How to Cite:
Anusuya, S., & Mallika, N. M. (2022). Web page recommendation system using bat optimization and weighted support vector machine algorithm for health care service. *International Journal of Health Sciences*, *6*(S2), 5296–5317.
https://doi.org/10.53730/ijhs.v6nS2.6333

Web page recommendation system using bat optimization and weighted support vector machine algorithm for health care service

S. Anusuya
Research Scholar, Department of Computer Science, Sri Vasavi College, Erode, Tamilnadu, India
Corresponding author email: anuciaa@gmail.com

Dr. N. M. Mallika
Assistant Professor in Computer Science, Sri Vasavi College, Erode, Tamilnadu, India
Email: nm.mallika@gmail.com

***Abstract*---**Web-page recommendation commands a predominant part in smartweb systems. Discovery of meaningfulinformation from web utilization data and improvedrepresentation of knowledge for resourcefulweb-page recommendations are important and a huge challenge. The availabletechniques faces problems with information over load when the users tryretrieving the exact information that end users require. In addition,it also experiences issues with meaningful semantic information and thereforeit does not yield satisfactory web based services for the end users. To deal with the above statedissues, in this research work, BAOs (Bat Agent Optimizations) and WSVMs (Weighted Support Vector Machines) are combined to form BAO+WSVM algorithm for web page recommendations that are effective and useful. The experiments are carried out on the health records of COPDs (chronic obstructive pulmonary diseases) collected from websites of government hospitals. Attention has been paid on BAO+WSVM to providetheoptimal webpages for patients. User bat agent forms the interface between the user and the system providing an active environment for the users. At first, web log pre-processing is carried out with the help of data cleaning, user identification, and session identification. Later, Semantic information extractions are carried out applying DOWLs (Domain OntologyWeb language) which helps construct the ontology associated with a website utilizing its structure. Next, FCM (Fuzzy C-Mans) clustering algorithm is presented for clustering web pages based on user sessions according to the time of their web page visits. At the end, WSVMs are studied for web page recommendations that highlight on developing recommender

Manuscript submitted: 27 Feb 2022, Manuscript revised: 18 March 2022, Accepted for publication: 09 April 2022

systems and performances along with accuracies of recommender systems are enhanced. It can be inferred from experimental results that performance of proposed BAO+WSVM excels in terms of improved accuracies, precisions, sensitivities, specificities, hit ratios and reduced MAE metrics

***Keywords*---**web page recommendation, bat optimization, vector machine algorithm.

Introduction

Fast paced progresses of WWWs (World Wide Webs) and internet have resulted in massive developments of online applications including academia. With the persistent increase in access to these applications made easier, users experience the massive task of gettingexciting information matching to their requirements[1]. As a result, users are forced toexpend tremendous amount of time to look for their preferred targets and the search process has also led to the drastic rise in the usage of system resources.

Web page recommendations are gained much popularity, and are displayed as links to the associated web page, relevant image, or well known pages at websites. After the user inputs a request to web server, the session is established for the user. While the session is in progress and the user browses a website, the list of pages that the user has visited is saved as a session data. The primary objective of recommender systems is to forecast web pages of for user sessions and data of relevance to the user [2]. The core attributes of recommender systems involves learning from the past data of the current user and the other users. Recommender systems determine the domain of the current user using the past data of the userand later the pages are predictedon the basis of the user's domain.

One more significant step to make an improvement in the efficiency of recommendation system includes the extraction of the semantic information from the web page. Semantic annotation is a useful approachaimed at understanding the semantics beneath the document [3]. It yieldssome more information as metadata and it is helpfulin processing the documents smartly. The issue faced with semantic annotation is that these annotations are not global e.g. the semantic annotation for a document in a specific domain might mean different in another domain. Hence, domain oriented knowledge is utilized for semantic annotation and ontologies provide this domain basedinformation. The primary concern with Semantic Annotation is the ontology availability for the domain. Ontology consists of concept and correlations [4] [5]. In ontology, a concept may be atomic in nature or specifiedusing a set of characteristics. This group of characteristicscategories the concepts in ontology.

Later, clustering algorithm is usedfor classifying the pages. The pages are mappedonto a set of ontological item sets, and therefore pages are clustered employing semantic similarities existing between individuals in ontology.In [6], is presented a technique, used for clustering the user sessions mapped onto set of objects. In the technical work, K means clusteringis used and this mechanism is adapted for clustering the pages and the clusteringresult is helpful in improving the recommendationaccuracy [7]. In the technique, even though clustering is brought into use, pages are clustered in place of sessions. As clustering depends on semantic similarity between the individuals of ontology, the page clustering result depicts the similarity in pages. The clustering result is helpful in identifying the pages that are irrelevant during the recommendation stage.

Recommender engines integrate analyses performed by clustering web pages and generate recommendations for users where user's navigational paths are compared using association rules mining for ultimate recommendations of web pages. As toevery page recommended, the page is verified to decidethe cluster in which it is found. Once the maximum number of clusters (clusters with a large number of pages) is defined, these cluster pages are added to final recommendations. Since, association rules are made up of individual ontology, user navigational histories get s transformed into ontology instance collections. In recommendation stages, navigated items become the initial search patterns and based on window_count, a parameter specifying maximum count of pages visited previously by users for recommendations of new pages [8]. Association rules and user's navigational histories where the rule's antecedent parts are equivalent to search patterns are extracted and added to recommendation sets. The ontology instances present in rules consequently get mapped as Web page addresses, incrementing recommendation set's pages.

The primaryobjective of this technical work involvesthe web page recommendation systemapplying theBAO+WSVM algorithm. Several research and techniques have been presented, however there is no considerable increase in performance. The availabletechniquesarebogged down by the accuracy and specificity values. To get over the above statedproblems, in this technical work, BAO+WSVMtechnique is introducedfor web page recommendation systemusing which the overall performance is improved. The importantachievement of this technical work includesweb page recommendation system. The proposed technique is helpful in rendering improved results applyingefficientmethodologies.

The organization of the other sections of the research article is described as given: Section 2 presents a shortoutline of few literature works in web page recommendation system. Section 3 elucidates the proposed technique for web page recommendation system. Section 4explains about the experimental results and performance analysis discussion. In Section 5, the conclusions are studied.

Literature survey

In [9], Göksedefet al (2010) investigated hybrid recommender systems that combined results of recommendations obtained by mining Web usages. The study's conducted extensive comparative analysis on diverse ensemble techniques and recommendation approaches in terms of hybrid recommender prediction accuracies. The obtained results were then discussed in relation to hybrid recommender systems in place of single recommender models. Their results showed that hybrid recommender systems were more effective in predictions of subsequent requests to be made by Web users.

Nguyenet al. (2013) presented in [10] innovative approaches that effectively rendered Web page recommendations using semantic enhancements where website domains and Web usage information were combined. In representations of domain knowledge, two novel models were used where the first model made use of ontology to represent domain knowledge while the second model made use of automatically generated singular semantic networks to represent domain terms, web pages, and their associations. Another conceptual prediction model helped in automatic generations of semantic networks pertaining to semantic Web usage information by integrating domain and Web usage knowledge. The study generated different queries to explore these knowledge bases. In response to these inquiries, a set of recommendations generated candidate Web-pages. The study made comparisons between their recommendation results and results obtained from improved WUM (Web Usage Mining) techniques showed that their proposed technique clearly outperformed WUM.

Wanaskaret al (2013) introduced in [11] an effective approach that was based on ARM (association rule mining) for web page recommendation issues. One main issue lies in giving equal weights to all pages though the importance of pages should be based on web page visit frequencies as well as the amount of time spent on pages. Moreover, recommendations for newly added web pages users have not yet visited are not added to recommendations. To address this issue, web usage logs are used for mining adaptively and personalizing recommendations based on ARM. The study's ARMs are entirely based on Apriori algorithm. However, a few unavoidable drawbacks existed in this method. The paper presented a novel methodology based on weighted ARM and text mining. The proposed scheme was capable of incorporating semantic knowledge into obtained results increasing its efficiency and in benchmarks it displayed superior performances when compared to other current techniques.

Gulzaret al (2018) introduced recommender systems [12] that suggested and guided learners in selecting courses based on their preferences. The study's hybrid technique was used in conjunction with ontology to retrieve meaningful information and make accurate recommendations which benefitted students in terms of improving their performances as well as increasing their satisfaction. The work by addressing shortcomings of fundamental individual recommender systems, performances of recommender systems can be enhanced.

Havenset al (2012) extended FCM clustering to Voluminous data in [13]. The study's comparisons of techniques relied on 1) sampling and ensuing non-iterative extensions; 2) incremental approaches that made singular successive passes through data subsets; and 3) kernelized variants of FCMs yielding approximations based on sampling which included three algorithms. Numerical experiments using both loadable and voluminous datasets were compared based on their time and space complexities, speeds, quality of approximations to batch loadable data (FCM), and evaluation of matches between groups and ground truths. Their empirical results showed that random sampling and extended FCM, bit-reduced FCM, approximating kernel FCM were desirable options for FCM approximations of voluminous data. They confirmed their projections based on voluminous dataset with 5 billion objects and introducing recommendations using the aforesaid FCM clustering approaches.

Kothariet al (2015) propose in [14] proposed better integrations of systems that relied on both contextual and non-contextual user choices. The work suggested use of MLTs (machine learning techniques) such as SVMs. SVMs assist in isolating data in the best possible way using hyperplanes which are then classified. Users' choices are further classified using training sets generated by applying SVM's classifications. They used practical datasets in their experiments to demonstrate their technique's proficiency in dealing with contextual and non-contextual choices of users in higher education.

Methodology

This work's proposed BAO+WSVM algorithm is introduced for achieving web page recommendations that are both useful and meaningful. It includes important steps like data collections, user agents employing BAOs, web log pre-processing, semantic information extractions, web page clustering and web page recommendations.

Data collection

Relevant records of COPDs for the tests were acquired from websites of government hospitals. The data encompasses details of COPD related hospitalisation readmissions within a month of the date indexed and causes. Demographic factors including ages, genders, geographic locations, rural/urban contexts, and insurance details were considered. Additional information on patient's sicknesses was obtained from videos and photos.

User agent using BAOs (Bat Agent Optimizations)

User agents serve as interfaces users and systems providing dynamic environments. User interest models were created from user's browsing histories and profiles. User agents change user page requests into exact formats and send changed page requests from users to agents for extracting semantic information and extending queries in their corresponding domains and key terms based on ontology. Full results from the agents are also used which are ranked person-wise and user wise results are produced. Also, User agent guidesthe production of a personalized profile for a new user.The agent develops the users' browsing orbehaviour evaluation is developed into a profile and consequently, a model is obtainedon the basis of on the user preference updated with date and time where user interfaces, databases, behaviours, and inference engines are all parts of user agents.

- Major goal of using modules in user agent environments is for considering user's inputs and outputs.
- Databases contain actual user's submission data.
- Moreover, this work's knowledge base contains descriptions of users' current knowledge including personal information.
- The behaviour models precisely predict user behaviours along with formatted data..

With the objective of replicating the steps in the user agents, Bat optimizations are used, which are swarm based algorithms simulating behaviour of bats. The initial populations are generated based on past values of browsing records of users in addition to data that registration yields. In bat algorithms, each user agent portrays a velocity and position. Bat algorithms find browsing history records having decreased search time positions that all bats identify in addition to best search time rates that bats identify during the collection process [15]. Bats regulate their velocities and locations for maintaining manageable searching times required for gathering users information from multi-dimensional spaces.

$$fi = fmin + (fmin - fmax)\beta \quad (1)$$
$$v_i^j = v_i^j(t-1) + [\hat{x}^j - x_i^j(t-1)fi \quad (2)$$
$$x_i^j(t) = x_i^j(t-1) + v_i^j(t) \quad (3)$$

Where β indicates arbitrary generated integer values in the range (0,1). The value of decision variable j for bats i at time step t is represented by xji (t). The results from fi control movement speeds and ranges of bats'. The variable $\hat{x}^j$represents current global best locations (solutions) for j and determined by comparing all solutions provided by m bats. The block diagram of the suggested system is shown in Figure 1.

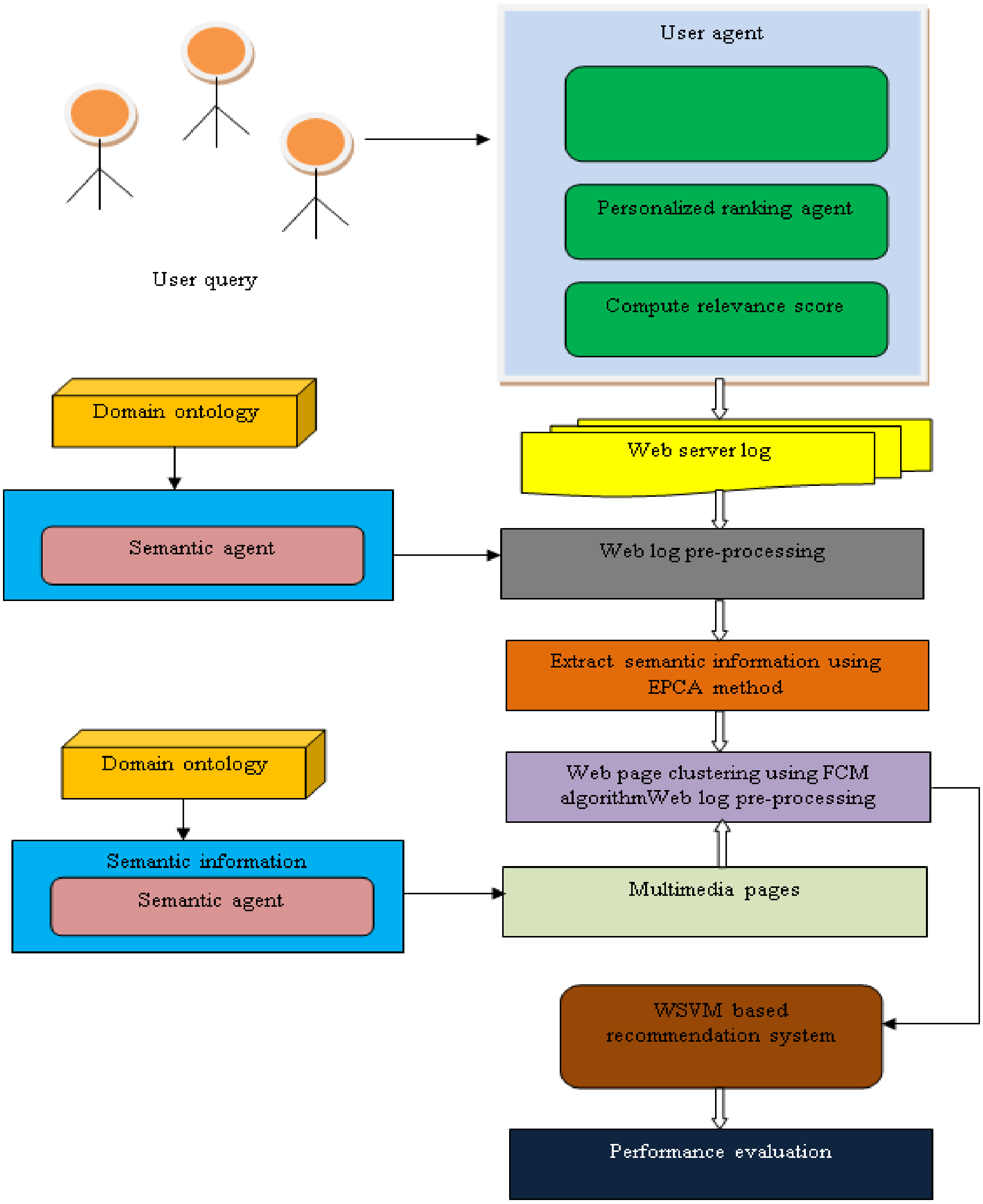

Fig 1 Architecture diagram of the proposed system

Based on the behaviour of the bats, the fitness values are computed to limit the irrelevant features effectively for the databases provided. It is intended to act as a group of bats that tracks the prey/foods with their echolocation capability. Simulated bat movements are given by updating their velocities and positions using Equations (1), (2) and (3).

In bat algorithms, customised rankings can be referred to as hearts of personalised systems based on multi-agents. New scores are determined based on user's selections, and rankings are applied on them where scores are computed based on the count of user clicks. This information gets transferred to user bat agents. This research work uses several agents and the defines iterations in its executions. (1) The max iterations counts are identified; (2) c1 and c2 are learning aspects for controlling social and cognitive element effects. The swarm sizes are noted; Bat positions on dimensions d is d.vei while d(t) refers to bat agent's velocity. Dimension d refers to the most suitable searching times for bat agents 'i' placed locally along dimension, d.ω indicates inertia weight which ensures points meet in bat algorithms. Position of bats are modified based on Equation (2) and bat agents rely entirely on locations (searching times) of algorithms. The modelling this algorithm, the following rules are followed:

1. All bats use echolocations for distance sensing and can estimate differences between foods/preys and obstacles.
2. To look for preys, bats bi fly at random velocities vi at positions xi with constant frequency $f$min, changing wavelength λ and loudness A0. Based on proximity of targets, bats change their wavelengths or frequencies of discharging pulses and modify their pulse emissions $r\in$ [0, 1],
3. Despite the fact that loudness can be measured in many ways, it is assumed that loudness varies from large (positive) A0 to smallest fixed values Ami.

Algorithm 1: BAO

Objective function (x), x = (x1, …,).
Initialize the bat population xi and vi, i = 1, 2, …,.
Specify pulse frequency fi at xi, ∀i = 1, 2,…,.
Initialize pulse rates ri and the loudness Ai, i = 1, 2,…,.

1. Input COPD database features
2. While t<T
3. For each bat bi, do
4. Generate new solutions applyingeq (1), (2) and (3)
5. If rand>ri, then
6. Extract the irrelevant feature from the given dataset
7. Reduce the dimension of the feature
8. Get the importantfeatures applying objective function values
9. Choose a solution among the optimal solutions
10. Produce local solution around the best solution
11. If rand<Ai and f(xi)<f(x') then
12. Admit the new solutions
13. Incrementri and decrement Ai
14. Sort the bats and get the current best x'

Bat's (bi) initial positions xi, velocities vi, and frequencies fi are initially determined. T denotes maximum iterations counts for time steps t. COPD database instances inputs, and characteristics are processed using pulse frequencies. Best fitness values are determined in optimum manners using objective functions. For bats that fit requirements, single solutions are picked from amongst current best solutions, and new solutions are constructed.

$$x_{new} = x_{old} + \epsilon \bar{A}(t) \quad (4)$$

in which $\bar{A}(t)$refers to the average loudness of all the bats at time t, and $\epsilon\in[-1,1]$ attempts to the direction and strength of the bat algorithm.

$$Ai(t+1)=\alpha Ai(t) \quad (5)$$

Web log pre-processing

Pre-processing is the initial phase and its main aim is to convert raw click stream data into user profile sets. Initial data gathered from online logs are insufficient and inappropriate for mining. Pre-processes are required to convert data into formats that can be used for finding patterns. Web log data sources contain undesired information and hence pre-processes follow data deletions, data cleansing, and data filtering [16]. These processes are helpful in sifting and systematizing only desirable information employing web mining algorithms which perform scheduling of web server logs. New server logs undergo cleaning, formatting, and then categorize. This stage consists of three sub steps, which are Data Cleaning, User recognition, and Session Identification

Data cleaning

Log data is maintained in the database for additional processing of data through queries and program.Data file obtainedis tremendous and it consumesnearly 80% of the overall time for the data mining. In data cleaning process, the unwanted information is eliminated from the log database.

User Identifications:

After cleaning data from log files, users need to be identified using either IP addresses or cookies. This research work used IP addresses to identify users. The IP addresses and browser types are checked from web pages. If the IP addresses in the two are different, users are assumed to be new. For same IP addresses, matching procedures are performed with users of the browser. When two browsers match, it implies users are not new else users are treated as new and users counts grows.

Session Identification

Session identificationspecifies identifying the sessions and refers to a set of pages that the same user has visited within the time of one specific visit to a web-site. A user session is regarded to have all the accesses to the page occurring during one visit to a Web site.

Semantic information extraction

As per this research,it is advantageous to both humans and machines to have the extraction of semantic information from web documents. Humans can browse and get documents in a semantic way while the machine are capable of easily processingthese structured forms. Also, integration of information extracted from several documents can yield the users, a global knowledge model of few domains.

Owing to the structure of human knowledge, the tasks of extracting semantic information in web documents

DOWLs used in this research work create ontology that corresponds to websites by utilising its structures. The information discovery gained from Web sites can be used to create ontology cores. The contents of web pages are considered when mapping concepts and instances. DOWL is again used for extracting the text's semantic features. An ontology performs the modelling of the domain knowledge structured on concept hierarchies, correlations (between concepts), and axioms. Ontology based semantic information extraction (i.e. exploiting meaning/context-associated features along withsyntax/grammar-relevant features) is capable of yielding improved performance compared to syntactic information extraction (i.e. information extractionperformed with only syntactic features), since the domain knowledge (defined in an ontology) could prove to be useful in identifying or differentiating domain-oriented terms and meanings [17].

The produced ontologies and the site address in the browsing history are mapped to various people in the ontology. This mapping is done to match the summary of the semantics of the web pages with the web objects that are present on those sites. Finally, instead of page URLs, transactions are made up of ontology of individuals.

Semantic bat agents focus on identifying semantics contained in requests made by users. To extract semantic characteristics, agents and ontology approaches are utilised which examine the links between user requests and current contents. The following phases are included in this step:

- Query pre-processing: unwanted terms in the text, such as articles, neuter pronouns, and symbolic meanings, are removed from the query.
- Semantic Analysis: This function looks for semantic components in the content of the request, such as subject, property, and objects, and analyses their relationships in a semantic manner.
- Semantic matching: User particle agents are used in conjunction with structured user requests to provide tailored recommender systems utilising semantic matching agents. This enhances user's request to meet user's specific purposes, and upgraded user requests are then submitted to the agents to search for semantics. They verify information retrieved, filter necessary data, and then processes results before displaying them to users. This component is responsible for tracing and extracting required results.

Web page clustering using FCM algorithm

FCM clustering algorithms cluster web pages in this technical work where they aid in clustering of user sessions corresponding to their time of visits on web pages. User sessions se_i are represented by multi-dimensional vectors of visited web pages in the search-space. FCM clusters session data where center points represent browsing patterns of clusters. The count of clusters, Fuzzification parameters, and error threshold values are specified based on web page information in this study. Computing mean vectors of user sessions to which clusters are assigned is a simple way to find cluster centres..

$$ce = We(con_j) = \frac{1}{|cl_t|} \sum_{se_i \in cl_t} We(con_j, se_i)$$

Clustered pages are carried out by means of computing the semantic similarities found among theindividuals of the ontology. E.g., including pages Pg1and Pg2 mapped tofew individuals.

- Pg1. Html is mapped into obj2, obj3, obj4
- Pg2. Html is mapped into obj8, obj15, obj17

Each Pg_i implies a unique web page, and Ob_{jk} a separate ontology in this study. The pages are clustered using FCM clustering, and results are utilized to enhance recommendation accuracies. Clusters are formed from pages instead of sessions and as previously stated, clustering is based on semantic similarities amongst ontology of individuals, and like hood of pages are determined based on the page clusters and where undesirable pages are filtered before recommendations. The spacing between ontology instances are computed using a different parameter. It is basically an iterative clustering approach used to improve the division of parameter through the reduction of the squared error objective function. The expression of the objective function J is given as below

$$J = \sum_{i=1}^{N} \sum_{j=1}^{cl} \mu_{ji}^{m} d^2(con_i, ce_j), \ \ m > 1 \qquad (6)$$

Let con_i refer to the i^{th} pages used for clustering, N indicates the overall number of allconceptions, ce_jsignifies the j^{th} cluster centre, cl refers to the number of clusters, uj_iindicates the amount of membership of the pages to the jth group and m specifies the fuzzycoefficient.The variable $d^2(con_i, ce_j)$represents thedistance between the conceptions $con_i$and also the cluster centrece_j. The fuzzy partitioning isachieved by reducing the objective function in the next iterative process.At first, ce_j is updated based on U and the expression becomes

$$ce_j{}^{(itr)} = \frac{\sum_{i=1}^{N} (u_{ji}{}^{(itr)})^m con_i}{\sum_{i=1}^{N} (u_{ji}{}^{(itr)})^m} \qquad (7)$$

Here '*itr*' signifies the iterative number. Next to getce_j, the fuzzy membershipmatrix is update based on the next equation

$$ce_j{}^{(itr+1)} = \frac{1}{\sum_{k=1}^{Cl} (\frac{d_{ji}}{d_{ki}})^{\frac{2}{m-1}}} \qquad (8)$$

The iterative process sees an end depending on the difference amid the present and past fuzzy membership matrices. The iterative process ends and thenext iterative process begins with gettingthece_j. up to date. At last, the centroid of acluster representing the average values of the full conceptions weighted by their amount ofbe the property of the cluster is achieved and used to calculate the level of correlation of each cluster with every webpage.Subsequently, the defuzzification process is undertaken to keep upwith the membership matrices so that the clusters having the highestdegree ofmembership in each webpage can be identified

$$CL_i = \arg_i\{\max\{u_{ji}\}\}\forall_j, \forall_k \quad (9)$$

Here CL_irepresents the cluster in which the ithconception is present. The distance amidst thetwo concepts is specified as a distance function and their properties. This constitutes the final leg ofthe proposed system. It combines examining the web-usage mining and also clusteringthe online pages and gets the user the recommended pages. The browsing path of a user getscompared with each one of thesuccessive association rules to yield recommendation of the pages. Eachsuggested page is validated to findwhich cluster it belongs to. Consequently,the maximum clusters (clusters having amassive amount of desirable pages added tothe final recommendation set are described.

Recommendation phase

In this research work, Weighted Support Vector Machine (WSVM) is introduced for the recommendation of web page. It highlights on developing recommender systems and also improves the performances along with accuracies of recommender systems. SVMs are strong MLTs which can be exploited for data classifications where efforts lie towards getting a linear separating hyperplanes having highest margins to isolate the datain a higher dimensional space. The training time of SVMs is high. To get over the problem, in this research work, weight based SVMs are presented.

During separation of different classes, WSVMs attain a level of separation that is close to optimal. WSVM performs the intrinsic embedding of data onto a high dimensional feature space, where the linear algebra and geometry may be helpful in differentiating the data, which only nonlinear rules can separate. Hyperplane is used for separating the maximum possible ratio of training data into the same class, when it increases the distance between either of the classes and the hyperplane [18]. To this end, WSVMs utilize different kernel functions allowing the inner products directly to be in feature space. In thiswork, the extracted features are used for training WSVMs for the classification process of considered datasets. WSVMs perform the separation in high-dimensional spaces by developing a hyperplane. The labeled vector of displacements of everysample weed image featureprovided is considered as the input to aWSVM classifier, yielding a model pertaining to the training data, which is later utilized for the dynamic classification ofunknown feature displacements. The testing process is performedwith the help of the training dataset model. WSVMs fall under maximal margin hyper plane classifiers achieving maximum classification accuracies in the training of sets exploiting meaningful features.

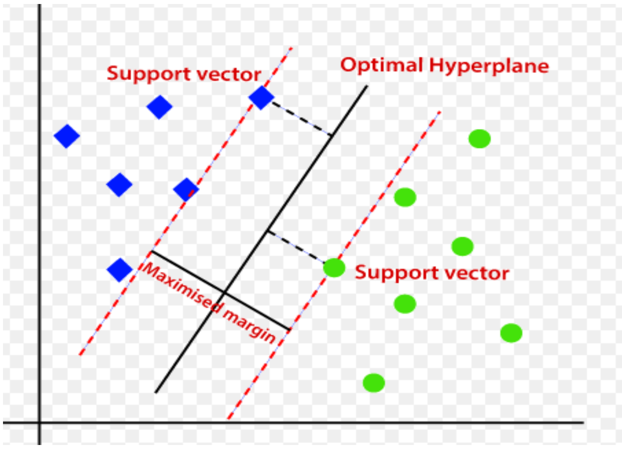

Fig 2 structure of SVMs

The fundamental concepts of WSVMs are in assigning unique weights to data points based on their relative significance in classes in way for data points to contribute differently in learning of decisions. Weights added while training datasets is:

$$\{(x_i, y_i, W_i)\}_{i=1}^{l} x_i \epsilon R^N, \quad y_i \epsilon \{-1, 1), \quad W_i \epsilon R \qquad (10)$$

where the scalar $0 \leq Wi \leq 1$refers to the weight set to data point x_i
WSVMs attempt to enhance generalisation ability by increasing separation margins and reducing classification errors, starting with development of cost functions. WSVM accomplish weighing of penalty terms such that their influences become less significant on data points used. The values assigned to C are constant, and all training data points are treated equally in training processes. WSVMs constrained optimizations can be deduced as

$$Minimize\ \Phi(w) = \frac{1}{2} w^T w + C \sum_{i=1}^{l} W_i \xi_i \qquad (11)$$

Subject to

$$y_i(\langle w, \phi(x_i) \rangle + b \geq 1 - \xi_i, \quad i = 1, \dots l \qquad (12)$$

$$\xi_i \geq 0\ i = 1, \dots l \qquad (13)$$

The weight $W_i$is set to the data point $x_i$in the expression above. In accordance, the dual formulation gets to be

$$W(\alpha) = \sum_{i=1}^{l} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{l} \alpha_i \alpha_j z_i z_j K(x_i x_j) \qquad (14)$$

In WSVMs, the upper limits of $\alpha_i$are constrained by constants C whereas they are limited by dynamic boundaries or their weights CWi. Assume a data point xi is an outlier placed on the incorrect side in kernel space. So the outlier may get to be a support vector and the upper limit of Lagrange multiplier αi is attained. In the case of WSVM training, $\alpha_i = CW_i$. The weight $W_i$fixed to this support vector (outlier) is much lesser than the rest of the support vectors. The proposed WSVMs enhance recognition accuracies.

Algorithm 2: WSVM

Input: multimedia database

1. Use Bat algorithm for optimum user page selection
2. Choose the most frequent pages using fitness values
3. Perform web log pre-processing
4. Carry out semantic information extraction using
5. Extract the most informative features
6. Perform WSVM classification process
7. Carry out training and testing process
8. Classify the input data
9. Apply (12) and (14) to obtain accurate classification results
10. Predict user's choices
11. Recommend the more accurate web pages

In recommendations, browsed items are regarded as usage patterns if they match wi_{count}, a parameter that specifies the max count of pages that have been visited previously and used for recommending a new page. Association rules and user's browsing histories are merged, and partially associated rules with antecedent parts that are similar to browsing patterns are extracted and added to recommendation sets. URLs are then mapped back to ontology objects resulting in incrementing recommended page counts. Undesirable pages are detected utilising information associated with page clustering, only e pages that are different are deleted. Prior approaches recommended pages to users based on page clustering information, and counts of these pages were retained in clusters. Maximum clusters were chosen, and pages classified under these clusters which were the recommendations for users. This work suggests WSVMs for developing multimedia database recommendation systems where the suggested systems have high levels of accuracy, and WSVMs based recommender systems help users identify and evaluate things of interest, since, WSVM's processes are highly dynamic and adaptable.

Experimental result

The trials are conducted using COPD patients' medical information obtained from government hospital websites. The obtained datasets covered all causes and COPD-related hospitalisation readmission rates for one month using the index date as starting points. It took into account demographic factors such as ages, genders, geographic locations, rural/urban environments, and insurance types. The northeast, north central, west, and south are among the regions evaluated. Multimedia related elements such as music, video, and photos will be included in the illness information obtained. Perspectives into personal health information, self-monitoring health values, education, and coaching to accomplish personal health-relevant goals make up key information in healthcare records together with multimedia data. The focus is on the first six weeks of log data following release. The existing Association Rule Mining (WARM) [19], PASOs (Particle Agent Swarm Optimizations) Web Page Recommendation (PASO-WPR) [20], and suggested BAO+WSVMs are used to measure performance metrics such as accuracies, specificities, precisions, sensitivities, hit ratios, and AMEs.

It has 12 variables, as indicated in Table 1. A total of 160 log files are taken into account. Nearly 25% of the cleaned sessions from each data set are chosen at random to be the testset, while the other sessions are utilised as the training set. The effects of using various recommender methods are investigated in this paper.

Table 1 Dataset description

S. no	Variable	Description
1	ip	IP-Address only the first three octets is considered formaintaining the distinctness. The fourth octetis masked with a 3 character string, which helps not show the complete identity of the IP
2	Uid	Users' identification number at the time of session
3	Date	Hospital log file date
4	Time	Hospital log file time
5	Zone	Hospital log file zone
6	AT	Kind of action that patient / user perform
7	doc	This variable gives the filename of the file requested, which also includes the document extension.In case of missing filename and just the file extension exists, then the filename becomes thedocument accession number
8	code	Hospital log file status code for the request
9	Filesize	document file size
10	OAT	optional extra information about the action
11	norefer	assigned a value of one if the Hospital log file referrer field is empty, and zero otherwise
12	noagent	assigned a value of one if the Hospital log file user agent field is empty, and zero otherwise
13	Multimedia files	Audio, video and images

$Recomment_{List}$refers to the initial 'n' web pages of session Se is usedin the form of the input of the recommendation engine and $Real_{List}$ specifies the second section is taken as the upcomingrequests (page visits), which are matched up with the results of recommender systems. The precision of a transaction is defined by the ratio of the number ofweb pages suitably predicted and the overall number of web pages predicted

Precision

The precision is formulated as below:

$$\text{Precision} = \frac{TP}{TP+FP} \tag{15}$$

Where TP- True Positive, FP-FalsePositive

Precision is a measure of precision or quality, whereas recall is a measure of completeness or quantity. A maximum precision usually means that an algorithm has returned a large number of relevant results rather than irrelevant ones. When it comes to classification tasks, the accuracy of a class is determined by the ratio of genuine positives to the total number of components designated as belonging to the positive class.

Sensitivity

Sensitivity is characterised by the ratio of genuine positives that are correctly detected, and it is also known as the true positive rate, recall, or likelihood of detection in a few domains (e.g., the percentage of ill individuals who are rightly identified to be affected with the condition)

$$\text{Sensitivity} = \frac{TP}{TP+FN} \tag{16}$$

Where TP stands for True Positive, FN indicates False Negative

Specificity

Specificity (also known as the genuine negative rate) is a measurement of the ratio between correctly diagnosed actual negatives (e.g., the percentage of healthy individuals who are rightlydetected to not be affected by the condition).

$$\text{Specificity} = \frac{TN}{TN+FP} \tag{17}$$

Where TN indicates the True Negative and FP refers to the False Positive

Accuracy

The overall correctness of the model determines accuracy, which is calculated by dividing the entire real classification parameters (TP+TN) by the sum of the classification parameters (TP+TN+FP+FN). The accuracy formula is as follows:

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{(TP+TN+FP+FN)}} \tag{18}$$

Where TP- True Positive, TN- True Negative, FP-FalsePositive and FN is False Negative

Mean Absolute Error (MAE)

MAE is given by the average absolute differentseen between the recommendedwebpages and the original webpages. MAE calls outeach of thewrong prediction using its distance tothe original webpages

$$MAE = \frac{\sum_{i=1}^{n} |y_i - x_i|}{n} \tag{19}$$

Where MAE =mean absolute error
y_i=prediction, x_i=true value$n$=total number of data points

Hit ratio

A hit is proclaimed when any one of the foursuggested pages forms the subsequent request$pg_{j+1}$of the user. The Hit-Ratio is given by theratio of number of hits and the over-all amount of recommendations that the system generates.

$$HR = \frac{number\ of\ hits}{n} \qquad (20)$$

Weighted

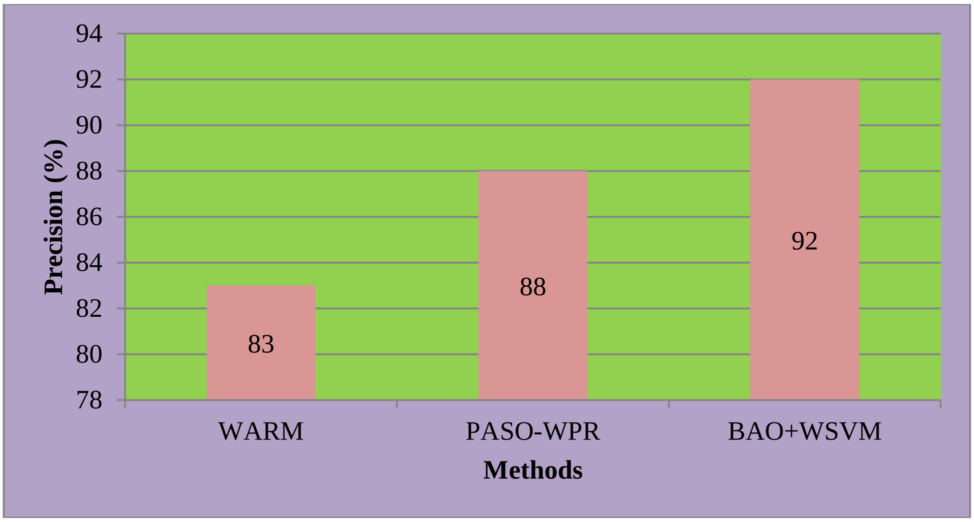

Fig 3 precision

The above Fig 3 shows the comparison analysis between the existing and proposed techniques carried out in terms of precision metric. The techniques are drawn along the x-axis and the precision value is taken along the y-axis. The precision of the current approaches including WARM and PASO-WPRalgorithms is much lesser while the proposed BAO+WSVM algorithm yields an increased value of precision for the dataset provided. It can be concluded from the result that the proposed BAO+WSVMimproves the web page recommendation system performanceby achieving semantic information extraction effectively.

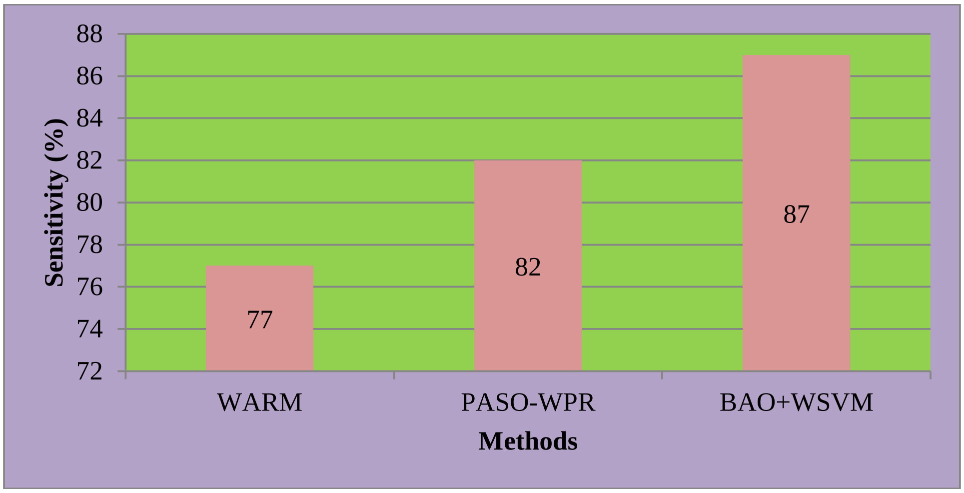

Fig 4 Sensitivity

The above Fig 4 demonstrates the comparison analysis between the current and proposed techniques in terms of sensitivity metric. The sensitivity of arecommender system measures the potential of a system of generating the pages,which the users can possibly visit.The currenttechniques likeWARM and PASO-WPR algorithms yieldreducedsensitivitywhile the proposed BAO+WSVM algorithm rendered increase in sensitivityvalue for the dataset provided. It can be confirmed from the result that the proposed BAO+WSVMimproves the web page recommendation system performance through the efficient semantic information extraction

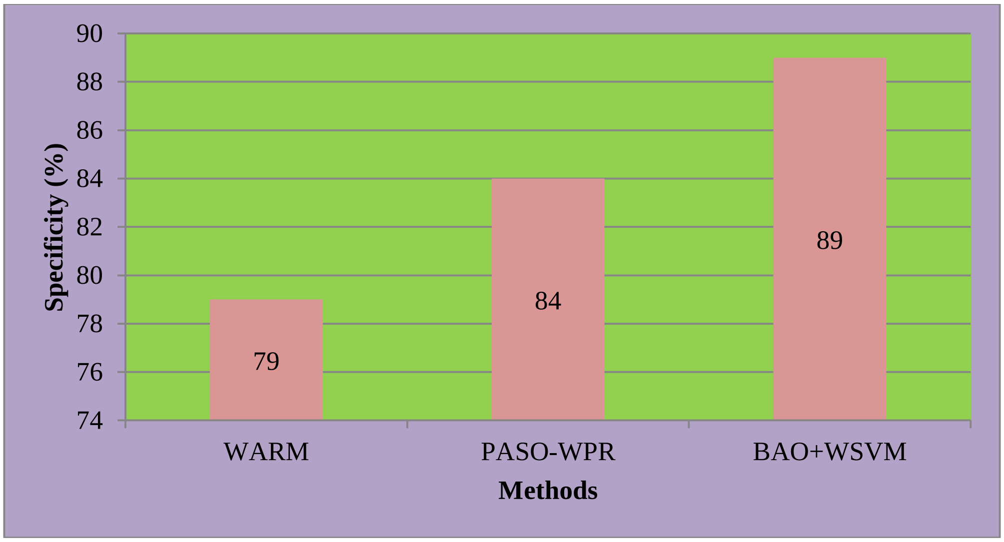

Fig 5 Specificity

In terms of specificity, the comparison measure is assessed using both current and new methods, as shown in Fig 5. The sensitivity of a recommender system calculates a system's ability to build each and every page that is likely to be viewed by users. Existing algorithms like WARM and PASO-WPR have lesser specificity, however the suggested BAO+WSVM approach has a greater specificity for the provided dataset. As a consequence, the suggested BAO+WSVM improve the performance of web page recommendation systems by efficiently extracting semantic information.

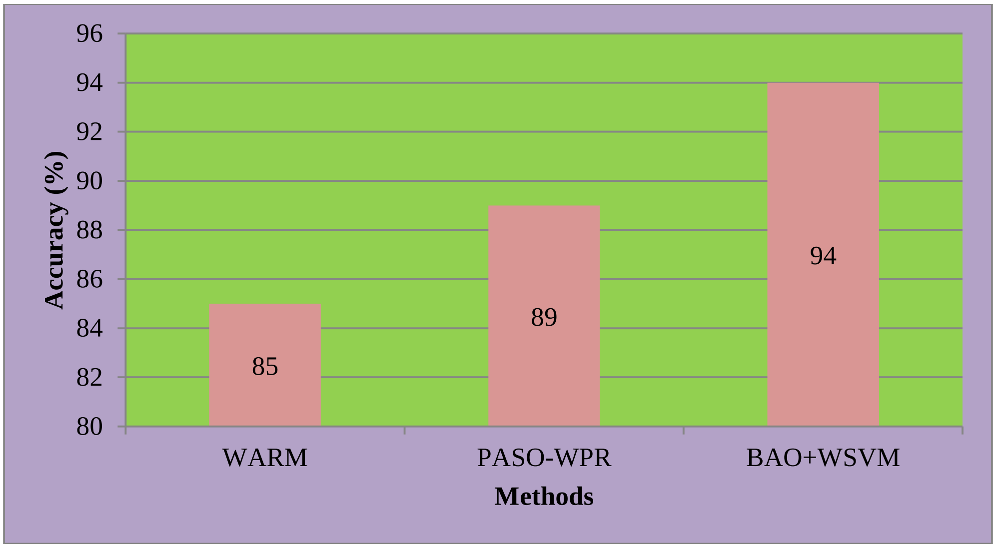

Fig 6 Accuracy

Fig 6 depicts the comparison analysis between the available and proposed techniques in terms of accuracy. The sensitivity of arecommender system helps measures the system's potential to generate each and every page,which the users can possibly visit.The availabletechniques likeWARM and PASO-WPR algorithms deliver reducedaccuracywhile the proposed BAO+WSVM algorithm renders improvedaccuracy for the dataset provided. It can be confirmed from the result that the proposed BAO+WSVM increase the web page recommendation system performance through the efficient semantic information extraction

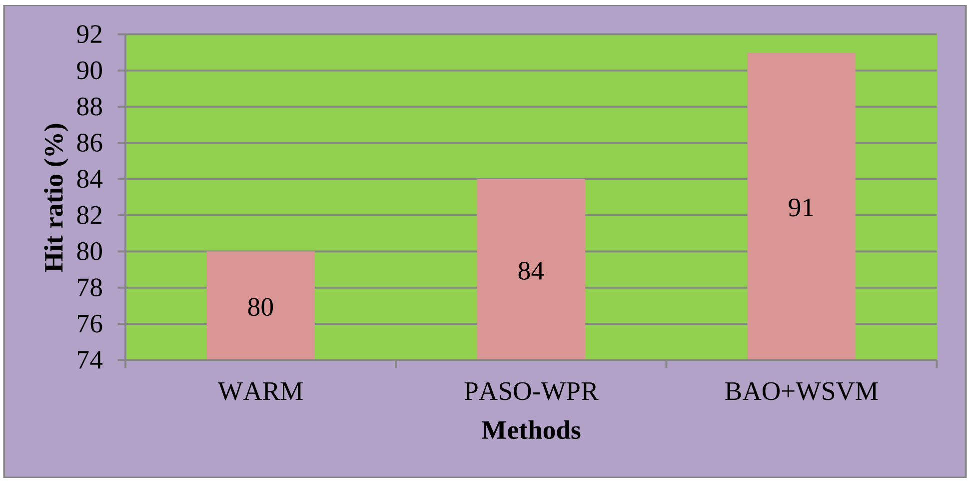

Fig 7 Hit ratio

The proposedresearch work introducesbat agents to improve the searching speed and the relevance score is alsotaken into consideration for improving the recommendation outcomes. When number of session's increases, the hit ratio is also increased using the proposed BAO+WSVM method rather than the existing WARM and PASO-WPR algorithms. It can be inferred from the result that the proposed BAO+WSVMimproves the web page recommendation system performance through the efficient semantic information extraction

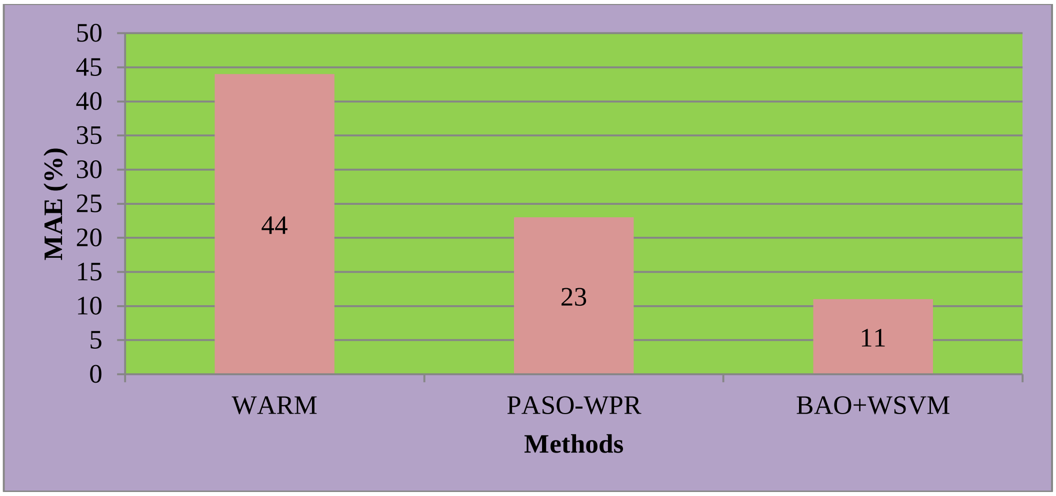

Fig 8 MAE

The above Fig 8 shows the comparison analysis carried out between the contemporary and proposed techniques in terms of MAE metrics. The proposed BAO+WSVM recommend the more accurate web pages for the users.The MAE of the current techniques like WARM and PASO-WPR algorithms arehigh while that of the proposed BAO+WSVM algorithm is much lesser for the dataset provided. It can be concluded from the results that the proposed BAO+WSVM increase the web page recommendation system performance through the efficient semantic information extraction

Conclusion

In this work, BAO+WSVM based web page recommendation is introduced with the aim toimprove the semantic web information on multimediadata. In the proposed work, BAO+WSVM method is developed for healthcare, identification of videoand other multimedia objects are complete through metadata, tags or titles. With the objectiveto achieve this purpose, semantic web are integrated into the generated patterns. In this work, preprocessing is carried outwhich contains data cleansing, user recognition, session identification and path completion. Then semantic information extraction is performed by using DOWL which is used to extract the more informative with semantic data for understanding the health status of the patients. After that FCM algorithm is applied for web page clustering then WSVM is proposed for web page recommendation. It is focused to build recommender systems and also to improve the performance as well as accuracy of recommender systems. From the experimental result, it can be confirmed that the proposed

BAO+WSVM yields superior performance in terms of higher accuracy, precision, sensitivity, specificity, hit ratio and lower MAE metrics.

References

1. Wu, Yi-Hung, Yong-Chuan Chen, and Arbee LP Chen. "Enabling personalized recommendation on the web based on user interests and behaviors." *Proceedings Eleventh International Workshop on Research Issues in Data Engineering. Document Management for Data Intensive Business and Scientific Applications. RIDE 2001*. IEEE, 2001.
2. Pushpa, C. N., et al. "Web page recommendations using radial basis neural network technique." *2013 IEEE 8th International Conference on Industrial and Information Systems*. IEEE, 2013.
3. Malik, Sanjay Kumar, and Syed AM Rizvi. "Information extraction using web usage mining, web scrapping and semantic annotation." *2011 International Conference on Computational Intelligence and Communication Networks*. IEEE, 2011.
4. Remi, Sa, and S. Ca Varghese. "Domain ontology driven fuzzy semantic information retrieval." *Procedia Computer Science* 46 (2015): 676-681.
5. Gu, Haiyan, et al. "An object-based semantic classification method for high resolution remote sensing imagery using ontology." *Remote Sensing* 9.4 (2017): 329.
6. Yilmaz, Hakan, and Pinar Senkul. "Using ontology and sequence information for extracting behavior patterns from web navigation logs." *2010 IEEE International Conference on Data Mining Workshops*. IEEE, 2010.
7. Park, Sungjune, Nallan C. Suresh, and Bong-KeunJeong. "Sequence-based clustering for Web usage mining: A new experimental framework and ANN-enhanced K-means algorithm." *Data & Knowledge Engineering* 65.3 (2008): 512-543.
8. Ali, Farman, et al. "Merged ontology and SVM-based information extraction and recommendation system for social robots." *IEEE Access* 5 (2017): 12364-12379.
9. Nguyen, ThiThanh Sang, Hai Yan Lu, and Jie Lu. "Web-page recommendation based on web usage and domain knowledge." *IEEE Transactions on Knowledge and Data Engineering* 26.10 (2013): 2574-2587.
10. Wanaskar, Ujwala, SheetalVij, and DebajyotiMukhopadhyay. "A hybrid web recommendation system based on the improved association rule mining algorithm." *arXiv preprint arXiv:1311.7204* (2013).
11. Kothari, Aansi A., and Warish D. Patel. "A novel approach towards context based recommendations using support vector machine methodology." *Procedia Computer Science* 57 (2015): 1171-1178.
12. Havens, Timothy C., James C. Bezdek, Christopher Leckie, Lawrence O. Hall, and MarimuthuPalaniswami. "Fuzzy c-means algorithms for very large data." *IEEE Transactions on Fuzzy Systems* 20, no. 6 (2012): 1130-1146.
13. Gulzar, Zameer, A. Anny Leema, and Gerard Deepak. "Pcrs: Personalized course recommender system based on hybrid approach." *Procedia Computer Science* 125 (2018): 518-524.
14. Göksedef, Murat, and ŞuleGündüz-Öğüdücü. "Combination of Web page recommender systems." *Expert Systems with Applications* 37.4 (2010): 2911-2922.

15. Qasim, Omar Saber, and Zakariya Y. Algamal. "Feature selection using different transfer functions for binary bat algorithm." *International Journal of Mathematical, Engineering and Management Sciences* 5.4 (2020): 697-706.
16. Langhnoja, Shaily, MehulBarot, and Darshak Mehta. "Pre-processing: procedure on web log file for web usage mining." *International Journal of Emerging Technology and Advanced Engineering* 2.12 (2012): 419-423.
17. Haribabu, Sairam, et al. "A novel approach for ontology focused inter-domain personalized search based on semantic set expansion." *2019 Fifteenth international conference on information processing (ICINPRO)*. IEEE, 2019.
18. Banjoko, A. W., et al. "Weighted support vector machine algorithm for efficient classification and prediction of binary response data." *Journal of Physics: Conference Series*. Vol. 1366. No. 1. IOP Publishing, 2019.
19. Etminani, Kobra, Amin RezaeianDelui, and Mahmoud Naghibzadeh. "Overlapped ontology partitioning based on semantic similarity measures." *2010 5th International Symposium on Telecommunications*. IEEE, 2010.
20. Manikandan, R., and V. Saravanan. "A novel approach on Particle Agent Swarm Optimization (PASO) in semantic mining for web page recommender system of multimedia data: a health care perspective." *Multimedia Tools and Applications* 79.5 (2020): 3807-3829.