

# REVIEW AND ANALYSIS OF MACHINE LEARNING AND SOFT COMPUTING APPROACHES FOR USER MODELING

Madhuri Potey<sup>1</sup> and Dr. Pradeep K Sinha<sup>2</sup>

<sup>1</sup>Research Scholar, College of Engineering Pune, India

<sup>2</sup>Adjunct Professor, College of Engineering, Pune, India

## ABSTRACT

*The adequacy of user models depends mainly on the accuracy and precision of information that is retrieved to the user. The real challenge in user modelling studies is due to the inadequacy of data, improper use of techniques, noise within the data and imprecise nature of human behavior. For the best results of user modelling, one should choose an appropriate way to do it i.e. by selecting the best suitable approach for the desired domain. Machine learning and Soft computing Techniques have the ability to handle the uncertainty and are extensively being used for user modeling purpose. This paper reviews various approaches of user modeling and critically analyzes the machine learning and soft computing techniques that have successfully captured and formally modelled the human behavior.*

## KEYWORDS

*User Model, personalization, relevance, context modeling, user profiles, relevance feedback, behavior modeling adaptive user model.*

## 1. INTRODUCTION

User Modeling is the process of building an user model. A user model contains the systems assumptions about all aspects of the user which are deemed relevant for tailoring the dialog behavior of the system to the user [11]. Figure 1 depicts the process of user modeling, considering all the aspects discussed in this paper. The main goal of user modeling is customization and adaptation of systems to the user specific needs. User model is a set of information structures designed to represent one or more of the following elements: (1) Representation of assumptions about the goals, plans, preferences, tasks and abilities and the knowledge about one or more types of users; (2) Representation of relevant common characteristics of users pertaining to specific user subgroups (stereotypes); (3) The classification of a user in one or more of these subgroups; (4) The recording of user behavior; (5) The formation of assumptions about the user based on the interaction history and/or (6) The generalization of the interaction histories of many users into stereotypes.[12] The information a user model has is directly proportional to the personalization content and presentation that is to be personalized. A user model is created through a process of User Modeling (UM) in which unobservable information about a user is inferred from observable information from that user; for example, using the interactions with the system [11].

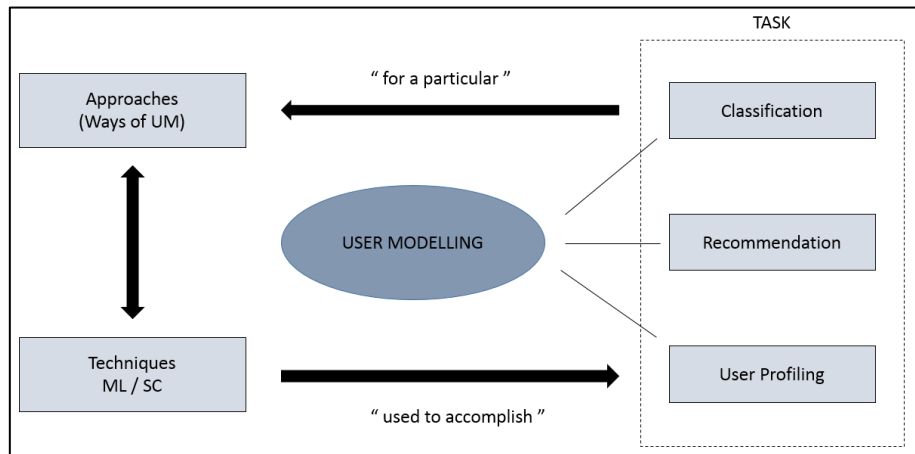


Figure 1: User Modeling

User models can be created using 1) Explicit approach where the user influences the model creation by explicitly stating interests, 2) Implicit approach where the user model is automatically created by capturing the user interests. The user-guided explicit approach produces adaptable services and adaptable user models, while the automatic (implicit) approach produces adaptive services and adaptive user models [13]. Main information that may need to be captured about users is - Personal information (name, address, telephone number, geographical data, age, gender, education, civil status etc), user’s knowledge about concepts, and relationships between concepts in an application domain, user abilities on how to perform the tasks in the domain, user interests and preferences, user goals and plans on the interaction with the system and usage data i.e. the way as a user interacts with the system.[14]

The organization of this paper is as follows: the paper first describes the concept of user modeling; second section discusses the role of domain modeling in various application domains. In third section the approaches of user modeling are discussed. In section four, the various techniques for user modeling are discussed.

## 2. USER MODELING FOR APPLICATION DOMAINS

A user model specifies features, needs, preferences and goals of end users. It can vary according to the domain or environment. Every domain model has a high level of abstraction that represents the formulation of a problem, knowledge or activity of the real world termed as Domain Model [14] as shown in figure 2.

Domain modeling is performed according to the following tasks also shown in figure 2:

*Goal modeling:* The general goal of the system can be identified by considering the problem that the system intends to solve, the refinement of which leads to specific goals. Product of task: Goal Model.

*Role modeling:* The responsibilities exercised by internal and external roles are identified for each specific goal. Then, the activities allowing the exercise of each responsibility are defined. During this refinement process, it can be identified that the same activity or a set of related activities are executed by several roles. In this case, it should be appropriate to create an independent role having the responsibility of executing these activities on behalf of the other

roles. The resources that an individual playing a role will need to execute his/her activities are also identified. Product of task: Role Model.

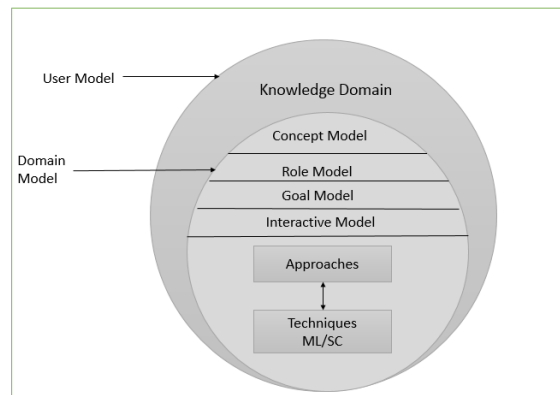


Figure 2: Domain Model

*Interaction modeling:* Through an analysis of their respective activities, the interactions between internal and external roles are identified. Product of task: Interaction Model.

*Concept modeling:* It is the process of identifying the domain, the information about domain, important concepts in the domain and relationships among these concepts. Existing software applications in the domain are also analyzed in order to identify commonalities and differences between them. Product of task: Conceptual Model.

User Modeling implements three types of basic tasks: Classification, Recommendation and User Profiling.

*Classification:* Classification is the process of building a model that maps or classifies data items into one of several predefined classes.

*Recommendation:* Recommendation is the capability of suggesting interesting elements to a user based on some extra information not based on the past behavior of the user; for example, from the items to be recommended or from the behavior of other users. There are two types of recommendation techniques viz. (a) Content-based (similarities between content) and (b) Collaborative (similarities between users).

*User Profiling:* A user profile is a collection of personal information[15]. The information is stored without adding further description or interpreting this information. It is comparable to a getting-setting mechanism of classes in object-oriented programming, where different parameters are set or retrieved. User profiles represent few characteristics of users such as cognitive skills, intellectual abilities, intentions, learning styles, preferences and interactions with the system. User behavior is the most intensively studied context information in the state-of-the-art literature. The user models are basically divided in four main domains, viz. (a)Information Retrieval and Filtering, (b) Intelligent Tutoring Systems, (c) Intelligent User Interfaces, (d) Adaptive Web Applications.

*Information retrieval and filtering* consists of user models that represent the users interests in a form of keywords, concepts or topics and provide relevant information to the user [8]. Some of the relevant work in information retrieval and filtering is: the user model that models user's interests by a multi-layer tree [10] with a structure that is dynamically changeable, the top layers

of which are used to model user interests on fixed categories, and the bottom layers are for dynamic events. Hochul Jeon *et. al.* [5] used Collaborative Filtering to propose an adaptive user profiling method using dynamic updating policy considering the change of the user preferences over time and domain. Xuehua Shen *et. al.* [7] proposed an approach to exploit implicit user modeling to intelligently personalize information retrieval and improve search accuracy. This approach is used to infer users interest from the users search context and use the inferred implicit user model for personalized search. Xuehua Shen *et. al.* developed an intelligent client-side web search agent (UCAIR- User-Centered Adaptive Information Retrieval) that can perform eager implicit feedback.

*Intelligent tutoring systems* consists of user models that select educational activities and deliver individual feedback that is most relevant to the user’s level of knowledge.

*Intelligent User Interfaces* are a special case of adaptive systems which instead of adapting contents of applications, adapt the actual interface through which the user interacts with the system. Intelligent interfaces can be - adaptive i.e the interfaces that automatically modifies the performance or adaptable i.e. user can modify the interface according to his/her preferences or a combination of both. Silvia Schiafino and Anala Amandi [1] discussed the main issues concerning user profiles: how a user profile is composed; how a user profile can be acquired; and how a user profile can be used. The Adaptive Web Applications are the user models that adapt the contents of a web page or provide navigation support to the user. Jae-wook Ahn *et. al.* [3] proposed a Web-based system for personalized news access to the user. Like many other adaptive news access systems, YourNews observes a user’s news reading behavior, constructs a user model (profile) representing user interests, and uses this model to recommend the most relevant news articles. Fabian Abel *et. al.* [2] developed a user modeling framework for Twitter and investigated how the different design alternatives influence the characteristics of the generated user profiles. David and William [4] reported Social Media Sources for Personality Profiling. They have discussed the causes that affect type and difficulty of pre-processing to extract appropriate text, the accuracy of training, the time period sampling for training texts with a special emphasis on Social Media Peculiarities and Implications for Personality Profiling.

### 3. USER MODELING APPROACHES

This section describes the ways of user modeling i.e. the approaches of UM which include Relevance feedback, Ontology-based, UM in Mobile Environments and Demographic. In particular, the chosen approaches focus on the role of user modeling for enhanced and personalized user support within interactive applications rather than providing an exhaustive review.

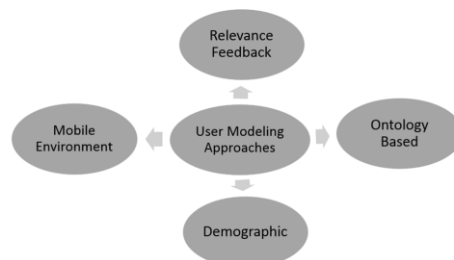


Figure 3: User Modeling Approaches

### 3.1. User Modeling based on Relevance Feedback

Relevance Feedback is the most popular query reformulation strategy where the user is presented with a list of the retrieved documents and marks that are relevant after examining them[16]. The general idea is to reformulate the original query vector  $Q_{org}$  into new query vector  $Q_{new}$  such that it gets closer to the term - weight vector of the relevant documents. For this purpose the Rocchio formulation is being widely used.

Gwizdka and Jacek [17] contributed their work in personalization for news stories relevance from user's perspective. They inferred relevance from eye-tracking data and observed if it is related to the cognitive effort involved in relevance judgments. Hassan *et. al.* [18] developed a Web browser add-in to monitor user search activity across search engines and collect explicit judgments of user satisfaction with the global search goal. Their model can also be used to optimize different search engine components. A method that uses task level success prediction to provide a better interpretation of click through data. Click through data has been widely used to improve relevance estimation. Their system was used to evaluate Search Results Quality and improve the relevance estimation. Jiang *et. al.* [19] proposed a novel method of Multimodal Pseudo Relevance Feedback (MMPRF) for event search in the video, which does not require search examples from the user. Their experiments proved that Pseudo Relevance Feedback has great potential in retrieval tasks, to tackle the event search task which is inherently multi modal. MMPRF technique takes the advantage of multiple modalities and multiple ranked lists to enhance event search performance in a principled way.

Table 1 : Examples of some user models based on relevance feedback

	Type of Feedback Supported	UM Representation	Dataset	Key Features
Gwizdka <i>et. al</i> [17]	Explicit	Statistical measures for relevance calculation	AQUAINT Corpus	<ul style="list-style-type: none"> <li>•Measure of statistical differences in reading patterns and in cognitive effort between documents</li> </ul>
Hassan <i>et. al</i> [18]	Explicit	Probabilistic models	Data from all major web search engines.	<ul style="list-style-type: none"> <li>•Evaluation of search results quality and improved relevance estimation</li> </ul>
Jiang <i>et. al</i> [19]	Explicit	MMRF	TREC VID Multimedia	<ul style="list-style-type: none"> <li>•Multi Modal Pseudo Relevance Feedback (MMPRF) for event search in video</li> </ul>
Bogdan <i>et. al</i> [20]	Implicit	multi-class SVM	Div400	<ul style="list-style-type: none"> <li>•hybrid approach for classification of users</li> </ul>

Boteanu *et. al.* [20] proposed a new hybrid approach that integrates both the automatization power of the machines and the intelligence of human observers via an optimized multi class Support Vector Machine (SVM) classifier-based relevance feedback. Their approach involved a classifier-based relevance feedback and consisted of two steps. The first step- an optimized multi-class Support Vector Machine (SVM) classifier-based relevance feedback with an objective to use user input to categorize the images in a number of distinct classes. The second step- the actual diversifier and consisted of an intra and inter-class image diversification strategy which was operated on the SVM class output confidence scores.

### 3.2. Ontology Based User Modeling

Ontology has been a basis for the construction of a user model [22] in several personalized systems ranging from information delivery systems to Intelligent Tutoring Systems [23]. In this section, we briefly discuss few important systems of this kind.

Razmerita et. al [24] proposed an ontology-based user modeling (OntobUMF) architecture. OntobUMF architecture consists of three layers namely- the user front layer, a middleware layer or a service layer and an ontology and a data layer. This ontology-based user modeling system integrates three ontologies: *User ontology* that includes different user's characteristics and their relationships. *Domain ontology* that captures the domain or application specific concepts and their relationships. *Log ontology* which represents the semantics of the user interaction with the system. In [25] the user profile is represented as hierarchy of concepts. The concepts are adopted from a reference ontology of around 4,400 concepts taking the top level categories from Magellan web site. myPlanet [26] is an ontology based personalized news delivery system. Simple relationships among the concepts inside the domain have been used to filter out information relevant to the user.

Table 2: Examples of some Ontology based User models

	Data Collection		User Mode Representation	Key Features
	Implicit/Explicit	Dataset		
Razmerita[24]	Explicit	Real time data	OntbUMF	<ul style="list-style-type: none"> <li>• Manage information about users</li> </ul>
Zhang et al. [28]	Implicit	Web Search logs	SWULPM	<ul style="list-style-type: none"> <li>• User model consists of personal ontology</li> <li>• Ontology represented through concept graph</li> </ul>
Yannis et. al [26]	Explicit	Web documents	YAGO general purpose ontology	<ul style="list-style-type: none"> <li>• Personalizes user's web searches</li> <li>• Improves quality of results</li> </ul>
Plumbaum et. al [29]	Implicit	Web documents	SWUM	<ul style="list-style-type: none"> <li>• Model fits to needs of the Social Web.</li> <li>• User Personal data form Social networks</li> </ul>

Zeng *et. al.* [27] proposed two different approaches for acquisition of users knowledge requirements about course content in an e-learning system. The ontology of course is represented as hierarchy of concepts. The first approach relies on interactive question-answer session and the historical session logs are analyzed to determine users requirements. The second approach is based on users reading behavior logs while reading e-documents. The actions considered by the user in this work are underline, highlight, circle, annotation and bookmark. Zhang et al. [28] proposed a system for constructing user models automatically by monitoring the users browsing behaviors in every session. The system tracks the usage logs by means of Semantic Web Usage Log Preparation Model (SWULPM). The user model consists of personal ontology which is represented through concept graph. Plumbaum *et. al.* [29] presented a new user model, the Social Web User Model (SWUM) that is fitted to the needs of the Social Web. Aroyo *et. al.* [30] provided a method for automatic acquisition of user knowledge through an ontology based dialog system. An ontology based dialog agent, called OWL-OLM, interacts with the user to analyze the current state of the user's knowledge according to the needs for a particular course task. Silvia Calegari and Gabriella Pasi [14] considered the problem to personalize user's web searches for improving the quality of results. They proposed a preliminary methodology that allowed to define a conceptual user profile based on the YAGO ontology.

### 3.3. UM in Mobile Environment

Lee *et. al.* [31] implemented the real-time activity recognition application on a Smartphone with the Google android platform. They attempted to recognize actions in real time on the Android platform, and recognize the user's activities from a recognized set of actions. In this, the uncertain time-series acceleration signal is analyzed by using hierarchical hidden markov models.

Table 3: Examples of some User models in Mobile environment

	Data Collection		UM Representation	Key Features
	Implicit/Explicit	Dataset		
Lee et. al.[30]	Explicit	Real time data	Hierarchical HMM and ANN	<ul style="list-style-type: none"> <li>•Integration of multimodal sensor data</li> <li>•Recognition of actions in real time data by BN</li> </ul>
Norman et. Al[31]	Implicit	LBE privacy guard	Personalized Classifiers	<ul style="list-style-type: none"> <li>•Analysis of people's privacy diverse preferences and customization</li> </ul>
Gauglitz et. al. [32]	Explicit	No standard dataset used	NCC based template matching	<ul style="list-style-type: none"> <li>•Improved Collaboration on tasks that involve physical environment</li> </ul>
Zhang et. al.[33]	Implicit	Mobile footprint of 22 users	Android App. Easytrack:	<ul style="list-style-type: none"> <li>•Mobile usage pattern captured from search logs</li> </ul>

The recognition models are designed as hierarchy of actions and activities. The structure consists of two steps of HMMs to analyze acceleration data and recognize a user's behavior. After acceleration data is collected from a three axis accelerometer on a Smartphone, it is transferred to low-level HMM to classify users' actions and then high-level HMM is used to recognize a user's activities from the set of actions. Continuous HMMs with Gaussian distribution are used to recognize a user's action from the acceleration data.

Norman *et.al.*[32] analyzed personnel's privacy preferences when it comes to granting permissions to different mobile apps. Their results are based on the analysis of settings of 4.8 million Smartphone users of a mobile security and privacy platform. They used LBE, a rooted version of the Android platform that has been in use by several million people and that has allowed its users to manually configure 12 particularly sensitive Android permissions. They differentiated between users who have passively accepted default settings selected by LBE on their behalf and those more active users who went through the trouble of modifying these settings. The LBE app organizes all API calls by permissions.

Gauglitz *et. al.*[33] described a framework and prototype implementation for unobtrusive mobile remote collaboration on tasks that involve the physical environment. The system used the Augmented Reality paradigm and marker less, model-free, visual tracking to facilitate decoupled, live updated views of the environment and world-stabilized annotations while supporting a moving camera and unknown, unprepared environments. They used a multi-level, active search patch tracker with normalized cross-correlation (NCC)-based template matching and key frame-based recovery.

Zhang *et. al.*[34] proposed an approach that used multimodal mobile sensor and log data to build a novel user modeling framework called mFingerprint that can effectively and uniquely depict users. They analyzed multimodal mobile usage data to extract simple yet effective statistics that can uniquely represent mobile users. They particularly focused on designing

frequency and entropy based statistical features to capture mobile device usage patterns. DBSCAN clustering algorithm was applied to the stationary segments in order to identify important locations.

### 3.4. Demographic User Modeling

Demographic user modeling is the process of creating user models based on the user's demographic attributes such as age, gender, race, etc. It is very important to understand the variations in the user's online behavior based on these attributes. Goel *et. al.* [35] surveyed who does what on web i.e. which demographics people are engaged in what activities. The relevant literature on user modeling, capturing search behavior and information retrieval based on demographic attributes are discussed here.

Bilal *et. al.* [36] investigated children's cognitive, effective, and physical behaviors as they use the Yahoo! Search engine to find information on a particular search task. The participants of this project were twenty-two seventh-grade science children from a middle school located in Knoxville, Tennessee. Their cognitive and physical behaviors were captured using Lotus ScreenCam (a Windows-based software package) that captures and replays activities recorded in Web browsers. Their affective states were captured via a one-on-one interview. A new measure called Web Traversal Measure was developed to measure children's weighted traversal effectiveness and efficiency scores, as well as their quality moves in Yahoo! Search. Children's prior experience in using the Internet/Web and their knowledge of the Yahoo! Search interface were gathered via a questionnaire.

Sergio Duarte Torres *et. al.* [37] employed toolbar logs from a commercial search engine to characterize the browsing behavior of young aged users, particularly for understanding the activities on the Internet that trigger search activity. The proportion of browsing and search activity in the toolbar sessions estimated the likelihood of a user to carry out search on the Web vertical and multimedia verticals (i.e. videos and images) given the previous event is another search event or a browsing event which was quantified. It was even observed that these metrics clearly demonstrate an increased level of confusion and unsuccessful search sessions among children. They also found a clear relation between the reading level of the clicked pages and characteristics of the users such as age and educational attainment.

Bernard J. Jansen *et. al.* [38] evaluated the effect of gender targeted advertising on the performance of sponsored search advertising. They used the Microsoft ad Center Labs Demographics Prediction Tool [74]. This application takes a given search phrase and provides the probability that the query is male or female-oriented (i.e., provides a probability for both within range of 0-1 inclusive).

J Hu. *et al.* [39] made a first attempt to predict users gender and age from their Web browsing behaviors, where the Webpage view information is treated as a hidden variable to propagate demographic information between different users.

Weber and Castillo [40] presented a query logs study on how search differs in users with different demographic attributes. To describe search patterns and behaviors they used demographic information that was derived from the US-census and user profile information for population segments with different demographic characteristics. They also employed an analogous methodology to show that the reading level of the urls clicked by children also varies across demographic features.



Table 4: Examples of some user modeling techniques based on demographic attributes

	Data Collection		UM Representation	Demographics captured and Key Features
	Implicit/Explicit	Dataset		
Yuxiao et. al[42]	Explicit	1 billion call & text-message events	Double Dependent-Variable Factor Graph Mode	<ul style="list-style-type: none"> <li>• Social Attributes</li> <li>• Inference of demographics in social network analysis.</li> </ul>
J. Hu. Et al. [39]	Implicit	Yahoo Search logs	SVM and SVD	<ul style="list-style-type: none"> <li>• Users age and gender prediction</li> </ul>
Liat Antwarg[41]	Implicit	Real Dataset website BGU	Attribute driven HMM	<ul style="list-style-type: none"> <li>• Online user behaviour using user attributes like Age, Gender.</li> </ul>

Liat Antwarg *et. al.*[41] introduced a novel approach to generate an intention prediction model of user interactions with system. As part of this new approach, they included personal aspects in the system, user characteristics, that can increase prediction accuracy. The model is automatically trained according to the fixed attributes of user(e.g., demographic data such as age and gender) and the users sequences of actions in the system.

Yuxiao *et. al.*[42] studied human interactions on demographics by investigating a country-wide mobile communication network. Considering this, they discovered a set of social strategies stemming from human communications.

The literature discussed so far focuses on the precise retrieval of results, demographics is considered for an appropriate user interface in use. Gossen *et. al.*[43] presented a new user interface called Knowledge Journey. It is a web search user interface for young users in the primary school age.

#### 4. KR Formalisms for user modeling

A user model consists mainly of knowledge about the individual preferences which determine the user's interaction behavior. Like in the most knowledge based approaches, two are the main issues that user modeling faces: The user model representation and acquisition. The process of generation of user models using machine learning or soft computing techniques can be seen as a standard process of extracting knowledge from data where UM is used as a wrapper for the entire process.

Table 5 gives the general characteristics of the two techniques presented along seven dimensions viz. Knowledge Representation (KR) which states how knowledge is represented in most of the cases, input given to the techniques, Uncertainty-the ability of the techniques to handle uncertainty in modeling human behavior i.e., to produce a user model that takes into account the inherent fuzziness of UM, Dynamic Modeling- which indicates the suitability of the technique to change a user model on-the-fly, the interpretability of the results- how easy it is for a human to understand the knowledge captured, (considered a critical dimension), the human aspect of techniques and the state of output.

Table 5: Characteristics of Machine Learning and Soft Computing based Models

	<b>KR</b>	<b>Input</b>	<b>Uncertainty Supported</b>	<b>Dynamic Modelling</b>	<b>Interpretability</b>	<b>Human Aspects</b>	<b>Output</b>
ML	Hierarchical representation/ Predicate logic	Observations and assumptions	Yes	Yes	Low	Behavior	Decision
SC	Predicate Logic	Assumptions	Yes	yes	High	Mentalist & Behavior	Decision & Assumption

The field of machine learning and soft computing is vast, versatile, and fascinating. It can be viewed as judicious mixture of computational intelligence and applied statistics. The machine learning methods are Supervised and Unsupervised. Supervised learning requires the training data to be pre classified. This means that each training item is assigned a unique label, signifying the class to which the item belongs and includes techniques like k-NN, SVM, Decision tree classifiers, probabilistic models and classifiers. The unsupervised learning methods do not require preclassification of the training examples and form clusters of examples, which share common characteristics. Prominent techniques are Association Rules, K- means clustering and Self Organizing Maps.

The following section presents how these learning techniques have been used for UM: which knowledge can be captured with each technique, examples of applications, and its limits and strengths.

#### 4.1. Machine Learning based User Models

Machine learning (ML) methods have been applied to user modeling, mainly for acquiring models of individual users interacting with an information system[44][45]. In situations like this, the use of the system by an individual is monitored and the collected data are used to construct the model of the user, i.e., his/her individual requirements. ML offers a suite of powerful techniques either for user model acquisition or for user community induction. ML techniques support complex decision making tasks and improve the prediction quality of a UM software [46].

The basic belief about machine learning techniques is that there is a process that explains the data we observe. Though we do not know the details of the process underlying the generation of data, we know that it is not completely random. Machine learning has two approaches, Content based and Collaborative.

##### 4.1.1. SVM

Support Vector Machines (SVMs) are pattern classifiers based on a novel statistical learning technique proposed by Vapnik[47]. Primarily SVM is a learning machine intended for binary classification problem. Following are few examples of SVM classifiers for this context. Tang *et. al.* [9] used SVM for identifying the homepage of a person, and to define features where whether the title of the page contains the person name and whether the URL address (partly) contains the person name. Gershman *et. al.* [48] contributed their work in personalization of news for the user where SVM identifies positive and negative examples of news. The experiments were performed using SVM\_pref and SVM\_Rank which are available in the open

source. Boteanu *et. al.* [20] proposed a multi-class SVM classification framework wherein individual binary SVM classifier is trained so that each classifier generates a confidence score for each of the output classes. The final fusion of these scores achieves the multi-class attribution.

#### **4.1.2. Bayesian Approach**

Interest has been growing steadily in the application of Bayesian representations and inference methods for modeling the goals, preferences, and needs of users. The Bayesian approach consists of Bayesian network and Bayesian classifier. A Bayesian network (BN) is a compact, expressive representation of uncertain relationships among variables of interest in a domain. A BN is a directed acyclic graph where nodes represent random variables and arcs represent probabilistic correlations between variables. The absence of edges in a BN denotes statements of independence. A BN also represents a particular probability distribution, the joint distribution over all the variables represented by nodes in the graph. This distribution is specified by a set of conditional probability tables (CPT). Each node has an associated CPT that specifies the probability of each possible state of the node given each possible combination of states of its parents.

Probabilistic model, the Bayesian hidden score (BHS) model can be used for dynamic pairwise learning[6]. Kuenzer *et. al.*[51] evaluated Dynamic Bayesian Networks for predicting the future user events. Hongning and Zhai *et. al.*[52] studied the problem of user modeling in the search log data and proposed a generative model, dpRank, within a non-parametric Bayesian framework. Tedesco *et. al.*[53] investigated the integration of distributed, fragmented user models by means of Bayesian Networks.

#### **4.1.3. Decision Tree**

Decision tree learning [55] is a method for approximating discrete-valued functions with disjunctive expressions. Decision tree learning is generally best suited to problems where instances are represented by attribute-value pairs and the target function has discrete output values. Classification rules are an alternative representation of the knowledge obtained from classification trees. They construct a profile of items belonging to a particular group according to their common attributes. In the context of UM, decision trees can be used to classify users and/or documents in order to use this information for personalization purposes. Decision trees can also handle noisy data and/or data with missing parameters, which makes them very useful for creating user models due to the noisy and imprecise nature of the data available. Classification trees are typically used to implement Classification tasks. In this case, the Classification trees are used to construct user models to personalize the user experience [56]. Due to their ability to group users with similar characteristics, Classification trees can also be used to implement recommendation tasks [57]. Classification rules are widely used to model user behavior as they provide a straight forward framework to represent knowledge. The readability of the output knowledge is a great advantage of this approach.

### **4.2. Soft Computing based User Models**

Machine learning techniques have some limitations for modeling human behavior, mainly the lack of any reference to the inherent uncertainty that human decision-making has. This problem can be partially solved with the introduction of Soft Computing (SC) for UM. SC is an innovative approach to building computationally intelligent systems that differs from conventional (hard) computing in that it is tolerant of imprecision, uncertainty and partial truth.

The guiding principle of soft computing is to exploit the tolerance for imprecision, uncertainty and partial truth to achieve tractability, robustness and low solution cost.

#### **4.2.1. Neural Networks**

A Neural Network (NN) is an information processing paradigm composed of a large number of highly-interconnected processing elements (neurons) working in unison to solve special problems. NNs are able to derive meaning from complicated and/or imprecise data and to extract patterns that are too complex to be noticed by other computational techniques. No initial knowledge about the problem that is going to be solved is required. These characteristics make NNs a powerful method to model human behavior and useful technique to create user models for hypermedia applications. NNs have been extensively used for user modeling, mainly for classification and recommendation in order to group together users with the same characteristics and create profiles. Lemoine and Piat *et. al.*[58] uses NN to classify user navigation paths. Self Organizing Maps (SOM) is a type of unsupervised NN that has also been extensively used for recommendation, because it transforms high dimensional data into a two dimensional grid, grouping elements with the same characteristics. Goren-Bar *et. al.*[59] use SOM to classify documents based on a subjectively predefined set of clusters in a specific domain. Roh *et. al.*[60] used SOM to create a recommendation system for movies and Changchien *et. al.*[61] used SOM to create a recommendation system for ecommerce.

#### **4.2.2. Evolutionary and Genetic Representations**

Genetic Algorithms (GAs) and Evolutionary Algorithms (EAs) [62] are search algorithms based on the mechanics of natural selection. They begin with a set of potential solutions called the population. Solutions from one population are taken and used to form a new population, which are closer to the optimum solution to the problem at hand. The idea behind this process is the survival of the fittest. In general, GAs and EAs have been used for Recommendation in the form of rules, which can capture user goals and preferences, because they perform a global search and cope better with attribute interaction than algorithms used in data mining, where the search is more local. Romero *et. al.* [63] proposed an approach for student modeling, and Lee *et. al.* [59] presented an approach for capturing users preferences for improvement of web searches. They have also been applied for filtering [65] and for classification, as in [66], which uses GAs to model bankruptcy prediction of companies. This approach is suitable for searching vast, complex, and multi modal problem spaces but may have some limitations with respect to its potential for dynamic modeling and its computational complexity.

#### **4.2.3. Fuzzy Logic**

Fuzzy Logic (FL) [67] [68] defines a framework in which the inherent ambiguity of real information can be captured, modeled and used to reason under uncertainty. A key concept in FL theory is the notion of the fuzzy set. A fuzzy set expresses the degree of membership of an element in that set. When compared to traditional binary or multi-valued logic, in which the degree of truth takes values from a discrete finite set, in fuzzy logic the degree of truth can take continuous values between [0,1]. This characteristic allows capturing the uncertainty inherent to real data. FL is not a machine learning technique; nevertheless, due to its ability to handle uncertainty, it is used in combination with other machine learning techniques in order to produce behavior models that are able to capture and manage the uncertainty of human behavior. Some examples of these combinations are Fuzzy Clustering, Fuzzy Association Rules, and Fuzzy Bayesian Networks. Another alternative is to capture user models with a machine learning technique (possibly with some kind of representation of uncertainty) and use FL inference to implement the personalization engine. A traditional FL inference system processes

knowledge in three steps: fuzzifies the input data, conducts fuzzy inference based on fuzzy information and defuzzifies the fuzzy decisions to produce the final outcome. FL in UM does not necessarily realize all of the three steps, but maybe only a subset of them.

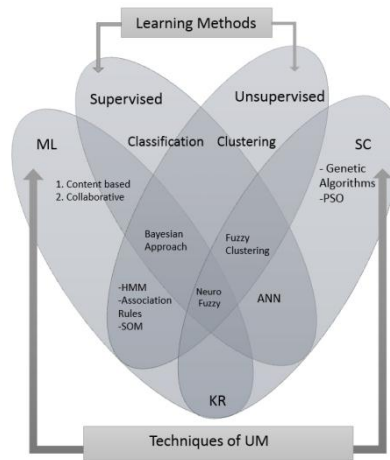


Figure 4: Techniques for User Modeling

Hybrid approaches have also been a keen research area for modeling human behavior. The most successful part of examples of applications reviewed already is used in modeling behavior considering their own advantages in the respective domain. Figure 4 shows the techniques that are commonly used for this purpose, and also states if they fall into which learning method. Hybrid approaches facilitates the model to deal with more uncertainties considering the complexity of an application. Neural networks and fuzzy logic [69], or neural networks and genetic algorithms[66], clustering and fuzzy logic[70][72] or genetic algorithms and rule extraction show great potential for UM. The combination of these SC techniques among themselves and with other machine learning techniques e.g. Bayesian approach with fuzzy logic [71] provide a useful framework to efficiently model the natural complexity of human behavior.

Table 6: Criteria of selection of Techniques

Task	Criteria	Technique
Classification	Two class classification	SVM [48,49]
	Data with missing parameters	Decision trees[55]
	Multi-class classifier	Bayesian classifier [20], SVM[54]
Recommendation	Users with the similar characteristics	Decision tree [56] NN[58]
	Uncertainties in user characteristics	Fuzzy approach [72]
User Profiling	Belief of users background and actions	Bayesian network[73]

## 5. CONCLUSIONS

Data deluge in recent past has caused the User to be more demanding for relevant and personalized services. Formal modeling of User is key for intelligent automated services that gave way to broader research interest in user models. This paper has presented a detailed review of existing approaches for user modeling. This paper also critically analyzed the machine learning and soft computing techniques that have tried to overcome the challenges in human behavior modeling due to various forms of uncertainty. Use models based on demographics, relevance feedback (implicit & explicit) are studied and approaches are compared. We also

reviewed user modeling for smart devices as user behavior is significantly different in mobile environment. We analyzed the significance of formalizing the user models with ontologies as it facilitates the automated user modeling which can also be adaptive with respect to time. Finally we discussed the machine learning and soft computing techniques along with how the knowledge about user is formalized and represented (using KR formalisms) and reasoned (classification, Recommendation).

## ACKNOWLEDGEMENTS

The authors are deeply indebted to all the researchers in the field and anonymous reviewers for their useful comments.

## REFERENCES

- [1] Silvia Schiafino and Analfia Amandi. Intelligent user profiling. In *Artificial Intelligence An International Perspective*, pages 193-216. Springer, 2009.
- [2] Fabian Abel, Qi Gao, Geert-Jan Houben, and Ke Tao. Analyzing user modelling on twitter for personalized news recommendations. In *User Modelling, Adaption and Personalization*, pages 1-12. Springer, 2011.
- [3] Jae-wook Ahn, Peter Brusilovsky, Jonathan Grady, Daqing He, and Sue Yeon Syn. Open user profiles for adaptive news systems: help or harm? In *Proceedings of the 16th international conference on World Wide Web*, pages 11-20. ACM, 2007.
- [4] David N Chin and William R Wright. Social media sources for personality profiling. In *Proceedings of the 2nd Workshop Emotions and Personality in Personalized Services(EMPIRE 2014)*, Aalborg, Denmark, pages 613-0073, 2014.
- [5] Hochul Jeon, Taehwan Kim, and Joongmin Choi. Personalized information retrieval by using adaptive user profiling and collaborative filtering. *AISS*, 2(4):134-142, 2010.
- [6] Daxin Jiang, Jian Pei, and Hang Li. Mining search and browse logs for web search: A survey. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 4(4):57, 2013.
- [7] Xuehua Shen, Bin Tan, and ChengXiang Zhai. Implicit user modelling for personalized search. In *Proceedings of the 14th ACM international conference on Information and knowledge management*, pages 824{831. ACM, 2005.
- [8] Kazunari Sugiyama, Kenji Hatano, Masatoshi Yoshikawa, and Shunsuke Uemura. Adaptive web search based on users implicit preference. *DEWS2004*, 2004.
- [9] Jie Tang, Limin Yao, Duo Zhang, and Jing Zhang. A combination approach to web user profiling. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 5(1):2, 2010.
- [10] Jue Wang, Zhiwei Li, Jinyi Yao, Zengqi Sun, Mingjing Li, and Wei-ying Ma. Adaptive user profile model and collaborative filtering for personalized news. In *Frontiers of WWW Research and Development-APWeb 2006*, pages 474-485. Springer, 2006.
- [11] Matthias Schneider-Hufschmidt, Uwe Malinowski, and Thomas Kuhme. *Adaptive user interfaces: Principles and practice*. Elsevier Science Inc., 1993.
- [12] Enrique Frias-Martinez, Sherry Y Chen, and Xiaohui Liu. Survey of data mining approaches to user modelling for adaptive hypermedia. *Systems, Man, and Cybernetics, PartC: Applications and Reviews*, IEEE Transactions on, 36(6):734-749, 2006.
- [13] Enrique Frias-Martinez, G Magoulas, S Chen, and R Macredie. Modelling human behavior in user-adaptive systems: *Recent advances using soft computing techniques*. *Expert Systems with Applications*, 29(2):320-329, 2005.
- [14] Rosario Girardi and C Faria. An ontology-based technique for the specification of domain and user models in multi-agent domain engineering. *CLEI electronic journal*, 7(1):7, 2004.
- [15] Nora Parcus de Koch. *Software engineering for adaptive hypermedia systems*. PhD thesis, PhD Thesis, Verlag Uni-Druck, Munich, 2001.
- [16] Joseph John Rocchio. Relevance feedback in *information retrieval*. 1971.
- [17] Jacek Gwizdka. News stories relevance effects on eye-movements. In *Proceedings of the Symposium on Eye Tracking Research and Applications*, pages 283-286. ACM, 2014.

- [18] Ahmed Hassan, Yang Song, and Li-wei He. A task level metric for measuring web search satisfaction and its application on improving relevance estimation. In *Proceedings of the 20th ACM international conference on Information and knowledge management*, pages 125-134. ACM, 2011.
- [19] Lu Jiang, Teruko Mitamura, Shou-I Yu, and Alexander G Hauptmann. Zero-example event search using multimodal pseudo relevance feedback. In *Proceedings of International Conference on Multimedia Retrieval*, page 297. ACM, 2014.
- [20] Bogdan Boteanu, Ionut Mironica, and Bogdan Ionescu. A relevance feedback perspective to image search result diversification. In *Intelligent Computer Communication and Processing (ICCP), 2014 IEEE International Conference on*, pages 47-54, 2014.
- [21] Heikki Keskustalo, Kalervo Jarvelin, and Ari Pirkola. The effects of relevance feedback quality and quantity in interactive relevance feedback: A simulation based on user modelling. In *Advances in Information Retrieval*, pages 191-204. Springer, 2006
- [22] Stuart E Middleton, Harith Alani, and David C De Roure. Exploiting synergy between ontologies and recommender systems. *arXiv preprint cs/0204012*, 2002.
- [23] Dirina Dicheva and Lora Aroyo. An approach to intelligent information handling in webbased learning environments. 2001.
- [24] Liana Razmerita, Albert Angehrn, and Alexander Maedche. Ontology-based user modelling for knowledge management systems. In *User Modelling 2003*, pages 213-217. Springer, 2003.
- [25] Alexander Pretschner and Susan Gauch. Ontology based personalized search. In *Tools with Artificial Intelligence, 1999. Proceedings. 11th IEEE International Conference on*, pages 391-398. IEEE, 1999.
- [26] Yannis Kalfoglou, John Domingue, Enrico Motta, Maria Vargas-Vera, and Simon Buckingham-Shum. Myplanet: an ontology-driven web-based personalised news service. 2001.
- [27] Qingtian Zeng, Zhongying Zhao, and Yongquan Liang. Course ontology-based users knowledge requirement acquisition from behaviors within e-learning systems. *Computers & Education*, 53(3):809-818, 2009.
- [28] Hui Zhang, Yu Song, and Han-Tao Song. Construction of ontology-based user model for web personalization. In *User Modelling 2007*, pages 67-76. Springer, 2007.
- [29] Till Plumbaum, Songxuan Wu, Ernesto William De Luca, and Sahin Albayrak. User modelling for the social semantic web. In *SPIM*, pages 78-89, 2011.
- [30] Lora Aroyo, Ronald Denaux, Vania Dimitrova, and Michael Pye. *Interactive ontology-based user knowledge acquisition: A case study*. Springer, 2006.
- [31] Young-Seol Lee and Sung-Bae Cho. Activity recognition using hierarchical hidden markov models on a smartphone with 3d accelerometer. In *Hybrid Artificial Intelligent Systems*, pages 460-467. Springer, 2011.
- [32] Bin Liu, Jialiu Lin, and Norman Sadeh. Reconciling mobile app privacy and usability on smartphones: could user privacy profiles help? In *Proceedings of the 23rd international conference on World wide web*, pages 201-212. International World Wide Web Conferences Steering Committee, 2014.
- [33] Steffen Gauglitz, Cha Lee, Matthew Turk, and Tobias Hollerer. Integrating the physical environment into mobile remote collaboration. In *Proceedings of the 14th international conference on Human-computer interaction with mobile devices and services*, pages 241-250. ACM, 2012.
- [34] Haipeng Zhang, Zhixian Yan, Jun Yang, Emmanuel Munguia Tapia, and David J Crandall. mfingerprint: Privacy-preserving user modelling with multimodal mobile device footprints. In *Social Computing, Behavioral Cultural Modelling and Prediction*, pages 195-203. Springer, 2014.
- [35] Sharad Goel, Jake M Hofman, and M Irmak Sirer. Who does what on the web: A largescale study of browsing behavior. In *ICWSM*, 2012.
- [36] Dania Bilal. Children's use of the yahoorigans! web search engine: I. cognitive, physical, and affective behaviors on fact-based search tasks. *Journal of the American Society for information Science*, 51(7):646-665, 2000.
- [37] Sergio Duarte Torres, Ingmar Weber, and Djoerd Hiemstra. Analysis of search and browsing behavior of young users on the web. *ACM Transactions on the Web (TWEB)*, 8(2):7, 2014.
- [38] Bernard J Jansen, Kathleen Moore, and Stephen Carman. Evaluating the performance of demographic targeting using gender in sponsored search. *Information Processing & Management*, 49(1):286-302, 2013

- [39] Jian Hu, Hua-Jun Zeng, Hua Li, Cheng Niu, and Zheng Chen. Demographic prediction based on user's browsing behavior. In *Proceedings of the 16th international conference on World Wide Web*, pages 151-160. ACM, 2007.
- [40] Ingmar Weber and Carlos Castillo. The demographics of web search. In *Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval*, pages 523-530. ACM, 2010.
- [41] Liat Antwarg, Lior Rokach, and Bracha Shapira. Attribute-driven hidden markov model trees for intention prediction. *Systems, Man, and Cybernetics, Part C: Applications and Reviews*, IEEE Transactions on, 42(6):1103-1119, 2012.
- [42] Yuxiao Dong, Yang Yang, Jie Tang, Yang Yang, and Nitesh V Chawla. Inferring user demographics and social strategies in mobile social networks. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 15-24. ACM, 2014.
- [43] Tatiana Gossen, Marcus Nitsche, and Andreas Nurnberger. Knowledge journey: A web search interface for young users. In *Proceedings of the Symposium on Human-Computer Interaction and Information Retrieval*, page 1. ACM, 2012.
- [44] Eric Bloedorn, Inderjeet Mani, and T Richard MacMillan. Machine learning of user profiles: Representational issues. *arXiv preprint cmp-lg/9712002*, 1997.
- [45] Bhavani Raskutti and Anthony Beitz. Acquiring user preferences for information filtering in interactive multi-media services. In *PRICAI'96: Topics in Artificial Intelligence*, pages 47-58. Springer, 1996
- [46] Christos Papatheodorou. Machine learning in user modelling. In *Machine Learning and Its Applications*, pages 286-294. Springer, 2001.
- [47] Simard P Scholkopf, V Vapnik, and AJ Smola. Improving the accuracy and speed of support vector machines. *Advances in neural information processing systems*, 9:375-381, 1997.
- [48] Anatole Gershman, Travis Wolfe, Eugene Fink, and Jaime Carbonell. News personalization using support vector machines.
- [49] S Bidel, L Lemoine, F Piat, T Artieres, and Patrick Gallinari. Statistical machine learning for tracking hypermedia user behavior. In *2nd Workshop on Machine Learning, Information Retrieval and User Modelling*, 9th Int. Conf. in UM, 2003.
- [50] Zheng Chen, Fan Lin, Huan Liu, Yin Liu, Wei-Ying Ma, and Liu Wenyan. User intention modelling in web applications using data mining. *World Wide Web*, 5(3):181-191, 2002.
- [51] Alexander Kuenzer, Christopher Schlick, Frank Ohmann, Ludger Schmidt, Holger Luczak, et al. An empirical study of dynamic bayesian networks for user modelling. In *Proc. of the UM2001 Workshop on Machine Learning for User Modelling*, pages 1-10, 2001.
- [52] Hongning Wang, ChengXiang Zhai, Feng Liang, Anlei Dong, and Yi Chang. User modelling in search logs via a nonparametric bayesian approach. In *Proceedings of the 7th ACM international conference on Web search and data mining*, pages 203-212. ACM, 2014.
- [53] Roberto Tedesco, Peter Dolog, Wolfgang Nejdl, and Heidrun Allert. Distributed Bayesian networks for user modelling. In *World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education*, volume 2006, pages 292-299, 2006.
- [54] Ingo Schwab, Wolfgang Pohl, and Ivan Koychev. Learning to recommend from positive evidence. In *Proceedings of the 5th international conference on Intelligent user interfaces*, pages 241-247. ACM, 2000.
- [55] P Winston. Learning by building identification trees. *Artificial intelligence*, 423, 1992.
- [56] Joseph E Beck, Peng Jia, June Sison, and Jack Mostow. Predicting student help-request behavior in an intelligent tutor for reading. In *User Modelling 2003*, pages 303-312. Springer, 2003
- [57] Georgios Paliouras, Vangelis Karkaletsis, Christos Papatheodorou, and Constantine D Spyropoulos. *Exploiting learning techniques for the acquisition of user stereotypes and communities*. Springer, 1999.
- [58] S Bidel, L Lemoine, F Piat, T Artieres, and Patrick Gallinari. Statistical machine learning for tracking hypermedia user behavior. In *2nd Workshop on Machine Learning, Information Retrieval and User Modelling*, 9th Int. Conf. in UM, 2003.
- [59] Dina Goren-Bar, Tsvi Kuik, Dror Lev, and Peretz Shoval. Automating personal categorization using artificial neural networks. In *User Modelling 2001*, pages 188-198. Springer, 2001.



- [60] Tae Hyup Roh, Kyong Joo Oh, and Ingoo Han. The collaborative filtering recommendation based on som cluster-indexing cbr. *Expert Systems with Applications*, 25(3):413-423, 2003
- [61] S Changchien and Tzu-Chuen Lu. Mining association rules procedure to support on-line recommendation by customers and products fragmentation. *Expert systems with applications*, 20(4):325-335, 2001.
- [62] Hans-Paul Paul Schwefel. *Evolution and optimum seeking: the sixth generation*. John Wiley & Sons, Inc., 1993.
- [63] Cristobal Romero, Sebastian Ventura, Paul De Bra, and Carlos De Castro. Discovering prediction rules in aha! courses. In *User Modelling 2003*, pages 25-34. Springer, 2003.
- [64] Wei-Po Lee and Tsung-Che Tsai. An interactive agent-based system for concept-based web search. *Expert systems with Applications*, 24(4):365-373, 2003.
- [65] Weiguo Fan, Michael D Gordon, and Praveen Pathak. Personalization of search engine services for effective retrieval and knowledge management. In *Proceedings of the twenty first international conference on Information systems*, pages 20-34. Association for Information Systems, 2000.'
- [66] Kyung-Shik Shin and Yong-Joo Lee. A genetic algorithm application in bankruptcy prediction modelling. *Expert Systems with Applications*, 23(3):321-328, 2002.
- [67] KLIR GEORGE J and Yuan Bo. *Fuzzy sets and fuzzy logic, theory and applications*. - 2008.
- [68] Jun Yan. *Using fuzzy logic: towards intelligent systems*. 1994.
- [69] S.K. Pal, V. Talwar, and P. Mitra. Web mining in soft computing framework: relevance, state of the art and future directions. *Neural Networks, IEEE Transactions on*, 13(5):1163- 1177, Sep 2002.
- [70] Raghu Krishnapuram, Anupam Joshi, Olfa Nasraoui, and Liyu Yi. Low-complexity fuzzy relational clustering algorithms for web mining. *Fuzzy Systems, IEEE Transactions on*, 9(4):595-607, 2001.
- [71] Han-Saem Park, Ji-Oh Yoo, and Sung-Bae Cho. A context-aware music recommendation system using fuzzy bayesian networks with utility theory. In *Fuzzy systems and knowledge discovery*, pages 970-979. Springer, 2006.
- [72] Maria N Moreno, Joel P Lucas, and Vivian F Lopez. Fuzzy data-mining hybrid methods for recommender systems. In *Soft Computing for Business Intelligence*, pages 327-344. Springer, 2014.
- [73] Eric Horvitz, Jack Breese, David Heckerman, David Hovel, and Koos Rommelse. The lumiere project: Bayesian user modelling for inferring the goals and needs of software users. In *Proceedings of the Fourteenth conference on Uncertainty in artificial intelligence*, pages 256-265. Morgan Kaufmann Publishers Inc., 1998.
- [74] <http://adlab.microsoft.com/DemographicsPrediction/DPUI.aspx>

## Authors

Madhuri Potey is pursuing PhD at University of Pune in Government College of Engineering, Pune, India. She acquired BE(1993) and MBA(1996) from Amaravati University, India and ME(2006)from COEP, Pune, India. She is also a faculty at D Y Patil College of Engg., Akurdi, Pune, India. Her research interests are Information Retrieval, Advanced Databases, Cryptography and Software Engineering. She is ACM Professional member since 2009.



Dr. P.K. Sinha is Senior Director(High Performance Computing) in Pune, India and Adjunct Professor at College of Engg. Pune, India. He is ACM Distinguished Engineer(2009) and IEEE Fellow(2013) with several research papers to his credit. He has contributed significant books including “Distributed Computing” and “Electronic Health Record: Standards, Coding Systems, Frameworks and Infrastructures”. He is actively involved in academic excellence at College of Engineering, University of Pune and mentoring several Research Projects at C-DAC, India.

