

의사결정나무법을 이용한 귀납적 학습방법에 의한 정보시스템 수용자 세분화

Identifying Early Adopters of Information Systems by Inductive Learning Using Decision Tree Method

이 민 수 (Minsoo Lee)

최 영 찬 (Young Chan Choe)

유 병 준 (Byungjoon Yoo)

농촌진흥청 농촌자원개발연구소 연구원

서울대학교 농업생명과학대학 농경제사회학부 교수

고려대학교 경영대학 경영학과 교수, 교신저자

요 약

한 정보시스템을 소개 및 전파 시에, 만약 정보시스템의 공급자가 조기 수용성향을 가진 사용자 집단을 식별해낼 수 있다면, 그 집단에 대한 집중적인 권장 및 홍보를 통하여 더욱 효율적으로 시스템을 전파할 수 있을 것이다. 또한, 도입 이후 그 조기수용자들은 타 수용대상 집단들에 대한 시스템의 자발적 홍보와 수용 시의 도움들을 통해 시스템 도입을 더욱 용이하게 하는 효과를 가져오게 할 수 있다.

이와 같은 조기 수용자군 식별 목적을 위하여, 본 논문은 기존 주로 사용되어 온 이산선택기법 대신, 의사결정나무를 이용한 식별기법을 제안한다. 기존의 이산선택기법에 비하여 이 기법은 집단 예측에 있어서 더욱 정확하며, 비선형적으로 세분화된 그룹을 식별할 수 있는 장점도 가지고 있다. 이와 같은 효과의 검증을 위하여, 본 논문은 농업분야의 한 정보시스템 도입시의 실증데이터를 이용하여 목표집단 식별상의 의사결정나무기법의 우수성을 입증하였다. 본 기법은 앞으로 조기 수용자 집단 식별목적에 실질적으로 시스템 개발 공급자에게 이용될 수 있을 기법으로 기대된다.

키워드 : 의사결정나무, 정보시스템 수용, 세분화

I. 서 론

Rogers(2003)에 의하면 혁신전파의 과정의 5단계(knowledge > persuasion > decision > implementation > confirmation) 중 2번째 단계인 설득 단계(persuasion stage)에서 개인은 혁신기술이나 아이디어에 대해 우호적이거나 비우호적인 태도를 형성한다. 개인은 혁신의 기대된 성과에 대한 불

확실성을 줄이기 위해 혁신에 대한 정보나 메시지를 찾는다. 따라서 이 단계에서 개인에게 적절한 정보와 메시지를 제공하여 혁신에 우호적인 태도를 갖도록 하는 변화 촉진자(change agent)의 역할과 설득을 위한 효과적인 커뮤니케이션 전략이 매우 중요하다.

다수 대중을 대상으로 한 효과적인 커뮤니케이션은 다양한 채널을 통해 차별화된 메시지를 전

달하기 위해 응답자(receiver)를 유사한 집단으로 세분화하는 전략에 항상 의존해왔다(Dawson, 1989, Frankenberger and Sukhdial, 1994). 사회적, 경제적, 상황적 특징에 대한 분석을 기반으로 한 응답자에 대한 지식은 설득과 혁신 적응과정에 가장 중요한 요소이며[Albrecht and Bryant, 1996], 세분화는 커뮤니케이션을 더 효과적이고 효율적으로 만들며, 이를 통해 특정 문제에 대한 관심과 지식의 수준을 변화시키고, 태도를 변화시켜, 궁극적으로는 응답자의 행동을 변경시킨다(Rogers, 2003).

혁신기술을 조기에 수용하는 사용자의 특성을 규명하고, 조기수용 사용자들을 대상으로 혁신기술을 조기에 보급한다는 측면에서 보면, 혁신성을 가진 사용자의 특성을 구분하는 연구도 세분화의 일종으로 볼 수 있다. 혁신수용 연구를 포괄적으로 검토한 Feder and Umali(1993)의 연구를 보면 대부분 로짓(logit)과 프로빗(probit)모형을 이용하여 혁신적인 사용자의 사회경제적인 특성을 규명하는 연구였다. 이는 혁신성을 가진 사용자의 특성을 규명한다는 측면에서 보면 적절하지만, 동질적인 다양한 집단으로 구분하여 커뮤니케이션의 효과와 비용효과적인 전략을 개발하고자 하는 세분화의 관점에서는 부적절하다. 왜냐하면 로짓(logit)과 프로빗(probit) 분석은 기술수용가능성이 높은 사용자를 다양한 동질적인 집단으로 구분할 수는 없기 때문이다.

현재 마케팅 분야에서 시장 세분화는 가장 핵심적인 전략이다(Levin and Zahavi, 2001). 마케팅 분야에서 개발된 세분화 기법은 보건, 환경 등에 대한 공공의 태도와 행동을 변화시키기 위한 사회적 마케팅(social marketing)에도 활발히 적용되고 있다(Albrecht and Bryant, 1996). 예를 들면, 보건 분야에서는 세분화를 통해 다양한 건강증진 프로그램과 질병예방 프로그램을 개발하여 대중들에게 제공함으로써 비용을 절감과 함께 커뮤니케이션 효과가 증대되고 있다(Brown, 1992).

이런 경향을 볼 때, 마케팅 분야에서 개발된 세분화 방법을 사용하여 기술수용 가능성이 있

는 사용자 집단을 다양하게 세분화한 후, 각 집단에 적절한 기술보급 전략을 수립하면 더욱 효과적일 것이다. 정보기술의 수용 및 확산에 있어서 영향을 미치는 요인을 파악하고, 분석하는 것은 경영정보 연구 분야에서도 매우 중요한 연구 주제이다(양희동, 문윤지 2005, 이동만, 장명희, 유지영 2003).

정보기술의 보급을 촉진시키는 데 있어서 공공부문의 투자는 매우 결정적인 역할을 한다. 그러나 현재 정보기술을 보급하기 위한 정부 지원 예산은 매년 줄어들고 있다(박덕병, 이민수, 2002). 따라서 제한된 예산 하에서 정보기술을 국가차원에서 효율적으로 보급하기 위해서는 세분화에 기초한 새로운 기술보급전략이 필요하다.

본 연구의 목적은 정보시스템 수용가능성이 높은 사용자 집단을 세분화하여, 각 집단의 특성을 고려한 기술보급 전략을 제시하는 것이다. 최근 마케팅 분야에서는 데이터에 기반하여 자동으로 고객을 세분화하고, 특정 제품을 특정고객에게 판매하기 위한 표적화 의사결정(targeting decisions)에 의사결정나무(decision tree)를 사용하여 큰 효과를 보고 있다(Levin and Zahavi, 2001). 정보시스템 분야 연구에서도 의사결정나무를 이용한 고객 세분화 방법론의 효율성이 제시되고 있다(장남식, 2005). 본 연구는 의사결정나무 방법의 효율성을 평가하기 위하여, 국내 전업양돈 농가를 대상으로 이루어진 농림부의 설문조사를 이용하여 정보시스템 수용가능성이 높은 농가를 세분화해보고, 그 예측력을 실증 분석하고 있다.

다음 장에서는 고객세분화에 사용되는 변인과 기술수용연구에서 사용되는 변인이 어떤 차이가 있는지를 검토하고, 본 연구에서 사용되는 의사결정나무기법이 고객 세분화 기법에서 차지하는 위치에 대해 논의한다. 3장에서는 의사결정나무를 소개하고, 4장에서는 연구방법을 제시한다. 5장에서는 연구방법에 따른 연구결과를 제시하고, 6장에서는 연구결과를 토대로 세분화된 집단에 대한 기술보급 전략을 제시하며, 그

제시에 있어 볼 수 있는 의사결정나무 방법론의 장점을 분석해 본다.

II. 시장세분화

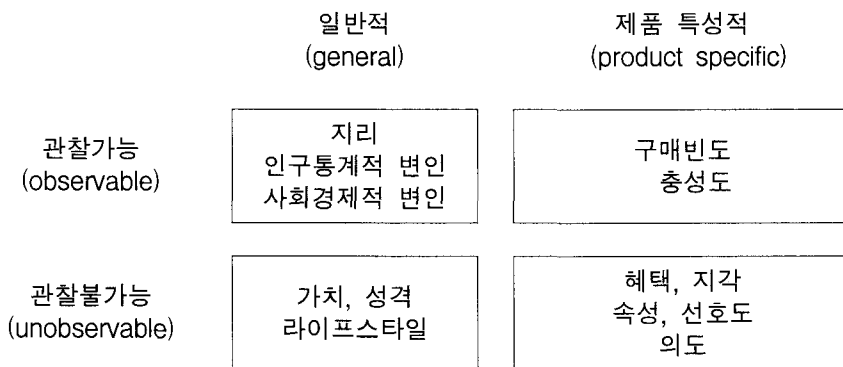
2.1 시장세분화 변인과 기술수용 변인과의 관계

고객들은 동일하지 않고, 선호와 욕구 등이 다르다. 따라서 구매행동이 유사할 것으로 기대되는 비슷한 요구와 특징을 가진 집단으로 고객으로 분류하는 것이 시장세분화이다. 이상적으로는 세분화를 위해서 제품과 서비스에 대한 고객의 태도가 직접 반영되는 변수를 사용해야 하지만, 이는 대부분 사전에 알려져 있지 않다. 따라서 고객의 구매습관과 행동을 가장 잘 반영하는 데이터를 대리변수로 사용한다. 세분화를 위해 사용되는 변인을 Weinstein(1994)는 5가지 차원으로 나누었다. 첫 번째 차원은 지리적 차원이다. 지역, 도시의 규모, 기후 등의 변인이 이 차원에 속한다. 두 번째 차원은 사회경제학적 차원이다. 소득, 교육 등의 변인이 이 차원에 속한다. 세 번째는 심리학적 차원이다. 사회적 계층, 라이프 스타일, 개성 등이 이 차원에 속한다. 네 번째는 제품사용 차원이다. 사용률, 사용자 상태, 상표 충성도 등이 이 차원에 속한다. 다섯 번째는 편

의 차원이다. 제품에 대한 문제해결, 제공되는 구체적인 편익 등의 변인이 이 차원에 속한다.

Wedel and Kamakura(1998)는 변인성격과 분석 단위에 의해 시장세분화 변인을 구분하였다. 즉 분석단위 일반적인 제품영역으로 할 것인지, 특정 제품만을 대상으로 할 것인지를 한 축으로 하고, 관찰 가능한 변수를 사용했는지와 그렇지 않은지를 한 축으로 하여 4개의 차원으로 구분하였다(<그림 1>). 첫 번째 차원은 일반적 관찰가능(general, observable)한 차원이다. 지리적 변인, 인구통계학적 변인, 사회경제적 변인 등이 이 차원에 속한다. 두 번째 차원은 일반적 관찰불가능(general, unobservable)한 차원이다. 개인의 가치, 성격, 라이프스타일이 이 차원에 속한다. 세 번째 차원은 제품특성적 관찰가능(product specific, observable)한 차원이다. 사용자의 구매빈도, 충성도 등이 이 차원에 속한다. 네 번째 차원은 제품특성적 관찰불가능(product specific, unobservable)한 차원이다. 제품에 대한 지각, 제품 속성, 선호도 등이 이 차원에 속한다.

농업기술 수용연구에서는 혁신성을 가진 농가를 규명하기 위해서는 다양한 차원의 개인차를 드러내는 변수를 모형에 포함하였으며, 이 경우 어떤 변수를 택할 것인가가 근본적 문제였다(Bhattacharyya et al., 1997). 농업기술 수용에 영향을 미치는 요인에 대한 연구에서 혁신기술의 성격과 연구자



<그림 1> 시장세분화 변인(Wedel and Kamakura, 1998:7)

에 따라 다른 변수들을 사용하였다(Feder, *et al.*, 1985, Feder and Umali, 1993).

Kagwanja(2001)는 변인들을 4가지 차원 경제적, 기관적, 개인적, 물리적으로 요인을 구분하였다. 그는 경제적 요인(economic factors)들로는 소득(income), 신용(access to credit), 노동력(labor), 토지 소유권(land tenure) 등을 제시하였으며, 기관적 요인 (institution factors)들로는 지도소에 대한 접근성정도, 기술 훈련 프로그램 참여정도 등을 제시하였다. 개인적 요인(personal factors)로는 영농 경력, 연령, 교육 등을 제시하였으며, 물리적 요인(physical factors)으로는 토지비옥도, 토지오염도 등을 제시하였다. Negatu and Parikh(1999)는 기존의 연구들이 사용한 변수는 4가지, 즉 자연적인 자원(resource) 차원, 사회경제적(socio economic) 차원, 인구학적(demographic) 차원, 지도기관에 대한 접근 차원으로 분류 가능하다고 주장하였다. Shrestha(1993)은 경제적(economic) 차원, 물리적(physical) 차원, 위치적(location) 차원으로 구분하였다. Bhattacharyya *et al.*(1997)은 인적 요인(human factors)과 생산요인(product factors)으로 구분하였다.

이처럼 기술수용에 사용된 변수들을 적절한 차원으로 구분하는 것은 기술과 연구자마다 다르다. 그러나 기존의 기술수용 문헌들을 검토하여 중요 변인을 도출한 Feder(1985)의 연구와 Foltz(2003)의 연구에 의하면 농업경영자의 특성(연령, 교육 수준, 태도, 혁신성 등)과 농가 특성(영농규모, 영농형태, 부채비율 등)으로 구분할 수 있다.

마케팅분야의 시장세분화 변인과 농가의 기술 수용 연구에서의 변인을 비교해보면, 농가의 기술 수용 연구변인은 Weinstein(1994)이 제시한 차원 중 사회경제학적 차원과 심리학적 차원과 밀접하게 관련되어 있으며, Wedel and Kamakura(1998)의 마케팅 분야의 일반적(general) 차원과 관련되어 있음을 알 수 있다.

Wedel and Kamakura(1998)은 인구통계적 변인이나 사회경제적 변인과 같은 관찰 가능한 변인만을 사용할 경우에는 개인이 어느 집단에 속하

는 지에 대해 알 수 있지만, 대상자에게 구체적으로 어떤 마케팅 프로모션(promotion)을 할 것인가가 불명확하다고 지적하였다. 예를 들면 30대 고학력의 여성이 구매량이 많다는 것을 세분화를 통해 분석했지만, 이 여성들이 추구하는 가치를 모른다면 어떤 메시지로 프로모션해야 할지 알 수 없다. 따라서 향후 농가를 대상으로 좀 더 정밀하게 기술보급사업 전략을 수립하기 위해서는 관찰불가능한 변수를 고려할 필요가 있다.

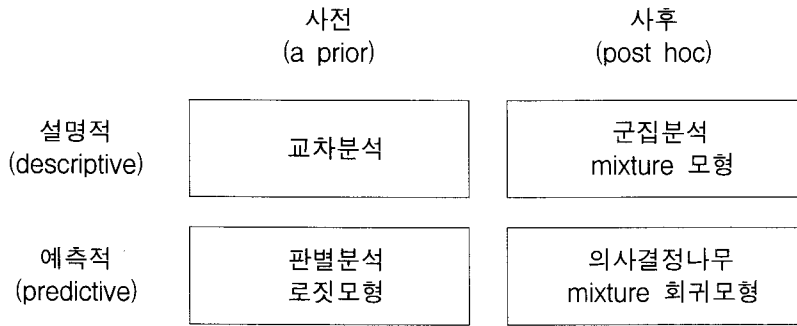
2.2 시장세분화 방법

Wedel and Kamakura(1998)은 시장세분화 방법을 시장세분화가 이루어지는 시기와 예측가능성을 고려하여 4가지 차원으로 구분하였다. 시장세분화가 이루어지는 시기와 관련된 시장세분화 방법은 세분화를 미리 할 것인가 혹은 나중에 할 것인가를 기준으로 사전세분화(a priori segmentation)와 사후세분화(post hoc segmentation)로 구분하는 것이다. 사전세분화(a priori segmentation)란 시장에 대한 조사를 실시하기에 앞서 세분화를 위한 적절한 기준을 미리 결정하는 방법이다. 이에 반해 사후 세분화(post hoc segmentation)란 실행된 시장조사의 결과에 근거하여 소비자를 세분집단으로 구분하는 방법을 말한다.

예측가능성과 관련된 시장세분화 방법은 목표(종속)변수의 존재여부에 따라 설명적 방법(descriptive methods)과 예측적 방법(predictive)으로 구분된다. 설명적 방법 (descriptive methods)은 변수들을 독립변인과 종속변인으로 구분하지 않고 분석하는 방법이다. 예측적 방법(predictive)은 일반적으로 하나의 변인을 종속변수로 설정하고 나머지 변인을 독립변인으로 설정하여 분석하는 방법이다.

4가지 차원에 따른 대표적 분석방법은 <그림 2>와 같다.

사전적(a prior) 방법은 연구자의 주관에 의한 오류가능성이 존재하며, 종속변인으로 사전에



〈그림 2〉 시장세분화 방법(Wedel and Kamakura, 1998:18)

분류된 집단이 세분화에 의미가 없을 가능성이 있다. 그러나 사전조사나 경험을 통하여 시장구분을 위한 가장 핵심이 되는 기준을 미리 알고 있다면 사전세분화방법이 더욱 유용하다(Wedel and Kamakura, 1998). 그러나 새로운 신제품에 대한 세분화의 경우에는 사후세분화방법이 시장구분을 위한 핵심기준을 결정하는데 더 유용할 가능성이 크다(McGlone and Calantone, 1992).

Wedel and Kamakura(1998)의 구분에 의하면 의사결정나무는 사후 예측적(post hoc predictive) 방법이다. 의사결정나무는 로짓모형이나 판별분석과 동일하게 종속변수를 가장 잘 예측하는 독립변인을 투입하여 모형을 생성한다. 그러나 로짓모형과 판별분석의 경우에는 분석 후에도 사전에 설정한 종속변인으로 설정된 집단의 수로 고정되는 반면에 의사결정나무는 분석 후에 다양한 하위 집단이 생성된다는 측면에서 사후적(post hoc) 방법에 속한다. 따라서 의사결정나무는 새로운 혁신기술 수용에 대해 세분화하고자 할 경우에는 더 유용한 결과를 제공할 수 있다.

III. 의사결정나무(decision tree)

3.1 의사결정나무 알고리즘

데이터에 기반해 자동으로 종속변수의 값을 동질적인 집단으로 분리하기 위한 여러 의사결정나무

알고리즘이 개발되었다. 대표적인 알고리즘으로는 AID(Automatic Interaction Detection; Sonquist, Baker and Morgan, 1971), CHAID(Chi square AID; Kass, 1980), CART(Classification and Regression Trees; Breiman, Friedman, Olshen, and Stone, 1984), ID3(Quinlan, 1986), C4.5(Quinlan, 1993) 등이 있다(Murthy, 1998).

위의 의사결정나무 알고리즘은 공통적인 구조를 가지고 있다. 전체데이터를 가진 뿌리마디(root node)로 부터 시작해 체계적인 방법으로 가지들(branches)과 잎들(leaves)을 가진 나무(tree)를 생성한다. 각 알고리즘은 부모마디(father node)로부터 분리기준(splitting criteria)에 근거하여 자식마디들(children nodes)을 생성한다. 분리된 자식마디는 자신이 부모마디가 되어 위의 과정을 재귀적으로 반복한다. 분리된 자식마디는 종료규칙(termination rules)에 의해 어떤 마디들은 더 이상 분리가 되지 않고 끝마디(terminal node), 즉 잎(leaf)이 되며, 나머지 자식마디는 자신이 부모마디가 되어 위의 과정을 재귀적으로 반복한다. 이 과정은 모든 노드가 끝마디가 되었을 때 멈추게 된다.

의사결정나무 알고리즘 개발에서 주요하게 고려해야 할 3가지 사항은 다음과 같다. 첫 번째 고려사항은 나무의 생성(growing the tree)이다. 나무의 생성은 전체데이터를 가진 뿌리마디(root node)를 가지들(branches)과 잎들(leaves)을 가진 나무

(tree)를 생성하는 과정이다. 하나의 마디는 한번에 여러 변수들에 의해 분리될 수도 있고(예를 들면 학력과 연령의 조합), 변수들의 함수에 의해서도 분리될 수 있다(예를 들면 학력과 소득을 표준화하여 더하는 함수). 독립변수들이 많을 경우 마디를 분리하는 방법은 무한하다. 또 독립변수가 소수일 경우에도 연속형 변수가 포함되면 분리방법은 무한해진다. 따라서 대부분의 알고리즘은 위의 문제를 피하기 위해 다음과 같은 규칙을 적용한다(Levin and Zahavi, 2001).

- i) 모든 연속형 변수는 나무의 생성 이전에 일정한 수의 범위로 나누어 범주화한다.
- ii) 마디는 단지 한번에 하나의 변수에 의해서만 분리한다..
- iii) 부모마디에서 생성되는 자식마디는 2~3개로 제한한다.
- iv) 분리는 욕심장이기법(greedy algorithm)에 기초에 이루어진다. 즉 분리는 순차적으로 매 단계에서 가장 좋은 경우를 선택하며, 이전의 결정을 다시 고려하지 않는다.

두 번째는 최적분리(the best split)의 결정이다. 어떤 변수를 선택하고 이 변수를 어떻게 분리하는가에 따라 마디는 매우 다양하게 분리된다. 이때 어떻게 분리하는 것인 최적인가가 문제가 된다. 분리기준(splitting criteria)은 알고리즘마다 다르며, 어느 알고리즘이 더 우수한지를 판단할 수 있는 하나의 기준은 없다. 대표적인 분리기준은 엔트로피(entropy), 지니(GINI), 카이스퀘어 검정이다.

세 번째 고려사항은 종료규칙(termination rules)이다. 이론적으로 의사결정나무는 끝마디가 모두 동일한 값을 갖도록 분리할 수 있으며, 단 하나의 관측치만 가지는 마디로 분리할 수도 있다. 그러나 이렇게 생성된 나무는 통계적으로 유의미한 동질의 집단을 분리하고자 의사결정나무 분석의 목적에 적합하지 않게 된다. 또 위의 경

우처럼 너무 큰 나무를 생성할 경우에는 훈련데이터는 매우 잘 판별하지만, 훈련에 사용되지 않은 새로운 데이터는 거의 판별하지 못하는 과도적합(overfitting)의 문제가 발생한다(Levin and Zahavi, 2001). 따라서 나무의 생성이 적절하게 멈추도록 하는 규칙을 적용하여 나무의 크기를 적절하게 만들 필요가 있다. 이 단계에서 설정되어야 하는 종료규칙으로는 i) 분리에 필요한 최소/최대 관측치 개수, ii) 자식마디로 분리될 마디의 최소 관측치 개수, iii) 나무의 깊이(depth of tree), iv) 분리에 적용할 유의수준 이다.

3.2 의사결정나무 접근법의 장단점

의사결정나무 모형은 연속형 데이터를 처리하는 능력이 로짓분석과 같은 통계적 기법에 비해 떨어지며, 모형구축에 사용되는 표본의 크기에 민감하다는 단점이 있다. 그러나 의사결정나무 접근방법은 기존에 주로 사용되어 왔던 요인분석과 군집분석을 통한 시장세분화에 비해 분류와 예측의 측면에서 강력한 힘을 가지고 있으며, 최근에는 다양한 분야의 연구에서 응용되고 있다(Berry & Linoff, 2003). 의사결정나무 접근법이 시장세분화에서 타 접근법에 비해 가지는 장점은 다음과 같다.

첫째, 의사결정나무 접근법은 연구 기준(criterion of interest)에 유의한 변수를 토대로 시장을 세분화할 수 있다. 전통적으로 시장세분화에 가장 자주 사용된 군집분석에 의한 시장세분화 방법은 연구자의 주관에 의한 오류가능성이 존재한다. 따라서 세분화에 유의하지 않는 변수가 사용될 가능성이 크며, 이 경우 최적의 동질적 군집화가 불가능하다. 군집분석에서는 특정한 목표나 기준에 의해 군집들이 분류되는 것이 아니라, 관측치들의 거리에 따라 분류된다. 따라서 변수의 투입이나 제거에 따라 다른 결과가 나타날 가능성이 크며, 이 경우 어느 변수를 통해 분류된 군집이 더 옳은지를 판단할 수 있는 근거

가 없다. 군집분석을 이용한 전통적인 접근방법은 시장세분화에서 널리 사용되고 있기는 하지만, 서술적 방법, 혹은 객관적 기준이 없는 방법(non criterion method)이라고 비판받고 있다(Chen 2003). 그러나 의사결정나무 접근법의 경우에는 모형구축과정에서 분류에 영향을 미치지 않는 변수들은 자동으로 제외되고 종속변수를 가장 잘 예측하는 변수들에 의해 세분화가 이루어진다.

둘째, 의사결정나무 접근법은 연구 기준(criterion of interest)에 의해 분류하는 예측적 시장세분화 접근법인 로짓모형과 판별분석에 비해 여러 가지 통계적 가정으로부터 자유로울 수 있다. 판별분석, 로짓모형 등은 기본적으로 입력변수 및 모형에 대한 특성 및 자료 구조를 가정하고 있다. 예를 들어 판별분석의 경우 i) 각 집단은 다변량 정규분포(multi variate normal distribution)이며, ii) 각 집단의 분산 공분산 행렬 (variance covariance matrices)은 동일하며, iii) 사전확률을 알 수 있다 라는 세 가지 가정을 가지고 있다. 그러나, 많은 경우 모형에 사용되는 변인은 이를 충족시키기 어렵다. 이에 비해, 의사결정나무 모형은 독립변수간의 상관관계, 변수의 분포에 대해 특별한 가정을 하고 있지 않으므로, 예측 모형의 성과를 향상시킬 수 있고 모형의 일반화(generalization)에도 장점을 가지고 있다.

셋째, 의사결정나무 접근법은 연구 기준(criterion of interest)에 의해 분류하는 예측적 시장세분화 접근법인 로짓모형과 판별분석에 비해 좀더 세밀하게 집단을 세분화하는 것이 가능하다. 왜냐하면, 로짓모형과 판별분석의 경우에는 분석 후에도 사전에 설정한 종속변인으로 설정된 집단의 수로 고정되는 반면에 의사결정나무는 분석 후에 다양한 하위 집단이 생성되기 때문이다.

넷째, 의사결정나무 접근법은 간단한 분류 로직에 의해 세분집단이 표현되므로, 분류나 예측의 근거를 이해하기 쉬우며, 특정 세분집단의 분류에 영향을 미친 변인을 쉽게 파악할 수 있다는 장점이 있다.

IV. 연구방법

4.1 자료수집

본 연구는 2001년 농림부와 대한양돈협회가 실시한 ‘전국 전업규모 양돈농가 경영실태 조사서’의 설문결과를 분석에 사용하였다. 이 조사는 2001년 9월 1일을 기준으로 500두 이상의 규모화된 양돈농가를 대상으로 조사하였으며, 최종적으로 4,976농가의 데이터가 회수되었다. 조사내용은 농장 경영주의 일반사항(성별, 연령, 학력 등), 농장의 일반사항(돈사면적, 사육두수, 출하두수 등), 농장성적(평균 산자두수, 이유두수, PSY 등), 경영실태(소득, 부채, 사료구입량, 생산비 등), 정보화현황(컴퓨터보유여부, 컴퓨터사용목적, 사용자 등), 기타(애로사항, 건의사항 등) 등으로 총 45문항으로 구성되어 있다. 본 연구에서는 총 4,976농가 중 무응답치가 많은 208농가를 제외한 4,768농가의 데이터를 분석에 이용하였다.

4.2 변수선정

4.2.1 목표 변인

농업정보시스템의 수용여부가 목표 변인이다. 인터넷이나 컴퓨터를 이용해 축산정보를 수집하거나 축산경영관리에 컴퓨터를 사용하면 이 값은 1이 되고, 그렇지 않으면 0이 된다.

4.2.2 입력 변인

정보시스템 수용과 관련된 입력변인을 도출하기 위해서 정보시스템수용과 관련된 선행연구를 토대로 변인을 선정하였다. 이를 통해 선정된 변인은 개인적 특성차원에서는 연령, 양돈종사기간, 교육수준, 겸업여부, 사육규모확대정도, 보험가입정도가 선정되었다. 농가특성 차원에서는 사육규모, 판매두수, 노동력, 농가형태, 사료정산방법, 축사표준설계적요, 인공수정여부, 가축 방역 정도가 선정되었다.

외부지원 차원에서는 가입단체수가 선정되었다. 선정된 변인에 대한 자세한 설명은 <표 1>에 있다.

<표 1> 변인설명

차원	변인명	변인설명		
개인적 특성	Age	연령(세)		
	JobLength	양돈종사기간 (1=1년이내, 2=~2년, 3=~3년, 4=~5년, 5=~10년, 6=~20년, 7=20년이상)		
	Education	교육수준 (1=무학, 2=초졸, 3=중졸, 4=고졸, 5=전문대졸, 6=대학이상)		
	OJobSale	축산물판매업 (0=미겸업, 1=겸업)		
			OJobProcessing	축산물가공업 (0=미겸업, 1=겸업)
	Enlarge	사육규모확대정도(%) (최저는 -100%, 최고는 제한없음)		
	독립변인	Insurance	보험가입 (1=가축·건물가입, 2=가축가입, 3=건물가입, 4=미가입)	
		TotalPigs	사육규모(두)	
		SalesPigs	판매두수(두)	
Labors		노동력 (1=가족, 2=1인고용, 3=2~3인, 4=4인이상, 5=공동운영)		
			FarmType	농가형태 (1=일괄사육, 2=자돈생산, 3=자돈구입비육돈생산, 4=비육돈위탁)
농가 특성		FeedAccount	사료정산방법 (1=선금, 2=현금, 3=1달이내상환, 4=2달이내상환, 5=3달이내상환, 6=3달초과상환)	
				Pigpen
		AI	인공수정여부 (1=외부AI센터이용, 2=자가AI, 3=미실시)	
				Epidemic
외부 지원		JoinGroups	가입단체수 (0=미가입, 1=1개, 2=2개이상)	
	종속변인		농업정보시스템수용여부 (0=미수용, 1=수용)	

4.3 자료의 전처리(preprocessing)

데이터를 본격적으로 분석하기 전에 이상치(outlier)의 제거, 변수의 정규성 검정, 결측치 처리 등의 작업이 필요하다. 특히 머신러닝 연구자들의 경우는 대부분 예측결과의 정확성을 높이기 위해서 분석 전에 다음과 같은 데이터 전처리 과정을 거친다.

- i) 결측치 처리
- ii) 이상치(outlier) 처리
- iii) 데이터분할 : 훈련데이터와 유효성검정데이터의 분리
- iv) 데이터변화 : 로그함수 등을 사용하여 원래 변수를 정규분포에 가까운 변수로 변환

본 연구에서는 사용된 데이터는 무응답이 많거나, 잘못된 값으로 추정되는 설문지가 사전에 제외된 것이다. 따라서 이상치 제거과정은 거치지 않았다. 또 의사결정나무의 경우에는 비모수적 방법으로 분포에 민감하지 않으므로 데이터 변환 과정도 거치지 않았다. 의사결정나무에서는 결측치도 나무의 생성시 하나의 정보로 받아들여이므로 결측치도 처리되지 않았다.

의사결정나무에서 가지치기와 모형의 평가를 위해서 데이터를 두 부분 훈련데이터(training data set)와 유효성검정데이터(validation data set)으로 나누었다. 4,768개의 데이터 중 67%인 3,195개의 데이터를 무작위(random)로 추출한 후 이 데이터를 의사결정나무의 훈련데이터(training data set)로 사용하였다. 나머지 33%(1,573개)의 데이터는 의사결정나무의 유효성검정데이터(validation data set)와 모형을 평가하기 위한 데이터로 사용하였다.

4.4 규칙 설정

4.4.1 분리기준(splitting criteria)

본 연구에서는 최적분리기준으로 카이스퀘어

검정을 사용하였다. 이 분리기준은 부모마디로부터 분리되는 자식마디들이 최대한 서로 다르도록 만드는 것이다. 이 검정을 위해 사용되는 통계량은 아래와 같다.

$$Y = \sum_{splits} \frac{Y_i - \bar{Y}}{\bar{Y}} \quad (식 1)$$

이 통계량은 분포를 가진다. 이 값은 자식마디들이 서로 다를수록 커진다. 분포는 자유도에 따라 달라지므로, 최적분리는 검정을 통해 계산된 p값(p-value)을 통해 결정한다. 즉 가장 적은 p값을 갖는 분리(split)를 선택한다.

4.4.2 종료규칙 설정

하나의 마디에 너무 많은 관측치가 있거나 너무 적은 관측치가 있을 경우에는 오류가 증가한다. 만약 마디가 너무 적은 관측치를 가지고 있을 경우에는 통계적인 유의성의 부족으로 오류가 발생할 가능성이 많아진다. 또 너무 많은 관측치를 가질 경우에는 더 좋은 마디가 제거되었을 가능성(1종 오류)과, 더 나쁜 마디가 생성되었을 가능성(2종 오류)이 존재한다(Levin and Zahavi, 2001). 따라서 마디에 포함될 관측치의 개수에 대한 설정이 필요하다. 이와 함께 설정에 따라 최종 나무의 크기가 작을 수도 있고 클 수도 있다. 만약 너무 작은 나무가 생성되면 유용성이 떨어지며, 너무 많은 나무는 해석이 어려운 문제가 있다(Levin and Zahavi, 2001). 본 연구에서는 설정한 종료규칙으로 아래와 같다.

- i) 분리된 마디는 최소한 30개 이상의 관측치를 가져야 한다.
- ii) 마디의 관측치가 100개 이하일 경우에는 더 이상 분리하지 않는다.
- iii) 나무의 깊이(depth of tree)의 최대값은 6으로 설정하였다.
- iv) 검정시 p값이 0.1 이상일 경우에는 더 이상

분리하지 않는다.

4.4.3 가지치기(pruning)

종료규칙을 토대로 생성된 의사결정나무의 경우도 훈련에 사용되지 않은 새로운 데이터는 잘 판별하지 못하는 과도적합(overfitting)의 문제가 존재할 수 있다. 따라서 유효성검정데이터(validation data)를 이용하여 가지치기(Pruning)을 실시한다. 가지치기 방법은 다음과 같다.

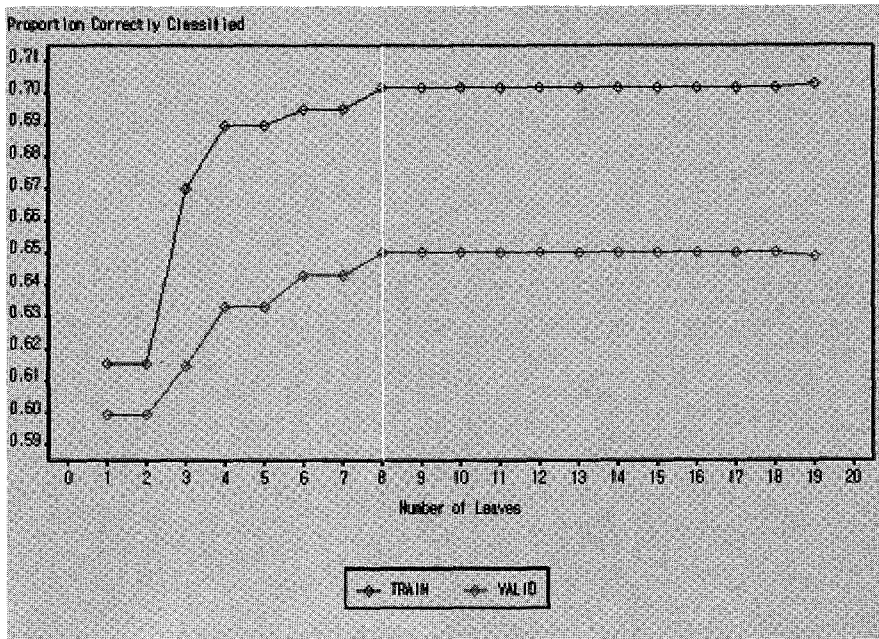
- i) 종료규칙에 의해 최종 의사결정나무를 생성한다.
- ii) 최종 의사결정나무로부터 하위 의사결정나무들(subtrees)을 생성한다. 하위 의사결정나무는 최종 의사결정나무의 잎(leaf) 수를 순차적으로 하나씩 줄여가면서 만든다. 이 경우 동일한 잎 수를 가진 여러 하위 의사결정나무가 나타날 수 있다. 이 경우 유효성검정 데이터를 적용하여 가장 잘 분류하는 하나의 나무만 선택한다.

- iii) 잎 수가 하나인 하위 의사결정나무로부터 시작하여 최종 의사결정나무까지 유효성검정 데이터를 적용하여 분류정확도를 계산한다.
- iv) 유효성검정데이터의 분류정확도가 더 이상 증가하지 않거나 감소하는 시점에서 가지치기를 한다.

V. 결과 및 분석

5.1 결과 기술

전처리 과정에서 논의 되었듯이 총 4,768개의 관측치 중 67%인 3195개의 관측치는 의사결정나무 모형을 생성하는데 사용되었다. 그리고 나머지 33%(3195개)는 이 모형을 평가하는데 사용되었다. 의사결정나무의 생성은 (3.2)절의 의사결정나무 분석에서 논의된 방법에 따라 진행되었다. 종료규칙에 의해서는 20개의 잎(leaf)를 가진 나무가 생성되었다. 과도적합의 위험을 피하기 위해 가지치기(pruning)를 실시하였다. <그림 3>은 훈련



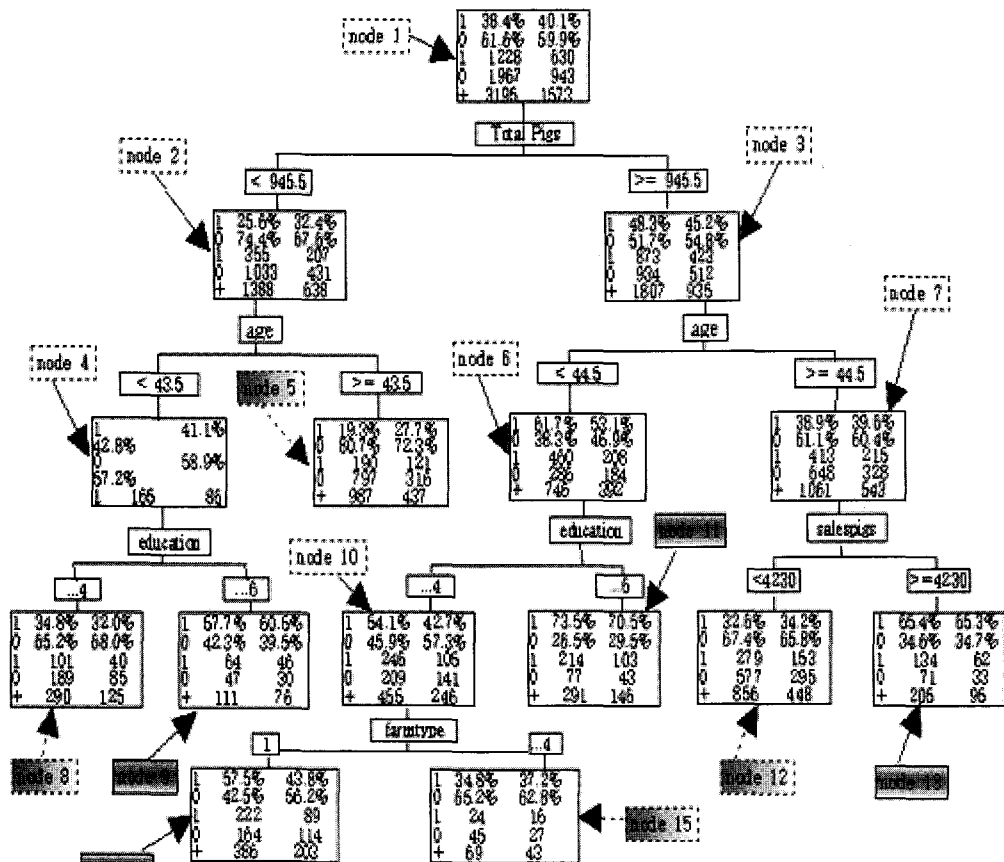
<그림 3> 훈련데이터(train)와 유효성검정데이터(valid)의 분류정확도

데이터와 유효성검정데이터의 분류정확도를 보여주고 있다. 이를 보면 유효성검정데이터의 분류정확도는 앞이 8개인 시점에서 더 이상 증가하지 않고 감소하고 있다. 따라서 최종적으로 <그림 4>와 같이 8개의 잎을 가진 의사결정나무가 생성되었다.

의사결정나무 분석결과는 <그림 4>과 같다.

<그림 4>에는 각 마디에 속하는 농업정보시스템 수용자(MIS=1)의 빈도와 비율, 미수용자(MIS=0)의 빈도와 비율이 나타나있다. 여기서 왼쪽은 훈련데이터(training data)이며, 오른쪽은 훈련데이터를 통해 생성한 의사결정나무 규칙에 유효성검정데이터(validation data)를 적용한 것이다. 뿌리마디에는 총 3195개의 훈련데이터가 있다. 이 중 농업정보시스템 수용자는 38.4%이다.

이 뿌리마디는 사육규모(totalpigs)에 의해서 두 개의 마디-사육규모가 945.5두 미만인 마디 2 (node 2)와 945.5이상인 마디 3(node 3)-로 분리되었다. 이렇게 분리된 마디2와 마디3은 분리기준(splitting criteria)에 의해 또 다시 분리된다. 이와 같은 과정을 재귀적으로 반복하면서 종료규칙과 가지치기를 통해 최종적으로 8개의 잎이 생성되었다. 이 8개의 잎을 가진 의사결정나무를 생성하는 데 투입된 변수는 사육규모(totalpigs), 연령(age), 교육(education), 출하두수(salespigs), 농가형태(farmtype)이다. 이 중 상대적으로 큰 영향을 미친 변인은 사육규모, 연령, 교육이었다. 사육규모의 경우는 약 1000두를 기준으로 분리되었으며, 연령의 경우는 약 45세를 기준으로, 학력은 고졸 이하와 대졸 이상으로



<그림 4> 정보시스템 수용에 대한 의사결정나무 분석 결과

분리되었다. 분리된 잎들은 확률 값 50%를 기준으로 농업정보시스템 수용자 집단과 미수용자 집단으로 구분하였다. 따라서 마디 9, 11, 13, 14는 농업정보시스템 수용율이 50% 이상이므로 수용자 집단으로 분류되었다. 마디 5, 8, 12, 15는 미수용자 집단으로 분류되었다(<표 III 2>).

유효성검정데이터를 통해 각각의 잎(leaf)들을 평가해보면(<그림 III 4>, <표 III 2>), 마디 14는 14%, 마디 5는 8%의 차이를 나타냈다. 그 외는 0~3% 정도의 차이를 보여 분류결과가 매우 우수한 것으로 나타났다. 마디 5의 경우에는 8%의 차이가 나타나지만 미수용자의 비율이 70% 이상으로 미수용자 집단으로 분류하는데 큰 문제가 없다. 그러나 마디 14는 유효성검정데이터 입장에서 보면 미수용자 집단으로 분류된다. 따라서 마디 14는 미수용자 집단을 수용자 집단으로 잘못 분류하는 오류(Type II error)의 가능성이 크다.

수용자 집단인 4개의 마디들은 다음과 같이 분류된다.

- i) 마디 9
사육두수가 적고(945두미만), 젊고(43.5세이하), 교육수준이 높다(전문대이상)
- ii) 마디 11
사육두수가 많고(945두이상), 젊고(44.5세이

하), 교육수준이 높다(전문대이상)

- ii) 마디 13
사육두수가 많고(945두이상), 연령이 많고(44.5세이상), 출하량(4230두이상)이 많다.
- ii) 마디 14
사육두수가 많고(945두이상), 젊고(44.5세이하) 교육수준이 낮고(고등학교이하), 영농형태가 복잡(일괄사육)하다.

이를 통해 다음과 같은 해석이 가능하다.

- i) 사육규모와 관계없이 젊고 교육수준이 높은 농가의 경우는 정보시스템 수용확률이 높다.
- ii) 연령이 많으면 일반적으로 정보시스템 수용확률이 낮다. 그러나 연령이 많은 경우에도 사육규모가 크고 출하량이 많은 농가는 정보시스템 수용확률이 높다.
- iii) 교육수준이 낮은 경우는 일반적으로 정보시스템 수용확률이 낮다. 그러나 교육수준이 낮은 경우에도 사육규모가 크고, 젊고, 일괄 사육하는 농가의 경우는 정보시스템 수용확률이 높다.

유효성검정 데이터를 적용하여 생성된 의사

<표 2> 의사결정나무 잎(leaf)들의 농업정보시스템 수용비율과 분류

마디번호 (node ID)	N	훈련데이터 MIS=1(%)	유효성검정데이터 MIS=1(%)	농업정보시스템 수용여부예측
9	111	57.66	60.5	수용자
11	291	73.54	70.5	수용자
13	205	65.37	65.3	수용자
14	386	57.51	43.8	수용자
5	987	19.25	27.7	미수용자
8	290	34.83	32.0	미수용자
12	856	32.59	34.2	미수용자
15	69	34.78	37.2	미수용자

<표 3> 의사결정나무의 실제값과 예측값의 분할표

실제값 \ 예측값	0	1	전체
0 행백분율	723 76.7%	220 23.3%	943 100.0%
1 행백분율	330 52.4%	300 47.6%	630 100.0%
전체 행백분율	1053 66.9%	520 33.1%	1573 100.0%
정확도	(723+300)/1573 * 100 = 65.0%		

결정나무가 예측한 값을 실제값과 비교한 분할표는 <표 3>과 같다. 실제 0인 경우를 0으로 정확히 예측한 76.7%였으며, 실제 1인 경우를 1로 예측한 경우는 47.6%로 나타났다. 이를 통해 모형의 정확도(accuracy)를 구하면 65.0%로 나타났다.

5.2 결과 분석 및 적용

의사결정나무 분석을 이용해 농업정보시스템 수용자를 세분화하였다. 의사결정나무를 생성하는 데 사용된 변수는 사육규모, 연령, 교육, 출하두수, 농가형태이다. 이 중 상대적으로 큰 영향을 미친 변인은 사육규모, 연령, 교육이었다. 사육규모의 경우는 약 1000두를 기준으로 분리되었으며, 연령의 경우는 약 45세를 기준으로, 학력은 고졸이하와 대졸이상으로 분리되었다.

농업정보시스템을 수용할 가능성이 높은 집단은 4개로 분류되었다. 각 집단의 특성과 정보시스템 수용을 증대시키기 위한 전략을 도출하면 다음과 같다.

첫 번째 집단은 사육두수가 많고(945두 이상), 젊고(44.5세 이하), 교육수준이 높은(전문대이상) 경우이다. 이 집단은 기존의 로짓분석을 통한 선행연구 결과와 동일하며, 정보시스템 수용율도 가장 높았다. 즉 이 집단은 농업정보시스템을 가장 적극적으로 수용하는 집단으로 새로운 농업

정보시스템이 도입될 경우 초기수용자가 될 가능성이 높다. 혁신전파이론에 따르면 새로운 기술의 확산에는 초기 수용자가 매우 중요하다. 따라서 새로운 정보시스템을 보급하고자 할 경우 규모가 크고 젊으며, 교육수준이 높은 농가를 대상으로 할 경우 더 효과적일 것이다.

두 번째 집단은 사육두수가 많고(945두 이상), 연령이 많고(44.5세 이상), 출하량이 많은(4230두 이상) 경우이다. 연령이 많으면 컴퓨터를 사용하는 방법을 배우기가 어려우며(Idding and Apps, 1990), 컴퓨터 사용을 위한 학습비용에 비해 향후 사용기간이 짧아 기대이익이 크지 않으므로 정보시스템 수용률이 떨어진다(Doye et al. 2000). 그러나 이 집단에서 보듯이 연령이 많을 경우에도 사육규모가 크고, 출하량이 많은 경우에는 수용률이 높다. 따라서 이 집단에 속하는 농가의 경우에는 학습능력을 고려하여 복잡하지 않고 사용하기 쉬운 농업정보시스템을 제공하고, 지속적인 학습기회를 제공하는 전략이 필요하다.

세 번째 집단은 사육두수가 많고(945두 이상), 젊고(44.5세 이하) 교육수준이 낮고(고등학교 이하), 영농형태가 복잡(일괄사육)한 경우이다. 일반적으로 교육수준이 낮을 경우에는 농업정보시스템 수용률이 낮지만, 이 집단에서 보듯이 교육수준이 낮은 경우에도 사육두수가 많고, 젊고, 영농형태가 복잡할 경우에는 정보시스템 수용확률

이 높다. 이 집단의 경우에는 교육수준이 낮아 정보시스템에 대한 학습능력이 떨어질 가능성이 많다. 따라서 두 번째 집단과 마찬가지로, 농업정보시스템을 제공하고, 지속적인 학습기회를 제공하는 전략이 필요하다. 그러나 두 번째 집단에 비해 이 집단은 젊은 농가들이므로, 지속적인 학습이 이루어지면 학습능력이 제고될 가능성이 크다. 따라서 이 집단에는 학습능력 제고에 더 많은 투자가 필요하다. 그러나 이 집단의 경우에는 해석에 주의가 필요하다. 왜냐하면 유효성 검정데이터를 적용한 경우에는 미수용자의 집단으로 분류되었기 때문이다.

네 번째 집단은 사육두수가 적고(945두 미만), 젊고(43.5세 이하), 교육수준이 높은(전문대이상) 경우이다. 일반적으로 사육규모가 적을 경우에는 정보시스템 수용률이 낮다. 그러나 이 집단에서 보듯이 사육규모가 적은 경우에도, 젊고 교육수준이 높으면 수용률이 높다. 사육규모가 적을 경우에는 정보시스템사용으로 인한 기대이익이 크지 않다(Kuhlmann and Brodersen, 2001). 그럼에도 수용률이 높다는 것은, 이 집단이 정보시스템 자체를 선호하는 경향이 있음을 보여준다. 따라서 새로운 정보시스템이 개발될 경우 이 집단에게는 초기에 제공하여 사용할 수 있는 기회를 지속적으로 제공할 필요가 있다.

VI. 결 론

본 연구는 의사결정나무 분석을 이용해 정보시스템 수용자를 세분화해 보고, 그 효율성을 검증해 보았다. 의사결정나무를 생성하는 데 수용자의 특성 변수들을 이용하였고, 그 결과로 실증 검증 대상이 된 농업정보시스템의 수용 가능성이 높은 집단 4개 집단이 확인되었다. 5장에서 확인된 초기 수용자 집단의 성향에 따른 수용정책의 상이한 접근법은 의사결정나무 방법론의 효율적 활용가능성을 보여주고 있다.

또한, 4개의 집단 중 제 2, 3, 4 집단의 경우처

럼 일부 확산에 저해되는 성향(소규모,고령, 저학력)을 가지고 있음에도 불구하고, 확산이 효과적인 집단으로 확인됨으로써, 의사결정나무 분석법은 기존의 로짓 및 프로빗 분석법에 비해 비선형적인 상이한 성격의 집단이 초기 수용자 집단으로 확인이 가능한 장점을 보여주고 있다. 그리고, 의사결정나무 분석법은 매우 간단한 논리로 표현되므로 실무에 적용하기가 매우 편리하다(예를 들면, 전문대이상의 학력을 가진 44세이상의 경영주는 정보시스템수용 가능성이 높음). 이는 전문지식을 가지지 않은 일반인에게도 직관적으로 방법론과 결과의 이해가 용이한 장점을 가지고 있다.

본 연구의 실증연구에서 확인한 바와 같이, 의사결정나무 방법론은 앞으로 정보시스템의 수용확산을 위한 초기 수용 집단의 확인과 분석에 있어 아주 유용한 방법론이 되리라고 생각한다. 더욱이 본 방법론이 기존의 방법론과 배타적이지 않은 방법론으로서 동시에 분석될 경우 분석이 상호 보완적으로 사용될 수 있는 장점을 가지고 있다. 앞으로 실증 분석을 통한 방법론들의 결과 직접 비교 및 본 방법론을 응용하여 학습효과로 이용한 발전적 분석 모형의 모색 등이 추후 연구로 추진될 수 있을 것으로 기대된다.

참 고 문 헌

- 박덕병, 이민수, “농업지식체계 접근에 의한 농업연구, 지도 연계를 위한 당면과제”, 한국농촌지도학회지, 제9권, 2002, pp. 199-214.
- 양희동, 문윤지, “정보기술 수용에 있어서 사용자 특성과 정보기술 특성에 따른 사회적 영향의 차이”, 경영정보학연구, 제15권, 제2호 2005, pp. 97-120.
- 이동만, 장명희 유지영, “기업의 웹 사이트 채택 시기에 영향을 미치는 요인 - 혁신확산이론 관점에서”, 경영정보학연구, 제13권, 제4호, 2003 pp. 257-277.
- 장남식, “사전세분화를 통한 고객 분류모형의

- 효과성 제고에 관한 연구”, *Information Systems Review*, 제7권, 제2호, 2005, pp. 23-40.
- Albrecht, T.L., and Bryant, C.(1996), Advanced in Segmentation Modeling for Health Communication and Social Marketing Campaigns, *Journal of Health Communication*, Vol.1, pp. 65-80.
- Berry, M. J.A. and Linoff, G.S.(2003), *Data Mining Techniques*, Indianapolis, IN: Wiley.
- Bhattacharyya, A., Harris, T.R., Kvasnicka, W.G., and Vesperat, G.M.(1997), Factors Influencing Rates of Adoption of Trichomoniasis Vaccine by Nevada Range Cattle Producers, *Journal of Agricultural and Resource Economics*, Vol.22, pp. 174-190.
- Breiman, L., J. Friedman, R. Olshen, and C. Stone, Classification and Regression and Regression Trees, Belmont, CA:Wadsworth, 1984.
- Brown, J.D.(1992), Benefit Segmentation of the Fitness Market, *Health Marketing Quarterly*, Vol.9, pp. 19-28.
- Chen, J.(2003), Market segmentation by tourists' sentiments, *Annals of Tourism Research*, Vol. 30, pp. 178-193.
- Dawson, S.(1989), Health Care Consumption and Consumer Social Class: A Different Look at the Patient, *Journal of Health Care Marketing*, Vol.9, pp. 15-25.
- Doye, D., Jolly, R., Hornbaker, R., Cross, T., King, R.P., Lazarus, W.F., and Yeboah, A.(2000), Farm Information systems: Their Development and Use in Decision Making, Iowa State University: Research Bulletin 601, North Central Regional Research Publication No.345.
- Feder, G., “Adoption of Agricultural Innovations in Developing Countries: A Survey”, *Economic Development and Cultural Change*, Vol.33, 1985, pp. 255-298.
- Feder, G. and Umali, D.L.(1993), “The Adoption of Agricultural Innovations: A Review”, *Technological Forecasting and Social Change*, Vol.43, pp. 215-239.
- Foltz, J.D.(2003), The Economics of Water Conserving Technology Adoption in Tunisia: An Empirical Estimation of Farmer Technology Choice, *Economic Development and Cultural Change*, Vol.51, pp. 359-373.
- Frankenberger, K.D., and Sukhdial, A.S.(1994), Segmenting Teens for AIDS Preventive Behaviors with Implications for Marketing Communications, *Journal of Public Policy and Marketing*, Vol.13, pp. 133-150.
- Iddings, R.K., and J. W. Apps , “What Influences Farmers' Computer Use?”, *Journal of Extension*, Vol.28, No.1, 1990 (<http://www.jce.org/jce/1990spring/a4.html>.2004/10/1).
- Kass, G., “An Exploratory Technique for Investigating Large Quantities of Categorical Data”, *Applied Statistics*, Vol.29, 1980, pp. 119-127.
- Kuhlmann, F., and Brodersen, C.(2001), Information technology and farm management: developments and perspectives, *Computers and Electronics in Agriculture*, Vol.30, pp.71-83.
- Lehmann, Donald, Sunil Gupta, and Joel Steckel. Marketing Research. Reading, Mass.: Addison Wesley, 1998.
- Levin, N. and J. Zahavi, “Predictive Modeling Using Segmentation”, *Journal of Interactive marketing*, Vol.15, 2001, pp. 2-22.
- McGlone, T.A., and Calantone, R.J.(1992), A Goal Programming Model for Effective Segment Determination: A Comment and Application, *Decision Sciences*, Vol.23, pp. 1231-1239.
- Murthy, K.S.(1998), Automatic Construction of Decision Trees from Data: A Multi disciplinary Survey, *Data Mining and Knowledge Discovery*, Vol.2, pp. 345-389.

- Quinlan, J.R.(1986), Induction of Decision Trees, *Machine Learning*, Vol.1, pp. 81-106.
- Quinlan, J.R.(1993), C4.5: Program of Machine Learning. CA., Morgan Kaufman Publishing.
- Rogers E.M.(2003), Diffusion of Innovation, Free Press, NY.
- Rogers, E.M., and Storey, J.D.(1987), Communication Campaigns. In C.R. Berger and S.H. Chaffee(Eds.), *Handbook of Communication Science*(pp. 817-846). Newbury Park, CA:Sage.
- Shrestha, R.B., and Gopalakrishnan, C.(1993), Adoption and Diffusion of Drip Irrigation Technology: An Econometric Analysis, *Economic Development and Cultural Change*, Vol.41, pp. 407- 418.
- Sonquist, J., E. Baker, and J.N. Morgan, Searching for Structure, Survey Research Center, Ann Arbor: University of Michigan, 1971.
- Wedel & Kamakura, 2000, M. Wedel and W.A. Kamakura, Market Segmentation: Conceptual and Methodological Foundations., Kluwer, Dordrecht(2000).
- Wedel, M., and Kamakura, W.C.(1998). Market Segmentation: Conceptual and Methodological Foundations, International Series in Quantitative Marketing, Kluwer Academic Publishers, Dordrecht, the Netherlands.
- Weinstein, A.(1994), Market Segmentation. New York: Irwin Professional Publishing.

Information Systems Review

Volume 9 Number 1

April 2007

Identifying Early Adopters of Information Systems by Inductive Learning Using Decision Tree Method

Minsoo Lee* · Young Chan Choe** · Byungjoon Yoo***

Abstract

In diffusing an information systems (IS), the provider of the IS can be more effective if they can identify user groups who can adopt the system early. By focusing on the user groups, system providers can encourage them to adopt the IS. After the early adopters adopt an IS, the diffusion of the system to other groups can be easier by early adopters' voluntary advertisement and help in adopting the IS.

Instead of discrete choice methods which are usually used for this purpose, we suggest a decision tree method. Compared to discrete choice methods, this method is more accurate for prediction and can easily identify non-linear segments of groups. By testing the data of adopters of an IS in agricultural business, we show the excellence of this method in identifying target groups to focus on. This method would help system providers to diffuse their systems by starting from early adopters.

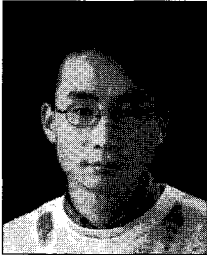
Keywords: Decision Tree, System Adoption, Segmentation

* Researcher, Rural Resources Development Institute

** Professor, Department of Agricultural Economics and Rural Development, Seoul National University

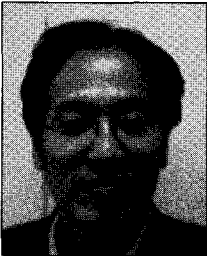
*** Professor, Korea University Business School, Corresponding author

◎ 저자 소개 ◎



이민수 (minsoo.lee@gmail.com)

서울대학교 농촌사회교육학과를 졸업하고, 동대학원에서 교육학 박사 학위를 취득하였다. 현재 서울대학교 농업생명과학연구원 연구원을 거쳐, 농촌자원개발 연구소 연구원으로 재직 중이다. 주요 관심 연구분야는 농업에서의 정보기술 수용 및 효과, 농산물전자상거래, 농촌개발과 정보화 등이다.



최영찬 (aggi@korea.ac.kr)

서울대학교 농과대학을 졸업하고, Michigan State University에서 농업경제 석사와 박사학위를 취득하였다. 현재 서울대학교 농경제사회학부 교수로 재직 중이다. 주요 관심 연구분야는 농산물전자상거래 전략 및 시스템계획, 농업정보시스템 개발 및 효과분석, 농업분야의 정보화 전략 등이다.



유병준 (byoo@korea.ac.kr)

서울대학교 경영학과를 졸업하고, University of Arizona에서 MIS 석사학위를 취득하고, Carnegie Mellon University에서 경영학 박사 학위를 취득하였다. 현재 홍콩 과학기술대 교수를 거쳐, 고려대학교 경영대학 교수로 재직 중이다. 주요 관심 연구분야는 전자상거래의 디지털 콘텐츠 가격 전략, 기업간거래 전자 시장, 전자 경매 등이다.

논문접수일 : 2006년 10월 27일
1차 수정일 : 2007년 03월 30일

게재확정일 : 2007년 04월 02일