

Review of Web Personalization

Zeeshan Khawar Malik, Colin Fyfe

School of Computing, University of The West of Scotland

Email: {zeeshan.malik,colin.fyfe}@uws.ac.uk

Abstract—Today the modern phase of the internet is the personalize phase where the user is able to view everything that matches his/her interest and needs. Nowadays, Web users are relying totally on the internet in relation to all the problems they have in their daily life. If someone wants to find a job he/she will look on the internet, similarly if someone wants to buy some product/item the best preferred platform will be the internet so due to large numbers of users on the internet and also due to the large amount of data on the internet people starts preferring those platforms where they can find what they need in as minimum time as possible. The only way to make the web intelligent is through personalization. Web Personalization has been introduced more than a decade ago and many researchers have contributed to make this strategy as efficient as possible and also as convenient for the user as possible. Web personalization research has a combination of many other areas that are linked with it and includes AI, Machine Learning, Data Mining and natural language processing. This report describes the whole era of web personalization with a description of all the processes that have made this technique more popular and widespread. This report has also thrown light on the importance of this strategy and also the benefits and limitations of the methods that are introduced in this strategy. This report also discusses how this approach has made the internet world more facilitating and easy-to-use for the user.

Index Terms— Web Personalization, Learning, Matching and Recommendation

I. INTRODUCTION

In the early days of internet technology, people used to suffer a lot while browsing and finding data as per their interest and needs due to the richness of information available online. The concept of web personalization has to a very large extent enabled the internet users to find the most appropriate and best information as per their interest. This is one of the major contributions on the internet derived from the first and foremost concept of Adaptive Hypermedia which becomes more popular by giving a major contribution to adaptive web-based hypermedia in teaching systems [1], [2] and [3]. Adaptive Hypermedia was derived by observing the browsing habits of different users on the internet where people faced a lot of difficulty in choosing links out of many links available at one time. Based on this linking system, this concept of adaptive hypermedia was introduced which provides the most appropriate links to the users based on their browsing habits. This concept became more popular when it was introduced in the area of educational hypermedia

[4]. Web personalization [11, 22, 72, 16] is to some extent closely linked with adaptive hypermedia in the way that the former most of the time works on an open corpus hypermedia whereas the latter mostly worked and became popular on closed corpus hypermedia. The basic objective of personalization is to some extent similar to adaptive hypermedia which is to help users by giving them the most appropriate information for their needs. The reason why web personalization has become more popular than adaptive hypermedia is due to its frequent implementation in commercial applications. Very few areas of the internet are left where this concept has not yet reached. Most areas of the internet have adopted this method including e-business [5], e-tailing [6], e-auctioning [7], [8] and others [9]. User adaptive services and personalization features both are basically designed for enabling users to reach their targeted needs without spending much time in searching.

Web Personalization is divided into three main phases

- 1) Learning [10]
- 2) Matching [11] and
- 3) Recommendation [5], [12]

as shown in Fig. 1 and Fig. 2 in detail. Learning is further subdivided into two types 1) Explicit Learning and 2) Implicit Learning. There is one more type of learning method mentioned most frequently nowadays by different researchers called behavioural learning [13] that also comes under the Implicit Learning category. The next stage is the matching phase. There is more than one type of matching or filtration techniques proposed by different researchers which primarily include

- 1) Content-Based Filtration [14]
- 2) Collaborative Filtration [15], [16] [17], [18]
- 3) Rule-Based Filtration [19] and
- 4) Hybrid Filtration [20]

These prime categories further include sub-categories mentioned later that are based on the prior mentioned categories but are used to further enhance the performance and efficiency of this phase. There is still a lot of weaknesses to the efficiency and performance of this phase. Many new ideas are currently being proposed all over the world to further improve the performance in finding the nearest neighbours in the shortest possible time and producing more accurate results. The last phase is the recommendation [21] phase which is responsible for displaying the closest match to the interest and personalized choice of users. In this report a detailed review of web personalization is made by taking into account the following major points.

This paper covers the whole review of web personalization technology
This work was supported by University of The West of Scotland.

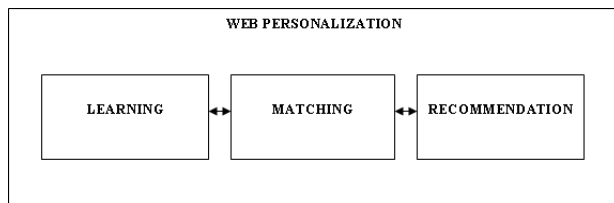


Figure 1. Three Stages of Web Personalization.

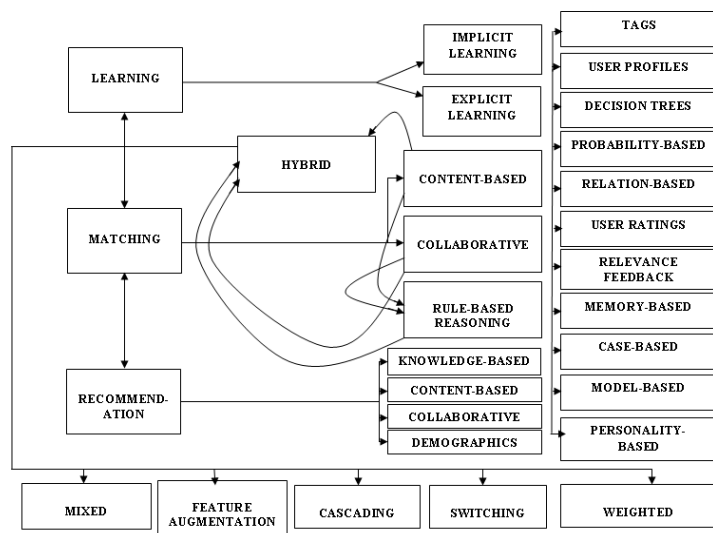


Figure 2. Web Personalization Process.

- 1) What is Web Personalization? What are the main Building Blocks of Web Personalization?
- 2) What are the major techniques that are involved in each phase of web personalization?
- 3) Description of each phase with complete overview of all the major contributions that are made in each phase of web personalization.

II. WHAT IS WEB PERSONALIZATION?

Web Personalization can be defined as a process of helping users by providing customized or relevant information on the basis of Web Experience to a particular user or set of users [22].

A Form of user-to-system interactivity that uses a set of technological features to adapt the content, delivery, and arrangement of a communication to individual users explicitly registered and/or implicitly determined preferences [23].

One of the first and foremost companies who had introduced this concept of personalization was Yahoo in 1996 [24]. Yahoo has introduced this feature of personalizing the user needs and requirement by providing different facilitating products to its users like Yahoo Companion, Yahoo Personalized Search and Yahoo Modules. Yahoo experienced quite a number of challenges which include scalability issues, usability issues and large-scale personalization issues but summing up as whole find it quite a successful feature as far as the user needs and requirements were concerned.

Similarly Amazon, one of the biggest companies in the internet market, summarizes the recommendation system with three common approaches

- 1) Traditional Collaborative Filtering
- 2) Cluster Modelling and
- 3) Search-Based Methods

as described in [25]. Amazon has also incorporated this method of web personalization and the most well-known use of collaborative filtering is also done by Amazon as well.

Amazon.com, the poster child of personalization, will start recommending needlepoint books to you as soon as you order that ideal gift for your great aunt. (<http://www.shorewalker.com>)

Web Personalization is the art of customizing items responding to the needs of users. Due to the large amount of data on the internet, people often get so confused in reaching their correct destination and spend so much time in searching and browsing the internet that in the end they get disappointed and prefer to do their work using traditional means. The only way to help internet users is by providing an organized look to the data and personalizing the whole decoration of items to satisfy the individual's desire and in doing this the only way is to embed features of web personalization. Everyday a user has a different mood when browsing the internet and based on that day's particular interest the user browses the internet, but definitely a time comes when the interest starts becoming redundant day by day and at that particular situation if the historical transactional record [5], [12] is maintained properly and the user behaviour is recorded [13] properly then the company can take benefit in filtering the record based on a single user or a group of users and can recommend useful links according to the interest of the user.

Web Personalization can also be defined as a recommendation system that performs information filtering.

The most important layer on which this feature is strongly dependent is the data layer [26]. This layer plays a very important role in recommendation. The system which is capable of storing data from more than one dimension is able to personalize the data in a much better way. Hence the feature of web personalization has a pretty closed relationship with web mining

Web Personalization is normally offered as an implicit facility to the user: whereas some websites considered it optional for the user, most websites do it implicitly without asking the user. The issues that are considered very closely while offering web personalization is the issue of high-scalability of data [27], lack of performance issues [19], correct recommendation issues, black box filtration issues [28], [29] and other privacy issues [30]. Black box filtration is defined as a scenario where the user cannot understand the reason behind the recommendation and is unable to control the recommendation process. It is very difficult to cover the filtration process for a large amount of data which includes pages and products while maintaining a correct prediction and performance

accuracy and this normally happens due to the sparsity of data and the incremental cost of correlation among the users [31], [32].

This feature has a strong effect on internet marketing as well. Personalizing users needs is a much better way of selling items without wasting much time. This feature further pushes the sales ratio and helps merchants convince their customers without confusing them and puzzling them [33]. The internet has now become a strong source of earning money. The first step towards selling any item or generating revenue involves marketing of that item and convincing the user that the items which are being offered are of a superior quality and nobody can give them this item with such a high quality and at such a low price. In order to make the first step closer to the user, one way is by personalizing the items for each user regarding his/her area of interest. It means personalization can easily be used to reduce the gap between any two objects which can be a user and a product, a user and a user, a merchant and consumer, a publisher and an advertiser [34], a friend and an enemy and all the other combinations that are currently operating with each other on the internet.

In a recent survey conducted by [23] in City University London, it is found that personalization as a whole is becoming really very popular in news sites as well. Electronic News platforms such as WSJ.com, NYTimes.com, FT.com, Guardian.co.uk, BBC News online, Washington-Post.com, News.sky.com, Telegraph.co.uk, theSun.co.uk, TimesOnline.co.uk and Mirror.co.uk which has almost completely superseded traditional news organizations are right now considered to have one of the highest user viewership platforms globally. Today news sites are highly looking towards these personalization features and trying to adopt both explicit and implicit ways that includes email newsletters, one-to-one collaborative filtering, homepage customization, homepage edition, mobile editions and apps, my page, my stories, RSS feeds, SMS alerts, Twitter feeds and widgets as a former and contextual recommendations/ aggregations, Geo targeted editions, aggregated collaborative filtering, multiple metrics and social collaborative filtering as ways for personalizing the information just to further attract a users attention and to enable users to view specific information according to their interest. Due to the increasing number of viewers day by day these news platforms are becoming one of the biggest sources of internet marketing as well and most of the advertisers from all over the world are trying very hard to offer maximum percentage in terms of PPC (Pay Per Click) and PPS (Pay Per Sale) strategy to place their advertisement on these platforms to increase their sales and to generate revenue. So it is once again proved that personalization is one of the most important features that give a very high support to internet advertising as well.

While discussing internet advertising the most popular and fastest way to promote any product or item on the internet is through affiliate marketing [35]. Affiliate Marketing offers different methods as discussed in [36]

for the affiliates to generate revenue from the merchants by selling or promoting their items. Web Personalization is playing an important role in reducing the gap between affiliates and advertisers by facilitating affiliates and providing them an easy way of growing with the merchant by making their items sell in a personalized and specific way.

With the growing nature of this feature it is proved as confirmed by [37] that the era of personalization has begun and further states that

people what they want is a brittle and shallow civic philosophy.

It is hard to guess what people really want but still researchers are trying to reach as close as possible. Further in this report the basic structure of web personalization is explained in detail.

III. LEARNING

This phase is considered one of the compulsory phases of web personalization. Learning is the first step towards the implementation of web personalization. The next two phases are totally dependent on this phase. The better this phase is executed, the better and more accurate the next two phases will execute. Different researchers have proposed different methods for learning such as Web Watcher in [38] which learns the user's interest using reinforcement learning. Similarly Letizia in [39] behaves as an assistant for browsing the web and learns the user's web behaviour in a conventional web browser. A system in [40] is described as a system that learns user profiles and analyses user behaviour to perform filtered net-news. Similarly in [41] the author uses reinforcement learning to analyse and learn a user's preferences and web browsing behaviour. Recent research in [42] proposed a method of semantic web personalization which works on the content structure and based on the ontology terms learns to recognize patterns from the web usage log files.

Learning is primarily the process of data collection defined in two different categories as mentioned earlier

- 1) Explicit Learning and
- 2) Implicit Learning

Which are further elaborated below:-

A. IMPLICIT LEARNING

Implicit learning is a concept which is beneficial since there is no extra time consumption from the user point of view. In this category nobody will ask the user to give feedback regarding the product's use, nobody will ask the user to insert product feedback ratings, nobody will ask the user to fill feedback forms and in fact nobody will ask the user to spend extra time in giving feedback anywhere and in any form. The system implicitly records different kinds of information related to the user which shows the user's interest and personalized choices. The three most important sources that are considered while getting implicit feedback for a user includes 1) Reading time of the user at any web page 2) Scrolling over the same page again and again and 3) behavioural interaction with the system.

1) *GEO LOCATIONS*: Geolocation technology helps in finding the real location of any object. This is very beneficial as an input to a personalization system and hence most of the popular portals like Google implicitly store geographical location of each user using Google search engine and then personalize the search results for each user according to the geographical location of that user. This concept is becoming very popular in other areas of the internet as well which primarily includes internet advertising [43]. Due to the increase in the mobility of internet spatial information is also becoming pervasive on the web. These mobile devices help in collecting additional information such as context information, location and time related to a particular user's transaction on the web [44]. Intelligent techniques [45], [46] and [47] are proposed by researchers to record the spatial information in a robust manner and this further plays an additional role of accuracy in personalizing the record of the user. It is evident from the fact that many services on the internet require collection of spatial information in order to become more effective with respect to the needs of the user. Services such as restaurant finders, hospital finders, patrol station finders and post office finders on the internet require spatial information for giving effective recommendations to the users.

2) *BEHAVIORAL LEARNING*: In this category the individual behaviour of a user is recorded by taking into consideration the click count of the user at a particular link, the time spent on each page and the search text frequency [13], [40]. Social networking sites are nowadays found to extract the behaviour of each individual and this information is used by many online merchants to personalize pages in accordance with the extracted information retrieved from social sites. An adaptive web is mostly preferred nowadays which changes with time. In order to absorb the change, the web should be capable enough to record user's interest and can easily adapt the ever increasing changes with respect to the user's interest in terms of buying or any other activity on the web. Many interesting techniques have been proposed to record user's behaviour [48], [49] and adapt with respect to the changes by observing the dynamic behaviour of the user.

3) *CONTEXTUAL RELATED INFORMATION*: There are many organizations like ChoiceStream, 7 Billion People, Inc, Mozenda and Danskin that are working as product development companies and are producing web personalization software that can help online merchants filter records on the basis of this software to give personalized results to their users. Some of these companies are gathering contextual related information from various blogs, video galleries, photo galleries and tweets and based on these aggregated data are producing personalized results. Apart from this since the origin of Web 2.0 the data related to users is becoming very sparse and many learning techniques are proposed by different researchers to extract useful information from this high amount of data by taking into account the tagging behaviour of the user, the collaborative ratings of the user and to record

social bookmarking and blogging activities of the user [50], [51].

4) *SOCIAL COLLABORATIVE LEARNING*: Online Social Networking and Social Marketing Sites [52] are the best platforms to derive a user's interest and to analyse user behaviour. Social Collaborative filtering records social interactions among people of different cultures and communities involved together in the form of groups in social networking sites. This clustering of people shows close relationship among people in terms of nature and compatibility among people. Social Collaborative Learning systems learn a user's interests by taking into account the collaborative attributes of people lying in the same group and give benefit to their users from these socially collaborative data by personalizing their needs on the basis of the filtered information they extract from these social networking sites. This social networking site introduces many new concepts that portray the feature of web personalization like facebook Beacon introduced by Facebook but removed due to privacy issues [53].

5) *SIMULATED FEEDBACKS*: This is the latest concept discussed by [54] and [55] in which the researchers have proposed a method for search engine personalization based on web query logs analysis using prognostic search methods to generate implicit feedback. This concept is the next generation personalization method which the popular search engines like Google and yahoo can use to extract implicitly simulated feedbacks from their user's query logs using AI methods and can personalize their retrieval process. This concept is divided into four steps 1) query formulation 2) searching 3) browsing the results and 4) generating clicks. The query formulation works by selecting a search session from user's historical data and sending the queries sequentially to the search engine. The second steps involves retrieval of data based on the query selected in the previous step. The browsing result session is the most important step in which the patience factor of the user is learned based on the number of clicks per session, maximum page rank clicked in a session, time spent in a session and number of queries executed in each session. The last step is the scoring phase based on the number of clicks the user made on each link in every session. This is one of the dynamic ways proposed to get simulated feedback based on insight from query logs and using artificial methods to generate feedbacks.

B. EXPLICIT LEARNING

Explicit Learning methods are considered more expensive in terms of time consumption and less efficient in terms of user dependency. This method includes all possible ways that merchants normally adopt to explicitly get their user's feedback in the form of email newsletter, registration process, user rating, RSS Twitter feeds, blogs, forums and getting feedbacks through widgets. Through explicit learning sometimes the chance of error becomes greater. Error arises because sometimes the user is not in a mood to give feedback and therefore enters bogus information into the explicit panel [56].

1) *EMAIL NEWSLETTER*: This strategy of being in touch with your registered users is getting very popular day by day [23] and [57]. The sign up process for this strategy will help the merchant find their user's interest by knowing which product's update the user want in his/her mailbox regularly. This strategy is the best way of electronic marketing as well as finding the interest of your customers. There are many independent companies like Aweber and Getresponse that are offering this service to most merchants on the internet and people are getting a lot of benefits in terms of revenue generation and building a close personalized relationship with their customers. Tools like iContact [58] have a functionality to do message personalization as well. Message personalization is a strategy through which certain parameters in the email's content can be generalized and is one more quick and personalized way of explicitly getting feedback by just writing one generic email for all the users.

2) *PREFERENCE REGISTRATION*: This concept is incorporated by content providing sites such as news sites to get user preferences through registration for the recommendation of content. Every person has his own choice of content view so these news sites have embedded a content preference registration module where a user can enter his/her preference about the content so that the system can personalize the page in accordance with the preferences entered by the users. Most web portals create user profiles using a preference registration mechanism by asking questions of the user during registrations that identify their interest and reason for registering but on the other hand these web portals also have to face various security issues in the end as well [59]. The use of a web mining strategy has reduced this technique of preference registration system [60].

3) *SMS REGISTRATION*: Mobile SMS service is being used in many areas starting from digital libraries [61] up to behavioral change intervention in health services [62] as well. Today mobile technology is getting popular day by day and people prefer to get regular updates on mobiles instead of their personal desktops inbox. Buyers who till now only expect location-based services through mobile are also expecting time and personalization features in mobile as well [63]. Most websites like Minnesota West are offering SMS registration through which they can get personalized interests of their users explicitly and can send regular updates through SMS on their mobiles regarding the latest news of their products and packages.

4) *EXPLICIT USER RATING*: Amazon, one of the most popular e-commerce based companies on the internet has incorporated three kinds of rating methods 1) A Star Rating 2) A Review and 3) A Thumbs Up/Down Rating. The star rating helps the customers judge the quality of the product. A Review rating shows the review of existing customer after buying the product and a Thumbs Up/Down Rating gives the customer's feedback after reading the reviews of other people related to that product. These explicit user rating methods are one of the biggest sources to judge customer's needs and desires

about the product and Amazon is using this information for personalization purposes. Explicit user rating plays a vital role in identifying user's need but extra time consumption of this process means that sometimes the user feels very uncomfortable to do it or sometimes the user feels very reluctant in doing it unless and until some benefit is coming out of it [64]. However still websites have incorporated this method to gather data and identify user's interest.

5) *RSS TWITTER FEEDS*: RDF Site Summary is used to give regular updates about the blog entries, news headlines, audio and video in a standard format. RSS Feeds help customers get updated information about the latest updates on the merchant's site. Users sometimes feel very tired searching for their interest related articles and this RSS Feeding feature help users by updating them about the articles of interest. To them this feature of RSS Feeding is very popular among content-oriented sites such as News sites and researchers are trying to evolve techniques to extract feedbacks from these RSS feeds for recommendations [65]. This concept is also being used by many merchants for the personalization process by getting user's interests with regards to the updates a user requires in the form of RSS. Similarly the twitter social platform is becoming very popular in enabling the user to get updated about the latest information. Most merchants' sites are offering integration with a user's twitter account to get the latest feeds of those merchants' product on the individual's twitter accounts. Almost 1000+ tweets are generated by more than 200 million people in one second which in itself is an excellent source for recommender systems [66]. These two methods are also used by many site owners especially news sites so that they can use this information for personalizing the user page.

6) *SOCIAL FEEDBACK PAGES*: Social feedback pages are those pages which companies usually build on social-networking sites to get comments from their customers related to the discussion of their products. These product pages are also explicitly used by the merchants to derive personalized interest of their users and to know the emotions of their customers with their products [67]. It has now become a trend that every brand, either small or large before introducing itself into the market, first uses the social web to get feedback about their upcoming brand directly from the user and then based on the feedback introduce their own brand into the market [68]. Although the information on the page seems to be very large and raw but still it is considered a very useful way to extract user's individual perception regarding any product or service.

7) *USER FEEDBACK*: User Feedback plays a vital role to get a customer's feedback about the company's quality of services, quality of products offered and many other things. This information is collected by most merchants to gather a user's interests so that they can give a personalized view of information to that user next time when the user visits their site. It is identified in [69] that most of the user feedbacks are differentiated in terms of

explicitness, validity and acquisition costs. It is identified that especially for new users explicit customer requirements as in [70] are also considered as a useful source of user feedbacks for personalization. Overall user feedbacks plays a very important role for recommendation but it is proved that in most of the systems, gradually the explicit user feedbacks decreases with time and sometimes it shows a very negative effect on a user's behaviour [71].

8) **BLOGS AND FORUMS:** Blogs and forums play a vital role in creating a discussion platform where a user can share his views about the product or services he has purchased online. Most e-companies offer these platforms to their customers where customers explicitly give their feedback regarding the products by participating in the forum or by giving comments on articles posted by the vendor related to the products or services. This information is used by the merchant for personalizing their layouts on the basis of user feedbacks from these additional platforms. Semantic Blogging Agent as in [72] is one of the agents proposed by researchers that works as a crawler and extracts semantic related information from the blogs using natural language processing methods to provide personalized services. Blogging is also very popular among mobile users as well. Blogs not only contains the description of various products, services, places or any interesting subject but also contains user's comments on each article and with mobile technology the participation ratio has increased a lot. Researchers have proposed various content recommendation techniques in blogs for mobile phone users as in [73] and [74].

IV. MATCHING

The matching module is another important part of web personalization. The matching module is responsible for extracting the recommendation list of data for the target user by using an appropriate matching technique. Different researchers have proposed more than one matching criterias but all of them lie under three basic categories of matching 1) Content-Based Filtration Technique 2) Rule-Based Reasoning Technique and 3) Collaborative Filtration Technique.

A. CONTENT-BASED FILTRATION TECHNIQUE

Content-Based filtration approach filters data based on a user's previous liking based stored data. There are different approaches for the content-based filtration technique. Some merchants have incorporated a rating system and ask customers to rate the content and based on the rating of the individual, filter the content next time for that individual [75]. There is more than one content-based page segmentation approach introduced by researchers through which the page is divided into smaller units using different methods. The content is filtered in each segment and then the decision is made whether this segment of the page is incorporated in the filtered page or not [76], [77]. Content-based filtration technique is feasible only if there is something stored on the basis of content that

shows the user's interest for e.g. it is easy to give a recommendation for the joke about a horse out of many horse related jokes stored in the database on the basis of a user's previous liking but it is impossible to extract the funniest joke out of all the jokes related to horses; for that one has to use collaborative filtration technique. In order to perform content filtration the text should be structured but for both structured and unstructured data one has to incorporate the process of stemming [78] especially news sites which contains news articles which are example of unstructured data. There are different approaches used for content filtration as mentioned in figure 3.

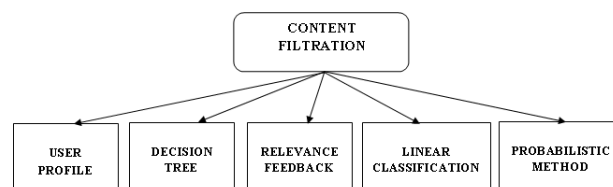


Figure 3. Methods Used in Content Filtration Technique.

1) **USER PROFILE:** The profile of user plays a vital role in content filtration [79]. The profiles mainly consist of two important pieces of information.

1) The first consists of the user's preferred choice of data. A user profile contains all the data that shows a user's interest. The record contains all the data that shows a user's preference model.

2) Secondly it contains the historical record of the user's transactions. It contains all the information regarding the ratings of users, the likes and dislikes of the users and all the queries typed by the user for record retrieval. These profiles are used by the content filtration system [80] for displaying a user's preferred data which will be personalized according to the user's interest.

2) **DECISION TREE:** A decision tree is another method used for content filtration. Decision tree is created by recursively partitioning the training data as in [81]. In decision trees a document or a webpage is divided into subgroups and it will be continuously subdivided until a single type of a class is left. Using decision trees it is possible to find the interests of a user but it works well on structured data and in fact it is not feasible for unstructured text classification [82].

3) **RELEVANCE FEEDBACK:** Relevance feedback [83] and [84] is used to help users refine their queries on the basis of previous search results. This method is also used for content filtration in which a user rates the documents returned by the retrieval system with respect to their interest. The most common algorithm that is used for relevance feedback purposes is Rocchio's algorithm [85]. Rocchio's algorithm maintains the weights for both relevant and non-relevant documents retrieved after the execution of the query and on the basis of a weighted sum incrementally move the query vector towards the cluster of relevant documents and away from irrelevant documents.

4) **LINEAR CLASSIFICATION:** There are numerous linear classification methods [86], [87] that are used for

text categorization purposes. In this method the document is represented in a vector space. The learning process will produce an output of n-dimensional weight vector whose dot product with an n-dimensional instance produces a numeric score prediction that leads to a linear regression strategy. The most important benefit of these linear classification approaches is that they can be easily learned on an incremental basis and can easily be deployed on web.

5) **PROBABILISTIC METHODS:** This is one more technique used for text classification and the method primarily used in it is the Naive Bayesian Classifier [88]. The two most common methods of Bayesian Classifier that are used for text classification are the multinomial model and multivariate Bernoulli as described in [89]. Some probabilistic models are called generative Models.

B. COLLABORATIVE FILTERING

Most online shops store records related to the buying of products by different customers. It is true that many products can be bought by many customers and it is also true that a single product can be bought by more than one customer but in order to predict which product the new customer should buy it is important to know the number of products that have been bought by other customers with the same background and choice and for this purpose collaborative filtration is performed. Collaborative filtration [27], [90], [91] is the process through which one can predict based on collaborative information from multiple users the list of items for the new users. Collaborative Filtration has some limitations as well that come with the increase in the number of items because it is very difficult to scale this technique to high volume of data while maintaining a reasonable prediction accuracy however apart from these limitations collaborative filtering is the most popular technique that is incorporated by most merchants for personalization. Many collaborative systems are designed on the basis of datasets on which these systems have to be implemented. The collaborative system designed for one dataset where there are more users than items may not work properly for any other type of datasets. The researchers in [92] perform a complete evaluation of collaborative systems with respect to the datasets being used, the methods of prediction and also perform a comparative evaluation of several different evaluation metrics on various nearest-neighbour based collaborative filtration algorithms. There are different approaches used for collaborative filtration as mentioned in Fig. 4.

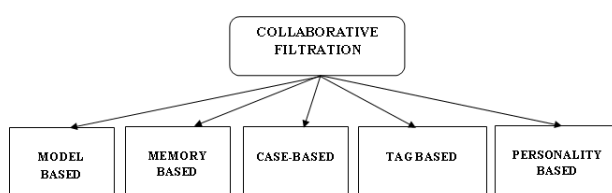


Figure 4. Methods Used in Collaborative Filtration Technique.

1) **MODEL-BASED APPROACH:** Model-based approaches such as [93] classify the data based on probabilistic hidden semantic associations among co-occurring objects. Model-based approaches divide the data into multiple segments and based on a user's likelihood [94] move the specific data into atleast one segment based on the probability and threshold value. Most of the model-based approaches are computationally very expensive but most of them gather user's interest and classify them into multiple segments [95], [96] and [97].

2) **MEMORY-BASED APPROACH:** Clustering Algorithms such as K-means [98] are considered as the basis for memory-based approaches. The data is clustered and classified based on local centroid of each cluster. Most of the collaborative filtration techniques such as [99] work on user profiles based on their navigational patterns. Similarly [100] performs clustering based on sliding window of time in active sessions and [101] presents a fuzzy concept for clustering.

3) **CASE-BASED APPROACH:** Most of the times one problem has one solution which represents a case in the case-based reasoning approach. In case-based reasoning [102] if a new customer comes and needs a solution to his/her problem then depending upon the previously stored problems that are linked with at least one case solution, the one which is nearest to the customer's problem will be considered as the case solution to his/her problem. Case-based recommender are closer to user requirements and work more efficiently and intelligently than normal filtering approaches in a way that every case works as a perfect match for a subset of users and so the data for consideration becomes less as compared to normal filtration approaches which resulted in an increase in performance as well as accuracy. Overall case-based reasoning always helps in improving the quality of recommendations [103].

4) **TAG-BASED APPROACH:** A Tag-based approach as in [104] was introduced in collaborative filtering to increase the accuracy of the CF process. Usually two persons like one item based on different reasons such as one person may like a product as he is finding that product funny whereas another user likes that item as he is finding that product entertaining, so a tag is an extra-facility to write a user's views in one or two short words in the form of a tag that shows his/her reason for his interest and will help in finding the similarity and dissimilarity among user's interest using collaborative filtration. tag-based filtration sometimes are dependent on additional factors such as popularity of tag, representation of tag and affinity between user and tags [105].

5) **PERSONALITY BASED APPROACH:** A Personality-based approach [106] was introduced to add the emotional attitude of the users to the collaborative filtration process which became further useful in reducing the high computational processing in calculating the similarity matrix for all users. User attitude plays a vital role in deriving the likes and dislikes of users so by using a big five personality model [106] the researcher explicitly derive the interest of the user that makes the

collaborative filtration process more robust and accurate.

6) *RULE-BASED FILTRATION*: This approach is one more method that is used for personalization purposes. The concept of rule-based approach is elaborated as all the business rules that are created by merchants either on the basis of transactions or on the basis of expert policies to further facilitate or create attraction in their online business. Rule-based approach such as a merchant offers gold membership, silver membership or bronze membership to its customers based on specific rules. Similarly a merchant offers discount coupons to its customers who make purchases on weekends. These rule-based approaches [107] are created in different ways as template-driven rule-based filtering approach, interestingness-based rule filtering approach, similarity-based rule-filtering approach and incremental profiling approach. Rules are also identified using mining rules as Apriori [108] which is used to discover association rules; similarly Cart is a decision tree [109] used to identify classification rules. The only limitation in the rule-based approach is the creation of invalid or unreasonable rules just on one or two transactions which makes the data very sparse and complex to understand. A rule-based approach is very much dependent on the business rules and a sudden change in any rule will have a very high impact on the whole data as well.

7) *HYBRID APPROACHES*: A single technique is not considered enough to give a recommendation taking into account all the dynamic scenarios for each user. It is true that each user has his own historical background and his own list of likes and dislikes. Sometimes one method of filtration is not enough for one particular case for example collaborative filtration process is not beneficial for a new user with not enough historical background but is proved excellent in other scenarios, similarly a content-based filtration process is not feasible where a user has not enough data associated with it that shows his likes or dislikes. Taking into account these scenarios researchers have proposed different hybrid methods [26] that include more than one technique [110] for filtration to be used for personalization purpose which could be used on the basis of a union or intersection for recommendation.

WEIGHTED APPROACH: In this approach [111] the results of more than one method for filtration are calculated numerically for recommendation.

MIXED APPROACH: In this approach [112] the results of more than one approach are displayed based on ranking and the recommender's confidence of recommendation.

SWITCHING APPROACH: In this approach [21] more than one method for filtrations is used in a way that if one is unable to recommend with high confidence the system will be switched to the second method for filtration and if the second as well is unable to recommend with high confidence, the system will switch to the third recommender.

FEATURE AUGMENTATION: In this approach [113] a contributing recommendation system is augmented with the actual recommender to increase its performance in terms of recommendation.

CASCADING: In this approach [114] the primary and secondary recommenders are organized in a cascading way such that on each retrieval both recommenders break ties with each other for recommendation.

V. RECOMMENDATION

Recommendation is considered the final phase of personalization whose performance and work is dependent wholly upon the previous two stages. Recommendation is the retrieval process which functions in accordance with the learning and matching phase. The review of all the methods which are discussed in learning and matching phase recommendation is primarily and conclusively based on four main methods that include content-based recommendation, collaborative-based recommendation, knowledge-based recommendation and based on user-demographics or user demographic profiles.

VI. FUTURE DIRECTIONS

The overall objective of reviewing the whole era of web personalization is to realize its importance in terms of the facilities it provides to the end-users as well as giving a precise overview of the list of almost all the methods that have been introduced in each of its phases. One more important aim of this review is to give a brief overview of web personalization to those researchers working in other areas of the internet so that they are able to use this feature to evolve some intelligent solutions which match human needs in their areas as well. Some of the highlighted areas of the internet for future directions with respect to web personalization are:-

1) Internet marketing is the first step towards any product or service recognition on the internet. Through web personalization one is able to judge to some extent the browsing needs of the user and if a person is able to see advertisements of those products or services which he/she is looking for then the chances of that person's interest in buying or even clicking that advertisement's link will rise. Researchers are already trying to use personalizing features for doing improved social web marketing as in [52] and helping customers in decision making using web personalization [115].

2) Internet of Things [69] is a recent development of internet. The internet of things will make all the identifiable things communicate with each other wirelessly. This concept of web personalization can offer many applications to IoT (Internet of Things) like personalizing things to control and communicate as per users interest, helping the customer in selecting the shop within a pre-selected shopping list, guidance in interacting with things of the user related to their interest and enabling the things learn from users personalized behaviour.

3) Affiliate Networks are the key platform for both the publishers and advertisers to interact with each other. There is a huge gap [34] between the publisher and advertiser in terms of selecting the most appropriate choice based on similarity. This gap can be reduced

using web personalization by collaboratively filtering the transactional profiles of publisher and advertiser and giving recommendations on the basis of best match to both of them.

4) The future of mobile networking also requires personalization, ambient awareness and adaptability [116] in its services. All services need to be personalized for each individual in his or her environment and in accordance with his or her preferences and different services should be adapted on a real time basis.

In other words in every field of life which includes aerospace and aviation, automotive systems, telecommunications, intelligent buildings, medical technology, independent living, pharmaceutical, retail, logistics, supply chain management, processing industries, safety, security and privacy requires personalization in them to enable these technologies more user specific and in compatible with human needs.

VII. CONCLUSION

Every second there is an increment of data on the web. With this increase of data and information on the web the adoption of web personalization will continue to grow unabated. This trend has now become a need and with the passage of time this trend will enter every field of our life and so in the future we will be provided with everything that we actually require. In this paper we have briefly describe the various research carried out in the area of web personalization. This paper also states how the adoption of web personalization is essential for users to facilitate, organize, personalize and to provide exactly needed data

ACKNOWLEDGMENT

The authors would like to gratefully acknowledge the careful reviewing of an earlier version of this paper which has greatly improved the paper.

REFERENCES

- [1] P. Brusilovsky and C. Peylo, "Adaptive and intelligent web-based educational systems," *Int. J. Artif. Intell. Ed.*, vol. 13, no. 2-4, pp. 159–172, Apr. 2003.
- [2] W. N. Nicola Henze, "Adaptability in the kbs hyperbook system," in *Proceedings of the 2nd Workshop on Adaptive Systems and User Modeling on the WWW*, 1999.
- [3] N. Henze and W. Nejdl, "Extendible adaptive hypermedia courseware: Integrating different courses and web material," in *Proceedings of the International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems*, ser. AH '00. London, UK, UK: Springer-Verlag, 2000, pp. 109–120.
- [4] R. Zeiliger, "Concept-map based navigation in educational hypermedia : a case study," 1996.
- [5] W.-P. Lee, C.-H. Liu, and C.-C. Lu, "Intelligent agent-based systems for personalized recommendations in internet commerce," *Expert Systems with Applications*, vol. 22, no. 4, pp. 275 – 284, 2002.
- [6] A. D. Smith, "E-personalization and its tactical and beneficial relationship to e-tailing," 2012.
- [7] C. Oemarjadi, "Web personalization in used cars e-commerce site," 2011.
- [8] C. Bouganis, D. Koukopoulos, and D. Kalles, "A real-time auction system over the www," 1999.
- [9] O. Nasraoui, "World wide web personalization," in *In J. Wang (ed), Encyclopedia of Data Mining and Data Warehousing*, Idea Group, 2005.
- [10] H. Hirsh, C. Basu, and B. D. Davison, "Learning to personalize," *Commun. ACM*, vol. 43, no. 8, pp. 102–106, Aug. 2000.
- [11] C. Wei, W. Sen, Z. Yuan, and C. Lian-Chang, "Algorithm of mining sequential patterns for web personalization services," *SIGMIS Database*, vol. 40, no. 2, pp. 57–66, Apr. 2009.
- [12] Q. Song, G. Wang, and C. Wang, "Automatic recommendation of classification algorithms based on data set characteristics," *Pattern Recogn.*, vol. 45, no. 7, pp. 2672–2689, July 2012.
- [13] M. Albanese, A. Picariello, C. Sansone, and L. Sansone, "Web personalization based on static information and dynamic user behavior," in *Proceedings of the 6th annual ACM international workshop on Web information and data management*, ser. WIDM '04. New York, NY, USA: ACM, 2004, pp. 80–87.
- [14] W. Chu and S.-T. Park, "Personalized recommendation on dynamic content using predictive bilinear models," in *Proceedings of the 18th international conference on World wide web*, ser. WWW '09. New York, NY, USA: ACM, 2009, pp. 691–700.
- [15] S. Gong and H. Ye, "An item based collaborative filtering using bp neural networks prediction," *Intelligent Information Systems, IASTED International Conference on*, vol. 0, pp. 146–148, 2009.
- [16] J. L. Herlocker, J. A. Konstan, A. Borchers, and J. Riedl, "An algorithmic framework for performing collaborative filtering," in *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval*, ser. SIGIR '99. New York, NY, USA: ACM, 1999, pp. 230–237.
- [17] J. A. Konstan, B. N. Miller, D. Maltz, J. L. Herlocker, L. R. Gordon, and J. Riedl, "GroupLens: applying collaborative filtering to usenet news," *Commun. ACM*, vol. 40, no. 3, pp. 77–87, Mar. 1997.
- [18] U. Shardanand and P. Maes, "Social information filtering: algorithms for automating word of mouth," in *Proceedings of the SIGCHI conference on Human factors in computing systems*, ser. CHI '95. New York, NY, USA: ACM Press/Addison-Wesley Publishing Co., 1995, pp. 210–217.
- [19] C.-H. Lee, Y.-H. Kim, and P.-K. Rhee, "Web personalization expert with combining collaborative filtering and association rule mining technique," *Expert Systems with Applications*, vol. 21, no. 3, pp. 131 – 137, 2001.
- [20] M. Ubuk, "Hybrid recommendation engine based on anonymous users," *Eindhoven : Technische Universiteit Eindhoven*, 2009, 2009.
- [21] D. Billsus and M. J. Pazzani, "User modeling for adaptive news access," *User Modeling and User-Adapted Interaction*, vol. 10, no. 2-3, pp. 147–180, Feb. 2000.
- [22] S. Anand and B. Mobasher, "Intelligent techniques for web personalization," in *Intelligent Techniques for Web Personalization*, ser. Lecture Notes in Computer Science, B. Mobasher and S. Anand, Eds. Springer Berlin / Heidelberg, 2005, vol. 3169, pp. 1–36.
- [23] S. Thurman, N. Schifferes, "The future of personalization at news websites: lessons from a longitudinal study," 2012.
- [24] U. Manber, A. Patel, and J. Robison, "Experience with personalization of yahoo!" *Commun. ACM*, vol. 43, no. 8, pp. 35–39, Aug. 2000.

- [25] G. Linden, B. Smith, and J. York, "Amazon.com recommendations: item-to-item collaborative filtering," *Internet Computing, IEEE*, vol. 7, no. 1, pp. 76–80, jan/feb 2003.
- [26] R. Burke, "Hybrid recommender systems: Survey and experiments," *User Modeling and User-Adapted Interaction*, vol. 12, no. 4, pp. 331–370, Nov. 2002.
- [27] A. S. Das, M. Datar, A. Garg, and S. Rajaram, "Google news personalization: scalable online collaborative filtering," in *Proceedings of the 16th international conference on World Wide Web*, ser. WWW '07. New York, NY, USA: ACM, 2007, pp. 271–280.
- [28] J. L. Herlocker, J. A. Konstan, and J. Riedl, "Explaining collaborative filtering recommendations," in *Proceedings of the 2000 ACM conference on Computer supported cooperative work*, ser. CSCW '00. New York, NY, USA: ACM, 2000, pp. 241–250.
- [29] J. ZASLOW, "If tivo thinks you are gay, here's how to set it straight what you buy affects recommendations on amazon.com, too; why the cartoons?" 2002.
- [30] E. Toch, Y. Wang, and L. Cranor, "Personalization and privacy: a survey of privacy risks and remedies in personalization-based systems," *User Modeling and User-Adapted Interaction*, vol. 22, pp. 203–220, 2012.
- [31] "Clustering items for collaborative filtering," 2001. [Online]. Available: <http://citeseer.ist.psu.edu/connor01clustering.html>
- [32] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl, "Analysis of recommendation algorithms for e-commerce," in *Proceedings of the 2nd ACM conference on Electronic commerce*, ser. EC '00. New York, NY, USA: ACM, 2000, pp. 158–167.
- [33] C. Allen, B. Yaeckel, and D. Kania, *Internet World Guide to One-to-One Web Marketing*. New York, NY, USA: John Wiley & Sons, Inc., 1998.
- [34] Z. Malik, "A new personalized approach in affiliate marketing," *International Association of Development and Information Society*, 2012.
- [35] B. B. C., "Complete guide to affiliate marketing on the web," in *Complete Guide to Affiliate Marketing on the Web*, B. B. C., Ed. Atlantic Publishing Co, 2009, pp. 1–384.
- [36] J. K. R. Bandyopadhyay, Subir Wolfe, "Journal of academy of business and economics," *Int. A. Bus. Econ. Ed.*, vol. 9, no. 4, Apr. 2003.
- [37] P. Eli, "The filter bubble," 2011.
- [38] T. Joachims, D. Freitag, and T. Mitchell, "Web-watcher: A tour guide for the world wide web," in *IN PROCEEDINGS OF THE FIFTEENTH INTERNATIONAL JOINT CONFERENCE ON ARTIFICIAL INTELLIGENCE*. Morgan Kaufmann, 1997, pp. 770–775.
- [39] H. Lieberman, "Letizia: an agent that assists web browsing," in *Proceedings of the 14th international joint conference on Artificial intelligence - Volume 1*, ser. IJCAI'95. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1995, pp. 924–929.
- [40] M. Morita and Y. Shinoda, "Information filtering based on user behavior analysis and best match text retrieval," in *Proceedings of the 17th annual international ACM SIGIR conference on Research and development in information retrieval*, ser. SIGIR '94. New York, NY, USA: Springer-Verlag New York, Inc., 1994, pp. 272–281.
- [41] Y.-W. Seo and B.-T. Zhang, "Learning user's preferences by analyzing web-browsing behaviors," in *Proceedings of the fourth international conference on Autonomous agents*, ser. AGENTS '00. New York, NY, USA: ACM, 2000, pp. 381–387.
- [42] R. G. Tiwari, M. Husain, V. Srivastava, and A. Agrawal, "Web personalization by assimilating usage data and semantics expressed in ontology terms," in *Proceedings of the International Conference & Workshop on Emerging Trends in Technology*, ser. ICWET '11. New York, NY, USA: ACM, 2011, pp. 516–521.
- [43] R. Lixăndroi, "Customizing web advertisements based on internet users' location," in *Proceedings of the 11th WSEAS international conference on mathematics and computers in business and economics and 11th WSEAS international conference on Biology and chemistry*, ser. MCBE'10/MCBC'10, Stevens Point, Wisconsin, USA, 2010, pp. 273–278.
- [44] E. Gabber, P. B. Gibbons, D. M. Kristol, Y. Matias, and A. Mayer, "Consistent, yet anonymous, web access with lpwa," *Commun. ACM*, vol. 42, no. 2, pp. 42–47, Feb. 1999.
- [45] M.-H. Kuo, L.-C. Chen, and C.-W. Liang, "Building and evaluating a location-based service recommendation system with a preference adjustment mechanism," *Expert Systems with Applications*, vol. 36, no. 2, Part 2, pp. 3543–3554, 2009.
- [46] M.-H. Park, J.-H. Hong, and S.-B. Cho, "Location-based recommendation system using bayesian users preference model in mobile devices," in *Ubiquitous Intelligence and Computing*, ser. Lecture Notes in Computer Science, J. Indulska, J. Ma, L. Yang, T. Ungerer, and J. Cao, Eds. Springer Berlin / Heidelberg, 2007, vol. 4611, pp. 1130–1139.
- [47] Y. Yang and C. Claramunt, "A hybrid approach for spatial web personalization," in *Web and Wireless Geographical Information Systems*, ser. Lecture Notes in Computer Science, K.-J. Li and C. Vangenot, Eds. Springer Berlin / Heidelberg, 2005, vol. 3833, pp. 206–221.
- [48] W. Z. J. J. M. kilfoil, A.Ghorbani and X.Xu, "Towards an adaptive web: The state of the art and science," 2003.
- [49] D. Kelly and J. Teevan, "Implicit feedback for inferring user preference: a bibliography," *SIGIR Forum*, vol. 37, no. 2, pp. 18–28, Sept. 2003.
- [50] Q. Wang and H. Jin, "Exploring online social activities for adaptive search personalization," in *Proceedings of the 19th ACM international conference on Information and knowledge management*, ser. CIKM '10. New York, NY, USA: ACM, 2010, pp. 999–1008.
- [51] M. A. Muhammad Nauman, Shahbaz Khan and F. Hus-sain, "Resolving lexical ambiguities in folksonomy based search systems through common sense and personalization," 2008.
- [52] B. Cugelman, "Online social marketing: website factors in behavioural change."
- [53] L. Story, "Feedback retreats on online tracking," 2012.
- [54] N. Kumar M and V. Varma, "An introduction to prognostic search," in *Behavior Computing*, L. Cao and P. S. Yu, Eds. Springer London, 2012, pp. 165–175.
- [55] N. K. M, "Generating simulated feedback through prognostic search approach," *Search and Information Extraction Lab*, pp. 1–61, 2010.
- [56] A. STENOVA, "Feedback acquisition in web-based learning," 2012.
- [57] K. Y. Tam and S. Y. Ho, "Web personalization: Is it effective?" *IT Professional*, vol. 5, pp. 53–57, 2003.
- [58] R. Allis, "Best practices for email marketers," *Broadwick Corps.*, 2005.
- [59] C.-H. Lee, Y.-H. Kim, and P.-K. Rhee, "Web personalization expert with combining collaborative filtering and association rule mining technique," *Expert Systems with Applications*, vol. 21, no. 3, pp. 131–137, 2001.
- [60] B. Mobasher, H. Dai, T. Luo, Y. Sun, and J. Zhu, "Integrating web usage and content mining for more effective personalization," in *Electronic Commerce and Web Technologies*, ser. Lecture Notes in Computer Science. Springer Berlin / Heidelberg, 2000, vol. 1875, pp. 165–176.

- [61] A. Parker, "Sms its use in the digital library," in *Asian Digital Libraries. Looking Back 10 Years and Forging New Frontiers*, ser. Lecture Notes in Computer Science, D. Goh, T. Cao, I. Slvberg, and E. Rasmussen, Eds. Springer Berlin / Heidelberg, 2007, vol. 4822, pp. 387–390.
- [62] C.-H. Lee, Y.-H. Kim, and P.-K. Rhee, "Web personalization expert with combining collaborative filtering and association rule mining technique," *Expert Systems with Applications*, vol. 21, no. 3, pp. 131 – 137, 2001.
- [63] A. Dickinger, P. Haghiriian, J. Murphy, and A. Scharl, "An investigation and conceptual model of sms marketing," in *System Sciences, 2004. Proceedings of the 37th Annual Hawaii International Conference on*, jan. 2004, p. 10 pp.
- [64] M. Claypool, D. Brown, P. Le, and M. Waseda, "Inferring user interest," *IEEE Internet Computing*, vol. 5, pp. 32–39, 2001.
- [65] J. J. Samper, P. A. Castillo, L. Araujo, and J. J. M. Guervós, "Nectarss, an rss feed ranking system that implicitly learns user preferences," *CoRR*, vol. abs/cs/0610019, 2006.
- [66] O. Phelan, K. McCarthy, M. Bennett, and B. Smyth, "Terms of a feather: Content-based news recommendation and discovery using twitter," in *Advances in Information Retrieval*, ser. Lecture Notes in Computer Science, P. Clough, C. Foley, C. Gurrin, G. Jones, W. Kraaij, H. Lee, and V. Mudoch, Eds. Springer Berlin / Heidelberg, 2011, vol. 6611, pp. 448–459.
- [67] S. Carter, "Get bold: Using social media to create a new type of social business)," in *Get Bold: Using Social Media to Create a New Type of Social Business*, P. K. Brusilovsky, Ed. Pearson PLC, 2012.
- [68] C.-H. Lee, Y.-H. Kim, and P.-K. Rhee, "Web personalization expert with combining collaborative filtering and association rule mining technique," *Expert Systems with Applications*, vol. 21, no. 3, pp. 131 – 137, 2001.
- [69] L. McGinty and B. Smyth, "Adaptive selection: An analysis of critiquing and preference-based feedback in conversational recommender systems," *International Journal of Electronic Commerce*, pp. 35–57, 2006.
- [70] M. Zanker and M. Jessenitschnig, "Case-studies on exploiting explicit customer requirements in recommender systems," *User Modeling and User-Adapted Interaction*, vol. 19, pp. 133–166, 2009.
- [71] G. Jawaheer, M. Szomszor, and P. Kostkova, "Characterisation of explicit feedback in an online music recommendation service," in *Proceedings of the fourth ACM conference on Recommender systems*, ser. RecSys '10. ACM, 2010, pp. 317–320.
- [72] K. T. Wolfgang Woerndl, Georg Groh, "Semantic blogging agents: Weblogs and personalization in the semantic web," *aaai.org*, 2010.
- [73] P.-H. Chiu, G. Y.-M. Kao, and C.-C. Lo, "Personalized blog content recommender system for mobile phone users," *International Journal of Human-Computer Studies*, vol. 68, no. 8, pp. 496 – 507, 2010.
- [74] D.-R. Liu, P.-Y. Tsai, and P.-H. Chiu, "Personalized recommendation of popular blog articles for mobile applications," *Information Sciences*, vol. 181, no. 9, pp. 1552 – 1572, 2011.
- [75] P. Resnick and J. Miller, "Pics: Internet access controls without censorship," *Commun. ACM*, vol. 39, no. 10, pp. 87–93, Oct. 1996.
- [76] K. Kumppusamy and G. Aghila, "A personalized web page content filtering model based on segmentation," *International Journal of Information Science and Techniques*, vol. 2, no. 1, pp. 1–51, 2012.
- [77] C. Kohlschütter and W. Nejdl, "A densitometric approach to web page segmentation," in *Proceedings of the 17th ACM conference on Information and knowledge management*, ser. CIKM '08. ACM, 2008, pp. 1173–1182.
- [78] M. Porter, "An algorithm for suffix stripping," 2006.
- [79] B. Mobasher, H. Dai, T. Luo, and M. Nakagawa, "Discovery and evaluation of aggregate usage profiles for web personalization," *Data Mining and Knowledge Discovery*, vol. 6, pp. 61–82, 2002.
- [80] S. Gauch, M. Speretta, A. Chandramouli, and A. Micarelli, "User profiles for personalized information access," in *The Adaptive Web*. Springer Berlin / Heidelberg, 2007, vol. 4321, pp. 54–89.
- [81] J. R. Quinlan, "Induction of decision trees," *Machine Learning*, vol. 1, pp. 81–106, 1986.
- [82] Y. Yang, "A comparative study on feature selection in text categorization," 1997.
- [83] "Relevance feedback and personalization: A language modeling perspective," *Systems in Digital*, 2001.
- [84] Y. Hijikata, "Implicit user profiling for on demand relevance feedback," in *Proceedings of the 9th international conference on Intelligent user interfaces*, ser. IUI '04. New York, NY, USA: ACM, 2004, pp. 198–205.
- [85] J. Rocchio, "Relevance feedback in information retrieval," 1971.
- [86] T. Zhang and F. Oles, "Text categorization based on regularized linear classification methods," *Information Retrieval*, vol. 4, pp. 5–31, 2001.
- [87] D. D. Lewis, R. E. Schapire, J. P. Callan, and R. Papka, "Training algorithms for linear text classifiers," in *Proceedings of the 19th annual international ACM SIGIR conference on Research and development in information retrieval*, ser. SIGIR '96. New York, NY, USA: ACM, 1996, pp. 298–306.
- [88] P. E. H. Richard O. Duda, "Pattern classification and scene analysis," in *Pattern Classification and Scene Analysis*. Wiley-Blackwell, 1973, pp. 1–512.
- [89] A. McCallum and K. Nigam, "A comparison of event models for naive bayes text classification," 1998.
- [90] D. M. Pennock, E. Horvitz, S. Lawrence, and C. L. Giles, "Collaborative filtering by personality diagnosis: a hybrid memory- and model-based approach," in *Proceedings of the Sixteenth conference on Uncertainty in artificial intelligence*, ser. UAI'00. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2000, pp. 473–480.
- [91] B. M. Kim and Q. Li, "Probabilistic model estimation for collaborative filtering based on items attributes," in *Proceedings of the 2004 IEEE/WIC/ACM International Conference on Web Intelligence*, ser. WI '04. Washington, DC, USA: IEEE Computer Society, 2004, pp. 185–191.
- [92] J. L. Herlocker, J. A. Konstan, L. G. Terveen, and J. T. Riedl, "Evaluating collaborative filtering recommender systems," *ACM Trans. Inf. Syst.*, vol. 22, no. 1, pp. 5–53, Jan. 2004.
- [93] T. Hofmann, "Unsupervised learning by probabilistic latent semantic analysis," *Mach. Learn.*, vol. 42, no. 1-2, pp. 177–196, Jan. 2001.
- [94] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the em algorithm," *JOURNAL OF THE ROYAL STATISTICAL SOCIETY, SERIES B*, vol. 39, no. 1, pp. 1–38, 1977.
- [95] J. S. Breese, D. Heckerman, and C. Kadie, "Empirical analysis of predictive algorithms for collaborative filtering," in *Proceedings of the Fourteenth conference on Uncertainty in artificial intelligence*, ser. UAI'98. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1998, pp. 43–52.
- [96] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl, "Item-based collaborative filtering recommendation algorithms," in *Proceedings of the 10th international conference on*

- World Wide Web*, ser. WWW '01. New York, NY, USA: ACM, 2001, pp. 285–295.
- [97] G. Shani, R. I. Brafman, and D. Heckerman, “An mdp-based recommender system,” in *Proceedings of the Eighteenth conference on Uncertainty in artificial intelligence*, ser. UAI'02. Morgan Kaufmann Publishers Inc., 2002, pp. 453–460.
- [98] A. Ahmad and L. Dey, “A k-mean clustering algorithm for mixed numeric and categorical data,” *Data amp; Knowledge Engineering*, vol. 63, no. 2, pp. 503 – 527, 2007.
- [99] F. B. K. J. F. Cyrus Shahabi, Adil Faisal, “Insite: A tool for real-time knowledge discovery from user web navigation,” 2000.
- [100] T. S. C.P. Sumathi, R. Padmaja Valli, “Automatic recommendation of web pages in web usage mining,” 2010.
- [101] A. K.Suresh, R.Madana Mohana, “Improved fcm algorithm for clustering on web usage mining,” 2011.
- [102] C. Hayes, P. Cunningham, and B. Smyth, “A case-based reasoning view of automated collaborative filtering,” in *Proceedings of the 4th International Conference on Case-Based Reasoning: Case-Based Reasoning Research and Development*, ser. ICCBR '01. London, UK, UK: Springer-Verlag, 2001, pp. 234–248.
- [103] B. Smyth, “Case-based recommendation,” in *The Adaptive Web*, ser. Lecture Notes in Computer Science. Springer Berlin / Heidelberg, 2007, vol. 4321, pp. 342–376.
- [104] J. M. S. U. Reyn Nakamoto, Shinsuke Nakajima, “Tag-based contextual collaborative filtering,” 2008.
- [105] F. A. Durão and P. Dolog, “A personalized tag-based recommendation in social web systems,” *CoRR*, vol. abs/1203.0332, 2012.
- [106] J. T. Marko TKalcic, Matevz Kunaver, “Personality based user similarity measure for a collaborative recommender system,” 2009.
- [107] G. Adomavicius and A. Tuzhilin, “Expert-driven validation of rule-based user models in personalization applications,” *Data Mining and Knowledge Discovery*, vol. 5, pp. 33–58, 2001.
- [108] C. Aggarwal and P. Yu, “Online generation of association rules,” in *Data Engineering, 1998. Proceedings., 14th International Conference on*, feb 1998, pp. 402 –411.
- [109] L. Breiman, “Classification and regression trees,” in *Classification and regression trees*. Wads-worth Brooks/Cole Advanced Books Software, 1984.
- [110] N. Taghipour and A. Kardan, “A hybrid web recommender system based on q-learning,” in *Proceedings of the 2008 ACM symposium on Applied computing*, ser. SAC '08. New York, NY, USA: ACM, 2008, pp. 1164–1168.
- [111] T. Miranda, M. Claypool, M. Claypool, A. Gokhale, A. Gokhale, T. Mir, P. Murnikov, P. Murnikov, D. Netes, D. Netes, M. Sartin, and M. Sartin, “Combining content-based and collaborative filters in an online newspaper,” in *In Proceedings of ACM SIGIR Workshop on Recommender Systems*, 1999.
- [112] B. Smyth and P. Cotter, “A personalised tv listings service for the digital tv age,” *Knowledge-Based Systems*, vol. 13, no. 23, pp. 53 – 59, 2000.
- [113] R. J. M. Prem Melville, “Content-boosted collaborative filtering for improved recommendations,” 2002.
- [114] R. Burke, “Hybrid recommender systems : A comparative study,” *cdmdepauledu*, 2006.
- [115] S. H. Ha, “Helping online customers decide through web personalization,” *Intelligent Systems, IEEE*, vol. 17, no. 6, pp. 34 – 43, nov/dec 2002.
- [116] S. Arbanowski, P. Ballon, K. David, O. Droegehorn, H. Eertink, W. Kellerer, H. van Kranenburg, K. Raatikainen, and R. Popescu-Zeletin, “I-centric com-

munications: personalization, ambient awareness, and adaptability for future mobile services,” *Communications Magazine, IEEE*, vol. 42, no. 9, pp. 63 – 69, sept. 2004.



Zeeshan Khawar Malik is currently a PhD Candidate at the University of The West of Scotland. He received his MS and BS(honors) degree from University of The Central Punjab, Lahore Pakistan, in 2003 and 2006, respectively. By profession he is an Assistant Professor in University of The Punjab, Lahore Pakistan currently on Leave for his PhD studies.



Colin Fyfe is a Personal Professor at The University of the West of Scotland. He has published more than 350 refereed papers and been Director of Studies for 23 PhDs. He is on the Editorial Boards of 6 international journals and has been Visiting Professor at universities in Hong Kong, China, Australia, Spain, South Korea and USA.