Effects of Fuzzy Theoretic Inference Strategy and Similarity Measure on a Recommender System

Dr. Azene Zenebe[1]; Dr. Anil Khatri[2]; and Dr. David Anyiwo[3]
Department of Management Information Systems, Bowie State University, Bowie, MD 20715, USA
[1]azenebe@bowiestate.edu;[2]akhatri@bowiestate.edu;[3]danyiwo@bowiestate.edu

Abstract: There is a need for modeling and reasoning on the subjective, incomplete, imprecise and vague nature of item features, user behavior and their relationships in machine learning algorithms. This paper presents results of analysis of effects of fuzzy theoretic inference/recommendation strategies and fuzzy theoretic similarity measures on a recommender system's accuracy- measured in terms of precision, recall and F1-measure. The study uses the MovieLens benchmark dataset on movies that are widely used in recommender system research, and conduct simulation runs. Finally, a guideline for recommender system designer to choose a combination of inference strategy and similarity measure for personalized recommender systems is presented based on statistical analysis of the simulation results.

Key words: Knowledge Representation, Inference Method, Fuzzy Logic, Similarity Measure, Recommender System

I. Introduction

The recommendation methods refer to the approaches and algorithms used in identifying and generating assumptions about users for making recommendation. One of the most important issues in recommender system algorithms is computing similarity between users, and between objects (items, events, etc.). This in turns highly depends on the appropriateness and accuracy of the methods used for users and objects representation in terms of their corresponding attributes, as well as the inference strategy used.

There are various empirical studies on the effects of recommendation strategies and similarity measures for recommender systems. A few examples are [1, 2]. These studies use crisp set representation of and classical logic on reasoning user behavior and item features. They do not take into account the subjective, incomplete, imprecise and vague nature of item features, user behavior and their relationships.

In previous study [3], we have showed the need for modeling and reasoning on the subjective, incomplete, imprecise and vague nature of item features, user behavior and their relationships using fuzzy set and logic. After developing a simulated fuzzy theoretic recommender system for users of movies, we have obtained better recommendation performance when compared to results of state-of-the-art recommender systems researches that use crisp set and traditional logic. The fuzzy theoretic recommender system is comprising of representation method for items' feature and user feedbacks using fuzzy set, and an algorithm for analogical reasoning based on four fuzzy theoretic similarity measures (fuzzy theoretic extension of the Jaccard Index, Cosine, Proximity and Correlation) and two inference strategies (maximum-minimum and weighted-sum recommendation strategies). Moreover, significance interaction effects on performance by similarity measure and recommendation strategy are found [3].

This paper presents results of further analysis of effects of recommendation approach and similarity measures on a recommender system's performance - precision, recall and F1-measure. Moreover, it attempts to provide a guideline for recommender system designer to choose a combination of recommendation approach and similarity measure for personalized recommender systems.

II. Methodology

The representation scheme, inference strategies, similarity measures and the simulation recommender systems are described in this section.

A. Item Representation using Fuzzy Set

For an item described with multiple attributes, more than one attribute can be used for recommendation. Moreover, some attributes can be multi-valued involving overlapping or non-mutually exclusive possible values. For example, movies are multi-genres and multi-actors [4]. These values of multi-valued attributes in an item can be represented more accurately with fuzzy set framework than with crisp set framework.

Let an item I_j ($j = 1 \ldots M$) is defined in the space of an attribute X $=\{x_1, x_2, x_3, \ldots x_L\}$, then I_j can take multiple values such as x_1, x_2, …, and x_L. These values of X can be sorted in the decreasing order of their presence in the item I_j expressed by degrees of membership. The membership function of item I_j to value x_k ($k = 1 \ldots L$) is denoted by $\mu_{x_k}(I_j)$, which can be obtained either heuristically from domain analysis or empirically from the data. Hence, a vector $X_j=\{(x_k, \mu_{x_k}(I_j)), k= 1\ldots L\}$ is formed for I_j. $\mu_{x_k}(I_j)$ can be interpreted as the degree of similarity of I_j to a hypothetical (or prototype) pure x_k type of the item; or as the degree of presence of value x_k in item I_j.

We use movie as the domain and movie genre as the attribute for formalizing and applying the heuristic. Given the definition of a movie in the space of genre (G) a movie can have one major genre denoted by g_1 and one or more minor genres g_2, g_3, and so on, in the decreasing order of their degrees of presence in a movie. The degree of membership of movie m_j ($j = 1 \ldots M$) to genre g_k ($k = 1 \ldots N$) is denoted by $\mu_{g_k}(m_j)$, which can be obtained either heuristically from domain analysis or empirically from the data. Hence, for m_j, we can form a vector $G_j=\{(g_k, \mu_{g_k}(m_j)), k= 1\ldots N\}$. Since membership function in fuzzy set theory is deliberately designed to treat the vagueness and imprecision in the context of the application [5], we take the following steps:

Step 1: Sort g_k in descending order of $\mu_{g_k}(m_j)$. In IMDB (www.imdb.com[1]) the genres of movie m_j are presented in the order of significance. For example, movie 'King Kong (2005)' has Action as a major genre, and Adventure as the 1st minor, Drama the 2nd minor, Fantasy as the 3rd minor, and Thriller as 4th minor genres.

Step 2: Assign higher degrees of membership value to more important genres of a movie. For instance,

If m_j has only one genre, then $\mu_{g_1}(m_j) = 1$ and $\mu_{g_k}(m_j)=0$ for all k=2 to N.

If m_j has two genres, then $\mu_{g_1}(m_j) = 0.8$, $\mu_{g_2}(m_j) = 0.2$ and $\mu_{g_k}(m_j)=0$ for all k=3 to N.

If m_j has three genres, then $\mu_{g_1}(m_j) = 0.70$ and $\mu_{g_2}(m_j) = 0.30$, $\mu_{g_3}(m_j)=0.10$ and $\mu_{g_k}(m_j)=0$ for all k=4 to N.

[1] "IMDb History," vol. 2004: Internet Movie Database Inc., n.d.http://www.imdb.com/

Based on the heuristics illustrated for movie, the possibility for item I_j to take different values of X varies, and the membership function should meet the following three criteria: 1) assigning higher degree of membership to major values than minor values; 2) assigning 0 to values that are not associated with the item; 3) degrees of membership should be normalized to the range of [0,1]; 4) the same value of X at same rank positions between different items should have varying degrees of membership values if the numbers of values of X associated with the items are different. We propose to represent the generalization of this type of heuristic rules with a Gaussian-like fuzzy set membership function, as shown in (1).

$$\mu_{x_k}(I_j) = r_k / 2^{\sqrt{\alpha * |L_j| (r_k - 1)}} \qquad (1)$$

where $N=|L_j|$ is the number of values of X associated with I_j and r_k ($1 \le r_k \le |L_j|$) is the rank position of value x_k, and $\alpha > 1$ is a constant used as a threshold to control the difference between consecutive values of X in I_j.

For example, with α set to 1.2, movie 'King Kong (2005)' is represented in terms of genres as follows:

$|L_j| = 5$ and x_j = {(Action, 1), (Adventure, 0.366), (Drama, 0.272), (Fantasy, 0.211), (Thriller, 0.168)}.

B. User Feedback Representation using Fuzzy Set

User rating is the most widely used user feedback in recommender systems. It is a proxy variable used for measuring user degree of interest in an item. In the popular collaborative algorithms (e.g. [6],[7]) user ratings are represented and interpreted in binary value- those liked or disliked. In 5-scale ratings, above 3 are considered as liked. However, intrinsically user rating is imprecise as user demonstrates to give different ratings to same item at different time and situation due to the difficulty to make a distinction between rating 4 and 5, and 1 and 2 by the users, as well as due to mood and taste change. Moreover, the same rating say 4 in scale of 5 given by two users do not necessarily imply equal degree of interest to an item. For pessimist user, rating of 4 may mean strongly liked but for optimist rater it may mean somewhat liked. Is the difference between ratings 3 and 4 equals to difference between 4 and 5? These all contribute to fuzziness that arises from human thinking processes instead of randomness associated with user rating.

Therefore, user interest based on user rating is treated as fuzzy variable and its uncertainty is represented using possibility distribution function (π). π is defined to be numerically equal to membership function [5]. For the fuzzy variable degree of interest in an item (DI) associated with user rating (R) expressed in continuum from Minimum value (Min) to Maximum value (Max), e.g., 1 to 5. Then the proposition 'User U has strongly liked an item I' has the possibility distribution function $\pi_R(I)=\mu_{SL}(R=r)$, for r between Min and Max. Where DI comprising of Strongly Liked (SL), Liked(L), Indifferent(I), Disliked (D), and Strongly Disliked (SD) fuzzy values.

Under this interpretation or semantics of the fuzzy variable DI, user rating is represented and reasoned using possibility distribution function by treating the rating as fuzzy number. For instance a rating 4 in 5 scale which refers to strongly liked is represented in terms of its possibility distribution function values={$\mu_{SL}(R=4)/4$, $\mu_L(R=4)/4,\ldots,\mu_{SD}(R=4)/4$}. Different membership functions can be used to represent DI such as triangular and bell-shaped membership function. Without losing generalization, a half triangular fuzzy number, which is the simplest models of uncertain quantity, is used to represent the degree of positive experience a user has in relation to an item. Equation 2 defines the half triangular fuzzy number membership function.

$$\mu_A(I_i) = (r - Min)/(Max - Min) \qquad (2)$$

for user rating r on I_i є [Min to Max] and A is a fuzzy set on DI

As result, a set of items liked by a user denoted by E is defined as: {$I_i : \mu_A(I_i) > 0.5$, i.e., $I_i : r > (Min+Max)/2$}. For r є [1 to 5], (2) reduces to (r - 1)/4.0; and a set of items liked by a user denoted by E is defined as: {$I_i : \mu_A(I_i) > 0.5$, i.e., $I_i : r>3$}. Due to current practice and available datasets, the range of rating is restricted to 1 to 5. Extending the possible range of rating, e.g., from -100 to 100, and representing within fuzzy set framework expect to produce more accurate measurement of user interest in an item, and in turn improves effectiveness of recommender systems.

C. Inference Strategies and Similarity Measures

Based on the representation scheme defined for items and user feedback to these items, the recommendation strategies and similarity measures are defined.

1) Inference Strategies

There are different alternatives for inference and making recommendation within fuzzy and possibility theory framework. The two alternatives that regard both user feedbacks such as previous user ratings and similarity between previously liked items and new items are considered.

a) Weighted-Sum inference method

For each inexperienced or new item I_j, calculate a confidence score using the weighted sum as:

$$R_1(I_j) = \sum_{I_k \in E} \mu_E(I_k) S(I_k, I_j) \qquad (3)$$

E is a set of positively experienced (liked) items by users, and $\mu_E(I_k)$ is the membership of the item I_k to fuzzy sets E. Furthermore, $S(I_k, I_j)$ is the similarity between I_j and I_k computed using the techniques defined next. A normalized evaluation score for each I_j's recommendation confidence is obtained using $NR_1(I_j) = R_1(I_j) / \underbrace{Max}_{k}[R_1(I_k)]$

b) Maximum-Minimum inference method

For each inexperienced or new item I_j, calculate a confidence score as:

$$R_2(I_j) = Max\left[\underbrace{Min}_{I_k \in E}\left(S(I_j, I_k), \mu_E(I_k)\right)\right] \qquad (4)$$

A normalized evaluation score is obtained by dividing each I_j's recommendation confidence by the maximum of recommendation confidence scores. Finally, the recommendation confidence scores are degree of support to recommend I_j, and can be interpreted as how much the recommender system assumes this user will like the item from the set of items un-experienced by the user, which portrays the strength of the recommendation.

2) Similarity Measures

In fuzzy set and possibility framework, similarity of users or items is computed based on the membership functions of the fuzzy sets associated to the users or items features. The set-theoretic, proximity-based and logic-based are the three classes of measures of similarity [8]. Based on the results of the study by Cross and Sudkamp [8] those measures that relevant for items recommendation application are considered.

I_j and I_k are defined as {(x_i, $\mu_{x_i}(I_j)$), i= 1… N} and {(x_i, $\mu_{x_i}(I_k)$), i= 1… N}; and represent the possibility distribution of the item I_j and I_k in the space of X with cardinality N. The similarity measure between I_j and I_k, denoted by $S(I_k, I_j)$, are defined as Fuzzy Set Theoretic Similarity in (5), Fuzzy Theoretic Cosine Similarity in (6), Fuzzy Theoretic Proximity (Minkowski's distance based) similarity in (7) and Fuzzy Theoretic Correlation-like similarity in (8).

$$S_1(I_k, I_j) = \sum_i \left|\mu_{x_i}(I_k) \cap \mu_{x_i}(I_j)\right| \Big/ \sum_i \left|(\mu_{x_i}(I_k) \cup \mu_{x_i}(I_j)\right| \qquad (5)$$

$$S_2(I_k, I_j) = \frac{\sum_i \mu_{x_i}(I_k) * \mu_{x_i}(I_j)}{\sqrt{\left(\left(\sum_i (\mu_{x_i}(I_k))^2\right)\right)}\sqrt{\left(\left(\sum_k (\mu_{x_i}(I_j)^2\right)\right)}} \qquad (6)$$

$$S_3(\mathbf{I_k},\mathbf{I_j}) = 1 - \left[\mathbf{d}_2(\mathbf{I_k},\mathbf{I_j}) \Big/ \underbrace{\mathbf{Max}}_{\mathbf{i}} \left\{ \mu_{\mathbf{x_i}}(\mathbf{I_k}), \mu_{\mathbf{x_i}}(\mathbf{I_j}) \right\} \right] \quad (7)$$

Where

$$\mathbf{d}_2(\mathbf{I_k},\mathbf{I_j}) = \left(\sum_{\mathbf{i}=1}^{\mathbf{L}} \left| \mu_{\mathbf{x_i}}(\mathbf{I_k}) - \mu_{\mathbf{x_i}}(\mathbf{I_j}) \right|^2 \right)^{\frac{1}{2}}$$

$$S_4(I_k, I_j) = 1 - \frac{2}{Z_{I_k} + Z_{I_j}} \left(d_2(I_k, I_j) \right)^2 \quad (8)$$

Where

$$Z_{I_k} = \sum \left(\left(2 * \mu_{x_i}(I_k) \right) - 1 \right)^2$$

$$Z_{I_j} = \sum \left(\left(2 * \mu_{x_i}(I_j) \right) - 1 \right)^2$$

D. Simulation System

Simulation recommender system for movies is developed. It comprising of representation method for items' feature and user feedbacks, and an algorithm for analogical reasoning based on the four fuzzy theoretic similarity measures and two recommendation strategies defined above. The simulation system works as follows:

a. For each user, it randomly splits the ratings dataset into a training set and a test set.
b. Using the training set, it computes similarity indices and recommendation confidence scores for each item in the test set using the different similarity and inference strategy.
c. For each user, using the movies in the test set, it computes Top-N recommendations and the recommendation accuracy metrics – precision, recall, and F1–measure.
d. 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, and the average results are reported; (ii) 100 out of 943 users are selected randomly; (iii) normalization is applied on the recommendation scores; and (iv) all statistical testes are done at 5% level, and in all referred literature and in this paper, movies with ratings 4 and 5 are considered as movies liked by users.

III. EVALUATION

A. Experimental Design

Performance measures precision, recall, and F1-measure are the dependent variables. On each of this dependent variable, a 2 x 4 x 6 x 12 factorial design is conducted. In this design the fixed factors are inference strategy (with values Weighted-Sum and Maximum-Minimum), similarity measure (with values Fuzzy Set-theoretic, Fuzzy Set Based Cosine, Fuzzy set Based Proximity-based, Fuzzy Set Based Correlation-like), training size (with values 5, 10, 15, 20, 25, 30) and recommendation size (with values 3,4, 5, 10,15, 20, 25, 30, 35, 40, 45, 50). Univariate analysis of variance using the General Linear Model (GLM) is used to assess the effects of the four similarity measures and two recommendation strategies on the performance of the simulated movie recommender system, see Section IV. The effect of testing size is also studied as covariate variable.

B. Dataset

The benchmark dataset from MovieLens at the University of Minnesota (http://movielens.umn.edu), which has been widely used in recommendation research, is employed in this study. The dataset includes movie attributes, user ratings, and simple user demographic information. The dataset consists of 100,000 ratings (1-5) from 943 users on 1682 movies; each user has rated at least 20 and at most 737 movies. In the dataset, movies are described with: movie id, movie title, release date, video release date, IMDb URL, and 20 genres including action, adventure, animation, children's, comedy, crime, documentary, drama, fantasy, film-noir, horror, musical, mystery, romance, sci-fi, thriller, war, western, family, and unknown. Genres in the MovieLens dataset are represented with binary values, which do not reflect the true content of movies in the genre space. Therefore, we use the proposed representation scheme by incorporating information about movie genres from the Internet Movie Database (imdb.com). The IMDB is a large database consisting of comprehensive information about past, present and new coming movies.

C. Evaluation Metrics

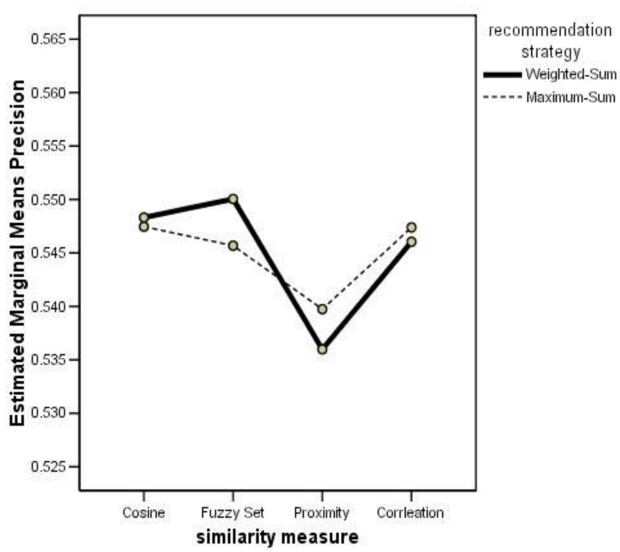
The accuracy metrics are computed for measuring the performance. Precision measures the ratio of correct recommendations being made. Recall reflects the coverage or hit rate of recommendations. In conventional retrieval system, in order to make a complete evaluation of the performance, precision and recall need to be considered together because they are inversely related. A single metric that combines precision and recall is defined as F1-Measure= (2*precision*recall)/(precision + recall).

IV. RESULTS

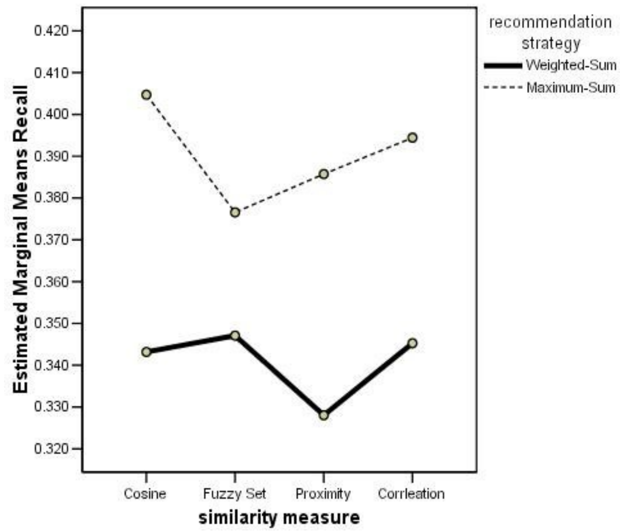
The number of records or evaluation data obtained from the 10 simulation runs is 189,862 cases for each recommendation strategy, and about 95000 for each similarity measure, or about 475000 for each similarity by inference strategy. For precision, the significant factors are: similarity measure, training size, recommendation size, approach by similarity interaction, similarity by training size interaction, approach by recommendation size interaction, and training size by recommendation size interaction. For recall and F1-measure, the significant factors are: all single factors, all two-way interaction effect between two single factors, three-way interaction effect among inference strategy, training size and recommendation size [3]. The means of the recommendation accuracy are presented in Table 1.

Table 1: Means of recommendation accuracy by inference strategy and similarity measure

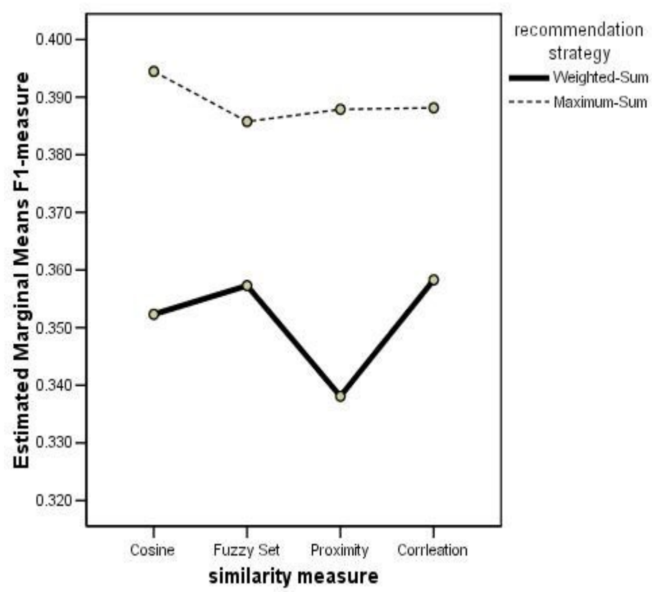
Approach id	Similarity Measure	Precision Mean	Recall Mean	F1-Measure Mean
Weighted-Sum	Cosine	0.550	0.317	0.338
	Fuzzy Set	0.552	0.321	0.343
	Proximity	0.539	0.299	0.323
	Correlation	0.549	0.318	0.343
Maximum-Sum	Cosine	0.550	0.379	0.381
	Fuzzy Set	0.548	0.351	0.371
	Proximity	0.543	0.356	0.371
	Correlation	0.551	0.371	0.375



(a)



(b)



(c)

Figure 1: Means Plot of (a) precision, (b) recall and (c) F1-measure for the different similarity measure by inference strategy

Table 2: Multiple Means Comparisons for Weighted-Sum

	Similarity		Weighted-Sum	
Dependent Variable	(I)	(J)	Mean Difference (I-J)	Sig.
Precision	Cosine	Fuzzy Set	-.00171	1.0
		Proximity	.01079(*)	.00
		Correlation	.00095	1.0
	Fuzzy Set	Proximity	.01250(*)	.00
		Correlation	.00266	.52
	Proximity	Correlation	-.00984(*)	.00
Recall	Cosine	Fuzzy Set	-.00444(*)	.02
		Proximity	.01770(*)	.00
		Correlation	.00149	1.0
	Fuzzy Set	Proximity	.02214(*)	.00
		Correlation	.00294	.28
	Proximity	Correlation	-.01918(*)	.00
F1-measure	Cosine	Fuzzy Set	-.00519(*)	.00
		Proximity	.01511(*)	.00
		Correlation	-.00564(*)	.00
	Fuzzy Set	Proximity	.02030(*)	.00
		Correlation	-.00045	1.0
	Proximity	Correlation	-.02075(*)	.00

Similarity measure and inference strategy interaction effects on the three accuracy metrics are shown in Figure 1. In order to draw further conclusions about which of the pairs are different and establish orders, Pair-wise multiple comparisons test using Bonferroni test are conducted between each pair of means, and yield a matrix where asterisks indicate significantly different group means at an alpha level of 0.05, summarized in Table 2 and Table 3. The Bonferroni test, based on Student's t statistic, adjusts the observed significance level for the fact that multiple comparisons are made. It is commonly used multiple comparison tests, and more powerful for a small number of pairs.

Table 3: Multiple Means Comparisons for Max-Min

	Similarity		Maximum-Minimum	
Dependent Variable	(I)	(J)	Mean Difference (I-J)	Sig.
Precision	Cosine	Fuzzy Set	.00191	1.0
		Proximity	.00650(*)	.00
		Correlation	-.0010012	1.0
	Fuzzy Set	Proximity	.00450(*)	.02
		Correlation	-.00299	.26
	Proximity	Correlation	.00750(*)	.00
Recall	Cosine	Fuzzy Set	.02863(*)	.00
		Proximity	.02302(*)	.00
		Correlation	.00811(*)	.00
	Fuzzy Set	Proximity	-.00561(*)	.00
		Correlation	-.02058(*)	.00
	Proximity	Correlation	-.01491(*)	.00
F1-measure	Cosine	Fuzzy Set	.00967(*)	.00
		Proximity	.00978(*)	.00
		Correlation	.00602(*)	.00
	Fuzzy Set	Proximity	.00011	1.0
		Correlation	-.00365(*)	.00
	Proximity	Correlation	-.00376(*)	.00

V. CONCLUSION AND FUTURE RESEARCH

The results of multiple pair wise means comparison show that there are significant different among the different alternative combination of fuzzy theoretic similarity measures and recommendation strategies in their recommendation accuracy. Hence, depending on tasks and goals of recommender systems, it is practical to choice one of combination of similarity measures and inference strategies that would produce optimal performance using the results are summarized in Table

4. For precision, fuzzy theoretic set and cosine similarity measures perform better for both inference strategies. For recall and F1-measure, correlation and cosine similarity measures performer better with maximum minimum inference strategy. In all cases, distance based similarity measure performs poor.

Future studies include: (i) test with additional datasets and domain applications to see the generalization of the results; (ii) further analysis on variations in sensitivity to training size, recommendation size and their interaction of a recommender system among the four similarity measures within each inference strategy can lead to identifying optimal values.

Table 4: Summary of ordered performance by the inference strategies and similarity measures

Metric	Inference strategy	Order
Precision	Weighted-Sum	Prox < Cos=FS=Corr
	Maximum-Minimum	Prox < Cos=FS Corr < Prox Corr=Cos=FS
Recall	Weighted-Sum	Prox < Cos < FS FS=Corr Cos=Corr
	Maximum-Minimum	FS < Prox < Corr < Cos
F1-Measure	Weighted-Sum	Prox < Cos < Corr=FS
	Maximum-Minimum	Prox=FS < Corr < Cos
Notation: Prox=Proximity or Distance-based; Cos=Cosine; FS=Fuzzy Set-base; Corr=Correlation-like		

V. References

[1] J. A. Konstan and J. Riedl, "An Empricial Analysis of Design Choices," *Information Retrieval*, vol. 5, no.4, pp. 287-310, 2002.

[2] B. M. Sarwar, G. Karypis, J. A. Konstan, and J. Riedl, "Analysis of Recommender Algorithms for E-Commerce," Proc. Proceedings of the ACM E-Commerce 2000 Conference, pp. 158-167, 2000.

[3] A. Zenebe and A. F. Norcio, "Evaluation Framework for Fuzzy Theoretic-Based Recommender System," Proc. 11th International Conference on Human-Computer Interaction, 2005.

[4] R. Altman, *Film/Genre*. London: British Film Institute, 1999.

[5] S.-M. Hsu, C. Wu, and T.-W. Tien, "A Fuzzy Mathematical Approach for Measuring Multi-facet Consumer Involvement in the Product Category Evaluation," *Marketing Research On-Line*, vol. 3, pp. 1-19, 1998.

[6] B. M. Sarwar, G. Karypis, J. A. Konstan, and J. Riedl, "Item-based collaborative filtering recommendation algorithms," Proc. 10th International World Wide Web Conference (WWW10), pp. 285-295, 2001.

[7] M. Deshpande and G. Karypis, "Item-Based Top-N Recommendation algorithms," *ACM Transactions on Information Systems*, vol. 22, no.1, pp. 143-177, 2004.

[8] V. V. Cross and T. A. Sudkamp, *Similarity and Compatibility in Fuzzy Set Theory: Assessment and Applications*. Heidelberg; New York: Physica-Verlag, 2002.