An Empirical Study of Dynamic Bayesian Networks for User Modeling

Alexander Kuenzer¹, Christopher Schlick², Frank Ohmann¹, Ludger Schmidt¹, and Holger Luczak¹

Institute of Industrial Engineering and Ergonomics, Aachen University of Technology, Bergdriesch 27, 52062 Aachen, Germany
{a.kuenzer, f.ohmann, l.schmidt, h.luczak}@iaw.rwth-aachen.de
² Research Institute for Communication, Information Processing and Ergonomics, Neuenahrer Strasse 20, 53343 Wachtberg-Werthhoven, Germany schlick@fgan.de

Abstract. Six topologies of dynamic Bayesian Networks are evaluated for predicting the future user events: (1) Markov Chain of order 1, (2) Hidden Markov Model, (3) autoregressive Hidden Markov Model, (4) factorial Hidden Markov Model, (5) simple hierarchical Hidden Markov Model and (6) tree structured Hidden Markov Model. Goal of the investigation is to evaluate, which of these models has the best fit for modeling the prediction of rule-based interaction behavior for a real domain. Case study of the experiments is a multimodal user interface for supervisory control of advanced manufacturing cells. A group of experienced users were observed while executing a typical task, to build a data basis for the evaluation. The results show that the number of user cases has high influence on the prediction quality and that there are no significant differences in using Markov Chain of order 1, factorial or tree structured Hidden Markov Models.

1 Introduction

This study deals with theoretical aspects of software ergonomic design of human-machine systems. According to the ISO 9241 standard (part 10), seven ergonomic principles of dialog design are distinguished. In the following sections the sixth design principle 'suitability for individualization' is analyzed in detail. The suitability for individualization deals with adapting structure and dynamics of the user interface to human information processing. With regard to Rasmussen's (1986) model of human information processing three levels of cognitive control can be taken into account: skill-based behavior, rule-based behavior, and knowledge-based behavior. We focus on a dynamic individualization shaped by rule-based behavior, which corresponds to the syntactic layer of the user interface and represents the lower symbolic level of cognitive control, which is characterized by conscious state recognition and access to stored rules from past work scenarios.

2 Background

2.1 Syntactic User Modeling

Modeling human information processing has a long tradition in human-machine interaction (Sheridan and Ferrell 1974, Pew and Baron 1983). Compared to other facets of human performance modeling the state of scientific knowledge of models of skill-based behavior is excellent concerning both, the quality of experimental data and the maturity of the mathematical formalism (details in Stein 1992 and Sheridan 1992).

Models of rule-based behavior or syntactic user models do not have reached a comparable state of scientific knowledge yet. A survey of Rouse et al. (1989) reviews a few symbolic approaches, but they deal with a deterministic behavior description with the help of switching functions, finite-state machines or pushdown automata, and therefore do not consider stochastic approaches to machine learning. To cope with behavioral uncertainty due to individually preferred interaction paths or due to human errors often empirical input data as a set of symbolic time series representing event-driven interaction sequences is used. In the process theory the technical term 'traces' applies for them (e.g. Schneider 2000). Popular models of stochastic processes are: (1) Markov Chains (Cassandras and Lafortune 1999), (2) Hidden Markov Models (Bengio 1999) and (3) Dynamic Bayesian Networks (Dean and Kanazawa 1988, Ghahramani 1998). A survey of models of stochastic processes is given in Minka (2000).

With regard to human-machine interaction Zukerman et al. (1999) use Markov Chains of first and second order to predict user request for Web pages. Cadez et al. (2000) link first order Markov Chains with a clustering technique. Their research aims at identifying behavioral homogeneous user groups to ease the customer-centered administration of Web sites. Borges and Levene (1999) identify navigation patterns of Web users with the help of Markov Chains of first up order and up to fifth order. Gorniak and Poole (2000) predict future user action when interacting with unmodified educational applications with a k-nearest neighbors algorithm.

Yang et al. (1997) use Hidden Markov Models in order to learn and transfer human skills when interacting with robots. Lane (1999) also investigates behavioral user models with the help of this model. He aims at discovering anomalies when using the UNIX shell for intruder detection. Takada et al. (1994) introduce a Markovian Programmer Activity Monitor which uses the logged key clicks of programmers when interacting with shell, editor, compiler, debugger, etc. Finally, Hidden Markov Models are a standard method of computer-supported speech processing (Cole et al. 1996).

There are numerous examples of Bayesian Network approaches to declarative or static representation of user models. Some references are Jameson (1996), Conati et al. (1997) or Horvitz et al. (1998). However, an approach to dynamic user modeling with Bayesian Networks is introduced by Schaefer and Weyrath (1997) dealing with flexible user interface design in a time critical environment such as a firefighter department. Lau and Horvitz (1999) use Bayesian Networks with discrete temporal

variables in order to model search patterns of users when refining request for Web search engines.

2.2 Dynamic Bayesian Networks (DBN)

The following sections are bound to modeling and predicting the observations of an stationary interaction process with dynamic Bayesian Networks (DBN) and do not deal with event timing. The network graph of DBN represents event flow forward in time (Ghahramani 1998). Therefore the random variables are assigned a time index (t=1,...,T) and the dependencies among time slices are modeled with the help of directed edges. The dynamic model is "clamped" by root nodes and consecutive time slices are 'enrolled' horizontally. Each time slice can hold observable variables as well as non-observable variables, which might be vertically linked. Hence, DBN are able to model interaction sequences with both temporal and spatial structure at multiple resolutions.

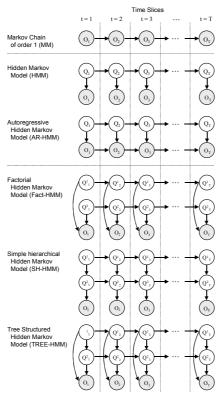


Fig. 1. Investigated topologies of dynamic Bayesian Networks (Hidden Nodes are white)

A semi-formal and layered approach to DBN topologies is the following (figure 1):

- ☐ A single layer DBN is equivalent to a Markov Chain as shown for an order 1 model (MM).
- ☐ A double layer DBN might integrate a hidden stochastic process for increased expressiveness, where in each time slice a single observable variable is linked with a single non-observable variable. Two topologies are derived: first, the classic Hidden Markov Model (HMM) and second, the autoregressive HMM (AR-HMM).
- □ A triple layer DBN might integrate two hidden stochastic processes and therefore has the potential to capture the underlying state space of the interaction process more efficiently than double layer models. Two simple topologies are shown in: first, the factorial HMM (FACT-HMM) and second, the simple hierarchical HMM (SH-HMM).
- Moreover, complex hierarchical DBN topologies can be designed such as the tree structured HMM (TS-HMM) which is of interest, because it is an approach to probabilistic decision trees with Markovian dynamics (Jordan et al. 1994).

Details of the selected model topologies are described in Luczak et al. 2001. For a more formalistic view of dynamic Bayesian Networks see e.g. Friedman et al. 1998.

To parameterize the selected models a predefined DBN topology is needed. A computation task is then to iteratively adopt the conditional probability tables according to the interaction cases. Murphy (2000) introduces a variety of estimation methods for fixed model topology. The first principle behind these methods is a maximum likelihood estimation with hidden variables called Expectation-Maximation. In the following sections an estimation method of Boyen and Koller (1998) is preferred, which was implemented by Murphy (2000).

3 Investigation of Syntactic User Models

3.1 Goals and Methodology

Goal of the investigation is to evaluate, which of the introduced stochastic models has the best fit for modeling the prediction of rule-based interaction behavior for real interaction tasks. The interaction task was derived from a case study of an user interface for supervisory control of future manufacturing systems and an refers to an earlier simulation study (Luczak et al. 2001).

3.2 Case Study: ACTIVE-UI as a Multimodal User Interface

Case study is a self-developed multimodal user interface (ACTIVE-UI, see fig. 2) for supervisory control of future 3D-laser welding cells (details in Schlick 2000). The pragmatic user interface layer is designed to span a metaphoric interaction space representing a virtual laser welding cell. With the help of ACTIVE-UI the operator is supported at the following manufacturing tasks: (1) configuration of welding robot; (2) setup of networked sensors and numerical control; (3) selection of manufacturing order; (4) workshop-oriented simulation of material processing; (5) control of real welding process; (6) ex-post process diagnosis.

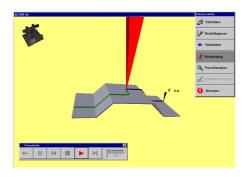
It is important to point out, that the logged user events represent the abstract interaction level mentioned in the task network in figure 3. Therefore our application deals with functional and object based events (such as Document.Print) and not with atomic events like key presses (click button "Print"), mouse movements or window events which are normally supplied by the user interface toolkit. Thus an implicit filtering process was implemented (see survey of Hilbert and Redmiles 1999).

3.3 Subjects and Interaction Tasks

Totally 30 subjects (9 female and 21 male) with experience in ACTIVE-UI were chosen. Their age extends from 21 to 42 years (mean=26,4 years). They all had to

complete the typical task "single workpiece processing" (see fig. 3) without time pressure.

Experienced users were confronted with the task 'single workpiece processing'. A part of the corresponding hierarchic-sequential task action network (according to Hacker 1973) of the task is shown in figure 3. Several process values had to be checked in order to give some necessary subgoals as default. The first subgoal is to initialize and configure two sensors and load the adequate numerical control program. The next subgoal is aiming at the workshop-oriented simulation of the numerical control program. After this the needed process values can be checked. The following subgoal deals with configuring and executing the real material processing with the help of the welding robot. Finally some further process values can be checked after the sensor-scales are parameterized. The sequences of most subgoals are not predefined within the dialog model and were expected with regard to human action variability expressing preferences.



display, where animation shows processing of workpiece is processed

Fig. 2. Screenshot of ACTIVE-UI's micro- Fig. 3. Part of the hierarchically-sequentially the task action network (HTA) of the task "single workpiece processing"

3.4 **Experimental Design**

To evaluate the introduced stochastic models the dependent variable 'prediction accuracy' (PA) of user's next procedural subgoal is considered. The PA is defined according to equation 1 as the average relative number of correctly predicted subgoals regarding the whole set of interaction cases for a single prediction lead. Similar computations of the PA can be found in other works (e.g. Zukerman et al. 1999b).

$$PA(\{\vec{o}_{L}\}, \lambda) = \frac{1}{L} \sum_{i=1}^{L} \frac{1}{|\vec{o}_{l}|} \sum_{i=1}^{|\vec{o}_{l}|} \begin{cases} 1 & P_{\lambda}(o_{l,i} \mid o_{l,1:i-1}) = \max_{\hat{o} \in O} (P_{\lambda}(\hat{o} \mid o_{l,1:i-1})) \\ 0 & otherwise \end{cases}$$
(1)

Independent variable is the stochastic model. With regard to the background section the primary independent variable has six levels: (1) Markov Chain of Order 1 (MM), (2) HMM, (3) autoregressive HMM (AR-HMM), (4) factorial HMM (FACT-HMM), (5) simple hierarchical HMM (SH-HMM) and (6) tree structured HMM (TS- HMM). All models had thirty-four hidden states on each level, which represent the number of distinguishable interaction events in the case study, except for the factorial and the simple hierarchical HMM, which both had two states on the upper level. It is important to point out that these six model structures can be regarded as minimal structures for their topological class concerning the case study. The resulting complexity of the transition tables and the prior vectors depending on the number of different events is shown in figure 4. Additionally the complexity of our earlier simulation study (Luczak et al. 2001) is pointed out in this figure for comparison.

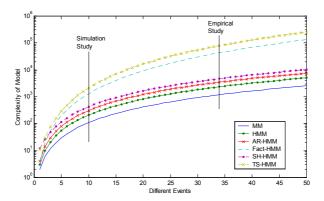


Fig. 4. Complexity as size of transition tables and prior vectors for the used stochastic models (Please note that the complexity is logarithmic).

The upper bound and the lower bound of the dependent variable PA can be defined easily: lower bound is a pure random process generating actual interaction events independently from prior events and identically distributed over the event set.

The random process has a PA, which is equivalent to the reciprocal value of the number of modeled interaction events. ACTIVE-UI case study encodes 34 linked events. Hence, the lower bound is $PA_{LowerBound} = 1/34 \approx 2.9\%$. Upper bound is an optimal prediction model, that is $PA_{UpperBound} = 1$.

3.5 Procedure

The logged interaction events of the users were used to build and to evaluate the 6 stochastic models for predicting the rule-based interaction behavior. This was achieved by using a cross-validation. The 30 traces were randomly spread into 6 groups, each containing 5 traces. In an estimation phase 5 groups of traces were used to estimate the parameters of the six stochastic models. The model parameters were iterated with the help of the Expectation-Maximation algorithm until the logarithmic likelihood increased less than one per mill. Prior probability vectors and the transition probability tables were initialized with random values. In the following evaluation phase the generated model base was assessed with regard to the *PA*. Therefore, the *PA* of the 5 traces of the remaining group was computed for each of the stochastic models in the model base. To avoid the effect of an unsuitable distribution of the traces in the

groups the computation of both the estimation and evaluation phase was repeated for 5 replications. Each replication contained of an randomly permutation of the traces of the data pool.

The statistical procedure was the following: a one way analysis of variance (ANOVA) of the dependent variable PA for 5 replications was computed. Null hypothesis was that the mean PA of the stochastic models are equal. The alternative hypothesis postulates a significant difference of at least a single pair of models. A significance level of $\alpha_{ANOVA} = 0.05$ was chosen. If the null hypothesis of the ANOVA can be rejected, the next question will be, which pairs of models show significant differences. These partial null hypothesis will be tested with the post-hoc Tukey test ($\alpha_{Tukey} = 0.05$).

3.6 Results and Discussion

The average duration of the complete task took 346 seconds (208 to 598 seconds). During this time they made 14 to 29 steps (median is 24 steps) where they used 14 to 26 different actions (median is 22 different actions). In general the actual version of ACTIVE-UI provides 34 different actions.

The results of the explorative investigation are shown in figures 5 and 6. Both the boxplot and the confidence-diagram indicate a superior PA of the MM (PA = 0.4826), the FACT-HMM (PA = 0.4720) and the TREE-HMM (PA = 0.4657). The HMM (PA = 0.4345) and the SH-HMM (PA = 0.4188) show an intermediate PA. The AR-HMM (PA = 0.3325) shows an inferior PA. Surprising is the good PA of the simple Markov Chain of order 1, which can be regarded as a minimal stochastic model. An explanation for this phenomenon is the relatively small sample size. Therefore an effect of 'model overfitting' seems to come into play. These explanations can be confirmed if the results of a previous simulation study are regarded, in which the other models could predict the interaction pattern in a better way than the stochastic model characterized by the simple Markov Chain of order 1. If the complexity of the models is taken into account, then especially the PA of the FACT-HMM and the TREE-HMM is good, even though the almost significant bigger number of model parameters (figure 4) is compared with the simple hierarchical HMM.

Although the differential effects of model overfitting and model underfitting are easy to distinguish, the absolute effects are rather weak, because the boxplot also indicates a surprising high level of the PA range for the complete set of stochastic models. The PA range is approximately centered around 46% and has a lower bound of 20% and an upper bound of 60%. These values emphasize that the small number of user cases is important for the PA of the models, because in earlier results the margin of PA was significant smaller (lower bound 67%, upper bound 82%) and therefore quite more robust.

The one way ANOVA computed a F-value of $F_{ANOVA} = 16.745$ (df = 5). Since the probability value of the F-test exceeds the critical threshold ($p_{ANOVA} < 0.001$) the null hypothesis was rejected. Hence, the set of models has significant PA differences and a post-hoc analysis was computed. The Tukey test indicates that the differences among MM and HMM, TREE-HMM, FACT-HMM are not significant but that the MM has a

significantly better mean PA ($\alpha_{Tukey} = 0.05$) than AR-HMM and SH-HMM. In addition the SH-HMM, HMM, TREE-HMM and FACT-HMM also do not demonstrate significant accuracy differences. The auto regressive Hidden Markov Model has significantly the worst PA.

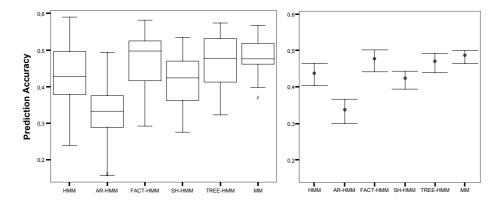


Fig. 5. Boxplots of prediction accuracy. For each Fig. 6. Diagram of the confidential model the lower and upper lines are the 25th and 75th intervals (95%) of the mean prediction percentiles, the line in the middle is the sample accuracy for the different models. median. Circle indicate outliers, which are more than 1.5 times the interquartile range.

4 **Conclusion and Outlook**

In the preceding sections an empiric approach to syntactic user modeling with discrete Bayesian Networks was presented. We used the stochastic models to predict user's rule-based behavior when interacting with software systems. Therefore an interaction scenario from computer-aided manufacturing in conjunction with a corresponding typical execution task model was used, which relies on simple observable variables and does not take more complex user or task parameters into account. A one way ANOVA in conjunction with statistical post-hoc tests demonstrated the significant prediction superiority of three different models of dynamic Bayesian Networks, namely the simple first order Markov Model, the factorial and the Tree structured Hidden Markov Model. It is remarkable, that especially the PA of the first order Markov Model is so good compared to the complexity of the other two models (whose complexity is about 67 times bigger than the first order Markov Model's). We suppose that this effect could be influenced by "model overfitting". Comparing to a previous simulated investigation of this interaction scenario we saw that the structure of the model and also the number of interaction cases and events had a significant effect to the predication accuracy. Thus it is necessary to balance the appropriate model for a special prediction domain regarding the particular circumstances.

Future research should investigate whether the significant different PA of the stochastic models also implies efficiency differences when dynamically individualizing ACTIVE-UI in real interaction scenarios. For this we want to focus more on the characteristics of manipulated objects. Additionally we want to study models, which support these demands and also can be learned by smaller numbers of user cases. One possibility could be that the prevention of over-fitting is quite enough to increase the prediction quality, another possibility could be the manual construction of the models to avoid the standard learning algorithms, which could by too nonspecific. Otherwise we want to examine the models in a different domain where it should be easier to gain higher number of user cases. Another goal for research will be problems with user expectations when predicting the appropriate initiative time to communicate predefined sequences of interaction events (user interface design principle 'strive for consistency'). This is an important open question, which was intendedly excluded from our study. Moreover, there is a need to investigate which dialog mode should be preferred for the dialog initiative. Artificial characters are an innovative dialog mode which is often mistakenly associated with interface agents, but these characters can be very inefficient for expert users.

Nevertheless, DBN are a powerful technique for modeling and predicting user's informational state and the dynamic individualization of user interfaces appears to be a fruitful domain for theory-driven and utility-driven research in ergonomics.

5 Acknowledgements

Parts of the research are funded by the German Research Foundation (Deutsche Forschungsgemeinschaft) according to the collaborative research center 368: Autonomous Production Cells. In addition, Kevin Murphy deserves a special acknowledgement for the cost-free supply of his Bayes Net Toolkit.

References

- 1. Bengio, Y.: Markovian Models for Sequential Data. Neural Computing Surveys 12, (1999) 129-162.
- Borges, J. and Levene, M.: Data Mining of User Navigation Patterns. Proceedings of the Workshop on Web Usage Analysis and User Profiling (WEBKDD'99) 31-36, August 15, San Diego, CA, URL: http://www.cs.ucl.ac.uk/staff/J.Borges/Home.html (1999)
- 3. Boyen, X. and Koller R, D. 1998: Tractable Inference for Complex Stochastic Processes. In: Proceedings Uncertainty in Artificial Intelligence (UAI)
- 4. Cadez, I., Heckerman, D., Meek, C., Smyth, P. and White, S.: Visualization of Navigation Patterns on a Web Site Using Model Based Clustering. Microsoft Research Technical Report MSR-TR-00-18 (2000)
- Cassandras, C.G. and Lafortune, S.: Introduction to Discrete Event Systems. Kluwer, Boston, MA (1999)
- Cole, R., Mariani, J., Uszkoriet, A., Zaenen, A. and Zue, V.: Survey of State of the Art in Human Language Technology. URL: http://cslu.cse.ogi.edu/HLTsurvey/HLTsurvey.html (1996)
- Conati, C, Gertner, A.S., Vanlehn, K. and Druzdel, M.J. 1997: Online Student Modeling for Coached Problem Solving Using Bayesian Network. In: Jameson, A., Paris, C., Tasso, C. (ed.): User Modeling -Proceedings of the Sixth International Conference UM 1997. Springer, New York (1997)

- Dean, T. and Kanazawa, K.: Probabilistic temporal reasoning. Proceedings of the Seventh National Conference on Artificial Intelligence. American Assoc. for Artificial Intelligence, Sta. Paul, MN (1988), 524-528.
- Ghahramani, Z.: Learning Dynamic Bayesian Network. In: Giles, C.L. and Gori, M. (Eds.): Adaptive Processing of Temporal Information. Springer, New York, MA (1998)
- Gorniak, P. and Poole, D.: Predicting Future User Actions by Observing Unmodified Applications. Proceedings of the Seventeenth National Conference on Artificial Intelligence. AAAI Press, Cambridge, MA (2000)
- 11. Hacker, W.: Allgemeine Arbeits- und Ingenieurpsychologie. Huber, Bern (1973)
- Hilbert, D.M. and Redmiles, D.F.: Extracting Usability Information from User Interface Events. Technical Report UCI-ICS-99-40. Department of Information and Computer Science, University of California, Irvine (1999)
- 13. Horvitz, E., Breese, J., Heckermann, D., Hovel, D. and Rommelse, K.: The Lumiere Project: Bayesian User Modeling for Inferring the Goals and Needs of Software Users. Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence. Morgan Kaufman, San Francisco, CA (1998)
- 14. ISO 9241, Part 10: Ergonomic Requirements for Office Work with Visual Display Terminals-Dialogue principles. International Organization for Standardization, Geneve (1996)
- Jameson, A. 1996: Numerical Uncertainty Management in User and Student Modeling: An Overview of Systems and Issues. User Modeling and User-Adapted Interaction 1 (5) (1996) 193-251.
- Jordan, M., Ghahramani, Z. and Saul, K.L.: Hidden Markov decision trees. MIT Computational Science Technical Report 9805 (1994)
- Lane, T. 1999: Hidden Markov Models for Human/Computer Interface Modeling. In: Proceedings of the IJCAI-99 Workshop on Learning about Users (1999) 35-44.
- Lau, T. and Horvitz, E.: Patterns of Search: Analyzing and Modeling Web Query Refinement. Microsoft Research Report, http://www.microsoft.research.com (1999)
- Luczak, H., Schlick, C., Kuenzer, A., Ohmann, F.: Syntactic User Modeling with Stochastic Processes. Theoretical Issues in Ergonomics Science, Vol. 2, No.1, (2001)
- 20. Murphy, K. 2000: Dynamic Bayesian Networks. URL: http://www.cs.berkeley.edu/~murphyk/-Bayes/usage_dbn.html (2000)
- 21. Pew, R.W. and Baron, S.: Perspectives on Human Performance Modeling. Automatica, 19 (6) (1983) 663-676
- Rasmussen, J.: Information Processing and Human-Machine Interaction: An Approach to Cognitive Engineering. North-Holland, New York (1986)
- 23. Rouse, W.B., Hammer, J.M. and Lewis, C.M.: On Capturing Human Skills and Knowledge: Algorithmic Approaches to Model Identification. IEEE Transactions on Systems, Man, and Cybernetics, 19 (3) (1989) 558-573
- 24. Schaefer, R. and Weyrath, T. 1997: Assessing Temporally Variable User Properties with Dynamic Bayesian Networks. In: Jameson, A., Paris, C. and Tasso, C. (eds.): User Modeling - Proceedings of the Sixth International Conference UM. Springer, New York (1997)
- 25. Schlick, C. 2000: Simulation of Rule-Based Behavior for a Multimodal Interaction Task with Stochastic Petri Nets. Proceedings of the XIVth Triennal Congress of the IEA and 44th Annual Meeting of the HFES 604-607. Human Factors and Ergonomics Society, Santa Monica, CA
- 26. Schneider, S.: Concurrent and Real-time Systems. John Wiley & Sons Chichester (2000)
- 27. Sheridan, T. B. and W. R. Ferrell, W. R.: Man-Machine Systems. MIT Press (1974)
- Sheridan, T.B.: Telerobotics, Automation, and Human Supervisory Control. The MIT Press Cambridge, MA (1992)
- Stein, W.: Mensch-Maschine-Systeme. Analyse, Simulation und Gestaltung. Vorlesungsskript des Instituts für Automatisierungstechnik der Ruhr-Universität Bochum (1992)
- 30. Takada, Y., Matsumoto, K. and Torii, K.: A Programmer Performance Measure Based on Programmer State Transitions in Testing and Debugging Process. Proceedings of the 16th International Conference on Software Engineering, (1994) 123 – 132
- 31. Yang, J., Xu, Y. and Chen, C.S.: Human Action Learning via Hidden Markov Model. IEEE Transactions on Systems, Man, and Cybernetics 27 (1), (1997) 34-44
- Zukerman, I., Albrecht, D.W. and Nicholson, A.E.: Predicting Users' Requests on the WWW. Proceedings of the Seventh International Conference on User Modeling (1999)
- Zukerman, I., Nicholson, A.E. and Albrecht, D.W. 1999b: Evaluation methods for learning about users.
 In: A. Rudstrom (ed): Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence (IJCAI'99) Stockholm, Sweden, 31 July 6 August 1999, 83-90.