

비통계적 링크분석을 활용한 보험사기의 효과적 적발방법 연구*

A Study of an Effective Insurance Fraud Detection Method Using Nonstatistical Link Analysis

김 현 수**
Kim Hun-Soo

본 연구의 목적은 첫째로 보험사기의 최근 현황과 보험사기 적발과 관련된 문헌을 정리하고, 둘째로 사회 네트워크 이론의 링크분석(Link Analysis)의 개념과 기본적인 방법론을 소개하고, 마지막으로 링크분석의 Emergent Group Algorithm와 Step Link기법으로 보험회사 보험금 청구 데이터를 분석하여 링크분석의 효과성을 검증해 보려는 것이다. 보험사기 적발 노력이 강화될수록 소비자 피해가 증가할 수 있다는 부작용을 극복하기 위해서는, 보험사기 적발 기법은 효과적일 뿐만 아니라 더욱 과학적이고 객관적이어야 한다. 그러나 과거 보험사기 적발 기법 연구는 Derrig(2002)에서 나타난 바와 같이 대부분은 속성(attribute) 데이터를 중심으로 보다 정교한 지수모형(scoring model) 개발에 초점을 두고 있으며 링크분석과 같이 보험사기자를 적발하는데 직접적으로 도움을 줄 수 있는 적발모형(detection model)에 관한 연구는 적었다. 링크분석은 속성중심 데이터가 아닌 관계(relation) 데이터를 중심으로 보험사기를 조사하는 적발모형의 한 방법인데 지수모형보다 효과적이고 직접적인 도움을 SIU에 제공해 줄 수가 있다는 장점이 있다. 본 연구에서 실제 보험사 자료를 분석해 본 결과 링크분석의 Emergent Group Algorithm과 Step Link Analysis는 경성 보험사기 적발에 상당한 가능성이 있다는 것을 보여주었다.

※ Key Word : 보험사기, 보험사기 적발방법, 데이터중심적 접근방법, 링크분석

* 본 연구에 도움을 준 익명의 심사위원과 노틸러스 효성 정수용 차장, 그리고 자료를 제공해 준 K보험회사에 감사함을 전한다. 본 연구는 순천향대학교 2001 교내연구비 지원을 받았다.

** 순천향대학교 금융보험학과 교수(e-mail : briank@sch.ac.kr)

I. 연구의 배경과 목적

1. 보험사기 현황

보험회사의 수익창출 과정은 크게 세 단계로 나눌 수 있다. 먼저 보험계약 단계에서 보험료는 충분히(adequate) 부과하면서도 우량위험을 선택하는 '입구전략'이다. 둘째로 수입보험료 등을 운영해서 수익을 창출하는 자산운용전략이다. 마지막으로 보험사기 등 보험금누수의 차단을 통해서 수익을 증대하려는 '출구전략'이 있다. 그런데 자유화 또는 규제완화가 진전되면서 입구전략과 자산운용전략을 통해서 이익을 확보하는 것이 점점 어려워지고 있다. 즉 가격자유화에 따른 가격경쟁에 따라서 보험료인하 압력이 보험사를 압박하고 있어 '입구전략'에서 잠재수익을 확보하는 것이 어려워지고 있다. 나아가 이자율 등 자본시장의 자유화와 금융자산의 시가평가 등 규제완화가 가속화되면서 자산운용 리스크가 점점 커질 뿐만 아니라 재무건전성에 대한 기준도 강화되어 안정적으로 자산운영수익을 확보하는 것이 점점 어려워지고 있다.

따라서 보험사들은 과거 소극적으로 대처했던 '출구전략', 즉 보험금누수 차단에 상대적으로 많은 노력을 기울이고 있고 보험감독 당국도 보험시스템의 안정성차원에서 보험사기에 적극적으로 대응하고 있다. 특히 1997년 IMF 외환위기 직후 생계형 보험사기가 급증하면서 보험감독당국과 보험회사들의 보험사기에 대한 관심은 급증하였다. 그후 보험금누수 차단과 보험사기 방지가 보험사의 이익증대에 직결된다는 것을 인식한 보험사들은 보험사기특별조사반(SIU) 등을 설치하는 등, 보험사기에 적극적으로 대처하고 있다. 금융감독원도 보험사기가 보험시스템의 근간을 흔들 수 있다는 인식하에서 보험사기조사실을 설치하고 적발시스템 도입하는 등 보험사기에 다각적으로 대처하고 있다.

우리나라 보험사기의 추이를 살펴보면 <표 1>과 같다. IMF 외환위기 전후인 1997년 이후 보험사기 적발건수는 급증했다가 최근 그 증가세는 둔화되었다. 보험사기 관련금액은 1999년까지 급격히 증가했지만 2000년에는 급락하여 2002년에

는 1999년 수준보다 좀 낮은 수준을 유지하고 있다.

〈표 1〉 연도별 보험사기 적발현황

(단위: 건, 백만원)

구 분	1997	1998		1999		2000		2001		2002	
			증가율		증가율		증가율		증가율		증가율
적발건수	1,951	2,684	37.6%	3,876	44.4%	4,726	21.9%	5,749	21.6%	5,757	0.1%
관련금액	25,336	29,587	16.8%	44,273	49.6%	31,421	-29.0%	40,440	28.7%	41,132	1.7%
건당금액	12.986	11.023	-15.1%	11.422	3.62%	6.649	-41.8%	7.034	5.79%	7.145	1.58%

자료 : 금융감독원(2003)

보험사기건수가 급격히 증가한 것과는 달리 건당 보험사기금액은 꾸준히 감소하고 있다. 건당 보험사기금액은 1997년 천삼백만원 수준에서 2002년에는 그 절반인 칠백만원 수준으로 하락하였다. 이는 청구금액이 큰 건수에 대한 보험사의 감시가 철저해졌고 보험사기자들도 대형 건수보다는 중소형 건수에 집중하는 데 기인한 것이라고 추측할 수 있다.

보험사기 적발에 대한 보험사의 노력이 강화되면서 부작용도 증가하고 있다. 보험사기에 대한 과도한 대응이 소비자나 보험가입자의 권익을 침해하는 사례가 종종 보도되고 있다(소비자보호원(2003)). 문제는 이러한 부작용이 일화적인(anecdotal) 경우가 아니라 상당히 광범위하다는 사실이다. 〈표 2〉에서 보듯이 보험사기 의심자에 대한 사법처리 현황을 보면 1999년 보험사기 의심자 중에서 구속자 비율은 48.3%였지만 2002년에는 25.1%로 크게 감소한 반면 보험사기 의심자 중 불구속자 비율은 1999년 12.1%에서 2002년 40.5%로 급격하게 증가하였다.

〈표 2〉 보험사기 의심자에 대한 사법처리 현황

(단위 : 명, %)

구 분	1999	2000		2001		2002	
			증가율		증가율		증가율
구 속	332 (48.3)	275 (45.5)	△17.2	545 (27.2)	98.2	479 (25.1)	△12.1
불구속	83 (12.1)	145 (24.0)	74.7	734 (36.7)	406.2	772 (40.5)	5.2
수 배	86 (12.5)	68 (11.3)	△20.9	166 (8.3)	144.1	189 (9.9)	13.9
수사중	50 (7.3)	24 (4.0)	△52.0	55 (2.7)	129.2	157 (8.2)	185.5
무혐의	136 (19.8)	88 (14.6)	△35.3	339 (16.9)	285.2	296 (15.5)	△12.7
파악불가	1 (0.1)	4 (0.7)	300.0	162 (8.1)	3,950.0	15 (0.8)	△90.7
합 계	688 (100.0)	604 (100.0)	△12.2	2,001 (100.0)	231.3	1,908 (100.0)	△4.6

자료 : 형사정책연구원(2003), 손해보험협회

소비자보호원에 피해를 호소한 보험가입자들의 숫자도 상당히 많다. 〈표 3〉의 소비자보호원 통계를 보면 보험관련 소비자 민원 접수건수는 연평균 약 천건에 이른다. 2002년 통계를 보면 보험소비자들이 접수한 867건 중 배상(보험금 지급), 계약이행, 환급 등 소비자가 피해를 구제 받은 건이 45.5%(394건)를 차지하였다. 소비자보호법상 처리제외 또는 청구인이 취하한 건 등을 제외할 경우 소비자 피해구제로 처리된 건은 56.4%(699건 중 394건)에 달해 지금도 다수의 보험소비자들이

보험금 청구와 관련하여 피해를 보고 있을 가능성이 높다고 추측할 수 있다.

〈표 3〉 소비자보호원 보험관련 민원 처리결과

(단위 : 건, %)

구 분	합 의 성 립						조정 요청	정보 제공	상담 기타	취하 증지	계
	환 급	계약 이행	계약해제 해지	배 상	부당행위 시정	소 계					
2002	50 (5.8)	65 (7.5)	2 (0.2)	260 (30.0)	17 (2.0)	394 (45.5)	35 (4.0)	143 (16.5)	127 (14.6)	168 (19.4)	867 (100.0)
2001	66 (6.4)	66 (6.4)	7 (0.7)	376 (36.6)	12 (1.2)	527 (51.3)	54 (5.3)	135 (13.1)	159 (15.5)	152 (14.8)	1,027 (100.0)

자료 : 소비자보호원(2003)

보험회사와 감독당국이 보험사기 적발노력을 강화하면 할수록 소비자 피해는 증가할 가능성이 높다. 소비자 피해는 보험회사의 개인정보 침해가능성, 보험금 지급의 지연 및 거절, 나아가 선의의 보험가입자에 대한 인권침해 등 다양한 형태를 나타낼 수 있다. 그렇다면 보험가입자의 권익을 보호하면서 보험사기에 철저히 대처하는 방안은 없을까? 이 고민은 국민의 인권을 철저히 보호하면서도 범죄에 대해서는 철저히 대응해야 하는 경찰이나 검찰의 고민과 유사하다. 여기서 '철저히 대응한다'는 의미는 과학적이고, 객관적인 수사로 소비자피해와 범죄의 빈도와 심도를 동시에 감소시킨다는 리스크관리전략과 맥을 같이 한다.

소비자의 권익을 보호하면서도 보험사기를 철저히 적발·예방하기 위해서는 보험사기에 대한 보험회사와 감독당국의 대응방식도 최대한 과학화, 객관화, 지속화되어야 한다. 이를 위해서 효과적인 사기적발 시스템을 개발하고 이 도구를 조사전문가가 적절하게 활용하도록 하는 접근방식이 필요할 것이다. 이러한 맥락에서 본 연구는 사기적발시스템의 한 축이 될 수 있는 보험사기 적발방법의 실질적인 대안을 제시하려는 시도이다.

2. 연구의 목적과 구성

보험사기에 과학적이며 객관적으로 대응하는 방법에는 크게 두 가지 접근방식이 있다. 이 두 접근방식은 전문가중심적인 접근방식(expert-driven approach)과 데이터중심적 접근방식(data-driven approach)이다. 전문가중심적인 방식은 보험사기 적발에 가장 널리 활용되는 방법이다. 이 방법은 ‘형사 콜롬보’ 식으로 보험사기 혐의가 높은 용의자를 밀착 조사하는 방법으로 현재 보험회사에서 널리 사용되고 있다. 보험회사는 전직 경찰 등 조사전문가를 중심으로 보험사기 특별조사팀(SIU: special investigation unit)을 구성하여, 현업에서 의뢰가 들어온 의심스러운 보험금 청구사건을 밀착·집중 조사한다. 데이터중심적 접근방식은 아직 일반화 되지 않았지만 축적된 데이터를 통해서 일정한 규칙성과 징후를 발견하려는 방법인데, 여기에는 보험사기 가능성을 점수화 또는 인덱스화해서 평가하는 지수모형(scoring model)과 사기용의자를 관계데이터를 통해서 추적·발견하는 적발모형(detection model)이 있다.

본 연구의 목적은 먼저 보험사기 현황과 문헌분석을 통해서 보다 객관적이고 효과적인 보험사기 적발모형의 필요성을 확인하고, 둘째로 데이터중심적 보험사기 대응기법중 적발모형(detection model)의 한 종류인 링크분석의 개요와 특징을 소개하고, 마지막으로 링크분석을 이용한 실제 보험회사의 보험금 청구 데이터를 분석하여 링크분석의 효과성을 검증하는 데 있다.

본 논문의 구성은 다음과 같다. II 장에서는 보험사기 적발과 관련된 기존문헌을 검토하고, III 장에서는 링크분석과 VDM(visual data mining)을 이론적으로 소개한다. 그리고 IV 장에서는 링크분석을 실제 데이터에 적용하여 링크분석의 효과성을 분석하고, V 장에서는 논문의 시사점을 논한다.

〈표 4〉 보험사기 적발기법의 비교

구 분	전문가중심적 접근방법	데이터중심적 접근방법	
		지수모형(scoring model)	적발모형(detection model)
개 요	의뢰가 들어온 건수를 사기 조사팀(SIU)을 중심으로 직접적으로 조사함.	과거 데이터를 기준으로 모형을 개발하여 이를 근거로 개별건수의 보험사기 가능성을 지수화 함.	DB의 데이터로부터 이상징후를 발견하고 이를 근거로 보험사기 용의자를 추출하여 SIU에 통보함.
필요정보	혐의자에 대한 직접적인 조사정보	개체의 속성(attribute) 정보	개체간의 관계(relation) 정보
비용측면	높은 인건비 비용구조	시스템 구축비용 중심적	시스템 구축비용 중심적
조사방법	'콜롬보식' SIU 개인 조사방식	전체 데이터를 대상으로 한 계량적 모형을 구축하여 활용	'사례추적 방식'을 통한 조사
활용방향	용의자에 대한 집중적인 조사를 통해 사기 여부 최종 판단	용의자를 사전 스크리닝(screening)하는데 효과적	용의자를 사후 적발(detection)하여 SIU에 정보 제공

Ⅱ. 보험사기 적발모형에 대한 문헌연구

보험사기가 도덕적 해태(moral hazard)의 결과라는 측면에서 Arrow(1963)는 보험사기 연구의 출발로 볼 수 있으며, 범죄경제학 관점에서 보험사기를 고찰하면 Becker(1974)는 보험사기 대응전략의 이론적 틀을 제시한 연구다. Dionne(1984)은 도덕적 해태 연구와 구별된 보험사기 자체에 대한 초기 연구인데 이 연구는 건강 보험 보험사기 문제를 분석하면서 공급자 사기(provider fraud)에 대한 이론적 틀을 제시하였다.

보험사기 대응에 관한 본격적인 이론연구는 Mookherjee and Png(1989)를 출발로 볼 수 있다. 그들은 보험사기를 대리인문제라고 인식, 게임이론적 방식으로 접

근하였고 그 후 Picard(1996)도 게임이론 방식으로 보험사기를 분석, 보험사기를 완전히 차단하는 방식은 존재하지 않는다는 것을 이론적으로 밝혔다. Boyer(2000)는 보험금에 세금을 부과하는 것이 보험료에 부과하는 것보다 보험사기를 막는데 더 효율적이라고 결론지었다. 최근 Boyer(2001)는 업계 공동으로 설립한 보험사기조사국(Insurance Fraud Bureau)의 평균비용이 업계평균을 초과하지 않는 경우에는 공동으로 보험사기조사국을 운영하는 것이 소비자를 위해서 바람직하다는 이론을 제시하였다. 이 결론은 효율적인 보험사기 조사를 위해서 중앙 집중적 방식이 더 효율적일 것이라는 시사점을 제공한다.

보험사기적발에 관한 실증적 연구는 보험관련기관에서 먼저 시작되었다. 미국 메사추세츠의 보험사기국(Insurance Fraud Bureau of Massachusetts)에서 1985년부터 1989년 사이에 자동차보험의 대인배상 보상 분석자료를 중심으로 다양한 연구결과를 발표하였다(Weisberg and Derrig(1991, 1992)). 캐나다 보험국에서도 유사한 연구가 진행되었다(Dionne and Belhadji(1996); Caron and Dionne(1996)). Weisberg and Derrig(1991, 1992)는 1985년부터 1989년 사이에 자동차보험의 대인배상 손해를 조사하였다. 이 연구는 보험사기적(fraudulent)의 의미는 보험사기나 보상전문가들에게 개인의 경험과 지식, 인식 등에 따라 대단히 주관적이며 따라서 명확하게 보험사기라고 해석할 수 있는 사례에 비해서 의심은 가지만 확실하지 않은 회색지대에 있는 사례가 대부분이라는 것을 지적하였다. 그리고 보상의 과정이 보험금 지급과 거부라는 이분적인 단순한 의사결정이 아니고 매우 복잡한 의사결정 과정을 포함하고 있다는 사실을 지적하였다. 또한 이 연구는 보험사기를 경성 보험사기(hard insurance fraud)와 연성 보험사기(soft insurance fraud)인 과다청구(build-up)로 구분하여 조사하였는데 대인배상(bodily injury)으로 보험금을 지급한 건수중 약 10%는 경성 보험사기와 약 50%는 연성 보험사기와 관련이 있다고 주장하면서 보험사기로 인한 보험금누수 심각성을 경고했다.¹⁾

보험사기 적발모형과 관련된 대표적인 실증연구으로는 Weisberg and Derrig(1995)가 있다. 이 연구는 65개의 보험사기 관련 변수를 먼저 찾고, 이 독립변수들 간의 상관계수를 근거로 상위 25개의 보험사기 연관변수를 선별하였다. 그

리고 이 선별된 25개 변수를 독립변수로 하고, 보험금 청구건수(claim file)의 보험사기혐의 정도를 종속변수로 하여 다변량회귀분석을 시도하였다.

한편 Cummins and Tennyson(1990)은 미국 자동차보험시장에서 보험사기에 부작용과 문제점을 경제학적 분석을 통해서 보여주었다. Derrig et al(1994)은 보험사기자 행태에 보다 초점을 맞추어서 보험사기를 분석하였고, Derrig and Ostaszewski(1995)는 종래의 통계학적 군집모형과 퍼지군집(fuzzy clustering)을 이용하여 보험사기를 예측한 결과 퍼지군집모형이 우월하다는 것을 보였다.

Brockett, Xia and Derrig(1998)에서는 Kohonen의 Self-Organizing Map이라는 'unsupervised learning' 기법을 시도하여 보험사기를 분류하였는데 이 방법이 보상전문가의 주관적인 지식에 의한 보험사기 분류 방식보다 우월하다는 것을 보였다. Belhadji and Dionne(2000)은 probit모형을 이용하여 보험사기 모형을 제시하였다. 이 연구에서는 각 회사의 손해사정인들은 자사의 과거 사고에 대한 사고파일(claim file)을 면밀히 검사하게 한 후 각 건별로 관련되었다고 생각하는 변수를 제시된 50개의 보험사기 징후변수에서 표시하는 방법을 취하였다. 거기에 따라서 각 변수별 조건부확률을 구하고 Probit회귀분석을 통해서 각 사고 건별 보험사기 확률을 도출하였다. 이후는 각 보험회사의 자체적인 판단에 따라서 전체 사고건수 중 보험사기 확률이 높은 상위 10% 또는 5%를 정밀 조사할 수 있다는 것이다.

본 연구와 궤를 같이 하는 선행연구는 Major and Riedinger(1992)이다.²⁾ 이 연구는 보험사기를 암시하는 27가지의 비계량적인 행위정보(behavior-heuristic information)를 통계적으로 정량화하여 보험사기를 보다 효과적으로 적발하려는 시도다. 본 연구는 EFD(Electronic Fraud Detection)란 이름으로 Travelers보험사에서 실제 실행하고 방법을 보여주었는데, 이 방법을 이용해서 건강보험에서 의료공

1) Weisberg and Derrig(1991, 1992) 연구에서 주장한 경성사기율 10%는 SIU직원이 과거 보험금 청구건을 검토하여 사기여부를 추정한 결과로 확인된 경성보험사기율은 아니다. 이 연구외에도 통상 SIU나 보험회사를 대상으로 예정 보험사기율을 설문하는 방식은 보험공급자의 인식만을 반영하여 주관적이라는 한계가 있다. 만일 소비자단체가 보험수요자를 대상으로 조사했다면 경성사기율은 달라질 것이다. 배심원이 결정적인 역할을 하는 사법제도 때문에 미국의 보험사기율을 우리나라보다 월등히 높을 것으로 짐작한다.

2) 이 연구는 *The Journal of Risk and Insurance*(2002)에 채수록 되었다.

급자(healthcare providers)의 보험사기를 효과적으로 적발하였다고 보고하였다.

국내에서 조해균(1989, 1990)을 시작으로 조수웅(1993), 박일용·안철경(1999), 안철경 외(2002) 등이 보험사기의 문제점과 대응방안을 논하였고 김용덕·안철경(2002)은 보험사기조사에 대한 투자의 타당성을 CAIF의 자료로 분석하였다. 보험사기에 대한 이론적 연구로 지홍민(2001)은 최적계약을 통해서 보험사기를 해결이 가능하다는 것을 보였다. 이운호(2000)는 모니터링효과의 국제비교를 시도하였고 이운호(2002)는 보험사기의 대부분을 차지하는 연상사기에 대응하기 위해서는 보험회사에 수사권이 부여되어야 한다고 주장했다. 본 연구와 직접적으로 관련이 되는 보험사기 적발모형에 관한 연구로는 김광용(1997)과 김헌수(1999, 2000)가 있는데, 김광용(1997)은 퍼지와 인공신경망모형을 이용해서 보다 효율적인 보험사기 적발모형을 개발할 수 있다는 것을 시뮬레이션으로 보여주었고, 김헌수(1999)는 AHP를 이용하여 보험사기 적발모형을 개발할 수 있다는 것을 이론적으로 보여주었고, 김헌수(2000)는 실제 자동차보험 보험사기 데이터를 이용해서 AHP 보험사기적발모형의 prototype을 보여주었다.

2002년 *The Journal of Risk and Insurance*는 보험사기 연구에 대한 집중조명이 있었다.³⁾ Derrig(2002)은 보험사기에 대한 현재까지 연구를 총괄적으로 검토·요약하고 향후 연구방향을 제시하였다. Tennyson and Salsas-Forn(2002)은 최적계약에 대한 이론적 연구를 통해서 보험금 지급후 최적 감사전략을 제시한 이론 연구방향과 보험사기 적발모형을 통계적으로 제시하는 실증연구 방향을 연결하려는 시도이다. 즉 보험회사의 보험금 지급후 감사 자료를 분석해서 이론적 연구들이 제시한 예측들을 선형적으로 검증하여 이론적 연구들의 결론을 지지하는 결과를 발표하였다.

Artis et al(2002)은 보험사기 적발모형에서 비사기(honest claims)라 분류된 건수중 상당수는 사기(fraudulent claims)일 수 있다는 가정에서 출발한다. 이 연구는 최우추정(Maximum likelihood estimation)을 통해서 분류오류(misclassification error)를 측정하고, 이 오류를 고려한 경우와 고려하지 않은 경

3) *The Journal of Risk and Insurance*(2002)에는 총 6개의 보험사기 관련 논문이 출판되었다.

우의 보험사기 적발모형의 차이가 심각하다는 것을 지적하였다. Brockett et al(2002)은 PRIDIT(Principal Component Analysis of Riduals)이란 일종의 ‘unsupervised learning’ 기법을 통해서 전통적인 ‘supervised learning’ 형식의 단점을 극복할 수 있음을 보였다. 즉, 전통적인 보험사기 score model(회귀분석, 로짓분석, 판별분석, 인공신경망분석) 개발에서 전제 조건으로 필요한 훈련표본(training sample)과 종속변수 없이 현실적으로 활용가능한 보험사기 모형을 제시하였다. Viaene et al(2002)은 로지스틱회귀분석, C4.5 의사결정나무 및 K-근접 이웃모형 등 전통적인 의사결정모형뿐 아니라 베이지언인공신경망모형(Bayesian learning multilayer perception neural network)과 최소자승벡트기계학습법(least square vector machine learning) 등 최신 통계학적 기법들의 예측성과를 서로 비교하여 최적의 보험사기적발모형을 찾는 연구를 진행하였다.

Ⅲ. 링크분석(Link Analysis)

1. 링크분석 개요

가. 링크분석의 정의와 이론적 배경

링크분석(Link Analysis)은 “인간은 관계를 통해서 존재하며, 만일 인간이 누구와도 아무런 관계도 없다면 그곳에는 아무 것도 없다”라는 사회적네트워크이론에 기초한다. 이는 모든 인간은 혈연, 지연, 학연의 기본적인 관계뿐 아니라 일거수일투족이 타자와 어떤 관계를 갖는다는 것이다(Wasserman and Faust(1994)). 이러한 무수한 관계중 흔적이 남지 않는 관계도 있지만 그 흔적이 데이터로 남는 관계도 많다. 예를 들어 아침에 일어나서 세면을 했다고 했을 때 ‘물과의 관계’는 시간이 지나면 곧 사라지지만 출근길에 휴대폰으로 동료에서 전화를 했다면 ‘전화로 연결된 관계’의 흔적은 남는다. 따라서 링크분석이나 네트워크 관점에서 보면 어떤 개체의 속성(attribute)보다는 개체간의 관계(relation)가 보다 중요하며, 사회현상

을 분석함에 있어서도 현상의 속성보다는 현상과 관련된 것을 분석한다.

Byrne(2002)에 의하면 '링크분석은 개체간의 관계에서 발생한 데이터 속에서 패턴 또는 움직임을 추출하기 위해 서로 연관되어 되어 있는 개체들의 관계를 네트 워크화하는 과정' 이라 정의하였다. 링크분석은 실무에서 ERD(Entity-Relationship Diagrams), Nodes-and-Links, Directed Graph 등의 용어로 사용 되기도 한다.

〈표 5〉 링크분석 적용 분야

분 야	데 이 터	적용분야의 예
금융거래	금융거래 데이터, profile자료	자금세탁(Money Laundering)
통신 네트워크	통신 내역, profile자료	통신사기(Telecommunication Fraud)
사업관련분석 (Business Intelligence)	고객, 서비스, 공급자, 파트너, 거래내역	비공식 네트워크 분석, 서브그룹 분석
웹 사이트 검색	페이지, 콘텐츠	Hyperlink-Induced Topic Search
GIS	교통흐름 데이터	병목구간조사, 최단구간조사
범죄·사기 조사 자료: AAAI(1998)	거래내역, 통화정보, 관련자 정보	보험사기, 범죄자 적발

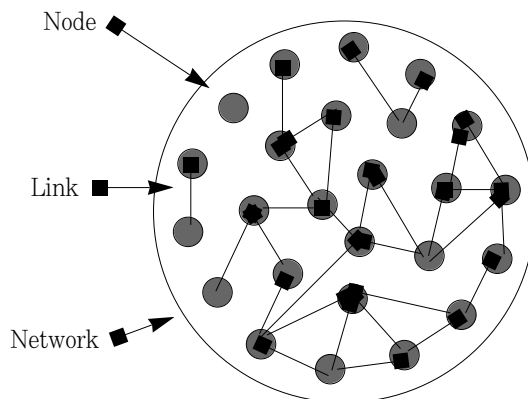
링크분석은 범인 추적 및 조직내부 상호작용 분석 등 다양한 분야에서 활용되고 있다. 예를 들면 웹사이트 연관분석(Search engines, Site maps), 네트워크 트래픽 분석(Web caching, Public transport), 데이터마이닝분석(E-commerce, Telecommunication services), 사회적 네트워크 분석(Social structures, Policy-making Decision), 텍스트 분석(Co-reference, Co-citation), 의사결정지원분석(Decision-making, Logistics), 사기적발(Money laundering, Calling cards fraud)등에 활용되고 있다(AAAI(1998)). 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 서로 연결되어 있음을 생각할 수 있기 때문이다.

나. 링크분석(Link Analysis)의 구성과 특징

링크분석은 Link, Node, Value, Direction, Network로 구성되어 있다. 여기서 Node는 임의의 Entity로 표현되어 있는데, 이는 데이터베이스의 속성을 Entity로 지정할 수 있음을 의미한다. 그러나 정량적인 숫자는 Entity가 될 수 없으며, 일반적으로 문자형 타입(전화번호, 주소, 차량번호 등)을 가지는 속성값이 Entity로 사용된다. 또한 Value는 두 개의 Entity간의 관계의 정도(빈도) 및 양을 의미하는 것으로 전화통화 횟수, 통화시간이 좋은 예라 할 수 있다. 또한 Direction은 두 개의 Entities간의 방향을 의미하는 것으로, 전화통화에서 발신지와 착신지를 나타내는 화살표를 생각할 수 있다. 그리고 Network은 노드와 노드가 연결되어 집단을 이룬 것으로 정의할 수 있다.

〈그림 1〉 링크분석의 기본 구조



2. 링크분석과 Visual Data Mining

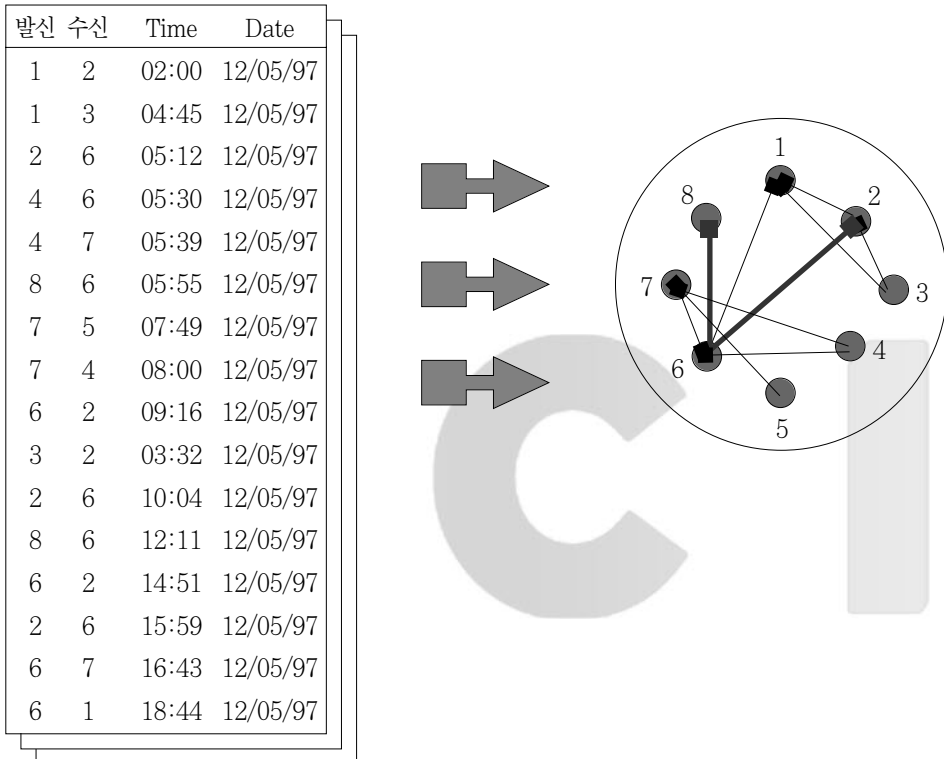
대부분의 과학적 분석기법은 정량적기법 즉, 데이터중심적인 접근방법이 주를 이루고 있었지만 이러한 방법들은 문제의 영역(domain)이 비체계적이고 문제나 대상이 수시로 변형, 발전되는 경우 특정 패턴을 파악하기가 매우 어렵다(Davison (1993)). 보험사기 적발도 마찬가지다. 단순한 보험사기가 아닌 계획적인 공모에 의한 보험사기의 경우 변수의 가중치로 추정하게 되는 회귀분석이나 인공신경망기법으로는 추적하기가 어렵다. 이렇게 문제영역이 정형화되어 있지 않고 문제가 수시로 변형되는 경우 전문가가 직접 자신의 지식을 이용해서 데이터의 패턴을 효율적으로 찾아내는 방법이 VDM(visual data-mining)이다.

범죄처럼 분석대상이 복잡한 경우 자료를 그림으로 보여주는 것이 텍스트나 표로 나타내는 것보다 훨씬 효과적이다. 기존 연구에 의하면, 문제 관련 자료가 복잡하고 정형화가 덜 될수록 전통적 방법보다 VDM 접근방식이 쉽고(easy to learn), 결과는 안정적이며(reliable), 대단히 비용효과적(highly cost-effective)이라고 보고하고 있다(Andrew and Peterson(1990)). VDM에서는 자료 스스로가 정보를 제시하고 이를 바탕으로 조사자는 자료를 다양한 방법으로 재구성할 수 있도록 직접 명령한다. 이 과정은 데이터중심적 분석과는 달리 블랙박스가 아니며 조사자에게 왜 그런 결과가 나왔는지 시각적으로 보여준다.

VDM은 인간의 인지적 능력을 극대화하는 시각적 방법에 의존하는 것인데 여기에는 링크분석(link Analysis) 외에도 군집화(clustering), 계층화(hierarchies), 자가구성지도(self-organizing map) 등의 분석방법이 있다. Tamassia 외(2001)에 의하면 Visualization은 복잡한 네트워크 구조와 기대하지 않은 패턴의 명확화에 있어 인간 의사결정에 매우 효과적임을 증명하였다.

〈그림 2〉처럼 많은 대상(object)의 전화내용을 표 대신에 링크로 시각화한다면 다음 사실을 쉽게 알 수 있다. 8번과 5번은 교신대상이 하나인 고립대상(isolated object)이고 6번은 이 집단의 중심이 되는 인물로 특히 2번과의 교신이 가장 많아 6번과 2번은 어떤 특별한 관계가 있을 것이라는 것을 추정할 수 있다(Westphal and Blaxton(1998)).

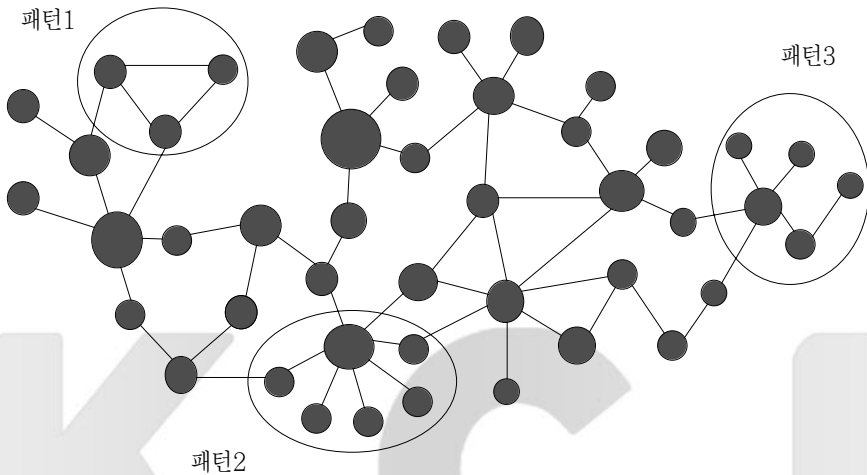
〈그림 2〉 전화내용을 시각화한 결과



3. 링크분석에 사용되는 분석패턴

링크분석에서 나타나는 대표적인 패턴은 상호연관성(Interconnectivity)에 의해서 파악된다. 대상상호간의 링크를 통하여 자료의 상호연관성을 파악하고 이를 통해서 특정한 패턴을 보이는 집단을 발견하는 것이다. 예를 들면 〈그림 3〉에서 패턴 1은 세 대상들이 비슷하다는 것을 보여주는 반면 패턴2는 특정 대상이 중심이 되어 있다는 것을 알 수 있다.

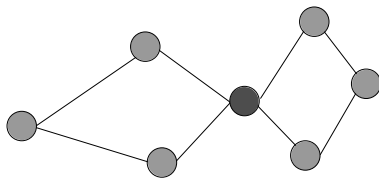
〈그림 3〉 링크분석을 통해서 나타난 패턴들



링크분석을 이용해서 범죄를 분석하는 경우 다양한 링크패턴이 등장하게 되는데 그 중 가장 일반적으로 나타나는 패턴을 살펴보면 다음과 같다(Westphal and Blaxton(1998)).

가. 관절점(Articulation Point)

관절점(Articulation Point)은 하위네트워크들을 연결하는 점으로 전체 네트워크



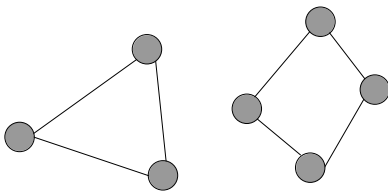
구조에서 중요한 역할을 하는 Entity를 말한다. 따라서 네트워크내에서 해당 Entity가 제거된다면(보험사기 네트워크에서 주모자가 체포되면) 전체적인 네트워크 구조에 큰 영향을 받을 수 있다는 것을 암시한

다. 이 관절점이 없다면 전체 그룹에 어떤 영향을 미치는지는 시나리오 분석을 통해서 관절점의 영향력을 분석할 수도 있다. 통신사기 예를 들면, 만일 특별한 전화 번호가 다른 일련의 전화번호 그룹들과 연결되어 있는 관절점이라면 이 특수 전화

번호는 통신사기 적발에 매우 유용한 단서가 될 것이다. 관절점과 유사한 것으로 공통요소(Commonality)란 것이 있는데 이는 두 개 이상의 Entities(subnetwork이 아닌)에게 공통요소에 의해서 연결된 것을 의미하는데 공통요소는 관절점보다 훨씬 자주 발견되는 특성이라고 할 수 있다. 이 공통요소 패턴은 다양한 조건을 만족하는 알고리즘을 조사자가 즉흥적으로 만들어서 발견할 수 있다. 예를 들어 보험사기의 비정상적인 대상을 추출하기 위해서 '피보험자가 지난 2년간 3번 이상 보험금을 받았고, 동일한 우편번호를 사용하면서 지난 2년 동안 주소가 3번 이상 바뀐 집단을 추출하라'는 알고리즘을 만든다면 이 패턴에 맞는 대상을 보여줄 것이다.

나. 분리 네트워크(Discrete Network)

전체 데이터에는 다양한 하위네트워크가 존재한다. 조직내 상호작용을 연구하는

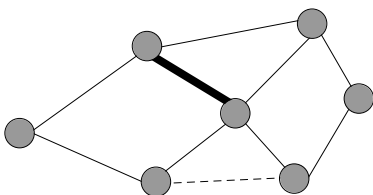


링크분석에서 보면 주류 네트워크와 분리된 네트워크는 조직내의 주류 의사소통 경로나 정보의 흐름에서 소외되었다고 볼 수 있다. 이 경우 분리 네트워크에 속한 대상들은 자기들만 의사소통을 하고 있어 독립

성을 유지하고 자체만의 정체성을 가진 경우가 많다. 범죄분석에서는 분리 네트워크는 독립적으로 움직이는 범죄집단을 의미할 수도 있을 것이다. 무연결점(Missing Connection)은 분리 네트워크(Discrete Network)의 특별한 형태로 단일 Entity로 하위 네트워크가 형성되어 있음을 나타낸다.

다. 링크의 상대적 강도

네트워크 내에 각 노드들간의 관계의 상대적 강도를 분석할 수 있도록 하는 방법

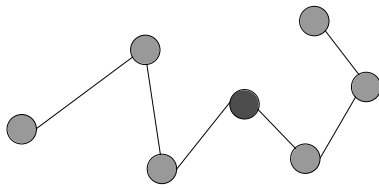


이다. 하나의 개체는 특정 개체와 지속적으로 수많은 관계성을 가질 수 있으며, 이러한 관계의 강약 정도는 연결횟수에 의해서 주로 표현된다. 예를 들면 9·11테러 후 주요 핵심 테러 용의자 적발을 위해 링크분

석을 활용한다고 가정하였을 경우, 테러주동자 A가 테러리스트 B와 은행 거래 관계에서 엄청난 금액이 오고 갔다면 둘 간의 관계는 매우 강하게 나타나게 된다. 즉, 테러주동자 A는 많은 테러리스트(B, C, D, E, F...)등과 관계를 가지고 있지만 테러리스트 B와 특히 관계가 깊다고 추정할 수 있을 것이다.

라. 경로분석(Pathway Analysis)

분석의 두 대상이 몇 개의 링크에 의해서 연결되는지, 그리고 분석 대상간 가장



간단한(복잡한) 연결체계는 어떠한 것인지를 파악하는데 유용한 분석틀이다. 예를 들어 뉴욕에서 LA로 자동차로 가는 경우 가장 거리가 짧은 길은 어떤 길인지, A라는 인물과 B라는 인물은 몇 단계를 걸쳐서 연

결되는지를 파악하는데 적합한 방법이다. 이 분석틀은 웹데이터마이닝에서 특정사이트에서 특정사이트로 연결되는데 그 경로를 분석하는 등 최근 활성화된 인터넷마케팅에서도 활용할 수 있는 도구이다.

Ⅳ. 링크분석(Link Analysis)의 효과성 검증

1. 효과성 검증 방법 및 과정

링크분석은 개별 건수의 보험사기 확률을 평가하는 지수모형(scoring model)이 아니고 혐의가 있는 건수를 추적하는 적발모형(detection model)이기 때문에 무작위추출 등으로 선택된 표본을 사용하기 어렵다. Random으로 일정한 데이터를 추출하였을 경우, 중요한 Node에 대한 관계 데이터의 유실로 관련자 또는 관련된 Node가 Link로 나타나지 않기 때문이다. 예를 들면 무작위추출의 경우 용의자 A 씨와 관계있는 사람이 대거 빠질 수 있기 때문이다. 따라서 링크분석 효과성을 검증

하기 위해서는 모집단, 즉 보험사의 전체 데이터를 대상으로 검증하여야 한다. 그러나 수 백만건 이상인 보험회사의 전체 데이터를 사용한다는 것은 현실적으로 너무나 많은 비용과 시간이 수반되므로 우리는 특정지역(XX시)에서 발생한 보험금 청구 건수만을 이용하여 링크분석의 효과성을 검증하도록 한다.⁴⁾ 이 경우 보험사고 용의자가 타지역에서 보험사고를 저질렀다면 그 용의자는 데이터에서 제외된다는 한계는 있다. 링크분석의 효과성을 검증하기 위하여 K보험사의 2000~2002년 특정지역 자동차보험 보험금 지급 데이터를 사용하였다.⁵⁾ 입수한 데이터를 분류하면 아래 <표 6>과 같이 사고관련 정보, 내부 직원 정보, 피보험자(물) 정보, 보험금 수취인 관련 정보 및 병원 및 정비소로 5가지로 분류할 수 있다.

<표 6> 보험사기 관련정보의 분류

데이터(노드) 분류	내 용
사고관련 정보	사고발생시점, 사고자, 경찰신고 등 관련 데이터
내부 직원 정보	사고담당 보상직원, 설계사 등과 정보
피보험자(물) 정보	피보험자와 피보험물에 관련된 정보
보험금 수취인 관련정보	보험금을 받은 수취인과 관련된 정보
병원 및 정비소 정보	사고자를 치료한 병원과 사고차량 수리한 정비소

링크분석을 위해서 필요데이터 지정과 관계데이터 선정을 위해서 문헌연구에서 정리한 국내외 연구자료외에도 보험사의 SIU팀, 금융감독원의 보험사기조사실 그리고 손해보험협회의 보험범죄팀을 면담하였다. 그 과정은 문헌연구와 SIU팀 면담을 토대로 필요데이터를 선정하고, 보험사에서 확보된 자료를 토대로 관계데이터를 선정하였다.

4) 본 보고서의 범위를 벗어나 생략되었지만 링크분석을 위해서는 데이터 정지작업 및 기존 데이터를 관계형 데이터 형태로 재구성하는 등 상당한 데이터 준비작업이 필요하다.

5) 데이터를 제공해준 K보험사의 요청으로 데이터수집 대상 지역 등의 정보를 구체적으로 명시할 수 없음을 밝힌다.

확보한 데이터 중에서 링크분석에 유용한 관계데이터로 청구번호, 계약자 및 청구자(사고자)를 포함하는 이름, 주소, 전화번호, 자동차 번호, 계약번호 등 11개가 설정되었으며, 링크는 이들 노드간의 관계에서 유의하다고 판단되는 10개가 설정되었다. 이렇게 설정된 총 분석용 데이터의 현황은 청구건수가 96,819건, 계약건수가 885,558건, 당사자(사고자, 청구자 등)가 156,019건, 지불건수가 258,664건, 해당 자동차 수는 42,234건으로 나타났다. 본 링크분석은 Netmap for Claim이란 소프트웨어를 사용하였다.

〈표 7〉 실증분석에 활용된 Node와 Link

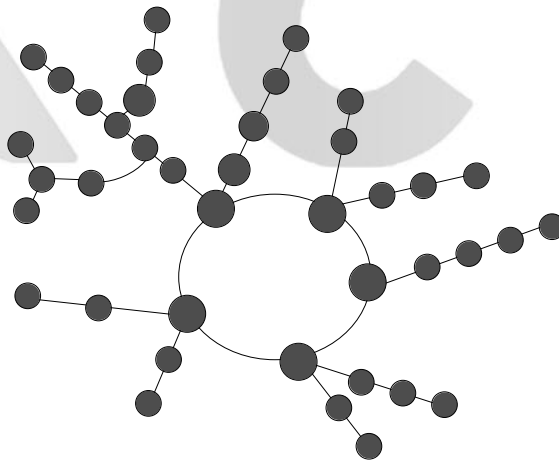
Node	Link
청구번호(Claim) 계약자 주민번호(Name) 청구자 주민번호(Name) 사고자 주민번호(Name) 집주소(Address) 전화번호(Phone) 자동차번호(VIN) 계약번호(Policy) 병원번호(Name) 정비소번호(Name) 사고당시 역할(Role)	Claim/Address Claim/Name Claim/VIN Name/Address Name/Name Name/Phone Name/Role Name/VIN Policy/Claim Policy/Name

링크분석에서는 지정된 조건을 충족하는 대상을 한꺼번에 찾아내는 Emergent Group Algorithm이라는 기법과 한 특정 노드를 중심으로 링크를 단계적으로 추적하는 Step Link 분석 방법이 있는데 우리는 두 분석방법을 전부 사용하였다.

2. Emergent Group Algorithm을 통한 링크분석 효과성 검증

Emergent Group이란 '서로 연관되어 있으면서 매우 긴밀하게 관련되어 있는 특수한 그룹'을 의미하며 Emergent Group Algorithm이란 Emergent Group을 컴퓨터로 신속하게 찾아내는 분석알고리즘을 의미한다. 통상 보험사기 공모집단 전체를 찾아내기 위해서는 서너명의 SIU가 경찰 등 공권력을 빌려서 몇 달 동안 수사를 하는 것이 보통이다. 그러나 이 링크분석의 Emergent Group Algorithm은 아무런 단서 없이 단 몇 분 이내에 공모사기 의심자를 찾을 수 있어 매우 효율적이다. Emergent Group Algorithm은 Quantitative Data Mining의 특성 중의 하나인 자동분류방법(Automatical classification method)을 링크분석에 적용한 것이다.

〈그림 4〉 Emergent Group의 한 예

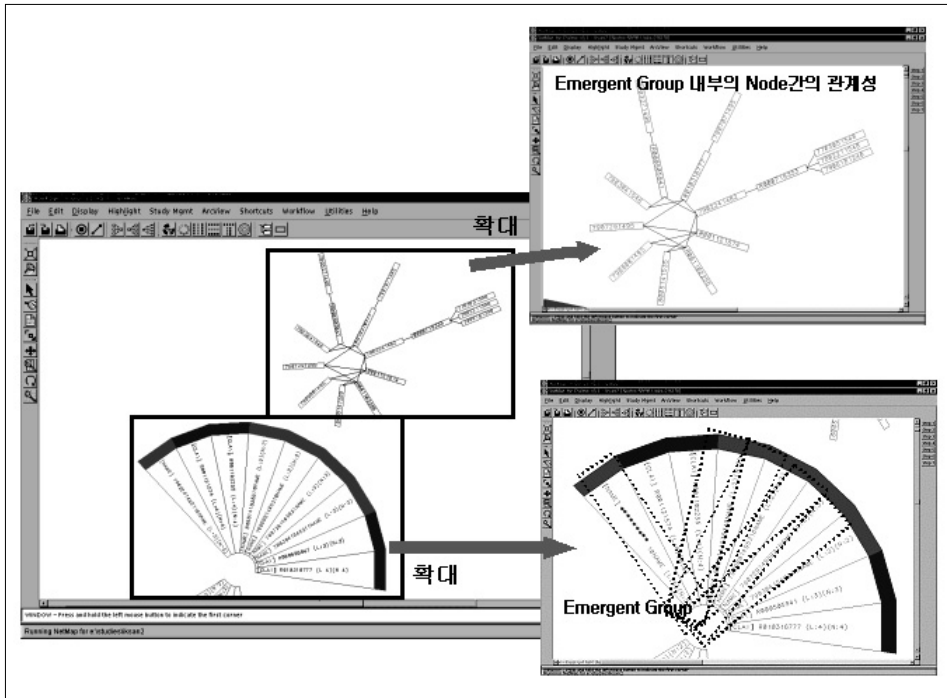


우리는 보험사기에 대한 어떠한 사전 정보 없이 Emergent Group Algorithm을 사용해서 사기 혐의자를 추측해보고 이를 K보험사가 보유하고 있는 실제 사기 혐의자 리스트와 비교해 보기로 하여 링크분석의 Emergent Group Algorithm의 정확성을 평가하고자 하였다. 본 분석에서는 다음 3가지 조건을 만족하는 Emergent

Group을 추출하였다.⁶⁾

- 한 Group내에 3개 이상의 노드가 존재할 것.
- 하나의 노드는 해당 Group내에 최소한 2개 이상의 다른 노드와 링크되어 있을 것.
- 하나의 노드는 해당 Group내의 전체 링크중 최소한 1/2 이상을 다른 노드와 연결되어 있을 것.

〈그림 5〉 Emergent Group을 찾은 링크분석 화면



6) Emergent Group Algorithm의 조건조합은 변경할 수 있으나 본 연구에서는 상업용 Full License Version이 아닌 데모용 버전을 사용한 관계로 연구자가 새롭게 EGA 조건 조합을 제시할 수는 없었다.

EGA를 사용한 결과 <표 8>에서 보듯이 4명의 보험사기 의심자가 적발되었는데 K보험사가 보유한 보험사기 리스트와 비교할 때 3명이 일치하여 EGA는 보험사기 적발에 상당한 잠재력이 있다는 것을 알 수 있었다. 경성보험사기 형태는 다양하게 변형될 가능성이 있으므로 실무에서는 EGA와 같은 방법으로 새로운 조건조합을 제시하여 이를 충족하는 집단을 우선적으로 수사하도록 할 수 있을 것이다.

<표 8> Emergent Group Algorithm에 의한 의심자 대 실제 보험사기자

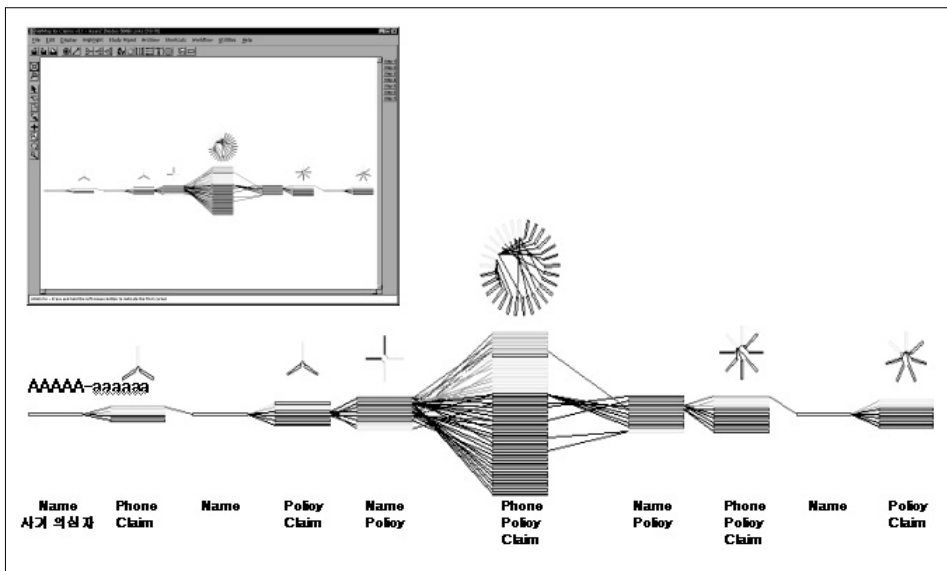
구 분	Emergent Algorithm에 의한 의심자 리스트	K보험사의 실제 보험사기범 리스트와 일치여부(○, ×)
Emergent Group Algorithm으로 발견된 사기공모 적용	혐의자A	○
	혐의자B	○
	혐의자C	○
	혐의자D	×

3. Step Link 분석을 통한 링크분석 효과성 검증

Step Link 분석은 혐의점이 있는 특정 노드를 출발점으로 하여 이 출발점과 연결된 링크와 노드를 차례차례 단계적으로 추적하는 링크분석의 일반적인 방법이다. Step Link 분석의 정확성을 평가하기 위해서 K보험회사에서 제공한 보험사기 의심자 1명(A씨)의 자료에서 출발하여 관련된 보험사기 의심자를 전부 파악해보기로 한다. 그리고 Step Link 분석을 파악된 보험사기 의심자와 K보험회사에게 이미 파악하고 있는 보험사기 의심자 리스트를 비교해서 Step Link 분석의 효과성을 검증하기로 한다.

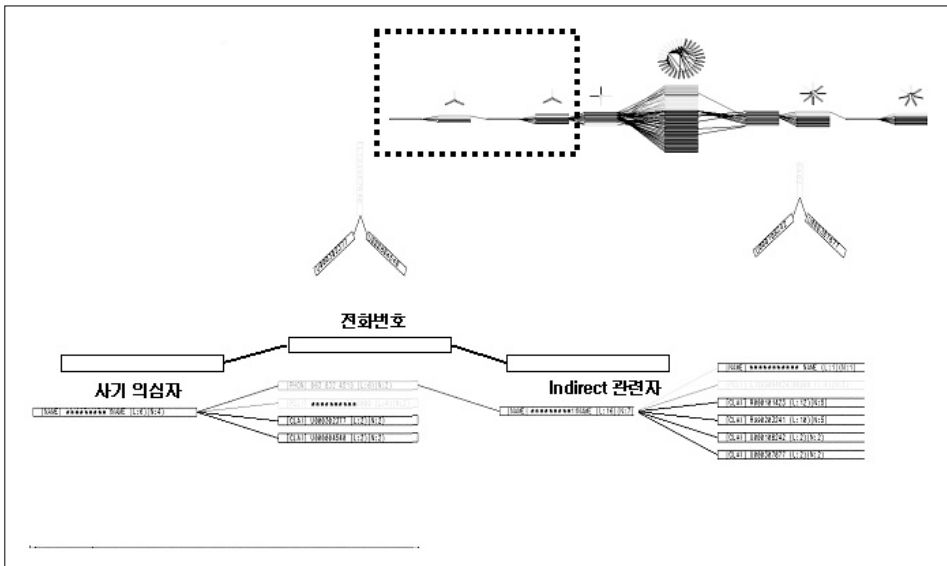
Step 1: A씨와 관련된 링크는 예상대로 매우 복잡한 관계를 보여주고 있는데 맨 왼쪽의 사기 의심자 A씨(Aaaaaa-aaaa)는 여러 가지 노드(정보)와 복잡하게 연결되어 있다는 것을 볼 수 있다. 아래 <그림 6>부터 단계별로 이 링크가 어떤 노드와 연결되어 있는 지 실제 컴퓨터화면을 중심으로 살펴본다.

<그림 6> Step 1: A씨(Aaaaaa-aaaa)와 관련 링크 Overview



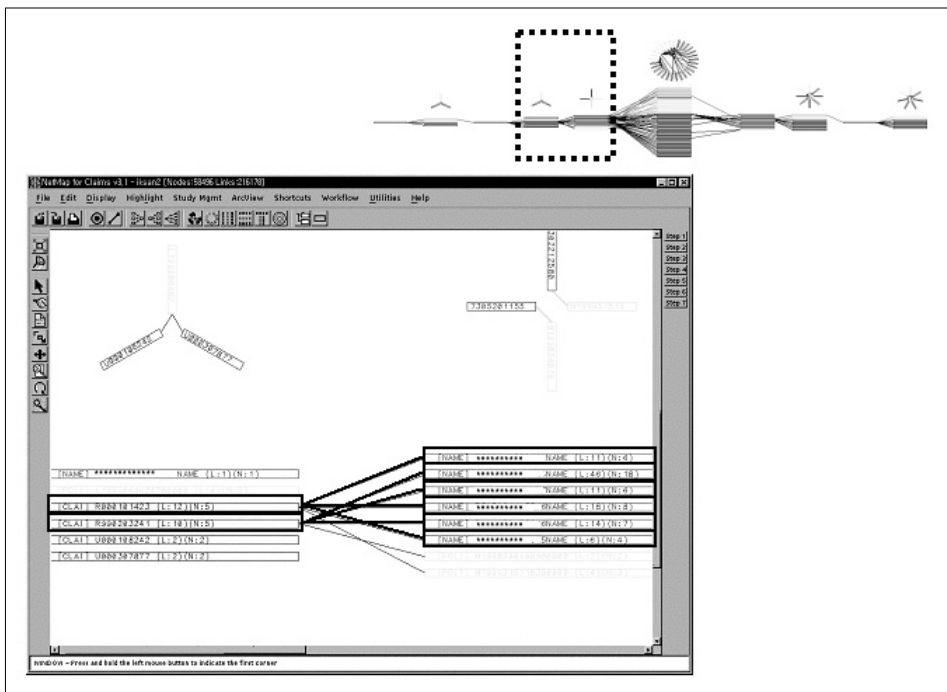
Step 2: <그림 7>은 보험사기 혐의자 A씨가 네 가지 정보(node) 즉, 전화번호, 보험계약번호, Claim청구번호1, Claim청구번호2와 연결되었다는 것을 보여준다. 나아가 A씨와 관련된 전화번호가 B씨와 공유하고 있다는 사실을 보여준다. 그런데 새롭게 등장한 B씨(Indirect 관련자)는 한 개의 계약과 네 개의 Claim청구에 관련되어 있다는 사실을 알 수 있다.

<그림 7> Step 2: A씨에 대한 링크로 Indirect 관련자 발견



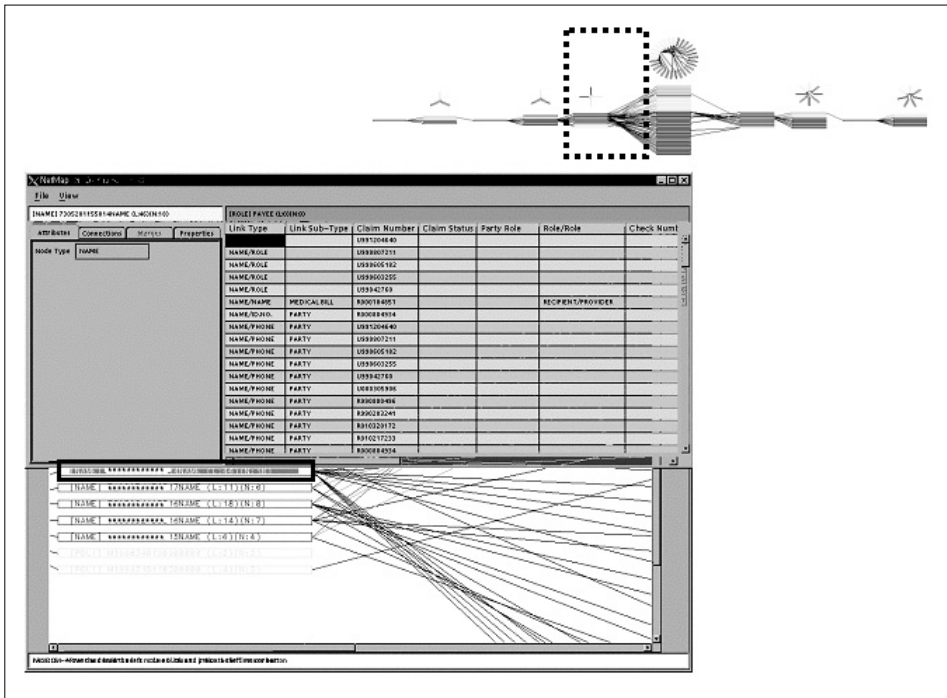
Step 3: <그림 8>에서 보듯이 Indirect 관련자 B씨와 연결된 보험금 청구건수 4개 중 2개의 보험금 청구건수는 새로운 인물과 복잡하게 관련되어 있다. 이렇게 B씨와 관련된 보험금 청구건수는 6명의 새로운 인물과 관련되어 있어 B씨는 뭔가 보험사기와 관련성이 있을 수 있다는 것을 암시한다.

<그림 8> Step 3: Indirect 관련자 B씨와 관련된 링크



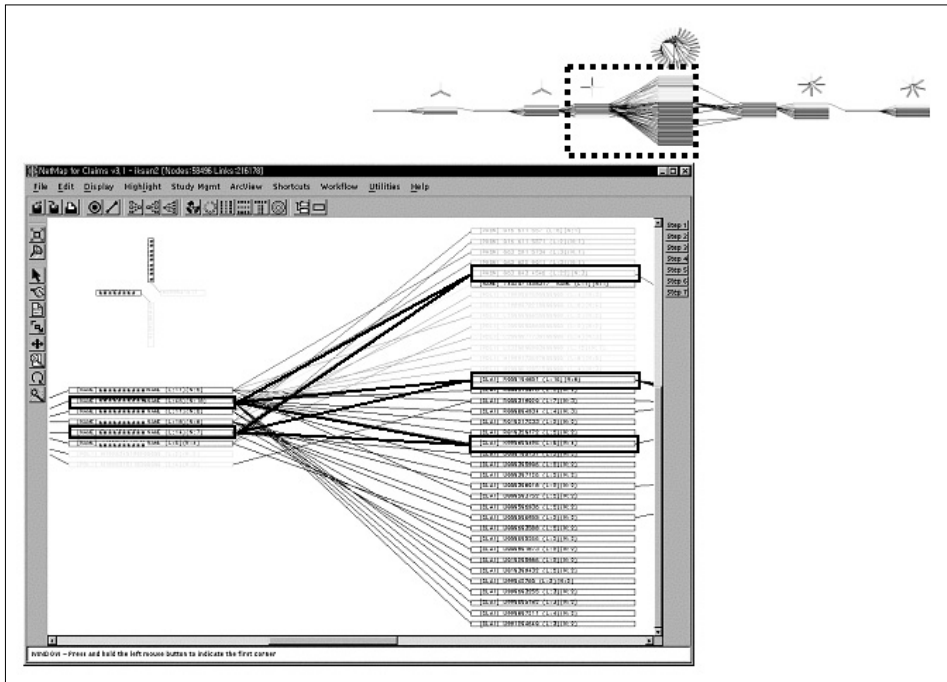
Step 4: <그림 9>에서는 앞 그림에서 등장한 새로운 인물중 C씨의 경우 18개의 노드와 46개의 링크를 가진 것으로 나타나 극히 비정상적인 인물로 추측할 수 있다. 또한 C씨는 이미 수십번 보험금을 청구한 적이 있는 것으로 나타나 보험사기와 관련이 있을 가능성이 매우 높은 것으로 판단할 수 있다.

<그림 9> Step 4: 6명의 보험사기 의심자에 대한 무수한 링크 발견



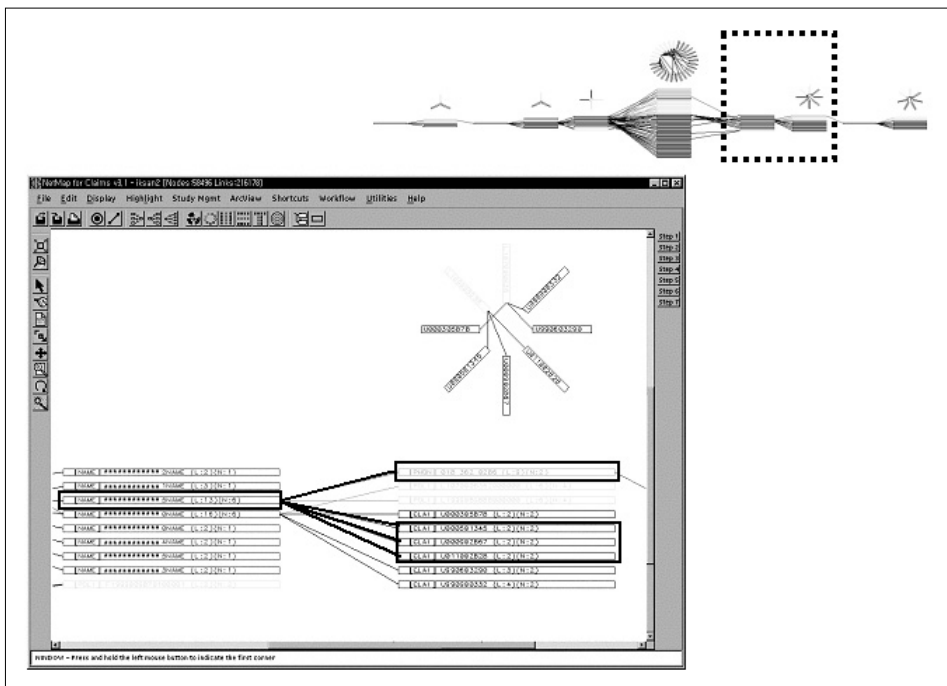
Step 5: 특히 앞에서 나타난 6명 중 D씨와 G씨는 세 가지 노드에 대해 공통적으로 링크되어 있다는 것을 <그림 10>을 통해서 알 수 있다. 특히 하나의 전화번호와 두 건의 보험금 청구에 대해서 D씨와 G씨가 같이 연계되어 있어, 이러한 관계가 우연이 아니며 보험금 청구와 관련해야 보험사기 공모 가능성이 있다는 것을 암시한다.

<그림 10> Step 5: D씨와 G씨가 관련된 사기혐의 높은 노드 발견



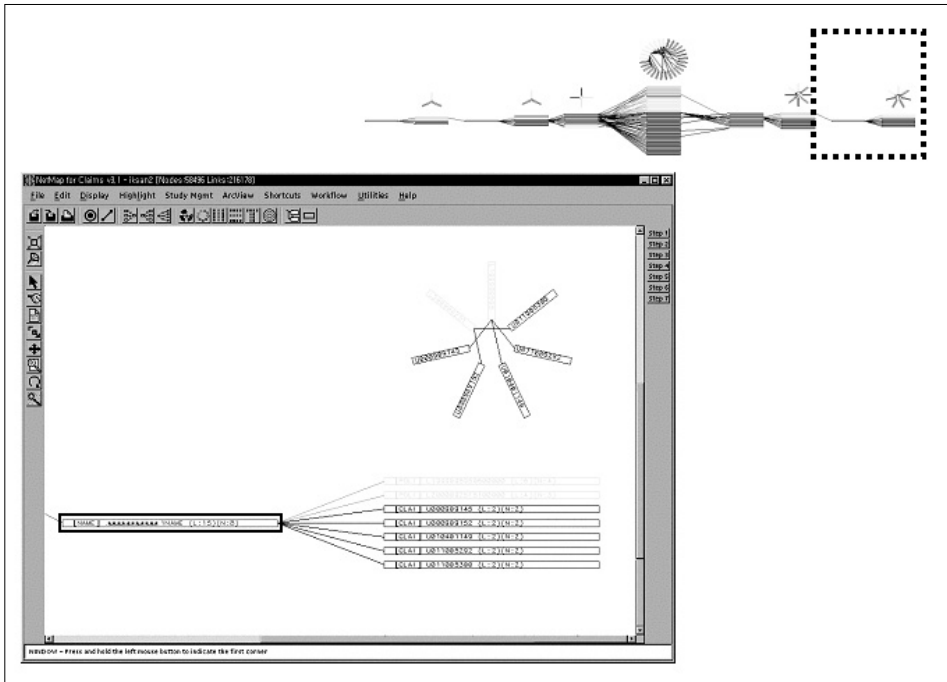
Step 7: 이전에는 혐의가 있는 링크를 발견할 수 없었던 인물 K씨가 다시 새로운 전화번호와 세 개의 보험금청구와 연결되어 있다는 것을 <그림 12>가 보여주고 있다. 그런데 K씨와 연결된 이 전화번호는 다른 노드와 비정상적으로 많이 연결되어 있어 K씨 역시 보험사기 공모 참가자일 가능성이 있다는 것을 뒷받침해주고 있다.

<그림 12> Step 7: 새롭게 등장한 K씨의 링크 현황



Step 8: <그림 13>에서 보듯이 새로운 전화번호는 다시 2건의 보험계약과 5건의 보험금청구에 연결되어있다는 것을 알 수 있다. 즉, 2건의 보험계약에 이 전화번호가 연락처로 기재되어 있을 뿐만 아니라 이 전화번호는 5건의 보험금 청구에도 관련되었다는 것이다. 사기조사팀은 이 5건의 보험금 청구가 합법적이었는지 아니면 보험사기의 혐의가 있는지 밝혀야 할 것이다.

<그림 13> 의심스러운 전화번호가 제시한 새로운 링크



라. Step Link 분석 결과

Step Link 분석을 통해서 총 15명의 보험사기 공모 의심자를 추출할 수 있었다. Step Link 분석을 통해서 찾은 보험사기 의심자 15명을 보험회사가 보유하고 있는 보험사기자 실제 리스트와 비교해 본 결과, 15명의 의심자중 9명이 실제 보험사기자 리스트와 일치하였다. 이는 아무런 사전정보 없이 오직 데이터만을 이용한 Step Link 분석의 보험사기 적발능력이 상당히 정확하다는 것을 보여준다. 링크분석 결과 보험사기 의심자로 파악되었지만 실제 보험사기자 리스트에 없었던 나머지 6명도 증거불충분 등으로 보험사기자로 입건이 되지 않았을 가능성이 있어 이들 역시 보험사기 공모에 가담되었을 가능성을 배제할 수 없다.

〈표 9〉 Step 분석 효과성 검증 : A씨 관련 공모혐의자 대 실제 적발자

	스텝분석에 의한 공모 혐의자	실제 적발된 자와 일치여부 (○, ×)
A씨 관련 보험사기 공모의심자	혐의자1	○
	혐의자2	○
	혐의자3	×
	혐의자4	×
	혐의자5	○
	혐의자6	○
	혐의자7	○
	혐의자8	×
	혐의자9	○
	혐의자10	○
	혐의자11	○
	혐의자12	○
	혐의자13	×
	혐의자14	×
	혐의자15	×

4. 효과성 검증 결과에 대한 해석

EGA(Emergent Group Algorithm)와 Step Link 분석 결과 링크분석은 적발 정확성(detection accuracy) 측면에서 상당한 가능성을 보여주었다. EGA의 경우 파악한 4명의 의심자 중 실제 보험사기자와 동일한 사람이 3명이었고, Step Link의 경우는 추출한 15명의 의심자 중 실제 보험사기자와 동일한 사람은 9명으로 나타나 링크분석이 사기적발에 상당히 효과적일 수 있다는 것을 시사하였다.⁷⁾

그러나 현 단계에서 링크분석 결과를 기존의 통계적방법 결과와는 직접 비교하기는 어려운데 이는 사과와 오렌지의 비교가 될 수밖에 없기 때문이다. 링크분석은 주어진 단서와의 관계를 통해서 그와 관련된 보험사기를 찾는 비통계적 '사례추적방식'인 반면 통계적방법은 자료 전체를 대상으로 사기와 비사기를 한꺼번에 분류하는 통계적 '분류방식'이다. 즉, 링크분석은 주어진 사례와 단서를 근거로 그것과 링크(관련)되어 있는 의심자를 추적하는 방식이므로 데이터내에 모든 사기 의심자를 전부 추출하기 위해서는 다양한 링크분석 알고리즘으로 보험사기 관련 모든 단서와 사례를 반복적으로 전부 사용하여야 한다(exhaustive application). 그러나 본 연구에서 활용할 수 있는 단서와 사례는 Emergent Group Algorithm 분석과 Step Link 분석에 각각 한 가지일 뿐이다.⁸⁾

EGA의 경우 특정 조건조합을 충족시키는 노드를 파악하는 것이며 전체 데이터

7) 본 분석의 대상인 보험금청구건수는 96,819건인데 이 중 경성사기건수가 10% 정도라는 기존 연구(Weisberg and Derrig(1991, 1992))의 주장을 적용하면 경성사기건수가 몇 천건이 되어야하고 보험사기자도 수백명 이상이 될 것이므로 링크분석의 결과가 보잘것 없는 것이 아니라는 주장이 있을 수 있다. 그러나 링크분석은 '사례추적' 방식이므로 데이터에 있는 보험사기를 전부 예측하는 통계적방법과 전혀 다르다. 그리고 문헌연구에서 설명한 것처럼 경성사기율 10%라는 높은 수치는 배심원에 의한 징벌적 처벌(punitive damage)로 이익을 쉽게 편취할 수 있는 미국의 사법제도의 영향을 받았을 것이고 조사 대상이 보험회사였기 때문에 보험사기율을 과대평가했을 가능성도 높다. 만일 보험소비를 대상으로 경성사기율을 조사 했다면 그리고 조사장소가 한국이라면 경성보험사기율은 1%를 초과하지 못할 것이라는 것이 연구자의 생각이다.

8) 전체 데이터에 대한 보험사기 적발 정확성을 평가하기 위해서 과거에는 실제 사기를 정확하게 예측한 비율인 성공확률(Percentage Correctly Classified)을 사용했으나 성공확률과 실패확률을 동시에 고려한 ROC(Receiver Operating Characteristic) 곡선이나 AUROC(the Area under ROC)을 사용하는 것이 바람직하다(Viaene외(2002)).

에서 한꺼번에 ‘모든 보험사기범’을 찾아내는 방식은 아니다. 따라서 EGA의 경우 다양한 보험사기 특성에 따라서 새로운 조건을 제시하고 이를 충족하는 노드를 찾아야 한다. 예를 들면, 병원공모 보험사기와 자동차추돌사고 조작 보험사기는 보험사기를 찾는 EGA 조건조합이 다를 수 있을 것이다. 본 연구는 EGA의 하나의 조건조합만을 적용하여 4명의 보험사기 의심자를 찾았고 이 중 3명이 실제 보험사기자였기 때문에 보험사기 적발에 상당히 효과적일 수 있다는 질적인 평가는 할 수 있으나 정확성을 통계적으로 나타내기는 어렵다.

Step Link 분석도 마찬가지다. 한 사람의 최초 의심자를 출발로 링크된 노드를 추적하는 방식이므로 전체 데이터를 한꺼번에 분류·스크리닝하는 통계적 모형과는 다르다. 하나의 사례를 출발로 그와 관련된 링크와 노드를 찾아가는 비통계적 사례추적방식이기 때문이다. 즉 Step Link를 통해서 찾아낸 15명의 보험사기 의심자 중 9명이 실제 보험사기자로 드러나 Step Link 분석도 질적으로 보험사기 적발에 상당히 효과적일 것이라고 평가할 수 있지만 정확성을 통계적으로 나타내기는 어렵다는 것이다.

V. 연구결과의 시사점과 활용 방향

IMF 이후 보험사기 건수의 증가율은 둔화되었으며 건수 보험사기 금액은 감소하였다. 그러나 소비자의 피해나 불만은 증가하는 경향을 보이고 있다. 이는 보험사기 근절을 위해 보험회사와 감독당국의 공격적인 대응이 선의의 소비자에게 대한 보험금 지급의 지연 또는 거부로 나타나기 때문인 듯하다. 보험사기에 대한 보험회사와 감독당국의 대응해법은 보험사기 적발 및 예방이라는 목표(objective)와 소비자의 권익보호라는 제약조건(constraints)이 있는 방정식이어야 한다. 이 방정식의 해는 보험사기 조사방법의 선진화를 의미한다. 본 연구에서는 선진화된 조사기법 중 하나인 링크분석의 기본적인 개념을 소개하고 보험회사 데이터를 사용하여 링크분석의 효과성을 검증하였다. 분석 결과 링크분석은 사기적발에서 상당히 효과적일 수 있다는 가능성을 발견하였다.

링크분석을 국내 보험사에서 적용하기 위해서는 먼저 보험회사 내부 데이터베이스를 관계중심 데이터베이스로 활용할 수 있어야 하며 추가적인 외부 데이터베이스 확보도 필요하다. 예를 들면, 관련자들의 관계를 파악할 수 있는 인구통계적 관계정보(졸업학교, 군경력 정보) 및 금융·신용정보 등이 필요한데 이 정보들은 보험가입자의 동의를 얻어서 합법적으로 확보하도록 해야 할 것이다. 나아가 기존의 속성중심 데이터베이스와 새로운 관계중심 데이터베이스와 연계해서 사용할 수 있도록 하는 것도 중요하다. 특히 인력측면에서 링크분석 등 새로운 조사기법을 잘 아는 전문가를 보상지원팀에 두고 SIU를 지원하든지 아니면 교육을 통해서 SIU요원이 직접 링크분석 기법을 익히는 것도 필요할 것이다.

링크분석은 보험사기 의심자 적발에 매우 효과적이지만, SIU(special investigation unit)기능을 대체할 수는 없다. 링크분석은 축적된 관계데이터를 통해서 보험사기 의심자를 수십분 이내에 추출하는 뛰어난 속도와 상당한 정확성을 가지고 있지만, 보험사기 의심자를 보험사기자라고 단정할 만한 구체적인 법적 증거를 확보하는 업무와는 별개다. 즉, 보험사기 의심자를 집중 조사해서 궁극적으로 보험사기자로 결정하는 것은 SIU의 고유업무이다. 따라서 링크분석이 SIU와 유기적으로 연계될 때 그 효과가 극대화될 수 있을 것이다.

본 연구와 연계된 향후 연구과제는 우리나라의 보험사기율 조사이다. 미국의 경성보험사기율이 보험금청구건수의 10%라는 주장도 있지만 이는 미국적 사법제도 환경과 보험회사의 시각에서 본 결과이다. 따라서 우리나라의 제도와 관습 하에서 소비자와 보험회사가 동시에 조사에 참여하는 경우 보험사기율은 상당히 다를 것이라고 전망한다. 국내 보험회사를 대상으로 한 설문 결과를 보면 보험사기 때문에 보험료의 3~6%가 누수된다고 하지만 과학적 근거는 적다. 향후 보험소비자와 독립 손해사정인과 보험회사 소속의 SIU를 포함한 FGI(focus group interview)나 심층 면접기법을 활용하여 보험사기율을 면밀히 과학적으로 조사할 필요가 있을 것이다.

링크분석이 보험사기 적발을 위한 절대적인 도구는 아니지만 기존의 보험사기 적발 업무를 획기적으로 개선할 수 있다. 링크분석을 활용하여 향후 보험사기에 보다 효과적으로 대응하기 위해서 다음의 3단계 대응전략에 대한 연구가 필요하다. 먼저, 회귀분석이나 인공지능망분석 등을 이용한 지수모형(scoring model)으로 보험

사기 가능성을 사전적으로 평가하고 걸러준다(screening process). 두번째, 링크분석을 이용한 보험금 지급 결과 보험사기 의심자를 추출한다(detection process). 마지막으로 SIU에서는 적발모형을 통해서 추출된 보험사기 의심자를 직접 조사하여 최종적으로 보험사기를 확인하고 사법처리를 하도록 한다(decision process).

보험사기 외에도 링크분석의 보험분야 적용 범위는 넓다. 관계에 의한 판매가 주를 이루는 생명보험 영업에서 링크분석의 활용가능성은 매우 높다. 예를 들면, 잠재고객의 계약체결가능성 분석이다. 다양한 잠재고객 중 어떤 링크를 가진 경우에 계약체결율이 높은 지를 파악할 수 있을 것이다. 그리고 고객의 추천에 의한 계약체결에서 어떤 링크를 가진 고객의 추천이 계약으로 연결될 가능성이 높은지를 분석하는 것도 흥미로울 것으로 생각된다.

참 고 문 헌

- 금융감독원, 보도자료, 2003. 4.
- 김광용, 「보험사기 적발을 위한 전문가시스템의 개발 : 퍼지이론과 AHP를 중심으로」, 『보험개발연구』 제 18호, 1996.
- 김용덕 · 안철경, 「보험사기조사의 효과성에 관한 실증연구: 미국의 주 보험사기국을 중심으로」, 『보험학회지』 61집, 2002.
- 김현수, 「보험사기 조기적발 모형에 관한 소고」, 『손해보험』, 1999.
- 김현수, 「보험전문가의 지식을 이용한 보험사기 조기경보모형의 개발에 관한 연구」, 『리스크관리연구』 11(1), 2000.
- 소비자보호원, 보도자료, 2003. 4. 10.
- 재정경제부, 『보험업법 개정안』, 보도자료, 2002. 7.
- 안철경 · 박일용, 『보험사기 적발 및 방지방안』, 보험개발원 보험연구소 1999.
- 안철경 · 조혜원 · 김경환, 『국내외 보험사기관리 실태 분석 : 선진사례 및 설문분석을 중심으로』, 보험개발원 보험연구소, 2002.
- 이윤호, 「보험사기에 대한 모니터링효과의 국제비교」, 『리스크관리연구』 11(1), 2000.
- 이윤호, 「보험사기에 대한 수사권 문제와 최적제재에 관한 연구」, 『보험학회지』 61집,

2002.

- 조수웅, 「보험범죄와 그 대책」, 『손해보험』, 1993. 3.
- 조해균, 「보험범죄의 발생원인과 그 대처방안에 관한 연구」, 『보험학회지』 35집, 1990.
- 조해균, 「도덕적위험과 보험범죄」, 『보험조사월보』 제141호, 1989. 11.
- 지홍민, 「보험사기와 최적보험계약」, 『리스크관리연구』 12(2), 2001.
- AAAI, *Symposium on Artificial Intelligent and Link Analysis*, 1998.
- Artis, Manuel, Mercedes Ayuso & Montserrat Guillen, “Detection and Automobile Insurance Fraud with Discrete Choice Models and Misclassified Claims”, *Journal of Risk and Insurance* 68. 2002.
- Artis, M., M. Ayuso, and M. Guillen, “Modelling Different Types of Automobile Insurance Fraud Behavior in the Spanish Market, Insurance”, *Mathematics and Economics*, 24(1-2), 1999, pp.67~81
- Becker, G., “Crime and Punishment”, *Journal of Political Econpmy*, 76(2), 1974.
- Belhadji E. B., and Dionne, G., & Tarkhani, F. “A Model for the Detection of Insurance Fraud”, *The Geneva Papers on Risk and Insurance* 24(4), 2000.
- Boyer, M. “Centralization Insurance Fraud Investigation”, *Geneva Papers on Risk and Insurance Theory* 25, 2001.
- Boyer, M. “Insurance Taxation and Insurance Fraud”, *Journal of Public Economics* 63, 1997.
- Brockett, P. L., R. A. Derrig, L. L. Golden, A. Levine, and M. Alpert, “Fraud Classification Using Principal Component Analysis of RIDITs”, *Journal of Risk and Insurance*, 69(3), 2002, pp.341~372.
- Brockett, P. L., Xia, S., and Derrig, R. A., “Using Kohonen’s Self-Organizing Feature Map to Uncover Automobile Bodily Injury Claims Fraud”, *Journal of Risk and Insurance*, 65, 2, 1998, pp.245~274.
- Browne, Mark J., & Well, Brenda P. “Claims Adjudication in the Personal Automobile Insurance Residual Market”, *Journal of Risk and Insurance*, 66(2), 1999, pp.275~290.
- Byrne, Ciara, “Intelligent Fraud Detection”, *Communication*. Vo. 5, No. 2 , 2002
- Caron, L., and G. Dionne, “Insurance Fraud Estimation: More Evidence From

- Quebec Automobile Insurance Industry, in: C. Laberge-Nadeau and G. Dionne, eds., *Automobile Insurance: Road Safety, New Drivers, Risks*, *Insurance Fraud and Regulation*(Boston, Mass.: Kluwer), 1999.
- Cummins, J. D., and S. Tennyson, "Controlling Automobile Insurance Costs", *Journal of Economic Perspectives*, 6, 1992, pp.95~115.
- Cummins, J. David & Tennyson, S. "Moral Hazard in Insurance Claiming: Evidence from Automobile Insurance", *Journal of Risk and Uncertainty*, 12, 1996, pp.29~50.
- Derrig, R. A., "Insurance Fraud", *Journal of Risk and Insurance*, 69, 2002, pp.271~287.
- Derrig, R. A., and L. Krauss, "First Step to Fight Workers Compensation Fraud", *Journal of Insurance Regulation*, 12, 1994, pp.390~415.
- Derrig, R. A., and K.M. Ostaszewski, "Fuzzy Techniques of Pattern Recognition in Risk and Claim Classification", *Journal of Risk and Insurance*, 62, 1995, pp.447~482.
- Derrig, R. A., H. I. Weisberg, and X. Chen, "Behavioral Factors and Lotteries Under No-Fault With a Monetary Threshold: A Study of Massachusetts Automobile Claims", *Journal of Risk and Insurance*, 61, 1994, pp.245~275.
- Derrig, R. A., and V. Zicko, "Prosecuting Insurance Fraud : A Case Study of the Massachusetts Experience in the 1900s, Working Paper", *Insurance Fraud Bureau of Massachusetts*, 2002.
- Dionne, G., "The Effect of Insurance on the Possibilities of Fraud", *Geneva Papers on Risk Insurance—Issues and Practice*, 9, 1984, pp.304~321.
- Dionne, G., P. St.-Michel, and Charles Vanasse, "Moral Hazard, Optimal Auditing and Workers' Compensation", Kingston, Canada: IRC Press, 1993.
- Major, J. A., and D. R. Riedinger, EFD: A Hybrid Knowledge/Statistical-Based System for the Detection of Fraud, *International Journal of Intelligent Systems*, 7:687-703(reprinted this volume), 1992.
- Mookherjee, D & Png, I. "Optimal Auditing, Insurance and Redistribution",

- Quarterly Journal of Economics* 63, 1989.
- Picard, P., "Auditing Claims in Insurance Market with Fraud: The Credibility Issue", *Journal of Public Economics* 63, 1996.
- Picard, P., *Economics Analysis of Insurance Fraud*, In: G. Dionne, ed., *Handbook of Insurance*, 2000, pp.315~363.
- Tamassia, Roberto, Ulrik Brandes, David Krackhardt and Dorothea Wagner., *Link Analysis and Visualization*, 2001.
- Tennyson, S., and P. "Salsas-Forn, Claims Auditing in Automobile Insurance : Fraud Detection and Deterrence Objectives", *Journal of Risk and Insurance*, 69(3), 2002, pp.289~308.
- Viaene, S., R. A. Derrig, B. Baesens, and G. Dedence, "Comparison of State-of-the-Art Classification Techniques for Expert Automobile insurance Fraud Detection", *Journal of Risk and Insurance*, 69(3), 2002, pp.373~421.
- Wasserman, S. and Faust, K., *Social Network Analysis*, Cambridge University Press, 1994.
- Watt, R., "Curtailling Ex-Post Fraud in Risk Sharing Arrangements, Universidad Autonoma de madrid", *Working Paper*, February, 2000.
- Weisberg, H. I. & Derrig, R. A. "Fraud and Automobile Insurance: A Report on the Baseline Study of Bodily Injury Claims in Massachusetts", *Journal of Insurance Regulation*, 9, 1991, pp.427~541.
- Weisberg, H. I. & Derrig, R. A. "Massachusetts Automobile Bodily Injury Tort Reform ", *Journal of Insurane Regulation*, 10, 1992, pp.384~440.
- Weisberg, H. I., and R. A. Derrig, *Identification and Investigation of Suspicious Claims*, in : *AIB Cost Containment/Fraud Filing*(DOI Docket R95-12) (Boston, mass.: Automobile Insurers Bureau of Massachusettes), 1995.
- Weisberg, H. I., and R. A. Derrig, "Quantitative Methods for Detecting Fraudulent Automobile Bodily Injury Claims", *Risques*, July-September: 35, 1998, pp.75~99.
- Westphal, C., and T. Blaxton, *Data Mining Solutions*, New York, John Wiley & Sons, Inc., 1998.

Abstracts

The objectives of this paper are, first to document recent insurance fraud trend and insurance fraud literature, second to introduce the concept and basic rationale of the Link Analysis originated from the Social Network Theory, and finally to evaluate the effectiveness of the analysis method by empirically analyzing real claim data. With strong cracking down efforts by insurers against insurance fraud, insurance consumers complaints appear to increase. In order to control this side effect of the anti-fraud efforts, fraud detection techniques have to be not only effective but also scientifically persuasive. Most of previous literature on insurance fraud detection like Derrig(2002), however, have not explore more scientifically persuasive and more practical 'link analysis' approach but mainly focus on difference kinds of scoring models. Comparing to scoring models using attribute data, the link analysis employing relation data could be more effective and practical in assisting SIU(special investigate unit). Based on the empirical analysis both Emergent Group Algorithm and Step Link Analysis have demonstrated its effectiveness to detecting hard insurance fraud.

※ Key Word : SIU, Emergent Group Algorithm, Step Link Analysis