

# 데이터마이닝 기법을 활용한 의료보험 진료비청구 삭감분석시스템 개발 및 구현에 관한 연구

## A Study on the Development and Implementation of a Data-mining Based Prototype for Hospital Bill Claim Reduction System

유 상 진 (Sangjin Yoo)

계명대학교 경영대학 경영정보학과

박 문 로 (Mun Ro Park)

계명대학교 동산의료원 전산실

### 요 약

경제의 세계화와 지식정보화 사회로의 진입과 함께 초래된 경영환경의 급속한 변화는 의료기관들에게도 경쟁력강화를 위한 변신을 강요하게 되었다. 다시 말하면, 의료기관들은 선진 의료기술의 확보, 환자들에 대한 서비스제고와 함께 경영의 효율성 증대라는 세가지 목표를 동시에 달성해야만 하는 상황에 놓이게 된 것이다.

본 연구는 의료기관들이 당면하고 있는 이러한 세가지 과제 중 병원의 경영효율성 증대를 위한 한가지 대안으로 진료비 청구삭감의 빈도 및 발생 가능성을 낮추기 위한 해법의 마련이 시도되었다. 진료비청구삭감이란 의료기관들이 환자들에 대한 의료서비스에 대한 진료비 중 의료보험으로 인해 환자들이 감면 받은 진료비를 건강보험심사원에 청구하면, 심사원이 의료기관의 청구내역의 적정여부를 심사하여 적정하지 않은 내용에 대한 청구금액을 삭감하는 제도를 이른다. 청구금액에 삭감이 발생하면 해당 의료기관의 수입이 감소하는 것은 물론 원인분석이나 재청구 작업등에 비용과 인력이 이중으로 투입되게 되어 의료기관의 경영에 부담을 주게 되고, 이러한 상황이 빈발하게 되면 해당 의료기관에 대한 환자과 건강보험심사평가원의 신뢰에 문제가 발생하게 된다. 그러므로, 효과적인 진료비 청구삭감분석시스템에 의한 사전대비의 필요성이 높아지게 되는 것이다.

이를 위하여 본 연구에서는 진료비 청구삭감분석을 위한 프로토타입의 개발이 시도되었다. 프로토타입은 데이터마이닝 기법 중 연관분석 알고리즘을 적용하여 개발 되었으며, 이렇게 개발된 프로토타입을 D의료원에서 10개월간 발생한 실제 진료데이터를 사용하여 성능을 시험하였다.

**키워드 :** 데이터마이닝, 의료보험, 진료비청구, 삭감분석시스템

## I. 서 론

1980년대 이후 경제의 세계화와 함께 기업의

경영환경은 이전에 경험하지 못했던 새로운 장을 맞이하게 되었다. 이러한 새로운 경영환경 속에서 기업들은 차별화와 경쟁력 강화를 최우

선 순위에 두게 되었고, 새로운 경쟁무기의 확보를 생존의 필요충분조건으로 인식하게 되었다. 이러한 기업들의 인식은 정보기술의 발달과 함께 다양한 경쟁무기의 출현을 낳게 되었다. 다시 말하면, 정보기술의 도입도 과거의 업무효율성의 증대라는 도입목적을 뛰어넘어 경쟁력 강화와 생존무기라는 새로운 차원의 의미를 갖게 된 것이다. 정보기술은 나름대로의 활용목적에 따라 각기 다른 가치를 갖게 되는데, 이러한 맥락에서 데이터베이스도 과거의 대용량 자료 저장 장소라는 기능에서 벗어나 경쟁무기로써의 역할을 부여받게 되었고 데이터웨어하우스(Data Warehouse)와 데이터마이닝(Data Mining)의 출현도 이러한 상황변화의 소산인 것이다.

의료기관들의 경우도 경영환경의 변화에서 예외는 아니어서 경쟁력 강화를 위하여 차별화된 선진 진료기술의 확보 이외에도 경영의 효율성 증대라는 과제를 맞이하게 된 것이다. 그리하여, 의료기관들도 과거에 구축된 대용량 데이터베이스를 실제 업무에서의 활용도를 높이는 한편 환자들에 대한 서비스를 제고함으로써 환자들의 만족도를 높이고, 이들 데이터들로부터 부가가치를 창출함으로써 궁극적으로는 의료기관의 경쟁력을 높일 수 있도록 데이터웨어하우스나 데이터마이닝에 많은 투자를 하고 있다. 특히, 국민들의 의료혜택을 높이기 위하여 도입된 의료보험제도는 의료기관들의 업무형태에 커다란 변화를 가져왔다. 즉, 의료보험제도의 도입과 함께 의료기관들은 전자청구시스템(EDI)을 통하여 건강보험심사평가원에 진료비를 청구하게 되었으며, 의료기관으로부터 진료비 청구가 접수되면 건강보험심사평가원은 이를 심사하여 진료비를 지급하게 된다. 진료비 심사과정에서는 다양한 사유로 인하여 의료기관이 청구한 진료비에 삭감이 발생하게 된다. 진료비가 삭감되면 의료기관의 수입이 감소하는 것은 물론 원인분석, 재청구 작업등에 비용과 인력이 이중으로 투입되게 되어 의료기관의 경영에 부

담을 주게 되고, 환자들이 부담한 본인부담금의 환불로 인한 환자들의 의료기관에 대한 불신초래, 그리고 삭감의 규모가 크거나 빈도가 높으면 의료기관이 실사대상기관으로 선정될 가능성이 높으므로 의료기관의 입장에서는 삭감의 원인을 분석하여 재발을 방지할 수 있는 삭감분석시스템의 도입이 바람직하다고 볼 수 있다.

본 연구에서는 이상과 같은 관점에서 의료기관들이 청구비 삭감의 발생가능성을 최소화 하고, 발생한 삭감에 대한 원인분석을 용이하게 할 수 있도록 데이터마이닝기법을 활용한 진료비 삭감분석시스템의 구축을 시도하였다. 이를 위하여 본 연구에서는, 첫째: 의료기관의 효과적인 정보 분석 업무를 위한 지표를 제시하기 위하여 데이터마이닝의 기법 중 일반적으로 적용되는 알고리즘을 분석, 분류하여 의료기관에 적합한 알고리즘을 제시하고, 둘째: 의약분업 및 보험재정악화로 인한 경영상의 어려움, 그리고 2005년도의 의료개방으로 인하여 더욱 악화될 가능성이 높은 경영상의 어려움에 대한 대비책의 일환으로써 본 연구에서 제시하고 있는 연관규칙 알고리즘을 바탕으로 한 삭감분석시스템을 구축함으로써 삭감의 정확한 원인분석을 통하여 삭감발생에 대한 대처방안을 제시함으로써 의료기관 수입의 90% 이상을 차지하고 있는 청구업무의 수익기여도를 높이고자 한다.

## II. 데이터웨어하우스 및 데이터 마이닝의 이론적 배경

### 2.1 데이터웨어하우스의 개념 및 정의

데이터웨어하우스란 기업 내 여러 곳에서 분산, 운영되는 트랜잭션 위주의 시스템들로부터 필요한 정보를 추출 및 통합 후, 중앙 집중화된 저장소에 모아놓고, 이를 여러 계층의 사용자들이 필요로 하는 정보를 시계열적으로, 분석적으로, 시각적으로 그리고 통계적으로 표현하는 것

을 지원하는 역할을 하는데, 일반적으로는 데이터의 저장고란 의미로 사용된다. 데이터웨어하우스란 용어는 Bill Inmon(1996)에 의하여 그 개념이 구체화 되었는데, 그는 “데이터웨어하우스는 조직의 의사결정을 지원하기 위해 축적되는 주제 중심적이고 통합적이며, 시계열 적이고 장기 존속하는 데이터의 집합”이라고 정의하였다. 이외에도 다양한 정의(Kelly Sean, 1994; Hammergren, 1996)가 존재하는데, 이들 정의에 공통적으로 담겨있는 용어가 ‘의사결정’임을 감안해 보면 데이터웨어하우스의 궁극적인 목표가 의사결정 지원임을 알 수 있다.

의료기관에 있어서 데이터웨어하우스는 중대형 병원에서 표출되는 다양한 의사결정 문제들에 대해서 의료 현상들과 관련 데이터 및 그 가공된 형태를 시계열적(Time Series)으로, 분석적(Analytic)으로, 시각적(Visual)으로 그리고 통계적(Statistical)으로 표현함으로써 얻어지는 환자 관련 지식이 병원의 모든 활동들을 계획, 집행, 통제할 수 있도록 지원하는 지식 인프라이다. 지금까지 개발 및 활용되고 있는 의료정보시스템은 병원 내 기본적 거래들의 입력, 저장, 출력 기능만을 가지는 거래처리시스템(Transaction Processing System)의 구현에 치중해 왔다. 이들 기존 시스템의 목표는 대기시간 단축, 관리 효율 향상, 원무 중심의 병원 수익 극대화에 초점을 맞추고 있으나, 다음과 같은 이유들로 인하여 데이터웨어하우스의 구축 필요성이 크게 증가하고 있다. 첫째, 점차 병원 업무가 다원화되고 복잡해지면서 통합적이고 전체적인 시각으로 진료활동 수행을 지원할 수 있는 정보 요구가 커지면서 병원 내 환자, 의사, 간호사, 원무 관리자, 최고 관리층 등의 효과적인 의사결정을 지원할 수 있는 정보기술의 획득과 관리, 활용이 절실히 요구되고 있다. 둘째, 의료기관의 정보화가 빠른 속도로 증가하면서 이들 의료정보시스템에서 발생 및 누적되는 각종 의료 데이터들을 효과적으로 이용하여 임상/연구/교육 등에 활용

하려는 노력이 활발해지고 있다. 셋째, 기존의 처방전달시스템(Order Communication System), 의료영상저장전송시스템(Picture Archive and Communication System), 전자 의무기록시스템(Electronic Medical Records) 등과의 보완적 결합을 통하여 이들 시스템에 누적된 데이터들을 효과적으로 활용하여 환자에 대한 분류, 군집화와 이들의 패턴(Pattern) 추출을 통해 확보되는 환자 관련 정보를 병원의 제반 업무활동에 적시에 적절한 형태로 공급함으로써 공급자 중심의 병원을 환자 중심의 병원으로 탈바꿈시키려는 노력이 증가하고 있다. 이러한 이유들로 인해서 도입의 필요성이 증대하고 있는 의료기관의 데이터웨어하우스는 업무의 특성상 다른 산업에서 구축되는 데이터웨어하우스와는 다른 특징을 나타내고 있는데, <표 1>은 타 산업에서 활용중인 전형적인 데이터웨어하우스와 의료 데이터웨어하우스의 주요 특징을 비교하여 보여 주고 있다.

<표 1> 기존 데이터웨어하우스 vs. 의료 데이터웨어하우스

기 준	전통 데이터 웨어하우스	의료 데이터 웨어하우스
통합성	높음	높음
상세 데이터 활용도	보통	높음
휘발성	낮음	없음
과거 데이터 활용도	가끔	항상
데이터 모델 복잡성	낮음	높음
시간차원의 중요성	보통	높음
분류의 난이도	낮음	높음
연속성 데이터 활용	없음	있음
차원요약	없음	있음
고도로 복잡한 데이터 여부	없음	있음
고도의 비즈니스 룰	없음	있음
데이터마이닝	낮음	높음

자료출처: 지식창출 병원으로의 혁신을 위한 의료데이터웨어하우스, 병원신문, 2001년 8월 2일, 11면.

한편, 이상과 같은 특징을 가지고 있는 의료 데이터웨어하우스의 기대효과는 다음과 같이 6가지로 요약될 수 있다(병원신문, 2001).

첫째, 의료 데이터웨어하우스는 기존의 공급자 중심의 병원이 환자 중심의 병원으로 전이되는 것을 촉진하며, 적시에 적절한 형태로 공급되는 환자 관련 정보는 병원의 의사결정 구조를 바꿀 수 있다. 예를 들어, 상병을 여러 개 가지거나 치료전력이 많은 환자를 효과적으로 치료하기 위해서 병원내 협진 체제 구축은 필수적인데, 데이터웨어하우스를 이용한 그룹별, 팀별, 클리닉 별 편익/비용 분석은 협진 체제 구축을 촉진하여 이러한 변화를 가능하게 한다.

둘째, 환자, 의사, 간호사, 투자자, 회계감사자 등 병원과 관련된 이해 관계자들 사이의 의사소통을 원활하게 하며, 이들의 효과적인 의사결정을 가능하게 한다. 이는 궁극적으로 병원의 투명한 경영을 가능하게 하는 정보인프라의 구축을 가능하게 하여, 여기에서 산출되는 정보를 바탕으로 병원전략 수립, 마케팅 계획, 재무계획, 인사 계획 등 병원 전반의 경영과정을 혁신할 수 있다.

셋째, 진료의 품질을 고양시킨다. 예를 들어, MRI, CT, 초음파 등으로부터 만들어지는 이미지 데이터, 원무 관리과의 환자 신상데이터, 의사의 처방데이터, 약국의 투약 데이터 등 기존에는 산재 되고 연동되지 않아 전혀 관계를 갖지 못했던 데이터들을 통합할 수 있도록 함으로써 환자 자신은 물론 환자의 가족병력까지 파악이 가능하게 되어 환자관련 데이터의 다각적인 분석을 통하여 진료의 정확성을 높일 수 있다.

넷째, 기존의 처방전달시스템, 의료영상저장 전송시스템, 전자의무기록시스템을 통해 쌓여진 데이터들은 데이터웨어하우스를 통해 정제되고 변환되어 부가가치 높은 정보들로 가공됨에 따라, 각 부서의 데이터입력 업무가 대폭 감소하게 되어 업무의 생산성을 높이고 인력수요를 감소시키며 기존 Legacy 시스템의 작업량을 상당

부분 줄일 수 있다.

다섯째, 웹 기반 데이터웨어하우스가 여러 병원에서 구축되면 병원과 병원간 온라인 또는 오프라인 형태의 B to B 네트워크 또는 병원간정보시스템(IOS, Interorganizational Information System)의 구축을 통하여, 진료관련 경험 및 정보를 서로 교환할 수 있게 되어 효과적인 병원간 협력 및 협업체계가 가능해진다. 이렇게 구축된 웹기반 데이터웨어하우스는 환자관계관리(PRM, Patient Relationship Management)시스템의 구축에 핵심적인 기반이 된다.

여섯째, 임상, 연구, 교육의 유기적 순환 관계를 정립할 수 있다. 다시 말하면, 임상과정을 통해 획득된 환자에 대한 각종 데이터는 연구과정을 통해 지식화 되고 검증된 후 교육에 활용된다. 그리고, 의료인은 교육을 통해 획득된 지식을 바탕으로 다시 임상에서 활용함으로써 지식의 창출, 공유, 확산, 재창출의 Spiral Process를 구성하게 되어 의료지식이 지속적으로 확대 재생산되는 토대가 마련된다.

## 2.2 데이터마이닝의 개념 및 정의

1980년대에 접어들면서 경영환경이 급변하고 불확실성이 높아지면서 기업들은 이러한 환경변화에 대응할 수 있는 대안이 필요하게 되었고, 기업들은 신속정확한 의사결정이 불확실성을 효과적으로 낮출 수 있는 방안임을 깨닫게 되었다. 상당수의 기업들이 방대한 양의 데이터를 보유하고 있었으나 신속정확한 의사결정은 데이터의 양보다는 특정 문제나 상황에 맞는 데이터를 얼마나 신속하게 효과적으로 의사결정자에게 지원해 줄 수 있는가에 달려있다. 그러나, 당시의 정보통신 기술수준이나 데이터베이스의 성능만으로는 이러한 기업의 욕구를 효과적으로 만족시켜주기에는 한계가 있어 기업들은 새로운 도구에 대한 필요성을 절실하게 느껴왔다.

데이터마이닝은 이러한 필요에 의해 등장하

게 되었다. 데이터마이닝에 대한 정의는 다양하게 내려지고 있는데 대표적인 것을 살펴보면 다음과 같다.

“데이터마이닝이란 방대한 양의 정보로부터 쉽게 드러나지 않는 유용한 정보의 패턴을 추출하는 과정이라고 말할 수 있다. 기업이 보유하고 있는 거래자료, 고객자료, 상품자료와 기타 외부자료를 포함하는 사용 가능한 데이터를 기반으로 감춰진 지식, 기대하지 못했던 패턴, 새로운 법칙 등을 발견하고 이를 실제 경영의사결정 등을 위한 정보로 활용하고자 하는 것이 바로 데이터마이닝이다.”(데이터베이스월드, 1998)

“데이터마이닝이란 자동화되고 지능을 갖춘 데이터베이스 분석기법으로 90년대 초반부터 지식발견(KDD: Knowledge Discovery in Databases), 정보발견(Information Discovery), 정보수확(Information Harvesting) 등의 이름으로도 소개되어 왔는데 일반적으로 ‘대량의 데이터로부터 새롭고 의미 있는 정보를 추출하여 의사결정에 활용하는 작업이다’”(Peter Adrians and Dolf Zantinge, 1996).

한편, 가트너그룹은 “대량의 데이터로부터 패턴인식 기술과 통계기법, 수학적 기법을 이용하여 의미 있는 새로운 상관관계, 패턴 그리고 추세를 발견하는 과정”이라고 정의했으며, 매타그룹은 데이터마이닝을 “대용량 데이터베이스에서 알려져 있지 않은 실행 가능한 정보를 추출하는 지식발견의 과정”이라 정의했다.

이상을 종합해 보면 데이터마이닝은 “대량의 데이터로부터 패턴인식, 통계적 기법, 인공지능 기법 등을 이용하여 숨겨져 있는 데이터 간의 상호 관련성, 패턴, 경향 등 유용한 정보를 추출하여 실제의 의사결정에 적용하는 전 과정”이라고 정의할 수 있다.

## 2.3 데이터마이닝 기법

데이터마이닝에는 다양한 기법들이 활용되고

있으나, 본 논문에서는 연관규칙(Association Rule), 클러스터링(Clustering), 신경망(Neural Network), 의사결정나무(Decision Tree), 유전자 알고리즘(Genetic Algorithm), 사례-기반 추론(Case-based Reasoning) 등 6가지 종류의 기법들을 간단히 살펴보기로 한다.

### 2.3.1 연관규칙(Association Rule)

데이터 안에 존재하는 항목간의 연관관계를 발견하고자 할 때 사용되는 데이터마이닝 기법이다. 연관규칙이란 상품을 구매하거나 서비스를 받는 등의 일련의 거래나 사건들의 연관성에 대한 규칙으로서, 마케팅에서는 손님의 장바구니에 들어 있는 품목간의 관계를 알아본다는 의미에서 장바구니분석(Market Basket Analysis)이라 부르기도 한다(Jiawei Han and Micheline Kamber, 2001).

장바구니 분석이라 부르는 이유는 다음과 같다. 슈퍼마켓에서 쇼핑을 마친 고객의 장바구니를 살펴보면 고객이 무엇을 구매했는지 알 수 있다. 특정 고객에 대한 장바구니만을 살펴본다면 그 고객의 쇼핑 취향에 대한 대략적인 정보만을 알 수 있으며, 그 결과를 일반화하기 어려우나 모든 고객의 장바구니 안에 들어 있는 물건을 전부 알 수 있다면 훨씬 더 구체적이고 다양한 정보를 얻을 수 있을 것이다. 왜냐하면, 모든 고객들의 성향이 모두 동일한 것은 아니며, 고객에 따라 구매 상품의 종류를 비롯하여 구매한 양과 구매시점 또한 다를 것이기 때문이다. 연관규칙이란 특정한 상품을 구매한 고객이 어떤 부류에 속하는지, 그들이 왜 그러한 상품을 구매했는지, 어떤 상품들이 함께 팔리는 경향이 있는지를 알아보기 위하여 고객들이 구매한 상품에 대한 자료를 분석하는 것이다. 물론, 이러한 분석결과는 매장 진열 및 배치의 변경, 어떤 상품들을 패키지로 묶을 것인지, 어떤 상품들을 교차 판매할 지, 그리고 어느 상품을 기획 상품으로 삼아야 할지를 결정하는데 중요한 역할을

하게 된다.

### 2.3.2 클러스터링(Clustering)

클러스터링(Clustering)이란 주어진 데이터 집합을 서로 유사성을 가지는 몇 개의 클러스터로 분할해 가는 과정으로, 하나의 클러스터에 속하는 데이터 점들 간에는 서로 다른 클러스터 내의 점들과는 구분되는 유사성이 존재하게 된다는 점에서 기인한다(Michael J. A. Berry and Gordon Linoff, 1997). 클러스터링은 대용량 데이터를 분석할 때 효과적인데, 대용량 데이터에서는 개개의 데이터를 요약하는 것보다는 전체를 유사한 데이터들을 클러스터로 묶어 복잡한 전체보다는 그를 잘 대표하는 클러스터들을 관찰함으로써 전체 데이터에 대한 의미있는 정보를 얻어낼 수 있을 것이라는 믿음에 근거하고 있다(강현철, 한상태, 최중후, 김은석, 김미경, 2001). 여러 클러스터링(Clustering)방법 중에 대용량 데이터를 빠르게 처리할 수 있으며, 그 알고리즘도 비교적 간단한 K-평균 군집(K-means Clustering)방법을 가장 많이 사용한다.

### 2.3.3 신경망(Neural Network)

신경세포는 크게 정보처리의 핵심이 되는 신경세포와 기억의 핵심이 되는 신경세포를 연결해 주는 매우 좁은 간극, 즉 시냅스(Synapse)의 두 부분으로 이루어져 있으며, 신경세포는 다시 신경세포체(Cell body), 입력선인 수상돌기(Dendrites), 출력선인 축색돌기(Axon)의 세 부분으로 나누어진다. 한편, 인공 신경망은 생물학적 뉴런을 모델링한 유닛(Unit)들과 그 유닛 사이의 가중치의 연결들로 이루어지며 각 신경망 모델에 따라 다양한 구조와 각기 독특한 학습 규칙을 갖는다. 일반적으로 말하는 인공 신경망은 컴퓨터가 사람의 학습 기능을 갖게 하기 위해 고안되었다.

인공 신경망은 계층별로 그룹화된 뉴런들의 집합으로 구성되어 있는데, 이들은 크게 입력

(Input Layer), 중간(Hidden Layer), 출력(Output Layer)의 세 계층으로 이루어져 있으며, 입력과 출력계층 사이에는 여러 개의 계층이 존재하도록 할 수도 있다(<http://chumsungdae.com/korea/research/research.html>). 인공 신경망 기술의 가치는 불완전하고 잡음이 많은 입력의 해석뿐만 아니라 패턴인식(Pattern Recognition), 학습, 분류, 일반화, 추상화 등에 활용할 수 있으며 의사결정지원시스템이나 전문가시스템의 논리적인 기법을 활용해서도 시뮬레이션하기 어려운 비구조적인 문제해결에도 효과적으로 활용될 수 있다는 점이다.

### 2.3.4 의사결정나무(Decision Tree)

의사결정나무(Decision Tree)는 의사결정규칙(Decision Rule)을 나무구조로 도표화하여 분류(Classification)와 예측(Prediction)을 수행하는 분석방법이다. 의사결정나무가 어떻게 새로운 레코드의 해당 부류 값을 예측하는가는 어렵게 느껴질 때 자주 하던 ‘스무고개’ 놀이를 떠올려 보면 이해하기 쉽다. 이 놀이에서는 한 사람이 특정한 사물이나 사건, 또는 인물이나 장소와 같은 개체 한 가지를 마음속으로 생각한다. 그러면 나머지 사람들은 “그것이 사람입니까?”와 같은 질문을 반복하여 정답을 맞힐 때까지 최고 20회의 질문 기회를 갖는다. 그러나 놀이에 익숙한 사람들은 20회의 질문 기회를 모두 사용하기 보다는 정답을 유도할 수 있는 예리한 질문을 몇 번 던지고는 정답을 맞히는 경우가 많다. 이 놀이와 마찬가지로 의사결정나무도 새로운 레코드의 부류 값을 예측하기 위해 이미 만들어진 의사결정나무가 지시하는 바에 따라 레코드의 속성 값을 질문하는 작업을 반복적으로 수행한다. 특히 결정적인 질문을 던지게 되면 다른 모든 속성의 값을 묻지 않고도 레코드의 부류 값을 정확히 예측할 수 있다. 따라서 레코드를 분류하고 예측할 수 있는 나무를 얼마나 잘 만드느냐가 의사결정나무 기법의 핵심이다(이경영, 2002).

### 2.3.5 유전자 알고리즘(Genetic Algorithm)

유전자 알고리즘(Genetic Algorithm)은 신경망과 같이 자연적 현상을 모델로 하여 해(Solution)를 탐색하는 알고리즘이다. 즉, 신경망이 인간의 두뇌를 모델링한 것이라면, 유전자알고리즘은 자연의 진화과정을 모델링한 것이다(도용태, 김일곤, 김종환, 박창현, 2001). Holland에 의해 기본적인 개념이 확립된 유전자알고리즘의 기본적인 개념은 주어진 문제에 대해 가능한 해들을 정해진 형태로 표현한 후 이들을 변화시키고 그 결과들 중 적합도가 높은 것들을 선택하여 변화를 계속함으로써 보다 좋은 해에 도달하는 것이다(J. H. Holland, 1975). 이는 마치 부모의 유전인자 중 우성의 형질을 지닌 개체가 계속 번성해간다는 다윈의 진화론적 적자생존의 법칙(Survival of the Fittest)과 유사하다.

### 2.3.6 사례-기반 추론(Case-based Reasoning)

사람들은 새로운 문제가 발생하면 이를 해결하는 방안의 하나로 이전에 경험했던 유사한 상황을 떠올린다. 사례기반추론의 기본적인 논거는 사람은 특정 문제를 해결 하는 과정 중에 과거에 발생했던 유사한 문제에 적용했던 해결 방식을 유추하여 현재의 문제에 대한 해결책을 모색하는 경우가 많다는 것이다(Kolodner, 1993). 의사나 변호사 같은 전문가들도 복잡한 문제를 해결하기 위해 유사한 문제를 해결했던 경험에 의한 추론을 한다. 예를 들면, 의사가 새로운 환자의 질병을 치료하기 위해 이전의 유사한 증상과 처방을 떠올려 더욱 빠르고 안전하게 그 병에 대한 처방을 내리는 경우가 이에 해당한다. 물론, 이런 처방이 항상 타당하다고는 볼 수 없으나 새롭게 발생한 문제에 대하여 쉽고 빠르게 적용해 볼 수 있는 실마리를 제공할 수는 있다. 이러한 문제해결 방식을 사례기반 추론이라 한다.

사례기반 추론을 이용해 문제의 해결하는 경

우, 새로운 문제가 발생하면 우선 발생한 문제와 가장 유사한 사례를 찾아내어 그 유사한 사례에 대한 해를 발생한 문제에 대한 해결책으로 대입해 본다. 그러나 과거의 유사한 사례와 새로 발생한 문제 사이에는 차이가 있을 수 있기 때문에 유사한 과거 사례의 해를 고집하지 말고, 발생한 문제와 과거의 유사한 사례 사이에 존재하는 차이를 고려하여 과거 유사한 사례에 대한 해를 수정내지는 보완하여 새로 발생한 사례의 해로 삼는다(Watson, I., 1997).

## Ⅲ. 청구삭감분석업무의 이론적 배경

### 3.1 진료비 심사제도

진료비 심사제도는 의료보험 제도 하에서 의료기관이 보험자단체에 청구하는 진료비가 적정한가를 심사·확인하여 불필요하거나 지나친 보험급여비를 억제하여 보험재정을 안정시키며, 진료질서를 확립하고 적정진료를 유지하여 의료보험제도를 성공적으로 발전시키는데 중요한 역할을 하는 제도이다(김수배, 1998). 우리나라의 의료보험 진료비 심사제도는 1979년부터 시작되었으며(박내희, 1995), 진료비 심사의 법적 근거는 국민의료보험법 제38조(요양의 비용 등) 제1항에 두고 있다. 구체적인 내용을 살펴보면, 보험자는 요양기관으로부터 요양급여나 분만급여에 관한 비용의 청구가 있을 때 보건복지부령이 정하는 바에 의하여 이를 심사·지급하되, 심사 결과 이미 납부한 본인 일부부담금이 과도한 경우 요양기관에 지급할 금액에서 과다하게 납부된 금액을 공제하여 이를 피보험자에게 반환하여야 한다고 명시되어 있다. 건강심사평가원의 정관에는 진료비 심사의 목표가 진료비의 적정성 및 타당성을 판정하는데 있다고 규정하고 있다(건강심사평가원, 1999).

결국 진료비 심사의 중요한 목표는 의료보험 제도를 운영하는데 있어서 의료기관의 부당한

진료비 청구가 있을 경우에는 이를 조정하여 진료질서의 확립을 유도하고 의료보험재정을 보호하는 한편, 각 의료기관의 진료서비스가 적정하게 이루어지도록 하는데 있다. 즉, 심사를 통하여 적정진료의 유도와 보험재정의 안정을 도모하고자 하는데 의의가 있다(유승훈, 1990). 그러나 현 의료보험제도하에서 진료비심사제도는 과잉진료, 부당청구, 산정착오 등의 명목으로 요양기관에서 청구된 진료비를 삭감하고 있으며, 이러한 진료비 삭감은 보험자에 의해 거의 일방적으로 이루어지고 있다. 그리고 2001년 7월의 약분업 실시로 개국 이래 초유의 의료대란으로 국가의 보험재정은 더욱 악화되고 의료기관의 진료비 과잉 삭감으로 병원업계는 존폐의 위기를 맞이하고 있다. 따라서 진료비 심사는 진료의 사회적 책임과 적정진료의 자율성을 유도·정립하고 이를 실체화하는 사후 확인행위로서 의료보험의 사회적 공동부담 비용에 대한 연대적 책임을 상호 인식케 함으로써 불필요한 서비스를 억제하여 보험재정, 나아가서는 국민전체의 의료비를 안정화시키는 동시에 의료서비스의 질이 저해 받지 않도록 하여 가장 효율적으로 의료자원을 배분하는데 그 의의가 있다고 하겠다(김정수, 1990).

### 3.2 진료비 심사업무

진료비 심사는 진료수가 기준, 약가기준, 요양급여 기준, 심사기준, 보건복지부의 행정지시 등에 의하여 이루어진다. 간호사, 약사, 의료기사로 구성된 심사요원들은 일차적으로 환자의 인적사항 등 일반 기재 사항을 점검하고 적용착오, 산정착오, 기재누락 등을 대조·검토한다. 또한 심사요원들은 요양급여 기준이나 진료수가 기준, 약가기준, 복지부 행정지시 등의 심사기준에 위배되었는지 여부를 확인하고 의문이 있는 사항에 대하여는 심사요원 소견서를 첨부하여 상급 심사위원회에 이를 심사·의뢰한다.

이 첫 번째 단계에서 모든 심사 물량의 95%가 처리된다. 두 번째 단계의 심사는 의사, 치과의사, 약사들로 구성된 상급 심사위원이 심사요원들로부터 심사·의뢰된 건 및 고액, 특수 장기진료의 타당성 여부를 판단 심사하며, 상급 심사위원의 재량으로 처리하지 못하는 고도의 전문적 판단을 요하는 사항에 대하여는 진료과목별로 구성되어 있는 분과위원회에 회부하여 심의 처리한다(김석일, 1994).

### 3.3 진료비 삭감분석

#### 3.3.1 진료비 삭감분석의 목적

진료비 삭감 분석이란 보험자 단체인 건강보험심사평가원에서 심사로 인해 조정된 진료비에 대하여 의료기관에서 조정원인과 유형을 분석하는 것을 의미한다. 즉 건강보험심사평가원에서 심사 후 통보되는 “진료비(약제비) 심사결과 통보서”상 정확한 심사조정 항목과 사유가 통보되므로, 심사결과통보서에 대한 항목에 따라 진료행위별, 조정사유별로 내용을 확인하고 전체적인 조정(삭감)내역을 분석하는 과정으로 이의신청의 기초 자료로 활용되며 병원자체 진료상의 문제점과 보험자단체의 심사경향 등을 파악하여 적정진료를 유도하는데 그 목적이 있다(김수배, 1998).

조정내역분석에 따라 도출된 문제점은 즉시 해당 진료과에 통보하여 진료에 반영토록 하며 유사한 문제로 인한 삭감재발을 방지하고, 진료지침서 등을 작성하여 진료에 활용할 수 있어야 한다. 심사결과 통보서의 내용 중 납득하기 곤란한 심사조정은 관련자료 보완 및 의사소견서 등을 첨부하여 이의신청을 하도록 한다. 그러나, 심사조정내역 분석 후 진료에 반영하는 것 보다 진료기준에 맞게 진료를 할 수 있도록 의사, 간호사 및 사무직 근무자들이 의료보험에 대한 취지와 급여기준을 사전에 이해하고 적용하는 것이 바람직하다.

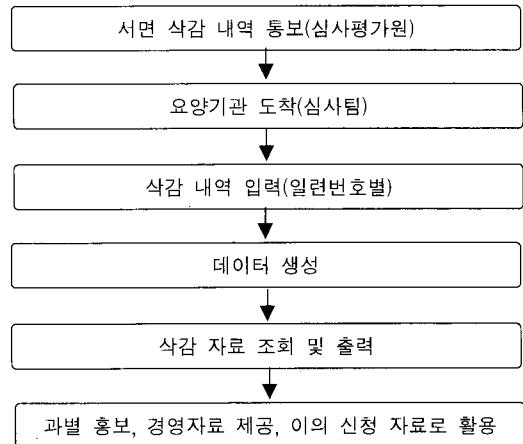


### 3.3.2 심사조정 분석 방법

심사조정분석은 심사결과 통보서와 의료기관에서 청구한 데이터, 그리고 원무발생 데이터를 자체 전산을 이용하여 비교분석하는 방법이다. 다시 말하면, 심사 지불기관인 건강보험심사평가원에서 EDI 청구 데이터를 심사하고 그 결과를 Sam File 형식의 진료비 심사결과통보서를 의료기관에 송부하면, 의료기관은 이를 근거로 심사조정내역을 분석하게 된다. 이 과정에서 진료비 심사결과 통보서상에 데이터가 없는 것은 청구 진료비가 전액 인정된 경우이며, 조정사유코드 또는 지급불능사유코드가 발생되어 있으면 심사조정이 발생한 자료이다. 청구된 진료비가 심사과정에서 조정된 경우에는 청구한 금액에서 조정된 금액만을 차감하고 진료비가 입금된다. 따라서 항목별로 자세한 심사조정 내역을 알아보려면 진료나 청구상의 문제점을 발췌하여 진료비 심사조정 내역을 분석하면 된다.

진료비 심사지급기관에서는 사회보험의 특성을 살린 경제적이며 보편타당한 진료를 요구하는 입장에서 심사를 하는 바 심사 후 통보되는 삭감의 원인과 유형을 분석, 파악하여 진료에 이를 적극 반영토록 하며, 의료기관 자체적으로는 의료의 질관리 및 경영분석의 기초 자료로 활용한다(안옥희, 1998). 진료비삭감 사유별 유형은 대체적으로 진료적 착오(요양급여기준 적용 착오 즉, 과잉진료, 비급여 등)와 행정적 착오(수가적용 착오, 재료대 과다청구, 계산착오, 증빙자료 미제출, 산정 착오)로 분류할 수 있다.

분석의 단순화, 표준화를 추구하였지만 EDI 청구가 아닌 서면 청구였기 때문에 삭감분석시스템을 구현하는데 많은 제약사항이 있었다. 건강보험심사평가원에서 삭감내역을 진료과별·청구서일련번호별·진료비항목별로 서면으로 요양기관에 청구하면 요양기관에서는 청구내역을 참조하여 진료과별·진료비 항목별로 삭감내역을 분석한 후 청구서의 일련번호별로 데이터를 입력하여 자료를 생성 시켰다. 이렇게 입력된 자료를 기초로 하여 진료과별·항목별·환자별로 삭감내역을 생성하고 그 자료를 출력하여 경영분석의 기초 자료로 활용한다. 기존 진료비 청구 삭감 분석시스템은 크게 데이터 추출, 정제, 통합의 세 가지 모듈로 나눌 수 있는데, <그림 1>은 이러한 기존 진료비 청구 삭감 분석시스템의 전체적인 구조를 보여주고 있다.



<그림 1> 기존 진료비 청구 삭감 분석시스템 업무 흐름도

## IV. D의료원의 기존 진료비청구 삭감 분석시스템

### 4.1 기존 진료비청구 삭감분석시스템

기존의 삭감분석방법은 사용 중인 서식들을 분석하고 삭감담당자들의 요구사항을 토대로 전산화에 적합하도록 업무 방법의 개선과 삭감

#### 4.1.1 데이터 추출 과정

진료비 청구 삭감 분석시스템에서의 데이터의 추출과정은 원무 발생 데이터와 청구 데이터를 통합시켜 분석에 용이한 데이터로 만들기 위한 과정으로써 가장 첫 단계이자 많은 시간을 요하는 단계이다. 수집된 데이터는 데이터의 크

기, 데이터 입력시의 오류 등의 문제로 인하여 청구삭감 분석업무에 곧바로 사용될 수 없으므로 불필요한 데이터 및 필드는 소거시킴으로써 데이터의 크기를 줄이고 데이터 수집 과정에서 발생한 입력 오류는 제거하거나 적절한 값으로 대체하는 작업을 통하여 청구삭감 분석업무에 적합한 형태로 바꿔주어야 한다.

#### 4.1.2 데이터 정제 과정

청구 심사결과는 건강보험심사평가원에서 서면으로 데이터가 제공되므로 직접 데이터베이스에 저장하는 것이 불가능하며 삭감내역을 수작업으로 입력해야 한다. 그러나, 삭감 내역을 수작업 입력하는데 많은 시간이 낭비되고 삭감담당자의 업무 부담이 늘어나는 문제점이 발생되고 있다.

#### 4.1.3 데이터 통합 과정

데이터 추출 및 정제 작업이 완료되면 삭감담당자가 요구하는 리스트에 맞게 데이터를 통합시키는 작업이 이루어져야 한다. 이렇게 통합된

자료는 진료과별, 의사별, 거래처별로 자료를 출력하여 진료과 홍보 및 경영분석의 기초 자료로 활용되는데, <그림 2>는 기존 진료비 청구삭감 분석시스템의 데이터베이스의 구조를 보여주고 있다.

### 4.2 기존 진료비청구 삭감분석시스템의 문제점

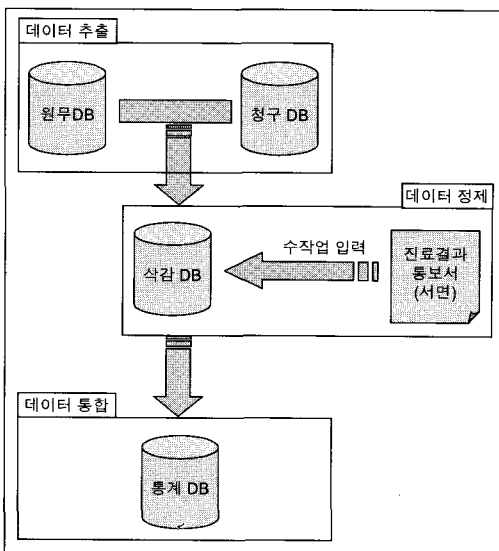
기존 진료비 청구 삭감 분석시스템은 비교적 단순한 통계적인 보고서만 지원하는 수준이며, 삭감 분석 자료의 업무 활용도가 매우 낮은 상태이다. 다시 말하면, 삭감된 약품이나 재료에 대한 단순한 삭감 통계치만 조회될 뿐, 진료과와 상병 그리고 의사, 삭감사유, 진료수가 등 다양하게 비교 분석하여 삭감이 되는 근본적인 문제를 찾아낼 수 없다. 예를 들어 어떠한 의사는 어떠한 상병에 대한 처방을 내릴 경우, 항상 “A”라는 약제가 삭감된다는 사실은 단순한 통계치로는 분석이 불가능하다. 기존 삭감 분석시스템의 장단점은 다음과 같이 요약될 수 있으며, <표 2>는 영역별 문제점을 요약하여 보여주고 있다.

#### (1) 장 점

- 삭감담당자가 프로그램 사용자 교육을 받지 않아도 쉽게 운영할 수 있다.

#### (2) 단 점

- 삭감금액을 전부 수작업으로 입력해야 함으로 소요시간이 많고 잘못 입력될 수 있다. 잘못 입력이 된 경우 금액이 맞지 않아 삭감담당자가 잘못 입력된 금액을 일일이 찾아서 수정작업을 해야 한다.
- 청구 심사결과 통보서가 건강보험심사평가원에서 서면으로 데이터가 제공되기 때문에 1차로 청구 심사결과 통보서를 분석한 다음 삭감금액을 입력해야 한다. 그러므로



<그림 2> 기존 진료비 청구 삭감 분석시스템 데이터베이스 구조도

삭감담당자의 주관이 많이 개입된다. 또 청구 심사결과 통보서에서 분석이 안 된 내용은 처리가 불가능하여 수작업 입력에서 누락되며, 이런 경우 삭감금액이 큰 경우 분석이 불가능하다.

- 삭감분석이 완료되었다 할지라도 삭감 입력금액과 삭감금액이 일치하지 않을 경우 하나하나 체크하면서 잘못된 부분을 찾아야 한다.
- 삭감 자료입력과 삭감분석이 덜된 경우 어디까지 삭감분석이 되었는지 삭감담당자가 관리해야 한다.

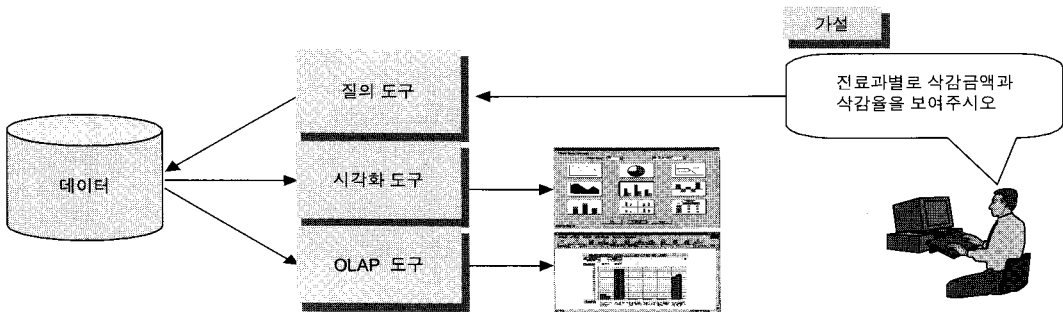
### 4.3 문제점의 해결방안으로써의 데이터 마이닝

데이터의 분석과 활용을 위해 사용되어진 질의나 OLAP 도구, 그리고 이들을 기반으로 만들

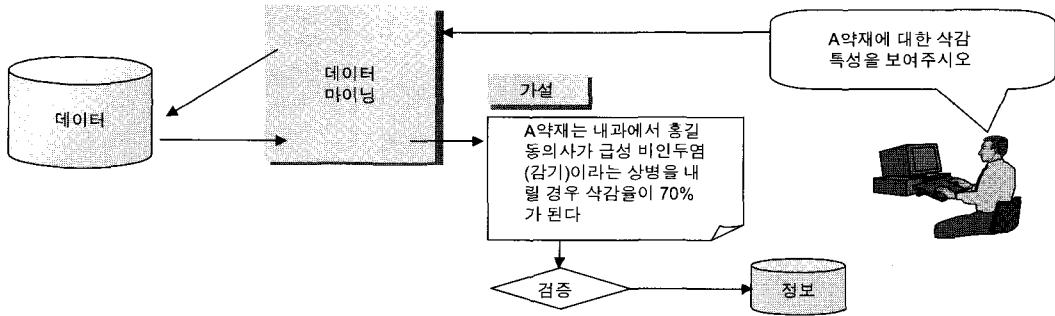
어진 의사결정시스템(Decision Support Systems)이나 중역정보시스템(Executive Information Systems) 등은 <그림 3>에서 보여주듯이 원칙적으로 미리 가설을 세우고 데이터를 통해 그 가설을 확인하는 방식이다. 즉 사용자가 자신의 경험에 비추어 가설을 세우고, 여기에 맞추어 질의를 만들어 실행하며, 결과를 검토하여 가설이 맞는지 틀리는지를 확인하는 것이다. 만약 가설이 틀리다는 결과가 나올 때까지, 혹은 사용자가 가설이 틀렸다고 생각할 때까지 질의를 반복한다. 그러나 이러한 방식은 적절한 가설을 세울 수 없는 상황에서는 적용하기 곤란하며, 가설이 수용되거나 거부되는 것 외에는 새로운 정보가 거의 발견되지 않는다. 그러나, 사용자가 미리 가설을 세우지 않아도 컴퓨터시스템이 가능한 가설을 스스로 생성하고 이를 검증하는 <그림 4>와 같은 데이터마이닝 방식은 사용자의 가이드를 거의 받지 않고 필요한 정보를 찾는다.

<표 2> 기존시스템의 영역별 문제점

영역	기존의 문제점
진료	◦ 리포트를 각 진료과에 전달하기 때문에 각 진료과에서는 분석하기 어렵다.
삭감분석	◦ 보고서 및 통계 결과를 산출하는데 최소 2~3개월이 소요된다. ◦ 수작업 입력 시 오류 발생 빈도가 높다. ◦ 청구 심사결과 통보서가 건강보험심사평가원에서 서면으로 데이터가 제공되기 때문에 1차로 청구 심사결과 통보서를 분석하는데 어렵다.
전산실	◦ 분석시 전산실에서 운영계 데이터를 추출/가공하여 삭감담당자에게 전달하기 때문에 전산실 직원의 업무가 과다하다.



<그림 3> 가설 확인 중심의 기존 조회 방식



〈그림 4〉 가설 발견 중심의 데이터마이닝

예를 들면, 삭감분석을 하기 위해 진료과, 의사, 상병, 삭감사유, 진료수가 등과 같은 데이터들을 수집하고, 이들을 분석하여 “A약제는 내과에서 홍길동의사가 급성 비인두염(감기)이라는 상병을 내릴 경우 삭감율이 70%가 된다.”라고 얻은 결론이 정보다. 그리고 이러한 정보와 과거의 경험을 근거로 “삭감될 확률이 60% 이상이면 A약제 대신 다른 약제로 처방하라!”라는 지식을 도출할 수 있다. 정보는 그 자체로서도 기업 경쟁력의 원천이 되지만, 지식으로 발전하게 되면 경쟁력은 더욱 강화된다. 이러한 분석 작업을 지원하는 정보기술이 바로 데이터마이닝이다.

## V. 데이터마이닝을 활용한 진료비청구 삭감분석시스템의 프로토타입 (Prototype)

### 5.1 진료비청구 삭감분석시스템의 구성

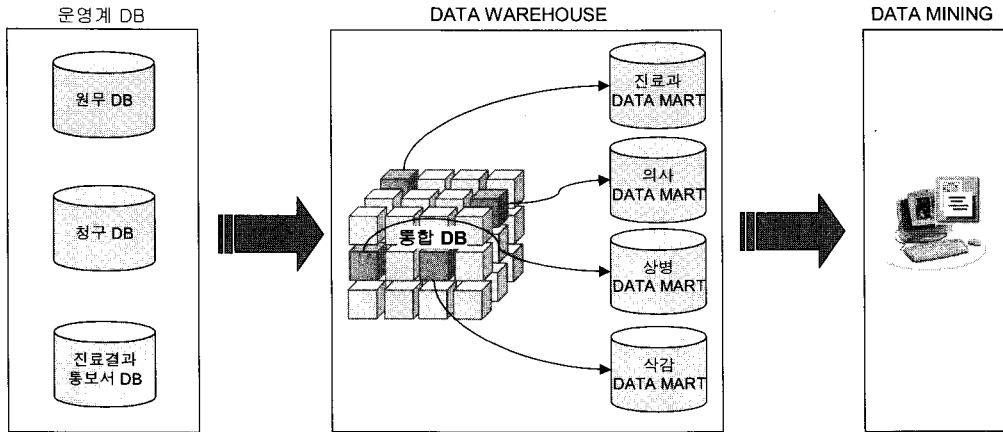
#### 5.1.1 진료비청구 삭감분석시스템의 범위

국내 병원의 경영환경이 급속도로 악화되고 있는 가운데, 경쟁력 강화를 위해 많은 병원들이 다양한 방안을 마련하고 있다. 데이터웨어하우스와 데이터마이닝을 활용한 진료비 청구삭감분석시스템의 구축은 이러한 시도 중의 하나로써 원시 데이터로부터 주제 중심적이고, 통합

적이며 시간성을 가지는 비취발성 자료의 집합인 데이터웨어하우스를 구축한 후 다양한 데이터마이닝 알고리즘으로 청구 삭감분석을 가능하게 하는 시스템이다. 병원은 진료비 청구 삭감 분석시스템을 구축함으로써 정확한 삭감분석을 통한 각종 진료 Order 발생 통계와 효과적인 투자에 의한 경영상태 개선, 또한 다양한 형태의 임상분석 작업 등의 효과를 기대할 수 있다. 진료비 청구 삭감 분석시스템을 병원 경영에 적용할 수 있는 분야는 크게 다음과 같은 두 가지로 나타난다. 첫째, 기존의 O.C.S(Order Communication System: 처방전달시스템)에서 발생한 데이터와 청구삭감 데이터를 이용하여 이를 다차원으로 분석하여 경영분석에 활용할 수 있다. 둘째, 원무데이터와 청구데이터 그리고 건강보험심사평가원 삭감데이터를 활용하여 빈번하게 발생하는 삭감원인을 분석, 삭감 재발을 방지하는 자료 등에 이용될 수 있다. 본 연구에서는 원시자료가 원무데이터와 청구데이터 그리고 건강보험심사평가원 삭감데이터로 한정되어 있기 때문에 진료비 청구 삭감 분석시스템의 구축으로 보여줄 수 있는 분야는 후자에 국한하기로 한다.

#### 5.1.2 진료비청구 삭감분석시스템의 구성

본 진료비 청구 삭감 분석시스템에 필요한 하드웨어는 크게 Web Server, Mining Client Server 그리고 Data Warehouse Server로 구성되어 있



〈그림 5〉 데이터웨어하우스 구성도

다. 이 중 Web Server와 Mining Client Server의 운영체제는 Windows 2000이며, Data Warehouse Server는 Alpha Server 8200이며, 운영체제는 Digital Unix 버전 4.0d이다. 데이터베이스는 Oracle 7.3.4.5를 이용했으며, Mining Client Server의 프로그램 개발도구는 Visual Basic 6.0 Enterprise를 사용했다.

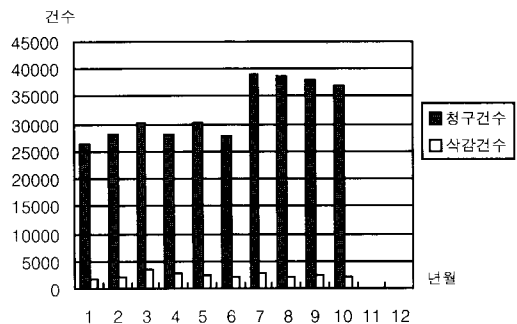
한편, 청구된 진료비의 삭감 원인을 발생 가능한 분야와 원인별로 분석하려는 개발목적을 달성하기 위하여 본 시스템에서는 삭감사유를 진료과, 의사, 상병, 진료수가별로 요약 및 상세화 된 데이터웨어하우스를 설계를 시도하였다. 이를 위하여 본 시스템에서는 기존의 운영계 데이터베이스와 건강보험심사평가원에서 제공하는 심사결과통보서에서 데이터를 추출, 변환 및 정제한 데이터베이스를 통합하여 특정 목적에 맞도록 항목별로 지역화 된 데이터마트를 구축하였는데, <그림 5>는 본 시스템의 데이터웨어하우스 구성도를 보여주고 있다.

## 5.2 진료비 청구 삭감 분석시스템의 프로토타입

### 5.2.1 데이터의 수집

D의료원에서 진료를 받은 환자 중에 건강보

험심사평가원에 청구한 2002년 1월~10월 데이터를 표본으로 사용하여, 운영계 데이터와 청구 데이터, 삭감 데이터(건강보험심사평가원 제공)로 추출, 변환, 정제하여, 데이터웨어하우스를 구축하고 삭감분석의 기초 자료로 사용하였다. 이 데이터에는 총 401,490건의 청구건수와 이들 진료 내역에 포함된 품목(진료과목, 의사, 상병, 삭감사유, 진료수가)이 기록되어 있다. 그러나 연관규칙은 2개 이상 품목의 종속관계를 발견하는 작업이므로 단지 1개의 품목만을 분석하는 것은 대상에서 제외시켰다. <그림 6>은 청구건수 및 삭감건수를 포함하는 데이터의 분포를 나타내고 있다.



〈그림 6〉 청구 건수 및 삭감건수 분포

### 5.2.2 데이터웨어하우스의 구축

데이터웨어하우스를 구축하는 가장 중요한 이유는 다양한 정보의 제공에 있으므로 구축대상과 범위를 정확하게 파악하여야 한다. 그러나 운영계 DB에서 취급하는 모든 데이터를 데이터웨어하우스의 범위로 파악한다면, 주어진 시간 및 자원으로 구현하기가 어려울 뿐 아니라, 비중을 두어야 하는 데이터와 중요하지 않은 데이터가 섞여있는 상황이므로, 필요한 정보를 도출하는 것이 모래알 속에서 좁쌀을 찾는 것과 같은 어려운 작업이 된다. 그러므로 운영계 DB의 데이터 중에서 어떤 데이터를 어느 정도까지 데이터웨어하우스에 포함시킬 것인가를 결정하는 것은 아주 중요한 일이다.

데이터웨어하우스의 대상을 결정짓기 위해서 각 부문별로 현재 사용되고 있거나 제공하기를 원하는 보고서 양식을 수집하고, 각 부문별로 삭감담당자/관리자/의사 등을 대상으로 인터뷰를 실시하여 필요한 정보요건을 도출하였다. 그러나, 현재 운영계 DB에서 사용하는 업무 코드는 장기적인 관점에서 설정한 것이 아니고, 일

시적인 필요에 의하여 추가된 경우가 많으며, 코드의 명칭이나 형태 등이 표준화되어 있지 않은 경우가 대부분이므로 이러한 업무 코드를 그대로 사용한다면 데이터웨어하우스로부터 정확한 정보를 도출하기 어려우며, 도출한 정보의 진정한 의미를 파악하기도 어렵다. 특히, 동일한 업무코드가 부문별로 다른 형태로 사용되는 경우가 많아 모든 코드를 수집하여 일관된 시각에 의하여 재구성하였다.

데이터웨어하우스 DB를 설계하면서 가장 어려움을 겪었던 사항은 유사한 테이블을 병합할 것인지, 분리할 것인지에 대한 것이었다. 예를 들어 외래 테이블과 입원 테이블, 청구 테이블과 삭감 테이블을 동일한 테이블로 병합하여 관리하는 것이 유리한지 혹은 별도로 독립하여 관리하는 것이 유리한지를 결정해야 할 필요가 있는 것이다. 본 연구에서는 유사한 테이블을 통합관리 하였는데, <표 3>은 운영계 데이터베이스에서 가장 핵심이 되는 테이블들의 항목과 형식을 보여주고 있다. <표 3>에서 하나의 레코드는 한 개의 품목에 대한 진료정보이며, PANO

<표 3> 운영계 데이터베이스 테이블의 항목과 형식

외래 테이블	입원 테이블	청구 테이블	삭감결과통보서 테이블	심사결과통보서 테이블
( ACTDATE DATE	( ACTDATE DATE	( WRIND NUMBER(8)	( SMSGCHSU VARCHAR2(8)	( SMSGCHSU VARCHAR2(8)
PANO CHAR(8)	PANO CHAR(8)	SEQNO1 NUMBER(4)	JUPJNO VARCHAR2(7)	JUPJNO VARCHAR2(7)
BI CHAR(2)	BI CHAR(2)	SEQNO2 VARCHAR2(6)	MUKNO VARCHAR2(5)	MUKNO VARCHAR2(5)
BDATE DATE	BDATE DATE	SUNEXT CHAR(6)	CSEQNO VARCHAR2(3)	CSEQNO VARCHAR2(3)
ENTDATE DATE	ENTDATE DATE	BUN CHAR(2)	MSEQNO VARCHAR2(5)	JGIGAN VARCHAR2(16)
ENTTIME CHAR(5)	ENTTIME CHAR(5)	BUN NUMBER(8,3)	LINENO NUMBER(3)	TIMNAME VARCHAR2(35)
SUNEXT CHAR(8)	SUNEXT CHAR(8)	NAL NUMBER(4)	BUN12 VARCHAR2(1)	JONAME VARCHAR2(35)
BUN CHAR(2)	BUN CHAR(2)	PRI CE NUMBER(7)	JUSAV02 VARCHAR2(2)	DJNAME VARCHAR2(26)
NJ CHAR(2)	NJ CHAR(2)	GENGT CHAR(1)	JJAMT NUMBER(8)	TEL VARCHAR2(20)
REALQTY NUMBER(5,2)	REALQTY NUMBER(5,2)	GBG SUL CHAR(1)	RSR01U VARCHAR2(2)	MSEQNO VARCHAR2(5)
QTY NUMBER(5,2)	QTY NUMBER(5,2)	GBCHILD CHAR(1)	COMENTS VARCHAR2(1750)	SWANE VARCHAR2(20)
NAL NUMBER(2)	NAL NUMBER(2)	AMT NUMBER(8)	ROW10KEY VARCHAR2(18)	BHK1HO VARCHAR2(8)
BASEAMT NUMBER(8)	BASEAMT NUMBER(7)	UPCHECK CHAR(1)		PNAME VARCHAR2(20)
GBSPC CHAR(1)	GBSPC CHAR(1)	XLINENO NUMBER(3)		BHJUN VARCHAR2(1)
GBNGT CHAR(1)	GBNGT CHAR(1)	XIND VARCHAR2(2)		JRTYPE VARCHAR2(2)
GBIMV CHAR(1)	GBIMV CHAR(1)	XIND VARCHAR2(2)		JUMIND VARCHAR2(13)
GBDIV NUMBER(2)	GBDIV NUMBER(2)	XCODE VARCHAR2(1)		JCONF VARCHAR2(2)
GBCHILD CHAR(1)	GBCHILD CHAR(1)	XCODE VARCHAR2(10)		NECONF VARCHAR2(26)
GBSELF CHAR(1)	GBSELF CHAR(1)	XDTY NUMBER(8,3)		CGJ_TAMT NUMBER(1,.)
GBGSUL CHAR(1)	GBGSUL CHAR(1)	XPRI CE NUMBER(9,1)		CGJ_BAMT NUMBER(1,.)
DEPTCODE CHAR(2)	DEPTCODE CHAR(2)	XNAL NUMBER(4)		CGJ_YAMT NUMBER(1,.)
DRCODE CHAR(6)	DRCODE CHAR(6)	XAMT NUMBER(8)		CGJ_BKBMAMT NUMBER(1,.)
WARDCODE CHAR(2)	WARDCODE CHAR(2)	XCHANGE DATE VARCHAR2(8)		SSA_GLEBN VARCHAR2(4)
SUCODE CHAR(8)	SUCODE CHAR(8)	XSFYAKGUBUN CHAR(1)		SSA_TAMT NUMBER(1,.)
GBSLIP CHAR(1)	GBSLIP CHAR(1)	XSFYAKGATE VARCHAR2(8)		SSA_BAMT NUMBER(1,.)
GBHOST CHAR(1)	GBHOST CHAR(1)	XSFYAKGIWAN VARCHAR2(8)		SSA_LDKAMT NUMBER(1,.)
AMT1 NUMBER(8)	AMT1 NUMBER(8)	XSEKPTGUBUN CHAR(2)		SSA_JYAMT NUMBER(1,.)
AMT2 NUMBER(8)	AMT2 NUMBER(8)	XHCODE VARCHAR2(50)		SSA_SBAMT NUMBER(1,.)
SEQNO NUMBER(4)	SEQNO NUMBER(4)	XJNAME VARCHAR2(140)		SSA_CJAMT NUMBER(1,.)
GELCODE VARCHAR2(8)	GELCODE VARCHAR2(8)	XB00M1 VARCHAR2(32)		SSA_WJAMT NUMBER(1,.)
SENDYYMM CHAR(4)	SENDYYMM CHAR(4)	XGHOST CHAR(1)		SSA_SAMT NUMBER(1,.)
REMARK VARCHAR2(50)	REMARK VARCHAR2(50)	XDDCHECK CHAR(1)		SSA_JYAMT NUMBER(1,.)
PART CHAR(1)	PART CHAR(1)	XBD01CAUSE VARCHAR2(1)		JJAMT2 NUMBER(3)
BONKIND CHAR(1)		MINDATE VARCHAR2(8)		NKNT NUMBER(3)
				TOYAKCNT NUMBER(3)
				ORDERCNT NUMBER(3)
				YSTARTDATE VARCHAR2(8)
				SPODCE VARCHAR2(4)
				LLODCE VARCHAR2(5)
				COMMENTS VARCHAR2(1750)
				ROW10KEY VARCHAR2(18)

<표 4> 데이터웨어하우스의 항목과 형식

발생 테이블		청구/삭감 테이블		운영계/청구/삭감 테이블	
YYMM	CHAR(6)	YYMM	CHAR(6)	YYMM	CHAR(6)
IPDOPD	CHAR(1)	IPDOPD	CHAR(1)	IPDOPD	CHAR(1)
JOHAP	CHAR(1)	JOHAP	CHAR(1)	JOHAP	CHAR(1)
DTNO1	CHAR(1)	DTNO1	CHAR(1)	DTNO1	CHAR(1)
SEQNO	CHAR(8)	SEQNO	CHAR(8)	SEQNO	CHAR(8)
JUPSUNO	VARCHAR2(7)	JUPSUNO	VARCHAR2(7)	JUPSUNO	VARCHAR2(7)
DEPTCODE	CHAR(2)	DEPTCODE	CHAR(2)	DTNO	CHAR(2)
DRCODE	CHAR(6)	DRCODE	CHAR(6)	DEPTCODE	CHAR(2)
B1	CHAR(2)	B1	CHAR(2)	DRCODE	CHAR(6)
ILLCODE	VARCHAR2(6)	ILLCODE	VARCHAR2(6)	ILLCODE	VARCHAR2(6)
SUNEXT	CHAR(8)	SUNEXT	CHAR(8)	ANALCODE	CHAR(2)
BUN	CHAR(2)	BUN	CHAR(2)	ANALCODEEXT	CHAR(1)
NU	CHAR(2)	NU	CHAR(2)	SUNEXT	CHAR(8)
SOUPRICE	NUMBER(9)	CHUPRICE	NUMBER(9)	BUN	CHAR(2)
SOUQTY	NUMBER(7,2)	CHUQTY	NUMBER(7,2)	NU	CHAR(2)
SOUNAL	NUMBER(4)	CHUNAL	NUMBER(4)	SOUPRICE	NUMBER(9)
SOUAMT	NUMBER(9)	CHUAMT	NUMBER(9)	SOUQTY	NUMBER(9,2)
GBNGT	CHAR(1)	ANALPRICE	NUMBER(9)	SOUAMT	NUMBER(9)
GBGISUL	CHAR(1)	ANALQTY	NUMBER(7,2)	CHUPRICE	NUMBER(9)
GBCHILD	CHAR(1)	ANALAMT	NUMBER(9)	CHUQTY	NUMBER(9,2)
ANALFLAG	CHAR(1)	ANALSAYOU	VARCHAR2(2)	CHUAMT	NUMBER(9)
		GBNGT	CHAR(1)	ANALPRICE	NUMBER(9)
		GBGISUL	CHAR(1)	ANALQTY	NUMBER(7,2)
		GBCHILD	CHAR(1)	ANALAMT	NUMBER(9)
		REMARK	VARCHAR2(100)	REMARK	VARCHAR2(60)
		COMMENTS	VARCHAR2(1750)	ANALFLAG	CHAR(1)
		ANALFLAG	CHAR(1)	OLD_DTNO1	CHAR(1)
				OLD_DTNO	CHAR(2)
				OLD_DEPTCODE	CHAR(2)
				OLD_DRCODE	CHAR(6)

(병록번호), WRTNO(일련번호)는 한 명의 환자 진료내역에 대하여 부여된 순번이다. 예를 들어, 어느 한 환자가 한번에 세 개의 진료 Order가 발생되었다면 테이블에는 세 개의 레코드로 나뉘어 기록되며, PANO(병록번호)는 동일하게 부여된다. 그러나 이와 같은 데이터에는 각 테이블의 모든 정보가 포함되어 있지 않기 때문에 특정 삭감 분석을 한번에 파악하기 위해서는 삭감분석에 관련된 데이터를 별도의 데이터웨어하우스로 구축하여야 한다. <표 4>는 <표 3>의 데이터가 데이터웨어하우스로 가공된 데이터의 형식을 보여주고 있다.

### 5.2.3 분석 대상

본 연구에서는 ‘삭감사유’의 연관관계를 진료과별, 의사별, 상병별, 삭감사유별, 진료수가별로 나누어 분석하였는데, 이를 위하여 삭감사유를 삭감 분석할 항목의 값에 따라 ‘진료과’, ‘의사’, ‘상병’으로 구분하였고 삭감사유는 ‘삭감코드’ 항목을 이용하여 각각 구분하였다. <표 5>, <표 6>은 각각 진료과별, 의사별, 상병별 삭감 건수를 보여주고 있는데 전체 삭감건수를 보면

C(요양급여 기준 적용 착오)가 전체의 80% 이상을 차지하고 있다. 즉, 요양급여 기준 적용 착오로 삭감되는 건수가 대부분이다. 그리고 진료과별로는 <표 5>에서 보듯이 ‘GI(소화기내과)’에서 13%가 삭감되어 가장 많이 삭감되었으며, 의사별 삭감건수는 <표 6>에서 보는 바와 같이 ‘921024’에서 10%가 삭감되어 의사 중에는 가장 많이 삭감되었음을 알 수 있다.

### 5.2.4 데이터마이닝 기법 선정

진료비 청구 삭감 분석시스템에 대한 효율을 높이기 위해서는 당연히 분류예측의 정확도가 높은 알고리즘을 사용하여야 한다. 일반적으로 신경망 알고리즘은 예측 정확도는 높지만 결과에 대한 해석이 불가능하다. 반면, 의사결정나무 분석은 결과 해석이 명쾌하나 정확도 측면에서는 예측력이 떨어진다. 이와 같이 알고리즘마다 분명한 특징이 있으므로 데이터마이닝 분석을 할 때 사용할 알고리즘의 선택은 분석 결과를 어떠한 부서에서 어떤 목적으로 사용하는가에 따라 달라져야 한다. 예를 들어 병원 이탈 가능성이 높은 환자에 대한 특성을 파악하여 캠페

인 전략을 세워야하는 마케팅 부서에서는 의사 결정나무 분석 알고리즘을 선호하는 반면 진단 예측에 대한 정확성이 업무 효율에 직접적인 영향을 미치는 진료부서에서는 신경망과 같은 알

고리즘을 선호 할 것이다.

본 연구에서 소개하는 진료비 청구 삭감 분석 시스템의 삭감 예측의 경우 신경 이러한 원칙에 따라 선별된 삭감 데이터에 연관규칙 알고리즘

<표 5> 진료과별 삭감건수

	AN	CS	CV	DM	DT	EN	ER	FM	GI	GS	GU	GY	HO	NE	NP	NS	OB	OL	OS	OT	PD	PG	PL	PS	PY	RE	RH	TR	합계	%
A	0	0	1	0	0	8	0	0	0	2	0	0	0	1	10	0	2	0	1	0	0	0	2	0	0	0	0	27	0.54	
B	2	17	19	19	16	2	39	1	46	94	9	53	30	20	34	36	44	7	19	26	29	14	16	12	9	20	3	2	638	12.75
C	2	109	465	30	1	93	78	1	531	232	102	273	118	417	249	153	355	281	107	34	180	5	237	13	36	25	10	2	4139	82.75
E	0	0	1	0	3	0	21	0	6	11	2	1	1	1	3	1	1	1	2	0	0	0	2	2	0	1	2	0	66	1.32
F	0	0	1	0	0	0	0	0	1	4	0	1	1	1	0	1	0	0	2	0	0	0	0	0	0	0	0	0	17	0.34
I	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	0.20
L	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0.02
P	0	0	3	0	0	0	4	0	0	2	0	1	0	0	0	4	0	48	0	0	0	0	0	0	0	0	0	0	63	1.24
U	0	0	0	0	0	0	23	0	2	0	3	0	0	0	1	0	0	3	0	0	0	0	0	0	0	0	0	0	42	0.84
합계	4	134	490	49	20	103	167	2	586	345	117	329	150	440	297	195	402	340	131	60	228	19	255	29	45	46	15	4	5002	100
%	0.1	2.7	9.8	1.0	0.4	2.1	3.3	0.0	11.7	6.9	2.3	6.6	3.0	8.8	5.9	3.9	8.0	6.8	2.6	1.2	4.6	0.4	5.1	0.6	0.9	0.9	0.3	0.1	100	

<표 6> 의사별 삭감건수

	921024	000127	990045	980011	000161	000194	010176	760016	951208	860181	820379	780121	780063	941675	850184
A	1	0	0	0	0	0	1	0	0	0	5	5	0	0	0
B	37	12	7	5	14	1	18	6	11	8	7	10	22	11	16
C	267	226	207	157	146	157	109	106	96	98	94	86	67	72	60
E	1	0	2	0	2	0	0	3	0	2	2	0	0	0	1
F	0	0	0	0	0	0	0	1	0	0	0	0	1	1	1
I	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
L	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
P	0	0	0	24	0	0	0	0	1	0	0	0	0	0	2
U	0	0	0	3	9	0	0	0	0	0	0	0	0	0	0
합계	306	238	216	189	171	158	128	116	108	108	108	101	90	84	80
%	10.3	8.0	7.3	6.4	5.8	5.3	4.3	3.9	3.6	3.6	3.6	3.4	3.0	2.8	2.7

	800082	000151	760111	910069	940865	820402	010200	860440	010199	850811	840832	881216	합계	%
A	0	0	0	0	1	0	0	0	0	1	0	0	14	0.47
B	0	4	11	8	26	9	3	5	2	3	3	3	262	8.82
C	78	72	64	63	43	56	65	61	63	59	45	50	2667	89.77
E	0	0	0	0	2	0	0	1	0	0	4	0	20	0.67
F	0	0	0	0	0	0	0	0	0	0	0	3	7	0.24
I	0	0	0	2	0	4	0	0	0	0	0	0	6	0.20
L	0	0	0	0	0	0	0	0	0	1	0	0	1	0.03
P	0	0	0	0	0	0	0	0	2	0	0	0	29	0.98
U	0	0	0	0	0	2	0	0	0	0	7	0	21	0.71
합계	78	76	75	73	72	71	68	67	67	64	59	0	2971	100
%	2.6	2.6	2.5	2.5	2.4	2.4	2.3	2.3	2.3	2.2	2.0	0.0	100	



을 적용하여 새로운 패턴을 발견하는 작업을 진행하였다. 그 결과, 몇 가지 중요한 패턴이 발견되었는데 그 중에서 일부는 <표 5>, <표 6>에 나타난 결과와 동일하게 진료과에서는 'GI(소화기내과)'에서 가장 많이 삭감되었고, 의사는 '921024'가 가장 많이 삭감된 것과 같이 이미 알려진 사실들이다. 그러나 매우 새롭고 흥미로운 정보들도 발견되었는데 예를 들어, 이비인후과(OL)에 의사(980011)가 H663(기타 만성 화농성 중이염) 상병으로 발생시킨 Order 중 J656361(임피단스(고막운동성계측)), J656363(임피단스(등골근반사소실검사)) 항목이 신뢰확률 100%로 이 항목들이 동시에 발생할 경우 삭감율이 높게 나타나고 있다. 이러한 결과를 삭감담당자가 분석해 보면 특정 상병에 대하여 과잉진료가 이루어졌는지 여부를 판단할 수 있으며, 이러한 분석결과는 의사들의 진료에 유익한 정보로 활용되어 삭감을 줄이는데 크게 기여할 수 있다.

## VI. 결 론

### 6.1 연구결과

데이터마이닝은 대용량의 데이터베이스에 저장된 데이터로부터 관심있는 지식 또는 규칙이나 패턴을 추출하는 일련의 과정이다. 일반적으로 데이터마이닝은 데이터베이스에서 특정지식을 발견하는 중요한 과정으로 이해된다. 이러한 데이터마이닝은 효과적인 기법들이 출현하면서 응용분야가 확대되고 이들 분야에서 긍정적인 활용성효과가 보고 되면서 기업들의 관심이 고조되고 있다. 그러나, 데이터마이닝은 가능성과 기대효과 못지않게 극복해야 할 과제들이 남아있다. 예를 들면, 데이터마이닝 프로젝트는 대규모 데이터를 다루어야 하며, 여러 분야의 관련자들이 함께 참여하여야 하므로 장기적이고 구체적인 계획이 필요하며, 통합된 데이터가 필요하다. 또한 현재 사용되고 있는 알고리즘은 모든 경우

에 일반화하여 적용시키기에는 많은 문제점들이 남아 있다. 이러한 한계점에도 데이터마이닝은 대량의 데이터로부터 특정 의사결정이나 문제를 해결하는데 필요한 지식이나 정보를 추출해 내는데 상당히 중요한 역할을 하고 있으므로 효과적인 알고리즘만 만들어 낼 수 있다면 다양한 분야에서 엄청난 역할을 할 수 있을 것으로 기대된다.

본 연구에서는 진료비 청구삭감분석을 하기 위하여 연관규칙 알고리즘을 활용하여 데이터마이닝을 시도하였다. 이를 위하여, 청구삭감분석을 위한 프로토타입을 개발 한 후 D의료원의 실제 데이터를 이용하여 실증적 연구를 시행하였다. 본 연구에서 개발된 청구삭감분석을 위한 프로토타입 시스템의 구현을 통해서 얻어진 결과는 다음과 같다.

첫째, 기존의 삭감시스템에서는 보통 2~3개월의 시간이 소요되었지만 새로운 진료비 청구 삭감 분석시스템을 구현함으로써 삭감담당자의 삭감 분석업무가 1주일 이내에 가능하게 되었다.

둘째, 진료비 청구 삭감 분석시스템에서 80% 정도는 자동으로 삭감 분석 작업이 이루어지기 때문에 삭감담당자의 삭감 분석업무의 부담을 줄일 수 있게 되었다.

셋째, 통계학적 방법에 의해서 발견된 결과인 진료과에서는 'GI(소화기내과)'에서 가장 많이 삭감되며, 의사는 '921024'가 가장 많이 삭감된다는 등의 이미 알려진 사실들도 연관규칙 알고리즘을 적용한 데이터마이닝을 통하여 다시 한번 확인되었다.

넷째, 이비인후과(OL)에 의사(980011)가 H663(기타 만성 화농성 중이염) 상병으로 발생시킨 Order 중 J656361(임피단스(고막운동성계측)), J656363(임피단스(등골근반사소실검사)) 항목이 신뢰확률 100%로 이 항목들이 동시에 발생할 경우 삭감율이 높게 나타나는 것으로 분석되는 등 일반적인 통계학적 방법으로는 발견하기 어려운 규칙들도 발견되었다.

## 6.2 연구의 한계점 및 향후 과제

본 연구는 데이터마이닝 기법 중 하나인 연관 규칙 알고리즘을 사용하여 청구삭감분석을 위한 프로토타입을 개발한 후, 실제 데이터를 이용하여 앞에서 언급한 결과들을 이끌어 내는 기여를 하였지만, 아직도 다음과 같은 한계점을 가지고 있다.

첫째, 본 연구에서 사용된 데이터마이닝 알고리즘 중 연관규칙 알고리즘을 사용함으로써 연관규칙을 도출하는데 컴퓨터 처리 시간이 많이 소요되었다.

둘째, 연관규칙 알고리즘은 연관규칙의 수가 삭감 분석 항목들의 수에 비례하여 기하급수적으로 늘어나기 때문에 항목의 수를 줄이는 작업이 필요하므로 삭감 분석 항목 선정이 어렵다.

셋째, 연관규칙 알고리즘에 의해 도출된 규칙들 중에서 불필요한 규칙을 제외하고 필요한 규칙들만을 선별하기가 어렵다.

넷째, 특정 상병에 대하여 이루어진 과잉진료의 경우 환자의 증상에 따라 불가피하게 이루어진 것인지, 아니면 수입증대를 위하여 이루어진 것인지를 분석할 수 없었다.

이상과 같은 한계점들을 극복하기 위해서는 삭감이 발생될 우려가 있는 데이터를 사전에 제거하여 삭감율을 최대한 낮출 수 있도록 사례기반추론 알고리즘에 기초한 프로토타입을 구축해서 활용해 보는 시도가 바람직하다. 또한, 올바른 삭감분석 항목을 선정할 수 있도록 진료의사, 삭감담당자, 전산담당자 사이에 충분한 협의가 이루어진다면 청구삭감분석시스템의 분석결과가 한층 더 정확해 질 수 있을 것이다.

## 참고 문헌

강현철, 한상태, 최중후, 김은석, 김미경, SAS Enterprise Miner 4.0을 이용한 데이터마이

닝, 자유아카데미, 2001.

건강심사평가원, 정관, 제7장, 제37조 제1항, 1999.

김석일, “일 상병의 진료비 심사모형 개발에 관한 연구”, 석사학위논문, 연세대학교, 1994.

김수배, 원무관리실무, 군자출판사, 1998.

김정수, “우리나라 의료보험 진료비심사 지급제도의 개선방안에 관한 연구”, 석사 학위논문, 경희대학교, 1990.

도용태, 김일곤, 김종환, 박창현, 인공지능 개념 및 응용, 사이텍미디어, 2001.

데이터베이스월드, 한국데이터베이스진흥센터, 1998년 8월.

박시우, “사례기반추론을 이용한 데이터웨어하우스의 설계방법론 개발”, 석사학위논문, 홍익대학교, 1997.

안옥희, “의료보험 진료비 삭감 및 이의신청 분석에 관한 조사 연구”, 석사학위논문, 경희대학교, 1998.

용환승, 데이터마이닝, 1998, 그린.

이경영, “데이터마이닝 기법을 이용한 의료자료 비교분석 및 의료정보시스템 구현”, 석사학위논문, 전북대학교, 2002.

장남식, 홍성환, 장재호, 데이터마이닝, 대청미디어, 2000.

장동인, 실무자를 위한 데이터웨어하우스, 대청미디어, 1999

지식 창출병원으로의 혁신을 위한 의료 데이터웨어하우스, 병원신문, 2001년 8월2일, 11면.

최국렬, 데이터마이닝 이론과 실습-보건의료데이터 중심, 2001, 청구문화사.

Agrawal. R. and Srikant. R., *Fast Algorithms for Mining Association Rules in Large Databases*, Santiago: 20th Int'l Conf. on VLDB, Santiago, Chile, 1994.

Ashely, K. and Rissland, E., “A Case-based Approach to Modeling Legal Experties”, *IEEE Expert. University of Massachusetts*, 1988, 3, Amherst, MA., pp.70-96.

- Berry, Michael J. A. and Linoff, Gordon S., *Data Mining Techniques: For Marketing, Sales, and Customer Support*, John Wiley & Sons. New York, 1997.
- Breiman, L., Friedman, J. H., Olshen, R. A. and Stone, R. A., *Classification and Regression Trees*, Wadsworth, Belmont, 1984.
- Hammergren, T., *Data Warehousing: Building the Corporate Knowledgebase*, International Thomson Computer Press, 1996.
- Han, Jiawei and Kamber, Micheline, *Data Mining: Concepts and Techniques*, Morgan Kaufmann Publishers, 2001.
- Holland. J. H., *Adaptation in Nature and Artificial Systems*, The University of Michigan Press, Ann Arbor, MI, 1975.
- Inmon, Bill, *Building the Data Warehouse*, John Wiley & Sons. Inc, 1996.
- Kass. G. V., "An Exploratory Technique for Investigating Large Quantities of Categorical Data", *Applied Statistics*, Vol. 29, 1980, pp.119-127.
- Kimball, R., *The Database Market Splits*, DBMS Magazine, September, 1995.
- Kolodner, J., *Case Based Reasoning*, Morgan Kaufmann, San Francisco, 1993.
- Quinlan, J. R., *C4.5 Programs for Machine Learning*, Morgan Kaufmann, San Francisco, 1993.
- Sean, Kelly, *Data Warehousing-The Key to Mass Customization*, John Wiley & Sons. Inc., New York, 1994.
- Watson, I., *Applying Case-based Reasoning: Techniques for Enterprise Systems*, Morgan Kaufmann, San Francisco, 1997.
- 건강심사평가원(<http://www.hira.or.kr>), 자료실, 2002.
- 대한병원협회(<http://www.kha.or.kr>), 자료실, 2002. <http://idflow.netian.com/신경망.htm>.
- <http://chumsungdae.com/korea/research/reserach.html>.
- [http://data-science.net/mining\\_how\\_ass.htm#마이닝기법%20\(연관규칙\)](http://data-science.net/mining_how_ass.htm#마이닝기법%20(연관규칙)).
- <http://chumsungdae.com/korea/research/reserach.html>.
- [http://data-science.net/mining\\_how\\_neu.htm](http://data-science.net/mining_how_neu.htm).

## **A Study on the Development and Implementation of a Data-mining Based Prototype for Hospital Bill Claim Reduction System**

Sangjin Yoo\* · Mun Ro Park\*\*

### **Abstract**

Changes in business environment caused by globalization of the world economy and the beginning of the knowledge society forced hospitals to equip with tools for the enhanced competitiveness. In other words, hospitals must aim three targets such as acquisition of advanced medical skills and equipments, improvement of service level for patients, and achievement of superior managerial performance simultaneously.

This study has been done to suggest a way to reduce the possibility of hospital bill claim reduction as an alternative for the achievement of superior managerial performance. If the reduction rate of hospital bill claim is high, it will put negative impact on the hospital's revenue stream and hospital's reliability. Thus, if they want to stay competitive, hospitals need to device ways to cut the reduction rate as much as possible.

In this study, a prototype system has been developed and implemented to check the possibility to cut the reduction rate through deep analysis of causes of reduction. The prototype first developed utilizing data mining techniques and the relation rules algorithm. Then the prototype was tested its performance using the D hospital's live data.

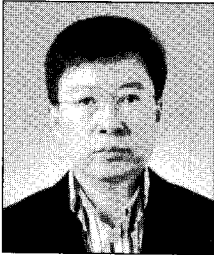
***Keywords: Data Mining, National Health Insurance, Hospital Bill, Claim Reduction System***

---

\* Professor, Department of MIS, Keimyung University

\*\* Division of Information Systems, Dong San Medical Institute

## ◎ 저 자 소 개 ◎



**유 상 진 (yoosj@kmu.ac.kr)**

서강대학교에서 물리학(이학사), 경영학(경영학사)을 복수 전공하였으며, 미국 University of Nebraska-Lincoln에서 MIS전공으로 박사학위를 취득하였다. 현재 계명대학교 경영정보학과 교수로 재직 중이며, 현직에 오기전에는 미국 Bowling Green State University에서 조교수로 재직하였다. 한국경영정보학회 부회장, 한국정보시스템학회 회장을 역임하였으며, 현재 대구경북 CIO협의회 회장, 대구경북 ECRC전문위원, 대한상사중재원 중재인으로 활동하고 있다. 주요 관심분야는 IS/IT의 전략적 활용, 경영혁신, 지식경영, 인터넷창업등이다.



**박 문 로 (munro@dsmc.or.kr)**

경일대학교에서 컴퓨터공학(학사), 동대학원에서 경영정보학(석사)을 졸업하고 학위를 취득하였다. 현재 계명대학교 동산의료원 전산운영팀 재직 중이다. 주요 관심분야는 DW(Data Warehouse), DM(Data Mining) 등이다.

논문접수일 : 2004년 11월 17일  
1차 수정일 : 2004년 12월 18일

게재확정일 : 2005년 3월 12일