

A Study on Conspired Insurance Fraud Detection Modeling Using Social Network Analysis

Tae-Ho Kim*, Jong-In Lim*

*Student, Graduate School of Information Security, Korea University, Seoul, Korea

*Professor, Graduate School of Information Security, Korea University, Seoul, Korea

[Abstract]

Recently, proving insurance fraud has become increasingly difficult because it occurs intentionally and secretly via organized and intelligent conspiracy by specialists such as medical personnel, maintenance companies, insurance planners, and insurance subscribers. In the case of car accidents, it is difficult to prove intentions; in particular, an insurance company with no investigation rights has practical limitations in proving the suspicions. This paper aims reveal that the detection of organized and conspired insurance fraud, which had previously been difficult, could be dramatically improved through conspiring insurance fraud detection modeling using social network analysis and visualization of the relation between suspected group entities and by seeking developmental research possibilities of data analysis techniques.

▶ **Key words:** social network analysis, insurance fraud detection, conspired insurance fraud, insurance fraud, fraud detection

[요 약]

최근 보험사기는 의료인, 정비업체, 보험설계사 등 분야별 전문가와 보험가입자가 공모하여 조직화, 지능화되는 형태로 계획적이고 은밀하게 이루어지고 있어 그 혐의를 입증하는 것이 점점 더 어려워지고 있다. 자동차사고의 경우 고의를 입증하기가 쉽지 않으며 특히 수사권이 없는 보험사가 혐의를 입증하기는 현실적으로 한계를 가지고 있다. 본 논문에서는 공모형 보험사기 적발을 위한 Social Network Analysis(SNA), 혐의그룹 개체간 관계 시각화 기법 등 공모형 보험사기 혐의자 적발 기법을 통해 기존에는 알아내기 어려웠던 조직형, 공모형 보험사기 파악이 획기적으로 개선될 수 있다는 점을 살펴보고 조직적 공모형 보험사기 적발을 위한 데이터 분석기법의 발전적 연구 방향을 모색해 보고자 한다.

▶ **주제어:** 사회관계망분석, 보험사기 적발, 공모형 보험사기, 보험사기, 사기적발

• First Author: Tae-Ho Kim, Corresponding Author: Jong-In Lim
*Tae-Ho Kim (taeho@fss.or.kr), Graduate School of Information Security, Korea University
*Jong-In Lim (jilim@korea.ac.kr), Graduate School of Information Security, Korea University
• Received: 2020. 02. 14, Revised: 2020. 03. 19, Accepted: 2020. 03. 19.

I. Introduction

일반적으로 보험범죄라 함은 보험가입자 또는 제3자가 받을 수 없는 보험급부를 대가 없이 받거나 부당하게 낮은 보험료를 지불하거나 부당하게 높은 보험금의 지급을 요구할 목적을 가지고 고의적이며 악의적으로 행동하는 것을 의미한다. 보험범죄가 구체적인 범법행위로 나타난 결과만을 지칭하는 반면에 보험사기는 보험가입자의 악의성을 포함하는 보다 광범위한 개념이라는 점에서 보험범죄보다 더 광의의 개념이다.[1] 2016. 3.29. 법률 제14123호로 제정된 보험사기방지 특별법에서 보험사기행위를 ‘보험사고의 발생, 원인 또는 내용에 관하여 보험자를 기망하여 보험금을 청구하는 행위를 말한다’고 보험사기에 대한 법률적 정의를 하였다. ’18년 중 보험사기 적발금액은 [Fig 1]과 같이 전년 대비 9.3% 증가한 7,982억원으로 역대 최고금액으로 집계되었고 적발인원은 총 79,179명으로 전년 대비 5.2% 감소하여 1인당 평균 사기금액은 1,010만원으로 점점 고액화 추세이다. 통상 살인, 방화 등의 강력범죄가 동반되는 것으로 인식되고 있으나, 실제로는 범죄행위라는 인식이 강하지 않은 형태의 허위, 과다사고 유형의 보험사기가 대부분을 차지하고 있다.[2] 보험사기의 적발금액은 전체 보험사기 추정 금액대비 약 10%로 내외로 실제 보험사기 규모에 비해 적발되는 보험사기 규모는 빙산의 일각으로 표현되고 있다. 특히, 조직형-공모형태의 보험사기 금액비중은 ’14년 약 940억(15.7%)에서 ’17년 약 1,300억(17.7%)으로 크게 증가하고 [Fig 2]와 같이 정비업체, 보험설계사 등 보험업종사자 또는 분야별 전문가와 보험가입자가 공모하여 조직화, 지능화되는 형태로 계획적이고 은밀하게 이루어지고 있어 그 혐의를 입증하는 것이 점점 더 어려워지고 있다. 자동차사고의 경우 고의를 입증하기가 쉽지 않으며 특히 수사권이 없는 보험사가 혐의를 입증하는 현실적으로 한계를 가지고 있다.[3] 이러한 보험사기로 인하여 지급된 보험금의 누수는 결국 보험료를 인상시켜 그 피해가 전체 보험계약자에게 돌아가게 된다.[4] 본 연구에서는 사회학 위주로 연구되고 있는 사회연결망 분석기법을 보험사기 조사에 대한 영역으로 확장하여 개별 데이터로는 인지할 수 없었던 공모형 보험사기 적발을 위한 Social Net work Analysis (SNA), 혐의그룹 개체간 관계 시각화 기법 등 보험사기 혐의자 그룹 적발 기법을 통해 기존에는 알아내기 어려웠던 조직형, 공모형 보험사기 파악이 획기적으로 개선될 수 있는 데이터 분석기법을 제시 하고자 한다.

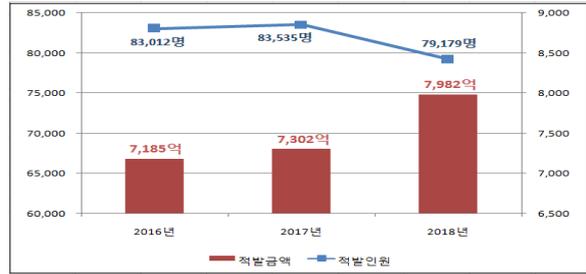


Fig. 1. Amount and number of people detected for insurance fraud (KRW 100 million won, people)

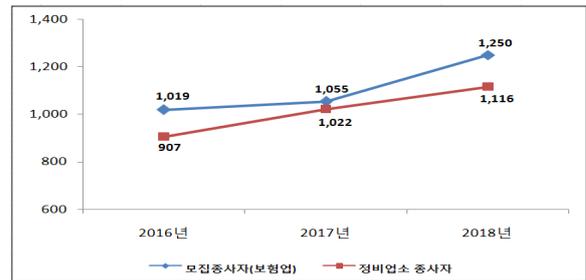


Fig. 2. Workers in insurance company and body shop detected in insurance fraud (people)

II. Preliminaries

1. Lack of research on methods for detecting conspired insurance fraud

보험업계에서는 심각한 보험범죄에 대응하기 위해서 각 보험회사에서는 자체적으로 보험사기 조사를 위한 특별조사팀(SIU, Special Investigation Unit) 구성하고 보험사 내에 보유하고 있는 보험계약 및 보험금 지급 내역 등을 바탕으로 보험사기 혐의가 농후한 보험사고를 분류하기 위해 보험금 지급총액, 사고건수, 중복계약 등 다수의 보험사가 관련된 고액의 사고 건을 중심으로 데이터 항목을 세분화하고 이를 대상으로 보험사기 혐의를 분석하는 방법으로 보험사기 조사를 실시하고 있다. 금융감독원에서는 내부에 각 보험사로부터 보험계약 및 보험금 지급정보를 데이터 베이스로 축적하고 과거 보험사기로 적발된 자(개인, 보험설계사, 병원 등으로 구분)들의 사기유형과 사고형태를 정형화하여 보험사기에 대한 자료를 개발하고 이에 대해 점수화하여 보험사기 혐의자를 적출하는 통계적인 기법을 활용[5]하여 큰 성과를 거두고 있다. 그러나 이러한 방식은 방대한 계약정보 및 사고정보 자료를 장기간 동안 분석해야 하므로 상당한 시간, 인력 및 비용이 소요되거나, 개별 혐의자 위주로 분석하는 방식으로 공모 형태의 조직적 보험사기에 효과적으로 대응하기는 곤란한 실정이다. 공모 형태의 조직적 보험사기는 주로 공모에 가담한 자의

제보에 의존하여 적발되는 경우가 대부분이다. 이미 언급한 바와 같이 최근 보험사기의 형태가 보험설계사 또는 보험가입자가 공모하여 조직화, 지능화되고 있으나 방대한 자료에 대한 분석을 통해 그 혐의를 입증하기는 현실적으로 한계를 가지고 있다. 아울러, 보험범죄 예방을 위해 보험사기방지 특별법이 제정되는 등 보험사기에 대한 사회적 경각심은 높아지고 있으나 공모 형태의 보험사기 적발을 위한 데이터 분석 기법에 대한 연구는 거의 전무한 실정이다. 본 논문에서는 제보 없이도 공모 형태의 보험사기 혐의그룹을 자동으로 추출하고 주 혐의자를 손쉽게 식별하는 등 보험사기 조사업무의 효율성을 획기적으로 개선할 수 있는 보험사기 혐의자 적발 모델링을 제시하고자 한다.

2. Social network analysis techniques suitable for analysis of conspired insurance fraud charges

SNA(Social Network Analysis, 사회연결망분석, 이하 SNA)는 일련의 개체와 관계집합으로 구성되어 있는 네트워크(Network) 형식의 데이터를 분석 대상으로 하는 방법론이다. SNA는 네트워크의 중심구조, 응집구조, 이웃구조 등과 같은 구조적 분석을 위한 분석 지표를 정의하고 있다. 1930년대부터 이러한 주요 지표들이 개발되어 왔으며 최근에는 분석을 위한 데이터와 계산 속도가 향상됨에 따라 사회의 다양한 현상을 이해하기 위한 접근 방법으로 여러 분야에서 활용되고 있으며 사회적 현상을 점(Node, 개체)과 선(Link, 관계)으로 구성되는 네트워크로 정의하여 분석하는 방법론이다.[6-8] 사회연결망 분석 방법론은 1980년대 초반 국내에 도입되었고 사람과 사람 사이의 연결관계를 시각화, 수치화하고 관계 속에 내재된 속성을 분석하였다. 그동안 주로 사회과학 분야에서 많은 연구가 진행되었으며 각 단위의 조직이나 집단에 존재하는 구성원별 역할을 정의하거나 비슷한 성향의 구성원을 군집화 하는 등 사회의 구조적 속성을 규명하는 분석도구로 활용[9]되고 있으나 SNA를 보험사기 조사에 활용하려는 시도는 없었다. 보험사기 조사를 위해 SNA 접근 방법을 사용하는 이유는 보험사기 패턴이 다양한 관련 개체들(사람, 차량, 병원, 정비소)이 서로 복잡한 관계(계약, 사고, 치료 등)를 형성하는 양태를 나타내기 때문이다. 즉, 보험사기의 패턴이 개체와 관계로 구성되는 네트워크로 표현될 수 있기 때문에, 이와 같은 네트워크의 구조적 분석에 가장 적합한 SNA 접근 방법을 사용하는 것이 효율적이기 때문이다[10]. 최근 머신러닝 기술의 발달에 따라 사고정보를 수집하고 머신러닝을 통해 사고정보를 학습하여 보험청구건에 대한 사고분석 및 손해사정을 하고 보험금을 지급하거나 보험금 청구 패턴을

군집화 및 다양한 분류 알고리즘(RandomForest, Decision Tree, SVM 등)으로 최적 모델을 도출하고 반복 학습하여 부당청구를 탐지하는데 활용하는 연구가 진행되고 있다. 데이터과학의 관점에서 이러한 머신러닝을 활용한 부당 보험청구 분석 기술은 개별 사고정보에 대한 반복적인 지도학습(정상/비정상)을 통해 개별 사건에 대한 부당 보험금 청구 패턴을 탐지 하는데 활용하고 있다. 또한 과거 사기적발시스템(Fraud Detection System)도 개별 사건에 대한 사기적발에 집중하고 있지만 SNA는 축적되어 있는 사고정보에 대한 개체간의 관계 구조 분석으로 보험사기 혐의그룹 추출이 가능하여 공모형태의 보험사기 적발에 상당히 효율적이다. 보험업에 사용되는 데이터는 다양한 개체들의 관계로 구성되어 있기 때문에 SNA를 활용함으로써 기존의 통계 분석이 제공하지 못하는 분석 결과를 얻을 수 있다. 보험과 관련된 구체적인 개체로는 사람(역할 구분: 설계사, 계약자, 피보험자, 운전자, 동승자 등), 차량, 병원, 정비소 등이 있으며, 이들 개체들은 보험 계약, 병원 치료, 사고, 동승, 지인 관계 등의 다양한 관계를 맺고 있다. 이와 같은 개체들간의 관계 구조의 특성을 분석함으로써 '보험사기' 혐의 집단을 도출하고자 하는 것이 SNA를 보험 데이터에 적용하는 목적이라고 할 수 있다.

III. Technical Elements for Social Network Analysis on Insurance Fraud Charges

보험사기는 고의로 사고를 발생시키거나 과잉 청구를 함으로써 부당한 이익을 편취하는 것이 목적이다. 이와 같은 고의 사고나 과잉 청구를 통해 부당 행위를 하기 위해서는 관련 당사자들간의 담합을 필요로 하며, SNA는 이러한 고의성이나 과잉성이 의심되는 혐의집단의 담합 현상을 기초로 혐의 의심 그룹(네트워크 하위 그룹)을 도출하는데 활용될 수 있다.[11] 이러한 SNA 혐의 분석을 위해 필요한 기술 요소에 대해 설명 하고자 한다.

1. Measurement of similarity relations

SNA와 같은 네트워크 분석에서 중요한 것 중 하나는 그래프의 하위 집단을 찾는 것이다. 하위집단은 동질적인 이해관계나 성격을 가진 네트워크 내에 부분집합으로 구성되며 하위집단에 대한 분석을 통해 구성원들 사이의 이해관계나 다양한 집단특성을 파악할 수 있다. 이는 그래프에서 유사한 노드(Node, 점, 개체, 이하 노드)를 찾아 그룹 집단을 추출하거나, 노드와 비슷한 연결을 갖는 콘텐츠

를 추천해주는 알고리즘 등에 응용되기도 한다. 이렇게 그룹 집단을 추출하기 위해서 연결 관계를 정의해야 하는데, 이를 위해서는 반드시 모든 노드 사이의 유사성을 계산해야 한다. 네트워크의 유사성을 계산하는 과정에서 네트워크를 구성하는 링크에 가중치가 있고 없함에 따라 사용하는 유사성 지표가 달라진다. 일반적으로 링크에 가중치가 없는 경우에는 jaccard coefficient가 많이 사용되고 있으며, 링크에 가중치가 있는 경우에는 cosine similarity가 많이 사용된다.

· Jaccard Coefficient

두 집합이 포함하는 데이터의 유사성을 비교하기 위한 통계량이다. 기본적인 산출 방식은 두 집합 사이의 공통적으로 갖는 데이터의 크기를 두 집합이 갖고 있는 모든 데이터의 크기로 나눈 것으로 아래와 같은 수식(1)으로 표현된다.[12] 예를 들어, 일반적으로 특정 피보험자가 방문한 병원 보다 방문하지 않은 병원의 수가 많다. 따라서 두 피보험자가 모두 방문하지 않았던 병원을 제외하고 피보험자간 방문한 병원의 유사도를 측정할 수 있게 된다.

$$J = \frac{M_{11}}{M_{01} + M_{10} + M_{11}} \dots\dots\dots(1)$$

- M_{11} : 공통으로 갖고 있는 데이터의 개수
- M_{10} : A가 단독으로 갖고 있는 데이터의 개수
- M_{01} : B가 단독으로 갖고 있는 데이터의 개수

· Cosine Similarity

다차원 양수공간에서의 유사도 측정에 자주 이용된다. 특히 네트워크는 이러한 다차원 구조(링크구조)를 갖고 있기 때문에, 링크의 가중치가 부여되어 있는 네트워크에 대해서 두 노드의 유사성을 측정하는 매우 유용한 방법이다.[13] 유사도가 '1'이면 매우 유사하고, 유사도가 '0'이면 유사하지 않음을 나타낸다.

$$S = \frac{\sum_{i=1}^n A_i \times B_i}{\sqrt{\sum_{i=1}^n (A_i)^2} \times \sqrt{\sum_{i=1}^n (B_i)^2}} \dots\dots\dots(2)$$

A_i or B_i : A 또는 B가 i 번 째 링크에 갖는 가중치

2. Extraction of network sub-groups

· Cycle Detection

SNA에서 cycle detection 알고리즘은 특정 관계를 통해 구성된 네트워크 내에서 방향성을 가진 cycle 등을 추출하는 구조 분석 방식이다. 자동차 보험은 생명, 상해 보험과는 달리 가해자, 피해자간 공모를 통해 자동차 사고를 인위적으로 조작하고, 특히, 피해차량에 다수가 탑승(동승)할 수 있어 단일사건으로 다수가 보상 받기 쉽다는 특성이 존재하기 때문이다. 일반적으로 다양한 분야에서 fraud detection을 탐지하는데 자주 사용되며 Cycle을 탐지하기 위해 사용되는 알고리즘은 매우 다양하지만, 일반적으로 많이 사용되는 알고리즘은 depth-first search(DFS)를 기반으로 하는 알고리즘(by Robert Sedgwick)이 사용되며, 시간복잡도(time complexity)가 가장 낮은 알고리즘은 Tarjan's strongly connected component 알고리즘을 이용한 알고리즘이지만, 본 분석에서 추출되는 cycle detection은 cycle의 크기 및 링크 종류의 제한 조건이 사용되기 때문에, DFS 기반의 알고리즘을 사용하는 것이 훨씬 효율적인 방법이다. Cycle Detection 알고리즘은 일반적으로 방향성이 있는 네트워크에 대해 분석을 수행하므로, 단순히 연결성만으로 cycle을 탐지하지는 않는다. 따라서 방향성을 무시한 cycle 탐지를 위해 모든 링크를 양방향화(symmetrize)하여 분석을 수행 할 수도 있다.

· Clique(complete graph)

네트워크는 노드(점, 객체)와 링크(선, 관계)로 구성되어 있고 네트워크의 특징을 살펴보는 방법 중의 하나로 노드들이 서로 링크로 연결되어서 응집된 그룹을 이루는 경우 그 응집된 그룹 안에 있는 노드들은 공통 특징이 있다고 볼 수 있으며, 네트워크 내의 하위구조에서 응집구조를 이루는 하위구조그룹을 추출하는 방법 중 한가지로 clique는 이러한 네트워크의 응집성을 수치화해서 계산하는 방법의 하나이다. [14] 응집구조란 노드 간의 연결 관계가 밀집되어진 구조를 의미하며, SNA에서의 응집 구조는 일반적으로 같은 기능을 하는 집단 또는 동일 function module의 집합을 의미하기도 한다. 특히 clique은 이러한 응집구조를 추출하는 방법 중 가장 엄격한 제한 조건을 갖는 방법으로써 모든 노드가 서로 간에 직접적으로 연결되어 있어야만 clique이 성립된다. 따라서 clique 내부의 링크 밀도는 1이 된다. 단 clique은 연결 구조에만 의존하기 때문에, 링크의 가중치에는 영향을 받지 않는다. 이는 연계도 분석에서 구조적 혐의그룹을 분류할 수 있도록 해 준다.

· Weak component

응집된 그룹을 정의하기 위한 방법중의 하나로 모든 노드들이 하나 이상의 경로를 통해 서로 연결되는 최대 크기의 그래프이다. Weak component는 링크의 방향과 상관 없이 모든 노드가 하나로 연결되어 있는 집단을 말한다. 가령 노드 a에서 출발하여 노드 b까지 연결된 링크가 존재한다면 a와 b는 하나의 컴포넌트 안에 있는 것이다. 이렇게 하나의 컴포넌트 안에 존재하는 노드는 직간접적으로 관련성이 있음을 의미한다. 특히 엄격한 조건을 갖는 링크 관계에서는 동일한 weak component를 갖는 노드 집합엔 상당한 관련성이 존재함을 의미 한다.

3. Visualization of object relations

개체와 관계가 생성하는 네트워크를 시각화함으로써 그 구조를 직관적으로 이해하기 위한 것이며, 특히 개체간 관계가 복잡한 패턴을 보이는 경우에 네트워크를 시각화함으로써 그 구조를 이해하는데 소요되는 시간을 절약할 수 있다. 예를 들어, 아래 [Table 1] 데이터는 소스와 타겟 개체의 관계를 나타내고 있는데, 이 자료만으로는 데이터에 내재되어 있는 패턴을 파악하는 것이 어렵다. 하지만 [Fig 4]와 같이 이를 시각화한 연계도를 보면, 데이터에 담겨 있는 패턴을 직관적으로 파악할 수 있게 된다. 어떤 개체를 주목해서 보고, 어떤 관계를 집중적으로 파악해야 하는지 분석의 방향을 잡는데 중요한 역할을 하게 된다.

Table 1. Source object and Target object Relation

Source	Target	Source	Target
0000	1116	1238	1216
2680	1001	2780	2304
2704	2381	2604	0381
2773	1001	1513	1001
2847	1004	2147	2234
...

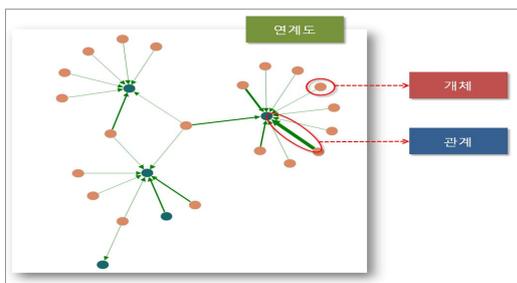


Fig. 3. Connection chart that visualizes objects and relations

IV. SNA Model for Insurance Fraud Charge

유사관계 측정, 네트워크 하위그룹 추출, 개체와 관계 시각화 기법 등을 활용하여 설계사와 병원, 설계사간 공모 보험사기 행위와 지인간 고의로 자동차 사고를 유발하고 보험금을 편취하는 공모형 보험사기 혐의 특성과 이를 적발하기 위한 보험사기 혐의 분석 모델을 제시하고자 한다.

1. Expansion of objects and relations

SNA 활용한 보험사기 조사를 위해서는 노드(객체)가 될 수 있는 요소를 정의하고 노드들간의 매트릭스(matrix)를 구성하여 관계(링크)를 도출하게 된다. 즉, 사람(계약자, 피보험자 등의 역할), 차량, 병원, 정비소 등을 개체로 하여 이들 개체가 맺는 보험의 계약, 사고, 치료, 수리 등 다양한 관계를 파악할 수 있어야 한다.

1.1 Types of objects

Table 2에서는 공모형 보험사기 조사를 위한 개체의 종류, 식별자와 파악 해야 할 개체간 관계를 나타내고 있다. 각 개체를 연계도상에서 모양과 색깔을 달리하여 표시할 수 있도록 하면 그 구조를 직관적으로 파악하기에 용이하다

Table 2. Types and identifiers of objects

Type	Identifier	Target object and relation
People	Resident Registration Number	Relations between policy holder, the insured, beneficiary, planner, driver, and passenger
Vehicle	Vehicle Registration Number	Relations between the vehicle at fault and damaged vehicle
Hospital	Unique code	Relationship between people and planner
Garage	Business registration number	Repair and other relations to the vehicles

1.2 Expansion of the relations

사람, 차량, 병원 등의 개체들과 연결될 수 있는 관계는 보험사기 조사 목적이나 특성에 따라 계약, 사고, 사고 차량, 병원치료, 보험금, 지인 등 다양한관계로 정의 할 수 있다. 이는 특정 개체를 연계도에 표시하고, 조사목적에 따른 관계 종류를 선택하여 시각화함으로써 그 구조를 직관적으로 이해할 수 있게 된다. 또한 이러한 연계도가 데이터 분석 업무에 효과적으로 사용하기 위해서는 지정한 일련의 개체를 기반으로 하여 특정한 관계로 연결된 개체를 연계도에 추가적으로 표시할 수 있도록 하는 '관계 확장'이 필요하다. 즉, 연계도에 특정 사람 1인이 표시되어

있다고 하면 이 사람을 기본 개체로 하여 설계사(From)→계약자(To) 관계를 확장한다는 것은 이 사람이 설계사일 때, 계약자였던 사람을 데이터로부터 검색하여 연계도에 추가하면서 기본 개체로 부터 추가된 사람에게로 [Fig 4]와 같이 선을 표시해 주는 것을 말한다. 하나의 연계도에 복합적 관계를 시각화할 수 있으나 관계의 개수가 많아지면 관계를 시각적으로 정확히 파악하기 어려워질 수 있다.

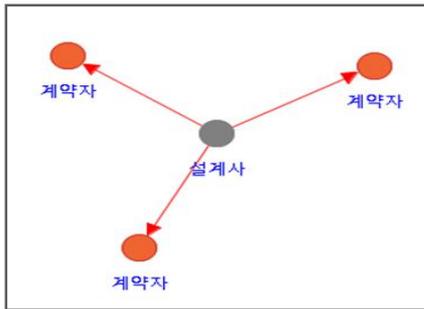


Fig. 4. Planner-contractor connection chart

2. Conspired insurance fraud by planner and hospital

설계사의 모집행위와 피모집 피보험자들의 병원 내원행위는 서로 독립적으로 발생하는 것이 일반적이다. 따라서 특정병원 진료를 사전에 예정하고 이루어진 모집행위나 병원 진료행위가 집중된 형태로 나타나는 형태를 의심 혐의 행위로 적발할 수 있으며, 이는 Jaccard Coefficient 지표를 통해 분석할 수 있다. 예를 들어 설계사와 병원사 무장이 공모하여 피보험자들을 모집하고 특정병원에서 진료하는 행위나 설계사가 특정 보험상품의 취약점을 피보험자들에게 교육 후 고지의무 등을 위반 또는 묵인하는 행위에 대한 혐의 행위를 적발할 수 있다.

2.1 Analysis overview

설계사와 병원이 공모하여 특정 피보험자 집단을 모집하고 진료하는 그룹을 인지하는 것으로 [Fig 5]와 같이 특정 설계사의 계약으로 인한 피보험자 집단이 특정 병원을 이용한 경향이 높은 설계사와 병원간의 관계를 추출하여 분석하는 모델이다.

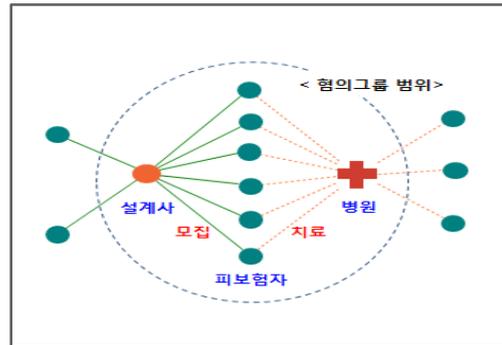


Fig. 5. Relation chart of conspiracy charge between planner and hospital

2.2 Analysis Model

일반적으로 분석대상기간은 보유 데이터량과 분석인력, 수사 용의성 등을 고려하여 일반적으로 3~5년 정도로 설정하고 혐의분석 대상은 보험금, 치료병원수, 보험계약수, 사고발생수 값과 병원과 설계사가 피보험자를 매개로 맺고 있는 관계 강도는 Jaccard Coefficient값으로 계산되며 값이 클수록 병원과 설계사간의 강한 연결 관계가 있음을 파악할 수 있다.

① (Extraction of the relations between the planner, the insured, and the hospital) Classified by the relations among objects, the setting of the target analysis period excluding group insurance, and payment items (including death, hazard, hospitalization, and surgery)



② (Exclusion of planners below the average generated benefits) Exclusion of the planners below the overall average generated benefits as the minimum criteria for the target planners for charge analysis



③ (Measurement of the jaccard coefficient between planner and hospital) Calculation of the jaccard coefficient based on the number of the insured



④ (Extraction of the relations with high jaccard coefficient) Extraction of the upper relations with high jaccard coefficient (the maximum number can be limited)



⑤ (Exclusion of the tertiary medical institutions) Exclusion of the relations that include medical institutions with no charge



⑥ (Cohesive structure analysis) Extracting suspicious groups connected in series by implementing weak components



⑦ (Analysis of unit index by suspicious group and detailed data on benefits payment) Extracting data with highly suspicious groups by analyzing the cumulated data on the number of the insured, planners, and the generated benefits (the maximum number in frequency and maximum amount of benefits)

3. Planners conspired insurance fraud

특정 그룹내 설계사 또는 피보험자들이 상호간에 병원 구성의 유사성을 가지는 그룹은 별도의 내부관계 없이는 생성되기 어려운 구조로 볼 수 있다. 이는 설계사들이 내부 공모 및 정보공유를 통해서 유사한 병원정보를 상호 공유하면서 보험금을 편취하는 행위나 가족이나 지인 등 피보험자들이 내부 공모를 통해 유사한 병원정보를 공유하며 빈번하게 입원 및 퇴원을 반복하는 혐의 행위를 적발할 수 있다. 이러한 유사성은 Cosine Similarity 지표를 활용해 혐의 그룹을 추출해 낼 수 있다

3.1 Analysis overview

[Fig 6]과 같이 피보험자인 설계사 집단이 병원 정보를 공유하여 여러 병원에 입퇴원을 반복하면서 보험금을 편취하는 혐의 패턴을 추출하는 것으로 설계사간에 부정공모를 위해 병원 정보를 공유하는 설계사 그룹을 인지하기 위한 모델이다.

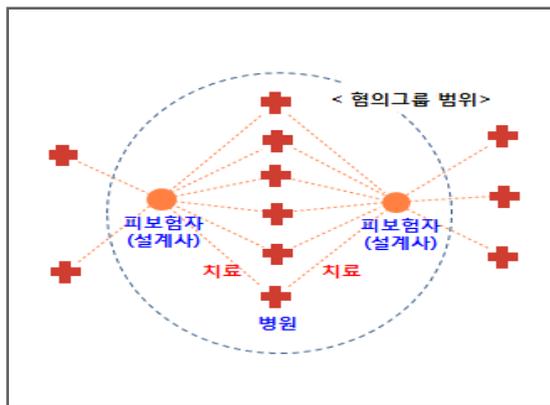


Fig. 6. Relation chart of conspiracy charge between planners

3.2 Analysis Model

피보험자가 병원을 매개로 맺고 있는 관계의 측정은 Cosine Similarity 값으로 계산되며, 값이 클수록 피보험자간에 강한 연결 관계가 있음을 파악할 수 있다. 최종적으로 개체와 관계가 연쇄적으로 연결되어 있는 집단을 혐의 그룹의 단위로 추출할 수 있다.

① (Extraction of planners and the insured) Setting relations between objects and target analysis period in life insurance and car insurance fields, excluding group insurance, and classified based on the payment item (including death, hazard, hospitalization, and surgery)



② (Extraction of charge analysis targets) Restriction of the maximum number of beneficiaries based on the payment item (including death, hazard, hospitalization, and surgery); the target criteria: premium × number of the hospitals that provide treatment × number of accidents × number of insurance contracts



③ (Measurement of similarity and number of common hospitals) Measurement of the number of common hospitals and cosine similarity based on the number of accidents (regarding the identical date of accidents and the identical main name of the disease as one accident) (excluding the tertiary medical institutions and increase in the number of hospitals by a certain standard deviation)



④ (Extraction of groups with mutual relations over a certain degree of similarity) Extraction of groups between planners that show high similarity (extraction of the clique directly connected to planners)



⑤ (Cohesive structure analysis) Extraction of suspicious groups connected in series by implementing weak components



⑥ (Analysis of unit index by suspicious group and detailed data on benefits payment) Extraction of data with highly suspicious groups by analyzing the cumulated data on the number of the insured, hospitals, planners, accidents, and the generated benefits (the maximum frequency and maximum amount)

4. Car Insurance Fraud Conspired Between Acquaintances

차량 운전자와 동승자는 서로 지인관계이므로 사고차량 운전자와 동승자 관계의 분석을 통해 지인간 자동차사고 공모 혐의 행위를 분석할 수 있다. 이는 연계도 분석상 운전자와 동승자의 사람을 기준으로 사이클(Cycle detection)형태로 나타나게 된다. 예를 들어 사고운전자와 동승관계의 연계도 구성에서 사고관계를 포함한 사이클 형태의 구조를 추출하게 되면 사고 공모 관계를 분석할 수 있다.

4.1 Analysis overview

정상적인 사고는 무작위 사건이라는 전제로부터 운전자와 동승자간 사고 관계를 분석했을 때 지인임이 확인 되거나

반복성이 나타날 경우 혐의점이 있는 고의 사고라고 판단할 수 있다. [Fig 7]과 같은 형태의 혐의 행위 그룹을 인지하는 것으로 운전자간 사고 관계 및 동승 관계에 의해 형성되는 사이클(링) 구조를 혐의 패턴으로 추출하는 모델이다.

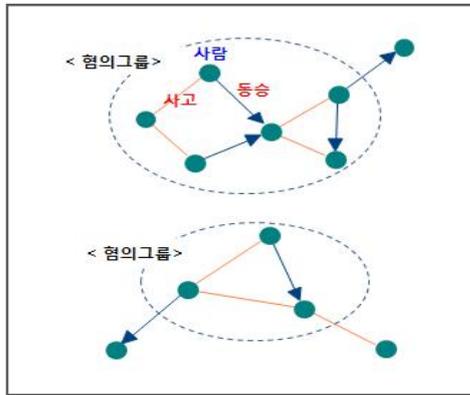


Fig. 7. Relation chart of conspired insurance fraud between acquaintances

4.2 Analysis Model

일반적으로 분석대상기간은 분석 데이터량, 수사기관의 수사용이성 등을 고려하여 최근 3년 이내로 설정하며, 사이클 구조를 구성하는 개체와 관계가 연쇄적으로 연결되어 있는 집단을 혐의그룹 단위로 추출할 수 있다.

① (Extraction of the relations between the drivers and passengers involved in accident) Extraction of object relations, analysis duration, accident relations among the drivers involved in accident (driver-driver), and passenger relations (driver-passenger)



② (Extraction of the cycle structure) Extraction of only the relations where vehicles at accident are linked by a cyclic structure (undirected cycle) (excluding the driver-passenger relations only)



③ (Cohesive structure analysis) Extraction of suspicious groups connected in series by implementing weak components, only the groups with three or more nodes (conspiring relationship)



④ (Analysis of unit index by suspicious group and detailed data on benefits payment) Extraction of data with highly suspicious groups by analyzing the data by group on the number of people, accidents, passenger relations, the cumulated generated benefits (the maximum frequency and maximum amount), and the duration in which the group was formed (the difference between the first accident date and the most recent accident date)

5. Car Insurance Fraud Conspired by the Driver at Fault and the Victim

자동차사고 가해차량 운전자와 피해차량 운전자가 역할을 바꿔가며 고의로 사고를 발생시키는 혐의 행위를 분석하는 기법으로 지인간 자동차사고 공모 분석방법과 유사하나 연계도상 가해차량 운전자와 피해차량 운전자의 방향성(가해자→피해자)을 고려하여 분석한다

5.1 Analysis overview

[Fig 8]과 같이 가해 운전자와 피해 운전자 간 가해자-피해자 관계에 의해 형성되는 사이클 구조를 혐의 패턴으로 추출하는 모델이다.

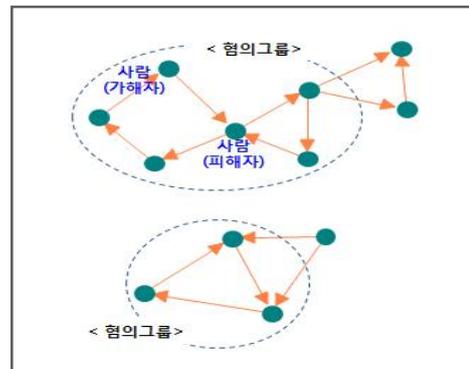


Fig. 8. Relation chart of conspiracy charge between the Driver at Fault and the Victim

5.2 Analysis model

사이클 구조 추출시 가해자와 피해자의 관계의 방향을 고려하여 추출한다. 혐의 그룹을 구성하는 개체간 사고 관계수가 많을수록 공모혐의가 높고, 혐의그룹 내에 포함되어 있는 최초사고일자와 최근사고일자 차이가 짧을수록 단기간에 혐의그룹이 형성되었음을 파악할 수 있다.

① (Extraction of the relations between the driver at fault and the victim) Setting relations between objects and target analysis period; extracting accident relations among the drivers involved in accident (driver at fault and victim)



② (Extraction of the cycle structure) Extraction of only the relations where vehicles at accident are linked by a cyclic structure (directed cycle)



③ (Cohesive structure analysis) Extraction of suspicious groups connected in series by implementing weak components, only the groups with three or more nodes (conspiring relationship)



④ (Analysis of unit index by suspicious group and detailed data on benefits payment) Extraction of data with highly suspicious groups by analyzing the data by group on the number of people, accidents, the cumulated generated benefits (the maximum frequency and maximum amount), and the duration in which the group was formed (the difference between the first accident date and the most recent accident date)

6. Conspired Insurance Fraud between Vehicles and Body Shops

여러명이 다수의 차량을 서로 공유하면서 고의로 사고를 유발하는 행위와 차량을 정비하는 정비업체가 함께 공모한 혐의 행위를 분석하는 기법으로 관계도를 확장하여 공모 혐의그룹을 추출하는 기법이다.

6.1 Analysis overview

차량간 사고 관계에 의해 형성되는 사이클링 구조를 혐의 패턴으로 추출하여 1차 혐의그룹을 분석하고 1차 혐의그룹 (Fig 9 참조)에 포함된 2대 이상의 차량이 수리를 받은 것으로 연결되는 정비업체를(Fig 10 참조) 공모 혐의그룹으로 추출하는 모델이다. [Fig 11]는 차량간 공모 혐의그룹 추출 후 정비소를 포함하여 관계도를 확장·분석하여 매개가 되는 정비업체를 공모 혐의그룹으로 추출과정을 단계별 그림으로 나타내었다. 즉, 정비업체와의 공모 관계 분석을 위해서는 차량 간 사고 관계에 의해 형성되는 공모 혐의 그룹 추출 작업이 선행되어야 한다.

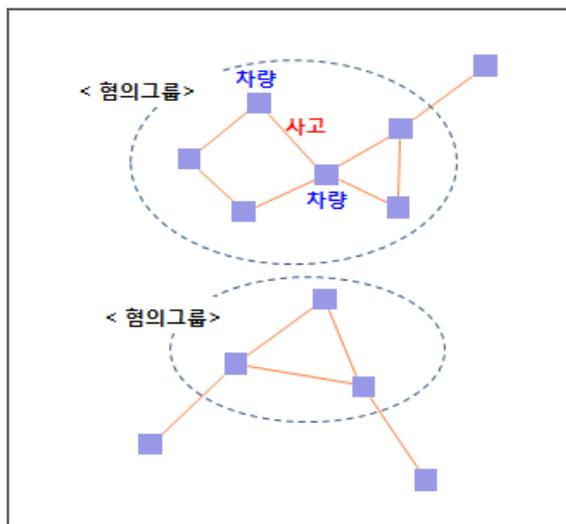


Fig. 9. Relation chart of conspiracy charge between vehicles at accident

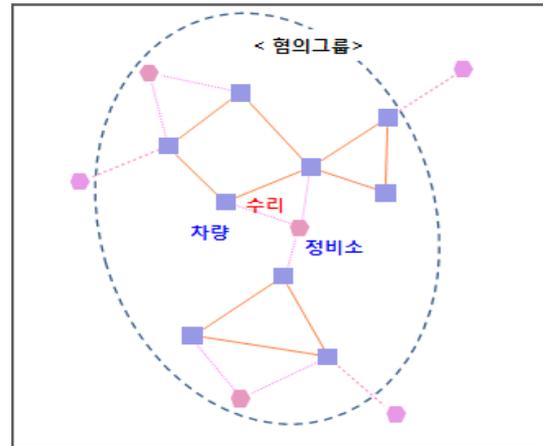


Fig. 10. Relation chart of the vehicle accident conspired by body shops

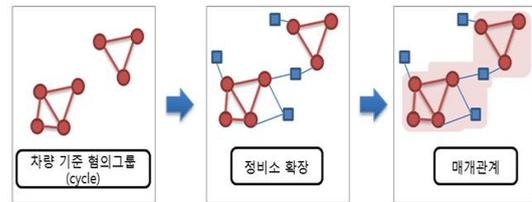


Fig. 11. Connection chart of the conspired charge between vehicle and body shops

6.2 Analysis model

차량간 사고관계에서 사이클 구조를 형성하는 그룹을 추출하여 1차 혐의 분석을 한다. 1차 혐의그룹 관계도에 정비업체를 포함하여 관계도를 확장하고 사고 차량과 정비업체와의 연결이 연쇄적으로 이루어진 집단을 혐의그룹으로 추출한다.

① (Extraction of accident relations between vehicles) Setting relations between objects and target analysis period; extracting accident relations among the vehicles (directionality is not considered)



② (Extraction of the cycle structure) Extraction of only the relations where vehicles at accident are linked by a cyclic structure (undirected cycle)



③ (Cohesive structure analysis) Extraction of suspicious groups connected in series by implementing weak components, only the groups with three or more nodes (conspiring relationship) - primary suspicious group



④ (Expansion of the relations including body shops) Extraction of body shops (using weak components) where vehicles in the primary suspicious group were repaired



㉟ (Analysis of unit index by suspicious group and detailed data on benefits payment) Extraction of data with highly suspicious groups by analyzing the cumulated data by the primary suspicious group on the number of people, accidents, and the cumulated generated benefits (the maximum frequency and maximum amount) and by the secondary suspicious group on the number of vehicles, body shops, repair relations, and the duration in which the group was formed (the difference between the first accident date and the most recent accident date)

V. Conclusions

SNA는 사고정보에 대한 개체간의 관계 구조 분석으로 보험사기 공모 혐의그룹 축출이 가능하여 조직형-공모형 보험사기에 효과적으로 대응이 가능한 분석기법이다. 기존 통계적 접근 방법이 각 개체의 특성 자체를 분석하는데 초점이 있다면, SNA는 개체간의 관계 구조를 분석하는데 초점이 있다. 보험업에 사용되는 데이터는 다양한 개체들의 관계로 구성되어 있기 때문에 SNA를 활용함으로써 기존의 통계 분석이 제공하지 못하는 분석 결과를 얻을 수 있다. 그 동안 공모형태의 보험사기 혐의자에 대한 적발은 제보 등을 통해 접수된 혐의자를 중심으로 보험사고 관계를 분석하는 등 조사자의 판단에 크게 의존해 왔다. 그러나 SNA를 활용한 보험사기에 대한 분석이 이루어질 경우 제보 없이도 혐의그룹을 자동으로 추출하고 주 혐의자를 식별해내고, 특정 그룹의 유사성, 응집도 등을 분석하여 공모형태나 조직형 보험사기에 대한 조사업무의 효율성을 획기적으로 개선할 수 있고 궁극적으로 보험금 누수를 차단하여 이로 인한 혜택이 보험계약자에게 돌아갈 것으로 기대된다. 다만, SNA 활용한 보험사기 혐의 분석을 하더라도 보험사기가 성립하기 위해서는 고의에 의한 재산적 이득을 얻었음을 입증하여야 하므로 보험금 지급 내역 등에 대한 추가 분석과 수사가 필요하다. 서두에 언급한 바와 같이 최근 보험사기는 의료인, 보험설계사 등 분야별 전문가와 보험가입자가 공모하여 계획적이고 은밀하게 이루어지고 있어 그 혐의를 입증하는 것이 점점 더 어려워지고 있다. 이러한 보험사기는 보험금을 수령하기 위해 혈연, 지연 등의 인적 신뢰관계를 악용하여 귀중한 생명과 재산을 고의적으로 살상하고 훼손할 뿐 만 아니라 모방 범죄를 유발하여 불건전한 사고를 사회에 전염시킴으로써 건전한 윤리의식과 생명존중의 가치관을 파괴하고 있다. 최근 빈발하고 있는 조직형-공모 형태의 보험사기에 효과적으로 대응하기 위해서는 보험사기 혐의 주체간 상호연

관성을 분석하여 혐의그룹을 추출하는 SNA 기법에 대한 활용 및 연구를 강화 할 필요가 있다.

REFERENCES

- [1] Gwangju metropolitan police agency, "Insurance crime white paper", 2017.12
- [2] Financial supervisory service press release, " Insurance fraud detected 800 billion won and reported reward 2.4 billion won in 2018", 2019.4.23.
- [3] Mi-Suk Park, "The Reform Measures for Investigation in the Area of Financial Unlawful Act", Korean Institute of Criminology, 2002.12.
- [4] Jong-Seong Eun, "A study on the aspects of insurance and measures to cope with", Korea Business Law Association, pp. 13-38, 2001.10.
- [5] Semin Park "A Study on the Analysis on the Current Countermeasure for the Insurance Fraud and Its Reform Measure" Korean legal Center pp. 140-177, Jun. 2009
- [6] Seong-Hee Kim and Rho-Sa Chang, "The Study on the Research Trend of Social Network Analysis and the its Applicability to Information Science" Journal of The Korean Society for Information Management, pp. 71-87, Dec. 2010
- [7] Sang-kuk Kim "Social Network Analysis and Industrial engineering" Industrial Engineering Magazine, pp. 24-32, Mar. 2011
- [8] Uchang Park , "Relational Approach for Social Network Analysis" Journal of KIIT, pp. 143-151, Sep. 2012
- [9] Sung-Je Park, Je-Uk Lee "Big Data era of social network analysis techniques and the use of the sports field strategy " The Korea Journal of Sports Science," pp. 933-946, Vol. 23. 2014
- [10] Jimin Park and Seong-Woong Yoon and Sang-Hoon Lee "Finding Relevant Peoples in the Gated Community based on Social Network Analysis" Korea Information science society, pp. 150-152, Dec. 2014
- [11] Min-Sik Lee , "Co-Offending Structure in Drug Crime : A Network Approach " Korean Institute of Criminology , pp. 231-263, Jun. 2007
- [12] Soojung Lee, "Improving Performance of Jaccard Coefficient for Collaborative Filtering", Journal of the Korea Society of Computer and Information 21(11), pp.121-126, 2016.11
- [13] Minjae Kim, Sangjin Lee, "Measures of Abnormal User Activities in Online Comments Based on Cosine Similarity", Journal of the Korea Institute of Information Security & Cryptology 24(2), pp. 335-343, 2014.4
- [14] Sung-Je Park, Jea-Woog Lee, "Structure and use knowledge map of sports field through text mining and social network analysis",

Korean Journal of Sports Science 26(4), pp.639-653, 2017.8

- [15] Jay-Bok Lee, "Legislative Measures to Combat for the Prevention of Life Insurance Crime", Korea Business Law Association, pp. 111-143, 2001.10
- [16] Wasserman, Stanley, Faust, Katherine, Granovetter, Mark, "Social Network Analysis- Methods and Applications", Cambridge Univ, 1994.11

Authors



Tae-Ho Kim received the M.S. degrees in the Department Information security from Korea University, in 2009. Currently he is a Ph. D. student of Graduate School of Information Security at Korea University. His current

research interests include Information Security Policy, Financial Fraud Detection, Electronic Financial Security.



Jong-in Lim received the BS, MS, and Ph.D. degrees in the Department of Mathematics at Korea University. Seoul in 1980, 1982 and 1986. Currently he is a professor of the Graduate School of Information Security at

Korea University. Also, he is a former Special Advisor to the President for National Security, Republic of Korea. His main research interest include National Cyber Security, Cyber Warfare, Convergence Security, Privacy and Cryptography.