Auto Insurance Fraud Detection with Multimodal Learning

Jiaxi Yang[1], Kui Chen[1], Kai Ding[1], Chongning Na[1†] & Meng Wang[2]

[1]Financial Technological Research Center, Zhejiang Lab, Hangzhou 361005, China

[2]School of Computer Science and Engineering, Southeast University, Nanjing 211189, China

Keywords: Auto Insurance Multi-modal Learning; Fraud detection; Ensemble learning

Citation: Yang, J.X., Chen, K., Ding, K., et al.: Auto insurance fraud detection with multimodal learning. Data Intgelligence 5(2), 388-412 (2023). doi: 10.1162/dint_a_00191
Received: Feb. 10, 2022; Revised: Apr. 20, 2022; Accepted: July 1, 2022

ABSTRACT

In recent years, feature engineering-based machine learning models have made significant progress in auto insurance fraud detection. However, most models or systems focused only on structural data and did not utilize multi-modal data to improve fraud detection efficiency. To solve this problem, we adapt both natural language processing and computer vision techniques to our knowledge-based algorithm and construct an Auto Insurance Multi-modal Learning (AIML) framework. We then apply AIML to detect fraud behavior in auto insurance cases with data from real scenarios and conduct experiments to examine the improvement in model performance with multi-modal data compared to baseline model with structural data only. A self-designed Semi-Auto Feature Engineer (SAFE) algorithm to process auto insurance data and a visual data processing framework are embedded within AIML. Results show that AIML substantially improves the model performance in detecting fraud behavior compared to models that only use structural data.

1. INTRODUCTION

According to the insurance industry development report issued by China Insurance Regulatory Commission (CIRC) in April 2021, until the end of 2020, there are in total 235 insurance companies with total assets of 23 trillion RMB, among which the income from insurance premiums is 4.53 trillion RMB, making China the second largest insurance market across the world. Conservatively speaking, China's auto insurance fraud leakage accounts for at least 20% of the total compensation amount [1]. The estimate of China's auto

[†] Corresponding author: Chongning Na (Email: na@zhejianglab.com; ORCID: 0000-0003-2680-5774).

insurance compensation is 472.55 billion RMB in 2020, correspondingly, the loss caused by insurance fraud leakage is more than 90 billion RMB [2]. The huge amount of losses has led to great efforts spent on auto insurance fraud detection. Besides, concealed crime and gang crime also make it challenging in investigation, evidence collection and automatic identification of fraud information.

Many methods have been developed to analyze and predict insurance fraud behaviors, such as bayesian modelling, clustering analysis, data mining and random forest etc. [3, 4, 5, 6, 7, 8] Most existing models only rely on structural tabular data and are highly likely to have over-fitting issues and bad performance in real data due to sparse feature, poor label quality and missing data in other modality. Different types of data are collected in different stages of the insurance claim, such as structural data, photos of accident scenes, invoices and letters of responsibility etc, and those provide a promising means of automatically detecting auto insurance fraud with multi-modal information by using deep learning models. Extracting information from multi-modal data would provide useful anti-fraud insights for professionals in insurance industry and provide entry point in questioning high-risk cases. It also reduces the loss of insurance fraud and the cost of repeated investigation.

This paper proposes an ensemble learning method, Auto Insurance Multi-modal Learning (AIML). The system of AIML includes feature extraction from multi-modal data, feature engineering and tree-based classification. Computer vision and natural language processing models are necessary to extract factors in the form of structural data from images and texts that may be correlated with auto insurance fraud behaviors. AIML will be examined by its capability of detecting fraud behavior in real data from an auto insurance company. Our research will answer the following questions:

1. How to build AI models that could precisely predict high-risk cases?
2. How to use AI to make maximum utilization of multi-modal data that are collected during the insurance business?
3. How to use AI to extract risk factors from different types of data, will these factors be helpful in predicting insurance fraud?

Results show that AIML could extract risk factors from multi-modal data efficiently and improve the model performance to predict auto insurance fraud behavior. Compared to baseline machine learning model that only uses structural data, the ensemble model in AIML increases the AUC by 12.24% in predicting fraud behavior with multi-modal data. The rest of the paper is organized as follows. Section 2 outlines the related work of auto insurance fraud detection and the state-of-the-art methods of multi-modal data processing. Section 3 describes details of the experimental dataset and the design of our evaluation. Section 4 shows the results and model performances based on our design. Section 5 concludes and discusses possible future topics.

2. RELATED WORK

In this section, we summarize related work in two main areas: auto insurance fraud detection and multi-modal data processing methods.

2.1 Auto Insurance Fraud Detection

Insurance fraud detection can be treated as a binary classification or multiple classification problem. Many researchers have adapted machine learning models to auto insurance fraud detection and have achieved solid results. Viaene et al. [3], Kašćelan et al. [4] and Li et al. [5] examined the performance of Bayesian modelling, clustering analysis, data mining and random forest in auto insurance fraud detection. David et al. [9] achieved features and characteristics of population with high-risk in fraud behavior by analyzing the age variable of insurance holder. He et al. [6], Guo et al. [7] and Wang et al. [10] further explored the potential of deep learning models in fraud detection. Subudhi et al. [11] and Majhi et al. [12] built mixture models that could detect auto insurance fraud effectively. Tuo et al. [13] and Liu et al. [14] first discussed and studied the game theory of insurance fraud in China. Gui et al. [15] have reviewed and classified literature on moral hazard of auto insurance. Zhao et al. [16], Tang et al. [17] and Wang et al. [18] applied traditional machine learning methods to model insurance fraud behavior based on Chinese auto insurance market data. It is not until recently that Yan et al. [19, 20], Yu et al. [1] and Xu et al. [21] started to analyze insurance fraud problem with deep learning models and mixture models and made progress in the field of auto insurance detection.

Although different methods have been proposed to analyze different types of data generated from the business process of auto insurance, few multi-modal data-oriented models have been built in the field of auto insurance fraud detection. More high-risk factors await to be extracted from the multi-modal data, e.g., images, texts, to detect fraud behavior.

2.2 Multi-modal Data Processing

Multi-modal data processing has been widely adapted in the scenario of multimedia [22], disaster monitoring [23] and intelligence analysis [24]. The representative work is GAIA proposed by Li et al. [22]. The GAIA system consists of a text knowledge extraction branch and a visual knowledge extraction branch and thus enables seamless search of complex graph queries, and retrieves multimedia evidence including text, images and videos.

In the aspect of machine learning in multi-modal processing, Ngiam et al. [25] adopted the idea of shared representation learning to extend the idea of unsupervised learning of auto-encoders to the field of multi-modal learning, aiming to map data from different modalities to a uni-dimensional space. The core idea is to use noise degrading auto-encoders to represent each modality separately and then use another auto-encoder to fuse them into a multi-modal representation at the neural network fusion layer. Another method is the shared representation learning, whose idea is to project each modality into independent but constrained spaces for representation. For example, Wang et al. [26] proposed a compact hash coding method for multi-modal expression. In their work, a deep learning model is designed to generate hash-codes based on the inter-modal and intra-modal correlation constraints, and then the redundancy of hash coding features is reduced based on orthogonal regularization method. Peng et al. [27] proposed the concept of cross-media intelligence. It refers to the function of human brains across different sensory information, such as sight, hearing, language and other cognitive features of the outside world. It mainly studies the techniques and application of multi-modal learning in cross-media reasoning analysis, including fine-grained image classification, cross-media retrieval, text-generated image and video description generation, etc. Wu et al. [28] proposed a neural network that combines both visual information and text information to recognize and disambiguate entities in short texts, whose core idea is to connect visual and text information through embedding generated representation learning and to introduce a common concern mechanism for fine-grained information interaction. Experiments show that this method is superior to methods that only rely on text information.

In the aspect of knowledge engineering, a representative work is from Mousselly et al. [29], where they constructed a unified knowledge embedding based on visual features, text features and structural features of symbolic knowledge. Compared with traditional structure-based knowledge graph representation learning, their performances in link prediction and entity classification tasks were improved. Xie et al. [30] later proposed an improved model IKRL, whose core idea is to conduct joint modeling of visual features and structural features of knowledge graph, so as to generate multi-modal knowledge graph embedding with higher quality through connections between different types of modality. Chen et al. [31] explored how to effectively jointly mapping and modeling cross-modal semantic information in the knowledge graph, thus laying an important foundation for supporting intelligent application services for multi-modal content. Guo et al. [32] further explored the entity alignment task of multi-modal knowledge graph, which mainly extended the multi-modal entity alignment task from Euclidean space to hyperbolic space.

Since there are many relatively mature algorithms for each type of data, digging more information and factors from both text data and visual data in the scenario of auto insurance is practical and promising.

3. FRAMEWORK

In this section, the multi-modal insurance fraud detection framework of AIML is explained in detail. Overall, our framework includes three modules as shown in the Figure 1.

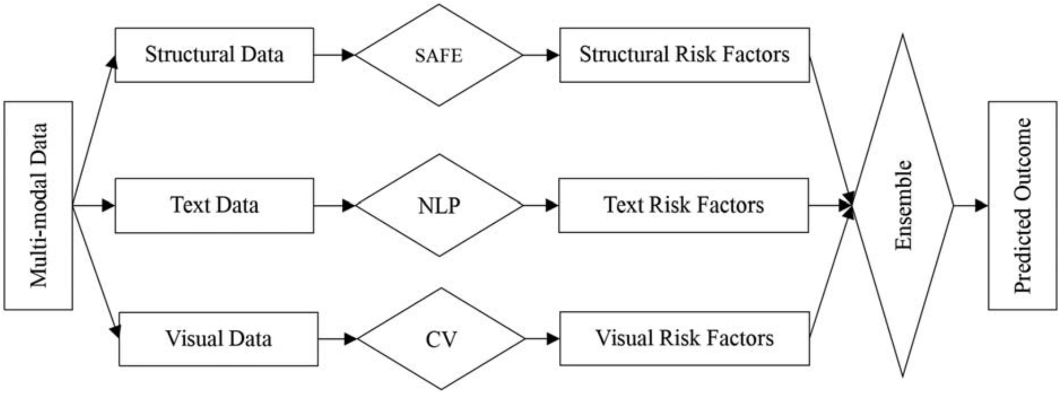

Figure 1. AIML workflow.

Structural data will be cleaned and processed by a feature engineering model to extract and generate risk factors for fraud behavior. Both text and visual data will be processed by systems that are embedded with Natural Language Processing (NLP) models and Computer Vision (CV) algorithms to extract risk factors correspondingly. Finally, the ensemble factors will be assigned to a machine learning model to predict fraud behavior.

3.1 Structural Data and Baseline Model

The workflow of baseline model in AIML is:

1. Data are collected based on cases and stages from insurance companies, including case reporting stage, investigating stage and loss verification stage (All data are labeled and verified by experts and professionals from insurance companies).
2. Collected data are then cleaned and pre-processed, i.e., cases with more than 50% missing information will be removed, categorical variables will be one-hot encoded.
3. New features are generated with feature engineering algorithms from original features.
4. New features are fed to a machine learning model to achieve predicted outcomes.

During the predicting process, feature engineering is an essential part in the process of predicting problems for real case scenarios. It is divided into feature classification and feature derivation, among which feature classification refers to the classification of original features based on their distributions; feature derivation refers to feature synthesis based on classified features in order to obtain richer feature combinations. After comparing multiple popular machine learning and deep learning methods, AIML uses the combination of Semi-Auto Feature Engineering (SAFE) for automated feature engineering, which is a self-designed and semi-automatic method for feature engineering, and eXtreme Gradient Boosting tree (XGB) [33] to predict whether the case is fraud or not.

3.2 Unstructured Text Data Processing

Extracting risk factors from auto insurance case description texts is treated as NLP text mining tasks. There are in total six text data mining tasks in AIML, i.e., recognizing driving status, type of accident, type of roads, cause of accident, number of cars and parties involved in the accident.

Table 1 illustrates the key information we extracted from the unstructured text data:

Table 1. Example of text data.

Descriptions	Num. Of Cars	Cause of Accident	Driving Status	Type of Accident	Type of Roads	Parties involved
标的车与三者车高速公路行驶相撞，两车受损	双车事故	疏忽	行驶状态	撞伤	高速公路	车/车
Insured car crashed into a third-party car when driving on the highway. Both cars were damaged.	Two cars accident	Negligence	Driving	Crashed	Highway	Car/Car
标的车与障碍物高速公路行驶相撞，本车受损	单车事故	疏忽	行驶状态	撞伤	高速公路	车/障碍物
Insured car crashed into an obstacle when driving on the highway. Insured car was damaged.	Single car accident	Negligence	Driving	Crashed	Highway	Car/Object

AIML uses multi-task classification framework to achieve the goal of risk factor mining, wherein a common backbone representative learning model is shared by the six test mining tasks. The advantage of multi-task learning is that it could reduce computational complexity and cost of training, while taking into account different levels of correlation between tasks. Specifically, feature extraction layer is fully shared, based on Bidirectional Encoder Representations from Transformers (BERT) pre-trained model, combined with multi-task loss linear fusion fine tuning and Conditional Random Fields (CRF) method to achieve multi-task learning. The multi-task model is shown in Figure 2, including input layer, encoding layer, fully connection layer (FC layer), activation layer, CRF layer and output layer.

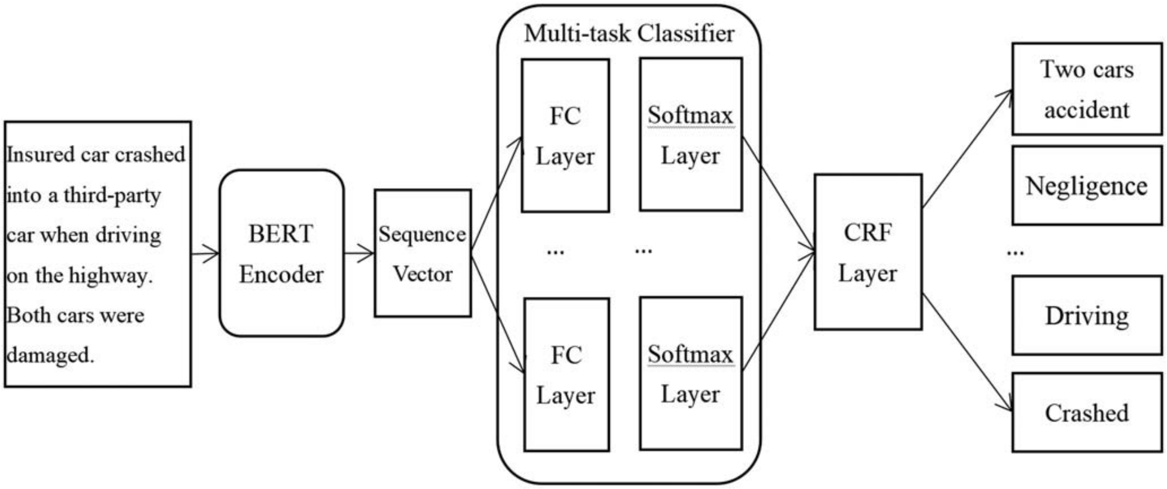

Figure 2. Unstructured text data processing.

First, AIML treats the description of accident as input and chooses BERT (Chinese Version) as pre-trained encoder. BERT could dynamically represent the meaning and relevance of characters in context by using powerful multi-directional self-attention mechanism combined with self-supervised learning, so as to construct a vector that represents the semantic feature of the whole sentence after weighted combination. Additionally, BERT pre-trained model uses massive data including Wikipedia and other knowledge as training corpus to ensure its applicability to insurance text. BERT could still achieve a rather nice classification accuracy by adding a full connection layer to the output layer, even without fine-tuning of model parameters.

Second, AIML uses multiple classifiers to extract multi-event factors, taking the sequence vector output by BERT as their inputs. A cost function is defined for each classification task and each task was considered independent to each other. Parameters of the newly added FC layer and BERT sequence output layer are tuned by multi-task loss linear fusion method.

Finally, based on the correlation between computing tasks, AIML uses CRF to calculate the maximum joint probability for multiple classification results. CRF is most commonly used in the field of sequential annotation in NLP, using joint probability to calculate the co-occurrence relationship between text and annotation to optimize the overall accuracy of sequential annotation. Here we use a similar mechanism to optimize the overall accuracy of multi-task prediction with a CRF layer. The original CRF must satisfy two prerequisites: Exponentially distributed and. Only adjacent elements are correlated. The input for CRF is the output sequence vector for multi-classification task, presented as:

$$P(Y|X) = P(y_i, y_{i-1}, \ldots, y_0, X) = \frac{1}{Z(x)} \exp\left(f\left(y_i, y_{i-1}, \ldots, y_0, X\right)\right) \tag{1}$$

$$\left(y_i, y_{i-1}, \ldots, y_0, X\right) = h\left(y_i, X\right) + g\left(y_i, y_{i-1}\right) + \ldots + h\left(y_0, X\right) + g\left(y_1, y_0\right) \tag{2}$$

where P indicates probability function, Z indicates normalization factor, h indicates the mapping function between single output and global input, g indicates the function for local correlation between output elements, y indicates the single output element and X indicates the global input.

3.3 Visual Data and Processing

Based on the scenarios of auto insurance fraud detection, this paper mainly focuses on three techniques, namely, Object Detection, Optical Character Recognition (OCR) and Pedestrian Re-identification (ReID). We design a systematic approach for AIML, as shown in Figure 3, to process and extract risk factors from visual data, i.e., photos and pictures of car accidents.

Raw visual data are stored in folders with case ID as folder names. The first step is to classify pictures into seven categories, i.e., accident scene, car components, invoices, driver license, driving license, photos of inspectors and cars and others. A ResNet classification model is trained on 413 cases with 1,392 well labeled pictures. Then AIML adapts the trained model to a much larger test set with 22,385 pictures to make a rough classification as those 22,385 pictures were originally unlabeled. Flowing a manually fine classification, all pictures with correct categories are used to re-train the classification model.

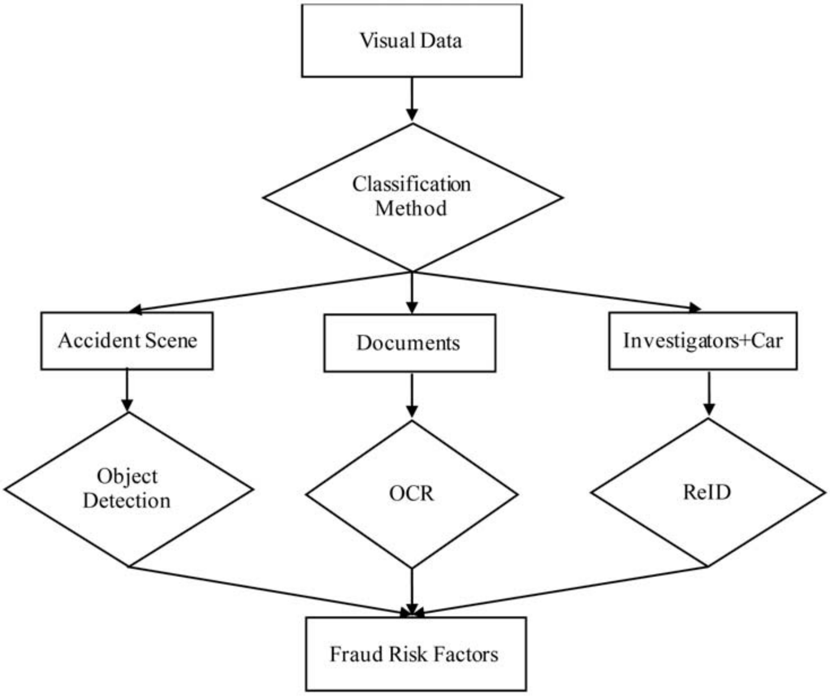

Figure 3. Visual data processing.

The second step is to extract risk factors from each category of pictures. For pictures in categories of accident scene and car components, a Yolov5 model is used to extract risk factors from photos to identify damage conditions of cars and a ResNet model is used to extract scene information such as daytime or nighttime. For pictures that contain text information, AIML uses OCR to recognize information from licenses and invoices. For pictures that contains both investigators and cars, AIML uses ReID to identify different investigators and check anomalies, i.e., if they appeared in previously detected fraud cases or if they appear in multiple cases.

In the last step, factors extracted from visual data will be merged with structural data by case ID to improve model performance.

4. RESULTS

In this section, we report our experimental results on a real-world auto insurance dataset. The results of baseline model will be firstly presented. Then, risk factors extracted from text data and visual data will be added and show the effectiveness of multi-modal learning in improving the fraud detection capability.

4.1 Dataset

Experimental data are collected and resampled from 4,946 auto insurance cases from November 10, 2014, to October 26, 2020, among which 3,613 are non-fraud cases and 1,333 are confirmed fraud cases. Data are organized in a per-car basis, i.e., cases containing multiple car accidents are treated as multiple data samples, indicated by a compound Case Unique ID (CaseUID, including both case ID and car plate). Therefore, number of samples in the entire dataset is slightly larger than the number of cases, i.e., including 5,034 non-fraud Case Unique IDs and 1,413 fraud Case Unique IDs respectively. There are in total 216 fields of data containing information collected from the case reporting stage, investigating stage and loss verification stage. Variables with over 70% missing information will be excluded; variables with information that are not suitable for fitting into XGB model will be excluded (e.g. ID-type variable, names etc); only structural data, i.e., mainly categorical and numerical data are used in the baseline model.

4.2 Results of Baseline Models

After the original variables are pre-processed by SAFE, i.e., our self-designed feature selection and feature interaction tool, there are in total 1,155 features, which are generated from the original 216 variables combined, one-hot encoded, interacted, added and subtracted according to their types. One special Boolean feature named 'Compensation Type_Normal Case' is excluded, because it is a huge giveaway in predicting fraud cases. In order to evaluate the performance of model comprehensively, four criteria, precision, recall, F_1-score and Area Under the Curve (AUC) will be used to evaluate model performance.

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} \tag{3}$$

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative} \tag{4}$$

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall} \tag{5}$$

All 6,447 subjects were randomly separated into train and test set with the ratio of 80%/20% and all 1,154 features are fed to the XGB model. The trained model has an overall accuracy of 0.8364 with precision equals to 0.7095, recall equals to 0.4441 and F_1 score as 0.5462. The plots of the ROC and PR curves are presented in Figure 4 and Figure 5 respectively.

Based on the results for baseline model, the model performance is rather moderate in predicting fraud behavior in auto insurance cases.

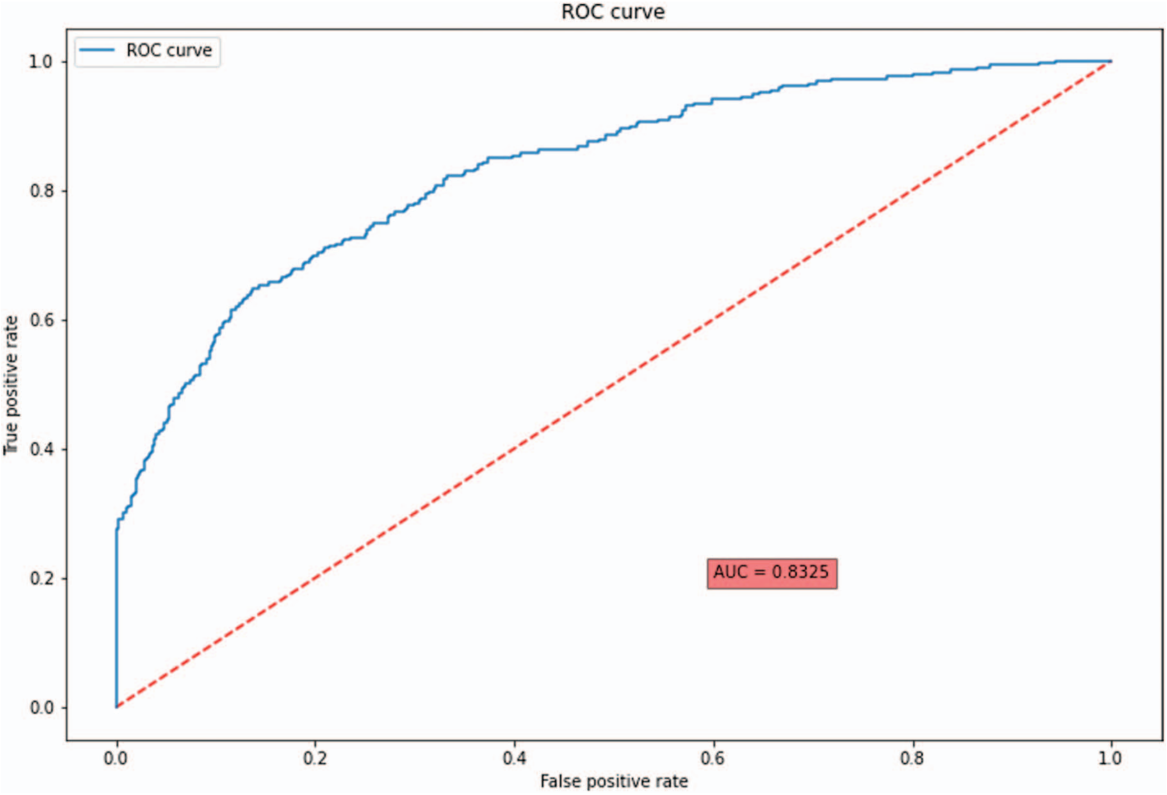

Figure 4. ROC curve for baseline model.

Table 2. Feature importance baseline model.

Ranks	Extracted factors from text	Feature importance
1	Type of Cases	0.136462
2	Type of Compensation	0.068291
3	Gender of Driver Unknown (T/F)	0.025964
4	Case Only Contain Automobile Damage Insurance	0.022832
5	Case Contain 4S Store (T/F)	0.014042
6	Third Party Insurance	0.012947
7	Province A	0.011575
8	Valid Date for Auto Insurance Cases	0.010736
9	Time Length for Damage Assessment	0.009672
10	Case Contain Automobile Damage Insurance (T/F)	0.009474
11	Province B	0.009463
12	Province C	0.008983
13	Object Damage (T/F)	0.008516
14	Unknown Auto Insurance Type (T/F)	0.007255
15	Province D	0.007206

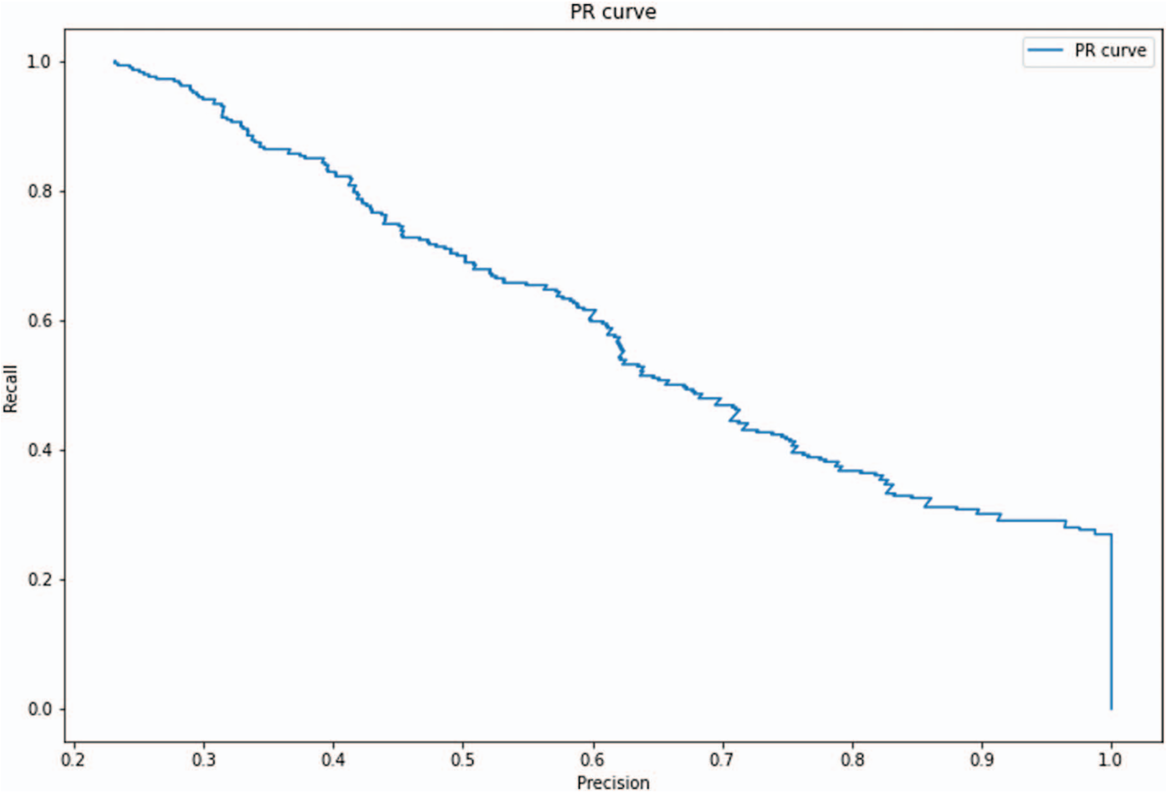

Figure 5. PR curve for baseline model.

4.3 Results of Unstructured Text Data Processing

In order to extract information that is relevant to fraud detection from text data, we formulated five text classification tasks, wherein each one is a multi-class classification task. For each task, we defined text labels, i.e., 12 types of accidents, types of driving status, 11 types of cause of accident, 4 types of car numbers and 5 types of roads. We manually labeled each accident description text with those five types of labels. To simplify the effort of the labeling work. we firstly selected 750 relatively uncorrelated samples and labeled them manually. The uncorrelation is achieved by clustering the texts and select text in different clusters. Then for each type of a label within as single task, we ensure at least 35 samples from those 750 labeled data samples. Afterall, we achieved a small data set to train a coarse classfier for each of the five tasks. Then the coarse classifier is used to categorize all text samples. Incorrect categorization results a manually adjusted. Final classification results for those five tasks are shown in Table 3 below.

Table 3. Precision in categorizing text data with BERT.

Criterion	Driving status	Type of accident	Type of roads	Cause of accident	Number of cars
F_1-Score	0.93	0.84	0.79	0.85	0.94

Five new features were generated from text data. After one-hot encoding, there are 45 new boolean factors. The trained XGB model with factors extracted from text has overall accuracy of 0.8481 with precision equals to 0.7473, recall equals to 0.4755 and F_1 score as 0.5812.

According to Table 4, although the number 45, i.e., the number of features extracted and derived from text data is relatively small compared to the original number of features, i.e., 1,154. There is significant improvement in model performance. Both recall and F_1 score increase by around 6–7%. The plots for ROC and PR curves are presented in Figure 6 and Figure 7 respectively.

Table 4. Model performance for baseline model and model with text factors.

Criterion	Accuracy	Precision	Recall	F1-Score	AUC
Baseline Model	0.8364	0.7095	0.4441	0.5462	0.8325
Model with Text Factors	0.8481	0.7473	0.4755	0.5812	0.841
Increase	1.40%	5.33%	7.07%	6.41%	1.02%

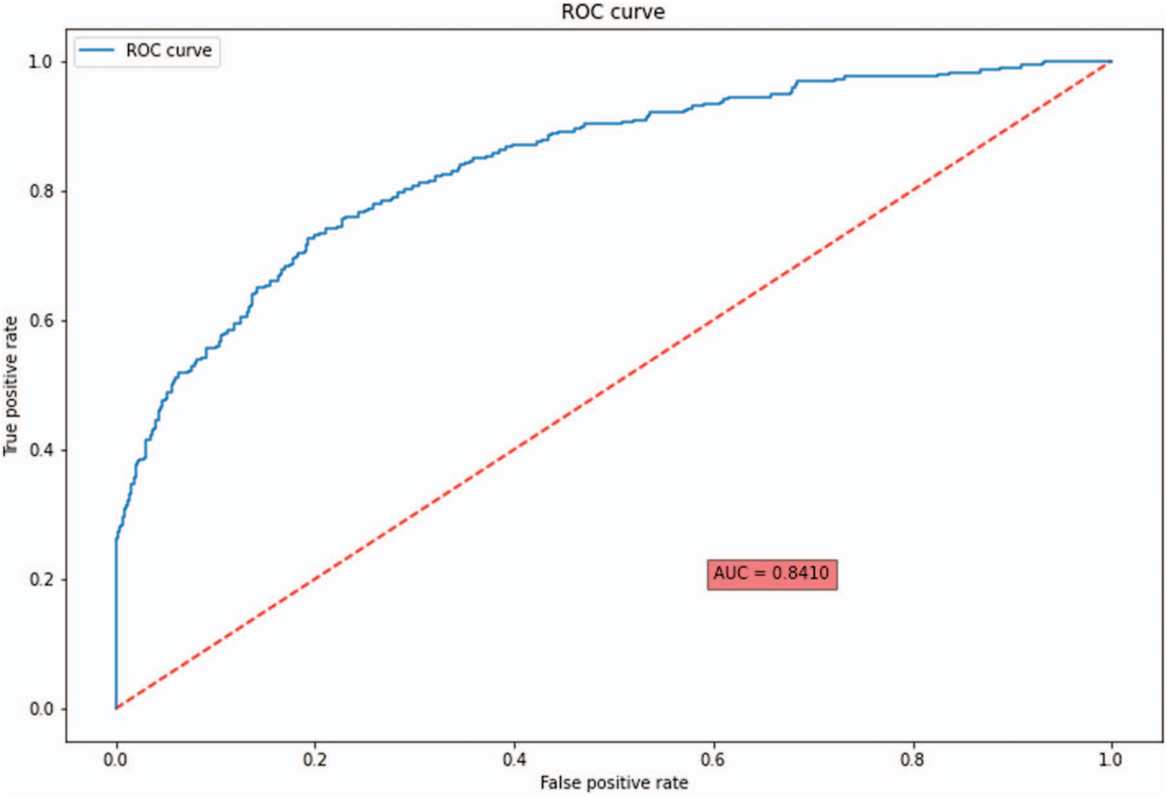

Figure 6. ROC curve for model with text factors.

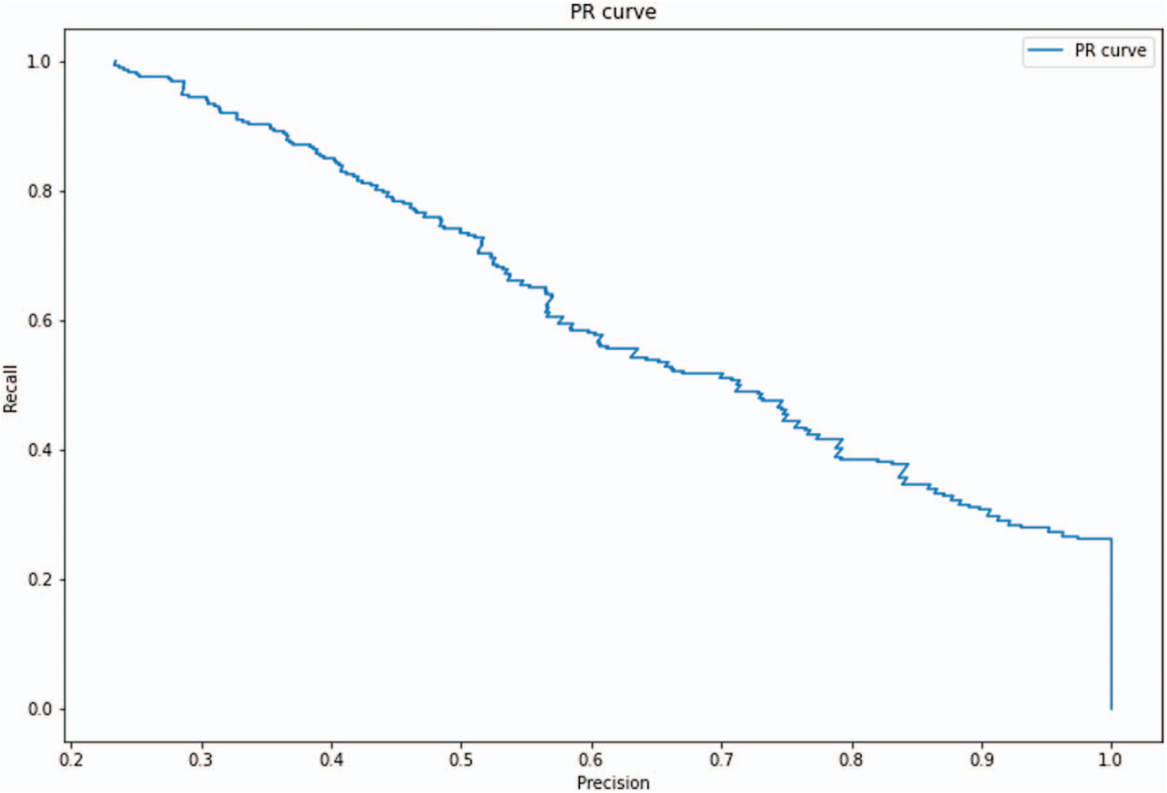

Figure 7. PR Curve for fraud detection with text factors.

The feature importance of partial extracted factors from text data is listed in Table 5, along with their rankings.

Table 5. Feature importance for text factors.

Ranks	Extracted factors from text	Feature importance
24	Single_Car_Accident	0.005487
30	Negligence	0.005257
48	Transportation_Facility	0.004256
53	Other_Minicars	0.004139
61	Rear_End	0.003830
73	Auxiliary_Buildings	0.003349
95	Third-party_Responsibility	0.002961

4.4 Results of Visual Data Encoding

As mentioned in Section 3.3, the first task of visual data mining is the categorization of the raw data. The accuracy of the automatic multi-label classification algorithm is listed in Table 6 for each category, wherein most categories are well classified.

Table 6. Categories of visual data.

Category	Number of pictures	Misclassified pictures	Accuracy rate
Accident scene	2,821	333	88.20%
Car component	17,451	276	98.42%
Invoice	309	12	96.12%
Driver license	426	1	99.77%
Driving license	866	30	96.54%
Inspector + Car	500	9	98.2%
Other	12	0	100%

Then risk factor extraction was carried out according to the scheme described in Section 3.3. Some extraction accuracy results are presented in Table 7, wherein car parts and damage detection accuracy are relatively low due to the imbalance issue in corresponding data

Table 7. Overall accuracy for CV tasks.

Task	Model	Overall accuracy	Categories	Training pictures
Classification	ResNet	97.00%	7 Categories	22,385
Re-identification	ReID	83.23%	79 persons	453
Invoice recognition	OCR	86.58%	8 Features	65
Driver license Recognition	OCR	80.67%	9 Features	403
Driving license Recognition	OCR	76.37%	8 Features	436
Day/Night recognition	ResNet	99.64%	Day or Night	3,321
Scene recognition	ResNet	62.67%	5 Categories	3,321
Car plate recognition	Yolov5/LPRNet	84.76%	NA	3,983
Car parts detection	Yolov5	62.39%	8 Parts	8,059
Damage detection	Yolov5	21.30%	5 Categories	11,023

Detailed risk factors extracted from visual data are listed in Table 8.

Table 8. Visual risk factors descriptions.

High-risk Factors	Extracted Factors from Pictures	Algorithms	Processing
Correlation between investigators	ReID 0/1	ReID	Recognition and matching between multiple face images
Cost	Cost of repair	OCR	Recognition of invoice
Car brand	Car brand		Recognition of driving license
License type	Car type		Recognition of driver license
Stone	Boolean 0/1	Yolov5	Object Detection
Cars involved	Car count		Object Detection
Recognition of damage	Scratch, break, deformation etc.		Object Detection
Location of damage	Nearest car parts to damage spot		Object Detection
Daytime/Nighttime	Boolean 0/1	ResNet	Environmental identification
Road condition	Categorical variable		Environmental identification

All factors are designed and defined based on previous expert knowledge and reports in detecting fraud cases. However, due to the quality of pictures, only 10 variables are relatively complete (less than 30% missing) and were extracted from documentary pictures. After one-hot encoding for categorical variables, there are 29 new features from visual data. The trained XGB model with factors extracted from text has overall accuracy of 0.8736 with precision equals to 0.724, recall equals to 0.6107 and F_1 score as 0.6625.

According to Table 9, we can see that there is a significant improvement in model performance with these visual features. Both recall and F1-score increase by over 20% which may be because visual data contain key information that is not included in structural data. The plots for ROC and PR curves are presented in Figure 8 and Figure 9 respectively.

Table 9. Model performance for baseline model and model with visual factors.

Criterion	Accuracy	Precision	Recall	F_1-score	AUC
Baseline model	0.8364	0.7095	0.4441	0.5462	0.8325
Model with visual factors	0.8837	0.7456	0.6489	0.6939	0.9288
Increase	5.66%	5.09%	46.12%	27.04%	11.57%

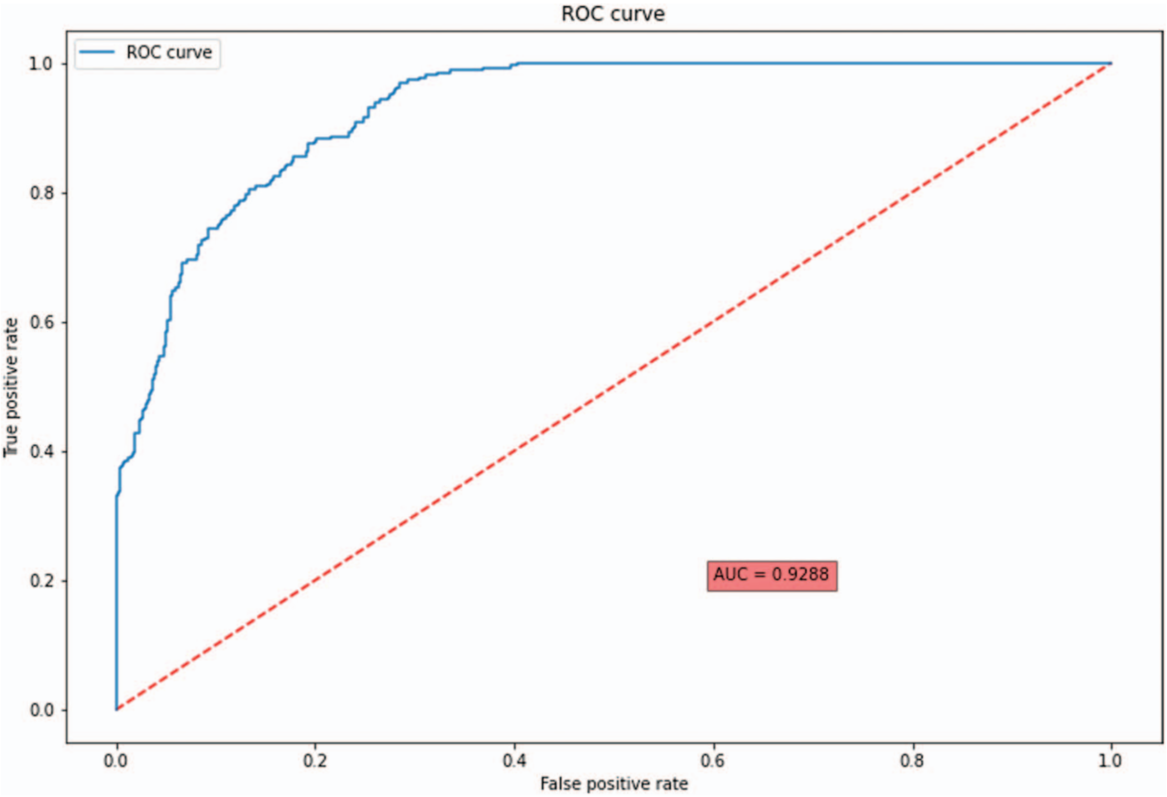

Figure 8. ROC curve for fraud detection model with visual factors.

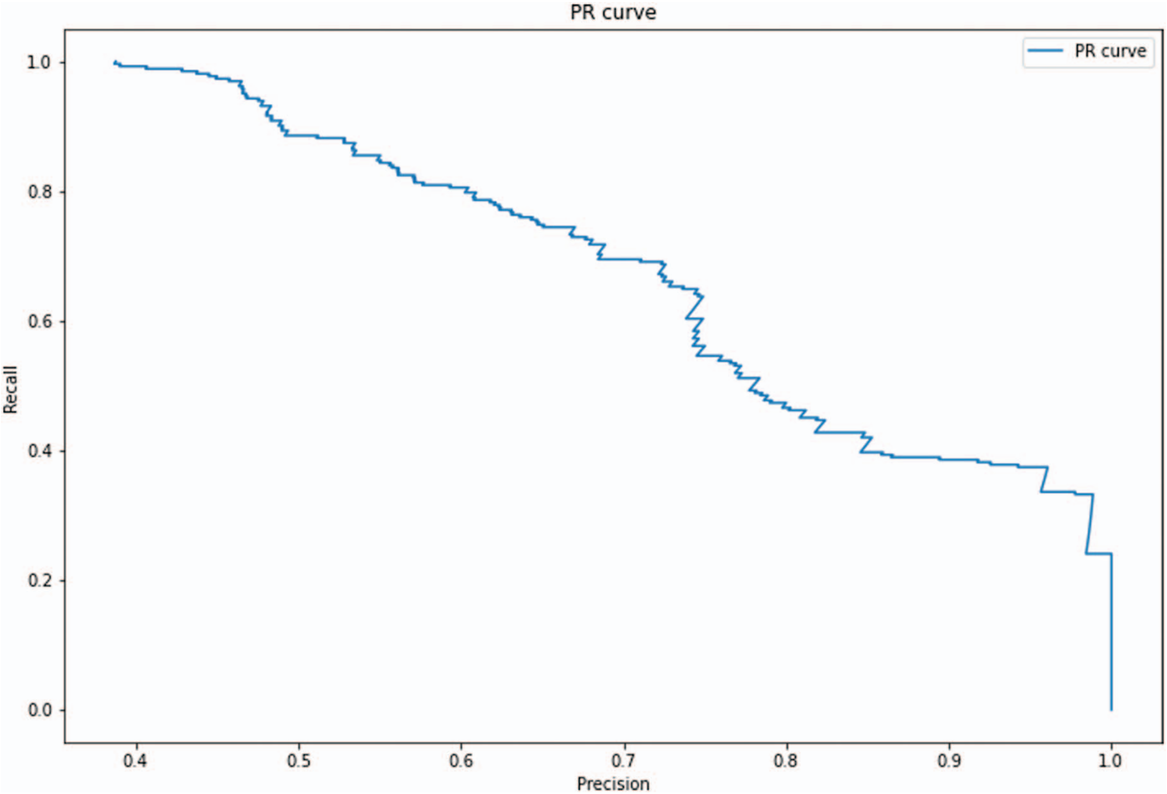

Figure 9. PR curve for fraud detection model with visual factors.

In order to be more specific, some anonymised visual data are shown in Figure 10 to Figure 14.

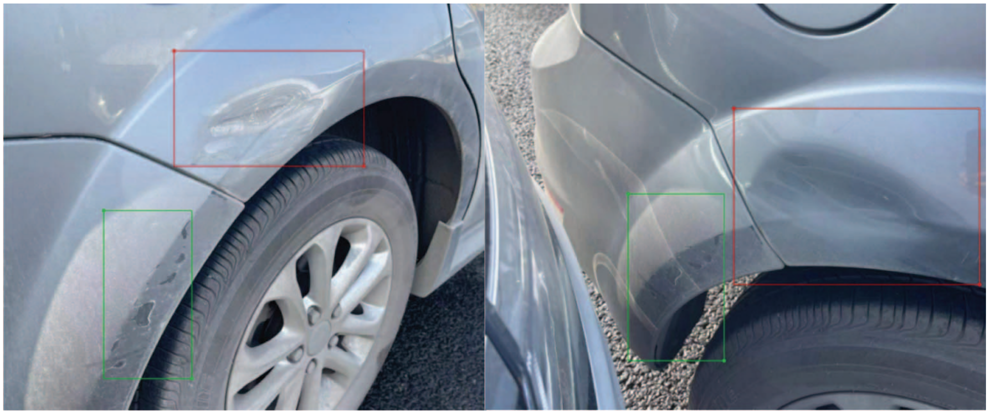

Figure 10. Car component pictures.

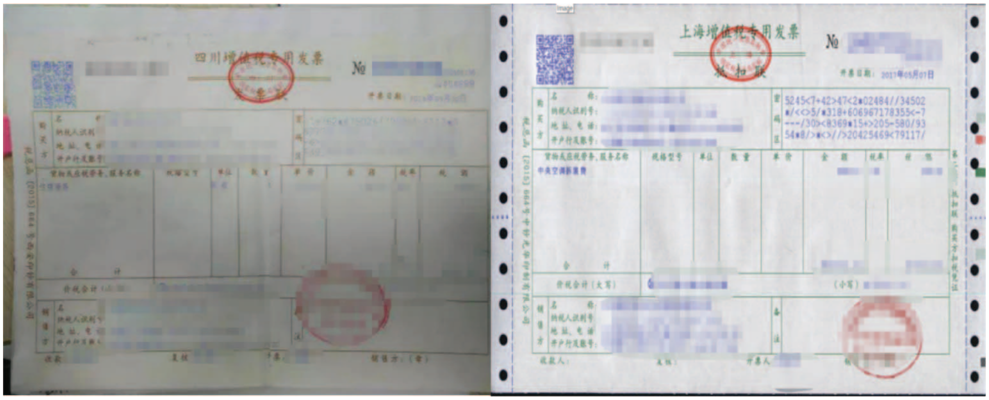
四川增值税专用发票
上海增值税专用发票
№

Figure 11. Invoice pictures.

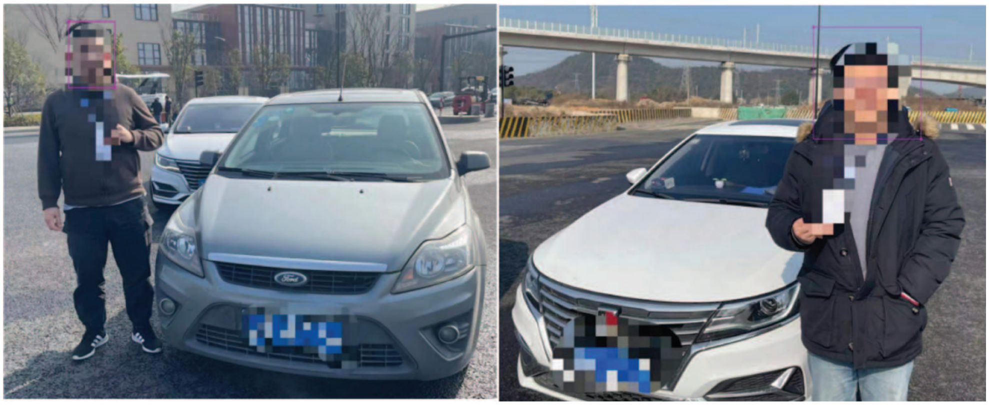

Figure 12. Inspectors + Cars pictures.

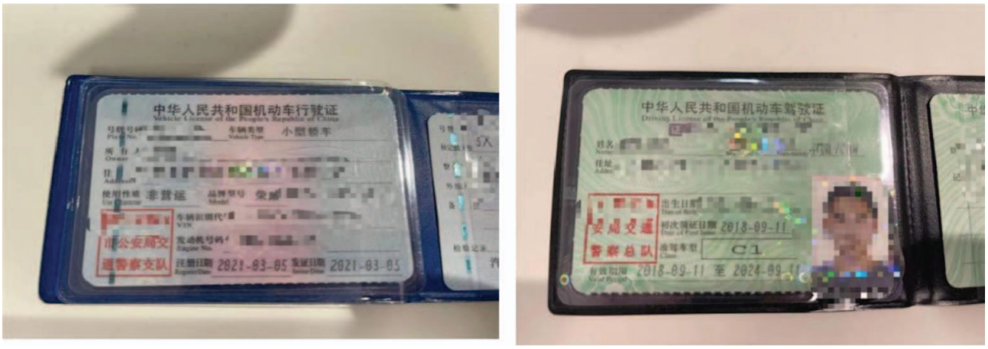
中华人民共和国机动车行驶证
Vehicle License of the People's Republic of China
小型轿车
非营运
2021-03-05
2021-03-05
中华人民共和国机动车驾驶证
Driving License of the People's Republic of China
2018-09-11
C1
2018-09-11 至 2024-09-11

Figure 13. Driver and driving license pictures.

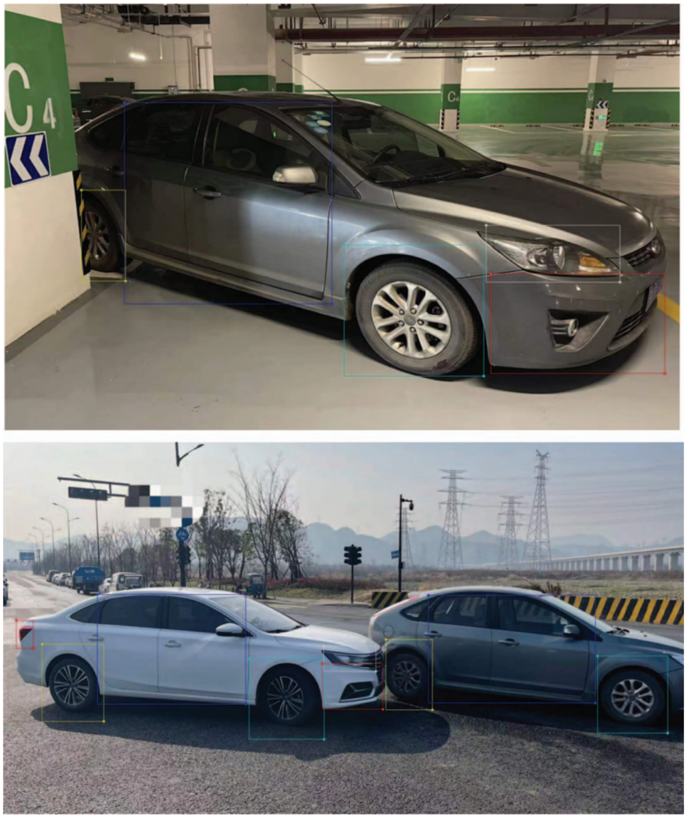

Figure 14. Accident scene pictures.

The rectangle annotation marks different parts of cars, damage on cars and inspectors hired by insurance companies, which will be then converted to structural features as risk factors from visual data.

4.5 Results for Ensemble Model

Finally, we combine the high-risk factors extracted from both text data and visual data to our baseline model in order to check the improvement of model performance brought by multi-modal data.

Table 10. Model performance for ensemble model.

Criterion	Accuracy	Precision	Recall	F_1-Score	AUC
Baseline model	0.8364	0.7095	0.4441	0.5462	0.8325
Model with text factors	0.8481	0.7473	0.4755	0.5812	0.841
Model with visual factors	0.8837	0.7456	0.6489	0.6939	0.9288
Ensemble model	0.8713	0.7143	0.6107	0.6584	0.9344
Overall increase compared to baseline model	4.17%	0.68%	37.51%	20.54%	12.24%

The results in Table 10 show that there is a substantial increase in model performance after adding factors extracted from multi-modal data in auto insurance cases. Compared to baseline model, the performance increases by 12.24% in AUC after adding 45 text features and 29 visual features. The ROC and PR curves for ensemble model are presented in Figure 15 and Figure 16 respectively.

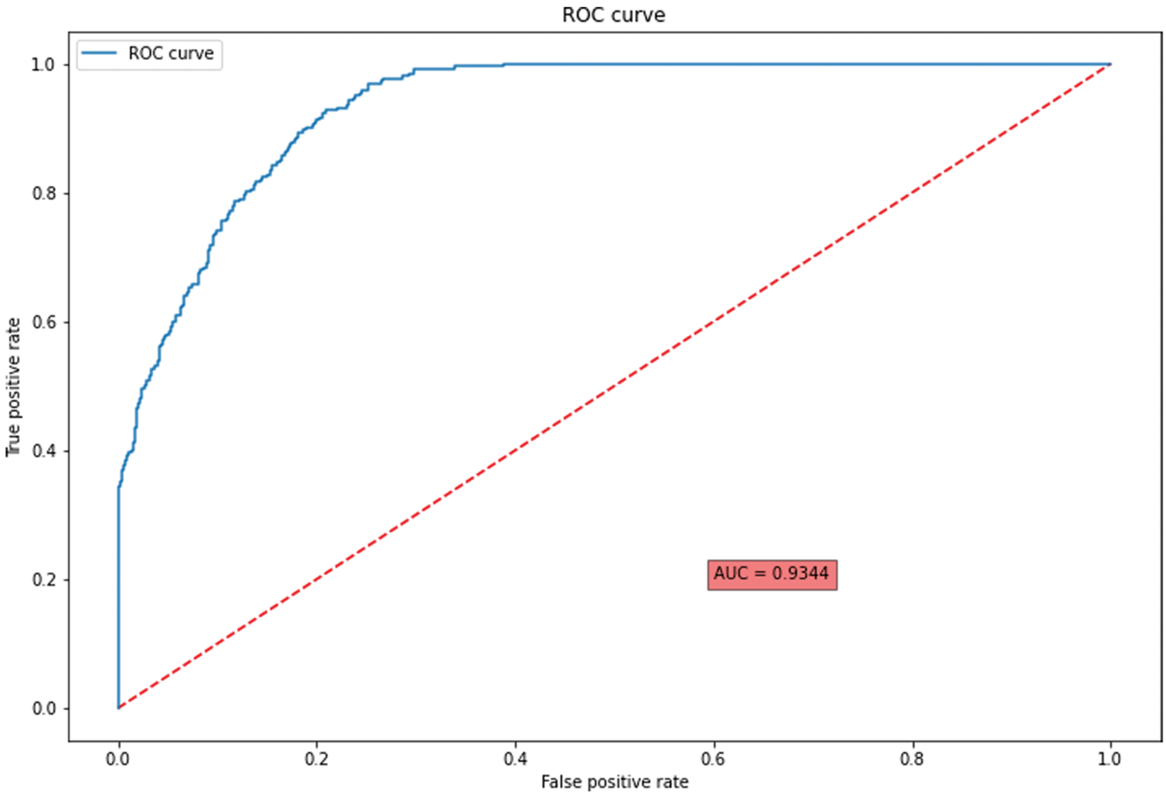

Figure 15. ROC curve for ensemble model.

4.6 Limitation Analysis

Although we have achieved rather nice model performance by adding factors extracted from the multi-modal data, we still observe some limitations of the current scheme. Firstly, categorization of text data is extremely imbalanced. For example, main causes of accidents are driver's fault and third-party's responsibility, while other causes, e.g., bad weather, are not adequately present in current dataset. Additionally, the consequences caused by driver's fault are also imbalanced and varied. It is easy to be misclassified when the consequence of one accident is semantically close to another. Examples are shown below.

1) When driving through the water section, vehicle flameout. Car damaged. ---Single car accident ---Flooding ---Driving
行驶到积水路段，车辆熄火。本车有损；---单车事故---水淹---行驶状态

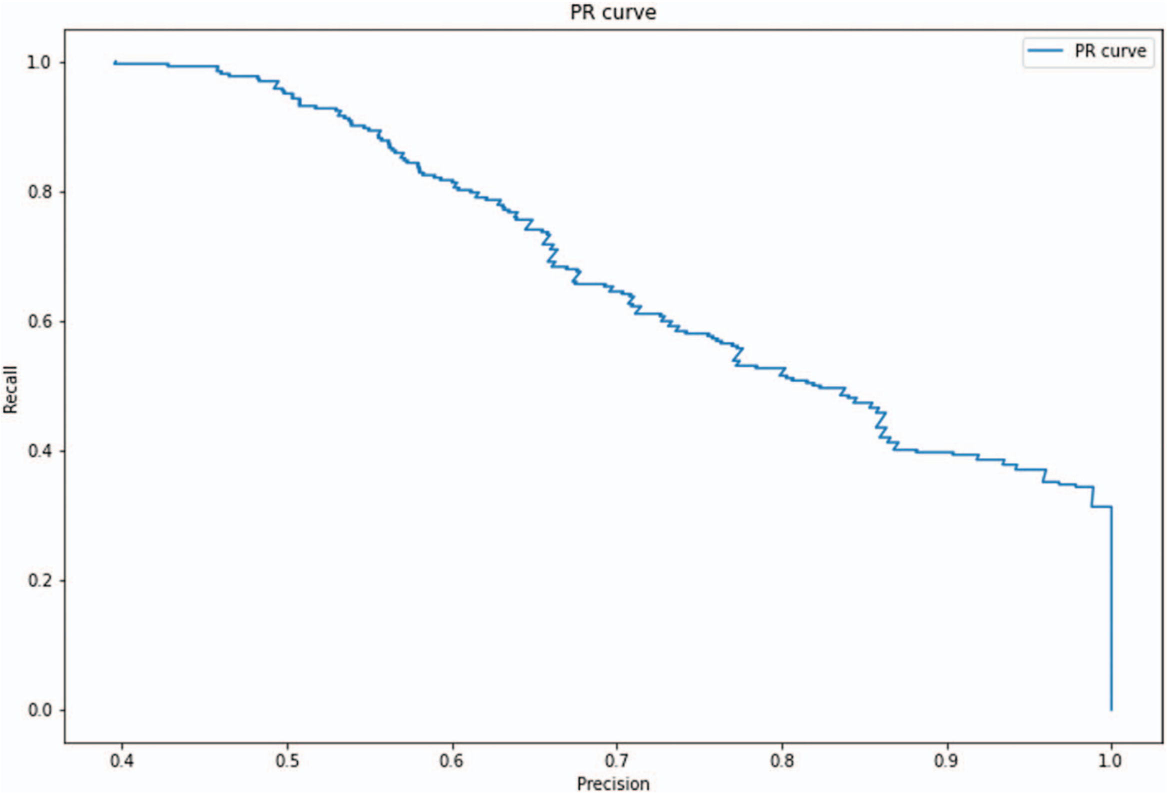

Figure 16. PR curve for ensemble model.

2). Heavy rain flooded car in the parking lot. Car damaged. ---Single car accident ---Flooding ---Parking
暴雨积水，车辆在停车场被淹。本车有损；---单车事故---水淹---停放

The cause of the first accident should be driver's fault while the cause of the second accident should be bad weather. However, due to the imbalanced number of samples relevant to those two different causes in the training data, the model may easily misclassify the cause of first case as weather.

Also, due to the limitation of BERT (mainly refer to its adaptation of the specific context of car accident scenarios), semantic ambiguity or limited data samples, the causal relationship or the sequence of multiple elements in one sentence cannot be identified clearly. There are still a big portion of text data left unused. For example, the wounded information for drivers or people involved in the accident, the description of injuries from doctors' notes and traffic police report, etc.

Due to the poor quality for many visual data, only 10 variables were extracted from visual data with satisfactory accuracy. Pictures for some cases cannot be detected or recognized and thus lead to lots of missing data, which will bring data leakage issue to some extent consequently. Because of the small quantity and bias issue, the performance of damage detection model are limited as well. More visual data are needed to train the fine-grained images or parts. Additionally, there should be a better way to annotate the visual data. Rectangle annotation is relatively rough when marking tiny or irregular damage. Semantic segmentation is worth trying for the next step research.

6. CONCLUSION AND FUTURE WORK

In this paper, we ensemble a structural data feature engineering algorithm, a natural language processing model and a processing framework for visual data together with a machine learning model to handle the task of auto insurance fraud detection based on multi-modal data. We first design an auto insurance multi-modal learning (AIML) framework to analyze multi-modal data collected during the auto insurance business. With AIML, we can utilize multi-modal data efficiently and improve the model performance to predict auto insurance fraud behavior. We also design a text mining algorithm and a framework to process visual data. Both of them have achieved significant improvements in predicting fraud behavior. Experimental results show the high quality of AIML, and the effectiveness of applying AIML to auto insurance fraud detection on multi-modal data.

As we have achieved substantial increase in model performance based on multi-modal data mining with real-world dataset, constructing a real-time system or pipeline will be an appealing topic for the next step to introduce multi-modal data mining in auto insurance industry. One possible challenge could be multi-modal big data. As the amount of data increases, there will exist a bottleneck for each branch that processes different types of data. Potential solution may consider distributed system with load balance between algorithms handling different types of data, e.g., NLP for text data and CV for visual data. Considering the potential of further performance improvement, one may consider using knowledge graph to connect and represent multi-modal data in a more structured way.

ACKNOWLEDGEMENTS

This research was supported by “Research on intelligent Computing technology in Financial Risk Control and Anti-fraud”, funding code 2020NF0AC01, Zhejiang Lab, leaded by Dr. Chongning Na. We thank our colleagues from Zhejiang Lab who provided insight and expertise that greatly assisted the research.

We would also like to show our gratitude to the auto insurance companies for sharing their pearls of wisdom and data with us during the course of this research.

AUTHOR CONTRIBUTIONS

C.N. Na (ORCID: 0000-0003-2680-5774, na@zhejianglab.com) and J.X. Yang (ORCID: 0000-0002-5055-8729, jiaxiyang@zhejianglab.com) conceived of the presented idea and designed the framework of AIML system. J.X. Yang wrote the manuscript with the help of K. Chen (ORCID: 0000-0002-7925-9968, chenkui@zhejianglab.com), K. Ding (ORCID: 0000-0003-4534-2904, dingkaid@zhejianglab.com) and

M. Wang (ORCID: 0000-0002-2293-1709, meng.wang@seu.edu.cn). J.X. Yang, K. Chen and K. Ding designed and carried out the experiment. C.N. Na and M. Wang proofread and edited the manuscript. All authors provided critical feedback and helped shape the research and analysis.

REFERENCES

[1] Yu, W., Feng, G., Zhang, W.: A Research on Fraud Detection System and Gang Identification of Vehicle Insurance. Insurance Studies (2), 63–73 (2017)

[2] The Joint Research Team on Anti-Vehicle Frauds. A Research on Vehicle Insurance Frauds and Anti-fraud Issues and Regulatory Suggestions. Insurance Studies (06), 3–10 (2021)

[3] Viaene, S., Dedene, G., Derrig, R.A.: Auto claim fraud detection using Bayesian learning neural networks. Expert Systems with Applications 29(3), 653–666 (2005)

[4] Kašćelan, L., Kašćelan, V., Novović-Burić, M.: A Data Mining Approach for Risk Assessment in Car Insurance: Evidence from Montenegro. International Journal of Business Intelligence Research (IJBIR) 5(3), 11–28 (2014)

[5] Li, Y., Yan, C., Liu, W., Li, M.: A principle component analysis-based random forest with the potential nearest neighbor method for automobile insurance fraud identification. Applied Soft Computing 70, 1000–1009 (2018)

[6] He, X., Chua, T.S.: Neural factorization machines for sparse predictive analytics. In Proceedings of the 40th International ACM SIGIR conference on Research and Development in Information Retrieval, pp. 355–364 (2017, August)

[7] Guo, J., Liu, G., Zuo, Y., Wu, J.: Learning sequential behavior representations for fraud detection. In 2018 IEEE international conference on data mining (ICDM), pp. 127–136 (2018, November)

[8] Wang, R., Fu, B., Fu, G., Wang, M.: Deep & cross network for ad click predictions. In Proceedings of the ADKDD'17, pp. 1–7 (2017)

[9] David, M., Jemna, D.V.: Modeling the frequency of auto insurance claims by means of poisson and negative binomial models. Analele stiintifice ale Universitatii "Al. I. Cuza" din Iasi. Stiinte economice/Scientific Annals of the "Al. I. Cuza" (2015)

[10] Wang, Y., Xu, W.: Leveraging deep learning with LDA-based text analytics to detect automobile insurance fraud. Decision Support Systems 105, 87–95 (2018)

[11] Subudhi, S., Panigrahi, S.: Use of optimized Fuzzy C-Means clustering and supervised classifiers for automobile insurance fraud detection. Journal of King Saud University-Computer and Information Sciences 32(5), 568–575 (2020)

[12] Majhi, S.K.: Fuzzy clustering algorithm based on modified whale optimization algorithm for automobile insurance fraud detection. Evolutionary intelligence 14(1), 35–46 (2021)

[13] Tuo, G., Duan, J.: Game Theory Analysis of Insurance Fraud. Journal of Capital University of Economics and Business (3), 51–54 (1999)

[14] Liu, X., Jin, J.: The Insurance Fraud Game and Insurance Contract Based on Optimal Game Strategies. Systems Engineering—Theory & Practice 24(2), 19–24 (2004)

[15] Gui, P., Hu, Q.: A Literature Review of Auto Insurance Moral Hazard at Home and Abroad. Insurance Studies (6), 121–127 (2011)

[16] Zhao, G., Wu, H.: Is There Moral Hazard in Chinese Automobile Insurance Market?—Evidence from Dynamic Renewal Policies. Journal of Financial Research (6), 175–188 (2010)

[17] Tang, J., Mo, Y.: Construction of Auto Insurance Anti-fraud System Based on Data Mining Technology. Journal of the Postgraduate of Zhongnan University of Economics and Law (5), 80–87 (2013)

[18] Wang, H.: A Research on Chinese Insurers' Moral Hazard Screening in Operation: From the Big Data Hadoop Clustering Analysis Technology Perspective. Insurance Studies (2), 59–67 (2016)

[19] Yan, C., Li, Y., Sun, H.: A Research on Automobile Insurance Fraud Identification Based on Random Forest Model and Ant Colony Optimization Algorithm. Insurance Studies (6), 114–127 (2017)

[20] Yan, C., Li, M., Zhou, X.: Improved genetic algorithm for vehicle insurance fraud identification model based on BP neural network. Journal of Shandong University of Science and Technology (Natural Science) 38(5), 72–80 (2019)

[21] Xu, X., Wang, Z., Wang, M.: An Empirical Study of Auto Insurance Fraud Identification Model Based on Deep Learning Technology. Shanghai Insurance (8), 53–58 (2019)

[22] Li, M., Zareian, A., Lin, Y., Pan, X., Whitehead, S., Chen, B., ... & Freedman, M.: GAIA: A fine-grained multimedia knowledge extraction system. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations, pp. 77–86 (2020, July)

[23] Zhang, B., Lin, Y., Pan, X., Lu, D., May, J., Knight, K., Ji, H.: Elisa-edl: A cross-lingual entity extraction, linking and localization system. In Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations, pp. 41–45 (2018, June)

[24] Li, M., Lin, Y., Hoover, J., Whitehead, S., Voss, C., Dehghani, M., Ji, H.: Multilingual entity, relation, event and human value extraction. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics (Demonstrations), pp. 110–115 (2019, June)

[25] Ngiam, J., Khosla, A., Kim, M., Nam, J., Lee, H., Ng, A.Y.: Multimodal deep learning. In ICML, pp. 689–696 (2011, January)

[26] Wang, D., Cui, P., Ou, M., Zhu, W.: Learning compact hash codes for multimodal representations using orthogonal deep structure. IEEE Transactions on Multimedia 17(9), 1404–1416 (2015)

[27] Peng, Y.X., Zhu, W.W., Zhao, Y., Xu, C.S., Huang, Q.M., Lu, H.Q., ... & Gao, W.: Cross-media analysis and reasoning: advances and directions. Frontiers of Information Technology & Electronic Engineering 18(1), 44–57 (2017)

[28] Wu, Z., Zheng, C., Cai, Y., Chen, J., Leung, H.F., Li, Q.: Multimodal Representation with Embedded Visual Guiding Objects for Named Entity Recognition in Social Media Posts. In Proceedings of the 28th ACM International Conference on Multimedia, pp. 1038–1046 (2020, October)

[29] Mousselly-Sergieh, H., Botschen, T., Gurevych, I., Roth, S.: A multimodal translation-based approach for knowledge graph representation learning. In Proceedings of the Seventh Joint Conference on Lexical and Computational Semantics, pp. 225–234 (2018, June)

[30] Xie, R., Liu, Z., Luan, H., Sun, M.: Image-embodied knowledge representation learning. In Proceedings of the 26th International Joint Conference on Artificial Intelligence, pp. 3140–3146 (2017, August)

[31] Chen, L., Li, Z., Wang, Y., Xu, T., Wang, Z., Chen, E.: MMEA: Entity Alignment for Multi-modal Knowledge Graph. In International Conference on Knowledge Science, Engineering and Management, pp. 134–147 Springer, Cham. (2020, August)

[32] Guo, H., Tang, J., Zeng, W., Zhao, X., Liu, L.: Multi-modal Entity Alignment in Hyperbolic Space. Neurocomputing 461(1), 598–607 (2021)

[33] Chen, T., Guestrin, C.: Xgboost: A scalable tree boosting system. In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, pp. 785–794 (2016, August)

AUTHOR BIOGRAPHY

Jiaxi Yang, postdoctoral researcher of Financial Technological Research Center at Zhejiang Lab. He received his M.S/Ph.D. in Statistics, in 2016/2021 from Columbia University and B.A. in Economics in 2014 from Chu Kochen Honors College, Zhejiang University. He was enrolled in the international postdoctoral exchange & introduction program funded by China Postdoctoral Science Foundation. His research interests include causal inference, machine learning and multi-modal learning, and applications in area such as finance, economics and education. He is now focusing on the research of financial risk management, multi-modal data generation/imputation and improving model robustness.
ORCID: 0000-0002-5055-8729

Kui Chen, a postdoctoral fellow in Fintech Research Center of Zhejiang Lab. The main research interests include theoretical analysis and research of machine learning, model construction and optimization, data governance, automatic feature engineering and scenario solution construction. He graduated from Shanghai University majoring in mathematics and applied Mathematics with a doctor's degree. During the doctoral period, the research work focused on dimension reductions and integrabilities of high-dimensional semi-discrete integrable system, and completed seven SCI academic papers, two of which were listed as highly cited papers by Web-of-Scicence. After graduation, he went to the School of Mathematics and Science of Fudan University to engage in full-time postdoctoral research. The main research content is the algebraic structure of constrained high-dimensional semi-discrete integrable systems.
ORCID: 0000-0002-7925-9968

Ding Kai is currently a senior researcher of Financial Technology Center in ZheJiang Lab. He received his Ph.D. degree in School of Automation Science and Electronic Engineering at Beihang University, China. His research interests include multimedia information retrieval, nature language process, and deep learning.
ORCID: 0000-0003-4534-2904

Chongning Na received his B.E. degree in 2002 from Tsinghua University, M.S. and Ph.D. degrees in Electrical Engineering, in 2005 and 2010 from the Technical University of Munich. He is now a research expert at FinTech Research Center, Zhejiang Lab. His research interests include machine learning, probabilistic inference, graphical models and their applications in various areas e.g., industrial automation, telecommunications and FinTech. He was with Siemens Corporate Research and NTT DOCOMO Research Lab, and contributed to the R&D of ProfiNet, 4G/5G physical layers in the aspects of high-performance algorithm development and associated standardization, e.g., IEEE 1588, 4G LTE/LTE-A, 5G NR. He is now focusing on the research of intelligent computing technologies in financial risk management, using multi-modal-machine-learning based and knowledge-based financial data mining techniques.

ORCID: 0000-0003-2680-5774

Meng Wang is an assistant professor in the Knowledge Graph & AI Research Group, School of Computer Science and Engineering, Southeast University, China. He obtained the doctoral degree from the Department of Computer Science and Technology, Xi'an Jiaotong University in 2018. He was a visiting scholar in the DKE lab at University of Queensland, Australia in 2016. His research area is in the knowledge graph (KG), semantic search, NLP, and cross-modal data.

ORCID: 0000-0002-2293-1709