Automobile Insurance Fraud Detection using the Evidential Reasoning Approach and Data-Driven Inferential Modelling

Xi Liu, Jian-Bo Yang, Dong-Ling Xu
Decision and Cognitive Sciences Research Centre
The University of Manchester
Manchester, UK
xi.liu-2; jian-bo.yang; Ling.Xu@manchester.ac.uk

Karim Derrick, Chris Stubbs, Martin Stockdale
Kennedys Law LLP, 3rd floor, Churchgate House, 56 Oxford St
Manchester, UK
Karim.Derrick; Chris.Stubbs; Martin.Stockdale@kennedyslaw.com

***Abstract*—Automobile insurance fraud detection has become critically important for reducing the costs of insurance companies. The majority of insurance companies use expert knowledge to detect fraud. Experience-based knowledge are interpretable and re-usable but the simplistic way that this knowledge is used in practice, often leads to some degree of misjudgment. This paper aims to establish a unique Evidential Reasoning (ER) rule that combines independent evidence from both experience based indicators and probabilities of fraud obtained from historical data. Each piece of evidence is weighted and then combined conjunctively with the weights optimised using a maximum likelihood evidential reasoning (MAKER) framework for data-driven inferential modelling. Based on a real-world insurance claim dataset, our experimental results reveal that the proposed approach preserves the interpretability and usability of expert detection system, and anticipates the changes in fraud practices by tracking the trend of the weights of experience-based indicators. Furthermore, the experimental results show that the proposed approach outperforms a number of widely used machine learning models, such as logistic regression and random forests.**

***Index Terms*—Evidential Reasoning Rule, Fraud Detection, Decision Making**

I. Introduction

Every year insurance fraud costs businesses billions. The Insurance Fraud Taskforce calculate the total cost to be in excess of £3bn, with undetected fraud costing more than £2bn. In addition there is also the cost of those measures required by insurers to fight it, with the ABI reporting this to be in excess of $200m per year. These statistics illustrate the significant problem fraud presents. Winning the fight against fraud will never be an end game where fraud stops existing. Instead, there is an acceptance that measures to protect against fraud will always be necessary and part of operational best practice.

In respect of fraud detection, most insurance companies predominantly adopt rules-based approaches by pre-defining a set of red flags (fraud indicators), which are interpretable and reusable. However, progress has been slow with such an approach, moreover, such a rules-based system can have expert knowledge biases as it is generated based on the subjective experience of a claim being of higher fraud risk. Furthermore, the underwriting process of high-risk claims is very detailed and individualistic which strongly depends on the experience or knowledge of underwriters.

Those valuable pieces of knowledge from fraud specialists are the centre of the fraud prevention system, hence, in order to help the fraud analysts be more efficient in their work we need technology to augmenting the expert rules. In this paper, we develop a data-driven fraud prevention and detection service which requires a robust system to inference and back-test the underwriters' decisions to minimize unrecognised biases, errors and time.

There have been various research efforts conducted in this space to reduce the fraud activity using advanced data analytics; see the details from [1]–[3]. In practice, the whole system is expected to be interpretable and as transparent as possible so that all involved parties during the handling process will be aware of the cause of the decision. Traditional interpretable ML approaches, such as logistic regression [4], [5] and among others, tree-based approaches [6], [7] and among others, rely on the certainty of the consequence of interest. However, during the prolonged claim handling process, quite a few claims will be ended up with ambiguous results or drop out without any conclusion. Hence, the imprecision or even incorrectness of classification results is likely to be caused and how to handle different types of uncertainties becomes a practical problem.

From the perspective of interpretable and uncertain information processing, Dempster–Shafer evidence theory (DST) can provide a mechanism to deal with the classification imprecision. In fact, different kinds of uncertainty may coexist in real systems, e.g., fuzzy information may coexist with ignorance, leading to the induction of knowledge without certainty but only with degrees of belief or credibility regarding a hypothesis [8]. Recently, the evidential reasoning (ER) rule has been established to advance the seminal DST [9]–[12] and the original ER algorithm [13]–[16]. It has been proved in [17] that (1) the Dempster's combination (DC) rule is a special case of the ER rule when each piece of evidence is fully reliable, and (2) the original ER algorithm is also a special case when the weights of all pieces of evidence being normalized are

978-1-7281-6932-3/20/$31.00 ©2020 IEEE

equal to their respective reliabilities .

Compared with the DC rule, the main advance of the ER rule is to propose a novel concept of weighted evidence (WE) and extend to WE with Reliability (WER) in order to characterize evidence in complement a basic belief assignment (BBA), i.e. a belief distribution (BD) function introduced in the DST. The most important property of the ER rule is that it constitutes a generic conjunctive probabilistic reasoning process, or a generalized Bayesian inference process which can be implemented on the power set of Frame of Discernment (FoD). Moreover, the evidence reasoning procedure which consists of the belief structure can model various types of uncertainty, such as incomplete information, probabilistic uncertainty, etc. [18], [19].

Given that the ER rule has explicitly generalized the DST and the original ER algorithm, it becomes perfectly logical and suitable to handle various types of uncertainty in general. This paper presents a unique ER rule [17], [20] to combine multiple pieces of independent evidence from both experience based indicators and probabilities of fraud obtained from historical data. Each piece of independent uncertain evidence is profiled as a belief distribution and then combined conjunctively with the weights optimised by using a maximum likelihood evidential reasoning (MAKER) framework for data-driven inferential modelling. A MAKER framework can be treated as an extended Bayesian inference process, where the reliability of evidence and probabilistic prediction can be learnt from historical data by maximising the likelihood of true state. Other approaches such as those based on either on rough set or on Choquet integral would not be applicable to our particular case. Specifically, to address the common class-imbalanced issue within the data, we conduct a cost-sensitive objective function to avoid changing the prior of the data, which can be treated as independent evidence and used to incorporate additional information into the evidence set. An optimization model using Sequential Quadratic Programming (SQP) is proposed, available in the R package *nloptr* [21].

The rest of paper is structured as follows. Section II briefly covers the essentials of the real world automobile insurance dataset built into the ER framework. In Section III, we briefly review the ER framework and elaborate the inference process with a detailed workflow illustrated in Section IV. A UCI benchmark dataset as well as a real-word example of insurance fraud detection in Section V are proposed to illustrate the outperformance of the proposed framework. Section VI presents the conclusions and discussions.

II. Input & Output

A. Input

Kennedys Law LLP in conjunction with the University of Manchester have committed to adopting an automated approach by embedding the subject matter expertise of fraud specialists, and flex the rules to meet unique fraud threats faced by various clients in order to add: relevant fraud indicators and a cross-industry watch list analysis and detection methods to data matching. Those relevant fraud indicators are generalized based on the expert knowledge and internal & external data, integrated with key industry and public data sources, as well as consented data and Kennedys own data. During claim handling process, fraud specialists assess multiple factors in a claim, including but not limited to time of accident; medical attention/treatment; vehicle occupancy; claim chronology, etc.

Kennedys has provided a dataset $\mathbf{D}$ that includes $N = 718$ entries of insurance claim record that concerns with car accident affairs, associated with $K = 49$ key performance indicators that Kennedys wish to improve. The provided dataset is of the format shown in Table I. The first 477 samples are used as training data for model estimation and the remaining 241 samples are left as test data for the out-of-sample evaluation. Both of them share similar prior compared to the original whole dataset, shown in Table II.

TABLE I
An example of the provided inputs.

Indicator	Ind_1	Ind_2	...	Ind_{48}	Ind_{49}	Fraud
Record 1	Y	N	$\cdots$	Y	Y	Yes
Record 2	Y	N	$\cdots$	Y	Y	No
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\vdots$
Record N	N	Y	$\cdots$	Y	N	No

Datasets for fraudulent claims are usually imbalanced, as it is shown in Table II. Imbalanced data may greatly affect the performance of classification algorithms. The prediction will be biased towards the majority class present in the dataset. The traditional way of handling this issue of imbalance is to utilise a re-sampling methodology, such as undersampling the majority class; oversampling the minority class; SMOTE (Synthetic Minority Over-sampling Technique), etc. [22]. The significant shortcomings with the re-sampling approach are that the optimal class distribution is always unknown and the criterion in selecting instances is uncertain; furthermore under-sampling may increase information loss and over-sampling may lead to overfitting or overgeneralization of the model constructed [23]. Moreover, such re-sampling approaches change the prior of the original data, which can be treated as an independent evidence and used to incorporate additional information into the evidence set.

TABLE II
Comparison of the prior between the training/validation and original dataset.

	Legitimate	Fraud
Whole sample	84.26%	15.74%
Training sample	85.95%	14.05%
Test sample	80.91%	19.09%

Given this principle, an example-dependent cost-sensitive learning will be appropriate, which takes example-dependent costs into account and make predictions that aim to minimize the overall costs instead of minimizing misclassifications [6], [24]. The details of cost-sensitive learning will be introduced in Section IV-E.

B. Output

Most companies predominantly adopt rules-based approaches to fraud detection using IF-THEN rules. Red flags and indicators are necessary parts of fraud detection. However, a strict IF-THEN rules format will lead to a higher rate of false positives. Hence, in this paper we will report on the optimised IF-THEN rule-based output in a belief way (Table III), which protects innocent people from false accusations and everyone from the cost of insurance fraud.

TABLE III
REPRESENTATION OF BELIEF RULE IN EXPERT KNOWLEDGE RULE-BASE.

IF: (Ind_1 is Triggered) & (Ind_2 is Triggered) & $\cdots$ & (Ind_k is NOT Triggered)
THEN: Decision is {(High risk, 0.65), (Low risk, 0.35)}

III. OUTLINE OF THE EVIDENTIAL REASONING RULE

In this section, the ER rule [17], [20], [25] is briefly reviewed. Suppose $\mathbb{H} = \{h_1, h_2, \cdots, h_c\}$ is a set of mutually exclusive and collectively exhaustive hypotheses. $\mathbb{H}$ is referred to as a frame of discernment. The power set of $\mathbb{H}$ consists of all its subsets, denoted by $P(\mathbb{H})$ or $2^{\mathbb{H}}$. A piece of evidence on high risk of fraud is profiled by a belief distribution as follows:

$$e_i = \left\{ (h, p_{h,i}) | \forall h \subseteq \mathbb{H}, \sum_{h \subseteq \mathbb{H}} p_{h,i} = 1 \right\}, \tag{1}$$

where $(h, p_{h,i})$ is an element of evidence e_i, represents that the evidence points to proposition h with the degree of p_h, i referred to as probability or degree of belief in general. h can be any subset of $\mathbb{H}$ or any element of $P(\mathbb{H})$ except for the empty set. $(h, p_{h,i})$ is referred to as a focal element of e_i if $p_{h,i} > 0$.

In the ER rule, reliability r_i and weight w_i of evidence e_i are defined. The former indicates the ability of the attribute or its evidence to provide correct assessment, which is the inherent property of the evidence; while the latter reflects the relative importance of evidence in comparison with other evidence when they need to be combined and determined according to who uses the evidence. This means that weight w_i can be subjective and different from reliability r_i in situations where different pieces of evidence are generated from different sources and measured in different ways. A so-called weighted belief distribution with reliability (WBDR) can be defined as follows:

$$m_i = \left\{ (h, \tilde{m}_{h,i}) | \forall h \subseteq \mathbb{H}, (P(\mathbb{H}), \tilde{m}_{P(\mathbb{H}),i}) \right\}, \tag{2}$$

where $\tilde{m}_{h,i}$ measures the degree of support for h from e_i with both the weight and reliability of e_i taken into account, defined as follow

$$\tilde{m}_{h,i} = \begin{cases} 0 & h = \emptyset \\ C_{rw,i} m_{h,i} & h \subseteq \mathbb{H}, \; h \neq \emptyset \\ C_{rw,i}(1 - r_i) & h = P(\mathbb{H}) \end{cases} \tag{3}$$

where $m_{h,i} = w_i p_{h,i}$ and $C_{rw,i} = 1/(1 + w_i - r_i)$ is a normalization factor such that $\sum_{h \subseteq \mathbb{H}} \tilde{m}_{h,i} + \tilde{m}_{P(\mathbb{H}),i} = 1$ given $m_{h,i} = w_i p_{h,i}$ and $\sum_{h \subseteq \mathbb{H}} p_{h,i} = 1$. A so-called weighted belief distribution (WBD) can be treated as a special case of WBDR if e_i and other evidence are acquired from the same data source hence with $r_i = w_i$. Note that, in this paper all the calculation are generated via a weighted ER system based on such assumption so that $r_i = w_i$.

If two pieces of evidences e_1 and e_2 are independent, the combined degree of belief to which e_1 and e_2 jointly support proposition h, denoted by $p_{h,e(2)}$, can be generated by the ER fusion rule as follows

$$\begin{aligned} p_{h,e(2)} &= \begin{cases} 0 & h = \emptyset \\ \frac{\hat{m}_{h,e(2)}}{\sum_{D \subseteq \mathbb{H}} \hat{m}_{D,e(2)}} & h \subseteq \mathbb{H}, \; h \neq \emptyset \end{cases} \\ \hat{m}_{h,e(2)} &= [(1 - r_2) m_{h,1} + (1 - r_1) m_{h,2}] \\ &\quad + \sum_{B \cap C = h} m_{B,1} m_{C,2}, \quad \forall h \subseteq \mathbb{H} \end{aligned} \tag{4}$$

The recursive formulae of the ER rule to combine multiple pieces of evidence in any order are also given in [17].

$$\begin{aligned} m_{h,e(i)} &= [m_1 \oplus \cdots \oplus m_i](h) \\ &= \begin{cases} 0 & h = \emptyset \\ \frac{\hat{m}_{h,e(i)}}{\sum_{D \subseteq \mathbb{H}} \hat{m}_{D,e(i)} + \hat{m}_{P(\mathbb{H}),e(i)}} & h \subseteq \mathbb{H}, \; h \neq \emptyset \end{cases} \\ \hat{m}_{h,e(i)} &= \left[(1 - r_i) m_{h,e(i-1)} + m_{P(\mathbb{H}),e(i-1)} m_{h,i} \right] \\ &\quad + \sum_{B \cap C = h} m_{B,e_{i-1}} m_{C,i}, \quad \forall h \subseteq \mathbb{H} \\ \hat{m}_{P(\mathbb{H}),e(i)} &= (1 - r_i) m_{P(\mathbb{H}),e(i-1)} \end{aligned} \tag{5}$$

It is proven that Dempster's rule is a special case of the above ER rule when each piece of evidence e_i in question is assumed to be fully reliable, or r_i = 1 for all i. Moreover, it has been proven in [20] that Bayes' rule is also a special case of ER in likelihood inference if likelihoods are normalised for mapping observations from sample space to hypothesis space. The relationship among Bayes' rule, Dempster's rule and the above ER is displayed in Figure 1.

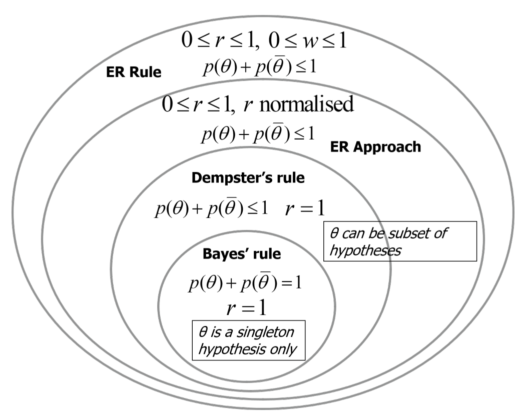

Fig. 1. The relationship among Bayes' rule, Dempster's rule and Evidential Reasoning rule

IV. THE DATA-DRIVEN ER METHODOLOGY FOR AUTOMOBILE FRAUD DETECTION

In this section, we elaborate on the inference process of ER associated with the application of automobile fraud detection.

A. Data-driven likelihood calculation for single indicator

The following data-driven likelihood of single indicator has been conducted with an inferential modelling process [26]. Let $e_{f,k}$ denote the f^{th} piece of evidence from the k^{th} input variable Ind_k at $\text{Ind}_k = \text{Ind}_{f,k}$ and $e_{f,k}(h)$ be an element of evidence $e_{f,k}$ pointing exactly to proposition h. The evidence from the single input variable Ind_k can be transformed to a basic probability distribution for all $h \subseteq \mathbb{H}$. Let $l_{h,f,k}$ be the likelihood of observing f^{th} piece of evidence from the k^{th} input variable V_k given proposition h, the basic probability, obtained from a normalized likelihood, is given by:

$$p^{(D)}_{h,f,k} = p(e_{f,k}(h)) = \frac{l_{h,f,k}}{\sum_{X \subseteq \mathbb{H}} l_{X,f,k}} \quad \forall\ h \subseteq \mathbb{H}, \tag{6}$$

which is a one-dimensional evidence acquisition process, where the likelihood is obtained by generating a contingency table.

B. Experience-based likelihood of a single indicator

For incorporating the expert-defined score of each fraud indicator, we convert it into normalised percentile format $score_{h,f,k} \rightarrow p^{(E)}_{h,f,k}$ with Figure 2 displaying the relationship visually.

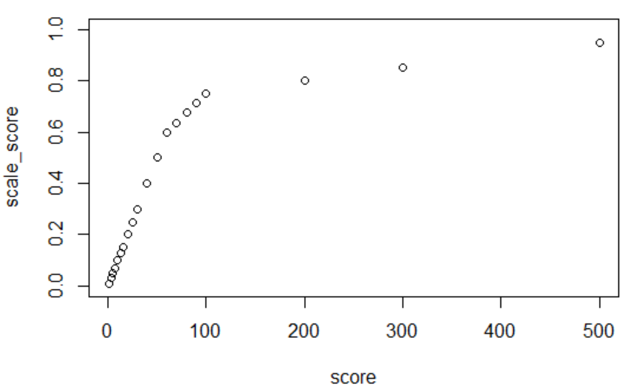

Fig. 2. Relationship between the expert-defined scores and scaled normalised score.

C. Likelihood Fusion

After deriving both experience-based and data-driven likelihood for single indicators, we adopt the below logics to obtain the combined likelihood $p_{h,f,k}$ for the k^{th} input variable Ind_k at $\text{Ind}_k = \text{Ind}_{f,k}$. Since the dataset is imbalanced, the data-driven likelihood for single indicator can only be adopted when the size of fraud cases is statistically significant. Hence, under the first scenario, we only use the experience based likelihood as a representative when the number of fraud cases under the chosen indicator is less than a certain amount value, a. The second scenario shows the situations where the number of fraud cases is beyond a relative large number b, we omit the effect of the expertise and make full use of the data-derived likelihood. The third scenario is an intermediate case, where ER is used to combine the two pieces of independent evidence (experienced-based and data-driven likelihood) for generating their joint support for a single indicator.

TABLE IV
COMBINE EXPERIENCED & PRACTICAL LIKELIHOOD FOR SINGLE INDICATORS (REGULARISED).

	Condition	Combined Likelihood
1	$\#h_{f,k} < a$	$p_{h,f,k} = p^{(E)}_{h,f,k}$
2	$\#h_{f,k} > b$	$p_{h,f,k} = p^{(D)}_{h,f,k}$
3	$\#h_{f,k} \in [a, b]$	$p_{h,f,k} = p_{h,e(2)}$ calculated via Eq.(4)

Notes. (1) $\#h_{f,k}$ denotes the number of fraud cases triggered with k^{th} input variable at $\text{Ind}_k = \text{Ind}_{f,k}$; (2) a and b refer to confidential cut points.

D. ER Algorithm

Algorithm 1 elaborates the ER rule for insurance fraud detection.

E. Training of ER-based parameters using SQP

Here we discuss how to determine the significant parameters in the ER framework, namely the reliability and weight of each piece of evidence, that is, r_i and w_i in Eq.(4), these are the parameters that need to be assigned for inference.

Datasets for fraudulent claims are usually imbalanced, which may greatly affect the performance of classification algorithms. The prediction will be biased towards the majority class present in the dataset. Here we conduct an example-dependent cost-sensitive learning, which take example dependent costs into account and make predictions that aim to minimize the overall costs instead of minimizing misclassifications [6], [24]. In our particular automobile fraud case, Table V shows the following costs associated with each prediction scenario.

TABLE V
THE DATA SETS USED FOR EXPERIMENTATION CONFUSION AND COST MATRIX.

	Actual Fraud ($p^{(m)} = 1$)	**Actual Legitimate** ($p^{(m)} = 0$)
Predicted Fraud ($\hat{p}^{(m)} = 1$)	**True Positive** $C_{TP} = 0$	**False Positive** $C_{FP} = admin$
Predicted Legitimate ($\hat{p}^{(m)} = 0$)	**False Negative** $C_{FN} = compensation$	**True Negative** $C_{TN} = 0$

To evaluate a model's performance in terms of costs, the sum of all costs resulting from the predictions based on each scenario is calculated as follows:

$$\text{Cost} = \sum_{m=1}^{M} p^{(m)}(h) \left(\hat{p}^{(m)}(h) C_{TP_i} + (1 - \hat{p}^{(m)}(h)) C_{FN_i}\right) + \left(1 - p^{(m)}(h)\right) \left(\hat{p}^{(m)}(h)\ C_{FP_i} + (1 - \hat{p}^{(m)}(h))\ C_{TN_i}\right)$$

Algorithm 1: ER Algorithm

Input: D

Output: Likelihood to Fraud under ER rule p_h

Calculate $p^{(D)}_{h,f,k}$, and $p^{(E)}_{h,f,k}$

Calculate $p_{h,f,k}$ according to Table IV

if *Claim is triggered with only one indicator* **then** $p_h = p_{h,f,k}$

else

  repeat

    Let claim be triggered with $k(k > j)$ indicators

    Initialisation:

$$val(i) = p_{h,e(i)}$$
$$W(i) = \prod_{i=1}^{i}(1 - w_i)$$

    $i = 1$

    Compute $p_{h,e(i+1)}$

$$p_{h,e(i+1)} = \frac{(1 - w_{i+1})\ val(i) + W(i)\ p_j + val + p_{i+1}\ (1 - val(i))}{(1\ -\ w_{i+1} + W(i))}$$

    Update val and w

$$val(i+1) = p_{h,e(i+1)}$$
$$W(i+1) = \prod_{i=1}^{i+1}(1 - w_i)$$

    $i = i + 1$

  until $i = k - 1$

In order to take into account the different costs during the training of the algorithm, we develop a cost-sensitive objective function inherited from cost-sensitive logistic regression [27]:

$$J^c(\delta) = -\frac{1}{2M}\sum_{m=1}^{M} p^{(m)}(h)\left(\hat{p}^{(m)}(h)C_{TP_i} + (1-\hat{p}^{(m)}(h))C_{FN_i}\right)$$
$$+\left(1 - p^{(m)}(h)\right)\left(\hat{p}^{(m)}(h)\ C_{FP_i} + (1-\hat{p}^{(m)}(h))\ C_{TN_i}\right)$$
$$\hat{\delta} = \arg\min_{\delta} J^c(\delta).$$
$$s.t.\ \ r_i, w_i \in \Omega$$

where, $\hat{p}^{(m)}(h)$ and $p^{(m)}(h)$ are the real probability and the estimated probability to which a proposition h is true given in the mth observation, respectively. Ω is the feasible space of parameters, with constraints such as $0 < w_i \leq 1$.

F. ER work flow

Algorithm 1 elaborates the ER rule for insurance fraud detection. Figure 3 displays the work flow of the automated automobile fraud detection system deployed in Kennedys Law LLP. After acquiring and preprocessing the knowledge that either experienced based or data-related following Sections IV-A - IV-C, we store them in a knowledge base for further training based on an inference reasoning engine elaborated in Sections III and IV-D. By optimising the whole process, the proposed system is required to learn from insurer companies' feedback and claim outcomes and moves away from expert bias towards in Section IV-E, all the time reducing false positives and increasing the prospects of successfully challenging fraudulent claims.

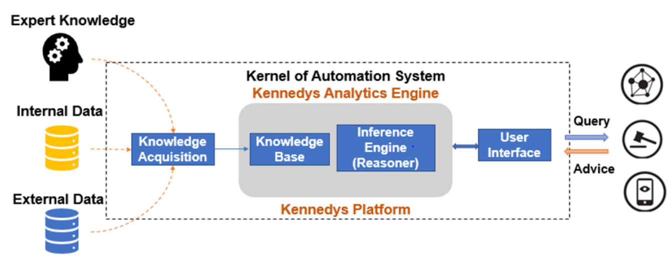

Fig. 3. Work flow of the automated automobile fraud detection system deployed in Kennedys Law LLP.

V. NUMERICAL RESULTS

In this section, a UCI benchmark dataset and a real-world example of automobile insurance data have been adopted to illustrate the utility of the proposed ER approach. A relevant study on this model, such as scalability and robustness by taking into account ambiguous data is available from [28].

A. Benchmark Data Analysis

We present here an analysis of a benchmark dataset of balloon from UCI Machine Learning Repository [29]. This balloon dataset contains 78 records of inflated or not, associated with 4 independent attributes age, act, colour, size. We have split the data into 56 training samples and 18 test samples with 10-fold cross validation. Conventional machine learning approaches, such as Logistic Regression (LR), Neural Network (NN), Support Vector Machine (SVM), Random Forest (RF), Decision Tree (DT) and Naive Bayes (NB) are compared with the proposed ER framework.

Furthermore, in order to demonstrate that our ER framework outperforms the aforementioned conventional ML approaches, two measurements are employed to evaluate the prediction accuracy for the numerical example, namely, the accuracy and the F1 score. Also, we show the results of AUC (Area Under The ROC Curve) value, which is invariant to prior probabilities or class prevalence in the data, to demonstrate the robustness of each approach.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$$
$$F_1 \text{ Score} = \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$$

where precision $= \frac{TP}{TP+FP}$ and recall $= \frac{TP}{TP+FN}$, with TP = True Positives, TN = True Negatives, FP = False Positives, and FN = False Negatives.

Table.VI reports the validation results and shows the outperformance of the proposed ER framework in terms of both accuracy and robustness.

TABLE VI
THE VALIDATION RESULTS OF DIFFERENT APPROACHES ON UCI BALLOON DATASET.

	Accuracy	F1 score	AUC
ER	**0.8333**	**0.8000**	**0.844**
ANN	0.7778	0.7500	0.756
RF	0.7222	0.7059	0.762
DT	0.6667	0.7500	0.631
LR	0.7222	0.6667	0.788
NB	0.7222	0.6667	0.775

$*$ Configuration of each algorithm: (1) ANN: R package *nnet*, with 5 hidden layers; (2) RF: R package *randomForest* with 500 trees to grow; (3) DT: R package *rpart*, with splitting index as 'information'; (4)LR: R package *VGAM*, with link as 'logit'; (5) NB: R package *e1071*.

B. Kennedys - Automobile Insurance Fraud Detection

Previously, Kennedys had employed a rule-based method, developed by a team of specialist fraud analysts, with each rule contributing to an overall score. When a score crosses a predetermined threshold, the claim is tagged as fraudulent. The ER framework instead makes use of statistical methods, to help our fraud analysts be more efficient in their work by augmenting our expert rules. The approach calculates the maxim likelihood of fraud given a set of observable features and has been developed to be as transparent as possible. At each step, the observables and their respective contributions to the overall fraud likelihood are explicit. In this section, we illustrate the efficacy of our ER framework on the real world automobile insurance data.

1) Cost-Sensitive Decision Tree: Besides the aforementioned conventional ML approaches, a new cost-sensitive decision-tree learning algorithm has been proposed in [6], which selects the splitting variable of a node, if a split is possible, based on the reduction of the total misclassification cost instead of reduction of impurity. Four methods are defined in [6] to calculate the total misclassification cost in the case of assigning the transactions of the node as fraudulent (C_P) and the total misclassification cost in the case of assigning the transactions as normal (C_N). Here we take the CS – Class Probability method as an example, the relative frequency of the classes (class probabilities) are integrated in the cost calculation functions to add the effect of the class distributions to the node costs, and hence favor the class with higher frequency in the node:

$$C_N = \left(\sum_{i=1}^{f} (C_{FN})_i\right) * \left(\frac{f}{n+f}\right)$$
$$C_P = \left(\frac{n}{n+f}\right) * n * C_{FP}$$

where, there are f fraudulent records and n normal (legitimate) records those falling into a node where $N = f + n$. Detailed explanations of the other three calculation methods are elaborated in [6].

2) Validation Results: The validation results for all methods are given in Table.VII.

TABLE VII
PERFORMANCES OF MODELS ON REAL WORLD AUTOMOBILE INSURANCE DATASET.

	Accuracy	F1 score	Sensitivity	AUC
ER	0.6390	**0.4238**	0.6957	**0.6814**
ANN	0.7884	0.1356	0.3077	0.5943
RF	0.8133	0.0426	0.0217	0.5692
LR	0.8091	0.0417	0.5000	0.6719
NB	0.1950	0.3217	1.0000	0.5000
DT	0.8091	NA	0.0000	0.5000
CS - DT	0.7759	0.3250	0.2826	0.5668

$*$ Configuration of each algorithm: same as in Tab. VI.

As can be seen from the table, the proposed ER outperforms other ML approaches in terms of F1 score and AUC value, which implies ER can obtain robust and accurate predictions independent of the decision threshold. Note that, the NB models show the best performance in terms of frauds caught or sensitivity. However, the performance of the model should be evaluated according to the misclassification costs which means that the common performance metrics such as accuracy or precision (or True Positive Rate – TPR) are not suitable to evaluate the performance of models where varying misclassification costs is in question such as this case. Moreover, cost-sensitive DT outperforms the traditional DT which indicates that for data sets with class imbalance and unequal misclassification costs, a well-defined cost-sensitive learning algorithm can improve the performance.

VI. CONCLUSION AND DISCUSSION

In this paper, we introduce a transparent intelligent automation system which can provide a decision on detecting fraudulent claims by considering complex experienced-based rules and by analysing and managing a large amount of insurance data. In this system, a rules-based inference methodology is developed by pre-defining optimal relationships between inputs and outputs through probabilistic inference and prediction. The whole system can be fine-tuned by combining expert knowledge and insurance data. This intelligent automated decision-making system not only reduces the underwriters decision cycle time but also improves the quality of decision making. Increases in data and reduction in computing costs allows us in identifying trends and patterns efficiently to help companies to improve their relationship with clients, process optimization, resource administration and increase the profits.

One potential improvement of the proposed system is to extend the ER rule based systems to a hierarchical belief rule based (BRB) system, which employs the informative belief structure to represent various types of information and knowledge with uncertainties and shows the capability of approximating any linear and nonlinear relationships across the fraud indicators of interest. Furthermore, maximum likelihood evidential reasoning (MAKER) framework can be adopted to define detailed relationships between inputs (antecedent attributes) and outputs (consequences) through probabilistic predictions, when deal with both categorical and numerical inputs described by an interval of values, see [30].

REFERENCES

[1] C. Phua, V. Lee, K. Smith, and R. Gayler, "A comprehensive survey of data mining-based fraud detection research," *arXiv preprint arXiv:1009.6119*, 2010.

[2] S. Viaene, R. A. Derrig, B. Baesens, and G. Dedene, "A comparison of state-of-the-art classification techniques for expert automobile insurance claim fraud detection," *Journal of Risk and Insurance*, vol. 69, no. 3, pp. 373–421, 2002.

[3] D. Yue, X. Wu, Y. Wang, Y. Li, and C.-H. Chu, "A review of data mining-based financial fraud detection research," in *Wireless Communications, Networking and Mobile Computing, 2007. WiCom 2007. International Conference on*. Ieee, 2007, pp. 5519–5522.

[4] S. Viaene, M. Ayuso, M. Guillen, D. Van Gheel, and G. Dedene, "Strategies for detecting fraudulent claims in the automobile insurance industry," *European Journal of Operational Research*, vol. 176, no. 1, pp. 565–583, 2007.

[5] J. H. Wilson, "An analytical approach to detecting insurance fraud using logistic regression," *Journal of Finance and Accountancy*, vol. 1, p. 1, 2009.

[6] Y. Sahin, S. Bulkan, and E. Duman, "A cost-sensitive decision tree approach for fraud detection," *Expert Systems with Applications*, vol. 40, no. 15, pp. 5916–5923, 2013.

[7] Y. Li, C. Yan, W. Liu, and M. Li, "A principle component analysis-based random forest with the potential nearest neighbor method for automobile insurance fraud identification," *Applied Soft Computing*, vol. 70, pp. 1000–1009, 2018.

[8] E. Binaghi and P. Madella, "Fuzzy dempster–shafer reasoning for rule-based classifiers," *International Journal of Intelligent Systems*, vol. 14, no. 6, pp. 559–583, 1999.

[9] A. P. Dempster, "Upper and lower probabilities induced by a multivalued mapping," in *Classic Works of the Dempster-Shafer Theory of Belief Functions*. Springer, 2008, pp. 57–72.

[10] ——, "A generalization of bayesian inference," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 30, no. 2, pp. 205–232, 1968.

[11] G. Shafer, *A mathematical theory of evidence*. Princeton university press, 1976, vol. 42.

[12] G. Shafer and J. Pearl, *Readings in uncertain reasoning*. Morgan Kaufmann Publishers Inc., 1990.

[13] J.-B. Yang, "Rule and utility based evidential reasoning approach for multiattribute decision analysis under uncertainties," *European journal of operational research*, vol. 131, no. 1, pp. 31–61, 2001.

[14] J.-B. Yang and D.-L. Xu, "On the evidential reasoning algorithm for multiple attribute decision analysis under uncertainty," *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, vol. 32, no. 3, pp. 289–304, 2002.

[15] D.-L. Xu, J.-B. Yang, and Y.-M. Wang, "The evidential reasoning approach for multi-attribute decision analysis under interval uncertainty," *European Journal of Operational Research*, vol. 174, no. 3, pp. 1914–1943, 2006.

[16] J.-B. Yang, Y.-M. Wang, D.-L. Xu, and K.-S. Chin, "The evidential reasoning approach for mada under both probabilistic and fuzzy uncertainties," *European journal of operational research*, vol. 171, no. 1, pp. 309–343, 2006.

[17] J.-B. Yang and D.-L. Xu, "Evidential reasoning rule for evidence combination," *Artificial Intelligence*, vol. 205, pp. 1–29, 2013.

[18] J.-B. Yang and M. G. Singh, "An evidential reasoning approach for multiple-attribute decision making with uncertainty," *IEEE Transactions on systems, Man, and Cybernetics*, vol. 24, no. 1, pp. 1–18, 1994.

[19] D.-L. Xu, "An introduction and survey of the evidential reasoning approach for multiple criteria decision analysis," *Annals of Operations Research*, vol. 195, no. 1, pp. 163–187, 2012.

[20] J.-B. Yang and D.-L. Xu, "A study on generalising bayesian inference to evidential reasoning," in *International conference on belief functions*. Springer, 2014, pp. 180–189.

[21] J. Ypma, "Introduction to nloptr: an r interface to nlopt," Tech. rep, Tech. Rep., 2014.

[22] G. Haixiang, L. Yijing, J. Shang, G. Mingyun, H. Yuanyue, and G. Bing, "Learning from class-imbalanced data: Review of methods and applications," *Expert Systems with Applications*, vol. 73, pp. 220–239, 2017.

[23] P. Cao, O. Zaiane, and D. Zhao, "A measure optimized cost-sensitive learning framework for imbalanced data classification," in *Biologically-Inspired Techniques for Knowledge Discovery and Data Mining*. IGI Global, 2014, pp. 48–75.

[24] A. C. Bahnsen, A. Stojanovic, D. Aouada, and B. Ottersten, "Cost sensitive credit card fraud detection using bayes minimum risk," in *2013 12th international conference on machine learning and applications*, vol. 1. IEEE, 2013, pp. 333–338.

[25] X. Xu, J. Zheng, J.-b. Yang, D.-l. Xu, and Y.-w. Chen, "Data classification using evidence reasoning rule," *Knowledge-Based Systems*, vol. 116, pp. 144–151, 2017.

[26] J.-B. Yang and D.-L. Xu, "Inferential modelling and decision making with data," in *2017 23rd International Conference on Automation and Computing (ICAC)*. IEEE, 2017, pp. 1–6.

[27] A. C. Bahnsen, D. Aouada, and B. Ottersten, "Example-dependent cost-sensitive logistic regression for credit scoring," in *2014 13th International Conference on Machine Learning and Applications*. IEEE, 2014, pp. 263–269.

[28] X. Liu, S. Sachan, J.-B. Yang, and D.-L. Xu, "Maximum likelihood evidential reasoning-based hierarchical inference with incomplete data," in *2019 25th International Conference on Automation and Computing (ICAC)*. IEEE, 2019, pp. 1–6.

[29] M. Pazzani, "UCI machine learning repository," 1991. [Online]. Available: UCI Machine Learning Repository

[30] J.-B. Yang, J. Liu, J. Wang, H.-S. Sii, and H.-W. Wang, "Belief rule-base inference methodology using the evidential reasoning approach-rimer," *IEEE Transactions on systems, Man, and Cybernetics-part A: Systems and Humans*, vol. 36, no. 2, pp. 266–285, 2006.