In: Proc. Fifth European Conference on Machine Learning (EWSL-91), pp457-462,
Ed: Yves Kodratoff, Berlin: Springer-Verlag, (1991)
http://www.cs.utexas.edu/users/pclark/papers

Applications of Machine Learning: Notes from the Panel Members

Peter Clark (Turing Institute, Glasgow, UK)
Bojan Cestnik (Jozef Stefan Institute, Ljubljana, Yugoslavia)
Claude Sammut (Univ. New South Wales, Australia)
Joachim Stender (Brainware GmbH, Berlin, Germany)

October 25, 1994

1 Introduction

Machine learning (ML) is devoted to the study and computer modelling of learning processes in their multiple manifestations. Although ML research started with the advent of computers, it is only relatively recently that its results have left the research laboratories and found their way into real-world applications.

The motivation for applying ML techniques to real-world tasks is strong: the problems of manually engineering a knowledge base are now well known, and ML offers a technology for assisting in this task; there is vast potential for automatically discovering new knowledge in the recent explosion of available on-line databases, too large for humans to manually sift through; and the ability of computers to automatically adapt to changing expertise would offer huge benefits for the maintenance and evolution of expert systems.

Despite this, the success of ML applications varies tremendously. There are some spectacular successes to its credit, but also the number of mature real-world applications reported in the literature is limited.

The fact ML has been highly successful in some instances deserves emphasising, to dispell the myth that ML has yet to achieve serious real-world success. Two notable examples are GASOIL and BMT, expert systems engineered using inductive tools with massive time savings (see Table 1). GASOIL was constructed in 1986 [Slocombe et al., 1986], and reported to be in regular use at four company sites in 1987 [Hertz, 1987]. BMT was constructed by Brainware GmbH, and is currently being deployed [Hayes-Michie, 1990]. A third example is Westinghouse's process control system [Leech, 1986], constructed with the aid of the inductive tool ExpertEase. By using ExpertEase along with other traditional means of analysis, Westinghouse achieved increased throughput in one important factory to the extent of increasing business volume by more than ten million dollars per annum, a result considered unlikely to have been achieved without ExpertEase [Barry, 1984]. Brief reports on around 25 other real-world ML applications can be found in [Hayes-Michie, 1990] and [Mowforth, 1986], and the number continues to grow.

System	Application	No. of rules	Development (man-years)	Maintenance (man-years)	Inductive tools
Mycin	Medical diagnosis	400	100	n/a	n/a
XCon	VAX computer configuration	8,000	180	30	n/a
GASOIL	Hydrocarbon separation system configuration	2,800	1	0.1	Yes
BMT	Configuration of fire-protection equipment in buildings	>30,000	9	2.0	Yes

Table 1: Tabulation from [Slocombe et al., 1986] and augmented by Michie with BMT data [Michie, 1990]

These achievements are substantial. However, it is also true that the number of such successes reported in the literature is limited. While this is partly because companies are either reluctant or see no reason to make public their progress, part of the reason must also be attributed to the need for ML research to mature further to extend the class of real-world problems that can be handled. It is thus worth reflecting on the current state of ML application: What type of applications have been successfully tackled by ML? Which ML techniques have been most successful in application? How well does the current research in ML match the requirements of application-builders? and what directions of further research are pointed to by those experienced in building applications?

The panel, all who have had substantial experience of constructing real-world ML applications, will be discussing these issues. Here, as a prelude to the panel meeting, some initial comments are offered from the panel members.

2 Comments

2.1 from Joachim Stender

Machine Learning research has resulted mainly in algorithms which handle synthetic data. This data has the following properties:

- The 'fields' are either carefully separated beforehand into conditions and conclusions by human experts or this can easily be done afterwards. This separation is semantically unique or is at least interpreted as such
- The examples are always carefully selected with the help of a human expert
- Very frequently these examples show the semantic properties of rules (through the extensive use of "Don't Care" or wildcard characters "*")
- The scope of the data samples is relatively small and it is easy to get an overview

- The data are typically free of noise; "Clashing Examples" are as a rule interpreted as input errors

- Very frequently, the examples cover the complete state space

Such synthetic databases are mostly artificial. i.e., they have been carefully constructed at a specific point in time.

The solution of real world problems usually requires the handling of analytic data. This data has the following properties:

- The state space is not completely covered by the given data samples

- The data are usually very noisy

- The hidden regularities within the data are not known by a human expert

- Conventional methods of data analysis cannot successfully be used due to (a) the complexity of the relations to be considered and (b) the size of the dataset

- Popular machine learning algorithms (in particular ID3-based algorithms) cannot be used due to the size of the dataset and the degree of noise in the data

- The separation of the 'fields' into conditions and conclusions is not simple and there is no single 'correct' way to do so

We call datasets with such properties analytic. Such datasets typically evolve through "natural evolution" which has normally taken place through years of additions and changes. Examples are customer databases or machine-acquired data.

Some ML researchers have responded to this discrepancy by attempting to extend their algorithms so that they can handle analytic data more easily (Tree pruning, statistical criteria etc). However, this is the wrong approach! Such algorithms were originally designed for something different, namely for the solution of artificial problems and through necessity have been extended. These extensions frequently do not reflect the original structure of the algorithm and sometimes represent a completely different philosophy, namely that of statistics.

Instead, the required approach is designing algorithms for the handling of large-scale, analytic data. These algorithms regard synthetic data as a special case.

2.2 from Bojan Cestnik

I would like to make two general points about the relation of ML research to ML applications. Firstly, most of the ML methods developed so far (e.g. ID3 and its successors, AQ, CN2, DUCE, FOIL) are general in the sense that they can be applied to a variety of different domains. On the other hand, the incorporation of domain-specific properties in the acquired knowledge has not been sufficiently considered yet in the ML research. Among the domain-specific properties, I would like to specifically mention the monotonicity of an attribute that can, when encountered, serve many useful purposes. It can be, for example, used in checking consistency of learning examples, or in generating additional learning examples if they are required. And more importantly, it is reasonable to believe that the quality of acquired knowledge can be substantially improved by taking the monotonicity into account within a learning method.

Secondly, for the applications of machine learning methods in real-world domains it seems important that several methods are applied and combined together. Since every single method generates only an approximation of domain knowledge, the main benefits of the combination are expected to be the following:

- Due to different methods, the acquired domain knowledge can be observed form different perspectives, in different knowledge representation formalisms, etc,; therefore the understandability of acquired knowledge from a user viewpoint can be substantially improved

- The performance of a reasonable combination of applied methods, in terms of classification accuracy, can be improved with respect to every single method.

2.3 from Claude Sammut

2.3.1 Where have practical ML successes occurred?

Almost certainly, the most successful learning program so far has been Quinlan's C4. What have been the reasons for its success?

- It can cope with noisy data

- It's representation allows continuous-valued attributes.

- The domains in which it has been use only require simple feature vector representations.

Donald Michie has conjectured that feature vectors seem to be a characteristic representation for one of the major application areas of expert systems, viz, finance. However, scientific and engineering domains tend to require more structured representations. Why should this be so?

2.3.2 Representations and structure

Perhaps one explanation is that many current expert systems capture 'seat of the pants' knowledge that is not governed by any deep model but purely by experience. However, if we are trying to discover laws governing natural phenomena then it is likely that a deep model is required. This implies that research in constructive induction and learning relations is central to learning in scientific domains. We might also mention the importance of qualitative models since they are useful in capturing relations.

Does the above also endorse research in explanation-based learning, since it too can deal with structured representations? I do not believe so. EBL (or more correctly, automatic programming) seeks to 'operationalise' an already acquired concept. While there is good reason to wish to do this, learning is not one of them since it is not possible to make any justifiable generalisation from one example. Learning by induction is still likely to be the most productive method for acquiring knowledge. This leads us to a consideration of how data for learning are collected.

2.3.3 Noisy data and non-determinism

The real world is messy! It is noisy, non-deterministic and not at all well behaved. We are familiar with noise in data sets such as those obtained from pathology laboratories, etc. However, even this data has been extensively filtered before being input to a learning program. Useful attributes and even higher-level features are chosen for the program.

A much more difficult task for machine learning is to place a learning program in control of an agent in a reactive environment and have the program learn how to achieve some desired result. In this case the program alone is responsible for collecting data. This situation arises in manufacturing and robotics when we would like some processes to tune itself automatically. The difficulties in this kind of task are: finding an appropriate representation, especially finding threshold values for variables; dealing with non-deterministic behaviour when the representation is inadequate.

Very little research into learning in this kind of domain has been conducted and it is vital if learning programs are to be of any use in real- time applications. Neural-net enthusiasts will immediately claim that they are making progress in this area. However, most research has been in the form of designing programs with almost no understanding of how or why they should work. A theory for this style of learning is essential. Interestingly, this is likely to bring statistics and learnability theory even more into mainstream machine learning.

2.4 from Peter Clark

It is tempting to try and generalise about which applications are most suitable for a ML approach. However, while some guidelines can be given, it should also be noted that the success of a ML application depends as much on the skill of the 'ML expert' as on the ML technique used. For example, in order to apply ID3, a suitable way of converting the problem into a classification-of-examples task must be selected; a suitable way of describing examples must be designed, requiring careful consideration and imagination to ensure the essential information is captured in the attributes; appropriate class values must be chosen; and initial runs of ID3 might reveal the data needs to be reorganised, or more data collected, or the definition of some attributes changed. This process may iterate several times until an acceptable solution is produced.

Thus the skills required to apply ML are quite considerable. For those seeking a ML solution to a real-world problem, the cost of obtaining the necessary ML expertise to use the ML algorithm frequently exceed the costs of the algorithm itself. In a nutshell, the 'unsupervised learning' of a tool such as ID3 is never unsupervised.

The implication of this for someone interested in a ML application to their problem is that ML expertise as well as the algorithm itself should be sought. The implication for ML research is that there is considerable activity in 'learning' to solve a problem which is not yet conducted by machines. Study of these extra areas of activity must be essential for further progress to be made in addressing real-world problems. Two practical ways forward for this are as follows. Firstly, critical study of the application as well as the development of ML techniques should be encouraged, and should be viewed as an important area of research. Secondly, as the ML expert's skill involves the use of substantial domain knowledge, the integration of ML with other knowledge-rich AI techniques (eg. qualitative modelling, causal reasoning, handling uncertainty) will become increasingly more important.

References

Barry, R. Personal communication, 1984.

Hayes-Michie, J. E., editor. *Pragmatica: Bulletin of the Inductive Programming Special Interest Group*, volume 1. Turing Institute Press, Glasgow, UK, 1990.

Hertz, D. B., editor. *Applied AI Reporter*, volume 4 (4). Univ. of Miami, NJ, 1987.

Leech, W. J. A rule-based process control method with feedback. *Advances in Instrumentation*, 41:169–175, 1986.

Michie, D. Machine executable skills from 'silent' brains. In Addis, T. R. and Muir, R. M., editors, *Research and Development in Expert Systems VII*, pages 1–24. Cambridge Univ. Press. (Proc. ES90, the 10th BCS Specialist Group on Expert Systems), 1990.

Mowforth, P. Some applications with inductive expert system shells. TIOP 86-002, Turing Institute, Glasgow, UK, 1986.

Slocombe, S., Moore, K., and Zelouf, M. Engineering expert system applications. In *Annual Conference of the BCS Specialist Group on Expert Systems*, 1986.

References

[Barry, 1984] Barry, R. (1984). (Personal communication).

[Hayes-Michie, 1990] Hayes-Michie, J. E., editor (1990). *Pragmatica: Bulletin of the Inductive Programming Special Interest Group*, volume 1. Turing Institute Press, Glasgow, UK.

[Hertz, 1987] Hertz, D. B., editor (1987). *Applied AI Reporter*, volume 4 (4). Univ. of Miami, NJ.

[Leech, 1986] Leech, W. J. (1986). A rule-based process control method with feedback. *Advances in Instrumentation*, 41:169–175.

[Michie, 1990] Michie, D. (1990). Machine executable skills from 'silent' brains. In Addis, T. R. and Muir, R. M., editors, *Research and Development in Expert Systems VII*, pages 1–24. Cambridge Univ. Press. (Proc. ES90, the 10th BCS Specialist Group on Expert Systems).

[Mowforth, 1986] Mowforth, P. (1986). Some applications with inductive expert system shells. TIOP 86-002, Turing Institute, Glasgow, UK.

[Slocombe et al., 1986] Slocombe, S., Moore, K., and Zelouf, M. (1986). Engineering expert system applications. In *Annual Conference of the BCS Specialist Group on Expert Systems.*