International Journal of Hybrid Information Technology
Vol.8, No. 5 (2015), pp. 41-50
http://dx.doi.org/10.14257/ijhit.2015.8.5.05

Predictor Variables' Influence on Classification Outcome in Insurance Fraud Detection

Saliu Adam Muhammad, Xiangtao Chen and Liao Bo

Department of Computer Science and Technology,
School of Information Science and Electronic Engineering,
Hunan University, Changsha, Hunan Province,
410082, P.R. China
madasbinlaf@gmail.com, boliao@yeah.net, and &xtchen2009@qq.com

Abstract

In fraud detection paradigms, the role of predictor variables cannot be overemphasized particularly when analytical tools – such as statistical, machine learning and artificial intelligent tools are employed. These variables or attributes are used to organize records of data in database tables. The combination of the values of these attributes usually affects which class of a target variable a record or an observation would belong. In this paper, we propose an algorithm and employ spreadsheet 'count' 'count if' and 'filtering' functionalities (techniques) to take toll on how the individual attribute may affect the prediction of the class of an observation in an insurance dataset of 5000 observations. The analysis showed that indeed, the individual predictor attribute affects the outcome of the target variable (legal or fraudulent) differently.

Keywords: *predictor variable, insurance, fraud detection, target variable, observation /instance, premium*

1. Introduction

A popular saying goes thus: "Prevention is better than a cure." Paraphrasing this saying into the realm of fraud scams, it is better to prevent a fraud from occurring than to detect or discover and cure it. But prevention is almost impossible, at least as at now, and so detection becomes highly necessary to complement the effort of preventive mechanisms (efforts put in place to stop frauds from occurring). In fraud detection paradigms, the role of predictor variables cannot be overemphasized particularly when analytical tools – such as statistical, machine learning and artificial intelligent tools are employed. These variables or attributes are used to organize records of data in database tables. The combination of the values of these attributes usually affects which class of a target variable a record or an observation would belong.

Globally, financial fraud is a known problem, ranging from insurance fraud, credit card fraud, telecommunications fraud, and check forgery [1-2]. Fraud is defined as a wrongful or criminal ploy for financial or personal gains [3]. Various bodies: security outfits, regulatory agencies, anti-fraud bodies, governments and so on, have reported cases of high profiles of financial fraud. Falsified techniques that appear to be true in the face of the victims are usually used to perpetrate these frauds [2].

Commissions of insurance frauds have been notoriously noted. Cases of a rodent in a bowl of soup, a tricked-out street racer at the bottom of a lake, a foreclosure property mysteriously set ablaze and a woman who died twice have been reported. These are all part of the outrageous world of insurance scams, in which cash-strapped policyholders, phony or unscrupulous insurance agents, desperate business owners and sundry con artists conspire to defraud insurers by filing inflated or outright bogus claims. James Quiggle, a spokesman for the nonprofit coalition of insurers and consumer groups, says, "The

ISSN: 1738-9968 IJHIT

Coalition against Insurance Fraud", conservatively estimates that insurance fraud costs $80 billion a year in stolen claims, not including the social costs [4].

Happily, data mining and statistical methods have been proved to successfully detect fraudulent activities such as money laundering, e-commerce credit card fraud, telecommunications fraud, insurance fraud, and computer intrusion [5]. Researchers have employed a number of these techniques to detect these fraudulent activities where prevention measures failed. They employed such data mining techniques as Decision Trees, Bayesian Network, Rule Base Network, Support Vector Machines, Neural Network, and so on, in their various attempts to develop models that would help in detecting these frauds in credit card system, telecommunication systems, banking systems, insurance systems and so on.

Apparao, *et. al,* in their analysis of financial fraud statement, discovered that out of the twenty-six (26) data mining based techniques for financial fraud detection, all the 26 (100%) techniques are found to be classification techniques [5]. Ngai *et al.*, reviewed financial fraud detection on the basis of classification framework. They discovered that classification models are mostly used on insurance fraud detection [6].

Clearly, classification techniques of data mining algorithms are seen to be most engaged by researchers in their fraud detection efforts. These classification (in the case categorical data) or prediction (in the case of numerical data) methods all make use of predictor variables in database to do their classifications or predictions. This development underscores the importance of predictor variables to classification, and hence, our interest in seeking how these variables could individually influence classification of insurance fraud data analysis.To this end, we engaged the procedure of 'filtering', 'count' and 'countif' functionalities of spreadsheet. We sort the instances individually based on *sex, previous claim, ticket,* and *attorney* attributes as they independently affect the outcome or target (whose value is legal or fraudulent) attribute. To the best of our knowledge, we have not come across a literature that would individually examine the influence of each predictor attribute in a dataset as we have done in this work. We used an insurance dataset of 5000 instances, discovered that the individual predictor attribute affects the outcome of the target variable (legal or fraudulent) differently.

The organisation of the rest of this paper is as follow: Some insurance scams and few previous works on insurance fraud classifications are reviewed in Section 2. In Section 3 we present the methodology used in this research work, while in Section 4, the results are showcased. We presented conclusion and future direction in Section 5.

2. Insurance Fraud Review

Insurance can be stated as a treaty (policy) in which an individual or entity receives financial protection or reimbursement against losses from an insurance company [2, 3]. It is widespread in automobile, travel and telecommunication industries, and money laundery. Insurance is a way of managing risks. When you buy insurance, you transfer the cost of a potential loss to the insurance company in exchange for a fee, known as the premium [7]. It is an agreement where, for a stipulated payment called the premium, one party (the insurer) agrees to pay to the other (the policyholder or his designated beneficiary) a defined amount (the claim payment or benefit) upon the occurrence of a specific loss [8]. The purpose of insurance is to protect against the event of a financial loss [9]. For example:

(a) Auto insurance could pay the cost of repairs to your vehicle if you have an accident.
(b) Crop insurance could provide finance against loss of crops.
(c) Health insurance could pay for the cost of health problems

But, despite the good intent of this system, some fraudsters (at various levels and categories) are seizing the chance to defraud the insurers. This is why researchers/scientist

and software practioners are up and doing in developing methods and software products to help combat these fraudsters.

Scamming an insurance company can be done in various ways, from phony slip-and-falls to faked deaths. Ultimately, honest consumers and businesses pay for it. From suspicious tainted-food claims to phony slip-and-falls, faked deaths to real murder, the following are a review of some of the most desperate and devious insurance frauds from the coalition's Insurance Fraud Hall of Shame [4]:

1. Burning down the house - for profit

Marc Thompson, the executive of Chicago grain futures was deeply indebted out of high living. In desperation, he torched his home for the $730,000 in insurance money. Now Thompson's own future is secure - for 190 years in federal prison. "People in small but increasing numbers were burning down their homes," says Quiggle.

2. Fake slip-and-falls gain traction

Parker, the septuagenarian queen of the slip-and-fall scam, prostrated herself in department stores, supermarkets and liquor stores dozens of times for claims totaling at least $500,000 during her long career, a sad by-product of her gambling addiction.

Others include are vehicle give-ups: total loss, total fraud, fake insurance, fraudulent repairs Waiter, there's a mouse in my meal, and life insurers anticipate phony deaths, and 'Pill mills' and insurance fraud.

Many more of these instances are polarizing the field of insurance company causing insurers to pay illegal amounts to fraudsters. Generally, detecting frauds is not an easy task as the criminals themselves are knowledgeable and smart using such tools as Internet facilities. Fanning and Cogger [10] argued that detecting management fraud is difficult because: firstly, there is a shortage of knowledge concerning the characteristics of managing frauds; secondly, most auditors lack the experience necessary to detect such frauds; finally, financial managers and accountants are deliberately trying to deceive the auditors. For such managers, with limited knowledge of an audit system, standard auditing procedures may be insufficient. These limitations suggest the need for additional analytical procedures for the effective detection of false financial statements [11].

Insurance frauds have been tackled by various efforts. For instance, Wei *et al.*, [17] employed rough set reduction to generate a set of reductions from which randomly subsets of reductions were constructed. They then used these subsets to produce neural network classifiers based on the insurance data, which are subsequently combined using ensemble strategies. The investigational results show that the proposed model outperforms each of the single classifiers and other models used in the evaluation. Rekha [1], employed classification tasks of decision tree and Bayesian Network techniques of data mining to detect frauds in an auto insurance company. The performance of the model was tested on classification metrics of accuracy, recall, and precision derived from the confusion matrix, and claimed that the model is strong with respect to class skew, making it a reliable performance metric in many important fraud detection application areas. GA-Kmeans (combination of K-means algorithm with genetic algorithm (GA)) and The MPSO-Kmeans (combination of K-means algorithm with Momentum-type Particle Swarm Optimization (MPSO)) were used in the prediction of insurance fraud data [12]. The proposed GA-Kmeans and MPSO-Kmeans were employed to determine the optimal weights and final cluster centers for attributes. The accuracy of prediction for test set is then computed on the basis of the optimal weights and the final cluster centers. The results of the two algorithms showed significant improvement over those of the pure K-means algorithm. Ngai *et al.*, [18], reviewed financial fraud detection on the basis of classification framework. They discovered that classification models including neural network, are mostly used on insurance fraud detection.

All these approaches are based on predictor attributes which are used to organize data in databases, with the values of these attributes being employed in classification or prediction of fraudulent activities.

3. Materials and Methods

3.1. Data Set

An insurance dataset [2, 12] of 5000 observations or instances is used in this work. Tables 1 and 2 respectively and partially represent the real and the normalized data forms of the database or dataset. The data set is made up of six predictor attributes and one target attribute with two values (fraud and agree/legal). After normalization [13], legal is represented by zero, while fraud is represented by one.

Table 1. A Real Sample Dataset for Insurance Fraud

Age	Gender	Claim	tickets	prior claims	atty	outcome
59	0	1700	0	0	Atty	Agree
35	0	2400	0	0	Atty	Agree
39	1	1700	0	0	Atty	Agree
18	1	3000	0	0	Atty	Agree
24	1	1600	0	0	Atty	Agree

Table 2. A Normalized Sample Dataset for Insurance Fraud

Age	Gender	Claim	Tickets	Claims	Atty	outcome
1	0	0.66	1	1	0	0
0.75	0	0.52	1	1	0	0
0.95	1	0.66	1	1	0	0
0	1	0.4	1	1	0	0
0.2	1	0.68	1	1	0	0

3.2. The Spreadsheet Function

A Spreadsheet program provides structure for data representation as cells of an array, organized in rows and columns. This makes spreadsheets particularly suitable for organizing data into tables of databases. Besides performing basic arithmetic and mathematical functions, modern spreadsheets provide built-in functions for common financial and statistical operations and analyses. Such calculations as net present value count or standard deviation can be applied to tabular data with a pre-programmed function in a formula. Spreadsheet programs also provide conditional expressions, functions to convert between text and numbers, and functions that operate on string of text.

Here, we employ some Spreadsheet functions of Microsoft Excel 2010 to perform some analyses in our proposed work.

3.2.1. Filter Function

This describes a function in Data Analysis Expressions (DAX), a formula expression language used to define calculations in PowerPivot in Excel [14]. Returns a table that represents a subset of another table or expression. The syntax is FILTER (<table>, <filter>), where table, which is the table (or an expression that results in a table) to be

filtered, and filter, which is a Boolean expression that is to be evaluated for each row of the table.

3.2.2. Count Function

The COUNT function counts the number of cells that contain numbers, and counts numbers within the list of arguments [15]. The syntax of COUNT function is:COUNT (value1, [value2] ...), where *Value1* is the first item, cell reference, or range to count, and value2, ... is additional items, cell references, or ranges to count

3.2.3. Count if Function

The COUNTIF function counts the number of cells within a range that meet a single criterion that is specified. The syntax of COUNTIF function is [16]:COUNTIF (range, criteria), where *range* is one or more cells to count, including numbers or names, arrays, or references that contain numbers, and *criteria* is a number, expression, cell reference, or text string that defines which cells will be counted.

3.3. The Proposed Method

In this paper, we proposed and applied the algorithm presented in Algorithm 1. We analyzed some of the predictor attributes as they affect the classification of whether an observation is legal of fraudulent. In the scenario, the worth of ticket, previous claims, attorney and sex predicting attributes are investigated as they affect the determination of whether evidence is legal or fraudulent.

Algorithm 1: Determining the Influence of Individual Predictor on the Outcome of Classification

1. Start
2. No. of Observations = 5000
3. Find the No. of Legals
4. Find the No. of Fraudulents
5. Determine the ratio of Legals to Observations
6. Determine the ratio of Fraudulents to Observations
7. Choose a Predictor
   - 7.1 Filter by each value
   - 7.2 Count the No of each value
   - 7.3 Find the ratio of each value to its corresponding Legals and Fraudulents
8. If more Predictors then
   - 8.1 Repeat 7 else
   - 8.2 Goto 8
9. Stop

From Algorithm 1, our algorithm starts (Step 1) and then proceeds to determine the total number of observations in the dataset (Step 2). The algorithm then finds the totals of legal and fraudulent observations in the dataset (Steps 3 and 4). Thereafter, the ratios of legal and fraudulent instances are determined in relation to the total number of instances in the database (Steps 5 and 6). A predictor is then selected. The number of each value of the selected predictor is then counted with the ratios determined in relation to the legal and fraudulent instances (Step 7). At Step 8, a request is made as whether there are more

predictors of interest. If there more predictors for consideration, Step 7 is repeated, otherwise the algorithm stops at Step 9.

4. Result and Analysis

In this work, we applied three functions of FILTER, COUNT and COUNTIF of Excel spreadsheet (Section 3.2) to implement our algorithm (3.3), to analyze some of the predictor attributes of a dataset of 5000 observations, as they individually affect the outcomes (*legal* or fraudulent) of the target attribute. In the scenario, the worth of *ticket, previous claims, attorney* and *sex* predictor attributes are investigated.

After applying the method to the insurance dataset, we were able to find that the dataset has as many as a total of 4918 cases of legal claims representing 98.4% of the total population of the dataset. Only 82 (1.6%) cases of the entire population are fraudulent claims. We say 82 cases are only quantitatively, not in terms of impact. No amount of fraud cases is small in terms of negative impact it brings on an individual, organization or economy.

The results of the analyses of the considered predictor variables (*Previous Claims (Claims), Attorney (Atty), Tickets* and *Sex* (Gender) predictors), are presented as follow:

4.1. Ticket Predictor Attribute

Table 3. Tickets Predictor's Influence on the Outcome of Classification

	Legal	Fraud	
Policyholders with Tickets	499	4	**503**
Policyholders without Tickets	4419	78	**4497**
	4918	**82**	

In the dataset, 503 policyholders were previously issued with tickets. This constitutes 10.1 % of the entire population of the dataset. The remaining 89.9 % (4497 policyholders) observations do not have previous tickets.

The analysis shows that only 0.8% (four observations) of the policyholders with tickets is fraudulent, which implies that 99.2% (499 observations) claims of the policyholders with tickets are legal. Policyholders with tickets constitute 1.7% (78 observations) legal claims, while 98.3% (4419 observations) policyholders without tickets constitute legal claims.

Only 4.9 % (four observations) of fraudulent claims are policyholders with tickets, leaving policyholders without tickets with the remaining 95.1% (78 observations) of the frauds. Of legal claims, 10.1% (499 observations) are policyholders with tickets, while the remaining 89.9% (4419 observations) are policyholders without ticket.

The analysis here shows that, although policyholders with tickets are expected to make more claims, owing to their previous records and payment of higher premiums (because insurers consider them to be riskier), their claims may not necessarily be fraudulent. That is, ticket-carrying policyholders might not necessarily be riskier customers of the insurers.

4.2. Sex Predictor Attribute

Table 4. Sex Predictor's Influence on the Outcome of Classification

	Legal	Fraud	
Female	2472	40	**2512**
Male	2446	42	**2488**

The result of the analysis shows that 2512 females are policyholders of this dataset. This constitutes 51.2 % of the entire population of the dataset while 48.8 % (2488) are men. This shows that, (even though, more men are involved in driving in general) more women participate in insurance policy than men – a clear indication of more risk-mitigation tendency on the part of females than men.

Of 2512 females, only 1.6% (40) is fraudulent, while the remaining 98.4% (2472) is legal. About 1.7% (42) of policyholders who are males are fraudulent, while 98.3% (2446) are legal.

48.8 % (40) of fraudulent claims are females while 51.2 % (42) are men. Females constitute 50.3% (2472) of legal claims while the male counterparts constitute 49.7% (2446).

We can deduce here that, even though it is marginal, men are still more involved in sharp practices than their women equals.

4.3. Sex Previous Claims Attribute

Table 5. Previous Claims Predictor's Influence on the Outcome of Classification

	Legal	Fraud	
Policyholders with Previous Claims	638	25	**663**
Policyholders without Previous Claims	4280	57	**4337**
	4918	**82**	

379 of 5000 observations have 1 previous claim each, 284 have a minimum of two previous claims, which gives a total of 663 policyholders with previous claims cases. The remaining 4337 observations or policyholders are without previous claims (See Table 5).

3.8% (25) of the 663 policyholders with previous claims have fraudulent claims, while the left 96.2% (638) have legal claims. The policyholders without previous claims have 1.3% (57) and 98.7% (4280) of their claims to be fraudulent and legal respectively.

About 30.5% (25) of fraudulent claims are perpetrated by policyholders with previous claims; the remaining 69.5% (57) of frauds come from policyholders without previous claims. The analysis also indicates that a total of 87.0% of legal claims are of policyholders without previous claims, while the ones with previous claims only constituted 13.0% of legal claims.

4.4. Attorney Predictor Attribute

Table 6. Attorney Predictor's Influence on the Outcome of Classification

	Legal	Fraud	
Accompanied by Attorney	4851	64	**4915**
Not Accompanied by Attorney	67	18	**85**
	4918	**82**	

A total of 4915 (98.3%) policyholders is all accompanied by attorneys. Only 85 (1.7%) policyholders in the entire dataset are not accompanied by an attorney.

98.7% (4851) of the policyholders accompanied by attorney are legal while the remaining 1.3% (64) is fraudulent. 67 (78.8%) and 18 (21.2%) policyholders of the total of 85 who are not accompanied by attorney are legal and fraudulent respectively.

64 (78.0% of the total frauds) policyholders accompanied by attorneys are fraudulent in claims, while 18 (22.0% of the total frauds) were perpetrated by policyholders not accompanied by attorney. Attorney-accompanied policyholders' claims 4851 (98.6% of the total legal claims) are legal, while the remaining 67 legal claims (1.4% of the legal claims) are of policyholders not accompanied by attorney.

We can observe that in terms of fraudulent cases, those accompanied by attorneys are more, possibly to help them fight the legal battles that might come up when attempting to make their fraudulent claims.

5. Conclusion and Recommendation

In this paper, we show that the influence of the individual attributes on the target attribute of the dataset is different. The different number of occurrences of the attributes in relation to the output of legal or fraud is a proof of this. The percentages of the individual predictor attributes show the degree of discrepancies among the predictor attributes as they affect the out of the come target attribute.

The result of our analysis show that, although, policyholders with tickets are known to pay higher premiums because the insurance companies see them to be of higher risks, their claims may not necessarily be fraudulent ones. We also deduce from our analysis that females are less fraudulent and more risk-conscience than their male counterparts. Most of the fraudulent claims are accompanied by attorneys, since most of such claims may involve intense legal battle. Our result also shows that policyholders with previous claims are not an indication of more claims from the insurers.

We might, therefore, say that individual predictor has its implication different from another predictor of the same dataset, which in a way influence how an observation may be classified. Future research work could be an analysis of interdependence among the predictor attributes as they jointly affect the outcome of classification or prediction. We may conclusively state that even though interdependence exists among the predictor attributes of insurance dataset, independent analysis of each of the attributes still hold a paramount place in the anal of insurance fraud detection analysis, as more facts are deducible from such examination.

References

[1] B. Rekha, "Detecting Auto Insurance Fraud by Data Mining Techniques", Journal of Emerging Trends in Computing and Information Sciences, vol. 2, no. 4, **(2011)**, pp. 156 – 162.
[2] A. M. Saliu, "Fraud: The Affinity of Classification Techniques to Insurance Fraud Detection", International Journal of Innovative Technology and Exploring Engineering (IJITEE), vol. 3, no. 11, **(2014)**, pp. 62 – 66.
[3] H. S. Lookman and T. Balasubramanian, "Survey of Insurance Fraud Detection Using Data Mining Techniques", International Journal of Innovative Technology and Exploring Engineering (IJITEE), vol. 2, no. 3, **(2013)**, pp. 62 – 65.
[4] Extreme Cases of Insurance Fraud (http://money.msn.com/insurance/8-extreme-cases-of-insurance-fraud-bankrate.aspx?cp-documentid=6833441).
[5] G. Apparao, S. Arun, G. S. Rao, B. L. Bhavani, K. Eswar and D. Rajani, "Financial Statement Fraud Detection by Data Mining", Int. Journal of Advanced Networking and Applications (IJANA), vol. 1, no. 3, **(2009)**, pp. 159 - 163 .
[6] E. W. T. Ngai, H. Yong, Y. H. Wong, C. Yijun and S. Xin, "The Application of Data Mining Techniques in Financial Fraud Detection", A Classification Framework and an Academic Review of Literature, Decision Support Systems, Elsevier, vol. 50, **(2011)**, pp. 559 - 569.
[7] http://www.cooperators.ca/en/Answer-Centre/how-does-insurance-work/why-do-we-need-insurance.aspx
[8] F. A. Judy and L. B. Robert, (FSA), "Education and Examination Committee of the Society of Actuaries Risk and Insurance", Copyright 2005 by the Society of Actuaries, P-21-05 Printed in U.S.A. Second Printing.
[9] FCAC, Financial Consumer Agency of Canada, Understanding Insurance Basics, February, **(2011)**.

[10] K. Fanning and K. Cogger, "Neural Network Detection of Management Fraud Using Published Financial Data", International Journal of Intelligent Systems in Accounting, Finance & Management, vol. 7, no. 1, **(1998)**, pp. 21 – 24.
[11] E. Turban, J. E. Aronson, T. P. Liang and R. Sharda, "Decision Support and Business Intelligence Systems", 8thed, Pearson Education, **(2007)**.
[12] L. Jenn-Long, C. Chien-Liang, and Y. Hsing-Hui, "Efficient Evolutionary Data Mining Algorithms Applied to the Insurance Fraud Prediction", International Journal of Machine Learning and Computing, vol. 2, no. 3, **(2012)**, pp. 17 - 22.
[13] D. Olson, and Y. Shi, "Introduction to Business Data Mining", McGraw-Hill Education, (2008).
[14] FILTER Function (DAX), https://support.office.com/en-us/article/FILTER-Function-DAX-17a618de-b295-4a7f-8ffc-440fa4f0479a?ui=en-US&rs=en-US&ad=US#__toc319685033
[15] COUNT Function, http://office.microsoft.com/en-001/excel-help/count-function-HP010342338.aspx
[16] COUNTIF Function, https://support.office.com/en-us/article/COUNTIF-function-4764f197-0127-49fa-9f5a-b188177b6db4?ui=en-US&rs=en-US&ad=US
[17] X. Wei, W. Shengnan, Z. Dailing and Y. Bo, "Random Rough Subspace Based Neural Network Ensemble for Insurance Fraud Detection", Fourth International Joint Conference on Computational Sciences and Optimization, IEEE, Computer Society, pp. 1276 – 1280, Kunming and Lijiang City, China.
[18] E. W. T. Ngai, H. Yong, Y. H. Wong, C. Yijun and S. Xin, "The Application of Data Mining Techniques in Financial Fraud Detection", A Classification Framework and an Academic Review of Literature, Decision Support Systems, Elsevier, vol. 50, **(2011)**, pp. 559 - 569.

Authors

Saliu Adam Muhammad, received B. Tech. Mathematics/ Computer Science from Federal University of Technology, Minna, Niger State- Nigeria, MSc. Computer Science from Abubakar Tafawa Balewa, Bauchi, Bauchi State Nigeria. He is currently a PhD. Student in Computer Science & Technology Department, School of Information Science and Electronic Engineering, Hunan University, Changsha, Hunan Province – PR. China. He was a lecturer in the Department of Mathematics/Computer Science and currently a lecturer in the Department of Computer Science, School of Information & Communication Technology, Federal University of Technology, Minna, Niger State – Nigeria. He has authored and co-authored eleven papers in Journals (National & International). He has also participated in ten Conferences(all national) – with four papers in Book of Proceedings, three presentations and three without presentation.

Xiangtao Chen, received the BSc and PhD degrees from Central South University, Changsha, China, in 2000 and 2004, respectively. He is an Associate Professor in the College of Computer Science and Electronic Engineering at the Hunan University, Changsha, China. His research interests are data mining, machine learning, Large-Scale Information Network Analysis, Social Network and Link Analysis.

Liao Bo, received the PhD degree in Computational Mathematics from Dalian University of Technology, China, in 2004. He is currently a Professor at Hunan University. He was at the Graduate University of Chinese Academy of Sciences as a Post-Doctorate from 2004 to 2006. His current research interests include Bioinformatics, Data Mining and Machine Learning.

