Middle-East Journal of Scientific Research 15 (2): 266-271, 2013
ISSN 1990-9233

DOI: 10.5829/idosi.mejsr.2013.15.2.11024

Unlearning Machine Learning: The Challenge of Integrating Research in Business Applications

[1]Anand N. Asthana and [2]Sangeeta Khorana

[1]CENTRUM Católica Graduate Business School,
Pontificia Universidad Católica del Perú, Lima, Perú
[2]School of Management and Business, Aberystwyth University, UK

Abstract: New applications of Machine Learning, a branch of Artificial Intelligence, have earned billions of dollars for businesses that have cleverly used them. Business professionals in Finance, Marketing, Operations and Logistics and even Strategy are avidly using machine learning applications. In a range of activities from planning for a new product to reorganising supply chains, businesses have a huge stake in research in this field. Yet there has been little evaluation of usefulness of current research and the direction in which it is proceeding. The use of ever more complex algorithms has made machine learning a high-tech highbrow area of research, but robustness and external validity are yet to be established. Focusing on the elegance of the solution is leading researchers towards less-important questions, while there is an unfulfilled need to look at areas which are probably more relevant from the point of view of business.

Key words: Machine Learning · learning machines · Business Analytics · Data mining · KDD

INTRODUCTION

Recent years have seen spectacular applications of machine learning. Many smart applications are already in the hands of delighted consumers and have generated billions of dollars for businesses. Siri in iPhones predicts the meanings of human voices and tries to provide the desired answers. Photo album in Facebook recognises faces to be tagged in photos. LinkedIn predicts who the users want to connect with. In the recently released MS Office 2013, the Excel program can comb very large amounts of data to find meaningful patterns. For example, with bi-directional access to live Twitter, it can scan millions of Twitter posts and create charts to show which product is getting the most buzz. A new version of MS Outlook will employ machine learning to review the e-mail habits of users to see whether a user wants to read each message that comes in. Casinos are able to use face recognition to bar entry to card counters. Driverless cars are not the stuff of science fiction any more. These cars pioneered by Google are ferrying its staff to their offices. A recent Netflix contest, which offered $1 million to any team of researchers whose algorithm could improve the company's movie recommendation system by 10%, was a battle of wits waged with the weapons of machine learning. Less glamorous but equally useful applications are found in the field of operations and logistics like supply chain demand forecasting. For example, forecasting the manufacturer's demand under information asymmetry – the bullwhip effect – can be tackled through advanced machine learning techniques [1].

Big business is placing big bets on machine learning applications. IBM, seeing an opportunity in data-hunting services, created a Business Analytics and Optimization Services group in April 2009. Microsoft is designing machine-learning software that can trawl internal corporate computer systems with a view to predict which software applications are most likely to fail when seemingly unrelated programs are tweaked. In the long term, Microsoft hopes to combine even more machine learning with its cloud computing system, called Azure, to rent out data sets and algorithms so that businesses can build their own prediction engines [2]. Drug companies are using machine learning to understand spread of diseases. The machine learning approach is being found useful for finding business partners and building

Corresponding Author: Anand N. Asthana, CENTRUM Católica Graduate Business School, Pontificia Universidad Católica Del Perú.

reciprocal relationships [3]. There is hardly any area of business which is untouched by machine learning. It is viewed as the foundation of a better and smarter future. In the academic world of computer science, enthusiasm for machine learning is growing at all levels. In 2010, the Turing Award, recognised as the Nobel Prize of Computing, went to Leslie Valiant, an innovator in this field for his transformative contributions to the theory of computation, including the theory of probably approximately correct (PAC) learning. Next year, at Udacity (a start-up launched by Stanford University) an online course on machine learning, taught by Peter Norvig, Google's director of research, attracted a record 160,000 students [4].

Machine learning programmes may not be able to answer questions relating to meaning of life but are getting better at answering queries relating to business. The programmes, however, are yet to understand nuances of languages and machine translation is improving at a very slow place. Even so, confidence in the capabilities of machine learning is growing. The idea of algorithms sucking up and sorting through the otherwise useless garbage of data is appealing and it is easy to be attracted to the mystique of machine learning. It seems like a brave new world where thinking need not be constrained by what may or may not be possible. Machine learning seems to be becoming like motherhood and apple pie. Nobody is opposed to it. The reality, however, is a bit more complex. There is a huge unmet demand for applications whereas much of the output is either not being used or is unusable. Netflix paid out the prize money but never used the winning algorithm because extraneous factors came into play. Like any breakthrough technology, machine learning involves some forethought and discipline before being let loose in the enterprise. The key questions to ask are: "What kinds of business problems do we want to solve?" "What questions do we want to be able to answer?" and finally, "How much would they be worth if we could answer them?" This article looks at the cutting edge applications of machine learning, where the research in this field has gone off-track from the point of view of the user and it also delineates possibilities for the future.

What Is Machine Learning?: Can machines think? The question has baffled the philosophers and laymen alike for a long time [5]. At the dawn of the computer age, British mathematician Alan Turing, the designer of programmable computing device, began a classic article titled 'Computing Machinery and Intelligence' with this question [6]. He did not answer the question and discarded it as too meaningless to deserve discussion. Turing was not committing himself to the view that to think means thinking like a human [7]. In a prescient section entitled 'Learning Machines', Turing successfully anticipated machine learning in 1950 [8].

Machine learning, a branch of artificial intelligence, relates to the construction of systems that can learn from data thus giving computers the ability to learn without being explicitly programmed. Machine learning lies at the intersection of computer science, engineering, statistics and often appears in other disciplines, especially business. Machine learning techniques can be applied to many business problems that need to interpret and act on data. Arthur Samuel, whose papers in the 1950's on the subject are still worth studying, defined machine learning as a field of study that gives computers the ability to learn without being explicitly programmed [9]. Samuels wrote a checkers playing program, had the program play over ten thousand games against itself and work out which board positions were good and bad depending on wins and losses. Forty years later, Tom Mitchell formalised this definition as: "A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E" [10: 2]. In the case of the checkers learning problem, task T is playing games, performance measure P is proportion of games won and experience E is playing practice games. Thus, searching for definitions in cognitive terms favoured by philosophers, e.g., "Can Machines think?" has given way to formulating definitions in operational terms, i.e., "Can it do what we, the thinking entities, can do?"

Machine learning and data mining have so much overlap that it is not easy to distinguish between the two disciplines. Machine learning is more focused on prediction, based on known properties learned from the training data. Data mining focuses on the discovery of (previously) unknown properties in the data. But, machine learning also employs methods like "unsupervised learning" where there is no training data and data mining uses many machine learning methods, albeit sometimes with a slightly different goal in mind. In machine learning, performance is usually evaluated with respect to the ability to reproduce known knowledge, while in Data mining, the key task is the discovery of previously unknown knowledge. The research in the two fields is converging. The best known conferences in the two fields, the European Conference on Machine Learning (ECML) and European Conference on Principles and

Table 1: Examples of common Machine Learning applications in Business

	Applications		
ML Problem	Finance	Marketing	Business (other)
Classification: identify the maximally distinguishing attributes using training data set	Risk classification (RI)	Market segmentation (RI)	Churn management Call tracking (CBR)
Prediction: Finding probable future values or distributions of attributes	Forecasting default (RI)	Customer reaction to promotions (GA)	Network behaviour (NN)
Detection: Identifying causes of irregular patterns	Suspicious transactions (NN)		Software cost estimation (NN)
Association: identifying rules governing relationships		Market basket Analysis (VS)	Similarity assessment (ILP)

Note: Techniques most commonly used in brackets. CBR - Case Based Reasoning; GA - Genetic Algorithms; ILP - Inductive Logic Programming; NN - Neural Networks; RI - Rule Induction; VS - Visualisation

Table 2: Examples of clash of terminology

Meaning/ example	Business Analytics	Machine Learning
$(X_1, Y_1, Z_1), (X_2, Y_2, Z_2), \ldots.. (X_n, Y_n, Z_n)$	Data	Training sample
X_i's	Covariates	Features
Using data to estimate an unknown quantity	Estimation	Learning
Predicting a discrete Y from X	Classification	Supervised learning
Putting data into groups	Clustering	Unsupervised learning
Map from covariates to outcomes	Classifier	Hypothesis
Multivariate distribution with given conditional independence relations	Directed acrylic graph	Bayes net
Uniform bounds on probability of errors	Large deviation bounds	PAC learning

Practice of Knowledge Discovery in Databases (PKDD), were co-located in 2001. In 2008 the conferences were merged into one conference and the division into traditional machine learning topics and traditional data mining topics has been removed. In the 1980's the field of machine learning was largely empirical and ad hoc. Over a period of quarter of century, a series of models have been developed that combine technical depth and broad applicability, thus giving the field theoretical adequacy as also effective applications in business. An illustrative (not comprehensive) list of popular business applications and the machine learning techniques therein are given in Table 1.

Lost In Translation: In business research 'data mining' refers to lack of academic integrity where a researcher 'mines' the data that fits well in the model and ignores the rest of the available data. In computer science, 'data mining' has no such connotation and data mining is a respectable field. However, of late the practitioners of data mining seem to prefer a fancy name for data mining: Knowledge Discovery in Databases (KDD). Machine Learning and KDD are based on statistics, but neither admits to being a branch of statistics. Statistics works in the background and the user rarely sees the statistical output. This is for the benefit of the uninitiated who have a poor opinion of statistics. "Lies, damned lies, and statistics" is a phrase popularised by Mark Twain describing the persuasive power of numbers [11: 471]. The most widely read Statistics book in the history of the world is titled 'How to Lie with Statistics' [12]. The terms used to describe the same or similar phenomenon in statistics and machine learning are often different [13]. For example, terms like confidence interval are avoided in machine learning. Some examples of different terminology are given in Table 2.

The discordance in terminology has gone too far and the increasing jargon is becoming an impediment to mutual understanding. Researchers in business analytics who can apply the methods from Physics, Mathematics and Mechanical Engineering to the problem of their choice cannot simply apply machine learning. It is difficult to phrase the problem, decide which features to use and how to search over parameters because machine learning solutions come "packaged in a Ph.D." [14]. For a larger machine learning impact, it is imperative for the machine learning community not only to restrict jargon but to simplify machine learning algorithms and tools making them mature and robust.

There is the added problem of translation of the results, i.e., interpreting the results obtained from the dataset to the wider world. Here one can see the difference in the approach of scholars in machine learning and those of business analytics. Machine Learning being a young field, conferences are very important and journals less so. In business, as the saying goes, the difference between a conference and a holiday is that in case of the former, your employer pays the bill. Getting a paper published in an ISI accredited Economics or Business journal is difficult not the least because of the requirements of replicability. The authors are expected to show whether their results have external validity. This is not the case in machine learning conferences. What is required there is development of a new algorithm or the explication of a novel theoretical analysis. Unfortunately, the same yardstick has been adopted by the two ISI accredited journals in the field, viz. *Machine Learning* and *Journal of Machine Learning Research*. The researchers burrow into the huge archives of datasets and attempt prove that their algorithm performs better than others. How does it translate to the actual business application? Neither the referees nor the editors seem to ask. There are some papers written in collaboration with domain experts, but their number is very small. The communication barrier could make the field inward looking and much of the research in the field is for its own sake.

Unlearning And Relearning: Machine Learning may seem like any other scientific discipline: scientific methods, competing paradigms, evaluative criteria, prestigious master's programmes, conferences at exotic locations and archival publication. Concepts in learning are better defined than in other branches of Artificial Intelligence, e.g., reasoning, expert systems, planning, natural language, robotics and vision [15]. But, it does not seem to be headed for continued evolutionary progress in the standard mode. It is full of uncertainty and excitement. Some of the excitement is due to the recent explosion of digital data from new realms – sensor signals, surveillance tapes, social network chatter, opening of public records and so on. Data overload which is seen as a problem in some disciplines is considered a boon by scholars in the field of machine learning as they hunt for meaningful patterns in these troves.

Data and Metrics: Machine learning papers mostly describe a new algorithm and the research relies heavily on repositories - collections of databases, domain theories, and data generators for empirical analysis of machine learning algorithms. The new algorithm's behaviour is illustrated on synthetic data sets and then the paper reports results on a collection of standard data sets available in an archive. The most commonly used archive is UCI maintained by University of California, Irvine. This repository focuses on tasks like classification and regression and has made cross-domain studies in this area straightforward and commonplace leading to more than a thousand papers citing this repository. At the same time, it has had a harmful effect in the long term as the researchers are neglecting complex tasks like reasoning, problem solving and language understanding [16]. There is no consensus within the machine learning community about what role UCI data set serves other than helping scholars churn out papers. They are less useful than synthetic data, since the researchers do not control the process through which data are generated and they cannot be called real world data as they are not associated with real world users, experts or operational systems. Worse, they have depreciated the value of formulating problems and defining features [14].

Problems of data management are of a different type in businesses and most businesses do not know how best to use accumulated data for improving business decisions. With the exponential growth of technology, businesses not only need better tools to understand the data they currently have, but also to prepare themselves for the data they will have. Data has piled up for a long period of time and executives are afraid to get rid of it. They know that much of the data is useless but have no idea what use could be made of it in the future. Apart from normal business decision making, the data could also be used in litigation and government investigation. This has led to information governance-related paralysis. Of late there is a realisation that the days of storing everything indefinitely are over as there is simply too much data. The cost of data storage is significant if one includes the cost of all the overheads of managing the different systems and the clogging of the networks as data moves back and forth. Machine learning technologies combined with an iterative workflow that leverages a small amount of human input to identify relevant information can be used to address data hoarding. On the legal side, some software applications have come in the market using which businesses can quickly find information that is safe for deletion and feel confident that they have made the right decision. As data accumulation continues to accelerate, more research is required in this direction. Some kind of data-management scheme has to be applied as data arrives so that wasteful volumes of data do not start accumulating and useful data do not fall through the cracks.

Experimental evaluation revolves around performance metrics. The race to improve performance metrics, however abstract, has meant that there is little interest in research on knowledge-generating mechanisms. At the same time availability of large data sets meant lack of interest in using background knowledge to improve learning rate, marginalising research on learning faster from fewer experiences. A researcher claims victory when her algorithm makes an improvement in say accuracy in classification across certain datasets. The methodologies of showing improvement are open to question. Moreover, an improvement of x% can have very different impact in different data sets [14]. What the meaning of that improvement is in a real world situation is rarely examined. Moreover, mindless competition among the algorithms reveals little about the sources of power or the effects of domain characteristics [16]. Businesses need to use machine learning diagnostically (what data to use) and if need be opportunistically (finding new data sources). Global businesses operating across countries need to be careful about aggregate data from different countries, especially in the field of human resource management, because management practices in emerging and developed economies are significantly different [17].

Integration: For quite some time research in business has been becoming increasingly quantitative and business scholars in the domains of Economics, Finance, Human Resource, Marketing and Industrial Organisation, among others, are aspiring to achieve the same standards of academic excellence that hard disciplines demand [18]. For this reason, business scholars and practitioners sailed into computer age seamlessly. However, machine learning has not attracted attention of professionals or even hard core business analytics and the job is left to consultants with little oversight. Businesses are getting good at using machine learning to attract new customers but encounter problems in promoting new products to existing customers. This is so because the entire lifecycle of the customer is not known to the marketers, who typically focus on acquisition of new customers and lack access to up to date information that would enable them to market to existing customers effectively. The marketing management system that sends out flyers, the call centre software that receives phone calls from customers and the Customer Relationship Management (CRM) system that is used when the customer visits the shop or the office are not integrated. To make optimal use of machine learning, not only should these systems be integrated with each other, they should also be integrated with click-stream information, such as a web log that captures the fact that the customer was looking around for a related product.

Interdisciplinary integration is even more difficult as can be seen from application of machine learning in insurance industry. Increasing human longevity is a foreseeable long-term trend but the future in terms of morbidity is uncertain [19]. A modern approach to engineering systems, instead of focusing on the risk of things going wrong focuses on what keeps systems functional. A tenet of this approach, called resilience engineering is that a system is safe if it can adjust its functioning before, during, or after disturbances, so that the required operations are sustained under expected and unexpected conditions [20]. Today's underwritten market is highly competitive and a large number of companies are trying to gain sustainable competitive advantage by creating intellectual property that can help them understand numbers in finer details. Focusing on survival rather than death and data-mining the domain of social psychology, we can now define a resilience distribution to reflect the marked population variability in cognitive and social functioning and well-being [21]. There is a significant communication gap between various branches of business research. For example, in case of consumer choice, insights from market science are yet to be fully integrated in business economics [22]. A discourse that includes machine learning and the relevant branches of business is not easy to construct but this difficult task needs to be undertaken to make machine learning more relevant to business.

CONCLUSION

Machine Learning offers many useful ways to solve business problems. Much of the current research is detached from these problems as researchers prefer to withdraw into their private world and work in isolation with data sets to improve performance of algorithms [14]. Concern with learning in the context of intelligent systems – learning methods within sophisticated systems that carry out various types of complex cognitive activities like multi-step reasoning, heuristic problem solving and language understanding - has been replaced by component learning algorithms where the performance element is often simple and usually borrowed from earlier work. Researchers are showing no interest in the outside world, not even in other branches of Artificial Intelligence, which is a great loss [16]. Since publication in conference proceedings is the end of the process, the outside world has no way of knowing whether the results are valid or useful. According to Leslie Valiant, the most critical choice for a scientist is what problems to work on and the successful ones find problems of significance [23]. If machine learning looks inward, there could be many great accomplishments acknowledged by the peer group but of no use outside. This would be a tragedy because the opportunities for real impact are unlimited. The challenge of machine learning is to recover the discipline's original breadth of vision and continue to develop learning mechanisms that cover a range of business activities.

REFERENCES

1. Carbonneau, R., K. Laframboise and R. Vahidov, 2008. Application of Machine Learning techniques for supply chain demand forecasting. European Journal of Operational Research, 184(3): 1140-1154.
2. Hardy, Q., 2012. Microsoft Seeks an Edge in Analysing Big Data. New York Times, pp: 29.
3. Mori, J., Y. Kajikawa, H. Kashima and I. Sakata, 2012. Machine learning approach for finding business partners and building reciprocal relationships. Expert Systems with Applications, 39(12): 10402-10407.
4. The Economist 2012. Learning new lessons Dec 2.
5. Mays, W., 1952. Can Machines Think? Philosophy, 27: 148-162.
6. Turing, A.M., 1950 Computing Machinery and Intelligence. Mind, LIX, (236): 433-460.
7. Denett, D.C., 2008. Can Machines think? In Levitin, D. J. (Ed.), Foundations of Cognitive Psychology: Core Readings (pp: 35-54). Pearson.
8. Harnad, S., 2008. The Annotation Game: On Turing (1950) on Computing, Machinery and Intelligence. In Epstein, R., G. Roberts and G. Beber, (Eds.), Parsing the Turing Test: Philosophical and Methodological Issues in the Quest for the Thinking Computer. Evolving Consciousness, (pp: 23-66). Springer.
9. Samuel, A., 1959. Some Studies in Machine Learning Using the Game of Checkers. IBM Journal of Research and Development, 3(3): 210-229.
10. Mitchell, T.M., 1997. *Machine Learning*, McGraw Hill.
11. Twain, M., 1907. Chapters from My Autobiography - XX. North American Review, 185(618): 465-474.
12. Steele, J.M., 2005. Darrell Huff and Fifty Years of How to Lie with Statistics. Statistical Science, 20(3): 205-209.
13. Van Iterson, M., H.H.B.M. van Haagen and J.J. Goeman, 2012. Resolving confusion of tongues in statistics and Machine Learning. Proeteomics, 12: 543-549.
14. Wagstaff, K.L., 2012. Machine Learning that Matters. Proceedings of the 29th International Conference on Machine Learning, Edinburgh.
15. Hoffman, L., 2011. A Lifelong Learner. Communications of the ACM, 54(6): 127-128.
16. Langley, P., 2011. The changing science of machine learning. Machine Learning, 82(3): 275-279.
17. Nigam, R. and Z. Su, 2011. Management in Emerging versus Developed Countries: A Comparative Study from an Indian Perspective. Journal of Centrum Cathedra, 4(1): 122-133.
18. Asthana, A.N., S. Mohan and S. Khorana, 2012, Physics Envy and Natural Experiments in Business and Economics. World Applied Sciences Journal, 20(3): 464-469.
19. Beard, J.R., S. Biggs, D.E. Bloom, L.P. Fried, P. Hogan, A. Kalache and S.J. Olshansky, 2012, Global Population Ageing: Peril or Promise. World Economic Forum.
20. Hollnagel, E., D.D. Woods and N. Leveson, 2006. Resilience Engineering: Concepts and Precepts. Ashgate Publishing.
21. Woo, G., 2013. Of wealth and health. The Actuary, pp: 28-29.
22. McFadden, D.L., XXXX. The New Science of Pleasure. Working Paper 18687. NBER.
23. Anthes, G., 2011. Beauty and Elegance. Communications of the ACM, 54(6): 15.