Volume 5, Issue 11, November 2015 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

TSVM Approach for Classification of the Behavior Profiles of Customers in Ecommerce

B. Uma Maheswari
Research Scholar in Bharathiyar University,
Coimbatore, Tamil Nadu, India

Dr. P. Sumathi
Asst. Professor, Govt. Arts College,
Coimbatore, Tamil Nadu, India

Abstract-The Web mining field encompasses a wide array of issues, primarily aimed at deriving actionable knowledge from the Web, and includes researchers from information retrieval, database technologies, and artificial intelligence. Most data used for mining is collected from Web servers, clients, proxy servers, or server databases, all of which generate noisy data. Predicting the behaviour of customers is challenging, but important for service oriented businesses. Data mining techniques are used to make such predictions, typically using only recent static data. In this paper, behavior based classification approach using Transductive Support Vector Machines are employed. In this research work, in numerous applications, labeling instances might cost huge human reports, while huge amounts of unlabeled information are often willingly available offering some further information. To address the problem Transductive Support Vector Machines (TSVM) is established with only a small division of the labels are available to the customer. TSVM suggest best result when compared with SVM. This research approach has been tested on data collected from the live website, where a summary is obtained and analysis of accuracy and execution time demonstrates the proposed TSVM method yields better results.

Keywords-Web usage mining, customer behaviour, classifier, support vector machine, transductive support vector machine.

I. INTRODUCTION

Web has newly become a dominant platform for, not only retrieving information, but also discovering knowledge from web data. Web mining can be defined as the application of data mining techniques to the web related data. Many webs mining algorithm is available to retrieve the webpages. Web Usage Mining consists of three main steps: data preprocessing, knowledge extraction, and results analysis. Raw data is highly susceptible to noise, missing values. The quality affects the data mining results. In order to improve the data quality, that the data is preprocessed. In order to collect the data for preprocessing, much research has been done so far, e.g. the cookies or the remote agent recognize the user session can do help to user identification, session identification and path completion.

A fundamental component in search engine personalization is a good user profiling strategy. Several personalization methods is related on the construction of one single profile for a user and user have to indicate his search interest in that and based on that interest, the results will be retrieved. If search interest changes, have to update that manually. Different queries from the user should be handled differently because a user's preferences may change across queries. Based on the changing behavior of the user, profiles should be updated automatically.

The problem is that it requires too much insight understanding and effort from the user to recognize the opportunity to employ an agent take the initiative to create it supply the agent with explicit knowledge and maintain the underlying rules or scripts over time. The basic assumption in this technique is that the data's are grouped from the corresponding environment can be converted into a sequence of events. But in traditional approaches are not evaluating the abnormal user behaviors with time varying queries. To address this problem introducing an approach called Transductive Support vector machine.

SVM is a new erudition technique developed in recent years based on the foundations of statistical learning theory. By taking a Transductive approach instead of an inductive one in support vector classifiers, the working set can be used as an additional source of information about margins. Compare with traditional inductive support vector machines, TSVM is often more powerful and may give better performance. In transduction, one estimates the classification function at points within the working set using information from both the training and the working set data. This will help to improve the generalization performance of SVMs, especially when training data are inadequate

This paper is summarized as follows: an overview of related work is given in Section 2. The main performance of the proposed technique is provided in Section 3. Experimental results are discussed in section 4. Conclusion and future work provided in section 5.

II. LITERATURE SURVEY

Kalelkar et al (2014), Their study is an attempt to recognized the actives of the main factors that consumers seem into any of these e-tailors and be familiar with the chief measurement that assists in fastening the consumers with these portals. Chen et al (2014) explores the roles of routine reconfiguration in traditional companies' e-commerce strategy implementation. Depicting on the conceptualization of "routine as trajectory", in which routines are viewed as routes of interdependent actions through which organizations accomplishes to examine the interacts between components of a trajectory. Kondratovich et al (2013) conveys that Transductive Support Vector Machines a semi-supervised large-margin categorization technique. Srivastava et al (2010) provides an original knowledge technique, SVM is useful on dissimilar information (Diabetes data, Heart Data, Satellite Data and Shuttle data) which have two or multi class. Li et al (2010) propose an original two-view transductive SVM that obtain advantages of both the abundant quantity of unlabeled information and their multiple representations to get better the presentation of classifiers.

Bottou et al (2010), related a technique for training a transductive support vector machine. The support vector machine is sophisticated based on labeled training data and unlabeled test data. SVMs are particularly interesting in the remote sensing field due to their capability to simplify well even with restricted training samples, a general drawback for remote sensing applications are given by Mountrakis et al (2011) and a similar work done by Yu et al (2005) to identify the buying interest of the visitor. Carneiro et al (2014) proposed an electronic commerce Website is successful if it achieves the purpose for which it has been created. The analysis is conducted using web access logs and sales data. Pavlou et al (2003), their plans to calculate consumer acceptance of e-commerce through proposing a set of key drivers for appealing consumers in online operations.

Liu et al (2000), the research framework was derivative from information classifications and marketing journalism. Webmasters from Fortune 1000 organizations were used as the target group for a survey. A study of the data offers expensive management connotations for Web site victory in the context of e-commerce. Stoll et al (2013) suggested a creative approach to this difficulty, which sort outs characteristics generated from HTML tag attributes with an e-commerce specific white list. They estimate their findings on an autonomous dataset and on orientation shop sites. Di et al (2014), their research associated to e-commerce product image has typically focused on excellence at perceptual level, but not the excellence of content, and the method of presentation. Profiling the behavior of programs can be a useful reference for detecting potential intrusions against systems. Ghosh et al (1999) presented three anomaly detection techniques for profiling program behavior that evolve from memorization to generalization. Thus, it is essential to take into account these changes in any behavior recognition system. A general approach to the classification of streaming data which represents a specific agent behavior based on evolving systems are given by Iglesias et al (2010).

III. RESEARCH METHODOLOGY

Electronic commerce refers to the buying and selling of information, products and services through computer network. This section introduces the proposed approach for classification of the behavior profiles of customers. Due to changing in the customer's behaviour, it is not rigid. But this proposed approach, can be applied to all behavior which is represented by event sequence, this chapter may use a command-line interface environment. The classifier of SVM and TSVM is used to evaluate the performance of the customer.

A. Support Vector Machine

Support Vector Machines (SVMs) have urbanized from Statistical Learning Theory. SVMs purpose through nonlinearly analytical of the training data in the put in space to a characteristic liberty of superior dimension as a result of a kernel function. Every function has exclusive parameters that have to be resolute proceeding to classification and they are also frequently determined through a cross validation process. By nature SVMs are essentially binary classifiers, conversely, there exist strategies by which they can be modified to multiclass assignments. Two special properties of SVM are- (1) it attain elevated simplification by maximizing the margin and (2) support an capable purchasing of non-linear functions by kernel trick. SVMs were originally developed for classification and have been extensive for regression and favorite purchasing. The original form of SVMs is a binary classifier where the production of educated function is either positive or negative. SVM are adopted respectively to construct the classification behavior model of logistics quality and service satisfaction of e-commerce. Using SVM, each principle component and the entire logistics satisfaction item of the auction buyer is further associated to become the sample data of constructing classification predict model of service satisfaction. From these description, this work know that SVM uses pivot loss function, so this algorithm is less robust, in other words, sensitive to noises. This is the common drawback of SVM method using hinge loss function. In order to avoid this drawback, this research considers TSVM idea to improve the algorithm by using to modify error variables.

B. Proposed Transductive Support Vector Machine based Classification

Presently, the major significant work of transductive deduction in the field of support vector purchasing is Transductive Support Vector Machine (TSVM), which will be briefly discussed in the remainder of this division .Given a set of autonomous, identically distributed labeled examples

$$(x_1, y_1), \dots, (x_n, y_n), x_i \in R^m, y_i \in \{-1, +1\} \qquad (1)$$

and a further set of unlabeled examples from the similar sharing,

$$x_1^*, x_2^*, x_3^*, \dots, x_k^* \qquad (2)$$

In a common linearly non-separable information case, the education procedure of Transductive SVM can be formulated as the following optimization trouble:

$$(y_1^*, \dots, y_k^*, w, b, \xi_1, \dots, \xi_n, \xi_1^*, \dots, \xi_k^*) \qquad (3)$$

$$\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}\xi_i + C^*\sum_{j=1}^{k}\xi_j^* \tag{4}$$

Subject to:

$$\begin{gathered}\forall_{i=1}^{n}: y_i t[w.x_i + b] \geq 1 - \xi_i \\ \forall_{j=1}^{k}: y_j{}^*\left[w.x_j^* + b\right] \geq 1 - \xi_j{}^* \\ \forall_{i=1}^{n}: \xi_i \geq 0 \\ \forall_{j=1}^{k}: \xi_i{}^* \geq 0\end{gathered} \tag{5}$$

Training Algorithm TSVM:
Input: -training examples ($\vec{x}$1, y1),…., ($\vec{x}$n,yn)
-test examples $\vec{x}^*_1,\ldots,\vec{x}^*_k$
Parameters: -C,C*: parameters from OP(2)
-num+: number of test examples to be assigned to class +
Output: -predicted labels of the test examples $\vec{y}^*_1$- $\vec{y}^*_k$ ($\vec{w}$,b, $\vec{\xi}$, _) :=solve_ svm_qp([($\vec{x}$1, y1)…($\vec{x}$n, yn)], [],C, 0, 0);
Classify the test examples using < ~w; b >. The num+ test examples with the highest value of $\vec{w} * \vec{x}_j^* + b$ are assigned to the class + ($\vec{y}_j^*$:= 1);
the remaining test examples are assigned to class - ($\vec{y}_j^*$:= -1).
C_-^*:= 10-5; // some small number
C_-^*:= 10-5 $* \frac{\text{num}_+}{\text{k}-\text{num}_+}$;
while (($C_-^* < C^*$) ‖ ($C_-^* < C^*$)) {
// Loop1
($\vec{w}, b, \vec{\xi}, \vec{\xi}$*):= solve_sum_qp ([$\vec{x}$1, y1)… ($\vec{x}$n, yn)],[($\vec{x}^*_1$, $\vec{y}^*_1$),($\vec{x}^*_k$, $\vec{y}^*_k$)],C, C_-^*, C_+^*);
while ($\exists$m, l: (y_m^* * y_1^*<0)&(ξ_m^* >0)&(ξ_1^*>0)&(ξ_m^* *ξ_1^*>2)) { // Loop2
y_m^* := -y_m^*; // take a positive and a negative test
y_1^*:= -y_1^*; // example, switch their labels, and retrain
($\vec{w}, b, \vec{\xi}, \vec{\xi}$*):= solve_sum_qp ([$\vec{x}$1, y1)… ($\vec{x}$n, yn)],[($\vec{x}^*_1$, $\vec{y}^*_1$),($\vec{x}^*_k$, $\vec{y}^*_k$)],C, C_-^*, C_+^*);
}
C_-^*:= min(C_-^* *2, C*);
C_+^*:= min(C_+^* *2, C*);
}
return($\vec{y}^*_1$-$\vec{y}^*_k$);

Where C and C* are user-specified parameters. C* is called the ‘‘effect factor’’ of the unlabeled examples and C*ξ_i^* is called the ‘‘effect term’’ of the ith unlabeled instance in the objective function. Training the TSVM quantity to solving procedure of the above optimization difficulty. TSVM training algorithm can be approximately outlined as the subsequent steps:

Step 1**:** Identify the parameter C and C*, carry out an initial learning with inductive learning using all labeled instance, and construct an original classifier. Identify a number N as the predictable amount of the positive-labeled examples in the unlabeled examples.

Step **2:** Calculate the decision function charges of all the unlabeled illustration with the original classifier. Label N examples with the major judgment function values as constructive, and the others as negative. Set a provisional achieve factor C_{tmp}^* .

Step 3: Retrain the support vector machine overall the examples. For the recently yielded classifier, change labels of one couple of different-labeled unlabeled illustration using a certain regulation to build the value of the objective function diminish as greatly as feasible. This step is repeated until no pair of examples satisfying the switching situation is established.

Step 4**:** Enlarge the value of C_{tmp}^* slightly and return to Step 3. When $C_{tmp}^* \geq C^*$, the algorithm is completed and the result is output.

The marker switching procedure in Step 3 assurance that the objective function will diminish after switching. The successive enlarge of the provisional achieve factor in Step 4 tries to pursue a practical error manage by calculation the effect of the unlabeled examples on the objective function little by little. The algorithm can conclude after restricted iterations because C* specified in Step 1 is a restricted quantity.

IV. EXPERIMENTAL RESULTS

In this section describe the efficient results of proposed TSVM method. It evaluates the concurrent results for every customers present in the data base. Experimental result is implemented by using MATLAB.

Table 1: Attributes used for e commerce

Attributes	**Description**
A1	Look for product offers
A2	Price details
A3	Read sub pages
A4	Product details
A5	Visit all pages for few minutes
A6	Visit web page regularly

The characteristics of the attributes are discussed in table 1. It contains six attributes to find the interested and non interested buyers.

Table 2: Rate for interested and non-interested customer

Customer Class	**Attributes**	**Rate**
With purchase interest	A1	0,0,1,0,1,1
	A2	1,1,1,1,1,0
	A3	0,2,1,3,0,1
	A4	1,0,1,0,1,1
Without purchase interest	A5	0,1,0,0,1,0
	A6	0,0,0,0,1,0

Table 2 shows the values for with purchasing and without purchasing interest. The interested buyers may have maximum value of above one for all the attributes, but non interest buyers have maximum value as zero. Table 2 taken a six customer from huge amount of customers and show the result.

Table 3: Classification Rate (Percent) of Different Classifiers in the customer reading interest using Different Subsequence Lengths

Number of commands for training	**Classifier Rate (%)**			
	With purchase interest		**Without purchase interest**	
	SVM	**TSVM**	**SVM**	**TSVM**
100	19.3	20.1	18.5	20.2
	21.7	28.6	21.9	33.5
	33.5	40.9	36.3	65.8
	41.1	52.8	32.8	68.1
	45.6	65.7	37.7	70.4
500	43.8	65.9	38.2	73.7
	45.2	66.2	35.7	71.3
	45.8	67.4	39.8	73.8
	46.5	70.5	41.3	75.6
	49.3	71.3	43.8	75.9
1000	50.8	73.8	44.6	69.2
	51.2	74.6	44.9	75.0
	55.3	77.5	45.6	79.5
	56.8	79.3	46.2	82.7
	57.4	81.6	46.8	85.6

These results show that, the proposed classifier can compete well with offline approaches. Nevertheless, the proposed environment needs a classifier able to process streaming data in online and in real time. In addition, the learning in TSVM is performed in single pass and a significantly smaller memory is used. Then the number of attributes is very large in the proposed environment and it changes frequently, TSVM is the most suitable alternative.

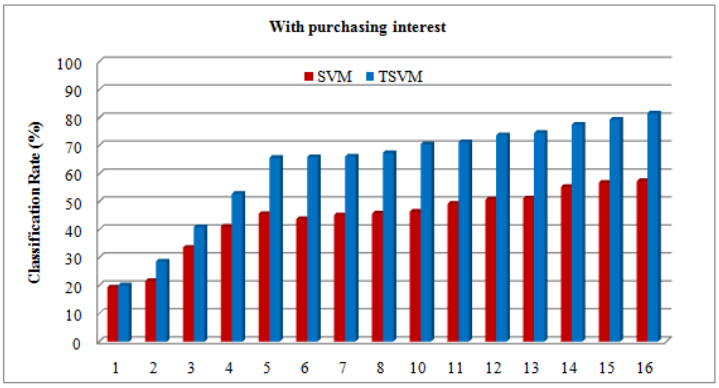

Figure 1: Classification rate for interested customer

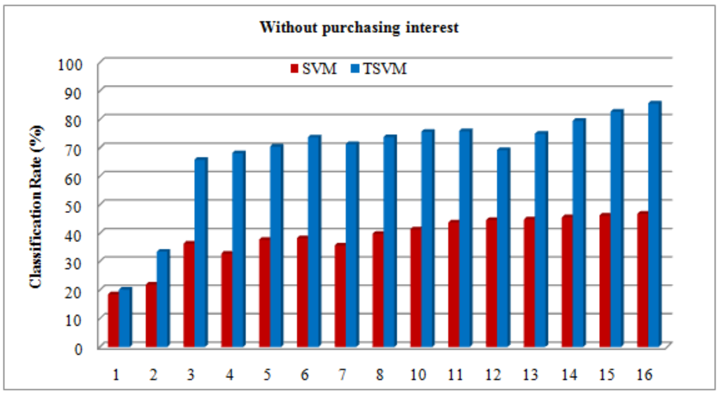

Figure 2: Classification rate for non-interested customer

Figure 1 and figure 2 show the interested and non-interested customers for purchasing in e-commerce. Thus the proposed method of TSVM has better performance in classification rate when compare with the SVM.

V. CONCLUSION

In this research work, standard web site presents well known TSVM methods on the level of trust in e-commerce and thus provides experiential verification suggesting that e-commerce web sites should present customer behavior model. Moreover, the findings here suggest that it would be more effective to have tributes, particularly, if the e-commerce site sells expensive products. And in finale, evidences would be influential if the site caters for inexperienced e-shoppers. It is significant to enlarge an understanding concerning the shopping cart, as the cart is an essential piece of software for e-commerce businesses. Only if customer understands that they will increases their shopping cart adaptation. Thus, proposed TSVM presents in online purchasing interest and without purchasing interests of the customers in accurate calculation using nine attributes. In future work, supplier spirit improves their own web site by adding coupons and wish list so; the buyer can have some of things for free.

REFERENCES

[1] Kalelkar, Gaurav R., Gaurav Kumbhare, Varun Mehta, and Arpan Kumar Kar 2014, ‘Evaluating E-Commerce Portals from the Perspective of the End User–A Group Decision Support Approach’, In Advances in Signal Processing and Intelligent Recognition Systems, pp. 107-117.

[2] Chen, Jing Elaine, Shan L. Pan, and Tao Hua Ouyang 2014, ‘Routine reconfiguration in traditional companies’e-commerce strategy implementation: A trajectory perspective’, Information & Management 51, no. 2: 270-282.

[3] Li, Guangxia, Steven CH Hoi, and Kuiyu Chang 2010, ‘Two-View Transductive Support Vector Machines’, In SDM, pp. 235-244.

[4] Kondratovich, Evgeny, Igor I. Baskin, and Alexandre Varnek 2013, ‘Transductive support vector machines: Promising approach to model small and unbalanced datasets’, Molecular Informatics 32, no. 3: 261-266.

[5] Bottou, Leon, Ronan Collobert, and Jason Weston 2010, ‘Method and apparatus for transductive support vector machines’, U.S. Patent 7,778,949.

[6] Mountrakis, Giorgos, Jungho Im, and Caesar Ogole 2011, ‘Support vector machines in remote sensing: A review’, ISPRS Journal of Photogrammetry and Remote Sensing 66, no. 3 : 247-259.

[7] Yu, J.X.; Yuming Ou; Zhang, C.; Zhang, S., "Identifying interesting visitors through Web log classification," Intelligent Systems, IEEE , vol.20, no.3, pp.55,59, May-June 2005 doi: 10.1109/MIS.2005.47

[8] Srivastava, Durgesh K., and Lekha Bhambhu 2010,‘Data classification using support vector machine’.

[9] Carneiro, Ana Ribeiro, Alípio Mário Jorge, Pedro Quelhas Brito, and Marcos Aurélio Domingues 2014, 'Measuring the Effectiveness of an E-Commerce Site Through Web and Sales Activit', In Modeling, Dynamics, Optimization and Bioeconomics I, pp. 149-162.
[10] Pavlou, Paul 2003, 'Consumer acceptance of electronic commerce: integrating trust and risk with the technology acceptance model', International journal of electronic commerce 7, no. 3: 101-134.
[11] Liu, Chang, and Kirk P. Arnett 2000, 'Exploring the factors associated with Web site success in the context of electronic commerce', Information & management38, no. 1: 23-33.
[12] Stoll, Kurt Uwe, and Martin Hepp 2013, 'Detection of E-Commerce Systems with Sparse Features and Supervised Classification', In e-Business Engineering (ICEBE), 2013 IEEE 10th International Conference on, pp. 199-206.
[13] Di, Wei, Anurag Bhardwaj, Vignesh Jagadeesh, Robinson Piramuthu, and Elizabeth Churchill 2014, 'When relevance is not Enough: Promoting Visual Attractiveness for Fashion E-commerce', arXiv preprint arXiv: 1406.3561.
[14] Ghosh, Anup K., Aaron Schwartzbard, and Michael Schatz. "Learning Program Behavior Profiles for Intrusion Detection." In Workshop on Intrusion Detection and Network Monitoring, vol. 51462. 1999.
[15] Iglesias, Jose Antonio, Plamen Angelov, Agapito Ledezma, and Araceli Sanchis. "Creating evolving user behavior profiles automatically." Knowledge and Data Engineering, IEEE Transactions on 24, no. 5 (2012): 854-867.

