

사용편의성 모델수립을 위한 제품 설계 변수의 선별방법 : 유전자 알고리즘 접근방법

A Method for Screening Product Design Variables for Building A Usability Model : Genetic Algorithm Approach

양희철* · 한성호*

ABSTRACT

This study suggests a genetic algorithm-based partial least squares (GA-based PLS) method to select the design variables for building a usability model. The GA-based PLS uses a genetic algorithm to minimize the root-mean-squared error of a partial least square regression model. A multiple linear regression method is applied to build a usability model that contains the variables selected by the GA-based PLS. The performance of the usability model turned out to be generally better than that of the previous usability models using other variable selection methods such as expert rating, principal component analysis, cluster analysis, and partial least squares. Furthermore, the model performance was drastically improved by supplementing the category type variables selected by the GA-based PLS in the usability model. It is recommended that the GA-based PLS be applied to the variable selection for developing a usability model.

Keywords: Usability model, Screening Variables, Genetic Algorithm, Partial Least Squares

* 포항공과대학교 산업공학과
주소 : 790-784 경상북도 포항시 남구 포항공과대학교 산업공학과
전화 : 054-279-2855
E-mail : velocir@postech.ac.kr

1. 서론

과거에는 제품의 기능적 측면이 제품 설계 과정에서 가장 우선적으로 고려되었으나, 최근에는 사용자 인터페이스의 사용편의성(Usability)이 제품의 시장 경쟁력을 좌우하는 중요한 요인이 되었다. 제조회사에서도 제품의 시장경쟁력을 확보하기 위해 사용편의성 향상에 많은 노력을 기울이고 있다 (Myers et al, 1992).

사용편의성이란 단순히 사용자의 작업 수행도를 극대화하려는 소극적인 정의(Nielsen et al, 1994)에서 벗어나 사용자의 감성적인 측면까지 만족시킬 수 있는 포괄적인 개념이다 (한성호 외, 1998). 사용편의성을 구성하는 사용자의 수행도나 감성만족도의 개념은 매우 복합적인 내용들을 포함하고 있기 때문에 이들을 단순한 개념들의 집합으로 분할 정의하고, 분할된 각 세부 개념들을 사용편의성 요소로 정의하였다. 이를 활용할 경우, 객관적인 측면과 주관적인 측면을 동시에 고려하

는 사용편의성의 체계적인 평가가 가능하다.

사용편의성이 향상된 제품을 개발하기 위해서는 제품 설계의 초기 단계부터 체계적으로 사용편의성을 평가하고, 예측하는 과정이 필요하다. 즉, 사용편의성에 영향을 미치는 제품 설계 변수를 추출하고, 이들 간의 유기적인 관계를 분석하는 방법이 마련되어야 한다. 이를 위해 사용편의성 요소와 제품 설계 변수 간의 연관 관계를 다중회귀분석 기법의 수리적 모델로 예측하고, 상호 관련성을 해석하고자 하는 노력이 이루어졌다 (한성호외, 1998). 이를 사용편의성 모델이라 하며, 전자제품의 예를 들어 도식화하면 그림 1과 같다.

위와 같은 사용편의성 모델이 제품 설계 과정에 활용될 경우, 제품의 사용편의성에 영향을 주는 중요한 설계변수를 체계적으로 분석할 수 있으며, 제품의 사양 변경에 따른 사용편의성의 변화정도를 예측하여 해당 제품의 구체적인 설계 대안을 제시할 수 있다. 이전 연구 (한성호 외, 1998; Han et al, 1999)에서 개발된 사용편의성 모델은 70~90%에

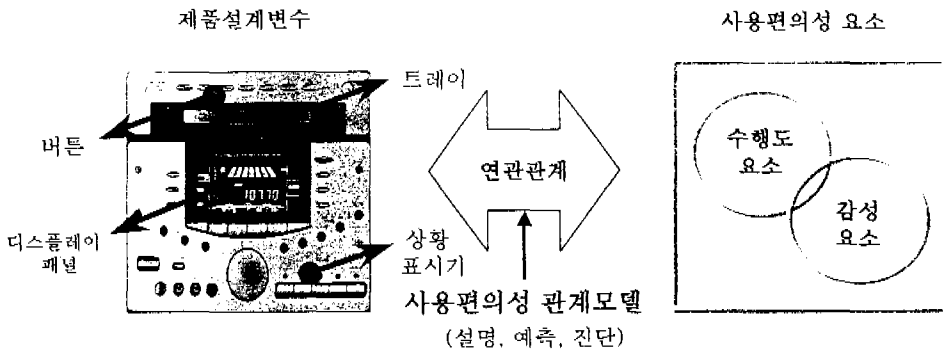


그림 1. 사용편의성 모델 개념도(Han et al., 2000)

이르는 예측성능을 보였으나, 모델에 포함될 초기 변수의 선정과 관련하여 두 가지 문제점이 존재한다.

첫째, 이전 연구에서는 전문가 의견을 통해 초기 제품 설계 변수를 선별하여 모델링에 적용하였다. 이는 변수 선별 과정의 객관성이 결여될 수 있으며, 많은 지식과 노하우(Know-how)를 지닌 다수의 전문가가 필요하다는 문제점이 있다. 이를 해결하기 위해 실험 데이터에만 의존하여 초기변수를 선별하는 주성분분석, 군집분석, 부분최소자승법 등의 통계적 기법을 활용하는 방안이 제시된 바 있다 (김중서, 2000).

둘째, 평가 대상이 되는 제품은 측정형(Measurement Type), 평가형(Rating Type)의 연속변수와 범주형(Category Type)의 불연속변수로 구성되는 제품의 설계 특성으로 분석된다 (한성호 외, 1998; Han et al, 1999). 범주형 변수를 다중회귀분석 모델에 포함할 경우, 각 범주형 변수별로 범주의 가지 수 만큼의 더미 변수(Dummy Variable)가 추가되어야 하므로 많은 수의 실험 데이터가 요구된다. 이전 연구에서는 한정된 수의 실험 데이터를 활용할 수 있도록 연속형 제품설계변수만을 포함하여 모델을 개발하였다. 이로 인해 범주형의 불연속 변수인 제품의 외관, 재질, 색상 등 제품의 기본적인 설계 변수들을 고려하지 않는 문제점이 존재한다. 실제로 범주형 제품 설계 변수가 중요한 영향을 미치는 색감, 형태감, 투명감 등의 사용편의성 요소들에 대한 모델의 경우, 예측 성능이 낮아지는 경향을 확인할 수 있다 (한성호 외, 1998; 한성민 외, 1998; Han et

al., 1999). 이러한 문제점은 앞서 언급된 통계적 기법을 이용한 변수 선별 방법으로는 효과적으로 해결되지 못한다.

따라서, 본 연구에서는 전문가에 의한 변수 선별과정과 변수 형태에 따라 데이터 부족을 야기시키는 기존 연구의 모델 개발 과정의 미비점을 해결하기 위해 유전자 알고리즘(Genetic Algorithm) 접근 방법을 제안하고자 한다. 또한, 선별된 변수를 활용하여 사용편의성 모델을 개발하고, 기존 연구를 통해 개발된 모델과 비교 분석함으로써 개선된 사용편의성 모델과 모델 개발 방법을 제시하고자 한다.

2. 유전자 알고리즘을 이용한 변수 선별

2.1 변수선별방법

사용편의성 모델을 개발하기 위해서는 사용편의성에 많은 영향을 미치는 적합한 제품설계변수를 선별하는 과정이 필요하다. 그러나, 모델에 포함되어야 하는 적합한 변수들을 결정하는 것은 다중회귀분석을 통한 모델링에 있어서 중요하지만 매우 어려운 문제이다. 어떠한 변수들이 중요한지 확인할 수 없고, 변수들에 대한 측정값이 많을 때 이는 더욱 더 어려워진다. 변수들을 모두 포함하여 모델을 수립하는 방법은 그 어려움을 피할 수 있는 하나의 방법이다. 그러나, 그럴 경우, 다음과 같은 데이터 분석과 관련된 문제를 야기시킨다 (Broadhurst et al, 1996). 첫째, 측정

된 변수들 중 일부는 모델의 목적과 무관한 것일 수 있다. 이러한 변수들이 모델에 포함될 경우 다른 변수들 사이에 존재하는 의미있는 관계조차 퇴색시킬 가능성이 있다. 둘째, 의미있는 모델의 파라미터 추정치를 얻기 위해서는 관측수가 변수의 수보다 많아야 하지만, 그렇지 못한 경우가 빈번히 발생한다. 셋째, 변수들간의 상관관계(Correlation)가 높을 경우, 정보가 중복되어 나타날 수 있어 많은 정보를 잃어버릴 수 있다. 넷째, 산포가 큰 변수들이 모델에 포함될 경우 모델의 성능이 떨어진다. 따라서, 변수의 수가 많을수록 모델링 과정에 앞서 중요한 변수들을 선별하는 과정이 절실히 필요하다. 변수선별방법은 크게 모델링 단계에서 실시하는 방법과 모델링 전단계에서 실시하는 방법으로 구분될 수 있다.

가장 사용이 편하고, 널리 쓰이는 변수선별 방법으로는 모델링 단계에서 실시하는 전진선택법(Forward Selection), 후진제거법(Backward Elimination), 단계적 선택법(Stepwise Selection)이 있다. 이 방법들 또한, 변수들이 많으면 각 단계에서 선택되는 변수들의 조합은 이전 단계에서 선택된 변수들의 조합에 의해 많은 영향을 받기 때문에, 부분 최적조합을 찾는 데 그칠 수 있다. 뿐만 아니라, 최종적으로 선택되는 결과가 단 하나의 변수조합으로 표현되기 때문에 다른 우수한 변수들의 조합을 간과할 수 있는 문제점이 있다(Broadhurst et al, 1996).

모델링 전단계에서 실시하는 변수선별 방법으로는 앞서 언급한 전문가 의견이 있다. 각각의 독립 변수들이 종속변수와 얼마나 관련

이 있는지 중요도를 전문가가 평가하는 방법이 주로 이용된다 (한성호 외, 1998). 그러나, 전문가 의견에 의한 변수선별방법은 전문가의 주관적 의견의 개입으로 객관성을 상실할 수 있다. 또한, 변수선별과정에 필요한 전문가는 많은 지식과 노하우를 필요로 하며, 이러한 조건을 만족하는 전문가를 구하기가 어렵다. 모든 설계변수와 종속변수 간의 연관 관계를 고려해야 하므로 모델링을 위한 전처리에 많은 시간과 노력이 요구된다는 단점도 지적된다.

모델링 전단계에서 실시하는 변수선별 방법으로는 전문가의견 이외에 객관성을 유지하기 위해 측정 데이터에만 의존하는 통계적인 방법인 주성분분석(PCA), 군집분석(Cluster Analysis), 부분최소자승법(Partial Least Squares) 등을 이용한 방법이 있다(김종서, 2000). 주성분분석은 원래의 변수 개수보다 훨씬 적고, 서로 상관관계가 없는 주성분을 도출하여 모델의 변수로 이용하는 방법이다. 군집분석은 독립 변수들을 상호 배타적인 몇 개의 군집으로 분리한 후, 각 군집으로부터 대표 변수를 하나씩 도출하여 모델링의 변수로 이용한다. 부분최소자승법에 의한 변수선별은 독립변수 뿐만 아니라 종속변수를 함께 사용하여 변동 요인을 동시에 고려하는 잠재변수(Latent Variables)를 도출한다. 잠재변수의 수는 원래 변수의 수보다 항상 적으며, 상관관계가 적으므로, 독립변수의 복잡도를 감소시킬 수 있다. 기존 연구(김종서, 2000)에서는 부분최소자승법을 통해 추출한 잠재 변수들 중에서 평균 이상 (VIP: Variable Importance in Projection > 1)

의 영향을 미치는 변수들을 선별하여 모델링 과정에 추가하는 방법을 이용하였다. 그러나, 잠재변수들 중에는 사용편의성에 미치는 영향도가 작은 것들도 존재하기 때문에, 사전에 이러한 잠재 변수들을 제거하는 것이 보다 바람직하다.

본 연구에서는 모델링 전단계에서 부분최소자승법을 통해 도출되는 잠재 변수들을 포함한 회귀모델(PLSR: Partial Least Square Regression)의 성능을 최대화 시키는 잠재변수의 조합으로부터 변수들을 찾는 방법을 이용하였다. 이 방법은 영향도가 작은 잠재변수를 사전에 제거함으로써, 모델에 불필요한 변수들을 미리 제거할 수 있으며, 독립변수의 수가 아무리 많아도 자유도 결여 등의 문제가 없다는 장점을 가지고 있다. 그러나, 변수선별을 위한 부분최소자승법 모델 수립과정에서 역시 전진선택법, 후진제거법 등을 이용할 경우 선별되는 잠재변수들은 국소최적해(Local Optimum)일 가능성이 많기 때문에, 보다 전역적인 최적의 변수조합을 찾기 위한 방법이 요구된다. 이러한 조건을 만족시킬 수 있는 확률적 최적화 방법으로 유전자 알고리즘(Genetic Algorithm)이 제안된다.

2.2 유전자 알고리즘(Genetic Algorithm)을 이용한 변수선별

유전자 알고리즘은 인공지능 검색기법 중 하나으로써, 유전적 계승과 생존경쟁이라는 자연적 현상을 모델링한 확률적인 검색 방법이다 (Michalewicz, 1996). 유전자 알고리즘은 변수 및 주성분을 선별하는 문제와 군집분

석에도 성공적으로 적용되어 왔으며(Barros, et al, 1998; Cowgill, et al, 1999; McShane et al, 1999), 모델의 최적화 등 감성공학 분야에서도 사용된 바 있다(Nishino et al, 1994; Nagamachi, 1995).

유전자 알고리즘은 많은 변수들을 포함한 모델을 수립할 때, 적합한 변수를 선별하는 방법에 응용될 수 있다. 유전자 알고리즘을 이용한 변수선별 과정을 자세히 설명하면 다음과 같다 (Michalewicz, 1996; Mitchell, 1996).

먼저, n개의 변수들로 구성된 변수 조합을 임의로 특정 개수만큼 생성한다. 각 변수조합은 독립변수 개수 만큼의 1과 0으로 이루어져 있으며, '1'은 변수가 모델에 선택되었음을 의미하고, 0은 선택되지 않았음을 의미한다. 예를 들어, 9개의 변수로 구성된 '001101101'이라는 하나의 집합을 생각해 볼 수 있다. 3, 4, 5, 6, 8번째 변수가 선택되었으며, 1, 2, 5, 7번째 변수는 선택되지 않았음을 의미한다.

각 변수조합에 대한 목적함수 값을 구하고 최적에 가까운 변수조합이 확률적으로 더 많이 선택될 수 있도록, 확률적인 무작위 추출을 한다. 목적함수는 모델의 성능을 나타내는 결정계수(Adjusted R²), 평균제곱오차의 제곱근(Root Mean Squared Error, RMSE) 등이 다양하게 사용될 수 있다. 초기 개체집단으로부터 일부 모델의 성능이 낮은 열성 변수조합은 제거되고, 우성 변수조합은 여러 번 선택된다.

선택된 우성의 변수조합 중 두 개를 임의로 선택하고, 그 중 일부 배열을 일정한 확률로

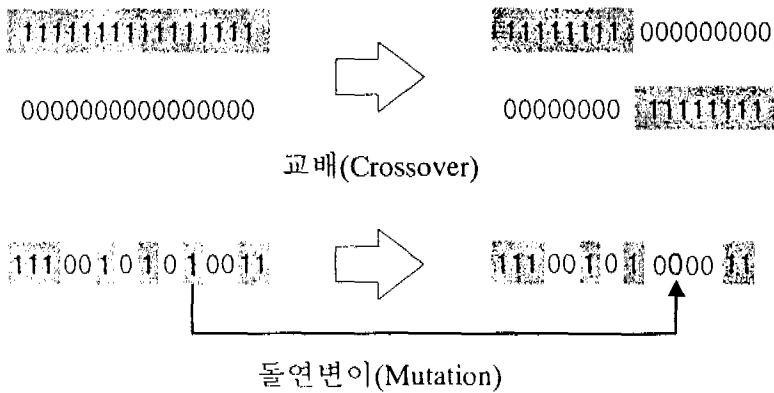


그림 2. 교배 및 돌연변이

서로 교환하여 교배(Crossover)를 수행한다. 각 변수조합에서 일정한 확률로 '0'은 '1'로, '1'은 '0'으로 변경함으로써 돌연변이(Mutation)를 수행한다 (그림 2 참조). 교배와 돌연변이를 통한 재생산 (Reproduction) 과정을 n 개의 새로운 변수 조합이 생성될 때까지 반복한다. 새로운 염색체들에 대한 평가함수의 목적함수 값을 구하고, 앞의 모든 과정을 다시 반복한다.

마지막으로, 특정한 기준에 도착할 때까지 앞의 과정을 반복된다. 주어진 적합도 값에 도달했거나, 미리 정해 놓은 반복회수가 지났거나, 유사한 변수조합이 특정수준 이상 존재할 경우 변수선별 과정을 멈춘다.

유전자 알고리즘과 다중회귀분석을 결합한 변수선별방법이 다중선형모델을 수립하는 합리적인 방법으로 많은 분야에서 사용되고 있다 (Wasserman, 1994). 그러나, 다중회귀분석과 유전자 알고리즘을 결합한 방법에서도 변수의 수가 많고, 데이터의 수가 부족할 때는 자유도 결여의 문제로 적용이 불가능하다.

본 연구에서는 변수선별과정에서 자유도 결여의 문제점이 없는 유전자 알고리즘 기반 부분 최소 자승법(GA-based PLS, Hasegawa et al, 1997)을 이용하여 변수를 선택함으로써, 보다 전역 최적에 가까운 변수들의 조합을 찾아낼 수 있도록 하였다.

3. 변수선별

3.1 실험 데이터 전처리

모델의 입력 데이터를 구하기 위해 제품 측정과 사용편의성 평가의 두 부분으로 나누어 실험을 수행하였다. VCR, CDP, Mini-component, DVD 등 총 36종의 국내외 오디오/비디오 전자제품이 실험 대상으로 선정되었다. 측정 실험에서는 88항목의 제품설계 변수(독립변수)가 이용되었으며, 5명의 전문가에 의해 측정실험이 진행되었다. 제품 평가 실험에서는 주어진 제품에 대해서 33항목(종

속변수)의 사용편의성 요소를 일반 소비자 60명이 평가하도록 하였다 (한성호 외, 1998).

범주형 변수가 포함된 회귀모델 수립을 위하여 모든 범주형 변수에 대한 측정치를 더미 변수(Dummy Variable)를 이용하여 표현하였다. 24개의 범주형 변수가 97개의 더미 변수(Dummy Variable)로 표현되었으며, 결과적으로 총 88개이던 독립변수가 최종 161개로 늘어났다. 유전자 알고리즘을 통한 변수 선별방법을 적용하기 위해 33개의 사용편의성 요소 중 대표적인 사용자 감성인 고급감, 볼륨감, 품위감과 범주형 변수에 의해 많은 영향을 받을 것 같은 견고성, 색감, 정교성, 역동감, 투명감, 형태감 등 9개의 사용편의성 요소를 선정하였다. 각 사용편의성 요소의 개념은 표 1에 결과와 함께 제시하였다. 고급감, 볼륨감, 품위감은 전문가 의견, 주성분 분석 등의 다른 변수선별방법과의 비교를 위해 연속 변수만을 포함하여 변수선별 과정이 진행되었다. 나머지 6개의 사용편의성 요소는 범주형 변수가 배제된 기존 모델에서 낮은 성능을 보였으므로, 본 연구에서는 범주형 변수를 추가한 변수선별과정을 거쳐 모델을 개발하였다.

3.2 유전자 알고리즘을 이용한 변수 선별 과정

본 연구에서는 주어진 데이터로부터 부분 최소 자승법(PLS)을 적용하여 잠재 변수들을 도출한 후, 잠재변수에 대한 회귀모델(PLSR)에서 평균제곱오차의 제곱근 (RMSE)을 최소

화하는 잠재 변수들을 유전자 알고리즘을 사용하여 선별하였다. 마지막으로, 선택된 잠재 변수로부터 최적의 변수조합을 선별하였다. 이 과정을 도식화하면 다음 그림 3과 같다.

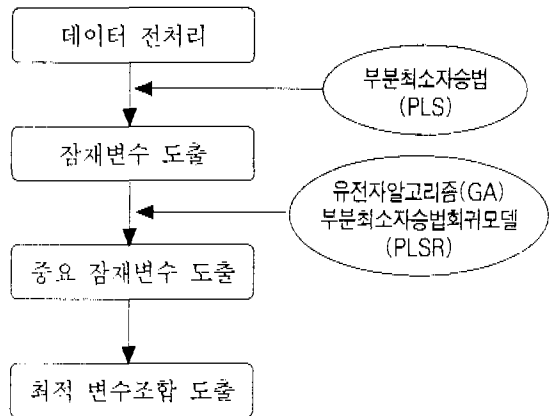


그림 3. 유전자 알고리즘 기반 부분최소자승법을 통한 변수선별 과정

변수선별과정에 대한 알고리즘 수행을 위해 소프트웨어로는 Matlab™을 사용하였다 (Sigmond, 1992). Matlab™을 이용할 경우, 경우에 따라 간단한 프로그래밍에 의해 목적 함수를 부분 최소자승법 회귀 모델에서의 평균 제곱 오차의 제곱근(RMSE)뿐만 아니라, 결정계수(Adjusted R²), 예측 잔차 제곱합 (Predicted Residual Sum of Squares, PRESS) 등으로 다양하게 변경가능하며, 목적함수의 복합적 사용이 가능하다.

변수선별과정에 유전자 알고리즘을 적용시키기 위해서는 유전자 알고리즘 수행에 필요한 몇 가지 매개변수 들의 값을 결정해야 한다. 먼저 총 변수의 수는 연속변수만을 포함한 경우에는 64개이며, 범주형 변수도 모두 고려할 경우 161 개이다. 초기 개체집단의 크기는 알

고리즘 수행 속도, 성능 등을 고려하여 일반적으로 많이 사용되는 100개로 결정하였다 (Broadhurst et al, 1996). 초기에 선택되는 변수의 수는 161개의 변수 중 20%로 하였다. 교배는 0.25의 확률로 이루어지며, 돌연변이는 0.005의 확률로 발생하게 하였다. 최대 세대 수 100 (범주형 변수 불포함)이나 200 (범주형 변수 포함)에 도달하거나, 전체 100개의 변수조합 중 20%, 30%, 40% 혹은 50%가 같은 변수조합으로 구성될 경우를 유전자 알고리즘 수행 정지기준으로 사용하였다. 이와 같은 매개변수 값을 바탕으로 5-6회 반복하여 최종적으로 가장 좋은 변수들의 조합을

선별하였다(양희철, 2000).

3.3 유전자 알고리즘을 이용한 변수선별 결과

전문가 의견이나 주성분분석, 군집분석, 부분최소자승법 등의 다른 변수선별 방법에 의해 선택된 변수들과의 비교를 위해 연속 유형의 변수만을 포함하여 유전자 알고리즘에 의한 변수선별 과정을 거쳤다. 다음 그림 4는 고급감에 대해 유전자 알고리즘을 이용한 변수선별 과정을 거친 후 마지막 세대에서의 결과를 표현한 것이다.

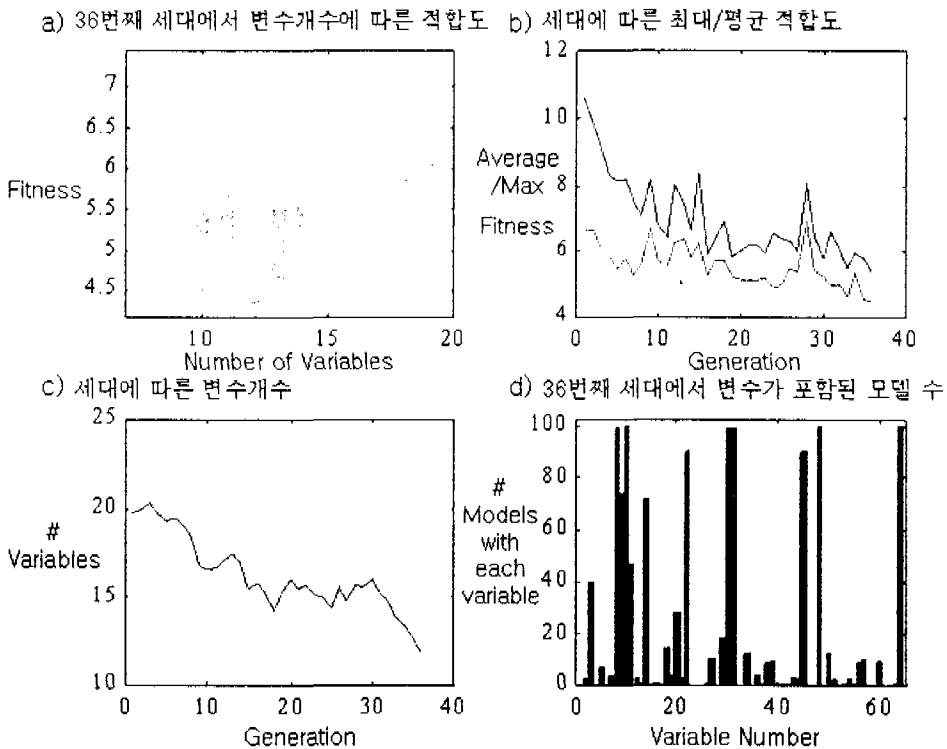


그림 4. 고급감 변수선별 프로그램 수행 결과

고급감 변수선별과정에서 100번째 세대에 도달하기 전 36번째 세대의 개체 집단에서 총 100개의 변수조합 중 40% 이상이 동일하기 때문에 36번째 세대에서 변수선별과정이 멈춘 것을 알 수 있다. 그림 4.a는 마지막 36번째 세대에서의 100개의 변수조합 중 포함된 변수의 개수에 따른 적합도(Fitness) 함수 값, 즉, 모델의 평균제곱오차의 제곱근 값(RMSE)을 나타낸다. 변수가 10~15개 정도 선택된 경우가 가장 많으며, 평균제곱오차의 제곱근 값은 5.0에서 5.5 정도 사이에 집중되어 있다. 그림 4.b는 각 세대별 평균제곱오차 제곱근의 평균값과 최소값을 나타내며, 세대가 거듭될수록 평균 및 최소 평균제곱오차의 제곱근 값이 점점 작아지는 경향을 알 수 있다. 그림 4.c는 각 세대별 개체집단에서 선택된 변수들의 평균 개수를 나타낸다. 처음에는 전체 64개의 연속변수 중 약 20개가 선택되었으나, 점점 줄어드는 경향을 보여

마지막 세대에는 약 12개 정도의 변수가 선택되었다. 마지막으로 그림 4.d는 마지막 36번째 세대의 100개 변수조합 중 64개의 변수들이 각각 몇 번 선택되었는지를 표현하는 것이다. 6개 정도의 변수가 100개의 변수조합 모두에 포함되어 있으므로, 고급감에 중요한 영향을 미치는 변수이다. 일부 변수는 100개의 변수조합에 한 번도 포함되지 않으므로, 상대적으로 고급감에 영향을 적게 미치는 변수로 간주할 수 있다.

이와 같은 유전자 알고리즘 기반 변수선별 방법을 통해 9개의 사용편의성 요소에 대한 변수 선별과정을 거쳤다. 고급감, 볼륨감, 품위감의 3개의 사용편의성 요소에 대해서는 기존 연구의 변수선별방법과의 비교를 위해 범주형 변수를 배제한 채 유전자 알고리즘 기반 변수선별과정을 거쳤다. 그 결과 64개 중 17~19%정도인 약 10여 개의 변수가 선별되었다. 나머지 6개의 사용편의성 요소의 경

표 1. 사용편의성 요소의 정의 및 유전자 알고리즘을 통해 선별된 변수의 수

범주형 변수 포함 여부	사용편의성 요소	정 의	변수 개수	비율 (%)
범주형 변수 포함하지 않은 경우	고급감	제품의 고급스러움이나 화려하고 사치스러운 느낌	11	17.2
	볼륨감	제품의 크기와 부피, 입체적 형태 등에 관한 감성	11	17.2
	품위감	제품으로부터 느껴지는 고상하고 우아한 느낌	12	18.6
범주형 변수 포함한 경우	견고성	제품이 단단하고 튼튼하게 보이는 정도를 평가하는 감성	25	15.5
	색 감	색에 관한 이미지와 온도감을 나타내는 감성	28	17.4
	역동감	제품이 갖는 동적인 느낌을 표현하는 감성	38	23.6
	정교성	제품의 섬세하고 꼼꼼함을 평가하는 감성	38	23.6
	투명감	제품의 맑고 투명한 정도를 나타내는 감성	33	20.5
	형태감	제품의 길이, 두께, 굵기, 선, 요철 등에 의해 형성되는 감성	34	21.1

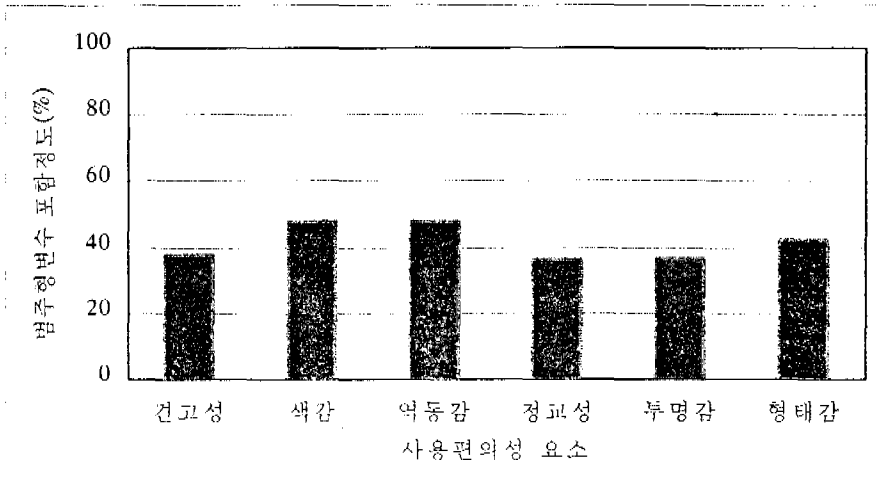


그림 5. 선별된 변수 중 범주형 변수의 비율

우에는 범주형 변수를 포함한 모델 개발을 위해 범주형 변수를 포함한 후, 유전자 알고리즘 기반 변수선별방법을 거쳤다. 그 결과, 총 161개의 변수 중 16~24%에 해당하는 25~38개 정도가 선별되었다. 범주형 변수의 포함여부에 따라 각 사용편의성 요소별로 선별된 변수들의 수는 표 1과 같다.

범주형 변수를 포함한 후 유전자 알고리즘 기반 변수선별방법을 거친 6개의 사용편의성 요소에 대하여, 선별된 변수 중에서 범주형 변수가 차지하는 비중을 살펴보면 다음 그림 5와 같다. 88개의 제품설계변수 중 40% 내외의 범주형 제품설계변수가 포함되어 있다. 이는 범주형 변수를 배제한 기존 사용편의성 모델은 불완전한 모델일 수밖에 없다는 사실을 간접적으로 시사하고 있다. 범주형 변수를 포함하여 선별한 변수들을 사용한 모델과 비교해 봄으로써, 범주형 변수의 추가가 얼마나 모델의 성능 향상에 도움을 주었는지 비교가

가능하다. 또한, 범주형 변수가 중요한 영향을 미칠 것으로 판단되는 사용편의성의 경우, 유전자 알고리즘 기반 변수선별방법의 우수성을 검증할 수 있다.

3.4 기존 연구의 변수선별결과와의 비교

기존 연구의 변수선별방법과 유전자 알고리즘에 의한 변수선별방법을 비교하기 위해 범주형 변수를 배제한 채 변수선별과정을 거친 고급감, 볼륨감, 품위감 3개의 사용편의성 요소에 대하여 각 변수선별방법에 의해 선별된 변수들이 어느 정도 일치하는지 살펴보았다 (그림 6 참조). 고급감의 경우, 기존 통계적인 변수 선별 방법 연구 (김중서, 2000)에서는 고려되지 않았다.

고급감의 경우에는 유전자 알고리즘에 의해 선별된 변수들 중 45%정도만이 전문가의견에 의해 선별된 변수들과 일치하였다. 유전자

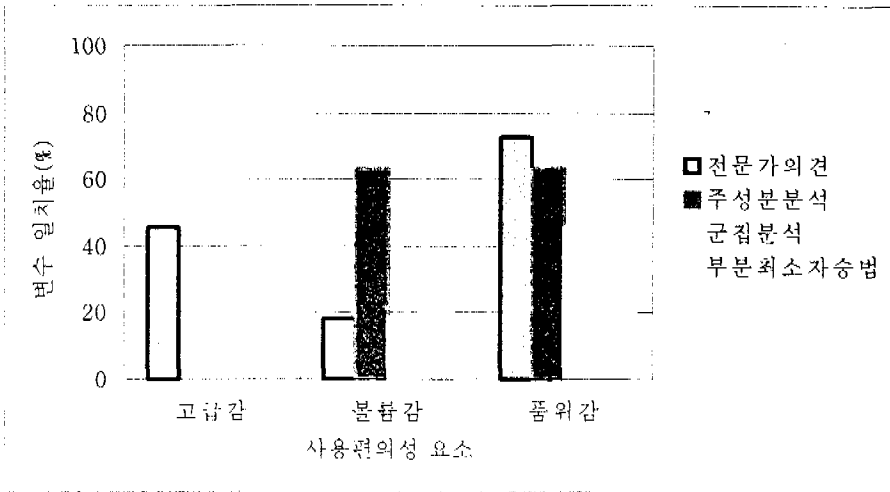


그림 6. 유전자 알고리즘에 의해 선별된 변수와의 일치정도

알고리즘에 의해 선별된 12개 중 5개만이 전문가 의견에 의해 선별된 26개의 변수에 포함되어 있어 일치정도가 매우 낮음을 알 수 있다. 즉, 전문가의견에서 중요하지 않은 것으로 간주된 변수들이 유전자 알고리즘을 통한 변수선별방법에서는 통계적으로 중요한 변수로 간주되었다. 이처럼 유전자 알고리즘에 의해 선별된 변수들 중 20~60%정도만이 전문가 의견, 주성분분석, 군집분석, 부분최소자승법에 의해 선별된 변수들과 일치하는 것으로 나타났다. 이는 연속변수만을 포함한 경우에도 제품설계변수의 개수가 64개에 이르며, 그들 간에 상관관계가 크기 때문인 것으로 판단된다. 결과적으로 각 변수선별방법마다 다른 변수조합이 의미가 있는 것으로 도출되었다. 따라서, 어떤 선별방법이 효과적인지를 검증하기 위해 각 변수선별방법을 통해 선별된 변수들을 이용한 사용편의성 모델의 성능을 비교하였다.

4. 사용편의성 모델 개발

유전자 알고리즘 기반 변수 선별 방법을 이용하여 선별된 변수들을 포함하여 제품설계변수와 사용편의성 요소간의 관계를 설명하는 모델을 개발하였다. 이를 위하여, 사용편의성 요소의 평가 결과를 종속변수로, 제품설계변수의 측정결과를 독립변수로 설정하여 다중회귀분석을 실시하였다. 도출된 모델을 기존 변수선별방법을 거친 모델과 비교하고, 범주형 변수를 추가한 모델을 범주형 변수를 배제한 모델과 비교해 봄으로써, 유전자 알고리즘에 의한 변수선별과정의 효과를 검증하였다.

4.1 사용편의성 모델 개발과정

모델 수립에 앞서 선별된 변수의 제공항 및 교호작용항의 추가과정을 거쳤다. 이는 제품

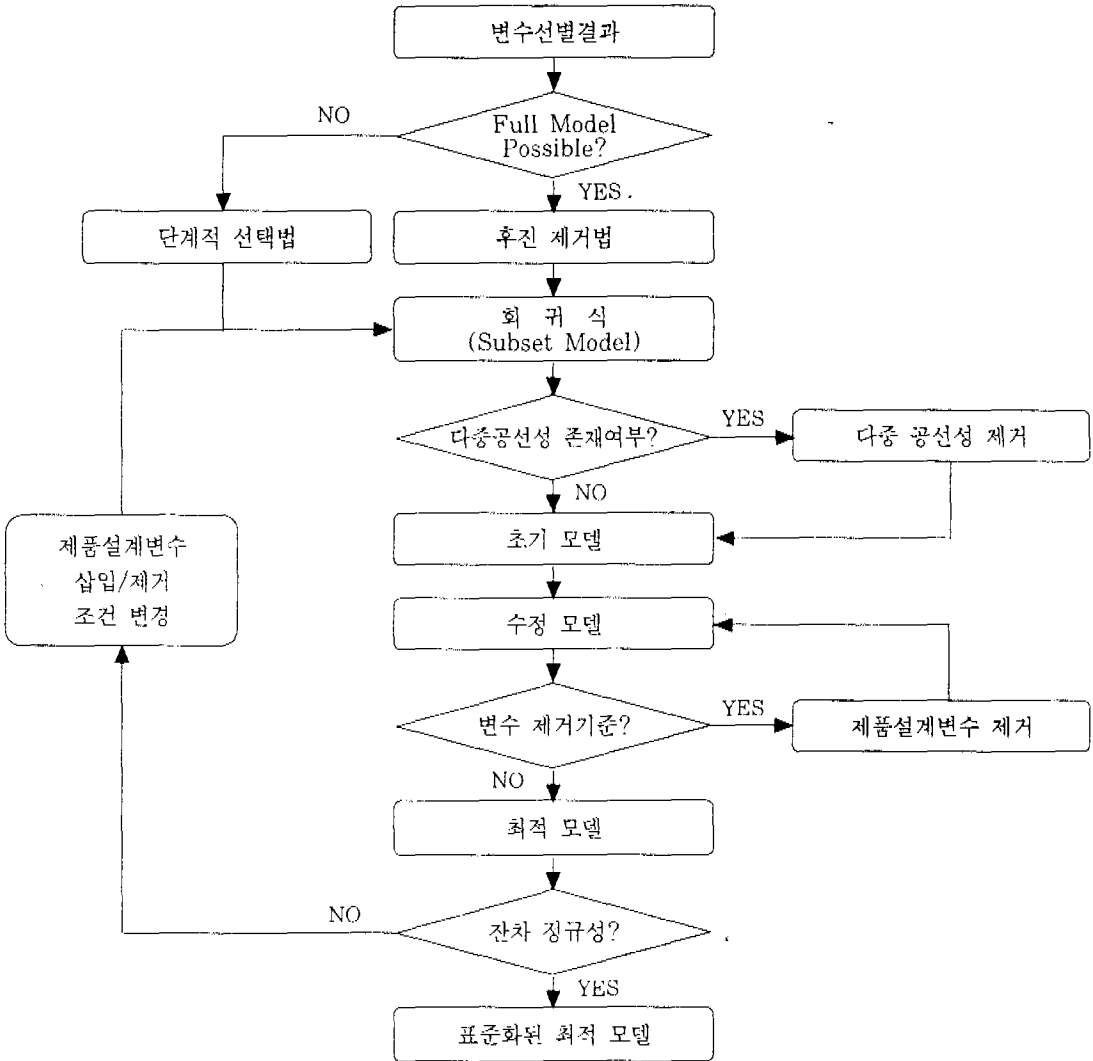


그림 7. 사용편의성 모델의 개발 절차 순서도 (한성호 외, 1998)

설계 변수들에 대한 측정 결과와 감성은 단순한 선형(Linear) 관계가 아닐 수 있으며, 각 제품 설계 변수들이 함께 상호 작용하여 감성에 영향을 줄 수도 있다고 판단했기 때문이다. 모든 변수와 제공항 및 교호작용항을 모델링에 포함할 정도로 실험에서 얻은 자료의 수가 많지 않기 때문에 단계적 선택법(Stepwise

Procedure)을 이용하여 다중회귀분석을 수행하였다. 그 결과 생성되는 모델에서 일차적으로 심각한 다중공선성(Multicollinearity)을 제거하여 초기 모델(Initial Subset Model)을 생성하였다. 초기 모델에서 분산팽창계수(Variance Inflation Factor, VIF)의 값이 가장 큰 제품 설계 변수를 하나씩 제거하는 작

표 2. 각 변수선별 방법간 모델 성능 비교

	변수선별방법	고급감	불륨감	품위감
결정 계수 (Adj. R ²)	전문가 의견	0.7215	0.8593	0.9880
	주성분 분석		0.9348	0.9107
	군집분석		0.9503	0.9717
	부분최소자승법		0.9136	0.8860
	유전자 알고리즘	0.8992	0.9538	0.9456
변수 개수	전문가 의견	17	6	16
	주성분 분석		4	9
	군집분석		12	13
	부분최소자승법		3	7
	유전자 알고리즘	7	7	11
예측잔차 제공합 (PRESS)	전문가 의견	2799.5	997.7	127.6
	주성분 분석		336.18	447.56
	군집분석		490.31	408.08
	부분최소자승법		448.16	554.9
	유전자 알고리즘	479.8	266.14	353.02

업을 반복하면서 수정 모델(Revised Subset Model)들을 생성하고, 가장 적합한 모델을 최적 모델(Best Subset Model)로 선정하였다. 그림 7은 사용편의성 모델 도출과정을 순서도 형식으로 나타낸 것이다.

4.2 기존 연구의 사용편의성 모델과의 성능 비교

그림 7에 제시한 순서에 따라 유전자 알고리즘에 의한 변수선별방법의 효과를 다른 변수선별방법과 비교하기 위해 사용편의성 모델을 개발하였다. [표 2]에서는 모델링 결과 중 모델의 성능과 직접적으로 관련이 있는 결정

계수(Adjusted R²), 변수 개수, 예측잔차 제공합(PRESS) 등을 제시하였다. 고급감의 경우 기존의 통계적인 변수선별방법에서는 적용되지 않았기 때문에 이들에 대한 자료가 없음을 밝힌다.

표 2에 나타난 바와 같이, 유전자 알고리즘에 의한 변수선별방법은 결정 계수와 예측잔차 제공합의 측면에서 전문가 의견에 의한 변수선별방법에 비해 비교적 우수한 결과가 도출되었다. 변수 개수의 경우에도 유전자 알고리즘이 전문가 의견보다 비교적 좋은 결과를 보이고 있다. 이는 모델에 포함된 변수의 숫자가 적으면 적을수록 모델의 해석이 용이하다는 것을 가정으로 도출된 결론이다. 주성분

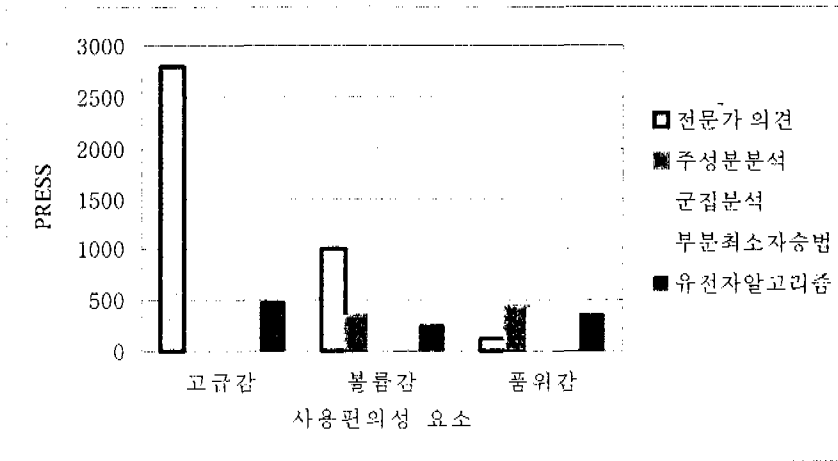


그림 8. 변수선별방법에 따른 모델의 예측잔차 제곱합

표 3. 모델 수립 목적별 변수 선별 방법 제안

모델 수립 목적	변수 선별 방법
예측력이 우수한 모델	유전자 알고리즘, 군집 분석, 주성분 분석
안정성이 우수한 모델	유전자 알고리즘, 주성분 분석
독립변수가 적은 모델	부분 최소 자승법, 주성분 분석, 유전자 알고리즘

분석과 부분최소자승법의 경우에는 유전자 알고리즘 결과에 비해 모델에 포함된 변수의 개수가 다소 작은 장점을 가지지만, 결정 계수 값이 많이 낮고, 예측잔차 제곱합이 높은 것을 알 수 있다. 군집분석은 결정 계수의 경우는 비슷하지만, 변수의 개수가 다소 많으며, 예측잔차 제곱합 값이 높은 단점이 있다.

특히 그림 8을 살펴보면, 모델의 안정성을 나타내는 예측잔차 제곱합의 경우, 유전자 알고리즘에 의한 변수선별방법을 거쳐 개발된

사용편의성 모델이 기존 연구의 변수선별방법을 거친 사용편의성 모델보다 더 우수한 성능을 보이고 있다. 이는 유전자 알고리즘에 의한 변수선별방법을 거친 모델이 데이터가 바뀌어도 예측력이 떨어지는 수준이 작다는 것을 의미한다. 즉, 유전자 알고리즘에 의한 변수선별방법을 거친 후 모델링을 실시할 경우, 모델의 안정성이 더욱 향상될 수 있다.

위의 결과를 종합하면, 모델수립 목적에 따라 다양한 변수선별방법을 적용할 수 있다고 판단된다. 각 모델수립 목적에 적합한 변수선별방법을 표 3에 종합 제시하였다.

결정 계수 값이 높아 예측력이 우수한 모델을 수립하기 위해서는 유전자 알고리즘, 군집 분석, 주성분 분석 등의 변수선별방법이 적합하다. 데이터를 추가하더라도 안정성이 우수한 모델수립을 위해서는 유전자 알고리즘, 주성분 분석 등이 적합하다. 포함된 독립변수의 개수가 적어 해석이 용이한 모델 수립을 위해서는 부분최소자승법, 주성분 분석, 유전자

알고리즘 등의 변수선별방법이 적합하다. 이를 종합해 볼 때, 유전자 알고리즘에 의한 변수선별방법은 모델의 예측력, 안정성, 독립변수의 개수 등의 모든 측면에서 성능이 전반적으로 우수하여, 어떠한 목적의 사용편의성 모델 개발 과정에도 적합하게 사용될 수 있다고 판단된다. 즉, 많은 변수가 포함된 모델 개발 과정에서 유전자 알고리즘 기반 부분최소자승법은 변수선별방법으로써 전반적으로 우수한 모델을 개발하는데 효율적으로 사용될 수 있

음을 의미한다.

4.3 범주형 변수 포함 여부에 따른 모델 성능 비교

기존 연구에서의 연속 변수만을 포함시켜 전문가 의견을 통해 선별된 변수들을 이용하여 개발된 모델과 범주형 변수를 포함시킨 후 유전자 알고리즘에 의해 선별된 변수들을 이용한 모델의 성능을 비교하였다. 표 4에 각

표 4. 범주형 변수 포함 여부에 따른 모델링 결과 비교

감성요소	범주형변수 불포함(전문가 의견)			범주형변수 포함(유전자 알고리즘)		
	Adj. R ²	변수개수	PRESS	Adj. R ²	변수개수	PRESS
견고성	0.462	17	2766.9	0.9152	8	79.73
색 감	0.7389	5	1032.6	0.9107	13	367.40
역동감	0.6399	15	1180.6	0.9808	11	59.40
정교성	0.411	18	4032.8	0.9582	9	139.5
투명감	0.7262	6	1493.8	0.9903	13	75.02
형태감	0.8347	3	1184.4	0.9733	9	173.15

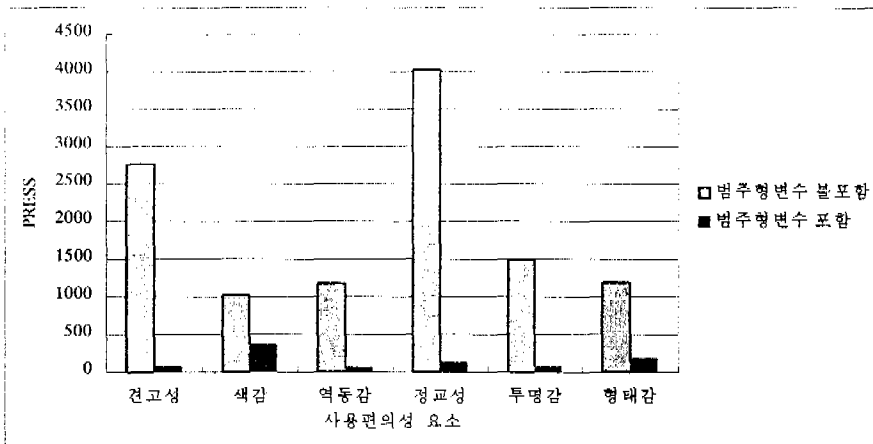


그림 9. 범주형 변수 포함 여부에 따른 예측잔차 제곱합 비교

방법에 의한 모델의 성능을 비교한 결과를 제시하였다.

표 4에 나타난 바와 같이, 범주형 변수를 포함하여 개발된 사용편의성 모델이 그렇지 않은 모델보다 성능이 매우 향상되었음을 알 수 있다. 모델 수립에 사용된 6개의 모든 사용편의성 요소에 대해 설명력이 90% 이상으로 예측 성능이 매우 높다. 특히, 모델의 안정성을 의미하는 예측잔차 제곱합(PRESS)이 기존 모델 보다 현저히 향상되는 것을 알 수 있다(그림 9 참조). 모델에 포함된 변수의 개수에 있어서는 많은 범주형 변수가 추가되었음에도 불구하고 최종적인 사용편의성 모델에 포함된 변수의 수에는 차이가 많지 않다. 즉, 색감이나 투명감, 형태감의 경우 변수의 수가 범주형 변수를 포함하지 않은 경우보다 증가하는 경우가 있으나, 이는 이들 사용편의성 요소가 범주형 변수에 의해 많은 영향을 받는다는 사실을 의미한다. 이들 결과를 종합하면, 범주형 변수를 추가할 경우, 보다 성능이 우수한 모델개발이 가능하다. 따라서, 범주형 변수가 많은 영향을 미치는 사용편의성 요소의 경우에는 기존 연구와는 달리 실험 데이터의 수에 제한이 없어 범주형 변수를 얼마든지 포함할 수 있는 유전자 알고리즘 기반 변수 선별 방법이 제안된다.

5. 결론 및 추후연구

본 연구에서는 확률적인 검색 방법인 유전자 알고리즘을 이용한 변수선별과정을 제안하였다. 유전자 알고리즘의 최적화 목적 함수로

는 부분최소자승법 회귀 모델 (PLSR)의 적합도(Fitness)를 나타내는 평균 제곱오차의 제곱근(RMSE)을 사용하였다.

기존의 전문가 의견이나 통계적인 변수선별 기법과의 비교를 위해 범주형 변수를 제외한 데이터를 가지고 본 연구에서 제안된 유전자 알고리즘에 의한 변수선별 과정을 수행하였다. 선별된 제품 설계변수들은 기존 방법에 의해 선별된 제품 설계변수들과 많은 차이를 나타냈다. 이는 변수들간의 상관관계가 매우 높기 때문인 것으로 판단된다. 따라서, 각 변수선별 방법을 통해 선별된 변수들을 이용한 사용편의성 모델을 개발하여 성능을 비교해 봄으로써, 모델 개발 목적에 따라 우수한 변수선별 방법을 제안하였다. 유전자 알고리즘에 의해 선별된 제품 설계변수들을 활용하여 개발된 사용편의성 모델은 예측능력, 안정성, 독립 변수 개수 등의 모델 성능면에서 기존 연구의 사용편의성 모델보다 전반적으로 우수한 것으로 나타났다. 특히, 새로운 데이터가 추가될 때, 모델의 예측성능의 변화정도를 나타내는 안정성(PRESS)면에서는 다른 변수선별방법보다 매우 우수한 것으로 나타났다.

기존의 모델링 과정에서 배제되었던 범주형 변수를 추가함으로써 보다 완전한 사용편의성 모델을 개발하였다. 결과적으로, 사용편의성 모델의 예측 성능이 90% 이상으로 나타났다. 이는 40~80% 정도의 예측성능을 보이던 기존 연구보다 매우 향상된 결과이다. 뿐만 아니라, 모델의 안정성, 독립변수 개수 등의 전체적인 성능이 매우 향상되는 결과를 보였다.

본 연구에서는 유전자 알고리즘 접근방법에 의해 선별된 제품설계변수를 활용하여 사용편

의성 모델을 개발할 경우, 모델의 성능을 크게 향상시킬 수 있음을 입증하였다. 또한, 색감, 형태감, 재질감 등 사용자의 감각적 감성에 많은 영향을 미치는 빔주형 제품설계변수를 포함하여 모델의 완성도를 극대화하였다. 따라서, 본 연구에서 제안된 유전자 알고리즘에 의한 변수선별방법은 객관성 결여되고, 빔주형 변수를 고려하지 못하는 기존 연구의 한계점을 극복할 수 있을 것으로 기대된다.

본 연구에서 제안된 유전자 알고리즘에 의한 변수선별방법의 활용성을 높이기 위해 요구되는 추후 연구 내용은 다음과 같다. 본 연구에서는 총 33개의 사용편의성 요소(한성호 외, 1998) 중 9개의 사용편의성 요소에 대해서만 유전자 알고리즘에 의한 변수선별방법을 적용하여 사용편의성 모델을 개발하였다. 본 연구의 최종 결과물인 사용편의성 모델이 제품 설계 과정에서 활용되기 위해서는 24개의 사용편의성 요소에 대한 모델들이 추가로 개발되어야 한다. 또한, 유전자 알고리즘 접근방법이 기존의 변수선별방법과 서로 병행될 경우, 각 접근방법의 장/단점을 보완하고, 모델의 성능을 보다 향상시킬 수 있을 것으로 예상된다. 예를 들어, 유전자 알고리즘의 객관성과 전문가 의견의 전문성을 최대한 활용할 경우, 통계적으로 검증된 제품설계특성과 사용편의성 간의 연관관계에 대한 설명이 용이한 사용편의성 모델을 개발할 수 있다.

참고 문헌

김중서, (2000) 사용편의성 평가 모델을 위한

제품 설계변수 선별방법 개발, 석사학위논문, 포항공대.

양희철, (2000) 사용편의성 모델 수립을 위한 제품 설계변수의 선별방법: 유전자 알고리즘 접근방법, 석사학위논문, 포항공대.

한성민, 김광재, (1998) 전자제품의 사용편의성 평가 모델 개발, 대한인간 공학회 추계학술대회, 321 - 324.

한성호, 윤명환, 김광재 외, (1998) 사용편의성 평가 기술 개발, 선도기술 사업보고서, 포항공대.

Barros, A. S. and Rutledge, D. N. (1998) Genetic algorithm applied to the selection of principal components, *Chemometrics and Intelligent Laboratory Systems*, 40, 65-81.

Broadhurst, D., Goodacre, R., Jones, A., Rowland, J. J. and Kell, D. B. (1997) Genetic algorithms as a method for variable selection in multiple linear regression and partial least squares regression with applications to pyrolysis mass spectrometry, *Analytica Chimica Acta*, 348, 71-81.

Cowgill, M. C. and Harvey, R. J. (1999) A genetic algorithm approach to cluster analysis, *Computers and Mathematics with Applications*, 37, 99-108.

Han, S. H., Kim, K. J., Yun, M. H., Kwahk, J., Hong, S. W. and Han, S. M. (1999) Usability Prediction Models Based on Human-Product Interface Elements, *International Society for*

- Occupational Ergonomics and Safety 1999 Conference.
- Hasegawa, K. and Funatsu, K. (1997) GA strategy for variable selection in QSAR studies: GAPLS and D-optimal designs for predictive QSAR model, *Journal of Molecular Structure (Theochem)* 425, 255-262.
- McShane, M. J., Cameron, B. D., Cote, G. L. Motamedi, M. and Spiegelman, C. H. (1999) A novel peak-hopping stepwise feature selection method with application to Raman spectroscopy, *Analytica, Chimica, Acta*, 388, 251-264.
- Michalewicz, Z. (1996) Genetic algorithms + data structures = evolution programs, New York: Springer-Verlag.
- Mitchell, M. (1996) An introduction to genetic algorithms, MIT Press, Cambridge.
- Myers, B. A. and Rosson, M. B. (1992) Survey on User Interface Programming. Proceedings of CHI Conference on Human Factors in Computing Systems, Masic Books, New York.
- Nagamachi, M. (1995) Kansei Engineering: A new ergonomic consumer-oriented technology for product development, *International Journal of Industrial Ergonomics*, 15, .3-11.
- Nielsen, J. and Levy, J. (1994) Measuring usability: Preference vs. performance, *Communications of the ACM*, 37(4), ACM, New York.
- Nishino, T., Nagamachi, M., Tsuchiya, T., Matsubara, Y. and Cooper, D. (1994) A Genetics-based Approach Design based on Kansei Engineering, Proceedings of The 3rd Pan-Pacific Conference on Occupational Ergonomics, Seoul, Korea, 162-166.
- Sigmond, K. (1992) Matlab Primer, 2nd Ed., [<http://www.cs.iastate.edu/~cs474/MATLAB/matlab-primer.html>]
- Wasserman, G. S. and Sudjianto, A. (1994) All subsets regression using a genetic search algorithm, *Computers & Industrial Engineering*, 27, 489-492.
-
- 저자 소개**
- ◆ 양희철
포항공과대학교 산업공학과 학사
포항공과대학교 산업공학과 석사
현재 포항공과대학교 산업공학과 박사과정
관심분야: HCI, 감성공학, 가상현실
- ◆ 한성호
서울대학교 산업공학과 학사
서울대학교 산업공학과 석사
Virginia Polytechnic Institute & State University, Dept. of Industrial and Systems Engineering, Ph.D.
-
- 논문접수일 (Date Received) : 2000/1/12
논문게재승인일 (Date Accepted) : 2001/5/22