

사례기반추론을 이용한 전통적 컨조인트 분석의 불완전한 응답 자료의 효용추정방법에 관한 연구*

서상윤(주저자)

경남대학교 경영학부

(syseo@kyungnam.ac.kr)

본 연구는 전통적 컨조인트 분석에서 서열척도로 상품 프로필을 평가할 때 발생하는 불완전한 응답 자료의 문제점을 사례기반추론을 이용하여 보완하고 그 응답자의 효용을 추정하는 방법에 대해 설명한다. 사례기반추론은 서열 변수들 간의 연관관계를 파악하는 켄달의 서열 상관계수와 스피어만의 서열 상관계수를 이용하여 추론에 활용할 자료를 검색하였다. 그리고 유사도가 높은 응답자의 속성 수준별 효용 값을 이용하여 불완전한 컨조인트 응답자료의 효용 값을 계산하고, 효용 값의 크기에 따라 불완전 응답 자료의 순위를 입력하여 효용을 추정하였다. 분석결과 불완전한 응답 자료와 추론에 사용된 자료 간의 유사성을 나타내는 상관계수 값이 클수록 자료의 복원을 통해 도출된 컨조인트 효용모형의 응답자 선택에 대한 예측정확도는 높아졌다. 그리고 유사성 검색에 사용된 방법 간에는 켄달의 서열 상관계수를 이용한 추론방법이 스피어만 서열 상관계수 보다 높은 예측정확도를 보였다. 끝으로 상관계수의 크기와 추론에 활용된 자료의 수는 상관계수의 크기와 상호작용하여 예측정확도를 더욱 높이는 것으로 나타났다.

주제어: 사례기반추론, 전통적 컨조인트, 불완전 응답자료, 켄달의 서열 상관계수, 스피어만 서열 상관계수

1. 서론

컨조인트 분석은 상품의 특성을 나타내는 프로필을 이용하여 상품 속성에 대한 소비자의 효용과 상품 평가시 중요하게 고려하는 속성을 파악하는 대표적인 방법으로 활용되고 있다. 그러나 컨조인트 분석의 이러한 유용성에도 불구하고 컨조인트 분석의 일부 문제점은 컨조인트 분석의 활용을 저해하고 있다. 이러한 제약사항 가운데 하나는 응답자들이 다수의 제품 프로필들에 대한 선호도 순위를 정확하게 평가하기가 어렵다는 점이다. 특히 컨조인트 분석은 프로필에 포함되는 속성과 속성수준이 증가함에 따라 응답자가 평가해야 하는 프로필의 숫자가 늘어난다(Green and Srinivasan 1990; Curry 1997;

Green and Krieger 1991, 1992). 이로 인해 프로필의 선호순위가 중복되거나 결측값이 포함되어 분석에 제대로 활용되지 못하는 경우가 자주 발생한다. 이러한 전체프로필법을 이용한 컨조인트 분석은 속성이 7개 이상 이거나 평가용 프로필 수가 20개 이상이면 응답자가 순위를 정하는데 어려움이 있으며, 속성과 수준의 수가 너무 많은 경우 선택대안들의 대표적인 특성을 부각하기 어렵고, 조사과정에서 응답자에게 지나친 부담을 주어 신뢰성 있는 조사결과를 얻기 어려울 수도 있다고 하였다(정영우·이은용 2008).

이러한 문제점으로 컨조인트 분석에는 일반적으로 속성은 5개 이하, 속성 수준은 3~4개 정도가 적절한 것으로 알려져 있다(Green and Srinivasan 1990). 일반적으로 5개 이하의 속성과 4개 이하의

속성수준을 이용하여 프로필을 직교설계 하는 경우 응답자가 평가하는 프로필은 보통 16개 이하로 나타난다. 그러나 이 역시 응답자들이 16개의 프로필에 포함되어 있는 속성수준을 비교해 가며 평가하는 작업은 부담스러운 일이다. 특히 전체 프로필에 대해 종합적으로 판단하여 선호순위(Rank)를 정하게 하거나, 선호하는 순서대로 카드를 정렬(Sequence)하도록 하는 것은 개별 카드의 선호점수(Rating)를 입력 하는 것 보다 더 힘든 작업일 수 있다.

그렇다면 선호점수 대신에 왜 선호순위를 이용하여 평가하도록 하는가? 그 이유는 선호점수로 평가하는 경우 선호 점수값의 차이가 크지 않을 수 있기 때문이다. 그러나 서열척도로 측정하는 경우 평가값은 차이를 갖기 때문에 효용의 추정에 있어 보다 효과적일 수 있다. 그러나 평가해야 하는 프로필의 숫자가 증가하면 그 순위를 정하는 것은 응답자에게 상당한 부담이 된다. 이로 인하여 순위척도(Rank, Sequence)로 컨조인트 응답 자료를 수집하는 경우 선호순위가 동일한 카드가 존재하거나, 중도 응답을 포기하는 경우 등 응답에 오류가 있는 불완전 응답 비율이 자주 발생되곤 한다.

본 연구에서는 이렇게 컨조인트 분석에서 응답자가 응답을 중도 포기하거나, 동점 순위가 발생하는 경우 등 컨조인트 자료가 불완전하게 수집되어 응답자의 효용을 파악할 수 없는 경우 그 응답자의 선호를 추정하는 방안에 대해 제시하고자 한다. 구체적으로 컨조인트 프로필의 선호순위를 유추할 수 있는 서열 상관분석을 이용하여 완전한 응답자의 프로필 선호 순위 자료와 불완전한 응답자의 선호순위 자료의 연관관계를 파악하고, 응답이 완전한 선호순위 자료를 활용하여 불완전한 응답자의 응답패턴을 유추한 다음 불완전한 응답자의 효용을 분석하는 방법에 대해 제시하고자 한다.

컨조인트 분석에 관한 이전 연구들은 크게 컨조인트 모형의 정확도를 높이기 위한 다양한 방법론과 컨조인트 분석 단계별 절차를 중심으로 여러 방법들이 서로 비교 연구되었다(Gustafsson, Herrmann, and Huber 2007). 그러나 컨조인트 분석에 대한 방법론적 발전에도 불구하고 연구자들은 여전히 불완전한 응답 자료의 분석에 대해서는 많은 관심을 두지 않았다. 이러한 자료는 분석에서 제외를 하는 것이 가장 일반적인 처리방법이었다. 그렇지 않는 경우 결측 자료를 보완하기 위한 방법으로 평균대체법이 많이 활용되었다. 그러나 평균대체법은 등간척도나 비율척도와 같이 계량척도(metric scale)를 이용하는 경우에는 적용할 수 있으나 서열척도(rank, sequence)로 자료를 수집한 경우에는 적용할 수 없는 단점이 있다.

최근 컨조인트 분석은 컴퓨터를 이용한 선택형 상호작용 컨조인트 분석을 많이 활용하고 있다(Johnson, Huber, and Bacon 2003). 그러나 선택형 상호작용 컨조인트 분석을 사용하기 위해서는 응답자와 컨조인트 디자인이 서로 상호작용할 수 있는 프로그램과 시스템 환경이 구비되어야 하는 관계로, 아직도 현장에서는 설문지를 이용한 전통적 컨조인트 분석 방법이 널리 활용되고 있다. 이러한 전통적 컨조인트 분석에서는 주로 서열척도를 이용하여 전체 프로필을 평가하도록 하고 있다. 그러나 서열척도로 프로필을 평가하는 경우 불완전한 컨조인트 응답을 평균값으로 대체하는 것이 불가능하다. 따라서 순위변수 간의 연관성을 파악할 수 있는 스피어만 상관계수와 켄달의 타우 값을 활용하여 응답이 불완전한 컨조인트 카드의 선호순위를 추론하고, 이들의 효용을 파악하는 방법에 대해 제안하고자 한다. 이러한 순위 상관관계를 이용한 추론 방법의 정확도는 추론을 통하여 복원된 컨조인트 효용 모형이 그 응답자

의 선택을 얼마나 잘 예측하는지, 예측타당성으로 평가하고자 한다. 그리고 사례기반추론 대신 무작위로 추출된 다른 응답자들의 평균 효용 값 정보를 활용하여 추정된 예측정확도와 비교하여 사례기반추론의 효과에 대해서도 살펴보고자 한다.

II. 이론적 고찰

2.1 컨조인트 모형

컨조인트 분석의 단점은 속성과 속성수준의 증가에 따른 응답자의 프로필 평가에 대한 부담의 가중이다. 이러한 문제점들을 해결하기 위해 다양한 컨조인트 분석방법이 제안되었다.

하이브리드 컨조인트(Hybrid Conjoint Analysis)는 자체설명식 접근법과 전통적 컨조인트 분석을 결합한 모형이다(Green, Goldberg, and Montemayor 1981). 응답자는 각 속성에 대한 중요도와 속성수준의 매력도를 평가하고, 일부 컨조인트 프로필에 대한 선호도를 평가한다. 속성수준별 부분가치 추정은 자체설명식 접근법에 의해 산출된 효용값과 속성수준을 더미 코딩한 변수 값을 독립변수로 입력하고, 응답자가 평가한 개별 컨조인트 프로필의 선호점수를 종속변수로 하는 회귀모형으로 분석된다. 응답자들은 전체 프로필 가운데 일부 프로필만을 평가하기 때문에 집단수준(aggregate level)의 부분가치만 추정되며 개별 응답자 수준에서 부분가치를 추정하기 위해서는 전통적 컨조인트 분석에서와 같이 전체 프로필을 모두 평가하여야 한다.

적응형 컨조인트(Adaptive Conjoint Analysis)은 최근 가장 널리 활용되고 있는 컨조인트 분석 방

법 가운데 하나이다(Johnson 1987). 적응형 컨조인트 분석은 PC를 이용하여 응답자의 반응에 따라 필요한 질문을 선택적으로 제시하기 때문에 적응형 컨조인트 라고 한다. 따라서 적응형 컨조인트 분석은 기본적으로 컴퓨터를 기반으로 진행된다. 적응형 컨조인트 분석은 많게는 30개 속성과 15개 속성수준 까지 분석에 사용할 수 있는 장점이 있으나, 이러한 여러 속성과 속성수준을 반영하기 위하여 자체설명식 접근법이 활용된다. 그러나 속성의 중요도 평가 방식이 “속성이 얼마나 중요한가?”와 같이 직접 질문하는 대신 다른 모든 조건이 서로 동일한 경우 두 비교 속성수준이 서로 다르다는 것이 얼마나 중요한지 물어서 평가한다는 점에 차이가 있다. 즉 중요한 속성일 경우 속성수준의 차이는 중요하다 생각할 것이며, 중요하지 않은 속성일 경우 속성수준의 차이는 크게 중요하지 않을 수 있다. 그리고 상품 컨셉에 대한 선호 대응비교(paired comparison)를 통해 이전 단계에서 수집한 속성수준별 선호점수와 속성의 중요도를 이용하여 응답자의 효용을 추정한다.

컨조인트 분석에 확률효과모형(random effect model)을 적용하면 응답자들의 선호체계에 대한 이질성(heterogeneity)을 반영하여 개별 응답자 수준에서 부분가치를 추정할 수 있다(Lenk, Desarbo, Green, and Young 1996). 식(1)과 식(2)는 개별 응답자들의 부분가치의 다양성을 반영하는 확률효과모형이다. 여기서 Y_i 응답자 i 의 프로필에 대한 선호도 평가를 의미하며, X_i 는 프로필 디자인 행렬, β_i 는 회귀계수(부분가치)행렬을 나타낸다. 식(2)는 개별 응답자 수준에서 부분가치의 이질성을 반영하는 식이다. 그러나 확률효과모형은 개별 응답의 회귀계수 혹은 부분가치에 대해 기본적으로 확률분포를 가정한다. 즉 개별 응답자의 부분가치를 나타내는 회귀계수의 분포에 대해 오차항(δ_i)은 상호 독립적

며 평균이 0이고, 공분산 행렬이 Λ 인 다변량 정규분포를 따른다고 가정한다($\delta_i \sim N(0, \Lambda)$).

$$Y_i = X_i\beta_i + \varepsilon_i \quad \text{for } i = 1, \dots, n \quad \text{식(1)}$$

$$\beta_i = \theta z_i + \delta_i \quad \text{for } i = 1, \dots, n \quad \text{식(2)}$$

따라서 확률효과모형을 이용하는 경우 응답자가 일부 킨조인트 프로필에 대해서만 응답을 하여도 개별 응답자 수준에서 효용추정이 가능하다. 즉 킨조인트 프로필에 결측자료가 포함되는 경우에도 부분가치 추정이 가능하다. 그러나 사례기반추론을 이용하는 경우 유사한 응답 자료로부터 부분가치에 대한 정보를 획득 하므로 확률효과모형과 달리 응답자간 부분가치의 분포에 대한 가정을 필요로 하지 않아 보다 현실적이다. 또한 하이브리드 킨조인트나 적응형 킨조인트의 경우 기본적으로 전체 수준에서 부분가치를 추정하기 때문에 일부 프로필에 대한 평가자료가 없는 경우 개별 응답자 수준에서 효용추정이 불가능할 수 있다.

2.2 불완전 응답 자료의 대체법

킨조인트 분석에서 불완전한 응답은 여러 가지 이유로 발생된다. 응답자의 실수에 의한 응답 누락이나 중복 순위의 부여, 평가의 어려움으로 인한 선택 회피 등 다양한 이유로 발생된다. 특히 비교대안에 대한 선호도가 비슷하여 응답자가 선호대안의 선택을 거부하는 선택 회피는 결측값(missing value)과는 다른 정보를 포함하기도 한다.

그러나 킨조인트 분석이 소개된 1970년대 이후 킨조인트 분석의 연구 흐름을 분석한 주요 연구들(Green, Krieger, and Wind 2001)을 살펴보면 불완전한 킨조인트 응답 자료의 처리방법에 관한 연

구는 본 연구자가 살펴본 바에 의하면 거의 연구가 이루어지지 않았다. 이는 불완전한 킨조인트 응답 자료 또한 일반적인 결측 자료의 처리방법을 적용할 수 있기 때문일 것이다.

따라서 킨조인트 분석에서 발생될 수 있는 불완전 응답 자료는 자료의 수집방법에 따라 여러 가지 방법으로 처리될 수 있다. 결측값을 삭제하고 완전한 응답 자료만을 이용하여 분석하는 방법(complete case analysis), 평균대체, 중앙값대체, 최근방대체법(nearest neighbor imputation), 회귀모형을 이용하여 결측값을 추정하는 방법(regression imputation) 등 다양하다.

완전한 응답 자료만 이용한 분석(complete case analysis)은 이용이 간편하고, 결측치가 완전히 무작위로 발생하는 경우 모수 추정치에 편향(bias)이 거의 발생하지 않은 장점이 있지만, 표본 수 감소에 따른 정보 손실과 검정력이 약화되며, 결측치가 완전히 무작위로 발생하지 않은 경우 모수 추정치에 편향의 발생가능성이 있다.

평균대체법은 표본을 몇 개의 집단으로 구분한 후 각 집단의 평균값으로 대체하는 방법이다. 이 방법은 편의성 측면에서 실제로 결측값을 대체하는 보편적인 방법으로 활용되지만 항목변수가 양적변수인 경우만 적용할 수 있고, 구하고자 하는 통계량 값이 평균값일 경우 유용하지만, 대체 후 평균값의 분포가 많아져 응답자들의 분포가 왜곡될 가능성이 높은 문제점을 가지고 있다. 따라서 대체군 집단 내에서 응답값의 분포와 결측값이 발생한 집단의 분포가 서로 동일하여야 한다.

최근방대체법(nearest neighbor imputation)은 결측 자료 처리를 위해 결측 자료 단위와 가장 유사한 응답 조사 단위를 찾아, 그에 대응되는 값 또는 유사 응답자들의 평균값으로 결측값을 대체하는 방법

이다. 이때 유사한 응답 단위는 거리함수를 이용하여 찾을 수 있다(Sande 1979; Vacek and Ashikago 1980). 즉 응답자간 유클리디언 거리를 측정하여 최근방 응답자의 자료로 대체하여 활용하는 방법이다(Chen, and Shao 2000). 따라서 최근방대체법을 적용할 수 있기 위해서는 자료들이 계량적 변수(metric variable)이어야 한다.

회귀모형을 이용한 대체법은 우선 자료를 완전한 응답 자료와 결측값을 가진 자료로 구분하여, 완전한 응답 자료를 이용하여 결측값에 해당하는 결과변수를 예측하는 회귀모형을 개발한다. 그리고 이 회귀모형을 결측값 자료가 포함된 응답자에 적용하여 결측값을 추정하는 방법이다. 이 방법은 결측값이 연속형 변수인 경우 일반 선형회귀모형을 이용하여 예측하고, 이산형 변수인 경우에는 다항 로짓모형을 이용하여 추정할 수 있다. 회귀모형을 이용한 결측값 대체법은 추정에 이용되는 독립변수와 결과변수 간의 연관성이 있는 높은 경우에 효과적인 방법이다. 그러나 전통적 컨조인트 분석과 같이 응답자들이 프로필의 선호순위를 평가하는 경우 프로필의 선호 순위 간에는 인과관계가 성립한다고 보기 어렵기 때문에 회귀모형을 이용한 결측값 대체법 역시 적용이 어렵다.

2.3 사례기반추론의 개념

불완전한 자료의 복원에 주로 활용되는 평균대체법, 회귀모형 대체법, 최근방 대체법 등은 서열척도로 측정된 자료의 결측값 대체에 여러 한계점이 있음을 알 수 있다. 반면 사례기반추론(Case-Based Reasoning)이란 과거에 있었던 사례(케이스)들의 결과를 바탕으로 새로운 사례의 결과를 예측하는 기법이다. 사례기반추론은 Exemplar-based Reasoning,

Instance-Based Reasoning, Memory-based Reasoning 등 다양한 용어로 사용되고 있다(이재식·명훈식 2008). 사례기반추론의 기본적인 사고는 유사한 사례의 문제는 유사한 사례의 해법으로 풀 수 있다는 것이고, 그와 유사한 사례는 여러 번 반복되어 나타날 수 있다는 것이다. 즉 과거의 경험을 바탕으로 그와 유사한 문제의 해결책을 추론할 수 있다는 것이다.

Aamodt and Plaza(1996)은 사례기반추론의 과정을 검색, 재사용, 수정, 유지의 네 단계로 제시하였다. 검색은 현재 문제와 가장 유사한 과거의 사례를 찾는 것이며, 재사용은 검색을 통해 찾은 유사한 사례의 해법과 결과를 현재의 문제해결에 사용 및 적용하는 것이다. 이때 현재 문제의 해결을 위해 활용되는 과거의 해법을 적당한 형태로 조정하는 수정 작업을 거치게 된다. 그리고 해결된 문제는 다시 새로운 문제의 해결을 위한 목적으로 사례기반 데이터 베이스에 저장하는 유지 단계를 거치게 된다.

현재 이와 같은 사례기반추론은 상품추천에서 널리 활용되고 있다. 예를 들면 아마존(Amazon)과 같은 온라인 서점이나, 온라인 쇼핑몰에서는 소비자들이 과거에 구매했던 내역을 바탕으로 그와 유사한 제품을 구매한 소비자에게 교차판매 상품을 추천해주고 있다. 따라서 사례기반추론이 효과적으로 활용되기 위해서는 검색에 활용되는 충분한 사례들이 확보 되어 있어야 하며, 발생한 문제와 유사한 사례를 검색할 수 있는 알고리즘이 필요하다. 즉 사례기반추론을 이용하기 위해서는 일반적으로 사례와 사례들 사이의 유사 정도를 측정하기 위한 유사도 척도가 준비되어야 한다. 본 연구에서는 이러한 유사 사례를 검색하는 알고리즘으로 순위자료 간의 일치 정도를 평가할 수 있는 서열 상관계수를 이용한다.

2.4 컨조인트 프로파일 선호순위 평가와 서열상관계수

전통적 컨조인트 분석은 가상의 상품 프로파일 설계로부터 실험이 진행된다. 컨조인트 실험과정은 Green and Srinivasan(1990) 연구에 잘 설명되어 있다. 전통적 컨조인트 분석의 일반적인 절차는 다음과 같다. 우선 컨조인트 실험용 프로필을 설계하기 위한 상품의 속성과 그 속성수준을 결정하는 것이다. 속성과 속성수준이 결정되면 직교설계를 이용하여 가상의 상품 프로필을 설계한다. 실험 참가자들은 가상의 상품 프로필에 대한 선호도나 효용을 순위(rank), 점수(rating) 척도를 이용하여 평가하도록 한다. 그리고 선호도 응답의 형태에 따라 통계적 모형을 적용하여 응답자의 효용 값들을 도출해 낸다. 본 연구에서는 상품 프로파일의 선호도 평가는 순위(rank)를 이용하여 측정한다.

이와 같이 상품 프로파일의 선호도를 순위로 평가하는 경우 평가가 누락되거나 중복순위 등이 발생하는 경우 그 응답자의 효용은 제대로 분석이 이루어지지 않게 된다. 따라서 사례기반추론을 이용하여 불완전 응답 자료의 문제를 해결하기 위해서는 그와 유사한 응답 패턴을 가지고 있는 사례를 검색해야 한다.

본 연구에서는 순위간의 연관관계를 파악할 수 있는 대표적인 두 가지 방법인 켄달의 순위 상관계수(Kendall, 1938)와 스피어만의 순위 상관계수(Spearman, 1904)를 중심으로 살펴본다. 켄달의 순위 상관계수는 두 순위 변수의 순서 쌍 (x_i, y_i) 와 순서 쌍 (x_j, y_j) 를 서로 비교하여 전체 조합 가능한 모든 순위 쌍 가운데 부호가 일치하는 쌍과 불일치하는 쌍의 차이 비율로 계산된다. 스피어만의 서열 상관계수는 모든 순위자료가 서로 중복되지 않은 정수인 경우 순위 변수 간의 피어슨 상관계수로 정의가 된다. 켄달의 순위 상관계수나 스피어만의 순위

상관계수 모두 두 자료의 순위가 일치할수록 “+1”에 가까운 값을 가지며, 반대 일수록 “-1”에 가까운 값을 갖는다.

컨조인트 프로필에 대한 응답자의 선호도 평가가 순위로 이루어진 경우 불완전한 응답 자료의 문제 해결을 위한 유사 사례의 검색방법으로 위에서 제시된 켄달의 순위 상관계수와 스피어만의 순위 상관계수 값이 활용될 수 있다. 즉 두 순위 상관계수 값이 +1에 가까울수록 불완전 응답 자료의 문제해결을 위해 활용될 유사한 사례는 상품 프로파일 선호 순위가 서로 일치할 가능성이 높기 때문이다.

III. 연구방법

3.1 컨조인트 실험용 프로파일 설계

본 연구는 컨조인트 실험에서 응답이 불완전한 응답자의 효용을 유사한 사례를 기반으로 추정하여 그 응답자의 선호를 예측하는 연구이다. 이러한 연구목적 달성을 위하여 학생들에게 친숙한 제품인 노트북 제품을 대상으로 컨조인트 실험을 위한 프로필을 직교 설계하였다. 프로파일 직교설계에 활용된 속성들은 응답자들이 노트북 컴퓨터를 구매할 때 중요하다고 생각하는 4가지 속성 CPU 속도, RAM 메모리 크기, 무게, 상표를 이용하였다. 프로파일 설계에 활용된 각 속성별 속성수준은 <표 1>과 같다.

직교설계 결과 총 16개의 실험용 프로필이 생성되었다. 그리고 응답에 대한 신뢰성을 평가하기 위한 검정용 카드 4개를 추가로 생성하였다. 이러한 16개의 실험용 프로필을 응답자에게 제시한 후 상품 프로필에 대한 선호도에 따라 가장 선호하는 것부터

〈표 1〉 컨조인트 실험 속성과 속성수준

속성	속성수준			
브랜드	삼성	LG	SONY	
CPU	코어 i3	코어 i5	코어 i7	
메모리	2G	4G	8G	16G
무게	1.2kg	1.5kg	2.0kg	2.5kg

선호하지 않는 것 까지 1위부터 16위 까지 순위를 부여하도록 하였다. 그리고 나머지 검정용 카드 4개에 대해서도 1위부터 4위 까지 선호순위를 부여하도록 하였다. 검정용 카드는 응답의 신뢰도를 평가하기 위한 것으로 앞에서 선택한 16개의 카드를 이용하여 소비자의 효용을 도출하고, 이를 바탕으로 4개의 검정용 카드에 대한 효용을 계산하여 순위를 정한다. 그리고 응답자가 정한 4개의 검정용 카드에 대한 순위와 비교하여 응답의 신뢰성을 평가하는 것이다.

3.2 실험조작 및 요인

본 실험에 사용할 불완전 응답 자료는 16개의 프로필에 대한 선호 순위를 평가한 자료 가운데 임의로 특정 순위 자료를 삭제하여 불완전 응답 자료를 생성하였다. 이와 같은 방식으로 응답자가 평가한 16개 순위 자료 가운데 순차적으로 제거하는 자료의 수를 증가시켜 가면서 2개부터 최종 7개 까지 자료를 제거하였다. 이를 바탕으로 사례기반추론에 활용되는 자료의 수에 따라 정확도가 어떻게 달라지는지 살펴본다. 그리고 이때 사용되는 불완전 응답 자료의 사례기반추론을 위한 자료 검색은 순위 상관계수 지표(켄달의 타우, 스피어만 서열 상관계수)를 이용하여 유사 사례를 찾아낸다. 따라서 예측정확도에 영향을 미치는 요인은 상관계수, 상관계수의 종류, 사례기반추론에 활용된 입력자료의 수 3가지 이다.

3.3 불완전한 컨조인트 응답 자료의 복원 방법

사례기반추론을 이용한 불완전 응답자료의 복원은 다음과 같은 방법으로 이루어진다. 예를 들어 컨조인트 응답자가 9개의 컨조인트 프로필 가운데 6개 카드에 대해서는 순위를 부여하였으나 나머지 3개 카드에 대해 순위를 부여하지 않고 중도에 포기하였다고 하자. 이러한 경우 응답자가 응답한 6개 카드의 정보를 활용하여, 6개 카드와 유사한 선호 패턴을 가지고 있는 응답자를 스피어만 서열 상관계수 또는 켄달의 순위 상관계수를 이용하여 검색한다. 그리고 검색된 응답자의 속성수준별 효용 값을 컨조인트 분석을 이용하여 파악한 다음 여기서 얻은 효용 값을 이용하여 불완전한 응답자가 누락한 프로필 3개 자료의 효용 값을 계산해 낸다. 3개 프로필의 효용 값 순서에 따라 누락된 나머지 순서를 차례대로 부여하여 준다. 이렇게 재추정된 선호순위 자료를 이용하여 불완전 응답자료의 속성수준별 효용을 추정한다.

응답자들의 부분가치 효용(part-worth utility) 분석은 추정모수의 분산이나 정확도에서 MONANOVA, LINMAP등과 유사한 결과를 보이는 OLS(Ordinary Least Square)방법으로 부분가치(part-worth)를 추정한다(Mishra, Umesh and Stern 1989).

3.4 사례기반추론을 이용한 추정 효용의 예측정확도

그리고 사례기반추론의 예측정확도를 산출하기 위하여 4개의 상품 프로필로 구성된 세트를 제시하고 그 가운데 가장 선호하는 상품 프로필을 한 개 선택하도록 하였다. 컨조인트 모형은 사례기반추론에 의해 불완전한 자료가 보완된 16개 프로필의 순위 정보를 이용하여 각 속성수준에 대한 효용 값을 파악하고, 이를 활용하여 앞에서 제시된 4개의 카드에 대한 효용 값을 각각 계산해 낸다. 그리고 4개 카드 중에 가장 효용 값이 높은 카드를 응답자가 선택할 것이라고 예측하는 것이다. 이때 컨조인트 모형이 계산한 가장 높은 효용 값을 갖는 카드와 응답자가 실제로 선택한 카드가 일치하면 컨조인트 모형은 응답자의 선택을 제대로 예측한 것이 된다. 그리고 이와 같이 선호 상품 프로필을 선택하도록 하는 실험을 총 네 번 반복하여 응답자가 실제로 선택한 카드와 사례기반추론에 의해 예측된 선호 카드가 서로 일치하는지 여부로 사례기반추론의 예측정확도를 측정한다. 즉 사례기반추론에 의해 예측된 선호카드와 응답자가 실제로 선택한 카드가 네 번 모두 일치하는 경우 예측정확도는 100%가 된다.

IV. 연구결과

4.1 자료수집

본 연구의 실험은 K대학교 대학생 및 대학원생을 대상으로 실시하였으며, 총 320명이 참가하였고, 응답의 신뢰도 평가를 위해 추가된 검정용 카드 4개의 피어슨 상관계수 지표가 0.9 이하인 응답자 10명을

제외하고 총 310명이 본 연구의 분석에 활용되었다. 응답자의 평균연령은 22.6세, 남성과 여성의 비율은 각 56%와 44%이었다.

4.2 서열 상관계수를 이용한 사례기반추론 분석결과

4.2.1 켄달의 서열 상관계수를 이용한 불완전 응답자의 효용 추정

불완전 응답자의 효용 추정은 다음과 같은 방법으로 진행되었다. 우선 310명 가운데 무작위로 1명을 추출하고, 이 응답자가 평가한 16개 카드의 선호순위 자료 가운데 7개 자료를 무작위로 삭제한 후 9개 카드의 선호 자료만을 보존하였다. 그리고 나머지 9개 카드에 대한 선호순위와 유사한 선호순서를 갖는 응답자를 검색하기 위하여 나머지 309명을 대상으로 켄달의 서열 상관계수 값을 계산하였다. 그리고 상관계수의 크기에 따라 검색된 유사성 순서대로 완전한 응답자의 속성수준의 효용값을 이용하여 삭제된 카드의 선호순위를 부여하여 준다. 예를 들면, 불완전한 응답자의 선호 순위 자료 가운데 15, 16 순위 자료가 생략되어 있는 경우, 나머지 14개 카드의 선호순위와 연관성이 높은 응답자의 속성수준의 효용 값을 이용하여 나머지 2개 카드에 대한 효용 값을 계산하고, 크기 순서에 따라 15등과 16등을 부여해 주는 것이다.

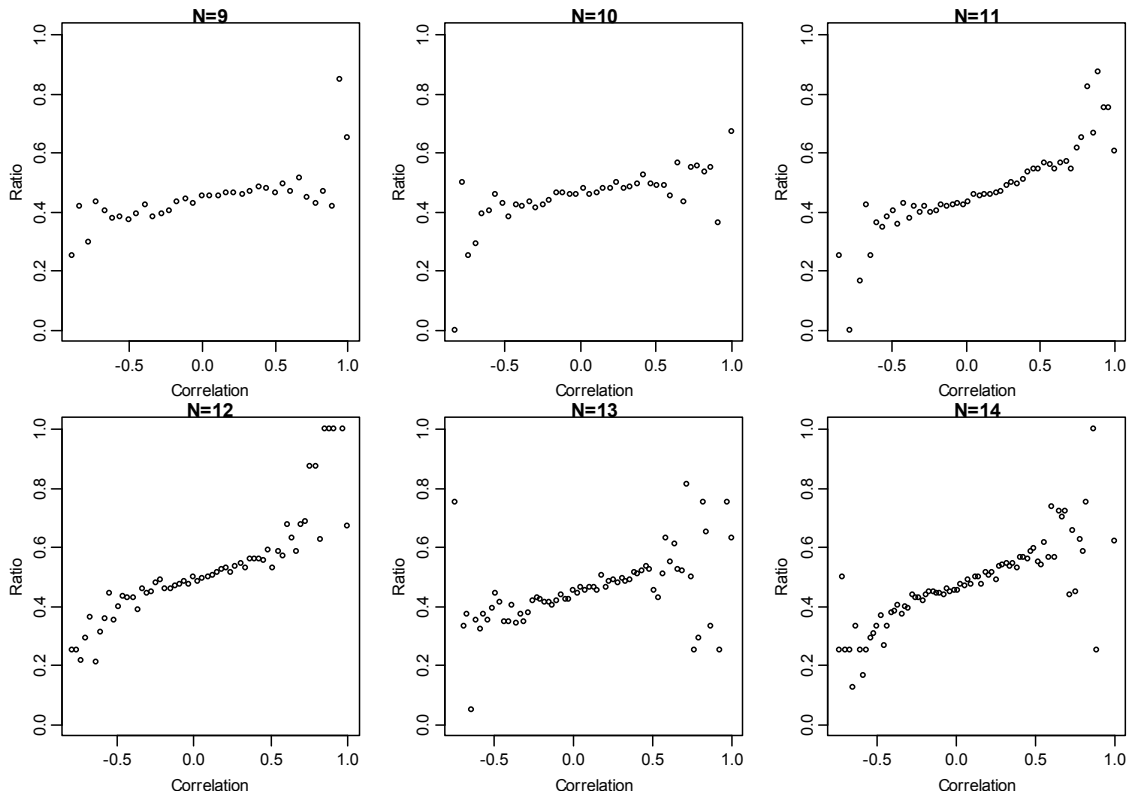
추론된 효용의 정확도는 예측정확도 측정을 위한 검정용 카드 선택 실험에서 불완전 응답자가 선택한 카드와 켄달의 서열 상관계수를 이용하여 추론한 효용 값으로 예측된 선호카드가 서로 일치하는지 여부로 예측정확도를 계산하여 평가하였다. 그리고 이러한 불완전한 응답 자료는 서로 다른 50명을 추출하여 동일한 방법으로 반복하였다. 따라서 총 50명 ×

309번 = 15,450번의 시뮬레이션이 이루어졌으며, 이러한 시뮬레이션을 9개 카드를 활용(7개 선호순위 삭제)하는 경우부터 14개의 카드를 활용(2개 선호순위 삭제)하는 경우 까지 입력자료의 수를 늘려가면서 시뮬레이션 하였다. 이와 같은 반복적인 시뮬레이션은 R 프로그램을 이용하여 수행하였다.

시뮬레이션 결과는 각 상관계수 값 별로 예측정확도의 평균값을 산출하여 산점도로 나타내었다. <그림 1>은 X축에 켄달의 서열 상관계수 값을, Y축에 사례기반추론을 바탕으로 예측한 불완전 응답자의 효용에 대한 예측정확도를 나타내고 있다. 분석결과를 <그림 1>에서 보는 바와 같이 사례기반추론에 활용

된 입력 자료의 수와 관계없이 모든 경우에 있어 상관계수 값이 커질수록 예측정확도가 증가하는 것을 알 수 있다. 그리고 입력자료의 수가 적을 때(N=9) 보다 입력자료의 수가 많을 때(N=14) 사례기반추론에 의한 예측정확도는 더욱 높아지고 있는 것으로 나타나고 있다.

상관계수 값에 따른 예측정확도의 차이를 보다 명확하게 비교하기 위하여 서열 상관계수의 크기에 따라 상관계수 값이 0 이하인 집단, 0~0.3 미만 집단, 0.3~0.6미만, 0.6~1.0 총 4개의 집단으로 구분하였다. 그리고 입력 자료의 수와 상관계수의 크기를 독립변수로 예측정확도를 종속변수로 하여 공



<그림 1> 켄달의 상관계수 값을 이용한 추론과 응답자 효용의 예측정확도

〈표 2〉 사례기반추론 활용카드 숫자와 켄달 상관계수의 예측정확도 상호작용 효과

Variables	Sum Sq	Df	Mean Sq	F	p-value
(1)	5.76	1	5.76	73.27	< 2e-16
(2)	105	3	34.98	445.29	< 2e-16
(1)×(2)	7	3	2.33	29.66	< 2e-16
residuals	7415	92692	0.08		

(1) the number of profile used in reasoning (2) kendall's coefficient categorization

변량 분산분석을 실시하였다. 〈표 2〉에서 (1) 추론에 사용된 프로파일 카드의 수는 9개~14개 까지 연속형 변수이기 때문에 공변량의 자유도가 1이 되었으며, (2) 켄달의 상관계수 자유도는 결과의 차이를 명확히 보여주기 위하여 상관계수 값에 따라 총 4개 구간으로 분할하였다(상관계수 값이 0 이하인 집단, 0~0.3 집단, 0~0.6 집단, 0.6~1.0 집단). 0 미만을 한 집단으로 구분한 이유는 상관계수가 0이하인 집단은 순서가 반대이거나 상관관계가 없기 때문에 추론정보로 적합하지 않기 때문이다. 상호작용 효과의 자유도는 공변량 자유도(1)×상관계수 집단 자유도(3) = 3 이다

분석결과는 〈표 2〉에 제시되어 있다. 통계적 검증 결과 역시 사례기반추론에 활용된 입력자료 수의 주효과는 통계적으로 유의하였으며($F=73.27, p<.000$), 상관계수의 주효과 역시 통계적으로 유의하였다($F=445.29, p<.000$). 특히 입력자료의 수와 상관계수는 서로 유의한 상호작용 효과가 존재하였다($F=29.66, p<.000$). 이는 입력자료의 수가 많고 상관계수가 높을 때 예측정확도가 더욱 상승함을 의미한다.

4.2.2 스피어만의 서열 상관계수를 이용한 불완전 응답자의 효용 추정

다음으로 스피어만 서열 상관계수를 이용한 사례기

반추론은 앞서 설명한 켄달의 서열 상관계수 값을 이용한 추론과 동일한 조건 하에서, 유사한 사례 검색 방법만 스피어만의 서열 상관계수를 이용하여 수행하였다.

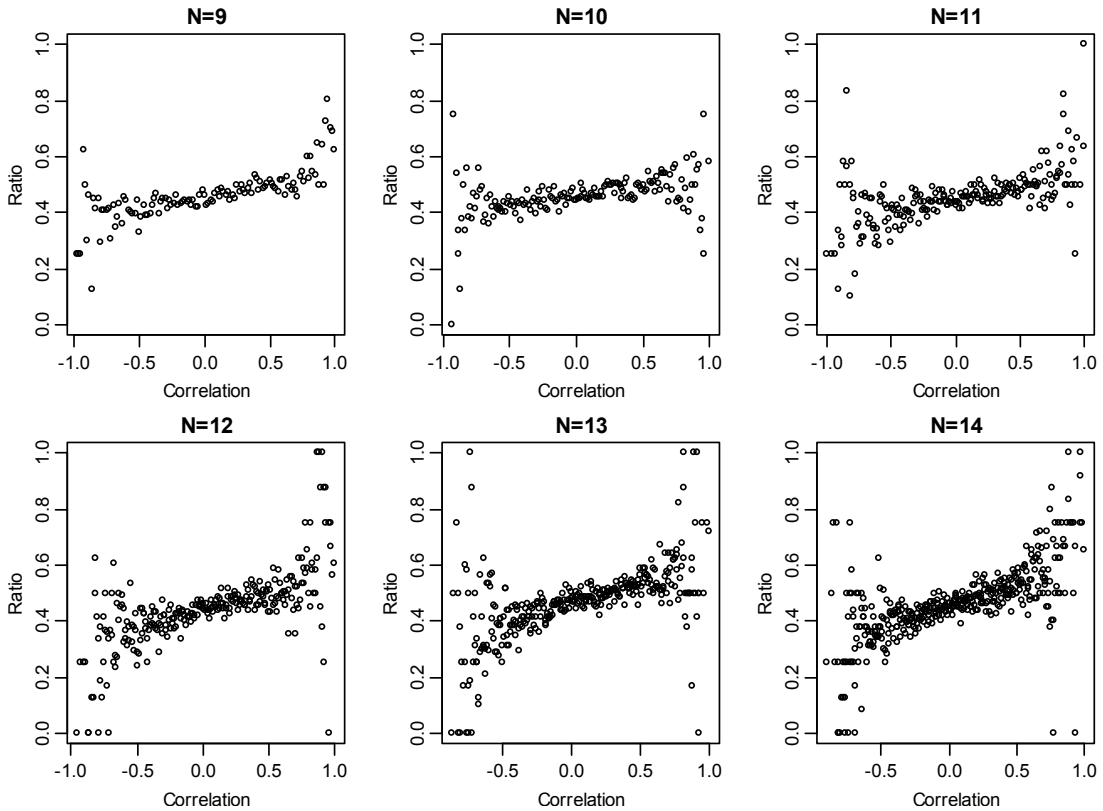
분석결과는 〈그림 2〉와 같이 X축에 서열 상관계수 값을 Y축에 사례기반추론에 의한 예측정확도를 표시하는 방법으로 산점도에 나타내었다. 〈그림 2〉의 분석결과는 켄달의 서열 상관계수를 이용한 방법과 유사한 형태를 보여주고 있다. 입력자료의 수와 상관없이 전반적으로 상관계수 값이 증가할수록 예측정확도도 상승하는 형태를 보이고 있다. 그리고 사례기반추론에 활용되는 입력자료의 수가 많아질수록 예측정확도의 증가 패턴이 더욱 명확하게 드러나고 있다. 상관계수 값이 -1과 +1사이에서 예측정확도의 분산이 크게 나타나는 것은 상관계수 절대값이 클수록, 즉 서로 일치하는 응답을 가진, 또는 정반대의 선호 서열을 가진 사례를 발견하기가 어렵고 이로 인해 사례기반추론에 이용되는 케이스가 적어 평균 예측정확도의 분산이 크게 나타나는 것이다. 즉 표본의 크기와 표본 평균의 분산은 반비례관계에 있기 때문이다.

이에 대한 통계적 검증 결과 역시 입력자료 수의 예측정확도에 대한 주효과는 통계적으로 유의하였으며($F = 13.72, p<.000$), 상관계수의 주효과 역시 통계적으로 유의하였다($F = 456.06, p<.000$).

그리고 입력자료 수와 상관계수는 상호작용 효과가 존재하여($F=20.64, p < .000$) 입력자료의 수가 증가하고 상관계수 값이 커짐에 따라 예측 정확도는 더욱 상승함을 의미한다(표 3).

4.2.3 사례기반추론의 검색 방법 간 비교

본 연구의 사례기반추론을 위해 검색에 적용된 두 가지 서열 상관계수의 효과를 비교하기 위하여 두 가



〈그림 2〉 스피어만 상관계수 값을 이용한 추론과 응답자 효용의 예측정확도

〈표 3〉 사례기반추론 활용카드 숫자와 스피어만 상관계수의 예측정확도 상호작용 효과

Variables	Sum Sq	Df	Mean Sq	F	p-value
(1)	1.03	1	1.03	13.7	0.000212
(2)	103	3	34.3	456.6	3.1E-293
(1)×(2)	5	3	1.55	20.5	2.67E-13
residuals	6994	92692	0.08		

(1) the number of profiles used in reasoning (2) spearman coefficient categorization

지 방법론을 추가적인 독립변수로 하여 공변량 분산 분석을 실시하였다. 즉 입력변수는 입력자료의 수, 상관계수의 크기(4개 집단), 추론 알고리즘의 유형(켄달 vs 스피어만)이고, 결과변수는 예측정확도이다.

분석 결과는 <표 4>와 같이 입력자료의 수($F=72.903, p<.000$), 상관계수 크기($F=882.108, p<.000$)와 추론 알고리즘의 유형($F=444.841, p<.000$)의 주효과가 모두 통계적으로 유의한 것으로 나타났다. 이러한 결과는 추론에 활용되는 카드의 수가 많을수록 예측정확도가 상승하며, 상관계수 값 역시 높을수록 예측정확도가 높아진다는 것을 의미한다. 추론 알고리즘의 주효과 역시 유의한 것으로 나타나 켄달의 상관계수를 이용한 추론($M=46.6\%$)이 스피어만의 상관계수를 이용한 방법($M=46.1\%$)보다 평균 예측정확도가 높게 나타났다. 두 방법 간 예측 정확도의 차이는 $0.5\%p$ 이며 통계적으로 유의한 것으로 나타났다($t = 3.69, p < .000$). 다만 켄달의 상관계수와 스피어만 상관계수는 그 계산 방법에 있어 서로 차이가 있기 때문에 동일한 상관계수 값에서 두 방법론의 효과를 비교하는 것은 타당성이 낮고 두 방법론 간의 예측정확도의 차이 비교만 의미가 있을 것으로 생각된다.

<그림 3>은 켄달의 서열 상관계수와 스피어만의 서열 상관계수를 이용하여 응답자의 효용을 산출하고 이를 바탕으로 계산한 예측정확도를 나타낸 것이다. 두 방법 모두 상관계수가 커질수록 예측정확도는 상승하고 있는 것으로 나타났다.

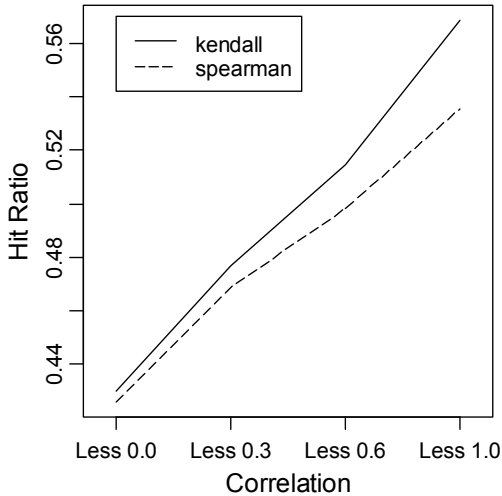
4.2.4 사례기반추론을 이용한 효용추정의 타당성 검토

사례기반추론을 이용한 불완전 응답자의 효용추정 방법의 효과성을 검토하기 위하여 프로필을 제거하지 않은 컨조인트 모형의 예측정확도와 사례기반추론으로 복원된 컨조인트 모형의 예측정확도를 서로 비교하였다. 시뮬레이션에 사용된 50명의 응답자 평균 예측정확도는 약 58%였으며, 전체 응답자 310명의 예측정확도는 64.7%, 표준오차 0.016으로 나타났다. <그림 1>과 <그림 2>에서 사례기반추론을 이용하여 복원된 자료로 도출된 효용값의 예측정확도를 프로필을 제거하지 않은 컨조인트 분석결과와의 예측정확도와 비교하면 사례기반추론에 사용된 입력자료의 수와 상관없이 켄달 또는 스피어만 상관계수가 0.5이하인 경우 사례기반추론을 이용하여 추정

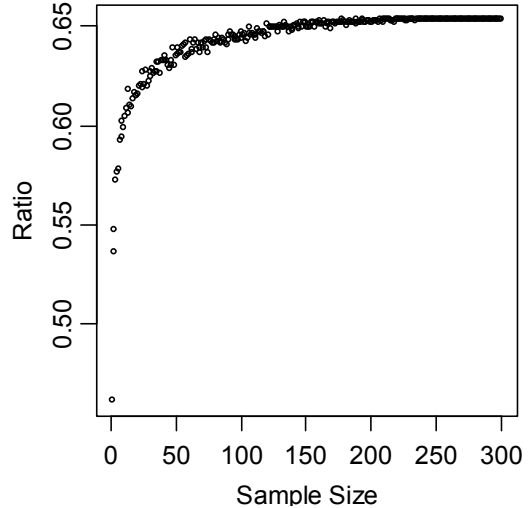
<표 4> 사례기반추론 입력자료의 수, 상관계수, 추론 알고리즘 유형의 상호작용 효과

Variables	Sum Sq	Df	Mean Sq	F	p-value
(1)	5.6	1	5.6	72.903	< 2e-16
(2)	203.5	3	67.82	882.108	< 2e-16
(3)	3.5	1	3.45	44.841	2.14E-11
(1) × (2)	9.8	3	3.28	42.669	< 2e-16
(1) × (3)	1.8	1	1.76	22.885	1.72E-06
(2) × (3)	2.4	3	0.79	10.231	9.88E-07
(1) × (2) × (3)	0.9	3	0.29	3.773	0.0101
Residuals	14831	185384	0.08		

(1) quantity of card (2) coefficient categorization (3) reasoning methodologies



〈그림 3〉 상관분석 유형에 따른 예측정확도



〈그림 4〉 표본의 크기에 따른 예측정확도 변화

효용값의 예측정확도는 평균 예측정확도를 하회하였으나, 사례기반추론에 사용된 자료가 12개 이상이고 (12/16=75%), 상관계수가 0.75 이상이면 대략 평균 예측정확도에 준하는 정도의 예측정확도를 보여주고 있다.

그리고 사례기반추론을 이용한 불완전 응답자 효용추정 방법의 타당성을 검토하기 위하여 추가로 다른 상품을 대상으로 동일한 알고리즘을 적용하여 불완전한 컨조인트 응답자의 효용을 추정하였다. 상품은 라면제품을 대상으로 하였고, 속성은 상표(S사, N사, O사: 실험에서는 실제 제조사명을 노출함), 맛(소고기맛, 해물맛), 매운정도(매운맛, 약간매운맛, 순한맛), 면의 종류(생면, 유당면)를 이용하였다. 총 16개의 가상의 상품 프로필이 직교설계에 의해 생성되었으며, 응답자들에게 16개 제품에 대한 선호순위를 정하도록 하였다. 그리고 효용추정 결과의 예측정확도를 측정하기 위해 4개의 제품으로 구성된 카드셋트를 준비하여 제시하고, 4개 중 가장 선호하는 제품을 한 개 선택하도록 하였다. 이러한 선택 실험

을 4회 반복하여 예측정확도를 측정하였다. 켄달의 상관계수와 스피어만 상관계수 값을 이용한 불완전 응답자의 추정효용에 대한 예측정확도는 부록 〈그림 5〉와 〈그림 6〉에 제시되어 있다. 상관계수의 크기에 따른 예측정확도의 변화는 본 실험의 결과와 유사한 결과를 보였다. 상관계수 높은 응답자의 자료를 활용하여 효용을 추정하는 경우 예측정확도가 높게 나타났다(F=1125.1, $p < .000$), 또한 사례기반추론에 사용된 자료의 수가 많을수록 정확도 역시 증가하였다(F=13.9, $p < .000$). 이와 같이 서열상관계수를 이용한 사례기반추론으로 불완전 응답자의 효용을 추정하는 방법은 다른 상품에서도 그 효용성이 있는 것으로 나타났다.

4.2.5 사례기반추론을 활용할 수 없는 경우

사례기반추론의 경우 응답자가 평가한 일부 자료를 활용하여 유사한 사례를 검색한 후 불완전한 응답자의 선호를 추정하게 된다. 그러나 응답자의 컨

〈표 5〉 사례기반추론을 이용한 효용추정결과의 재검증

Variables	Sum Sq	Df	Mean Sq	F	p-value
(1)	0.8	1	0.78	13.9	0.000188
(2)	189.3	3	63.11	1125.1	< 2e-16
(3)	3.8	1	3.81	67.9	< 2e-16
(1) × (2)	1.5	3	0.51	9.0	5.84E-06
(1) × (3)	0	1	0	0.0	0.840441
(2) × (3)	3.7	3	1.25	22.2	2.38E-14
(1) × (2) × (3)	0.6	3	0.21	3.7	0.01121
Residuals	1533.8	27344	0.06		

(1) quantity of card (2) coefficient categorization (3) reasoning methodologies

조인트 평가 자료가 전혀 없는 경우에는 다른 응답자들의 평균 효용을 이용하여 추정할 수밖에 없다. 즉 소비자들의 선호는 서로 비슷할 것이라고 가정하는 것이다.

이와 같은 경우 개별 응답자의 효용 추정시 전체 응답자 310명 가운데 일부를 표본으로 추출하여 그 표본의 평균 효용 값을 이용하여 개별 응답자들의 효용을 예측할 수 있다. 이러한 방법은 개별 응답자의 선호 이질성을 고려하지 않고, 시장에서 평균적으로 선호도가 높게 나타나는 속성 수준을 소비자에게 추천하는 방식으로 볼 수 있다.

본 연구에서는 표본의 평균 효용 값을 계산하기 위하여 전체 응답자 310명 가운데 표본의 크기를 1개부터 300개로 증가시켜 가며 추출하였다. 그런데 전체 310명 가운데 K개의 표본을 추출하는 경우는 총 ${}_{310}C_K$ 로 매우 다양하기 때문에 표본 크기에 따라 각 100번을 반복 추출하여 실험하였다. 따라서 표본의 크기 300개(1개~300개) × 100번 반복추출 = 총 30,000번의 시뮬레이션이 진행되었다. 이렇게 추출된 표본의 평균 효용 값을 이용하여 불특정 100명의 선택을 예측하였다. 선택의 예측정확도 계산은 앞선 방식과 동일하게 진행되었다. 즉 다른 응답자

들의 평균 효용 값을 이용해 계산된 가장 높은 효용 값을 갖는 카드와 응답자가 실제로 선택한 카드가 일치하는지 여부를 가지고 4번의 반복실험으로 예측정확도를 계산하였다.

〈그림 4〉는 각 응답자별로 예측정확도를 계산한 다음, 100명의 예측정확도 값을 평균 내어 표본의 크기에 따라 표시한 것이다. 〈그림 4〉에서 보듯이 추출된 표본의 수가 적은 경우에는 예측 정확도가 낮게 나타나지만, 표본의 크기가 50개 이상인 경우부터 예측 정확도가 60%를 상회하고 있으며, 표본으로 100명 이상을 추출하여 평균 효용 값으로 개별 응답자의 효용을 예측하는 경우 전체 응답자의 예측정확도 평균은 65%에 거의 수렴하는 것으로 나타났다.

이는 평균 효용값을 산출하기 위해 추출된 표본의 크기가 커질수록 계산된 효용값은 모집단 전체의 평균값으로 수렴할 것이며, 추정하고자 하는 불완전 응답자의 수가 커질수록 모집단과 일치하게 되므로, 예측 정확도 또한 모집단 전체의 예측정확도로 수렴할 것으로 예상할 수 있다. 따라서 평균값을 이용한 추정의 예측정확도가 약 65% 정도로 나타나고 있기 때문에 사례기반추론에 사용되는 입력자료의 수가 작거나, 응답 자료의 유사도가 낮은 경우에는 평

균값을 이용한 추정이 사례기반추론을 이용한 추정보다 정확할 수 있음을 시사한다. 그리고 응답자에 대한 선호 순위 정보가 전혀 없는 경우라도 나머지 다른 응답자들의 효용 평균값을 이용하여 선호 순위 정보가 없는 소비자의 선택을 예측할 수 있음을 시사한다. 이러한 방법으로 추정한 불완전 응답자들의 선택에 대한 예측정확도는 약 65% 수준으로 나타났다. 아무런 정보 없이 단순히 4개의 프로필 가운데 무작위로 1개를 선택하여 예측할 때 기대할 수 있는 최소한의 예측정확도는 1/4, 즉 25%이므로 효용의 평균값을 이용한 예측 방법의 정확도는 비교적 높은 수준인 것으로 생각된다.

V. 결론 및 시사점

5.1 요약 및 시사점

본 연구에서는 전통적 컨조인트 분석을 활용할 때 발생하는 문제점 가운데 하나인 불완전한 형태로 수집된 자료의 효용을 예측하기 위한 방안에 대해 제시하고자 하였다. 구체적으로 선호 순위를 이용하여 자료를 수집한 경우 발생하는 중도 응답 포기나, 중복 순위 발생과 같은 문제점 발생 시 이들의 효용을 사례기반추론에 기초하여 예측하고자 하였다. 사례기반추론에 있어 중요 절차인 유사성 자료의 검색은 순위 자료들 간의 선형적인 관계를 파악할 수 있는 서열 상관계수를 이용하여 유사한 사례를 검색하였고, 이들이 응답한 선호 순위 자료를 활용하여 불완전 응답자의 카드에 대한 선호순위를 추론한 후 효용 값을 분석하였다. 그리고 유사성 자료의 검색에 활용되는 방법론의 비교 뿐 만 아니라, 자료 검색에

활용되는 입력 자료의 수, 그리고 자료의 일치 정도 즉 상관계수의 크기가 사례기반추론의 정확도에 미치는 영향도 분석하였다.

분석결과를 요약하면 불완전 응답의 추론을 위해 활용된 입력 자료의 수가 많을수록, 서열 상관계수가 높아 유사성이 높을수록 불완전 응답자의 선호에 대한 예측정확도는 높아졌다. 그리고 사례기반추론에 활용된 자료 검색 방법에 있어 스피어만 서열 상관계수를 이용한 방법 보다 켄달의 서열 상관계수를 이용한 방법이 전반적인 예측정확도에 있어 근소하지만 0.5%p 정도로 높게 나타났다. 그리고 사례기반추론에 활용된 불완전 응답 자료의 수 역시 활용된 입력 자료의 수가 많을수록 예측정확도는 높아졌다.

본 연구에서는 컨조인트 분석에서 불완전 응답 자료를 보완하기 위한 방안으로 사례기반추론을 적용하여 효용을 추정하는 방법에 대해 제안하였지만, 컨조인트 분석 뿐 만 아니라 순위 정보를 수집하여 소비자의 선택을 예측하는 다양한 분야로도 적용이 가능할 것으로 생각된다. 예를 들면 오픈 마켓의 상품추천이나, 여행지 경로 설계에 있어 사례기반추론 방법에 근거하여 소비자의 선호도가 높게 나타날 것으로 예상되는 상품이나 경로를 소비자에게 추천하는데 활용할 수도 있다. 즉 소비자의 과거 행태를 분석하여 미래의 행위를 예측하는 다양한 방법에 적용시켜 볼 수 있을 것으로 생각된다.

5.2 향후 연구과제

본 연구의 컨조인트 분석에 활용된 상품의 속성들은 CPU 처리속도, 무게, 메모리와 같이 속성의 크기와 선호가 서로 선형적인 관계를 가지고 있는 연속적 속성들이 주를 이루었다. 이러한 연속적 속성들은 벡터모형을 적용하여 효용을 추정하면 부분이

치 모형을 적용하는 경우 보다 추정해야 하는 모수가 감소하여 모형의 자유도가 높아지는 장점이 있다. 또한 벡터모형은 속성수준과 효용이 선형적인 관계를 가지기 때문에 응답자들의 속성수준에 대한 선호의 이질성이 크게 나타나지 않을 수 있다. 반면 부분가치 모형에 활용되는 이산적 속성은 응답자들 간의 속성수준에 대한 선호가 다를 수 있어 속성수준과 효용의 증감이 일치하지 않을 수 있다(서상윤·이훈영·남양호 2009). 따라서 컨조인트 분석에 활용되는 속성이 브랜드나 색상, 원산지과 같이 이산적 속성이거나 경우 사례기반추론의 효용성에 대해 추가적인 연구가 필요할 것으로 생각된다.

또한 최근 컨조인트 분석은 다양하고 새로운 효용 추정 방법론들이 많이 제시되고 있다. 이러한 측면에서 전통적 컨조인트 분석 뿐 만 아니라, 선택형 컨조인트(Louviere et al., 2000; Toubia et al., 2004)나 하이브리드 컨조인트(Green, Goldberg, and Montemayor 1981; Eggers and Sattler 2009) 방법에서 사례기반추론의 적용 가능성에 대해서도 연구할 필요가 있겠다. 즉 전통적 컨조인트 분석의 경우 서열척도로 자료가 수집되면 서열 상관 계수를 이용하여 유사한 사례의 검색이 가능하지만, 선택형 컨조인트 분석을 수행하는 경우 응답자의 자료는 범주형 변수로 수집이 되기 때문에 유사도를 측정하는 다른 방법이 강구되어야 한다. 따라서 다양한 컨조인트 분석 방법에 따라 그에 맞는 사례기반추론 방법이 개발되어야 할 필요가 있겠다. 아울러 본 연구에서는 시뮬레이션을 통하여 사례기반추론을 이용한 응답자의 효용 추정방법의 유효성에 대해 실증분석이 이루어졌으나, 이론적 증명이 추가적으로 이루어진다면 연구의 완성도가 더욱 높아질 것으로 생각된다.

끝으로 추가적인 연구 방향으로 기존의 사례기반

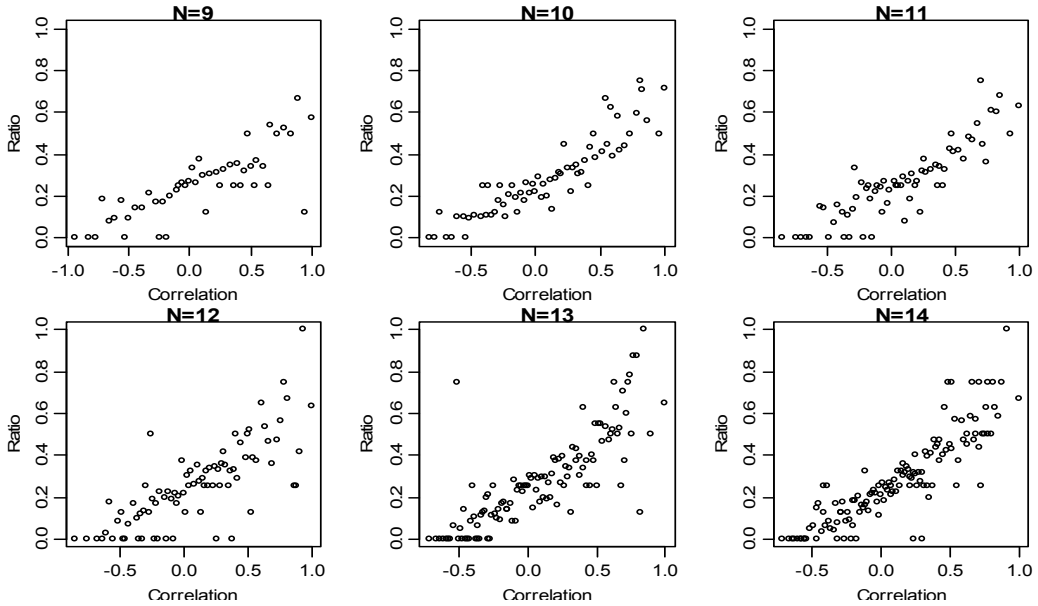
추론을 빅데이터 분석과 연계하는 것이다. 기존의 사례기반추론에 응답자의 SNS 감정표현, 개성, 제품 구매 정보 등 외생변수를 추가하여 데이터를 구성하고, 응답자들의 유사성에 따라 세분화한 후, 응답자들의 제품 구매 전 사전 정보만을 활용해서 소비자의 선택을 추론할 수 있는 모형을 개발하는 것이 보다 현실적인 활용방안이 될 것이다.

참고문헌

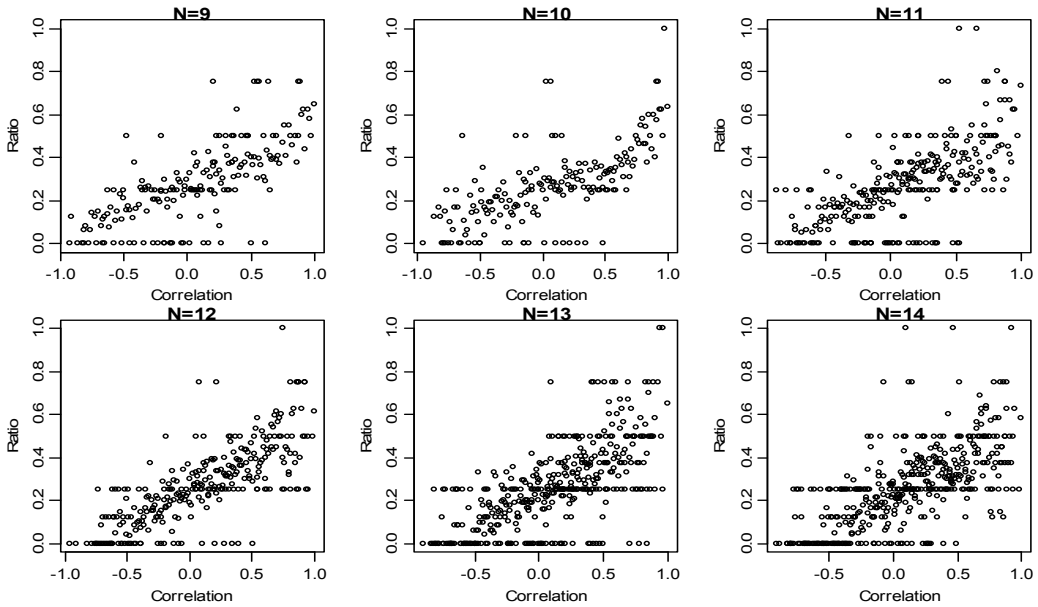
- 서상윤·이훈영·남양호(2009), "상품 속성의 특성과 자료 수집 방법이 컨조인트 모형의 효과에 미치는 영향에 관한 비교연구," *마케팅관리연구*, 14(1), 25-38.
- 이재식·명훈식(2008), "사례기반추론을 이용한 인터넷 서점의 서적 추천시스템 개발," *한국전자거래학회지*, 13(4), 173-191.
- 정영우·이은용(2008), "컨조인트 분석을 통한 푸드코트 선택 속성에 관한 탐색적 연구," *한국조리학회지*, 14(4), 106-118.
- Aamodt, A., and E. Plaza(1996), "Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches," *Artificial Intelligence Communications*, 4(3), 39-59.
- Chen, J., and J. Shao(2000), "Nearest Neighbor Imputation for Survey Data," *Journal of Official Statistics*, 16, 113-131.
- Curry, J.,(1997), "After the Basics : Keeping Key Issue in Mind Makes Conjoint Analysis Easier to Apply," *Marketing Research*, 9(Spring), 6-11.
- Eggers, F., and H. Sattler(2009), "Hybrid Individualized Two-level Choice-based Conjoint (HIT-CBC): A New Method for Measuring Preference Structures with Many Attribute

- Levels," *International Journal of Research in Marketing*, 26(June), 108-118.
- Green, P. E, S. M. Goldberg, and M. Montemayor (1981), "A Hybrid utility Estimation Model for Conjoint Analysis," *Journal of Marketing*, 45(1), 33-41.
- Green, P. E., and V. Srinivasan(1990), "Conjoint Analysis in Marketing : New Developments with Implications for Research and Practice," *Journal of Marketing*, 54(Oct), 3-16.
- Green, P. E., and M. Krieger(1991), "Product Design Strategies for Target Marketing Positioning," *Journal of Product Innovation Management*, 8(3), 189-202.
- Green, P. E., and M. Krieger(1992), "An Application of a Product Positioning Model to Pharmaceutical Products," *Marketing Science*, 11 (Feb), 117-132.
- Green, P. E., M. Krieger and Y. Wind(2001), "Thirty Years of Conjoint Analysis: Reflections and Prospects," *Marketing Engineering*, 31(3), S56-73.
- Green, P. E., and M. Krieger(2002), "What's Right with Conjoint Analysis?" *Marketing Research*, Spring, 24-27.
- Gustafsson, A., A. Herrmann, and F. Huber(2007), *Conjoint Measurement: Methods and Applications 4th*, Springer. 23.
- Johnson, R. M.(1987), "Adaptive Conjoint Analysis," in: Sawtooth Software Conference Proceedings, Ketchum, ID: Sawtooth Software, July, 253-65.
- Johnson, R., J. Huber, and L. Bacon (2003), *Adaptive Choice Based Conjoint Analysis*, Sawtooth Software.
- Kendall, M.(1938), "A New Measure of Rank Correlation," *Biometrika*. 30(1 - 2), 81 - 89.
- Lenk, P. J., W. S. Sesarbo, P. E. Green, and M. R. Young (1996), "Hierarchical Bayes Conjoint Analysis: Recovery of Partworth Heterogeneity from Reduced Experimental Designs," *Marketing Science*, 15(2), 173-191.
- Louviere, J. J., D. A. Hensher, and J. D. Swait(2000), *Stated Choice Methods: Analysis and Applications in Marketing, Transportation and Environmental Valuation*. Cambridge University Press, Cambridge.
- Sande, G.(1979), "Numerical Edit and Imputation," *International Association for Statistical Computing*, 42nd Session of the International Statistical Institute.
- Spearman, C.(1904), "The Proof and Measurement of Association between Two Things," *American Journal of Psychology*. 15, 72-101.
- Toubia, O., J. R. Hauser, and D. I. Simester(2004), "Polyhedral Methods for Adaptive Choice Based Conjoint Analysis," *Journal of Marketing Research*, 41(1), 116-131.
- Vacek, P. M., and T. Ashikaga(1980), "An Examination of the Nearest Neighbor Rule for Imputing Missing Value," *Proceedings of the Statistical Computing Section, American Statistical Association*, 326-331.

〈Appendix〉



〈그림 5〉 켄달의 타우 값을 이용한 추론과 응답자 효용의 예측정확도(라면제품)



〈그림 6〉 스피어만 상관계수 값을 이용한 추론과 응답자 효용의 예측정확도(라면제품)

A Study on the Predicting Utility for Incomplete Response in the Traditional Conjoint Analysis Based on the Case Based Reasoning

Sang Yun Seo*

Abstract

The conjoint analysis has been a representative methodology to find out consumer's preference. When carrying on the conjoint analysis, the most challenge task for the subjects is to evaluate more than eight profiles. When the respondent evaluate the profiles based on a ranking scale, as the predictive validity was greater compared to rating, this ranking scale have been usually applied with traditional conjoint analysis. However, several mistakes by respondents occurred such as the omitted ranking and same ranking answers. These incomplete answers are exempted from estimating utilities step. However, this study tried to find out the utility of subject who incompletely responds in traditional conjoint analysis based on the case based reasoning. The Kendall's and Spearman's ranking correlation coefficient are applied to search the similar cases to complement the incomplete response, also the utilities are estimated based on the preference data evaluated in regular sequence. The result shows that the hit-ratio predicted by data with higher correlation coefficient is higher than it done by the data with lower correlation coefficient. Also, in terms of the reasoning methodology, the Kendall's coefficient rather than the Spearman's coefficient shows higher predictive accuracy. In addition, the interaction between correlation coefficient and the number of profiles has positive impact on the predictive accuracy.

Key words: case based reasoning, traditional conjoint, incomplete response, Kendall's coefficient, Spearman coefficient

* School of Business Administration, Kyungnam University, First Author

- 저자 서상윤은 현재 경남대학교 경영학부 마케팅 전공 조교수로 재직 중이다. 경희대학교에서 경영학 학사, 석사, 박사를 취득하였다. 박사 학위 취득 이후에는 San Diego State University 박사 후 연수과정과 경희대학교 경영연구원을 거쳐 경남대학교에 부임하였다. 주요연구분야는 Marketing Modeling, Marketing ROI, Segmentation, Conjoint 등이다.