Abstract: -The main computational intelligence methods (CI) for personalization of web-based systems are identified and critical reviewed: Fuzzy Systems, Genetic Algorithms, Neural Networks, Artificial Immune Systems and Swarm Intelligence which includes Particle Swarm Optimization, Ant Colony Optimization, Bee Colony Optimization and Wasp Colony Optimization. Each method is discussed regarding its application to personalization of web-based systems. A taxonomy for personalization of web systems based on CI methods is proposed. It identifies two main approaches to personalize web-based systems as profile generation and profile exploitation which are further classified as either personalized navigation or personalized content. Future directions for application of CI modeling for personalization are discussed.

Key-words: -Web-based systems, Personalization, Computational intelligence, Modeling.

1 Introduction
The Internet grew from 2000-2009 at an estimated rate of 380% [1]. With this accelerated growth rate the size and complexity of many websites grow along with it. Millions of users continually face great difficulty interacting with web interfaces. They are bombarded with a world of information at each click. It is increasingly time consuming, confusing and frustrating for website visitors to find the information they are looking for.

Web personalization is a major part of user-centered design which addresses these user issues. It has been defined as any set of actions that can tailor the Web experience to a particular user or a set of users [2]. Personalization can be automatic or customized however extensive research is oriented towards automatic, dynamic or a combination of the two approaches over customized personalization [3].

Personalization can be applied to various aspects of a website which can be broadly classified as navigation personalization or content personalization. Personalized navigation provides automatic generation of user adapted navigation hierarchy for a set of web pages [4]. Content personalization involves user adaptation of information presented on a website [3]. Personalization techniques seek to learning about website users and create user profiles for personalization of navigation or content. A widely accepted technique to learn about users is Web Usage Mining (WUM). Raw web usage data contains hidden valuable information about user habits and preferences. The analysis of log data discovers valuable web usage patterns [5].

Processing web usage data to create accurate user models is crucial for effective personalization. A more accurate user model reflects better knowledge about the user and the ability to further predict user preferences. Information Filtering (IF), Information Extraction (IE), Information Retrieval (IR) and Collaborative Filtering (CF) are techniques for deriving useful information from user log data and model users.

Common problems in gathering, processing and analyzing web usage data using these techniques are scalability, processing time and accuracy of learning techniques. Ideally a perfect user model will be able to ‘think’ like the user by learning present user behavior, dynamically adjusting to changing user patterns and predicting future user preferences. To come as close to this ideal as possible techniques which artificially mimic the intelligence of users are applied. Personalization using Computational Intelligence (CI) techniques are “more suitable than standard techniques to approach these tasks” [6]. There is a lack of systematic review and taxonomy of CI models regarding their fit to personalize web-based systems. In the following eight identified CI methods are explained and their application to personalization of web-based systems, then taxonomy for this application is proposed followed by conclusions.
2 Computational intelligence models for personalization

CI has been defined as “the study of adaptive mechanisms to enable or facilitate intelligent behavior in complex and changing environments” [7]. This is an ongoing and evolving area of research since its term was coined by John McCarthy in 1956. Different CI models related to personalization are given in figure 1.

![Fig. 1: CI paradigms.](image)

*Fuzzy Systems* (FS) [8] and *Fuzzy Logic* (FL) mimic the concept the way people think, that is, with reasoning rather than precise. Fuzzy methods were found to be instrumental in web-based personalization when used with WUM data. User profiles are processed using fuzzy approximate reasoning to recommend personalized URLs [9]. Handling of user profiles with fuzzy concepts has been used by IR systems to provide users with personalized search engine results. Based on users web usage history data, fuzzy methods have been used to categorize or cluster web objects for web personalization [10]. Fuzzy logic was used with collective or collaborate data mining techniques to improve the quality of intelligent agents to provide personalized services to users [6].

*Evolutionary Algorithms* (EA) use mechanisms inspired by biological evolution such as reproduction, mutation, recombination and selection. One of the most popular EA is *Genetic Algorithms* (GA). They mimic the gene structure in humans based on evolutionary theory [11]. GA has been used to address some of the flaws of WUM and to tackle different problems such as personalized search, IR, query optimization and document representation [12]. GA was applied with user log mining techniques to get a better understanding of user preferences and discover associations between different URL addresses [13]. By GA was included randomness in content filtering rather than strict adherence to predefined user profiles. This is known as the element of serendipity in IR. This modified GA was introduced for optimal design of a website based on a multiple optimization criteria taking download time, visualization and product association level into consideration [14].

*Artificial Neural Networks* (ANN) or simply *Neural Networks* (NN) mimic the biological process of the human brain. A NN can be trained to group users into specified categories or into clusters. This is useful in personalization as each user group may possess similar preferences and hence the content of a web interface can be adapted to each group [15]. NNs can also be trained to learn the behavior of website users. Inputs for this learning can be derived from WUM data and CF techniques [16]. The learning ability of neural networks can also be used for real time adaptive interaction instead of only common content and static based personalization [16]. A NN was used to construct user profiles [17]. A NN was implemented to categorize e-mail folder [18].

*Swarm Intelligence* (SI) is based on the collective behavior of animals in nature such as birds, ants, bees and wasps. *Particle Swarm Optimization* (PSO) models the convergence behavior of a flock of birds [19]. PSO was used for analyzing unique behavior of web user for manipulation of web access log data and user profile data [5]. Personalized recommendation based on individual user preferences or CF data has also been explored using PSO. This was done by building up profiles of users and then using an algorithm to find profiles similar to the current user by supervised learning [20]. Personalized and automatic content sequencing of learning objects was implemented using PSO [21]. Research has also been done using PSO as a clustering algorithm but no use of this approach to clustering was found in relation to website personalization [23].

Another SI technique is *Ant Colony Optimization* (ACO) which models the behavior of ants that leave the nest to wander randomly in search of food and when it is found they leave a trail of pheromone when returning to the colony [24]. ACO resulted in the development of the shortest path optimization algorithms and has applications in routing optimization. ACO has been used to classify web users in WUM (cAnt-WUM algorithm) allowing personalization of the web system to each user class [25].

*Bees Colony Optimization* (BCO) is built on basic principles of collective bee intelligence [26]. It has been applied to web-based systems to improve the IR systems of search engines [27] incorporating WUM data [28], however the issue of personalization has not yet known to be directly addressed.
Wasp Colony optimization (WCO) or Wasp Swarm Optimization (WSO) has not yet been exploited in comparison to the other SI methods. It models the behavior of insect wasps in nature [29]. WCO has also been applied to the NP-hard optimization problem known as the Multiple Recommendations Problem (MRP). It occurs when several personalized recommendations are running simultaneously and results in churning where a user is presented with uninteresting recommendations [29]. Further research has to be done however, using WCO on real, scalable and dynamic data sets.

Artificial Immune Systems (AIS) mimic the functioning of the human immune system as the body learns to handle antigens by producing antibodies based in previous experience [30]. Applications of AIS have been solving pattern recognition problems, classification tasks, cluster data and anomaly detection. Already AIS has been applied to personalization of web-based systems. The human body is represented by a website, incoming web requests are antigens and learning is paralleled to the learning of the immune systems to produce the right antibodies to combat each antigen. Using this analogy and AIS based on WUM was used as a learning system for a website [31].

It is common practice to combine CI techniques to create a hybrid which seeks to overcome the weakness of one technique with the strength of another. Several hybrids were applied to personalization of web based systems.

NN was combined with FL to give a hybrid Neuro-Fuzzy strategy for Web personalization [32]. The topology and parameters of NN were used to obtain the structure and parameters of fuzzy rules. The learning ability of NN was then applied to this set of rules.

The ability of evolutionary techniques such as GA, to extract implicit information from user logs was combined with fuzzy techniques to include vagueness in decision making [34]. This FL-GA hybrid allows more accurate and flexible modeling of user preferences.

User data obtained from web usage data is the input for a NN. The weights and fitness functions derived from NN training is optimized using GA to derive classification rules to govern personalized decision making in e-Business [35].

A fuzzy-PSO approach was introduced to personalize Content Based Image Retrieval (CBIR). User logs were analyzed and used as the PSO input. Fuzzy principles were applied to the PSO velocity, position and weight parameters [36].

3 Personalization of web-based systems using CI models

Based on the eight major CI methods described above, it is noticed that WUM is the common input for all models. Data mining in a sense provides the fuel for personalization using CI methods. CI methods are comparable to taxonomy of intelligent agents for personalization [37]. Building on ideas from this approach taxonomy for personalization of web-based systems was proposed (cf. Fig. 2). Two main uses are identified for CI methods when applied to personalization: profile generation and profile exploitation. User profiles can further be used to personalize either the navigation or content of web-based systems.

3.1 Profile generation

Profile generation is the creation of user profiles based on both implicit WUM data and explicit user preferences [37]. User profiles can be generated either per individual or group users which appear to have similar previous web usage habits using CF techniques. Five CI methods found in previous work which were applied to user profile generation of web-based systems are: FL, NN, PSO, ACO and AIS.

FL models are constructed to identify ambiguity in user preferences [10] however there are many ways of interpreting fuzzy rules and translating human knowledge into formal controls can be challenging. NN was trained to identify similarities in user behavior [15] however for proper training the sample size must be large and the NN can be complex due to overfitting. Both PSO and GA were used to link users’ behavior by profile-matching [20] but PSO was found to outperform GA in terms of speed, execution and accuracy. ACO was used to model users with relative accuracy and simplicity [25]; however its computational complexity causes long computing time. PSO approach was found to be faster when compared to ACO [38]. AIS was used to dynamically adapt profiles to changing and new behaviors [31]. The theoretical concept of AIS is not fully sound however [33], since in reality other human systems support the functioning of the immune system and these are not modeled. The artificial cells in AIS do not work autonomously therefore the success or fail of one part of the system may determine the performance of the following step.
A hybrid method uses GA to optimize the input values of a NN, to maximize the output [35]. In this way the slow learning process of NN is helped with the optimization ability of GA.

3.2 Profile exploitation
Profile exploitation [37] personalizes various aspects of a web-based system by predefined user profiles. Two main approaches to personalize web-based systems were identified as personalization of navigation and personalization of content (cf. fig.2).

3.2.1 Personalized navigation
Personalized navigation includes WUM for personalized IR, such as search engine results, and URL recommendations. FL, BCO and GA were three main CI methods found for navigation personalization (cf. fig.2).

FL was used for offline processing to recommend URLs to users [39]. It is relatively fast, deal with natural overlap in user interests and suitable for real time recommendations. Various FL testing however showed slightly lower precision and harder to program for the fuzzy part.

GA was applied for search and retrieval [12] but is it known to be more general and abstract than other optimization methods and does not always provide the optimal solution.

BCO was used for IR [27] but it is not a widely covered area of research and currently there is a better theoretical than experimental understanding. ACO is similar to BCO and has seen more successful applications.

A hybrid between GA and FL was applied to this area. Fuzzy set techniques were used for better document modeling and genetic algorithms for query optimization to give personalized search engine results [34]. A Neuro-Fuzzy method combined the learning ability of NN with the representation of vagueness in Fuzzy Systems to overcome the NN black-box behavior and present more meaningful results than FL alone [33].

3.2.2 Personalized content
Personalized content refers to WUM for personalized web objects on each web page and sequence of content. FL, NN, GA, PSO and WCO were the main CI techniques found with applications in this area (cf. fig.2).

FL was used for a web search algorithm and to automate recommendations to eCommerce customers [41]. It was found to be flexible and able to support eCommerce application.

NN was used to group users into clusters for content recommendations [15] however overfitting problem still exists today.

GA was applied to devise the best arrangement of web objects [14]. It was found to be scalable, however it is suggest to be used in collaboration with other data mining tools.

PSO was used to sequence Learning Objects [21] and was chosen because of relative small number of parameters compared with other techniques such as GA. PSO parameter selection is also a well researched area [40]. Using a modified PSO for data clustering was found to give accurate results [22].

WCO was applied on the churning problem of uninteresting content recommendations to users [29]. This is mostly a theoretical concept, not well tested on real data and other biological inspired algorithms have found more success such as ACO.

Fuzzy-PSO was created to help improve the effectiveness of standard PSO particle movement in a content based system [36].
4 Conclusions

Eight main computational intelligence techniques were identified and critically reviewed regarding their application to personalization of web-based systems. A taxonomy with two main personalization categories as profile generation and profile exploitation was proposed. For profile generation FL, NN, PSO, GA, ACO and AIS were found to be the main CI techniques used and also a hybrid between GA and NN. Research studies show that PSO outperforms GA [20] and ACO [38]. PSO has relatively few parameters as compared to NN and has a more sound theoretical foundation than AIS. PSO is also relatively easier to program and easier to interpret than FL. Of the five methods presented above for profile generation using WUM data, PSO seems to be the more tested, useful and well rounded method.

For profile exploitation CI applications were found using FL, GA and BCO for navigation and FL, GA, WCO, NN and PSO for content personalization. Hybrid methods were also identified for each. Similarly as with profile generation, all the CI algorithms applied to profile exploitation thus far possess inherent strengths and weaknesses. Introduction of new CI techniques for personalization should address the weakness of the previous methods, but should also be validated with a sound theoretical background and testing. PSO seems to achieve this compromise. The functioning of PSO has been given superior results to other CI techniques and it is also widely tested unlike more recent techniques.

The hybrid techniques showed how the weakness of one method was aided by another method. PSO was credited with good performance in this paper as compared to the other methods however it is not without its flaws. Further work in this area may find useful combining PSO with other CI techniques to help with its inefficiencies. The more recent CI techniques methods also have a lot of room for exploration.

References:


