Volume 6, Issue 3, March 2016 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Optimized Feature Selection and Classification of Web Pages using Hybrid ACCS Algorithm

P. Malarvizhi
Research Scholar, Department of CSE
Karpagam University, Coimbatore,
Tamilnadu, India

N. Radhika
Professor, Department of CSE
Amrita Vishwa Vidyapeetham, Coimbatore,
Tamilnadu, India

Abstract*— Web is a huge information resource growing at an uncontrollable rate with hundreds of thousands of new web sites created every day. This massive increase of web sites introduces a challenge of keeping the web page directories up to date and creates a need for classification of web pages to facilitate searching and retrieval of information from it and is more difficult due to a large variety of noisy information embedded in web pages. To address the above problem, we have proposed an approach for classification of web pages based on optimized features of Hybrid ACCS algorithm. In the proposed approach, the web pages are collected from web by using the web crawler and N-gram entropy model is used for extracting some added features in the web pages. The Hybrid ACCS algorithm will optimize the web pages and select the optimum features and the Map Reducer the parallel programming model classifies the web pages based on the optimized features of Hybrid ACCS algorithm. On experimentation with four dataset of 3356 web pages and comparison of results with ACO and CS , the proposed approach for classification of web pages based on the optimized features of Hybrid ACCS algorithm performs well and more accurate in classifying web pages.

***Keywords*— Web page classification, Web crawler, N-gram entropy, Hybridization, MapReducer**

I. INTRODUCTION

Data mining is a process to extract interesting, implicit, previously unknown and potentially useful knowledge or patterns from data in large databases [1]. It is one of those latest technologies with potential to support or help organizations on the very vital processed data in their respective warehouse of data [2]. Web mining is the use of data mining techniques to automatically discover and extract information from Web documents and services [3].Automatic categorization or classification of Web pages have been studied extensively since the Internet has become a huge repository of information, in terms of both volume and variance [4].The availability of satellite imagery has expanded over the past few years, and the possibility to perform fast processing of massive databases comprising this kind of imagery data has opened ground-breaking perspectives in many different fields [5]. A relevant challenge for supervised classification techniques (which assume prior knowledge in the form of class labels for different spectral signatures) is the limited availability of labelled training samples, since their collection generally involves expensive ground campaigns [6]. Web-page classification is much more difficult than pure-text classification due to a large variety of noisy information such as advertisement banner and navigation bar embedded in Web pages [7].

HTML is a visual representation language, much useful information about the content organization is lost after the authoring step. If a pure-text classification method is directly applied to these pages, it will incur much bias for the classification algorithm making it possible to lose focus on the main topics and important content [8]. Traditional text classification is typically performed on structured documents written with consistent styles (e.g., news articles), while Web collections do not have such a property [9]. Using structured data such as Web pages rather than plain text as the input opens up new possibilities for extracting information from individual pages and large networks of pages [13]. Webpage classification can help improve the quality of web search and improve SEO skill. The techniques to pre-process natural language and extract the text semantics are too costly for large amounts of data. In addition, they are only effective with well-structured text, a thesaurus and other conceptual information [16]. A major problem of the web page classification is the high dimensionality of the feature space and it is necessary to select “good” subsets of features from the original feature space to reduce the dimensionality and to improve the efficiency and run time performance of the classification process [17]. However, in consequence of the usability and the physical nature of the Worldwide Web, several favorite specializations of HTML documents emerged. A document may contain many links (e. g. a link collection), a technical text (e. g. a research article), almost no text along with several pictures (e. g. an advertisement page), or a short answer to a particular question (e. g. a mail in a help forum) [10].

The first step in web page classification is preprocessing. After pre-processing and indexing, the important step of text classification is feature selection to construct vector space which improves the scalability, efficiency and accuracy

of a text classifier. The main idea of Feature Selection (FS) is to select subset of features from the original documents. FS is performed by keeping the words with highest score according to predetermined measure of the importance of the word [11]. This paves the path for research in finding an optimized feature extraction technique that works well with less number of features without lacking the performance. The prime objective of this paper is to find Optimum feature selection technique as an efficient automatic classification of Web pages with less computational time and improved accuracy. The proposed framework will be done in three stages. In the first phase Web Crawler will be used for collecting the sets of websites. The second phase is done in two steps which consists of feature extraction and optimum feature selection technique. Initially the salient features the keywords are extracted from the web pages. With the classical features, the proposed system uses the N-gram entropy model for extracting the added features and then the Hybrid ACCS feature selection technique which is the combination of Ant and Cuckoo optimization techniques is introduced for selecting the optimum features. Finally the Map Reducer acts as classifier for classifying the web pages based on the optimized features of Hybrid ACCS algorithm.

The paper is organised as follows. Section 2 concentrates on related works. Section 3 portrays about the proposed method for web page classification and the insights about Web crawler, N-gram entropy model , Hybrid ACCS algorithm and MapReduce are clarified in section 3.1, 3.2, 3.3, 3.4 respectively. Section 4 demonstrates the experimental evaluation and results. Section 5 ends with conclusion.

II. RELATED WORKS

L. Jayasimman and E.George Dharma PrakashRaj [12] proposed a Genetic Optimized Parallel MLP to Improve Classification Accuracy for Web Learning System. Here a questionnaire is prepared to identify the cognition of the student and his website preferences in the web learning environment. Then all the user preferences are classified using proposed neural network classifiers with an optimization technique. The cognitive attributes are used as the training input for the proposed genetically optimized neural network.

Lambodar Jena and Narendra Kumar Kamila [13] proposed a text mining approach for Data Extraction and Web page Categorization using boilerplate codes for classifying the individual text elements in a Web page. They enhanced it for date as well as content extraction from a web page by using text analysis process. Similarly classification of web page content is essential to many tasks in web information retrieval such as maintaining web directories, to organize the web information as well as focused crawling. Moreover the algorithm and code has improved for web page categorization. It categorizes all the web pages and extracts all the relevant dates enclosed within a web page with its relevant content as well as title. Here they are achieving a remarkable accuracy for categorization of web pages, date extraction, title extraction, as well as for comment extraction.

PikakshiManchanda *et al.* [14] carried out work on the automated classification of web pages using artificial neural network. They proposed a novel approach for web page classification that uses the HTML information present in a web page for its classification using Artificial Neural Network (ANN) which utilizes the information provided by HTML elements of a web page in order to identify its domain. Their work involves classifying web pages into a number of domains as specified by the user. Also, the work belongs to the category of multi-class classification, with multiple classes or categories such as entertainment, food, medicine, sports, education, etc.

SiniShibu *et al*. [15] proposed a combination approach for Web Page Classification using Page Rank and Feature Selection Technique in which they proposed a unique method for classification of web pages by applying feature selection technique along with page rank and the aim is to improve the classification and search time of web contents so as to provide accurate and speedy results to the users.

J. AlameluMangai *et al.* [16] proposed a novel approach for automatic web page classification using feature intervals in which a new web page classification algorithm using weighted voting of feature intervals known as WVFI is proposed. This classifier first discretizes the web page features using a supervised discretization algorithm which identifies the number of intervals each feature has to be discretized automatically. Each feature is then made to predict the class of the corresponding feature in the test web page using the class distribution of its intervals. The final class of the test web page is predicted by aggregating the weighted vote of each feature.

EsraSaraç and Selma AyGeOzel [17] proposed An Ant Colony Optimization Based Feature Selection for Web Page Classification and the aim of their study is to reduce the number of features to be used to improve runtime and accuracy of the classification of web pages in which they used an ant colony optimization (ACO) algorithm to select the best features and then applied the well-known C4.5, naive Bayes, and *k* nearest neighbor classifiers to assign class labels to web pages. They used the Web KB and Conference datasets in the experiments and showed that using the ACO for feature selection improves both accuracy and runtime performance of classification.

III. WEB PAGE CATEGORIZATION BASED ON OPTIMIZED FEATURES OF HYBRID ACCS ALGORITHM

The proposed method for web page classification based on optimized features of Hybrid ACCS algorithm is given in Fig.1 . and flowchart in Fig. 2 .

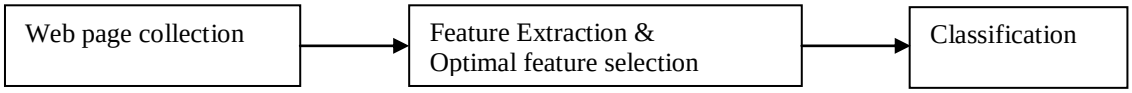

Fig 1. Framework of the proposed method for classification of web pages with optimized features

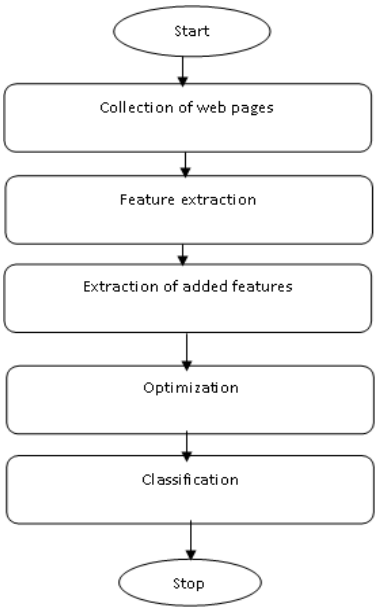

Fig 2. Flow chart of the proposed method for classification of web pages with optimized features.

Fig. 1 and Fig.2 shows the framework and flow chart of the proposed work for classifying web pages based on optimized features. The web crawler collects web pages from web and the features the keywords are extracted from these collected web pages after the preprocessing of stemming and stop word removal and the extracted features are then given to N-gram entropy model to extract some added features which are then given to Hybrid ACCS for optimization. Hybrid ACCS the feature selection technique is the Hybridization of Ant Colony and Cuckoo Search optimization technique optimizes the features extracted from N-gram entropy and these optimized features are then given to MapReducer for classification of web pages which classifies the webpages into six categories.

A. Web Crawler

Web crawling is the methodology utilized via internet searchers to collect pages from web and the function of traditional web crawler is given in Fig. 3

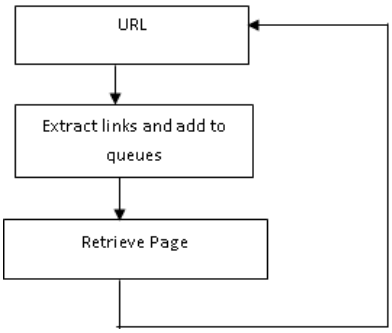

Fig 3. Traditional Web crawler

The web crawler begins with a single URL, downloads that page, extract the links from that page to others and add to the queue and this process is repeated with each of those pages till the stopping criteria of the web crawler are met which is the number of web pages downloaded or the total file size will be used as the indicators to stop crawling.

B. N-gram Entropy model

N-gram entropy is an N character slice of a more extending string. N-Gram is a statement expectation algorithm using probabilistic methods to predict next word after observing N-1 words. Therefore, computing the probability of the next word is closely related to computing the probability of a sequence of words. N-gram probabilities can be estimated by counting relative frequency on a training corpus and the basic probability of N-gram model can be expressed in (1)

$$P\left(x_1, x_2, x_3, \ldots\ldots\ldots x_m\right) = \prod_{i=1}^{m} P\left(x_i | x_{i-(n-1),\ldots\ldots} x_{i-1}\right) \qquad (1)$$

In the above equation it is observed that the probability of observing the i^{th} page is x_i and the n^{th} page is observed by n-1. The N-gram frequency count of web pages are calculated by a conditional probability as in (2)

$$P\left(x_i | x_{i-(n-1),\ldots\ldots} x_{i-1}\right) = \frac{Count\left(x_{i-(n-1),\ldots\ldots} x_{i-1}, x_i\right)}{Count\left(x_{i-(n-1),\ldots\ldots} x_{i-1}\right)} \qquad (2)$$

Typically the frequency count of web pages is not counted directly because of few issues and can use smoothing techniques before applying the probabilistic frequency count. The N-gram entropy model is used to evaluate the contents accessible in pages.

C. Hybrid ACCS Algorithm

Our proposed algorithm is Hybrid Ant Colony and Cuckoo Search algorithm which combines the advantages of Ant Colony Optimization and Cuckoo search. Though, both the Ant Colony and Cuckoo Search are good optimization techniques, but still combining them gives a great impact in classification. In ACO, the ants will walk through the path where the pheromone density is high to select the feature and this sequential feature selection process makes the entire system of feature selection slow. The major disadvantage in the ACO is that while trying to solve the combinatorial optimization problems the search has to perform much faster, but in ACO the performance will be slow. Hence Cuckoo search is used to overcome the drawback of slow performance of ACO in local search which is simple with a single parameter apart from the population size and the Hybrid ACCS algorithm is given in algorithm 1.

3.3.1) Ant colony Algorithm: Ant colony algorithm is a form of swarm intelligence where each ant lays down an amount of pheromone and the other ants are attracted to the strongest aroma to clarify the highway pattern they utilize while foraging for food and the ants tend to converge to the shortest path which is faster to transverse. If an equal amount of ants follow the long path and the short path, the ants that follow the short path will make more trips to the food and back to the colony and will deposit more pheromone over a given distance when compared to the longer path.

3.3.2) Cuckoo Search Optimization: It is another algorithm of optimization which is developed by Xin-She Yang and Suash Deb in 2009 [18]. It was dully motivated by the brood parasitism of cuckoo species. This is done by laying their respective eggs in another host birds' nest. Out of all host birds, some could hold straight divergence with the pushy cuckoos. Some out of all cuckoo species have become part in the way that female of species of parasitic cuckoos are very particular in the imitation especially in various colors and pattern of their eggs of host species which are chosen. Out of so available optimization methods, this Cuckoo Search proposed such behaviour works of breeding at almost of all spaces. Cuckoo Search finds the various expressions: Where each egg shows a solution in particular nest and so each egg is nothing but a solution provided by cuckoo search. The primary and most important aim of egg replace is to provide most optimized solutions with the simplicity of one egg per nest. From the study of various optimization techniques in various areas, it is very clear that cuckoo search perform better than ant colony, particle swarm optimization, etc., and its success make it applicable for more tedious and complex cases. [19]

1. Begin
2. Initialize the values
3. Update the initial values by calculating Fitness function f_p
4. Choose a value randomly say k
5. If (f_p < k)
6. Replace k by new value
7. End if
8. Worst values are eliminated by a fraction Ps
9. Rank the solutions and keep the best solution
10. End while
11. End

Algorithm 1. Hybrid ACCS Algorithm

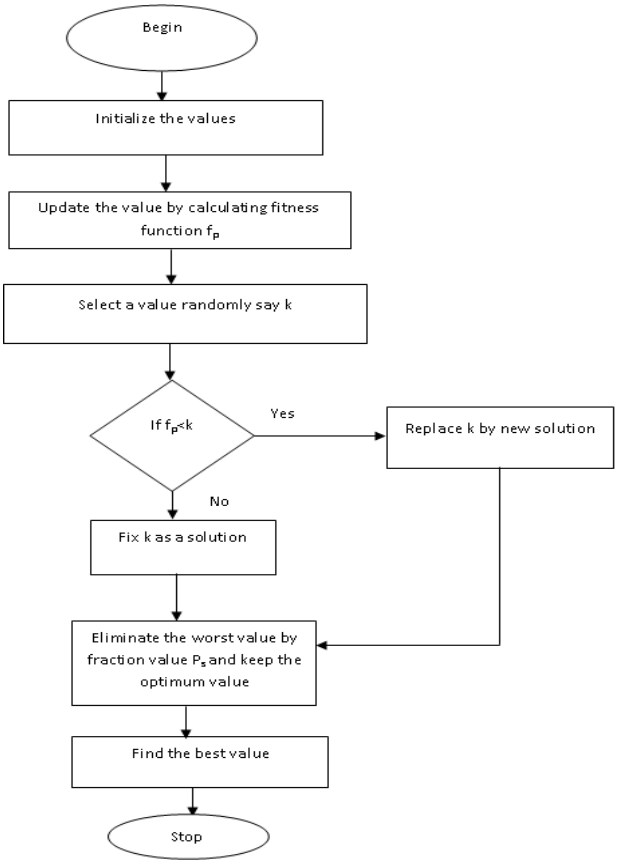

Fig 4. Flowchart of the Hybrid ACCS algorithm

Algorithm 1. gives the Hybrid ACCS technique and Fig 4. shows the flowchart for the Hybrid ACCS algorithm. The Hybrid ACCS algorithm is initialized with the values the features extracted from N-gram entropy model as in (3)

$$S_{ij}(0) \leftarrow \Delta S_{ij} \quad (3)$$

Where ΔS_{ij} is the number of web pages collected from web
The initialized value is then updated using the fitness function f_p which is calculated as in (4)

$$f_p = \begin{cases} 1 + |f|, & f < 0 \\ \frac{1}{1+f}, & f \geq 0 \end{cases} \quad (4)$$

After calculating the fitness value, a web page say 'k' is randomly selected and the fitness value is compared with another value as in (5)

$$If \ f_p < k \begin{cases} yes, \ f_p \ is \ the \ solution \\ otherwise \ k \ is \ the \ solution \end{cases} \quad (5)$$

If f_p is the solution, the randomly selected value 'k' is replaced by a new solution and the worst values are eliminated by characterizing a fraction value P_s which was calculated as in (6) and only the optimum values are retained.

$$P_s(s_n) = P_{smax} - \frac{s_n}{NI}(P_{smax} - P_{smin}) \quad (6)$$

Where NI is the number of total iterations and s_n is the Number of current iterations.

D. Map Reduce

Map reduce is a simple powerful execution engine, which can be done with other data storage and management components for classifying web pages based on the optimized features. It is a parallel programming model with functions map and reduce and classifies the web pages based on the optimized features. The map function takes the key, value as input and generates an intermediate key, value list as output and is given to the reduce function. The reduce function performs the specified process and gives the output is shown in fig. 5

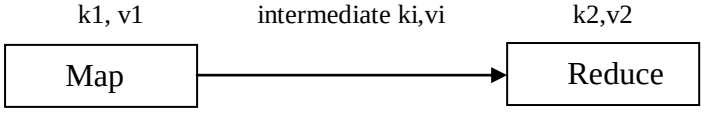

Fig 5. Map Reduce

Map: (k1,v1) → (ki, vi list)
Reduce: (ki, vi list) → (k2,v2)

IV. RESULTS AND DISCUSSION

The proposed approach for web page classification was implemented using Dot net frame work with a data set of 3356 web pages as test data and the performance of the proposed approach was evaluated on test data using the evaluation metrics of precision, recall and accuracy.

A. Data Set

Four data set are created with the web pages collected from web by the web crawler of data set-1, data set-2, data set-3 and data set-4 is given in Table I . For each dataset we mine a set of web pages from the source and totally 3356 web pages were collected for classification where each data set is classified into six category of arts, business, computers, mailing, social networking and sports.

Table I . Representation of Data Set

Data set	No. of web pages
Data set-1	640
Data set-2	893
Data set-3	1557
Data set-4	266

B. Performance Measures

The performance of the proposed algorithm was evaluated by calculation precision, recall and accuracy. The data set extracted from the web crawler was given to the N-gram entropy. The N-gram entropy framework reduces the web pages based on the occurrences and gives some weightage to the URLs. The output from the N-gram entropy is given to Hybrid Ant Colony and Cuckoo Search algorithm. The algorithm will optimize the web pages based on the fitness function and probabilistic function and choose the best URLs. Finally, the output was given to the map reduce function that reduce the URLs and provide appropriate category for each web pages. The evaluation metrics used defined as below Precision is the positive predictive value which is the fraction of retrieved web pages that are relevant.

$$Precision = \frac{\{relevant\ pages\} \cap \{retrieved\ pages\}}{\{retrieved\ pages\}}$$

Recall is the negative predictive value which is the fraction of web pages that are relevant to the query that are successfully retrieved.

$$Recall = \frac{\{relevant\ pages\} \cap \{retrieved\ pages\}}{\{relevant\ \ pages\}}$$

Accuracy is the degree of closeness of measurement based on the quality of the system.

$$Accuracy = 100 - \frac{(1 - F_A) + (1 - F_R)}{2}$$

Where F_A is the recall and F_R is the precision. The accuracy of each data set is calculated separately and compared with the existing system and given as a graph in Fig. 8

C. Comparison of performance measures

The proposed Hybrid ACCS was analyzed with Ant colony optimization and cuckoo search algorithm in terms of precision, recall and accuracy measures and the results showed that the proposed hybrid method has outperformed the existing techniques. Table II, III, IV and V shows the comparison of performance measures of Hybrid method with the other two for Data set 1, 2, 3 and 4 respectively.

Table II. Performance Measures Comparison with Existing Optimization Techniques for Data Set- 1

Data Set-1	ACO Algorithm	CS Algorithm	Hybrid ACCS
Precision (%)	97.5	98	99.5
Recall (%)	0.003	0.004	0.005
Accuracy (%)	98	98.5	99.503

Table III. Performance Measures Comparison with Existing Optimization Techniques for Data Set-2

Data Set-2	ACO Algorithm	CS Algorithm	Hybrid ACCS
Precision (%)	98	98.5	100
Recall (%)	0.0006	0.0007	0.0008
Accuracy (%)	96	97	99.501

Table IV. Performance Measures Comparison with Existing Optimization Techniques for Data Set- 3

Data Set-3	ACO Algorithm	CS Algorithm	Hybrid ACCS
Precision (%)	98	98.5	100
Recall (%)	0.0001	0.0002	0.0003
Accuracy (%)	97	97.5	99.501

Table V Performance Measures Comparison with Existing Optimization Techniques for Data Set- 4

Data Set-4	ACO Algorithm	CS Algorithm	Hybrid ACCS
Precision (%)	97.5	98	99.5
Recall (%)	0.004	0.005	0.006
Accuracy (%)	97	98	99.5025

Fig 6, 7 and 8 shows the comparison graph for precision, recall and accuracy values of the optimization techniques ACO, CS and Hybrid ACCS .

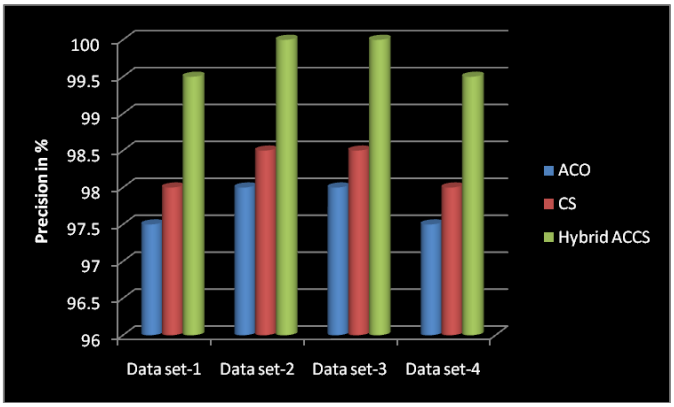

Fig 6. Comparison graph of precision values of ACO, CS and Hybrid ACCS optimization methods

In Fig.6, the comparison between proposed and the existing optimization techniques was depicted based on the precision value. We defined four data set and for each data set the precision value was high compared with the existing one and the proposed hybrid method gave higher percentage for precision. Therefore we conclude that the Hybrid ACCS method is efficient.

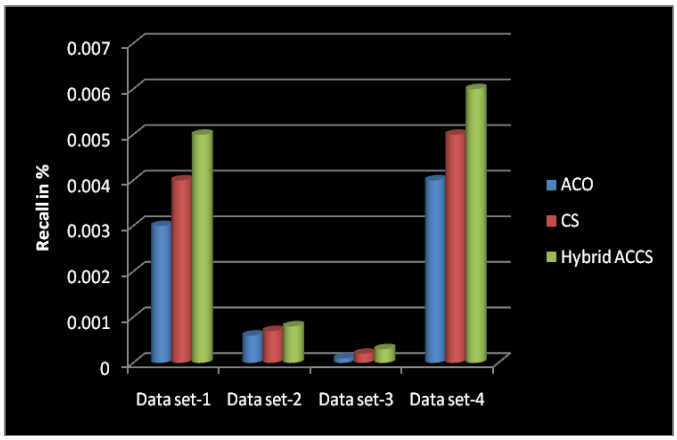

Fig 7. Comparison graph of recall values of ACO, CS and Hybrid ACCS optimization methods

In Fig.7, the comparison between proposed and the existing optimization techniques was depicted based on the recall value. We defined four data set and for each data set the recall value was low compared with the existing one and the proposed hybrid method gave lower percentage for recall. Therefore we conclude that the Hybrid ACCS method is efficient.

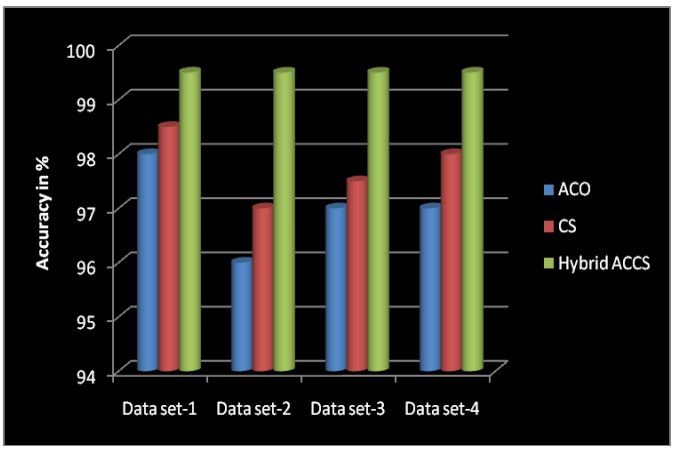

Fig 8. Comparison graph of accuracy values of ACO, CS and Hybrid ACCS optimization methods

In Fig.8, the comparison between proposed and the existing optimization techniques was depicted based on the accuracy. We defined four data set and for each data set the accuracy was high compared with the existing one and the proposed hybrid method gave higher percentage of accuracy. Therefore we conclude that the Hybrid ACCS method is efficient.

Fig.9 shows the performance of the proposed approach on test data.

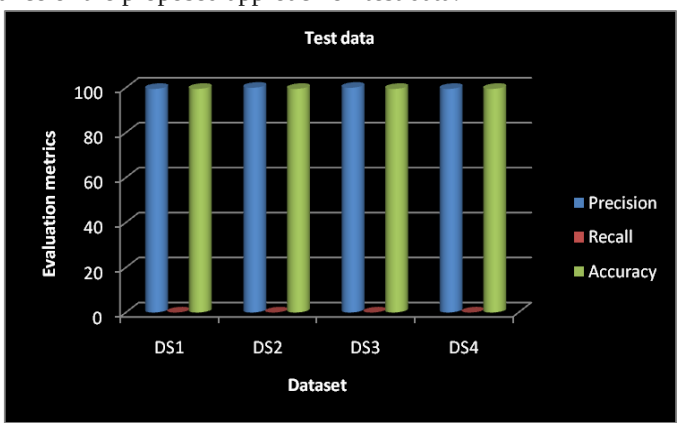

Fig 9. Performance of the proposed approach on test data

We defined four data set and for each data set the precision value was high and the proposed method gave 99 % percentage for precision. Therefore we conclude that the classification done was accurate. For best classification recall value should be low and the recall value extracted is highly accurate. Therefore we conclude that ACCS algorithm performs a better web page classification. The accuracy is calculated based on the precision and recall of classification process. From the graph, we came to know that the proposed system performs well and the classification done was accurate.

V. CONCLUSION

An algorithm for feature extraction and web page classification was proposed with three phases. In the first phase web crawler is used to collect the set of websites, nearly 3356 web pages were collected for classification. In the next phase features are extracted and N-gram entropy model is used for extracting the added features and then the Hybrid ACCS feature selection technique which is the combination of Ant and cuckoo Optimization techniques is introduced for selecting the optimum features. Finally, the Map Reducer acts as classifier for classifying the webpages based on the optimized features of Hybrid ACCS algorithm.. The performance evaluation is done by calculating the precision, recall, accuracy and is compared with the existing one and the simulation results shows that the proposed algorithm achieves high efficiency and less computation time.

REFERENCES

[1] J. Han and M. Kamber, "Data mining: Concepts and techniques", China Machine Press, vol. 8, pp. 3-6 2001.

[2] Nikita Sahu, Dr. R. K. Kapoor, "A Review on Optimization in Web Page Classification", International Journal of Advance Foundation and Research in Computer (IJAFRC) Volume 1, Issue 9, September 2014. ISSN 2348 – 4853.

[3] Raymond Kosala and HendrikBlockeel, "Web Mining Research: A Survey", ACM SIGKDD, Volume 2, Issue 1, pp.1-14, 2000.

[4] Hwanjo Yu, Jiawei Han, and Kevin Chen-Chuan Chang, "PEBL: Web Page Classification without Negative Examples", IEEE transactions on knowledge and data engineering, Vol. 16, pp. 1-12, 2004.

[5] AngelFerran and Sergio Bernabe, "A Web-Based System for Classification of Remote Sensing Data", IEEE journal of selected topics in applied earth observations and remote sensing, Vol. 6, pp. 1934-1948, 2013.

[6] Jun Li, Paolo Gamba, and Antonio Plaza, "A Novel Semi-Supervised Method for Obtaining Finer Resolution Urban Extents Exploiting Coarser Resolution Maps", IEEE Journal of selected topics in applied earth observations and remote sensing, Vol. 7, No. 10, October 2014.

[7] Dou Shen, Zheng Chen and Qiang Yang, "Web-page Classification through Summarization", ACM Digital Library, pp.242-249, 2004.

[8] PrabhjotKaur, "Web Content Classification: A Survey**",** International Journal of Computer Trends and Technology (IJCTT), volume 10, pp.97-101, 2014**.**

[9] Xiaoguang Qi and Brian D. Davison, "Web Page Classification: Features and Algorithms",ACM Computing Surveys, Vol. 41, 2009.

[10] Sven Meyer zuEissen and Benno Stein, "Genre Classification of Web Pages—User Study and Feasibility Analysis", Springer**,** Vol. 3238, pp 256- 269, 2004.

[11] Bhumika, Prof Sukhjit Singh Sehra and Prof AnandNayyar, "A review paper on algorithms used for text classification**",** International Journal of Application or Innovation in Engineering & Management (IJAIEM), Vol.2, pp.90-99, 2013**.**

[12] L. Jayasimman and E.George Dharma Prakash Raj, "A Genetic Optimized Parallel MLP to Improve Classification Accuracy for Web Learning System**",** International Journal of Emerging Trends & Technology in Computer Science (IJETTCS),Vol. 2, pp.315-320, 2013.

[13] Lambodar Jena and Narendra Kumar Kamila, "Data Extraction and Web page Categorization using Text Mining**",** International Journal of Application or Innovation in Engineering & Management (IJAIEM),Vol. 2, pp. 67-76, 2013**.**

[14] PikakshiManchanda, Sonali Gupta and Komal Kumar Bhatia, "On The Automated Classification of Web Pages Using Artificial Neural Network", IOSR Journal of Computer Engineering (IOSRJCE), Vol. 4, PP 20-25, 2012.

[15] SiniShibu, AishwaryaVishwakarma and NiketBhargava, "A combination approach for Web Page Classification using Page Rank and Feature Selection Technique", International Journal of Computer Theory and Engineering, Vol.2, pp.897-900, 2010.

[16] J. AlameluMangai, Dipti D. Kothari and V. Santhosh Kumar**, "**A Novel Approach for Automatic Web Page Classification using Feature Intervals", IJCSI International Journal of Computer Science Issues, Vol. 9, pp.282-287, 2012.

[17] EsraSaraç and Selma AyGeOzel, "An Ant Colony Optimization Based Feature Selection for Web Page Classification**",** Hindawi Publishing Corporation 優e Scientific World Journal, pp.1-16, 2014.

[18] Yang X.-S. and Deb S. "Cuckoo search via Levy flights", World Congress on Nature and Biologically Inspired Computing (NaBIC 2009), IEEE Publication, USA. pp.210–214.

[19] "Novel 'Cuckoo Search Algorithm' Beats Particle Swarm Optimization",http://www.scientificcomputing.com/news-DA-NovelCuckoo- Search-Algorithm-Beats-Particle-SwarmOptimization-060110.aspx.

