Application of Reinforcement Learning in Detecting Fraudulent Insurance Claims

Jung-Moon Choi†, Ji-Hyeok Kim††, and Sung-Jun Kim†††
jmchoi@wise.co.kr, jhkim78@wise.co.kr, jun1977@nsu.ac.kr
† Department of Research and Planning WISEiTECH, Seongnam-si, Republic of Korea
††Department of Research and Planning WISEiTECH, Seongnam-si, Republic of Korea
†††Department of Bigdata-Content Convergence, Namseoul University, Cheonan-si, Republic of Korea

Summary
Detecting fraudulent insurance claims is difficult due to small and unbalanced data. Some research has been carried out to better cope with various types of fraudulent claims. Nowadays, technology for detecting fraudulent insurance claims has been increasingly utilized in insurance and technology fields, thanks to the use of artificial intelligence (AI) methods in addition to traditional statistical detection and rule-based methods. This study obtained meaningful results for a fraudulent insurance claim detection model based on machine learning (ML) and deep learning (DL) technologies, using fraudulent insurance claim data from previous research. In our search for a method to enhance the detection of fraudulent insurance claims, we investigated the reinforcement learning (RL) method. We examined how we could apply the RL method to the detection of fraudulent insurance claims. There are limited previous cases of applying the RL method. Thus, we first had to define the RL essential elements based on previous research on detecting anomalies. We applied the deep Q-network (DQN) and double deep Q-network (DDQN) in the learning fraudulent insurance claim detection model. By doing so, we confirmed that our model demonstrated better performance than previous machine learning models.

Keywords:
insurance fraud detection, unbalanced data, reinforcement learning, DQN, DDQN.

1. Introduction

Due to the advancement of technology, data construction, and the improved processing environment, there has been a growing need in various business fields to process data to solve problems. For example, there is such a need in the insurance industry, and various methods have been explored, along with its core business technology. One of the industry's goals relates to the technology for detecting fraudulent insurance claims. Damages incurred due to fraudulent insurance claims are passed onto current or prospective insurers. Thus, the insurance industry researches the methods to minimize unfair insurance claims and prevent damages by detecting such claims in advance. The industry has begun to utilize data analysis technology more than traditional rule-based methods.

Manuscript received September 5, 2021
Manuscript revised September 20, 2021
https://doi.org/10.22937/IJCSNS.2021.21.9.17

The number of detected fraudulent insurance claims and claim payments has increased. This trend is found not only in South Korea but also in countries around the world. In order to detect and prevent fraudulent insurance claims, the insurance industry has utilized various artificial intelligence (AI) technologies. These are applied to various insurance technology fields, such as predicting fraudulent insurance claims, product design, and insurance policy reviews.

This study developed a fraudulent insurance claims detection model using the reinforcement learning (RL) method, one of the latest AI technologies. The RL method is an AI technology, including machine learning (ML) and deep learning (DL) technologies. Thus, it is necessary to define its essential elements.

There are few cases where the RL method has been applied for the detection of fraudulent insurance claims. On the contrary, the method has been used in business fields such as gaming, stocks, intelligent robots, and autonomous vehicles. This study applied the RL method to improve the performance of the fraudulent insurance claims prediction model, comparing how its performance in detecting fraudulent claims differs from that of the previous models, based on the small amount of unbalanced data.

From this section, input the body of your manuscript according to the constitution that you had. For detailed information for authors, please refer to [1].

1.1 Fraudulent Insurance Claims Detection

The number of cases utilizing insurance-related big data analysis has increased. The number of technology studies and cases that attempt to identify the characteristics of those who make fraudulent insurance claims, using data analysis technology to extract new values and preemptively prevent insurance fraud, has increased. In the previous study [1], we applied four labels by detecting secondary clustering anomalies and performed secondary clustering to supplement any inaccurate parts of the normal clustering. We supplemented the performance of primary clustering,

although there was a difference in the degree depending on the data types. We endeavored to back up an efficient selection of fraudulent insurance claims in suspicious groups. As a result, fraud data, which was very limited compared to normal data, were classified in detail as a suspicious group, subdividing and limiting detection targets.

There have been increasing cases of applying various ML and DL methods to detect fraudulent insurance claims for unbalanced data. The number of research papers on detecting fraudulent insurance claims using Medicare data from open data sources has grown. We conducted our research by applying the model developed in previous studies to Medicare data, and we confirmed its effectiveness for anomaly pattern detection [2]. We attempted Medicare fraud detection using CatBoost. CatBoost is an open-source library developed by Yandex, which provides a gradient boosting framework [3]. We examined gradient boosting decision tree algorithms and compared XGBoost and CatBoost-HCPCS models.

In addition to the methods mentioned above, there is also a case of approaching fraud detection problems by incorporating new technologies [4]. In this study, we endeavored to solve the insurance fraud problem in health care by applying blockchain technologies. The study concluded that the development of blockchain technology would be suitable for pharmaceutical and medical industries and complex data-sharing requirements. The study explains that using blockchain technology for storing decentralized data on health rather than centralized data is an effective way to prevent fraudulent claims.

1.2 Application of the RL Method

The academic areas which are known to apply the RL method are limited. There are few cases where the RL method has been applied in the fraudulent insurance claims detection field. Most cases of RL relate to gaming, robotics, and autonomous driving systems. However, there is a growing trend of attempts to apply the RL to other fields.
In the insurance industry, the RL method has been used to optimize price policies [5]. That is an example of applying the RL method for adjusting insurance renewal prices when two conflicting conditions must be considered—retaining existing customers and increasing insurance revenues. Our paper applied the model-free algorithms to solve the issue of adjusting prices and proposed applying the Markov Decision Process (MDP) and constrained MDP.

Examining literature on the RL application for anomaly detection, we have confirmed the cases of applying it to detect network anomalies and anomalies in unbalanced data [6]. Because the RL method demonstrated better performance for unbalanced data and anomaly detection, we attempted to apply it to detect fraudulent insurance claims.

1.3 Research Aim and Process

We propose a method to improve fraudulent insurance claims detection by letting the fraudulent insurance claims detection model learn by applying the RL algorithm to the previously researched data on fraudulent insurance claims. We used the previous study's [1] insurance data to test our RL method. We compared the prediction results of the model reliant on the RL method with those of the previous ML models and evaluated the model's performance. We constructed an RL environment, focusing on unbalanced data and anomaly detection, and defined the issue as a binary classification problem.

2. Research Methods

This section discusses the RL method, its elements, and its use for detecting fraudulent insurance claims. It also examines the selection of the RL algorithm for the fraudulent insurance claims detection model.

2.1 RL Method

The RL method is an ML learning method. In the RL method, an agent analyzes the current state in a fixed environment and learns to maximize rewards among selectable actions or learns an optimal sequence of actions. Unlike in the existing ML methods, the RL elements must be defined to apply the method. Thus, RL research in gaming has been conducted. An environment and agents are necessary for the RL method. The environment is a fixed environment where the simulation for RL occurs, and the agents are subjects that act within that environment. The agents learn by themselves as they explore goals within the environment. Table 1 summarizes the essential elements of the RL method.

The RL environment is mainly given as the MDP [7]. The MDP is a modeled decision-making process using probability and graphs. Our study approached fraudulent insurance claims by defining them as an MDP problem. Figure 1 demonstrates the process defined as the MDP problem.

Table 1. RL Essential Elements

Element	Description of each element
State	- An agent assesses by referring to data. - Variables for learning data are defined. - Sufficient information for assessing a situation should be provided to an agent.
Action	- A set of actions is determined based on the agent's assessment of a specific state. - An optimal behavior for RL purposes is selected.
Reward	- Reward values that can be obtained when a specific action is selected in a particular state are set. - The role of guiding the learning path is established. - Reward values to support exploration and optimization are set. - Rewards can have a significant impact on the performance of RL models.

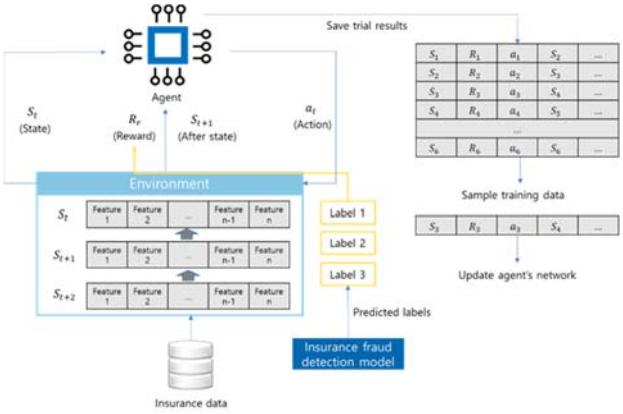

Fig. 1 Fraudulent Insurance Claims as the MDP.

In Figure 1, state (S) is one instance of data with the characteristics of a fraudulent claim, and action (a) denotes predicting normal and abnormal data based on the agent's judgment. The reward (R) is defined by comparing the predicted insurance fraud detection model's labels.

The agent analyzes a given state in the environment, takes action, evaluates a label accordingly, and receives a reward. The agent saves the trial-and-error results in the replay memory and learns by extracting learning data samples by repeating the process. The agent learns following the path of maximizing the sum of rewards through repetition of this sample learning. Therefore, the definition of reward and establishing a process for obtaining the best reward is important.

2.2 Fraudulent Insurance Claims Detection in the RL Model

In order to apply the RL method to detecting fraudulent insurance claims, each element of the RL model must be defined. The process differed from previous studies [1][2] in that it did not use under-sampling or over-sampling to deal with the imbalanced labels. However, a reward element was added to design the maximal cumulative reward based on the data sample status.

Table 2. Definition of the RL Fraudulent Claim Detection Model

Predicted action	Real label	
	Normal	Fraud
Fraud	0	1
Normal	P = Fraud/Normal (P < 1)	0

Reward settings were defined as normal-normal, normal-fraud, fraud-fraud, and fraud-normal by dividing the combination of agent actions and real labels into four groups, considering that normal data is more unbalanced than fraud data. If there were more normal data so that reward values were the same, all data could have been predicted as normal. To mitigate this, we applied different label-specific rewards. When the fraud was predicted based on real abnormal data, the reward was defined as +1. The reward for normal data predicted as normal was defined as less than 1 (< 1). It reflected the ratio (P) between the number of abnormal data instances and normal data instances (Table 3).

Table 3. Reward Definition in the Fraudulent Insurance Claim Detection Model

Element	Fraudulent insurance claim detection model
State	- Defined as existing learning model variables (32) - Applying the min-max scaling and data shuffling; not applying sampling
Action	- Defining labels as normal/fraud—the same as in the fraudulent insurance claims model
Reward	- Setting rewards by dividing the cases into those where the agent's actions match the existing detection model's labels and those where they do not match - Applying differentiation of rewards for unbalanced data structures

2.3 Selecting the RL Algorithm

A model-free-based RL algorithm was selected for the fraudulent insurance claims detection model since its action space is relatively simple with only two criteria—normal

and abnormal (fraudulent). Additionally, fraudulent insurance claims detection models that used a value-based algorithm demonstrated the highest accuracy, recall, and F1-score performance. The results were confirmed by having the model learn the deep Q-learning network (DQN) and the double deep Q-learning network (DDQN). The DQN algorithm can save enormous computational resources by approximating the Q-function for motion and state in a combination of Q-learning and a DL neural network model [8].

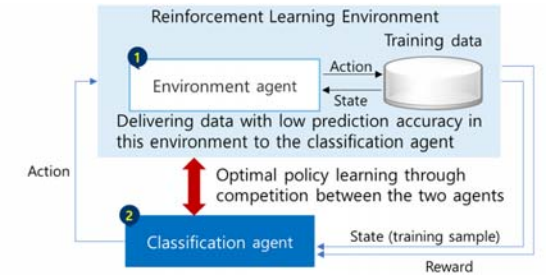

Figure 2. Reinforcement Learning Process in an Adversarial Environment.

As Figure 2 shows, the two agents learn competitively. The environment agent (attacker) quickly updates and stores simulation results to form an environment for reinforcement learning. The attacker allows the classification agent (defender) to learn from less accurate training samples from the simulation results as a priority. The defender updates slower compared to the attacker and can learn consistently [6]. This adversarial relationship controls normal (easy) and fraud (difficult) problems to be solved in a balanced manner.

3. Results of Applying the RL Method to a Fraudulent Insurance Detection Model

We developed the DQN and DDQN models to detect fraudulent insurance claims and confirmed training results by applying the models to the same data. The learning parameters were set as 1,000 for the total number of episodes (num_episodes = 1,000), and 100 for training iterations (iterations_episodes = 100). In addition, we tried the RL method by changing variables such as epsilon, gamma, and decay rate. In the case of the DDQN, the attacker and defender were given different parameter values to learn through competition. We checked the results for rewards and losses in each episode, confirming that learning was completed properly (Figure 3, Figure 4).

After comparing the results of the existing ML-based model with the performance of each RL model, we found that the DDQN model demonstrated the highest performance with 88% accuracy and 88% recall. Compared with the DQN model, the recall rate of the DDQN model was 30% higher. The details of the comparison of the performance results for each model are presented in Figure 5.

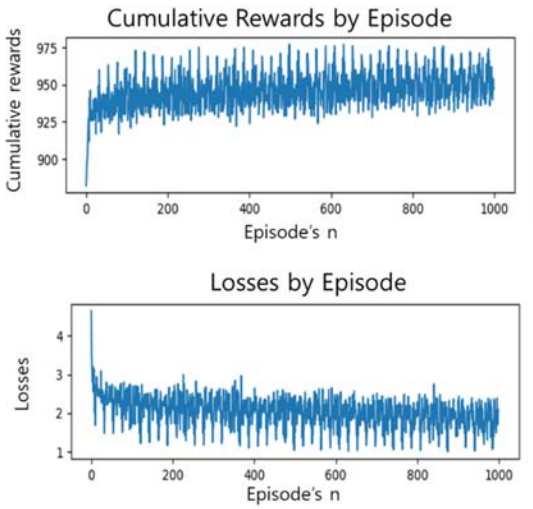

Fig. 3 Rewards/Losses Learning Processes for Each Episode of the DQN Model.

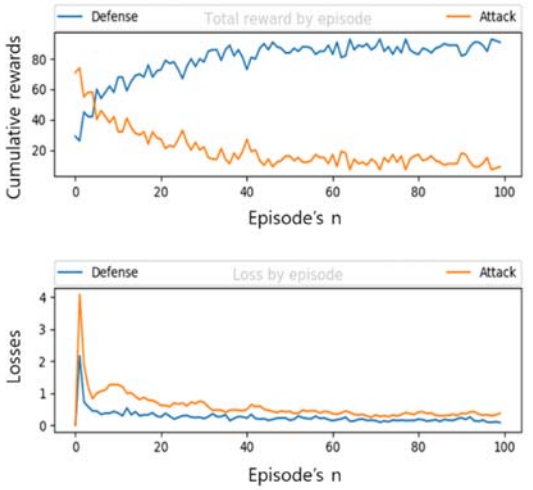

Fig. 4 Rewards/Losses Learning Processes for Each Episode of the DDQN Model.

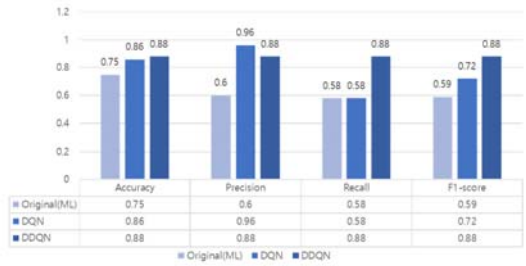

Fig. 5 Comparison of the Performance of Fraudulent Insurance Claims Detection Models.

The DDQN fraudulent insurance claims detection model demonstrated higher scores for accuracy (13%), precision (28%), recall (30%), and F1-score (29%) than the other models. After comparing the performance of normal and fraud labels, we found that the DDQN model for the fraud label had lower precision than the DQN model, but its recall was 29% higher than that of the DQN model.

The insurance claims data in this study were unbalanced. Due to this, precision and recall for each label showed opposite tendencies. Therefore, to evaluate the performance of the insurance claim detection models, F1-score was selected as an evaluation index. The F1-scores for the DQN and DDQN models were 0.72 and 0.82, respectively, demonstrating a 10% difference.

After applying the RL method in the fraudulent insurance claims models, recall and F1-score for the fraud label in the DDQN model were 87% and 0.82, respectively. Thus, learning in the DDQN model using unbalanced data was better than the DQN model, which showed a 58% recall and 0.72 F1-score.

Table 4. Performance Comparison of the RL Methods of Insurance Fraud Detection Models

	DQN model			DDQN model		
	Prec. (%)	Recall (%)	F1	Prec. (%)	Recall (%)	F1
Normal	84	99	0.91	94	88	0.91
Fraud	96	58	0.72	77	87	0.82

Note. Prec. = precision; F1 = F1-score.

Table 5 below shows a comparison of the existing fraudulent insurance claims model with the prediction results of the RL model. Because the RL prediction results were binary—classified into normal and fraud labels, the previous results classified as normal suspect and abnormal suspect, were substituted with normal and fraud when calculating accuracy.

Table 5. Comparison of the Results of Existing Fraudulent Insurance Claims Model and the RL Model

	Existing fraudulent insurance claim model				RL model			
	Prec. (%)	Recall (%)	F1-score	Accuracy (%)	Prec. (%)	Recall (%)	F1-score	Accuracy (%)
Normal	82	83	0.82	91	94 (+12%)	88 (+5%)	0.91 (+0.09)	90 (-1%)
Fraud	60	58	0.59	87	77 (+17%)	87 (+29%)	0.82 (+0.23)	90 (+3%)
Normal suspect	-	-	-	31	-	-	-	89 (+58%)
Abnormal suspect	-	-	-	26	-	-	-	80 (+24%)

When comparing the results of the finally selected RL model with the ML-based fraudulent insurance claims detection model, its performance increased in most areas, except for the normal label accuracy, which was 1% lower. Also, the fraud label's performance was higher in precision by 17% and in recall by 29% compared with the existing model. It could be confirmed that fraud detection performance regarding suspected cases (normal suspects and abnormal suspects) classified in the secondary anomaly detection models in the previous research also improved significantly.

4. Conclusion

In this study, reinforcement learning methods were applied to improve the detection of fraudulent insurance claims based on unbalanced data. Insurance claims data sets' definitions were tailored to the reinforcement learning elements, and models were implemented by configuring a reinforcement learning environment. Also, the study compared performance indicators such as accuracy, recall, precision, and F1-score of the previous ML model and the newly applied RL model. This study used DQN and DDQN

reinforcement learning algorithms. Experimental results showed that compared to the existing ML model, the F1-score of the reinforcement learning model was higher. The DDQN model demonstrated the highest performance. In the case of the DQN model in the reinforcement learning experiment, caution was required due to the possibility of overestimation. This limitation could be overcome through the adversarial learning environment of the DDQN model. This study is significant because the reinforcement learning methods previously used in a limited number of fields were applied to detect fraudulent insurance claims. Also, it was possible to improve the learning performance on the minority of unbalanced data. This confirmed that the reinforcement learning method is effective for anomaly detection, such as detecting fraudulent claims.

The limitations of this study are related to using existing datasets for the experiment, generalization of additional data, and optimal model evaluation. The DDQN model was chosen as a fraudulent insurance claims detection model to obtain stable results even with new data. In the future, additional research is required to expand the application of this method to the fraudulent receipt of subsidies detection model and insurance subscribers' contract review model.

Acknowledgments

This work was supported by the ATC program of the Ministry of Trade, Industry and Energy (MOTIE) and Korea Evaluation Institute of Industrial Technology (KEIT) (Assignment No. 10077293—Intelligent Fraudulent Claim Detection System's Technology Development that Improves the Fraudulent Claim Detection Rate by Over 40% Through Early Prediction of Fraudulent Insurance Claims).

References

[1] Choi, J., Kim, J.: *Developing an abnormal pattern classification model based on secondary abnormal detection*. International Journal of Advanced Science and Technology *28*(16), 91–105 (2019). http://sersc.org/journals/index.php/IJAST/article/view/1663

[2] Choi, J., Kim, J., Lee, J.: *A study on the application of the secondary anomaly pattern detection model based on unsupervised learning: Medicare service fraud detection*. International Journal of Advanced Science and Technology *29*(4), 10551–10562 (2020). http://sersc.org/journals/index.php/IJAST/article/view/33571

[3] Hancock, J., Khoshgoftaar, T. M.: *Medicare Fraud Detection Using CatBoost*. 2020 IEEE 21st International Conference on Information Reuse and Integration for Data Science (IRI), 97–103 (2020 September). doi:10.1109/IRI49571.2020.00022

[4] Saldamli, G., Reddy, V., Bojja, K. S., Gururaja, M. K., Doddaveerappa, Y., Tawalbeh, L.: *Health Care Insurance Fraud Detection Using Blockchain*. 2020 Seventh International Conference on Software Defined Systems (SDS), 145–152 (2020 July). doi:10.1109/SDS49854.2020.9143900

[5] Krasheninnikova, E., García, J., Maestre, R., Fernández, F.: *Reinforcement Learning for Pricing Strategy Optimization in the Insurance Industry*. Engineering Applications of Artificial Intelligence *80*, 8–19 (2019). doi:10.1016/j.engappai.2019.01.010

[6] Caminero, G., Lopez-Martin, M., Carro, B.: *Adversarial Environment Reinforcement Learning Algorithm for Intrusion Detection*. Computer Networks *159*, 96–109 (2019). doi:10.1016/j.comnet.2019.05.013

[7] Bellman, R.: *A Markovian Decision Process*. Journal of Mathematics and Mechanics *6*(5), 679–684. (1957). https://www.jstor.org/stable/24900506

[8] Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M. et al.: *Human-level Control through Deep Reinforcement Learning*, Nature *518*(7540), 529–533 (2015). doi:10.1038/nature14236

Jung-Moon Choi received the B.S. and B.L.A. degrees from Seoul Women's University in 2014. She is a senior researcher in the Department of Research and Planning at WISEiTECH. Her research has pioneered and focused on every aspect of data analytics, optimization, and predictive modeling. As a senior researcher, she has led a project to develop predictive analytics for insurance fraud detection using artificial intelligence. In addition, her research interests include machine learning. She explores neural network applications and tries to apply deep learning models based on prior knowledge obtained from different projects.

Ji-Hyeok Kim received his B.E. and M.E. degrees from Soongsil University in 2003 and 2005, respectively. He received the Dr. Eng. degree from Soongsil University in 2010. From 2016 to 2018, he worked as a chief researcher in the Department of Research and Planning at WISEiTECH. Since 2018, he has been the head of the research center at WISEiTECH. His research interests include software engineering, ML/DL, and XAI.

Sung-Jun Kim earned his Ph.D. in Law from Dongguk University in 2009. He has been an associate professor at the Department of Bigdata-Content Convergence, Namseoul University, since 2015. His research interests include deep learning, Digital Twin, and Mydata.