

# CART 방법론을 사용한 클라우드 컴퓨팅 도입 의사 결정 모델링

백승현 · 장병윤\*

## Cloud Computing Adoption Decision-Making Modeling Using CART

Seung Hyun Baek · Byeong-Yun Chang\*

### ABSTRACT

In this paper, we conducted a study on place-free and time-free cloud computing (CC) adoption decision-making model. Panel survey data which is collected from 65 people and CART (classification and regression tree) which is one of data mining approaches are used to construct decision-making model. In this modeling, there are 2 steps: In the first step, significant questions (variables) are selected. After that, the CART decision-making model is constructed using the selected variables. In the variable selection stage, the 25 questions are reduced to 5 ones. The benefits of question reduction are quick response from respondent and reducing model-construction time.

**Key words** : CART, Cloud Computing, Panel Data, Data Mining, Variable Selection

### 요약

본 논문에서는 장소와 시간의 제약을 받지 않는 클라우드 컴퓨팅 도입 의사 결정 모델링에 대한 연구를 진행하였다. 연구에서는 65명의 응답자에게 수집된 패널데이터와 데이터마이닝 방법 중 하나인 CART(회귀분류나무)를 사용하여 의사결정 모델을 구축하였다. 모델링에는 2단계로 진행되는데 첫 번째 단계에서는 패널데이터를 사용하여 도입 의사를 결정하는데 영향을 미치는 문항들을 선택하고 2 번째 단계에서는 선택된 문항을 사용하여 도입 의사 결정 모델을 구축하였다. 문항 선택을 통하여 설문지 수집 문항수를 25개에서 5개로 줄일 수 있어 응답자에게 빠른 답변을 얻을 수 있고 데이터의 사이즈가 작기 때문에 모델 구축 시간을 줄일 수 있는 장점을 보여주었다.

**주요어** : CART, 클라우드 컴퓨팅, 패널데이터, 데이터마이닝, 변수선택

## 1. 서론

최근 장소와 시간의 제약을 받지 않는 클라우드 컴퓨팅이 개인, 기업, 공공 및 교육 등의 다양한 공간에서 활성화 되고 있다. 컴퓨팅 환경은 최초 개인용 컴퓨터 환경에서 서버-클라이언트 환경으로 바뀐 후에 최근까지 클라우드 컴퓨팅 환경으로 지속적으로 진화하고 있다. 컴퓨팅 환경은 1세대는 1980년대 그리드 컴퓨팅으로 2세대는 1990년대의 유틸리티 컴퓨팅으로 진전되었고 2001년에 3세대인 SaaS(Software as a Service)로 그리고 최근에는

클라우드 컴퓨팅으로 발전되었다. 그리드 컴퓨팅 시대에는 여러 대의 컴퓨터를 병렬로 사용하여 특정한 문제를 푸는 것을 목표로 하였다. 유틸리티 컴퓨팅은 가상 플랫폼을 제공하기 위해 클러스터를 사용했으며 계량화된 비즈니스 모델을 사용하였다. SaaS에서는 가상화 뿐만 아니라 애플리케이션의 가치에도 의미를 부여하였다. 이러한 세 단계의 발전과정을 거쳐서 클라우드 컴퓨팅으로 정립되어 언제 어디서든 접근 할 수 있는 환경을 조성하였다 (Sim, 2009). 인터넷 기반의 컴퓨팅 기술인 클라우드 컴퓨팅은 1965년 미국의 컴퓨터 학자인 존 매카시에 의해서 제시된 컴퓨터 환경의 공공성에 대한 개념에서 유래되었다. 클라우드 컴퓨팅 서비스 이용자는 지속적으로 증가하고 있으며 기업이나 공공기관 등의 수요도 계속 증가 중이다. IHS iSuppli의 2011년 보고서에 따르면 퍼블릭 클라우드 서비스로 이용으로 인한 기업들의 지출이 2010

**Received:** 10 November 2014, **Revised:** 4 December 2014,  
**Accepted:** 22 December 2014

**\*Corresponding Author:** Byeong-Yun Chang  
E-mail: bychang@ajou.ac.kr  
School of Business, Ajou University

년 230억 달러에서 2015년 1100억 달러로 약 5배 정도의 증가가 전망된다고 한다. 또한 서비스별로 2013년에는 클라우드 컴퓨팅 시장을 분석하면 SaaS(Software as a Service)에서 298억달러, IaaS(Infrastructure as a Service) 123억달러, 그리고 PaaS(Platform as a Service)에서 53억달러의 규모가 될 것으로 판단된다(Kang, 2012). 이러한 폭발적인 증가추세에 따라 본 논문에서는 기업의 클라우드 컴퓨팅 도입 의사와 관련하여 설문조사를 통한 패널데이터를 획득하여 데이터 마이닝 방법론중 하나인 CART(Classification and Regression Tree) 방법을 사용하여 도입의사와 관련한 요인분석을 하였고 도입의사결정을 위한 모델을 제시하였다. 또한 설문분석을 위해 사용된 문항들은 25개로 기업의 클라우드 컴퓨팅 도입과 관련된 업무를 하는 직장인이나 연구자들에게 많은 시간을 할애하도록 하여 정확한 응답을 가져오지 못하는 부분이 존재할 것으로 판단되어 요인분석을 통해 가장 중요한 핵심 문항만 선택하려고 한다.

본 논문의 구성은 2장에서는 클라우드 컴퓨팅 및 관련 실증연구에 대한 고찰, 3장에서는 CART(회귀분류나무) 방법에 대해 기술한다. 4장에서는 데이터에 대한 설명과 실험에 대한 내용이 설명되어 있으며 5장에서 본 논문을 결론지었다.

## 2. 클라우드 컴퓨팅 및 관련 실증연구 고찰

### 2.1 클라우드 컴퓨팅

클라우드 컴퓨팅은 응용프로그램을 처리하기 위해 로컬 서버나 개인용 컴퓨팅 장치들을 가지지 않고 인터넷 상에서 IT 기반을 바탕으로 서버, 저장장치 또는 응용 프로그램과 같은 컴퓨팅 자원을 공유하는 안정된 컴퓨팅 환경을 제공하는 것을 말한다. 클라우드 컴퓨팅 환경에서는 소프트웨어는 웹브라우저 상에서 클라이언트 사용자에게 유료로 제공된다. 클라우드 컴퓨팅의 이념은, 2007 년에 통신 채널의 급속한 발전과 자신의 정보시스템을 확장하는 비즈니스 및 개인사용자의 요구의 기하 급수적 증가 덕분에 인기를 얻었다. 클라우드라는 용어의 출현은 2008 년 주제 인터넷 학회중 하나에서 논의되기 시작하였다. 이러한 논의의 결과로 클라우드 컴퓨팅의 다른 버전들이 제안되었다. 이러한 버전들 중 하나 때문에, 처음으로 구글 회장인 에릭 슈미트에 의해 사용되었고 후에 대중매체를 통해서 전파되었다(Arutynov, 2012). 클라우드 컴퓨팅은 군대나 연구 목적의 슈퍼 컴퓨팅이나 고성능 컴퓨팅 처리 능력에 적용, 재무 포트폴리오, 소비자 마케팅 분석

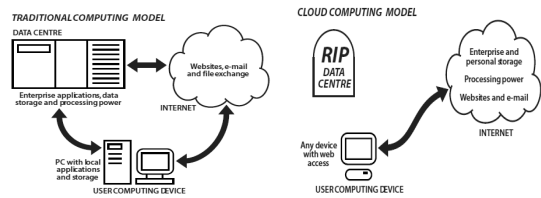


Fig. 1. Traditional and Cloud Computing (Barnatt, 2010, Kim 외, 2012)

과 같은 고객 관련 응용 분야에서 필요한 거대한 계산에 사용, 그리고 데이터 저장소를 제공하고 대규모의 컴퓨터 게임 등에 활용되도록 만들어졌다. 이러한 컴퓨팅 능력을 제공하기 위해서 클라우드 컴퓨팅은 개인용 컴퓨터를 하나의 서버로 이용하는 거대한 네트워크를 이용하여 처리 능력을 향상 시킨다. 이러한 시스템은 IT 기반시설을 이용하여 각기 다른 장소에 있는 자원을 하나로 모아서 공동 관리 체제로 엮는 것을 말한다. Fig. 1에서는 기존컴퓨팅과 클라우드 컴퓨팅을 구분하여 보여준다. Fig. 1에서 볼 수 있듯이 기존의 컴퓨팅에서는 사용자 컴퓨터에 소프트웨어 등 애플리케이션이 개인용 컴퓨터의 저장장치에 저장되는 반면에 클라우드 컴퓨팅에서는 인터넷 환경의 가상공간에 소프트웨어가 존재하고 사용자는 인터넷 접속만으로 그것들을 사용할 수 있다. 기존 컴퓨팅에서는 사용하고자 하는 소프트웨어가 많다면 개별 사용자별로 저장장치를 키워야 하는 단점이 있지만 클라우드 컴퓨팅에서는 소프트웨어를 서로 공유하기 때문에 저장장치의 제한이 없다. 클라우드 컴퓨팅은 서비스 운용 형태에 따라 세가지 종류가 포함되는데 이 세가지는 퍼블릭 클라우드, 프라이빗 클라우드, 하이브리드 클라우드 이다. 퍼블릭 클라우드는 불특정 다수를 대상으로 하여 제공되는 것으로 여러 서비스 이용자가 사용하는 서비스이다. 프라이빗 클라우드는 IT 부서의 통제 하에 기업 방화벽 내에서 구현되는 클라우드 컴퓨팅 플랫폼을 설명하는데 사용되는 것을 말한다. 프라이빗 클라우드는 퍼블릭 클라우드 시스템의 특징과 이점을 똑같이 제공하지만 기업과 고객 데이터에 대한 제어, 보안에 대한 걱정, 규정준수와 관련된 문제들을 포함하는 컴퓨팅 장애요인을 제거한다. 하이브리드 클라우드는 퍼블릭 클라우드와 프라이빗 클라우드를 결합한 형태의 클라우드 서비스로 공유를 원하지 않는 일부의 데이터나 서비스에 대해 프라이빗 정책을 적용하여 각 사용자의 개별 보안을 제공하는 서비스이다. 클라우드 컴퓨팅은 서비스 유형에 따라 IaaS(Infrastructure

as a Service), PaaS(Platform as a Service), SaaS(Software as a Service)로 나눌 수 있다. IaaS는 이용자에게 서버, 데스크톱, 스토리지 등의 IT자원을 통합, 가상화하여 필요로 하는 이용자에게 네트워크를 통해 하드웨어만을 제공하는 것을 말한다. PaaS는 서비스 이용자에게 소프트웨어 개발에 필요한 플랫폼(개발환경)을 제공하는 것을 말한다. 그리고 SaaS는 서비스 이용자에게 소프트웨어를 제공하는 것을 말한다(Kang, 2012).

## 2.2 클라우드 컴퓨팅 관련 실증연구 고찰

국내 클라우드 컴퓨팅 관련 연구들은 많이 진행되고 있지만 본 논문에서 시행한 설문조사를 통한 패널데이터를 이용한 분석은 드물게 진행된 것으로 판단된다. 현재 국내에서 진행된 클라우드 관련 설문조사 방법을 사용하여 실증분석을 진행한 연구들을 살펴보면 다음과 같다.

Hong 외(2012)는 국내 포털 서비스의 대표주자인 네이버와 다음의 클라우드 저장장치의 사용성을 평가하고 네이버의 N드라이버와 다음의 클라우드의 서비스를 비교/평가하여 사용자의 편의성 측면에서 연구를 진행하였다. Kim 외(2012)는 클라우드 컴퓨팅의 특성요인을 인식하고 관련 이론을 토대로 컴퓨팅 서비스의 적극적 도입 및 활용에 대한 인관관계를 검증하고자 기업의 IT 의사결정자 또는 의사결정지원자를 대상으로 설문하고 통계적으로 검증하였다. Lim 외(2012)는 기업의 클라우드 컴퓨팅 도입 의사결정에 영향을 미치는 요인, 즉 클라우드 컴퓨팅의 혜택, 클라우드 컴퓨팅의 위험, 외부환경의 영향력, 내부 환경의 영향력을 알아보기 위해 클라우드 컴퓨팅을 이용했거나 도입을 준비하는 기업과 클라우드 컴퓨팅 서비스를 제공하는 기업을 대상으로 설문조사를 실시한 후에 실증분석을 제공하였다. Kim & Kim(2011)는 개인의 모바일 클라우드 사용에 영향을 주는 요인으로 개인적 특성(개방성, 목적 지향성, 모바일 자기 효능감), 서비스 특성(신축성, 네트워크 접근성, 적절성), 시스템 특성(가용성, 데이터 융합성, 호환성) 과 모바일 신뢰의 조절효과를 제시했으며 설문조사로 수집된 자료를 가지고 구조방정식을 사용하여 그 관계를 규명하는 실증분석을 진행하였다. Kim 외(2012)는 도서관 사서들을 대상으로 하여 클라우드 컴퓨팅 서비스 인식에 대한 설문조사를 실시하여 도서관에서 클라우드 서비스를 적용, 적용에 적합한 도서관 업무 분야, 클라우드 컴퓨팅 서비스를 도서관 시스템에 적용 시 핵심적으로 고려해야 할 사항에 관한 결과를 도출하였다. Woo 외(2011)는 클라우드 컴퓨팅 환경에서 개인이 프로젝트를 수행할 때 나타나는 성과에 대한 실증

분석을 진행하였다. 실제 사용되는 상용 클라우드 시스템인 GoogleDocs 사용경험을 중심으로 연구가 진행되었다. 기업이나 공공기관의 클라우드 컴퓨팅의 적합성을 보기 보다는 개인의 클라우드 컴퓨팅 적합성을 평가하였다. Lim 외(2012)는 클라우드 환경에서 음악 스트리밍 서비스를 이용하는 사용자들의 음악 행태의 변화를 알아보기 위해 스트리밍 방식과 다운로드 방식을 비교하여 연구를 실시하였고 설문과 인터뷰 방식을 사용하여 클라우드 환경에서의 콘텐츠 소비 형태를 예측할 수 있도록 실증분석을 진행하였다.

Tara 외(2011)는 750명의 기본적인 컴퓨터 사용능력을 가진 커뮤니티 칼리지 재학생들을 연구에 참여시켜 커뮤니티 칼리지의 클라우드 컴퓨팅의 도입과 사용량에 대한 연구를 진행하였다. Wu(2011)는 공급업체와 기업사용자를 위한 SaaS 도입에 영향을 미치는 중요 요인을 탐색하였다. 분석 프레임으로 기술 수용 모델(TAM)과 러프 집합 이론(RST)을 사용하였다. 실증 분석은 타이완의 IT/MIS 기업을 대상으로 진행되었다. Low 외(2011)는 첨단산업에 속한 기업들의 클라우드 컴퓨팅의 채택에 영향을 미치는 요인을 조사하였다. 이 연구에서 조사한 여덟 가지 요인은 상대적 이점, 복잡성, 호환성, 최고경영관리지원, 기업규모, 기술 준비, 경쟁압박 그리고 거래 파트너의 압력이다. 실증분석을 위한 설문조사에서는 대만의 하이테크 산업에 속하는 111개의 기업으로부터 데이터를 수집하였고 로지스틱 회귀분석을 사용하여 분석하였다.

## 3. CART 방법론

의사결정 나무중의 하나인 CART(분류회귀나무)는 의사결정 규칙을 도표화하는 방법으로 독립변수에 의하여 종속변수를 예측하거나 분류하고자 할 때 사용하는 방법 중 하나이다. 분석과정이 나무구조로 표현되기 때문에 판별분석이나 회귀분석등과 같은 방법들에 비해서 쉽게 이해하고 설명할 수 있다는 장점이 있어서 분석의 정확도 보다는 분석과정의 설명이 필요한 경우에 사용하기 좋은 방법이다. 의사결정 나무가 사용될 수 있는 응용분야는 세분화(관측개체를 몇 개의 그룹으로 분할), 분류(독립변수에 근거하여 종속변수를 몇 개의 그룹으로 분류), 예측(독립변수에 근거하여 종속변수를 예측), 차원축소 및 변수선택(종속변수를 예측할 때 큰 영향을 주는 변수들을 다수의 독립변수에서 선택사용), 교호작용효과와 파악(여러 독립변수들이 서로 결합하여 종속변수에 적용되는 교호작용 파악), 범주의 병합 또는 연속형 변수의 이산화(범

주형 종속변수를 소수 몇 개로 병합하거나 연속형 종속변수를 범주화 하고자 할 때 사용)에 사용될 수 있다(Choi 외, 1998; Choi & Seo, 1999). CART는 범주형 종속 변수인 경우에는 지니지수를 사용하고 연속형 종속 변수인 경우에는 분산의 감소량을 사용하여 최적의 분류를 수행한다(Quinlan, 1993). CART의 구성요소는 뿌리마디(root node), 자식마디(child node), 부모마디(parent node), 끝마디(terminal node), 중간마디(internal node), 가지(branch), 깊이(depth)이다. 뿌리마디는 시작되는 마디로 전체 자료를 포함하며 자식마디는 하나의 마디로부터 분리되어 나간 2개 이상의 마디를 말한다. 부모마디는 주어진 마디의 상위마디이고 끝마디는 자식마디가 없는 마디를 말한다. 중간마디는 부모마디와 자식마디가 모두 있는 마디를 말하며 가지는 뿌리마디로부터 끝마디까지 연결된 마디들을 말한다. 깊이는 뿌리마디로부터 끝마디까지의 중간마디들의 수를 말한다. CART에는 출력변수가 연속형인 회귀나무(regression tree)와 범주형인 분류나무(classification tree)가 있다. CART의 형성은 성장, 가지치기, 타당성평가, 해석 및 예측으로 이루어진다. 성장단계에서는 적절한 분할 규칙과 정지규칙을 활용하는 재귀적 분할을 사용하여 나무를 성장시키는 과정이다. 가지치기에서는 오차가 큰 가지나 부적절한 추론 규칙을 가진 가지, 성능 향상에 기여하지 못하는 가지를 제거하는 단계이다. 가지치기는 과적합(overfitting)을 피하기 위하여 사용된다. 타당성평가단계에서는 이익도표나 위험도표를 활용하여 형성된 나무모형을 평가하게 된다. 해석 및 예측단계에서는 만들어진 나무모형을 해석하고 예측하는 데 사용하는 단계이다. 분할규칙이나 정지규칙에 따라서 다양한 나무모형이 형성된다(Choi & Seo, 1999; Yang & Lee, 2004). CART 모델의 장점을 보면 다음과 같다 : 1) 대용량의 데이터를 분석할 때 유용하고 모델 생성 후에도 대량의 샘플을 처리하는 것이 쉽다.; 2) 자동화하기가 쉽고 이상치의 영향을 적게 받으며 결측치를 쉽게 다룰 수 있다.; 3) 변수를 변환시킬 필요가 없고 변수 부분 선택이 자동으로 이루어진다. CART 모델의 단점은 데이터의 변화에 민감하기 때문에 데이터의 작은 변화에 완전히 다른 모형을 형성한다는 점이다(Shmueli 외, 2010).

## 4. 데이터 및 실험결과

### 4.1 데이터

데이터는 IT 업종에 종사하는 전문가들에게서 수집되었다. 총 26개의 문항이 존재하며 수집 분야는 총 5가지

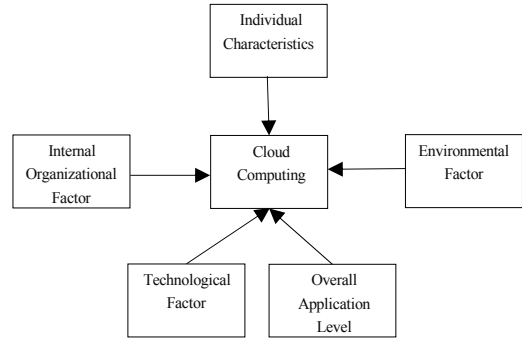


Fig. 2. Conceptual Model for Cloud Computing Adoption

로 구성되는데 여기에는 개인적 요소, 내부조직요소, 기술적 요소, 환경적 요소, 전반적인 적응수준이 속한다. 개인적 요소의 설문내용에는 최고경영층의지(Top Management Support: TMS)가 포함되며, 내부조직요소에는 공식화(Formalization: FZ), 하부구조의 유효성(Infrastructure Availability: IA), 조직의 크기(Organizational Size: OS)가 포함된다. 또한, 기술적 요소에는 기술의 복잡성(Technological Complexity: TC), 상대적 이점(Relative Advantage: RA)에 대한 설문내용이 포함되어 있으며 환경적 요소에는 경쟁자의 압력(Competitive Pressure: CP), 거래파트너의 압력(Trading Partners Pressure: TPP)에 대한 내용이 포함되며 전반적인 적응수준에서는 회사의 클라우드 컴퓨팅 채택여부가 포함된다. Fig. 2는 클라우드 컴퓨팅의 채택과 관련한 본 논문의 개념적인 모델을 보여준다.

총 65명의 응답자로부터 데이터를 획득하였으며 자료 취급을 위하여 IBM에서 제공하는 SPSS와 마이크로소프트의 엑셀을 사용하였다.

데이터의 분석을 위해서 CART 방법을 사용하였으며 사용한 소프트웨어는 MATLAB를 사용하였다. 우선 데이터를 사용하여 2단계 분석 방법을 사용하였으며 1단계에서는 CART를 활용하여 의미 있는 요인을 선택하였고 2단계에서는 선택된 요인을 이용하여 예측할 수 있는 의사결정모형을 도출하였다. 우선 25개의 변수(문항)를 독립변수로 사용하였으며 전반적인 적응수준에 관련한 한 문항을 종속변수(1: 채택, 2: 불채택)로 하였습니다. Table 1은 각 변수들과 연관된 질문들의 키워드를 보여준다. 궁극적인 목적은 회사의 클라우드 컴퓨터 채택에 영향을 미치는 요인이 무엇인지를 알고자 한다. 또한 선택된 요인을 사용하여 클라우드 컴퓨터 채택 예상 모형을 만들고자 한다. 학력사항 질문을 빼 나머지 문항은 5지선다형으로 구성되어 있다. 또한 학력사항은 대학원 레벨(고등학교,

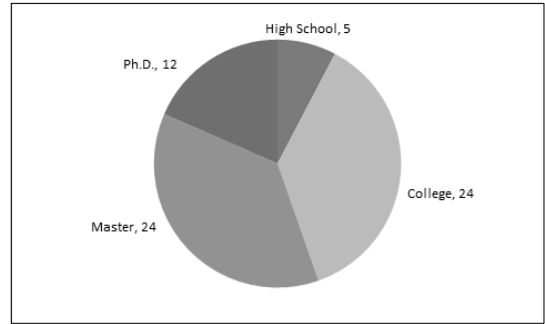
**Table 1.** Variables & Keyword of Corresponding Questionnaires

CC	Variables	Keyword of Corresponding Questionnaires
	X1	Education Level
TS	X2	Belief of Top Executives
	X3	Support of Top Management
FZ	X4	Employee's Acceptance
	X5	Rules & Regulations of Company
	X6	Existence of Formal Bureaucracy
IA	X7	Consistency of IT Infrastructure
	X8	Need for New IT Resource
OS	X9	Presence of Adequate Staff
	X10	Company Size
TC	X11	Application Difficulty of CC
	X12	Requirement of Special Skill in CC Use
	X13	Technological Complexity of CC
RA	X14	Improvement of Communication Speed
	X15	Time Efficiency
	X16	Improvement in Customer Service Level
	X17	Automatic Scalability of Storage
	X18	Accessibility of Market Information
	X19	Cost Saving
CP	X20	Positive Business Image from CC
	X21	Competition
	X22	Competitive Pressure in Industry
TPP	X23	Partner Communication Improvement
	X24	Partner Business Collaboration
	X25	CC Adoption Level of Partner
	Y	Adoption of CC

대학교, 석사, 박사)까지 포함하여 4지선다형으로 되어 있다. Fig. 3은 교육수준에 따른 응답자의 빈도를 보여준다.

#### 4.2 실험

실험은 총 4단계의 과정을 거친다. 1단계에서 데이터를 독립변수와 종속변수로 나누고 2단계에서 훈련 집합과 시험 집합으로 데이터를 분리하여 모델을 만드는데 훈련 집합을 사용하고 결정변수를 선택하는데 시험 집합을 사용하려고 한다. 3단계에서 훈련 집합과 시험 집합을 사용하여 의미 있는 결정변수를 찾고 4단계에서 선택된 결정변수 만으로 모델을 구한다.



**Fig. 3.** Respondents Frequency in Education Level

**Table 2.** Determinants & Model Selection Procedures

Step	Procedures
1	Data Division as Independent Variables and a Dependent Variable
2	Set Training and Test Sets
3	Run CART to Select Important Determinants (Variables)
4	Based on the Selected Determinants, Run CART

Table 2는 종속변수에 영향을 미치는 결정변수들을 찾고 최적의 모델을 찾는 방법을 순서대로 나타내고 있다. 보통 실험을 위해서 데이터를 훈련 집합과 시험 집합으로 분리한다. 훈련 집합을 사용하여 모델을 구축하고 주어진 모델의 정확성은 시험 집합을 이용하여 확인한다. 본 연구에서는 훈련 집합과 시험 집합이 동일하다. 이러한 모델링 과정을 통해서 설문 조사 시 다양한 질문으로 시간을 소비하기 보다는 종속변수에 영향을 미치는 독립변수 문항을 찾아 설문응답자의 지루함을 덜고 클라우드 컴퓨팅 도입여부를 결정하는 입장에서는 좀 더 핵심 모델을 구해 미리 예측할 수 있는 시스템을 만들자는 의도이다.

```

1 if X5<2 then node 2 elseif X5>=2 then node 3 else 2
2 class = 1
3 if X13<4.5 then node 4 elseif X13>=4.5 then node 5 else 2
4 class = 2
5 if X10<4.5 then node 6 elseif X10>=4.5 then node 7 else 2
6 if X9<4.5 then node 8 elseif X9>=4.5 then node 9 else 2
7 if X22<4.5 then node 10 elseif X22>=4.5 then node 11 else 2
8 class = 2
9 class = 1
10 class = 2
11 if X17<4.5 then node 12 elseif X17>=4.5 then node 13 else 1
12 class = 2
13 class = 1
    
```

위의 내용은 모든 독립변수를 사용하여 CART를 실행하여 얻은 결과이다.

ART에 사용된 변수를 확인해 보면, X1은 학력사항이며 X2, X3는 최고경영자의 지지, X4, X5, X6는 공식화, X7, X8은 하부구조의 유효성, X9, X10은 조직의 크기, X11, X12, X13은 기술의 복잡성, X14, X15, X16, X17, X18, X19는 상대적 이점, X20, X21, X22는 경쟁자의 압력, X23, X24, X25는 거래파트너의 압력을 나타낸다. 앞의 실험결과를 보면, X5, X13, X10, X9, X22, X17이 모델을 구축하는데 사용되는 것으로 나타난다. 위의 선택된 변수들을 자세히 살펴보면, X5는 회사의 규정, X13은 기술의 복잡성, X9는 직원의 수, X10은 회사의 크기, X17은 저장장치의 확장성과 컴퓨팅 능력, X22는 경쟁자의 압력을 나타내며 이러한 사항들이 클라우드 컴퓨팅을 채택여부를 나타내는 종속변수(Y)에 영향을 미치는 것으로 판명되었다.

선택된 변수들을 사용하여 CART를 실행해서 최적의 모델을 찾아보면 아래와 같다.

```

1 if X5<2 then node 2 elseif X5>=2 then node 3 else 2
2 class = 1
3 if X13<4.5 then node 4 elseif X13>=4.5 then node 5 else 2
4 class = 2
5 if X10<4.5 then node 6 elseif X10>=4.5 then node 7 else 2
6 class = 2
7 if X22<4.5 then node 8 elseif X22>=4.5 then node 9 else 2
8 class = 2
9 if X17<4.5 then node 10 elseif X17>=4.5 then node 11 else 1
10 class = 2
11 class = 1

```

앞의 결과에서 선택된 변수 중 최적의 모델에 포함되지 않은 변수는 X9으로 나타난다. 위의 결과로서 확인할 수 있는 것은 클라우드 컴퓨팅의 최적의 모델은 X5, X10, X13, X17, X22의 5개의 설문문항으로 확인되며 이 설문문항을 이용하여 기업의 채택여부를 확인 가능하며 설문데이터가 수집된다면 위의 결과에 넣어 확인해 보면 기업의 클라우드 컴퓨팅 채택여부가 판명된다.

또한 전체변수를 사용할 때의 모델과 선택변수를 사용하여 만든 최적의 모델의 정확도를 비교해 보면 전체변수 사용 시에는 86% 이고 선택변수를 사용할 시에는 85%의 결과를 보여줌으로써 전체 항목을 수집하지 않더라도 선택된 5문항으로 채택여부를 결정하는데 큰 차이를 보여주지 않는 것으로 보여준다.

## 5. 결론

본 연구에서는 기업의 클라우드 컴퓨팅을 채택하는지에 대한 여부를 판단하는 의사결정모형과 데이터 마이닝

기법 중 하나인 CART를 사용해서 패널데이터 분석을 하였다. 위의 결과들을 살펴보면 패널데이터 분석에 데이터 마이닝 방법론 중 하나인 CART가 유용함을 볼 수 있었고 클라우드 컴퓨팅 도입여부를 판별하는 데에 사용되는 데이터 수집 시 응답자에게 25개 문항을 제공하는 것보다 5개의 문항을 제공해서 빠른 답변을 요구할 수 있도록 제시하였고 두 결과의 정확도 부분에서는 큰 차이가 없다는 것을 보여주었다. 다시 말해서, 5개의 문장만으로 기업의 클라우드 컴퓨팅 도입여부를 알 수 있다. 또한 연구결과를 통해 알 수 있듯이, 기존의 패널데이터 수집에 어려움을 가진 연구자들에게 CART 요인분석 및 의사결정 모델을 제시하여 보다 정확한 데이터 수집 및 모델링 기법의 가능성을 보여주었다.

## References

- Arutyunov, V. V. (2012), "Cloud Computing: Its History of Development, Modern State, and Future Considerations", Scientific and Technical Information Processing, Vol. 39, No. 3, pp. 173-178.
- Tara, S. B., Eric, N. W., Jennifer, E. L., Emily, C. J. (2011), "Cloud Computing Adoption and Usage in Community Colleges", Behaviour & Information Technology, Vol. 30, No. 2, pp. 231-240.
- Wu, W. (2011), "Mining Significant Factors Affecting the Adoption of SaaS Using the Rough Set Approach", Journal of Systems and Software, Vol. 84, No. 3, pp. 435-441.
- Low, C., Chen, Y., Wu, M. (2011), "Understanding the Determinants of Cloud Computing Adoption", Industrial Management & Data Systems, Vol. 111, No. 7, pp. 1006-1023.
- Breiman, L., Friedman, J. H., Olshen, R. A., Stone, C. J. (1984), "Classification and Regression Trees", Belmont: Wadsworth.
- Quinlan, J. R. (1993), "C4.5 Programs for Machine Learning", Mateo: Morgan Kaufmann.
- Shmueli, G., Patel, N. R., Bruce, P. C. (2010), "Data Mining for Business Intelligence-Concepts, Techniques, and Applications in Microsoft Office Excel with XLMiner", New York: Wiley.
- Kang, W. (2012), "Recent Trends of Cloud Computing Services", Internet & Security Issue, Vol. 12, No. 6, pp. 3-30.
- Choi, J., Han, S., Kang, H., Kim, E. (1998), "Data Mining Decision Tree Analysis Using AnswerTree", Seoul: SPSS Academy.

10. Choi, J., Seo, D. (1999), "Decision Trees and Its Applications", Statistics Analysis Study, Vol. 4, No. 1, pp. 61-83.
11. Hong, Y., Kang, S., Kim, S. (2012), "Evaluation of the Usability of Local Web Portal Cloud Storage Services - Focused on Naver NCloud and Daum Cloud", Journal of Digital Design, Vol. 12, No. 1, pp. 165-175.
12. Kim, D., Lee, J., Park, Y. (2011), "A Study of Factors Affecting the Adoption of Cloud Computing", Journal of Society for e-Business Studies, Vol. 17, No. 1, pp. 111-136.
13. Rim, S., Kong, D., Shim, S., Han, Y. (2012), "A Study on the Significant Factors Affecting the Adoption of Enterprise Cloud Computing", Journal of the Korea Society of IT Services, Vol. 11, No. 1, pp. 173-196.
14. Kim, S. Kim, G. (2011), "An Empirical Study on the Factors Affecting the Adoption of Mobile Cloud and the Moderating Effect of Mobile Trust", The e-Business Studies, Vol. 12, No. 1, pp. 281-310.
15. Kim, S., Kim, S., Oh, S. (2012), "A Study on Librarians' Perspectives of Cloud Computing Service", Journal of Korean Library and Information Science Society, Vol. 43, No. 1, pp. 241-259.
16. Woo, H., Shim, J., Lee, J. (2011), "A Study on Project Performance in Cloud Computing : Focus on User Experience of GoogleDocs", Journal of Society for e-Business Studies, Vol. 16, No. 1, pp. 71-100.
17. Lim, M., Kim, M., Chun, Y. (2012), "Music Listening Behavior in the Cloud", HCI 2012, Alpensia Resort, pp. 1034-1036.
18. Barnatt, C. (2010), "A Brief Guide to Cloud Computing", London: Robison.
19. Shim, Y. (2009), "Technology Trends and Virtualization based Management Techniques for Cloud Computing", KNOM Review, Vol. 12, No. 1, pp. 20-32.
20. Yang, S., Lee, J. (2004), "Improved Segmentation Method of Decision Tree Using Fuzzy Theory", Proceedings of 2014 Korean Institute of Industrial Engineers, pp. 44-51.



**백 승 현** (sbaek4@hanyang.ac.kr)

2000 명지대학교 산업공학과 공학사  
 2002 미국 조지아공과대학교 산업공학 공학석사  
 2010 미국 테네시대학교 산업공학 공학박사  
 2010~2012 삼성전자 반도체연구소 책임연구원  
 2012~현재 한양대학교 ERICA 캠퍼스 경영학부 조교수

관심분야 : 데이터마이닝, Big Data Analytics, 품질경영, SCM(공급사슬관리), 생산관리, CRM(고객관계관리), 가치투자분석, 부동산 투자 및 상권 분석



**장 병 운** (bychang@ajou.ac.kr)

1996 성균관대학교 산업공학과 학사  
 2000 Georgia Tech. Operations Research 석사  
 2002 Georgia Tech. Applied Statistics 석사  
 2004 Georgia Tech. Industrial and Systems Engineering 박사  
 2004~2006 Georgia Tech. Post Doc.  
 2006~2009 KT 네트워크 연구소 선임연구원  
 2009~현재 아주대학교 경영대학 경영학과 부교수

관심분야 : SCM, 운영관리, 정보통신경영, BPM, OR/OM, Simulation, Applied Statistics