

## 인터넷 쇼핑몰에서 구매품목을 이용한 고객의 예측모델 설계

지혜영<sup>1</sup>, 조완현<sup>2</sup>

<sup>12</sup>전남대학교 통계학과

접수 2008년 12월 8일, 수정 2008년 12월 23일, 게재확정 2008년 12월 26일

### 요약

최근 인터넷 기술의 획기적인 발달에 따라 인터넷 이용자 및 이용률 등의 양적 성장뿐만 아니라 인터넷을 활용한 다양한 제품의 판매와 영업, 광고 등 질적 성장까지 인터넷 사용이 사회 전반으로 광범위하게 확대되고 있다. 본 연구는 급속히 변화하는 인터넷 이용환경을 바탕으로 업계의 비즈니스 전략 수립 및 학계의 연구 활동 등에 광범위하게 활용될 수 있는 인터넷 쇼핑 고객의 세분화 전략에 대한 방안과 해결방법에 대한 정보를 제공하는데 목적이 있다. 본 논문에서는 인터넷 쇼핑몰을 통해 고객들이 구매한 품목들을 통계적 분석기법인 포지셔닝과 대응분석을 통해 구매품목들 간의 유사성을 비교 분석하고, 기존 고객들을 세분화하기 위한 예측 모델을 설계하고자 하였으며 이를 통해 고객 개인의 특성에 맞는 마케팅 프로모션을 제공함으로써 기업은 안정된 고객층으로부터 수익을 창출하고 고객들은 해당 기업으로부터 더 많은 혜택을 받게 하는데 목적이 있다.

주요용어: 고객 세분화, 대응분석, 인터넷 쇼핑, 포지셔닝.

### 1. 서론

오늘날 정보통신기술의 급속한 발달은 정보혁명이라는 용어를 탄생시킬 정도로 우리의 생활을 급속하게 변화시키고 있다. 이러한 변화 가운데서도 가장 두드러지게 나타나는 현상은 온라인 및 인터넷의 급격한 성장이다. 특히, 인터넷의 사용이 급증하면서 매우 빠른 속도로 우리의 소비생활 전반에 영향을 미치고 있으며, 인터넷의 확산과 더불어 등장한 인터넷 쇼핑몰은 소비자들의 생활패턴과 소비방식까지 변화시키고 있다.

인터넷 이용인구의 증가에 따른 인터넷 쇼핑몰 성장을 보면 2007년 연간 거래 규모는 총 15조7,656억 원으로 전년 13조4,596억 원에 비하면 2조3,060억 원 (17.1%)이 증가한 것으로 집계되었다 (통계청의 "연간 사이버 쇼핑몰 통계 조사" 결과). 또한 2007년 12월 대상사업체수는 4,508개로 전년 동월의 4,531개에 비해서 23개 감소한 것으로 조사되었지만 기업 소비자간 거래액은 총 10조2,258억 원으로 전년도에 비해 12.0% 증가한 것으로 나타났다. 그림 1.1에서도 알 수 있듯이 대상사업체수의 전년 동월비는 감소하는 양상을 보이는 반면, 상품 및 서비스 거래액은 꾸준히 증가하는 양상을 보이고 있다. 즉, 인터넷 쇼핑몰 하나하나의 규모가 커져가고 있음을 알 수 있다.

현재 인터넷을 통한 잠재적 소비자가 급증함에 따라 보다 많은 소비자에게 제품을 판매하기 위해서는 고유의 특성을 고려하지 않고 기존의 마케팅 전략을 그대로 적용한다면 전략적 효과를 거두기 힘들 것이다. 따라서 인터넷 쇼핑몰의 성공적 수행을 위해서는 인터넷 상에서의 소비자의 특성을 이해하고 이에 맞는 전략을 수립할 필요가 있다. 기업은 고객 이탈을 최소화하면서 기존 고객을 잘 관리하여 평균 가치

<sup>1</sup> (500-757) 광주시 북구 용봉동 300, 전남대학교 통계학과, 석사과정.

<sup>2</sup> 교신저자: (500-757) 광주시 북구 용봉동 300, 전남대학교 통계학과, 교수. E-mail: whcho@chonnam.ac.kr

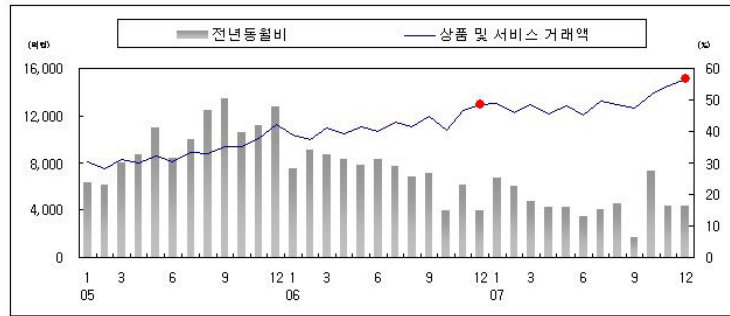


그림 1.1 인터넷 쇼핑몰의 사업체수와 거래액의 변화추이

를 높이기 위해 가장 기여도가 높은 고객들을 찾아내고, 지속적으로 고객들을 관리함으로써, 개별 고객의 수익성을 극대화하기 위한 방법을 모색하기 위한 대안으로 등장한 고객관계관리(CRM)에 대한 연구가 필요하게 되었다. 고객관계관리는 고객과의 관계에 중점을 둔 경영을 의미하는 것으로서 고객관련 데이터를 분석하여 고객의 특성에 기초한 마케팅 활동을 계획, 실행, 평가 하는 것으로 가치 있는 고객을 파악, 획득, 유지하여 한번 고객은 평생 고객이 될 수 있도록 하여 고객평생가치를 극대화하는 것이다. 따라서 인터넷 쇼핑몰도 마찬가지로 고객관계관리를 위한 전략 수립이 요구되고 있으며 이러한 연구는 인터넷 쇼핑몰에서 우수 고객을 관리하는데 필수적인 요소라고 볼 수 있다.

이러한 취지에서 본 논문은 인터넷 쇼핑몰 고객을 대상으로 쇼핑몰 이용자들의 성격을 파악하고 우수 고객을 유지하고 관리할 수 있는 고객세분화에 관한 연구를 수행하고자 한다. 본 연구의 방안으로 고객의 포지셔닝, 구입 품목의 성격 파악 및 분류, 우수 고객의 세분화 분석을 통해 국내 인터넷 쇼핑몰의 경쟁력 강화와 수익성 증대를 도모할 수 있는 이론적 방법과 통계적 기법을 제공하고자 한다.

## 2. 선행 연구

본 연구의 이론적 배경은 고객관계관리(CRM), 포지셔닝, 대응분석, 고객세분화전략과 관련되어 있으므로 이들 방법에 관한 연구결과들을 간략히 고찰해보고자 한다.

첫째, 고객관계관리에 관한 연구로는 조범석(2002)은 금융권 CRM 구축에서의 데이터 마이닝을 위한 고객 분류를 위해 RFM 개념을 토대로 한 군집분석을 이용하였으며, 노태협과 한인구(2002)는 인터넷 비즈니스 환경 하에서의 고객관계 관리기법의 현안과 시장 분석을 정리하고 인터넷 비즈니스의 변화된 환경 하에서의 CRM과 eCRM을 비교 분석하였다. 정영수와 강경화(2004)가 인터넷 쇼핑몰 사이트의 CRM 사례 분석에 연관성규칙을 적용하였으며, 이현우(2006)는 콜센터의 방대한 통화 자료를 데이터마이닝을 이용하여 분석하고 인바운드 콜센터의 ROI를 올릴 수 있는 방법을 제시하였다. 이외에도 인터넷 쇼핑몰에 관한 다른 연구로는 이승희와 정진원(2005)의 인터넷 중독의 매개효과를 중심으로 한 인터넷 쇼핑 중독구매에 관해 연구하였으며, 김선남과 이영원(2007)은 수용자의 인터넷 쇼핑 동기에 관한 연구를 수행하였다.

둘째, 쇼핑몰에서 각 제품에 대한 포지셔닝이란 개념은 소비자들의 마음속에 우리 제품을 어떻게 차별적으로 인식시킬 것인가에 대한 전략으로 개발된 제품을 소비자들의 지각 속에 적절히 위치시키도록 노력을 기하는 제품개발 및 촉진활동과 밀접한 관련이 있는 활동을 말한다. 포지셔닝과 관련된 선행연구로는 Meler와 Ruzic(1999)의 크로아티아국가의 관광상품과 시장포지셔닝에 관한 연구가 있으며, 김성섭 등(2002)의 해외여행지의 포지셔닝 분석 및 시사점에 관하여 연구하였다.

셋째, 대응분석은 범주형으로 관측된 두 변수사이의 관계를 분석하는 방법으로 두 변수의 각 범주들을 공간 (이차원 공간)상에 점으로 표현하여 범주들 사이의 관계를 분석하는 방법으로, 이와 관련된 선행연구로 김차용과 이철화 (1995)는 도시의 기혼여성들을 대상으로 스포츠 참여종목과 참여 동기간의 대응 분석을 실시하였다.

마지막으로 고객세분화의 문제는 일반적으로 성별, 연령, 소득, 취미 등과 같은 비슷한 성향과 욕구를 가진 집단을 나누는 작업을 말한다. 고객을 세분화하는 목적은 고객의 다양한 욕구를 파악하고 손쉽게 비즈니스전략을 수립하는 데 있다. 김규일 등 (2007)은 데이터마이닝 기법을 이용하여 국내 K대학교 합격자들 중 타 대학으로 이탈한 학생들의 성향을 분석하고 이탈자 예측모형을 개발하였다. 진서훈 (2005)은 고객정보의 전략적 활용을 위하여 신용카드고객의 신용카드 이용행태에 근거하여 고객세분화를 생성해 내는 과정을 소개하였으며, 이병엽 등 (2005)은 통신 산업에서 고객 분류를 위해 데이터마이닝 기법의 하나인 인공신경망, 로지스틱 회귀분석을 이용해 예측모형을 설계하였다.

### 3. 인구통계학적 특성과 구매 품목간의 연관성

본 연구의 모집단은 '인터넷 쇼핑물을 이용하는 모든 고객'으로 정의하였고, 이 연구는 한국인터넷진흥원으로부터 사전에 조사된 약 500명의 인터넷 쇼핑물 고객들의 신상자료를 이용하여 인구통계학적 변수들과 쇼핑물 구매품목간의 연관성 분석과 이를 바탕으로 쇼핑물 구매품목별 고객 세분화 분석을 실시하였다. 본 조사에 사용된 23개의 변수들은 고객들의 인구학적 특성으로 성별, 나이, 직업, 월평균 소득, 월평균 인터넷 쇼핑금액이 이용되었고, 각 구매자들의 구입품목에 대한 유형이 파악되었다. 인구통계학적 변수들과 구매품목간의 교차분석 수행결과는 표 3.1~표 3.4와 같다.

첫째, 성별에 따른 구매 품목의 교차분석의 결과 남성이 267명, 여성이 233명으로 조사되었다. 표 3.1의 교차표를 살펴보면, 정보기술은 484명으로 가장 높게 나타났고, 의류는 164명으로 가장 낮게 나타났다. 전체적으로 보면, 남성이 여성에 비해 구매품목이 많은 것으로 분석되었다.

표 3.1 구매품목과 성별의 교차표

		구매품목																합계		
		도서	의류	화장품	컴퓨터	게임	소프트웨어	금융서비스	식품	정보기술	영화	음악	사진	예약예매	여행	가전제품	아동유아		자동차용품	기타
성별	남성	151	113	216	201	193	240	252	241	255	174	122	237	178	234	211	243	241	242	267
	여성	118	51	100	215	202	226	226	203	229	157	119	219	178	212	201	190	227	209	233
	합계	269	164	316	416	395	466	478	444	484	331	241	456	356	446	412	433	468	451	500

둘째, 연령에 따른 구매 품목의 교차분석 결과 10대가 52명, 20대가 120명, 30대 192명, 40대 107명, 50대 23명, 60대 5명, 70대 1명으로 조사되었다. 구매품목에 따라 10대는 금융서비스, 식품 등 대부분의 구매품목들을 선호하는 것으로 나타났으며, 20대는 금융서비스, 30대는 정보기술, 40대는 기타, 50대는 아동유아와 소프트웨어를 선호하는 것으로 나타났고, 반면에 60대와 70대는 저조한 구매를 보이는 것으로 조사되었다.

셋째, 학력에 따른 구매품목의 교차분석을 실시한 결과, 중학생 18명, 고등학생 24명, 대학생 39명, 초졸이하 4명, 중졸 13명, 고졸 176명, 대졸이 226명으로 조사되었으며, 초졸이하를 제외한 모든 학력들이 대부분의 구매품목들을 선호하는 것으로 나타났다.

마지막으로, 월평균 소득에 따른 구매품목의 교차분석 결과를 살펴보면, 100만원 미만 22명, 100 ~ 200만원 미만 81명, 200 ~ 300만원 198명, 300 ~ 400만원 121명, 400만원 이상 78명으로 조사되었다.

표 3.2 구매품목과 연령의 교차표

	구매품목																		합계	
	도서	의류	화장품	컴퓨터	게임	소프트웨어	금융서비스	식품품	정보기술	영화	음악	사진	예약예매	여행	가전제품	아동유아	자동차용품	기타		
10대	33	21	40	46	27	51	52	52	25	9	45	46	51	51	52	41	52			
20대	57	25	58	100	94	110	117	111	112	61	36	108	77	97	98	104	114	98	120	
30대	100	65	120	153	160	171	183	169	188	140	109	175	134	173	158	148	180	177	192	
만연령	40대	57	41	78	91	88	105	98	90	104	82	69	102	73	98	81	100	96	107	107
	50대	17	8	14	21	21	23	22	17	22	19	15	21	21	20	23	21	22	23	
	60대	4	3	5	5	4	5	5	4	5	4	3	4	4	5	4	5	5	5	5
	70대	1	1	1	0	1	1	1	1	1	0	0	1	1	1	0	1	0	1	1
합계	269	164	316	416	395	466	478	444	484	331	241	456	356	446	412	433	468	451	500	

표 3.3 구매품목과 학력의 교차표

	구매품목																		합계
	도서	의류	화장품	컴퓨터	게임	소프트웨어	금융서비스	식품품	정보기술	영화	음악	사진	예약예매	여행	가전제품	아동유아	자동차용품	기타	
중학생	13	11	17	16	8	18	18	18	18	9	3	15	17	18	17	18	18	15	18
고등학생	14	7	16	22	15	23	24	24	24	11	4	23	22	24	24	24	24	19	24
대학생	19	9	21	34	25	34	39	37	37	18	8	34	26	34	34	38	39	30	39
학력	초졸이하	3	3	4	3	4	4	4	4	2	2	3	3	4	4	3	4	4	4
	중졸	8	8	7	12	10	12	10	13	10	7	12	11	12	12	12	13	12	13
	고졸	108	51	117	152	148	173	168	155	174	129	99	160	137	163	145	152	163	176
	대졸	104	75	134	177	185	202	213	196	214	150	118	210	140	192	176	185	204	226
합계	269	164	316	416	395	466	478	444	484	331	241	456	356	446	412	433	468	451	500

구매품목에 따라 모든 월평균 소득 수준에서 정보기술을 선호하는 것으로 나타났다.

표 3.4 구매품목과 월평균 소득의 교차표

	구매품목																		합계
	도서	의류	화장품	컴퓨터	게임	소프트웨어	금융서비스	식품품	정보기술	영화	음악	사진	예약예매	여행	가전제품	아동유아	자동차용품	기타	
100만원미만	13	6	13	20	18	20	22	20	21	13	8	20	19	20	20	21	22	19	22
100 200만원미만	53	27	53	69	60	74	79	74	80	52	33	71	60	72	60	69	76	67	81
200 300만원미만	112	58	121	165	160	189	189	172	190	138	103	182	150	183	161	166	181	183	198
300 400만원미만	53	42	76	96	99	111	114	105	118	81	66	112	78	110	104	106	114	110	121
400만원이상	38	31	53	66	58	72	74	73	75	47	31	71	49	61	67	71	75	72	78
합계	269	164	316	416	395	466	478	444	484	331	241	456	356	446	412	433	468	451	500

### 4. 인터넷 쇼핑몰 고객의 세분화 예측 모형

#### 4.1. 구매 품목의 유사성 비교분석

인터넷 쇼핑몰 구매 품목들 간의 유사성을 분석하기 위해 응답자들이 선호하고 있는 인터넷 쇼핑몰 구매 품목들 간의 지각도를 작성하는 접근방법을 이용하였다. 표 4.1은 의류와 도서의 유사성 계수를 구하는 과정을 나타낸 분할표로서 예시로 제시하였다. 두 구매항목 의류와 도서에 대하여 총 500명의 고객들이 상품을 구매 혹은 구매하지 않았을 경우 구매품목의 유사성을 계산하는 방법은 두 상품을 구매한 총 500명을 대상으로 구매여부에 따라 2 × 2행렬로 계산한 뒤, 두 상품의 유사도는  $161 / (161 + 70 + 175)$ 로 계산되었다. 각 구매품목의 유사성 분석은 구매품목간의 좌표를 중심으로 거리간격을 측정하여 평가하였으며 거리가 가까울수록 높은 상관관계임을 의미하고, 반대로 거리가 멀수록 낮은 상관관계를 나타낸다.

표 4.1 의류와 도서 구매여부 분할표

	의류		전체
	구매안함	구매	
도서	구매안함	94	175
	구매	70	161
전체	164	336	500

유사도 행렬을 이용하여 각 구매품목들의 유사성을 분석하기 위해 계산된 유사도를 구매품목들간의 유사성을 나타내는 거리로 간주하여 이차원 공간상에 표현하는 다차원 척도법을 실시하고 표 4.2의 결과를 얻었다.

표 4.2 구매품목들 간의 유사도 행렬

	도서	의류	화장품	컴퓨터	게임	소프트웨어	금융서비스	식품	정보기술	영화	음악	사진	예약	여행	가전제품	자동차	기타	
도서	-																	
의류	0.34	-																
화장품	0.33	0.44	-															
컴퓨터	0.202	0.18	0.15	-														
게임	0.175	0.151	0.112	0.118	-													
소프트웨어	0.095	0.053	0.048	0.146	0.121	-												
금융서비스	0.05	0.053	0.051	0.082	0.024	0.077	-											
식품	0.125	0.14	0.182	0.138	0.066	0.059	0.083	-										
정보기술	0.065	0.038	0.047	0.064	0.034	0.136	0.056	0.075	-									
영화	0.338	0.275	0.239	0.177	0.292	0.115	0.006	0.066	0.063	-								
음악	0.35	0.381	0.259	0.171	0.286	0.102	0.052	0.064	0.042	0.456	-							
사진	0.07	0.061	0.06	0.076	0.064	0.068	0.031	0.053	0.053	0.081	0.082	-						
예약	0.33	0.263	0.233	0.194	0.164	0.163	0.085	0.136	0.039	0.257	0.263	0.068	-					
여행	0.159	0.117	0.144	0.113	0.097	0.158	0.056	0.1	0.129	0.18	0.134	0.114	0.222	-				
가전제품	0.2	0.2	0.204	0.203	0.097	0.1	0.068	0.152	0.083	0.152	0.161	0.073	0.221	0.246	-			
자동차	0.133	0.175	0.195	0.102	0.082	0.063	0.06	0.171	0.051	0.078	0.113	0.067	0.111	0.1	0.24	-		
기타	0.065	0.07	0.059	0.115	0.038	0.065	0.102	0.1	0.067	0.069	0.07	0.07	0.093	0.117	0.121	0.065	-	
기타	0.077	0.078	0.099	0.056	0.085	0.025	0.029	0.029	0.032	0.107	0.145	0.011	0.06	0.051	0.046	0.045	0.025	-

각 구매품목들의 유사성 분석을 위한 다차원 척도법 수행 결과, 구매품목들간의 유사도가 높을수록 가까운 거리에 위치하고 있음을 알 수 있다. 의류와 도서, 가전제품과 컴퓨터는 거리가 매우 가까우므로 상관관계가 높음을 알 수 있었고, 음악은 다른 품목들과 가장 거리가 멀기 때문에 낮은 상관관계를 가지고 있음을 알 수 있었다.

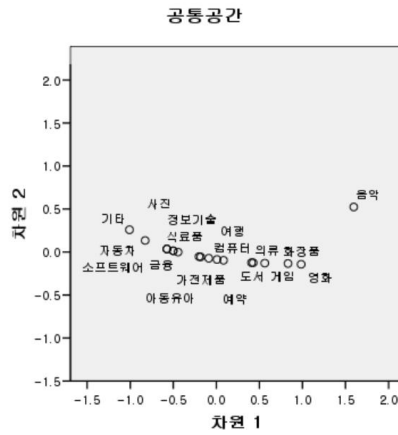


그림 4.1 구매품목들 간의 연관관계

4.2. 구매품목 별 대응분석

본 연구는 인구통계학적 특성에 따른 구매품목 별 고객 세분화를 목적으로 하기 때문에 한국 인터넷진 흥원으로부터 사전에 조사된 500명의 인터넷 쇼핑물 고객들의 자료를 바탕으로 인구통계학적 변수에 따 라 고객들이 어떤 제품을 구매하는지 분석하기 위해 단순대응분석 방법을 사용하였다.

첫째, 그림 4.2에서 성별에 따른 인터넷 쇼핑물 구매품목의 대응분석 결과를 살펴보면, 제1좌표축을 기준으로 열범주 B (의류)와 C (화장품)을 제외한 모든 항목들이 상대적으로 가까운 거리에 위치하고 있으며 성별에 따라 인터넷 쇼핑물 고객들의 구매품목에 뚜렷한 차이가 없음을 알 수 있다.

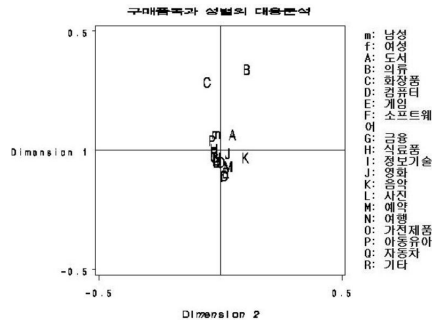


그림 4.2 성별에 따른 구매 품목 대응분석 결과

둘째, 그림 4.3은 연령에 따른 구매 품목에 대해 대응분석을 실시한 결과이다. 그림 4.3은 대응분석 결과를 살펴보면 행범주 g (70대)는 다른 행범주들과 다른 방향으로 상대적으로 멀리 떨어져 있음을 알 수 있다. 이런 행범주의 상대적인 위치는 연령에 따라 인터넷 쇼핑물 고객들의 구매품목에 차이가 있음을 의미한다. 실제로 a (10대)와 g (70대)를 제외한 행범주들이 열범주들과 모두 대응하고 있음을 알 수 있다.

셋째, 그림 4.4는 학력에 따른 구매 품목에 대해 대응분석을 실시한 결과이다. 인터넷 쇼핑물 이용 고

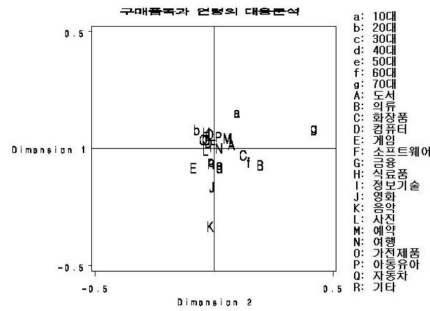


그림 4.3 연령에 따른 구매품목 대응분석 결과

객의 학력에 따른 대응분석 결과를 살펴보면 a (중학생), b (고등학생), c (대학생)는 다른 행범주들과 다른 방향으로 상대적으로 멀리 떨어져 있음을 알 수 있다. 이런 행범주의 상대적인 위치는 학력에 따라 인터넷 쇼핑물 고객들의 구매품목에 차이가 있음을 의미한다. 실제로 행범주 e (중졸), f (고졸), g (대졸) 만이 열범주인 구매품목들과 대응하는 특징을 발견할 수 있다.

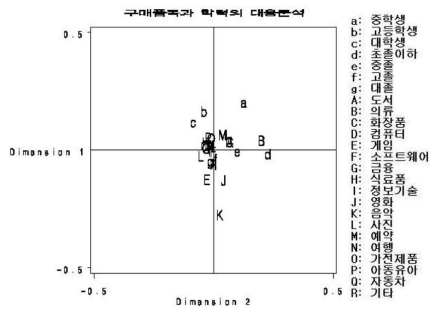


그림 4.4 학력에 따른 구매품목 대응분석 결과

넷째, 그림 4.5은 월평균 소득에 따른 구매 품목에 대해 대응분석을 실시한 결과이다. 인터넷 쇼핑물 이용 고객들의 월평균 소득에 따른 대응분석 결과를 살펴보면 a (100만원 미만), b (100 200만원 미만), c (200 300만원 미만), d (300 400만원 미만), e (400만원 이상)이 모두 거의 근접한 거리에 위치하고 있음을 알 수 있다. 월평균 소득에 따라서 인터넷 쇼핑물 고객들의 구매 품목 간에는 크게 차이가 없음을 예측할 수 있다.

### 4.3. 우수고객의 세분화

인터넷 쇼핑물을 이용하는 고객 데이터를 통계적 분석기법인 교차분석, 다차원 척도법, 대응분석을 이용하여 고객 그룹들의 속성을 파악할 수 있었다. 이를 바탕으로, 인터넷 쇼핑물을 이용하는 기존 고객들을 우수고객, 비 우수 고객으로 세분화하기 위한 예측 모델을 설계하기 위하여 SAS 데이터마이닝 툴인 엔터프라이즈 마이너를 이용하여 로지스틱 회귀분석과 의사결정나무분석을 수행하였으며, 데이터마이닝 모형 구성도는 그림 4.6과 같다.

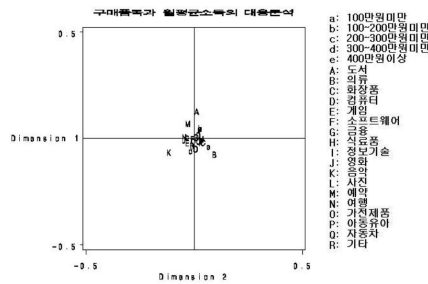


그림 4.5 월평균 소득에 따른 구매품목 대응분석 결과

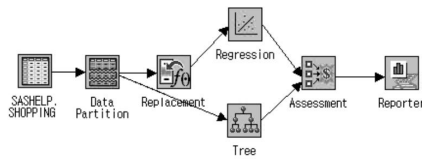


그림 4.6 데이터마ining 모형구성도

4.3.1. 변수 선정

로지스틱 회귀분석과 의사결정나무 분석을 수행하기 전에 유의한 변수들을 선택하기 위하여 카이제곱 통계량을 이용하였다. 인터넷 쇼핑몰 이용고객의 월평균 인터넷 쇼핑금액이 상위 30% 그룹 (월 평균 인터넷 쇼핑 금액이 45,000원 이상인 그룹)을 우수 고객이라 할 때, 카이제곱 통계량은 목표변수인 고객의 우수 여부 변수와 각 입력변수들 간의 교차분석표를 통해 계산되었으며 그 결과는 표 4.3과 같다.

카이제곱 통계량과 유의확률을 통해 총 23개의 변수중 화장품, 컴퓨터, 여행, 이동유아, 자동차, 기타 구매품목 변수들이 유의한 변수로 선택되었다.

4.3.2. 로지스틱 회귀분석

인터넷 쇼핑몰 이용고객의 월평균 인터넷 쇼핑금액이 상위 30% 그룹 (월 평균 인터넷 쇼핑 금액이 45,000원 이상인 그룹)을 우수 고객이라고 할 때 데이터마ining 모형을 통한 고객 분류 적중률을 분석하기 위해 로지스틱 회귀분석을 수행하였다. cutoff 값이 0.2인 경우 훈련용이 68.75%, 검증용이 69%로 가장 좋은 예측력을 보였으며 결과는 표 4.4과 같다.

로지스틱 회귀분석결과 고객의 우수여부를 예측하는데 유의한 변수로 총 23개 변수 중 화장품, 컴퓨터, 여행, 이동유아, 자동차, 기타의 6개 변수가 의미 있는 변수로 선택되었으며, 그 중 '이동유아'가 우수고객 여부에 가장 큰 영향을 미치는 것을 알 수 있다. 그 다음으로 여행, 자동차용품 순이었다. 결과를 해석해보면, 이동유아 상품을 구매하는 고객은 그렇지 않은 고객에 비해 우수고객일 확률이 2.3 (=exp (0.8205))배이며, 여행상품을 구매하는 고객은 그렇지 않은 고객에 비해 우수고객일 확률이 1.9 (=exp (0.6485))배 더 높음을 알 수 있다. 또한, 자동차 용품을 구매하는 고객은 그렇지 않은 고객에 비해 우수고객일 확률이 2.2 (=exp (0.7784))배 더 높음을 알 수 있다.

4.3.3. 의사결정 나무분석

의사결정 나무분석 수행결과, cutoff 값이 0.2인 경우 훈련용 집합에서는 우수고객을 우수하다고 예측



표 4.3 카이제곱 통계량 선택 변수

변수 명	카이제곱	유의확률 (p)
성별	0.2413	0.6233
만 연령	1.0859	0.2974
학력	6.1416	0.4075
소득	5.5777	0.2330
도서	0.2049	0.6508
의류	0.1455	0.7029
화장품	9.85	0.0017
컴퓨터	7.81	0.0052
컴퓨터게임	2.0348	0.1537
소프트웨어	0.0173	0.8955
금융서비스	0.3568	0.5503
식료품	2.0206	0.1552
정보기술	0.8939	0.3444
영화	0.0256	0.8729
음악	0.0000	0.9957
사진	0.5594	0.4545
예약	1.1584	0.2818
여행	13.36	0.0003
가전전자	0.9521	0.3292
아동유아	20.76	0.0001
자동차용품	10.83	0.0010
기타	3.87	0.0491
월평균인터넷쇼핑금액	3.21	0.0731

표 4.4 로지스틱 회귀모형 수행결과

훈련용 집합				검증용 집합			
합계	비 우수고객		총합	합계	비 우수고객		총합
	0	1			0	1	
0	171 42.75 64.29 85.07	95 23.75 35.71 47.74	266 66.50	0	44 44.00 65.67 84.62	23 23.00 34.33 47.92	67 67.00
1	30 7.50 22.39 14.93	104 26.00 77.61 52.26	134 33.50	1	8 8.00 24.24 15.38	25 25.00 75.76 52.08	33 33.00
총합	201 50.25	199 49.75	400 100.00	총합	52 52.00	48 48.00	100 100.00

표 4.5 로지스틱 회귀모형 성과표 (CUTOFF=0.2)

	훈련용 집합의 적중률	검증용 집합의 적중률
우수고객 표본	77.61%	75.76%
비 우수고객 표본	64.29%	65.67%
전체	68.75%	69%

하고 비 우수고객을 비 우수고객으로 예측할 확률이 68%로 나타났으며, 검증용 집합에서는 우수고객을 우수하다고 예측하고 비 우수 고객을 비 우수고객으로 예측할 확률이 69%로 가장 좋은 예측력을 보였으

표 4.6 로지스틱 회귀모형 선택변수

변수 명	회귀계수	표준오차	카이제곱	유의확률 (p)
절편	0.8621	0.3760	5.26	0.0219
화장품	0.3889	0.1239	9.85	0.0017
컴퓨터	0.4127	0.1476	7.81	0.0052
여행	0.6485	0.1774	13.36	0.0003
아동유아	0.8205	0.1801	20.76	0.0001
자동차용품	0.7784	0.2365	10.83	0.0010
기타	-0.4761	0.2420	3.87	0.0491

며 그 결과는 다음과 같다.

표 4.7 의사결정 나무모형 수행결과

훈련용 집합				검증용 집합			
합계	0		합계	합계	0		합계
	1	합계			1	합계	
0	196 49.00 73.68 77.17	70 17.50 26.32 47.95	266 66.50	0	50 50.00 74.63 78.13	17 17.00 25.37 47.22	67 67.00
1	58 14.50 43.28 22.83	76 19.00 56.72 52.05	134 33.50	1	14 14.00 42.42 21.88	19 19.00 57.58 52.78	33 33.00
총합	254 63.50	146 36.50	400 100.00	총합	64 64.00	36 36.00	100 100.00

표 4.8 의사결정 나무모형 성과표 (CUTOFF=0.2)

	훈련용 집합의 적중률	검증용 집합의 적중률
우수고객 표본	56.72%	57.58%
비우수고객 표본	73.68%	74.63%
전체	68%	69%

분류기준으로는 카이제곱 검정을 사용하였으며 구축한 의사결정나무분석의 결과는 그림 4.7과 같다. 의사결정 나무분석 결과, 인터넷 쇼핑물 구매품목 중 '컴퓨터 (i48\_4)'가 가장 큰 영향을 미쳤으며, 그 다음으로는 '화장품 (i48\_3)', '여행상품 (i48\_14)', '아동유아 (i48\_16)' 순이었다.

#### 4.3.4. 분석 모형의 결과 비교

데이터마이닝 고객 세분화 결과 두 모형의 정분류율은 다음과 같으며, 훈련용 집합의 경우 우수고객 예측률은 로지스틱 회귀분석 68.75%, 의사결정 나무분석 68%이며. 검증용 집합의 경우 로지스틱 회귀 분석 69%, 의사결정 나무분석 69%로 로지스틱 회귀분석이 더 우수한 것으로 나타났다.

### 5. 결론

본 연구에서는 인터넷 쇼핑물 고객들의 구매품목 데이터를 통계적 분석기법인 다차원 척도법과 대응

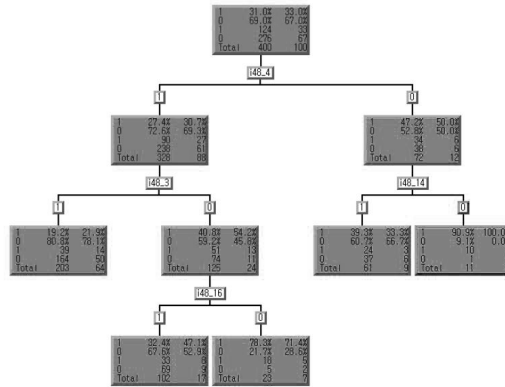


그림 4.7 의사결정나무 모형

표 4.9 로지스틱 회귀분석과 의사결정 나무분석 비교

적중률	로지스틱 회귀분석	의사결정 나무분석
훈련용 집합의 적중률	68.75%	68%
검증용 집합의 적중률	69%	69%

분석을 통해 구매품목들과 인구통계학적 변수들 간의 유사성을 비교분석하였다. 또한, 이를 바탕으로 기존고객들을 우수고객, 비 우수고객 집단으로 세분화하기 위한 예측모델을 설계하기 위해 로지스틱 회귀분석과 의사결정나무 분석을 이용해 데이터마이닝 예측모형을 구축하였다. 그리고, 고객 개개인의 특성에 맞는 마케팅 프로모션을 기업에 제공하기 위해 연구를 실시하게 되었으며 다음과 같은 결과를 얻었다.

첫째, 인구통계학적 변수들과 구매품목간의 연관성을 분석한 결과, 남성이 여성에 비해 구매품목이 많은 것으로 분석되었다. 10대는 대부분의 품목들을 선호했으며, 20대는 금융서비스, 30대는 정보기술, 40대는 기타, 50대는 아동유아와 소프트웨어를 선호하는 것으로 나타났고 60, 70대는 저조한 구매양상을 보였다. 학력이 초졸이하를 제외한 모든 학력들이 대부분의 구매품목을 선호했으며. 모든 월평균 소득수준에서 정보기술을 선호하는 것으로 나타났다.

둘째, 각 구매품목들의 유사성을 이차원 공간상에 시각적으로 나타내기 위해 다차원 척도법과 대응분석을 수행하였다. 다차원 척도법 수행결과, 구매품목들 간의 유사도가 높을수록 가까운 거리에 위치하고 있었으며, 의류와 도서, 가전제품과 컴퓨터는 거리가 매우 가까우므로 상관관계가 높음을 알 수 있었고, 음악은 다른 품목들과 가장 거리가 멀기 때문에 낮은 상관관계를 가지고 있음을 알 수 있었다. 대응분석 수행결과, 성별과 월평균 소득에 따라 구매품목들 간에는 큰 차이가 없었으며 연령에 따른 구매품목의 대응분석 결과 70대를 제외한 모든 연령들이 구매품목들과 대응하고 있었다. 학력에 따른 대응분석 결과 중졸, 고졸, 대졸인 경우 구매품목들과 대응하고 있었으며 다른 학력들은 상대적으로 멀리 떨어져 있었다.

마지막으로, 인터넷 쇼핑몰 고객세분화를 위해 이용한 데이터마이닝 분석 기법인 로지스틱 회귀분석과 의사결정나무 분석결과는 다음과 같다. 로지스틱 회귀분석을 수행한 경우 최대 68.75%의 예측 적중률을 보였으며, 의사결정나무 분석을 수행한 경우 최대 68%의 예측 적중률을 보였다.

향후 연구과제는 다차원 척도법과 대응 분석 간의 관계를 명확히 규명하고 더욱 방대한 인터넷 쇼핑

물 고객 데이터를 확보하여 예측력이 높은 고객 세분화모형을 구축하여 신규고객보다는 기존고객 유지에 비중을 두는 고객 유지 마케팅을 마련할 필요가 있다.

### 참고문헌

- 김규일, 김승한, 김은영, 김현, 양재완, 조장식 (2007). K대학의 입학시험 1차 합격자들에 대한 이탈 분석. <한국데이터정보과학회지>, **18**, 1-10.
- 김선남, 이영원 (2007). 수용자의 인터넷 쇼핑 동기에 관한 연구. <인문과학연구>, **7**, 81-110.
- 김성섭, 변효근, 김호재 (2002). 해외신혼여행지의 포지셔닝 분석 및 시사점에 관한 연구. <관광레저연구>, **14**, 2002.
- 김차용, 이철화 (1995). 도시 기혼여성의 스포츠 참여종목과 참여동기의 대응분석. <한국사회체육학회지>, **4**, 193-206.
- 노대협, 한인구 (2002). 인터넷 비즈니스 환경 하에서의 고객관계관리기법. <한국통신학회>, **12**, 2002.
- 이병엽, 조규하, 송석일, 유재수 (2005). 통신 산업의 고객 분류를 위한 예측 모델 설계. <한국콘텐츠학회논문지>, **6**.
- 이승희, 정진원 (2005). 인터넷 쇼핑 증독구매에 관한 연구모형. <한국의류학회지>, **29**, 167-176.
- 이현우 (2006). 국내 콜센터의 데이터마이닝 응용. <한국데이터정보과학회지>, **17**, 335-344.
- 정영수, 강경화 (2004). 데이터마이닝 기법을 이용한 인터넷 쇼핑몰 사이트의 CRM 사례분석. <경영경제연구>, **27**.
- 조범석 (2002). 금융권 CRM 구축에서의 데이터 마이닝 사례연구. <대은경제리뷰>, 2002.
- 진서훈 (2005). 데이터마이닝에 의한 고객세분화 개발. <응용통계연구>, **18**, 555-565.
- 통계청. <2007년 12월 및 연간 사이버쇼핑물 통계조사>.
- Meler, M. and Ruzic, D. (1999). Marketing identity of the tourism product of the Republic of Croatia. *Tourism Management*, **20**, 635-643.

## Predictive model plan of customer using purchasing items in internet shopping mall

Hye-Young Ji<sup>1</sup> · Wan-Hyun Cho<sup>2</sup>

<sup>1,2</sup>Department of Statistics, Chonnam National University

Received 8 December 2008, revised 23 December 2008, accepted 26 December 2008

### Abstract

Recently, according to the epoch-making advancement of the Internet technique, internet using is widely expanded to the social whole not only until quantitative growth but also until qualitative growth. This research is aimed to offer plan about segmentation strategy which could be applied to business strategic establishment and the academic research, and information about solution method. In this paper, we compared similarity among purchasing products by using statistical methods such as positioning and correspondence analysis, and we tried to design predictive model for segmentation of existing customer. In conclusion, we have objective that enterprises create a benefit from the stabilized customer and customers have a benefit from enterprises by providing marketing promotion which fits the property of each person.

*Keywords:* Correspondence analysis, customer segmentation, internet shopping, positioning.

---

<sup>1</sup> Graduate student, Department of Statistics, Chonnam National University, Gwangju 500-757, Korea.

<sup>2</sup> Corresponding author: Professor, Department of Statistics, Chonnam National University, Gwangju 500-757, Korea. E-mail: whcho@chonnam.ac.kr