

Survey Paper on Fuzzy Logic Based Recommender System

Rimpal Unadkat¹, Prof. Mehul Barot²

¹ Research Scholar, CE-LDRP-ITR, KSV University Gandhinagar, Gujarat, India

² Lecturer, CE-LDRP-ITR, KSV University Gandhinagar, Gujarat, India

ABSTRACT

With current projections regarding the growth of Internet sales, online retailing raises many questions about how to market on the Net. A Recommender System (RS) is a composition of software tools that provides valuable piece of advice for items or services chosen by a user. Recommender systems are currently useful in both the research and in the commercial areas. Recommender systems are a means of personalizing a site and a solution to the customer's information overload problem. Recommender Systems (RS) are software tools and techniques providing suggestions for items and/or services to be of use to a user. These systems are achieving widespread success in ecommerce applications now days, with the advent of internet. This paper presents a categorical review of the field of recommender systems and describes the state-of-the-art of the recommendation methods that are usually classified into four categories: Content based Collaborative, Demographic and Hybrid systems. To build our recommender system we will use fuzzy logic and Markov chain algorithm.

Keywords: - Recommender System, Information Filtering, Prediction, Classification, User based, Item base, Fuzzy Logic.

1. INTRODUCTION

Web discovery applications like Stumble Upon, Reddit, Digg, Dice (Google Toolbar) etc to name a few are becoming increasingly popular on the World Wide Web. Information on the Internet grows rapidly and users should be directed to high quality Websites those are relevant to their personal interests. However, there is no way to Judge these web pages. Displaying quality content to users based on ratings or past Search results are not adequate. There's a lacking of powerful automated process combining human opinions with machine learning of personal preference.

The goal of this project is to study recommendation engines and identify the shortcomings of traditional recommendation engines and to develop a web based recommendation engine by making use of user based collaborative filtering (CF) engine and combining context based results along with it using fuzzy logic and markov chain algorithm.

The system makes use of numerical ratings of similar items between the active user and other users of the system to assess the similarity between users' profiles to predict recommendations of unseen items to active user. The system makes use of Pearson's correlation to evaluate the similarity between users.

The results show that the system rests in its assumption that active users will always react constructively to items rated highly by similar users, shortage of ratings of some items, adapt quickly to change of user's interest, and

identification of potential features of an item which could be of interest to the user. The System would benefit those users who have to scroll through pages of results to find relevant content.

2. PROBLEM STATEMENT

While studying recommender system, there were some hints at the problems that these companies have to overcome to build an effective recommender system.

2.1 Lack of Data

Perhaps the biggest issue facing recommender systems is that they need a lot of data to effectively make recommendations. It's no coincidence that the companies most identified with having excellent recommendations are those with a lot of consumer user data: Google, Amazon, and Netflix. A good recommender system firstly needs item data (from a catalog or other form), then it must capture and analyze user data (behavioral events), and then the magic algorithm does its work. The more item and user data a recommender system has to work with, the stronger the chances of getting good recommendations.

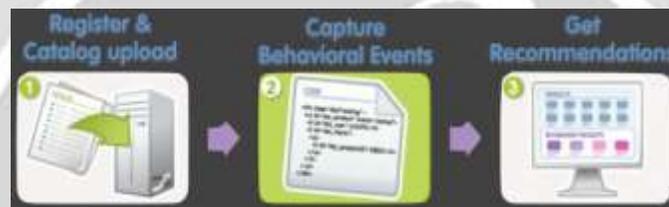


Fig. Data Gathering for Recommender System

2.2 Changing User Preferences

Again suggested by Paul Edmunds, the issue here is that while today I have a particular intention when browsing e.g. Amazon – tomorrow I might have a different intention. A classic example is that one day I will be browsing Amazon for new books for myself, but the next day I'll be on Amazon searching for a birthday present for my sister. On the topic of user preferences, recommender systems may also incorrectly label users.

2.3 This Stuff is Complex!

Below slide illustrates that it takes a lot of variables to do even the simplest recommendations. So far only a handful of companies have really gotten recommendations to a high level of user satisfaction – Amazon, Netflix (although of course they are looking for a 10% improvement on their algorithm), Google are some names that spring to mind. But for those select few success stories, there are hundreds of other websites and apps that are still struggling to find the magic formula for recommending new products or content to their users.

Social Recommender API	
Capturing behavior in a retail site	Item visited Recommended item visited Item reviewed Item added to shopping cart Item added to favorites Item added to wishlist Item purchased
Capturing behavior in a content site	Content visited Recommended content visited Content reviewed Content searched Content uploaded Content downloaded
Getting recommendations	Get item recommendation Get user recommendation Explain a recommendation
Session start/end	Session start Session end
Rating items	Rate an item Show item rating
Tagging users and items	Set user tags Get user tags Set item tags Get item tags

Fig: Variables needed for Recommendation

3. OBJECTIVE

Current recommender systems have a clear main objective: to guide the user to useful/interesting objects. It is very noticeable that this objective is composed of two different tasks:

1. To generate suggestions to be accepted by the user.
2. To filter useful/interesting objects. The first task has to do with the most external and interactive behaviour that any recommender directly reveals to the user. The second task is related to the known task “find good items”, with a more internal and less inter metrics published to date, it is difficult to identify these two tasks together, as a whole objective, on them. Moreover, a certain research bias towards the second part of the objective could be noticed, while frequently losing the first part.

4. SCOPE OF RECCOMENDER SYSTEM

Recommender System is the software engines and approaches for providing suggestion of products to the user which might be most probably matched to the user’s choice. Usually the recommender system is a technology which filters out the information to envision in case a particular user will like a specific item; this is usually called as prediction problem, or to identify N set of items that will be of certain users interest called as Top N recommendation problem. From past few years the use of recommender System is being gradually increasing in various different applications, for instance application for recommending books, CDs and other products at different search engines like amazon.com , Netflix.com, ebay.com and so on. Even the Microsoft suggests many additional software’s to user, to fix the bugs and so forth. When a user downloads some software, a list of software is provided

by the system. All the above examples would be result of diverse service, but all of them are categorized into a recommendation System, Identifying web-pages that will be of interest, or even implying backup ways of searching for information's.

5. LITERATURE SURVEY

In the last sixteen years, more than 200 research articles were published about research-paper recommender systems. I found that more than half of the recommendation approaches applied content-based filtering (55%). Collaborative filtering was applied by only 18% of the reviewed approaches, and graph-based recommendations by 16%.The use of efficient and accurate recommendation techniques is very important for a system that will provide good and useful recommendation to its users.

Serial No.	Year Of Pub & Journal Info	Paper Info	Technology Used	Conclusion
1.	IRACST - International Journal of Computer Science and Information Technology & Security (IICSITS), ISSN: 2249-9555 Vol. 2, No.5, October 2012	A Fuzzy Logic Based Personalized Recommender System	Fuzzy Near Compactness (FNC) concept	Personalized attribute-based recommender system as a solution to less frequently purchased products. The system also has the potential of increasing sales for online businesses.

Table 1- Literature survey-1

Serial No.	Year Of Pub & Journal Info	Paper Info	Technology Used	Conclusion
2.	February 9–12, 2011, Hong Kong, China. Copyright 2011 ACM 978-1-4503-0493-1/11/02	Recommender Systems with Social Regularization	Low-Rank Matrix Factorization	Design an effective algorithm to identify the most suitable group of friends for different recommendation tasks.
3.	IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 1, January 2011	FARS: Fuzzy Ant based Recommender System for Web Users	fuzzy recommender system based on collaborative behavior of ants (FARS)	By applying FCM and ant based clustering algorithms users are grouped in appropriate clusters.

Table 2- Literature survey-2

5. RECCOMENDATION SYSTEM

Recommendation system is an information filtering technique, which provides users with information, which he/she may be interested in.

5.1 Classification of Recommendation Systems:

Most of the recommendation systems can be classified into either User based collaborative filtering systems or Item based collaborative filtering systems. In user based collaborative filtering a social network of users sharing same rating patterns is created. Then the most similar user is selected and a recommendation is provided to the user based on an item rated by most similar user. In item based collaborative filtering relationship between different items is established then making use of the active user's data and the relationship between items a prediction is made for the active user (Machine, 2008).

5.2 Methodologies

The proposed system makes use of Pearson's correlation to implement User based collaborative filtering, and context, Synonym Finder to implement Context based filtering techniques to generate recommendations for the active user.

Following are the methodologies used/researched so far:

- **Taste:**

Taste is a flexible, fast collaborative filtering engine for Java. It takes the users' preferences for items and The engine takes users' preferences for items ("tastes") and recommends other similar items (Sean, 2008).

- **Vogoo:**

Vogoo is a php based collaborative filtering and recommendation library. It recommends items to users, which matches their tastes. It calculates similarities between users and creates communities based on them. The figure below shows the results of using vogoo to generate similar taste sharing users and recommendations made by the most similar users (Droux, 2008).

- **Fuzzy Logic:**

Here I tried to make use of fuzzy logic to calculate similar users.

Following is the currently used approach:

User Request:

User makes a request for recommendation by clicking on the recommendation menu. User is asked to provide contextual information.

Server:

The information provided by the user is send to the server. The server is composed on 2 sub engines: user based collaborative filtering engine, and context based engine. The server sends users request to both the sub engines.

User based collaborative filtering engine: - calculates similar users based on the numerical ratings of common items rated by the active users and other users of the system. The system achieves this by making user of the Pearson's correlation

Pearson's Correlation:

It is a way to find out similar users. The correlation is a way to represent data sets on graph. Pearson's correlation is x-y axis graph where we have a straight line known as the best fit as it comes as close to all the items on the chart as possible. If two users rated the books identically then this would result as a straight line (diagonal) and would pass through every books rated by the users. The resultant score in this case is 1. The more the users disagree from each other the lower their similarity score would be from 1. Pearson's Correlation helps correct grade inflation. Suppose a user 'A' tends to give high scores than user 'B' but both tend to like the book they rated. The correlation could still give perfect score if the differences between their scores are consistent.

Inaccurate queries: We have user typically domain specific knowledge. And users don't include all potential Synonyms and variations in the query, actually user have a problem but aren't sure how to phrase.

7. PROPOSED ARCHITECTURE

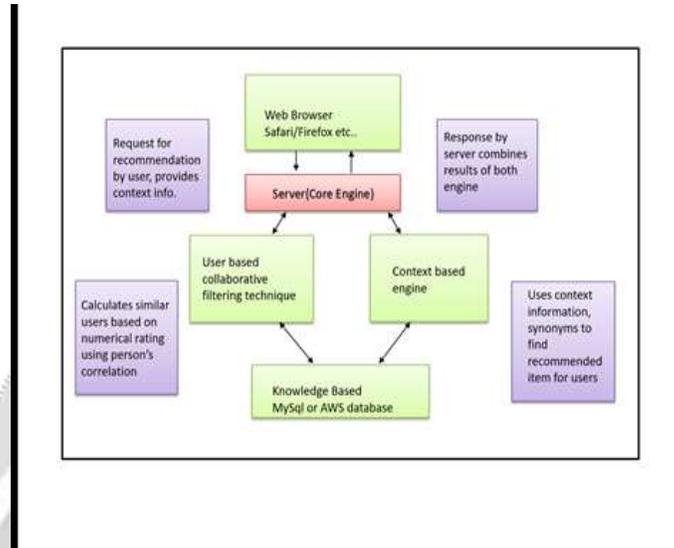


Fig. Proposed Architecture

Description:

1. User types in the URL for the system on a Web Browser.
2. User logs into the system using his `userid`.
3. The user chooses from amongst the type 2 different types of recommendation systems available.
4. If the user chose 'Collaborative Filtering' option, the system calculates similar users making use of engineering algorithms, and then recommends items to the users based on the most similar user.
5. If the user chose 'Context based Filtering' option, the system then makes use of the context information, and Synonym Finder to make predictions.

8. PROPOSED METHOD

Fuzzy Set and Fuzzy Logic

Fuzzy set theory consists of mathematical approaches that are flexible and well-suited to handle incomplete information, the un-sharpness of classes of objects or situations, or the gradualness of preference profiles. Fuzzy set theory and logic provide a way to quantify the uncertainty due to vagueness and imprecision. Membership functions, a building block of fuzzy sets, have possibilistic interpretation, which assumes the presence of a property and compares its strength in relation to other members of the set. A fuzzy set A in X is characterized by its membership, which is defined as: $(x) : x \in X \quad [0,1] \mu A$, where X is a domain space or universe of discourse. Alternatively, A can be characterized by a set of pairs:

$$A = \{(x, \mu_A(x)), X \}$$

According to the context in which X is used and the concept to be represented, the fuzzy membership function, $(x \in A)$, can have different interpretations. As a degree of similarity, it represents the proximity between different pieces of information. For example, movie x in the fuzzy set of "electronics" can be estimated by the degree of similarity.

As degree of preference, it represents the intensity of preference in favor of x , or the feasibility of selecting x as a value of X . For instance, a product rating of 4 out of 5 indicates the degree of a user's satisfaction or liking with x based on certain criteria.

Formalism of the Representation and Inference Methods

The proposed fuzzy theoretic content-based approach is based on a user's previous feedbacks, and features of the new items and features of the set of items for which the user has provided feedback. The rationale of this method is users are more likely to have interest in item like movie that is similar to the items like movies they have experienced and liked. This approach is useful for new item like movie with no or few user ratings and purchase. It is solely based on one user's previous interest expressed by ratings. The representation scheme, inference engine consisting of recommendation strategies and similarity measures, and the algorithm of the proposed method are presented in this section.

Items 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 not mutually exclusive possible values. For example, products are multi-categorized and multi-functional. These values of multi-valued attributes in an item can be represented more accurately with in a fuzzy set framework than with in a crisp set framework. Let an item I_j ($j = 1 \dots M$) be defined in the space of an attribute $X = \{x_1, x_2, x_3, \dots, x_L\}$, then I_j can take multiple values such as x_1, x_2, \dots , and x_L . If these values of X can be sorted in the decreasing order of their presence in the item I_j expressed by degrees of membership, then the membership function of item I_j to value x_k ($k = 1 \dots L$), denoted by $\mu(I_j, x_k)$, can be obtained heuristically. Hence, a vector formed for I_j :

$$X_j = \{(x_k, \mu(I_j, x_k)), k = 1 \dots L\}$$

$\mu(I_j, x_k)$ 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 .

Fuzzy Theoretic Similarity Measures

One of the most important issues in recommender systems research is computing similarity between users, and between objects (items, events, etc.). This in turns highly depends on the appropriateness and accuracy of the methods of representation. 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. Similarity is studied and applied in taxonomy, psychology and statistics. Similarity is subjective and context dependent. The set-theoretic, proximity-based and logic-based are the three classes of measures of similarity. Based on the results of the study those measures that are relevant for items recommendation application are adapted.

9. ANALYSIS

In below table essential parameters are discussed below along with their respective meanings and possible values i.e Yes: Y, No: N, Not Discussed: ND.

Serial No	Parameters	Meanings	Possible Values
1	<i>Efficiency</i>	A level of performance that describes a process that uses the lowest amount of inputs to create the greatest amount of outputs in minimum time & memory	Yes, No, Not Discussed

2	Accuracy in terms of prediction	Accuracy of algorithms should be maintained. Error rate should be minimized.	Yes, No, Not Discussed
3	User stratification	User specifications and needs must be satisfied	Yes, No, Not Discussed
4	Automatable	Methods describes are automatable which reduces manual work	Yes, No, Not Discussed
5	Robustness	Specification are may not be covered and but appropriate performance of a system	Yes, No, Not Discussed
6	Integration	Integration of system allows combination of 2 concepts such that system is capable of producing better results.	Yes, No, Not Discussed
7	Flexible	Flexibility refers to designs that can adapt when external changes occur.	Yes, No, Not Discussed
8	Performance	Backtracking or recovery process defines the performance of the system	Yes, No, Not Discussed
9	Satisfaction of the Recommendation Provider	User should be satisfied with recommended results. Relevant information should be provided in order to win out user satisfaction.	Yes, No, Not Discussed
10	Diversity	The concept of diversity encompasses acceptance and unique.	Yes, No, Not Discussed
11	Timing constraint	Appropriate timing is associated with every algorithm	Yes, No, Not Discussed
12	Effortless	The quality of a system that makes the user to use it easily	Yes, No, Not Discussed

13	Optimization	Optimization is the process of modifying a system to make some features of it work more efficiently or use fewer resources	Yes, No, Not Discussed
14	Error rate	It measures the total number of incorrect predictions against the total number of predictions	Yes, No, Not Discussed
15	Precision	It is defined where datasets are much unbalanced.	Yes, No, Not Discussed
16	Recall	It is the proportion of the number of data items that system selected as the positive	Yes, No, Not Discussed
17	F1-Score	For optimization F1 score combines both recall and precision with equal importance into a one parameter.	Yes, No, Not Discussed
18	Receiver Operating Characteristic (ROC) graph	It is technique to organize, visualize, and select classifiers that depend on their performance in 2D space.	Yes, No, Not Discussed

Table-3 Evaluation parameters for product recommendations

10. PRIOR AND RELATED WORK

As merchandisers gained the ability to record transaction data, they started collecting and analyzing data about consumer behavior. The term *data mining* is used to describe the collection of analysis techniques used to infer rules from or build models from large data sets. One of the best-known examples of data mining in commerce is the discovery of association rules—relationships between items that indicate a relationship between the purchase of one item and the purchase of another. These rules can help a merchandiser arrange products so that, for example, consumer purchasing ketchup sees relish nearby.

More sophisticated temporal data mining may suggest that a consumer who buys a new charcoal grill today is likely to buy a fire extinguisher in the next month. More generally, data mining has two phases. In the learning phase, the data mining system analyzes the data and builds a model of consumer behavior (e.g., association rules). This phase is often very time-consuming and may require the assistance of human analysts. After the model is built, the system enters a use phase where the model can be rapidly and easily applied to consumer situations. One of the challenges in implementing data mining within organizations is creating the organizational processes that successfully transfer

the knowledge from the learning phase into practice in the use phase. Automatic recommender systems are machine learning systems specialized to recommend products in commerce applications.

11. POPOSED ALGORITHM

Fuzzy Preference Tree-Based Recommendation Approach

In this algorithm it intends to cover both the user's intentionally expressed preference and their extensionally expressed preference from the user items. It form the structure based on two factors such as matching the corresponding the parts and rating by prediction on user targeted item using user preference aggregations. Next the two trees are mapped using conceptual similarities between the corresponding parts of two trees with fuzzy preference.

- Here the user's fuzzy preference tree is mentioned as and the item tree
- The maximum conceptual similarity tree mapping between and
- Then both trees are weighed equally and constructed by merging into
- The tree operation is done and the merging is defined by
- Next the function $pr()$, that takes the fuzzy preference tree and similarity tree mapping as input

Then it works as follows:

```

1   $mc \leftarrow MatchedChildren(t_u[j], M_{u,j})$ 
2  if  $v(t_u[j]) = null$  and  $mc = null$ 
3    return 0;
4  else if  $v(t_u[j]) \neq null$  and  $mc = null$ 
5    let the preference value be  $\tilde{p}_{uj} = \{f_{1,uj}, f_{2,uj}, \dots, f_{r,uj}\}$ 
6    return  $\sum_{k=1}^r k \cdot f_{k,uj}$ 
7  else if  $v(t_u[j]) = null$  and  $mc(t_u[j]) \neq null$ 
8    return  $\sum_{t_u[j_x] \in mc} w_x \cdot pr(t_u[j_x], M_{u,j})$ 
9  else if  $v(t_u[j]) \neq null$  and  $mc(t_u[j]) \neq null$ 
10   return  $\beta_j \cdot \sum_{k=1}^r k \cdot f_{k,uj}$ 
      +  $(1 - \beta_j) \cdot \sum_{t_u[j_x] \in mc} w_x \cdot pr(t_u[j_x], M_{u,j})$ 

```

Collaborative Filtering

The collaborative filtering approach originates in human behavior: people searching for an interesting item they know little of, such as a movie to rent at the video store, tend to rely on friends to recommend items they tried and liked. The person asking for advice is using a (small) community of friends that know her taste and can therefore make good predictions as to whether she will like a certain item. Over the net however, a larger community that can recommend items to our user is available, but the persons in this large community know little or nothing about each other. Conceptually, the goal of a collaborative filtering engine is to identify those users whose taste in items is

predictive of the taste of a certain person (usually called a neighborhood), and use their recommendations to construct a list of items interesting for her. To build a user's neighborhood, these methods rely on a database of past users interactions with the system. Early systems used explicit ratings. In such systems, users grade items (e.g., 5 stars to a great movie, 1 star to a horrible one) and then receive recommendations. Later systems shifted toward implicit ratings. A common approach assumes that people like what they buy. A binary grading method is used when a value of 1 is given to items the user has bought and 0 to other items. Many modern recommender systems successfully implement this approach. Claypool et al. (2001) have suggested the use of other implicit grading methods through a special web browser that keeps track of user behavior such as the time spent looking at the web page.

The Sequential Nature of the Recommendation Process

Most recommender systems work in a sequential manner: they suggest items to the user who can then accept one of the recommendations. At the next stage a new list of recommended items is calculated and presented to the user. This sequential nature of the recommendation process, where at each stage a new list is calculated based on the user's past ratings, will lead us naturally to our reformulation of the recommendation process as a sequential optimization process. There is yet another sequential aspect to the recommendation process. Namely, optimal recommendations may depend not only on previous items purchased, but also on the order in which those items are purchased. Zimdars et al. (2001) recognized this possible dependency and suggested the use of an auto-regressive model (a k -order Markov chain) to represent it. They divided a sequence of transactions X_1, \dots, X_T (for example, product purchases, web-page views) into cases $(X_{t-k}, \dots, X_{t-1}, X_t)$ for $t = 1, \dots, T$. They then built a model (in particular, a dependency network) to predict the last column given the other columns, under the assumption that the cases were exchangeable.

Factorizing Personalized Markov Chains (FPMC)

First, we introduce MC for sequential set data and extend this to personalized MCs. We discuss the weakness of Maximum Likelihood Estimates for the transition cubes. To solve this, we introduce factorized transition cubes where information among transitions is propagated. We conclude this section by combining both ideas into FPMCs.

12. COMPARATIVE ANALYSIS OF ALGORITHM

These algorithms show varied performance under different training sets. The dimensionality-reduction algorithm requires the highest runtime. The item-based algorithm works slowly for the large number of products. The spreading-activation and link-analysis algorithms require less number of iterations to achieve acceptable recommendation quality. The generative model algorithm is very efficient as it needs a small number of hidden classes for quality recommendations. The spreading-activation algorithm is especially fast. The link-analysis algorithm usually performs the best

Recommendation Strategy	Advantages	Disadvantages
Content Based Recommendation	User independence Transparency	Limited content analysis Over-specialization
Trust Based Recommendations	Avoid cold start problem Alleviate the sparsity problem.	Difficult to develop a trust network
Collaborative Filtering	Easy to create	It totally depends on human ratings.

Table 4 Comparative Analysis of Recommendation Strategies

[11] "[Fuzzy Logic](#)". *Stanford Encyclopedia of Philosophy*. Bryant University. 2006-07-23. Retrieved 2008-09-30.

[12] Zadeh, L.A. (1965). "Fuzzy sets". *Information and Control*. 8 (3): 338–353. doi:10.1016/s0019-9958(65)90241-x.

[13] P. Brusilovsky, A. Kobsa, and W. Nejdl (Eds.): *The Adaptive Web*, LNCS 4321, pp. 291 – 324, 2007. © Springer-Verlag Berlin Heidelberg 200

[14] Research-Paper Recommender Systems: A Literature Survey

[15] Bela Gipp, Jöran Beel, and Christian Hentschel. Scienstein: A Research Paper Recommender System. In

[16] Proceedings of the International Conference on Emerging Trends in Computing (ICETiC'09), pages 309–315,

[17] Virudhunagar (India), January 2009. Kamaraj College of Engineering and Technology India, IEEE

[18] https://en.wikipedia.org/wiki/Markov_chain

[19] WWW 2010, April 26–30, 2010, Raleigh, North Carolina, USA. ACM 978-1-60558-799-8/10/04

[20] https://en.wikipedia.org/wiki/Markov_chain

[21] <http://setosa.io/ev/markov-chains/>

[22] https://en.wikipedia.org/wiki/Fuzzy_logic

[23] <http://whatis.techtarget.com/definition/fuzzy-logic>

[24] <http://in.mathworks.com/help/fuzzy/what-is-fuzzy-logic.html?requestedDomain=in.mathworks.com>

BIOGRAPHIES

	<p>Zalak Prajapati is a Student in the Master of Computer Engineering Department. She is Pursuing Master of Computer Engineering (ME-CE) degree from LDRP-ITR,KSV University,Gandhinagar,Gujarat,India. Her research interests are Data Mining and Web Usage Mining, Artificial Intelligence, Information Retrieval.</p>
	<p>Prof. Mehul Barot is Lecturer in Computer Department in LDRP-ITR,KSV University,Gandhingar,Gujarat,India. His Research interest are Data mining,Computer Network, Artificial Intelligence, Information Retrieval.</p>