Journal of Theoretical and Applied Information Technology
20th August 2014. Vol. 66 No.2
© 2005 - 2014 JATIT & LLS. All rights reserved·
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

SUBJECTIVE SYMMETRICAL SUPPORT VECTOR MACHINE FOR WEB TRAFFIC MINING

[1]S.G.SHRINIVAS, [2]DR. N.M. ELANGO

[1]Asst. Professor, Dept. of Computer Applications, Chettinad College of Engineering & Technology, Karur District, Tamilnadu - 639114, India.

[2]Professor & Head, Dept. of Computer Applications, RMK Engineering College, Tiruvallur District, Tamil Nadu- 601 206, India.

[1]Email: sgshrinivas2014@gmail.com

ABSTRACT

Traffic classification has received significant attention due to the capability of blocking unwanted transfer of complex information. One of the important decisions that have to be made while constructing a classification model is to employ learning approach. Hierarchical Distributed Peer-to-Peer (HP2PC) architecture grouped to form higher level neighborhoods but elaborated technique were not adopted to handle bidirectional traffic and was not extended to dynamic structure. Cluster-Adaptive Distance Bound (CADB) based on separating hyper plane boundaries of Voronoi clusters facilitate resourceful spatial filtering, with relatively small preprocessing storage. The storage overhead was in a way addressed using the Euclidean and Mahalanobis similarity measures but effective attribute selection was not applied to solve the issue related to distance bounds. To overcome the bidirectional traffic problem with dynamic structure Subjective Symmetrical Support Vector Machine (SS-SVM) mechanism is developed. A hybrid attribute selection algorithm is designed which pre-filters (i.e., Classifies) the attributes with improved (refers to subjective symmetry) Support Vector Machine and solve the related distance bounds problem. After classifying the attributes with SS-SVM, the best attributes are further selected and then generates the attribute value. SS-SVM algorithm assigns higher values to the attributes that are capable to generate data report from minority and majority class. Moreover, SS-SVM for flow-based attribute selection in traffic classification is applied. Subjective Symmetrical mechanism experimented with factors such as classification rate, true positive rate, attribute selection efficiency, memory consumption, and report generation effectiveness. Subjective Symmetrical mechanism improves averagely the attribute selection up to 6 % when compared with the state-of-art methods.

Keywords: *Web Traffic classification, Support Vector Machine, Subjective Symmetry, Bidirectional traffic, Hybrid Attribute Selection Algorithm*

1. INTRODUCTION

With the rapid development of Internet, huge amount of information are accessible using World Wide Web (WWW). According to the content required by the user, web mining classifies one or more websites information to provide appropriate information to the user. The web pages are simply the plain text documents that contain the information about the products which is inherently accessed by various customers. To make the customer more satisfied, web classification is performed using hyperlinks and Hyper-Text Markup Language (HTML) labels and from which the context features of web pages are further analyzed.

The web enabled people with different goals and characteristics are the result of an ever-growing amount of information. The web designer and web administrator extracts the information of a website as demonstrated in [3] by measuring the link connections on different webpage. Due to this the weblog files are pre-processed and then a route investigation technique was carried out to inspect the URL information.

On the other hand, web semantic search does not permit the semantic processing of web search queries. The web search queries are examined based on visit count of web pages by the customers. The web pages with respect to keyword return accurately the semantic appropriate pages to a query. But, current standard web search does not

© 2005 - 2014 JATIT & LLS. All rights reserved·
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

estimate complex web search queries which involve certain level of reasoning over the web. Fast Nearest Neighbor Search with Keywords as illustrated in [6] constructed a new access method called the spatial inverted index for broadening the predictable inverted index to manage multidimensional types of data.

Probabilistic threshold keyword queries (PrTKQ) as described in [9] with XML information symbolize the results with the probable globe semantics. PrTKQ designed a probabilistic inverted (PI) index which was used to rapidly revisit the qualified answers and filter out the unnecessary ones based on the lower and upper bounds. Forum Crawler under Supervision (FoCUS), a supervised web-scale forum crawler as described in [12] designed an analogous implicit navigation route. The routes connected by precise URL further helped the users from entry pages to thread pages. But to an extent, FoCUS failed to identify fresh threads and avoided the crawled threads in a well-timed manner.

In the development process of web mining developing, the research of modeling and prediction for web traffic is all the way along listen by the people. The web traffic model concerns with the assessment of network performance analysis and designing. With high-quality model and predictive methods more important for web protocol designing, web management diagnose, high-powered web devices designing and web Qos enhancing for the next generation network. The continuous scale-up of the web traffic amplify different kinds of the services on the network. The web traffic is regularly analyzed on the basis of time series measured at regular intervals in addition to data obtained from the hour, month, minute and second. Web traffic in actual environment is examined according to the similarity of user query and therefore a model of precise prediction is now remains the focus for web traffic research.

A particular problem related to web mining is to categorize the documents into a set of user defined categories based on the web content. Web traffic mining introduced the Support Vector Machine (SVM) in detailed with vector dimension where the vector dimensions included distinct number of keywords. Support Vector Machine (SVM) is an influential classification method which includes classification in practice. However, the dimensionality is condensed during feature extraction algorithms. An SVM model was constructed based on the extracted features from the training data set resulting in significant computational difficulty.

Correct recognition of aliases for person name with binary code is highly required for different types of web related tasks. Robust alias detection system as constructed in [5] used the lexical pattern based approach to professionally extract out a huge set of candidate aliases from the patterns that were leftover. The retrieved patterns from web search engine described numerous ranking scores for different types of candidate aliases.

Taxonomy-Aware Catalog Integration (TACI) processing as described in [10] regulated the outcome of text-based classifier to make the best classification of the products. The products were very much secured together in the supplier taxonomy which remained close in the main taxonomy. But TACI failed to be familiar with candidate products for labeling different products. Three important factors that have been examined in present experiential studies [15], included gender differences, preceding information, and cognitive styles. These applications are at the same time worn by varied population of users with mixed backgrounds, in terms of their information, skills, and necessities but reliability and validity of such questionnaires was not further examined.

Adaptive Distance Bound (CADB) based on separating hyper plane boundaries in [2] facilitated resourceful spatial filtering, with a relatively small preprocessing storage. Many indexes have index specific free parameters which were tuned according to different performances. The storage overhead was made appropriate using Euclidean and Mahalanob similarity measures but effective attribute selection was not directed to solve the related distance bounds. Attribute based encryption systems as demonstrated in [18] applied superior cryptosystem for data sharing. But at the same time, the attribute based encryption system failed to encrypt the multimedia content and does not solve problems related to fully distributed approach to update the user secret key without revealing the user attribute Information.

Probabilistic spatiotemporal model for target event as demonstrated in [11] identified the midpoint of the event location but still advanced algorithms were not developed for query development. With the Twitter user being observed as a sensor and every tweet as sensory information, these virtual sensors were assigned as social sensors with different types of characteristics. The

© 2005 - 2014 JATIT & LLS. All rights reserved·
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

clustering-based pre-fetching system as illustrated in [7] included graph-based clustering algorithm with ranking score to evaluate the user requirements. The algorithm recognized clusters of connected web pages based on the users' prototype. The scheme was then incorporated easily into a web proxy server for improving its performance but the added similarity metrics were not further used for directed test generation.

The sentinel mining problem by bitmap operations as demonstrated in [20] uses the bitmapped encoding named signal streams. Sentinel level is extremely imaginative for large datasets and also found only if the information contains the statistically important associations. Multi-dimensional environment by having a sentinel mining does not robust the aggregation level on dimensions as well as the location and shape of the data area is not chosen.

HP2PC partitions the clustering problem in a modular way which transverse through the neighborhoods, and resolved every element separately using distributed K-means variant where the K-means consecutively combined the clustering up form with hierarchy level. Hierarchical Distributed Peer-to-Peer (HP2PC) architecture as described in [1] grouped together for outlining the advanced level neighborhoods. But elaborated technique was not developed to handle bidirectional traffic and also failed to extend a dynamic structure.

Precise categorization of traffic flow according to the application type is extensively required for IP network management and security monitoring. For example, it can assist Internet Service Providers (ISPs) and network administrators to appreciate the traffic work of art and prioritize certain level of bandwidth-sensitive traffic such as Voice over IP (VoIP) and video conferencing. Traffic classification is also functional for blocking unwanted or attack transfer for complex security. In order to provide solutions to the class imbalance problem while traffic classification, in this work a method called Subjective Symmetrical Support Vector Machine (SS-SVM) is developed. The attribute classification in SS-SVM selects the best attributes and then generates the attribute value. Further, it obtains the input data and performs the classification based on SS where the SS selects the highest value of attributes based on the improved SVM. The Hybrid Attribute Selection algorithm in SS-SVM works with dynamic generation of report where applies flow-based attribute selection during classification of traffic. Subjective Symmetrical Support Vector Machine also helps to handle the bidirectional traffic for different bounds of web traffic mining.

The structure of the paper is as follows. In Section 1, the web traffic mining and their drawbacks is described. In Section 2, the Subjective Symmetrical Support Vector Machine to handle the bidirectional traffic on different environment in web is demonstrated. Section 3 the experimental performance of SS-SVM mechanism with input data is explained. Section 4 provides table and graph based results and Section 5 illustrates the related work. Finally Section 6 offers the conclusion.

2. SUBJECTIVE SYMMETRICAL SUPPORT VECTOR MACHING FOR WEB TRAFFIC MINING

Subjective Symmetrical Support Vector Machine main objective is to solve the bidirectional traffic based on the Hybrid Attribute Selection Algorithm. Attribute selection metric in SS-SVM mechanism are used to get better feasibility result of machine learning based on traffic categorization methods as shown in Figure 1.

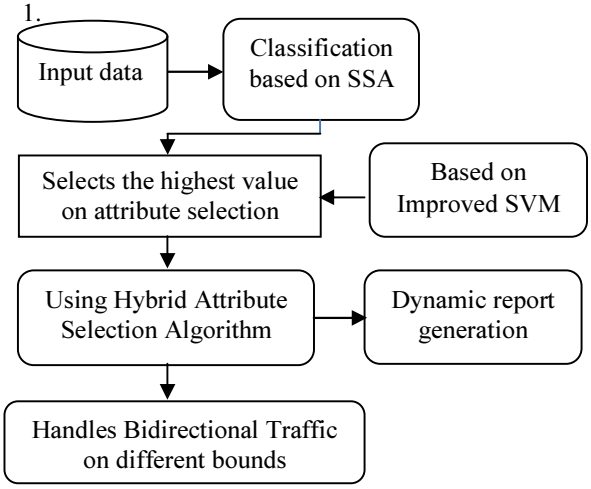

Figure 1 Architecture diagram of SS-SVM

Figure 1 describes the Subjective Symmetrical SVM using the input excel sheet data. Initially, SS-SVM classifies the attribute from the excel sheet based on the improved support vector machine. The classified attribute uses the Hybrid Attribute Selection Algorithm to handle bidirectional traffic with dynamic structure. Once the removal of bidirectional traffic is accomplished, SS-SVM selects the highest value during attribute selection.

© 2005 - 2014 JATIT & LLS. All rights reserved·
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Using the highest attributes value retrieved a dynamic report is generated the bidirectional traffic is handled for different bounds.

Moreover, in order to select a robust and stable attributes from input excel data sheet, SS-SVM design algorithm is developed. As a result, the SS-SVM algorithm is capable of selecting robust attributes from the results obtained by improved SVM. Furthermore, SS-SVM evaluates the selected attributes from three aspects using the occurrence frequencies of the selected attributes, the variance and mean of metric values of the selected attributes. Considering the impact of group inequity (i.e.,) minority and majority, the selected attributes are presented and generates the report. Experimental results show that three widespread flow based attributes selected from completely different traces have marked as a discriminative power.

2.1 SS-SVM Formulation Procedure

To address the group inequality problem, Subjective Symmetrical model is employed based on loaded entropy with the entropy obtained using prior probability distribution from input data source information. However, SS-SVM subjective entropy measures the information using both the probabilities of elementary events (i.e.,) majority and minority with a probabilistic experiment. The loaded qualitative data express the importance of one event with group attribute G, then the total sample of real data R with the loaded value is calculated as,

$$l_i = 1 - \frac{r_i}{R} \quad \text{..... Eqn (1)}$$

wherer_i is the number of real data samples obtained from the input data sheet that is assigned to the groupg_i. The loaded entropy of an attribute 'A' is then defined as

$$SSA_l(A) = -\sum_i \sum_j l_i SVM(g_i, a_j) log_2 SVM(a_j) \quad \text{..... Eqn (2)}$$

Eqn (2) describes the loaded entropy of attribute 'A' with an improved support vector machine form where the SVM solve the related distance bounds on loaded attribute. The loaded conditional entropy of G for attribute 'A' is defined for selecting the attribute from input data sheet as,

$$SSA_l(G/A) = -\sum_i (l_i SVM(g_i, a_j) log_2 SVM(g_i, a_j) \sum_j SVM(a_j) \quad \text{..... Eqn (3)}$$

$SSA_l(G/A)$ select the attribute from the classified group using SS-SVM loaded conditional entropy. With the Eqn (2) and (3) combined together attain the highest value for the attribute being selected and accordingly generates the report.

$$AVG = SSA_l(G/A) - SSA_l(A) \quad \text{..... Eqn (4)}$$

Eqn (4) denotes Attribute Value Generation (AVG), for the loaded conditional entropy of G after observing whether the attribute 'A' is greater than the loaded entropy of an attribute 'A'. SS-SVM actually makes use of the input data information of group distribution while scoring each potential attribute value.

2.2 Improved SVM based Attribute Selection & Report Generation

Once the classification of most of the attributes with SS-SVM metric is accomplished, it is necessary to select the optimal attributes. The selection of optimal attributes with the help of a specific classifier achieves best performance on the real worlds input data sheet. The characteristics of attribute selection are described as search association, generation of report and assessment measure. Improved SVM dominate the computation and solves the distance bound problem. The transformation of attributes brings several benefits such as validated report generation in SS-SVM, so that the true positive rate is improved. In improved SVM, the evaluation of distance boundary function is described as given below

$$SVM\ (\alpha) - \sum_{i=1}^{l} \alpha_i + \frac{1}{2} \sum_{i=1}^{l} \sum_{j=1}^{l} a_i a_j \alpha_i\, \alpha_j k(g_i, g_j) \quad \text{..... Eqn (5)}$$

Support vector machine $\frac{1}{2} \sum_{i=1}^{l} \sum_{j=1}^{l} a_i a_j \alpha_i\, \alpha_j$ calculates half the distance bound of the overall area and αdenotes the vector iterations for generating the value of attribute. The improved support vector machine implies that the classifier achieves the best performance and highest value on imbalanced data. The input real world data evaluates the classifier without making any specific assumption for group distribution. For this purpose, Improved Support Vector Machine is used to solve the related distance bounds.SS-SVM mechanism categorizes the distance boundary function using Improved SVM form to maximize the attribute value generation. The stepwise attribute selection in SS-SVM is shown in Fig 2.

Journal of Theoretical and Applied Information Technology
20th August 2014. Vol. 66 No.2
© 2005 - 2014 JATIT & LLS. All rights reserved·
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

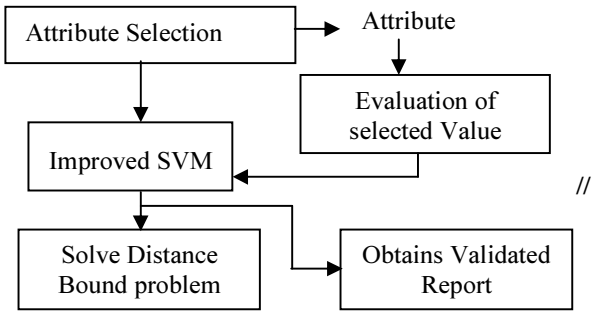

Figure 2: Steps on Attribute Selection in SS-SVM

An additional feature of improved SVM is that its distance boundary function is described by real input data closest to the SS-SVM solution point. Using Eqn (5), the vector α is computed for each element which specifies the load of each data. The improved SVM follows the data whose corresponding α is greater than zero. The other real world input data selects the optimal attribute distance function and evaluates the improved SVM boundary function to attain the best performance.

2.3 Hybrid Attribute Selection Algorithm

SS-SVM classification based on attributes, SS-SVM attribute selection, value generation and report generation are combined together in Hybrid attribute selection algorithm, to achieve the robust result while evaluating the data report. As the dimension of attributes for classifying traffic is always high, SS-SVM classifies most of the attributes with improved SVM metric. Improved SVM identifies an attribute set where the attributes are highly associated with a group but not highly associated with any of the other group attributes. After classification, redundant and irrelevant attributes are sorted out based on the values achieved for each attribute classification subsets and then it determines the best attribute subset that produces the highest value. The hybrid attribute selection algorithm is described below.

Begin Loop

Input: Excel sheet data row (r) * column (c) data separated by comma (,)
Output: Report Generated 'RD (i)'

// SS-SVM Classification Based on attributes

Step 1: For each c_j inr_i
Step 2: Find Non duplicate value DD (r) ($c_j$in $r_1, r_2, r_{3\ldots\ldots} r_i$)
Step 3: Find DD(r) support count in ($c_j$in $r_1, r_2, r_{3\ldots\ldots} r_i$)
Step 4:Find improved SVM (r_i , c_j) from each c_j values support count for $r_1, r_2, r_{3\ldots\ldots} r_i$

// SS-SVM Attribute Selection

Step 5:Find Attributes heading AT (j) from (improved SVM (r_i , c_j))
Step 6:List out all AT (j) to user for constraints based report
Step 7:Collect User Selected Attributes USA (k) from AT (j)
Step 8:Stored USA (k) and collect values from (improved SVM (ri, cj))

//SS-SVM Attribute value generation

Step 9: Get DD(r) for each USA (k)

```
for (int g0=0,g0<k;g0++)
{
for (int g1=0;g1<r;g1)
  {
    Value User Selected Attributes
    (VUSA)(a,b)=DD(gg0,gg1)
  }
}
```

//SS-SVM Report generation

Step 10: Apply each VUSA on attribute a_i
Step 11: Generate count (VUSA) level
Step 12: No of level=count (VUSA (a,b))
Step13:ReportRD (i) generated: Improved SVM (VUSA (a,b) fromattribute a_i)

End Loop

In the above algorithmic step, most of the attributes are classified with loaded value which is shown in the Eqn (2) and (3). The algorithm first evaluates the value of loaded entropy between each attribute and the group. The higher value with speed up attributes selection process improves the classification rate.

In the second step, the SS-SVM algorithm selects the best attributes with improved SVM metric for a specific classifier. It extracts the attributes from the input list one by one and finds a locally optimal attribute subset that produces the highest improved SVM value. Although SS-SVM presents an approach that handles group balancing in traffic flows, then the selected value of the attributes generate the report. The generated report is different due to different distribution of group on the real world input data sheet, which decides the attribute taken into practice.

© 2005 - 2014 JATIT & LL
ISSN: 1992-8645 www.jatit.o

3. EXPERIMENTAL EVALUATION OF SS-SVM WITH IMRPOVED SVM

Performance experiments are conducted with various conditions using JAVA platform. Initially, real world data sheet is used to generate dynamic report. The report generation using the JAVA programming, first classifies the attributes, selects the attribute and generates the value. After value generation, reports are generated from excel sheet data. For instance, input excel sheet is shown below

At1	At2	At3	At4	At5	At6	At7	At8	At9	At10
1	A1	Google	Samsung	BMW	X1	H1	P1	0.1	2004
2	B1	Facebook	Apple	Ford	Y2	H3	P1	0.1	2004
1	A1	Facebook	Sony	Honda	X2	H1	P1	0.3	2004
3	C1	Yahoo	Sony	Benz	X3	H4	P1	0.1	2004
1	A1	Facebook	Sony	Ford	Y2	H3	P1	0.1	2004

The above displayed excel sheet is a sample example used to demonstrate the Subjective Symmetrical (SS-SVM) mechanism with Improved Support Vector Machine. Subjective Symmetrical (SS-SVM) mechanism is compared against the hierarchically distributed Peer-to-Peer (HP2PC) architecture and Cluster-Adaptive Distance Bound (CADB). The performance factors such as classification rate, true positive rate, running time, attribute selection efficiency memory consumption and the effectiveness of report generation are taken for experimental evaluation.

The classification rate in SS-SVM refers to the categorization of attribute based on the highest value performed using improved SVM and measured in terms of percentage (%).The true events occurred on removing the web traffic is termed as the true positive rate and is evaluated as given below,

$$True\ Positive\ Rate = \frac{True\ Positive\ \ report\ generation\ ratio}{True\ Positive\ ratio + False\ Negative\ ratio}$$

True positive rate is the ratio of true positive report generation to the sum of true positive and false negative ratio. The running time measures the time taken to perform the overall operation to generate the report using the web information.

$$Runtime = Instruction * clock\ (seconds)$$

The Instructions of the web users vary depending on the needs and requirements of the user. The attribute selection is the most significant part in web traffic analysis and measured in terms of percentage. The attribute selection efficiency calculates a score value for each attribute, and then selects only the attributes that have the best scores for bidirectional traffic handling. The memory consumption is the power consumed for recalling or generating the report to the users query through associative mechanisms. The memory consumption in SS-SVM mechanism is reduced and measured in terms of Kilo Bytes (KB).

$$Report\ Generation\ Effeciency = \frac{Cloumn\ count * runtime}{No.of\ web\ users}$$

Report generation in SS-SVM is performed using both simple and complex information and is highly capable of generating effectual report. The web users are taken as '100' in experimental work. Consumers report is viewed by end users in order to read and retrieve data. Web based product report are the output format that is displayed in a human or machine-readable report. Overall Report generation efficiency is measured in terms of percentage (%).

4. RESULT ANALYSIS

Subjective Symmetrical Support Vector Machine (SS-SVM) is compared against the Hierarchical Distributed Peer-to-Peer (HP2PC) and Cluster-Adaptive Distance Bound (CADB).The table given below (table 1) shows the experimental values and graph illustrates the effective handling of SS-SVM mechanism when compared with the HP2PC process [1] and CADB mechanism [2].

Information Size (KB)	Classification Rate (%)		
	HP2PC	CADB	SS-SVM
500	49	51	55
1000	56	58	61
1500	65	69	73
2000	67	73	79
2500	71	80	85
3000	72	82	87
3500	76	85	90

Table 1 Tabulation of Classification Rate

Table 1 describes the classification rate based on the information size. The information is measured in Kilo Bytes (KB). As the information size increased, the classification rate is also increased gradually.

Journal of Theoretical and Applied Information Technology
20th August 2014. Vol. 66 No.2
© 2005 - 2014 JATIT & LLS. All rights reserved·
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

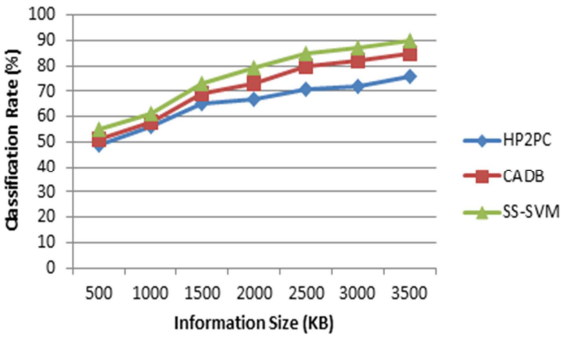

Figure 3 Measure of Classification Rate

Figure 3 illustrates the classification rate based on the size of information which ranges from 500 KB to 3500 KB. As the dimension of attributes for traffic classification is high in SS-SVM, the method, SS-SVM classifies most of the attributes with improved SVM metric obtained from Attribute Value Generation by solving the related distance bound. This is turn attains higher classification rate. Improved SVM find an attribute set in which attributes are highly associated to the group and improves classification by 8 – 20 % when compared with the HP2PC [1] mechanism and 5 – 8 % when compared with the CADB [2].

Table 2 Tabulation for True Positive Rate

Dataset Index	True Positive Rate (%)		
	HP2PC	CADB	SS-SVM
1	65	71	75
2	66	72	78
3	68	73	81
4	73	77	82
5	76	82	87
6	78	85	89
7	79	88	93

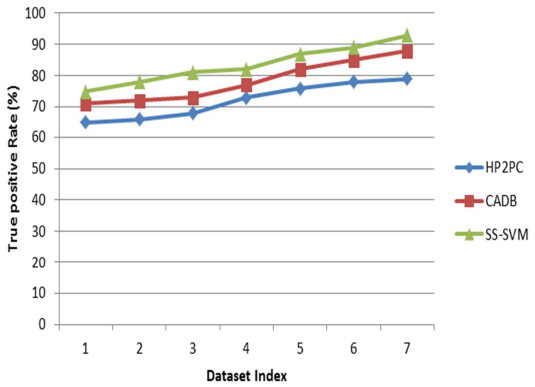

Figure 4 True Positive Rate Measure

The true positive rate as illustrated in figure 4 is measured based on the Dataset index extracted from the data in the excel sheet. With the selection of optimal attributes and with the transformation of attributes into the validated report, SS-SVM improves the true positive rate. Improved SVM dominate the computation by solving the distance bound problem and attains improved true positive rate by 12 -19 % when compared with the HP2PC [1] and 5 – 10 % improved when compared to the CADB system [2].

Table 3 Tabulation of Runtime

No. of instructions	Runtime (sec)		
	HP2PC	CADB	SS-SVM
100	120	100	90
200	280	250	220
300	600	510	450
400	840	780	720
500	1240	1160	1050
600	1560	1450	1350
700	1950	1790	1680

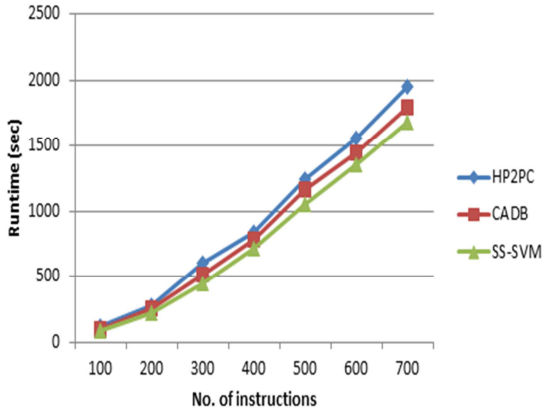

Figure 5 Measure of Runtime

Figure 5 describes the runtime based on the instruction count given by the web users in the range of 100 to 700. In SS-SVM method, with the introduction of improved Support vector machine, the distance boundary function is computed that calculates half the distance bound of the overall area and easily identifies the result. The vector iterations performed for generating the value for attribute reduces the runtime by 13 – 25 % when compared with the HP2PC [1]. With the introduction of improved support vector machine it implies that the classifier achieves the best performance with 6 – 13 % reduced runtime when compared with the CADB [2].

Journal of Theoretical and Applied Information Technology
20th August 2014. Vol. 66 No.2
© 2005 - 2014 JATIT & LLS. All rights reserved·
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 4 Tabulation Of Attribute Selection Accuracy

Attribute Collection Weight	Attribute Selection Accuracy (success %)		
	HP2PC	CADB	SS-SVM
100	78	82	87
200	79	85	88
300	83	88	93
400	84	89	94
500	85	90	95
600	87	92	96
700	88	93	97

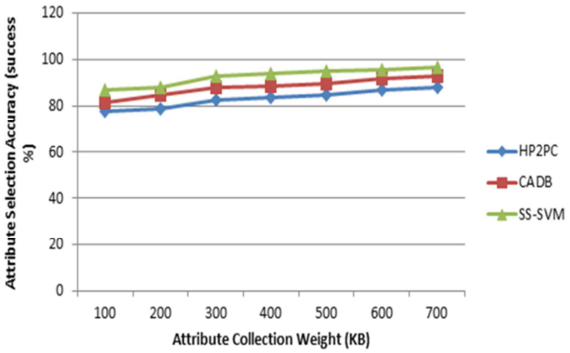

Figure 6 Measure Of Attribute Selection Accuracy

Figure 6 illustrates the attribute selection accuracy based on the weight of the attribute collection. The attribute weight is computed in terms of Kilo Bytes (KB).The attribute selection accuracy is comparatively increased using SS-SVM when compared to the two other methods. This increase in accuracy is due to the fact that with the application of SS-SVM algorithm, it further assigns higher values to the attributes that generate the data report from minority and majority class, present the selected attributes and generates the dynamic report with 10 – 12 % improved result when compared with the HP2PC [1]. Flow based attributes selected from completely different traces have marked has an effective attribute selection by 3 – 6 % when compared with the CADB [2].

Table 5 Tabulation Of Memory Consumption

Record Set	Memory Consumption (KB)		
	HP2PC	CADB	SS-SVM
Record_Set1	104	97	91
Record_Set2	201	193	171
Record_Set3	317	293	257
Record_Set4	436	386	346
Record_Set5	535	404	448
Record_Set6	667	625	590
Record_Set7	838	873	925

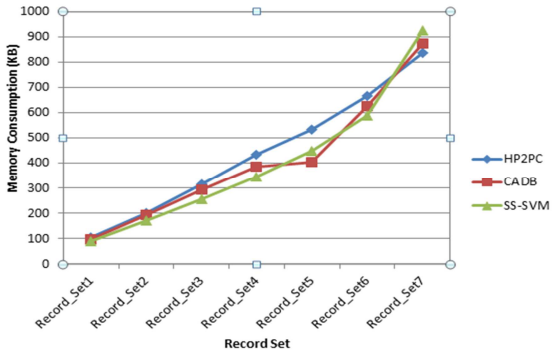

Figure 7 Performance Of Memory Consumption

Figure 7 describes the memory consumption based on the record set count. The Subjective Symmetrical based on loaded entropy obtains the prior probability distribution of input data source. As a result, the memory consumption is reduced by 10- 20 % when compared with the HP2PC [1]. With the flow based attributes in the server port, where the entire number of bytes are sent in the original window with the least amount of section size are only observed reducing the memory consumption by 5 – 12% when compared with the CADB [2].

Table 6 Tabulation For Report Generation Effectiveness

Column Count	Report Generation Effectiveness (result count)		
	HP2PC	CADB	SS-SVM
10	8	8.5	9
20	40	42	44
30	108	131	135
40	235	272	288
50	458	505	525
60	715	769	810
70	1052	1075	1176

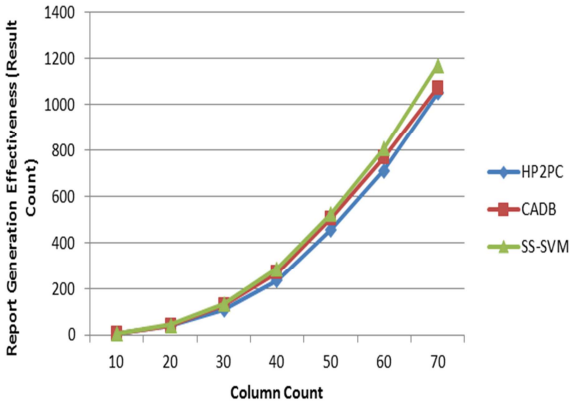

Figure 8 Report Generation Effectiveness Measure

© 2005 - 2014 JATIT & LLS. All rights reserved·
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Figure 8 shows the effectiveness of report generated with respect to the column count. As illustrated in the figure with the increase in column count the report generation effectiveness is also improved. This is because of the fact that only optimal attributes that make the classifier achieve the best performance and also with the application of distance boundary function, the generation of attribute value is maximized which improves the effectiveness of report being generated. The report generation effectiveness is improved by 3 – 9 % when compared with HP2PC [1]. The report generation result is also improved with specified load of data (i.e.,) column count by 10- 25 % when compared with the CADB [2].

Finally, attribute classification in SS-SVM further selects the best attributes and generates the attribute value.SS-SVM mechanism obtains the input data and performs the classification based on the SS. Hybrid Attribute Selection Algorithm works with dynamic generation of report and flow-based attribute selection is performed in SS-SVM classification.

5. RELATED WORK

Web People Search approach with the help of connection analysis in [13] clustered the web pages based on the relationship of different people. Web People searched the information based on semantic information. The semantic data was taken out from web pages, such as name unit, hyperlinks, disambiguate bounded by namesakes on the web pages. The information stored in the top-k web pages were not up to the maximum quality mark. The conceptual prediction model as illustrated in [14] generated a semantic network for the semantic web usage knowledge. Web usage knowledge is the combination of domain knowledge and web usage knowledge but the extreme comparisons on semantic query web-page recommendation systems were not performed.

An automatic annotation approach as described in [17] supports the data units on a consequence page into different groups such that the data in the same group have the same semantic. Subsequently, for each group annotate it forms different features and aggregate the different annotations to forecast an ultimate annotation. But the annotated label failed to integrate other features for expanding the performance result in the wrapper. Semantic Knowledge-Based as presented in [19] demonstrated abstract presentation from the raw real-world information using step by step resources obtained from semantic technologies. The framework triggered the information and replaced the status and examined the agents but failed to apply the approach for more complex scenarios involving other agents. Semantic Knowledge-Based framework does not deal with the acoustic communication limits connected to the submerged environment.

ML-based methodology as shown in [16] built an application capable of recognizing and broadcasting the semantic relations but added source of information were not integrated. Identifying and classifying medical-related information on the web was not effective in providing valuable information to the research community and also to the end user.

The enhanced method as presented in [4] built a Knowledge Base (KB) for the automatic enhancement of the semantic relation network. A rule based method using WordNet's glossaries and a proposition method provided effective web search result. These areas included semantic document indexing, document topic detection, query development, ontology expansion, semantic information retrieval and knowledge integration. Two learning methods as illustrated in [8] discovered the primary relations between images and texts based on small training datasets. Image-Text Associations between the image and textual features routinely sum up the linked features with position of data patterns.

6. CONCLUSION

In this paper, we have proposed a novel method, called SS-SVM classification, for classifying the attributes from the excel sheet based on the improved support vector machine. Furthermore, we have proposed hybrid attribute selection algorithm that classifies most of the attributes using subjective symmetry to solve the related distance bound problems. In SS-SVM method, we first classify the attributes from the excel sheet using improved support vector machine where the classified attributes uses the Hybrid Attribute Selection Algorithm to handle bidirectional traffic with dynamic structure. Then we utilized the highest value during attribute selection to produce a dynamic report in order to handle the bidirectional traffic for different bounds. A series of experiments were conducted to evaluate the performance of the proposed methods. The experimental results show that SS-SVM method achieves high-quality attributes selected with minimum memory consumption and the proposed

© 2005 - 2014 JATIT & LLS. All rights reserved·
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

SS-SVM method obtains highly precise results for web traffic mining. Meanwhile, with the application of hybrid attribute selection algorithm, a dynamic report generation is obtained that reduces the traffic occurrences on the web without modifying the distribution of training data. The results obtained through improved SVM, shows that the SS-SVM build effective mechanism for web traffic mining by handling bidirectional traffic with different patterns. SS-SVM is an efficient identifier in terms of 13-25 % minimal runtime and provides utmost 10-12 % attribute selection accuracy for web traffic mining than the state-of-the-art methods. Experiment demonstrates that the true positive rate and classification rate for web traffic mining are increased and as a result traffic classification rate has been improved.

REFERENCES:

[1] Khaled M. Hammouda. and Mohamed S. Kamel., "Hierarchically Distributed Peer-to-Peer Document Clustering and Cluster Summarization," IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 21, NO. 5, MAY 2009

[2] SharadhRamaswamy., and Kenneth Rose., "Adaptive Cluster Distance Bounding for High-Dimensional Indexing," IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 23, NO. 6, JUNE 2011

[3] Resul Das., Ibrahim Turkoglu., "Creating meaningful data from web logs for improving the impressiveness of a website by using path analysis method," Expert Systems with Applications, Elsevier Journal., 2009

[4] Myunggwon Hwang., Chang Choi., and Pankoo Kim., "Automatic Enrichment of Semantic Relation Network and Its Application to Word Sense Disambiguation," IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 23, NO. 6, JUNE 2011

[5] DanushkaBollegala., Yutaka Matsuo., and Mitsuru Ishizuka., "Automatic Discovery of Personal Name Aliases from the Web.," IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 23, NO. 6, JUNE 2011

[6] Yufei Tao., Cheng Sheng., "Fast Nearest Neighbor Search with Keywords," IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING., 2013

[7] George Pallis ., Athena Vakali ., JaroslavPokorny., "A clustering-based prefetching scheme on a Web cache environment," Elsevier Journal on Computers and Electrical Engineering, 2008

[8] Tao Jiang., and Ah-HweeTan., "Learning Image-Text Associations," IEEE TRANSACTION ON KNOWLEDGE AND DATA ENGINEERING. Volume:21 , Issue: 2, 2009

[9] Jianxin Li., Chengfei Liu., Rui Zhou., and Jeffrey Xu Yu., "Quasi-SLCA based Keyword Query Processing over Probabilistic XML Data," arXiv:1301.2362v1 [cs.DB] 11 Jan 2013

[10]Panagiotis Papadimitriou., Panayiotis Tsaparas, ArielFuxman., and LiseGetoor., "TACI: Taxonomy-Aware Catalog Integration," IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 25, NO. 7, JULY 2013

[11] Takeshi Sakaki, Makoto Okazaki, and Yutaka Matsuo., "Tweet Analysis for Real-Time Event Detection and Earthquake Reporting System Development," IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 25, NO. 4, APRIL 2013

[12] Jingtian Jiang., Xinying Song., Nenghai Yu., and Chin-Yew Lin., "FoCUS: Learning to Crawl Web Forums," IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 25, NO. 6, JUNE 2013

[13] Dmitri V. Kalashnikov., Zhaoqi (Stella) Chen.,SharadMehrotra, and RabiaNuray-Turan., "Web People Search via Connection Analysis," IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 20, NO. 11, NOVEMBER 2008

[14] ThiThanh Sang Nguyen., Hai Yan Lu, Jie Lu., "Web-page Recommendation based on Web Usage and Domain Knowledge," IEEE, 2013

[15] Sherry Y. Chena,b., Robert Macrediea., "Web-based interaction: A review of three important human factors.," Springer International Journal of Information Management., 2010

[16] OanaFrunza., Diana Inkpen., and Thomas Tran., "A Machine Learning Approach for Identifying Disease-Treatment Relations in Short Texts," IEEE TRANSACTIONS O``N KNOWLEDGE AND DATA ENGINEERING, VOL. 23, NO. 6, JUNE 2011

© 2005 - 2014 JATIT & LLS. All rights reserved·
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[17]Yiyao Lu., Hai He., Hongkun Zhao., WeiyiMeng., and Clement Yu., "Annotating Search Results from Web Databases," IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 25, NO. 3, MARCH 2013

[18]M. Pratheepa., R. Bharathi., "Improving Security and Efficiency in Attribute Based Data Sharing," International Journal of Science and Research (IJSR) ISSN (Online): 2319-7064., 2014

[19] Emilio Miguelanez., Pedro Patron., Keith E. Brown., Yvan R. Petillot., and David M. Lane., "Semantic Knowledge-Based Framework to Improve the Situation Awareness of Autonomous Underwater Vehicles," IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 23, NO. 5, MAY 2011

[20] Morten Middelfart., Torben Bach Pedersen., and Jan Krogsgaard., "Efficient Sentinel Mining Using Bitmaps on Modern Processors," IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING., 2013