

Data mining 기술을 활용한 보험사기 적발에 관한 연구

Applying data mining technologies to insurance fraud detection

방 명 하 *

< 차 례 >

I. 서 론

II. Data mining에 대한 이론적 고찰

III. 보험사기와 사기적발 시스템에 대한 이론적 고찰

IV. 데이터마이닝 기반의 보험사기 적발시스템의 활용 및 기대효과

V. 결 론

I. 서론

1.1 배경

최근에는 데이터마이닝이란 단어가 아직 보편화되어 있지만 그렇다고 아주 생소하지만
도 않다. 오늘날 경영환경의 복잡성과 불확실성은 그 어느 때보다도 더욱 심각한 지경이다. 실례로
최근의 테러와 각국의 이해관계로 인한 전쟁은 세인들을 충격의 도가니로 몰아넣을 뿐이 아니라 기
업인들의 경제활동마저 심각하게 위축시켜 결과적으로 불확실성 패닉 상태로 까지 몰고 가고 있다.

어느 마케팅 회사는 과거의 데이터를 이용한 모형을 가지고 대 고객에 대한 연구를 게을리 하지
않고 있다. 최근에 재정부 산하 한 조직에서는 누적되어지는 재무거래의 내용을 그 어떤 유사점 또
는 유사패턴을 찾아 자금세탁 또는 마약사업과 같은 범죄관련 재무거래인지 여부를 찾아내고 있
다. 이러한 모든 예측 및 적발 활동들의 효율성과 정확성을 높이기 위해 이미 데이터마이닝을 사용
하고 있다.

정보기술의 발달은 기존에 처리하기 어려웠던 내용량의 데이터 또는 복잡한 문제에 데이터마이
닝의 활용을 더욱 가속화하는 계기가 되었으며, 적용되지 않는 분야가 없을 정도로 그 활용범위 및

* 경영학부 교수

한계의 벽이 빠른 속도로 허물어지고 있는 실정이다. 매일 산더미처럼 쌓여지는 기업의 데이터로부터 기업의 전략적 의사결정을 위해 필요한 옥석같은 정보를 추출해내는 것을 지원하는 기술을 말하며, 이는 정보 활용의 최고 형태라는 이유 때문에 널리 각광을 받고 있다.

데이터마이닝은 다양한 기법을 활용하여 얻어지는 결과를 많은 경영활동 등에 활용되어 지는데 데이터마이닝의 그 목적은 데이터에 숨어 있는 패턴 또는 지식을 추출하는 일련의 과정으로서, 의사결정자의 문제해결 및 의사결정을 돕기 위한 기계적 접근법이라 할 수 있다.

특히, 데이터마이닝의 한 분야이면서 국내에 거의 잘 알려지지 않은 Visual data mining이야말로 데이터 내 숨겨진 의미 있는 패턴 및 정보를 찾아가는 일련의 프로세스에 시각적 기법을 적용하는 새로운 기법이다. 이는 보다 복잡한 관계로 표현되는 데이터 간의 연관성을 보다 종합적이며 그리고 체계적으로 정보를 제공하는 유용한 기법이다.

보험사기 적발과 같은 작업에 Visual data mining과 같은 기법이 효율적이며 효과적일 지를 알아 보려면 우선 보험사기와 보험사기 적발에 관한 지식이 Visual data mining에 대한 지식과 함께 필요할 것이다. 우선적으로, 보험사기 적발 업무는 보험사기에 대한 전문가의 오랜 경험으로 획득한 지식과 보험사기를 인정하게 만드는 수월한 논리력과 냉철한 분석력 및 판단이 상당히 많이 요구되어 지는 속성이 있다. 특히, 보험사기는 일반적으로 단독 범행으로 행하여지는 경우보다는 병원, 정비소, 설계사 등이 함께 공모하여 사기를 저지르는 게 보편화 되어 있는 게 현실이다. 종래의 적발 방법만으로는 사기에 연관되어진 복잡한 데이터들을 활용하기에 매우 제한적이며, 사기혐의자들을 적발하기에도 한계점이 있을 수 있다.

1.2 논문의 목적

본 연구는 데이터마이닝의 한 분야이지만 국내에는 거의 알려지지 않은 Visual data mining을 활용한 보험사기적발 시스템을 심도깊게 이해하는데 그 목적을 두고 있다. Visual data mining은 다양한 Visualization의 기법을 사용할 수도 있고 다양한 알고리즘을 또한 사용할 수도 있다. 다만, 데이터들의 관련성을 분석함에 있어 시각화의 기능을 극대화하는 방법 중 가장 대표적인 방법인 연관성 분석 (Link Analysis)알고리즘을 활용한 Visual data mining을 활용하며 이는 데이터 속에 숨어 있는 패턴이나 추세를 추출하기 위함이다. 아울러, 보험사기 적발과 같은 분야를 채택한 이유는 다른 어떤 분야보다도 복잡한 패턴이나 지식이 요구되어지는 분야이다. 보험사기는 개인이나 집단의 범죄이기 이전에 탈세와 같은 사회적인 범죄이며 보험사기는 결국 보험사의 손해를 과 보험 요율에 지대한 영향을 미치기 때문에 사회의 구성원에게 피해를 미칠 수 있는 속성 때문이다.

II. Data mining에 대한 이론적 고찰

2.1. 지식발견

2.1.1 지식발견의 의의

기업에 있어서의 기업 경쟁력은 그 기업이 보유하는 다른 많은 요소들도 있겠지만 그 중에서도 인간의 두뇌를 통해 얻어지는 지식이 새로운 핵심 요소로 부상하며 이는 곧 기업과 개인의 경쟁력에 변화를 미친다.

자본주의 역사상에 등재했던 갑부들은 21세기에는 거의 그 자취를 감추었다. 더 이상의 철강왕보다는 마이크로소프트와 같은 회사의 빌 게이츠 등이 갑부 리스트의 상위에 올라가 있다. 이는 소프트웨어나 정보통신과 같은 정보기술 분야의 기업주만 올라가 있는 것만은 아니다. 판매 정보와 마케팅 관련 지식의 실시간 공유로 유통업계의 신화를 창조한 월마트의 샘 월튼도 갑부로 등재되어 있다. 피터 드럭커(1980)는 지식사회에서 기업의 가치는 지식을 전략적으로 획득, 생성, 분배 및 적용할 수 있는 능력에 의해 영향을 받으며 이러한 지식이 얼마나 부가가치를 창출하느냐에 따라 주식시장에 반영이 된다고 했다.

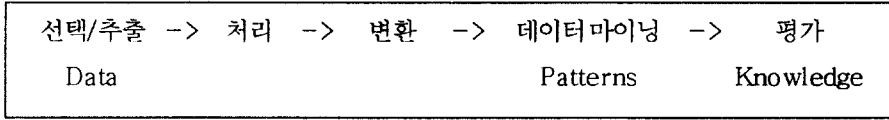
지식발견(Knowledge Discovery in Databases, KDD)은 데이터로부터 유용한 정보를 추출하는 프로세스의 전 과정을 일컫는다(장남식, 홍성완, & 장재호, 2000). 지식발견과 데이터마이닝에 관한 국제학술대회(Knowledge Discovery & Data mining, 1996)에 의하면 지식발견은 'non-trivial process of identifying patterns in data'라고 규정하며 데이터는 통계적으로 유효해야 하며, 사용자에게 의해 선호되는 새로운 시스템, 특정 목적에 유용한 그리고 이해할 수 있는 속성을 지녀야 한다.

2.1.2 지식발견의 프로세스

일반적으로 지식발견은 데이터로부터 유용한 정보 및 패턴을 발굴하는 일련의 과정을 거쳐야 하는데 그 일반적인 과정은 필요로 하는 정보를 충분히 정의하고 이를 위해 요구되는 원천 데이터의 성격과 소재, 그리고 충실도를 파악하는 준비 작업을 거친 후 진행 되어진다. 다음 그림2.1은 지식발견 프로세스의 단계를 도식화하여 보여주는 그림이다. 지식발견 프로세스는 한 번 거치고 나면 마는 그런 일회성 과정이 아니라 양방향이며 반복적인 작업의 연속이다. 다시 말하면, 데이터마이닝 단계에서 데이터에 오류 값이나 결손 값이 의외로 많을 경우에는 반복적인 데이터의 정제작업이 요구될 수 있으며, 데이터의 특성을 대표할 만한 속성들이 모자란다고 판단될 경우에는 데이터

의 재보완을 고려해야 한다.

[그림 2-1] 지식발견 프로세스



첫 번째 프로세스는 주어진 데이터에서 목표 데이터(Target data set), 데이터 항목(속성)(Data attributes), 데이터 샘플(Samples) 등을 추출하고 정의하는 단계이다. 지식을 추출하기 위해 선행되어지며 매우 중요한 과정이다. 예를 들면, 충성고객(loyal customer)의 특성을 분석하는 것이 지식발견의 목표라면, 과연 충성고객의 정의는 무엇을 할 것이며, 더 나아가 고객은 어떻게 정의할 것인가에 대한 부분을 포함한다. 또한 정의된 충성고객의 특성을 나타내는 데이터는 거래, 프로파일, 신용 등의 관련 테이블에서 어떤 항목(속성)을 이용하여야 할 것인가를 판단하며, 전체 데이터를 모집단에서 전수 조사를 할 것인지 샘플링을 추출된 샘플을 통해 지식발견을 수행할 것인지를 규정하게 되는 단계라 할 수 있다.

지식발견 프로세스를 통해 산출되는 정보의 품질은 정보의 원천이 되는 통합 데이터의 충실도에 매우 영향을 받는다. 일반적으로 데이터의 충실도란 데이터의 정확도, 데이터의 양(레코드의 수), 그리고 데이터의 깊이(항목의 수)에 의해 좌우되는데, 처리단계에서는 목표 데이터 및 항목들의 노이즈(noise)를 최소화하는 작업이 수행된다. 데이터의 정제 작업과 보완이 주로 이루어지는 단계로 데이터의 정확도뿐만 아니라 데이터의 양과 깊이를 높이므로 정보의 품질을 향상시키는데 기여를 하는 단계이다.

데이터 변환 단계에서는 원하는 목적을 달성하기 위해서 정의된 데이터에서 유용한 항목의 특성 파악 및 데이터에 불필요한 레코드와 항목을 삭제하는 작업이 시행된다. 그러나 결손 값을 다수 포함하는 레코드나 항목들의 삭제가 이루어지는데, 이는 향후 도출되는 정보에 지대한 영향을 미칠 수 있으므로 최종 결정에 앞서 충분한 검토가 미리 이루어져야 한다. 즉, 충성고객의 특성을 분석하기 위한 각각의 항목 값들이 어떠한 분포를 가지고 있으며, 충성고객 분석을 위해 가장 효율적인 데이터의 수 등을 주로 점검하게 된다.

데이터마이닝 단계는 지식발견의 핵심적 과정으로 분석 주제에 맞는 데이터 집단(Data set)에서 관심 있는 특정 패턴 및 지식을 찾아내는 과정으로 대표적인 데이터마이닝 작업으로는 분류(Classification), 예측(Estimation), 군집분석(Clustering), 연관관계분석(Association) 등을 들 수 있다. 사용자들은 데이터 분석의 목적, 수집된 데이터의 성격, 산출되는 정보의 설명력, 사용의

용이성 등 여러 가지 기준을 고려하여 적절한 기법을 선정하여야 한다.

지식발견의 마지막 프로세스인 모형의 평가는 데이터마이닝 기법을 이용하여 구축한 모형이 과연 실제로 현장업무에 적용하기 적합한가를 판단하는 작업이라 할 수 있다.

2.1.3 지식발견과 전통적인 전문가시스템

인공지능의 한 분야로서 각광을 받았던 전문가시스템은 특정 분야에서 전문가의 오랜 경험과 지식을 토대로 하여 시스템을 만든 것으로, 세미 전문가일지라도 필요할 때 마치 전문가처럼 시스템을 사용하도록 해 주는 소프트웨어이다.

이러한 시스템은 1960년대에 만들어진 Dendral 전문가시스템이 효시로 여겨지며 그 목적은 화합물의 구조를 추정하기 위함이었다. 그 후로도 다 수의 전문가시스템이 시장에 소개되었는데 의료진단시스템인 MYCIN, 광물탐사시스템인 PROSPECTOR 등이 1970년대에 만들어졌었다. 특히, 기계 고장 진단, 설계 지원, 유전 진단, 손해 배상 판정 등 산업계의 전 분야에서 광범위하게 소개되었다.

다만, 70년대만 하더라도 정보기술은 일부 연구소나 학계 등 특정 조직에서만 주로 사용이 되었고, 또한 데이터를 저장하고 처리하는 성능 또한 한계가 있었다. 하드웨어 및 소프트웨어 측면에서의 한계 때문에 전문가시스템의 보편화는 다소 시간을 요구했다.

그러던 것들이 오늘날에는 일반 사용자들이 사용할 수 있는 개인용 컴퓨터가 많이 보급되었고, 기가 바이트급 저장장치나 300MHz이상의 CPU가 개인용 컴퓨터에 장착되고 있다. 하드웨어의 발전과 더불어 지능을 갖춘 데이터 분석시스템을 개발하기 위한 노력이 계속되어 왔다. 미리 정해진 순서와 절차에 따른 문제 해결을 하는 전통적 프로그래밍 방식과는 다른 주어진 데이터로부터 추론이나 직관적 판단을 통해 스스로 문제 해결을 하는 학습능력(learning ability)을 갖춘 알고리즘이나 프로그램들이 주요 대상이었다.

전문가시스템의 핵심 요소인 지식베이스(Knowledge-base)에는 전문가와 인터뷰나 연구논문이나 전문서적과 같은 간접 경로를 통해 지식을 획득했다면, 오늘날의 기업이 안고 있는 문제는 다 방면에서의 전문적인 지식이 필수인데, 각각의 해당 분야에 통달한 전문가를 찾기가 점차 어려워졌으며, 지식획득 작업마저 장시간이 소요될 뿐만이 아니라 많은 비용이 발생하는 어려운 문제에 봉착하였다. 이것을 소위 '지식획득의 병목현상(Knowledge acquisition bottle-neck)'이라고 하며, 이로 인해 데이터를 분석하여 자동적으로 지식을 추출해 내는 방안에 대한 요구가 증가하는 추세이다 (장남식, 홍성완, & 장재호, 1999).

2.1.4 지식발견과 데이터마이닝

지식의 중요성에도 불구하고 데이터의 홍수와 지식의 빈곤이라는 환경에 처한 기업들은 정보기술을 이용하여 데이터를 여과하고, 분석하며, 결과를 해석하는 데이터 분석 방안에 높은 관심을 가지고 있다.

데이터마이닝은 지식발견 프로세스 중에서 데이터로부터 정보를 추출하기 위해 기법을 적용하는 핵심 단계이다. 특히 데이터마이닝 기법을 선정하여 데이터에 적용하는 과정은 지식발견프로세스 중에서도 가장 흥미로운 단계임이 분명하다. 지금까지 알려진 데이터마이닝 기법들은 그 종류가 매우 다양하며 최근의 관심도 증가로 새로운 기법들이 대학과 연구소 등을 통해 지속적으로 연구되고 있다. 이 중에서도 이론적인 검증을 거쳐 상품적 가치를 인정받은 일부 기법들은 소프트웨어 개발자에 의해 상용화되어 시장에 진출하고 있다.

데이터마이닝 기법 중 가장 적절한 기법을 자신의 과제에 적용하는 데이터마이닝 기법을 선택하는 작업은 그리 쉽지 않은 않다. 그 이유는 데이터마이닝 작업 유형에 관계없이 가장 탁월한 성능을 제공하는 특정 기법이 존재하는 것도 아니고 유사 기법이라고 하여도 분석대상이 되는 데이터의 특성이나 도출하고자 하는 정보의 성격에 따라 상이한 결과를 낼 수 있기 때문이다.

종래의 데이터마이닝 기법으로 판별분석(Discriminant Analysis), 군집분석(Clustering Analysis), 회귀분석(Regression Analysis)과 같은 통계적 분석기법과 인공신경망(Neural Network), 사례기반추론(Case Based Reasoning), 의사결정트리(Decision Tree) 등과 같은 인공지능 기법이 널리 사용되고 보편화되어 있다.

그 중에서도 가장 보편화되어 있는 기법들은 회귀분석(regression analysis)과 판별분석(discriminant analysis) 등이 있으며 이러한 전통적인 통계기법은 이미 널리 알려져 있다. 최근에는 인공지능(artificial intelligence)에 기반을 둔 기법들 중에서 의사결정나무(decision tree)와 신경망 기법(neural networks) 등이 데이터마이닝의 대표적인 기법으로 연구가 활성화 되고 있다. 실제로 의사결정나무나 신경망과 같은 기계학습(machine learning)에 근거한 기법들에 대한 활발한 연구가 시작된 원인 중의 하나가 전통적인 통계기법을 통한 데이터 분석의 한계이다. 물론 전통적 통계기법들은 오랜 기간동안 검증된 기법으로 변수들 간의 상관관계를 발견하고, 목표변수에 영향을 미치는 주요 속성들을 선별하고 각각의 속성이 목표변수에 대한 영향력을 해석 가능한 수식으로 제공하는 능력 및 이를 설명하는 능력은 타 기법에 비해 월등하다. 반면 데이터 값들이 정규분포, 공분산(covariance) 등과 같은 여러 통계학적 가정을 요구하는 경우가 많고, 데이터에 범주형(categorical) 변수가 포함되어 있는 경우 이들을 가변수(dummy variable)로 변환시켜야 하는 과정에서 발생할 수 있는 치우침(bias) 등의 한계를 가지고 있다.

2.2 Visual data mining

2.2.1 지식발견과 Visualization

최근 들어 데이터마이닝에 대한 관심과 수요가 늘어남에 따라 다양한 분야에서 많은 연구가 이루어지고 있다. 지식발견은 기존의 조회 도구나 전통적인 데이터마이닝 기법을 통해서도 가능하지만 알고리즘을 통해 밝혀진 패턴이나 결과는 사용자의 화면에 텍스트 형태로 주로 표현되어 있는데, 일련의 연관성 정보를 제공한다든지 하는 패턴과 관련된 정보를 표현하는 데는 매우 제한적일 수밖에 없다. 따라서 패턴과 관련된 일련의 정보를 시각화(Visualization) 하는 것이 매우 바람직할 수 있다. 시각화는 데이터와 정보 및 데이터 속의 패턴을 사용자에게 전달하는 사후 프로세싱(post-processing)의 커뮤니케이션 채널로 활용된다. 결국, 효과적인 시각화는 지식발견 과정을 통해 제시되는 지식을 보다 질적으로 향상 시키는 데 기여를 할 수 있다. 특히, 지식발견 과정에서의 시각화 기법의 활용은 산출된 패턴의 질적 이해를 높일 수 있으며, 현존하는 문제 영역에 대한 조사를 보다 광범위하지만 철저하게 할 수 있게 되어 결과물에 대한 신뢰를 높일 수 있다.

2.2.2 Visual data mining과 Link Analysis

Visual data mining은 지식발견 과정에서 해석 가능한 패턴을 추출하기 위해 사용자와 컴퓨터 사이에 의사소통의 채널로 시각화(Visualization)를 사용하는 것이다 (Ankerst,2000). Visual data mining은 매우 강력한 분석방법으로서 패턴이나 추세를 파악하는데 있어 기존의 비시각적인 방법에서 놓칠 수 있는 지식을 발견할 수 있는 면에서 보면 그 효과는 매우 탁월하다 (Westphal & Blaxton, 1998).

Visual data mining은 다양한 Visualization의 기법을 사용할 수도 있고 다양한 알고리즘을 또한 사용할 수도 있다. 다만, 데이터들의 관련성을 분석함에 있어 시각화의 기능을 극대화하기 위해 데이터 속에 숨어 있는 패턴이나 추세를 추출하기 위해 연관성 분석 (Link Analysis) 알고리즘을 활용한 Visual data mining을 활용한다.

Visual Data Mining을 Visualization에 중점을 두기보다는 알고리즘을 중심으로 한 Visual Data Mining을 소개 하고자 한다. 특히, Quantitative Data Mining과 대조적으로 명목형 data를 활용하고 속성값이 중심이 아닌 이들의 관련성을 분석 및 시각화함으로써 데이터 속에 숨어 있는 패턴 또는 지식을 추출하기 위한 Link Analysis를 활용한 Visual Data Mining을 소개하고자 한다.

1) Link Analysis에 대한 이론적 고찰

(1) Link Analysis의 배경 및 정의

Link Analysis는 Social Network Theory에 근간을 두고 있는 분석 기법으로 ‘인간은 다른 사람과의 관계가 없다면 존재할 수 없으며, 만일 주변에 어떠한 사람도 없고 다른 주변 사람과 관계를 가지고 있지 않다면 그곳에는 아무것도 없다’란 사상에 기본적인 배경을 두고 있다. 특히, 링크(Link) 또는 네트워크(Network)는 개체의 속성보다는 개체간의 관계가 보다 중요하며, 세상의 모든 현상은 링크 또는 네트워크적 문제라는 것이다.

위와 같은 Social Network Theory에 근간을 둔 Link Analysis는 ‘데이터 속에서 패턴 또는 흐름을 추출하기 위해 서로 연관되어 있는 개체(objects)들의 관계(Relationships)를 네트워크화 하는 과정이라 정의되며, ERD(Entity-Relationships Diagrams), Nodes-and Links, Directed Graph등의 용어로 사용되기도 한다.

임의로 정의된 엔티티(Entities)간의 연관성을 분석하는 Link Analysis는 다양한 분야에서 활용되고 있다. 예를 들면, 웹사이트 링크분석(Search engines, site maps), 네트워크 트래픽 분석(Web caching, public transport), 데이터마이닝(e-commerce, telecommunication services), 사회적 네트워크 분석(Social structures, policy marketion), 텍스트 분석(conference, cocitation), 의사결정지원(financial marketing, logistics), 사기적발(money laundering, calling cards)등에 활용되고 있다. 이러한 Link Analysis가 다양한 분야에 사용되는 것은 대용량 데이터 상에서 각 Entities간의 연결관계(Link)를 탐색하고, 명확히 정의하는 수단으로 매우 효과적이기 때문이다. 즉, 특정 필드(Field)에서 동일한 값을 가지고 있는 모든 레코드(Records)는 서로서로 연결되어 있음을 생각 할 수 있기 때문이다.

[표 2-1] Link Analysis활용 분야

적용분야	데이터	적용분야 예	연구자
금융 거래	금융거래 데이터, 프로파일	Money Laundering	Henry 외(1998년)
통신 네트워크	통신 내역, 프로파일	Telecommunication Fraud	Ciara(200년)
business Intelligence	고객, 서비스, 공급자, 파트너, 거래내역	비공식 네트워크 분석, 서브그룹	Ming 외(2000년)
웹사이트 검색	페이지, 콘텐츠	Hyperlink-Induced Topic Search	-
GIS	교통흐름 데이터	병목구간, 최단구간	-
범죄/사기 조사	거래내역, 통화정보	보험사기, 범죄자 적발	-

(2) Link Analysis의 구성과 특징

Link Analysis는 Link와 Node, Value, Direction으로 구성되어 있다. 여기서 Node는 임의의 Entities로 표현되어 있는데, 이의 의미는 데이터베이스의 Attribute를 Entities로 지정할 수 있음을 의미한다. 다만 정량적인 숫자는 Entities가 될 수 없으며, 일반적으로 문자형 타입(우편번호, 주소, 이름 등)을 가지는 속성값이 Entities로 사용된다. 또한 Value는 두개의 Entities 간의 관계의 정도(빈도) 및량을 의미하는 것으로, 전화통화 횟수, 시간이 좋은 예라 할 수 있다. 또한 Direction은 두 개의 Entities 간의 방향을 의미하는 것으로, 전화통화에서 발신지와 착신지를 나타내는 화살표를 생각할 수 있다.

[표 2-2] Link Analysis의 구성 요소

구 성	의 미
Node Set	임의의 Entities의 Set
Link Set	짝을 이룬 노드들의 Set
Link Value	가중치
Kirection of Links	순서 혹은 방향

Social Network Theory에 이론적 배경을 두고 있는 Link Analysis는 각 개체, 각 개체의 속성에 관심을 두기보다는 각 개체의 관계에 중점을 두고 있어 랜덤 샘플링이 가능하지 않으며, 각 개체는 독립성을 가지고 있지 않는 데이터의 특성이 있다. 이러한 Link Analysis의 특성을 기존의 Quantitative Data Mining의 대표적인 방법론인 통계와 비교하면 다음과 같다.

[표 2-3] 통계와 Link Analysis의 비교

통 계	Link Analysis
Attributer(속성값)	Relational(관계)
Correlation(상관관계)	Topololgy(기하학, 위상)
Statistial(통계적)	Mathematical(수학적)
Random Sampling	Not Random Sampling

Link Analysis에 근간을 둔 독립적이지 않은 즉, 네트워크(Network) 데이터들의 관련성을 분석하는 것으로, Visual Techniques을 활용하여 결과를 분석하는 것이 매우 효과적인 것으로 나타나고 있다. 즉, Visualization은 복잡한 네트워크 구조와 기대하지 않은 패턴의 명확화에 있어 인간 의사결정에 매우 효과적임을 증명하였다. 다음의 그림은 전화 통화내역의 데이터를 시각적으로 표현한 것으로 테이블로는 분석하기 힘들었던 전화통화가 많았음을 시각적으로 인지할 수 있다.

[그림 2-2] Link Analysis와 Visual Data Mining

(3) Link Analysis를 활용한 Visual Data Mining의 방법

Link Analysis를 활용한 Visual Data Mining의 방법은 Link Analysis의 Node와 Node들간의 관계(Links)를 시각화(Visualization)한다는 의미뿐만 아니라, 분석적 관점에서 패턴의 추출 및 결과의 해석을 포함하는 방법을 의미한다.

① Articulation Points

만일 하나의 엔티티(Entity)가 두 개 혹은 그 이상의 하위 그룹들과 연결되어 있다면 이러한 Articulation Points는 전체 네트워크 구조에서 매우 중요한 역할을 하고 있음을 짐작할 수 있으며, 해당 Entity가 제거된다면 전체적인 구조에 상당한 영향을 미치게 된다. 이러한 Articulation Points는 Bottleneck을 의미하는 것으로, What-If Analysis를 통해 시나리오 분석 또는 그룹 내에 미치는 영향 정도를 분석할 수 있으며, 해당 Point의 중요성은 데이터를 표현하는 모델로부터 결정되게 된다. 통신사기를 예를 들면, 전화의 Call 데이터를 모델링하고 Link Analysis를 하였을 경우, 만일 특별한 전화번호가 다른 일련의 전화번호 그룹들과의 연결되어 있는 Articulation Points라면 이 특수 전화번호는 통신사기 적발에 매우 유용한 단서가 될 것이다.

② Discrete Network/Missing Connections

데이터 셀에 포함되어 있는 하위 그룹들을 분류하는 방법으로 통계적 분석의 Clustering과 유사한 개념적 성격을 가지고 있다. 즉, 일련의 복잡하고 큰 네트워크 구조를 가지고 있는 조직 내부에는 다양한 하위 그룹들이 포함되어 있게 된다. 예를 들면 하나의 기업체에는 다양한 부서, 직군 등이 포함이 되어 있다. 만일 분석하고자 하는 주제의 특성에 따라 기업체는 부서별 군집과 직군별 군집 등 다양한 형태의 하위 그룹들로 표현이 될 수 있다. 또는 전체 데이터 셀에서 데이터 셀이 어떠한 특성으로 구성되어 있는지를 분석할 수 있어 특정 목표가 아닌

탐색적 분석에도 매우 가치 있게 활용될 수 있다. Missing Connection은 Discrete Network의 특별한 형태로 단일 Entity로 하위 그룹이 형성되어 있음을 나타낸다.

③ Strong/Weak Linkage

연결관계의 패턴을 분석함에 있어 네트워크 내에 각 노드들 간의 관계성의 정도를 분석 가능하게 하는 방법이다. 하나의 개체는 특정 개체와 지속적으로 수많은 관계성을 가질 수 있으며, 이러한 관계성은 관계 횟수 및 관계량 등을 표현하게 된다. 예를 들면, 911테러 직후, 주요 핵심 테러용의자 적발을 위해 Link Analysis기법을 활용한다고 가정하였을 경우, 테러자(A)가 테러지도자(B)와 은행 거래 관계를 가지고 있으며, 한번의 은행거래에서 상당한 금액이 오고 갔다면 둘 간의 관계는 매우 강하게 나타나게 된다. 즉, 테러지도자(B)는 테러자(A)뿐만 아니라 상당히 많은 테러자(C, D..) 등과 관계를 가지고 있다고 했을 경우, 주요 관심의 대상은 거래 금액이 큰 테러자(A)가 빠른 시간내 분석이 가능 할 것이다.

④ Pathway Analysis

분석의 목적인 시계열적인 연관 관계를 분석하는 게 주요한 관심사일 경우, 사용되는 방법으로 특정 사이트의 특정 고객에 대한 사용행태(Web Usage Mining)에 적합한 방법이다. 또는 특정 지역에서 특정지역으로의 여행을 계획하였을 경우, 여행 목적지를 최단기간에 찾아가는 방법을 쉽게 알 수 있게 된다.

⑤ Emergent Groups

Emergent Groups은 서로서로 연관되어 있으면서 매우 긴밀하게 관련되어 있는 특수한 그룹을 명확하게 하는 것으로, 수많은 개체들로 구성된 그룹 내에서 핵심적인 개체를 찾아내는 방법을 의미한다. 이러한 Emergent Groups은 일반적으로 다음과 같은 기준을 따르게 된다.

그룹은 적어도 3개 이상의 Node들로 구성된다.

그룹 내 각각의 Entity는 다른 Entity와 최소 2번 이상의 관계를 가지고 있어야 한다.

그룹 내의 하나의 Entity는 해당 그룹 내의 Link중 최소한 일정 비율(50%)이상의 다른 노드와 연결되어야 한다. 특히 Emergent Groups 알고리즘은 Quantitative Data Mining의 특성중의 하나인 자동적인 분류(Automatical Classification)가 Link Analysis방법을 사용한 Visual Data Mining에서 가능하게 하는 중요한 방법으로 사용될 수 있다. 따라서 본 연구에서도 Emergent Groups 알고리즘을 이용하여, 실제 사기 혐의자를 추출하고 이들의 혐의성을 분석하는 검증법도 수행하고자 한다.

III. 보험사기와 사기적발 시스템에 대한 이론적 고찰

3.1 보험사기에 대한 이론적 고찰

3.1.1 보험사기의 역사

세계 최초로 기록이 된 보험범죄는 「이네스」사건으로 이야기되는데, 이는 이네스라는 한 남자가 자신의 양녀를 피보험자로 하여 보험사인 에퀴터블사와 1천 파운드의 생명보험계약을 체결한 뒤에 양녀를 독살하고, 자신을 유산상속인으로 하는 양녀의 자필 유서를 제출한 것에서부터 출발한다. 결국에 유서의 증인 2명 가운데 1명이 위조된 유서임을 폭로해 범행사실이 발각됨으로써, 본 사건은 일단락되었다 (송기철, 1983).

우리나라 최초의 보험사기는 1924년 3월에 발생한 사건으로 당시 보험 모집인 이었던 조모씨가 정모씨 등 2명과 공모하여 대리진단을 통해 중병을 앓고 있던 이모씨의 보험계약을 체결하고, 수개월이 지나도록 사망하지 않자 1923년 10월경 당국에 허위 사망신고를 한 다음 우리나라 최초의 생명보험회사인 조선생명보험으로부터 5천원의 보험금을 편취한 사건을 들 수 있다 (이병희, 2001).

이렇게 보험사기는 부당하게 보험금을 수취하기 위해서 보험회사를 기만하는 행위를 의미한다. 이러한 보험사기는 그 도덕적 위험 (Moral Hazard)의 문제가 보험회사뿐만 아니라 대다수 선량한 보험 가입자들에게 큰 피해를 미치게 되며, 결과적으로는 한나라의 국가 경제에 돌이킬 수 없는 큰 타격을 줄 수 있기 때문에 자본주의 사회의 ‘고요한 대재난 (the Quiet Catastrophe)’으로도 표현되곤 있다.

3.1.2 보험사기와 보험범죄의 개념상 차이

쓰키타리 카즈키요가 쓴 “보험과 범죄 (1997.10)”의 영향으로 보험사기 혹은 보험범죄 연구자들 대부분은 보험사기가 보험범죄보다 넓은 개념 즉, 보험가입시의 악의성을 포함하는 보다 광범위한 개념으로 이해하고 있다 (김형기, 손성동, & 김상의, 1998). 그들은 고지의무위반의 경우 보험범죄는 되지 못하지만 보험사기는 된다고 한다. 그러나 고지의무위반의 구성요건을 보면, 중요사항의 불고지나 부실고지에 있어 악의적인 고지의무위반은 사기에 해당하지만 중대한 과실에 기인한 경우 보험사기를 성립시키지 못하므로 고지의무위반 그 자체가 보험사기가 되지는 못한다.

결국, 보험사기 자체가 범죄행위이므로 보험사기가 되면서 보험범죄가 되지 않는 경우란 없다고 보여 진다. 예를 들어 발목질 단사건처럼 보험금을 수령할 목적으로 범죄를 저질렀으나 보험금을

청구하기 전에 그 범행전모가 발각되어 보험금 청구를 하지 않은 경우 보험사기는 성립되지 않지만 보험범죄로 보는 것에는 대체로 의견을 같이 한다. 발목질단이라는 보험사고와 손해는 발생되었기 때문에 보험범죄에는 해당하나 보험금 청구행위를 하지 않았기 때문에 사기에는 해당하지는 않는다. 그러므로 발목질단을 한 택시운전자는 보험범죄자로 중상해 및 증거인멸 혐의로 처벌되나, 정작 당사자인 보험계약자는 보험사기죄로 처벌되지 않았다. 나아가 공갈, 폭력행위 등을 통해 보험금을 편취하려는 행위, 허위진단서 등을 발급하는 행위, 보험회사의 임직원 등이 보험업무를 수행함에 있어 배임수재를 하는 행위, 자동차 질도행위 등은 보험사기에는 해당되지 않지만 보험범죄에는 해당된다는 것이 일반적인 시각이라 할 수 있다.

1) 보험사기

현재 가장 많이 인용되는 보험사기의 정의는 “보험가입자 또는 제3자가 받을 수 없는 보험금부를 대가없이 받거나, 부당하게 낮은 보험료를 지불하거나 또는 부당하게 높은 보험금부의 지급을 요구할 목적을 가지고 고의적으로 악의적으로 행동하는 것”으로 보험범죄와 동일한 의미로 사용하였다 (조소웅, 1993). 그러나 쓰키타리 카즈키요(1997)의 영향으로 김형기·손성동·김상의(1998)에 의해 개념구분이 시도되었고, 이후 박일용·안철경(1995) 등에 의해 보다 세분화되고 구체화된 개념이 도입되었다. 그러나 대부분 연구자들은 보험사기가 보험범죄보다 상위의 개념으로 이해하고 있을 뿐만 아니라 선진국에서 보험사기로 보고 있는 허위 고지에 따른 부당한 보험료 면제 또는 할인, 보험사의 보험사기, 보험료 등의 횡령사고 등을 배제하고 있는 실정이다 (www.insurancefraud.org/facts.html).

보험사기가 되기 위해서는 우선 형법상의 사기죄 즉, 타인을 기만하여 재물의 교부를 받거나 재산상의 이익을 취득하는 경우 및 제3자로 하여금 재물의 교부를 받게 하거나 재산상의 이익을 취득하게 하는 죄(형법347조)가 성립되어야 한다. 이러한 보험사기는 ① 보험회사 ② 재산적 이득 ③ 기만행위 ④ 보험업무 ⑤ 보험관계인, 제3자 ⑥ 고의 등 6가지의 요소로 구성되어 있다. 그러므로 “보험사기는 부당한 방법으로 보험계약을 체결하거나, 보험업무와 관련하여 자기 또는 제3자에게 보험금의 형식으로 위법적인 이익을 보게 하는 행위”라고 정의할 수 있다.

2) 보험범죄

문헌상 보험범죄를 보험사기와 동일한 개념이 아니라 독립된 개념으로 파악한 조소웅(1993)은 “보험범죄를 보험계약자, 피보험자 또는 제3자가 보험금을 수령할 목적으로 인위적으로 보험사고를 야기 시키거나 보험자를 기만하여 부당하게 높은 보험금의 지급을 요구할 목적으로 고의적이며 악의적으로 행동하는 행위”로 규정하였다. 그 후 김형기·손성동·김상의(1998)은 “보험범죄는

보험금을 부당하게 수취하는 모든 행위로써 구체적인 범법행위로 나타나는 결과”로, 박일용·안철경(1999)은 “보험범죄는 보험계약자, 피보험자 또는 수익자가 보험제도의 원리상으로는 취할 수 없는 보험혜택을 부당하게 얻거나 보험제도를 역이용하여 고액의 보험금을 수취할 목적으로 고의적으로 악의적으로 행동하는 자의 인위적인 행위”로 규정하면서 보험범죄가 구체적인 범법행위로 나타난 결과만을 지칭하는 반면 보험사기는 보험가입시의 악의성을 포함하는 보다 광범위한 개념이라고 주장하였다.

한편, 위에서 정의한 보험범죄는 모두 보험금 수취행위를 전제로 하나, 산재보험과 같이 보험가입시 실제 임금과 다른 낮은 자료를 제출함으로써 적정보험료보다 부당하게 적게 납부하는 계약을 체결한 행위는(보험)사기죄가 됨에도 보험범죄에는 포함되지 않는 결과를 초래하게 된다. 또한, 의사 등에 의한 허위진료 기록서 작성, 공갈 또는 폭력행위 등 부정한 방법으로 보험금을 편취하는 행위 등은 보험범죄가 명백하나 이러한 기준으로는 포함되지 않는 결과를 초래한다. 따라서, 보험료 편취행위등을 포함하는 보다 포괄적이면서 단순한 정의가 필요하다고 생각된다.

보험범죄는 ① 보험회사 ② 재산적 이득 ③ 범법행위 ④ 보험업무 ⑤ 보험관계인, 제3자 ⑥ 고의 등 6가지의 요소로 구성되어 있으며, 특정 범법행위(사기)가 아니라 형법적 처벌대상으로서의 제반 위법행위를 대상으로 하고 있다는 점에서 보험사기와 다르다 이러한 점을 고려할 때, “보험범죄는 보험업무와 관련하여 자기 또는 제3자의 위법적인 이득을 위해 직·간접적으로 보험사에 행하는 제반 범법행위”로 규정할 수 있다.

3.1.3 보험사기 현황

최근 국내외적으로 보험시장의 보험범죄로 인한 손실액은 연간 1천억달러로 추정되고 있는데, 이는 수입보험료의 약 15%, 전체 지급보험금의 약 8%정도를 차지하는 상당한 금액이다. 보험범죄가 급격하게 증가하고 있으며, 특히 90년대 후반 보험사기 행위가 TV 등과 같은 언론매체에 알려지면서 그 심각성이 사회적으로 대두되기 시작하였으며, 1998년 12월, 외환위기로 인한 IMF가 시작되면서 그 규모는 급증한 것으로 나타났다.

현재 우리나라는 물론 전 세계적으로 각종 사기와 관련된 문제가 날로 심각성을 더해가고 있으며, 그 피해는 상상할 수 없을 정도로 매우 방대해져 가고 있다. 이미 생활에서 필수 불가결한 요소가 되어버린 보험업만 하더라도 한해 총 보험금 지급액이 60조를 넘고 있으며, 이중 보험사기로 예측되는 지급액은 총 지급액의 3%~5%로 추정되고 있다. 특히 미국이나 호주와 같은 선진국의 경우, 보험사기 규모는 전체 보험금 지급액의 10% 이상으로 추정되고 있으나, 실제 사기로 적발하는 경우는 0.5%에도 못 미치는 것으로 조사되고 있다.

이러한 보험청구 사기로 인한 누수액은 고스란히 보험금 산정에 반영되고 있으며, 보험회사는 지

불하지 않아도 되는 보험금을 지급함으로써 기업의 재정적인 부담으로 작용하고 있다. 또한 사기 수법이 매우 과학화, 지능화됨에 따라 사기의 경우를 파악하는 것도 매우 힘들뿐만 아니라, 사기로 의심이 가는 청구 건에 대해서도 자체 조사 인력의 부족, Database의 비효율성, 근거자료 확보의 어려움 등으로 실제 적발은 매우 어려운 것이 현실이다. 특히 전문 인력이 많이 필요한 조사자의 경우, 인력 확보도 어려울 뿐 아니라 정상적인 조사 능력을 갖추기까지 많은 비용 및 기업 자원이 투자되어야 하는 점은 사기 적발을 어렵게 만드는 가장 중요한 장애물이라 할 수 있다.

국내 보험사기의 적발규모는 1999년을 기준으로 총 433억원, 적발 건수는 3,876건으로 조사된 바 있으나, 실제 보험사기 규모에 대한 정확한 통계는 없으나 세 외국의 사례를 준용할 때 보험사기 규모는 공제를 포함한 연간 약 1조원에 달할 것으로 추정되고 있다.

[표 3-1] 연도별 보험사기 적발현황

구 분	1997	1998		1999		2000	
	건 수	건 수	증가율	건 수	증가율	건 수	증가율
적발건수	1,951	2,684	37.06	3,876	44.4	4,726	21.9
관련금액	25,336	29,587	16.8	44,273	49.6	31,421	-29.0

(단위 : 건, 백만원, %)

국내 손해보험에서 보험사기와 가장 밀접한 관계를 맺고 있는 것은 자동차보험이다. 아래의 [표 3-2]에서 나타나듯이 자동차보험은 전체 민영 손해보험 보험사기의 70%를 차지하고 있으나, 생명보험과 사회보험(의료보험, 산재보험, 고용보험, 국민연금)을 고려하는 경우 보험사기의 구성은 상당히 달라질 수 있다.

[표 3-2] 유형별 국내 보험사기 현황

		위장가공사고	교통사고악용	기타보험사기	합계
FY' 97	건 수	1,070	4,025	14	5,109
	금 액	8,719,719	21,809	1,997	32,525
	구성비	26.8	67.1	6.1	100
FY' 98	건 수	1,100	4,438	61	5,599
	금 액	8,462	20,116	148	28,726
	구성비	29.5	70	0.5	100

(건, 백만원, 백분율)

3.1.4 보험사기의 종류

일반적으로 보험사기를 분류하는 방법은 보험금 수취를 위해 조직적으로 보험사고를 위장, 조작하는 경성 보험사기(hard insurance fraud)와 사고발생 후 보험금을 과다청구하는 연성 보험사기(soft insurance fraud)로 나누는 방법이 주로 사용이 된다. 또한 최근에는 이러한 기존의 경성과 연성의 중간에 해당하는 보험사기가 지속적으로 증가하는 점을 감안, 다음과 같이 세 가지 종류로 분류하는 경향이 많다.

[표 3-3] 보험사기의 분류

사 기 내 용	사기수준	발생빈도수	사고당보험금 누수액
간헐적인 과다청구(연성 보험사기)	비전문가	많음	낮음
반복적인 과다청구(hard & soft insurance fraud)	반전문가	중간	중간
조직적이고 전문적인 행위(경성 보험사기)	전문가	적음	높음

자료: Conning & Company (1996)

3.1.5 보험사기의 특징

1) 저 위험 고수익(low risk - high return) 범죄

보험사기 및 범죄에 대한 관용적 사회 분위기에 편승하여 발각되면 경미하게 처벌되거나 관용하는데, 이러한 이유는 일반인이면 누구나 한번쯤 연상사기의 유혹에 빠지게 되기 때문에 강력한 처벌에 대해 부정적이고, 보험사도 영업중시정책 때문에 조용히 처리되기 바라기 때문이다. 아래 표는 보험사기자에 대한 처리현황인데, 검찰 등 수사기관에 의해 불구속 처리되거나 무혐의 처리된 사람은 227명으로 전체의 37.6%를 차지하고 있다. 이는 수사기관의 수사능력과 의지가 미흡하기 때문인 것으로 보이는데, 이러한 법 집행기관의 관대함은 보험사기 및 범죄를 촉발시키는 요인이 되고 있다.

[표 3-4] 보험사기자 처리현황

구 분	'97	'98		'99		'00	
			증가율		증가율		증가율
구 속	39(48.1)	27(44.3)	507.7	332(48.2)	40.1	277(45.9)	△16.6
불구속입건	8(9.9)	73(13.6)	812.5	83(12.1)	13.7	169(28.0)	103.6
수 배	8(9.9)	140(26.2)	1,650.0	86(12.5)	△38.6	72(11.9)	△16.3
수 사 중	3(3.7)	11(2.0)	266.7	50(7.3)	354.5	27(4.5)	△46.0
무 혐 의 등	23(28.4)	41(7.7)	78.3	136(19.8)	231.7	58(9.6)	△57.4
과악불가(사망등)	0(0.0)	33(6.2)	-	1(0.1)	△97.0	1(0.1)	-
합 계	81(100.0)	535(100.0)	560.5	688(100.0)	28.6	604(100.0)	△12.2

(단위: 명, %, () 안은 구성비)

한편, 유사보험 및 손보사의 적극적인 시장참여로 상해 및 질병보험 분야는 시장점유율 경쟁이 심화되고 있는 가운데, 저렴한 보험료와 고액 보장의 소위 복권형 보험개발과 판매채널 다양화라는 미명 하에 진행되고 있는 자판기식 보험판매 -예를 들어 인터넷 회원 가입 시 보험가입 혜택부여- 등은 보험사기 및 범죄를 저렴한 비용으로 손쉽게 할 수 있는 환경을 만들고 있다.

2) 혐의 입증의 난해성

보험사기가 성립되기 위해서는 고의에 의해 계산적 이득을 얻었음을 입증하여야 하는데, 예를 들어 자동차보험의 경우 중대한 과실과 고의를 구분하기가 수비지 않으며, 수사권이 없는 보험사가 고의를 입증하기는 현실적으로 매우 힘들다. 특히, 기업 또는 각 유관기관과의 정보수집 및 정보공유의 예로는 범죄혐의의 입증에 더욱 힘들게 하는 주요한 원인이 된다. 아울러 경·요추 염좌 등에 대한 진단의 주관성, 보험사기 및 범죄 전문수사 인력의 부재, 목격자 신고율 저하, 신고자 보상 및 보호체제의 전무 등도 혐의 입증에 어렵게 하는 요인들이다.

3) 지능적 범죄와 생계형 범죄의 공존

보험사기 중 경성사기는 대부분 지능적 범죄로서 악의적이고, 교묘한 수법이 특징이다. 특히, 공동범행을 통해 알리바이나 사고상황을 조작하는 등 조직적이고 치밀한 행동을 하며 경제상황과 무관하게 발생한다. 반면, 기회사기인 연성사기는 우발적이기 때문에 대부분 단독범행을 하고, 장기 불황이거나 IMF체제와 같은 경제 공황 하에서 많이 발생하게 된다.

4) 보상성 심리와 동조죄의 존재

보험사기 및 범죄에는 위험보장이라는 무형의 서비스에 대한 불만족과 소멸성 보험료에 대한 보상심리가 내재되어 있는데, 이러한 보상성 심리 때문에 나타나는 사기가 대개 연성사기이다. 이러한 연성사기를 경성사기 보다 건별 피해규모는 작지만 건수가 많아 사기피해의 대부분을 차지하고, 적발도 어렵고 적발비용 등이 과다하므로 방치되는 경향이 있다. 이러한 보상성 심리는 죄의식을 약화시켜 잠재적인 예비사기꾼을 양성시키는 토양이 된다.

5) 사기피해의 간접성, 광범위성

보험사기 및 범죄는 외견상 보험사에 직접적인 피해를 주는 것 같지만 보험사는 보험료 인상을 통해 피해 주체의 변경 즉, 현재의 보험계약자가 아니라 미래(보험료 인상시점)의 보험계약자로 피해당사자를 변경시킨다. 그러나 영업경쟁이 격화되고, 가격경쟁력이 시장 선점의 중요한 변수로 작용하는 보험환경 - 보험료 인상의 경직성이 존재하는 경우 -이도래하면, 보험사는 보험사기 및

범죄로 인한 직접적인 피해를 입게 된다. 아울러 현재 및 미래의 보험계약자도 순차적으로 보험료 인상이나 공적자금 부담 등의 간접적인 패해(hidden tax)를 입게 된다.

6) 범죄의 심각성

보험사기 및 범죄와 관련된 범죄는 반인륜적이고 위법적인 행위를 수반하는데, 다른 범죄의 결과로써 보험사기 및 범죄가 이용되기도 하지만 보험사기 및 범죄를 위해 다른 범죄를 저지르기도 하는 등 복합적인 성격을 띤다. 더욱이 조직적인 범죄가 성행하면서 범죄수법이 다양화되고, 최근에는 범행의 용이성 등을 이유로 차량을 이용한 범죄가 급증하고 있다.

7) 내부 종사자의 공모

보험사기 및 범죄의 지능화에 따라 내부종사의 목인, 방조, 공모행위가 많아지고 있다. 특히, 보상 내용을 이해하기 어려운 보험 상품의 특성상 상품정보와 보험사의 생리를 아는 모집종사자들의 보험사기 및 범죄 개입은 어찌면 당연한 결과인지도 모르겠다. 일부 지역의 보험사기 및 범죄건의 경우 모집인의 공모가 위험수준에 와 있음을 단적으로 보여주고 있는데, 실적경쟁과 영업내실화를 통한 사업비 절감 등은 모집인의 경제적 기반을 악화시키고, 보험모집인에 대한 형식적인 사내교육 및 구상제도의 미비 등은 모집관련 종사자들의 도덕적 해이를 부추기고 있는 실정이다.

3.2 보험사기 방지 및 적발 시스템에 대한 선행연구

미국의 보험사기 방지를 위한 많은 노력이 90년대부터는 정부차원의 법제정 및 적극적인 개입으로 인해 크게 부각되었다. 당국은 Database와 같은 정보기술을 서서히 활용하기 시작했으며 SIU(Special Investigation Unit) 과 같은 프로그램을 강력하게 보험사의 경영층에 요구하는 추세이다.

보험사기 예측모형에 최초의 연구는 Lemaire (1990)의 연구에서 시작되었다. 본 연구에서 그는 이러한 보험사기 조기경보 시스템이 보험이나 금융기관에 충분히 활용될 수 있다는 가능성을 보였으며, Derrig and Ostaszewski (1995)의 연구에서는 기존의 통계적 모형과 인공지능 기법을 비교함으로써 인공지능 기법이 보험사기를 예측하는데 보다 우월한 성능이 있다는 것을 증명한 바 있다. 그 후, Belhadji and Dionne (1997)의 연구에서는 보험회사내의 손해사정인을 대상으로 설문조사를 통해서 사고 징후변수(Red Flag)의 상대적 중요성을 추정하였다.

이에 반해서 국내의 보험범죄에 관한 연구의 시초로는 조해균(1995)의 연구를 들 수 있다. 그는 보험범죄의 발생원인, 효율적 관리방안 및 방지대책에 관한 포괄적 내용을 국내 최초로 다루었으

며, 이러한 조해균의 연구를 바탕으로 자동차보험을 중심으로 보다 구체적이고 실무적 사례연구가 나오고 김광용 (1996)은 각종 통계적인 모형과 인공지능 기법을 사용하여 실무보장전문가의 지식을 이용한 조기경보 모형 개발에 대한 기초연구를 수행하였다. 그 후, 안철경, 박일용 (1999)은 국내 보험사기 성향 및 규모추정 연구를 수행함으로써 국내 보험사기방지에 대한 실무적 차원의 법제적 대응방안과 보험사기에 대한 정책적 인식의 확대에 기여하였다. 이와 같이 보험사기에 대한 시스템적 연구는 거의 없으며, 특히 Visual Data Mining을 활용한 보험사기 적발 연구는 더욱 미비하다고 할 수 있다.

3.2.1 스코어링 시스템을 이용한 사기적발 시스템

스코어링 시스템은 조기경보 시스템으로도 불리는데, 각 개인 혹은 사건 별로 Score 값을 계산해서 점수를 부여함으로써, 특정 점수 이상(혹은 이하)의 사람 및 사건들을 사기 가능성이 높은 위험인들로 분류, 전문 조사관들에 의해서 특별히 관리할 수 있도록 해주는 시스템을 의미한다. 그러나 이러한 Score값을 이용한 사기적발 시스템은 각 개인이나 사건에 대한 평가만을 할 수 있지, 개인 간의 관련성을 파악할 수 없다는 커다란 약점이 존재하고 있다. 특히 최근 들어서 조직화, 대형화되어 가는 각종 사기범죄를 적발하기 위해서는 스코어링 시스템은 일반적인 데이터마이닝 과정에 따라 진행되나, 데이터마이닝의 특성상 보험사기에 대한 선행 연구 및 관련된 지식, 데이터의 축적이 미미한 상황에서는 정확도(Accuracy)에 대한 신뢰성이 부족할 수 있다.

3.2.2 Link Analysis를 이용한 사기적발 시스템

Link Analysis를 이용한 사기적발 시스템은 Link Analysis 하고 하는 Visualization Data Mining 기법을 기반으로, 사기적발에 있어 조사자로 하여금 직접적(Direct), 간접적(In-Direct)인 관계를 시각적으로 표현하여 사기성 유무를 파악하는 방법이다. 특히, 보험사기와 같은 금융사기의 경우 복잡한 DB로 인해서 데이터 간의 관련성을 파악하기 힘들지만, Link Analysis 기법을 사용함으로써 직접적인 관련성 외에도 간접적인 관련성을 파악하여 조사자로 하여금 복잡한 사기 관련성 및 혐의점을 포착할 수 있도록 도와주는 의사결정지원 시스템의 형태라 할 수 있다.

IV. 데이터마이닝 기반의 보험사기 적발시스템의 활용 및 기대효과

4.1 보험사기 유형별 보험사기 적발시스템 적용

보험사기에 대한 정부와 보험사의 노력은 기구를 설립, 운영하고 있으며 조사원들로 하여금 의심나는 청구건을 수사하고 있는 실정이다. 일부분에서는 정보기술의 일부인 데이터베이스로 데이터를 정리하여 그 정보를 이용하기도 한다.

그러나 앞서 살펴본 바와 같이 보험사기의 행태는 다양하게 구분됨에 따라 주먹구구식과 획기적인 사고의 전환이 없이는 보험사기의 증가에 따른 적절한 대응을 하기에 역부족일 수 밖에 없다. 따라서 보험사기를 적발하기 위한 방법은 보험사기 유형 및 특성별로 연구가 되어야 할 것으로 보인다. 하나의 접근 방법만으로는 다양한 행태의 보험사기를 적발하는 데는 그 한계점이 있을 수 있기 때문이며 새로운 정보기술을 적용한 방법으로 보험범죄에 대처해야 할 것이다.

연성 보험사기는 사고가 나기 전까지만 하더라도 전혀 속일 생각이 없었는데, 사고가 난 직후, 피해를 줄이거나 경미한 보험금을 추가로 더 타기 위해서 벌어지는 사기의 유형으로써, 이는 데이터마이닝과 시스템을 이용하더라도 적발을 하기에는 한계가 있을 것으로 보인다. 이러한 연성사기, 특히 운전자 바뀌치기의 경우 보험회사의 보상직원이 가지고 있는 경험에 의한 지식을 통해 상당 부분은 파악이 가능한 것으로 알려지고 있다.

하지만, 경성사기의 경우 연성사기와는 달리 악의를 가진 사람(주로 범죄 집단)에 의해서 계획적으로 이루어지는 보험사기를 의미하는데, 공모형 보험사기범죄 라고도 한다. 즉, 의도적으로 차를 가지고 논두렁에 빠진 후, 보험금을 청구하거나, 여러 명이 서로 공모를 한 후 서로 치고 받고 여러 번 사고를 낸 후 보험회사로부터 수차례 보험금을 타내는 수법, 그리고 가해자가 피해자인 것처럼 위장함으로써 생명보험금을 타내는 것과 같은, 사고가나기 전에 미리 치밀한 계획을 한 후 벌이는 사기의 유형을 의미한다.

이러한 경성사기는 앞의 연성사기에 비해서 데이터마이닝 기법을 이용해서 잡을 수 있는 확률이 높다. 이러한 경성사기를 다시 살펴보면 다음과 같이, 1인 단독으로 범행을 저지르는 경우와 2인 이상이 서로 짜고 사기를 저지르는 것으로 분류를 할 수 있다. 이렇게 경성사기를 두 가지로 분류를 하는 이유는 사기 유형에 따라 적용하는 데이터마이닝 기법이 달라지기 때문이다.

[표 4-1] 보험사기 유형과 데이터마이닝 방법

보험사기 유형	사기 적발 접근방법	데이터마이닝 방법
단독 사기	사기를 저지는 1사람의 위험도를 측정	스코어링 시스템
공모 사기	여러 사람간의 공모 관련성을 탐색	Link Analysis

4.1.1 단독사기 적발을 위한 스코어링 시스템

1인 단독사기를 적발하기 위해서는 어떠한 경우에 사기의 위험성이 높은지를 알 수 있는 보험 사기징후(Fraud Indicator: FI)를 파악하는 것이 중요하다. 이러한 사기징후로는 여러 가지가 있지만 그 중, 선진국에서 사용되는 대표적인 징후들을 살펴보면 다음과 같다.

[표 4-2] 대표적인 보험사기 징후

대표적인 보험사기 징후 (Fraud Indicator: FI)
보험계약이 월 마감일에 근접하여 체결 계약자의 월 납입보험료의 수준이 가입자들의 월 평균 납입 보험료를 상회 보험계약이 손해가 발생하기 전에 단기간으로 가입 사고목적자가 전혀 없음 고물차에 고보장형 상품을 가입 경미한 교통사고에 비해 과도한 치료비 청구

상기와 같이 사기를 사전에 감지할 수 있는 사기징후들을 파악한 후, 이러한 사기 징후들을 실제로 데이터마이닝을 하기 위한 데이터 변환을 하기 위해서는 아래와 같이, 각 사기 징후를 만들기 위해서 필요한 데이터를 확인해야만 한다. 이 경우, 어떤 사기징후는 데이터를 이용해서 만들 수 있을 것이고, 또 어떤 사기징후는 데이터가 없기 때문에 만들 수가 없을 것이기 때문이다.

[표 4-3] 보험사기 징후의 데이터 변환

보험사기 징후	사용 데이터
보험 계약이 월 마감일에 근접하여 체결	보험 계약일, 업무 마감일
계약자의 월 납입보험료의 수준이 가입자들의 월 평균 납입 보험료를 상회	계약자 월 납입보험료, 가입자 평균 납입 보험료
보험 계약이 손해가 발생하기 전에 단기간으로 가입	클래임 청구 전 각 월별 보험가입 여부
사고 목격자가 전혀 없다	사고 목격자 존재 여부, 목격자수, 목격자와 사고자와의 관계
고물차에 고보장형 상품을 가입	보장금액/차량가액 차량연식별 동급차량 보장금액
경미한 교통사고에 비해 과도한 진료비 청구	진료비/수리금액, 사고규모별 평균 진료비
서로 연관 없던 탑승객이 같은 변호사를 선임	차량 탑승객별 관련 제3자

이와 같은 방식으로 스코어링 시스템을 위한 입력변수를 생성한 후, 회귀분석, 의사결정나무 분석, 인공신경망 기법과 같은 데이터 마이닝 기법을 이용해서 사기적발 모형을 만들 수 있게 되는 것이다.

4.1.2 공모사기 적발을 위한 Link Analysis 기법

2인 이상의 공모사기를 적발하기 위해서는 보험사기자들끼리는 어떠한 데이터를 공유하고 있는지를 파악하는 것이 중요하다. 즉, 사고가 똑같은 사람끼리 여러 번 있는데 사고와 관련된 모든 사람이 같은 연령대이고, 같은 동네에서 태어나고 자랐으며, 같은 학교 출신이라면 보험사기 혐의가 포착이 될 수 있는 것이기 때문이다. 일반적으로 스코어링 시스템에서 사용하는 숫자와 관련된 데이터 형태보다는 문자형의 원천 데이터를 활용하게 되는데 아래와 같은 데이터들을 활용하게 된다.

[표4-4] Link Analysis활용 데이터

분류 항목	사용 데이터	
사고당사자	보험금청구 번호 전화번호 성별 등	관련자 이름 주민등록 번호
	계약번호 개시일 만기일 등	
계 약		보험 유형 효력일

실제 보험사기를 적발하는 보험회사의 내부 조직을 SIU(Special Investigation Unit)라고 하며, 이들이 실제로 한 건의 보험사기를 적발하기 위해서는 수많은 문서를 뒤져야 하고 데이터를 분석해야 하지만, 이러한 Link Analysis 기법을 통해서 보다 쉽고 빠르게 보험사기를 적발할 수 있다.

4.2 데이터마이닝 기반의 사기적발 시스템의 기대 효과

고객에게 신속한 보험금 지급 및 보험금 지급 이전에 사기성 유/무의 정도를 판단하여 적극적인 보험사기 예방 및 철저한 관리로 사기적발 누수액이 절감되는 효과가 있으며 이는 대다수 선의의 고객 보험료를 절감시켜 주는 궁극적으로 보험료가 저렴하게 되어 가격 경쟁력을 가지게 되는 효과가 크다. 따라서 보험금 누수예방 뿐만 아니라 사기적발로 보험금 지급을 최적화 할 수 있다. 이는 나아가 기업의 손익을 극대화하는 데 긍정적인 효과가 예상된다.

보험사기에 대한 수사를 더욱 효과적으로 수행할 수 있게 다양한 알고리즘을 사용해 그 결과를 시각적으로 보여주는 기능으로 인해서 대규모 공모를 한 보험사기 집단 식별에 매우 용이해서 적극적으로 활용될 소지가 높다. 아울러 보험사기 적발에 관련한 실무자들의 업무 효율성 향상으로 더 많은 사기성 클레임을 더 짧은 시간에 처리할 수 있을 것이다. 결과적으로 보험범죄에 시스템적인 대응으로 인한 적극적인 보험사기 예방 및 관리로 고객 신뢰성을 상승하는 효과가 기대된다.

IV. 결론

보험사기를 예방하고 적발하는 업무가 얼마나 복잡하고 다양한가는 이미 익히 알고 있을 것이다. 이러한 보험사기에 대한 종합적인 대책은 보험사의 노력만으로는 부족하며 보험범죄에 대한 당국의 인식과 이에 따른 시스템 구축 및 보험범죄 관련 법안과 같은 적극적인 노력과 당국의 지속적인 관심이 필요하다.

보험사기는 결국 보험사의 손해율과 보험 효율에 지대한 영향을 미치기 때문이다. 즉, 보험사의 보험금 지급을 최적화 시켜 비정상적인 손실을 방지하고 또는 절감하고 이러한 원가 절감을 통해 보험료 경쟁 우위를 확보해 결과적으로 고객의 회사에 대한 충성도를 높이는 중요한 역할을 한다고 믿고 있다. 허나, 여러 가지 이유에서 시스템적인 접근이 어려웠던 것만큼은 인정해야 하는데 그 중에 가장 큰 이유가 고객에 대한 정보 접근일 것이다. 이러한 문제는 각 기업의 고객 데이터베이스화 뿐 아니라 고객에 대한 여러 측면에서의 데이터 분석이 이루어지는 현실에서는 큰 문제가 될 수 없

다. 그러나 정작 데이터가 준비가 되어 있어도 수많은 의심 클레임과 관련된 데이터를 짧은 시간에 분석해서 유용한 결과를 가지고 보험금 지급한다는 것은 보통 어려운 일이지 아닐 수 없다.

최근에 접근하고 있는 방법은 본 논문에서 지적했듯이 데이터마이닝 기술을 활용해서 시스템적인 접근을 시도하고 있다. 즉 대량의 데이터 속에서 관련성 있는 데이터를 연관분석(Link Analysis)을 통하여 연결시키고 이를 Visual data mining 기술을 적용시켜 보험사기 적발업무를 매우 효율적이며 효과적으로 수행하는 데 큰 기여를 한다. 국내뿐만이 아니라 해외에서도 많은 시도가 이루어지고 있는 데이터마이닝 기반의 보험사기적발 시스템에 대한 연구가 지속적으로 이루어지길 기대해 본다.

[참고문헌]

- 김광용, 보험사기적발을 위한 전문가시스템이 개발 : 퍼지이론과 AHP를 중심으로,
[보험개발연구], 제8호(1996): 4-28
- 김기훈, “네트워크 분석기법을 활용한 데이터마이닝”, 한국데이터마이닝학회 춘계 컨퍼런스,
2001.
- 김형기, 손성동, 김상의, 보험범죄의 현황 및 대응방안, 삼성금융연구소, 1998.
- 박일용, 안철경, 보험사기 성향 및 규모추정, 보험개발원 보험연구소, 1999.7
- 송기철, 보험과 범죄, 보험학회지 22집 및 생활속의 보험, 보험사기와 보험범죄, 1993.
- 안철경, 박일용, 보험사기 적발 및 방지방안, 보험개발원 보험연구소, 1999.11.
- 이병희, 보험범죄론, 형설출판사, 2001.
- 조수용, 보험범죄와 그 대책, [손해보험], 1993.3.
- 조해균, 자동차보험에 있어서 도덕적 위험 방지대책, [손해보험], 1995.11.
- Belhadji, E. B, Doinne, Development of an Expert System for the Automatic Detection of
Automobile Insurance Fraud, Working paper, 1997
- Ciara Byrne., “Intelligent Fraud Detection”. Communication. Vol. 5, No. 2, 2002
1998 AAAL Fall Symposium on Artificial Intelligent and Link Analysis
- Christopher Westerphal, Teresa Blaxton. Data Mining Solutions : Methods and Tools for
Solution Real-World Problem”. John Wiley & Sons. 1998
- Conning, Co. Insurance Fraud: The Quiet Catastrophe, Conning & Co., 1996.
- Daniel A. Keim., “Information Visualization and Visual Data Mining”. IEEE. Vol. 8, No.
1, 2002
- Derrig, R. A. and Ostaszewski, K. M., Fuzzy techniques of Pattern Recognition in Risk and
Claim Classification, Journal of Risk and Insurance 62. 1995.
- Lemaire, M., Fuzzy Insurance, ASTIN BULLETIN, 20(1990): 33-35
- Roberto Tamassia. Ulrik Brandes. David Krackhardt. Dorothea Wagner., Link Analysis
and Visualization, 2001
- [Http://insu.co.kr/vohumkonwledge/world.enjoy/world_living003.htm](http://insu.co.kr/vohumkonwledge/world.enjoy/world_living003.htm).
- [Http://www.insurancefraud.org/facts.html](http://www.insurancefraud.org/facts.html)
- The Coalition Against Insurance Fraud, Insurance Fraud : The Hidden Tax,
[Http://www.insurancefraud.org/facts.html](http://www.insurancefraud.org/facts.html)