Predictive Statistical Models for User Modeling

INGRID ZUKERMAN and DAVID W. ALBRECHT
School of Computer Science and Software Engineering, Monash University, Clayton, VICTORIA 3800, AUSTRALIA, {ingrid,dwa}@csse.monash.edu.au

Abstract. The limitations of traditional knowledge representation methods for modeling complex human behaviour led to the investigation of statistical models. Predictive statistical models enable the anticipation of certain aspects of human behaviour, such as goals, actions and preferences. In this paper, we motivate the development of these models in the context of the user modeling enterprise. We then review the two main approaches to predictive statistical modeling, content-based and collaborative, and discuss the main techniques used to develop predictive statistical models. We also consider the evaluation requirements of these models in the user modeling context, and propose topics for future research.

1. Introduction

User modeling involves inferring unobservable information about a user from observable information about him/her, e.g., his/her actions or utterances. To perform this task, a user modeling system must deal with the uncertainty attendant to making inferences about a user in the absence of complete information. In particular, the area of plan recognition has been concerned with making inferences about a user's preferences, goals and forthcoming actions and locations.

Early user modeling systems relied on hand-crafted knowledge bases to make inferences from observations about users. In particular, early plan recognition systems used hand-crafted plan libraries to postulate a user's intentions or preferences from his/her utterances (Carberry, 2000). These knowledge bases were usually built by carefully analyzing several instances of the problem at hand which were deemed to be representative of this problem. However, these knowledge bases suffer from two main shortcomings: their construction is a resource-intensive process, and usually they are not adaptable or extendable. In Artificial Intelligence, the problem of obtaining such knowledge bases came to be called the *knowledge bottleneck problem*. This problem was further exacerbated with the advent of technologies that enabled novel applications such as automatic message forwarding, system-initiated assistance for document editing, or recommendations of films or WWW pages. These applications can generate large quantities of data (from electronic logs of many users) which are often vitiated by noise, such as interruptions and false starts (noise was typically omitted from the analyzed examples considered in the traditional approach). In addition, the users of these applications are often not actively cooperating with the system. Hand-building a knowledge base that is representative of many data points is clearly more labour intensive than hand-building such a knowledge base when only a few examples are being considered. Further, when the data are vitiated, the difficulty of this task is compounded. The vitiation of the data together with the possible lack of cooperation from users also increases the uncertainty associated with making predictions from observed data.

The need to address the knowledge bottleneck problem and the uncertainty in user modeling, plus the wealth of data generated by recent application domains pointed towards statistical models as a promising alternative to the traditional approach for user modeling. Statistical models are concerned with the use of observed sample results (which are observed values of random variables) in making statements about an unknown, dependent parameter (Larson, 1969). In *predictive statistical models for user modeling*, this parameter represents an aspect of a user's future behaviour, such as his/her goals, preferences, and forthcoming actions or locations.

The Artificial Intelligence areas of machine learning and reasoning under uncertainty have generated a variety of techniques that fall under the umbrella of predictive statistical models, such as decision trees, neural networks and Bayesian networks. The predictions made by these techniques have been used to adapt the behaviour of a system, e.g., modifying its dialogue strategy (Litman and Pan, 2000); to recommend objects a user may be interested in, e.g., news items (Jennings and Higuchi, 1993; Billsus and Pazzani, 1999) or films (Alspector et al., 1997); and to perform actions on behalf of a user, e.g., pre-sending WWW pages (Albrecht et al., 1999) or forwarding email (Macskassy et al., 1999). Two main approaches have been adopted to perform these tasks: content-based and collaborative. The former is based on the tenet that each user exhibits a particular behaviour under a given set of circumstances, and that this behaviour is repeated under similar circumstances. The latter is based on the tenet that people within a particular group tend to behave similarly under a given set of circumstances. Thus, in the content-based approach, the behaviour of a user is predicted from his/her past behaviour, while in the collaborative approach, the behaviour of a user is predicted from the behaviour of other like-minded people. Representative systems that implement these approaches are discussed in Section 2.

An important issue for user modeling systems in general, and for predictive statistical models in particular, pertains to the evaluation of these systems. Predictive statistical models for user modeling inherit evaluation norms and requirements from two disciplines: machine learning and user modeling. Typically, machine learning evaluations consist of dividing a data set into a training set and a test set, using the former to learn the model, and the latter to evaluate the model's performance. This methodology has been applied to the predictive user models developed to date with respect to several different measures: recall and precision, predicted probability and accuracy, and utility (Section 3). Once the predictive statistical models have passed the "machine learning test", the systems based on these models should be subjected to a "user modeling evaluation". However, contrary to machine learning evaluations, at present, there is no generally accepted methodology for the evaluation of systems which employ a user model.

Finally, the use of predictive statistical models for user modeling is a new and promising development. In Section 4, we consider limitations of the predictive statistical models developed to date, and suggest avenues of investigation towards the next generation of predictive statistical models.[1]

2. Current Research in Predictive Statistical Models

As indicated above, the use of predictive statistical models for user modeling is a relatively recent occurrence, prompted by the large quantities of electronically available data and by advances in machine learning. As a result, at present the field appears as a landscape dotted with a variety of representation methods and applications. We view this as an initial step in the process of gaining deeper insights regarding the application of predictive statistical models to user modeling.

In this section, we describe the two main approaches adopted to build predictive statistical models: content-based and collaborative. We then consider the main techniques currently used to represent predictive statistical models, drawing a distinction between their usage as content-based or collaborative tools. Finally, we discuss studies which compare

[1] Additional challenges posed to machine learning by user modeling applications are discussed in (Webb et al., 2000).

the performance of different predictive statistical models and the performance of the two main approaches.

2.1. Content-based Learning versus Collaborative Learning

Content-based learning is used when a user's past behaviour is a reliable indicator of his/her future behaviour. In this approach, a predictive model is built for a user using data from his/her past behaviour. For example, consider the domain of film recommendations, and imagine there is a user (Fred) who indicated he likes "Star Wars", "Raiders of the Lost Ark" and "Air Force One". The content-based approach learns the types of films enjoyed by Fred – action films starring Harrison Ford – and based on this will recommend another film of this type, e.g.,"Witness". Content-based models are particularly suitable for situations where users tend to exhibit an idiosyncratic behaviour. However, this approach requires a system to collect relatively large amounts of data from each user in order to enable the formulation of a statistical model.

Collaborative learning is used whenever one can assume that a user behaves in a similar way to other users. In this approach, a model is built using data from a group of users, and it is then used to make predictions about an individual user. Going back to the above film recommendation system, the collective approach finds that Fred's taste in films is similar to that of a particular group of users, and will recommend other films enjoyed by the users in this group. The collaborative approach is useful when trying to make a prediction about a new user (once s/he has been identified as a member of a group) or about a known user in a new situation (where the information known about the user does not support the formulation of a prediction). For instance, consider a user visiting a particular WWW site. In the absence of information about this user, a predictive model built with the collaborative approach will use its information regarding the habits of all visitors to the site in order to predict the WWW page the user is most likely to request next. Now, even if the site had a content-based model built on the basis of the WWW pages visited previously by this user, this model would not be able to make predictions if the user starts visiting a completely different set of WWW pages.

2.2. Predictive Statistical Models

Several different statistical models have been used under both the content-based and the collaborative approach. The main models are: linear models, TFIDF-based models, Markov models, neural networks, classification and rule-induction methods,[2] and Bayesian networks.

Linear models

Linear models have a simple structure, which makes them easily learnable, and also enables them to be easily extended and generalized. Linear models take weighted sums of known values to produce a value for an unknown quantity. For example, consider using the collaborative approach to build a linear model that predicts a user's rating for news articles. In this model, for each candidate article, the known values may be the ratings assigned to this article by other users, and the weights may be a measure of the similarity between the user in question and the other users. The resulting linear model is the weighted sum of the ratings (such a model is described in Resnick *et al.*, 1994). Linear models have also been

[2] Classification tasks are performed by unsupervised learning techniques, while rule-induction tasks are performed by supervised learning techniques.

used under the content-based approach, e.g., to predict the time intervals between a user's successive logins (Orwant, 1995), and to predict a user's ratings of films (Raskutti et al., 1997).

TFIDF-based models

The TFIDF (Term Frequency Inverse Document Frequency) method is a weighting scheme commonly used in the field of Information Retrieval to find documents that match a user's query (Salton and McGill, 1983). This method represents a document by a vector of weights, where each weight corresponds to a term in the document. The similarity between two documents (or between a document and a query) is then measured by the cosine of the angle at the origin which subtends the vectors corresponding to these documents. Balabanović (1998), Moukas and Maes (1998) and Basu *et al.* (1999) applied TFIDF-based models in content-based systems that recommend documents to a user based on other (similar) documents of interest to this user. Moukas and Maes extended this approach in that they used genetic algorithms to automatically adapt their recommender system to a user's (possibly changing) requirements.

Markov models

Like linear models, Markov models have a simple structure. This is due to their reliance on the *Markov assumption* to represent sequences of events (according to this assumption, the occurrence of the next event depends only on a fixed number of previous events). Given a number of observed events, the next event is predicted from the probability distribution of the events which have followed these observed events in the past. For example, when the task at hand consists of predicting WWW pages to be requested by a user, the last observed event could be simply the last visited WWW page or it could contain additional information, such as the link which was followed to visit this page or the size of the document. Bestavros (1996) and Zukerman *et al.* (1999) used Markov models under the collaborative approach in order to predict users' requests on the WWW. Bestavros' model calculated the probability that a user will ask for a particular document in the future, while Zukerman *et al.* (1999) compared the predictive performance of different Markov models which calculate the probability that a user will ask for a particular document in the following request. The predictions generated by these models were then used by systems which pre-send to a user documents s/he is likely to request (Bestavros, 1996; Albrecht et al., 1999).

Neural networks

Neural networks are capable of expressing a rich variety of non-linear decision surfaces. This is done through the structure of the networks, non-linear thresholds and the weights of the edges between the nodes. Jennings and Higuchi (1993) used neural networks under the content-based approach to represent a user's preferences for news articles. For each user, they learned a neural network where the nodes represent words that appear in several articles liked by the user and the edges represent the strength of association between words that appear in the same article.

Classification

Classification methods partition a set of objects into classes according to the attribute values of these objects. Given an n-dimensional space that corresponds to the attributes under consideration, the generated clusters or classes contain items that are close to each other in this space and are far from other clusters. Classification methods are unsupervised in the sense that there is no *a priori* information regarding the class to which each item belongs.

Under the collaborative approach, Perkowitz and Etzioni (2000) used a variation of traditional clustering in order to automatically create index pages which contain links to WWW pages that are related to each other (these are pages that users tend to visit during the same session). Their classification technique, which they called *cluster mining*, finds a small number of high-quality clusters (rather than partitioning the entire space of documents), and can place a document in several overlapping clusters.

Rule induction

Rule induction consists of learning sets of rules that predict the class of an observation from its attributes. The techniques used for rule induction differ from those used for classification in that during training, rule induction techniques require the class of each observation as well as its attributes.[3] The models derived by these techniques can represent rules directly, or represent rules as decision trees or in terms of conditional probabilities.

Rule-induction techniques have been used under both the content-based and the collaborative approach. Under the content-based approach, Morales and Pain (1999) used Ripper, a system that learns rules from set-valued features (Cohen, 1996), to learn rules that predict a user's next action in an experiment where the user has to balance a pole on a cart. Chiu and Webb (1998) combined C4.5, a rule-induction technique which builds decision trees (Quinlan, 1993), with Feature Based Modeling, an attribute-value modeling method designed for tutoring applications (Webb and Kuzmycz, 1996), to predict features of subtraction errors performed by students. Joerding (1999) used CDL4, a semi-incremental algorithm that learns rules (Shen, 1997), to learn users' media preferences for product presentations in a WWW shopping environment. Billsus and Pazzani (1999) applied a mixture of rule-induction methods and TFIDF-based and linear models to recommend news articles to a user. Their system used two models to anticipate whether a user would be interested in a candidate article. One model maintained a TFIDF vector representation of the articles in the system's knowledge base, and used only those articles that were similar to the candidate article in order to build a linear model that predicts whether the user will be interested in this article. This technique is particularly useful when building an initial model on the basis of limited data, since only a few news articles are required to identify possible topics of interest. The other model applied a naive Bayesian classifier (Duda and Hart, 1973) to a Boolean feature vector representation of the candidate article, where each feature indicates the presence or absence of a word in the article. This classifier calculates the probability that an item belongs to a particular class (e.g., the class of articles a user finds interesting) under the assumption that the attributes of the items in a given class are independent.

Under the collaborative approach, Basu *et al.* (1998) used Ripper to learn a set of rules which predict whether a user will like or dislike a film, and Litman and Pan (2000) used Ripper to learn a set of rules that adapt the dialogue strategy used by a spoken dialogue system. Gervasio *et al.* (1998) used ID3 (Quinlan, 1986) to learn a decision tree that predicts which action will be performed next by a user working on a scheduling problem.

Bayesian networks

Bayesian networks (BNs) (Pearl, 1988) and various extensions of BNs have steadily been gaining popularity in the Artificial Intelligence community, and have been used for a variety of user modeling tasks (Jameson, 1996). BNs are directed acyclic graphs where nodes correspond to random variables. The nodes are connected by directed arcs, which may be thought of as causal links from parent nodes to their children. Each node is associated with

[3] In principle, this class may be automatically derived using a classification method prior to performing rule induction.

a conditional probability distribution which assigns a probability to each possible value of this node for each combination of the values of its parent nodes. BNs are more flexible than the models discussed above in the sense that they provide a compact representation of any probability distribution, they explicitly represent causal relations, and they allow predictions to be made about a number of variables (rather than a single variable, which is the normal usage of the above models). In addition, BNs can be extended to include temporal information (dynamic Bayesian networks, Dean and Wellman, 1991) and utilities (influence diagrams, Howard and Matheson, 1984).

An important property of BNs is that they support the combination of the collaborative and the content-based approach. The collaborative approach may be used to obtain the conditional probability tables and the initial beliefs of a BN. These beliefs can then be updated in a content-based manner when the network is accessed by a user. This mode of operation enables a predictive model to overcome the data collection problem of the content-based approach (which requires large amounts of data to be gathered from a single user), while at the same time enabling the tailoring of aspects of a collaboratively-learned model to a single user.

BNs have been used to perform a variety of predictive tasks. Horvitz *et al.* (1998) used a BN to predict the type of assistance required by users performing spreadsheet tasks. Albrecht *et al.* (1998) compared the performance of several dynamic Bayesian networks which predict a user's next action, next location and current quest in a Multi-user Adventure Game. Lau and Horvitz (1999) built a BN which models search queries on the WWW and predicts the type of query-related action a user will perform next, e.g., generalize or further specify a query. Finally, Gmytrasiewicz *et al.* (1998) and Jameson *et al.* (2000) used influence diagrams to predict agents' behaviour. Gmytrasiewicz *et al.* considered various models that predict an agent's actions in an air-defense scenario, and incrementally updated the probability assigned to each model according to its predictive accuracy. Jameson *et al.* predicted the error rates of users when following instructions given in a certain style (e.g., "several together" versus "one at a time"), and selected an instruction style that minimizes this error rate.

2.3. Comparative Studies of Predictive Models

Ideally, one would like to determine the most suitable representation method for a particular application based on the features of the problem at hand (Section 4). However, in the absence of such information, an empirical comparison of the performance of different techniques is warranted. Such empirical studies have been performed in the framework of both the content-based and the collaborative approach.

Chiu *et al.* (1997), Davison and Hirsh (1998) and Macskassy *et al.* (1999) performed comparative studies of predictive models under the content-based approach. Chiu *et al.* compared the predictive performance of the models learned by two rule induction techniques, C4.5 (Quinlan, 1993) and FFOIL (Quinlan, 1996), in a system that anticipates the features of the result obtained by a student when performing subtraction. Davison and Hirsh compared the performance of the decision tree learned by C4.5 with that of a Markov model in a system that predicts a user's next UNIX command. An interesting feature of Davison and Hirsh's Markov model is that it was built incrementally, giving greater weight to more recent events in order to increase the system's sensitivity to changes in a user's behaviour. According to Chiu *et al.*'s study, the decision tree learned by C4.5 made more predictions and more accurate predictions than the rules learned by FFOIL, while Davison and Hirsh reported that their incrementally learned Markov model performed at least as well as the decision tree learned by C4.5. Macskassy *et al.* performed a preliminary comparison of the recommendations generated by a naive Bayes classifier, two TFIDF-based models, a rule-based model inferred using Ripper (Cohen, 1996), and a voting scheme in a system that determines which email messages should be forwarded to a user's personal pager. However, owing to the use of the pager (which is a prototype), only a few users could be involved in this study. Hence, its results are as yet inconclusive.

A collaborative recommender system for three different domains, WWW pages, television programs and films, is described in (Breese et al., 1998). This system was used as a platform for comparing the predictive performance of several linear models (with different weighting schemes), a BN and a naive Bayes classifier. Breese *et al.*'s results indicate that BNs outperform the other methods for a wide range of conditions.

A different type of comparative study was performed by Alspector *et al.* (1997) for the domain of film recommendations. They compared the performance of a recommender system built under the collaborative approach against that of a system built under the content-based approach. In addition, they considered two linear models under the collaborative approach, and linear networks (which are mixture of linear models) and decision trees under the content-based approach. Their results showed that the models obtained using the collaborative approach performed significantly better than those obtained using the content-based approach, and that among the content-based models the linear networks performed better than the decision trees. Alspector *et al.* also identified the following limitations of each approach: collaborative methods cannot be applied to new items (which no one has rated) nor to users which haven't been assigned to a group, while content-based methods require careful feature selection. These results led them to conclude that a (film) recommendation system should combine the content-based and the collaborative approach.

3. Evaluation Methods

To date, predictive statistical models used for user modeling have been evaluated using mainly the following techniques: recall and precision, which are borrowed from the field of Information Retrieval; and predicted probability, accuracy and utility, which are sourced from machine learning.

The recall and precision measures are particularly suitable for recommender systems (e.g., Raskutti et al. 1997; Basu et al. 1998; Billsus and Pazzani 1999). Recall measures the proportion of items of interest recommended by a system among the items of interest in the system's knowledge base, and precision measures the proportion of items of interest among the items recommended by the system. Thus, most of the predictive models that use these evaluation measures require users to provide ratings for all the items in the system's knowledge base. Ideally, a predictive model should have both high recall and high precision. However, current systems typically trade off these measures against each other.

Accuracy and predicted probability are measures used to evaluate models that predict a user's actions, locations or goals. Accuracy calculates the percentage of times the event that actually occurred was predicted with the highest probability (over several trials), while predicted probability returns the average of the probabilities with which this event was predicted (over several trials). Accuracy and variants thereof have been widely used (e.g., Breese et al. 1998; Chiu and Webb 1998; Gervasio et al. 1998; Davison and Hirsh 1998; Morales and Pain 1999), while both predicted probability and accuracy were used in (Albrecht et al., 1998) to compare the performance of different predictive models. The results obtained by Albrecht *et al.* show that predicted probability provides finer-grained information about the performance of a predictive model than accuracy. This is because for each trial, accuracy returns mainly a binary value (0 when the probability of the actual event is lower than that of any other event, and 1 when the probability of the actual event exceeds that of all the other events), while predicted probability returns the probability with which the actual event was predicted.

Finally, utility is a measure of the benefit derived from using a particular system – in our context, a system that uses a predictive model. This measure requires a function that represents the advantage resulting from a correct action performed by the system and the disadvantage resulting from a wrong action (Breese et al., 1998; Albrecht et al., 1999). Utility-based evaluations constitute an indirect evaluation of a predictive model, since they evaluate an action performed on the basis of the predictions made by such a model (and different protocols may be used to determine this action). In addition, utility-based evaluations are closer to user-based evaluations than the techniques described above, since they take into account at least some of the user's requirements.

As stated above, the only measures used to evaluate predictive statistical models for user modeling have been those inherited from machine learning. This is a first step in the evaluation of these models, since the validity of a model must be determined by means of intrinsic evaluations before usability studies can be conducted. However, even before such studies are considered, the different evaluation measures must be revisited in order to determine which features of a predictive model are evaluated by means of each measure. This will support the selection of a coherent suite of evaluation measures to assess different aspects of system performance.

4. Thoughts for the Future

An implicit working assumption of many predictive models is that of "persistence of interest" (Lieberman, 1995), whereby users maintain their behaviour or interests over time. This is true over the life time of user models intended for short term use, and also for certain user characteristics, e.g., love of classical music in an adult. However, as observed by Webb and his colleagues (Webb and Kuzmycz, 1996; Chiu and Webb, 1998), Davison and Hirsh (1998) and Moukas and Maes (1998), users also have interests and behaviours that change over time, and predictive models should be able to adapt to these changes. To this effect, these researchers incorporated the following incremental adaptations into their models. Both Webb *et al.* and Davison and Hirsh incorporated information regarding the recency of an event into their models, and Moukas and Maes used genetic algorithms. BNs also exhibit incremental adaptive behaviour, since they can update the probability distributions represented in the nodes as a result of interactions with a user. Other approaches to achieve adaptivity would involve retraining a predictive model whenever its performance deteriorates beyond a certain threshold, or retraining it periodically. In addition, hybrid adaptive processes which combine several approaches may be worth investigating.

At present, the predictive statistical models used for user modeling adopt two main learning approaches: content-based and collaborative. Each of these approaches has clear advantages and disadvantages. The content-based approach is ideal for tailoring a system's behaviour to the specific requirements of a particular user. However, this approach requires each user to provide relatively large amounts of data to enable the construction of a statistical model. In addition, the features selected when implementing this approach have a substantial effect on the usefulness of the resulting model. Features that are too specific yield a system that is useful only for repetitive behaviours, while features that are too general yield predictions of debatable usefulness.

The collaborative approach reduces the data collection burden for individual users, and can be implemented using the specific values of the data (without obtaining features with the “right” level of abstraction). However, since this approach makes predictions about the behaviour of a single user from observations of many users, it does not support tailoring a system to the requirements of a particular user. Further, most systems that implement this approach conflate all users into a model that represents an “average” user. This prevents them from modeling the behaviour of different types of users, e.g., undergraduate students versus researchers. To remedy this situation, predictive statistical models built with the collaborative approach must be extended to model the characteristics of groups of users, as described in (Alspector et al., 1997; Horvitz et al., 1998). These enhanced models can then make more accurate predictions about the behaviour of individual users by matching these users to a particular group. Nonetheless, even with the improved predictive accuracy of these models, the fundamental inability of the collaborative approach to represent the idiosyncrasies of individual users calls for a solution which combines both types of modeling approaches (Alspector et al., 1997; Delgado and Ishii, 1999).

Finally, an important issue that emerges from the preceding discussion pertains to the variety of techniques being applied to build predictive statistical models. At present, the only justification given for preferring a particular modeling technique is its empirical success, either in isolation or compared with that of other methods. While such a justification is acceptable, the time may be ripe for the field of machine learning to engage in a theoretical investigation regarding the suitability of different techniques for different operating conditions. The main question that must be answered in the framework of such an investigation is: which modeling technique is most suitable in light of the features of a given problem and the available data?

Acknowledgments

The authors would like to thank Sandra Carberry, Tony Jameson, Diane Litman and Geoff Webb for their thoughtful comments.

Authors’ Vitae

Ingrid Zukerman

Monash University, School of Computer Science and Software Engineering, Clayton, Victoria 3800, Australia

Ingrid Zukerman is an Associate Professor in Computer Science at Monash University. She received her B.Sc. degree in Industrial Engineering and Management and her M.Sc. degree in Operations Research from the Technion – Israel Institute of Technology. She received her Ph.D. degree in Computer Science from UCLA in 1986. Since then, she has been working in the School of Computer Science and Software Engineering at Monash University. Her areas of interest are discourse planning, plan recognition and agent modeling.

David Albrecht

Monash University, School of Computer Science and Software Engineering, Clayton, Victoria 3800, Australia

David Albrecht is a Lecturer in Computer Science at Monash University. He received his B.Sc and Ph.D. degrees in Mathematics from Monash University. Since 1995 he has been working in the School of Computer Science and Software Engineering. His areas of research include plan recognition and machine learning.

References

Albrecht, D. W., I. Zukerman, and A. E. Nicholson: 1998, 'Bayesian Models for Keyhole Plan Recognition in an Adventure Game'. *User Modeling and User-Adapted Interaction* **8**(1-2), 5–47.

Albrecht, D. W., I. Zukerman, and A. E. Nicholson: 1999, 'Pre-sending Documents on the WWW: A Comparative Study'. In: *IJCAI99 – The Sixteenth Joint International Conference on Artificial Intelligence*. Stockholm, Sweden, pp. 1274–1279.

Alspector, J., A. Koicz, and N. Karunanithi: 1997, 'Feature-Based and Clique-Based User Models for Movie Selection: A Comparative Study'. *User Modeling and User-Adapted Interaction* **7**(4), 279–304.

Balabanović, M.: 1998, 'Exploring versus Exploiting when Learning User Models for Text Recommendation'. *User Modeling and User-adapted Interaction* **8**(1-2), 71–102.

Basu, C., H. Hirsh, and W. Cohen: 1998, 'Recommendation as Classification: Using Social and Content-Based Information in Recommendation'. In: *AAAI98 – Proceedings of the Fifteenth National Conference on Artificial Intelligence*. Madison, Wisconsin, pp. 714–720.

Basu, C., H. Hirsh, W. Cohen, and C. Nevill-Manning: 1999, 'Recommending Papers by Mining the Web'. In: *Proceedings of the IJCAI99 Workshop on Learning about Users*. Stockholm, Sweden, pp. 1–11.

Bestavros, A.: 1996, 'Speculative Data Dissemination and Service to Reduce Server Load, Network Traffic and Service Time in Distributed Information Systems'. In: *Proceedings of the 1996 International Conference on Data Engineering*.

Billsus, D. and M. Pazzani: 1999, 'A Hybrid User Model for News Stories Classification'. In: *UM99 – Proceedings of the Seventh International Conference on User Modeling*. Banff, Canada, pp. 99–108.

Breese, J. S., D. Heckerman, and C. Kadie: 1998, 'Empirical Analysis of Predictive Algorithms for Collaborative Filtering'. In: *UAI98 – Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence*. Madison, Wisconsin, pp. 43–52.

Carberry, S.: 2000, 'Plan Recognition: Achievements, Problems, and Prospects'. *User Modeling and User-Adapted Interaction* **(this issue)**.

Chiu, B., G. I. Webb, and M. Kuzmycz: 1997, 'A Comparison of First-order and Zeroth-Order Induction for Input-Output Agent Modelling'. In: *UM97 – Proceedings of the Sixth International Conference on User Modeling*. Sardinia, Italy, pp. 347–358.

Chiu, B. C. and G. Webb: 1998, 'Using Decision Trees for Agent Modeling: Improving Prediction Performance'. *User Modeling and User-Adapted Interaction* **8**(1-2), 131–152.

Cohen, W. W.: 1996, 'Learning Trees and Rules with Set-valued Features'. In: *AAAI96 – Proceedings of the Thirteenth National Conference on Artificial Intelligence*. Portland, Oregon.

Davison, B. and H. Hirsh: 1998, 'Predicting Sequences of User Actions'. In: *Notes of the AAAI/ICML 1998 Workshop on Predicting the Future: AI Approaches to Time-Series Analysis*. Madison, Wisconsin.

Dean, T. and M. P. Wellman: 1991, *Planning and Control*. San Mateo, California: Morgan Kaufmann Publishers.

Delgado, J. and N. Ishii: 1999, 'Formal Models for Learning a User's Preferences, a Preliminary Report'. In: *Proceedings of the IJCAI99 Workshop on Learning about Users*. Stockholm, Sweden, pp. 13–20.

Duda, R. and P. Hart: 1973, *Pattern Classification and Scene Analysis*. New York, New York: John Wiley & Sons.

Gervasio, M., W. Iba, and P. Langley: 1998, 'Learning to Predict User Operations for Adaptive Scheduling'. In: *AAAI98 – Proceedings of the Fifteenth National Conference on Artificial Intelligence*. Madison, Wisconsin, pp. 721–726.

Gmytrasiewicz, P. J., S. Noh, and T. Kellog: 1998, 'Bayesian Update of Recursive Agent Models'. *User Modeling and User-Adapted Interaction* **8**(1-2), 49–69.

Horvitz, E., J. Breese, D. Heckerman, D. Hovel, and K. Rommelse: 1998, 'The Lumiere Project: Bayesian User Modeling for Inferring the Goals and Needs of Software Users'. In: *Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence*. Madison, Wisconsin, pp. 256–265.

Howard, R. and J. Matheson: 1984, 'Influence diagrams'. In: R. Howard and J. Matheson (eds.): *Readings on the Principles and Applications of Decision Analysis*. Menlo Park, California: Strategic Decisions Group, pp. 721–762.

Jameson, A.: 1996, 'Numerical Uncertainty Management in User and Student Modeling: An Overview of Systems and Issues'. *User Modeling and User-Adapted Interaction* **5**, 193–251.

Jameson, A., B. Großmann-Hutter, L. March, and R. Rummer: 2000, 'Creating an Empirical Basis for Adaptation Decisions'. In: *IUI2000 – Proceedings of the International Conference on Intelligent User Interfaces*. New Orleans, Louisiana, pp. 149–156.

Jennings, A. and H. Higuchi: 1993, 'A User Model Neural Network for a Personal News Service'. *User Modeling and User-Adapted Interaction* **3**(1), 1–25.

Joerding, T.: 1999, 'Temporary User Modeling for Adaptive Product Presentation in the Web'. In: *UM99 – Proceedings of the Seventh International Conference on User Modeling*. Banff, Canada, pp. 333–334.

Kobsa, A. and W. Wahlster: 1989, *User Models in Dialog Systems*. Springer-Verlag.

Larson, H. J.: 1969, *Introduction to Probability Theory and Statistical Inference*. Wiley International Edition.

Lau, T. and E. Horvitz: 1999, 'Patterns of Search: Analyzing and Modeling Web Query Refinement'. In: *UM99 – Proceedings of the Seventh International Conference on User Modeling*. Banff, Canada, pp. 119–128.

Lieberman, H.: 1995, 'Letizia: An Agent that Assists Web Browsing'. In: *IJCAI95 – Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence*. Montreal, Canada, pp. 924–929.

Litman, D. and S. Pan: 2000, 'Predicting and Adapting to Poor Speech Recognition in a Spoken Dialogue System'. In: *AAAI2000 – Proceedings of the Seventeenth National Conference on Artificial Intelligence*. San Antonio, Texas.

Macskassy, S. A., A. A. Dayanik, and H. Hirsh: 1999, 'EmailValet: Learning User Preferences for Wireless Email'. In: *Proceedings of the IJCAI99 Workshop on Learning about Users*. Stockholm, Sweden, pp. 45–50.

Morales, R. and H. Pain: 1999, 'Modelling of Novices' Control Skills with Machine Learning'. In: *UM99 – Proceedings of the Seventh International Conference on User Modeling*. Banff, Canada, pp. 159–168.

Moukas, A. and P. Maes: 1998, 'Amalthaea: An Evolving Multiagent Information Filtering and Discovery System for the WWW'. *Autonomous Agents and Multi-Agent Systems* **1**.

Orwant, J.: 1995, 'Heterogenous Learning in the Doppelgänger User Modeling System'. *User Modeling and User-Adapted Interaction* **4**(2), 107–130.

Pearl, J.: 1988, *Probabilistic Reasoning in Intelligent Systems*. San Mateo, California: Morgan Kaufmann Publishers.

Perkowitz, M. and O. Etzioni: 2000, 'Towards Adaptive Web Sites: Conceptual Framework and Case Study'. *Artificial Intelligence* **118**(1-2), 245–275.

Quinlan, J. R.: 1986, 'Induction of Decision Trees'. *Machine Learning* **1**, 81–106.

Quinlan, J. R.: 1993, *C4.5: Programs for Machine Learning*. San Mateo, California: Morgan Kaufmann Publishers.

Quinlan, J. R.: 1996, 'Learning First-order Definitions of Functions'. *Artificial Intelligence Research* **5**, 139–161.

Raskutti, B., A. Beitz, and B. Ward: 1997, 'A Feature-based Approach to Recommending Selections based on Past Preferences'. *User Modeling and User-Adapted Interaction* **7**(3), 179–218.

Resnick, P., N. Iacovou, M. Sushak, P. Bergstrom, and J. Riedl: 1994, 'GroupLens: An Open Architecture for Collaborative Filtering of Netnews'. In: *CSCW'94 Proceedings – the Conference on Computer Supported Collaborative Work*. pp. 175–186.

Salton, G. and M. McGill: 1983, *An Introduction to Modern Information Retrieval*. McGraw Hill.

Shen, W. M.: 1997, 'Active and Semi-Incremental Learning with Bordering Counterexamples'. Technical Report USC-ISI-97, Information Sciences Institute, Los Angeles, California.

Webb, G. I. and M. Kuzmycz: 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.

Webb, G. I., M. J. Pazzani, and D. Billsus: 2000, 'Machine Learning for User Modeling'. *User Modeling and User-Adapted Interaction* **(this issue)**.

Zukerman, I., D. W. Albrecht, and A. E. Nicholson: 1999, 'Predicting Users' Requests on the WWW'. In: *UM99 – Proceedings of the Seventh International Conference on User Modeling*. Banff, Canada, pp. 275–284.