Types of machine learning literature review example

Technology, Artificial Intelligence

\n[toc title="Table of Contents"]\n

\n \t

\n[/toc]\n \n

INTRODUCTION TO MACHINE LEARNING

Introduction

Machine learning refers to the development of learning algorithms, which enable computers to automatically execute tasks without the need for human intervention (Schapire 2003). With application learning techniques, computers are able to come up with their own programs for solving a problem, and these are based on the examples that the computers are provided.

In particular, some of the questions that the discipline of machine learning aims to answer are: 1.) How computer systems that automatically improve with experience can be built; and 2.) What the fundamental laws are that govern all learning processes (Mitchell 2006).

Machine learning also makes use of principles from both computer science and statistics where the computer science aspect is concerned with how

machines that solve problems can be built while the statistics aspect is concerned with what can be inferred from data, along with a set of modeling assumptions, and how reliable the inference is. However, even the study of human and animal learning in the fields of neuroscience, psychology, biology, and even economics and control theory can be related to machine learning.

Some of the learning tasks that are aimed to be performed with the use of machine learning techniques include how mobile robots can learn to navigate based on their own experience; how historical medical records can be used to determine which treatments will be most effective to certain types of patients; and how search engines will automatically customize to the interests of their users.

According to Mitchell (2006), " a machine learns with respect to a particular task T, performance metric P, and type of experience E, if a system reliably improves its performance P at task T, following experience E" (Mitchell 2006, p. 1). In this regard, the learning task may also be referred to as programming by example, database updating, autonomous discovery, or data mining.

The types of machine learning are supervised, unsupervised, and reinforcement learning (Marsland 2011).

In supervised learning, which is also referred to as learning from exemplars, the computer is provided with the correct responses or targets and based on this training set, the algorithm generalizes in order to correctly respond to all possible inputs (Marsland 2011). The data in this type of learning can be

represented as {X, Y} pairs where the Ys are actual labels of the various data elements in X (Adams 2011). As an example, each element xi € X is an image and yi € Y is a binary indicator that shows whether a cat is in the image xi or not. In this case, the main goal is to predict the labels Ynew for a new data set Xnew, which has no labels. This would then require the use of experience of the full pairs dataset for the prediction of the other labels. Examples of systems that are classified as supervised learning problems include recommender systems.

In unsupervised learning, the computer is not provided with the correct responses, but the algorithm tries to determine the similarities between the inputs so that the inputs with similar attributes are categorized together (Marsland 2011). With this type of learning, only the data elements X are known, and the labels Y are not provided. As such, the methods that are employed are those that use the data elements as basis for classifying the data. The data elements generally have strong natural properties that make them distinct from other types of datasets. An example is that an image is distinct from noise in that images have pixels that are strongly correlated whereas noise is an image where the pixels are independent random variables. In this case, an image can be differentiated from noise though the correlation among the pixels. This type of learning is employed using the statistical approach called density estimation (Marsland 2011). Pattern recognition problems are generally categorized as unsupervised learning.

On the other hand, reinforcement learning, also called learning with a critic, falls somewhere between supervised and unsupervised learning (Marsland 201). With this type of learning, the algorithm is informed about when the

answer is wrong but does not get informed about how it can be corrected. As such, the algorithm has to explore and try various possibilities until it is able to determine the right answer. Furthermore, this type of learning punishes or rewards a computer with regards to its accuracy. This type of learning does not make use of {X, Y} pairs but can be an effective technique for long-term learning processes.

Historical Overview

The history of machine learning can be divided into three periods, namely the exploration period in the 1950s and 1960s; the development of practical algorithms in the 1970s; and the exploration of research directions in the 1980s (Shavlik and Dietterich 1990).

Exploration Period

The work that started in the 1950s and 1960s was inspired by psychological, biological, and neurophysiological research. Research during this time focused on the development and testing of computational analogies of neurons, the most prominent of which was Rosenblatt's perceptron. More specifically, perceptrons referred to " a family of theoretical and experimental artificial neural net models" (Raza 2012, p. 7). On the other hand, other researchers during the exploration period conducted experiments on simulated evolution in order to test whether the processes of natural selection and random mutation would result in intelligent programs.

Development of Practical Algorithms

The 1970s saw the development of practical algorithms. In particular, Winston's concept of blocks-world learning began to make learning the focal point of artificial intelligence research (Shavlik and Dietterich 1990). This concept was demonstrated in the METADENDRAL system, which was used for mass-spectrometry prediction rules; in AQ11, which was intended for learning soybean disease diagnosis rules; and ID3, which was used for learning chess end-game rules. The MACROPS method was also developed for learning macro-operators in blocks-world learning (Shavlik and Dietterich 1990) while the AM program by Lenat showed that it was possible to automatically discover mathematical concepts through the use of a learning program.

The advantage of the ID3 algorithm is that it is simple. However, it results in decision trees rather than in production rules. As such, the production rules still need to be induced from the decision tree. On the other hand, the main advantage of the AQ algorithm is that negation may be easily expressed by the induced production rules. However, these rules may be of poor quality as the time complexity of this algorithm is exponential and because it uses the parameter MAXSTAR. With regards to the LEM2 algorithm, its main advantage is that it induces the minimal discriminant description. It is also of polynomial time complexity. However, because of this polynomial complexity, the induced description may not always be minimal.

Theoretical analysis started with the development of Mitchell's candidate elimination algorithm, as well as the concept of the bias of an inductive learning algorithm. In addition, this area of research led to the development

of PAC (probably approximately correct) learning. On the other hand, an example of a symbolic learning algorithm is Quinlan's ID3 learning algorithm for learning decision trees.

Connectionist networks are capable of being trained with specific learning rules based on the units and the network architecture used (Riloff and Scheler 1996). These can be used as numerical learning techniques for natural language processing. On the other hand, conceptual clustering involves the presentation of objects or observations, with each having its own set of features. This set is then divided into classes and subclasses, with objects that share similarities being grouped together. The conceptual clustering task also allows for both numeric and symbolic features.

Figure 3 A simple classification tree

One advantage of conceptual clustering, particularly hierarchical conceptual clustering, is its capability for having the greatest coverage with the smallest possible number of clusters. This implies that clusters are general enough to enable the description of all data while still allowing for the definition of individual concepts. Another advantage of conceptual clusters is their attribute of having big cluster descriptions. It should be noted that a cluster's inferential power increases as the more features the cluster has. As well, there is minimal overlap between clusters, which means that there are no disjoint concepts. Moreover, clearly defined concepts are required in conceptual clustering.

On the other hand, one of the limitations of conceptual clustering is its attribute value reorientations, which may limit their use with structural or relational information. In addition, this approach may have limitations when

used for learning inexact concepts.

With regards to explanation-based learning, this area of research mostly applied in the context of integrated problem-solving architectures. As well, explanation-based learning techniques are used in natural language understanding systems and learning-apprentice systems. Similarly, knowledge-guided learning techniques are used for the identification of defects in the explanation. These defects are then repaired through the use of inductive learning methods to ensure that the learning system is correct, complete, and computationally tractable. On the other hand, analogical and case-based learning aims to enable the transfer of knowledge from a task that is well-understood to a task that is less familiar. Finally, genetic algorithms aim to optimize searches, especially in large systems of classifiers. The effective application of genetic algorithms can greatly reduce the search effort with little effect on the quality of the obtained solution. However, with only the chains of rules receiving external evaluation, it can be quite difficult to determine whether to assign credit or blame to a specific subcomponent based on the global performance of a complex system.
Figure 2 A simple search tree

The Future of Machine Learning

Solomonoff (2009) opines that the field is not far from being able to seriously implement artificial intelligence. He particularly sees promise in Koza's system, as well as in Schmidhuber's OOPS system (Solomonoff (2009). Similarly, DARPA (Defense Advanced Research Projects Agency) has launched the PPAML (Probabilistic Programming for Advanced Machine

Learning) program, which aims to take machine learning to the same level as high-level programming languages so that machine learning projects can become more easily implemented and so that the present barriers to their development and implementation can be reduced; thus, giving way to increased effectiveness, productivity, and innovation (DARPA 2013).

With all the advances that are bound to occur in the field of machine learning, it is not improbable that many tasks that currently require human intervention will be performed by computers in the future. These would include not only the tasks that require mental processing on the part of humans, but even the tasks that would require physical exertion. Although this certainly has its benefits in that this can lead to the faster completion of tasks – hence, increased productivity – it also carries the threat of humans being replaced by robots where the value of human contributions in the various sectors may be diminished.

However, even before technology reaches that point, there's still the challenge of being able to cope with the vast amounts of information being amassed on a daily basis. As it is, there are already challenges in incorporating these large volumes of information into learning systems and learning analytics; so much more if it were translated into machine learning language. It's then important for humans to first be able to correctly and accurately interpret and make sense of the information that they gather before further developments in machine learning can be made. After all, how can man make the computer learn when the man doesn't have that learning yet?

Conclusion

The importance of machine learning as a field of study is increasing over time, particularly with regards to artificial intelligence. With information being the most important resource that enables our modern world to operate and with the vast amounts of information being gathered everyday, the only way to for humans to keep up with this information is through the automation of information processing tasks through machine learning. Indeed the outputs of machine learning benefit not only the field of computer sciences but even the fields of human sciences, such as psychology, medicine, biology, and even control theory and economics.

There are many approaches to machine learning. Since the advent of machine learning in the 1950s, many approaches have been developed, which can be classified into the three periods of machine learning history. These are the exploration period during the 1950s and 1960s; the development of practical algorithms in the 1970s; and the exploration of research directions in the 1980s.

The machine learning concepts developed during the 1950s and 1960s included Rosenblatt's perceptron and the experiments on simulated evaluation. On the other hand some of the practical algorithms developed in the 170s included the AQ11 and ID3 algorithms, as well as the AM program by Lenat. Finally, the 1980s saw researches being conducted in the areas of learning theory, symbolic learning algorithms, connectionist (neural network) learning algorithms, clustering and discovery, explanation-based learning, knowledge-guided inductive learning, analogical and case-based reasoning, and genetic algorithms

All of these approaches have their pros and cons, their weaknesses and their strengths. As such, it is important for research to continuously be conducted in order to find better means for enabling learning and making the most of these approaches. This will ensure that all of the information that our modern world gathers is analyzed and interpreted correctly so that they can be used for the betterment of our society.

References

Adams, R., 2011. Computer science 281: Notes from lecture. [online] Available at: [Accessed 30 March 2013].

DARPA, 2013. DARPA envisions the future of machine learning. [online] Available at: < http://phys. org/news/2013-03-darpa-envisions-future-machine. html> [Accessed 30 March 2013].

Langley, P. W. and Carbonell, J. G., 1984. Approaches to machine learning. Computer Science Department. Paper 1505. [online] Available at: [Accessed 30 March 2013].

Marsland, S., 2011. Machine learning: An algorithmic perspective. CRC Press.

Mitchell, T. M., 2006. The discipline of machine learning. [online] Available at: < http://www. cs. cmu. edu/~tom/pubs/MachineLearning. pdf> [Accessed 30 March 2013].

Raza, H., 2012. Fuzzy spiking neural networks. GRIN Verlag.

Riloff, E. and Scheler, G., 1996. Connectionist, statistical and symbolic approaches to learning

for natural language processing. Springer.

Schapire, R., 2003. COS 511: Foundations of machine learning. [online]

Available at: < http://www. cs. princeton.

edu/courses/archive/spring03/cs511/scribe_notes/0204. pdf> [Accessed 30

March 2013].

Shavlik, J. W. and Dietterich, T. G., 1990. Readings in machine learning.

Morgan Kaufman.