Final Program

HCI International

11th International Conference on Human-Computer Interaction

jointly with:

Symposium on Human Interface (Japan) 2005
6th International Conference on Engineering Psychology & Cognitive Ergonomics
3rd International Conference on Universal Access in Human-Computer Interaction
1st International Conference on Virtual Reality
1st International Conference on Usability and Internationalization
1st International Conference on Augmented Cognition
1st International Conference on Online Communities and Social Computing

July 22–27, 2005
Caesars Palace, Las Vegas, Nevada USA

Under the auspices of 9 distinguished international boards of 201 members from 31 countries

Evaluation Framework for Fuzzy Theoretic-Based Recommender System

Azene Zenebe and Anthony F. Norcio

Department of Information Systems
University of Maryland Baltimore County (UMBC)
Baltimore, MD 21250 USA
azenez1@umbc.edu; norcio@umbc.edu

Abstract

Empirical evaluation of user model based adaptive systems is challenging and intriguing due to the nature of user model and adaptive decision processes. Their success also depends on the subjective judgment of the user. It can be done by evaluating the constructed user model and the resulting adaptive decision made based upon the user model. Until recently, literature on comprehensive empirical evaluations of user modeling systems and adaptive systems are rare. The main reason is the lack of an evaluation methodology. Emerging literature attempts to address this gap in two fronts: by developing an evaluation methodology and by evaluating the user adaptive system. In this paper various evaluation methodologies are reviewed and adapted to fit the evaluation requirements of the fuzzy set theoretic user model driven recommender systems. Another related problem for evaluating the effectiveness of user models and adaptive recommender systems with regards to performance and usability is the absence of clear and effective evaluation metrics. In this paper evaluation metrics from the information retrieval community are adapted and discussed to evaluate effectiveness of the fuzzy theoretic-based adaptive recommender systems. Finally, in order to illustrate the application of the proposed evaluation framework, a simulation system for Fuzzy Theoretic-based Movie Recommender System is designed, developed and evaluated using statistical methods.

1. Introduction

Empirical evaluation of user model based adaptive systems is challenging due to the nature of user models and adaptive decision processes. Their success also depends on the subjective judgment of the users. It can be done by evaluating the constructed user models and the resulting adaptive decision made based upon the user models. Until recently, literature on comprehensive empirical evaluations of user modeling system and adaptive systems are rare (Chin, 2001). Emerging literature attempts to address this issue in two fronts: by evaluating user adaptive systems and by developing an evaluation methodology (Brusilovsky, Karagiannidis, & Sampson, 2001; Weibelzahl, 2001).

After observing the problem of treating the adaptation process as a 'monolithic' entity during evaluation, and focusing on the overall users' performance and satisfaction, Brusilovsky and et al.(Brusilovsky et al., 2001) propose a two-layered approach; and Weibelzahl (Weibelzahl, 2001) proposes a four layers process-based framework consists of evaluation of input data, evaluation of the inference mechanism, evaluation of the adaptation decision, and evaluation of the total interaction. Another similar work is by Paramythis and et al (Paramythis, Totter, & Stephanidis, 2001) that presents more detailed modular based evaluation frame work based upon the components of adaptation process to be evaluated along with appropriate methods and tools.

In this paper, layers based evaluation frameworks proposed by Weibelzahl (Weibelzahl, 2001) and Paramythis and et al (Paramythis et al., 2001) are reviewed and adapted to fit the evaluation requirements of the fuzzy set theoretic user model based recommender systems. Another related problem for evaluating the effectiveness of a user model and adaptive recommender systems with regards to performance and usability is the absence of clear and effective evaluation metrics. In this paper evaluation metrics from the information retrieval community are adapted and discussed to evaluate effectiveness of fuzzy theoretic-based adaptive recommender systems. Finally, in order to illustrate the application of the layered evaluation framework, a simulation system for Fuzzy Theoretic-Based Movie Recommender System (FTBMRS) is designed, implemented, and evaluated.

This paper is organized into five sections. Section 2 presents and discusses the architecture and components of adaptive (movie) recommender system. In section 3, a layered framework for evaluation of user modeling and adaptive recommender system is discussed. In section 4 evaluation aspects of the Fuzzy Theoretic-Based Movie Recommender System with respective to the four layers of the evaluation framework are presented along with results of the empirical evaluation of the movie recommender systems. Section 5 presents discussion of the statistical analysis results and conclusion of the paper.

1.2 User Model Based Recommender System

A user model is a collection of facts and assumptions about a user. It comprises of personal features such as demographic and other personal data such as job type; behavioral features such as interests, preferences and attitude; cognitive features such as goals, plans, beliefs, knowledge, ability and disability; and contextual features such as physical location, past interaction, hardware and software available, tasks and other users in the environment. User modeling is a knowledge engineering process for construction of user models, and it is comprises the acquisition, representation, reasoning, and maintenance (Chen & Norcio, 2001; Kobsa, 2001). At the acquisition stage, the sources of facets are the users, other applications or secondary resources like databases containing user model related data. Elicitation of facets from the user can be done either explicitly or implicitly.

User model representation refers to the process of organizing and representing the acquired user model, both individual and group user models. Reasoning in user modeling can have three stages. The primary reasoning is performed based on the user input during the acquisitions stage. Secondary reasoning is performed using the results of the acquisition stage (example domain knowledge, knowledge about other users) in order to augment further the user model to meet the higher demands of user adaptive application systems (Kobsa, Koenemann, & Pohl, 2001). Tertiary reasoning is required during user model maintenance.

The system architecture of a fuzzy theoretic user model based recommender system by taking movie recommendation as domain of application is presented in Figure 1. This architecture is used for developing the evaluation tasks and the metrics that are described in section 3. Its main components are:

- An user modeling subsystem that infers user's interests and preferences for a movie, movies of a given genre, etc. using the user and movie data
- An adaptive Decision Making Module or Adaptive Recommender subsystem (ARS) – makes recommendation decisions using the user model and inference mechanism. It performs the following :
    - Prepares a query using the user question, user browsing/interaction behavior and user model
    - Receives and filters retrieved movies using the user model and inference strategies
    - Formats and presents the selected movies to the user
- An user Interface Subsystem – is a gateway to the system for users
- Data Storage and Retrieval Subsystem – responsible for storage, retrieval and maintenance of user and movie data.

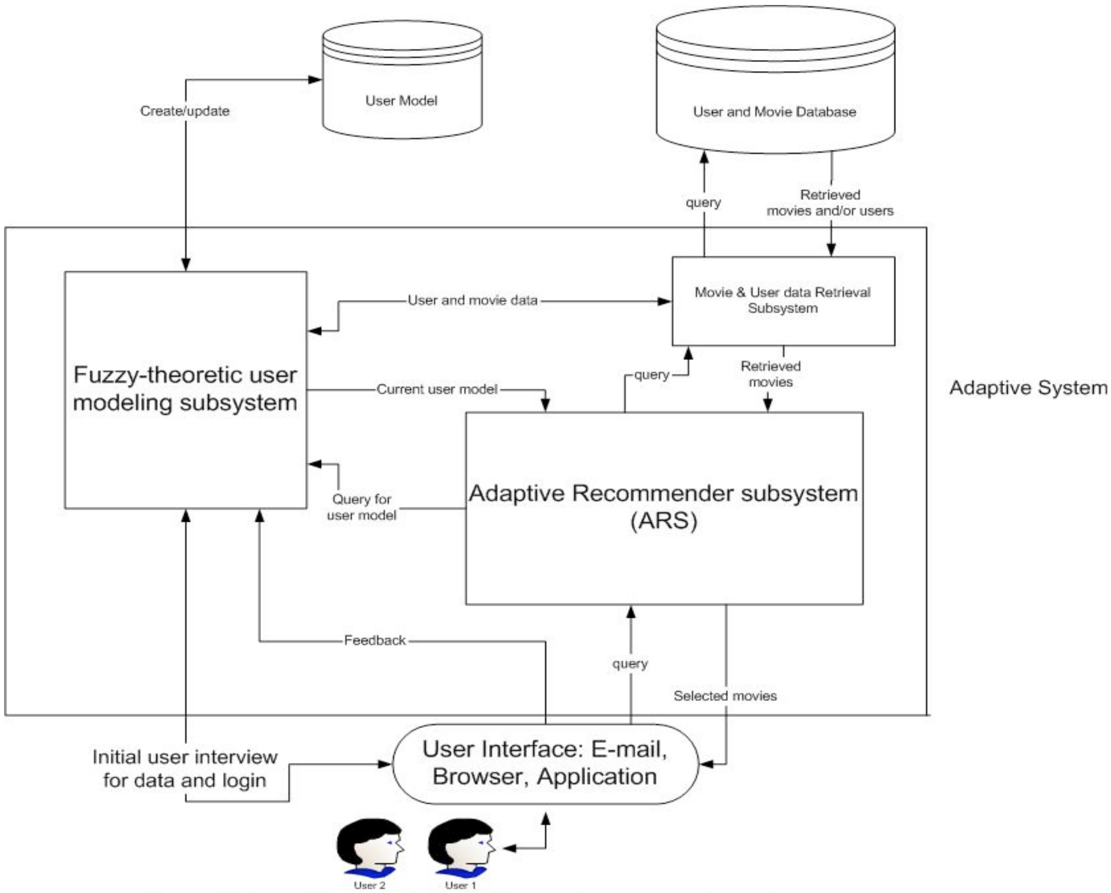

Figure 1: The architecture of an adaptive movie recommender system

2. Review of Literature

The literature in (Weibelzahl, 2001, 2003; Stephan Weibelzahl & G. Weber, 2003; Stephan Weibelzahl & Gerhard Weber, 2003) presents a summary of evaluation works and methods used in adaptive systems literature through 2001. In this paper a representative literature on and after 2000 are reviewed and presented in Table 2. Table 2 presents a summary of the evaluation studies in relation to the layered evaluation framework.

Table 2: A summary of recent literature on the evaluation of adaptive system

Citation	Layers Covered	Evaluation Metrics	Application domain	Algorithms/ Approaches used
(Billsus & Pazzani, 2000)	1, 2, 3, 4	IR measures	Online Daily News	Dynamic; t-test; Long-term (CB using naïve Bayesian), short-term(CB using K-NN) , hybrid
(Piyawat, 2002)	1,2	In house defined formulas	News Groups	Fuzzy set, OO
(Li & Kim, 2003)	1,2	Mean Absolute Error (MAE)	Movie Recommendation	Clustering; fuzzy set; combining CB and CF
(Rojsattarat & Nuanwan, 2003)	1,2	MAE and ROC(Receiver Operating Characteristic)	Movie Recommendation	Hybrid approach (CB & CF); naïve Bayesian; Support Vector Machines
(Deshpande & Karypis, 2004)	1,2,3	IR metrics	Movie Recommendation	Item-Based Top-N recommendation

3. Evaluation Methodology

3.1. Layered Evaluation Framework

The evaluation frameworks proposed by Weibelzahl (Weibelzahl, 2001) and Paramythis and et al (Paramythis et al., 2001) are the foundation of this evaluation methodology. These evaluation frameworks are summarized in Table 1.

The four layers in Table 1 are studied and applied for the evaluation of a fuzzy theoretic movie recommender system described in Figure 1. This layered evaluation framework can also be tailored and be applied for evaluating any user modeling and user adaptive system. The advantages of this framework are: (i) it provides insight into the success or failure of each stage of the adaptation; (ii) it facilitates the tackling of the problems; (iii) it enhances the reusability of successful results; and (iii) it enhances the generalization of the evaluation results.

At layer 1, statistical methods such as frequency distribution, cross-tabulation with chi-square and correlation coefficient can be used to study the attributes, and dataset used for building the user model. At layer 2 and layer 3, the effectiveness of the components and processes of the adaptive system are respectively evaluated in terms of effectiveness of the constructed user model using accuracy and reliability measures (Layer 2); and effectiveness of the resulting adaptation or recommendations are measured in terms of relevance of the recommendations (Layer 3).

Table 1: User Modeling and Adaptive System Evaluation Framework

Evaluation Layer	Evaluation Criteria	Methods and Tools	Components /Processes/purpose
Layer 1: Input data including primary reasoning	Objectivity, reliability and validity of data	Statistical methods such frequency distribution, cross-tabs with chi-square, correlation, etc.	Input data and domain analysis for constructing user models
Layer 2: Inference for modeling of the users (Secondary /Tertiary Reasoning)	Higher level: Validity of the inferences; Lower Level: Comprehensiveness, Redundancy, Accuracy, Sensitivity of the model (Verification of user model)	Exploratory studies using experimental data; use of simulations with hypothetical users; or experimental study to compare the inferred user model with the actual.	Construction of user models
Layer 3: Adaptation Decision	Necessity of adaptation; Appropriateness of adaptation; Acceptance of adaptation (adaptive system performance for users)	Compare the variants of adaptation including non-adaptation decisions in terms of different evaluation criteria, such as absolute mean error, recall, precision, etc.	Utility of user model for the adaptive system as well as adaptive strategies
Layer 4: Interaction	Usability; Timeliness; Obtrusiveness; User control of adaptation; Metrics for measuring the effects of adaptation such as increase profit, user satisfaction	Compare the different ways to adapt to users on a physical level, e.g., a recommendation vary in color, font size, shape and other properties. Also use objective metrics like task success and time	Usefulness of the system in short and long term in meeting the users needs and goals

Layer 3's main purpose is to determine whether the adaptation decisions made are the optimal ones given that the user's features are inferred correctly, and consists of two parts: determination of the necessity of adaptations and the required types of adaptations. For the same user models, it is possible to have no or several possibilities of adaptations. Given the specific interaction context, the necessity criteria considers whether adaptation is indeed required; where as the appropriateness criteria consider whether the established need for adaptation can satisfy the requirements posed by the current interaction context (e.g. to interrupt the user to provide a recommendation or change the interface while the user is working); and the acceptance criteria considers the user's perception of the established need for adaptation is both required and appropriate (Stephan Weibelzahl & Gerhard Weber, 2003). At layer 4, the focus is on the usability of the system for users in accomplishing their goals through the user interface. The system's importance, relevance, timeliness, etc., can be considered and measured in terms of users' task performance (success and time), increase profit and user satisfaction (usefulness of the system in long-term).

Another related challenging issue for evaluating the effectiveness of a user model and adaptive recommender systems is the absence of a clear and effective evaluation metrics, which is addressed in the next section.

3.2. Evaluation Metrics

The two broad evaluation metrics are metrics for the effectiveness of a user model and metrics for the effectiveness of adaptive decisions. The effectiveness of a user modeling system can be evaluated with respect to level of accuracy of the resulting user models, for instance in describing users interest in movies. When machine learning techniques are employed in the construction of user models, the standard machine learning metrics such as predictive accuracy and error measures that assess how training and testing data set fits the model can be used as measures of effectiveness.

From users' perspective, the system performance is judged by the relevance of the recommendations made by the adaptive system that uses user models. However, high user model predicative accuracy or low error measure does not mean high system performance in terms of its ability to recommend relevant and useful items that meet users needs (Billsus & Pazzani, 2000). Particularly, predictive accuracy metrics based on errors such as mean absolute error (MAE), mean square error (MSE) and root mean square error (RMSE) found to be less appropriate when the user task is to find 'good' items like the task in recommender systems and when the granularity of true value is small because predicting a 4 as 5 or a 3 as 2 makes no difference to the user (Herlocker, Konstan, Terveen, & Riedl, 2004) .

Recently, many studies (Table 2) started using common information retrieval (IR) performance metrics instead of MAE or MSE in evaluation of adaptive recommender systems. Retrieval effectiveness (Recommendation accuracy) relates to the appropriateness of the items that are retrieved to satisfy the information need of the user in response to a query. Based up on the level of judgment, there are two categories of measures: binary and n-ary. In binary measure items are accepted (relevant) or rejected (non-relevant). In n-ary measure different levels or degrees of relevance are assigned, with n=3, 5 or 7 being the most commons [Korfhage, 1997]. The definitions for the 5 levels are: 5=definitely relevant, 4=probably relevant, 3=barely relevant, neutral, or not judged, 2=probably not relevant and 1=definitely not relevant. Hence, the n-array (for n=5) effectiveness measures developed in information retrieval/filtering field including precision, recall and F-measure are used for evaluation of movie recommender system presented in section 4.

There is also an emerging demand for evaluating the usefulness and utility of recommender systems. Usefulness metrics include coverage, confidence level, model learning rate; and utility metric includes user satisfaction through user feedback about the recommender system (Herlocker et al., 2004). Some of these metrics require users to use the system for a period of time and hence require the deployment of the system.

3.3 Evaluation of the Fuzzy Theoretic Based Movie Recommender System

A summary of the evaluation aspects for the movie recommender system with respective to the four layered framework is presented in Table 3. In order to illustrate the application of the evaluation framework, a simulation system of the movie recommender system described in Figure 1 is implemented. Due to space limitation and focus of the paper, only the experimental design for analysis of results of simulation and brief overall results of the evaluations at layer 2 and layer 3 are presented. Since the system is not deployed for actual use, evaluation at layer 4 is not undertaken.

Table 3: A summary of the evaluation tasks for the movie recommender system in the perspective of the four layered framework

Evaluation Layer	Evaluation Criteria, Methods and Tools	Tasks /Evaluation metrics
Layer 1: Input data including primary reasoning	Statistical methods such frequency distribution, cross-tabs and correlation are used to explore the application domain, define fuzzy membership functions, and study relationships	User and movie data are obtained and preprocessed from MovieLens (http://www.grouplens.org/) and IMDb (http://www.imdb.com/) to extract user and movie features
Layer 2: Inference Modeling of users (Secondary /Tertiary Reasoning)	Exploratory studies using experimental data for inferring genres preference from user ratings; and experimental study to compare the inferred user model with the actual users by computing correlation between actual rating and recommendation confidence score	Using user previous ratings on movies, movies' features, similarity measures and recommendation strategies, user interest to movies are inferred with in fuzzy set and possibility theory
Layer 3: Adaptation Decision	Compare the variants similarity measures and recommendation strategies for the task "find top-N good movies" for a user	Top-N recall, precision, learning rate, confidence level coefficient are defined and performed
Layer 4: Interaction	Future Research	Future research

3.4 Experimental Design

Representation of movies' features using fuzzy set and possibility framework opens various research opportunities including studying users' genres preference inference. At layer 2, the result of analysis of variance using user and movie data for testing the hypothesis: There is genre preference by users and their genre preference can be inferred from their past movie watching behavior (expressed by ratings). Moreover, significant bi-variants correlations (Spearman's rank) are found between the variable actual user ratings for a user in the dataset and recommendation confidence scores for a user. These support the validity of the inference process using genres and user ratings.

At layer 3, the different effectiveness metrics be the dependent variables (particularly precision, recall and F-measure) and for the independent variables summarized in Table 4, a 2 x 5 x 6 x 12 factorial design based statistical analysis is conducted. The factors are recommendation approaches, similarity computation approaches, training size (no of training cases), and values of N in Top-N. The testing size is used as the covariate factor in general linear model (GLM). Moreover (i) using different random selection of the movies into testing and training sets 10 different runs are executed to avoid sensitivity to sampling bias; (ii) 100 users out of 943 are randomly selected; and (iii) normalization is applied on the recommendation scores.

The dependent variables are defined as:

- Top- N precision are computed

  Precision45 = (no of movies in the TOP-N recommendation with rating ≥ 4) / N
- Top-N recall – the proportion of the relevant movies rated as interesting by the user that is actually retrieved. It is similar to recall but based on recommendation evaluation restricted to rated movies. Let K be total number of movies rated as interesting form the test cases.

  Recall45 = (no of movies in the TOP-N with rating ≥ 4) / K
- F-measures - is a single metric that combines precision and recall.

  F1-ratio = (2*precision*recall)/(precision + recall)

Table 4: Independent Factors

Independent Factor	Description	Possible values
Recommendation Approach	Fuzzy reasoning as recommendation strategies	Weighted-Sum or Maximum-Min
Similarity Measure	As foundation for analogical reasoning using fuzzy set and possibility theory	Set-theoretic (Binary Jaccard Index), Fuzzy Set-theoretic (Modified Jaccard Index), Fuzzy Cosine, Fuzzy Correlation-like, Fuzzy Proximity-based
No of training cases (K most liked movies by a user)	Movies rated positively by users and are selected randomly	5,10,15,20,25 and 30
Top-N	Value of N in TOP-N	3,4, 5, 10,15, 20, 25, 30, 35, 40, 45, 50
No of testing cases (size of alternative items considered for recommendation)	Movies assumed not seen or experienced by users and are selected randomly	Total number of rated movies minus total movies considered for training by a user

4. Results and Discussion

Using the same MovieLens dataset and Item-Based Top-N recommendation algorithms within the crisp set framework, Deshpande and Karypis (Deshpande & Karypis, 2004) reported the top-10 Hit-Rate (equals to Average Top-10 recall) to be 0.131, 0.271 and 0.271 for Frequent (naïve), Cosine and Conditional probability based recommendation algorithms respectively. Similarly, 0.281 is the Top-10 Hit rate reported using the user-based algorithm. Compared to these results, the overall average Top-10 recall of Fuzzy theoretic Based Recommender Systems using maximum-minimum recommendation approach is greater (32%).

In order to assess the effects of the dependent and independent factors, Univariate analysis using the general linear model (GLM) is performed. The summary of the results are presented in Table 5, and at 5% level of significance,

- There is no significance difference in average precision between the two recommendation approaches.
- There is significance difference in average recall and average F1-ratio between the two recommendation approaches.
- There is significance difference in average precision, recall and F1-ratio by types of similarity measures.
- There is significance recommendation approach and similarity measure interaction effect on average precision, recall and F1-ratio. Table 5 presents the results of ANOVA for dependent variable precision.

Using Tamhane's T2 pair-wise means comparisons test based on a t test yield (SPSS Inc., 2003) the results in Table 6 and Table 7. With weighted-sum recommendation approach, all the fuzzy-theoretic similarity measures driven movie recommender system average precisions and F1-ratio are significantly greater than the corresponding crisp set-theoretic similarity measure driven movie recommender system. Moreover, there is no statistically significance difference between the two groups in their average top-N recall except the proximity-based similarity measure. With maximum-minimum recommendation approach, all the fuzzy-theoretic similarity measures driven movie recommender system average precisions are significantly greater than the corresponding crisp set-theoretic similarity measure driven movie recommender system. For average recall, , all the fuzzy-theoretic similarity measures driven movie recommender system average recalls are significantly less than the corresponding crisp set-theoretic similarity measure driven movie recommender system. Finally, except for the cosine similarity measure where there is no significance difference, all the fuzzy-theoretic similarity measures driven movie recommender system average recalls are significantly less than the corresponding crisp set-theoretic similarity measure driven movie recommender system.

Table 5: ANOVA for Precision, Recall and F1-Measure

	Precision		Recall		F1-Measure	
Source	F	Sig.	F	Sig.	F	Sig.
Corrected Model	41.806	.000	659.931	.000	609.465	.000
Intercept	676548.034	.000	1239921.109	.000	1860331.127	.000
testingsize	15341.207	.000	233028.663	.000	188502.121	.000
similarityid	842.576	.000	274.353	.000	88.781	.000
approachid	1.643	.200	7241.658	.000	7646.255	.000
topn	27.399	.000	22242.780	.000	20316.913	.000
trainingsize	1438.475	.000	167.477	.000	920.618	.000
similarityid * approachid	4.698	.001	129.732	.000	113.428	.000
similarityid * topn	4.915	.000	5.740	.000	26.218	.000
approachid * topn	2.681	.002	23.812	.000	109.383	.000
similarityid * approachid * topn	.098	1.000	1.699	.003	4.851	.000
similarityid * trainingsize	26.550	.000	79.646	.000	33.575	.000
approachid * trainingsize	.317	.903	10.045	.000	20.817	.000
similarityid * approachid * trainingsize	.609	.910	62.129	.000	38.187	.000
topn * trainingsize	10.788	.000	31.969	.000	49.924	.000
similarityid * topn * trainingsize	.395	1.000	2.563	.000	3.378	.000
approachid * topn * trainingsize	.146	1.000	1.997	.000	2.739	.000
similarityid * approachid * topn * trainingsize	.138	1.000	.568	1.000	.654	1.000

Table 6: Multiple Means Comparisons-with Weighted-Sum Recommendation Approach

Dependent Variable	(I) similarityid	(J) similarityid	Mean Difference (I-J)	Sig.
precision45	Crisp Set	Cosine	-.062655283(*)	.000
		Fuzzy Set	-.064364716(*)	.000
		Proximity	-.051860833(*)	.000
		Correlation	-.061702805(*)	.000
recall45	Crisp Set	Cosine	.004115424	.259
		Fuzzy Set	-.000328398	1.000
		Proximity	.021821275(*)	.000
		Correlation	.002638029	.822
F1-ratio	Crisp Set	Cosine	-.004470893(*)	.011
		Fuzzy Set	-.009662995(*)	.000
		Proximity	.010642737(*)	.000
		Correlation	-.010114670(*)	.000

Table 7: Multiple Means Comparisons-with Maximum-Minimum Recommendation Approach

Dependent Variable	(I) similarityid	(J) similarityid	Mean Difference (I-J)	Sig.
precision45	Crisp Set	Cosine	-.066318758(*)	.000
		Fuzzy Set	-.064321743(*)	.000
		Proximity	-.059817397(*)	.000
		Correlation	-.067318877(*)	.000
recall45	Crisp Set	Cosine	.012951678(*)	.000
		Fuzzy Set	.041586849(*)	.000
		Proximity	.035976127(*)	.000
		Correlation	.021066065(*)	.000
F1-ratio	Crisp Set	Cosine	.002493380	.466
		Fuzzy Set	.012160333(*)	.000
		Proximity	.012274586(*)	.000
		Correlation	.008511112(*)	.000

5. Conclusions

This paper discusses evaluation framework for adaptive systems that highly relies on user models. The four layers process-based framework for evaluation that is getting acceptance in the adaptive systems field are adapted and applied to a Fuzzy Set Theoretic Recommender System. The four layers in the framework are evaluation of input data, evaluation of the inference mechanism, evaluation of the adaptation decision, and evaluation of the total interaction. This framework provides insight into the success or failure of each stage of the adaptive process, enhances the reusability of results, and enhances the generalization of the evaluation results. Application of the evaluation framework, layer 1 to layer 3, is demonstrated using the Fuzzy Set Theoretic Based Movie Recommender System.

Evaluation at Layer 4 will be addressed in future research. Usefulness metrics including coverage, confidence level and model learning rate, and utility metric including user satisfaction will also be addressed in future research.

References

Billsus, D., & Pazzani, M. J. (2000). User Modeling for Adaptive News Access. *User modeling and user-adapted interaction, 10*, 147-180.

Brusilovsky, P., Karagiannidis, C., & Sampson, D. (2001, July 13, 2001). *The Benefits of Layered Evaluationof Adaptive Applications and Services.* Paper presented at the Workshop held at the Eighth International Conference on User Modeling, Sonthofen, Germany.

Chen, Q., & Norcio, A. F. (2001). Knowledge Engineering in Adaptive Interface and User Modeling. In Q. Chen (Ed.), *Human Computer Interaction: Issues and Challenges* (pp. 113-133). PA: Idea Group Publishing.

Chin, D. N. (2001). Empirical evaluation of user models and user adapted systems. *User modeling and user-adapted interaction, 11*(1-2), 181-194.

Deshpande, M., & Karypis, G. (2004). Item-Based Top-N Recommendation algorithms. *ACM Transactions on Information Systems, 22*(1), 143-177.

Herlocker, J. L., Konstan, J. A., Terveen, L. G., & Riedl, J. T. (2004). Evaluating Collaborative Filtering Recommender Systems. *ACM Transactions on Information Systems, 22*(1), 5-53.

Kobsa, A. (2001). Generic User Modeling Systems. *User modeling and user-adapted interaction, 11*, 49-63.

Kobsa, A., Koenemann, J., & Pohl, W. (2001). *Personalized Hypermedia Presentation Techniques for Improving Online Customer Relationships*. Retrieved, 2002, from the World Wide Web: http://citeseer.nj.nec.com/kobsa01personalized.html

Li, Q., & Kim, B. M. (2003, July 7, 2003). *An Approach for Combining Content-based and Collaborative Filters*. Paper presented at the Sixth International Workshop on Information Retrieval with Asian Languages (IRAL-2003), Sapporo Convention Center, Sapporo, JAPAN.

Paramythis, A., Totter, A., & Stephanidis, C. (2001). *A modular approach to the evaluation of adaptive user interfaces*. Paper presented at the Proceedings of workshop at the Eighth International Conference on User Modeling, UM2001, Freiburg, Germany.

Piyawat, T. (2002). *Fuzzy Object Oriented User Modeling: Theory, Representation, and Effectiveness.* Unpublished Dissertation, University of Maryland Baltimore County, Baltimore, MD.

Rojsattarat, E., & Nuanwan. (2003, March 2003). *Hybird Recommendation: Combining Content-Based Predicition and Collaborative Filtering.* Paper presented at the 4th International Conference, Intelligent Data Engineering and Automated Learning (IDEAL 2003), Hong Kong, China.

SPSS Inc. (2003). *SPSS 12.0 for Windows* [SPSS Help files]. Retrieved, 2005, from the World Wide Web:

Weibelzahl, S. (2001). *Evaluation of adaptive systems.* Paper presented at the User Modeling 2001: Proceedings of the Eighth International Conference, UM2001.

Weibelzahl, S. (2003). *Evaluation of adaptive systems.* Unpublished Dissertation, University of Trier, Germany.

Weibelzahl, S., & Weber, G. (2003). *Evaluating the Inference Mechanism of Adaptive Learning Systems.* Paper presented at the User Modeling: Proceedings of the Ninth International Conference.

Weibelzahl, S., & Weber, G. (2003). *Evaluating the Inference Mechanism of Adaptive Learning Systems.* Paper presented at the User Modeling 2003: Proceedings of the Ninth International, UM2003.