

Learning in Dynamically Changing Domains: Theory Revision and Context Dependence Issues

Charles Taylor¹ and Gholamreza Nakhaeizadeh²

¹ University of Leeds, Leeds LS2 9JT, UK

² Daimler-Benz Research and Technology, 89013 Ulm, Germany

Abstract. Dealing with dynamically changing domains is a very important topic in Machine Learning (ML) which has very interesting practical applications. Some attempts have already been made both in the statistical and machine learning communities to address some of the issues. In this paper we give a state of the art from the available literature in this area. We argue that a lot of further research is still needed, outline the directions that such research should go and describe the expected results. We argue also that most of the problems in this area can be solved only by interaction between the researchers of both the statistical and ML-communities.

1 Introduction

The ability to “learn” (or discover) knowledge from data in a more directed framework has had a long history both in Machine Learning (ML) and in Statistics; see [9] for a review and a list of references. In classification the data are generally of the form (*attributes, class*), and the objective is simply to learn a “rule” whereby a new observation can be classified into one of the classes using the information in the attributes.

For any rule, however, difficulties arise when the data are not static, and so the validity of patterns and rules (discovered ‘knowledge’) depends, in some unspecified way, on **time**, and it changes with time. In this case, learning and classification (or prediction) procedures need robust mechanisms for detecting and adapting to changes.

In the machine learning literature, the generic term for such time-dependent changes is **concept drift**. Concept drift may be caused by continuous changes of the world and the environment – for example, in economic problems, the presence of inflation causes a continual drift in the real value of money which means that rules based on absolute monetary values will quickly become out of date – or it may occur when the attribute values or the concepts depend on a certain (possibly unknown) *context*. For example, the relative proportions of vehicles made by each manufacturer will depend on the region (or country) of observation. Thus, the location can be seen as the context in which the data are collected – knowledge of the context (or a change of context) will aid the knowledge discovery process.

There are various ways in which concept drift can manifest itself: the distribution of the attributes in a class can change; new attributes become available

or existing attributes can no longer be measured (for example due to changes in the law); or a subset of the variables can change in their importance or ability to predict the class.

In the case of supervised learning (classification or prediction), the extracted rules or dependencies can be no longer valid over time due to different changes which can occur. However, this statement is probably valid for unsupervised and reinforcement learning as well.

Learning in dynamically changing domains has a very high industrial relevance. During 1990-1993, the authors of this paper were involved in the ESPRIT Project StatLog which completed an evaluation of the performance of Machine Learning, neural and statistical algorithms on complex commercial and industrial problems. The overall aim was to give an objective assessment of the potential for classification algorithms in solving significant commercial and industrial problems, and to widen the foundation for commercial exploitation of these and related algorithms [9]. This project showed the shortcomings of ML-algorithms in dealing with real industrial and commercial large-scale applications [10]. Specifically, it turned out that one serious shortcoming is that the readily available ML-algorithms can not be applied to dynamically changing domains. Two such examples were examined:

The first is about a complex problem (introduced by Daimler-Benz) dealing with fault diagnosis of automatic transmissions. The goal was to show the practicability of symbolic ML algorithms to the end users and to convince them that ML would be a cheaper and faster approach to develop the next generation of the fault diagnosis systems. A classification system based on a decision tree approach was implemented which provided the desired accuracy rates. The results are reported in [9] in detail. After a while, however, it turned out that the generated rules had become completely invalid due to the changes of the parameters of the diagnosis device.

The second example deals with a complex problem from the credit industry. Using data about the users of mobile telephone cards [4], it was observed that the payment behaviour of customers changes over time and learned classification systems based on static approaches could not handle such dynamic changes.

Clearly, learning in dynamically changing domains is relevant to a wide range of industrial and commercial applications. Further examples which are based on time-dependent data and involve **prediction or classification** include:

- the analysis and prediction of various time series, such as the stock market and currency exchange;
- the analysis of data concerning buyers of new cars or other articles, in order to determine changing patterns in customer behaviour;
- development of different HELP-DESK applications, specially for technical diagnosis

There have been some attempts in the literature to address the problem of structural change in concepts and the dynamic aspects of data in Machine Learning. But most of the ML-algorithms can still not deal with this problem.

As described above, research in this direction has very important practical implications because structural change in concepts occurs often in the real-world domain. By considering this issue Machine Learning has a better chance of acceptance in industry and commerce.

There are many contributions that Statisticians have (already) made to this field, but their communication is hampered by a different terminology. For example, ML uses “unfamiliar” terms such as context learning, dynamic learning and theory revision, whereas statistical approaches to these problems include significance tests, structural change (used by econometricians) and multivariate regression methods. The question which remains is:

Have the Statisticians done everything; What is the challenge for ML? In which directions the research should go?

In the rest of the paper we try to discuss some of these issues. In Sect. 2 we give a short summary of available approaches in the literature. In Sect. 3 we discuss the challenge and need for further research in dealing with dynamically changing domains. In Sect. 4 we suggest some ideas which could be useful in performing further research in this area and we outline the expected results.

2 The State of the Art

Most of the literature on on-line, incremental learning has more or less tacitly assumed *stationary* environments; dynamic long- or short-term changes were not considered. For example, [17], [18] has developed an incremental algorithm that produces the same decision tree as would be found if all the examples had been available at once. In this case the sequence of the observations does not affect the final outcome.

It is only rather recently that the notions of concept drift and the influence of changing contexts have been directly addressed in the machine learning and statistics communities. Some specialized methods have been developed that try to detect and track changes under certain circumstances (see [20] and the references therein). Approaches so far developed include:

- using Statistical Process Control (SPC) methods to detect changes and adapt or modify rules as appropriate [13];
- using machine learning and statistical methods to learn to detect contextual clues and react accordingly when a change in context is suspected [20];
- using a “forgetting factor” whereby the rule is constantly updated using the most recently available observations [12].
- *incremental* (or *on-line*) learning with flexible ‘forgetting operators’ [21, 20, 19];

There are also links with exemplar and predictor evaluation methods in instance-based learning [1] [2] and to well-known updating methods of stochastic approximation as well as more standard time series models which use exponential smoothing.

To deal with dynamic aspects there are essentially two problems (in addition to those normally associated with classification). The first problem is to *detect* any change in the situation. In econometrics this is known as *structural change* and there are a variety of methods to test for change-points (at least in the univariate case), and most of these approaches are model-based. For a recent review of developments in this area see [5], references therein and other papers in that volume. The second problem is how to react to any detected change. At one extreme it could be that the current rule must be completely re-learned since the population drift has been quite sudden; at the other extreme we may conclude that there has been no change, in which case an optimal approach may be to improve the current rule using incremental learning, for example. However, between these extremes we should be able to update the existing rule by giving giving appropriate weight to the more recent observations.

A closely related topic is that of *drifting concepts*. For example, early work was done by [16] with his STAGGER system. See also [15] and more recent work by [7]. When formulated in a logical framework [3] argue that incremental concept-learning can be understood as an instance of the more general problem of belief updating. This insight allows for potential interesting cross-fertilization between these areas. Some general principles were given by [6] for the development of classification systems for dynamic objects with variable properties that are observed under changing conditions. He outlines the following requirements for dynamic classification systems: to handle changing composition of features, different feature types, and varying accuracy and statistical characteristics of observations; to give precedence in time and quantitatively change proximity of the object to a particular class; to rely on a single prior description of objects or processes. The approach, which is fairly general, uses conditional probabilities. He concludes by recommending a hierarchical system for classification of dynamic objects, retaining a catalogue that stores the results of past observations, and generalized distributions of features on one of the classification levels.

3 The Need for Further Research

There is so far no comprehensive evaluation that shows if available algorithms can handle real industrial and commercial applications. Questions of robustness have not been sufficiently addressed. Moreover, *batch* incremental learning (in which new observations are naturally grouped in some way) has hardly been studied at all in this context, although it plays a very important role specially in the case of large-scale datasets.

Also, reliable and robust formal *measures* of change and concept drift are still largely missing. A recent step in this direction is the developed metric in [14] for “structural instability”, a measure of the change in the decision rule as a result of incremental learning updates due to new observations. More general criteria are needed that can measure the different relevant dimensions of concept drift (extent, speed, etc.), that can reliably distinguish real changes from *noise*, and that can also recognize potential *structure* in changes (*e.g.*, oscillations, cyclic recurrence of contexts, and other regularities).

Finally, and perhaps most importantly, the available literature does not provide a *general framework* that would provide reliable guidelines for dealing with dynamic aspects. There is no substantial body of practical experience with drift and context effects in realistic application environments, nor is there a comprehensive collection of general or specialized methods with well-understood characteristics. Due to this reason, it is necessary to both develop generally applicable methods and integrate them into a general framework which covers the dynamic aspects not only of symbolic learning algorithms but also of statistical and neural learning as well as knowledge discovery. Furthermore, further research is needed to realize some of the ideas presented in [11], [20] and [8].

4 Some Ideas for Performing Further Research and the Expected Results

4.1 Suggested Directions for Further Research

The above discussion shows that although some attempts have been made to handle the problem of learning in dynamically changing domains, there still remains a lot of questions which are open to both the statistical and ML-communities. In this section, we suggest some ideas to handle some of these issues.

Generally speaking, the main goal of the further research should be developing practical methods for classification, prediction, forecasting, and knowledge discovery that effectively deal with both gradual concept drift and more abrupt context changes. The common core of these tasks is inductive generalization, and the methods to be developed (or extended) belong to the fields of **machine learning** and **statistics**. Further research should aim at methods that are as general as possible, so that they will be useful to such diverse (though related) tasks as

- induction of classification rules;
- prediction (including forecasting) of numerical values;
- detection of logical patterns via ILP;
- detection of (other) associations (e.g., functional dependencies, clustering, interesting projections) within the data;
- description of interesting classes, clusters and their inter-relationships.

Generally, adapting to changing circumstances can be done by on-line learning and/or by updating our knowledge at appropriate times, *i.e.* when a change (in context) has been detected. One can identify various subproblems for which appropriate methods and algorithms could be developed:

Drift/change detection: An obvious way to detect concept drift is to use a form of Statistical Process Control (SPC). The methodology typically assumes that the data are grouped into “batches” which are monitored over time, and appropriate *warning* and *action* limits must be obtained for a

parameter of interest. In the case of monitoring the class means it is necessary to take account of the correlations amongst the attributes. In the case of non-Normal data standard distribution results will not be available and nonparametric methods will be needed to set appropriate limits.

Adaptation / classifier update: Once a change has been detected, adaptation of the classifier (or predictor) is needed. At this stage it may be helpful to discover in which attributes the distribution is changing, and/or discover the change in context which is the underlying cause. The adaptation can take the form of complete re-learning (in the case that the change has been “large”) or a modification of some of the existing rules (in the case that the change has been “small”). The precise meaning of “large” and “small” needs to be determined. Methods for assessing this from characteristics of the learning task will have to be developed. Adaptation and classifier updating functions need to be developed both for batch-incremental and truly incremental (on-line) learning.

Characterization and/or explanation of change: An important component in any adaptation is that there are suitable diagnostics to enable the description of the changes which have occurred. The explicit description or characterization of changes by the learning system makes it possible to detect “meta-level” patterns such as recurring contexts. When this is achieved then a second layer of learning could be formulated in which the appropriate updating method can be selected based on the type of change. The explicit description of changes is also desirable in applications directed towards knowledge discovery, where the goal is not so much to have an efficient classifier as to gain an understanding of the data or process under study. Inductive ‘meta-learning’ algorithms as described in [20] might be a first step in this direction and should be further developed.

Tackling all these subproblems in a comprehensive way will require a non-trivial combination of statistical methods with symbolic machine learning and knowledge discovery approaches. This should be one of the explicit premises of further research. One possible starting point might be the use of memory-based learning methods like *k-Nearest-Neighbor* (*k-NN*), which are well-known from the statistical literature and are now being studied with increased interest in machine learning. Such “locally weighted learning” algorithms [2] have been shown to have the potential to track concept drift. In further research this direction should be explored.

It should be mentioned that in some cases the further development or adaptation of existing methods is necessary. These methods should be tested on a variety of real industrial and commercial complex problems with the aim of determining which aspects are best suited for each type of data. The developed framework should enable a fair comparison of the different approaches. It is likely that certain mathematical properties can be derived which characterize the data for which these methods are applicable. In addition, characteristics which indicate the preferred method should be found. The testing and assessment of methods should use agreed criteria which will include objective measures (error rates, time to learn, ability to adapt) as well as theoretical considerations.

4.2 Anticipated Results of Further Research

The types of results and output which can be expected from further research:

- General tests and measures:** (statistical) tests and criteria for detecting concept drift or potential context changes as well as measures of characteristics like extent, speed, etc. of concept drift;
- Algorithms for classifier adaptation / update / knowledge revision:** in particular, methods that optimize learned rules or concepts to fit a changed environment;
- On-line algorithms for continual drift tracking:** especially for prediction and forecasting tasks, on-line learning algorithms will be provided which adapt to changing concepts incrementally;
- Diagnostics:** formalisms and algorithms for the characterization and explanation of context changes;
- General framework and procedures:** this will include the integration of various components described in the previous section; within this general framework a range of new algorithms which incorporate the processes of monitoring, adaptation, and diagnostics will be completed;
- Informative results of comparative experiments:** competing algorithms will be evaluated systematically on complex data sets from real applications;
- Characterization of domains of applicability** of developed methods; general guidelines as to the application of learning methods in tasks involving concept drift will be developed;

5 Conclusions

An important component in further research is the interaction of researchers working in the fields of statistics and computer science. Previous attempts to bring these groups together (in the non-dynamic setting) have been very fruitful; see for example [9] and [11].

Acknowledgements: The authors would like to thank Pavel Brazdil and Gerhard Widmer for useful discussion on some issues described in this paper.

References

1. Aha, D., Kibler, D., & Albert, M.K. (1991). Instance-Based Learning Algorithms. *Machine Learning*, 6(1), 37–66.
2. Atkeson, C.G., Moore, A.W. and Schaal, S. (1997) Locally Weighted Learning. To appear in a special issue of Artificial Intelligence Review on Lazy Learning.
3. De Raedt, L. and Bruynooghe, M. (1992). Belief updating from integrity constraints and queries. *Artificial Intelligence* 53, 291–307.

4. Dübon, K. (1997). Dynamic Aspects in Risk Classification. To appear.
5. Dufour, J. M. and Ghysels, E. (1996) Editors' introduction: Recent developments in the econometrics of structural change. *Journal of Econometrics* **70**, 1–8.
6. Filosofov, L. V. (1992). Dynamic classification systems with learning from observation catalogs. *Cybernetics and Systems Analysis* **28**, 368–376.
7. Kubat, M. and Widmer, G. (1995). Adapting to drift in continuous domains. In *Lecture Notes in Artificial Intelligence*, Volume 912, pp. 307–310.
8. Mannila, H. (1995) Aspects of Data Mining. Keynote talk in: *Statistics, Machine Learning and Knowledge Discovery in Databases*, Kodratoff, Y., Nakhaeizadeh, G. and Taylor, C.C. (eds.), Heraklion, Crete. pp. 1–12.
9. Michie, D., Spiegelhalter, D. J. and Taylor, C. C. (eds) (1994) *Machine Learning, Neural and Statistical Classification*, Ellis Horwood, Chichester.
10. Nakhaeizadeh, G. (1995) What Daimler-Benz has learned as an Industrial Partner from the Machine Learning Project StatLog? In: Aha, D. and Riddle P. (Eds.). Proceedings of the Workshop on Applying Machine Learning in Practice at the Twelfth International Machine Learning Conference.
11. Nakhaeizadeh, G. and Taylor, C. C. (eds) (1997) *Machine Learning and Statistics: the Interface*. Wiley, New York.
12. Nakhaeizadeh, G., Taylor, C. C. and Kunisch, G. (1996) Dynamic Aspects of Statistical Classification. In *Intelligent Adaptive Agents*, AAAI Technical report No. WS-96-04, pp. 55–64. AAAI Press, Menlo Park, CA.
13. Nakhaeizadeh, G., Taylor, C. C. and Kunisch, G. (1997) Dynamic Supervised Learning: Some Basic Issues and Application Aspects. In *Classification, Data Analysis, and Knowledge Organisation*, Klar, B. & Opitz, O. (eds.). Springer Verlag.
14. Rissland, E. E., Brodley, C. E. and Friedman, M. T. (1995) Measuring structural change in concepts. Technical report, Department of Computer Science, University of Massachusetts, Amherst, MA 01003.
15. Schlimmer, J. C. (1987). Incremental adjustment of representations for learning. In *Fourth International Workshop on Machine Learning*, pp. 79–90. Irvine, CA: Morgan Kaufmann.
16. Schlimmer, J. C. and Granger, R. (1986). Incremental learning from noisy data. *Machine Learning* **1**, 317–354.
17. Utgoff, P. E. (1989) Incremental learning of decision trees. *Machine Learning* **4**, 161–186.
18. Utgoff, P. E. (1994) An improved algorithm for incremental induction of decision trees. In *Proceedings of Eleventh Machine Learning Conference*, Rutgers University. Morgan Kaufmann.
19. Widmer, G. (1994) Combining Robustness and Flexibility in Learning Drifting Concepts. *Proceedings of the Eleventh European Conference on Artificial Intelligence* (pp. 468–472). Wiley, Chichester.
20. Widmer, G. (1997) Tracking Context Changes through Meta-Learning *Machine Learning* (to appear). Draft version available electronically at <ftp://ftp.ai.univie.ac.at/papers/oefai-tr-96-22.ps.Z>.
21. Widmer, G. and Kubat, M. (1993) Effective Learning in Dynamic Environments by Explicit Context Tracking. *Proceedings of the Sixth European Conference on Machine Learning* (pp. 227–243). Berlin: Springer-Verlag.