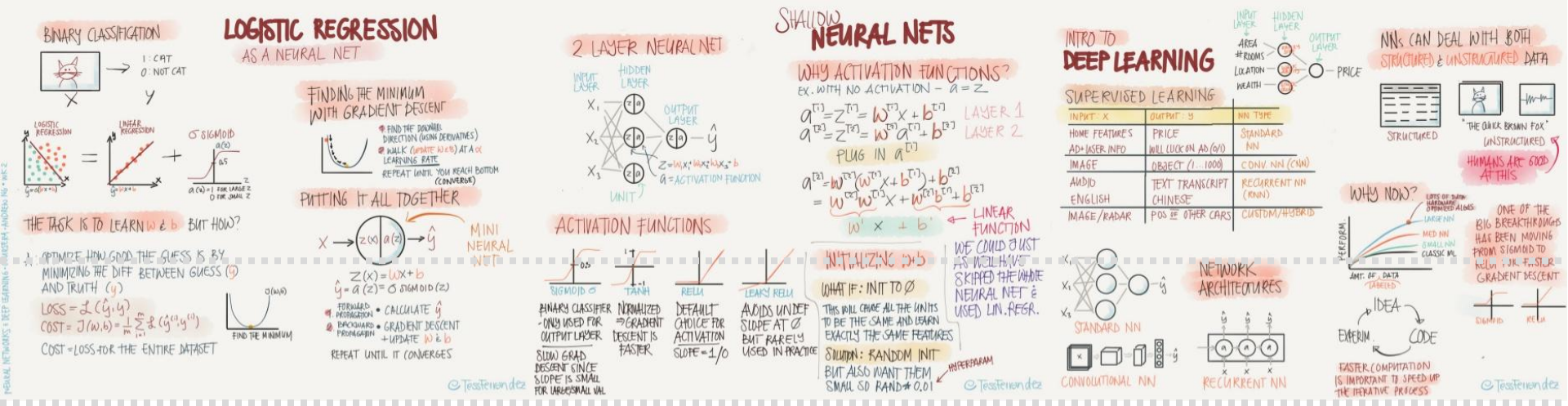


DATA SCIENCE в антифроде

– что предлагает Science?

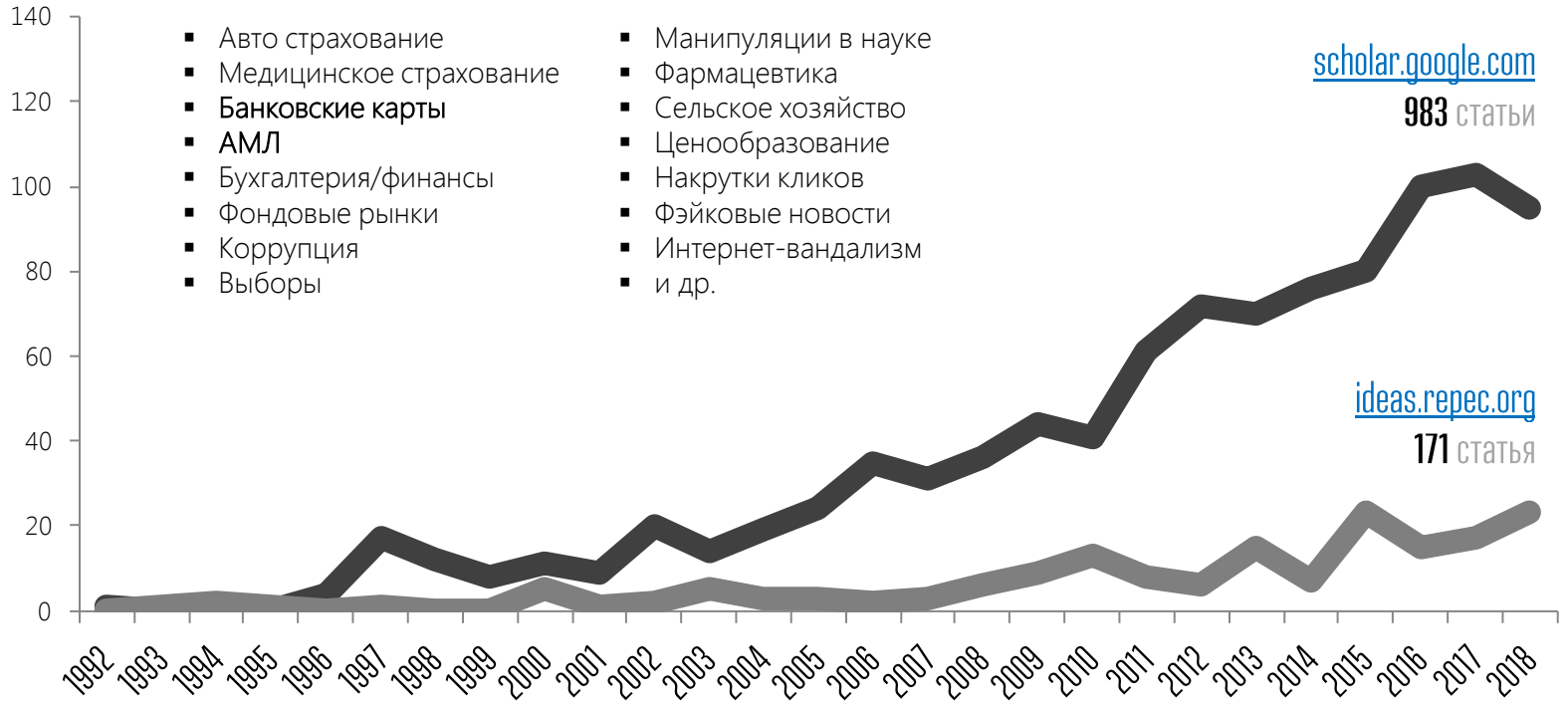
Афанасьев Сергей
КБ «Ренессанс Кредит»

22 ноября 2019 г.
Москва

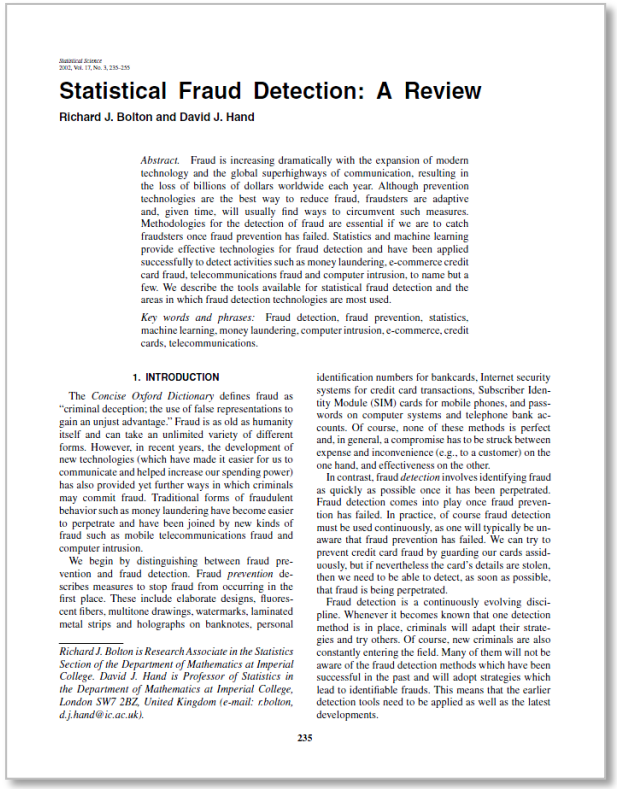
«Чтобы вы не придумали, был какой-то советский математик, который это уже сделал. А до него был какой-то французский математик, который это сделал. До него, какой-нибудь еврейский математик, который это сделал. И т.д.»

Сергей Бартунов, DeepMind

ML в антифроде — растет кол-во статей



1. С чего начать?



Statistical Science
2002, Vol. 17, No. 3, 235–255

Statistical Fraud Detection: A Review

Richard J. Bolton and David J. Hand

Abstract. Fraud is increasing dramatically with the expansion of modern technology and the global superhighways of communication, resulting in the loss of billions of dollars worldwide each year. Although prevention technologies are the best way to reduce fraud, fraudsters are adaptive and, given time, will usually find ways to circumvent such measures. Methodologies for the detection of fraud are essential if we are to catch fraudsters once fraud prevention has failed. Statistics and machine learning provide effective technologies for fraud detection and have been applied successfully to detect activities such as money laundering, e-commerce credit card fraud, telecommunications fraud and computer intrusion, to name but a few. We describe the tools available for statistical fraud detection and the areas in which fraud detection technologies are most used.

Key words and phrases: Fraud detection, fraud prevention, statistics, machine learning, money laundering, computer intrusion, e-commerce, credit cards, telecommunications.

1. INTRODUCTION

The *Concise Oxford Dictionary* defines fraud as "criminal deception; the use of false representations to gain an unjust advantage." Fraud is as old as humanity itself and can take an unlimited variety of different forms. However, in recent years, the development of new technologies (which have made it easier for us to communicate and helped increase our spending power) has also provided yet further ways in which criminals may commit fraud. Traditional forms of fraudulent behavior such as money laundering have become easier to perpetrate and have been joined by new kinds of fraud such as mobile telecommunications fraud and computer intrusion.

We begin by distinguishing between fraud prevention and fraud detection. Fraud *prevention* describes measures to stop fraud from occurring in the first place. These include elaborate designs, fluorescent fibers, multitone drawings, watermarks, laminated metal strips and holographs on banknotes, personal identification numbers for bankcards, Internet security systems for credit card transactions, Subscriber Identity Module (SIM) cards for mobile phones, and passwords on computer systems and telephone bank accounts. Of course, none of these methods is perfect and, in general, a compromise has to be struck between expense and inconvenience (e.g., to a customer) on the one hand, and effectiveness on the other.

In contrast, fraud *detection* involves identifying fraud as quickly as possible once it has been perpetrated. Fraud detection comes into play once fraud prevention has failed. In practice, of course fraud detection must be used continuously, as one will typically be unaware that fraud prevention has failed. We can try to prevent credit card fraud by guarding our cards assiduously, but if nevertheless the card's details are stolen, then we need to be able to detect, as soon as possible, that fraud is being perpetrated.

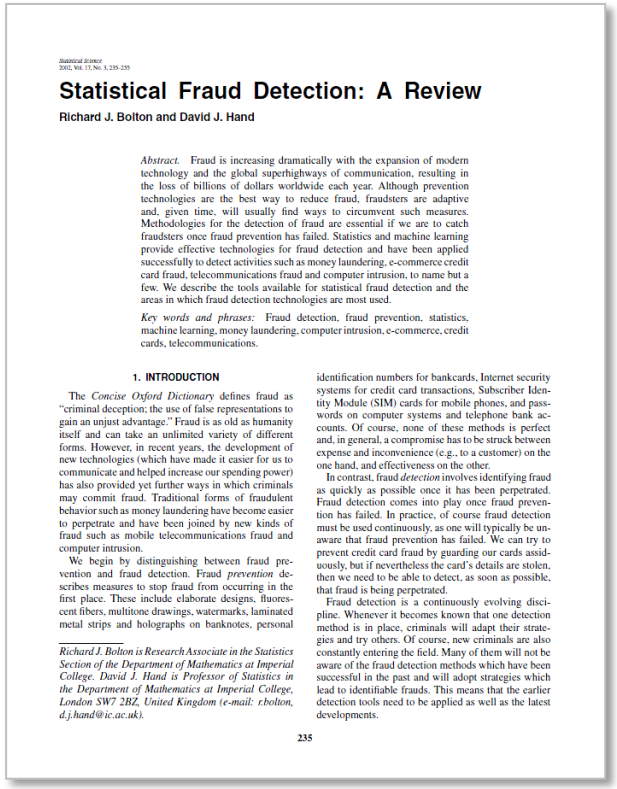
Fraud detection is a continuously evolving discipline. Whenever it becomes known that one detection method is in place, criminals will adapt their strategies and try others. Of course, new criminals are also constantly entering the field. Many of them will not be aware of the fraud detection methods which have been successful in the past and will adopt strategies which lead to identifiable frauds. This means that the earlier detection tools need to be applied as well as the latest developments.

Richard J. Bolton is Research Associate in the Statistics Section of the Department of Mathematics at Imperial College. David J. Hand is Professor of Statistics in the Department of Mathematics at Imperial College, London SW7 2BZ, United Kingdom (e-mail: r.bolton, d.j.hand@ic.ac.uk).

Richard J. Bolton
David J. Hand

2002

1. Мошенничество растет с развитием современных технологий
2. Превентивные меры рано или поздно обходятся => нужны детективные инструменты
3. Детективные алгоритмы тоже со временем обходятся => их надо постоянно развивать
4. Методики антифрода закрыты – это затрудняет исследование и развитие антифрода



Statistical Science
2002, Vol. 17, No. 3, 235–255

Statistical Fraud Detection: A Review

Richard J. Bolton and David J. Hand

Abstract. Fraud is increasing dramatically with the expansion of modern technology and the global superhighways of communication, resulting in the loss of billions of dollars worldwide each year. Although prevention technologies are the best way to reduce fraud, fraudsters are adaptive and, given time, will usually find ways to circumvent such measures. Methodologies for the detection of fraud are essential if we are to catch fraudsters once fraud prevention has failed. Statistics and machine learning provide effective technologies for fraud detection and have been applied successfully to detect activities such as money laundering, e-commerce credit card fraud, telecommunications fraud and computer intrusion, to name but a few. We describe the tools available for statistical fraud detection and the areas in which fraud detection technologies are most used.

Key words and phrases: Fraud detection, fraud prevention, statistics, machine learning, money laundering, computer intrusion, e-commerce, credit cards, telecommunications.

1. INTRODUCTION

The *Concise Oxford Dictionary* defines fraud as "criminal deception; the use of false representations to gain an unjust advantage." Fraud is as old as humanity itself and can take an unlimited variety of different forms. However, in recent years, the development of new technologies (which have made it easier for us to communicate and helped increase our spending power) has also provided yet further ways in which criminals may commit fraud. Traditional forms of fraudulent behavior such as money laundering have become easier to perpetrate and have been joined by new kinds of fraud such as mobile telecommunications fraud and computer intrusion.

We begin by distinguishing between fraud prevention and fraud detection. Fraud *prevention* describes measures to stop fraud from occurring in the first place. These include elaborate designs, fluorescent fibers, multitone drawings, watermarks, laminated metal strips and holographs on banknotes, personal identification numbers for bankcards, Internet security systems for credit card transactions, Subscriber Identity Module (SIM) cards for mobile phones, and passwords on computer systems and telephone bank accounts. Of course, none of these methods is perfect and, in general, a compromise has to be struck between expense and inconvenience (e.g., to a customer) on the one hand, and effectiveness on the other.

In contrast, fraud *detection* involves identifying fraud as quickly as possible once it has been perpetrated. Fraud detection comes into play once fraud prevention has failed. In practice, of course fraud detection must be used continuously, as one will typically be unaware that fraud prevention has failed. We can try to prevent credit card fraud by guarding our cards assiduously, but if nevertheless the card's details are stolen, then we need to be able to detect, as soon as possible, that fraud is being perpetrated.

Fraud detection is a continuously evolving discipline. Whenever it becomes known that one detection method is in place, criminals will adapt their strategies and try others. Of course, new criminals are also constantly entering the field. Many of them will not be aware of the fraud detection methods which have been successful in the past and will adopt strategies which lead to identifiable frauds. This means that the earlier detection tools need to be applied as well as the latest developments.

Richard J. Bolton is Research Associate in the Statistics Section of the Department of Mathematics at Imperial College. David J. Hand is Professor of Statistics in the Department of Mathematics at Imperial College, London SW7 2BZ, United Kingdom (e-mail: r.bolton, d.j.hand@ic.ac.uk).

Richard J. Bolton
David J. Hand

2002

5. Алгоритмы: «без учителя» – поиск аномалий; «с учителем» – выявление известных схем

6. Алгоритмы «без учителя» ловят в виде аномалий много операционных ошибок

7. Алгоритмы «с учителем» обучены на уже известных схемах => не ловят новые схемы

8. Проблема дисбаланса классов => алгоритмы много ошибаются => дорогой процесс

Statistical Science
2002, Vol. 17, No. 3, 235–255

Statistical Fraud Detection: A Review

Richard J. Bolton and David J. Hand

Comment

Foster Provost

The state of research on fraud detection recalls John Godfrey Saxe's 19th-century poem "The Blind Men and the Elephant" (Felleman, 1936, page 521). Based on a Hindu fable, each blind man experiences only a part of the elephant, which shapes his opinion of the nature of the elephant: the leg makes it seem like a tree, the tail a rope, the trunk a snake and so on. In fact, "...though each was partly in the right ... all were in the wrong." Saxe's poem was a criticism of theological debates, and I do not intend such a harsh criticism of research on fraud detection. However, because the problem is so complex, each research project takes a particular angle of attack, which often obscures the view of other parts of the problem. So, some researchers see the problem as one of classification, others of temporal pattern discovery; to some it is a problem perfect for a hidden Markov model and so on.

So why is fraud detection not simply classification or a member of some other already well-understood problem class? Bolton and Hand outline several characteristics of fraud detection problems that differentiate them [as did Tom Fawcett and I in our review of the problems and techniques of fraud detection (Faw-

Comment

Leo Breiman

This is an enjoyable and illuminating article. It deals with an area that few statisticians are aware of, but that is of critical importance economically and in terms of security. I am appreciative to the authors for the education in fraud detection this article gave me and to *Statistical Science* for publishing it. There are some interesting aspects that make this class of problems unique and that I comment on, running the risk of repeating points made in the article.

The analysis has to deal with a large number of problems simultaneously. For instance, in credit card fraud, the records of millions of customers have to be analyzed one by one to set up individual alarm settings. It is not a single unsupervised or supervised problem—a multitude of such problems have to be simultaneously addressed and "solved" for diverse data records. Yet the algorithm selected, modulo a few tunable parameters, has to be "one size fits all." Otherwise the on-line computations are not feasible. The alarm bell settings have to be constantly updated. For instance, as customers age and change their economic level and life styles, usage characteristics change. There are also serious database issues—how to structure the large databases so that the incoming streams of data are acces-

Foster Provost
Leo Breiman

2002

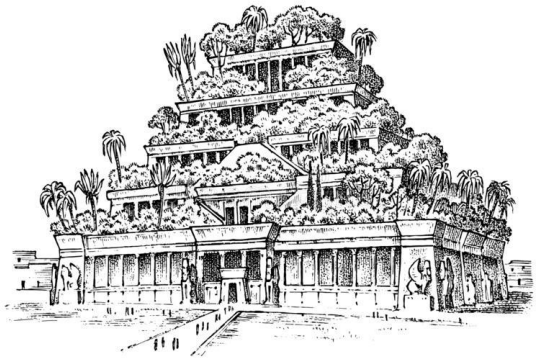
9. Мошенники быстро адаптируются => качество алгоритмов сильно падает после внедрения

10. Часто алгоритм настраивается под конкретную схему, хотя нужно думать о глобальной задаче

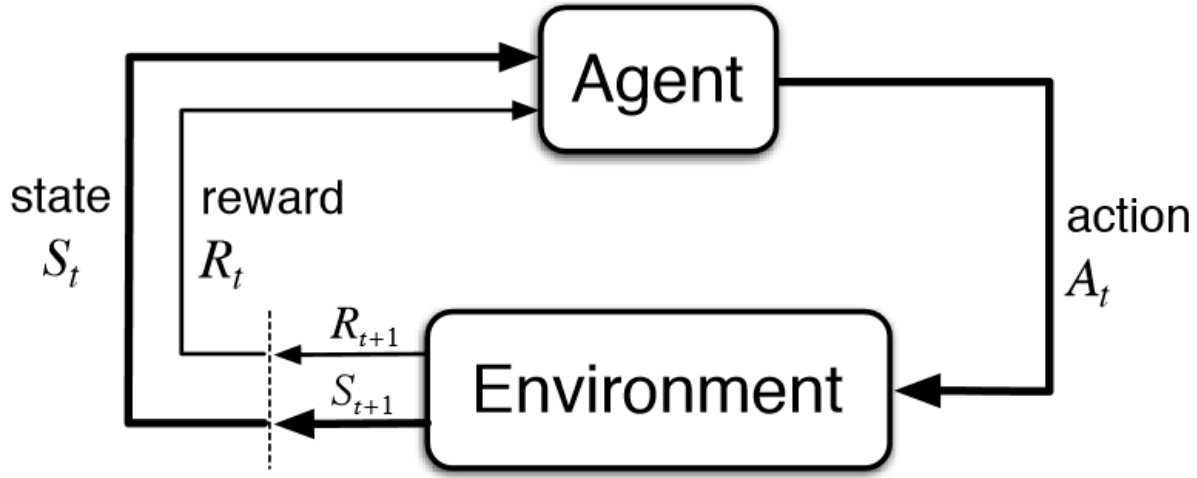
11. Для разработки эффективных алгоритмов нужны большие данные

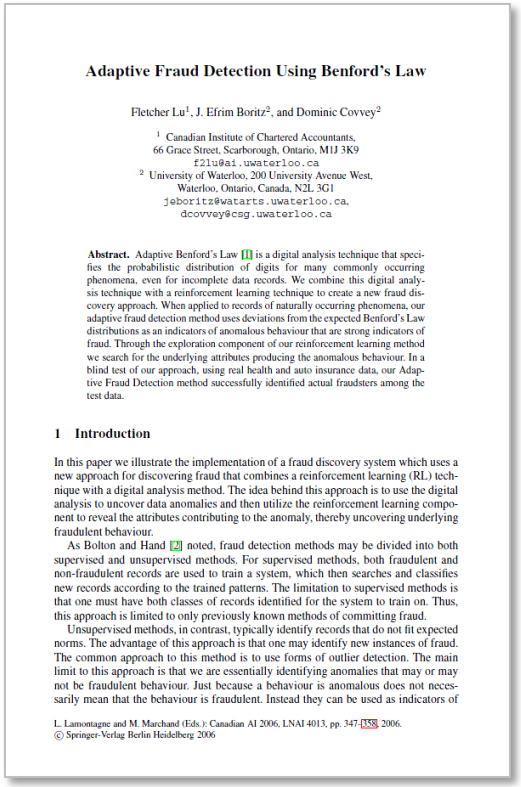
12. Выбор алгоритма не решает проблему, необходимо глубокое изучение данных

2. Обучение с подкреплением



RL — обучение с подкреплением





Adaptive Fraud Detection Using Benford's Law

Fletcher Lu[1], J. Efrim Boritz[2], and Dominic Covvey[2]

[1] Canadian Institute of Chartered Accountants,
66 Grace Street, Scarborough, Ontario, M1J 3K9
f2lu@ai.uwaterloo.ca
[2] University of Waterloo, 200 University Avenue West,
Waterloo, Ontario, Canada, N2L 3G1
jeboritz@watarts.uwaterloo.ca,
dcovvey@csg.uwaterloo.ca

Abstract. Adaptive Benford's Law [1] is a digital analysis technique that specifies the probabilistic distribution of digits for many commonly occurring phenomena, even for incomplete data records. We combine this digital analysis technique with a reinforcement learning technique to create a new fraud discovery approach. When applied to records of naturally occurring phenomena, our adaptive fraud detection method uses deviations from the expected Benford's Law distributions as an indicators of anomalous behaviour that are strong indicators of fraud. Through the exploration component of our reinforcement learning method we search for the underlying attributes producing the anomalous behaviour. In a blind test of our approach, using real health and auto insurance data, our Adaptive Fraud Detection method successfully identified actual fraudsters among the test data.

1 Introduction

In this paper we illustrate the implementation of a fraud discovery system which uses a new approach for discovering fraud that combines a reinforcement learning (RL) technique with a digital analysis method. The idea behind this approach is to use the digital analysis to uncover data anomalies and then utilize the reinforcement learning component to reveal the attributes contributing to the anomaly, thereby uncovering underlying fraudulent behaviour.

As Bolton and Hand [2] noted, fraud detection methods may be divided into both supervised and unsupervised methods. For supervised methods, both fraudulent and non-fraudulent records are used to train a system, which then searches and classifies new records according to the trained patterns. The limitation to supervised methods is that one must have both classes of records identified for the system to train on. Thus, this approach is limited to only previously known methods of committing fraud.

Unsupervised methods, in contrast, typically identify records that do not fit expected norms. The advantage of this approach is that one may identify new instances of fraud. The common approach to this method is to use forms of outlier detection. The main limit to this approach is that we are essentially identifying anomalies that may or may not be fraudulent behaviour. Just because a behaviour is anomalous does not necessarily mean that the behaviour is fraudulent. Instead they can be used as indicators of

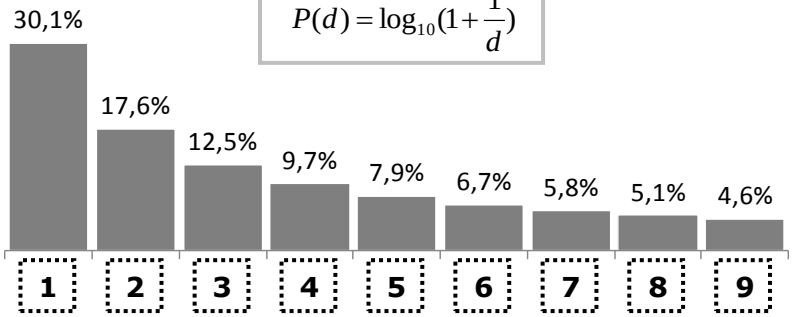
L. Lamontagne and M. Marchand (Eds.): Canadian AI 2006, LNAI 4013, pp. 347–358, 2006.

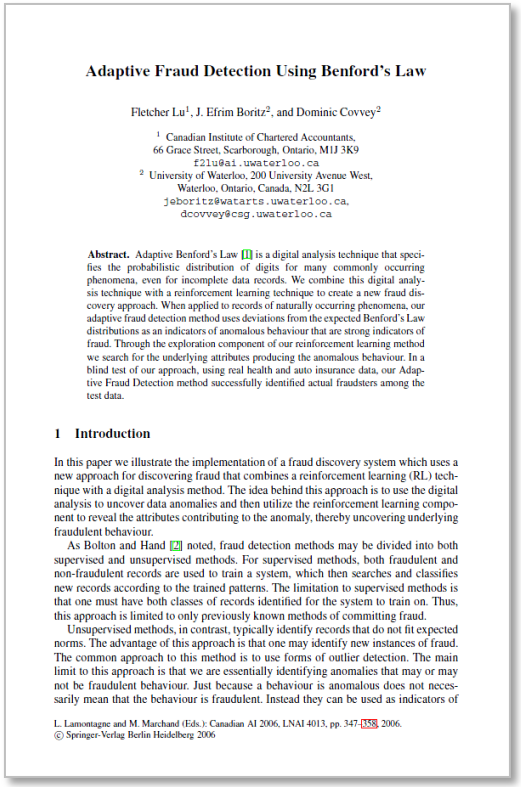
Fletcher Lu
J. Efrim Boritz & ko

2006

Benford's Low

$$P(d) = \log_{10}(1+\frac{1}{d})$$





Adaptive Fraud Detection Using Benford's Law

Fletcher Lu[1], J. Efrim Boritz[2], and Dominic Covvey[2]

[1] Canadian Institute of Chartered Accountants,
66 Grace Street, Scarborough, Ontario, M1J 3K9
f2lu@ai.uwaterloo.ca
[2] University of Waterloo, 200 University Avenue West,
Waterloo, Ontario, Canada, N2L 3G1
jeboritz@watarts.uwaterloo.ca,
dcovvey@csg.uwaterloo.ca

Abstract. Adaptive Benford's Law [1] is a digital analysis technique that specifies the probabilistic distribution of digits for many commonly occurring phenomena, even for incomplete data records. We combine this digital analysis technique with a reinforcement learning technique to create a new fraud discovery approach. When applied to records of naturally occurring phenomena, our adaptive fraud detection method uses deviations from the expected Benford's Law distributions as an indicators of anomalous behaviour that are strong indicators of fraud. Through the exploration component of our reinforcement learning method we search for the underlying attributes producing the anomalous behaviour. In a blind test of our approach, using real health and auto insurance data, our Adaptive Fraud Detection method successfully identified actual fraudsters among the test data.

1 Introduction

In this paper we illustrate the implementation of a fraud discovery system which uses a new approach for discovering fraud that combines a reinforcement learning (RL) technique with a digital analysis method. The idea behind this approach is to use the digital analysis to uncover data anomalies and then utilize the reinforcement learning component to reveal the attributes contributing to the anomaly, thereby uncovering underlying fraudulent behaviour.

As Bolton and Hand [2] noted, fraud detection methods may be divided into both supervised and unsupervised methods. For supervised methods, both fraudulent and non-fraudulent records are used to train a system, which then searches and classifies new records according to the trained patterns. The limitation to supervised methods is that one must have both classes of records identified for the system to train on. Thus, this approach is limited to only previously known methods of committing fraud.

Unsupervised methods, in contrast, typically identify records that do not fit expected norms. The advantage of this approach is that one may identify new instances of fraud. The common approach to this method is to use forms of outlier detection. The main limit to this approach is that we are essentially identifying anomalies that may or may not be fraudulent behaviour. Just because a behaviour is anomalous does not necessarily mean that the behaviour is fraudulent. Instead they can be used as indicators of

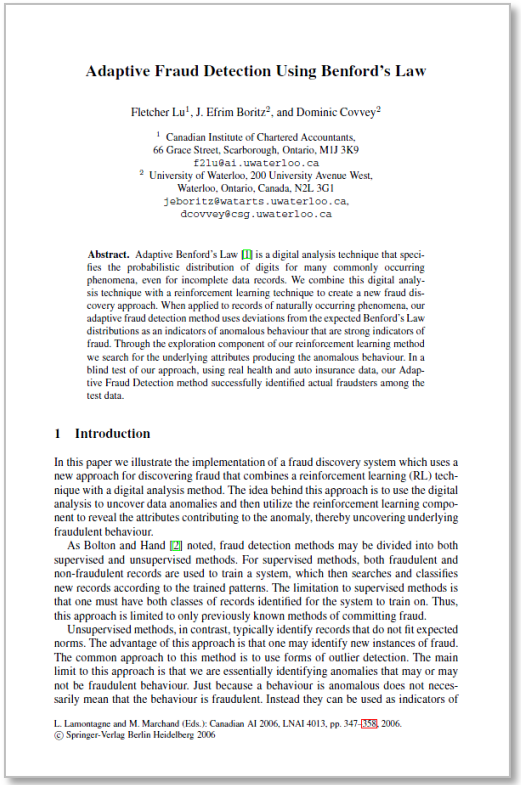
L. Lamontagne and M. Marchand (Eds.): Canadian AI 2006, LNAI 4013, pp. 347–358, 2006.

Fletcher Lu
J. Efrim Boritz & ko

2006

Reinforcement learning (RL)

State	Actions/Attributes				Digit Sequences	Rewards/ Magnitude of Anomalies
	Purchase Item	Store	Location	Form of Payment		
1	shoes	storeA	street15	credit	$52	1.6
2	hat	storeB	street12	cash	$38	3.2
3	hat	storeC	street12	debit	$22	6.2
4	TV	storeB	street12	cheque	$640	1.1



Adaptive Fraud Detection Using Benford's Law

Fletcher Lu[1], J. Efrim Boritz[2], and Dominic Covvey[2]

[1] Canadian Institute of Chartered Accountants,
66 Grace Street, Scarborough, Ontario, M1J 3K9
f2lu@ai.uwaterloo.ca
[2] University of Waterloo, 200 University Avenue West,
Waterloo, Ontario, Canada, N2L 3G1
jeboritz@watarts.uwaterloo.ca,
dcovvey@csg.uwaterloo.ca

Abstract. Adaptive Benford's Law [1] is a digital analysis technique that specifies the probabilistic distribution of digits for many commonly occurring phenomena, even for incomplete data records. We combine this digital analysis technique with a reinforcement learning technique to create a new fraud discovery approach. When applied to records of naturally occurring phenomena, our adaptive fraud detection method uses deviations from the expected Benford's Law distributions as an indicators of anomalous behaviour that are strong indicators of fraud. Through the exploration component of our reinforcement learning method we search for the underlying attributes producing the anomalous behaviour. In a blind test of our approach, using real health and auto insurance data, our Adaptive Fraud Detection method successfully identified actual fraudsters among the test data.

1 Introduction

In this paper we illustrate the implementation of a fraud discovery system which uses a new approach for discovering fraud that combines a reinforcement learning (RL) technique with a digital analysis method. The idea behind this approach is to use the digital analysis to uncover data anomalies and then utilize the reinforcement learning component to reveal the attributes contributing to the anomaly, thereby uncovering underlying fraudulent behaviour.

As Bolton and Hand [2] noted, fraud detection methods may be divided into both supervised and unsupervised methods. For supervised methods, both fraudulent and non-fraudulent records are used to train a system, which then searches and classifies new records according to the trained patterns. The limitation to supervised methods is that one must have both classes of records identified for the system to train on. Thus, this approach is limited to only previously known methods of committing fraud.

Unsupervised methods, in contrast, typically identify records that do not fit expected norms. The advantage of this approach is that one may identify new instances of fraud. The common approach to this method is to use forms of outlier detection. The main limit to this approach is that we are essentially identifying anomalies that may or may not be fraudulent behaviour. Just because a behaviour is anomalous does not necessarily mean that the behaviour is fraudulent. Instead they can be used as indicators of

L. Lamontagne and M. Marchand (Eds.): Canadian AI 2006, LNAI 4013, pp. 347–358, 2006.

Fletcher Lu
J. Efrim Boritz & ko

2006

Медицинское страхование

31 804 записи
94 признака

Авто страхование

17 640 записей
31 признак

Adaptive Fraud Detection Using Benford's Law

Fletcher Lu[1], J. Efrim Boritz[2], and Dominic Covvey[2]

[1] Canadian Institute of Chartered Accountants,
66 Grace Street, Scarborough, Ontario, M1J 3K9
f2lu@ai.uwaterloo.ca
[2] University of Waterloo, 200 University Avenue West,
Waterloo, Ontario, Canada, N2L 3G1
jeboritz@watarts.uwaterloo.ca,
dcovvey@csg.uwaterloo.ca

Abstract. Adaptive Benford's Law [1] is a digital analysis technique that specifies the probabilistic distribution of digits for many commonly occurring phenomena, even for incomplete data records. We combine this digital analysis technique with a reinforcement learning technique to create a new fraud discovery approach. When applied to records of naturally occurring phenomena, our adaptive fraud detection method uses deviations from the expected Benford's Law distributions as an indicators of anomalous behaviour that are strong indicators of fraud. Through the exploration component of our reinforcement learning method we search for the underlying attributes producing the anomalous behaviour. In a blind test of our approach, using real health and auto insurance data, our Adaptive Fraud Detection method successfully identified actual fraudsters among the test data.

1 Introduction

In this paper we illustrate the implementation of a fraud discovery system which uses a new approach for discovering fraud that combines a reinforcement learning (RL) technique with a digital analysis method. The idea behind this approach is to use the digital analysis to uncover data anomalies and then utilize the reinforcement learning component to reveal the attributes contributing to the anomaly, thereby uncovering underlying fraudulent behaviour.

As Bolton and Hand [2] noted, fraud detection methods may be divided into both supervised and unsupervised methods. For supervised methods, both fraudulent and non-fraudulent records are used to train a system, which then searches and classifies new records according to the trained patterns. The limitation to supervised methods is that one must have both classes of records identified for the system to train on. Thus, this approach is limited to only previously known methods of committing fraud.

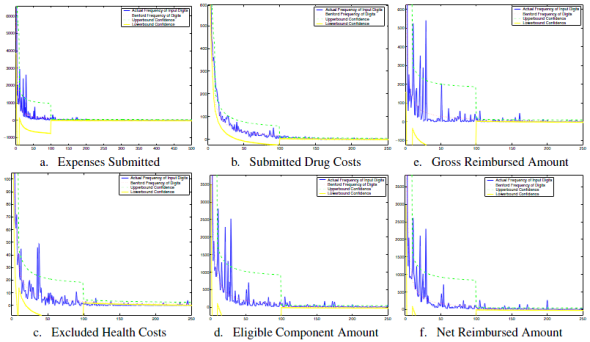
Unsupervised methods, in contrast, typically identify records that do not fit expected norms. The advantage of this approach is that one may identify new instances of fraud. The common approach to this method is to use forms of outlier detection. The main limit to this approach is that we are essentially identifying anomalies that may or may not be fraudulent behaviour. Just because a behaviour is anomalous does not necessarily mean that the behaviour is fraudulent. Instead they can be used as indicators of

L. Lamontagne and M. Marchand (Eds.): Canadian AI 2006, LNAI 4013, pp. 347–358, 2006.

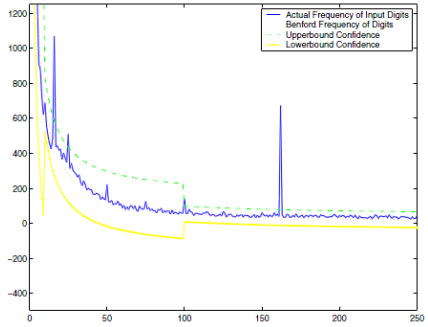
Fletcher Lu
J. Efrim Boritz & ko

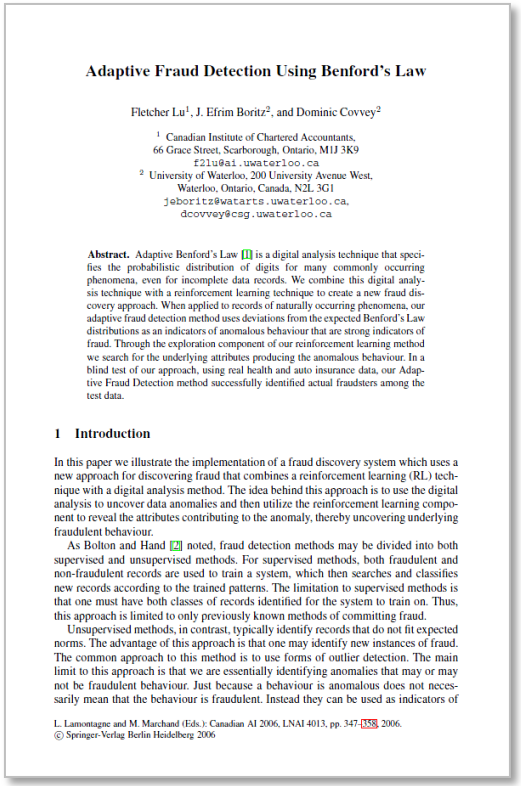
2006

Healthcare Digit Frequencies compared with their Benford Distributions



Auto Insurance Digit Frequencies





Adaptive Fraud Detection Using Benford's Law

Fletcher Lu[1], J. Efrim Boritz[2], and Dominic Covvey[2]

[1] Canadian Institute of Chartered Accountants,
66 Grace Street, Scarborough, Ontario, M1J 3K9
f2lu@ai.uwaterloo.ca
[2] University of Waterloo, 200 University Avenue West,
Waterloo, Ontario, Canada, N2L 3G1
jeboritz@watarts.uwaterloo.ca,
dcovvey@csg.uwaterloo.ca

Abstract. Adaptive Benford's Law [1] is a digital analysis technique that specifies the probabilistic distribution of digits for many commonly occurring phenomena, even for incomplete data records. We combine this digital analysis technique with a reinforcement learning technique to create a new fraud discovery approach. When applied to records of naturally occurring phenomena, our adaptive fraud detection method uses deviations from the expected Benford's Law distributions as an indicators of anomalous behaviour that are strong indicators of fraud. Through the exploration component of our reinforcement learning method we search for the underlying attributes producing the anomalous behaviour. In a blind test of our approach, using real health and auto insurance data, our Adaptive Fraud Detection method successfully identified actual fraudsters among the test data.

1 Introduction

In this paper we illustrate the implementation of a fraud discovery system which uses a new approach for discovering fraud that combines a reinforcement learning (RL) technique with a digital analysis method. The idea behind this approach is to use the digital analysis to uncover data anomalies and then utilize the reinforcement learning component to reveal the attributes contributing to the anomaly, thereby uncovering underlying fraudulent behaviour.

As Bolton and Hand [2] noted, fraud detection methods may be divided into both supervised and unsupervised methods. For supervised methods, both fraudulent and non-fraudulent records are used to train a system, which then searches and classifies new records according to the trained patterns. The limitation to supervised methods is that one must have both classes of records identified for the system to train on. Thus, this approach is limited to only previously known methods of committing fraud.

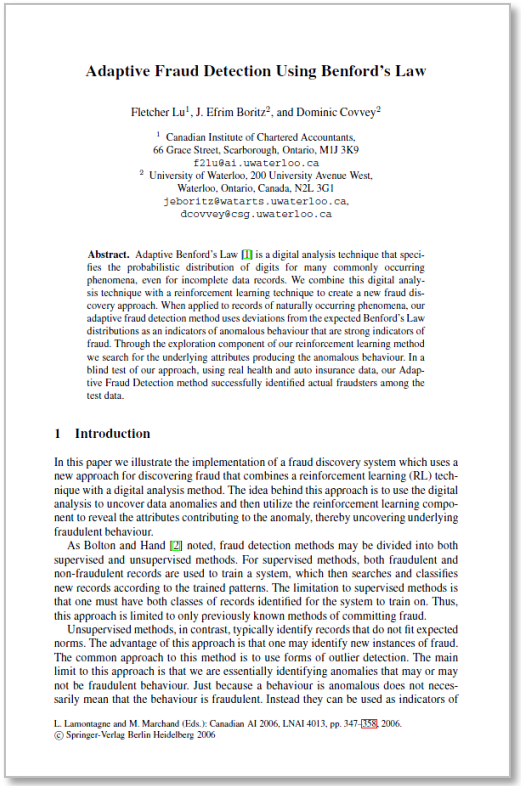
Unsupervised methods, in contrast, typically identify records that do not fit expected norms. The advantage of this approach is that one may identify new instances of fraud. The common approach to this method is to use forms of outlier detection. The main limit to this approach is that we are essentially identifying anomalies that may or may not be fraudulent behaviour. Just because a behaviour is anomalous does not necessarily mean that the behaviour is fraudulent. Instead they can be used as indicators of

L. Lamontagne and M. Marchand (Eds.): Canadian AI 2006, LNAI 4013, pp. 347–358, 2006.

Fletcher Lu
J. Efrim Boritz & ko

2006

- RL – узко специализированное направление. Главная книга по RL написана в 1998 году
- В теории RL много подводных камней и нерешенных проблем
- В статье отсутствует сравнение RL с классическими методами ML



Adaptive Fraud Detection Using Benford's Law

Fletcher Lu[1], J. Efrim Boritz[2], and Dominic Covvey[2]

[1] Canadian Institute of Chartered Accountants,
66 Grace Street, Scarborough, Ontario, M1J 3K9
f2lu@ai.uwaterloo.ca
[2] University of Waterloo, 200 University Avenue West,
Waterloo, Ontario, Canada, N2L 3G1
jeboritz@watarts.uwaterloo.ca,
dcovvey@csg.uwaterloo.ca

Abstract. Adaptive Benford's Law [1] is a digital analysis technique that specifies the probabilistic distribution of digits for many commonly occurring phenomena, even for incomplete data records. We combine this digital analysis technique with a reinforcement learning technique to create a new fraud discovery approach. When applied to records of naturally occurring phenomena, our adaptive fraud detection method uses deviations from the expected Benford's Law distributions as an indicators of anomalous behaviour that are strong indicators of fraud. Through the exploration component of our reinforcement learning method we search for the underlying attributes producing the anomalous behaviour. In a blind test of our approach, using real health and auto insurance data, our Adaptive Fraud Detection method successfully identified actual fraudsters among the test data.

1 Introduction

In this paper we illustrate the implementation of a fraud discovery system which uses a new approach for discovering fraud that combines a reinforcement learning (RL) technique with a digital analysis method. The idea behind this approach is to use the digital analysis to uncover data anomalies and then utilize the reinforcement learning component to reveal the attributes contributing to the anomaly, thereby uncovering underlying fraudulent behaviour.

As Bolton and Hand [2] noted, fraud detection methods may be divided into both supervised and unsupervised methods. For supervised methods, both fraudulent and non-fraudulent records are used to train a system, which then searches and classifies new records according to the trained patterns. The limitation to supervised methods is that one must have both classes of records identified for the system to train on. Thus, this approach is limited to only previously known methods of committing fraud.

Unsupervised methods, in contrast, typically identify records that do not fit expected norms. The advantage of this approach is that one may identify new instances of fraud. The common approach to this method is to use forms of outlier detection. The main limit to this approach is that we are essentially identifying anomalies that may or may not be fraudulent behaviour. Just because a behaviour is anomalous does not necessarily mean that the behaviour is fraudulent. Instead they can be used as indicators of

L. Lamontagne and M. Marchand (Eds.): Canadian AI 2006, LNAI 4013, pp. 347–358, 2006.

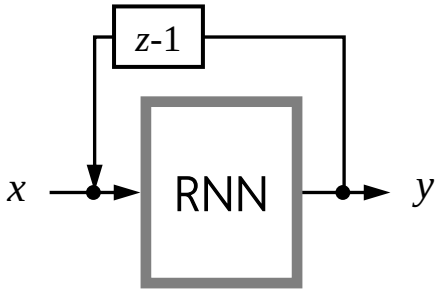
Fletcher Lu
J. Efrim Boritz & ko

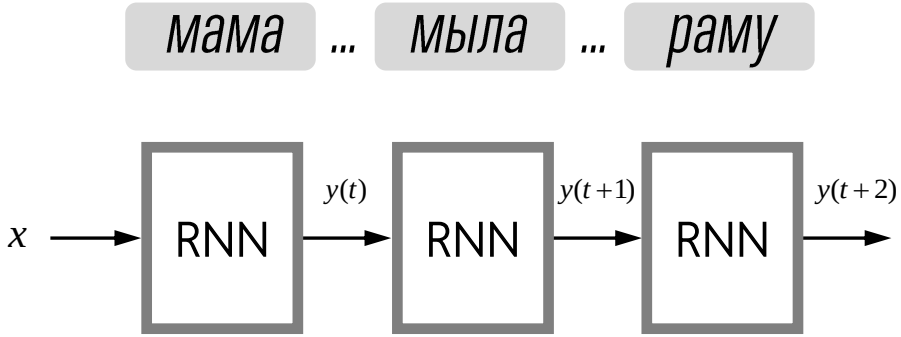
2006

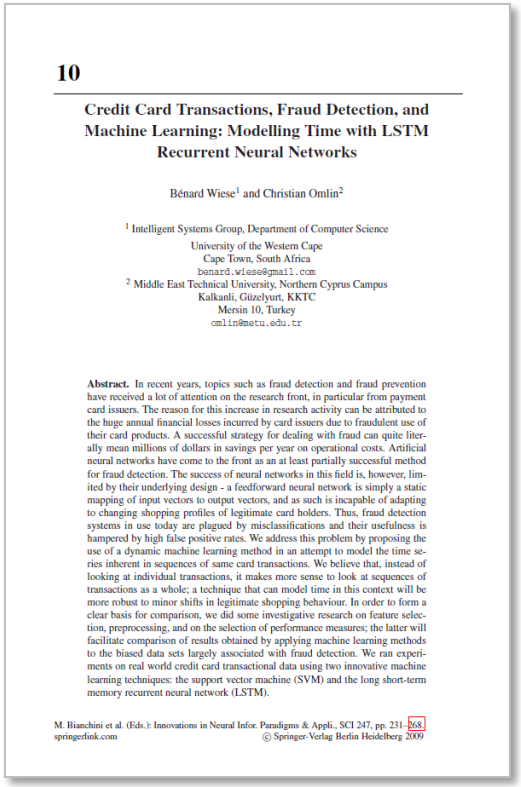
- Возможно **единственное исследование** о применении RL в антифроде
- В предложенном методе заложены **принципы работы аудитора**
- В адаптивном Бенфорд-методе **сняты ограничения** классического закона

3. Рекуррентные нейронные сети

RNN — рекуррентные нейронные сети







10

Credit Card Transactions, Fraud Detection, and Machine Learning: Modelling Time with LSTM Recurrent Neural Networks

Bénard Wiese[1] and Christian Omlin[2]

[1] Intelligent Systems Group, Department of Computer Science
University of the Western Cape
Cape Town, South Africa
benard.wiese@gmail.com
[2] Middle East Technical University, Northern Cyprus Campus
Kalkanli, Güzelyurt, KKTC
Mersin 10, Turkey
omlin@metu.edu.tr

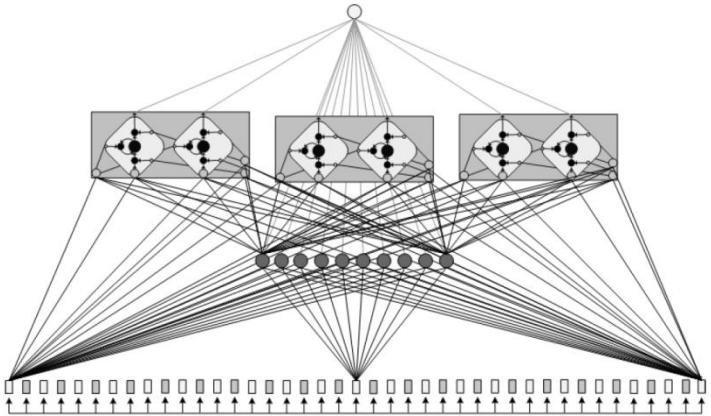
Abstract. In recent years, topics such as fraud detection and fraud prevention have received a lot of attention on the research front, in particular from payment card issuers. The reason for this increase in research activity can be attributed to the huge annual financial losses incurred by card issuers due to fraudulent use of their card products. A successful strategy for dealing with fraud can quite literally mean millions of dollars in savings per year on operational costs. Artificial neural networks have come to the front as an at least partially successful method for fraud detection. The success of neural networks in this field is, however, limited by their underlying design - a feedforward neural network is simply a static mapping of input vectors to output vectors, and as such is incapable of adapting to changing shopping profiles of legitimate card holders. Thus, fraud detection systems in use today are plagued by misclassifications and their usefulness is hampered by high false positive rates. We address this problem by proposing the use of a dynamic machine learning method in an attempt to model the time series inherent in sequences of same card transactions. We believe that, instead of looking at individual transactions, it makes more sense to look at sequences of transactions as a whole; a technique that can model time in this context will be more robust to minor shifts in legitimate shopping behaviour. In order to form a clear basis for comparison, we did some investigative research on feature selection, preprocessing, and on the selection of performance measures; the latter will facilitate comparison of results obtained by applying machine learning methods to the biased data sets largely associated with fraud detection. We ran experiments on real world credit card transactional data using two innovative machine learning techniques: the support vector machine (SVM) and the long short-term memory recurrent neural network (LSTM).

M. Bianchini et al. (Eds.): Innovations in Neural Infor. Paradigms & Appli., SCI 247, pp. 231–268.
springerlink.com

Benard Wiese
Christian Omlin

2009

LSTM Recurrent Neural Networks



10

Credit Card Transactions, Fraud Detection, and Machine Learning: Modelling Time with LSTM Recurrent Neural Networks

Bénard Wiese[1] and Christian Omlin[2]

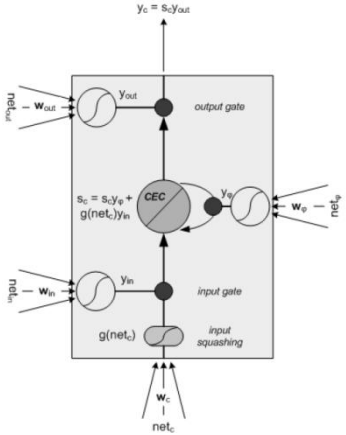
[1] Intelligent Systems Group, Department of Computer Science
University of the Western Cape
Cape Town, South Africa
benard.wiese@gmail.com
[2] Middle East Technical University, Northern Cyprus Campus
Kalkanli, Güzelyurt, KKTC
Mersin 10, Turkey
omlin@metu.edu.tr

Abstract. In recent years, topics such as fraud detection and fraud prevention have received a lot of attention on the research front, in particular from payment card issuers. The reason for this increase in research activity can be attributed to the huge annual financial losses incurred by card issuers due to fraudulent use of their card products. A successful strategy for dealing with fraud can quite literally mean millions of dollars in savings per year on operational costs. Artificial neural networks have come to the front as an at least partially successful method for fraud detection. The success of neural networks in this field is, however, limited by their underlying design - a feedforward neural network is simply a static mapping of input vectors to output vectors, and as such is incapable of adapting to changing shopping profiles of legitimate card holders. Thus, fraud detection systems in use today are plagued by misclassifications and their usefulness is hampered by high false positive rates. We address this problem by proposing the use of a dynamic machine learning method in an attempt to model the time series inherent in sequences of same card transactions. We believe that, instead of looking at individual transactions, it makes more sense to look at sequences of transactions as a whole; a technique that can model time in this context will be more robust to minor shifts in legitimate shopping behaviour. In order to form a clear basis for comparison, we did some investigative research on feature selection, preprocessing, and on the selection of performance measures; the latter will facilitate comparison of results obtained by applying machine learning methods to the biased data sets largely associated with fraud detection. We ran experiments on real world credit card transactional data using two innovative machine learning techniques: the support vector machine (SVM) and the long short-term memory recurrent neural network (LSTM).

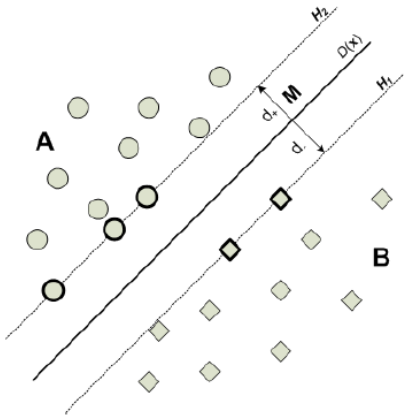
M. Bianchini et al. (Eds.): Innovations in Neural Infor. Paradigms & Appli., SCI 247, pp. 231–268.
springerlink.com

Benard Wiese
Christian Omlin

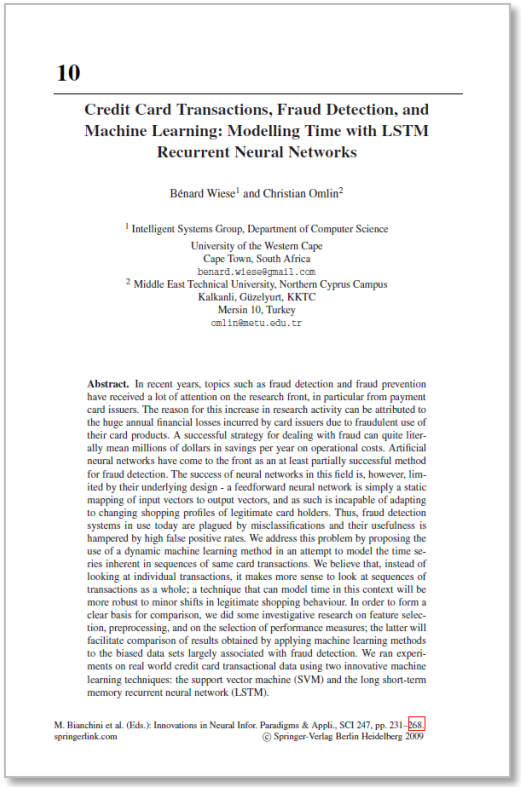
2009



LSTM



SVM



10

Credit Card Transactions, Fraud Detection, and Machine Learning: Modelling Time with LSTM Recurrent Neural Networks

Bénard Wiese[1] and Christian Omlin[2]

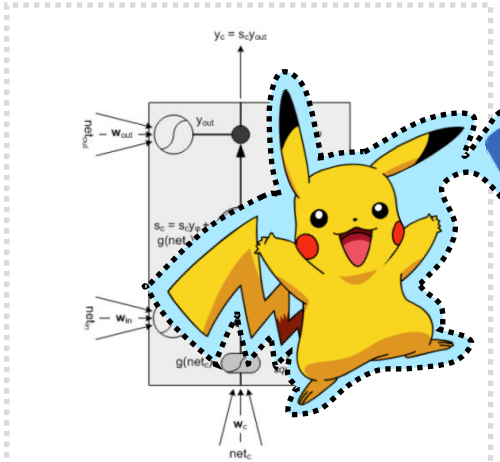
[1] Intelligent Systems Group, Department of Computer Science
University of the Western Cape
Cape Town, South Africa
benard.wiese@gmail.com
[2] Middle East Technical University, Northern Cyprus Campus
Kalkanli, Güzelyurt, KKTC
Mersin 10, Turkey
omlin@metu.edu.tr

Abstract. In recent years, topics such as fraud detection and fraud prevention have received a lot of attention on the research front, in particular from payment card issuers. The reason for this increase in research activity can be attributed to the huge annual financial losses incurred by card issuers due to fraudulent use of their card products. A successful strategy for dealing with fraud can quite literally mean millions of dollars in savings per year on operational costs. Artificial neural networks have come to the front as an at least partially successful method for fraud detection. The success of neural networks in this field is, however, limited by their underlying design - a feedforward neural network is simply a static mapping of input vectors to output vectors, and as such is incapable of adapting to changing shopping profiles of legitimate card holders. Thus, fraud detection systems in use today are plagued by misclassifications and their usefulness is hampered by high false positive rates. We address this problem by proposing the use of a dynamic machine learning method in an attempt to model the time series inherent in sequences of same card transactions. We believe that, instead of looking at individual transactions, it makes more sense to look at sequences of transactions as a whole; a technique that can model time in this context will be more robust to minor shifts in legitimate shopping behaviour. In order to form a clear basis for comparison, we did some investigative research on feature selection, preprocessing, and on the selection of performance measures; the latter will facilitate comparison of results obtained by applying machine learning methods to the biased data sets largely associated with fraud detection. We ran experiments on real world credit card transactional data using two innovative machine learning techniques: the support vector machine (SVM) and the long short-term memory recurrent neural network (LSTM).

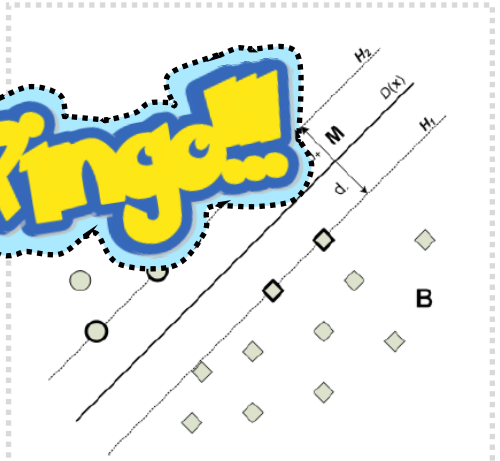
M. Bianchini et al. (Eds.): Innovations in Neural Infor. Paradigms & Appli., SCI 247, pp. 231–268.
springerlink.com

Benard Wiese
Christian Omlin

2009

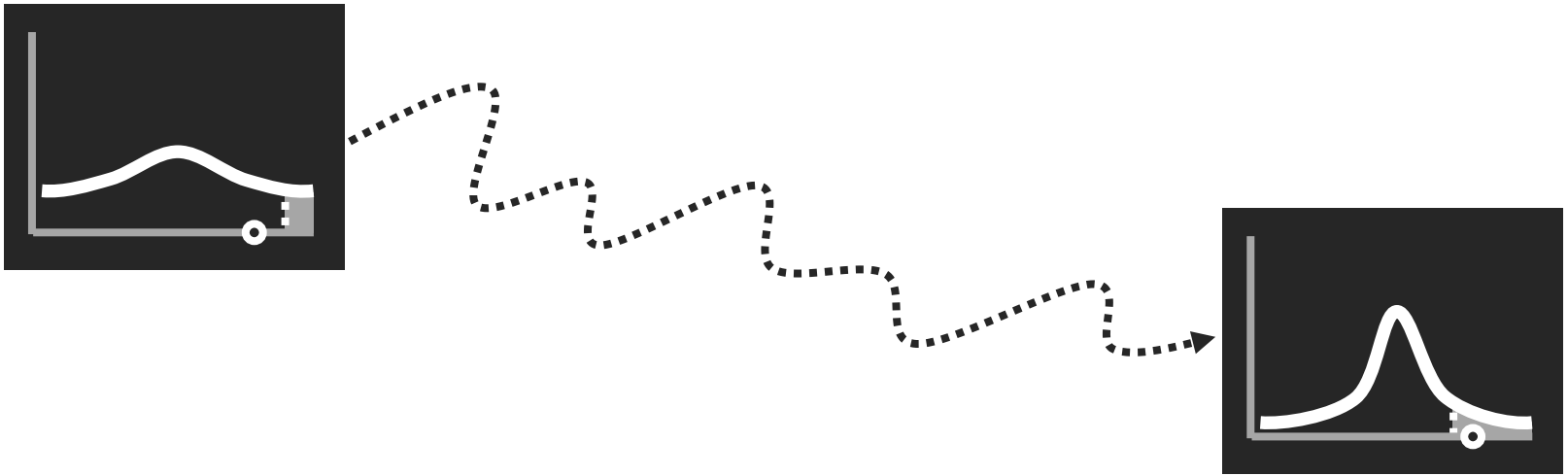


LSTM



SVM

P-hacking — бич современной науки



P-hacking — бич современной науки

P-hacking — бич современной науки

P-hacking — бич современной науки

P-hacking — бич современной науки

P-hacking — бич современной науки

Психология

64% психологических экспериментов не воспроизводятся

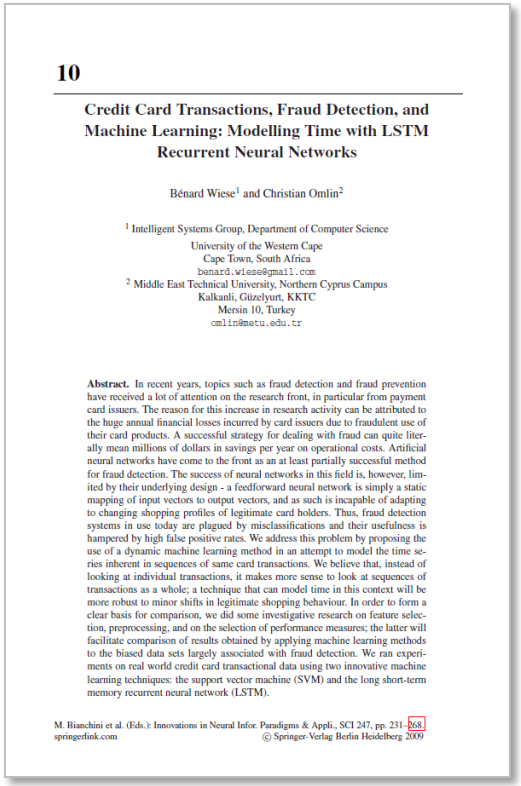
Социология

75% экспериментов в социальной психологии не воспроизводятся

Data Science

Какой % исследований в DS не реплицируется?

Пока не посчитали



10

Credit Card Transactions, Fraud Detection, and Machine Learning: Modelling Time with LSTM Recurrent Neural Networks

Bénard Wiese[1] and Christian Omlin[2]

[1] Intelligent Systems Group, Department of Computer Science
University of the Western Cape
Cape Town, South Africa
benard.wiese@gmail.com
[2] Middle East Technical University, Northern Cyprus Campus
Kalkanli, Güzelyurt, KKTC
Mersin 10, Turkey
omlin@metu.edu.tr

Abstract. In recent years, topics such as fraud detection and fraud prevention have received a lot of attention on the research front, in particular from payment card issuers. The reason for this increase in research activity can be attributed to the huge annual financial losses incurred by card issuers due to fraudulent use of their card products. A successful strategy for dealing with fraud can quite literally mean millions of dollars in savings per year on operational costs. Artificial neural networks have come to the front as an at least partially successful method for fraud detection. The success of neural networks in this field is, however, limited by their underlying design - a feedforward neural network is simply a static mapping of input vectors to output vectors, and as such is incapable of adapting to changing shopping profiles of legitimate card holders. Thus, fraud detection systems in use today are plagued by misclassifications and their usefulness is hampered by high false positive rates. We address this problem by proposing the use of a dynamic machine learning method in an attempt to model the time series inherent in sequences of same card transactions. We believe that, instead of looking at individual transactions, it makes more sense to look at sequences of transactions as a whole; a technique that can model time in this context will be more robust to minor shifts in legitimate shopping behaviour. In order to form a clear basis for comparison, we did some investigative research on feature selection, preprocessing, and on the selection of performance measures; the latter will facilitate comparison of results obtained by applying machine learning methods to the biased data sets largely associated with fraud detection. We ran experiments on real world credit card transactional data using two innovative machine learning techniques: the support vector machine (SVM) and the long short-term memory recurrent neural network (LSTM).

M. Bianchini et al. (Eds.): Innovations in Neural Infor. Paradigms & Appli., SCI 247, pp. 231–268.
springerlink.com

Benard Wiese
Christian Omlin

2009

Выборка	30 876 all (1% fraud)
Метрики	ROC AUC, MSE
Алгоритмы	LSTM vs. SVM

10

Credit Card Transactions, Fraud Detection, and Machine Learning: Modelling Time with LSTM Recurrent Neural Networks

Bénard Wiese[1] and Christian Omlin[2]

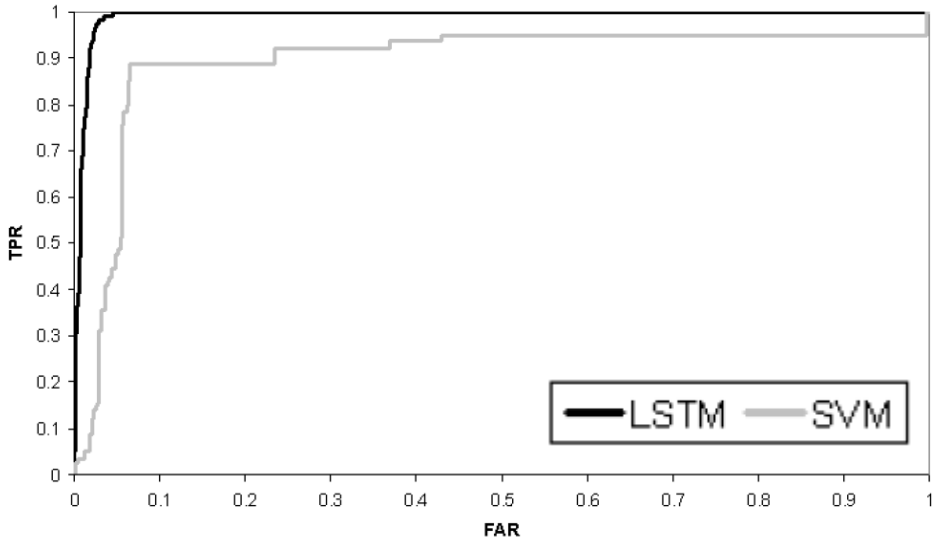
[1] Intelligent Systems Group, Department of Computer Science
University of the Western Cape
Cape Town, South Africa
benard.wiese@gmail.com
[2] Middle East Technical University, Northern Cyprus Campus
Kalkanli, Güzelyurt, KKTC
Mersin 10, Turkey
omlin@metu.edu.tr

Abstract. In recent years, topics such as fraud detection and fraud prevention have received a lot of attention on the research front, in particular from payment card issuers. The reason for this increase in research activity can be attributed to the huge annual financial losses incurred by card issuers due to fraudulent use of their card products. A successful strategy for dealing with fraud can quite literally mean millions of dollars in savings per year on operational costs. Artificial neural networks have come to the front as an at least partially successful method for fraud detection. The success of neural networks in this field is, however, limited by their underlying design - a feedforward neural network is simply a static mapping of input vectors to output vectors, and as such is incapable of adapting to changing shopping profiles of legitimate card holders. Thus, fraud detection systems in use today are plagued by misclassifications and their usefulness is hampered by high false positive rates. We address this problem by proposing the use of a dynamic machine learning method in an attempt to model the time series inherent in sequences of same card transactions. We believe that, instead of looking at individual transactions, it makes more sense to look at sequences of transactions as a whole; a technique that can model time in this context will be more robust to minor shifts in legitimate shopping behaviour. In order to form a clear basis for comparison, we did some investigative research on feature selection, preprocessing, and on the selection of performance measures; the latter will facilitate comparison of results obtained by applying machine learning methods to the biased data sets largely associated with fraud detection. We ran experiments on real world credit card transactional data using two innovative machine learning techniques: the support vector machine (SVM) and the long short-term memory recurrent neural network (LSTM).

M. Bianchini et al. (Eds.): Innovations in Neural Infor. Paradigms & Appli., SCI 247, pp. 231–268.
springerlink.com

Benard Wiese
Christian Omlin

2009



Credit Card Transactions, Fraud Detection, and Machine Learning: Modelling Time with LSTM Recurrent Neural Networks

by

Bénard Jacobus Wiese

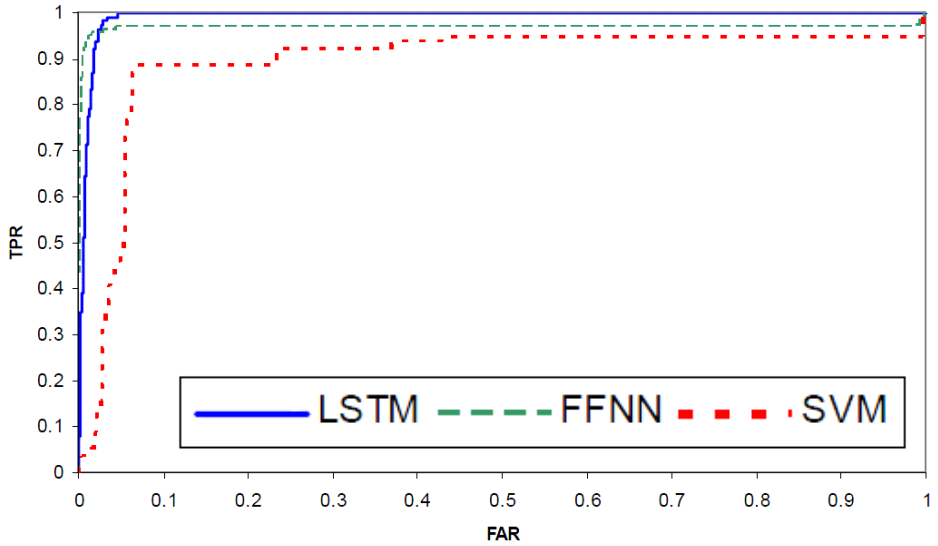
A thesis submitted in fulfilment of the requirements for the degree of
Magister Scientiae
in the Department of Computer Science,
University of the Western Cape.

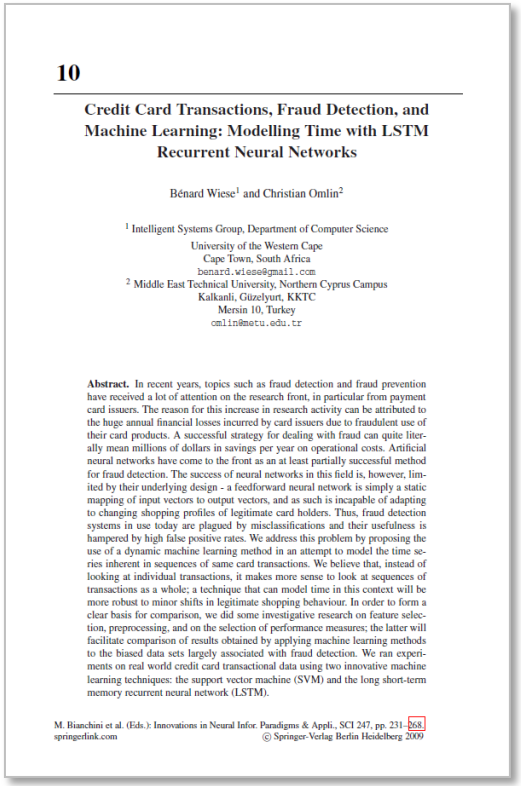
August 2007

Supervisor: Professor Christian W. Omlin

Benard Wiese
Christian Omlin

2007





10

Credit Card Transactions, Fraud Detection, and Machine Learning: Modelling Time with LSTM Recurrent Neural Networks

Bénard Wiese[1] and Christian Omlin[2]

[1] Intelligent Systems Group, Department of Computer Science
University of the Western Cape
Cape Town, South Africa
benard.wiese@gmail.com
[2] Middle East Technical University, Northern Cyprus Campus
Kalkanli, Güzelyurt, KKTC
Mersin 10, Turkey
omlin@metu.edu.tr

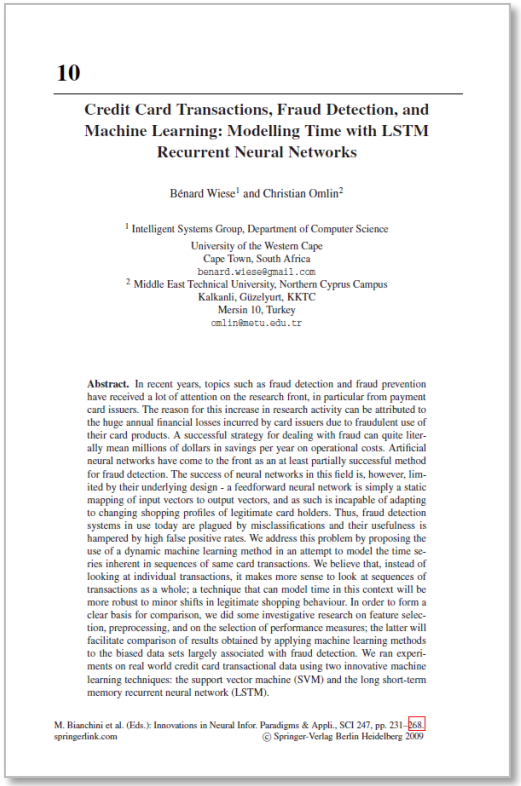
Abstract. In recent years, topics such as fraud detection and fraud prevention have received a lot of attention on the research front, in particular from payment card issuers. The reason for this increase in research activity can be attributed to the huge annual financial losses incurred by card issuers due to fraudulent use of their card products. A successful strategy for dealing with fraud can quite literally mean millions of dollars in savings per year on operational costs. Artificial neural networks have come to the front as an at least partially successful method for fraud detection. The success of neural networks in this field is, however, limited by their underlying design - a feedforward neural network is simply a static mapping of input vectors to output vectors, and as such is incapable of adapting to changing shopping profiles of legitimate card holders. Thus, fraud detection systems in use today are plagued by misclassifications and their usefulness is hampered by high false positive rates. We address this problem by proposing the use of a dynamic machine learning method in an attempt to model the time series inherent in sequences of same card transactions. We believe that, instead of looking at individual transactions, it makes more sense to look at sequences of transactions as a whole; a technique that can model time in this context will be more robust to minor shifts in legitimate shopping behaviour. In order to form a clear basis for comparison, we did some investigative research on feature selection, preprocessing, and on the selection of performance measures; the latter will facilitate comparison of results obtained by applying machine learning methods to the biased data sets largely associated with fraud detection. We ran experiments on real world credit card transactional data using two innovative machine learning techniques: the support vector machine (SVM) and the long short-term memory recurrent neural network (LSTM).

M. Bianchini et al. (Eds.): Innovations in Neural Infor. Paradigms & Appli., SCI 247, pp. 231–268.
springerlink.com

Benard Wiese
Christian Omlin

2009

- LSTM нейронную сеть **сравнивают со слабым алгоритмом** SVM
- Используется **не лучшая метрика** ROC AUC для несбалансированной выборки
- **Скрыли результаты** сопоставимой по качеству полносвязной нейронной сети



10

Credit Card Transactions, Fraud Detection, and Machine Learning: Modelling Time with LSTM Recurrent Neural Networks

Bénard Wiese[1] and Christian Omlin[2]

[1] Intelligent Systems Group, Department of Computer Science
University of the Western Cape
Cape Town, South Africa
benard.wiese@gmail.com
[2] Middle East Technical University, Northern Cyprus Campus
Kalkanli, Güzelyurt, KKTC
Mersin 10, Turkey
omlin@metu.edu.tr

Abstract. In recent years, topics such as fraud detection and fraud prevention have received a lot of attention on the research front, in particular from payment card issuers. The reason for this increase in research activity can be attributed to the huge annual financial losses incurred by card issuers due to fraudulent use of their card products. A successful strategy for dealing with fraud can quite literally mean millions of dollars in savings per year on operational costs. Artificial neural networks have come to the front as an at least partially successful method for fraud detection. The success of neural networks in this field is, however, limited by their underlying design - a feedforward neural network is simply a static mapping of input vectors to output vectors, and as such is incapable of adapting to changing shopping profiles of legitimate card holders. Thus, fraud detection systems in use today are plagued by misclassifications and their usefulness is hampered by high false positive rates. We address this problem by proposing the use of a dynamic machine learning method in an attempt to model the time series inherent in sequences of same card transactions. We believe that, instead of looking at individual transactions, it makes more sense to look at sequences of transactions as a whole; a technique that can model time in this context will be more robust to minor shifts in legitimate shopping behaviour. In order to form a clear basis for comparison, we did some investigative research on feature selection, preprocessing, and on the selection of performance measures; the latter will facilitate comparison of results obtained by applying machine learning methods to the biased data sets largely associated with fraud detection. We ran experiments on real world credit card transactional data using two innovative machine learning techniques: the support vector machine (SVM) and the long short-term memory recurrent neural network (LSTM).

M. Bianchini et al. (Eds.): Innovations in Neural Infor. Paradigms & Appli., SCI 247, pp. 231–268.
springerlink.com

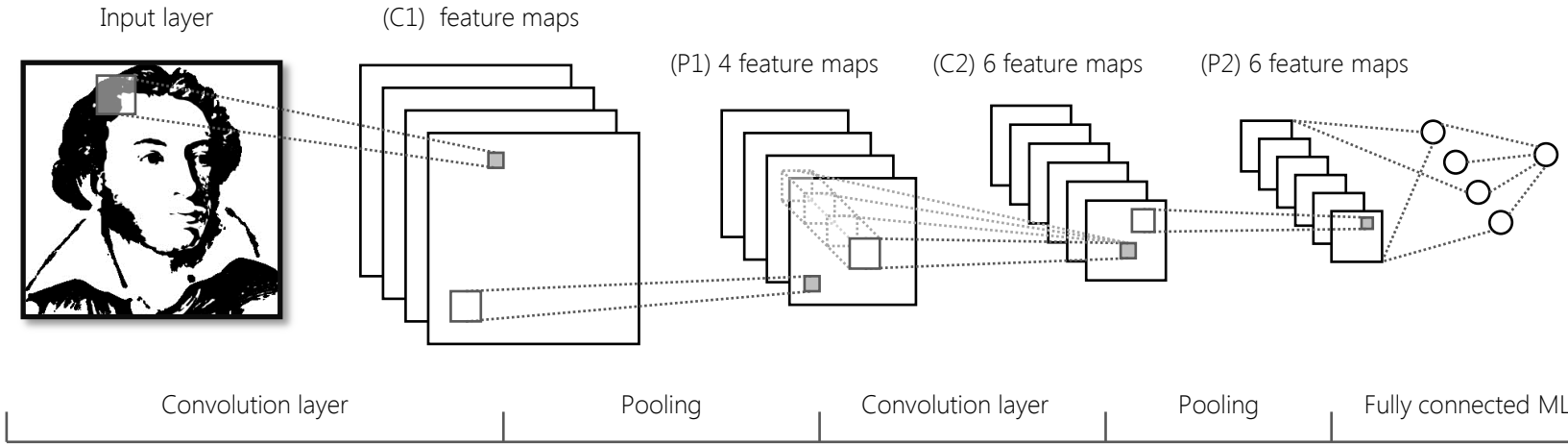
Benard Wiese
Christian Omlin

2009

- Первые опубликовали исследование о применении LSTM-сетей в антифроде
- Даны основы Feature Engineering для антифрод-моделей
- Описаны полезные алгоритмы преобразования фрод-переменных

4. Сверточные нейронные сети

CNN — свёрточные нейронные сети



Credit Card Fraud Detection Using Convolutional Neural Networks

Kang Fu, Dawei Cheng, Yi Tu, and Liqing Zhang(✉)

Key Laboratory of Shanghai Education Commission for Intelligent Interaction and Cognitive Engineering, Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai, China
{fukang1993,dawei.cheng,tuyi1991,lqzhang}@sjtu.edu.cn

Abstract. Credit card is becoming more and more popular in financial transactions, at the same time frauds are also increasing. Conventional methods use rule-based expert systems to detect fraud behaviors, neglecting diverse situations, extreme imbalance of positive and negative samples. In this paper, we propose a CNN-based fraud detection framework, to capture the intrinsic patterns of fraud behaviors learned from labeled data. Abundant transaction data is represented by a feature matrix, on which a convolutional neural network is applied to identify a set of latent patterns for each sample. Experiments on real-world massive transactions of a major commercial bank demonstrate its superior performance compared with some state-of-the-art methods.

Keywords: Credit card fraud · Convolutional neural network · Imbalanced data

1 Introduction

With the rapid development of economy globalization in recent decades, credit cards are much more popular in commercial transactions. The corresponding problem of the credit card fraud emerges accordingly. Machine learning approaches have been proposed to overcome these challenges. Kokkinaki [4] proposed the decision tree and boolean logic functions to characterize the normal transaction modes so as to detect fraudulent transactions. However, some of the fraudulent transactions similar to the legitimate trading patterns can not be identified. So neural networks and Bayesian networks have been employed. Ghosh [2] used a neural network to detect credit card frauds. Bayesian belief networks and artificial neural networks have been also introduced to tackle the problem [6]. These models have been criticized for being overly complex to detect frauds and there has been a high probability of being over-fitting. In order to reveal the latent patterns of fraudulent transactions and avoid the model over-fitting, we use a convolutional neural network to reduce the feature redundancy effectively.

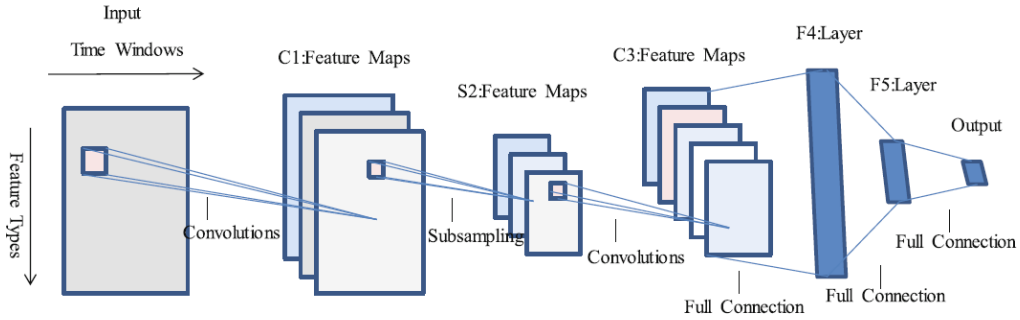
How to generate features of credit card transactions successfully is one of the major challenges to machine learning approaches. Some aggregation strategies

A. Hirose et al. (Eds.): ICONIP 2016, Part III, LNCS 9949, pp. 483–490, 2016.
DOI: 10.1007/978-3-319-46675-0_53

Kang Fu
Dawei Cheng & ko

2016

Convolutional Neural Networks



Credit Card Fraud Detection Using Convolutional Neural Networks

Kang Fu, Dawei Cheng, Yi Tu, and Liqing Zhang(✉)

Key Laboratory of Shanghai Education Commission for Intelligent Interaction and Cognitive Engineering, Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai, China
{fukang1993,dawei.cheng,tuyi1991,lqzhang}@sjtu.edu.cn

Abstract. Credit card is becoming more and more popular in financial transactions, at the same time frauds are also increasing. Conventional methods use rule-based expert systems to detect fraud behaviors, neglecting diverse situations, extreme imbalance of positive and negative samples. In this paper, we propose a CNN-based fraud detection framework, to capture the intrinsic patterns of fraud behaviors learned from labeled data. Abundant transaction data is represented by a feature matrix, on which a convolutional neural network is applied to identify a set of latent patterns for each sample. Experiments on real-world massive transactions of a major commercial bank demonstrate its superior performance compared with some state-of-the-art methods.

Keywords: Credit card fraud · Convolutional neural network · Imbalanced data

1 Introduction

With the rapid development of economy globalization in recent decades, credit cards are much more popular in commercial transactions. The corresponding problem of the credit card fraud emerges accordingly. Machine learning approaches have been proposed to overcome these challenges. Kokkinaki [4] proposed the decision tree and boolean logic functions to characterize the normal transaction modes so as to detect fraudulent transactions. However, some of the fraudulent transactions similar to the legitimate trading patterns can not be identified. So neural networks and Bayesian networks have been employed. Ghosh [2] used a neural network to detect credit card frauds. Bayesian belief networks and artificial neural networks have been also introduced to tackle the problem [6]. These models have been criticized for being overly complex to detect frauds and there has been a high probability of being over-fitting. In order to reveal the latent patterns of fraudulent transactions and avoid the model over-fitting, we use a convolutional neural network to reduce the feature redundancy effectively.

How to generate features of credit card transactions successfully is one of the major challenges to machine learning approaches. Some aggregation strategies

A. Hirose et al. (Eds.): ICONIP 2016, Part III, LNCS 9949, pp. 483–490, 2016.
DOI: 10.1007/978-3-319-46675-0_53

Kang Fu
Dawei Cheng & ko

2016

Magic Feature: Trading Entropy

$$-\sum_{i}^{K} p_i \log p_i + \sum_{j}^{K^{[+1]}} p_j^{[+1]} \log p_j^{[+1]}$$

$$p_i = \frac{AmountT_i}{TotalAmountT} \qquad p_j^{[+1]} = \frac{AmountT_j^{[+1]}}{TotalAmountT + NewTransaction} \qquad (K+1) \geq K^{[+1]} \geq K$$

Credit Card Fraud Detection Using Convolutional Neural Networks

Kang Fu, Dawei Cheng, Yi Tu, and Liqing Zhang(✉)

Key Laboratory of Shanghai Education Commission for Intelligent Interaction and Cognitive Engineering, Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai, China
{fukang1993,dawei.cheng,tuyi1991,lqzhang}@sjtu.edu.cn

Abstract. Credit card is becoming more and more popular in financial transactions, at the same time frauds are also increasing. Conventional methods use rule-based expert systems to detect fraud behaviors, neglecting diverse situations, extreme imbalance of positive and negative samples. In this paper, we propose a CNN-based fraud detection framework, to capture the intrinsic patterns of fraud behaviors learned from labeled data. Abundant transaction data is represented by a feature matrix, on which a convolutional neural network is applied to identify a set of latent patterns for each sample. Experiments on real-world massive transactions of a major commercial bank demonstrate its superior performance compared with some state-of-the-art methods.

Keywords: Credit card fraud · Convolutional neural network · Imbalanced data

1 Introduction

With the rapid development of economy globalization in recent decades, credit cards are much more popular in commercial transactions. The corresponding problem of the credit card fraud emerges accordingly. Machine learning approaches have been proposed to overcome these challenges. Kokkinaki [4] proposed the decision tree and boolean logic functions to characterize the normal transaction modes so as to detect fraudulent transactions. However, some of the fraudulent transactions similar to the legitimate trading patterns can not be identified. So neural networks and Bayesian networks have been employed. Ghosh [2] used a neural network to detect credit card frauds. Bayesian belief networks and artificial neural networks have been also introduced to tackle the problem [6]. These models have been criticized for being overly complex to detect frauds and there has been a high probability of being over-fitting. In order to reveal the latent patterns of fraudulent transactions and avoid the model overfitting, we use a convolutional neural network to reduce the feature redundancy effectively.

How to generate features of credit card transactions successfully is one of the major challenges to machine learning approaches. Some aggregation strategies

A. Hirose et al. (Eds.): ICONIP 2016, Part III, LNCS 9949, pp. 483–490, 2016.
DOI: 10.1007/978-3-319-46675-0_53

Kang Fu
Dawei Cheng & ko

2016

Features	1 day	2 days	7 days	30 days	...
Avg_Amount_T					
Total_Amount_T					
Bias_Amount_T					
Number_T					
Most_Country_T					
Most_Terminal_T					
Most_Merchant_T					
Trading_Entropy_T					

Credit Card Fraud Detection Using Convolutional Neural Networks

Kang Fu, Dawei Cheng, Yi Tu, and Liqing Zhang(✉)

Key Laboratory of Shanghai Education Commission for Intelligent Interaction and Cognitive Engineering, Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai, China
{fukang1993,dawei.cheng,tuyi1991,lqzhang}@sjtu.edu.cn

Abstract. Credit card is becoming more and more popular in financial transactions, at the same time frauds are also increasing. Conventional methods use rule-based expert systems to detect fraud behaviors, neglecting diverse situations, extreme imbalance of positive and negative samples. In this paper, we propose a CNN-based fraud detection framework, to capture the intrinsic patterns of fraud behaviors learned from labeled data. Abundant transaction data is represented by a feature matrix, on which a convolutional neural network is applied to identify a set of latent patterns for each sample. Experiments on real-world massive transactions of a major commercial bank demonstrate its superior performance compared with some state-of-the-art methods.

Keywords: Credit card fraud · Convolutional neural network · Imbalanced data

1 Introduction

With the rapid development of economy globalization in recent decades, credit cards are much more popular in commercial transactions. The corresponding problem of the credit card fraud emerges accordingly. Machine learning approaches have been proposed to overcome these challenges. Kokkinaki [4] proposed the decision tree and boolean logic functions to characterize the normal transaction modes so as to detect fraudulent transactions. However, some of the fraudulent transactions similar to the legitimate trading patterns can not be identified. So neural networks and Bayesian networks have been employed. Ghosh [2] used a neural network to detect credit card frauds. Bayesian belief networks and artificial neural networks have been also introduced to tackle the problem [6]. These models have been criticized for being overly complex to detect frauds and there has been a high probability of being over-fitting. In order to reveal the latent patterns of fraudulent transactions and avoid the model over-fitting, we use a convolutional neural network to reduce the feature redundancy effectively.

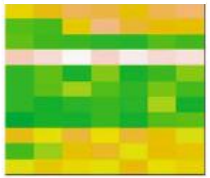
How to generate features of credit card transactions successfully is one of the major challenges to machine learning approaches. Some aggregation strategies

A. Hirose et al. (Eds.): ICONIP 2016, Part III, LNCS 9949, pp. 483–490, 2016.
DOI: 10.1007/978-3-319-46675-0_53

Kang Fu
Dawei Cheng & ko

2016

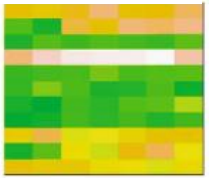
Fraud



Fraud

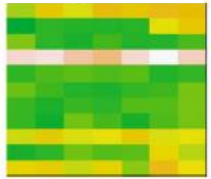
Fraud

Fraud



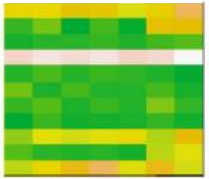
Legitimate

Legitimate



Legitimate

Legitimate



Credit Card Fraud Detection Using Convolutional Neural Networks

Kang Fu, Dawei Cheng, Yi Tu, and Liqing Zhang(✉)

Key Laboratory of Shanghai Education Commission for Intelligent Interaction and Cognitive Engineering, Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai, China
{fukang1993,dawei.cheng,tuyi1991,lqzhang}@sjtu.edu.cn

Abstract. Credit card is becoming more and more popular in financial transactions, at the same time frauds are also increasing. Conventional methods use rule-based expert systems to detect fraud behaviors, neglecting diverse situations, extreme imbalance of positive and negative samples. In this paper, we propose a CNN-based fraud detection framework, to capture the intrinsic patterns of fraud behaviors learned from labeled data. Abundant transaction data is represented by a feature matrix, on which a convolutional neural network is applied to identify a set of latent patterns for each sample. Experiments on real-world massive transactions of a major commercial bank demonstrate its superior performance compared with some state-of-the-art methods.

Keywords: Credit card fraud · Convolutional neural network · Imbalanced data

1 Introduction

With the rapid development of economy globalization in recent decades, credit cards are much more popular in commercial transactions. The corresponding problem of the credit card fraud emerges accordingly. Machine learning approaches have been proposed to overcome these challenges. Kokkinaki [4] proposed the decision tree and boolean logic functions to characterize the normal transaction modes so as to detect fraudulent transactions. However, some of the fraudulent transactions similar to the legitimate trading patterns can not be identified. So neural networks and Bayesian networks have been employed. Ghosh [2] used a neural network to detect credit card frauds. Bayesian belief networks and artificial neural networks have been also introduced to tackle the problem [6]. These models have been criticized for being overly complex to detect frauds and there has been a high probability of being over-fitting. In order to reveal the latent patterns of fraudulent transactions and avoid the model over-fitting, we use a convolutional neural network to reduce the feature redundancy effectively.

How to generate features of credit card transactions successfully is one of the major challenges to machine learning approaches. Some aggregation strategies

A. Hirose et al. (Eds.): ICONIP 2016, Part III, LNCS 9949, pp. 483–490, 2016.
DOI: 10.1007/978-3-319-46675-0_53

Kang Fu
Dawei Cheng & ko

2016

Выборка	260 млн (4 тыс. fraud)
Метрики	F1 score
Алгоритмы	CNN vs. SVM, NN, RF

Credit Card Fraud Detection Using Convolutional Neural Networks

Kang Fu, Dawei Cheng, Yi Tu, and Liqing Zhang(✉)

Key Laboratory of Shanghai Education Commission for Intelligent Interaction and Cognitive Engineering, Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai, China
{fukang1993,dawei.cheng,tuyi1991,lqzhang}@sjtu.edu.cn

Abstract. Credit card is becoming more and more popular in financial transactions, at the same time frauds are also increasing. Conventional methods use rule-based expert systems to detect fraud behaviors, neglecting diverse situations, extreme imbalance of positive and negative samples. In this paper, we propose a CNN-based fraud detection framework, to capture the intrinsic patterns of fraud behaviors learned from labeled data. Abundant transaction data is represented by a feature matrix, on which a convolutional neural network is applied to identify a set of latent patterns for each sample. Experiments on real-world massive transactions of a major commercial bank demonstrate its superior performance compared with some state-of-the-art methods.

Keywords: Credit card fraud · Convolutional neural network · Imbalanced data

1 Introduction

With the rapid development of economy globalization in recent decades, credit cards are much more popular in commercial transactions. The corresponding problem of the credit card fraud emerges accordingly. Machine learning approaches have been proposed to overcome these challenges. Kokkinaki [4] proposed the decision tree and boolean logic functions to characterize the normal transaction modes so as to detect fraudulent transactions. However, some of the fraudulent transactions similar to the legitimate trading patterns can not be identified. So neural networks and Bayesian networks have been employed. Ghosh [2] used a neural network to detect credit card frauds. Bayesian belief networks and artificial neural networks have been also introduced to tackle the problem [6]. These models have been criticized for being overly complex to detect frauds and there has been a high probability of being over-fitting. In order to reveal the latent patterns of fraudulent transactions and avoid the model over-fitting, we use a convolutional neural network to reduce the feature redundancy effectively.

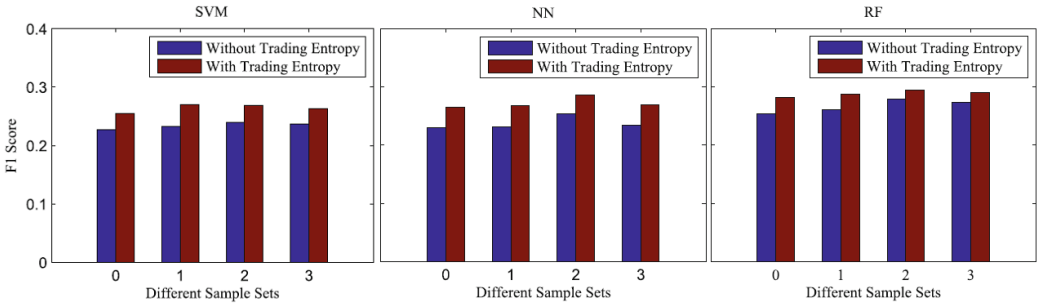
How to generate features of credit card transactions successfully is one of the major challenges to machine learning approaches. Some aggregation strategies

A. Hirose et al. (Eds.): ICONIP 2016, Part III, LNCS 9949, pp. 483–490, 2016.
DOI: 10.1007/978-3-319-46675-0_53

Kang Fu
Dawei Cheng & ko

2016

Trading Entropy – сильная фича



Credit Card Fraud Detection Using Convolutional Neural Networks

Kang Fu, Dawei Cheng, Yi Tu, and Liqing Zhang(✉)

Key Laboratory of Shanghai Education Commission for Intelligent Interaction and Cognitive Engineering, Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai, China
{fukang1993,dawei.cheng,tuyi1991,lqzhang}@sjtu.edu.cn

Abstract. Credit card is becoming more and more popular in financial transactions, at the same time frauds are also increasing. Conventional methods use rule-based expert systems to detect fraud behaviors, neglecting diverse situations, extreme imbalance of positive and negative samples. In this paper, we propose a CNN-based fraud detection framework, to capture the intrinsic patterns of fraud behaviors learned from labeled data. Abundant transaction data is represented by a feature matrix, on which a convolutional neural network is applied to identify a set of latent patterns for each sample. Experiments on real-world massive transactions of a major commercial bank demonstrate its superior performance compared with some state-of-the-art methods.

Keywords: Credit card fraud · Convolutional neural network · Imbalanced data

1 Introduction

With the rapid development of economy globalization in recent decades, credit cards are much more popular in commercial transactions. The corresponding problem of the credit card fraud emerges accordingly. Machine learning approaches have been proposed to overcome these challenges. Kokkinaki [4] proposed the decision tree and boolean logic functions to characterize the normal transaction modes so as to detect fraudulent transactions. However, some of the fraudulent transactions similar to the legitimate trading patterns can not be identified. So neural networks and Bayesian networks have been employed. Ghosh [2] used a neural network to detect credit card frauds. Bayesian belief networks and artificial neural networks have been also introduced to tackle the problem [6]. These models have been criticized for being overly complex to detect frauds and there has been a high probability of being over-fitting. In order to reveal the latent patterns of fraudulent transactions and avoid the model over-fitting, we use a convolutional neural network to reduce the feature redundancy effectively.

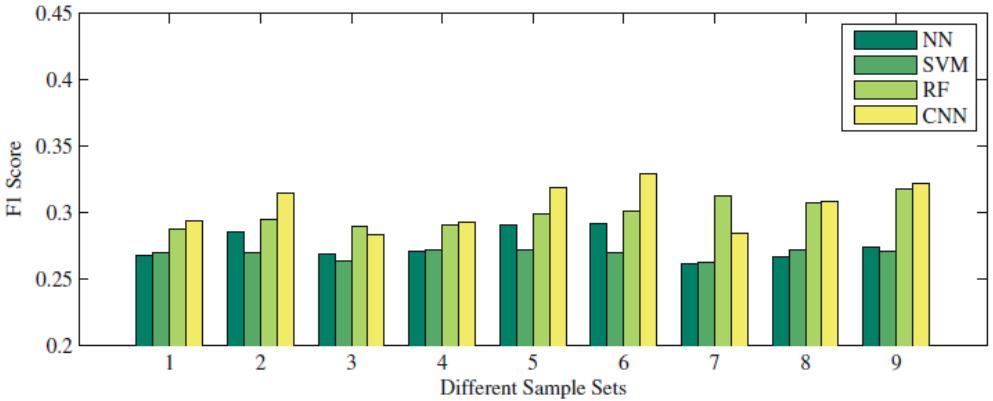
How to generate features of credit card transactions successfully is one of the major challenges to machine learning approaches. Some aggregation strategies

A. Hirose et al. (Eds.): ICONIP 2016, Part III, LNCS 9949, pp. 483–490, 2016.
DOI: 10.1007/978-3-319-46675-0_53

Kang Fu
Dawei Cheng & ko

2016

CNN лучше классических методов



Credit Card Fraud Detection Using Convolutional Neural Networks

Kang Fu, Dawei Cheng, Yi Tu, and Liqing Zhang(✉)

Key Laboratory of Shanghai Education Commission for Intelligent Interaction and Cognitive Engineering, Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai, China
{fukang1993,dawei.cheng,tuyi1991,lqzhang}@sjtu.edu.cn

Abstract. Credit card is becoming more and more popular in financial transactions, at the same time frauds are also increasing. Conventional methods use rule-based expert systems to detect fraud behaviors, neglecting diverse situations, extreme imbalance of positive and negative samples. In this paper, we propose a CNN-based fraud detection framework, to capture the intrinsic patterns of fraud behaviors learned from labeled data. Abundant transaction data is represented by a feature matrix, on which a convolutional neural network is applied to identify a set of latent patterns for each sample. Experiments on real-world massive transactions of a major commercial bank demonstrate its superior performance compared with some state-of-the-art methods.

Keywords: Credit card fraud · Convolutional neural network · Imbalanced data

1 Introduction

With the rapid development of economy globalization in recent decades, credit cards are much more popular in commercial transactions. The corresponding problem of the credit card fraud emerges accordingly. Machine learning approaches have been proposed to overcome these challenges. Kokkinaki [4] proposed the decision tree and boolean logic functions to characterize the normal transaction modes so as to detect fraudulent transactions. However, some of the fraudulent transactions similar to the legitimate trading patterns can not be identified. So neural networks and Bayesian networks have been employed. Ghosh [2] used a neural network to detect credit card frauds. Bayesian belief networks and artificial neural networks have been also introduced to tackle the problem [6]. These models have been criticized for being overly complex to detect frauds and there has been a high probability of being over-fitting. In order to reveal the latent patterns of fraudulent transactions and avoid the model overfitting, we use a convolutional neural network to reduce the feature redundancy effectively.

How to generate features of credit card transactions successfully is one of the major challenges to machine learning approaches. Some aggregation strategies

A. Hirose et al. (Eds.): ICONIP 2016, Part III, LNCS 9949, pp. 483–490, 2016.
DOI: 10.1007/978-3-319-46675-0_53

Kang Fu
Dawei Cheng & ko

2016

- Используется LeNet5 – **устаревшая архитектура** CNN (без residual)
- Не учтена **проблема инвариантности** при перестановке переменных
- Не решена **проблема разных длин** у транзакционных временных рядов

Credit Card Fraud Detection Using Convolutional Neural Networks

Kang Fu, Dawei Cheng, Yi Tu, and Liqing Zhang(✉)

Key Laboratory of Shanghai Education Commission for Intelligent Interaction and Cognitive Engineering, Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai, China
{fukang1993,dawei.cheng,tuyi1991,lqzhang}@sjtu.edu.cn

Abstract. Credit card is becoming more and more popular in financial transactions, at the same time frauds are also increasing. Conventional methods use rule-based expert systems to detect fraud behaviors, neglecting diverse situations, extreme imbalance of positive and negative samples. In this paper, we propose a CNN-based fraud detection framework, to capture the intrinsic patterns of fraud behaviors learned from labeled data. Abundant transaction data is represented by a feature matrix, on which a convolutional neural network is applied to identify a set of latent patterns for each sample. Experiments on real-world massive transactions of a major commercial bank demonstrate its superior performance compared with some state-of-the-art methods.

Keywords: Credit card fraud · Convolutional neural network · Imbalanced data

1 Introduction

With the rapid development of economy globalization in recent decades, credit cards are much more popular in commercial transactions. The corresponding problem of the credit card fraud emerges accordingly. Machine learning approaches have been proposed to overcome these challenges. Kokkinaki [4] proposed the decision tree and boolean logic functions to characterize the normal transaction modes so as to detect fraudulent transactions. However, some of the fraudulent transactions similar to the legitimate trading patterns can not be identified. So neural networks and Bayesian networks have been employed. Ghosh [2] used a neural network to detect credit card frauds. Bayesian belief networks and artificial neural networks have been also introduced to tackle the problem [6]. These models have been criticized for being overly complex to detect frauds and there has been a high probability of being over-fitting. In order to reveal the latent patterns of fraudulent transactions and avoid the model overfitting, we use a convolutional neural network to reduce the feature redundancy effectively.

How to generate features of credit card transactions successfully is one of the major challenges to machine learning approaches. Some aggregation strategies

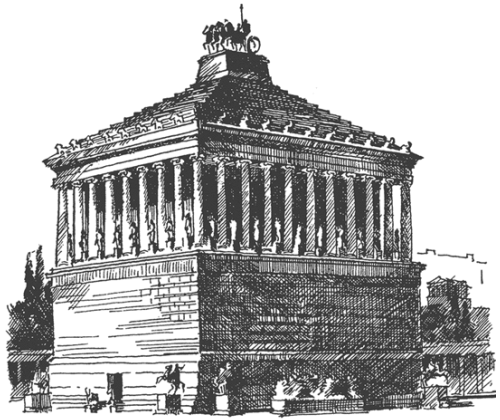
A. Hirose et al. (Eds.): ICONIP 2016, Part III, LNCS 9949, pp. 483–490, 2016.
DOI: 10.1007/978-3-319-46675-0_53

Kang Fu
Dawei Cheng & ko

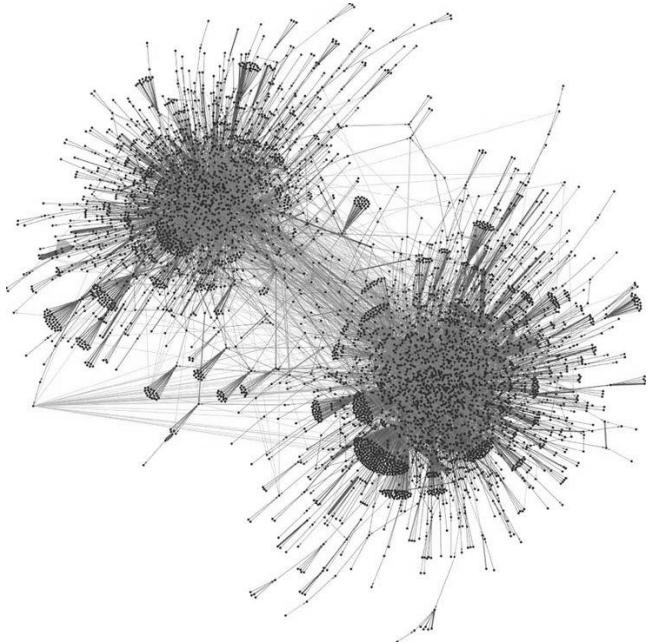
2016

- Первые опубликовали исследование о применении CNN в антифроде
- Предложена сильная переменная для транзакционного фрода – Trading Entropy
- Проведена всесторонняя проверка на обобщаемость полученных результатов

5. Deep Learning для графов



Графы популярны в крупных банках

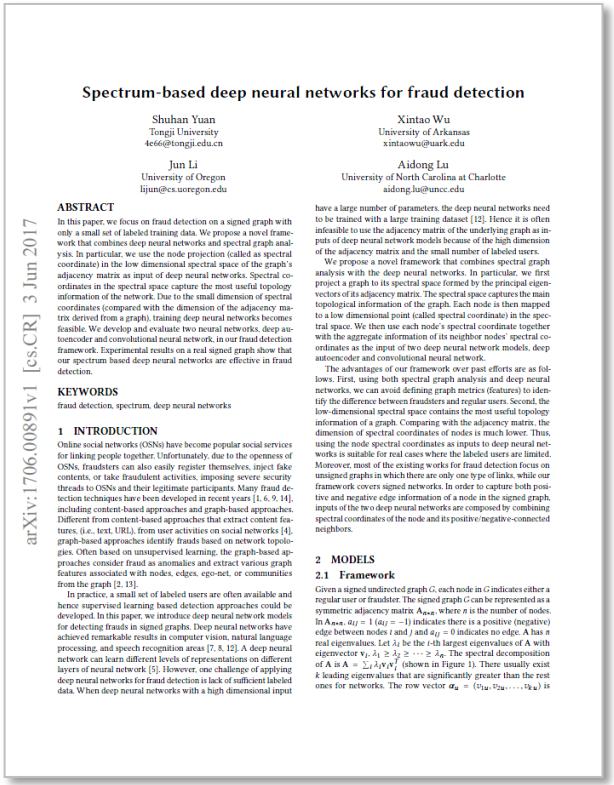


A1. Выберите правильный вариант

- ○ Граф
- ○ Греф
- ○ Оба варианта

Shuhan Yuan
Xintao Wu & ko

2017



arXiv:1706.00891v1 [cs.CR] 3 Jun 2017

Spectrum-based deep neural networks for fraud detection

Shuhan Yuan
Tongji University
4e66@tongji.edu.cn

Xintao Wu
University of Arkansas
xintaowu@uark.edu

Jun Li
University of Oregon
lijun@cs.uoregon.edu

Aidong Lu
University of North Carolina at Charlotte
aidong.lu@uncc.edu

ABSTRACT

In this paper, we focus on fraud detection on a signed graph with only a small set of labeled training data. We propose a novel framework that combines deep neural networks and spectral graph analysis. In particular, we use the node projection (called as spectral coordinate) in the low dimensional spectral space of the graph's adjacency matrix as input of deep neural networks. Spectral coordinates in the spectral space capture the most useful topology information of the network. Due to the small dimension of spectral coordinates (compared with the dimension of the adjacency matrix derived from a graph), training deep neural networks becomes feasible. We develop and evaluate two neural networks, deep autoencoder and convolutional neural network, in our fraud detection framework. Experimental results on a real signed graph show that our spectrum based deep neural networks are effective in fraud detection.

KEYWORDS

fraud detection, spectrum, deep neural networks

1 INTRODUCTION

Online social networks (OSNs) have become popular social services for linking people together. Unfortunately, due to the openness of OSNs, fraudsters can also easily register themselves, inject fake contents, or take fraudulent activities, imposing severe security threads to OSNs and their legitimate participants. Many fraud detection techniques have been developed in recent years [1, 6, 9, 14], including content-based approaches and graph-based approaches. Different from content-based approaches that extract content features, (i.e., text, URL), from user activities on social networks [4], graph-based approaches identify frauds based on network topologies. Often based on unsupervised learning, the graph-based approaches consider fraud as anomalies and extract various graph features associated with nodes, edges, ego-net, or communities from the graph [2, 13].

In practice, a small set of labeled users are often available and hence supervised learning based detection approaches could be developed. In this paper, we introduce deep neural network models for detecting frauds in signed graphs. Deep neural networks have achieved remarkable results in computer vision, natural language processing, and speech recognition areas [7, 8, 12]. A deep neural network can learn different levels of representations on different layers of neural network [5]. However, one challenge of applying deep neural networks for fraud detection is lack of sufficient labeled data. When deep neural networks with a high dimensional input have a large number of parameters, the deep neural networks need to be trained with a large training dataset [12]. Hence it is often infeasible to use the adjacency matrix of the underlying graph as inputs of deep neural network models because of the high dimension of the adjacency matrix and the small number of labeled users.

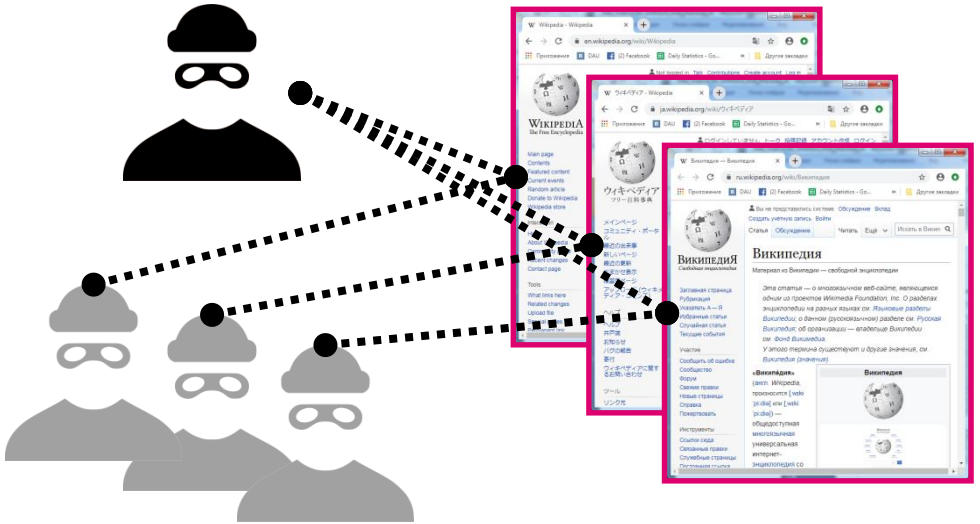
We propose a novel framework that combines spectral graph analysis with the deep neural networks. In particular, we first project a graph to its spectral space formed by the principal eigenvectors of its adjacency matrix. The spectral space captures the main topological information of the graph. Each node is then mapped to a low dimensional point (called spectral coordinate) in the spectral space. We then use each node's spectral coordinate together with the aggregate information of its neighbor nodes' spectral coordinates as the input of two deep neural network models, deep autoencoder and convolutional neural network.

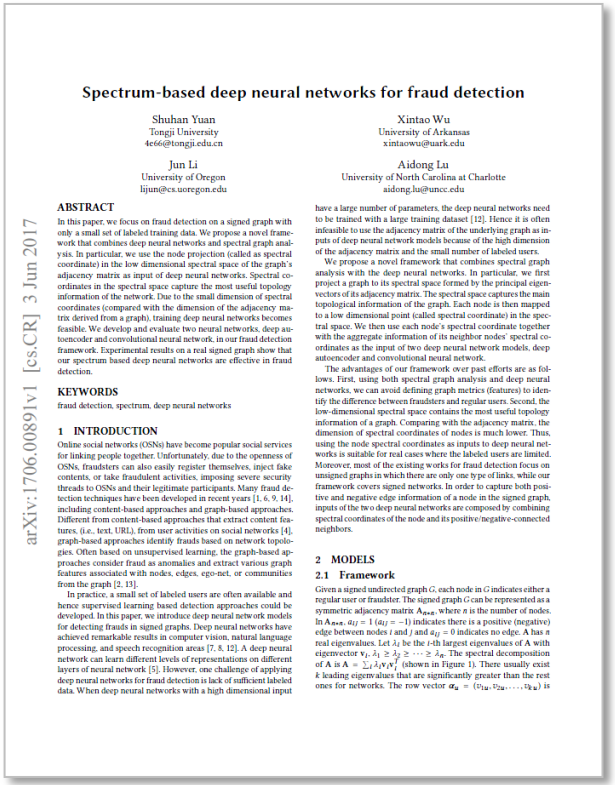
The advantages of our framework over past efforts are as follows. First, using both spectral graph analysis and deep neural networks, we can avoid defining graph metrics (features) to identify the difference between fraudsters and regular users. Second, the low-dimensional spectral space contains the most useful topology information of a graph. Comparing with the adjacency matrix, the dimension of spectral coordinates of nodes is much lower. Thus, using the node spectral coordinates as inputs to deep neural networks is suitable for real cases where the labeled users are limited. Moreover, most of the existing works for fraud detection focus on unsigned graphs in which there are only one type of links, while our framework covers signed networks. In order to capture both positive and negative edge information of a node in the signed graph, inputs of the two deep neural networks are composed by combining spectral coordinates of the node and its positive/negative-connected neighbors.

2 MODELS

2.1 Framework

Given a signed undirected graph G, each node in G indicates either a regular user or fraudster. The signed graph G can be represented as a symmetric adjacency matrix $\mathbf{A}_{n\times n}$, where n is the number of nodes. In $\mathbf{A}_{n\times n}$, $a_{ij} = 1$ ($a_{ij} = -1$) indicates there is a positive (negative) edge between nodes i and j and $a_{ij} = 0$ indicates no edge. A has n real eigenvalues. Let λ_i be the i-th largest eigenvalues of $\mathbf{A}$ with eigenvector $\mathbf{v}_i$, $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$. The spectral decomposition of $\mathbf{A}$ is $\mathbf{A} = \sum_i \lambda_i \mathbf{v}_i \mathbf{v}_i^T$ (shown in Figure 1). There usually exist k leading eigenvalues that are significantly greater than the rest ones for networks. The row vector $\boldsymbol{\alpha}_u = (v_{1u}, v_{2u}, \ldots, v_{ku})$ is





Spectrum-based deep neural networks for fraud detection

Shuhan Yuan
Tongji University
4e66@tongji.edu.cn

Xintao Wu
University of Arkansas
xintaowu@uark.edu

Jun Li
University of Oregon
lijun@cs.uoregon.edu

Aidong Lu
University of North Carolina at Charlotte
aidong.lu@uncc.edu

ABSTRACT

In this paper, we focus on fraud detection on a signed graph with only a small set of labeled training data. We propose a novel framework that combines deep neural networks and spectral graph analysis. In particular, we use the node projection (called as spectral coordinate) in the low dimensional spectral space of the graph's adjacency matrix as input of deep neural networks. Spectral coordinates in the spectral space capture the most useful topology information of the network. Due to the small dimension of spectral coordinates (compared with the dimension of the adjacency matrix derived from a graph), training deep neural networks becomes feasible. We develop and evaluate two neural networks, deep autoencoder and convolutional neural network, in our fraud detection framework. Experimental results on a real signed graph show that our spectrum based deep neural networks are effective in fraud detection.

KEYWORDS

fraud detection, spectrum, deep neural networks

1 INTRODUCTION

Online social networks (OSNs) have become popular social services for linking people together. Unfortunately, due to the openness of OSNs, fraudsters can also easily register themselves, inject fake contents, or take fraudulent activities, imposing severe security threads to OSNs and their legitimate participants. Many fraud detection techniques have been developed in recent years [1, 6, 9, 14], including content-based approaches and graph-based approaches. Different from content-based approaches that extract content features, (i.e., text, URL), from user activities on social networks [4], graph-based approaches identify frauds based on network topologies. Often based on unsupervised learning, the graph-based approaches consider fraud as anomalies and extract various graph features associated with nodes, edges, ego-net, or communities from the graph [2, 13].

In practice, a small set of labeled users are often available and hence supervised learning based detection approaches could be developed. In this paper, we introduce deep neural network models for detecting frauds in signed graphs. Deep neural networks have achieved remarkable results in computer vision, natural language processing, and speech recognition areas [7, 8, 12]. A deep neural network can learn different levels of representations on different layers of neural network [5]. However, one challenge of applying deep neural networks for fraud detection is lack of sufficient labeled data. When deep neural networks with a high dimensional input have a large number of parameters, the deep neural networks need to be trained with a large training dataset [12]. Hence it is often infeasible to use the adjacency matrix of the underlying graph as inputs of deep neural network models because of the high dimension of the adjacency matrix and the small number of labeled users.

We propose a novel framework that combines spectral graph analysis with the deep neural networks. In particular, we first project a graph to its spectral space formed by the principal eigenvectors of its adjacency matrix. The spectral space captures the main topological information of the graph. Each node is then mapped to a low dimensional point (called spectral coordinate) in the spectral space. We then use each node's spectral coordinate together with the aggregate information of its neighbor nodes' spectral coordinates as the input of two deep neural network models, deep autoencoder and convolutional neural network.

The advantages of our framework over past efforts are as follows. First, using both spectral graph analysis and deep neural networks, we can avoid defining graph metrics (features) to identify the difference between fraudsters and regular users. Second, the low-dimensional spectral space contains the most useful topology information of a graph. Comparing with the adjacency matrix, the dimension of spectral coordinates of nodes is much lower. Thus, using the node spectral coordinates as inputs to deep neural networks is suitable for real cases where the labeled users are limited. Moreover, most of the existing works for fraud detection focus on unsigned graphs in which there are only one type of links, while our framework covers signed networks. In order to capture both positive and negative edge information of a node in the signed graph, inputs of the two deep neural networks are composed by combining spectral coordinates of the node and its positive/negative-connected neighbors.

2 MODELS

2.1 Framework

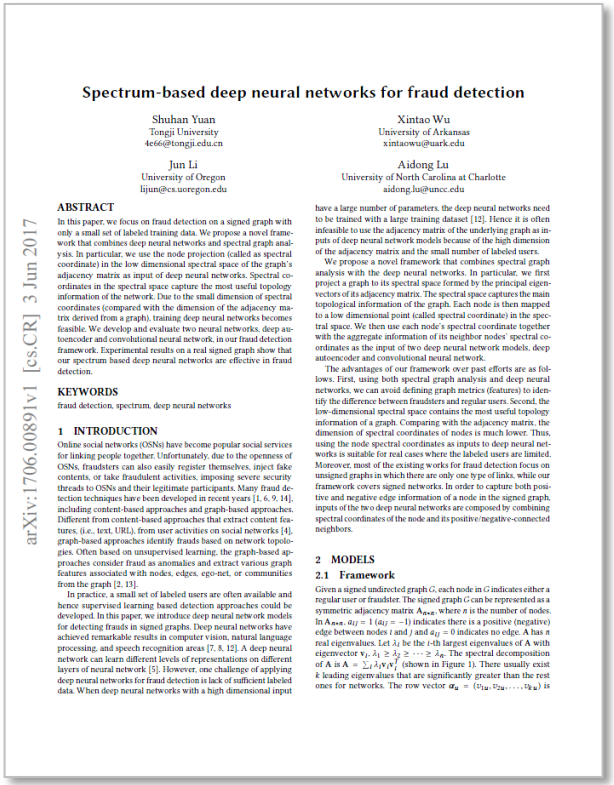
Given a signed undirected graph G, each node in G indicates either a regular user or fraudster. The signed graph G can be represented as a symmetric adjacency matrix $\mathbf{A}_{n\times n}$, where n is the number of nodes. In $\mathbf{A}_{n\times n}$, $a_{ij} = 1$ ($a_{ij} = -1$) indicates there is a positive (negative) edge between nodes i and j and $a_{ij} = 0$ indicates no edge. $\mathbf{A}$ has n real eigenvalues. Let λ_i be the i-th largest eigenvalues of $\mathbf{A}$ with eigenvector $\mathbf{v}_i$. $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$. The spectral decomposition of $\mathbf{A}$ is $\mathbf{A} = \sum_i \lambda_i \mathbf{v}_i \mathbf{v}_i^t$ (shown in Figure 1). There usually exist k leading eigenvalues that are significantly greater than the rest ones for networks. The row vector $\boldsymbol{\alpha}_u = (v_{1u}, v_{2u}, \ldots, v_{ku})$ is

Shuhan Yuan
Xintao Wu & ko

2017

Матрица смежности графа

$$A(G) = \begin{array}{c} \\ v_1 \\ \vdots \\ v_n \end{array} \begin{array}{c} \begin{array}{ccc} v_1 & \cdots & v_n \end{array} \\ \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix} \end{array}$$



Spectrum-based deep neural networks for fraud detection

Shuhan Yuan
Tongji University
4e66@tongji.edu.cn

Xintao Wu
University of Arkansas
xintaowu@uark.edu

Jun Li
University of Oregon
lijun@cs.uoregon.edu

Aidong Lu
University of North Carolina at Charlotte
aidong.lu@uncc.edu

ABSTRACT

In this paper, we focus on fraud detection on a signed graph with only a small set of labeled training data. We propose a novel framework that combines deep neural networks and spectral graph analysis. In particular, we use the node projection (called as spectral coordinate) in the low dimensional spectral space of the graph's adjacency matrix as input of deep neural networks. Spectral coordinates in the spectral space capture the most useful topology information of the network. Due to the small dimension of spectral coordinates (compared with the dimension of the adjacency matrix derived from a graph), training deep neural networks becomes feasible. We develop and evaluate two neural networks, deep autoencoder and convolutional neural network, in our fraud detection framework. Experimental results on a real signed graph show that our spectrum based deep neural networks are effective in fraud detection.

KEYWORDS

fraud detection, spectrum, deep neural networks

1 INTRODUCTION

Online social networks (OSNs) have become popular social services for linking people together. Unfortunately, due to the openness of OSNs, fraudsters can also easily register themselves, inject fake contents, or take fraudulent activities, imposing severe security threads to OSNs and their legitimate participants. Many fraud detection techniques have been developed in recent years [1, 6, 9, 14], including content-based approaches and graph-based approaches. Different from content-based approaches that extract content features, (i.e., text, URL), from user activities on social networks [4], graph-based approaches identify frauds based on network topologies. Often based on unsupervised learning, the graph-based approaches consider fraud as anomalies and extract various graph features associated with nodes, edges, ego-net, or communities from the graph [2, 13].

In practice, a small set of labeled users are often available and hence supervised learning based detection approaches could be developed. In this paper, we introduce deep neural network models for detecting frauds in signed graphs. Deep neural networks have achieved remarkable results in computer vision, natural language processing, and speech recognition areas [7, 8, 12]. A deep neural network can learn different levels of representations on different layers of neural network [5]. However, one challenge of applying deep neural networks for fraud detection is lack of sufficient labeled data. When deep neural networks with a high dimensional input have a large number of parameters, the deep neural networks need to be trained with a large training dataset [12]. Hence it is often infeasible to use the adjacency matrix of the underlying graph as inputs of deep neural network models because of the high dimension of the adjacency matrix and the small number of labeled users.

We propose a novel framework that combines spectral graph analysis with the deep neural networks. In particular, we first project a graph to its spectral space formed by the principal eigenvectors of its adjacency matrix. The spectral space captures the main topological information of the graph. Each node is then mapped to a low dimensional point (called spectral coordinate) in the spectral space. We then use each node's spectral coordinate together with the aggregate information of its neighbor nodes' spectral coordinates as the input of two deep neural network models, deep autoencoder and convolutional neural network.

The advantages of our framework over past efforts are as follows. First, using both spectral graph analysis and deep neural networks, we can avoid defining graph metrics (features) to identify the difference between fraudsters and regular users. Second, the low-dimensional spectral space contains the most useful topology information of a graph. Comparing with the adjacency matrix, the dimension of spectral coordinates of nodes is much lower. Thus, using the node spectral coordinates as inputs to deep neural networks is suitable for real cases where the labeled users are limited. Moreover, most of the existing works for fraud detection focus on unsigned graphs in which there are only one type of links, while our framework covers signed networks. In order to capture both positive and negative edge information of a node in the signed graph, inputs of the two deep neural networks are composed by combining spectral coordinates of the node and its positive/negative-connected neighbors.

2 MODELS

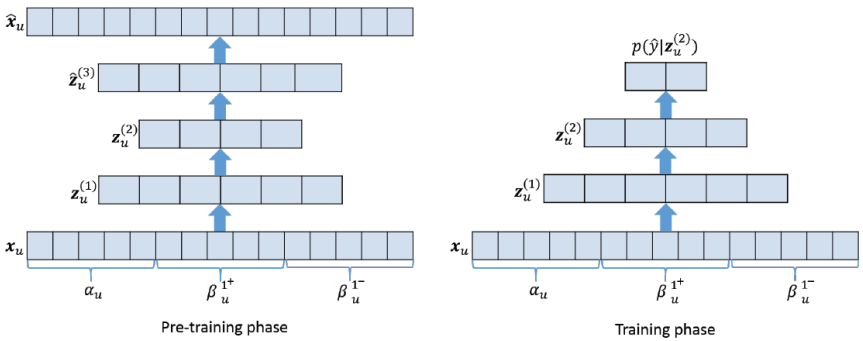
2.1 Framework

Given a signed undirected graph G, each node in G indicates either a regular user or fraudster. The signed graph G can be represented as a symmetric adjacency matrix $\mathbf{A}_{n\times n}$, where n is the number of nodes. In $\mathbf{A}_{n\times n}$, $a_{ij} = 1$ ($a_{ij} = -1$) indicates there is a positive (negative) edge between nodes i and j and $a_{ij} = 0$ indicates no edge. A has n real eigenvalues. Let λ_i be the i-th largest eigenvalues of $\mathbf{A}$ with eigenvector $\mathbf{v}_i$. $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$. The spectral decomposition of $\mathbf{A}$ is $\mathbf{A} = \sum_i \lambda_i \mathbf{v}_i \mathbf{v}_i^t$ (shown in Figure 1). There usually exist k leading eigenvalues that are significantly greater than the rest ones for networks. The row vector $\boldsymbol{\alpha}_u = (v_{1u}, v_{2u}, \ldots, v_{ku})$ is

Shuhan Yuan
Xintao Wu & ko

2017

DAE with spectral coordinates



Spectrum-based deep neural networks for fraud detection

Shuhan Yuan
Tongji University
4e66@tongji.edu.cn

Xintao Wu
University of Arkansas
xintaowu@uark.edu

Jun Li
University of Oregon
lijun@cs.uoregon.edu

Aidong Lu
University of North Carolina at Charlotte
aidong.lu@uncc.edu

ABSTRACT

In this paper, we focus on fraud detection on a signed graph with only a small set of labeled training data. We propose a novel framework that combines deep neural networks and spectral graph analysis. In particular, we use the node projection (called as spectral coordinate) in the low dimensional spectral space of the graph's adjacency matrix as input of deep neural networks. Spectral coordinates in the spectral space capture the most useful topology information of the network. Due to the small dimension of spectral coordinates (compared with the dimension of the adjacency matrix derived from a graph), training deep neural networks becomes feasible. We develop and evaluate two neural networks, deep autoencoder and convolutional neural network, in our fraud detection framework. Experimental results on a real signed graph show that our spectrum based deep neural networks are effective in fraud detection.

KEYWORDS

fraud detection, spectrum, deep neural networks

1 INTRODUCTION

Online social networks (OSNs) have become popular social services for linking people together. Unfortunately, due to the openness of OSNs, fraudsters can also easily register themselves, inject fake contents, or take fraudulent activities, imposing severe security threads to OSNs and their legitimate participants. Many fraud detection techniques have been developed in recent years [1, 6, 9, 14], including content-based approaches and graph-based approaches. Different from content-based approaches that extract content features, (i.e., text, URL), from user activities on social networks [4], graph-based approaches identify frauds based on network topologies. Often based on unsupervised learning, the graph-based approaches consider fraud as anomalies and extract various graph features associated with nodes, edges, ego-net, or communities from the graph [2, 13].

In practice, a small set of labeled users are often available and hence supervised learning based detection approaches could be developed. In this paper, we introduce deep neural network models for detecting frauds in signed graphs. Deep neural networks have achieved remarkable results in computer vision, natural language processing, and speech recognition areas [7, 8, 12]. A deep neural network can learn different levels of representations on different layers of neural network [5]. However, one challenge of applying deep neural networks for fraud detection is lack of sufficient labeled data. When deep neural networks with a high dimensional input have a large number of parameters, the deep neural networks need to be trained with a large training dataset [12]. Hence it is often infeasible to use the adjacency matrix of the underlying graph as inputs of deep neural network models because of the high dimension of the adjacency matrix and the small number of labeled users.

We propose a novel framework that combines spectral graph analysis with the deep neural networks. In particular, we first project a graph to its spectral space formed by the principal eigenvectors of its adjacency matrix. The spectral space captures the main topological information of the graph. Each node is then mapped to a low dimensional point (called spectral coordinate) in the spectral space. We then use each node's spectral coordinate together with the aggregate information of its neighbor nodes' spectral coordinates as the input of two deep neural network models, deep autoencoder and convolutional neural network.

The advantages of our framework over past efforts are as follows. First, using both spectral graph analysis and deep neural networks, we can avoid defining graph metrics (features) to identify the difference between fraudsters and regular users. Second, the low-dimensional spectral space contains the most useful topology information of a graph. Comparing with the adjacency matrix, the dimension of spectral coordinates of nodes is much lower. Thus, using the node spectral coordinates as inputs to deep neural networks is suitable for real cases where the labeled users are limited. Moreover, most of the existing works for fraud detection focus on unsigned graphs in which there are only one type of links, while our framework covers signed networks. In order to capture both positive and negative edge information of a node in the signed graph, inputs of the two deep neural networks are composed by combining spectral coordinates of the node and its positive/negative-connected neighbors.

2 MODELS

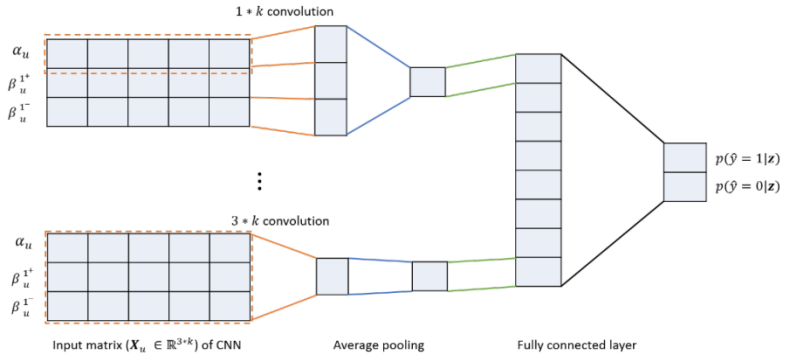
2.1 Framework

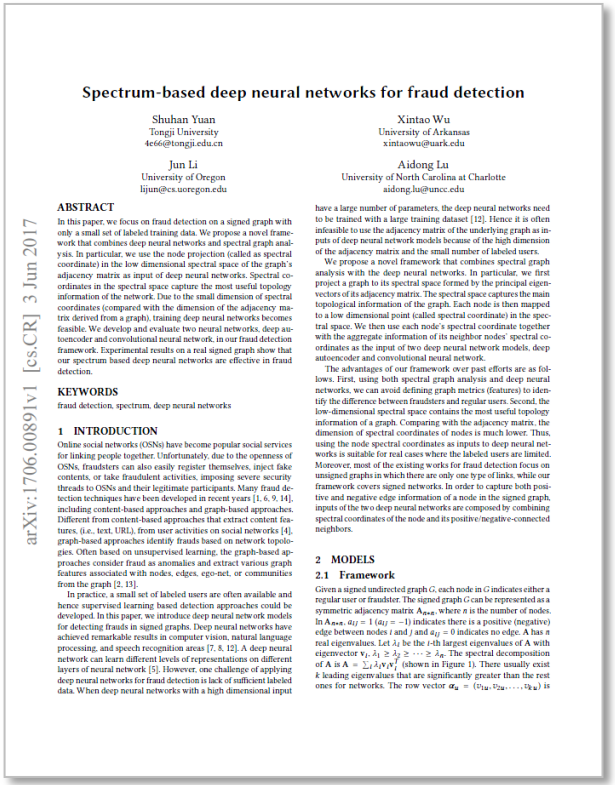
Given a signed undirected graph G, each node in G indicates either a regular user or fraudster. The signed graph G can be represented as a symmetric adjacency matrix $\mathbf{A}_{n\times n}$, where n is the number of nodes. In $\mathbf{A}_{n\times n}$, $a_{ij} = 1$ ($a_{ij} = -1$) indicates there is a positive (negative) edge between nodes i and j and $a_{ij} = 0$ indicates no edge. A has n real eigenvalues. Let λ_l be the l-th largest eigenvalues of $\mathbf{A}$ with eigenvector $\mathbf{v}_l$. $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$. The spectral decomposition of $\mathbf{A}$ is $\mathbf{A} = \sum_l \lambda_l \mathbf{v}_l \mathbf{v}_l^T$ (shown in Figure 1). There usually exist k leading eigenvalues that are significantly greater than the rest ones for networks. The row vector $\boldsymbol{\alpha}_u = (v_{1u}, v_{2u}, \ldots, v_{ku})$ is

Shuhan Yuan
Xintao Wu & ko

2017

CNN with spectral coordinates





Spectrum-based deep neural networks for fraud detection

Shuhan Yuan
Tongji University
4e66@tongji.edu.cn

Xintao Wu
University of Arkansas
xintaowu@uark.edu

Jun Li
University of Oregon
lijun@cs.uoregon.edu

Aidong Lu
University of North Carolina at Charlotte
aidong.lu@uncc.edu

ABSTRACT

In this paper, we focus on fraud detection on a signed graph with only a small set of labeled training data. We propose a novel framework that combines deep neural networks and spectral graph analysis. In particular, we use the node projection (called as spectral coordinate) in the low dimensional spectral space of the graph's adjacency matrix as input of deep neural networks. Spectral coordinates in the spectral space capture the most useful topology information of the network. Due to the small dimension of spectral coordinates (compared with the dimension of the adjacency matrix derived from a graph), training deep neural networks becomes feasible. We develop and evaluate two neural networks, deep autoencoder and convolutional neural network, in our fraud detection framework. Experimental results on a real signed graph show that our spectrum based deep neural networks are effective in fraud detection.

KEYWORDS

fraud detection, spectrum, deep neural networks

1 INTRODUCTION

Online social networks (OSNs) have become popular social services for linking people together. Unfortunately, due to the openness of OSNs, fraudsters can also easily register themselves, inject fake contents, or take fraudulent activities, imposing severe security threads to OSNs and their legitimate participants. Many fraud detection techniques have been developed in recent years [1, 6, 9, 14], including content-based approaches and graph-based approaches. Different from content-based approaches that extract content features, (i.e., text, URL), from user activities on social networks [4], graph-based approaches identify frauds based on network topologies. Often based on unsupervised learning, the graph-based approaches consider fraud as anomalies and extract various graph features associated with nodes, edges, ego-net, or communities from the graph [2, 13].

In practice, a small set of labeled users are often available and hence supervised learning based detection approaches could be developed. In this paper, we introduce deep neural network models for detecting frauds in signed graphs. Deep neural networks have achieved remarkable results in computer vision, natural language processing, and speech recognition areas [7, 8, 12]. A deep neural network can learn different levels of representations on different layers of neural network [5]. However, one challenge of applying deep neural networks for fraud detection is lack of sufficient labeled data. When deep neural networks with a high dimensional input have a large number of parameters, the deep neural networks need to be trained with a large training dataset [12]. Hence it is often infeasible to use the adjacency matrix of the underlying graph as inputs of deep neural network models because of the high dimension of the adjacency matrix and the small number of labeled users.

We propose a novel framework that combines spectral graph analysis with the deep neural networks. In particular, we first project a graph to its spectral space formed by the principal eigenvectors of its adjacency matrix. The spectral space captures the main topological information of the graph. Each node is then mapped to a low dimensional point (called spectral coordinate) in the spectral space. We then use each node's spectral coordinate together with the aggregate information of its neighbor nodes' spectral coordinates as the input of two deep neural network models, deep autoencoder and convolutional neural network.

The advantages of our framework over past efforts are as follows. First, using both spectral graph analysis and deep neural networks, we can avoid defining graph metrics (features) to identify the difference between fraudsters and regular users. Second, the low-dimensional spectral space contains the most useful topology information of a graph. Comparing with the adjacency matrix, the dimension of spectral coordinates of nodes is much lower. Thus, using the node spectral coordinates as inputs to deep neural networks is suitable for real cases where the labeled users are limited. Moreover, most of the existing works for fraud detection focus on unsigned graphs in which there are only one type of links, while our framework covers signed networks. In order to capture both positive and negative edge information of a node in the signed graph, inputs of the two deep neural networks are composed by combining spectral coordinates of the node and its positive/negative-connected neighbors.

2 MODELS

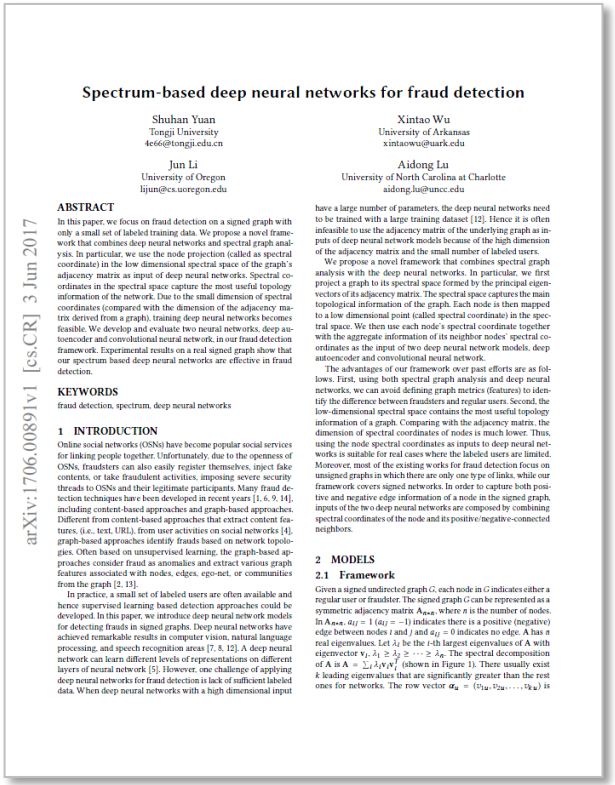
2.1 Framework

Given a signed undirected graph G, each node in G indicates either a regular user or fraudster. The signed graph G can be represented as a symmetric adjacency matrix $\mathbf{A}_{n\times n}$, where n is the number of nodes. In $\mathbf{A}_{n\times n}$, $a_{ij} = 1$ ($a_{ij} = -1$) indicates there is a positive (negative) edge between nodes i and j and $a_{ij} = 0$ indicates no edge. A has n real eigenvalues. Let λ_i be the i-th largest eigenvalues of $\mathbf{A}$ with eigenvector $\mathbf{v}_i$. $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$. The spectral decomposition of $\mathbf{A}$ is $\mathbf{A} = \sum_i \lambda_i \mathbf{v}_i \mathbf{v}_i^T$ (shown in Figure 1). There usually exist k leading eigenvalues that are significantly greater than the rest ones for networks. The row vector $\boldsymbol{\sigma}_u = (v_{1u}, v_{2u}, \ldots, v_{ku})$ is

Shuhan Yuan Xintao Wu & ko

2017

Выборка	17 015 fraud, 17 015 good
Метрики	Accuracy
Алгоритмы	DAE, CNN vs. k-NN, SVM



arXiv:1706.00891v1 [cs.CR] 3 Jun 2017

Spectrum-based deep neural networks for fraud detection

Shuhan Yuan
Tongji University
4e66@tongji.edu.cn

Xintao Wu
University of Arkansas
xintaowu@uark.edu

Jun Li
University of Oregon
lijun@cs.uoregon.edu

Aidong Lu
University of North Carolina at Charlotte
aidong.lu@uncc.edu

ABSTRACT

In this paper, we focus on fraud detection on a signed graph with only a small set of labeled training data. We propose a novel framework that combines deep neural networks and spectral graph analysis. In particular, we use the node projection (called as spectral coordinate) in the low dimensional spectral space of the graph's adjacency matrix as input of deep neural networks. Spectral coordinates in the spectral space capture the most useful topology information of the network. Due to the small dimension of spectral coordinates (compared with the dimension of the adjacency matrix derived from a graph), training deep neural networks becomes feasible. We develop and evaluate two neural networks, deep autoencoder and convolutional neural network, in our fraud detection framework. Experimental results on a real signed graph show that our spectrum based deep neural networks are effective in fraud detection.

KEYWORDS

fraud detection, spectrum, deep neural networks

1 INTRODUCTION

Online social networks (OSNs) have become popular social services for linking people together. Unfortunately, due to the openness of OSNs, fraudsters can also easily register themselves, inject fake contents, or take fraudulent activities, imposing severe security threads to OSNs and their legitimate participants. Many fraud detection techniques have been developed in recent years [1, 6, 9, 14], including content-based approaches and graph-based approaches. Different from content-based approaches that extract content features, (i.e., text, URL), from user activities on social networks [4], graph-based approaches identify frauds based on network topologies. Often based on unsupervised learning, the graph-based approaches consider fraud as anomalies and extract various graph features associated with nodes, edges, ego-net, or communities from the graph [2, 13].

In practice, a small set of labeled users are often available and hence supervised learning based detection approaches could be developed. In this paper, we introduce deep neural network models for detecting frauds in signed graphs. Deep neural networks have achieved remarkable results in computer vision, natural language processing, and speech recognition areas [7, 8, 12]. A deep neural network can learn different levels of representations on different layers of neural network [5]. However, one challenge of applying deep neural networks for fraud detection is lack of sufficient labeled data. When deep neural networks with a high dimensional input have a large number of parameters, the deep neural networks need to be trained with a large training dataset [12]. Hence it is often infeasible to use the adjacency matrix of the underlying graph as inputs of deep neural network models because of the high dimension of the adjacency matrix and the small number of labeled users.

We propose a novel framework that combines spectral graph analysis with the deep neural networks. In particular, we first project a graph to its spectral space formed by the principal eigenvectors of its adjacency matrix. The spectral space captures the main topological information of the graph. Each node is then mapped to a low dimensional point (called spectral coordinate) in the spectral space. We then use each node's spectral coordinate together with the aggregate information of its neighbor nodes' spectral coordinates as the input of two deep neural network models, deep autoencoder and convolutional neural network.

The advantages of our framework over past efforts are as follows. First, using both spectral graph analysis and deep neural networks, we can avoid defining graph metrics (features) to identify the difference between fraudsters and regular users. Second, the low-dimensional spectral space contains the most useful topology information of a graph. Comparing with the adjacency matrix, the dimension of spectral coordinates of nodes is much lower. Thus, using the node spectral coordinates as inputs to deep neural networks is suitable for real cases where the labeled users are limited. Moreover, most of the existing works for fraud detection focus on unsigned graphs in which there are only one type of links, while our framework covers signed networks. In order to capture both positive and negative edge information of a node in the signed graph, inputs of the two deep neural networks are composed by combining spectral coordinates of the node and its positive/negative-connected neighbors.

2 MODELS

2.1 Framework

Given a signed undirected graph G, each node in G indicates either a regular user or fraudster. The signed graph G can be represented as a symmetric adjacency matrix $\mathbf{A}_{n\times n}$, where n is the number of nodes. In $\mathbf{A}_{n\times n}$, $a_{ij} = 1$ ($a_{ij} = -1$) indicates there is a positive (negative) edge between nodes i and j and $a_{ij} = 0$ indicates no edge. A has n real eigenvalues. Let λ_i be the i-th largest eigenvalues of $\mathbf{A}$ with eigenvector $\mathbf{v}_i$. $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$. The spectral decomposition of $\mathbf{A}$ is $\mathbf{A} = \sum_i \lambda_i \mathbf{v}_i \mathbf{v}_i^t$ (shown in Figure 1). There usually exist k leading eigenvalues that are significantly greater than the rest ones for networks. The row vector $\boldsymbol{\sigma}_u = (v_{1u}, v_{2u}, \ldots, v_{ku})$ is

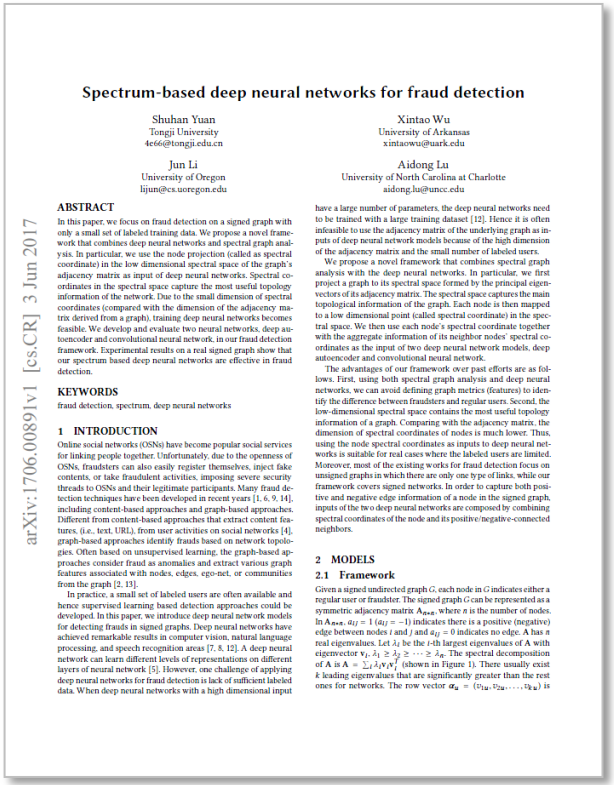
Shuhan Yuan Xintao Wu & ko

2017

Выборка 17 015 fraud, [illegible]5 good

Метрик[illegible] ccu[illegible]

Алгоритмы DAE, CNN vs. k-NN, SVM



Spectrum-based deep neural networks for fraud detection

Shuhan Yuan
Tongji University
4e66@tongji.edu.cn

Xintao Wu
University of Arkansas
xintaowu@uark.edu

Jun Li
University of Oregon
lijun@cs.uoregon.edu

Aidong Lu
University of North Carolina at Charlotte
aidong.lu@uncc.edu

ABSTRACT

In this paper, we focus on fraud detection on a signed graph with only a small set of labeled training data. We propose a novel framework that combines deep neural networks and spectral graph analysis. In particular, we use the node projection (called as spectral coordinate) in the low dimensional spectral space of the graph's adjacency matrix as input of deep neural networks. Spectral coordinates in the spectral space capture the most useful topology information of the network. Due to the small dimension of spectral coordinates (compared with the dimension of the adjacency matrix derived from a graph), training deep neural networks becomes feasible. We develop and evaluate two neural networks, deep autoencoder and convolutional neural network, in our fraud detection framework. Experimental results on a real signed graph show that our spectrum based deep neural networks are effective in fraud detection.

KEYWORDS

fraud detection, spectrum, deep neural networks

1 INTRODUCTION

Online social networks (OSNs) have become popular social services for linking people together. Unfortunately, due to the openness of OSNs, fraudsters can also easily register themselves, inject fake contents, or take fraudulent activities, imposing severe security threads to OSNs and their legitimate participants. Many fraud detection techniques have been developed in recent years [1, 6, 9, 14], including content-based approaches and graph-based approaches. Different from content-based approaches that extract content features, (i.e., text, URL), from user activities on social networks [4], graph-based approaches identify frauds based on network topologies. Often based on unsupervised learning, the graph-based approaches consider fraud as anomalies and extract various graph features associated with nodes, edges, ego-net, or communities from the graph [2, 13].

In practice, a small set of labeled users are often available and hence supervised learning based detection approaches could be developed. In this paper, we introduce deep neural network models for detecting frauds in signed graphs. Deep neural networks have achieved remarkable results in computer vision, natural language processing, and speech recognition areas [7, 8, 12]. A deep neural network can learn different levels of representations on different layers of neural network [5]. However, one challenge of applying deep neural networks for fraud detection is lack of sufficient labeled data. When deep neural networks with a high dimensional input have a large number of parameters, the deep neural networks need to be trained with a large training dataset [12]. Hence it is often infeasible to use the adjacency matrix of the underlying graph as inputs of deep neural network models because of the high dimension of the adjacency matrix and the small number of labeled users.

We propose a novel framework that combines spectral graph analysis with the deep neural networks. In particular, we first project a graph to its spectral space formed by the principal eigenvectors of its adjacency matrix. The spectral space captures the main topological information of the graph. Each node is then mapped to a low dimensional point (called spectral coordinate) in the spectral space. We then use each node's spectral coordinate together with the aggregate information of its neighbor nodes' spectral coordinates as the input of two deep neural network models, deep autoencoder and convolutional neural network.

The advantages of our framework over past efforts are as follows. First, using both spectral graph analysis and deep neural networks, we can avoid defining graph metrics (features) to identify the difference between fraudsters and regular users. Second, the low-dimensional spectral space contains the most useful topology information of a graph. Comparing with the adjacency matrix, the dimension of spectral coordinates of nodes is much lower. Thus, using the node spectral coordinates as inputs to deep neural networks is suitable for real cases where the labeled users are limited. Moreover, most of the existing works for fraud detection focus on unsigned graphs in which there are only one type of links, while our framework covers signed networks. In order to capture both positive and negative edge information of a node in the signed graph, inputs of the two deep neural networks are composed by combining spectral coordinates of the node and its positive/negative-connected neighbors.

2 MODELS

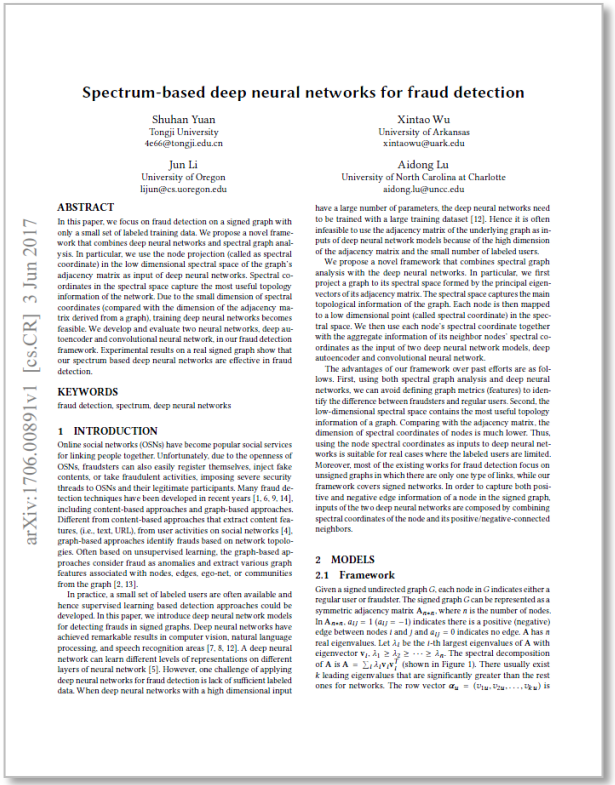
2.1 Framework

Given a signed undirected graph G, each node in G indicates either a regular user or fraudster. The signed graph G can be represented as a symmetric adjacency matrix $\mathbf{A}_{n\times n}$, where n is the number of nodes. In $\mathbf{A}_{n\times n}$, $a_{ij} = 1$ ($a_{ij} = -1$) indicates there is a positive (negative) edge between nodes i and j and $a_{ij} = 0$ indicates no edge. A has n real eigenvalues. Let λ_i be the i-th largest eigenvalues of A with eigenvector $\mathbf{v}_i$. $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$. The spectral decomposition of $\mathbf{A}$ is $\mathbf{A} = \sum_i \lambda_i \mathbf{v}_i \mathbf{v}_i^T$ (shown in Figure 1). There usually exist k leading eigenvalues that are significantly greater than the rest ones for networks. The row vector $\boldsymbol{\sigma}_u = (v_{1u}, v_{2u}, \ldots, v_{ku})$ is

Shuhan Yuan
Xintao Wu & ko

2017

Input	Algorithm	Ratio of the training dataset			
		5%	10%	15%	20%
A (adjacency matrix)	k-NN	66,16%	68,82%	69,66%	74,00%
	SVM	67,81%	67,82%	67,88%	67,92%
	DAE	76,31%	78,55%	79,56%	80,59%
	CNN	76,70%	78,95%	80,09%	81,33%
Xu (spectral coordinates)	k-NN	76,60%	77,38%	77,83%	78,19%
	SVM	71,60%	80,40%	80,82%	81,15%
	DAE	80,89%	81,13%	81,45%	81,92%
	CNN	80,57%	81,40%	82,02%	82,61%



Spectrum-based deep neural networks for fraud detection

Shuhan Yuan
Tongji University
4e66@tongji.edu.cn

Xintao Wu
University of Arkansas
xintaowu@uark.edu

Jun Li
University of Oregon
lijun@cs.uoregon.edu

Aidong Lu
University of North Carolina at Charlotte
aidong.lu@uncc.edu

ABSTRACT

In this paper, we focus on fraud detection on a signed graph with only a small set of labeled training data. We propose a novel framework that combines deep neural networks and spectral graph analysis. In particular, we use the node projection (called as spectral coordinate) in the low dimensional spectral space of the graph's adjacency matrix as input of deep neural networks. Spectral coordinates in the spectral space capture the most useful topology information of the network. Due to the small dimension of spectral coordinates (compared with the dimension of the adjacency matrix derived from a graph), training deep neural networks becomes feasible. We develop and evaluate two neural networks, deep autoencoder and convolutional neural network, in our fraud detection framework. Experimental results on a real signed graph show that our spectrum based deep neural networks are effective in fraud detection.

KEYWORDS

fraud detection, spectrum, deep neural networks

1 INTRODUCTION

Online social networks (OSNs) have become popular social services for linking people together. Unfortunately, due to the openness of OSNs, fraudsters can also easily register themselves, inject fake contents, or take fraudulent activities, imposing severe security threads to OSNs and their legitimate participants. Many fraud detection techniques have been developed in recent years [1, 6, 9, 14], including content-based approaches and graph-based approaches. Different from content-based approaches that extract content features, (i.e., text, URL), from user activities on social networks [4], graph-based approaches identify frauds based on network topologies. Often based on unsupervised learning, the graph-based approaches consider fraud as anomalies and extract various graph features associated with nodes, edges, ego-net, or communities from the graph [2, 13].

In practice, a small set of labeled users are often available and hence supervised learning based detection approaches could be developed. In this paper, we introduce deep neural network models for detecting frauds in signed graphs. Deep neural networks have achieved remarkable results in computer vision, natural language processing, and speech recognition areas [7, 8, 12]. A deep neural network can learn different levels of representations on different layers of neural network [5]. However, one challenge of applying deep neural networks for fraud detection is lack of sufficient labeled data. When deep neural networks with a high dimensional input have a large number of parameters, the deep neural networks need to be trained with a large training dataset [12]. Hence it is often infeasible to use the adjacency matrix of the underlying graph as inputs of deep neural network models because of the high dimension of the adjacency matrix and the small number of labeled users.

We propose a novel framework that combines spectral graph analysis with the deep neural networks. In particular, we first project a graph to its spectral space formed by the principal eigenvectors of its adjacency matrix. The spectral space captures the main topological information of the graph. Each node is then mapped to a low dimensional point (called spectral coordinate) in the spectral space. We then use each node's spectral coordinate together with the aggregate information of its neighbor nodes' spectral coordinates as the input of two deep neural network models, deep autoencoder and convolutional neural network.

The advantages of our framework over past efforts are as follows. First, using both spectral graph analysis and deep neural networks, we can avoid defining graph metrics (features) to identify the difference between fraudsters and regular users. Second, the low-dimensional spectral space contains the most useful topology information of a graph. Comparing with the adjacency matrix, the dimension of spectral coordinates of nodes is much lower. Thus, using the node spectral coordinates as inputs to deep neural networks is suitable for real cases where the labeled users are limited. Moreover, most of the existing works for fraud detection focus on unsigned graphs in which there are only one type of links, while our framework covers signed networks. In order to capture both positive and negative edge information of a node in the signed graph, inputs of the two deep neural networks are composed by combining spectral coordinates of the node and its positive/negative-connected neighbors.

2 MODELS

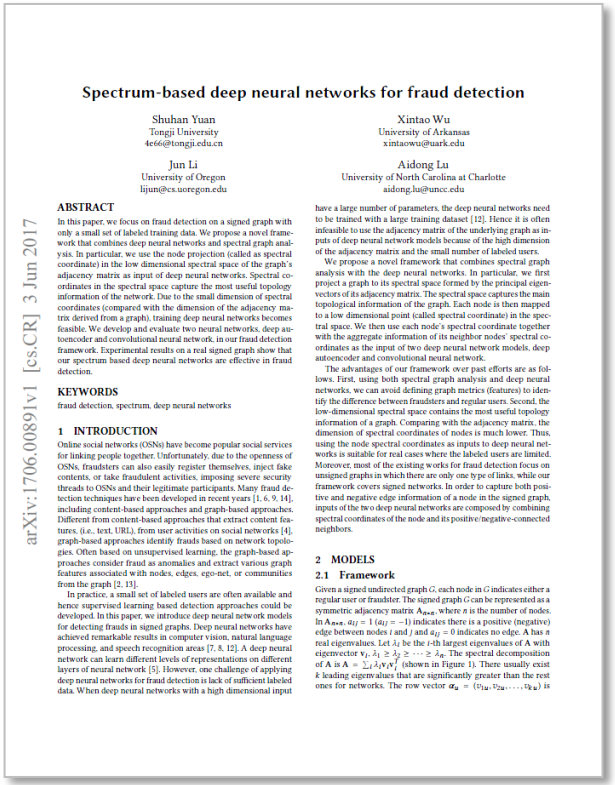
2.1 Framework

Given a signed undirected graph G, each node in G indicates either a regular user or fraudster. The signed graph G can be represented as a symmetric adjacency matrix $\mathbf{A}_{n\times n}$, where n is the number of nodes. In $\mathbf{A}_{n\times n}$, $a_{ij} = 1$ ($a_{ij} = -1$) indicates there is a positive (negative) edge between nodes i and j and $a_{ij} = 0$ indicates no edge. A has n real eigenvalues. Let λ_l be the l-th largest eigenvalues of $\mathbf{A}$ with eigenvector $\mathbf{v}_l$. $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$. The spectral decomposition of $\mathbf{A}$ is $\mathbf{A} = \sum_l \lambda_l \mathbf{v}_l \mathbf{v}_l^T$ (shown in Figure 1). There usually exist k leading eigenvalues that are significantly greater than the rest ones for networks. The row vector $\boldsymbol{\sigma}_u = (v_{1u}, v_{2u}, \ldots, v_{ku})$ is

Shuhan Yuan
Xintao Wu & ko

2017

- Используется **один датасет**, не ясна обобщающая способность алгоритмов
- Используется **маленькая выборка** для нейронных сетей (34030 примеров)
- Делается **спорный вывод** о том, что матрица смежности хуже спектральных координат



Spectrum-based deep neural networks for fraud detection

Shuhan Yuan
Tongji University
4e66@tongji.edu.cn

Xintao Wu
University of Arkansas
xintaowu@uark.edu

Jun Li
University of Oregon
lijun@cs.uoregon.edu

Aidong Lu
University of North Carolina at Charlotte
aidong.lu@uncc.edu

ABSTRACT

In this paper, we focus on fraud detection on a signed graph with only a small set of labeled training data. We propose a novel framework that combines deep neural networks and spectral graph analysis. In particular, we use the node projection (called as spectral coordinate) in the low dimensional spectral space of the graph's adjacency matrix as input of deep neural networks. Spectral coordinates in the spectral space capture the most useful topology information of the network. Due to the small dimension of spectral coordinates (compared with the dimension of the adjacency matrix derived from a graph), training deep neural networks becomes feasible. We develop and evaluate two neural networks, deep autoencoder and convolutional neural network, in our fraud detection framework. Experimental results on a real signed graph show that our spectrum based deep neural networks are effective in fraud detection.

KEYWORDS

fraud detection, spectrum, deep neural networks

1 INTRODUCTION

Online social networks (OSNs) have become popular social services for linking people together. Unfortunately, due to the openness of OSNs, fraudsters can also easily register themselves, inject fake contents, or take fraudulent activities, imposing severe security threads to OSNs and their legitimate participants. Many fraud detection techniques have been developed in recent years [1, 6, 9, 14], including content-based approaches and graph-based approaches. Different from content-based approaches that extract content features, (i.e., text, URL), from user activities on social networks [4], graph-based approaches identify frauds based on network topologies. Often based on unsupervised learning, the graph-based approaches consider fraud as anomalies and extract various graph features associated with nodes, edges, ego-net, or communities from the graph [2, 13].

In practice, a small set of labeled users are often available and hence supervised learning based detection approaches could be developed. In this paper, we introduce deep neural network models for detecting frauds in signed graphs. Deep neural networks have achieved remarkable results in computer vision, natural language processing, and speech recognition areas [7, 8, 12]. A deep neural network can learn different levels of representations on different layers of neural network [5]. However, one challenge of applying deep neural networks for fraud detection is lack of sufficient labeled data. When deep neural networks with a high dimensional input have a large number of parameters, the deep neural networks need to be trained with a large training dataset [12]. Hence it is often infeasible to use the adjacency matrix of the underlying graph as inputs of deep neural network models because of the high dimension of the adjacency matrix and the small number of labeled users.

We propose a novel framework that combines spectral graph analysis with the deep neural networks. In particular, we first project a graph to its spectral space formed by the principal eigenvectors of its adjacency matrix. The spectral space captures the main topological information of the graph. Each node is then mapped to a low dimensional point (called spectral coordinate) in the spectral space. We then use each node's spectral coordinate together with the aggregate information of its neighbor nodes' spectral coordinates as the input of two deep neural network models, deep autoencoder and convolutional neural network.

The advantages of our framework over past efforts are as follows. First, using both spectral graph analysis and deep neural networks, we can avoid defining graph metrics (features) to identify the difference between fraudsters and regular users. Second, the low-dimensional spectral space contains the most useful topology information of a graph. Comparing with the adjacency matrix, the dimension of spectral coordinates of nodes is much lower. Thus, using the node spectral coordinates as inputs to deep neural networks is suitable for real cases where the labeled users are limited. Moreover, most of the existing works for fraud detection focus on unsigned graphs in which there are only one type of links, while our framework covers signed networks. In order to capture both positive and negative edge information of a node in the signed graph, inputs of the two deep neural networks are composed by combining spectral coordinates of the node and its positive/negative-connected neighbors.

2 MODELS

2.1 Framework

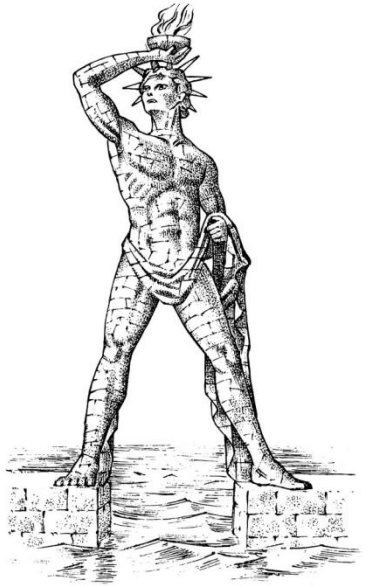
Given a signed undirected graph G, each node in G indicates either a regular user or fraudster. The signed graph G can be represented as a symmetric adjacency matrix $\mathbf{A}_{n\times n}$, where n is the number of nodes. In $\mathbf{A}_{n\times n}$, $a_{ij} = 1$ ($a_{ij} = -1$) indicates there is a positive (negative) edge between nodes i and j and $a_{ij} = 0$ indicates no edge. A has n real eigenvalues. Let λ_i be the i-th largest eigenvalues of $\mathbf{A}$ with eigenvector $\mathbf{v}_i$. $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$. The spectral decomposition of $\mathbf{A}$ is $\mathbf{A} = \sum_i \lambda_i \mathbf{v}_i \mathbf{v}_i^t$ (shown in Figure 1). There usually exist k leading eigenvalues that are significantly greater than the rest ones for networks. The row vector $\boldsymbol{\sigma}_u = (v_{1u}, v_{2u}, \ldots, v_{ku})$ is

Shuhan Yuan
Xintao Wu & ko

2017

- Показана универсальность CNN и DAE – глубокие сети хорошо работают и на графах
- Получены интересные результаты по упрощенному представлению графа
- Возможно поможет на собеседованиях в крупные банки

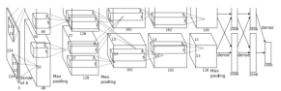
6. Гибридные нейронные сети



Можно строить разные архитектуры

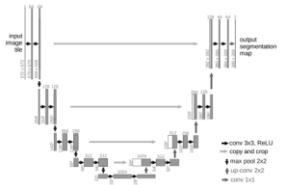
Для изображений

AlexNet
2012



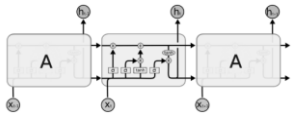
ResNet
2015

U-net
2015

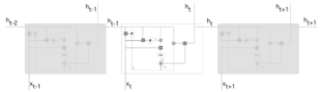


Для текстов

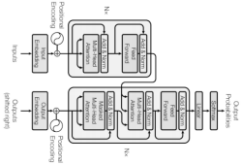
LSTM
1997



GRU
2014

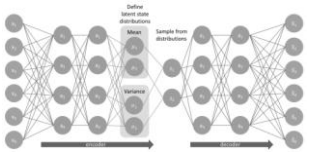


Transformer
2017

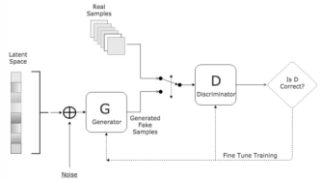


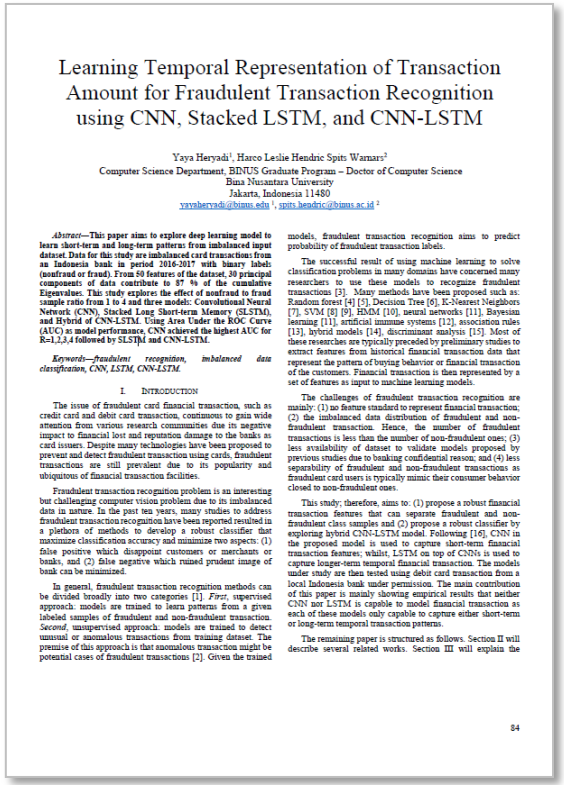
Для генерации

VAE
2014



GAN
2014





Learning Temporal Representation of Transaction Amount for Fraudulent Transaction Recognition using CNN, Stacked LSTM, and CNN-LSTM

Yaya Heryadi[1], Harco Leslie Hendric Spits Warnars[2]
Computer Science Department, BINUS Graduate Program – Doctor of Computer Science
Bina Nusantara University
Jakarta, Indonesia 11480
yayaheryadi@binus.edu [1], spits.hendric@binus.ac.id [2]

Abstract—**This paper aims to explore deep learning model to learn short-term and long-term patterns from imbalanced input dataset. Data for this study are imbalanced card transactions from an Indonesia bank in period 2016-2017 with binary label: (nonfraud or fraud). From 50 features of the dataset, 30 principal components of data contribute to 87 % of the cumulative Eigenvalues. This study explores the effect of nonfraud to fraud sample ratio from 1 to 4 and three models: Convolutional Neural Network (CNN), Stacked Long Short-term Memory (SLSTM), and Hybrid of CNN-LSTM. Using Area Under the ROC Curve (AUC) as model performance, CNN achieved the highest AUC for R=1,2,3,4 followed by SLSTM and CNN-LSTM.**

Keywords—fraudulent recognition, imbalanced data classification, CNN, LSTM, CNN-LSTM.

I. INTRODUCTION

The issue of fraudulent card financial transaction, such as credit card and debit card transaction, continuous to gain wide attention from various research communities due its negative impact to financial lost and reputation damage to the banks as card issuers. Despite many technologies have been proposed to prevent and detect fraudulent transaction using cards, fraudulent transactions are still prevalent due to its popularity and ubiquitous of financial transaction facilities.

Fraudulent transaction recognition problem is an interesting but challenging computer vision problem due to its imbalanced data in nature. In the past ten years, many studies to address fraudulent transaction recognition have been reported resulted in a plethora of methods to develop a robust classifier that maximize classification accuracy and minimize two aspects: (1) false positive which disappoint customers or merchants or banks, and (2) false negative which ruined prudent image of bank can be minimized.

In general, fraudulent transaction recognition methods can be divided broadly into two categories [1]. *First*, supervised approach: models are trained to learn patterns from a given labeled samples of fraudulent and non-fraudulent transaction. *Second*, unsupervised approach: models are trained to detect unusual or anomalous transactions from training dataset. The premise of this approach is that anomalous transaction might be potential cases of fraudulent transactions [2]. Given the trained models, fraudulent transaction recognition aims to predict probability of fraudulent transaction labels.

The successful result of using machine learning to solve classification problems in many domains have concerned many researchers to use these models to recognize fraudulent transactions [3]. Many methods have been proposed such as: Random forest [4] [5], Decision Tree [6], K-Nearest Neighbors [7], SVM [8] [9], HMM [10], neural networks [11], Bayesian learning [11], artificial immune systems [12], association rules [13], hybrid models [14], discriminant analysis [15]. Most of these researches are typically preceded by preliminary studies to extract features from historical financial transaction data that represent the pattern of buying behavior or financial transaction of the customers. Financial transaction is then represented by a set of features as input to machine learning models.

The challenges of fraudulent transaction recognition are mainly: (1) no feature standard to represent financial transaction; (2) the imbalanced data distribution of fraudulent and non-fraudulent transaction. Hence, the number of fraudulent transactions is less than the number of non-fraudulent ones; (3) less availability of dataset to validate models proposed by previous studies due to banking confidential reason; and (4) less separability of fraudulent and non-fraudulent transactions as fraudulent card users is typically mimic their consumer behavior closed to non-fraudulent ones.

This study; therefore, aims to: (1) propose a robust financial transaction features that can separate fraudulent and non-fraudulent class samples and (2) propose a robust classifier by exploring hybrid CNN-LSTM model. Following [16], CNN in the proposed model is used to capture short-term financial transaction features; whilst, LSTM on top of CNNs is used to capture longer-term temporal financial transaction. The models under study are then tested using debit card transaction from a local Indonesia bank under permission. The main contribution of this paper is mainly showing empirical results that neither CNN nor LSTM is capable to model financial transaction as each of these models only capable to capture either short-term or long-term temporal transaction patterns.

The remaining paper is structured as follows. Section II will describe several related works. Section III will explain the

84

Yaya Heryadi
Harco Warnars

2017

CNN	LSTM	CNN + LSTM
Выделяет краткосрочные локальные паттерны	Выделяет долгосрочные цепочки событий	Выделяет краткосрочные и долгосрочные паттерны

Learning Temporal Representation of Transaction Amount for Fraudulent Transaction Recognition using CNN, Stacked LSTM, and CNN-LSTM

Yaya Heryadi[1], Harco Leslie Hendric Spits Warnars[2]
Computer Science Department, BINUS Graduate Program – Doctor of Computer Science
Bina Nusantara University
Jakarta, Indonesia 11480
yayaheryadi@binus.edu [1], spits.hendric@binus.ac.id [2]

***Abstract*—This paper aims to explore deep learning model to learn short-term and long-term patterns from imbalanced input dataset. Data for this study are imbalanced card transactions from an Indonesia bank in period 2016-2017 with binary label: (nonfraud or fraud). From 50 features of the dataset, 30 principal components of data contribute to 87 % of the cumulative Eigenvalues. This study explores the effect of nonfraud to fraud sample ratio from 1 to 4 and three models: Convolutional Neural Network (CNN), Stacked Long Short-term Memory (SLSTM), and Hybrid of CNN-LSTM. Using Area Under the ROC Curve (AUC) as model performance, CNN achieved the highest AUC for R=1,2,3,4 followed by SLSTM and CNN-LSTM.**

Keywords—fraudulent recognition, imbalanced data classification, CNN, LSTM, CNN-LSTM.

I. Introduction

The issue of fraudulent card financial transaction, such as credit card and debit card transaction, continuous to gain wide attention from various research communities due its negative impact to financial lost and reputation damage to the banks as card issuers. Despite many technologies have been proposed to prevent and detect fraudulent transaction using cards, fraudulent transactions are still prevalent due to its popularity and ubiquitous of financial transaction facilities.

Fraudulent transaction recognition problem is an interesting but challenging computer vision problem due to its imbalanced data in nature. In the past ten years, many studies to address fraudulent transaction recognition have been reported resulted in a plethora of methods to develop a robust classifier that maximize classification accuracy and minimize two aspects: (1) false positive which disappoint customers or merchants or banks, and (2) false negative which ruined prudent image of bank can be minimized.

In general, fraudulent transaction recognition methods can be divided broadly into two categories [1]. *First*, supervised approach: models are trained to learn patterns from a given labeled samples of fraudulent and non-fraudulent transaction. *Second*, unsupervised approach: models are trained to detect unusual or anomalous transactions from training dataset. The premise of this approach is that anomalous transaction might be potential cases of fraudulent transactions [2]. Given the trained models, fraudulent transaction recognition aims to predict probability of fraudulent transaction labels.

The successful result of using machine learning to solve classification problems in many domains have concerned many researchers to use these models to recognize fraudulent transactions [3]. Many methods have been proposed such as: Random forest [4] [5], Decision Tree [6], K-Nearest Neighbors [7], SVM [8] [9], HMM [10], neural networks [11], Bayesian learning [11], artificial immune systems [12], association rules [13], hybrid models [14], discriminant analysis [15]. Most of these researches are typically preceded by preliminary studies to extract features from historical financial transaction data that represent the pattern of buying behavior or financial transaction of the customers. Financial transaction is then represented by a set of features as input to machine learning models.

The challenges of fraudulent transaction recognition are mainly: (1) no feature standard to represent financial transaction; (2) the imbalanced data distribution of fraudulent and non-fraudulent transaction. Hence, the number of fraudulent transactions is less than the number of non-fraudulent ones; (3) less availability of dataset to validate models proposed by previous studies due to banking confidential reason; and (4) less separability of fraudulent and non-fraudulent transactions as fraudulent card users is typically mimic their consumer behavior closed to non-fraudulent ones.

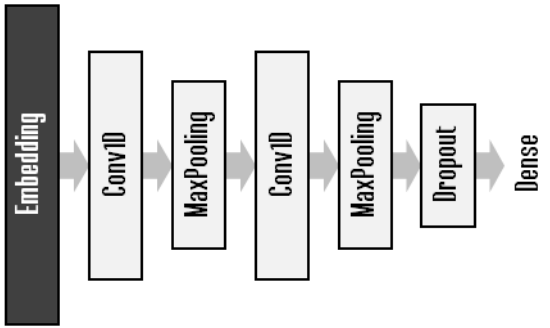
This study; therefore, aims to: (1) propose a robust financial transaction features that can separate fraudulent and non-fraudulent class samples and (2) propose a robust classifier by exploring hybrid CNN-LSTM model. Following [16], CNN in the proposed model is used to capture short-term financial transaction features; whilst, LSTM on top of CNNs is used to capture longer-term temporal financial transaction. The models under study are then tested using debit card transaction from a local Indonesia bank under permission. The main contribution of this paper is mainly showing empirical results that neither CNN nor LSTM is capable to model financial transaction as each of these models only capable to capture either short-term or long-term temporal transaction patterns.

The remaining paper is structured as follows. Section II will describe several related works. Section III will explain the

Yaya Heryadi
Harco Warnars

2017

CNN



Learning Temporal Representation of Transaction Amount for Fraudulent Transaction Recognition using CNN, Stacked LSTM, and CNN-LSTM

Yaya Heryadi[1], Harco Leslie Hendric Spits Warnars[2]
Computer Science Department, BINUS Graduate Program – Doctor of Computer Science
Bina Nusantara University
Jakarta, Indonesia 11480
yayaheryadi@binus.edu [1], spits.hendric@binus.ac.id [2]

***Abstract*—This paper aims to explore deep learning model to learn short-term and long-term patterns from imbalanced input dataset. Data for this study are imbalanced card transactions from an Indonesia bank in period 2016-2017 with binary labels (nonfraud or fraud). From 50 features of the dataset, 30 principal components of data contribute to 87 % of the cumulative Eigenvalues. This study explores the effect of nonfraud to fraud sample ratio from 1 to 4 and three models: Convolutional Neural Network (CNN), Stacked Long Short-term Memory (SLSTM), and Hybrid of CNN-LSTM. Using Area Under the ROC Curve (AUC) as model performance, CNN achieved the highest AUC for R=1,2,3,4 followed by SLSTM and CNN-LSTM.**

Keywords*—*fraudulent recognition, imbalanced data classification, CNN, LSTM, CNN-LSTM.

I. INTRODUCTION

The issue of fraudulent card financial transaction, such as credit card and debit card transaction, continuous to gain wide attention from various research communities due its negative impact to financial lost and reputation damage to the banks as card issuers. Despite many technologies have been proposed to prevent and detect fraudulent transaction using cards, fraudulent transactions are still prevalent due to its popularity and ubiquitous of financial transaction facilities.

Fraudulent transaction recognition problem is an interesting but challenging computer vision problem due to its imbalanced data in nature. In the past ten years, many studies to address fraudulent transaction recognition have been reported resulted in a plethora of methods to develop a robust classifier that maximize classification accuracy and minimize two aspects: (1) false positive which disappoint customers or merchants or banks, and (2) false negative which ruined prudent image of bank can be minimized.

In general, fraudulent transaction recognition methods can be divided broadly into two categories [1]. *First*, supervised approach: models are trained to learn patterns from a given labeled samples of fraudulent and non-fraudulent transaction. *Second*, unsupervised approach: models are trained to detect unusual or anomalous transactions from training dataset. The premise of this approach is that anomalous transaction might be potential cases of fraudulent transactions [2]. Given the trained models, fraudulent transaction recognition aims to predict probability of fraudulent transaction labels.

The successful result of using machine learning to solve classification problems in many domains have concerned many researchers to use these models to recognize fraudulent transactions [3]. Many methods have been proposed such as: Random forest [4] [5], Decision Tree [6], K-Nearest Neighbors [7], SVM [8] [9], HMM [10], neural networks [11], Bayesian learning [11], artificial immune systems [12], association rules [13], hybrid models [14], discriminant analysis [15]. Most of these researches are typically preceded by preliminary studies to extract features from historical financial transaction data that represent the pattern of buying behavior or financial transaction of the customers. Financial transaction is then represented by a set of features as input to machine learning models.

The challenges of fraudulent transaction recognition are mainly: (1) no feature standard to represent financial transaction; (2) the imbalanced data distribution of fraudulent and non-fraudulent transaction. Hence, the number of fraudulent transactions is less than the number of non-fraudulent ones; (3) less availability of dataset to validate models proposed by previous studies due to banking confidential reason; and (4) less separability of fraudulent and non-fraudulent transactions as fraudulent card users is typically mimic their consumer behavior closed to non-fraudulent ones.

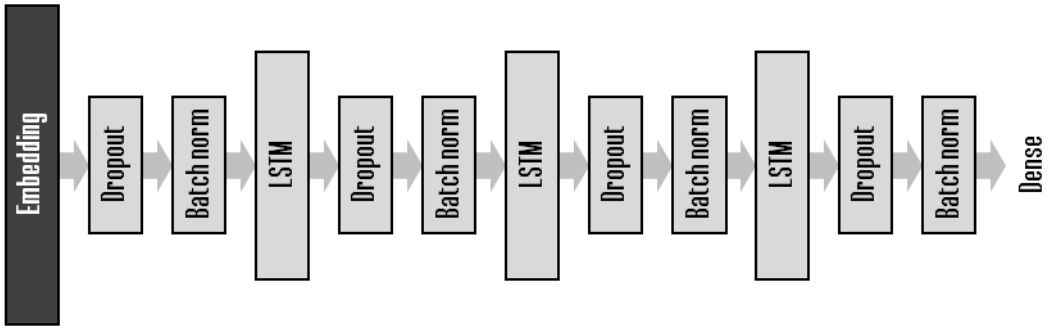
This study, therefore, aims to: (1) propose a robust financial transaction features that can separate fraudulent and non-fraudulent class samples and (2) propose a robust classifier by exploring hybrid CNN-LSTM model. Following [16], CNN in the proposed model is used to capture short-term financial transaction features; whilst, LSTM on top of CNNs is used to capture longer-term temporal financial transaction. The models under study are then tested using debit card transaction from a local Indonesia bank under permission. The main contribution of this paper is mainly showing empirical results that neither CNN nor LSTM is capable to model financial transaction as each of these models only capable to capture either short-term or long-term temporal transaction patterns.

The remaining paper is structured as follows. Section II will describe several related works. Section III will explain the

Yaya Heryadi
Harco Warnars

2017

Stacked LSTM



Learning Temporal Representation of Transaction Amount for Fraudulent Transaction Recognition using CNN, Stacked LSTM, and CNN-LSTM

Yaya Heryadi[1], Harco Leslie Hendric Spits Warnars[2]
Computer Science Department, BINUS Graduate Program – Doctor of Computer Science
Bina Nusantara University
Jakarta, Indonesia 11480
yayaheryadi@binus.edu [1], spits.hendric@binus.ac.id [2]

***Abstract*—This paper aims to explore deep learning model to learn short-term and long-term patterns from imbalanced input dataset. Data for this study are imbalanced card transactions from an Indonesia bank in period 2016-2017 with binary label (nonfraud or fraud). From 50 features of the dataset, 30 principal components of data contribute to 87 % of the cumulative Eigenvalues. This study explores the effect of nonfraud to fraud sample ratio from 1 to 4 and three models: Convolutional Neural Network (CNN), Stacked Long Short-term Memory (SLSTM), and Hybrid of CNN-LSTM. Using Area Under the ROC Curve (AUC) as model performance, CNN achieved the highest AUC for R=1,2,3,4 followed by SLSTM and CNN-LSTM.**

Keywords*—*fraudulent recognition, imbalanced data classification, CNN, LSTM, CNN-LSTM.

I. INTRODUCTION

The issue of fraudulent card financial transaction, such as credit card and debit card transaction, continuous to gain wide attention from various research communities due its negative impact to financial lost and reputation damage to the banks as card issuers. Despite many technologies have been proposed to prevent and detect fraudulent transaction using cards, fraudulent transactions are still prevalent due to its popularity and ubiquitous of financial transaction facilities.

Fraudulent transaction recognition problem is an interesting but challenging computer vision problem due to its imbalanced data in nature. In the past ten years, many studies to address fraudulent transaction recognition have been reported resulted in a plethora of methods to develop a robust classifier that maximize classification accuracy and minimize two aspects: (1) false positive which disappoint customers or merchants or banks, and (2) false negative which ruined prudent image of bank can be minimized.

In general, fraudulent transaction recognition methods can be divided broadly into two categories [1]. *First*, supervised approach: models are trained to learn patterns from a given labeled samples of fraudulent and non-fraudulent transaction. *Second*, unsupervised approach: models are trained to detect unusual or anomalous transactions from training dataset. The premise of this approach is that anomalous transaction might be potential cases of fraudulent transactions [2]. Given the trained models, fraudulent transaction recognition aims to predict probability of fraudulent transaction labels.

The successful result of using machine learning to solve classification problems in many domains have concerned many researchers to use these models to recognize fraudulent transactions [3]. Many methods have been proposed such as: Random forest [4] [5], Decision Tree [6], K-Nearest Neighbors [7], SVM [8] [9], HMM [10], neural networks [11], Bayesian learning [11], artificial immune systems [12], association rules [13], hybrid models [14], discriminant analysis [15]. Most of these researches are typically preceded by preliminary studies to extract features from historical financial transaction data that represent the pattern of buying behavior or financial transaction of the customers. Financial transaction is then represented by a set of features as input to machine learning models.

The challenges of fraudulent transaction recognition are mainly: (1) no feature standard to represent financial transaction; (2) the imbalanced data distribution of fraudulent and non-fraudulent transaction. Hence, the number of fraudulent transactions is less than the number of non-fraudulent ones; (3) less availability of dataset to validate models proposed by previous studies due to banking confidential reason; and (4) less separability of fraudulent and non-fraudulent transactions as fraudulent card users is typically mimic their consumer behavior closed to non-fraudulent ones.

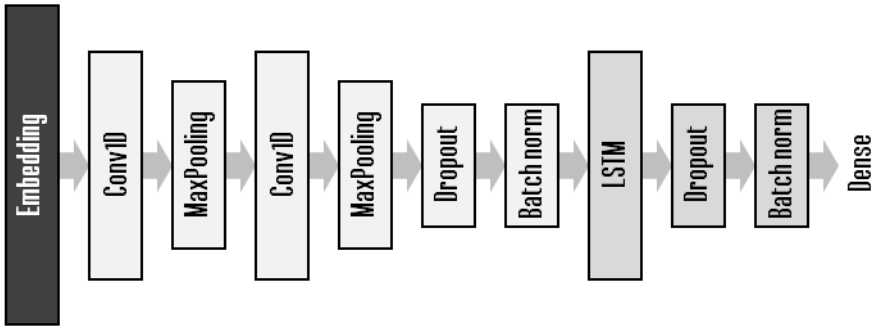
This study; therefore, aims to: (1) propose a robust financial transaction features that can separate fraudulent and non-fraudulent class samples and (2) propose a robust classifier by exploring hybrid CNN-LSTM model. Following [16], CNN in the proposed model is used to capture short-term financial transaction features; whilst, LSTM on top of CNNs is used to capture longer-term temporal financial transaction. The models under study are then tested using debit card transaction from a local Indonesia bank under permission. The main contribution of this paper is mainly showing empirical results that neither CNN nor LSTM is capable to model financial transaction as each of these models only capable to capture either short-term or long-term temporal transaction patterns.

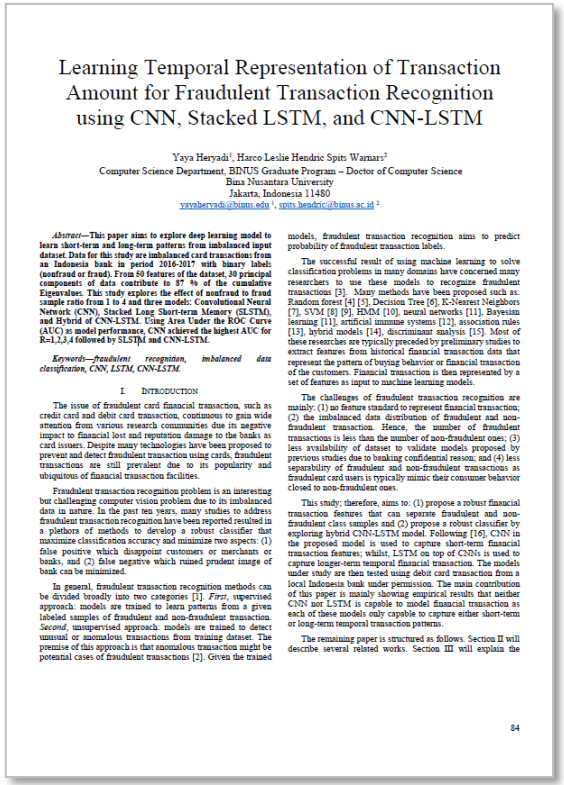
The remaining paper is structured as follows. Section II will describe several related works. Section III will explain the

Yaya Heryadi
Harco Warnars

2017

CNN-LSTM





Learning Temporal Representation of Transaction Amount for Fraudulent Transaction Recognition using CNN, Stacked LSTM, and CNN-LSTM

Yaya Heryadi[1], Harco Leslie Hendric Spits Warnars[2]
Computer Science Department, BINUS Graduate Program – Doctor of Computer Science
Bina Nusantara University
Jakarta, Indonesia 11480
yayaheryadi@binus.edu [1], spits.hendric@binus.ac.id [2]

***Abstract*—This paper aims to explore deep learning model to learn short-term and long-term patterns from imbalanced input dataset. Data for this study are imbalanced card transactions from an Indonesia bank in period 2016-2017 with binary label: (nonfraud or fraud). From 50 features of the dataset, 30 principal components of data contribute to 87 % of the cumulative Eigenvalues. This study explores the effect of nonfraud to fraud sample ratio from 1 to 4 and three models: Convolutional Neural Network (CNN), Stacked Long Short-term Memory (SLSTM), and Hybrid of CNN-LSTM. Using Area Under the ROC Curve (AUC) as model performance, CNN achieved the highest AUC for R=1,2,3,4 followed by SLSTM and CNN-LSTM.**

Keywords*—*fraudulent recognition, imbalanced data classification, CNN, LSTM, CNN-LSTM.

I. INTRODUCTION

The issue of fraudulent card financial transaction, such as credit card and debit card transaction, continuous to gain wide attention from various research communities due its negative impact to financial lost and reputation damage to the banks as card issuers. Despite many technologies have been proposed to prevent and detect fraudulent transaction using cards, fraudulent transactions are still prevalent due to its popularity and ubiquitous of financial transaction facilities.

Fraudulent transaction recognition problem is an interesting but challenging computer vision problem due to its imbalanced data in nature. In the past ten years, many studies to address fraudulent transaction recognition have been reported resulted in a plethora of methods to develop a robust classifier that maximize classification accuracy and minimize two aspects: (1) false positive which disappoint customers or merchants or banks, and (2) false negative which ruined prudent image of bank can be minimized.

In general, fraudulent transaction recognition methods can be divided broadly into two categories [1]. *First*, supervised approach: models are trained to learn patterns from a given labeled samples of fraudulent and non-fraudulent transaction. *Second*, unsupervised approach: models are trained to detect unusual or anomalous transactions from training dataset. The premise of this approach is that anomalous transaction might be potential cases of fraudulent transactions [2]. Given the trained models, fraudulent transaction recognition aims to predict probability of fraudulent transaction labels.

The successful result of using machine learning to solve classification problems in many domains have concerned many researchers to use these models to recognize fraudulent transactions [3]. Many methods have been proposed such as: Random forest [4] [5], Decision Tree [6], K-Nearest Neighbors [7], SVM [8] [9], HMM [10], neural networks [11], Bayesian learning [11], artificial immune systems [12], association rules [13], hybrid models [14], discriminant analysis [15]. Most of these researches are typically preceded by preliminary studies to extract features from historical financial transaction data that represent the pattern of buying behavior or financial transaction of the customers. Financial transaction is then represented by a set of features as input to machine learning models.

The challenges of fraudulent transaction recognition are mainly: (1) no feature standard to represent financial transaction; (2) the imbalanced data distribution of fraudulent and non-fraudulent transaction. Hence, the number of fraudulent transactions is less than the number of non-fraudulent ones; (3) less availability of dataset to validate models proposed by previous studies due to banking confidential reason; and (4) less separability of fraudulent and non-fraudulent transactions as fraudulent card users is typically mimic their consumer behavior closed to non-fraudulent ones.

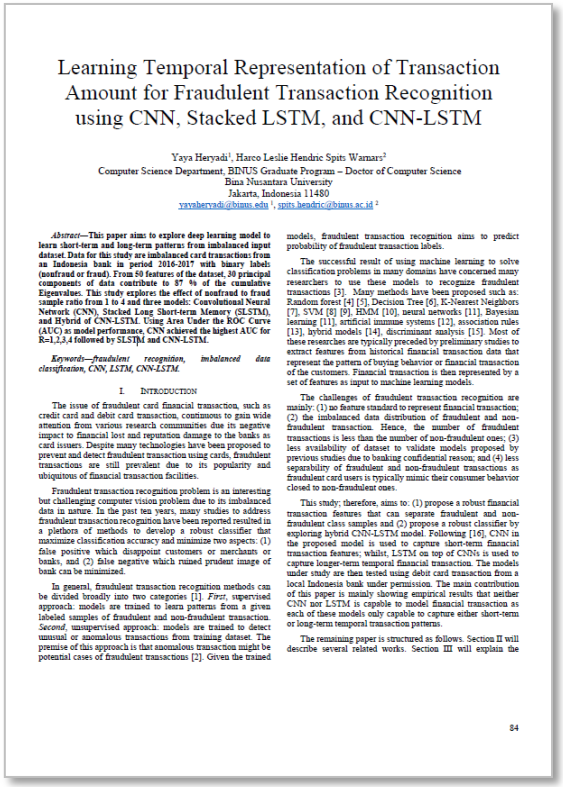
This study; therefore, aims to: (1) propose a robust financial transaction features that can separate fraudulent and non-fraudulent class samples and (2) propose a robust classifier by exploring hybrid CNN-LSTM model. Following [16], CNN in the proposed model is used to capture short-term financial transaction features; whilst, LSTM on top of CNNs is used to capture longer-term temporal financial transaction. The models under study are then tested using debit card transaction from a local Indonesia bank under permission. The main contribution of this paper is mainly showing empirical results that neither CNN nor LSTM is capable to model financial transaction as each of these models only capable to capture either short-term or long-term temporal transaction patterns.

The remaining paper is structured as follows. Section II will describe several related works. Section III will explain the

Yaya Heryadi
Harco Warnars

2017

Выборка	2 896 fraud, 11 584 good
Метрики	Accuracy, AUC ROC
Алгоритмы	SNN, SLSTM, SNN-LSTM



Learning Temporal Representation of Transaction Amount for Fraudulent Transaction Recognition using CNN, Stacked LSTM, and CNN-LSTM

Yaya Heryadi[1], Harco Leslie Hendric Spits Warnars[2]
Computer Science Department, BINUS Graduate Program – Doctor of Computer Science
Bina Nusantara University
Jakarta, Indonesia 11480
yayaheryadi@binus.edu [1], spits.hendric@binus.ac.id [2]

***Abstract*—This paper aims to explore deep learning model to learn short-term and long-term patterns from imbalanced input dataset. Data for this study are imbalanced card transactions from an Indonesia bank in period 2016-2017 with binary labels (nonfraud or fraud). From 50 features of the dataset, 30 principal components of data contribute to 87 % of the cumulative Eigenvalues. This study explores the effect of nonfraud to fraud sample ratio from 1 to 4 and three models: Convolutional Neural Network (CNN), Stacked Long Short-term Memory (SLSTM), and Hybrid of CNN-LSTM. Using Area Under the ROC Curve (AUC) as model performance, CNN achieved the highest AUC for R=1,2,3,4 followed by SLSTM and CNN-LSTM.**

***Keywords*—fraudulent recognition, imbalanced data classification, CNN, LSTM, CNN-LSTM.**

I. INTRODUCTION

The issue of fraudulent card financial transaction, such as credit card and debit card transaction, continuous to gain wide attention from various research communities due its negative impact to financial lost and reputation damage to the banks as card issuers. Despite many technologies have been proposed to prevent and detect fraudulent transaction using cards, fraudulent transactions are still prevalent due to its popularity and ubiquitous of financial transaction facilities.

Fraudulent transaction recognition problem is an interesting but challenging computer vision problem due to its imbalanced data in nature. In the past ten years, many studies to address fraudulent transaction recognition have been reported resulted in a plethora of methods to develop a robust classifier that maximize classification accuracy and minimize two aspects: (1) false positive which disappoint customers or merchants or banks, and (2) false negative which ruined prudent image of bank can be minimized.

In general, fraudulent transaction recognition methods can be divided broadly into two categories [1]. *First*, supervised approach: models are trained to learn patterns from a given labeled samples of fraudulent and non-fraudulent transaction. *Second*, unsupervised approach: models are trained to detect unusual or anomalous transactions from training dataset. The premise of this approach is that anomalous transaction might be potential cases of fraudulent transactions [2]. Given the trained models, fraudulent transaction recognition aims to predict probability of fraudulent transaction labels.

The successful result of using machine learning to solve classification problems in many domains have concerned many researchers to use these models to recognize fraudulent transactions [3]. Many methods have been proposed such as: Random forest [4] [5], Decision Tree [6], K-Nearest Neighbors [7], SVM [8] [9], HMM [10], neural networks [11], Bayesian learning [11], artificial immune systems [12], association rules [13], hybrid models [14], discriminant analysis [15]. Most of these researches are typically preceded by preliminary studies to extract features from historical financial transaction data that represent the pattern of buying behavior or financial transaction of the customers. Financial transaction is then represented by a set of features as input to machine learning models.

The challenges of fraudulent transaction recognition are mainly: (1) no feature standard to represent financial transaction; (2) the imbalanced data distribution of fraudulent and non-fraudulent transaction. Hence, the number of fraudulent transactions is less than the number of non-fraudulent ones; (3) less availability of dataset to validate models proposed by previous studies due to banking confidential reason; and (4) less separability of fraudulent and non-fraudulent transactions as fraudulent card users is typically mimic their consumer behavior closed to non-fraudulent ones.

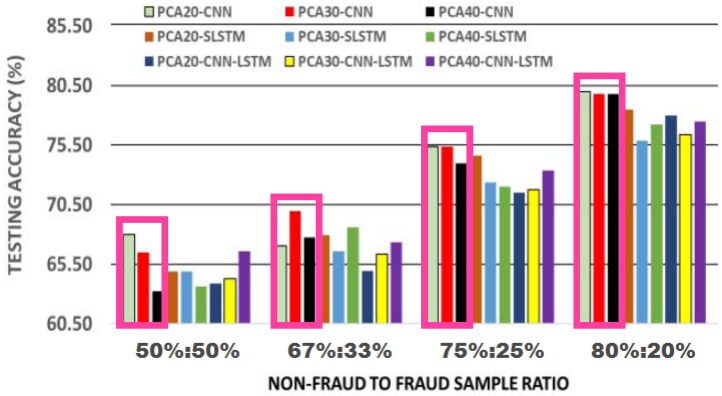
This study, therefore, aims to: (1) propose a robust financial transaction features that can separate fraudulent and non-fraudulent class samples and (2) propose a robust classifier by exploring hybrid CNN-LSTM model. Following [16], CNN in the proposed model is used to capture short-term financial transaction features; whilst, LSTM on top of CNNs is used to capture longer-term temporal financial transaction. The models under study are then tested using debit card transaction from a local Indonesia bank under permission. The main contribution of this paper is mainly showing empirical results that neither CNN nor LSTM is capable to model financial transaction as each of these models only capable to capture either short-term or long-term temporal transaction patterns.

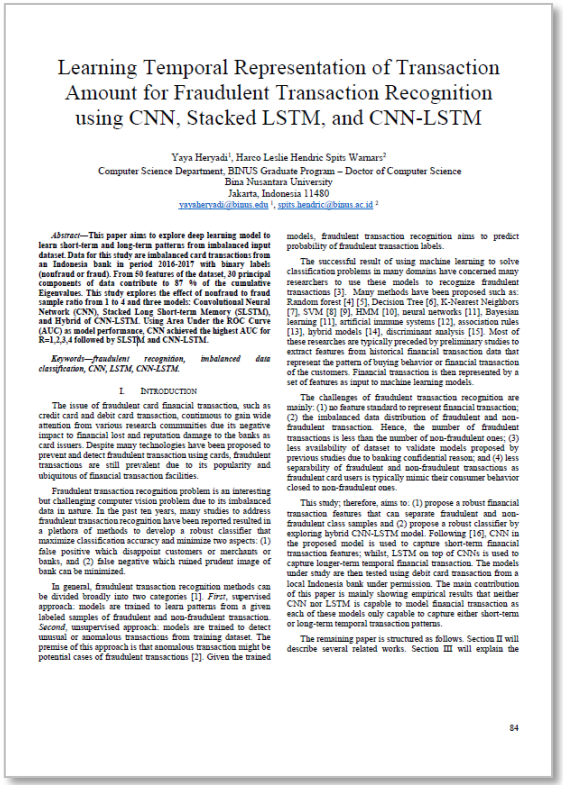
The remaining paper is structured as follows. Section II will describe several related works. Section III will explain the

Yaya Heryadi
Harco Warnars

2017

CNN лучше SLSTM и SNN-LSTM





Learning Temporal Representation of Transaction Amount for Fraudulent Transaction Recognition using CNN, Stacked LSTM, and CNN-LSTM

Yaya Heryadi[1], Harco Leslie Hendric Spits Warnars[2]
Computer Science Department, BINUS Graduate Program – Doctor of Computer Science
Bina Nusantara University
Jakarta, Indonesia 11480
yayaheryadi@binus.edu [1], spits.hendric@binus.ac.id [2]

Abstract—**This paper aims to explore deep learning model to learn short-term and long-term patterns from imbalanced input dataset. Data for this study are imbalanced card transactions from an Indonesia bank in period 2016-2017 with binary labels (nonfraud or fraud). From 50 features of the dataset, 30 principal components of data contribute to 87 % of the cumulative Eigenvalues. This study explores the effect of nonfraud to fraud sample ratio from 1 to 4 and three models: Convolutional Neural Network (CNN), Stacked Long Short-term Memory (SLSTM), and Hybrid of CNN-LSTM. Using Area Under the ROC Curve (AUC) as model performance, CNN achieved the highest AUC for R=1,2,3,4 followed by SLSTM and CNN-LSTM.**

Keywords—fraudulent recognition, imbalanced data classification, CNN, LSTM, CNN-LSTM.

I. INTRODUCTION

The issue of fraudulent card financial transaction, such as credit card and debit card transaction, continuous to gain wide attention from various research communities due its negative impact to financial lost and reputation damage to the banks as card issuers. Despite many technologies have been proposed to prevent and detect fraudulent transaction using cards, fraudulent transactions are still prevalent due to its popularity and ubiquitous of financial transaction facilities.

Fraudulent transaction recognition problem is an interesting but challenging computer vision problem due to its imbalanced data in nature. In the past ten years, many studies to address fraudulent transaction recognition have been reported resulted in a plethora of methods to develop a robust classifier that maximize classification accuracy and minimize two aspects: (1) false positive which disappoint customers or merchants or banks, and (2) false negative which ruined prudent image of bank can be minimized.

In general, fraudulent transaction recognition methods can be divided broadly into two categories [1]. *First*, supervised approach: models are trained to learn patterns from a given labeled samples of fraudulent and non-fraudulent transaction. *Second*, unsupervised approach: models are trained to detect unusual or anomalous transactions from training dataset. The premise of this approach is that anomalous transaction might be potential cases of fraudulent transactions [2]. Given the trained models, fraudulent transaction recognition aims to predict probability of fraudulent transaction labels.

The successful result of using machine learning to solve classification problems in many domains have concerned many researchers to use these models to recognize fraudulent transactions [3]. Many methods have been proposed such as: Random forest [4] [5], Decision Tree [6], K-Nearest Neighbors [7], SVM [8] [9], HMM [10], neural networks [11], Bayesian learning [11], artificial immune systems [12], association rules [13], hybrid models [14], discriminant analysis [15]. Most of these researches are typically preceded by preliminary studies to extract features from historical financial transaction data that represent the pattern of buying behavior or financial transaction of the customers. Financial transaction is then represented by a set of features as input to machine learning models.

The challenges of fraudulent transaction recognition are mainly: (1) no feature standard to represent financial transaction; (2) the imbalanced data distribution of fraudulent and non-fraudulent transaction. Hence, the number of fraudulent transactions is less than the number of non-fraudulent ones; (3) less availability of dataset to validate models proposed by previous studies due to banking confidential reason; and (4) less separability of fraudulent and non-fraudulent transactions as fraudulent card users is typically mimic their consumer behavior closed to non-fraudulent ones.

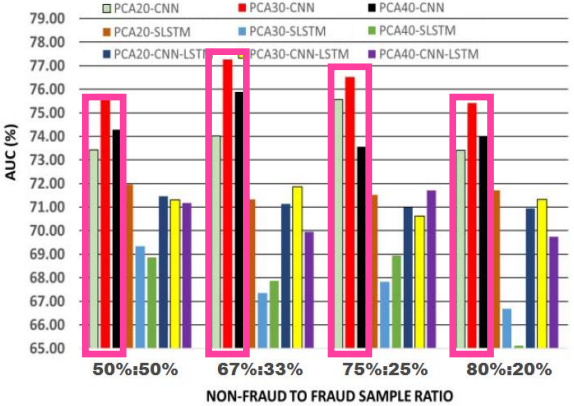
This study; therefore, aims to: (1) propose a robust financial transaction features that can separate fraudulent and non-fraudulent class samples and (2) propose a robust classifier by exploring hybrid CNN-LSTM model. Following [16], CNN in the proposed model is used to capture short-term financial transaction features; whilst, LSTM on top of CNNs is used to capture longer-term temporal financial transaction. The models under study are then tested using debit card transaction from a local Indonesia bank under permission. The main contribution of this paper is mainly showing empirical results that neither CNN nor LSTM is capable to model financial transaction as each of these models only capable to capture either short-term or long-term temporal transaction patterns.

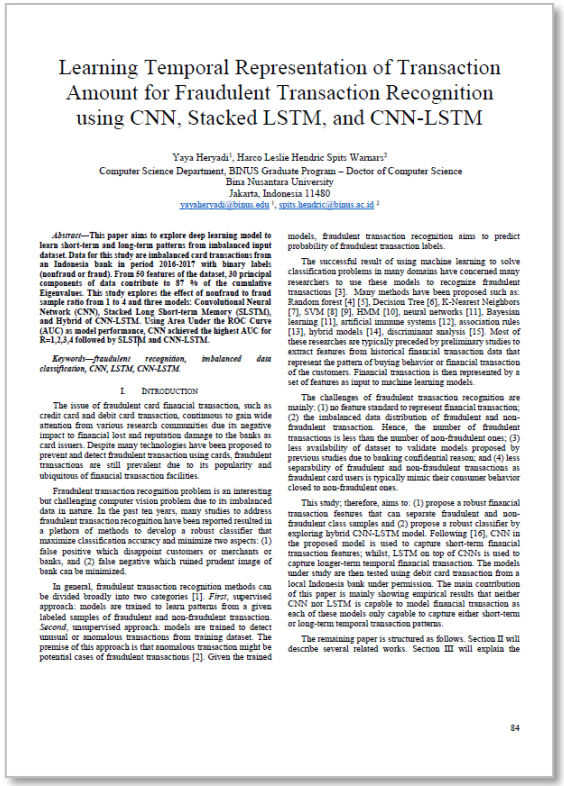
The remaining paper is structured as follows. Section II will describe several related works. Section III will explain the

Yaya Heryadi
Harco Warnars

2017

CNN лучше SLSTM и SNN-LSTM





Learning Temporal Representation of Transaction Amount for Fraudulent Transaction Recognition using CNN, Stacked LSTM, and CNN-LSTM

Yaya Heryadi[1], Harco Leslie Hendric Spits Warnars[2]
Computer Science Department, BINUS Graduate Program – Doctor of Computer Science
Bina Nusantara University
Jakarta, Indonesia 11480
yayaheryadi@binus.edu [1], spits.hendric@binus.ac.id [2]

Abstract—**This paper aims to explore deep learning model to learn short-term and long-term patterns from imbalanced input dataset. Data for this study are imbalanced card transactions from an Indonesia bank in period 2016-2017 with binary label: (nonfraud or fraud). From 50 features of the dataset, 30 principal components of data contribute to 87 % of the cumulative Eigenvalues. This study explores the effect of nonfraud to fraud sample ratio from 1 to 4 and three models: Convolutional Neural Network (CNN), Stacked Long Short-term Memory (SLSTM), and Hybrid of CNN-LSTM. Using Area Under the ROC Curve (AUC) as model performance, CNN achieved the highest AUC for R=1,2,3,4 followed by SLSTM and CNN-LSTM.**

Keywords—fraudulent recognition, imbalanced data classification, CNN, LSTM, CNN-LSTM.

I. INTRODUCTION

The issue of fraudulent card financial transaction, such as credit card and debit card transaction, continuous to gain wide attention from various research communities due its negative impact to financial lost and reputation damage to the banks as card issuers. Despite many technologies have been proposed to prevent and detect fraudulent transaction using cards, fraudulent transactions are still prevalent due to its popularity and ubiquitous of financial transaction facilities.

Fraudulent transaction recognition problem is an interesting but challenging computer vision problem due to its imbalanced data in nature. In the past ten years, many studies to address fraudulent transaction recognition have been reported resulted in a plethora of methods to develop a robust classifier that maximize classification accuracy and minimize two aspects: (1) false positive which disappoint customers or merchants or banks, and (2) false negative which ruined prudent image of bank can be minimized.

In general, fraudulent transaction recognition methods can be divided broadly into two categories [1]. *First*, supervised approach: models are trained to learn patterns from a given labeled samples of fraudulent and non-fraudulent transaction. *Second*, unsupervised approach: models are trained to detect unusual or anomalous transactions from training dataset. The premise of this approach is that anomalous transaction might be potential cases of fraudulent transactions [2]. Given the trained models, fraudulent transaction recognition aims to predict probability of fraudulent transaction labels.

The successful result of using machine learning to solve classification problems in many domains have concerned many researchers to use these models to recognize fraudulent transactions [3]. Many methods have been proposed such as: Random forest [4] [5], Decision Tree [6], K-Nearest Neighbors [7], SVM [8] [9], HMM [10], neural networks [11], Bayesian learning [11], artificial immune systems [12], association rules [13], hybrid models [14], discriminant analysis [15]. Most of these researches are typically preceded by preliminary studies to extract features from historical financial transaction data that represent the pattern of buying behavior or financial transaction of the customers. Financial transaction is then represented by a set of features as input to machine learning models.

The challenges of fraudulent transaction recognition are mainly: (1) no feature standard to represent financial transaction; (2) the imbalanced data distribution of fraudulent and non-fraudulent transaction. Hence, the number of fraudulent transactions is less than the number of non-fraudulent ones; (3) less availability of dataset to validate models proposed by previous studies due to banking confidential reason; and (4) less separability of fraudulent and non-fraudulent transactions as fraudulent card users is typically mimic their consumer behavior closed to non-fraudulent ones.

This study; therefore, aims to: (1) propose a robust financial transaction features that can separate fraudulent and non-fraudulent class samples and (2) propose a robust classifier by exploring hybrid CNN-LSTM model. Following [16], CNN in the proposed model is used to capture short-term financial transaction features; whilst, LSTM on top of CNNs is used to capture longer-term temporal financial transaction. The models under study are then tested using debit card transaction from a local Indonesia bank under permission. The main contribution of this paper is mainly showing empirical results that neither CNN nor LSTM is capable to model financial transaction as each of these models only capable to capture either short-term or long-term temporal transaction patterns.

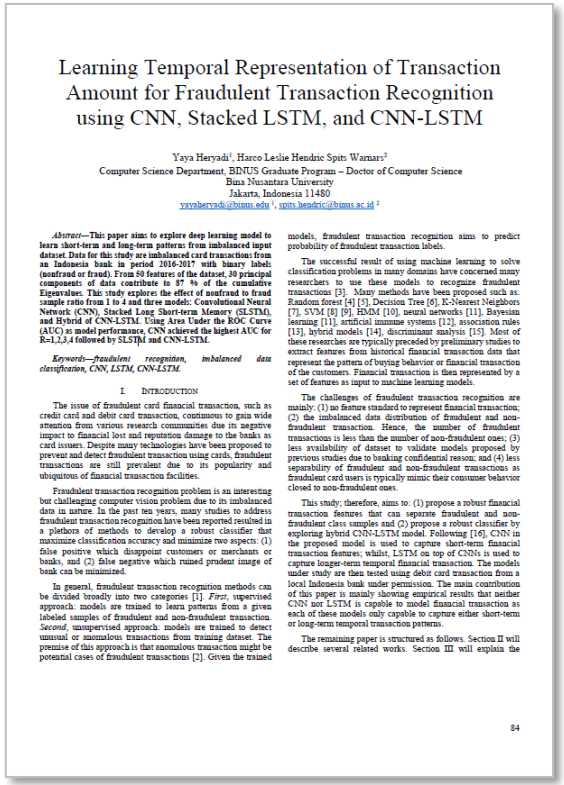
The remaining paper is structured as follows. Section II will describe several related works. Section III will explain the

84

Yaya Heryadi
Harco Warnars

2017

- Используют маленькие датасеты для глубоких нейронных сетей
- Намешали много техник без объяснения целесообразности их использования
- Получили противоречивые результаты для разных метрик



Learning Temporal Representation of Transaction Amount for Fraudulent Transaction Recognition using CNN, Stacked LSTM, and CNN-LSTM

Yaya Heryadi[1], Harco Leslie Hendric Spits Warnars[2]
Computer Science Department, BINUS Graduate Program – Doctor of Computer Science
Bina Nusantara University
Jakarta, Indonesia 11480
yayaheryadi@binus.edu [1], spits.hendric@binus.ac.id [2]

***Abstract*—This paper aims to explore deep learning model to learn short-term and long-term patterns from imbalanced input dataset. Data for this study are imbalanced card transactions from an Indonesia bank in period 2016-2017 with binary label: (nonfraud or fraud). From 50 features of the dataset, 30 principal components of data contribute to 87 % of the cumulative Eigenvalues. This study explores the effect of nonfraud to fraud sample ratio from 1 to 4 and three models: Convolutional Neural Network (CNN), Stacked Long Short-term Memory (SLSTM), and Hybrid of CNN-LSTM. Using Area Under the ROC Curve (AUC) as model performance, CNN achieved the highest AUC for R=1,2,3,4 followed by SLSTM and CNN-LSTM.**

***Keywords*—fraudulent recognition, imbalanced data classification, CNN, LSTM, CNN-LSTM.**

I. INTRODUCTION

The issue of fraudulent card financial transaction, such as credit card and debit card transaction, continuous to gain wide attention from various research communities due its negative impact to financial lost and reputation damage to the banks as card issuers. Despite many technologies have been proposed to prevent and detect fraudulent transaction using cards, fraudulent transactions are still prevalent due to its popularity and ubiquitous of financial transaction facilities.

Fraudulent transaction recognition problem is an interesting but challenging computer vision problem due to its imbalanced data in nature. In the past ten years, many studies to address fraudulent transaction recognition have been reported resulted in a plethora of methods to develop a robust classifier that maximize classification accuracy and minimize two aspects: (1) false positive which disappoint customers or merchants or banks, and (2) false negative which ruined prudent image of bank can be minimized.

In general, fraudulent transaction recognition methods can be divided broadly into two categories [1]. *First*, supervised approach: models are trained to learn patterns from a given labeled samples of fraudulent and non-fraudulent transaction. *Second*, unsupervised approach: models are trained to detect unusual or anomalous transactions from training dataset. The premise of this approach is that anomalous transaction might be potential cases of fraudulent transactions [2]. Given the trained models, fraudulent transaction recognition aims to predict probability of fraudulent transaction labels.

The successful result of using machine learning to solve classification problems in many domains have concerned many researchers to use these models to recognize fraudulent transactions [3]. Many methods have been proposed such as: Random forest [4] [5], Decision Tree [6], K-Nearest Neighbors [7], SVM [8] [9], HMM [10], neural networks [11], Bayesian learning [11], artificial immune systems [12], association rules [13], hybrid models [14], discriminant analysis [15]. Most of these researches are typically preceded by preliminary studies to extract features from historical financial transaction data that represent the pattern of buying behavior or financial transaction of the customers. Financial transaction is then represented by a set of features as input to machine learning models.

The challenges of fraudulent transaction recognition are mainly: (1) no feature standard to represent financial transaction; (2) the imbalanced data distribution of fraudulent and non-fraudulent transaction. Hence, the number of fraudulent transactions is less than the number of non-fraudulent ones; (3) less availability of dataset to validate models proposed by previous studies due to banking confidential reason; and (4) less separability of fraudulent and non-fraudulent transactions as fraudulent card users is typically mimic their consumer behavior closed to non-fraudulent ones.

This study; therefore, aims to: (1) propose a robust financial transaction features that can separate fraudulent and non-fraudulent class samples and (2) propose a robust classifier by exploring hybrid CNN-LSTM model. Following [16], CNN in the proposed model is used to capture short-term financial transaction features; whilst, LSTM on top of CNNs is used to capture longer-term temporal financial transaction. The models under study are then tested using debit card transaction from a local Indonesia bank under permission. The main contribution of this paper is mainly showing empirical results that neither CNN nor LSTM is capable to model financial transaction as each of these models only capable to capture either short-term or long-term temporal transaction patterns.

The remaining paper is structured as follows. Section II will describe several related works. Section III will explain the

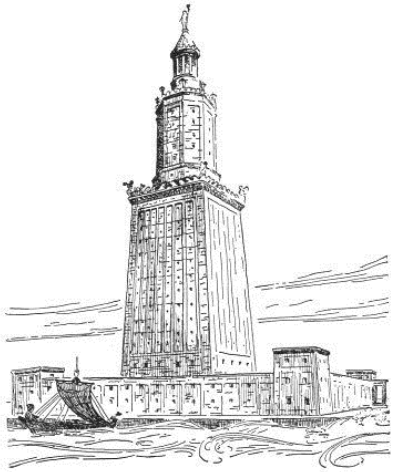
84

Yaya Heryadi
Harco Warnars

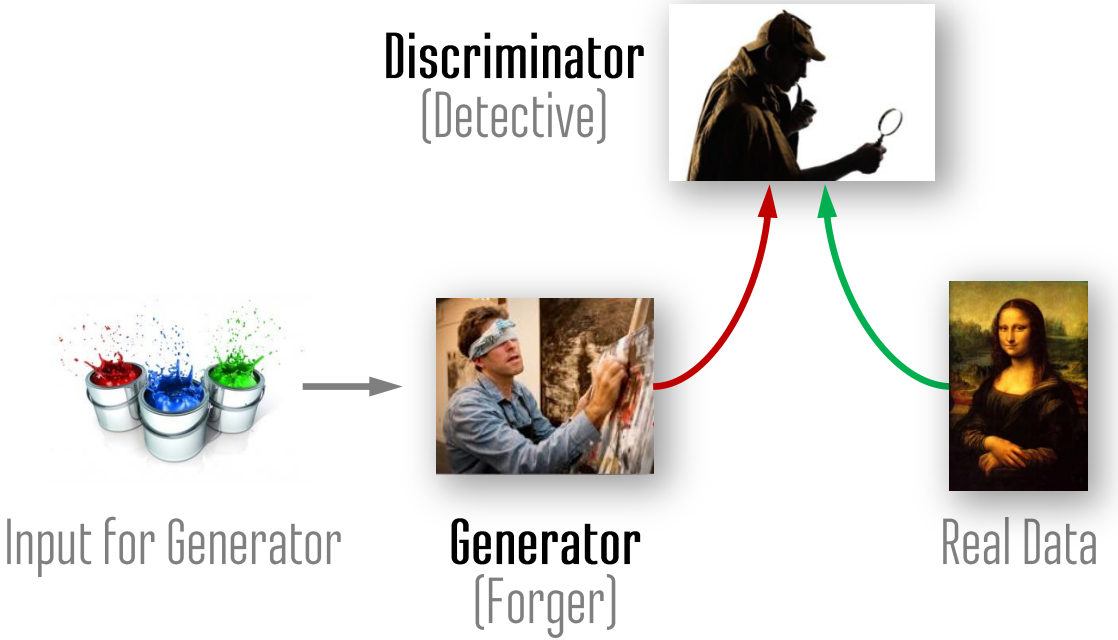
2017

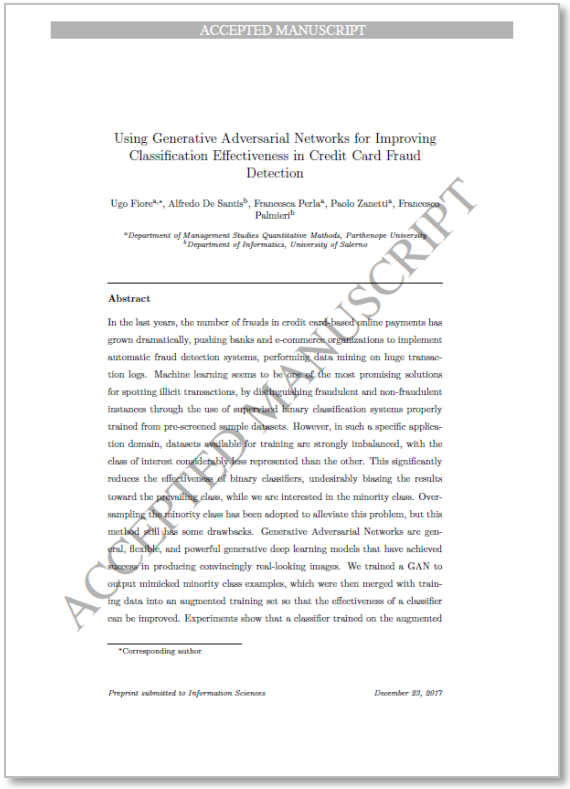
- На одном датасете **сравнили глубокие сети** из разных исследований
- **Провели сравнение** для разных метрик, дисбалансов и длин входных векторов
- Выявили, что **краткосрочные патерны встречаются чаще** долгосрочных (зефирный тест – мошенники не любят долго ждать)

7. GAN'ы для сэмплинга



GAN — Generative Adversarial Network





ACCEPTED MANUSCRIPT

Using Generative Adversarial Networks for Improving Classification Effectiveness in Credit Card Fraud Detection

Ugo Fiore[a,*], Alfredo De Santis[b], Francesco Perla[a], Paolo Zanetti[a], Francesco Palmieri[b]

[a]Department of Management Studies Quantitative Methods, Parthenope University
[b]Department of Informatics, University of Salerno

Abstract

In the last years, the number of frauds in credit card-based online payments has grown dramatically, pushing banks and e-commerce organizations to implement automatic fraud detection systems, performing data mining on huge transaction logs. Machine learning seems to be one of the most promising solutions for spotting illicit transactions, by distinguishing fraudulent and non-fraudulent instances through the use of supervised binary classification systems properly trained from pre-screened sample datasets. However, in such a specific application domain, datasets available for training are strongly imbalanced, with the class of interest considerably less represented than the other. This significantly reduces the effectiveness of binary classifiers, undesirably biasing the results toward the prevailing class, while we are interested in the minority class. Oversampling the minority class has been adopted to alleviate this problem, but this method still has some drawbacks. Generative Adversarial Networks are general, flexible, and powerful generative deep learning models that have achieved success in producing convincingly real-looking images. We trained a GAN to output mimicked minority class examples, which were then merged with training data into an augmented training set so that the effectiveness of a classifier can be improved. Experiments show that a classifier trained on the augmented

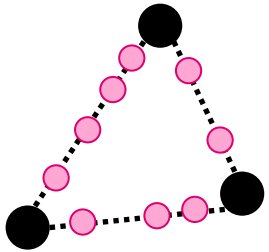
*Corresponding author

Preprint submitted to Information Sciences *December 22, 2017*

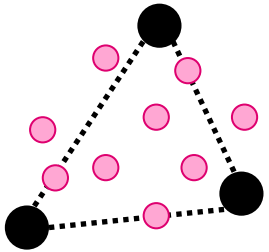
Ugo Fiorea
Alfredo De Santisb & ko

2017

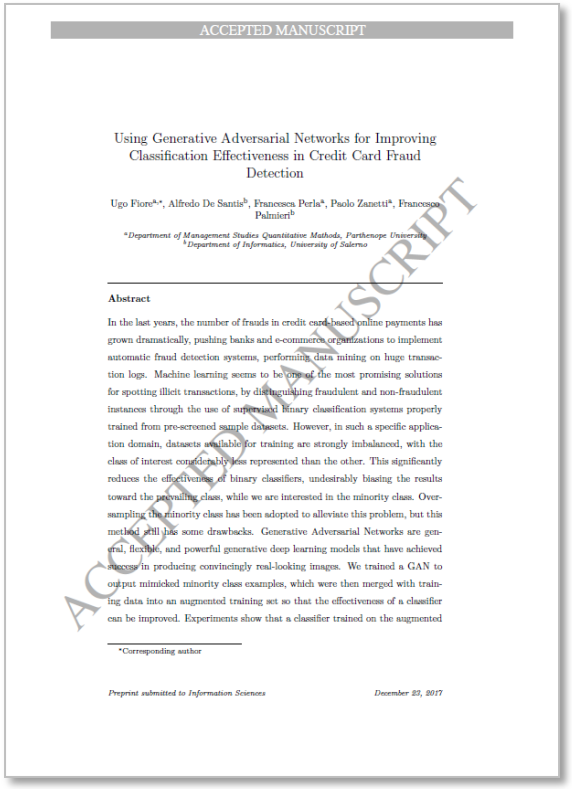
SMOTE



GAN



● Примеры из исходной выборки ● Синтезированные примеры



ACCEPTED MANUSCRIPT

Using Generative Adversarial Networks for Improving Classification Effectiveness in Credit Card Fraud Detection

Ugo Fiore[a,*], Alfredo De Santis[b], Francesca Perla[a], Paolo Zanetti[a], Francesco Palmieri[b]

[a]Department of Management Studies Quantitative Methods, Parthenope University
[b]Department of Informatics, University of Salerno

Abstract

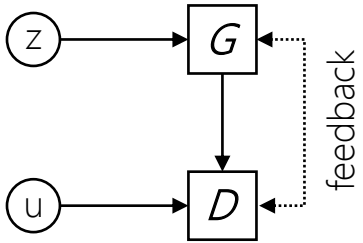
In the last years, the number of frauds in credit card-based online payments has grown dramatically, pushing banks and e-commerce organizations to implement automatic fraud detection systems, performing data mining on huge transaction logs. Machine learning seems to be one of the most promising solutions for spotting illicit transactions, by distinguishing fraudulent and non-fraudulent instances through the use of supervised binary classification systems properly trained from pre-screened sample datasets. However, in such a specific application domain, datasets available for training are strongly imbalanced, with the class of interest considerably less represented than the other. This significantly reduces the effectiveness of binary classifiers, undesirably biasing the results toward the prevailing class, while we are interested in the minority class. Oversampling the minority class has been adopted to alleviate this problem, but this method still has some drawbacks. Generative Adversarial Networks are general, flexible, and powerful generative deep learning models that have achieved success in producing convincingly real-looking images. We trained a GAN to output mimicked minority class examples, which were then merged with training data into an augmented training set so that the effectiveness of a classifier can be improved. Experiments show that a classifier trained on the augmented

*Corresponding author

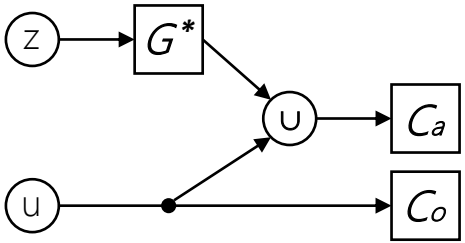
Ugo Fiorea
Alfredo De Santisb & ko

2017

GAN



GAN (fraud)



ACCEPTED MANUSCRIPT

Using Generative Adversarial Networks for Improving Classification Effectiveness in Credit Card Fraud Detection

Ugo Fiore[a,*], Alfredo De Santis[b], Francesca Perla[a], Paolo Zanetti[a], Francesco Palmieri[b]

[a]Department of Management Studies Quantitative Methods, Parthenope University
[b]Department of Informatics, University of Salerno

Abstract

In the last years, the number of frauds in credit card-based online payments has grown dramatically, pushing banks and e-commerce organizations to implement automatic fraud detection systems, performing data mining on huge transaction logs. Machine learning seems to be one of the most promising solutions for spotting illicit transactions, by distinguishing fraudulent and non-fraudulent instances through the use of supervised binary classification systems properly trained from pre-screened sample datasets. However, in such a specific application domain, datasets available for training are strongly imbalanced, with the class of interest considerably less represented than the other. This significantly reduces the effectiveness of binary classifiers, undesirably biasing the results toward the prevailing class, while we are interested in the minority class. Oversampling the minority class has been adopted to alleviate this problem, but this method still has some drawbacks. Generative Adversarial Networks are general, flexible, and powerful generative deep learning models that have achieved success in producing convincingly real-looking images. We trained a GAN to output mimicked minority class examples, which were then merged with training data into an augmented training set so that the effectiveness of a classifier can be improved. Experiments show that a classifier trained on the augmented

*Corresponding author

Ugo Fiorea
Alfredo De Santisb & ko

2017

Выборка	446 fraud, 283 726 good
Метрики	Sn (Rec), Sp, Prec, F1, Acc
Алгоритмы	GAN vs. SMOTE

ACCEPTED MANUSCRIPT

Using Generative Adversarial Networks for Improving Classification Effectiveness in Credit Card Fraud Detection

Ugo Fiore[a,*], Alfredo De Santis[b], Francesco Perla[a], Paolo Zanetti[a], Francesco Palmieri[b]

[a]Department of Management Studies Quantitative Methods, Parthenope University
[b]Department of Informatics, University of Salerno

Abstract

In the last years, the number of frauds in credit card-based online payments has grown dramatically, pushing banks and e-commerce organizations to implement automatic fraud detection systems, performing data mining on huge transaction logs. Machine learning seems to be one of the most promising solutions for spotting illicit transactions, by distinguishing fraudulent and non-fraudulent instances through the use of supervised binary classification systems properly trained from pre-screened sample datasets. However, in such a specific application domain, datasets available for training are strongly imbalanced, with the class of interest considerably less represented than the other. This significantly reduces the effectiveness of binary classifiers, undesirably biasing the results toward the prevailing class, while we are interested in the minority class. Oversampling the minority class has been adopted to alleviate this problem, but this method still has some drawbacks. Generative Adversarial Networks are general, flexible, and powerful generative deep learning models that have achieved success in producing convincingly real-looking images. We trained a GAN to output mimicked minority class examples, which were then merged with training data into an augmented training set so that the effectiveness of a classifier can be improved. Experiments show that a classifier trained on the augmented

*Corresponding author

Preprint submitted to Information Sciences *December 22, 2017*

Ugo Fiorea
Alfredo De Santisb & ko

2017

Ng	Sensitivity		Specificity		Precision		F-measure		Accuracy	
	GAN	SMOTE	GAN	SMOTE	GAN	SMOTE	GAN	SMOTE	GAN	SMOTE
0	.70229	.70229	**.99998**	**.99998**	**.97872**	**.97872**	.81778	.81778	**.99964**	**.99964**
79	.70229	.71247	.99997	.99997	.96842	.96552	.81416	**.81991**	.99963	.99964
158	.71756	.70229	.99994	.99998	.93069	.97872	.81034	.81778	.99961	.99964
315	.72519	**.72519**	.99992	.99994	.91346	.93137	.80851	.81545	.99960	.99962
630	**.73282**	.69466	.99994	.99997	.93204	.96809	**.82051**	.80889	.99963	.99962
945	.72519	.70229	.99994	.99996	.93137	.95833	.81545	.81057	.99962	.99962
1 260	.72519	.70229	.99994	.99998	.93137	.97872	.81545	.81778	.99962	.99964
2 520	.72519	.70229	.99994	.99998	.93137	.97872	.81545	.81778	.99962	.99964
3 150	.73028	.71756	.99994	.99996	.93182	.94949	.81883	.81739	.99963	.99963
6 300	.72519	.70229	.99995	.99998	.94059	.97872	.81897	.81778	.99963	.99964
31 500	.70229	.70229	.99996	.99998	.95833	.97872	.81057	.81778	.99962	.99964

ACCEPTED MANUSCRIPT

Using Generative Adversarial Networks for Improving Classification Effectiveness in Credit Card Fraud Detection

Ugo Fiore[a,*], Alfredo De Santis[b], Francesca Perla[a], Paolo Zanetti[a], Francesco Palmieri[b]

[a]Department of Management Studies Quantitative Methods, Parthenope University
[b]Department of Informatics, University of Salerno

Abstract

In the last years, the number of frauds in credit card-based online payments has grown dramatically, pushing banks and e-commerce organizations to implement automatic fraud detection systems, performing data mining on huge transaction logs. Machine learning seems to be one of the most promising solutions for spotting illicit transactions, by distinguishing fraudulent and non-fraudulent instances through the use of supervised binary classification systems properly trained from pre-screened sample datasets. However, in such a specific application domain, datasets available for training are strongly imbalanced, with the class of interest considerably less represented than the other. This significantly reduces the effectiveness of binary classifiers, undesirably biasing the results toward the prevailing class, while we are interested in the minority class. Oversampling the minority class has been adopted to alleviate this problem, but this method still has some drawbacks. Generative Adversarial Networks are general, flexible, and powerful generative deep learning models that have achieved success in producing convincingly real-looking images. We trained a GAN to output mimicked minority class examples, which were then merged with training data into an augmented training set so that the effectiveness of a classifier can be improved. Experiments show that a classifier trained on the augmented

*Corresponding author

Ugo Fiorea
Alfredo De Santisb & ko

2017

- Обучение GAN'ов **не является легкой задачей** (низкая стабильность)
- Для сравнения качества использовались только **пороговые метрики**
- **Сомнительные результаты**: 1) слабая балансировка; 2) по некоторым метрикам сэмплинг ухудшает качество

ACCEPTED MANUSCRIPT

Using Generative Adversarial Networks for Improving Classification Effectiveness in Credit Card Fraud Detection

Ugo Fiore[a,*], Alfredo De Santis[b], Francesca Perla[a], Paolo Zanetti[a], Francesco Palmieri[b]

[a]Department of Management Studies Quantitative Methods, Parthenope University
[b]Department of Informatics, University of Salerno

Abstract

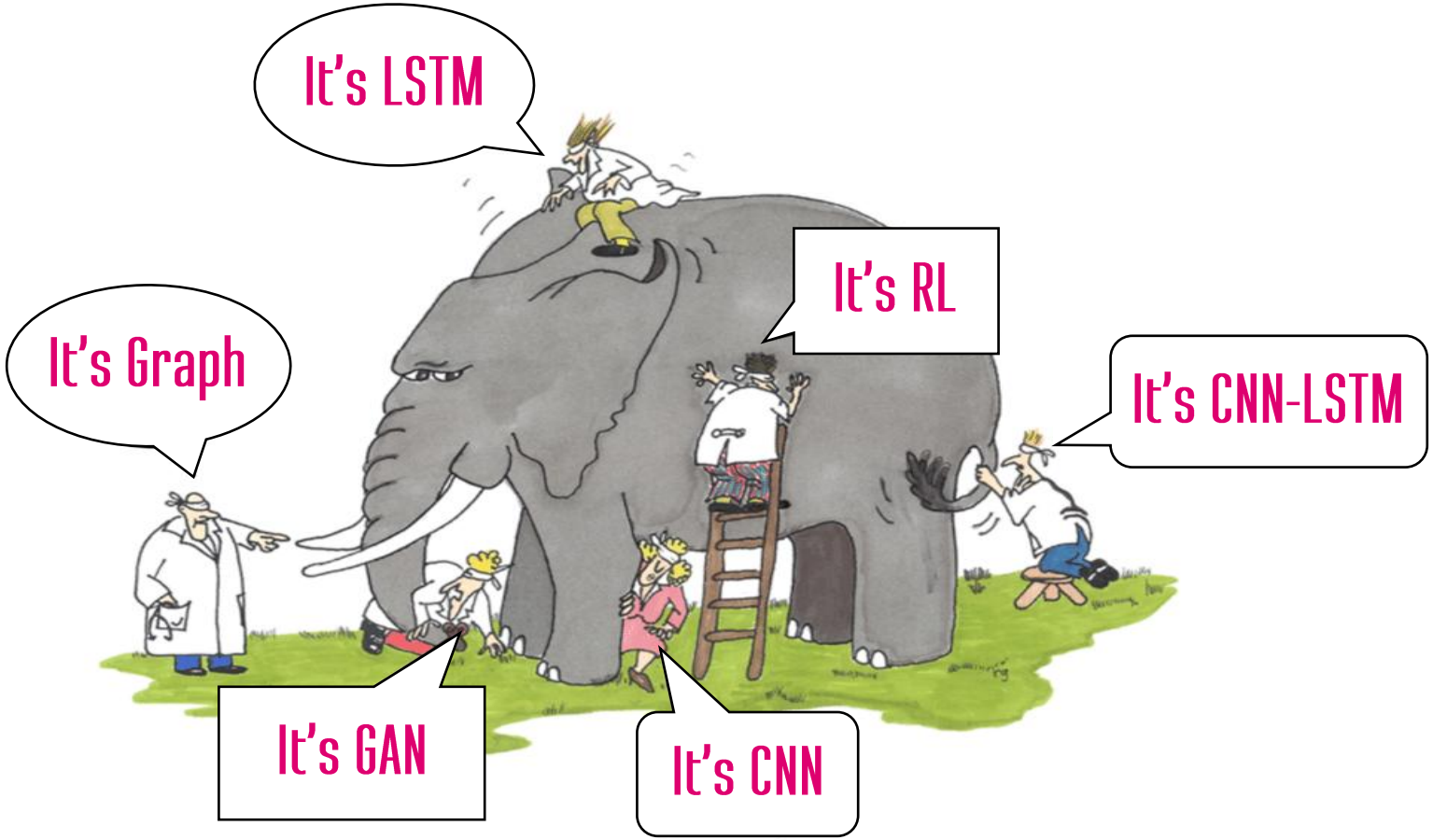
In the last years, the number of frauds in credit card-based online payments has grown dramatically, pushing banks and e-commerce organizations to implement automatic fraud detection systems, performing data mining on huge transaction logs. Machine learning seems to be one of the most promising solutions for spotting illicit transactions, by distinguishing fraudulent and non-fraudulent instances through the use of supervised binary classification systems properly trained from pre-screened sample datasets. However, in such a specific application domain, datasets available for training are strongly imbalanced, with the class of interest considerably less represented than the other. This significantly reduces the effectiveness of binary classifiers, undesirably biasing the results toward the prevailing class, while we are interested in the minority class. Oversampling the minority class has been adopted to alleviate this problem, but this method still has some drawbacks. Generative Adversarial Networks are general, flexible, and powerful generative deep learning models that have achieved success in producing convincingly real-looking images. We trained a GAN to output mimicked minority class examples, which were then merged with training data into an augmented training set so that the effectiveness of a classifier can be improved. Experiments show that a classifier trained on the augmented

*Corresponding author

Ugo Fiorea
Alfredo De Santisb & ko

2017

- Первые опубликовали статью о применении GAN'ов для сэмплинга фрод-выборки
- Качественно подобранная архитектура (настраивали много гиперпараметров)
- На миноритарном классе GAN быстро обучается и не тормозит общий процесс


It's LSTM
It's Graph
It's RL
It's CNN-LSTM
It's GAN
It's CNN

Спасибо за внимание!

Афанасьев Сергей

Исполнительный директор
Начальник управления статистического анализа

КБ «Ренессанс Кредит»

safanasev@rencredit.ru