A HYBRID APPROACH BASED ON RECLUSIVE AND COLLABORATIVE METHODS

Dissertation submitted to Jawaharlal Nehru University, in partial fulfillment of the requirement for the award of the Degree of

MASTER OF TECHNOLOGY
In
COMPUTER SCIENCE AND TECHNOLOGY

By
VIBHOR KANT

Under the Guidance of
PROF. K. K. BHARADWAJ

SCHOOL OF COMPUTER & SYSTEMS SCIENCES
JAWAHARLAL NEHRU UNIVERSITY
NEW DELHI -110067

JULY 2010

JAWAHARLAL NEHRU UNIVERSITY
NEW DELHI - 110067

CERTIFICATE

This is to certify that the dissertation entitled "**A Hybrid Approach Based on Reclusive and Collaborative methods**", being submitted by **Vibhor Kant** to the School of Computer and Systems Sciences, Jawaharlal Nehru University, New Delhi, in partial fulfillment of the requirement for the award of the Degree of **Master of Technology in Computer Science and Technology**, is a bona fide work carried out by him under the guidance and supervision of **Prof. K. K. Bharadwaj**.

The matter embodied in the dissertation has not been submitted for the award of any other Degree or Diploma.

Vibhor Kant

Vibhor Kant
(Student)

21-07-10

Prof. K.K.Bharadwaj
(Supervisor)

Prof. Sonajharia Minz

Dean, SC&SS

Jawaharlal Nehru University

New Delhi-67,

India.

Acknowledgement

I owe my deepest gratitude to those who have directly or indirectly, helped me in the completion of this academic attempt.

Above all, I would like to thank Prof. K. K. Bharadwaj, my supervisor at the Jawaharlal Nehru University in New Delhi. He has supported me in all my endeavors and provided encouragement and guidance throughout this work's lifetime. He has been an inspiration to me and it was really an unforgettable experience to work under his guidance. I would like to express gratefulness to him for developing in me independent problem solving approach.

I would also like to thank Prof. Sonajharia Minz, Dean of SC&SS, Jawaharlal Nehru University, New Delhi for providing me the necessary infrastructure to carry out this dissertation. I express my heart-felt indebtedness to my all lab-mates and appreciate their help appropriately in this dissertation. I wish to express my admiration to all of them for their love and affection and being with me through thick and thin.

At last, I would like to thank my parents for their love and support.

VIBHOR KANT

Abstract

Recommender systems have emerged as an important part of the solution to the information overload problem facing today's Web users. Combining ideas and techniques from information filtering, user modeling, artificial intelligence, user interface design and human-computer interaction, recommender systems provide users with proactive suggestions that are tailored to meet their particular information needs and preferences. The main techniques that are used for the recommendations are collaborative and content based filtering. Since human decisions employ a wide range of fuzzy terms and therefore a fuzzy theoretic reclusive method that employs content- based filtering came into the existence for better representation of user's behavior and items. The main strength of collaborative filtering(CF) is its cross-genre or 'outside the box' recommendation ability. However, CF suffers from new item ramp up problem, sparsity and loss of neighbor transitivity. Content based filtering(CBF) provides solution to the new item problem but it suffers from overspecialization.

In this work, we propose a hybrid recommender system using reclusive method (also known as fuzzy theoretic method) and collaborative filtering to overcome the overspecialization problem of CBF and new item problem of CF. First of all, we apply Pearson collaborative method on the available movie-lens dataset and then we apply fuzzy theoretic method. We used weighted scheme of hybridization of these two methods. The experimental results reveal that the proposed hybrid system outperforms the classical approaches.

Table OF Contents

page

List of Figures:

List of Tables:

Chapter 1

1. Introduction

"Although an application designer's first instinct is to reduce a noble human being to a mere account number for the computer's convenience, at the root of that account number is always a human identity" (Khare & Rifkin , 1997).

Internet, as the tool of transferring data and information, helps most industries to be more effective in their specializations. E commerce (Yager, 2000) as the next step after internet guides the discussion into retail and act of selling and buying (Schafer et al., 2001; Terveen & Hill, 2001)(making the act of purchasing via internet and electronic means, speedup the retail activities and avoid making huge losses of energy and cost). Nowadays, we live in the web based information society. The quality of new information available everyday (news, movies, scientific papers, songs, websites,) goes over our limited processing capabilities. So we need something able to suggest us only the beneficial information, recommender system (RS) fulfills our aim. RSs deal with information overload by signifying which information is most relevant to us. RS has affluence of useful and practical applications regarding online content and services. The techniques on improving the effectiveness of RS constitute a problem rich research area.

Content based filtering (CBF) and collaborative filtering (CF) are mainly used in recommender systems. RSs based on CF have some weaknesses: new user problem, new item problem and sparsity and also RSs based on CBF suffer problem of overspecialization. Accuracy and scalability are another issues for these techniques and constitute another research area in RSs. Several soft computing techniques such as fuzzy logic, rough set theory etc. are used in improving the accuracy of RSs. Normally, items representing and reasoning about user behavior and their relationships are the major problems in recommender systems. This is because; these representations are subjective, imprecise and ambiguous. Thus with the aid of fuzzy logic in RS, the performance of recommender system can be enhanced. Reclusive method (Yager, 2003) provides a fuzzy framework for better recommendation accuracy in content based RSs using fuzzy representation of items. Reclusive methods are based upon the item representation but not collaborators.

The following sections give a brief overview of recommender systems, various recommendation techniques, some applications of fuzzy logics in recommender systems, the problem statement and the organization of dissertation.

1.1 Recommender System

"What information consumes is rather obvious: it consumes the attention of its recipients. Hence a wealth of information creates a poverty of attention, and a need to allocate that attention efficiently among the over abundance of information sources that might consume it." (Herbert A. Simon)

Acting upon recommendations from other people is a normal part of life. We do it when we eat at restaurant on the advice of a friend, or we see a movie having read the review in the newspaper of our choice. In each case our decision to act upon a recommendation is based on essentially three premises: first, we trust the recommender; second, we assume that the recommender has sufficient knowledge of our tastes or of the tastes of people like us; third, we assume that the recommender has knowledge of the alternatives available.

Web access has changed from academic to commercial and this has meant that a vast amount of available information is not accessible for the users, because they are unfamiliar that it exists. This situation offers a very attractive framework for researching in the form of new accurate and efficient techniques designed to access this information (Campos et al., 2008). In this framework, recommender systems (RS) have emerged to help people deal with this information overload. (Vozalis & Margaritis, 2006; Waddington, 1996). Broadly speaking, an RS provides specific suggestions about items (or actions) within a given domain, which may be considered of interest to the user. Examples of such applications include recommending books, CDs and other products at Amazon.com (Linden et al., 2003; Sarwar et al., 1998), movies by MovieLens, books at LIBRA, etc.

There are three main steps for a typical RS:

1. The user provides some form of input to the system. These inputs can be both implicit and explicit (Resnick et al., 1994). Ratings provide by users are among explicit inputs whereas time spent reading a website and URLs visited by a user are among implicit inputs.
2. These inputs are used to form a representation of the user's likes and dislikes.

3. The system computes recommendations using these “user profiles”.

1.2 Methodologies in Recommender system

There are several recommendation techniques developed so far. These are mainly based on collaborative and content based filtering approach (Wei et al., 2005). The classification of various techniques of RSs depends upon the sources of data and the use to which that data (Burke, 2002) is put. On the basis of there are five different recommendation techniques as follows:

1.2.1 Collaborative Filtering(CF)

The collaborative filtering (CF) is probably the most familiar, widely implemented, and most mature of the information filtering technologies. Collaborative recommendation systems aggregate ratings or recommendations of objects, recognize commonalities between users on the basis of their ratings, and generate new recommendations based on inter-user comparisons. RSs based on Collaborative Filtering (CF) (Goldberg et al., 1992; Breese et al., 1998; Lathia et al., 2007) try to automate the “word of mouth" (Shardanand & Maes, 1995) process. The perceived in this way is as follows: when we have to decide about going to see a new movie for example, we often ask to some friends with similar movies tastes and then we act based on their recommendations. CF tries to automate this process to a world scale.

A typical user profile in a collaborative system consists of a vector of items and their ratings, continuously augmented as the user interacts with the system over time. Some systems use time-based discounting of ratings to account for drift in user interests. In some cases, ratings may be binary (like/dislike) or real-valued indicating degree of preference.

The CF approach is more interesting and creative. The recommender asks users to rate items so that it knows who likes what. Then, when asked for a recommendation for the current user , it identifies users similar to him/her (neighbors) and it suggests him/ her the items the neighbors have liked in past. The interesting point is that the algorithm doesn't need a representation of the items in term of features but it is based on the tastes of its users' community. e.g., 'Liu and Lucy like *Sleepless in Seattle*. Liu likes *You've Got Mail*. Lucy also might like *You've Got Mail*.' The greatest strength of collaborative techniques is that they are completely independent of any machine-readable representation of the objects being recommended. In addition, they work well for complex objects such as music and movies, where variations in taste are responsible for much of the variation in preferences. (Billsus & Pazzani, 1998)

CF recommendations consist of three steps:

1. In first step, users provide some ratings for those items, they have experienced before.
2. Second, the current user is grouped with other users which have same taste with the current user. To do so, similarity measures are used to compute similarity between the current user and other users in the system. There are some similarity measures are used such as Pearson, Jaccard and Cosine etc. After computing similarity, a neighborhood set is formed for the current user according to some threshold or "top – N" similar users.
3. In the last step, predictions are made for the items that the current user has not yet rated, but the neighbour has rated, (Resnick et al., 1994; Herlocker et al., 1999).

There are two approaches to design of CF recommender systems: memory based and model based. (Koren, 2008)

Memory Based CF

Memory-based algorithms utilize the entire user-item database to generate a prediction. These systems employ statistical techniques to find a set of users, known as *neighbors*, that have a history of agreeing with the target user (i.e., they either rate different items similarly or they tend to buy similar sets of items). Once a neighborhood of users is formed, these systems use different algorithms to combine the preferences of neighbors to produce a prediction or *top-N* recommendation for the active user (Sarwar et al., 2001; Drachsler et al., 2007). The techniques, also known as *nearest-neighbor* or user-based collaborative filtering are more popular and widely used in practice (Jameson et al., 2003). Two major measures for similarity computation are Pearson correlation coefficient and Cosine based. These approaches compute similarity between two users only on the items that both users have rated .There are some extensions of these correlation based and cosine based techniques for better performance(Adomavicius et al., 2005).Memory based CF algorithms offer more accuracy but not more scalable because of memory intensive.

Model Based CF

Model-based collaborative filtering algorithms provide item recommendation by first developing a model of user ratings in an off-line learning phase and algorithms in this category take a probabilistic approach and envision the collaborative filtering process as computing the expected value of a user prediction, given his/her ratings on other items. The model building process is performed by different *machine learning* algorithms such as Bayesian network (Manouselis & Costopoulou, 2007), clustering, (Kim & Ahn, 2008) and rule-based approaches through Apriori algorithm (Sullivan et al., 2002). The Bayesian network model (Breese et al., 1998) formulates a probabilistic model for collaborative filtering problem. The clustering model treats collaborative filtering as a classification problem (Basu et al., 1998; Breese et al., 1998; Ungar et al., 1998) and works by clustering similar users in same class and estimating the probability that a particular user is in a particular class C and from there computes the conditional probability of ratings. The rule-based approach applies association rule discovery algorithms to find association between co-purchased items and then generates item recommendation based on the strength of the association between items (Sarwar et al., 2000). Model based approaches are more scalable but not more accurate. Memory- based and model based approaches have been combined in (Al-Shamri & Bharadwaj, 2008) using a fuzzy genetic approach that retains the accuracy of memory-based CF and the scalability of model-based CF.

There are some strengths of CF which are as follows:

- Cross-genre niches
- No limited content analysis
- No overspecialization
- Domain free technique

But CF suffers some problems:

- Cold start problem(Adomavicius et al.,2005)
- Sparsity (Gřcar et al., 2006 ; Lee et al., 2004)
- First rater problem
- Loss of neighborhood transitivity
- Scalability(Bell et al., 2007)

1.2.2 Content Based Filtering(CBF)

The *content-based filtering* (CBF) is an outgrowth and continuation of information filtering research. It makes recommendation based on the correlation between difference resources. In content-based recommendation systems, resources are described as a vector of attributes. For example, text recommendation systems like the newsgroup filtering system NewsWeeder use the words contained in the documents and their frequencies as features.

Content based approaches to recommendation making are deeply rooted in information retrieval approach (Baudish, 2001). A content-based filtering system selects items based on the correlation between the content of the items and the user's preferences as opposed to a collaborative filtering system that chooses items based on the correlation between people with similar preferences.

There are four steps of CBF (Mooney et al., 2000)

1. To assemble content data about the items.
2. To ask user to provide some ratings for the items either randomly given, or can search and find any item he or she likes
3. To compile profile of the user using the content information extracted in the first step and rating provided in second step. Term frequency/inverse document frequency weighting (Lang., 1995) and Bayesian learning algorithm (Mooney et al., 2000) are some techniques for profile compilation.
4. To match unrated items' content with user profile obtained from third step and assigning values to the items depending on quality of the match.

There are some strengths of CBF

- No Sparsity problem

- Adaptiveness
- No new item problem
- Providing explanations of recommended items by listing content-features that caused an item to be recommended.

But CBF suffers some problems:

- Cold start problem
- Limited content analysis
- Overspecialization

1.2.3 Demographic Filtering

The demographic recommendation systems aim to categorize the user based on personal attributes and make recommendations based on demographic classes. Demographic RSs (Pazzani,1999) use prior knowledge on demographic features like age, sex, gender etc. about the users and classify users into different classes based on these features. Recommendations are then made based on these demographic classes. One of the most popular example of demographic RS is Lifestyle Finder (Krulwich, 1997). Demographic filtering can be combined with other filtering techniques generally CF in order to overcome their drawbacks.(Vozalis & Margaritis, 2007) This technique does not depend upon the user's past rating history thus it avoids the new user problem. Major disadvantage of this technique is that there may be several users who don't fall in any demographic class.

1.2.4 Knowledge Based Filtering

The *knowledge-based recommendation* attempts to suggest objects based on inferences about a user's needs and preferences (Burke, 2002). In some sense, all recommendation techniques could be described as performing some kind of inference. Knowledge-based approaches are distinguished in that they have functional knowledge. The user profile can be any knowledge structure that supports this inference. In another words knowledge-based RSs recommend items based on logical reasoning about user preferences. They use prior knowledge on how an item meets users' needs and can therefore reason about the relationship between a need and a possible recommendation. One example of this type RS is the restaurant recommender Entrée System.

1.2.5 Utility Based Filtering

Utility based RSs use utility based approaches like as multi-attribute utility theory and the aim of calculating a utility for all items for a specific user and thus are able to order the items on the recommendations (Manouselis and Costopoulou, 2007). Each Utility –Based system like Tete- a-Tete or e-commercial site personal logic (www.personalogic.aol.com/go/gradschools) has different technique for arriving at a user-specific utility function.

1.2.6 Hybrid Recommender System (Burke, 2002)

Hybrid recommender systems combine two or more recommendation techniques to gain better performance with fewer of the drawbacks of any individual one. The reason is to make use of their combined strengths and to leave out their corresponding weaknesses (Burke, 2002). Most of the hybrid approaches are geared towards unifying collaborative filtering with content based filtering under one single framework, absconding synergetic effects and justifying inherent deficiencies of either paradigm. Most commonly, collaborative filtering is combined with some other techniques. P-tango system, DailyLearner system and LIBRA system are some examples of hybridization of content based and collaborative filtering techniques. FAB is the first meta level hybridization for web filtering (Balabanovic & Shoham, 1997).

Burke lists an entire plethora of hybridization methods to compare of recommendation algorithms. Table 1.1 shows hybridization methods such as weighted, switching, mixed, feature combination, cascade, feature augmentation and meta level. This table also shows some examples of these hybridization methods. First four techniques are order insensitive and remainings are order sensitive.

Techniques	Description	Examples
Weighted	The scores (or votes) of several recommendation techniques are combined together to produce a single recommendation.	P.Tengo,Pazzani
Switching	The system switches between recommendation techniques depending on the current situation.	PTV
Mixed	Recommendations from several different recommenders are presented at the same time.	Daily-learner
Feature combination	Features from different recommendation data sources are thrown together into a single recommendation algorithm.	Basu ,Hirsh and Cohen
Cascade	One recommender refines the recommendations given by another.	Fab,EntreeC
Feature augmentation	Output from one technique is used as an input feature to another.	Libra, Grouplens
Meta-level	The model learned by one recommender is used as input to another.	Fab

Table 1.1 Different hybridization techniques (Burke, 2002)

1.3 Fuzzy Recommender System

Fuzzy logic is an explicit notional system of reasoning, deduction and computation in which the objects of discourse and analysis are associated with information which is, or is allowed to be, imperfect- imprecise, uncertain, vague, incomplete, unreliable, partially true (Zadeh , 1965) Everything is or is allowed to be a matter of degree (membership degree) in fuzzy logic. Generally, representation of items and user's behavior and reasoning about their relationships are major problems in recommender systems. This is because these representations are subjective, imprecise and vague. Thus with the aid of fuzzy logic in RS, the performance of recommender system can be enhanced. In the following subsections, we discuss some applications of fuzzy logic in RSs.

1.3.1 Reclusive Methods

Reclusive method (Yager, 2003) is based solely on the preferences of the single individual and makes no use of the collaborators .It also depends upon the representation of the objects. In this method, fuzzy set methods are used for the representation and subsequent construction of justifications and recommendation rules.

1.3.2 Recommender System for consumer electronic products

Because the consumer electronic products (such as laptops, digital cameras, etc.) are more expensive than those frequently-purchased commodities, (Yukun & Yunfeng ,2007) . We cannot argue on the personal preference of a common customer purchasing history and provide appropriate information services to meet his needs. Recommendation systems based on consumer preferences are not sufficient to recommend this type of products(Shih & Liu, 2008). Thus, it is necessary to construct a new recommendation system for consumer electronics products. In this study, there is a fuzzy-based recommendation system on the less frequently purchased products, including consumer electronics., When a common buyer goes for shopping, some distinctive features of consumer electronics may get him into some troubles .These features are as follows: First, compared with other products , the life of a new model consumer electronic is short about two years. Second , 'one get one free' scheme for newly-produced models , and also a great number of new techniques come forth to improve product functions. Last, because of the large gap in price among the different models of the same product and the continuously decreasing the price of a new model within its life cycle, (it is hard for a common consumer to know the prices of all product models). (Yukun & Yunfeng, 2007; Choi et al., 2006) The proposed study aims to assist a consumer to navigate characteristic space of the product interactively that the consumer has his own need for each feature dimension so that the client can find the best products according to his personal preferences. In this study, proposed recommender system is established for laptop recommendations.

Based on the proposed system, the consumer needs and characteristics of the candidate product may be expressed in an appropriate manner. In this approach, triangular fuzzy numbers are used to characterize consumer needs and product features. A triangular fuzzy number is a particular case of fuzzy sets. It has a triangle-shaped membership function, which can be viewed as possibility distribution. The linguistic terms are used to linguistically evaluate the importance of customer needs and ratings of product features so that a consumer can easily express his judgements and domain experts can easily evaluate the product features. . (Yukun & Yunfeng, 2007) Seven linguistic sets are allowable to describe the variables of one's subjective judgment: *(1) very low, (2) low, (3) medium low, (4) medium, (5) medium high, (6) high, (7) very high.*

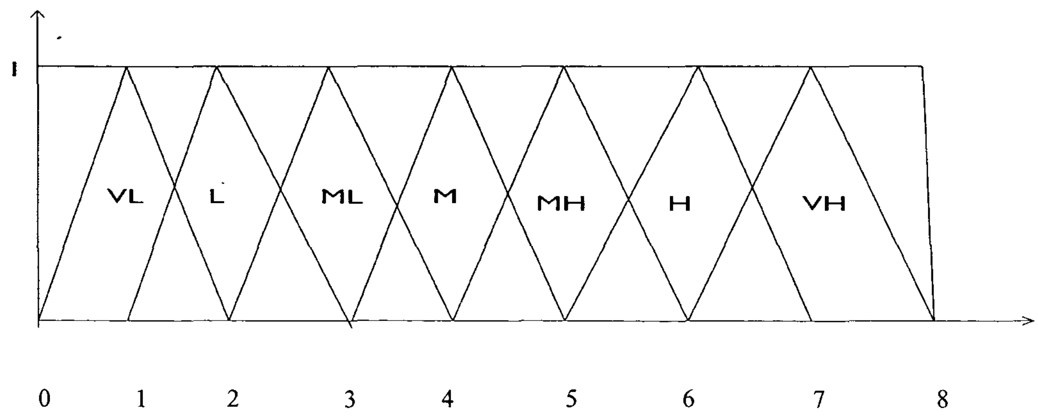

Figure 1.1 Membership functions for linguistic terms (Yukun & Yunfeng, 2007)

1.3.3 A Fuzzy-Genetic approach in Recommender Systems

CF techniques are based on memory based and model based approaches. Memory based CF are more accurate but not more scalable than model based CF. Fuzzy genetic approach maintain the accuracy of memory based CF as well as the scalability of model based CF. Fuzzy sets are applied to fuzzify age and genre interestingness measure .

Generally, human decisions are in ambiguous, imprecise in nature so the crisp description of the age and genre interestingness measure does not reflect the real case for man decisions. The distance, for example, the age of 6 and 10 is 4 between two users, while the users belong to the same age group, namely childhood. These features should be fuzzy taken into account when making comparisons between two and more users. The goal of this approach is to fuzzify the proposed model to get as close as possible a set of neighbors for active user.

(Al-Shamri & Bharadwaj, 2008). In this approach, firstly age is fuzzified into three fuzzy sets *young, middle-aged and old.*

Let **a** and **b** be the membership vectors correspond to two crisp values a and b for a given feature with l fuzzy sets. The fuzzy distance between a and b is defined as -

fd (a, b) =d (**a, b**)   d (a, b)

where d (a, b) is simply the difference operator, a and b are vectors of size l, and d (**a, b**) is Euclidean distance metric. In this approach genre interestingness measure (GIM) is also fuzzified in six fuzzy sets *very bad (VB), bad (B), average (AV), good (G), very good (VG), and excellent (E)* with the membership functions as given in the Figure 1.2 and Figure 1.3. (Al-Shamri & Bharadwaj, 2008)

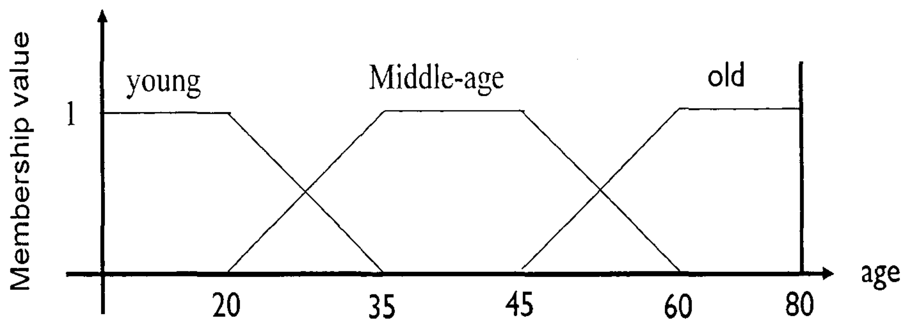

Figure 1.2 Membership functions for age feature(Al-Shamri & Bharadwaj, 2008)

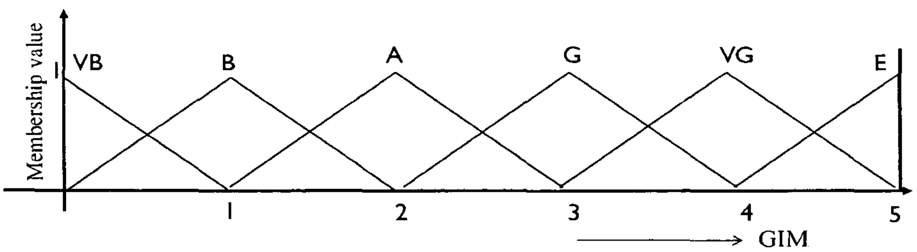

Figure 1.3 Membership functions for GIM feature (Al-Shamri & Bharadwaj, 2008)

1.3.4 Fuzzy computational models for trust and reputation systems

E-business consists of the buying and selling of products or services over electronic systems such as the Internet and other computer networks. People buy and sell goods , play, and recommend products for each other without knowing the remote partner in the online environment(shopping and online service providing). These are feasible only when if partners trust each other. (Bharadwaj & Al-Shamri, 2009) In real life, many persons believe the world of future will be based on reputation: reputation will become the only "currency" of our life. Reputation is a public view concept in our social circle. However, it is not always acceptable to depend solely on the public view and therefore a trust measure is required to give an illustrated view of the future encounters with a particular partner. Trust plays a crucial role in computer mediated transactions and processes. Trust and reputation systems represent a significant trend in decision support for Internet mediated service provision. In spite of the obvious importance of trust and reputation, transferring them into computer science is still a challenging issue. E-commerce seems to be the most benefitted area from the trust and reputation concepts amongst different areas of computer science like semantic web and security. RS has an important role in e-commerce. Reputation systems are collaborative in nature so these are related to collaborative RSs. (Massa & Avesani, 2004)

Nowadays, with the emergence of online communities, e-market places, weblogs and peer-to-peer communities, a new kind of information is available: rating expressed by a user on another user. Based on these ratings online communities provide a better solution to the consumer/user. Recently, many large electronic communities have been established and have grown very fast unfortunately, a major weakness of e-markets is the high level of risk associated with the loss of the notions of trust and reputation. Online service provision commonly takes place between parties who have never transacted with each other before, the consumer is generally forced to accept the "risk of return before", i.e. pay for services and goods before receiving, which can leave him in a vulnerable position (Bharadwaj & Al-Shamri, 2009). On the other hand, the service provider has much more information on the quality of products that the customer, as it receives payment before shipment in most cases. The inefficiencies resulting from this information asymmetry can be palliated through trust and reputation. The idea is that even if the consumer can not try the product or service in advance, he can be satisfied that it will be what he wants as long as he trusts the seller. A trusted seller therefore has a significant advantage in case the product quality cannot be verified in advance.

Demonstrations of trust are easy to recognize because we experience and rely on it every day, but at the same time trust is quite challenging to define because it manifests itself in many different forms. The concept of reputation is closely linked to that of trustworthiness, but it is evident that there is a clear and important difference. According to this model definitions of trust, reciprocity and reputations are as follows

Trust is a subjective expectation a partner has about another's future behavior based on the history of their encounters. It is context-dependent, asymmetric and dynamic.

Reciprocity is the mutual exchange of deeds (such as favor or revenge). Thus the reciprocity is a symmetric property.

Reputation is what is generally said or believed about a person's or thing's character or standing. It is indirect information. Since trust and reputation are often a gradual phenomenon, fuzzy set methods are the pre-eminent tools for modeling such networks.

Fuzzy computational model for trust

Trust is a complex concept that has several properties. Amongst these properties one is symmetric (reciprocity) whereas the other is asymmetric (experience). Reciprocity is the mutual favor or revenge and therefore to model it, we need to find the agreement (both individuals are Satisfied or unsatisfied) and disagreement (only one of them is unsatisfied) between two partners. (Bharadwaj & Al-Shamri, 2009) So there are two fuzzy subsets of each partner's rating – Satisfied and unsatisfied. The sum of the membership values of these two fuzzy sets is 1. The possible combinations of satisfied and unsatisfied fuzzy subsets define four values for any two partners, namely satisfied-satisfied SS, unsatisfied -unsatisfied UU, satisfied-unsatisfied SU and unsatisfied-satisfied US. Depending on these values the agreement and disagreement between two partners can be calculated easily.

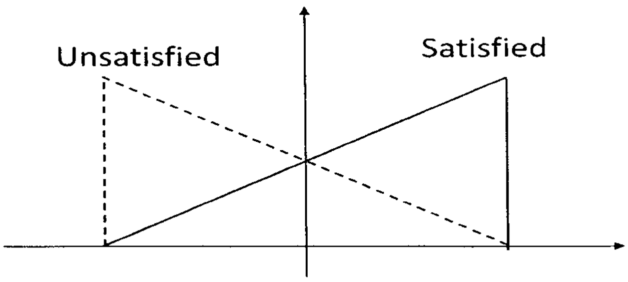

Figure 1.4 Membership functions for partner's rating (Bharadwaj & Al-Shamri, 2009)

Fuzzy computational model for reputation

To compute the reputation for a given partner, each member of the community has to give him a reputation score based on their previous encounters and this score from an individual should be independent of the others opinions. The overall reputation score is the aggregated with the help of OWA operator from all individuals' scores. The OWA guided by a linguistic quantifier '*usually*' is used for the aggregation process. (Bharadwaj & Al-Shamri, 2009). The individual reputation score needs to be associated with a reliability measure. This will measure how reliable is the score received from a given individual.

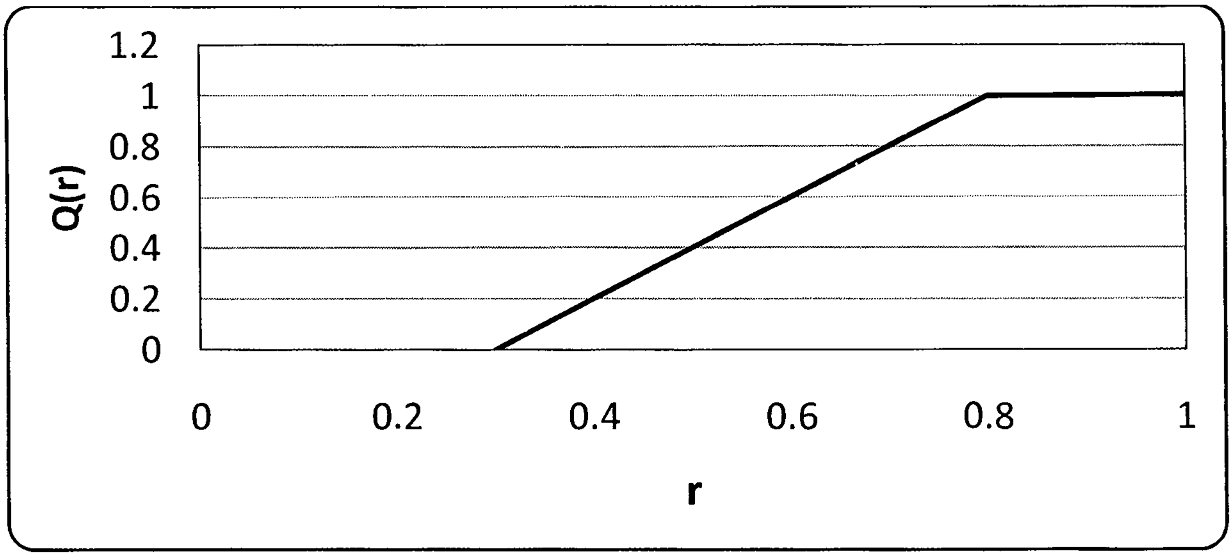

Figure 1.5 Membership function for 'usually' quantifier (Bharadwaj & Al-Shamri, 2009)

In this dissertation, we propose a hybrid recommender system which is based on a reclusive method for content based RS and a pore CF. In this work, we use weighted hybridization technique which gives better results .This hybridization is depends upon the use of fuzzy logic methods to improve the accuracy of the RSs. The experimental results support our theory and demonstrate that the proposed method is superior to the general fuzzy theoretic method: a reclusive method (Zenebe, 2009) and a pure CF.

1.4 Thesis Roadmap

The rest of the dissertation is organized as follows .Chapter 2 discusses the brief introduction of reclusive method. A hybrid method is proposed by using different similarity measures and aggregation methods in Chapter 3. Chapter 4 gives details of various experiments conducted and the analysis of results so obtained. Finally, Chapter 5 presents the conclusion and points out some directions for future work.

Chapter 2

Reclusive Methods: A Brief Overview (Yager, 2003)

In this chapter, we describe reclusive method, how is this method different from the general collaborative method? What are the requirements for reclusive methods? What are the ways of representation of items and user feedback? How are the different ways from which a user can express his preferences ?, How can a model be generated from users' preferences ?, and at last, after generating the model how can the recommendation rules be generated with the help of users' model and object representation ? In this chapter we discuss all these types of questions.

Recommender system is the most popular in e-markets. Generally, collaborative technique is used in constructing recommender system. Recommender system provides most appropriate suggestions to an active user with help of his profile. Collaborative method has some minuses so there is another important technique for recommender system is content based filtering. Normally a human talk is based on fuzzy observations. So reclusive method came into the existence. Reclusive method is a special type of content based filtering which is based on fuzzy logic methods.

Reclusive methods are based solely on the preferences of the individual for whom we provide the recommendation and not to use the preferences of other collaborators. An important characteristic of these reclusive methods from collaborative methods is that they require a representation of objects. The following subsections of reclusive method is based on (Yager, 2003).

2.1 Representation of users' behavior

Generally, in all types of recommender systems, primary requirement for constructing recommender system is current user's profile. For this, recommender system asks from user about some information .There are two modules of information in which a user can provide his preferences first, is extensional, and second, is intentional .

Extensionally expressed information This type of information is based upon the user's past behavior, actions and experiences.

For example - the movies, a user has seen in the past and some ratings for these movies provided by the user in movie domain.

Intentionally expressed information This type of information is based upon the interest of the user's need in object. In another way, we can say some specifications provided by a user that the user desires in objects of type under consideration. These specifications are associated with the attributes and features used in the representation of the objects. Figure 2.1 shows the information structure for recommender system

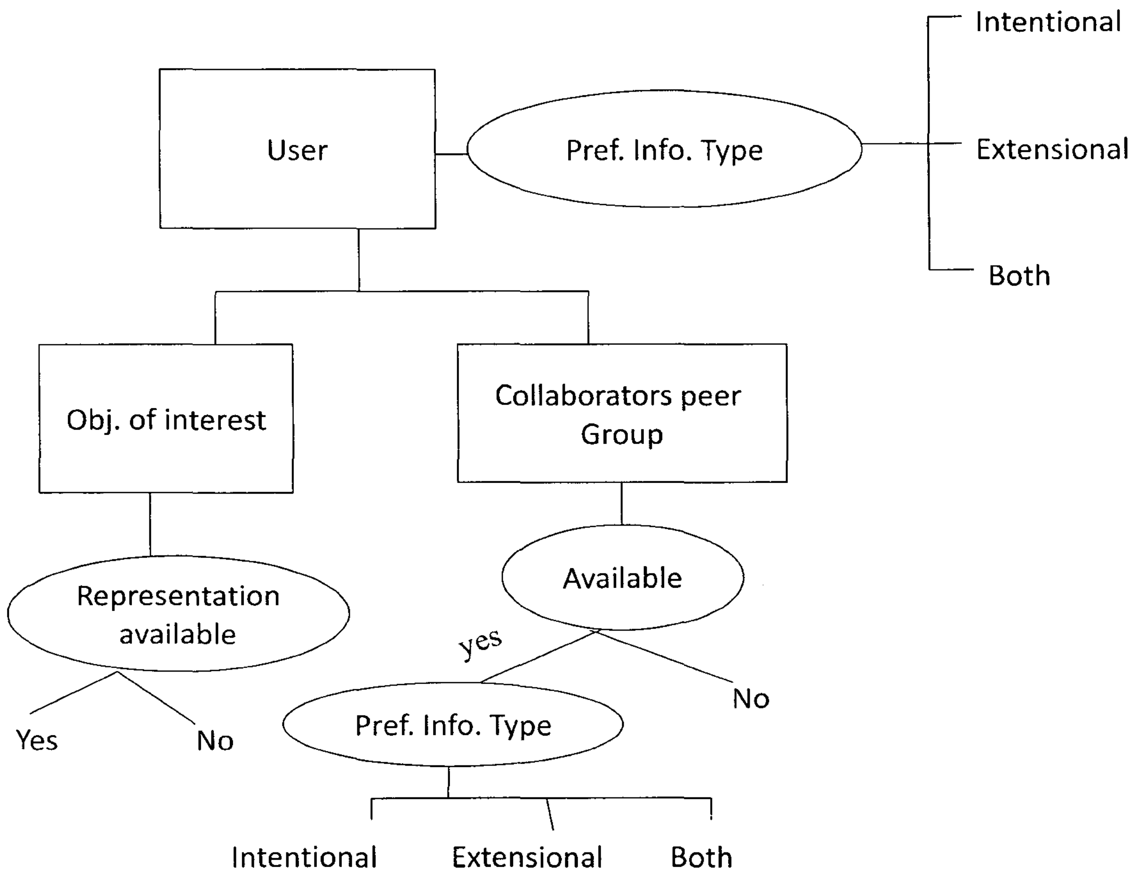

Figure 2.1 Information structure for recommender system (Yager, 2003)

For developing different types recommender systems, we need different types of the information expressed by a user. For constructing collaborative type recommender system, we need only extensional preference information; similarly for non- collaborative, reclusive type recommender systems, we need only intentional preference information and object representation.

Different types of information we can classify different types of recommender system. The Table 2.1 differentiates the different types of recommender systems according to different type of information.

RS typology	Reclusive Method	Collaborative filtering	Reclusive method	Reclusive method	Hybridization of both of these
Extensional preferences			√	√	√
Intentional preferences	√			√	√
Object representation	√		√	√	√
Collaborators		√			√

Table 2.1 Classification of different types of recommender systems (Yager, 2003)

Suppose D be the set of objects such as movies. For each object the user has experienced a representation R_i containing current user's rating a_i . For an unexperienced Object d we only have a representation R, The procedure for obtaining the degree of recommendation also involves following two steps:

Procedure for degree of recommendation:

- Obtain a degree of similarity $\widehat{S_i}$ of the object d with the experienced objects by combining R and R_i .
- Rating of d = Aggregation of weighted tuples ($\widehat{S_i}$, a_i).

2.2 Object representation

Normally, description of any object is depend upon its characteristics .These characteristics is described some specific attributes in a specific domain. So we can say that object representation is based upon the presence or absence of these attributes. In crisp representation, if an attribute is present in the object then it is given by one otherwise it is given by zero. In reclusive method there is different scenario of presence or absence of these attributes.

In this method, an attribute is characterized by the set of primitive assertions or combination of these assertions so object representation shall be based upon a set of primitive assertions . Associated with each object and each assertion is a value τ lying between the unit interval indicating the degree to which the assertion is valid for that object. Suppose A be the set of primitive or atomic assertions such as A= $\{A_1, . . . A_n\}$. Then A_j(d) indicates to which this assertion is satisfied by d. For some purposes object d can view as a fuzzy subset over the space A. Using this perspective the membership grade of A_j in d,

$$d(A_j) = A_j(d) = a_j.$$

An attribute is a subset of related assertions. Suppose F= $\{V_1, V_2, . . . V_q\}$ be a collection of objects. Each attribute V_j corresponds to a subset of assertions which can be seen as constituting the possible values for the attribute. If V_j is an attribute, then for a particular object d, the value of this attribute is obtained by

$V_j(d) = A(V_j) \cap d$, where d represents the fuzzy subset of A correspond to object d.

In movie domain, a movie can be represented in the terms of one and more genres (generally nineteen genres are considered for movie representation). Suppose a movie has three genres comedy, romance and drama and rest 16 genres are absent. Then , in crisp representation for the presence of a genre , one is used and absence of genre , zero is used. But for more specific representation for above example there are some questions in mind how much comedy , how much romance and how much drama is presented in this movie . Answers of all these types of questions can easily described by the help of fuzzy set theory. The degree of presence of particular genre in a movie is characterized by the membership degree of that genre.

Crisp representation of movies (objects) in space of genres:

Movie Id	Comedy	Romance	Drama
M 1	1	1	0
M 2	0	1	1

Table 2.2(a) Crisp representation of movies

Fuzzy representation of these movies in space of genres:

Movie Id	Comedy	Romance	Drama
M 1	1.0	0.8	0
M2	0	0.9	0.6

Table2.2 (b) Fuzzy representation of movies

2.3 Modeling of user expressed preferences

The basic function of a recommender system is to use justifications to produce recommendations for a user. By justification we designate a rational being to believe that a user may like an object. These justifications may be obtained either directly from the preferences expressed by users or induced using data on user experiences.

How can a user expresses his/her preferences in more specific ways .We have already talked about that the normal human talk is based upon the fuzzy observations so fuzzy degree of membership is characterized by linguistic quantifier . After expressing the preferences by the user, then OWA operator (Yager & Kacprzyk, 1997) is used for aggregation. In the next paragraph, we discuss all of these in detail.

The quality of the performance of a recommender system is highly dependent on the ability of the system to allow the user to effectively express their preferences intentionally. This capacity itself depends both on the attributes/assertions used to represent the object and the sophistication of the language provided to the user in which he /she expresses his/ser preferences regarding these attributes. A vocabulary of linguistic quantifiers $Q = \{Q_1, Q_2, \dots Q_q\}$ is provided to user. With the help of these quantifiers (Yager, 1985), a user can express his /her preferences effectively.

There are two main preference models for representation user preferences information in reclusive method - Primal Preference Model (PPM) or Basic Preference Model (BPM). OWA operator is used to aggregate user preferences. A user can express preference by describing what properties they are interested with respect to the class of objects in D and then using Q to capture the desired relationship between these properties. For example do they desire *all* or *most* or *some* or at *least one* of these requirements satisfied? If h is a PPM which is a module of the form $h = \langle A_1, \dots A_s : Q \rangle$.

In particular for object d we obtain the values $A_j(d)$ from our representation of d then use the OWA aggregation to evaluate it , $h(d) = OWA_q(A_1(d), A_2(d), \dots A_s(d))$. Here the weighting vector is determined from Q. (Yager, 2003)

A BPM is a module of the form $m = \langle C_1, C_2, \dots C_p : Q \rangle$ in which the C_i are called the components of the BPM . The only required property of these components is that they can be evaluated for each object. That is for any C_i we need to be able to obtain C_i (d). Once having this we can obtain using the OWA aggregation $m(d) = OWAq\ [C_1(d), \dots C_p(d)]$.

User profile is constructed by using these BPMs One part of the user profile is the user preference profile which consists of a collection of basic preference modules , m_j for j = 1 to k. Each m_j provides a description of a class of objects from D that the user likes. For any object d, $m_j(d)$ indicates the degree to which it satisfies the BPM m_j. If $M = [m_1, m_2, \dots m_k]$ is the preference profile for a given user, then for any object d in D we calculate M(d) = $Max_j[m_j(d)] \wedge \alpha_j$ as the degree of positive recommendation of this object to the user, Where α_j indicates the strength of these BPMs.

N (d) = $Max_i[n_i(d)] \wedge \beta_i$] as the degree of negative recommendation of this object to the user, where β_i indicates the strength of these basic rejection module.

Now overall degree of recommendation for an object d in one of the two different ways:

$$R(d) = (M(d) - N(d)) \vee 0$$

OR

$$R(d) = (1 - N(d)) \wedge M(d).$$

These are justifications deducted by the user model which is constructed with the help of BPMs or PPMs.

2.4 Guidelines for Recommendation

If a user expresses his preferences extensionally and fuzzy representation of objects are available then in this environment , reclusive method for RS tries to find those objects which are the most similar to that of user's past liked experiences. For this ,similarity measurement is used to find the similarity between un-experienced and experienced objects .These experienced objects are lying in the liked category for the active user , after finding similarity between objects , a recommendation score is calculated for an un-experienced object.

Let E be the set of those movies that a user has previously seen in which the user has provided some score lying in [0, 1] for each movie of E. A can be viewed as a fuzzy subset of E of the score or ratings for each object d_j in E such that A(d_j)=a_j. If R_j is a collection of circumstances for recommending objects where indicates the degree to which $d_i \epsilon M = D - E$ meets this condition then

$$R(d_i) = Max_j[R_j(d_i)] .$$

There are two guidelines for recommendation of an un-experienced object.

Rule 1: Recommend an object if there exists a *similar* object that the user *liked.*

$$R_1(d_i) = \underset{j \epsilon E}{Max} [S(d_i, d_j) \wedge A(d_j)],$$

where $S\ (d_i, d_j)$ *is the similarity between* experienced and unexperienced object.

Rule 2: Recommend an object for which we have *at least several comparable* objects which the user somewhat *liked*

$$R_2(d_i) = \underset{F \subseteq E}{Max} [Q(|F|) \Lambda \underset{d_j \in F}{Min} (\tilde{A}(d_j) \Lambda comp(d_i, d_j)],$$

where Q is the fuzzy subset for quantifier at least several , $\tilde{A}\ (d_j)$ is for somewhat likeness and comp(d_i, d_j) is the softening for similarity(Zadeh, 1983).

Chapter 3

A hybrid approach based on Reclusive and Collaborative Methods

In order to improve recommendation accuracy and to address some of the limitations (e.g., new item) of CF recommender systems, a hybrid recommender system can also be augmented by reclusive method. There are several hybridization techniques for this hybridization.

In this chapter, we have given a brief overview of fuzzy theoretic method for content based recommender system: a reclusive method, Collaborative filtering, all possible hybridization techniques for these two methods, dataset construction for this hybridization, performance measurement of this hybrid approach and main steps for proposed scheme.

All necessities of a RS, fuzzy theoretic method (FTM) provides us in different manner. Reclusive method is a fuzzy framework for constructing content based recommender system. FTM method is an advance empirical version of reclusive method. Representation of items and user feedback, similarity measurements and aggregation methods for recommendation, all of these are provided by FTM in fuzzy framework. All requirements of FTM follow the methodology of reclusive method.

A content based recommendation based on FTM requires data on the behavior of users and features of items. Its performance depends on the data and how this data is used. There are so many challenges about the representation of items and user feedback because features of items and user behavior are subjective, vague and imprecise. Such uncertainty is non-stochastic or non-random and is induced from subjectivity, vagueness and imprecision in data and domain knowledge. In relation to items, the uncertainty is associated to the extent (e.g. low to high) in which the items have some features. For instance, to what extent does a movie has comedy content or is it highly comedy related? In relation to the user's behavior such as interest, the uncertainty is associated to methods employed to measure and represent users' interest as precisely as possible. (Zenebe, 2009)

In fuzzy modeling, Membership functions in fuzzy set theory are intentionally designed to handle the vagueness and imprecision in the context of the application. FTM can address the representation and the challenges associated with the non stochastic uncertainty in recommendation systems. FTM provides representation of items and user's feedback, similarity measures and aggregation techniques.

3.1. Representation of items in FTM

An item I_j (j = 1,… M) is defined in the space of an attribute X $=\{x_1, x_2, \dots, x_L\}$, then I_j can take multiple values such as $x_1, x_2, \dots, x_L$. The membership function of item I_j *to value* $x_k(k = 1,2, \dots, L)$ is denoted by $\mu_{x_k}(I_j)$. Hence a vector $X_j = \{(x_k, \mu_{x_k}(I_j))\}$ is formed for item I_j.

In movie domain, a movie (item) M_j is represented in space of genres. A movie can have one major genre denoted by x_1 and one or more minor genres x_2, x_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,… M) to genre

$x_k(k = 1,2, \dots, N)$ is denoted by $\mu_{x_k}(M_j)$. (Zenebe, 2009)

Different values and membership functions of X for a movie M_j follows the following criteria:

(1) Assigning higher degree of membership to major values than minor values

(2) Assigning 0 to values that are not associated with the movie;

(3) Degrees of membership should be normalized to the range of [0, 1]; and

(4) The same value of X at similar rank positions between different movies should have varying degrees of membership values if the number of values of X associated with the movies are different.

For this there is a Gaussian-like membership function, (Zenebe, 2009)

$$\mu_{x_k}(M_j) = \frac{r_k}{2\sqrt{\alpha * |L_j| (r_k - 1)}} \qquad (3.1.1)$$

where $N = |L_j|$ is the number of values of X associated with M_j and r_k is the rank position of value, and

$\alpha > 1$ is a parameter used as a threshold to control the difference between consecutive values of X in M_j? Using this perspective a movie Race having five genres is represented as

x_j = {(Action, 1), (Adventure, 0.366), (Drama, 0.272), (Fantasy, 0.211), (Thriller, 0.168)}.

Movie Id	Comedy	Romance	Drama
M 1	1.0	0.8	0
M 2	0	0.9	0.6

Table 3.1 Representation of movies in FTM reclusive method

3.2. Users feedback representation in FTM

The primary requirement for an RS is user's behavior for items. Normally, user rating is the most popular user feedback in recommender systems. It is an alternate way to represent the user degrees of interest in an item. User ratings are represented as binary values-those liked or disliked. Those above 3 are considered as liked in five scale rating pattern. However, user rating is really imprecise as a user may give different ratings to the same item at different times and situations. User degrees of interest based on user rating is treated as a fuzzy variable and its uncertainty is represented using a possibility distribution function (Zadeh, 1978). This fuzzy variable degree of interest is classified into different fuzzy values such as- strongly liked (SL), liked (L), indifferent (I), disliked (D) and strongly disliked (SD). Degree of interest is associated with user rating (R) expressed in between from minimum value (Min) to maximum value (Max). Then, the proposition 'a user has strongly disliked an item I' has the possibility distribution function, (Zenebe, 2009)

$$\prod_R (I) = \mu_{SD}(R = r), \qquad min \leq r \leq max$$

A half triangular fuzzy number (Pedrycz & Gomide, 1998) is used to represent the degree of positive experience for a user in relation to an item

The half triangular fuzzy number membership function, for user rating r on $M_j \in [min, max]$ and for a fuzzy subset A on DI is defined as

$$\mu_A(M_j) = \frac{(r-min)}{(max-min)} \qquad (3.2.1)$$

3.3. Similarity measures in FTM

Computing similarity between users and items is the most important issue for a recommender system research. Similarity computation between items is calculated by Pearson correlation formula, cosine similarity. In this study some similarity measures are suggested to compute similarity between items (movies) in fuzzy environment. A movie is represented in space of genres.

For movies M_j and M_k that are defined as $\{(x_i, \mu_{x_i}(M_j)) : i = 1,2,\dots,N\}$ and $\{(x_i, \mu_{x_i}(M_k)) : i = 1,2,\dots,N\}$, a similarity between these two movies is denoted by (M_k, M_j) .

There are different similarity measures in this environment which are as follows

$$S_1(M_k, M_j) = \frac{\sum_i min(\mu_{x_i}(M_k), \mu_{x_i}(M_j))}{\sum_i max(\mu_{x_i}(M_k), \mu_{x_i}(M_j))} \qquad (3.3.1)$$

$$S_2(M_k, M_j) = \frac{\sum_i \mu_{x_i}(M_k) * \mu_{x_i}(M_j)}{\sqrt{((\sum_i(\mu_{x_i}(M_k))^2}\sqrt{((\sum_i(\mu_{x_i}(M_j))^2}} \qquad (3.3.2)$$

$$S_3(M_k, M_j) = 1 - \left(\frac{d_2(M_k, M_j)}{\underbrace{Max}_{i}\{\mu_{x_i}(M_k), \mu_{x_i}(M_j)\}}\right) \qquad (3.3.3)$$

$$where\ in\ formula\ (3.3.3)$$

$$d_2(M_k, M_j) = \sqrt{\sum_i (\mu_{x_i}(M_k) - \mu_{x_i}(M_j))^2}$$

Movie id	Crime	Horror	Mystery	Thriller	Drama
C_1	1	0	1	1	1
F_1	1	0	0.44	0.35	0.29
C_2	0	1	1	1	0
F_2	0	1	0.41	0.47	0

Table 3.2 Crisp and fuzzy representation of movies in FTM method (Zenebe, 2009).

C_1 = {Satya 1995: Crime/Mystery/Thriller/Drama}, and C_2= {Vastushashtra 2004: Horror/Thriller/Mystery} with their crisp (C_1 and C_2) and fuzzy (F_1 and F_2) representations. Using crisp set theoretic (similar to (1)), crisp cosine (similar to (2) except that 1 or 0 are used instead of membership degree) and crisp distance measures (similar to (3) except that 1 or 0 are used instead of membership degree) the similarity between C_1 and C_2 are 0.40, 0.58, and 0.73, respectively. Finally, using the corresponding fuzzy set theoretic similarity measures (1), (2), and (3) the similarity coefficients are 0.24, 0.25, and 0.55. (Zenebe, 2009).Fuzzy set theoretic measure is used in my proposed scheme.

3.4. Aggregation methods in FTM

The recommendation decision-making methods used in FTM are case-based decision. There are different recommendation score aggregation methods in order to make recommendation decision within the framework of fuzzy and possibility theory. Amongst these, the two alternatives are weighted-sum and max–minimum. In my proposed scheme, weighted- sum are used. (Dubois et al., 1999)

Weighted-sum

For each targeted item M_j calculate the weighted sum (weighted-sum) recommendation confidence score as (Zenebe, 2009)

$$R_1(M_j) = \sum_k \{\mu_E(M_k) S(M_k, M_j)\} \qquad (3.4.1)$$

where E is a set of positively experienced items by users, and $\mu_E(M_k)$ is the membership of item M_k to the set E.

Maximum–minimum

For each targeted item M_j calculate the maximum–minimum (max–min) recommendation confidence score as (Zenebe, 2009)

$$R_2(M_j) = \underbrace{max}_{k} \{min(S(M_k, M_j), \mu_E(M_k))\} \quad (3.4.2)$$

3.5 Performance metrics

After constructing recommender system, we select those triplets from the training dataset in which user's rating are known. Let T = {u, i, r} be the set of users, items and ratings .There are so many performance measure of the recommender system. Mean Absolute Error (MAE) and Root Mean Square Error (RMSE). But we used MAE.

Mean Absolute Error if T = {u, i, r} be the set of users, items and ratings for test.

Then $$MAE = \frac{1}{|T|} \sum_{(u,i,r) \in T} |p_{u,i} - r| \qquad (3.5.1)$$

where $|T|$ is the total no. of triplets for which we are testing, $p_{u,i}$ is the predictive rating by the system and r is actual rating by the active user.

3.6 Collaborative method

Collaborative filtering makes the recommendation by finding correlation among user of recommendation system. Collaborative recommender system refers to multiple users sharing recommendation in the form of ratings. From the study the recommendation technique, we know that the collaborative filtering in not to predict ratings but rather to improve the quality of recommendation by hybridization of content based with collaborative. Collaborative filtering user-based recommendation approaches try to identify neighborhoods' for each user based on similar features – *e.g.* demographics, psychographics, behavioral. Most collaborative techniques work based on ratings about items provided for users.

There are three main steps of collaborative filtering technique

First, **user profile information** users generally state their preferences by rating items to express their like or dislike for a particular item.

Second, **Neighborhood formation** based on the preferences the active user is matched with other users having similar tastes, to give a set of neighbors. In this work for CF, we used Pearson formula is used for similarity measurement.

Pearson similarity measurement

$$sim(x,y) = \frac{\sum_{s \epsilon S_{xy}} (r_{x,s} - \bar{r_x})(r_{y,s} - \bar{r_y})}{\sqrt{\sum_{s \epsilon S_{xy}} (r_{x,s} - \bar{r_x})^2 \, (r_{y,s} - \bar{r_y})^2}} \qquad (3.6.1)$$

where S_{xy} is the set of items which are rated by both users x and y.

and third, **Prediction** Ratings from all neighbors who have rated the item in question are aggregated to arrive at the predicted rating for the active user. Resnick formula is used for prediction in this approach.

Resnick formula

$$p_{a,i} = \bar{r_a} + \frac{\sum_{u=1}^{k}\left(r_{u,i} - \bar{r_u}\right) \times sim(a,u)}{\sum_{u=1}^{k}|sim(a,u)|} \qquad (3.6.2)$$

where a is the active user and u be a user of neighborhood set of a, sim (a, u) is the similarity between a and u which is computed by Pearson formula.

Performance of CF recommender system is calculated by same function MAE or RMSE but we are used only MAE.

3.7 Dataset construction

Normally, Movielens and Net-flix are available datasets in the recommender system research. In these datasets, items are movies and genres are used for movie representation. In movie-lens dataset item representation is crispy, but our work is related to fuzzy environment. So there is requirement to construct dataset for testing and training of recommender systems. In movie –lens dataset, if a movie has a particular genre, then it is represented by one otherwise it is zero. But in construction we gave it a value lying between 0 and 1 according to their ranks described above by using above formula (3.1.1). We will discuss movie lens dataset in next chapter. Here we describe only how did we develop our appropriate item representation. We have already described how can an item be represented in this environment in section 3.1.

3.8 Proposed Scheme

Collaborative filtering and content based filtering are the most popular techniques in recommender system. FTM reclusive method is content based technique for a recommender system in fuzzy environment. But CF and CBF also have some limitations such as new item problem for CF and overspecialization problem for CBF. In order to overcome these limitations and better performance, we have tried the hybridization of reclusive and collaborative methods using weighted hybridization scheme (Burke, 2002).

The main steps involved in the proposed scheme are as under:

Step 1: Compute the similarity between users

Compute PC similarity, sim(x,y), between all pairs of users, x and y using the formula (3.6.1).

Step 2: Compute the neighborhood set for an active user

After finding similarity between the active user and remaining all users, sort the similarity vector in decreasing order and find the top-N users for a particular active user. Let N be the set of neighbors for an active user.

Step 3: Obtain the predictive ratings using neighbors for an active user in CF method

Compute (using resnick formula (3.6.2)) the predictive rating for an unseen movie for an active user. Let $p_{a,i}$ is the predictive rating for an unseen movie i for a user a.

Step 4: Compute the similarity between movies

Compute the similarity between all pairs of unseen movies and user's experience movies, using the formula (3.3.1). Let $S_1(M_k, M_j)$ refers to the fuzzy set theoretic similarity between M_k and M_j as computed in this step.

Step 5: Obtain the membership value for each item in the set of experienced movies for an active user

The set E of experienced movies is the collection of movies seen by an active user. Compute the membership value for each movie in the set E by using formula (3.2.1).

Step 6: Obtain the normalized recommendation score for an unseen movie to an active user in F TM reclusive method

Compute the recommendation score, $R_1(M_j)$ for an unseen movie M_j for an active user by using weighted-sum aggregation formula (3.4.1). After finding recommendation scores for all unseen movies to an active user, compute the normalized score by dividing the highest recommendation score.

Step 7: Obtain the predictive ratings to an active user in reclusive method

Based on membership value for each movie, compute the predictive ratings according to the following rules

- If (normalized score$<$.125) then rating=1
- If (.125 $\leq$ normalized score $<$ .375) then rating=2
- If (.375 $\leq$ normalized score $<$ .62) then rating=3
- If (.62 $\leq$ normalized score $<$ .86) then rating=4
- If (normalized score $\geq$.86) then rating=5

Let $p^*_{a,i}$ is the predictive rating for an unseen movie i for an active user a.

Step 8: Weighted hybrid of CF and FTM reclusive method

Merge the predictive ratings by using weighted hybrid such as

$$p_{a,i}^{hybrid} = \alpha * p_{a,i} + (1 - \alpha) * p^*_{a,i}$$

In this formula, α is positive real number which describes the weight for the recommendation technique.

Step 9: Compute MAE

Compute MAE for active user a by using following formula:

$$MAE(a) = \frac{1}{|S_k|} \sum_{i=1}^{|S_k|} \left| p_{a,i}^{hybrid} - r_{a,i} \right|$$

where S_k is the set of test ratings of user a.

The overall MAE of all active users can be computed as

$$MAE = \frac{1}{N_A} \sum_{a \epsilon N_A} MAE(a)$$

MAE is used to compute the accuracy of proposed CF-FTM reclusive hybrid approach.

Chapter 4

Experiments and results

In this chapter we present the results of conducting experiments using the proposed method in this work. The experiments are conducted comparing the proposed weighted scheme with the FTM reclusive method and CF method. The performance of proposed method and its algorithms are evaluated using movies as items. The movie lens dataset, experimental setup and performance metrics are present below

4.1 Dataset

The benchmark dataset from MovieLens at the University of Minnesota (which has been widely used in recommender system research) is used in this approach. This dataset was collected by GroupLens Research Project at the University. For our purpose we made some modifications in the dataset accordingly which is discussed in the previous chapter. This dataset is properly arranged in separate files. The description of the MovieLens dataset files is given below:

Data File: This file consists of 100,000 ratings by 943 users on 1682 items and each user has rated at least 20 movies.

User File: Demographic features about each user are given in this file.

Genre File: It contains a list of genres.

Item File: This file consists of the nineteen fields, representing the genres. Boolean value are used for genres , 1 indicates whether a genre is present in a specific movie and 0 indicates that genre is not present in that movie; a movie can be described by more than one genre.

The dataset consists of 100,000 ratings (1–5) from 943 users on 1682 movies; and 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, and 19 genres including action, adventure, animation, children's, comedy, crime, documentary, drama, fantasy, film-noir, horror, musical, mystery, romance, sci-fi, thriller, war, western, family. All ratings follow the 1-bad, 2-average, 3-good, 4-very good, and 5-excellent numerical scale. Demographical information for each user such as age, gender, occupation and zip codes are included for all users. The different users' occupations are administrator, artist, doctor, educator, engineer, entertainment, executive, healthcare, homemaker, lawyer, librarian, marketing, none, other, programmer, retired, salesman, scientist, student, technician and writer.

Experiments are conducted to compare the predictive accuracy of the proposed weighting hybrid approach with CF and FTM reclusive method. This scheme is compared against CF and FTM reclusive method under different configurations enable comparison between CF and FTM under MAE.

4.2 Experimental setup

On the basis of movieLens, we considered only users who have rated at least 60 movies, 24 to build a user model and 36 for testing. Only 497 users satisfied this condition out of 943 users and contributed 84,596 ratings out of 1, 00,000. For our experiments, we choose six datasets, containing 100,150,200,250,300 and 350 users called ML100, ML150, ML 200, ML 250, ML 300, and ML 350 respectively. Also all these datasets satisfy this condition. This is to illustrate the effectiveness of the proposed scheme under varying number of participating users. Each of these sub datasets was randomly split into 40% training data and 60 % test data. The ratings of the items in the test set are treated as items unseen by the active user, while the ratings in the training set is used for neighborhood construction and for prediction of ratings. For each dataset the experiment was run 16 times to eliminate the effect any bias in the data. In this experiment, we have described the result in the following two manners.

(a) Comparison of overall MAE of the various methods of a particular dataset among these datasets with different runs.
(b) Also, comparison of user-wise MAE of the dataset ML350 with different methods.

The effectiveness of the proposed scheme is compared with FTM reclusive method and Pearson based CF.

4.3 Performance evaluation

There are several metrics to compare the prediction accuracy of recommender systems. Here we compare the various schemes via Mean Absolute Error (MAE) to compare the prediction accuracy. MAE measures the average absolute deviation of the predicted rating from the actual ratings.

4.4 Comparison of overall MAE

In order to test the effectiveness of the hybrid schemes in this framework we compare the overall MAE of the proposed scheme with CF and FTM methods. We test the proposed scheme with the CF and FTM with α set to a value which results in the least MAE. The best α is obtained empirically, by taking α values in the interval (0, 1) in increments of 0.05.

4.5 Analysis of the results

To demonstrate the ability of the proposed method to offer better prediction accuracy we compare the MAE with that FTM and CF. The results are shown as in table 4.1. The MAE is computed based on the average over 16 runs of the experiment over the different datasets. A lower value of MAE corresponds to a better performance. The results show that the proposed hybrid scheme outperforms other techniques FTM and CF for all datasets with respect to MAE. The MAE for different runs of the experiment for ML350 is shown in figure 4.1. We have also showed the user-wise comparison of average MAE for a particular dataset ML350 in sixth run. This is shown only to find the number of users who get good quality prediction under the various schemes. User-wise MAE is the average MAE over all predictions for an active user. Figure 4.2 through 4.8 show the comparison of user-wise MAE of hybrid approach with FTM and CF. The MAE comparison for the 350 users are plotted in groups of 50 users. Over all 288 users hybrid approach gave more accurate predictions. For these datasets α is found is 0.85 which gave me better MAE.

Datasets		FTM reclusive method	Pearson CF method	Proposed Hybrid Approach
ML100	MAE	1.28142	0.88900	0.86851
ML150	MAE	1.27473	0.87758	0.85508
ML200	MAE	1.26855	0.86881	0.84657
ML250	MAE	1.26768	0.87433	0.85172
ML300	MAE	1.25771	0.87428	0.85143
ML350	MAE	1.25726	0.88275	0.85983

Table 4.1: MAE comparison of the proposed hybrid scheme with FTM reclusive and CF methods.

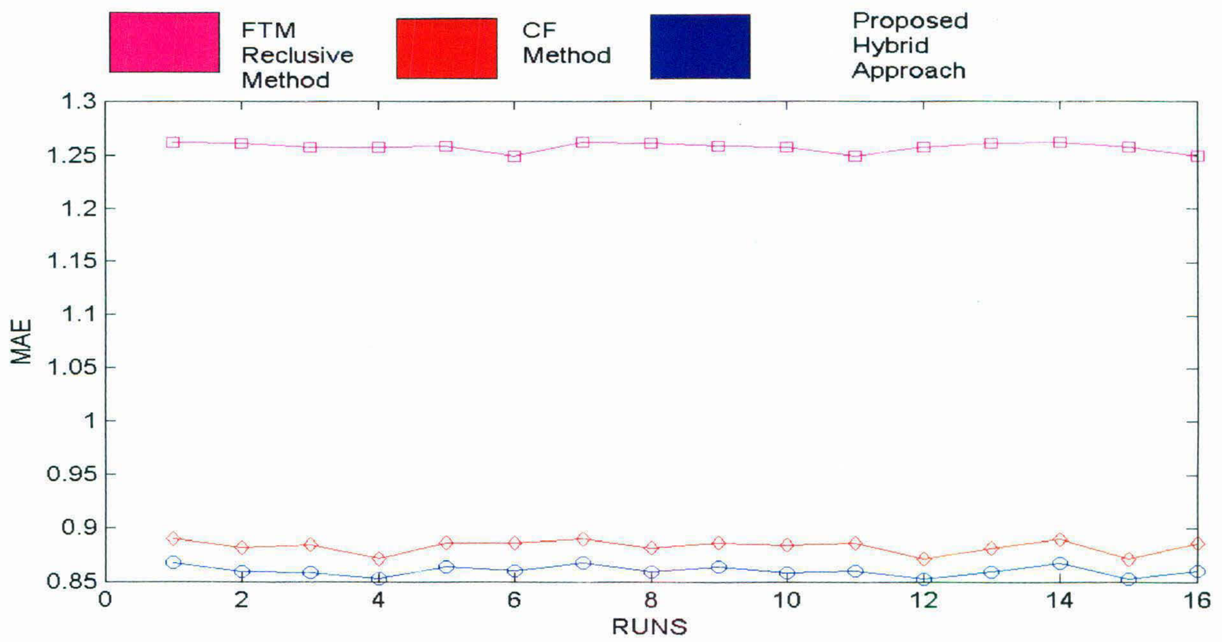

Figure 4.1: MAE for ML350 over 16 runs

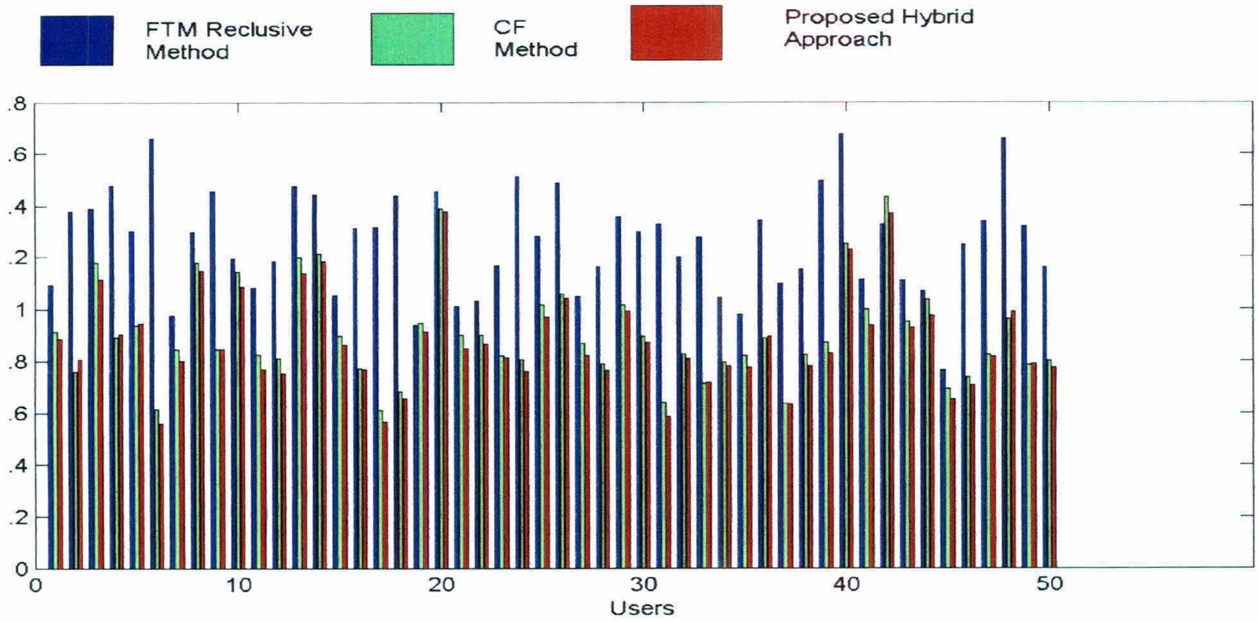

Figure 4.2: Mean Absolute Error for users 1-50

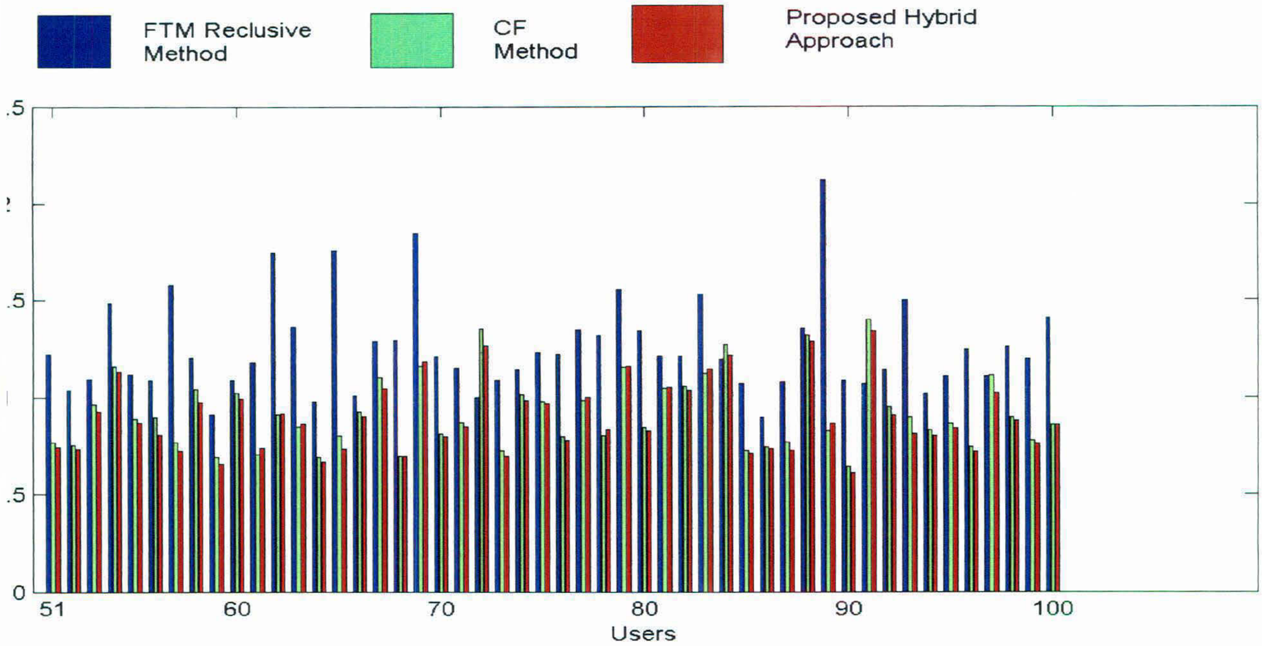

Figure 4.3: Mean Absolute Error for users 51-100

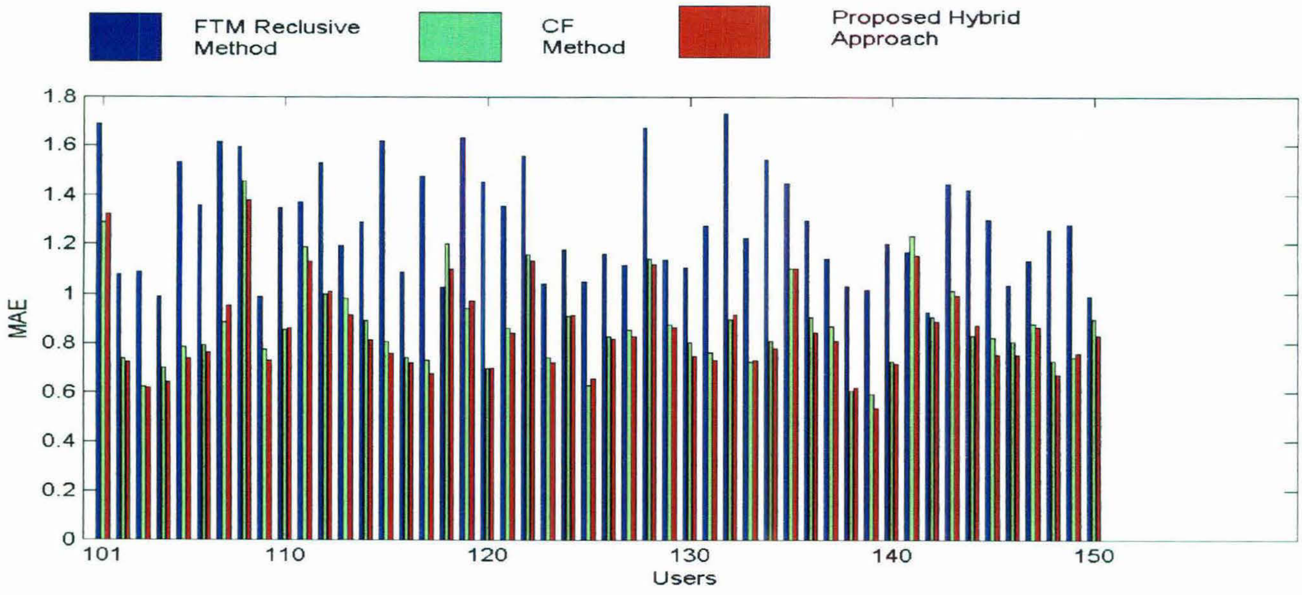

Figure 4.4: Mean Absolute Error for users 101-150

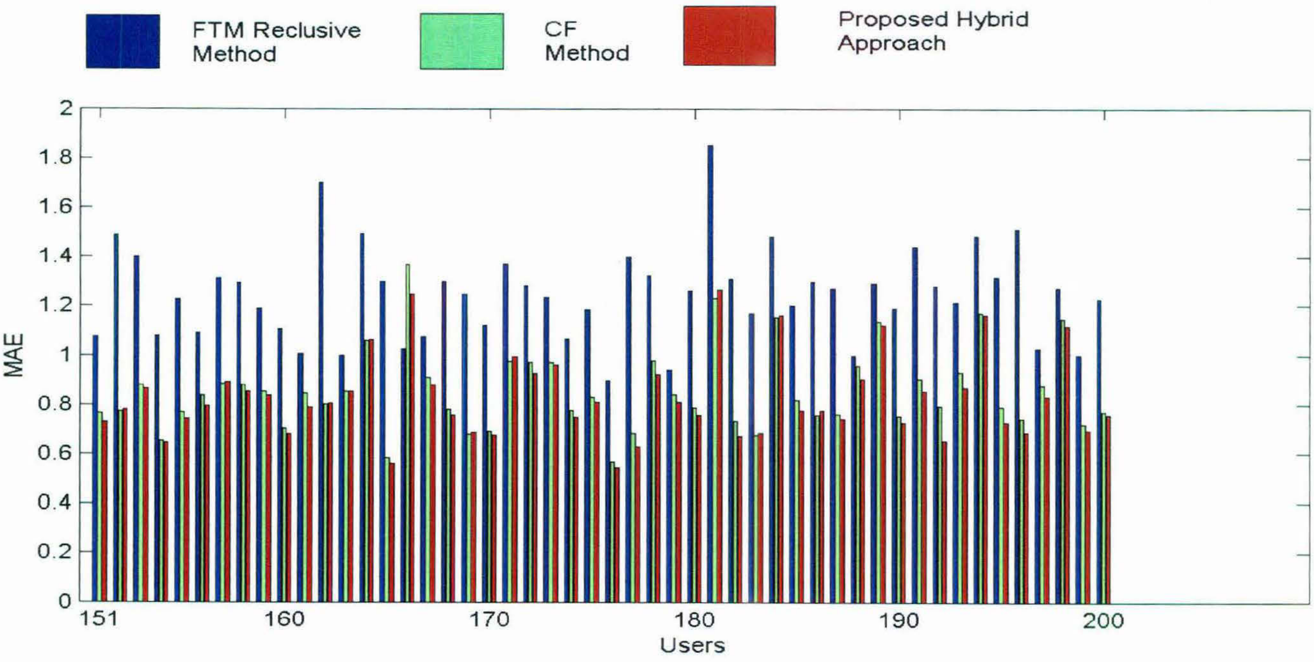

Figure 4.5: Mean Absolute Error for users 151-200

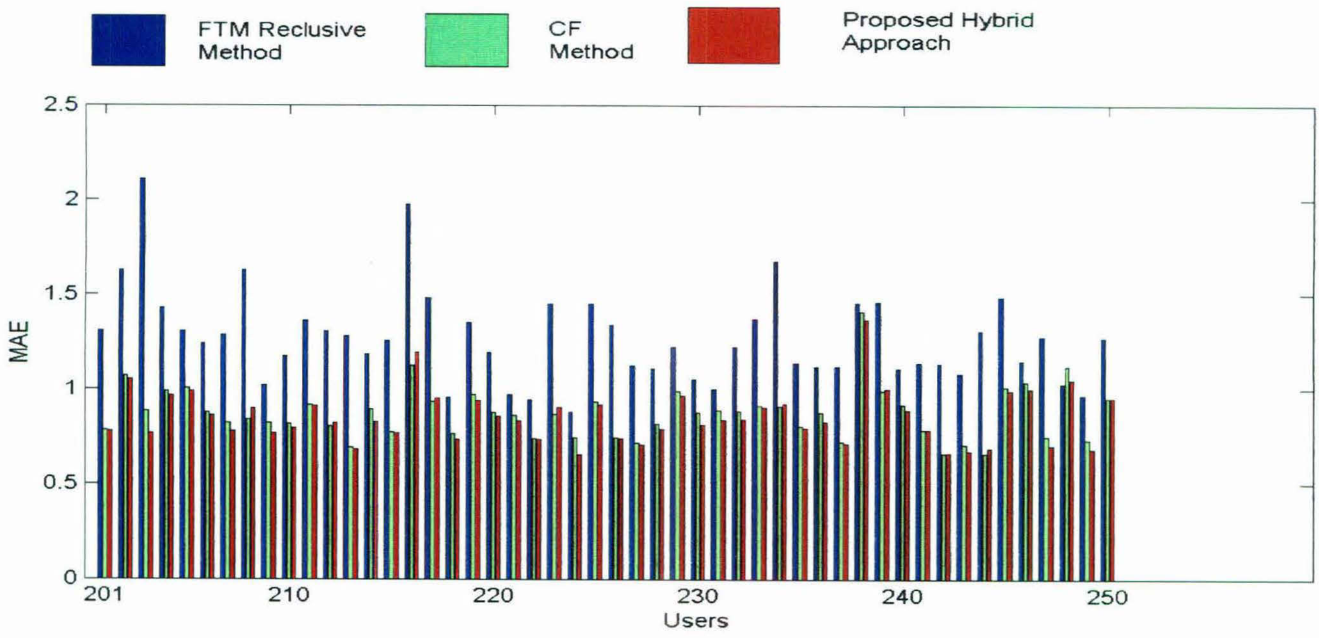

Figure 4.6: Mean Absolute Error for users 201-250

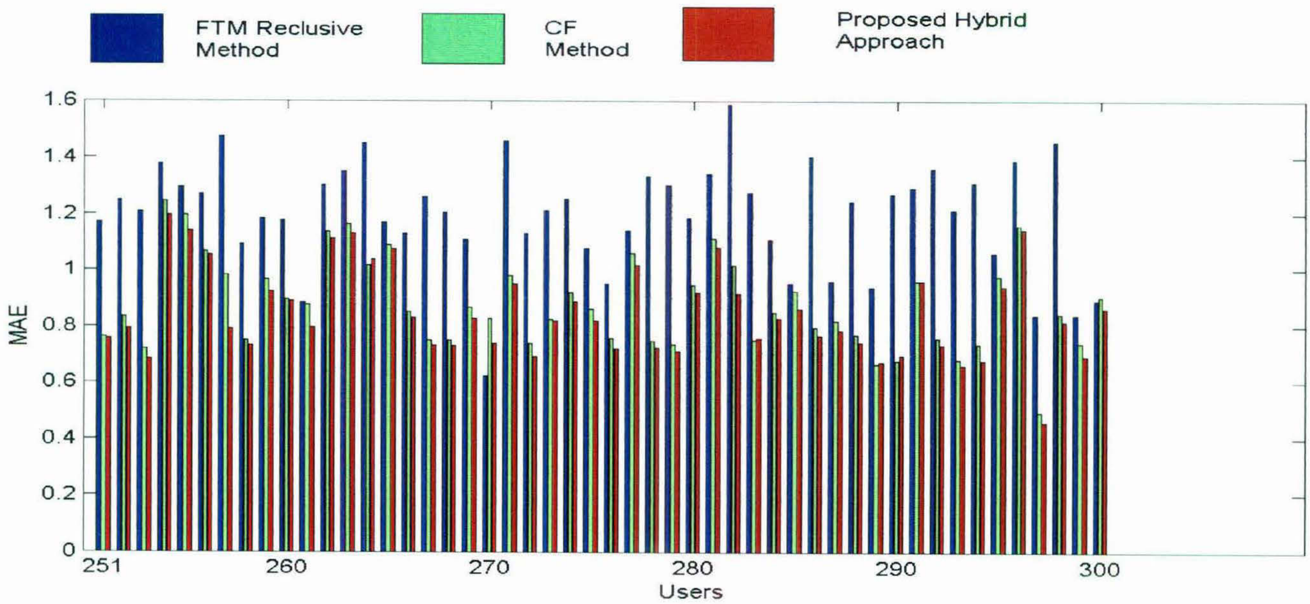

Figure 4.7: Mean Absolute Error for users 251-300

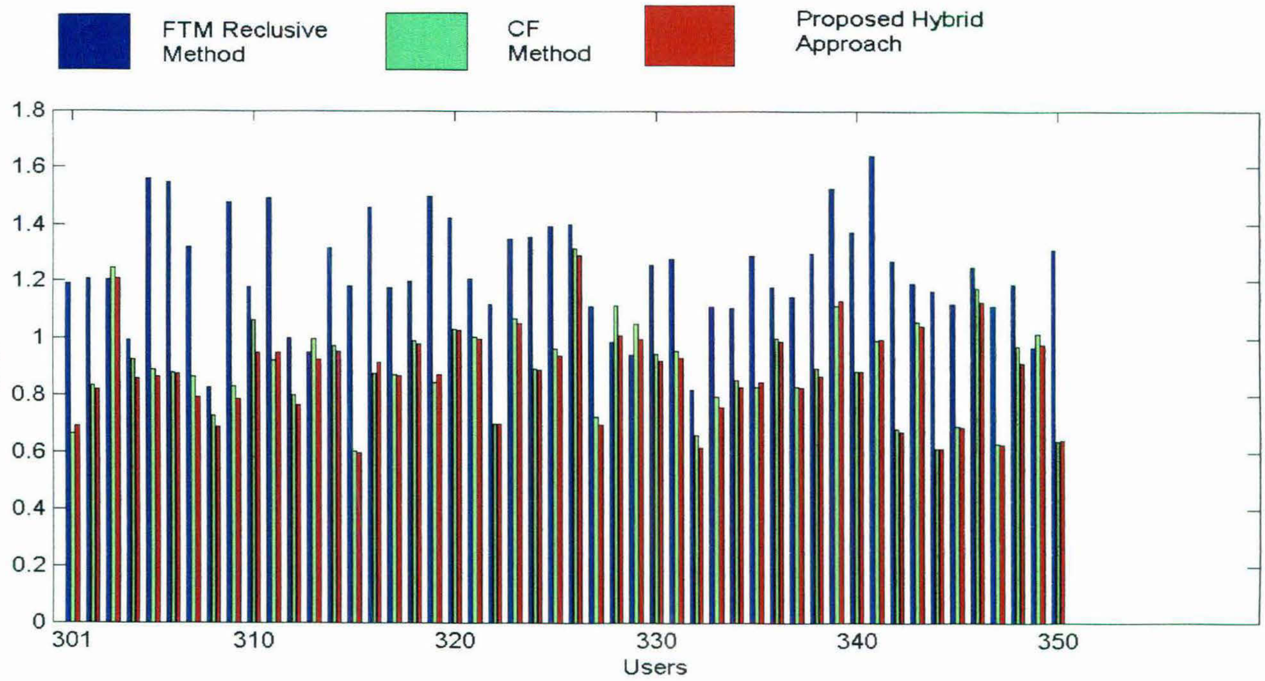

Figure 4.8: Mean Absolute Error for users 301-350

Chapter 5

Conclusion and Future Work

Conclusion

In spite of extensive research in the domain of recommender systems in the industry as well as in academia, the current generation of recommender systems face several challenges the key issue among them being accuracy. A pure CF is widely used technique in recommendation research. There are some issues related to CF- scalability, Sparsity and new item problem etc. Second important technique is CBF for RS but it suffers overspecialization, new user problem. FTM reclusive method is fuzzy based CBF. Hybridizing collaborative filtering with reclusive method can significantly improve the predictions of a recommender system. In this dissertation, we have provided an effective way of achieving this. This study provides a solution to the new item problem for a pure CF and removes the overspecialization of FTM reclusive method. The results of this study provide experimental evidence that supports the effectiveness of this hybrid approach in terms of better accuracy.

Future Work

In this study, we focused on the hybridization of pure CF and FTM reclusive method using weighted hybridizing technique. Therefore, one of the possible direction for future work would be to take into consideration other hybridization techniques. (Burke, 2002)

Further studies would be required to extend the proposed hybrid RS in several directions. First, incorporating trust and reputation concepts (Bharadwaj and Al-Shamri, 2009) at the pure CF level in the proposed system needs to be investigated. Second, different membership functions, different aggregation operators and several linguistic quantifiers needs to be considered for better performance of FTM. Third, inclusion of importance of actors, actresses and directors can be used to improve the accuracy of recommender system. Fourth, using GIM feature. (Al-Shamri & Bharadwaj, 2008) in FTM method the performance of this system can be enhanced.

REFERENCES

Adomavicius, G. & Tuzhilin, A. (2005). "Toward the next generation of recommender systems: A survey of the state-of-the-Art and possible extensions." *IEEE Trans. Knowledge and Data Engineering,* Volume *17, No. 6, 734–749.*

Al-Shamri, M.Y.H. & Bharadwaj, K.K. (2008). "Fuzzy-genetic approach to recommender system based on a novel hybrid user model." *Expert Systems with Applications, Volume 35, Issue 3, 1386-1399.*

Balabanovic, M. & Shoham, Y. (1997). "Fab: content-based, collaborative recommendation." *Communications of the ACM, Volume 40, No. 3, 66-72.*

Basu, C., Hirsh, H. & Cohen, H. (1998). "Recommendation as classification: using social and content-based information in recommendation." *In Proc .of the 15th Nat. Conf. on Artificial Intelligence, 714–720.*

Baudish, P. (2001). " Dynamic information filtering." *Ph.D. thesis, Technische Universitat Darmstadt, Germany.*

Bell, R. M., Koren, Y. & Volinsky, C. (2007). "Modeling relationships at multiple scales to improve accuracy of large recommender systems." *Proc. 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 95-108.*

Bharadwaj, K.K. & Al-Shamri, M.Y.H. (2009). "Fuzzy computational Models for trust and reputation systems." *Electronic Commerce Research and Applications, Elsevier, Volume 8, No. 1, 37-47.*

Billsus, D. & Pazzani, M. (1998). "Learning Collaborative Information Filters." *In Proceedings of the International Conference on Machine Learning, 46-54.*

Breese, J. S., Heckerman, D. & Kadie, C. (1998). "Empirical analysis of predictive algorithms for collaborative filtering." *In Proceedings of the Fourteenth Annual Conference on Uncertainty in Artificial Intelligence, 3–52.*

Burke, R. (2002). "Hybrid Recommender Systems: Survey and Experiments." *User Modeling and User-Adapted Interaction 12, Kluwer Academic Publishers, 331-370.*

Campos, Luis M. de, Fernández-Luna, J.M. & Huete, J.F. (2008). "A collaborative recommender system based on probabilistic inference from fuzzy observations." *Fuzzy Sets and Systems, 159, 1554-1576.*

Choi, S. H., Kang, S. & Jeon, J. Y. (2006). "Personalized recommendation system based on product specification values." *Expert Systems with Applications, 31, 607–616.*

Drachsler, H, Hummel, H. & Koper, R. (2007). "Recommendations for learners are different: Applying memory based recommender system techniques to lifelong learning." *In Proc. of the 1st workshop on Social Information Retrival for Technology- Enhanced Learning and Exchange SIRTEL 07, Crete, Greece.*

Dubois, D., Godo, L., Prade, H. & Zapico, A. *(1999).* "On the possiblistic decision model: from decision under uncertainty to case-based decision." *Int. J. Uncertainty Fuzziness Knowledge 7, 631–670.*

Goldberg, D., Nichols, D., Oki, B.M. & Terry, D. (1992). "Using collaborative filtering to weave an information tapestry." *Communications of the ACM, Volume 35, Issue 12, 61-70.*

Gřcar, M., Mladenic̆, D., Fortuna, B. & Grobelnik, M. (2006). "Data sparsity issues in the collaborative filtering framework." *Advances in Web Mining and Web Usage Analysis LNAI. Volume 4198, 58-76.*

Herlocker, J.L., Konstan, J.A., Borchers, AI. & Reidl, j. (1999). "An algorithm framework for performing collaborative filtering." *In Proc. of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, ACM press, 230-237.*

Jameson, A., Konstan, J. & Riedl, J. (2003). "AI techniques for personalized recommendation." *Tutorial SA1, Eighteenth International Conference on Artificial Intelligence.*

Kim, K. & Ahn, H. (2008). "A recommender system using GA K-means clustering in an online shopping market." *Expert Systems with Applications, Volume 34, Issue 2, 1200–1209.*

Khare R. & Rifkin A. (1997). "Weaving a web of trust." *World Wide Web Journal, 2(3):77-112.*

Koren, Y. (2008). "Tutorial on recent progress in collaborative filtering." *In Proceedings of the 2008 ACM Conference on Recommender systems (ACM Recsys'08), 333-334*

Krulwich, B. (1997). "Lifestyle Finder intelligent user profiling using large-scale demographic data." *AI magazine, Volume 18, no. 2, 37-46.*

Lang, K. (1995). "NewsWeeder: learning to filter netnews." *In Proc. of the 12th International Conferences on Access (Stanford, Calif).*

Lathia, N., Hailes, S. & Capra, L. (2007). "Private distributed collaborative filtering using estimated concordance measures". *In Proceedings of the 2007 ACM conference on Recommender systems, 1-8.*

Lee, S., Yang, J. & Park, S. (2004). "Discovery of hidden similarity on collaborative filtering to overcome sparsity problem." *Discovery Science, LNCS, Volume. 3245, 396-402.*

Linden, G., Smith, B. & York, J. (2003). "Amazon.com recommendations: item-to-item collaborative filtering." *Internet Computing IEEE, Volume.7, Issue 1, 76-80.*

Manouselis, N. & Costopoulou, C. (2007). "Analysis and classification of multi-criteria recommender systems." *World Wide Web: Internet and Web Information Systems, Volume 10 No. 4, 415-441.*

Massa, P. & Avesani, P. (2004). "Trust-aware collaborative filtering for recommender systems." *Lecture Notes in Computer Science, Volume 3290, 492-508.*

Mooney, R.J. & Roy, L. (2000). "Content- based book recommendation using learning for text categorization." *In Proceedings of the fifth ACM Conference on Digital Libraries, ACM press, 195-204.*

Pazzani, M. (1999). "A framework for collaborative, content-based, and demographic filtering." *Artificial Intelligence Review, Volume 13, Issue 5-6, 393-408.*

Pedrycz, w. & Gomide, F. (1998). "An introduction to Fuzzy Sets." *The MIT Press, Cambridge, MA.*

Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P. & Riedl, J. (1994). "Grouplens: an open architecture for collaborative filtering of netnews." *In Proceedings of ACM CSCW'94 Conference on Computer-Supported Cooperative Work, 175-186.*

Sarwar, B.M., Konstan, J.A., Borchers, A., Herlocker, J., Miller, B. & Riedl, J. (1998). "Using filtering agents to improve prediction quality in the grouplens research collaborative filtering system." *In Proceedings of ACM CSCW'98 Conference on Computer-Supported Cooperative Work, Social Filtering, Social Inuences, ACM Press, 345-354.*

Sarwar, B. M., Karypis, G., Konstan, J. & Riedl, J. (2000). "Analysis of recommender algorithms for e-commerce." *In Proceedings of the 2nd ACM E-Commerce Conference (EC'00), Minneapolis, MN, 158-167.*

Sarwar, B. M., Karypis, G., Konstan, J. A. & Riedl, J. (2001). "Item-based collaborative filtering recommender algorithms." *Accepted for publication at the WWW10 Conference.*

Schafer, J.B., Konstan, J.A. & Riedl, J. (2001). "E-commerce recommendation applications." *Data Mining and Knowledge Discovery, Springer Verlag, Volume 5, Issue 1-2, 115-153.*

Shardanand, U. & Maes, P. (1995). "Social information filtering: Algorithms for automating "word of mouth"." *In Proceedings of the ACM CHI'95 Conference on Human Factors in Computing Systems, 210-217.*

Shih, Y. Y. & Liu, D. R. (2008). "Product recommendation approaches: Collaborative filtering via customer lifetime value and customer demands." *Expert Systems with Applications, 35(1–2), 350–360.*

O'Sullivan, D., Wilson, D. & Smyth, B. (2002). "Using collaborative filtering data in Case-based recommendation." *In Proceedings of the Fifteenth International Florida Artificial Intelligence Research Society, AAAI, 121-125.*

Terveen, L. & Hill, W. (2001). "Beyond recommender systems: Helping people help each other." *In HCI in the New Millennium, J. Carroll, Ed. Addison-Wesley.*

Ungar, L. H. & Foster, D. P. (1998). "Clustering methods for collaborative filtering." *In Workshop on Recommender Systems at the 15th National Conference on Artificial Intelligence.*

Vozalis, E. & Margaritis, K.G. (2006). "Analysis of recommender systems' algorithms." *In Proceedings of the 2006 ACM symposium on Applied Computing, Dijon, France, 872-878.*

Vozalis, M.G. & Margaritis, K.G. (2007). "Using svd and demographic data for the enhancement of generalized collaborative filtering." *Information Sciences, Volume. 177, Issue 15, 3017-3037.*

Waddington, P. (1996). "Dying for information? A report on the effects of information overload in the UK and Worldwide." *Report, Reuters, United Kingdom.*

Wei, Y.Z., Moreau, L. & Jennings, N.R. (2005). "A market based approach to recommender systems." *ACM transactions on information systems, Volume 23, No. 3, 227-266.*

Yager, R.R. (1985). "Aggregating evidence using quantified statements." *Inform. Sci. 36 (1, 2) 179–206.*

Yager, R.R. & Kacprzyk, J. (1997). "The ordered weighted averaging operators: theory and applications." *Kluwer, Norwell, MA.*

Yager, R.R. (2000). "Targeted E-commerce marketing using fuzzy intelligent agents." *IEEE Intelligent System, 15 (6), 42–45.*

Yager, R.R. (2003). "Fuzzy logic methods in recommender systems." *Fuzzy Sets and Systems, 136, 133–149.*

Yukun Cao & Yunfeng Li. (2007). "*An intelligent* fuzzy-based *recommendation system for consumer electronic products." Source, Expert* Systems *with Applications*, *Volume 33, Issue 1, 230-240*

Zadeh, L. A. (1965). "Fuzzy sets." *Information Control, 8, 338–353.*

Zadeh, L.A. (1978). "Fuzzy sets as a basis for a theory of possibility." *Fuzzy Sets and System, 1 (1) 3–28.*

Zadeh, L.A. (1983). "A computational approach to fuzzy quantifiers in natural languages." *Computer. Math. Appl. 9, 149–184.*

Zenebe, A. and Norcio, A.F. (2009). "Representation, similarity measures and aggregation methods using fuzzy sets for content based recommender systems."*Fuzzy Set and Systems, 160, 76-94.*