

# A CATEGORICAL REVIEW OF RECOMMENDER SYSTEMS

RVVSV Prasad<sup>1</sup> and V Valli Kumari<sup>2</sup>

<sup>1</sup>Department of Computer Science & Engineering, Bhimavaram Institute of Engineering & Technology, Bhimavaram – 534243, India

ramayanam.prasad@gmail.com

<sup>2</sup>Department. Of Computer Science & Systems Engineering, AU College of Engineering, Andhra University, Vishakapatnam – 533003, India

vallikumari@gmail.com

## ABSTRACT

*As more and more information became available electronically, the need for effective information retrieval and implementation of filtering tools have become essential for easy access of relevant information. 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 e-commerce applications now a 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. This paper discusses the pro's and con's of the current categories as well as the trustworthiness of the recommender system in a new dimension as evaluating the evaluator for more appropriate recommendations. In the domain of recommender system, this work also put forward the use of agents as an enabling technology.*

## KEYWORDS

Collaboration, Demography, Information Filtering, Recommender System, Agent, Trust

## 1. INTRODUCTION

The advances in information technology (IT) and entertainment technologies have accelerated the availability of various alternative items in each and every domain in the online market like hundreds of books, songs etc. In addition to this, the emergence of World Wide Web has opened up new possibilities for the users to know the details/specifications of items seamlessly without visiting shops and outlets. It is an easy task for an individual to choose from limited number of available alternatives. When the collection becomes large, it is a tedious and time consuming task for any individual to really evaluate the features of items/products while purchasing quality, economic, and useful items. In such circumstances, people seek suggestions or recommendation from friends and experts who have knowledge about the items/products. The main purpose of the recommender systems is to provide tools to leverage the information hunting and gathering activities and interests of other people or groups of people. Recommendation systems have been an important application area and the focus of considerable recent academic and commercial interest. Generally, recommender systems are used online to suggest items that customers find interesting, thereby, benefiting both customer and merchant.

As more and more information became available electronically, the need for effective information retrieval and filtering tools became essential for easy access to relevant information. Explosive growth of information has resulted in the problem of information overload, the information available is greater than can be managed effectively, and creates a degree of stress for which the coping strategies are ineffective. Technological aids such as information retrieval systems, information filtering systems, intelligent agents, ranking algorithms, clustering

techniques, categorization techniques, data mining techniques, web mining techniques, personalization and recommender systems to tackle the problem of information overhead.

As e-commerce Web sites began to develop, a pressing need emerged for providing recommendations derived from filtering the whole range of available alternatives. Users were finding it very difficult to arrive at the most appropriate choices from the immense variety of items (products and services) that these Web sites were offering.

Recommender Systems have proved in recent years to be a valuable means for coping with the information overload problem. Ultimately a RS addresses this phenomenon by pointing a user towards new, not-yet-experienced items that may be relevant to the user's current task.

### 1.1 Recommendation Process

In general, every recommendation system follows a specific process [1] in order to produce product recommendations, see figure 1. The recommendation approaches can be classified based on the information sources they use. Three possible sources of information can be identified as input for the recommendation process. The available sources are the user data (demographics), the item data (keywords, genres) and the user-item ratings (obtained by transaction data, explicit ratings).

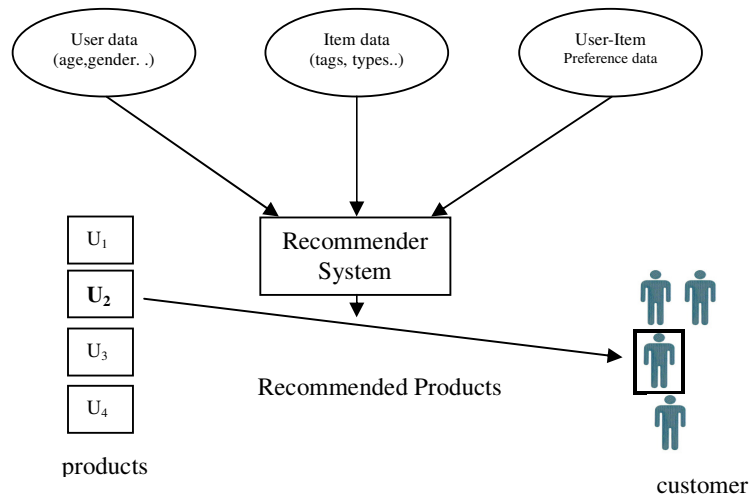


Figure 1. Recommendation Process

## 2. APPROACHES OF RECOMMENDER SYSTEMS

Recommender Systems (RS) are primarily directed towards individuals who lack sufficient personal experience or competence to evaluate the potentially overwhelming number of alternative items that a Web site, for example, may offer [3]. One of the popular web sites like Amazon.com employs a RS to personalize the online store for each customer [2]. Since recommendations are usually personalized, different users or user groups receive diverse suggestions. Personalized recommendations are offered as ranked lists of items. In this computational task RSs try to predict what are the most suitable products or services, based on the user's preferences and constraints. In addition there are also non-personalized recommendations which are much simpler to generate and are normally featured in magazines or newspapers. Typical examples include the top ten selections of books, songs etc.

Recommender systems can be used by various applications like E-commerce can be categorized in two ways:

- (i) Personalized recommender system
- (ii) Non-personalized recommender system.

Personalized recommender systems are used by E-commerce sites to suggest products to their customers. The products can be recommended based on the top sellers of a site, **demographics of the customer**, or analysis of the past buying behaviour of the customer as a prediction for future buying behaviour, for example eBay[4]. These techniques help the sites spread overall the World Wide Web to adapt itself to each customer requirements thus enabling individual personalization for each customer.

Non-personalized recommender systems recommend products to customers based on what other customers have said about the products on average. The recommendations are independent of the customer, so each customer gets the same recommendations. Non-personalized recommender systems are automatic, because they require little customer effort to generate the recommendations and are momentary. These recommendations are completely independent of the particular customer targeted by the recommender system. For example, Amazon.com and Moviefinder.com web sites are treated as non-personalized recommender systems.

### 3. FUNCTIONAL CLASSIFICATION OF RECOMMENDER SYSTEMS

Recommender systems can be classified broadly into several categories depending on the information they use to recommend items. However the following four categories are mostly followed in the literature:

- (i) **Content-based Filtering** Systems: Use **data about the items** and **information regarding the active user**.
- (ii) **Collaborative Filtering System**: Make recommendations to the active user using **information about a set of users and their relation with the item**.
- (iii) **Demographic Filtering Systems**: Use demographic information such as **age, gender, education**, etc. of people for **identifying types of users**.
- (iv) Hybrid Recommender Systems: With mixed approaches, attempting to keep the advantages of the combination of methods, and to reduce or take out disadvantages and problems. These systems are called **"Hybrid Recommender Systems"**, which are the most interesting and can be considered as the state-of-the-art systems.

This paper discusses four types of typical recommender systems based on their functionality. However, the majority of product recommender systems are developed using **content-based, collaborative-based** or **hybrid recommender systems**. By putting trustworthiness as a forward feature to **evaluate the credibility of the recommender systems**, social recommender systems have evolved and are addressed as **trust based recommender systems**. In the domain of recommender systems, this paper especially deals with **the use of agents** as an enabling technology.

#### 3.1 Content-based Filtering Recommender Systems

Content-based filtering systems recommend items based on **descriptions or content of items** rather than other **user's ratings** of the system as shown in figure 2. Instead of deriving a **user-to-item correlation** and defining methodologies, they use **item-to-item correlation for generating** recommendations. In these systems, the process of recommendation first starts by gathering content data about the items. For example, **title, author**, cost etc. for the books, are the some of the common content information. Most of these systems use **feature extraction** techniques and **information indexing** to extract the content data.

In content-based filtering, the system processes information from various sources and tries to extract useful features and elements about its content. Websites like BargainFinder and Jango try to collect information from many different web information sources. Different sources have different inputs (e.g., CGI scripts, Java applets) and presentation methods, so recommender systems have to adjust their interaction methods depending upon the web site.

As in content-based approaches, constraint-based filtering **uses features of items to determine their relevance**. However, unlike most feature-based techniques which access data in their native formats, constraint-based techniques require that the **problem and solution space be formulated in terms of variables, domains, and constraints**. Many problems can be formulated as a **Constraint Satisfaction Problem (CSP)** technique such as scheduling, planning, configuration, and machine vision problems. In PersonaLogic, CSP techniques are used in product brokering stage in Customer Buying Behavior (CBB) to evaluate product alternatives.

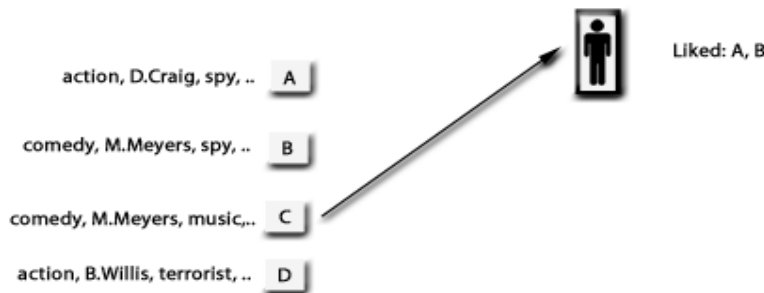


Figure 2 - Content-based approach

Advantages:

Content based approach does not require data of other users and has capability of recommending item to **users with unique taste** and **does not suffer from first rater problem** i.e., they are capable of recommending new item and unpopular item to every customer.

Disadvantages:

The feature extraction and representation can be achieved automatically e.g. news or papers. But human editors have to **manually insert the features for items** e.g. movies and songs. Content-based filtering techniques have no other method for finding something unexpected issues which are useful while searching. In content-based filtering systems, **items are limited to their initial descriptions or features** i.e. the features are specified explicitly.

### 3.2 Collaborative filtering Recommender Systems

The goal of collaborative filtering systems is to suggest new items or predict the utility of a certain item for a particular user based on **users past liking** and the **opinions of other like-minded users**, see figure 3.

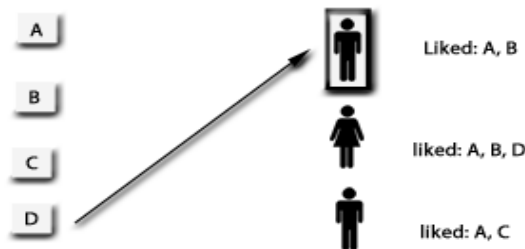


Figure 3 - Collaborative filtering approach

It is widely used and perhaps the most familiar recommendation technique implemented in several e-Commerce applications. In general, collaborative filtering systems collect the ratings for a list of items from a list of users. Opinions can be explicitly given by the user as a rating score or can be implicitly derived from the historical data of the user.

The collaborative filtering can adapted with Neighborhood methods, see figures 3(a) and 3(b), whose focus on relationships between items or, alternatively, between users. They are:

- (i) **User-based collaborative Filtering:** For each user, compute correlation with others users. For each item, aggregate the ratings of the users highly correlated with each user.

Problem: sparsity of data, bad correlation, easy to attack.



Figure 3(a) – User based Collaborative filtering approach

- (ii) **Item-based collaborative Filtering:** For each item, compute correlation with others items. For each user, aggregate her ratings of the items highly correlated with each item.

For items, sparsity of data is less important, less problems with cold start and attacks



Figure 3(b) – Item based Collaborative filtering approach

Advantages:

Collaboration Filtering Approach does not need a representation of items in terms of features but it is based only on the judgment of participating user community. Example websites, movies, songs, books, jokes, etc. Scalability of the items database is large because it does not require any human involvement. They can have cross type recommendations for predictions which are different to users and does not require any domain knowledge which saves time. The recommendations can be improved over a period of time.

Disadvantages:

The item cannot be recommended to any user until the item is either rated by another user(s) or correlated with other similar items of the database. In practice, in many e-commerce applications the active users rate only few items in spite of having a large item database which leads to very sparse results.

Due to diversity of opinions of large number of participants make this approach expensive and time consuming. It is common those individuals do not agree or disagree with other individuals. It recommends only popular items but cannot recommend unique tastes of a user.

Qing Li and Byeong Man Kim [5] proposed a collaborative recommender framework, in which clustering technique is applied, to consider semantic contents available from the user profiles. For this, they suggest methods to construct user profiles from rating information and attributes of items to accommodate user preferences. They explored the feature-based information of users from the user profiles and the atomic-item based rating information from the ratings and integrates the information of user profiles and user ratings to calculate the user-user similarity.

Zied Zaier et al. [6] introduce a long tail theory and its impact on recommender system. They provide a comprehensive review of the different datasets used to evaluate collaborative filtering recommender systems techniques and algorithms. They also compare the performance of different recommendation techniques based on different dataset distributions. Their study investigates a single-criterion rating datasets like most of the current collaborative filtering recommender systems. Multi criteria rating are considered as an added-value to current recommender system. Due to the nature of the information involved in recommender systems, they tried to develop anonymity and encryption techniques to protect individual data privacy in distributed recommender system.

Panagiotis Symeonidis et al. [7] construct a feature-weighted user profile to disclose the duality between users and features. Exploiting the correlation between users and features they reveal the actual reasons of their rating behavior. Based on Information Retrieval, they include the Term Frequency Inverse Document Frequency (TFIDF) weighting scheme in collaborative filtering. They also proposed the top-N list generation algorithm that counts the frequency of each positively rated item inside the found neighborhood by exploiting the item features. It takes into account the fact that, each user has own reasons for rating an item.

### 3.3 Demographic Filtering Systems

A general technique people use to build models of other people very quickly is the evocation of stereotypes or clusters of characteristics. A typecast is a collection of frequently occurring characteristics and knowledge of users or user groups. When a new user enters the system, he will be assigned into related typecast(s) if some of his characteristics match that of a particular typecast(s). The same concept is used in demographic recommender systems by the use of descriptions of people to learn the relationships between a single item and the class or type of people who liked it.

Demographic based recommender systems use prior knowledge on demographic information about the users and their opinions for the recommended items as basis for recommendations, see figure 4. Demographic systems are stereotypical, because they depend on the assumption that all users belonging to a certain demographic group have similar taste or preference.

Pazzani [8] uses machine learning techniques to obtain a classifier based on demographic data. The representation of demographic information in a user model can vary greatly. Grundy system, the first system evolved in its kind uses hand-crafted attributes with numeric confidence values, while Pazzani extracts features from users' home pages.

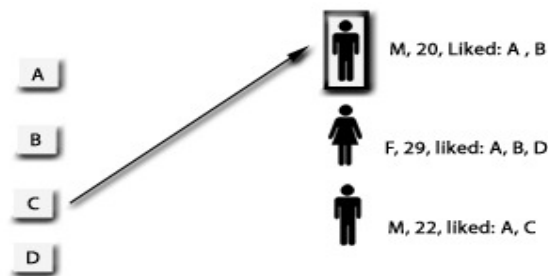


Figure 4 - Demographic-based recommendation based on popularity

**Advantages:**

The advantage of a demographic approach is that it does not require history of user ratings that are required by collaborative and content based techniques. This approach is quick, easy and straight forward for making results based on few observations.

**Disadvantages:**

Concerning the security and privacy issues, collection of complete user information is impractical for effective demographic filtering. Demographic clustering is mainly based on users interests are general which leads this system recommend the same item to users of same demographic profiles and results in giving recommendations which are too general.

**3.4 Hybrid Recommender Systems**

Hybrid recommender system is another category of recommender systems that tries to overcome the limitations of the other approaches. A Hybrid recommender system combines two or more recommendation techniques to gain better system optimization and fewer of the weaknesses of any individual ones. The most popular hybrid approaches are those of content-based and collaborative filtering. Usually, hybrid algorithms use both items attributes and the ratings of all users. Generally the hybrid recommender systems adapt Heuristic combination of content filtering and collaborative filtering methods.

There are different strategies by which hybridization can be achieved and they are broadly classified into seven categories that are summarized[1] in table 1:

<b>Hybridization method</b>	<b>description</b>
Weighted	The ratings of several recommendation techniques are combined together to produce a single recommendation
switching	The system switches between recommendation techniques depending on the current situation
Mixed	Recommendations from several different recommenders are presented at the same time
Feature combination	Features from different recommendation data sources are thrown together into a single recommendation algorithm
Cascade	One recommender refines the recommendations given by another
Feature augmentation	Output from one technique is used as an input feature to another
Meta-level	The model learned by one recommender is used as input to another

Table 1. Hybrid method categories

**4. TRUST BASED SOCIAL RECOMMENDER SYSTEMS**

Considering the concept of evaluating the evaluators, we focus on the trust of a recommender system. There are two different notions of trust: trust about the other users of the recommender and trust about a system's recommendations. With this notion and considers a class of recommender systems termed as "social recommender systems". These systems attempt to

generate more useful recommendations derived from information about user profiles and relationships between users that nowadays can be found virtually everywhere; e.g. in social networking sites such as Facebook, LinkedIn and MySpace.

The main claimed advantage is that users will be aware of the nature of the recommendations, i.e., how they have been identified, and will tend to place greater trust in these recommendations. Trust emerges as one out of seven roles that can be played by explanations in RSs. These roles [9] are:

- (i) transparency - explaining how the system works
- (ii) scrutability - allowing users to tell the system it is wrong
- (iii) trust - increasing user confidence in the system
- (iv) effectiveness - helping users make good decisions
- (v) persuasiveness - convincing users to try or buy
- (vi) efficiency - helping users make decisions faster and
- (vii) Satisfaction - increasing the ease of use or enjoyment.

Indeed, the literature about persuasion suggests that people are likely to accept recommendations from credible sources and we therefore conclude that the credibility of the RS is vital to increasing the likelihood of recommendation acceptance.

The ever-growing demand for peer to peer (P2P) networks for applications like file sharing, software downloading and on-line commerce become commonplace led to several concerns among which the issue of trust is predominant.

Abdul-Rahman and Hailes [10] discuss their vision on the inherent risk trustworthiness of each other that arises between the entities that mediate on-line transactions. They also explained why traditional network security mechanisms are incomplete in their function to manage trust. To minimize the risk they insisted on trust management which becomes an essential component in distributed network environment. Using the definition of trust, considering the issues which motivate need for complementary trust management they proposed a trust computing model. This model extends and generalizes security and trust management emphasizing on the following four goals:

- (i) To adopt a decentralized approach to trust management.
- (ii) To generalize the notation of trust.
- (iii) To lessen ambiguity by making trust statements more explicit.
- (iv) To facilitate the exchange of trust –related information via a common protocol.

In their model they defined agents as entities that are able to execute the recommendation protocol and trust relationships with following properties:

- (i) It is always between exactly two entities.
- (ii) It is non-symmetrical.
- (iii) It is non-transitive.

Ido Guy et al. [11] studied interestingly the personalized recommendations of software items based on social relations, including bookmarked web pages, blog entries, and communities. They focused on recommendations that are derived from the user's social network. Social network information is collected and aggregated across different data sources within their organization. The center of their approach is a comparison between recommendations that are based on the user's familiarity network and his/her similarity network. They also examine the effect of adding explanations to each recommended item that show related people and their relationship to the user and to the item.

Konstantin Pussep et al. [12] proposed a recommender system with privacy constraints. In their work, they analyze the feasibility of privacy conserving recommender systems in social



networks and presented a recommender system. They use existing social relationships to build a private social network for file sharing and recommender system on top of it. Instead of gathering all information about users at one place the users exchange information only with their social contacts. They presented a privacy-conserving recommender algorithm and compared its performance with two extremes: an algorithm that exchanges file lists only among neighbours and an algorithm that distributes file popularity rankings including user data in the entire system.

Trust on a peer is represented by a measure which combines the overall experience or satisfaction on transactions performed with that peer. Most of the current systems in literature prefer recommender systems for computing trustworthiness of a peer. RVVSV Prasad et al. [17] presented a method for reputation calculation based on credibility factor of a peer.

## **5. AGENT BASED RECOMMENDER SYSTEMS**

In the domain of recommender system [13], this paper especially deals with the use of agents as an enabling technology. Probably the most efficient solution would be to use agents as recommenders. This would mean that agents could either inspect the resources or try by themselves. However, existing quality measures and evaluation algorithms only work in some limited domain. For example, agents could be used to check the syntactic and some of the semantic validity of some information resources and to test some properties of executable resources, as programs and digital services.

It seems that there is a natural fit between agent technologies and recommender systems. Some research areas are important to both, such as machine learning, adaptivity support, user modelling and filtering techniques. For agent technologies, recommender systems offer a practical and important application domain with useful concepts. For recommender systems, agent research offer ways to manage autonomy, pro-activity, distribution, reputation and trust, etc.

Now, multi-agent systems researches have been extensively studying the notions of reputation and trust, as they form some of the key ingredients for reaching coordination and cooperation in a world of autonomous, even selfish entities. There's wealth of work considering both theoretical aspects of reputation and practical frameworks for implementing systems. The connections to the information retrieval-driven research in recommender and reputation systems are clear. It seems a very promising idea to combine these efforts. Agents can provide feedback on behalf of customers, extract useful recommendations from large data sets including inherent feedback information, reveal dishonest reporting by examining relationships between recommenders, offer platforms for open, distributed opinion management schemes, and so on.

To cope with the problem of information overload, Yan Zheng Wei et al.[14] initially developed a market-based recommender system that allows multiple agents to compete with one another to present their best recommendations to the user. They assumed that the market place coordinates between multiple recommender agents and ensures that the best recommenders are presented. To achieve this they developed a reinforcement learning and Boltzmann exploration strategy that the recommender agents can use for these tasks. Subsequently, a novel reinforcement learning strategy is developed to enable the agents to effectively and quickly learn the user's interests while still making good recommendations. They also showed that when all agents adopt this strategy, the market rapidly converges and makes good recommendations quickly and frequently.

Osmar R.Zaiane [15] suggests the use of web mining techniques to build such an agent that uses data mining techniques such as association rules mining in order to build a model that represents on-line user behaviors, and use this model to suggest activities or shortcuts. On the learner's side, it would be very useful if the system could automatically guide the learner's activities and

intelligently recommended on-line activities or resources that would favor and improve the learning using web mining in the context of e-learning.

Robert H.Guttman et al. [16] reviewed several AI technologies and focuses on issues and technologies concerning the next-generation agent mediated electronic commerce infrastructure like recommender systems, user interface approaches, negotiation mechanism and infrastructure, language, protocols.

## 6. CONCLUSION AND FURTHER APPROACH

Different approaches of recommender systems have been discussed in detail. Due to the overload of information on the World Wide Web, the necessity of recommender systems to generate efficient solutions have evolved. In the present scenario, finding the right recommender for evaluating the Credibility of recommender systems is an essential feature. Retrieval of information from huge volumes of data in diversified areas results in a tedious process. Hence, Agent mediated recommender systems have evolved to make the recommendation process trivial.

## REFERENCES

- [1]. Manisha Hiralall, (2011) Recommender systems for e-shops, Business Mathematics and Informatics paper, Vrije Universiteit, Amsterdam.
- [2]. Jannach, D. (2006) Finding preferred query relaxations in content-based recommenders. In: 3<sup>rd</sup> International IEEE Conference on Intelligent Systems, pp. 355–360.
- [3]. Resnick, P., Varian, H.R., (1997) Recommender systems. Communications of the ACM 40(3), 56–58.
- [4]. Francesco Ricci, Lior Rokach and Bracha Shapira, (2011) Introduction to Recommender Systems, Handbook, DOI 10.1007/978-0-387-85820-3\_1, © Springer Science+Business Media, LLC.
- [5]. Qing Li and Byeong Man Kim, (2004) Constructing User Profiles for Collaborative Recommender System, APWeb 2004, LNCS 3007, pp. 100-110, 2004, Springer-Verlag Berlin Heidelberg.
- [6]. Zied Zaier, Robert Goding and Lue Faucher, Evaluating Recommender Systems, International Conference on Automated solutions for Cross Media Content and Multi-channel Distribution, IEEE, 2008.
- [7]. Panagiotis Symeonidis, Alexandros Nanopoulos and Yannis Manolopoulos, (2007) Feature-Weighted User Model for Recommender Systems, LNAI 4511, pp. 97-106, Springer Verlag Berlin Heidelberg.
- [8]. M. J. Pazzani, (1999) A framework for collaborative, content-based and demographic filtering, Artificial Intelligence Review, 13(5-6):393-408.
- [9]. Kay, J, (2006) Scrutable adaptation: Because we can and must. In: Adaptive Hypermedia and AdaptiveWeb-Based Systems, 4th International Conference, AH 2006, Dublin, Ireland, June 21-23, 2006, Proceedings, pp. 11–19.
- [10]. Alfarez Abdul Rahaman and Stephen Hailes, (1997) Using Recommendations for Managing Trust in Distributed Systems, Proceedings of the IEEE International Conference on Communications.
- [11]. Ido Guy, Naama Zwerdling, David Carmel, Inbal Ronen, Erel Uziel, Sivan Yogev, Shila Ofek-Kolfman, (2009) Personalized Recommendation of Social Software Items Based on Social Relations, , ACM.

- [12]. Konstantin Pussep, Sebastian Kaune, Jonas Flick, Ralf Steinmetz, (2009) A Peer-to-Peer Recommender System with Privacy Constraints, IEEE.
- [13]. Olli Niinivaara, (2004) Agent-Based Recommender Systems, Helsinki 17.5.2004, Software Agent Technology Course Paper, Department of Computer Science, UNIVERSITY OF HELSINKI.
- [14]. Yan Zheng Wei, Lue Moreau and Nicholas R.Jennings, (2004) Learning Users' Interests in a Market-Based Recommendation Systems, IDEAL 2004, LNCS 3177, pp. 833-840, Springer Verlag Berling Heidelberg.
- [15]. Osmar R. Zaiane, (2002) Building Recommender Agent for e-Learning System, Proceedings in ICCE, IEEE.
- [16]. Robert H.Guttman, Alexandros G. Moukas and Pattie Maes, (1998) Agent Mediated Electronic Commerce : A Survey, Journal The Knowledge Engineering Review, Volume 13 Issue 2 Pages 147 – 159.
- [17]. RVVSV Prasad, Vegi Srinivas, V. Valli kumari and KSVSN Raju, "Credibility based reputation calculation in P2P networks", Springer-Verlag Berlin Heidelberg 2008, ICDCIT 2008, LNCS 5375, pp. 188-195,2008.

#### Authors

**RVVSV Prasad** received his Masters degree in computer applications from Madurai Kamaraj University and Master of Philosophy in Computer Science from Alagappa University. He is currently doctoral candidate in the Department of Computer Science & Engineering, Acharya Nagarjuna University, Nagarjuna Nagar, Guntur, India. He is working as Associate Professor in the Department of Computer Science & Engineering, Bhimavaram Institute of Engineering and Technology, Pennada, Bhimavaram, India. His research interests include recommender systems, security, privacy and trust in mobile agents and P2P networks.



**V.Valli Kumari** received her B.E. in Electronics and Communication Engineering and M.Tech. and PhD in Computer Science and Engineering all from Andhra University, India and is currently working as Professor in the same department. . She was awarded a gold medal for best research in 2008 in Andhra University. Her areas of research interest include Web, Data Engineering, Network Security, e-commerce and has 70 publications in conferences /journals of international and national repute. She is a member of IEEE, ACM, CRSI and CSI.

