Spring School in Machine Learning
Teaching experiences

Cécile Capponi Cecile.Capponi@lif.univ-mrs.fr
François Denis Francois.Denis@lif.univ-mrs.fr
Rémi Eyraud Remi.Eyraud@lif.univ-mrs.fr
Amaury Habrard Amaury.Habrard@lif.univ-mrs.fr
Liva Ralaivola Liva.Ralaivola@lif.univ-mrs.fr

Laboratoire d'Informatique Fondamentale
Centre de Mathématiques et Informatique
39 rue Joliot-Curie
F-13453 Marseille Cedex13 France
http://www.lif.univ-mrs.fr/spip.php?article7

Keywords: spring school organisation, attendance to a machine learning event, teaching experiences.

Abstract

The machine learning team of the "Laboratoire d'Informatique Fondamentale de Marseille" (ML-LIF) is a Joint Research Unit of the Université de Provence and the CNRS. All of the 5 professors/assistant professors of the team are involved in the organisation of a spring school in machine learning that will be held at the end of next May. The particularity of this school, compared to the machine learning summer schools usually supported by PASCAL, is that it is targeted to an attendance made of people coming from various scientific fields such as mere statistical inference, bioinformatics, machine vision. This short note summarizes some of the salient aspects regarding this school; it particularly stresses the diffusion of machine learning over a wide range of research areas.

A second item that is covered by this note deals with the teaching experience of the machine learning team members. It brings some clues as to how undergraduate and graduate students positively perceive machine learning and may fail to sustain their interest when the theoretical aspects of the field are tackled.

1. EPIT'08

The 36th Spring School in Theoretical Computer Science (namely EPIT 08, which stands for "Ecole de Printemps en Informatique Théorique") is devoted to statistical machine learning: it will cover the basic foundations of statistical machine learning, kernel methods, boosting, reinforcement learning and the connections between learning and wavelet modeling. It will take place from the 24th to the 29th of May 2008 on the island of Porquerolles (French Riviera). Most courses will be given in French by internationally renowned researchers.

Position	Number of participants
PhD candidate	38
Academic Researcher	23
Enterprise	3
TOTAL	66

Table 1. Position of participants. This table is made from the registration questionnaires.

Even though the school has not been held yet, two worthy facts may be singled out from questionnaires filled by registered participants. First, quantity-wise, it turns out that the school is very (and surprisingly) successful: a very short while after the registration was open and way before the early registration deadline, the whole 80 accommodations provided by the

school were booked up (66 for the participants, the rest for the organisation committee and speakers). It points out that machine learning is nowadays a very appealing scientific field. In Table 1, one can see that, though a small majority of participants are PhD candidates, a large part of the attendance will be formed by researchers.

Laboratory sector	Number
Math	5
Agronomy	6
Medicine	6
Bio-informatic	6
Computer science, multimedia & telecom	43

Table 2. Number of participants laboratories from each represented sector. This table is made from the registration questionnaires.

Secondly, the large broadcasting announcement of the school, and the fact that the call for participation was explicitly directed to every (young) researcher, brought in registrations from people having highly diverse backgrounds. Table 2 concerns the scientific sectors of participants' laboratories, while Table 3 summarises the attendance of the school based on questionnaires: they show that machine learning is of some interest to a large audience, part of which may have a limited background in the field. A further examination of the questionnaires especially shows that some of the attendees do their researches in apparently remote research areas but, for some purposes, usually related to classification/visualization problems, have to resort to machine learning techniques.

Field	number of participants
machine learning	16
Bio-informatic	3
Image processing	3
Statistic	2
Pattern recognition	1
TALN	1
Data mining	1
Complex systems	1
Modelisation	2
Signal treatment	1
...	

Table 3. Number of participants from each represented field. This table is made from partial results (25 over 60 participants) of a evaluation questionnaire sent to participants after their registration.

The first aim of the evaluation questionnaire is to evaluate average knowledge about each theme of the school, in order to help speakers to prepare their courses. We can then use the partial answers we receive at the moment we write down this article as a clue about which part of statistical machine learning is best known (and thus which may need further "advertising"). This should of course be nuances by the fact that obviously our school is far from covering all statistical machine learning subfields.

Table 4 sumerises declared knowledge about the different topics of the school[1]. It appears that kernels methods are the ones that are better known, especially from people who do not work in machine learning. Surprisingly, a confidential number of participants declares knowledge in Re-inforcement learning and boosting. The last course being about the link between statistical objects (wavelets) and learning, a quite unusual topic, it sounds reasonable that most participants will discover it during the school.

Topic	research subject	familiar	small knowledge	do not know
Statistical machine learning	7	17	7	1
Kernel methods	4	16	11	2
Re-inforcement learning		3	26	4
Boosting		7	14	10
Wavelet and learning	1	4	12	14

Table 4. A priori knowledge declared by the participants for each topic of the school.

2. Teaching experience

Our team is in charge of all the courses in machine learning and data mining taught at the University of Provence. These courses take place at undergraduate

[1]The inconsistencies between the sum of each line (and Table 3) is due to the fact that some of the questionnaires we got were partially uncomplete.

and graduate levels for students majoring in computer science. This teaching experience allows us to pinpoint some specific points as to how students perceive machine learning, which is essentially taught from the statistical standpoint.

At first, students seem very interested in machine learning: they have a rather good intuition on how machine learning might be useful in various applications such as spam filtering, autonomous driving, face recognition, search engines, data analysis, game programming... However, their enthusiasm usually fades away when the theoretical foundations of the field are tackled. This can be attributed to their lack of mathematical background, which is probably a consequence of how the studies in computer science and mathematics in our University (and more generally, in France) are split at the undergraduate level. This brings us to the following mismatch situation: on the one hand, the students that we teach to are precisely those who wanted to avoid mathematics and who are more interested in the practical aspects of computer science while, on the other hand, the students that could highly benefit from a course in machine learning, namely the students from the mathematics department, are afraid of the technicalities that may be inherent to a computer science-related course.

We can however nuance this statement as it is possible to get students in computer science to be interested in all of the aspects (including the theoretical ones) of machine learning. This particularly stands for some second year students undertaking a practical course aiming at doing a project in C that involves machine learning for face detection, game programming, digit recognition or robot battling. Most of these students really enjoy these kinds of projects, which may require the programming of kernel perceptrons, multilayer perceptrons, logistic regressions, and so on. Some of them are even willing to dig deeper into the field and ask for papers and various pointers about machine learning, while not being deterred by the mathematical aspects of it.

As a conclusion, the problem of teaching machine learning stems from the fact that it is not clear where to put machine learning: is it computer science ? applied mathematics ? something else ? Not being able to clearly identify what field this research area belongs to is probably one of the reasons of the relative inadequation between the courses and the attendance (in addition, of course to the structural problem of the way computer science and mathematics studies are parted in France).

Conclusion

From our experience in the organisation of a spring school and in teaching, two main conclusions can be drawn:

- Machine learning is a field that interests people from various horizons. This is true either for (young) researchers or students.
- A real effort is needed to be done about teaching and promoting machine learning in order to ensure a nice development of the field.

We hope that the Teaching Machine Learning workshop will be the first brick in the construction of this underdeveloped aspect of our field.

Relevant links

Webpage of EPIT'08: http://epit08.lif.univ-mrs.fr

Webpage of the computer science Master: http://www.lif.univ-mrs.fr/~liva/master

Webpage of the computer science Licence: http://www.lif.univ-mrs.fr/~egodard/L3Info