
User Modeling and User-Adapted Interaction **11:** 19–29, 2001.

Machine Learning for User Modeling

GEOFFREY I. WEBB[1], MICHAEL J. PAZZANI[2] and DANIEL BILLSUS[2]
[1] *School of Computing and Mathematics, Deakin University, Geelong, Victoria 3217, Australia*
[2] *Department of Information and Computer Science, University of California, Irvine, Irvine, California 92697, U.S.A.*

(Received: 12 November 1999; in revised form 22 May 2000)

Abstract. At first blush, user modeling appears to be a prime candidate for straightforward application of standard machine learning techniques. Observations of the user's behavior can provide training examples that a machine learning system can use to form a model designed to predict future actions. However, user modeling poses a number of challenges for machine learning that have hindered its application in user modeling, including: the need for large data sets; the need for labeled data; concept drift; and computational complexity. This paper examines each of these issues and reviews approaches to resolving them.

Key words: user modeling, machine learning, concept drift, computational complexity, World Wide Web, information agents

1. Introduction

The past decade has seen research into the use of machine learning to support user modeling (ML for UM) pass through a period of decline and then resurgence, with the research area at the close of the twentieth century more active and vibrant than at any previous time. It is tempting to identify the start of the ML for UM winter as being marked by the publication of Self's (1988) paper in which he asserted that a search problem that appeared to underlie a direct machine learning approach to inferring possible cognitive process models for a relatively simple modeling task was 'clearly intractable'. While the paper did not argue that student modeling was intractable per se, the phrase 'the intractable problem of student modeling', taken from the title of that paper, has been oft repeated, perhaps with less attention to the finer detail of the argument within the paper than might be desired. Without needing to ascribe causes to the ML for UM winter, it is notable that it was preceded by a decade of much activity. Notable examples from this period include the work of Brown and Burton (1978), Brown and VanLehn (1980), Gilmore and Self (1988), Langley and Ohlsson (1984), Mizoguchi et al. (1987), Reiser et al. (1985), Sleeman (1984), VanLehn (1982), and Young and O'Shea (1981), much of it in the area of student modeling.* In contrast, the period 1988–1994 saw relatively little activity in the area. A strong resurgence is evidenced however by a special issue of this

* We consider student modeling to be a form of user modeling.

journal devoted to the subject (volume 8, numbers 1–2, 1998) the number of recent workshops on the subject (Bauer et al., 1997; Bauer et al., 1999; Joachims et al., 1999; Rudstorm et al., 1999; Papatheodorou, 1999), and sessions in major conferences (Goettl et al., 1998; Jameson et al., 1997; Kay, 1999; Lajoie and Vivet, 1999. It is, perhaps, tempting to equate the start of the thaw with the presentation of the best paper award to Martin and VanLehn's (1993) paper on student modeling at the 1993 World Conference on Artificial Intelligence in Education.

While the field was initially dominated by research on student modeling, the demands of electronic commerce and the world-wide-web have led to rapid growth in research in the area of information retrieval. With vast quantities of information available to all users on the web, the need for technologies to personalize the web has arisen.

This paper provides a brief overview of the application of machine learning for user modeling and reviews four critical issues that are currently limiting the real world application of user modeling and looks at the current state of attempts to overcome them. The four issues addressed are:

- the need for large data sets;
- the need for labeled data;
- concept drift; and
- computational complexity.

2. Machine Learning and User Modeling

The forms that a user model may take are as varied as the purposes for which user models are formed. User models may seek to describe

(1) the cognitive processes that underlie the user's actions;
(2) the differences between the user's skills and expert skills;
(3) the user's behavioral patterns or preferences; or
(4) the user's characteristics.

Early applications of machine learning in user modeling focused on the first two of these model types, with particular emphasis paid to developing models of cognitive processes. In contrast, recent research has predominantly pursued the third approach, focusing on users' behavior, as advocated by Webb (1993), rather than on the cognitive processes that underlie that behavior. Applications of machine learning to discovering users' characteristics remain rare.

Another important dimension along which it is important to distinguish approaches is with respect to whether they model individual users or communities of users. Whereas much of the academic research in ML for UM concentrates on modeling individual users, many of the emerging applications of ML for UM in electronic commerce relate to forming generic models of user communities.

For example, very substantial increases in purchases are claimed for systems that recommend products to users of retail web sites using models based on purchases by other users (as exemplified by Ungar and Foster, 1998).

Situations in which the user repeatedly performs a task that involves selecting among several predefined options appear ideal for using standard machine learning techniques to form a model of the user. One example of such a task is processing e-mail by deleting some messages and filing others into folders (Segal and Kephart, 1999). Another example is to determine which news articles to read from a web page (Billsus and Pazzani, 1999). In such situations, the information available to the user to describe the problem and the decision made can serve as the training data for a learning algorithm. The algorithm will create a model of a user's decision making process that can then be used to emulate the user's decisions on future problems. At first glance, it may be tempting to consider such user modeling problems as straightforward standard classification learning tasks. However, user modeling presents a number of very significant challenges for machine learning applications. The following sections address some of the key challenges that it poses.

3. The Need for Large Data Sets

The Syskill and Webert system (Pazzani and Billsus, 1997) is a straightforward implementation of a machine learning algorithm (a simple Bayesian classifier) applied to the problem of recommending web sites. As a user browses the web, the user indicates whether a web page is interesting (by clicking on a 'thumbs up' button on the web browser) or not interesting (by clicking on 'thumbs down'). The system then annotates unseen links on the web pages with an assessment of whether the user would be interested.

One important limitation of the straightforward application of machine learning systems such as Syskill and Webert to real world user modeling tasks is that the learning algorithm does not build a model with acceptable accuracy until it sees a relatively large number of examples (e.g. 50). In most situations, it is natural that learning algorithms require many training examples to be accurate (Valiant, 1984) since there are typically a large number of alternative models to select from. This problem is addressed in a variety of ways:

– Knowledge-based learning approaches, such as theory refinement (Baffes and Mooney, 1996), create a new model by modifying an initial model. If an accurate model of the user is close to the initial model, few examples may be required to transform accurately the initial model into the user model. This may be the case in student modeling where the initial model is the 'correct' model, and the student model to be acquired is close to the correct model. This assumes, however, that there is a single 'correct' model that can serve as a suitable initial model. Attempting to model incorrect performance as a perturbation of a

'correct' model that does not correspond to the basic underlying strategy of the user or student may be seriously misleading. For instance, there are several substantially different 'correct' procedures for the relatively simple skill of elementary subtraction (see, for example, Young and O'Shea, 1981). Minor perturbations of each of these procedures may result in substantial differences in predictions about future performance.

- Some approaches to learning (e.g. nearest neighbor algorithms) can be fairly accurate with a few examples if the new examples are very similar to the training examples. NewsDude (Billsus and Pazzani, 1999) takes advantage of this to recommend news stories that follow up on stories the user read previously.
- In some cases, it is possible to structure the task so that a learned model need not exactly replicate the user's decision. For example, the SwiftFile system (formerly known as MailCat, Segal and Kephart, 1999) does not automatically file mail into users' folders, but rather puts the three most likely folders for a message on a prominent place on the screen. By having more than one option available and not hindering the user from taking actions that were not anticipated, the system does not have to have an accurate model to be useful.

4. The Need for Labeled Data

Another difficulty confronting direct application of machine learning to many user modeling tasks is that the supervised machine learning approaches used require explicitly labeled data, but the correct labels may not be readily apparent from simple observation of the user's behavior. Consider again the example of Syskill and Webert. It would be very difficult to infer from a web user's browsing behavior which web pages they found interesting and which they did not. However, Syskill and Webert requires these labels in order to be able to make recommendations. The solution in this case has been to require the user to explicitly label the data by clicking a 'thumbs up' or 'thumbs down' button. The user must perform additional work to provide explicit feedback to the system (by clicking on a button) but is not provided with an immediate reward. Users rarely provide information to the modeling system if they must go out of their way or if they see no immediate benefit.

One approach to this problem is to infer the labels from the user's behavior. For example, the Letizia system (Lieberman, 1995) infers that a user is interested in a web page if a variety of actions are performed (e.g. printing the page or creating a bookmark), while the user is not interested under other circumstances (e.g. by quickly hitting the back button). Such implicit feedback methods allow a large amount of data to be collected unobtrusively. One can imagine future systems that would use the user's facial expression, body language or other forms of implicit feedback for this purpose.

Another approach to the problem is to use a small initial body of labeled examples to infer labels for a larger body of examples which is then used to train

the learning algorithm. This technique is related to the information retrieval method of pseudo-feedback (Kwok and Chan, 1998) in which first the system finds documents similar to the user's query and then it finds documents similar to the retrieved documents. However, in the machine learning approach (Nigam et al., 1998), the process of inferring the label for unseen documents is repeated until a stable solution is found via a procedure known as expectation maximization. As well as circumventing the problem of training sets sizes, as discussed in the last section, this technique reduces the demand on the user to label training cases by reducing the number of labeled cases that are required. These approaches are currently in their infancy but are likely to have a big impact on the field into the future.

5. Concept Drift

Early approaches to the use of machine learning for user modeling tended to develop new, special purpose, and frequently ad hoc, machine learning techniques to support their specific needs. More recently, there has been a tendency to seek an adequate problem representation in the form of training examples and corresponding class labels in order to be able to draw on well-known algorithms and results from the vast literature on classification learning. A potential pitfall of this methodology is that it might lead to solutions that are not specifically geared towards the unique characteristics of user modeling applications. For example, user modeling is known to be a very dynamic modeling task – attributes that characterize a user are likely to change over time. Therefore, it is important that learning algorithms be capable of adjusting to these changes quickly. From a machine learning perspective, this is a challenging problem known as *concept drift* (Widmer and Kubat, 1996).

In order to emphasize the importance of this problem and further clarify the issues involved, we report on recent developments in the use of user models for Information Retrieval (IR) applications.

As part of the advent of the World Wide Web and the recent resurgence of machine learning for user modeling research, IR tasks have received much attention. This problem is well illustrated by the demands of user modeling for information retrieval. The main objective is to learn a model of the user's interests or information need, in order to facilitate retrieval of relevant information. Most work on content-based information filtering casts the automated acquisition of user profiles as a text classification task (for example, Pazzani and Billsus, 1997; Lang, 1995; Mooney and Roy, 1998). In these systems, a set of text documents rated by the user (e.g. interesting vs. not interesting) is used as the input for a learning algorithm, and the resulting classifier can be interpreted as an automatically-induced model of the user's interests. An underlying assumption often made is that more training data leads to improved predictive performance. However, if we take into account that a user's interests are dynamic and are likely

to change over time, this assumption does not hold. A classifier built from a large number of training documents that accurately reflect the user's past interests is of limited practical use and might perform substantially worse than a classifier limited to recent data that reflects the user's current interests. This example illustrates that a good text classification algorithm is not necessarily a useful user modeling algorithm.

As researchers have begun to take the importance of concept drift for user modeling applications into account, a few initial solutions have emerged in the literature. A straightforward approach is simply to place less weight on older observations of the user (for example, Webb and Kuzmycz, 1996). However, there is some evidence that the effectiveness of this simple approach is constrained (Webb et al., 1997). Klinkenberg and Renz (1998) explore windowing techniques similar to ideas proposed by Widmer and Kubat (1996) in the context of Information Retrieval. The central idea is to limit training data to an adjustable time window, where the window size depends on observed indicators such as sudden changes in term distributions.

Chiu and Webb (1998) have studied the induction of dual user models as an approach for handling concept drift in the context of student modeling. In general, user modeling is a task with inherent temporal characteristics. We can assume recently collected user data to reflect the current knowledge, preferences or abilities of a user more accurately than data from previous time periods. However, restricting models to recent data can lead to overly specific models, i.e. models that classify instances that are similar to recently collected data with high precision, but perform poorly on instances that deviate from data used to induce the model. To overcome this problem, Chiu and Webb use a dual model that classifies instances by first consulting a model trained on recent data, and delegating classification to a model trained over a longer time period if the recent model is unable to make a prediction with sufficient confidence.

Billsus and Pazzani (1999) propose a related idea for personalized recommendation of news stories. A nearest-neighbor text classification algorithm built from recent observations forms a short-term model of the user's interests in daily news stories. In cases where the short-term model cannot make a prediction with sufficient confidence, classification is delegated to a more general classifier based on observations collected over a longer period of time. This architecture allows a system to adjust to interest changes rapidly, without sacrificing the potential benefits of data collection over longer time periods. Furthermore, this system tries to automatically anticipate a special case of concept drift: news stories that are presented to the user are assumed to directly affect the user's information need. As a result, the system tries to prevent presenting similar information multiple times, as it is assumed that a certain piece of information is only interesting once, and that the *concept* of what is considered interesting *drifts* at that time.

While a start has been made on tackling this challenging problem, this is an area in which more progress is required if user modeling is to realize its full potential.

6. Computational Complexity

The current ML for UM resurgence has witnessed tremendous research activity. In contrast, the field still has a dearth of fielded applications. The resulting difference between research interest and commercially deployed systems is especially apparent in the field of Internet-based applications. The growth of the Internet has had a tremendous impact on the field of ML for UM over the past decade, as researchers have realized the potential of learning techniques for automated information retrieval assistance, resulting in a surge in research on intelligent information agents. However, the actual impact of this technology on the average web user has been fairly limited. We speculate that one reason for this effect is the computational complexity of many approaches proposed in academic research. While the Internet has paved the way for new opportunities to assist users through the use of detailed user models, the sheer amount of information available as well as the number of users online has created new challenges. It is not uncommon for big portal sites (e.g. Yahoo, Excite or Lycos) to receive millions of visits per day. Clearly, if every one of these users were to be assisted through the use of automatically acquired user models, computational complexity would play a major role in the viability of user modeling on the Internet.

In contrast, academic research in machine learning is often dominated by a competitive race for improved predictive accuracy. When a new algorithm is proposed, it is not uncommon that an empirically measured increase of a fraction of a percent in predictive accuracy is considered a success if the result is statistically significant. While we realize that there are domains where these subtle accuracy improvements make a crucial difference, we think that ML for UM is not such a domain. For example, an algorithm that recommends interesting information with a predictive accuracy of 78% might be preferred over an algorithm that achieves 80%, if the former algorithm requires considerably less CPU time, and therefore allows for deployment in high-volume real-world scenarios.

At a first glance, the constraints imposed by the need for efficient user modeling algorithms seem to exclude many computationally expensive learning algorithms and data analysis techniques from consideration for user modeling tasks. For example, reducing the need for labeled training data through expectation maximization (Nigam et al., 1998) leads to improved predictive performance, but causes a significant increase in CPU time. However, computationally expensive algorithms can still be utilized if they can be applied in scenarios where models can be learned offline, i.e. without real-time constraints that would require short response times. Initial work with a focus on computational complexity and suitability for large-scale deployment is starting to emerge in the literature. While not strictly a machine learning approach, Jester 2.0 is a collaborative filtering system that models a user's taste in humor, based on similarities to other users' ratings for jokes (Gupta et al., 1999). The underlying idea of the proposed algorithm is to speed up the recommendation process through the use of a preprocessing step based on

principal component- and cluster analysis. Since the preprocessing step can be performed offline, online recommendations can be computed efficiently. We believe that this is a step in the right direction and hope that future research in this field will be geared towards techniques that are directly applicable to real-world applications in order to make the benefits of ML for UM available to a broad audience.

7. Conclusion

ML for UM has awoken from the winter of the early nineties with renewed strength and vigor, fueled largely by the demands of the internet and other emerging information retrieval technologies. However, despite clear potential and demand for ML for UM technologies, they remain primarily in the research domain. We are yet to witness the widespread appearance of fielded applications.

In this paper we have outlined four major issues that must be overcome before widespread application of ML for UM will be possible:

- the need for large data sets;
- the need for labeled data;
- concept drift; and
- computational complexity.

While the difficulty of these problems should not be underestimated, as we indicate, approaches to overcoming them are being actively pursued and strong progress has been made. Looking forward it appears evident that ML for UM is a research area on the cusp of coming-of-age and that by the time of the twentieth anniversary of this journal, ML for UM will have taken a place as a core technology underlying the information economy.

Acknowledgments

Pazzani's research on machine learning has been supported by National Science Foundation Grant 9731990. The paper has benefited from comments and suggestions by Alfred Kobsa, Ingrid Zukerman, and David Albrecht.

References

Baffes, P. and Mooney, R.: 1996, Refinement-based student modeling and automated bug library construction, *Journal of Artificial Intelligence in Education* **7**, 75–116.

Bauer, M., Pohl, W. and Webb, G. (eds.): 1997, *UM97 Workshop: Machine Learning for User Modeling*. Online proceedings: http://www.dfki.uni-sb.de/~bauer/um-ws/.

Bauer, M., Gmytrasiewicz, P. and Pohl, W. (eds.): 1999, *UM99 Workshop: Machine Learning for User Modeling*. Online proceedings: http://www.dfki.de/~bauer/um99-ws/.

Billsus, D. and Pazzani, M.: 1999, A hybrid user model for news story classification. In: *User Modeling* (*Proceedings of the Seventh International Conference*). Banff, Canada, pp. 99–108.

Brown, J. S. and Burton, R. R.: 1978, Diagnostic models for procedural bugs in basic mathematical skills. *Cognitive Science* **2**, 155–192.

Brown, J. S. and VanLehn, K.: 1980, Repair theory: A generative theory of bugs in procedural skills. *Cognitive Science* **4**, 379–426.

Chiu, P. and Webb, G.: 1998, Using decision trees for agent modeling: improving prediction performance. *User Modeling and User-Adapted Interaction* **8**, 131–152.

Gilmore, D. and Self, J.: 1988, The application of machine learning to intelligent tutoring systems. In: J. Self (ed.): *Artificial Intelligence and Human Learning: Intelligent Computer-Aided Learning*. London: Chapman and Hall, pp. 179–196.

Goettl, B., Halff, H., Redfield, C. and Shute, V. (eds.): 1998, *Intelligent Tutoring Systems: Fourth International Conference, ITS'98*. Berlin: Springer.

Gupta, D., DiGiovanni, M., Narita, H.and Goldberg, K.: 1999, Jester 2.0 : A new linear time collaborative filtering algorithm applied to jokes. In: *Proceedings of the SIGIR-99 Workshop on Recommender Systems: Algorithms and Evaluation.* Berkeley, CA.

Jameson, A., Paris, C. and Tasso, C. (eds.): 1997, *User Modeling* (*Proceedings of the Sixth International Conference UM97*). New York: SpringerWien.

Joachims, T., McCallum, A., Sahami, M. and Ungar, L. (eds.): 1999, *IJCAI Workshop IRF-2: Machine Learning for Information Filtering*. New York: IJCAI Inc.

Kay, J. (ed.): 1999, *User Modeling: Proceedings of the Seventh International Conference UM99*. New York: SpringerWien.

Klinkenberg, R. and Renz, I.: 1998, Adaptive information filtering: learning in the presence of concept drift. In: *AAAI/ICML-98 Workshop on Learning for Text Categorization. Technical Report WS-98-05*. Madison, Wisc.

Kwok, K. and Chan, M.: 1998, Improving two-stage ad-hoc retrieval for short queries. In: *Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. Melbourne, Australia, pp. 250–256.

Lajoie, S. P. and Vivet, M. (eds.): 1999, *Artificial Intelligence in Education: Proceedings of the Ninth International Conference AIEd-99*. Amsterdam: IOS Press.

Lang, K.: 1995, NewsWeeder: Learning to filter news. In: *Proceedings of the Twelfth International Conference on Machine Learning*. Lake Tahoe, CA., pp. 331–339.

Langley, P. and Ohlsson, S.: 1984, Automated cognitive modeling. In: *Proceedings of the Second National Conference on Artificial Intelligence*, pp. 193–197.

Lieberman, H.: 1995, An agent that assists web browsing. In: *Proceedings of the International Joint Conference on Artificial Intelligence*. Montreal, Canada, pp. 924–929.

Martin, J. and VanLehn, K.: 1993, OLAE: Progress toward a multi-activity, bayesian student modeler. In: *Proceedings of the 1993 World Conference on Artificial Intelligence in Education*. Edinburgh, Scotland, pp. 410–417.

Mizoguchi, R., Ikeda, M. and Kakushu, O.: 1987, An innovative framework for intelligent tutoring systems. In: *Artificial Intelligence Tools in Education*. Amsterdam-New York, pp. 105–120.

Mooney, R., Bennet, P. and Roy, L.: 1998, Book recommending using text categorization with extracted information. In: *AAAI/ICML-98 Workshop on Learning for Text Categorization.* Madison, Wisc.

Nigam, K., McCallum, A., Thrun, S. and Mitchell, T.: 1998, Learning to classify text from labeled and unlabeled documents. In: *Proceedings of the 15th International Conference on Artificial Intelligence*. Madison, Wisc., pp. 792–799.

Papatheodorou, C. (ed.): 1999, *Machine Learning and Applications Workshop W03, Machine Learning in User Modeling*, Chania, Greece.

Pazzani, M. and Billsus, D.: 1997, Learning and revising user profiles: The identification of interesting web sites. *Machine Learning* **27**, 313–331.

Reiser, B. J., Anderson, J. R. and Farrell, R. G.: 1985, Dynamic student modelling in an intelligent tutor for LISP programming. In: *Proceedings of the Ninth International Joint Conference on Artificial Intelligence*. Los Angeles, CA, pp. 8–14.

Rudstorm, A., Bauer, M., Iba, W. and Pohl, W. (eds.): 1999, *IJCAI Workshop ML-4: Learning About Users*. IJCAI Inc.

Segal, R. and Kephart, M.: 1999, MailCat: An intelligent assistant for organizing e-mail. In: *Proceedings of the Third International Conference on Autonomous Agents*. Seattle, WA, pp. 276–282.

Self, J. A.: 1988, Bypassing the intractable problem of student modelling. In: *Proceedings of the Intelligent Tutoring Systems Conference*. Montreal, pp. 107–123.

Sleeman, D. H.: 1984, Inferring student models for intelligent computer-aided instruction. In: R. S. Michalski, J. G. Carbonell, and T. M. Mitchell (eds.): *Machine Learning: An Artificial Intelligence Approach*. Berlin: Springer-Verlag, pp. 483–510.

Ungar, L. H. and Foster, D. P.: 1998, Clustering methods for collaborative filtering. In: *AAAI-98 Workshop on Recommender Systems*. Madison, Wisc.

Valiant, L. G.: 1984, A theory of the learnable. *Communications of the ACM* **27**, 1134–1142.

VanLehn, K.: 1982, Bugs are not enough: empirical studies of bugs, impasses, and repairs in procedural skills. *Journal of Mathematical Behavior* **3**, 3–72.

Webb, G. I.: 1993, Feature based modelling. In: *Proceeding of AI-ED 93, World Conference on Artificial Intelligence in Education*. Edinburgh, Scotland, pp. 497–504.

Webb, G. I., Chiu, B. C. and Kuzmycz, M.: 1997, Comparative evaluation of alternative induction engines for Feature Based Modelling. *International Journal of Artificial Intelligence in Education* **8**, 97–115.

Webb, G. I. and Kuzmycz, M.: 1996, Feature Based Modelling: A methodology for producing coherent, consistent, dynamically changing models of agents' competencies. *User Modeling and User-Adapted Interaction* **5**(2), 117–150.

Widmer, G. and Kubat, M.: 1996, Learning in the presence of concept drift and hidden contexts. *Machine Learning* **23**, 69–101.

Young, R. M. and O'Shea, T.: 1981, Errors in childrens' subtraction. *Cognitive Science* **5**, 153–177.

Authors' Vitae

Geoff Webb is Professor of Computer Science at Deakin University, founder and director of G. I. Webb and Associates Pty. Ltd. and director of the Deakin University Priority Area of Research in Information Technology for the Information Economy. He received his B.A. and Ph.D. degrees in Computer Science from La Trobe University. He has worked in several areas of artificial intelligence, including machine learning, knowledge acquisition, and user modeling.

Michael Pazzani is a professor and the chair of the Information and Computer Science Department at the University of California, Irvine. His research interests include data mining and intelligent agents. He received his BS and MS in computer

engineering from the University of Connecticut and his PhD in computer science from UCLA. He is a member of the AAAI and the Cognitive Science Society.

Daniel Billsus received a diploma in computer science from the Technical University of Berlin, Germany, and M.S. and Ph.D. degrees from the University of California, Irvine. His research focus has been in the area of intelligent information access. He studied the use of machine learning techniques as part of various information agents, leading to his doctoral dissertation on ''User Model Induction for Intelligent Information Access''.