

# DNA: An Online Algorithm for Credit Card Fraud Detection for Games Merchants

Michael Schaidnagel  
D-72072 Tübingen, Germany  
Michael.Schaidnagel@web.de

Iliia Petrov, Fritz Laux  
Data Management Lab  
Reutlingen University  
D-72762 Reutlingen, Germany  
{Iliia.Petrov | Fritz.Laux}@reutlingen-university.de

**Abstract**—Online credit card fraud represents a significant challenge to online merchants. In 2011 alone, the total loss due to credit card fraud amounted to \$ 7.60 billion with a clear upward trend. Especially online games merchants have difficulties applying standard fraud detection algorithms to achieve timely and accurate detection. The present paper introduces a novel approach for online fraud detection, called DNA. It is based on a formula which uses attributes that are derived from a sequence of transactions. The influence of these attributes on the result of the formula reveals additional information about this sequence. The result represents a fraud level indicator, serving as a classification threshold. A systematic approach for finding these attributes and the mode of operation of the algorithm is given in detail. The experimental evaluation against several standard algorithms on a real life data set demonstrates the superior fraud detection performance of the DNA approach (16.25 % better fraud detection accuracy, 99.59 % precision and low response time). In addition to that, several experiments were conducted in order to show the good scalability of the suggested algorithm.

**Keywords**- binary classification, credit card fraud, online environment

## I. INTRODUCTION

The approximate global business volume of the computer gaming industry in total rose from \$ 20 billion in 2001 to \$ 65 billion in 2011 [1]. It is estimated to grow by 10.6 % in 2013. New technology developments, such as browser games and Massive Multiplayer Online Games have created new business models (based on micropayments) for online games merchants. Both, technology and business model affect the customers payment behavior. Their first choice for performing online payments is the credit card. The downside of this development is an increase in online credit card fraud, which continues to pose a big threat for online merchants. The total loss due to credit card fraud rose to \$ 7.60 billion in 2011 [2] and is supposed to increase further. Especially online games merchants have difficulties applying standard techniques for fraud detection. The reason for this is the lack of personal information about their customers (e.g., real names and postal address for address verification) as well as the need for real time classification.

*Problem definition:* There are a number of constraints, which make it difficult to apply the traditional algorithms for credit card fraud detection. Players do not feel comfortable to reveal their real names and addresses in an online gaming

environment. This lack of financial data in addition to the short transaction histories of players makes it difficult to apply standard techniques. Furthermore, the real time nature of business makes it necessary to be able to apply an algorithm in real time, or near real-time, in order to reject fraudulent transactions at authorization time. Most of the techniques proposed so far are bulk oriented and designed for offline batch processing.

*Contributions:* We present a novel algorithm, which is able to handle the scarce data situation by deriving attributes out of a sequence of transactions. These attributes are normalized, weighted and arranged in a way that enables a simple formula to recognize different fraud behavior patterns. In order to assess transactions without any history, a concept of cultural clusters was introduced to help classifying those transactions. In addition to that, a metric for assessing the suitability of attributes, their influence on the fraud level as well as the calculation of the threshold are introduced. The DNA approach performs 16.26 % better than the best standard method (Bayesian Net) and achieves an almost perfect 99.59 % precision. In addition, the DNA approach scales better than other approaches with increasing data volumes, while offering acceptable response/detection times.

The presented paper is structured as follows: Section II will give an overview of the related work, which describes different data mining algorithms normally suggested for this problem. They are also part of the experimental evaluation. Section III will detail our suggested method and describe the major components. The suggested method is applied to a real life data set in Section IV. Section V concludes the results and mentions a few points for future work.

## II. RELATED WORK

So far, there have been many data mining algorithms applied in order to detect credit card fraud [3]. Please note that we do not go into details here on how they work. All mentioned methods have been implemented and will be compared in terms of fraud detection performance in Section IV.

*Artificial Neural Network (ANN):* Gosh and Reilly [4] were the first ones to adapt Neural Networks on credit card fraud detection. Other authors such as Dorronsoro et al. [5], Brause et al. [6] and Maes et al. [7] have also implemented ANNs in real life applications. ANNs in general are too dependent on meaningful attributes, which should not

necessarily be available. The information gain from such attributes is too low to be utilized in ANNs.

**Bayesian Belief Network (BBN):** The first implementation for fraud detection was done by Ezawa et al. [8]. Other recent implementations are Lam et al. [9], Maes et al. [7] and Gadi et al. [10]. However, some data set do not provide enough attributes in order to construct a suitable network.

**Hidden Markov Model (HMM):** In recent years several research groups applied this model for fraud detection. Srivastava et al. [11] have conducted a very systematic and thorough research in their work. Other implementations were done by Mhamane et al. [12], Bhusari et al. [13] and Dhok [14]. A classic and comprehensive introduction to the topic of HMM was published by Rabbiner and Juang [15] and also Stamp [16] is worth reading for introductory purposes. HMMs in general are only able to utilize a single numeric attribute for their prediction, which is insufficient for a proper classification.

**Decision Tree (DT):** The biggest impact on how Decision Trees are built had Quinlan [17] in the late 90s. There have been some applications on fraud detection in recent years, e.g., Minegishi et al. [18]. Other mentionable fraud detection implementations are Sahin and Duman [19], Sherly et al. [20] and Gadi et al. [21]. DTs in general suffer the same insufficiencies as ANNs.

### III. PROPOSED METHOD

The algorithm is named after the famous Deoxyribonucleic Acid (DNA), which is the basis of all living organisms. The DNA is able to store very complex information with just four basic components (the so-called nucleotides). The crucial insight here is that not only the sequence of these nucleotides is important, but also their interconnection. Similar to nature, we derive attributes out of the sequence of transactions and then take a look at their influence to our labels (fraud, genuine). The remainder of this section will now detail the different parts necessary for building the DNA approach.

#### A. Cultural Clusters

The Cultural Clusters were introduced in order to help classifying transactions without any history. The idea behind it was to get as much information out of the given attributes as possible. These attributes include the origin of the user (IP country) and the origin of the credit card used in a transaction (BIN country – BIN is an abbreviation for Bank Identification Number: first 6 digits of a credit card number, enables to locate the card issuing bank of the cardholder). Cultural related countries form clusters; which are roughly based on continents. A range of weights is assigned to each cluster. Every country is assigned with a specific weight within its cluster’s range depending on its cultural distance to its cluster center and the risk of the county of being defrauded. The weight of a country within a certain cultural cluster is set empirically and can be subject for adaption, in case the fraudulent behavior changes. In other words: the weight of a country lies within the range of its cultural

cluster and is set by an initial value, based on the experience of a fraud expert. If cards from this country turn out to be defrauded frequently, the weight can be increased (within the limits of its cultural clusters). This will increase the risk value of a country pair, which can be calculated as it can be seen in equation (1):

$$risk = |weight(IP\ country) - weight(BIN\ country)| \quad (1)$$

This value will be low for country pairs within the own country cluster (e.g., a user from Sweden tries to use a card originated in Norway) or 0 if the user and the corresponding card are from the same country. On the other hand, this value increases if there is a suspicious country pair involved (cross-cultural cluster). This simple metric allows depicting complex risk relationships between several countries.

#### B. Building sequence based attributes

As briefly noted at the beginning of this section, the DNA approach does not only rely on the risk assessment of the involved country pairs of a transaction. Furthermore, multiple other sequential based attributes are used to enhance the fraud detection performance. The basic process starts by selecting an attribute which is used to identify associated transactions. An example for such an attribute could be the account number or email address of a user. In this work we use the term sequence to refer to all transactions belonging to a certain user email address. In our case, the attribute email address represents the sequencing attribute which is used to build grouped data entries. Please note that the transactions of a sequence are sorted by the transactions timestamp prior to aggregation. An excerpt of the grouped, intermediate data is shown in Figure 1.

timestamp	userEmail	transaction no	a <sub>1</sub>	a <sub>2</sub>	agg. Attrib. a <sub>1</sub>	agg. Attrib. a <sub>2</sub>	sequence identifier	label
01.01.2012	userEmail2	t1	value_X	value_A	1	10	userEmail1	fraud
03.01.2012	userEmail2	t2	value_Y	value_B	3	2	userEmail2	genuine
18.01.2012	userEmail2	t3	value_Z	value_A	4	9	userEmail3	genuine
					5	10	userEmail4	genuine
					2	4	userEmail5	genuine

used for parameter calculation

Figure 1. Schematic representation for the sequenciation (grouping) process

During aggregation, we used typical operators, such as sum or count of distinct values and calculated the following parameters of the aggregated attributes for both of our labels (fraud, genuine) and overall for all sequences (see Table I):

TABLE I. AGGREGATED PARAMETERS FOR EXAMPLE ATTRIBUTE SUMCREDITCARDTOKEN

parameter name	genuine	fraudulent	total
Maximum value	29	38	38
Minimum value	0	0	0
Average	1.19	2.61	1.34
Standard deviation	0.62	2.85	1.07

Based on these values, we are able to estimate whether an attribute is suitable for the DNA approach or not. As it can

be seen the parameter minimum alone, for example, is not able to sufficiently segregate both labels (the average of both labels yield more segregation). Adequate attributes have a great distance between the average fraud and average genuine value. The corresponding standard deviation boundaries should, if possible, not overlap. Figure 2 shows this based on the example attribute *sumCreditCardToken*. The attribute was part of the data set described in Section IV. It was created by counting up all distinct credit cards, which were used by a certain user\_email. For the given example case, we calculated the values for the aggregation parameters (as depicted in Table 1 column two to four).

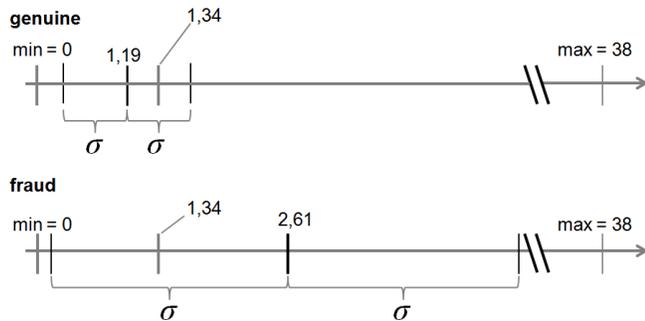


Figure 2. Example distribution for used credit cards per user for two classes of labels (fraud, genuine)

As it can be seen in Figure 2, the depicted attribute is not perfect, since the boundaries of the standard deviation of fraudulent transactions overlap with the area of standard deviation of genuine transactions. However, the distance between the average fraud value and average genuine value is suitable for classification. We can also see that the average fraud value is on the right hand side of the total average value and therefore tends to larger values in a fraudulent case. This indicates that the attribute *sumCreditCardToken* should be placed in the numerator of (4), since it will tend to its maximum if a sequence is fraudulent.

For our work, we were able to identify six suitable attributes:

- *sumCreditCard*: # of distinct credit cards per user
- *sumTransactionStats*: # of transaction status 'rejected'
- *sumSuccessfulTrans* # of completed transactions
- *avgDensity*: average distance between dates
- *sumCountries*: # of distinct countries involved
- *sumDates*: # of distinct dates involved

### C. Formula and Weighting

The above-described process is repeated for each of the n available attributes in the given data set [see also (2)] apart from the aggregation attribute and the label attribute. The training data set T consists of transactions  $t_i$  which are described by attributes  $a_i \in A$ , see also Formula (2). Thereby A consists of all original attributes of a transaction  $a_i$  and T denotes the set of available transactions with attributes from A. These transactions are grouped into sequences S, which consist of various combined attributes  $b_i$ , see also (3). These

combinations can either be done by functions  $b_i \in f(t_i(a_j))$  or by  $b_i \in f(a_i, a_j)$ .

The attributes are normalized with the min-max normalization [3, p. 114] to bring the different attributes on the same numerical level (ranging from 0 to 1). We distinguish between two types of attributes, as briefly described above. These attributes are referred as  $b_i \in F$ , where F is the set of derived attributes that tend to 1 if normalized and are summed up in the nominator of (4). The denominator, in contrary, is composed of a second type of attributes  $b_i \in E$ , which tend to 0 if normalized with min-max normalization. If the quotient of the normalization expression is not defined, it will be discarded.

$$T = \{t_i(a_1, a_2, \dots, a_n) | \forall i = 1, \dots, n: a_i \in A\} \quad (2)$$

$$S = \left\{ s(b_1, b_2, \dots, b_m) \left| \begin{array}{l} \forall k \in 1, \dots, m: b_k \in f(a_i, a_j) \quad i, j \in 1, \dots, n \\ \forall k \in 1, \dots, m: b_k \in f(t_i(a_j)) \quad \begin{array}{l} i \in 1, \dots, |T| \\ j \in 1, \dots, n \end{array} \end{array} \right. \right\} \quad (3)$$

$$riskLevel = \frac{\sum_{b_i \in F} \left( \frac{b_i - \min_S b_i}{\max_S b_i - \min_S b_i} \right) * \alpha_{b_i}}{\sum_{b_i \in E} \left( \frac{b_i - \min_S b_i}{\max_S b_i - \min_S b_i} \right) * \beta_{b_i}} \quad (4)$$

The next step is to weight the normalized attributes according to their importance for fraud detection. We use parameter  $\alpha_{b_i} > 0$  for F attributes and  $\beta_{b_i} > 0$  for E attributes to achieve this task. Two sets of parameters are necessary since attributes used for the denominator need to be scaled down, in order to increase their significance. The attributes in the nominator are increased in significance, if the  $\alpha_{b_i}$  is scaled up. The weight parameters  $\alpha_{b_i}$  and  $\beta_{b_i}$  are determined empirically.

### D. Threshold Selection

Formula (4) is an indicator on how prevalent fraudulent attributes are for a particular transaction. The last step for applying the DNA approach for fraud detection is to determine a threshold value whose violation will lead to the classification "fraudulent transaction". As mentioned above, this threshold is determined empirically by undertaking a series of tests with a set of thresholds (e.g., from 0 to 100). Accuracy metrics such as Precision P, Recall R and score F1 (5) are then determined for each assumed threshold. Thereby F1 represents the harmonic mean of Precision and Recall and is used to rank the performance of different methods in the experimental evaluation:

$$F1 = 2 * \frac{P * R}{P + R} \quad (5)$$

The development of these performance measures over different threshold values is depicted Figure 3 for an example case:

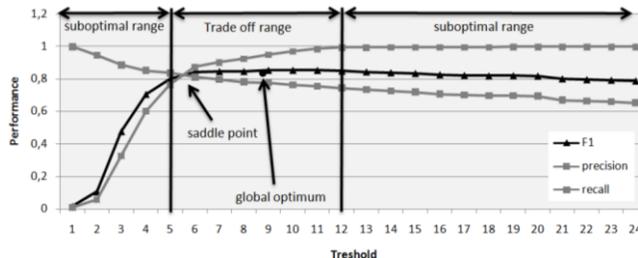


Figure 3. Determining threshold value for DNA approach

Accuracy indicators (Figure 3) are increasing fast until threshold value 5. It is not reasonable to select a threshold lower than 5, since the F1 is far from the global optimum. From threshold 5 on, there is an intersection point, which will keep the F1 near the global optimum. This second range is called “trade-off range” and spans up to threshold value 12, in the case depicted in Figure 3. Within this range the merchant can choose between detecting more fraudsters, including a higher rate of false positives or catching less fraudsters, but increase Precision and therefore avoid false positives. This choice can depend on the ability of the merchant to deal with false positives and on the merchants specific total fraud costs. In the context of fraud detection the term total fraud costs means the sum of lost value, scanning cost as well as reimbursement fees associated with a fraud case.

After a certain threshold value, in the shown case 12, the Precision is almost 1 and will only increase insignificantly. The Recall and consecutively F1, will decrease from that point. The Reason for this is the intrinsic mechanic in the DNA approach. Fraudulent transactions with a comparable low fraud profile will be assigned a lower risk level. This however, is still higher level than the risk level of genuine users. If however, the threshold is set high enough these lower profile fraud cases will be classified incorrectly as genuine, causing the Recall and F1 to drop. Therefore it makes no sense to choose a threshold greater than 12.

IV. EXPERIMENTAL EVALUATION

The performance of the proposed DNA approach is compared to the standard techniques, mentioned in the related work. This comparison is based on real credit card fraud data, which was thankworthy provided by a successful gaming company on the online games market.

A. Data Set

The given data set (referred as full data set) comprises of 156,883 credit card transactions from 63,933 unique users. The records in the data set have the schema as it can be seen in Table II. Due to the high number of occurrences in several columns as well as the lack of distinctive attributes, most of the standard techniques were not applicable on that data set. To overcome these obstacles and to get a fair comparison, several adaptations to the data set have been done. The resulting prepared data has a minimum sequence length of three (smaller sequences have been discarded) and four derived attributes (see Table III) were added.

TABLE II. FULL DATA SET SCHEMA

column Name	description
created	timestamp of the payment transaction
user_signup_time	
creditcard_token	identifies credit card, hashed
card_bin	Bank Identification Number
user_country	user’s land of origin
user_id	
user_email	hashed for privacy compliance
transaction_amount	
order_payment_status	

TABLE III. ADDITIONS PREPARED DATA SET

column Name	description
bin_country	2 letter country code derived from card_bin
days_since_signup	integer attribute calculated as difference from signup_time to created
total_count	denotes total transaction figure for a particular user_email
package	a single letter attribute ranging from A to E. It was derived from the offer_price attribute to reduce the cardinality of the offer_price attribute

The prepared data set comprised of 13,298 unique users which are accompanied by 46,516 transactions. The last transaction of each user was cut out in order to form the test data set. This procedure segmented the prepared data set into 71.4 % train data and 28.58 % test data.

B. Fraud Detection Performance

All tests in this section were performed using the prepared data set. We used the F1 score in order to rank the compared methods. As shown in Figure 4, the DNA approach is able to perform 16.25 % better than the best standard method, which is the Bayesian Net.

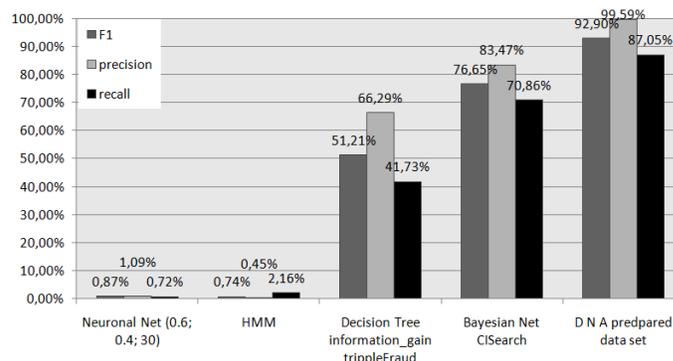


Figure 4. Fraud detection performance comparison

The DNA approach is also able to achieve an almost perfect 99.59 % Precision, which is especially valuable for online gaming merchants, since it reduces the risk of punish

genuine users and consecutively reduces the risk of reputation loss.

C. Scalability

The previous subsection showed the impressive fraud detection accuracy of the DNA approach. The next step was to take a deeper look into the run time behavior of the compared methods. Therefore, we prepared scaled data sets by multiplying the original prepared data (Subsection IV A) by 4, 8 and 16 times. The corresponding total execution times in seconds of the used methods can be seen in Figure 5.

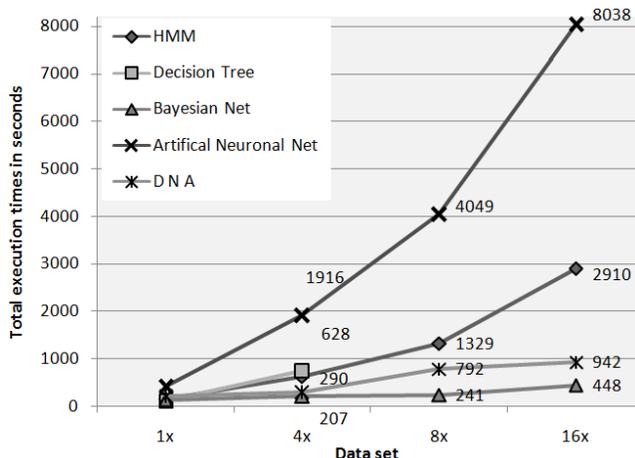


Figure 5. Execution time of all implemented methods (the used DTs produced out of memory exceptions for data set 8x and 16x)

The DNA approach uses simple operations and data structures like HashMaps or simple additions for calculating the risk level per sequence. This lowers the computational intensiveness and improves scalability significantly. The Bayesian Belief Networks obtain consistently the best results for static analysis (all transactions are classified in one batch). However, the BBN is an offline algorithm; therefore it cannot be applied in a real time online payment system on a transaction basis. To use offline algorithms in real applications, further timeslots need to be considered which were neglected in this work (data collection, transfer and preparation time).

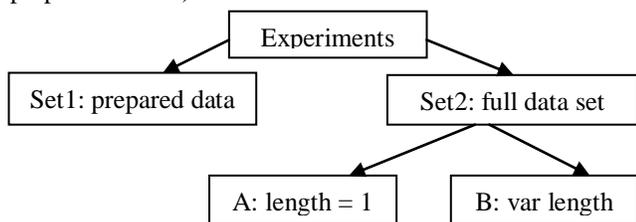


Figure 6. Overview experimental evaluation

The second set of experiments measures the execution time of the DNA approach using the full data set and varying the sequence length. Multiplications of the full data set (Section IV, A) were used for these experiments. In order to perform the measurements, the algorithm was adapted to

record the execution time per sequence. This enables us to measure the average execution time for each sequence length.

*Constant Sequence Length:* In the first round of experiments, we measured the execution time for the fixed sequence length = 1. This allows measuring the influence of the data set size on the execution time.

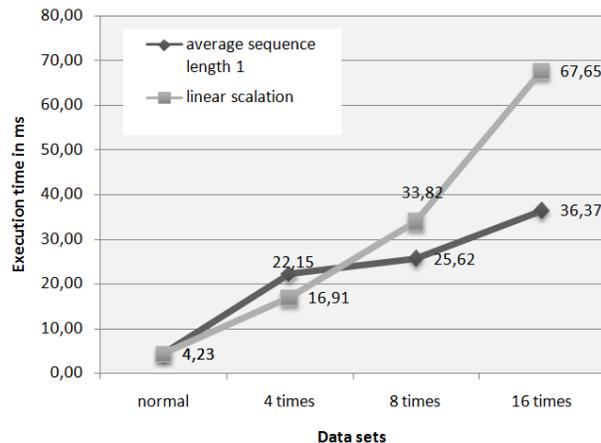


Figure 7. DNA results for fixed length sequence

In Figure 7, the grey line indicates the linear scaled value, based on the results of the normal data set. This means that a linear scaling algorithm, which needs 4.23 [ms] for a sequence of length 1 on the normal table, would need 16 times as long for the same calculation on a 16 times bigger table. As it can be seen the execution time for the DNA approach is less than that.

*Higher Sequence Length:* Throughout the second round of experiments the average execution times for all sequence lengths were averaged. Since the above mentioned data sets were created by multiplication of the full data set, not only the data set size but also the sequence lengths were increased. Figure 8 shows the experimental results.

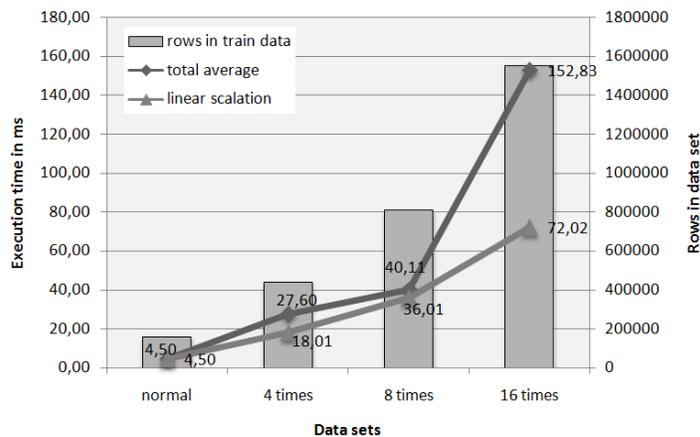


Figure 8. DNA results over higher sequence length

It is visible that the performance stays close to the expected linear response time behavior for the first three data

set sizes. There is a knee-point somewhere between the 8th and the 16th size of the data set, on which the execution time increases over proportionally. These results indicate that the sequence length has a greater influence on the execution time, than the data set size. This property does not necessarily impose a problem for online games merchants since the sequence length usually stays well below critical length.

All performance numbers presented in the experiment evaluation (Figure 7 - 8) reflect the pure response time of the DNA. The execution times of the other algorithms mentioned in Figure 5, display only the decision time of the trained algorithms. An additional overhead has to be taken into account, which reflects time for data collection, data transfer and preparation time. Hence, in practical terms, the reported performance delta is conservative. In real scenarios it will be higher.

## V. CONCLUSION

This work deals with the problem of fraud detection in online games. The problem is caused by the lack of useful financial data, the anonymity in online games as well as the comparably short transaction sequences. The problem is solved by introducing an algorithm that is able to find and use distinctive attributes within sequences. In addition, a concept of country clusters is used to evaluate the legitimacy of a transaction. The DNA approach performs 16.25 % better than the best standard method (Bayesian Net) and achieves 99.59 % Precision. The achieved Recall rate (87.05 %) reduced the probability for false negatives and therefore the need for human intervention is reduced. In addition, the DNA approach scales better than other approaches with increasing data volumes, while offering acceptable response/detection times. This allows the application of the DNA approach in a real time online fraud detection system.

*Future Work:* The algorithm suggested in this work is especially designed to overcome the limiting conditions given in the online fraud detection area. The experiments showed good fraud detection results. However, further development is needed in order to reduce the influence of the sequence length on the execution time. It would also be helpful to incorporate the sequence length into the algorithm. The algorithm may be susceptible to the sequence length due to the proposed additive technique depicted in (4). The used data set did not allow us to precisely quantify possible impacts. Another direction of development could be the abstraction of the proposed attribute construction technique and their application on other classification domains.

## REFERENCES

- [1] n.n. Video Games Wiki [http://vgsales.wikia.com/wiki/Video\\_game\\_industry](http://vgsales.wikia.com/wiki/Video_game_industry) [retrieved: June, 2013]
- [2] n.n. U.S. Leads the World in Credit Card Fraud Report <http://www.businesswire.com/news/home/20111121005121/en/U.S.-Leads-World-Credit-Card-Fraud-states> [retrieved: June, 2013]
- [3] J. Han, M. Kamber and J. Pei, "Data mining: Concepts and techniques" 3. edition, 2012
- [4] S. Ghosh, and D. Reilly, "Credit card fraud detection with a neural-network" In: Proceedings of the Twenty-Seventh Hawaii International Conference, 1994, p. 621–630
- [5] J. R. Dorronsoro, F. Ginel, C. Sgnchez and C. S. Cruz, "Neural fraud detection in credit card operations" In: IEEE Transactions on Neural Networks 8 (1997), No. 4, p. 827–834
- [6] R. Brause, T. Langsdorf and M. Hepp, "Neural Data Mining for Credit Card Fraud Detection", 2000
- [7] S. Maes, K. Tuyls, B. Vanschoenwinkel and B. Manderjack, "Credit Card Fraud Detection Using Bayesian and Neural Networks" In: Proceedings of NF2002, 2002
- [8] K. Ezawa and S. Norton, "Constructing Bayesian networks to predict uncollectible telecommunications accounts" In: IEEE Experts (1996), Vol 11, Issue 5, p. 45–51
- [9] W. Lam and F. Bacchus, "Learning Bayesian Belief Networks: An approach based on the MDL Principle" In: IJCI Vol 13, 1994, pp. 269–293
- [10] M. Gadi, X. Wang and A. P. Lago, "Comparison with Parametric Optimization in Credit Card Fraud Detection" In: Seventh International Conference on Machine Learning and Applications, 2008, p. 279–285
- [11] A. Srivastava, A. Kundu, S. Sural and A. K. Majumdar, "Credit Card Fraud Detection Using Hidden Markov Model" In: IEEE Transactions on Dependable and Secure Computing 5, 2008, No. 1, S. 37–48.
- [12] S. Mhamane and L. M. R. J. Lobo, "Fraud Detection in Online Banking Using HMM" In: International Conference on Information and Network Technology (ICINT 2012), 2012, Vol. 37.
- [13] V. Bhusari and S. Patil, "Application of Hidden Markov Model in Credit Card Fraud Detection" In: International Journal of Distributed and Parallel systems 2, 2011, No. 6, p. 203–211
- [14] S. Dhok, "Credit Card Fraud Detection Using Hidden Markov Model" In: International Journal of Soft Computing and Engineering (IJSCE), 2012, Vol. 2
- [15] L. Rabiner and B. Juang, "An introduction to hidden Markov models" In: IEEE ASSP Magazine 3, 1986, No. 1, p. 4–16.
- [16] M. Stamp, "A Revealing Introduction to Hidden Markov Models", 2012, URL <http://www.cs.sjsu.edu/~stamp/RUA/HMM.pdf> [retrieved: June, 2013]
- [17] J. R. Quinlan, "C4.5: Programs for machine learning", 1998
- [18] T. Minegishi and A. Niimi, "Detection of Fraud Use of Credit Card by Extended VFDT", 2011
- [19] Y. Sahin and E. Duman, "Detecting Credit Card Fraud by Decision Trees and Support Vector Machines" In: Proceedings of the International MultiConference of Engineers and Computer Scientists, 2011, Vol. I
- [20] K. K. Sherly and R. Nedunchezian, "BOAT adaptive credit card fraud detection system", 2010
- [21] M. F. A. Gadi, X. Wang and A. P. Lago, "Comparison with Parametric Optimization in Credit Card Fraud Detection" In: Seventh International Conference on Machine Learning and Applications, 2008, p. 279–285.