Health-Care Fraud Detection with Machine Committee

Yeolwoo An[a] and Hyunjung Shin[b,*]

[a] *Dept. of Industrial Engineering, Ajou University, Suwon, 443-749, South Korea*
[b] *Dept. of Industrial & Information Systems Engineering, Ajou University, Suwon, 443-749, South Korea*
Tel: +82-31-219-2417 E-mail: {youlwoo, shin}@ajou.ac.kr

ABSTRACT

Health care bill claim review process is heavily dependent on human assessment, which is easily likely to be subjective and error-prone. One way to reduce the possible human mistakes during the review process is to use an objective backup reference, for instance, the assessment result from a machine trained on the bill claim dataset. There have been various types of machines that can be employed in fraud detection problem, i.e., support vector machines, support vector data description, PCA data description, K-means data description, Gaussian density estimation, Mixture of Gaussian and so on. And if a consensus from a committee of different machines can be made, the assessment becomes more objective than that of a single machine since the individuals can complementarily catch the diverse patterns of novelties in the bill-claim dataset. In this paper, we evaluate the proposed committee on the dataset offered by HIRA, 2007.

Keywords: Medical Bill Claim, Hospital / Medical Fraud, Novelty Detectors, Ensemble method, Committee Network, Anomaly Detection

Acknowledgement

The authors would like to gratefully acknowledge support from Post Brain Korea 21 and the research grant from National Research Foundation of Korea (2009-0065043/2010-0028631)

* Corresponding author: Hyunjung (Helen) Shin, shin@ajou.ac.kr

Health-care Fraud Detection with Machine Committee

Yeolwoo An and Hyunjung Shin *
Data Mining Lab, Industrial Engineering Dept, Ajou University, South Korea

* Corresponding author: Hyunjung (Helen) Shin, shin@ajou.ac.kr

Acknowledgement

The authors would like to gratefully acknowledge support from Post Brain Korea 21 and the research grant from National Research Foundation of Korea (2009-0065043/2010-0028631).

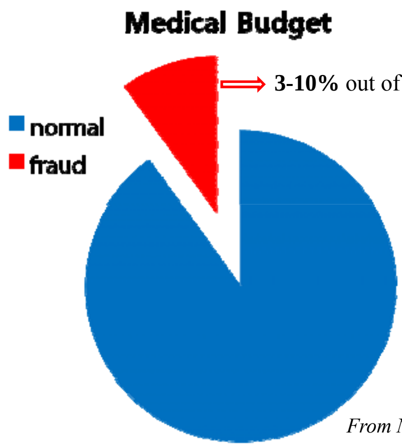
ABSTRACT

Health care bill claim review process is heavily dependent on human assessment, which can be subjective and error-prone. One way to reduce the possible human mistakes is to use an objective backup reference, i.e., the assessment result from a machine trained on the bill claim dataset. If a consensus from a committee of different machines can be made, the assessment becomes more objective than that of a single machine. We evaluate the proposed committee on the dataset offered by HIRA, 2007.

Contents

Introduction : The Background

Why should the medical care frauds be detected? : The Importance



⟹ **3-10%** out of the total: ***60 billion*** to ***160 billion dollars***

Furthermore, the pie is getting even bigger recently

From National Health Care Anti-Fraud Association, 2007

Case: The United States of America

Introduction : Global Issue

Australia

H. He*, et al.*, "Application of Genetic Algorithm and k-Nearest Neighbor Method in Medical Fraud Detection " *Lecture Notes in Computer Science,* vol. 1585, pp. 74-81, 1999.

H. He*, et al.*, "Application of Neural Network to Detection of Medical Fraud," *Expert Systems With Applications,* vol. 13, pp. 329-336, 1997.

G. J. Williams and Z. Huang, "Mining the Knowledge Mine: The hot spots methodology for mining large real world databases," *Lecture Note in Computer Science,* vol. 1342, pp. 340-348, 1997.

Chile

P. A. Ortega*, et al.*, "A Medical Claim Fraud/Abuse Detection System Based on Data Mining: A Case Study in Chile," in *International Conference on Data Mining*, Las Vegas, Nevada, 2006.

Introduction: Global Issue cont'd

USA

A. Shapiro, "The merging Neural Networks, Fuzzy Logic, and Genetic Algorithms.," *Insurance: Mathmatics and Economics,* vol. 31, pp. 115-131, 2002.

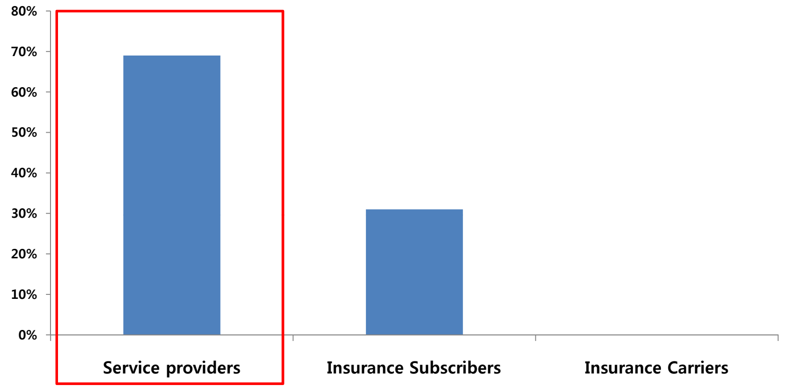
Taiwan

W.-S. Yang and S.-Y. Hwang, "A Process-mining Framework for the Detection of Healthcare Fraud and Abuse," *Expert Systems With Applications,* vol. 31, pp. 56-68, 2006.

C. L. Chan and C. H. Lan, "A Data Mining Technique Combining Fuzzy sets theory and Bayesian classifier—An application of auditing the health insurance fee," in *the International Conference on Artificial Intelligence*, 2001, pp. 402-408.

Introduction

Proportion of Three Types of Medical Fraud



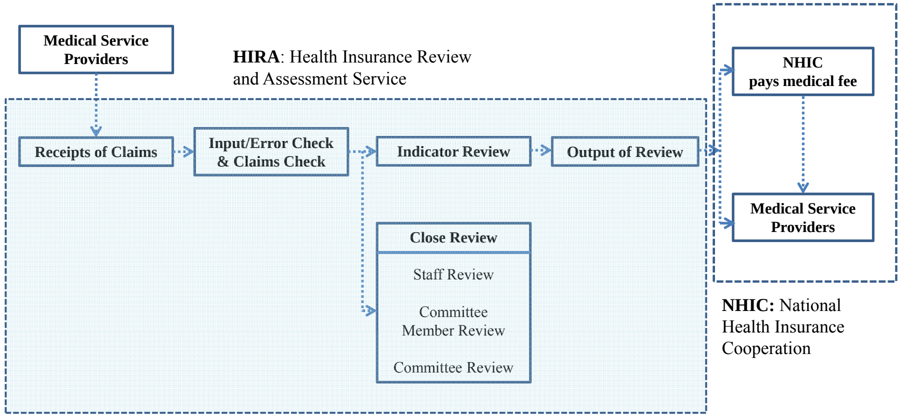
The bill-claim fraud by service providers takes the largest proportion which poses great damage to the quality of health care service

Introduction

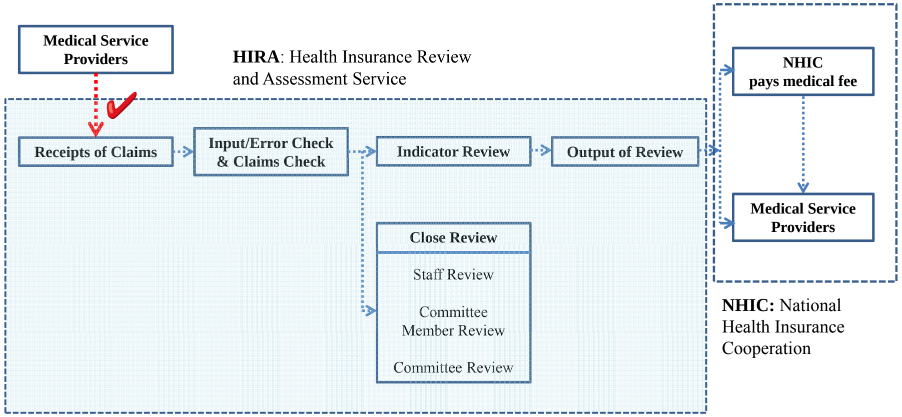
Medical Bill-Claim System
&
Health Care Frauds

(Case of Republic of Korea)

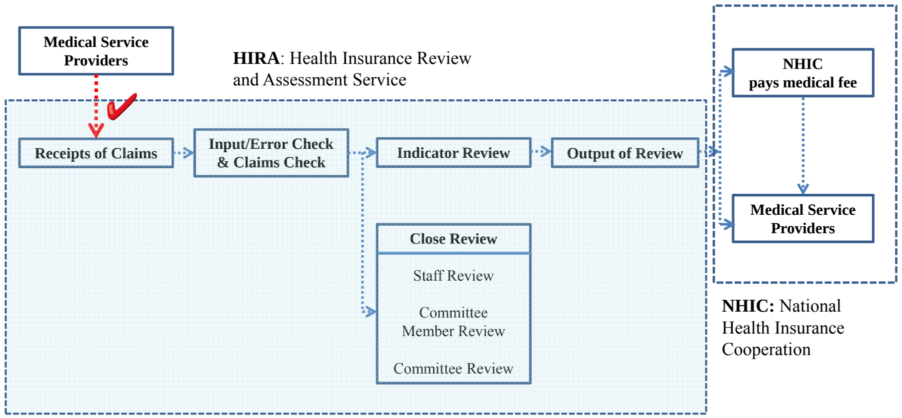
Introduction: Review Process



Introduction: Where the Medical Bill-Claim Frauds Occur

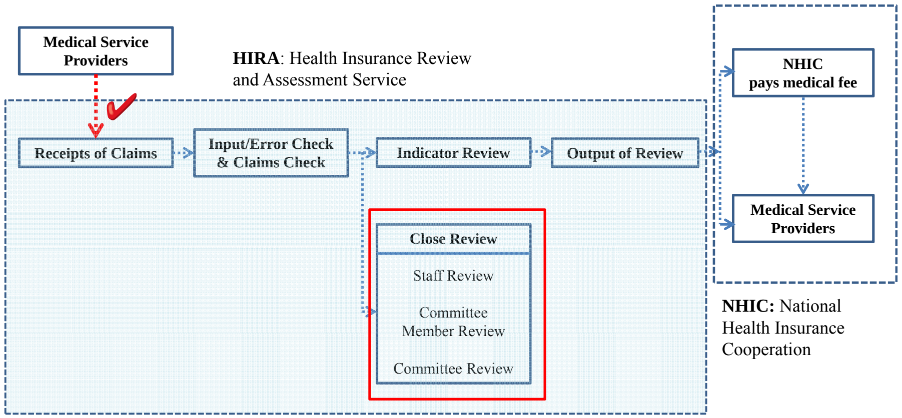


Introduction: What type of Frauds?



1. *Duplicate charges for a single service or item*
2. *Charges for services never performed*
3. *Incorrectly calculated room or pharmacy charges*

Introduction: The Problems of Current Process

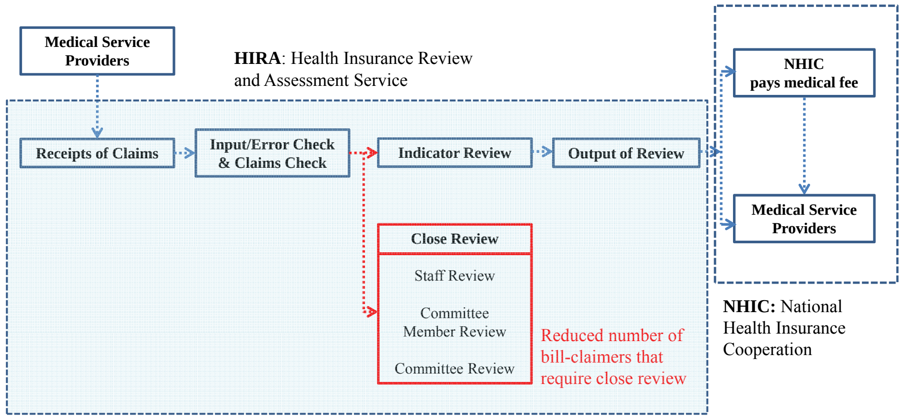


Human Committee suffers from high level of work load reviewing tremendous number of bill-claims which results in errors, and lack of time-efficiency.

Proposed Method

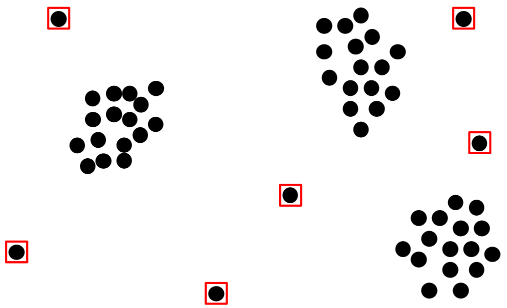
Develop 'filter system' based on consensus from various novelty detectors trained with data

Proposed Method: Goal



Reduce the work load in 'close review step' by providing human committee with pre-filtered suspicious bill-claims that have high probability that require further investigation.

Proposed Method: Novelty Detection Approach



Novelty Detection	Medical-Bill Claim Fraud Detection
Normal	The bill-claimers without intervention history
Abnormal	The bill-claimers with intervention history

Proposed Method: Novelty Detection Approach

Can we rely on single novelty detector's decision?

Choice of Novelty Detectors

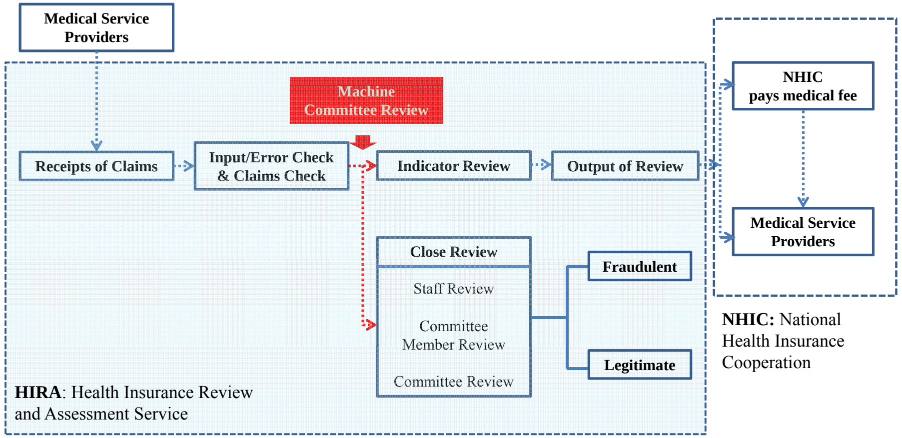
- Each novelty detector has its own strong points and weak points
- It is not usual to be informed of the characteristics of the given task in advance

Choice of Parameters

- Once the novelty detector to use is decided, parameter selection is an another concern

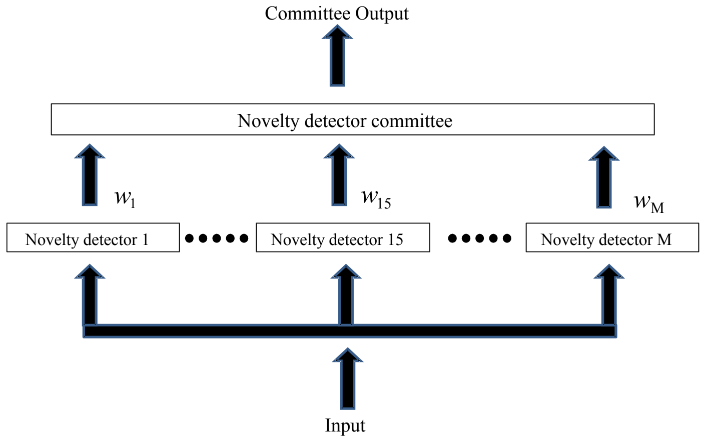
Novelty Detector Committee

Proposed Method: Machine Committee Review



The machine committee review prepare for the list of suspicious medical service providers so that the close review by human experts can be more efficiently performed focusing on reduced number of bill-claims that has high probability to be the intervention target.

Proposed Method: Ensemble Method



Diverse novelty detectors forms *Novelty Detector Committee* to decide whether a service provider shows normal or abnormal bill-claim behavior.

Proposed Method: Committee Members

1_SVM	SVDD	KmeansDD	PCADD	KnnDD	GaussDD	MoGDD
1 C = 0.1	6 σ=3	9 K = 9	12 n= 1	17 K=1	22	23 n=1
2 C = 0.2			13 n= 2	18 K=2		24 n=2
						25 n=3
3 C = 0.3	7 σ=4	10 K = 10	14 n= 3	19 K=3		26 n=4
4 C = 0.4			15 n= 4	20 K=4		27 n=5
						28 n=6
5 C = 0.5	8 σ=5	11 K = 11	16 n= 5	21 K=5		29 n=7
						30 n=8

Proposed Method: Ensemble Method

v(x): ***Majority Vote***

I.D	detector 1	detector 2	• • • •	detector 29	detector30	v(x)
1	+1	0		0	0	0
2	+1	+1		0	+1	+1
3	0	0		+1	+1	+1
4	0	0	• • • •	0	0	0
5	+1	+1		0	0	0
6	+1	+1		+1	+1	+1

Let $v_j = \begin{cases} 1 & \text{if the detector consider the data point as outlier} \\ 0 & \text{otherwise} \end{cases}$

$$v(x) = \frac{1}{2}\,sign\left(\sum_{j=1}^{M} v_j - \frac{M}{2}\right) + \frac{1}{2}$$

M: The committee size

Proposed Method: Ensemble Method Incorporating Weights

s(x): *Novelty Score* by Weighted Majority Vote

I.D	detector 1	detector 2	• • • •	detector 29	detector30	v(x)	s(x)
1	+ 1	0		0	0	0	4.6495
2	+ 1	+ 1		0	+ 1	+ 1	6.3759
3	0	0		+ 1	+ 1	+ 1	17.6453
4	0	0	• • • •	0	0	0	0.0000
5	+ 1	+ 1		0	0	0	5.4316
6	+ 1	+ 1		+ 1	+ 1	+ 1	19.3698

$$s(x) = \sum_{j=1}^{M} w_j v_j \text{, where} \quad w_i = \frac{F_j}{\sum_{j=1}^{M} F_j}$$

Proposed Method: Weights of Individual Machines

		Correct result	
		Not Intervened	Intervened
Obtained Result	Not Intervened	True Positive (TP)	**False Positive (FP)**
	Intervened	**False Negative (FN)**	True Negative (TN)

<Confusion Matrix>

Precision = TP / (TP + FP) : Positive Predictive Value
Recall = TP / (TP +FN) : Sensitivity

$$F = \frac{2 \cdot precision \cdot recall}{precision + recall}$$

Committee Members

Mechanisms of Various Novelty Detectors

Methodology

Various Novelty detectors

A ***Classification Based Novelty Detection Algorithms***
A.1 One class Support Vector Machine
A.2 Support Vector Data Description

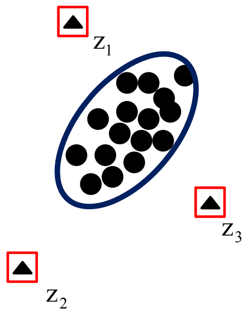
B ***Nearest Neighbor Analysis Detection Algorithms***
B.1 K-nearest Neighbor data description

C ***Clustering Based Novelty Detection Algorithms***
C.1 K-means Data Description

D ***Spectral Anomaly Detection Algorithms***
D.1 Principal Component Analysis Data Description

E ***Statistical Anomaly Detection Algorithms***
E.1 Gaussian density estimation
E.2 Mixture of Gaussian Data Description

A.1 One Class SVM



Training Phase

$$\min_{w\in F,\ \xi\in\Re^{\ell},\ \rho\in\Re} \frac{1}{2}\|\omega\|^2 + \frac{1}{\nu\ell}\sum_i \xi_i - \rho$$

subject to $(\omega\cdot\Phi(x_i)) \geq \rho - \xi_i,\ \ \xi_i \geq 0$

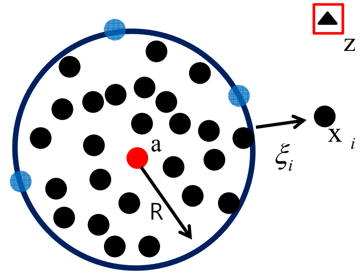
When solved w.r.t $\boldsymbol{\omega}$ *and* $\boldsymbol{\rho}$ this problem, the formulation leads to *decision function*.

Test Phase

$$f(\mathrm{z}) = \mathrm{sgn}((w\cdot\Phi(\mathrm{z})) - \boldsymbol{\rho})$$

Most of original normal patterns are in the estimated region. The data points in the estimated region are to be labeled as +1, and -1 for the ones outside the region.

A.2 Support Vector Data Description(SVDD)



Support vectors

The center of the sphere

Novelty

The sphere is characterized by a center a and radius R and the sphere contains all training objects χ^{tr}

Training phase

Define the structural error:

$$\varepsilon_{struct}(R, \mathrm{a}) = R^2$$

which has to be minimized with the constraints:

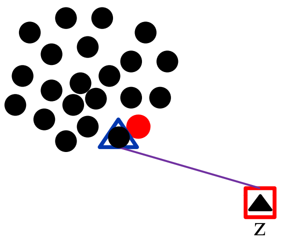
$$\|\mathrm{x}_i - \mathrm{a}\|^2 \leq R^2, \qquad \forall i$$

Test Phase

$$f(\mathrm{z}) = \|\mathrm{z}_i - \mathrm{a}\|^2 - R^2 \geq 0$$

B.1 *K*-NN Data Description

when k=1; Nearest Neighbor case



Training Instances

Test Instance

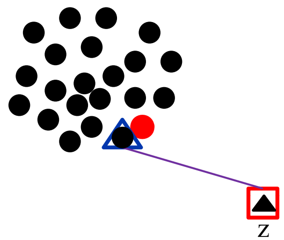
NN of the Test Instance

NN of the Training Instance

1. A test instance picks the nearest neighbor from training set
2. The selected one also selects the nearest neighbor from training set
3. If the distance from 1 is much greater than the distance from 2 the corresponding test instance is considered as an outlier

B.1 *K*-NN Data Description

when k=1; Nearest Neighbor case



Training Instances

Test Instance

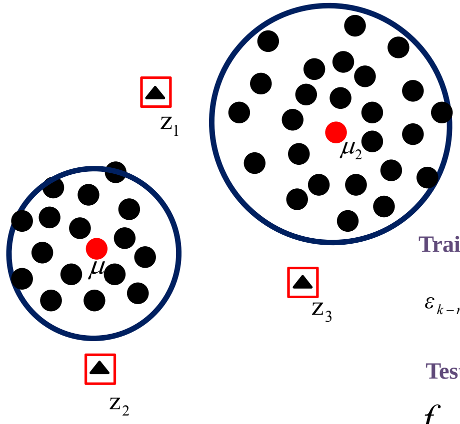
NN of the Test Instance

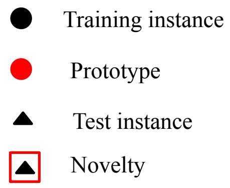
NN of the Training Instance

$$p_{NN}(z) = \frac{k/N}{V_k(\|z - NN_k^{tr}\|)}$$

$$f_{NN^{tr}}(z) = I\left(\frac{V(\|z - kNN^{tr}(z)\|)}{V\|(NN^{tr}(z) - NN^{tr}NN^{tr}(z))\|} \leq 1\right)$$

C.1 K-means Data Description





Training Phase

$$\varepsilon_{k-means} = \sum_i \left(\min_k \left\| x_i - \boldsymbol{\mu}_k \right\|^2 \right)$$

Test Phase

$$f_{k-means}(z) = \min_k \left\| z - \boldsymbol{\mu}_k \right\|^2 > \boldsymbol{\theta}$$

Thus, **the detection** can be performed based ***on the squared distance of an object z to the nearest prototype***

D.1 Principal Component Analysis Data Description

This method describes the target data by a linear subspace. This subspace is defined by the eigenvectors of the data covariance matrix.

Training Phase

The subspace is defined by the eigenvectors of the data covariance matrix Σ. Only k eigenvectors are used. assume they are stored in d by k matrix **W.**

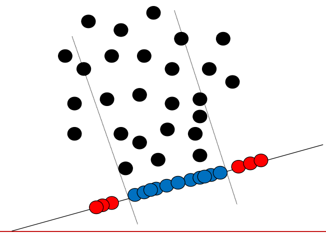
Test Phase

To check if a new object **x** fits the target space,
➔ Compute the ***reconstruction error***: the difference between the original object **x** and the projection of that object on to the subspace.

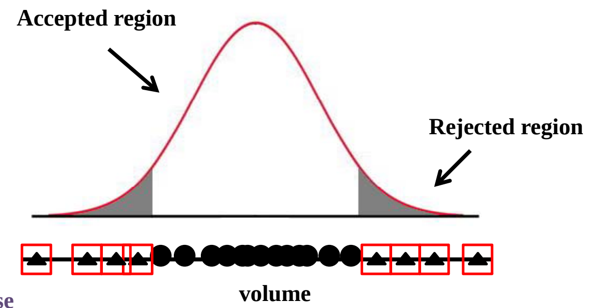
$$z_{proj} = W(W^T W)^{-1} W^T z$$

for the reconstruction error

$$f(z) = \left\| z - z_{proj} \right\|^2 > \boldsymbol{\theta}$$



E.1 Gaussian Density Estimation



Training Phase

The probability distribution for a d-dimensional object **x** is given by

$$p_N(x; \boldsymbol{\mu}, \Sigma) = \frac{1}{(2\boldsymbol{\pi})^{d/2} |\Sigma|^{1/2}} \exp\{-\frac{1}{2}(x - \boldsymbol{\mu})^T \Sigma^{-1} (x - \boldsymbol{\mu})\}$$

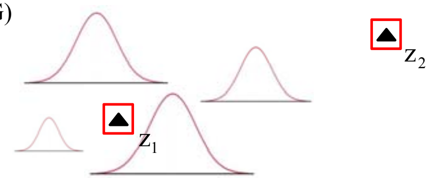
where $\boldsymbol{\mu}$ *is the mean and* Σ *is the* cov *matrix,* $\boldsymbol{\mu}$ *and* Σ *are sample estimate*

Test Phase

$f(z) = p_N(z; \boldsymbol{\mu}, \Sigma) < \boldsymbol{\theta}$ ⬌ **z** is outlier

E.2 Mixture of Gaussian Data Description

The Gaussian distribution assumes a very strong model of the data. It should be unimodal and convex. For most datasets these assumptions are violated. To obtain a more flexible density estimation, the normal distribution can be extended to a mixture of Gaussians(MoG)



Training Phase

$$p_{MoG}(x) = \frac{1}{N_{MoG}} \sum_j \boldsymbol{\alpha}_j p_N(x; \boldsymbol{\mu}_j, \Sigma_j)$$

where α_j are the mixing coefficients. P_i, μ_j, Σ_j are optimized using the EM algorithm.

Test Phase

$f(z) = p_{MoG}(z; \boldsymbol{\mu}, \Sigma) < \boldsymbol{\theta}$ ⬌ **z** is outlier

Experiments & Results

Experiments: Data

Medicus Gratus(MG)

	Input variables
1	Number of medicine
2	Costliness index
3	VI index
⋮	⋮
34	The 4th ranked CI
35	The 5th ranked CI

# of attributes : 35

# of data points : 3,694

Provided by HIRA 2007, 3/4

Experiment Setup: Committee Members

Total *30 machines* are employed ⇔ Total *30 committee members*

1_SVM	SVDD	KmeansDD	PCADD	KnnDD	GaussDD	MoGDD
1 C = 0.1	6 σ=3	9 K = 9	12 n= 1	17 K=1	22	23 n=1
2 C = 0.2			13 n= 2	18 K=2		24 n=2
3 C = 0.3	7 σ=4	10 K = 10	14 n= 3	19 K=3		25 n=3
4 C = 0.4			15 n= 4	20 K=4		26 n=4
5 C = 0.5	8 σ=5	11 K = 11	16 n= 5	21 K=5		27 n=5
						28 n=6
						29 n=7
						30 n=8

Experiment Setup: Cross Validation

5-fold Cross Validation

	Training Set	**Test Set**	**Total**
Number of **normal patterns**	2,670	667	3,337
Number of **abnormal patterns**	0	357	357
Sum	2,670	1,024	3,694

Abnormal patterns are used only in test phase not in training phase, since for training phase of novelty detection algorithm, only normal patterns are used.

Experiment Setup: Fraction of outliers

Fraction of Outliers

- Fraction of outliers is set before training phase begins.

- Fraction of outliers let the learner control how many data points to consider as outliers in test set.

- In this research, the fraction of outliers is set as 10%.

Ensemble Method

v(x): *Majority Voting*

I.D	detector 1	detector 2	• • • •	detector 29	detector30	v(x)
1	+ 1	0		0	0	0
2	+ 1	+ 1		0	+ 1	+ 1
3	0	0		+ 1	+ 1	+ 1
4	0	0	• • • •	0	0	0
5	+ 1	+ 1		0	0	0
6	+ 1	+ 1		+ 1	+ 1	+ 1

$$\text{Let } v_j = \begin{cases} 1 & \text{if the detector consider the data point as outlier} \\ 0 & \text{otherwise} \end{cases}$$

$$v(x) = \frac{1}{2} sign\left(\sum_{j=1}^{M} v_j - \frac{M}{2} \right) + \frac{1}{2}$$

M: The committee size

Ensemble Method Incorporating Weights

s(x): *Novelty Score* by Weighted Majority Voting

I.D	detector 1	detector 2	• • • •	detector 29	detector30	v(x)	s(x)
1	+ 1	0		0	0	0	4.6495
2	+ 1	+ 1		0	+ 1	+ 1	6.3759
3	0	0		+ 1	+ 1	+ 1	17.6453
4	0	0	• • • •	0	0	0	0.0000
5	+ 1	+ 1		0	0	0	5.4316
6	+ 1	+ 1		+ 1	+ 1	+ 1	19.3698

$$s(x) = \sum_{j=1}^{M} w_j v_j , \; where \quad w_i = \frac{F_j}{\sum_{j=1}^{M} F_j}$$

Weights of Individual Machines for Ensemble Learning

		Correct result	
		Not Intervened	Intervened
Obtained Result	Not Intervened	True Positive (TP)	**False Positive (FP)**
	Intervened	**False Negative (FN)**	True Negative (TN)

<Confusion Matrix>

Precision = TP / (TP + FP) : Positive Predictive Value
Recall = TP / (TP +FN) : Sensitivity

$$F = \frac{2 \cdot precision \cdot recall}{precision + recall}$$

Results : Machine Committee Reaches a Consensus

RANK	I.D	SCORES
1	668	19.3698
1	678	19.3698
1	681	19.3698
1	689	19.3698
1	693	19.3698
1	704	19.3698
⋮	⋮	⋮
100	847	8.1965
101	677	7.7288
102	577	7.7253

The scores are ***sorted in descending order***

The fraction of outliers: **10%**

# of instance in test set: **1,024**

➔1,024 * 0.1 = 102.4, rounded to **102**

So, the top 102 ranked medical bill claimers are selected to be considered as the target for intervention

Results : Machine Committee Reaches a Consensus

Test Set	
# of total test set	1,024
# of bill claimers without intervention(proportion)	667(65%)
# of bill claimers with intervention(proportion)	357(35%)

Intervention history	Results		
	Statistics	consensus score <threshold$_{10\%}$	consensus score > threshold$_{10\%}$
Non-intervention (n= 667)	# of bill-claimers	660 (98.95%)[a]	7 (1.05%)[b]
	average score	2.7561	13.1500
Intervention (n= 357)	# of bill-claimers	262 (83.39%)[c]	95 (26.61%)[d]
	average score	2.9670	16.4328

Most of the top ranked 102 bill claimers are from the ones with intervention history, where 7 of them are from the ones without intervention history.

Conclusion

Conclusion

MCRS as Reference System

In this research, we suggested the machine committee review system (MCRS) that facilitates the more objective and time/manpower efficient filtering process to provide human examiners with the most likely 'suspect list' in detecting health care fraudsters. The list provided by MCRS functions as a guide for human examiners.

Ensemble method: Weighted Majority Vote

The novelty detector committee reaches a consensus by weighted majority voting using f-measures of individual novelty detectors as corresponding weights. The highly scored service providers are selected as suspicious intervention target which require close review in detail, the medical bill claim fraudsters.

Human Committee Members Focus on Suspicious Bill Claims

The human committee members can focus on reduced number of bill claims so that possible erroneous reviews caused by heavy work load can be prevented.