an algorithm to learn using AQ learning and demonstrated the latest incarnation of his learning system. He finally presented a framework for concept learning including a hybrid approach with genetic algorithms.

Prof. Yves Kodratoff gave a lecture of ML vs. KDD. This was a very interesting lecture that centred in exploring the similarities and differences between machine learning and data mining. After the lecture I got the impression that the main difference is the measure that drives the search. Several measures of "interest" were presented in depth, each measure was presented along an informal explanation of what it means.

Prof. Pat Langley gave a recount of successes of machine learning in discovering new scientific facts. Prof. Langley argues that these stories share the intervention of scientists in several levels and not only as users of a system and for a system being able to produce new scientific knowledge this should happen. He also presented some results in probabilistic concept hierarchies.

Prof. Ramón López de Mantaras presented an overview of case based reasoning. It included comparisons between rule-based and case-based reasoning, induction and also the differences between a database and a case-base. Several representation cases were shown to make clear the uses of case base reasoning. Some issues in this field were discussed and an interesting example on the music domain was shown.

Prof. Ivan Bratko talked about handling noise in tree induction and inducing intermediate concepts. The first part was a survey in several approaches to handle noise. In the second part he discussed a novel approach to induce concepts. This approach is based on inducing intermediate concepts using Boolean decomposition.

Prof. Luc de Raedt gave a survey of the field of inductive logic programming. This included the framework on which it is based as well as some tools developed. These tools included FOIL, Tilde and Progol. Progol in particular has been very successful in solving problems in the toxicology domain.

Dr. Jon Shapiro presented relevant aspects of genetic algorithms. He presented some views on claims made by the GA community such as implicit parallelism, optimising and others. He also presented the concept of evolutionary learning and genetic programming.

Workshops

Machine learning for intelligent information access focused on the intelligent retrieval of information from two places: the web and a digital library. The main method to access the web was through agents. The papers presented on the web make a strong case to continue research on this domain although it seems that they are still too general and therefore specific applications and in a somewhat distant future. The digital library centred more on the representation of information as to make easy for a learning system to obtain knowledge from it.

In the workshop machine learning in user modelling interesting papers were discussed. Prof. Langley presented an adaptive model to fill out forms. Thus allowing the system to learn "how" to fill some forms. Two other papers focused in learning the profile of a user in order to help him find documents that are most interesting to him. The web was also a source for learning to model the users that browse a site.

The workshop pre and post processing in machine learning and data mining: theoretical aspects and applications unfortunately only included machine learning and no data mining papers. Two papers presented techniques to deal with numerical data for a particular technique, tree induction and genetic programming. Others presented a machine learning technique used to pre-process data. But my impression is the papers didn't reflect the importance of this area.

Finally the workshop data mining in economics, marketing and finance presented some interesting papers. This included a stock analysis tool, use of agents to gather financial information and several uses of rule induction in financial applications. It seems that rule induction in particular seems an interesting approach to this type of problems.

Conclusions

During the lectures there was an issue that was addressed by most of the speakers: comprehensibility. It seems that there is still a lot of effort being put into achieving ever greater accuracy but this hasn't make the solutions found by system more readable by its users. An example of this is the learning of induction trees aided with boosting. The accuracy achieved is very high but is impossible to extract any knowledge from the solution. Comprehensibility was very controversial. But the general consensus was that it should be topic of more research.

An important aspect that became apparent during the workshops is the little communication that exist between people doing research in the same field. It would be very useful to have more workshops like this allowing to people share experiences in the solution of related problems.

The are areas that seem will attract lot of research in the future. Most notably the web were there are a lot projects trying to obtain knowledge form the vast amounts of data generated in the Web. Data mining is a field that is very related with machine learning and it seems clear that research into the use of machine learning techniques in data mining will also be numerous.