AN APPROACH TO DETECT CREDIT CARD FRAUDULENCE USING HMM MODEL

Manjula.D

ABSTRACT

Online payments or electronic payments nowadays are provided with maximum security, even though, they are not free from flaws. One of the major e-transactions that face threats is while using credit/debit cards. If the cardholder does not know that his/her card is threatened by a third party, there is a chance that the particular card being misused. So, there is a dire need to reduce security threats. In this research SVM, and HMM algorithms are compared and analyzed.

I. INTRODUCTION

Credit card is the most common payment option nowadays, because of its ease of use for online shopping etc. While using physical such cards there are chances of fraudulent use. If a person loses his/her card, there is a huge chance of its misuse by a third party.

Electronic payment method started in the early 70's. It widely spread all over the world because of easy transaction through it by internet. Even though after the 90s quite a number of people started using credit cards cashless transaction also came into vogue.

In the electronic commerce field a lot of financial transactions are made with the data sent via internet. Security is the biggest problem while using electronic payment.

[1] Assistant Professor, Dept of CS, CA & IT, Karpagam Academy of Higher Education, Coimbatore

CLUSTERING

Clustering helps in gathering information into comparable groups, which enables uncomplicated recovery of information. The information in a solitary cluster contains close relationships with one another, while the information in distinctive groups doesn't present close affiliations.

Method of partitioning

Iterative movement calculation.

Input: The quantity of cluster K, and a database containing n objects. Output: An arrangement of K cluster, which minimizes a paradigm capacity J.

Step 1: Start with a beginning K focuses/circulations as the introductory arrangement.

Step 2: (Re)compute participations for the information focuses utilizing the present group focuses.

Step 3: Redesign some/all cluster focuses/dispersions as per new part tastes of the information focuses.

Step 4: Rehash from Step 2 until no change to J or no information focuses change group. Utilizing this structure and iterative routines figure out the assessments for cluster centers, which are fairly alluded to as models or centroids. The models are intended to be the most illustrative focuses for the cluster. The mean and middle are ordinary decisions for the appraisals. Then again, a few systems, for example, the EM-

calculation, gauge an arrangement of parameters that amplifies the probability of the picked appropriation model for an information. The best-known of the model-based calculations are K-means and K-medoids, while the EM-calculation is presumably the most mainstream appropriation-based calculation.

DETECTING CREDIT CARD FRAUDULENCE USING SVM

SVM is a twofold order and the exchanges are named either as false or honest. To handle the imbalanced dataset, the calculation utilizing diverse blunder cost for the positive (C+) and the negative (C-) classes is used.

Accessible information can't be straightforwardly utilized by the SVM. SVM utilizes just numerical information under characterized limits. Yet, the information introduced by the client will, for the most part, contain numerical and absolute traits. These ascribes must be changed into a numerical arrangement to be utilized by the SVM. Thus the preprocessing stage assumes an imperative part in the beginning stage. This can't be performed specifically by the framework, since straight out characteristics are included. The client must give comparable qualities to the clear-cut properties available in the information set.

For standardization process minimum and maximum normalization is used.

Range [C,D]

It is given by the recipe :

Standardized worth (B)=(A-Minimum)/(Maximum-Minimum)*(D-C)+CEq(3.1)

whereas is Min-Max Normalized data one, is pre-defined boundary, is the range of original data, is Minimum value of , is Maximum value of .

Each accessible information goes into the Normalization capacity to get its comparing standardized worthy. This procedure changes the information into an effortlessly justifiable and effectively interpretable configuration, and henceforth gets to be less demanding for the calculations amid the relative investigation of the information.

After this procedure, utilizing the standardized information, the preparation and testing information documents are made. The SVM requires an exceptional organization for perusing the information.

SVM Fraud Detection Algorithm

1. Obtain exchange information.
2. Preprocess information to change over absolute ascribes into numerical Properties.
3. Prepare information records using Support Vector Machine.
4. Prepare document to Support vector Machine
5. Set the qualities for C and ?.
6. Obtain results utilizing the present C and ? pair.
7. Continue 6 & 7 until the correct result is obtained.
8. Test the document.

DETECTION OF CREDIT CARD FRAUD USING HMM

Hidden Markov Model is a limited set of state. HMM will detect particular person's card behavior and his/her transaction.

In business side, every transaction is verified by detection section. The detection section will check all the card details like its number, cvv, type etc. Clustering will be done for the training set to identify particular card holder[8]

The detection mode will check the regular transaction type. If any variance happens in a transaction, an alarm will be raised and the bank will stop the process. The full details of a card holder are stored in security information. For the verification process name of mother, father, favorite food etc., will be checked[9].

HMM can be classified as:

The number of states in the model is N.

The set of states is S={S_1, S_2,… S_N}, where S_i, i=1,2,…. N is an individual state. The state at time instant t is referred by q_t.

1.The number of distinct observation symbols per state is M.

The set of symbols is V={V_1,V_2,… V_3}

where Vi, i=1;2;…..; Mis an individual symbol.

2.The state transition probability matrix A=[a_{ij}]

where_i_j=P(q_t+1 = S_j| q_t= S_i); 1≤i≤N; 1≤j≤N; t=1,2,… N

where a_{ij}> 0 for all i,j. Also $\sum_{j=1}^{N} a_{ij} = 1, 1 \leq i \leq N.$

4. The observation symbol probability matrix B=[b_j(K)],

Where $b_j(K)=P(V_k|S_j)$, 1 |≤ j≤ , 1≤ K ≤ M

and $\sum_{k=1}^{M} b_{ij}(K) = 1, 1 \leq j \leq N.$

5. The initial state probability vector π where π_i=P(q_i=S_i), 1≤$_j$≤$_N$ such that $\sum_{K=1}^{M} \pi_i = 1$

6. The observation sequence O=O_1,O_2, O_3….O_R where each observation

O_t is one of the symbols from V, and R is the number of observations in the sequence.

It is obvious that a complete specification of an HMM needs the estimation of 2 model parameters, N and M, and three probability distributions A, Band π. The notation λ=(A, B, π) is used to indicate the complete set of parameters of the model, where A and B implicitly include N and M.

Algorithm	Data samples	Fraud Case	Non-Fraud Case
SVM	500	320	180
	1000	520	480
	1500	850	650
	2000	1400	600
	2500	1700	800
HMM	500	400	100
	1000	700	300
	1500	1000	500
	2000	1700	300
	2500	1900	600

Table 4.1 shows the datasamples values for different algorithms like SVM, and HMM.

(i) Precision Rate

Performance offered by SVM and HMM are analysed and compared. The precision of the HMM is High.

Precision= Eq $\frac{t_p}{t_{p+f_p}}$ (4.1)

whereastpis true positive, $f_p$is false positive.

Table 4.2 Precision Rate by using SVM, HMM for data samples

S.No	Data Samples	SVM (%)	HMM (%)
1	500	73	82
2	1000	66	74
3	1500	56	62
4	2000	44	54
5	2500	37	47

Recall Rate:

Based on the comparison HMM works better than SVM.

Recall= $\frac{t_p}{t_{p+f_n}}$ Eq(4.2)

whereast$_p$is true positive, $f_n$is false negative.

Table 4.3 Recall Rate by using SVM, HMM for data samples

S.No	Data Samples	SVM (%)	HMM (%)
1	500	75	77
2	1000	68	76
3	1500	58	69
4	2000	46	56
5	2500	40	51

(ii)Accuracy Rate:

The performance offered by SVM and HMM are analysed and compared. The accuracy of the HMM is good.

Accuracy= Eq$\frac{t_p+t_n}{t_p+t_n+f_p+f_n}$ (4.3)

Table 4.4 Accuracy Rate by using SVM, HMM for data samples

S.No	Data Samples	SVM (%)	HMM (%)
1	500	74	83
2	1000	67	75
3	1500	57	66
4	2000	45	55
5	2500	39	49

(iii) Error Rate

The performance offered by SVM and HMM are analysed and compared. The error rate of the HMM is better.

Error Rate = $\frac{f_p+f_n}{t_p+t_n+f_p+f_n}$ Eq(4.4)

where as f_p is false positive, f_n is false negative, t_p is true positive, t_n is true negative.

Table 4.5 Error Rate by using SVM, HMM for data samples

S.No	Data Samples	SVM (%)	HMM (%)
1	500	75	67
2	1000	69	56
3	1500	48	48
4	2000	46	44
5	2500	34	32

CONCLUSION AND FUTURE ENHANCEMENT

In recent times credit cards have become most the popular means of payment and if credit card transactions increase, so do its fraudulent uses. This research has presented the classification of credit card challenges faced by cardholders.

Support Vector Machines (SVMs) clear problems relating to the classification of complex data. A Hidden Markov Model performs well to detect fraudulence in the use of credit card.

In this research, Precision, Recall, Accuracy and Error Rate metrics are utilized for analysis. Datasets are downloaded from UCI repository to analyse the performance of three algorithms. From this research, it can be concluded that HMM algorithm works better for Online Credit Card Fraud Detection when compared to SVM algorithm with high Precision and Accuracy.

REFERENCES :

1. Osuna. E, "Applying Support Vectors Machines to Face Detection", IEEE Intelligent Syst. Mag., Support Vector Machines, Vol. 13, No. 4, pp. 23-26, 1998
2. Tareq Allan and Justin Zhan, "Towards Fraud Detection Methodologies", IEEE Proceedings of the Fifth International Conference on Future Technology, Print ISBN:978-1-4244-6948-2, pp.1-6, 2010.
3. Dhanapal.R, "An Intelligent Information Retrieval Agent", Elsevier International Journal on Knowledge-Based systems, The Netherlands,ISSN: 0950-7051, Vol.21, Issue.6, pp. 466-470, 2008.
4. Cristianini N and Shawe-Taylor J, "An Introduction To Support Vector Machines and other Kernel-Based Learning Methods", Cambridge University Press, Cambridge, ISBN:0-521-78019-5, pp.189, 2000.
5. Vapnik, V.N, "The Nature Of Statistical Learning Theory", Springer, ISBN: 0-387-94559-8, 1995.
6. Sahin.Y and E.Duman, "Detecting Credit Card Fraud By Decision Trees and Support Vector Machines", Proceedings of the International Multiconferenceof Engineers and Computer Scientists, ISBN:978-988-18210-3-4, Print ISSN:2078-0958, Online ISSN:2078-0966, Vol.1, 2011.
7. Yang Zhang, Weiming Liu, "A Novel Pedestrian Detection Method Based on Cost-Sensitive Support Vector Machine and Chaotic Particle Swarm Optimization With T Mutation", PrzegladElektrotechniczny (Electrical Review), ISSN 0033-2097.R.88 Nr 1b/2012, 2012.
8. Phua, V. Lee, K. Smith, and R. Gayler, "A Comprehensive Survey of Data Mining-Based Fraud Detection Research,2007.
9. Rabiner, L. R, "A tutorial on hidden markov models and selected applications in speech recognition",ISBN:1-55860-124-4, pp.267-296, 1989
10. Kim.J.M and T.S. Kim, "A Neural Classifier With Fraud Density Map For Effective Credit Card Fraud Detection",Proc. International Conference on Intelligent Data Engineering and Automated Learning, Lecture Notes in Computer Science, Springer Verlag, Print ISBN:978-3-540-44025-3, No. 2412, pp. 378-383, 2002.

11. Brause, R, T Langsdorf, and M Hepp. "Neural Data Mining For Credit Card Fraud Detection", Proceedings of the 11th IEEEInternational Conference on Tools with Artificial Intelligence, ISBN:0-7695-0456-6, pp.103,1999.

12. D.Manjula, J.Thilagavathi "A Novel Approach for Behavior based Charge Card Fraud Detection using Support Vector Machines", IJSRD - International Journal for Scientific Research & Development, Vol. 3, Issue 06, 2015 | ISSN (online): 2321-0613.