저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:

- 저작자표시: 귀하는 원저작자를 표시하여야 합니다.
- 비영리: 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.
- 변경금지: 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리의 내용에 의하여 영향을 받지 않습니다.

이것은 이용허락규약(Legal Code)을 이해하기 쉽게 요약한 것입니다.

Disclaimer
이학석사 학위논문

Bradley-Terry model analysis
using varying coefficient

공변량 의존 브래들리-테리 모형을 이용한
평가자 편의 모형화 및 분석

2016년 8월

서울대학교 대학원
통계학과
정 유 빈
국문초록

쌍 비교를 통해 전체 개체를 서열화 하는 확률모형으로 브래들리-테리 모형이 있다. 비교를 통한 평가를 할 경우 객관적인 평가지표가 존재하지 않는 이상 평가자의 주관이 개입된다. 본 논문에서는 공정한 평가를 위해 기존의 브래들리-모형을 보완한 모형을 제안한다. 제안한 모형은 평가자와 평가대상자의 정보를 모두 포함하며, 평가자와 평가대상자의 상호작용을 평가자 의존 공변량으로 모형화하여 평가자의 주관을 효과적으로 제거시킨다. 또한 제안 모형으로 모의 실험을 진행하여 고정된 평가자와 평가대상자의 정보로 최적의 평가자 수와 평가자 당 평가 수를 찾는다.

주요어: 브래들리-테리 모형, 쌍 비교 분석, 평가자 의존 공변량

학 번: 2014-20283
Contents

1 서론 ........................................ 1

2 이론적 배경 .................................. 3
   2.1. Thurstonian model .......................... 3
   2.2. Bradley-Terry model ...................... 4
   2.3. Luce model ................................ 5
   2.4. 모수추정 .................................. 5

3 모형설명 ..................................... 7
   3.1. 기존의 브래들리-테리 모형 .................. 7
   3.2. 평가대상자의 정보가 포함된 브래들리-테리 모형 ...... 8
   3.3. 평가대상자, 평가자 정보가 포함된 브래들리-테리 모형 .. 9
   3.4. The Lasso ................................ 11

4 모의실험 .................................... 12
   4.1. 자료설명 .................................. 12
   4.2. 결과 ...................................... 15
       4.2.1. 모형비교 ................................ 16
       4.2.2. 모형 3.3이 유효한 평가자 수와 평가자 당 평가 수 ... 17
결론
List of Tables

4.1 역량 값의 관점에 따른 모형의 성능비교 .................. 17
List of Figures

4.1 $p_1, p_2$에 따른 모형 3.3 상관계수 .......................... 18
4.2 $p_1, p_2$에 따른 각 모형 간 상관계수 차이 .......................... 20
일상생활에서 비교를 통해 우열을 가리는 것은 흔히 있는 일이다. 이러한 비교를 수학적으로 분석을 가능케 하는 확률모형으로는 대표적으로 브래들리-테리 모형(Bradley-Terry model)을 들 수 있다. 브래들리-테리 모형은 쌍 비교(paired comparison)에 대한 자료가 있을 때 쌍의 선호도에 대한 순위를 비교할 수 있는 확률모형이다. 

$i$와 $j$라는 두 개의 아이템이 있을 때, $i$의 순위가 $j$보다 높을 확률 혹은 $i$가 $j$보다 선호 될 확률은 $Pr(i > j)$을 추정한다. 실생활에서는 운동경기의 승률을 예측하거나 정보검색(Information retrieval, IR)에서 문서의 랭킹을 매기는 용도로 쓰이고 있다.

순위를 비교하는 객관적인 기준이 존재하는 경우에는 쌍 비교에 대한 평가 결과로 순위를 세우는 기존의 브래들리-테리 모형이 합리적으로 쓰일 것이다. 그러나 객관적인 기준이 존재하지 않는 경우에는 평가자의 주관이 개입될 가능성이 매우 크지만 기 모형에서는 평가자의 주관을 제거하거나 조절할 수 없다. 따라서 후자의 경우 기존의 브래들리-테리 모형의 성능을 개선한 모형으로 평가를 해야 한다. 특히 사람이 사람을 평가한다면 평가자와 평가대상자 간 친소관계 등으로 인한 평가자의 주관을 제거하는 것은 필수적일 것이다.
따라서 본 논문에서는 평가자가 주관적인 평가를 할 경우에 초점을 맞추어 이를 보완한 모형 두 가지를 제안한다. 첫째, 평가대상자들의 기본적인 정보를 공변량으로 포함한 모형이다. 둘째, 평가를 하는 사람의 주관이 섞여 있음을 경우, 평가자의 객관적인 평가를 위해 평가자와 평가대상자의 정보를 모두 포함하며 평가자와 평가대상자의 상호작용을 평가자별 varying coefficient로 나타낸 모형이다. 두 번째 모형은 평가자의 주관을 배제한 평가대상자들의 순위를 구할 수 있다. 가장 기본적인 브레들리-테리 모형과 본 논문에서 제안한 두 모형에서 구한 평가대상자들의 순위를 실제 순위와 스피어만 순위 상관계수 (Spearman Correlation Coefficient)를 이용하여 비교한다. 또한 평가대상자들의 수와 개인의 속성의 수를 고정시켰을 경우, 평가자와 평가자 당 평가 수를 변화시키면서 최적의 평가자와 평가자 당 평가 수의 비율을 도출한다.

본 논문의 제2장에서는 쌍 비교에 대한 이론적인 배경에 대해 설명한다. 제3장에서는 논문에서 제안하는 모형과 모수 추정시 변수 선택 방법인 Lasso에 대해 간략히 소개한다. 이어서 제4장에서는 3장에서 소개한 모형을 이용한 모의실험 분석 결과에 대해 논의한다. 마지막으로 제5장에서는 결론을 도출하도록 한다.
Chapter 2

이론적 배경

선호도를 확률모형으로 제안한 인물에는 대표적으로 Louis Leon Thurstone, Ralph Allen Bradley, R. Duncan Luce가 있다. 이들이 제안한 각각의 모형의 특징과 브래들리-테리 모형이 갖는 이점에 대해서 살펴보도록 하겠다.

2.1. Thurstonian model


\[ Z_j = \lambda_j + \epsilon_j \quad \text{for} \quad j = 1, ..., p, \quad \epsilon_j \overset{iid}{\sim} F \quad \text{and} \quad \lambda_j \in \mathbb{R} \quad (2.1) \]

플레이어 j가 플레이어k보다 선호되는 사건을 \( \{ j \rightarrow k \} \)로 표현한다면 잠재변수로 표현한 사건은 아래와 같이 나타낼 수 있다.
\{j \to k\} = \{Z_j > Z_k\}

Thurstone은 (2.1)의 확률분포 \(F\)에 정규 분포를 사용하였으며, 모수 \(\lambda_j\)는 플레이어 \(j\)의 평균 선호도를 나타낸다.

2.2. Bradley-Terry model


플레이어 \(j\)와 \(k\)의 모수를 \(u_j, u_k\)이라 하자. \((u_j, u_k\)는 0보다 큰 실수) 플레이어 \(j\)가 플레이어 \(k\)보다 선호 될 확률(\(\theta_{jk}\))은 다음과 같이 나타내며 각각의 선호도는 독립이라 가정한다.

\[
\theta_{jk} = \frac{u_j}{u_j + u_k}
\]

플레이어 \(j\)가 \(k\)보다 선호 됨을 나타낸 지시변수를 \(y_{jk}\)라고 하고 각각의 \(y_{jk}\)는 독립이라고 가정하자. 가정에 따라 \(y_{jk}\)는 아래와 같이 나타낼 수 있다.

\[
y_{jk} \sim \text{Bernoulli}(\theta_{jk})
\]

따라서 \(y_{jk}\)가 1일 경우 플레이어 \(j\)가 선호되고, 0일 경우 플레이어 \(k\)가 선호된다고 할 수 있다.
2.3. Luce model

R. Duncan Luce(1925-2012)는 choice axiom과 Luce model을 제안하였다. Luce model은 인간의 선택 과정을 확률적으로 구축한 모형이다.  $j$개의 아이템이 있을 때 $i$번째 아이템을 선택할 확률은 다음과 같이 나타낸다(Luce, 1959).

$$\Pr(i) = \sum_{k=1}^{j} \frac{\exp(\lambda_i)}{\sum_{k \leq i} \exp(\lambda_j)}$$

Luce는 Thurstonian model에서 제안한 잠재변수를 이용하여 브래들리-테리 모형을 설명하여, 브래들리-테리 모형이 Luce model의 특별한 경우임을 보였 다. $\epsilon_j$가 표준 엘 mund 분포(standard Gumbel distribution)를 따른다고 가정하면 $\Pr\{j \rightarrow k\}$은 아래와 같이 나타낼 수 있다(Luce, 1959; Plackett, 1975).

$$\Pr\{j \rightarrow k\} = \frac{\exp(\lambda_j)}{\exp(\lambda_j) + \exp(\lambda_k)}$$

위의 식은 흔히 알고 있는 로지스틱 회귀분석의 확률 모형과 동일한 것으로 모수추정과 해석에 있어 매우 용이하다.

2.4. 모수추정

모수를 추정하는 방법은 다음과 같다. $y_{ijk}$가 플레이어 $j, k$의 선호도를 나타내는 지시변수라 하자. $y_{ijk}$는 독립이라 가정하면 가능도 함수는 다음과 같이 나타낼 수 있다.

$$L(\lambda_1, ..., \lambda_p) = \prod_{i,j,k} \left(\frac{\exp(\lambda_j)}{\exp(\lambda_j) + \exp(\lambda_k)}\right)^{y_{ijk}} \left(\frac{\exp(\lambda_k)}{\exp(\lambda_j) + \exp(\lambda_k)}\right)^{1-y_{ijk}}$$
모형이 참일 경우, MLE는 정칙조건 하에서 접근적으로 정규성을 만족하는 좋은 성질을 지닌 일치 추정량이 된다.

이처럼 잠재변수를 통해 나타낸 브래들리-테리 모형은 \( \epsilon_j \) 가 정렬분포를 따르는 Turstonian model과 동일한 모형으로 볼 수 있다. 해당 모형은 간단하게 로지스틱 회귀분석으로 나타낼 수 있다. 또한 모형이 가정이 참일 경우 접근적 성질로 인해 MLE를 쉽게 도출할 수 있어 쌍 비교 분석 시 브래들리-테리 모형이 널리 사용된다.
Chapter 3

모형설명

3.1. 기존의 브래들리-테리 모형

가장 기본적인 브래들리-테리 모형으로 개개인을 1:1로 비교하여 평강을 매긴다고 하자. 이때 사람들마다 고유한 하나의 역량 값이 존재한다고 가정하고, 그 값에 의해서만 승패에 대한 확률모형이 만들어진다. (예: 팔씨름, 역량 값-해당 선수의 힘, 역량 값(힘)이 셀수록 이길 확률이 높도록 확률 모델을 만들다.)

즉, j와 k라는 평가 대상자간의 있을 때 j가 k보다 선호 될 확률Pr(j > k)은 다음과 같이 나타낼 수 있다.

\[ Z_j = \lambda_j + \epsilon_i, \quad Z_k = \lambda_k + \epsilon_k \]  \hspace{1cm} (3.1)

\(Z_j\)와 \(Z_k\)는 \(j\)와 \(k\)의 고유 역량 값인 \(\lambda_j, \lambda_k\)에 대한 잠재변수이다. 연속형 변수이며 \(Z_i\)들은 독립이다.
이 모형은 오직 쌍에 대한 승패 정보만을 사용하게 된다. 즉, 모든 사람들을 오직 하나의 역량 값으로만 대표하기 때문에 사람들의 고유한 정보를 전혀 사용하지 않게 된다. (인구학적 정보 등) 따라서 이를 반영하여 사람들의 고유 정보를 공변량으로 반영하여 발전시킨 모형을 다음과 같다.

3.2. 평가대상자의 정보가 포함된 브래들리-테리 모형

기존의 브래들리-테리 모형에 사람들의 고유 정보까지 포함하여 만든 브래들리-테리 모형이다. 승패 정보뿐 아니라 사람들의 정보도 사용하기 때문에 좀 더 정확하고 좋은 결과를 만들 수 있다.

\[
Z(j, x_i) = \lambda_j' + \sum_{m=1}^{p} x_{jm}\beta_m + \epsilon_j' \\
Z(k, x_i) = \lambda_k' + \sum_{m=1}^{p} x_{km}\beta_m + \epsilon_k'
\]

\(Z_{j}\)와 \(Z_{k}\)는 앞서 언급했던 고유 역량 값에 대한 잔재변수이다. \(x_{jm}\)와 \(x_{km}\)는 평가대상자 정보이다. 평가대상자의 개인적인 속성을 나타내며 연속형 변수와 범주형 변수가 혼합되어 있다. 모의실험에서는 분석과 모형의 해석의 용이성을 위해 범주형 변수로 변환하여 분석을 진행한다. \(\beta_m\)은 평가대상자 효과이다. 평가대상자의 개인적인 속성(\(x_i\))가 추정된 순위에 영향을 미치는 정도를 수치로 나타낸다.
$$\Pr(Z_j > Z_k) = \Pr(\lambda'_j - \lambda'_k > \epsilon_k - \epsilon'_j)$$
$$= \frac{\exp(\lambda'_j - \lambda'_k) + \sum_{m=1}^{p} (x_{jm} - x_{km})\beta_m^{ij(j>k)}}{1 + \exp(\lambda'_j - \lambda'_k) + \sum_{m=1}^{p} (x_{jm} - x_{km})\beta_m} \quad (3.4)$$

하지만 이 모형 역시 제3자가 승패를 결정하는 상황에서는 적합하지 않을 수 있다. 두 사람의 고유한 정보 외에도 제3자의 주관적인 기준이 게임되기 때문이다. 따라서 제3자(평가자)가 승패를 결정하는 경우에는 그 사람이 중요하게 생각하는 기준 또한 모형에 포함되어야 한다.

### 3.3. 평가대상자, 평가자 정보가 포함된 브래들리-테리 모형

3.2의 브래들리-테리 모형에 평가자의 정보를 추가한 모형이다. 각 평가자들마다 평가 기준이 다르다는 가정 하에 3.2의 모형보다 더 좋은 예측 결과를 기대할 수 있다. 평가자에 따라 평가대상자의 어떤 점을 중요하게 생각하는지 모르기 때문에 가능한 모든 경우의 수를 고려하여 평가자들의 기준을 모형에 포함시킨다.

$$Z(j, x_i, \nu) = \lambda''_j + \sum_{m=1}^{p} x_{jm}\beta_m + \sum_{m=1}^{p} \sum_{n=1}^{q} (\nu_{hn} - \bar{\nu}_n)x_{jm}\delta_{mn} + \epsilon''_j \quad (3.5)$$
$$Z(k, x_i, \nu) = \lambda''_k + \sum_{m=1}^{p} x_{km}\beta_m + \sum_{m=1}^{p} \sum_{n=1}^{q} (\nu_{hn} - \bar{\nu}_n)x_{km}\delta_{mn} + \epsilon''_k$$

$Z_j$와 $Z_k$는 앞서 언급했던 고유 역량 값에 대한 잠재변수이다. $x_{jm}$와 $x_{km}$는 평가대상자 정보이며 $\beta_m$은 평가대상자 효과이다. $\nu_{hn}$는 평가자 $h$의 정보를
나타낸다. 모의실험에서는 평가대상자의 정보의 개수인 $x_i$와 일치하게 설정했으나 평가자의 자질을 평가하는 항목이 추가되거나 $x_i$보다 적은 항목으로 설정하여도 무방하다. 평가자 $h$는 평가대상자 $j,k$를 평가하는데 평가와 평가 대상자 사이의 상호작용을 평가자 효과 $(\delta_{mn})$로 나타낸다. $\delta_{mn}$의 크기에 따라 평가자가 평가대상자의 주관적 평가가 순위에 어떤 영향을 나타내는지 알 수 있다.

\[
\Pr(Z_j > Z_k \mid \lambda_h) = \Pr(A + B + C > \epsilon''_k - \epsilon''_j), \epsilon''_j, \epsilon''_k \sim \text{Gumbel dist.}
\]

\[
A = \lambda''_j - \lambda''_k
\]

\[
B = \sum_{m=1}^{p} (x_{jm} - x_{km})\beta_m
\]

\[
C = \sum_{m=1}^{p} \sum_{n=1}^{q} \{(\nu_{hn} - \bar{\nu}_n)\delta_{mn}\}(x_{jm} - x_{km})
\]

위의 모형의 경우 설명변수의 절대적인 수치는 중요하지 않고, 설명변수의 높 낮이면 중요하기 때문에 $\nu$에서 $\bar{\nu}_n$을 빼줌으로써 센터링을 시킨다. 이 과정을 통해 identifiable 하지 않은 모수가 identifiable 하게 된다. 하지만 이처럼 너무 많은 정보를 모형에 집어넣게 되면 앞의 모형들과 비교하여 너무 복잡하다는 단점이 생긴다. 모형을 결정하는 변수의 가짓수가 너무 많기 때문에 과적합 (overfitting)이 우려된다. 이를 위해 최신 통계 기법(Lasso)을 이용하여 모든 변수를 사용하지 않고, 유의한 변수들만을 선택하여 사용한다.
3.4. The Lasso

고차원 또는 대용량 자료를 분석에서 관측치의 수보다 설명변수의 수가 현저히 많은 경우 과적합이 일어난다. 과적합을 피하기 위해 변수 선택(Variable selection), 차원 축소(Dimension reduction techniques), 축소 추정법(Shrinkage method) 등의 해결 방법이 있다. 그중 Lasso (The least absolute shrinkage and selection operator)는 축소 추정법으로 범접한 함수(penalty function)를 통해 계수 추정 및 변수 선택을 동시에 수행한다. 3.3처럼 복잡한 모형의 모수를 추정할 때, 모든 변수를 추정하지 않고 중요한 변수를 골라 모수 추정을 하게 된다.

조절 모수(tuning parameter)를 \(\lambda \geq 0\)이라 할 때 Lasso의 범점함수는 아래와 같이 정의한다(Tibshirani, 1996).

\[
Pr_{\lambda}(\beta) = \lambda \| \beta \|_1 = \lambda \sum_{j=1}^{p} |\beta_j|
\]

Lasso의 손실 함수는 기존의 손실 함수에 범점함을 집어넣어 다음과 같이 정의하며 라그랑지 승수법 (The Method of Lagrange Multipliers)에 의해 아래의 두 식은 동치이다.

\[
l(\beta) = \sum_{i=1}^{n} l(y_i, X_i^T \beta) + \lambda \sum_{j=1}^{p} |\beta_j|
\]

equal to, minimize \[
\sum_{i=1}^{n} l(y_i, X_i^T \beta) \text{ subject to } \lambda \sum_{j=1}^{p} |\beta_j| \leq c
\]

조절 모수를 통해 \(\beta\)값을 조절할 수 있어 계수 축소 역할을 한다. 또한 범점함수가 0에서 미분이 불가능한 오목함수이기 때문에 미분이 불가능한 구간에서 \(\hat{\beta}\)는 0이 되고 이로 인해 Lasso는 성감성(Sparsity)을 갖는다. 즉 모형의 복잡도를 감소시키고 해석력을 높이게 된다. 본 논문에서는 통계 분석 프로그램 R의 glmnet 패키지를 이용하여 유의한 변수를 선택하고 모수를 추정하였다.
Chapter 4

모의실험

4.1. 자료설명

모의 실험을 위한 자료의 생성은 아래의 방법으로 하였다.

- \( N = 100 \) : 전체 평가대상자 수
- \( p = 20 \) : 개인의 속성 수
- \( M = N \times p_1 \) : 평가자 수
- \( r = N - 1 \times C_2 \times p_2 \) : 평가자 당 평가 수
- \( p_1 = \{0.05, 0.10, 0.15, 0.20, ..., 0.45\} \) : 평가자 수를 조절하는 비율
- \( p_2 = \{0.005, 0.010, 0.015, 0.020, ..., 0.100\} \) : 평가자 당 평가 대상자 수를 조절하는 비율
실제 평가가 있을 경우, 평가대상자와 개인의 속성은 고정 되는 수이므로 모의실험의 전체 평가대상자 $N$명과 $p$개의 개인 속성을 각각 100명과 20개로 고정시킨다. 전체 평가대상자 $N$에 포함되는 평가를 하는 인원과 평가자가 평가를 하는 수는 경제적인 선에서 조절이 가능하다. 따라서 평가자 수 $M$은 $N$에 $p_1$을 곱하여 나타낸다. 같은 방법으로 평가자 당 평가 수 $r$은 평가자를 제외한 가능한 모든 조합인 $N-1\binom{}{2}$에 $p_2$라는 조절 변수를 곱하여 $r$값을 조절한다. 조절 변수 $p_1$은 0.05에서 0.45까지 0.05를 간격으로 설정하였다. $p_2$는 0.005에서 0.1까지 0.05를 간격으로 설정하였다. $p_1$, $p_2$는 0에서 1까지의 값이 가능하나 과반수의 평가자가 평가를 하게 될 경우 전체 평가 수가 너무 많아지고 현실 적으로 불가능하기 때문에 $p_1$의 최댓값은 0.4로 하였다. 마찬가지로 한 평가자 당 너무 많은 조합의 평가는 어려우므로 $p_2$의 최댓값은 0.1로 하였다.

그 외의 모수의 샘플링과 설정은 다음과 같이 진행하였다.

1. 개인 속성 ($p$) 설정
개인의 속성은 총 20가지로 $X_1, X_2, ..., X_{20}$으로 나타낸다.

$$X_1 \sim N(0, \sigma_1)$$

$$X_2 \sim N(0, \sigma_2)$$

$$\vdots$$

$$X_{20} \sim N(0, \sigma_{20})$$

이때 $\sigma_i$는 감마 분포에서 추출하였다. ($\sigma_i \sim Gamma(1, 1)$)
2. 모수 샘플링: 역량 값($\lambda$), 평가대상자 효과($\beta$), 평가자 효과($\delta$)

$$\lambda \sim N(0, 1)$$
$$\beta \sim N(0, 0.1)$$
$$\delta \sim N(0, 0.05)$$

역량 값, 평가대상자 효과, 평가자 효과는 평균이 0인 정규분포에서 추출하였다. 분산은 모수의 중요도에 따라 영향을 가장 크게 미칠 것이라고 생각되는 모수의 분산을 크게 하여 역량 값은 1, 평가대상자 효과는 0.1, 평가자 효과는 0.05로 하였다. 모수의 중요도는 평가 시 임의로 결정한 것이며 실제 자료가 있을 경우에는 모수 추정량의 절댓값으로 판단할 수 있을 것이다.

3. 평가자, 평가자 당 평가 수 샘플링

평가자는 앞서 언급한 방법으로 $M$명을 선택하게 된다. 평가자는 100명 중 $M$명을 뽑는 방법 즉, $100C_M$으로 추출한다. 평가자 당 평가 수 역시 조합(Combination)으로 추출하는데, 평가자 본인을 제외한 99명의 평가 대상자 2명을 뽑는다. ($99C_2$) 평가자 당 평가 수가 $r$번 진행되므로 평가자 한 명당 $99C_2 \times r$번의 평가를 하게 된다. 같은 방법으로 전체 평가자가 $M$명이므로 모든 평가자에 의해 총 $99C_2 \times r \times M$번의 평가가 진행된다.

4. 평가 쌍에 대한 평가 결과 샘플링

평가 쌍이 ($j, k$)일 때, 평가자가 평가대상자 $j$를 선호할 확률, $\Pr(Z_j > Z_k)$은 모형 별로 식 (3.2), (3.4), (3.6)과 같이 나타낼 수 있다.

위의 확률들을 $\theta_{jk}$로 나타내고 $\theta_{jk}$는 베르누이 분포를 따른다고 가정하자. 평가자는 Bernoulli($\theta_{jk}$) 값이 1일 경우 평가대상자 $j$를 0일 경우 평가대상자 $k$를 선택하게 된다.
\[
\text{Bernoulli}(\theta_{jk}) = \begin{cases} 
1, \text{ 평가대상자 } j\text{ 선택} \\
0, \text{ 평가대상자 } k\text{ 선택}
\end{cases}
\]

4.2. 결과

모의 실험의 결과를 보기 앞서 성능비교에 사용된 측정지표와 개인의 역량 값을 정의하겠다.

1. 성능평가 지표: 스피어만 순위 상관계수
   앞서 소개한 세 모형은 모두 평가 쌍의 평가 정보를 통해 평가대상자들의 순위를 도출해낸다. 이처럼 측정값이 순위 척도로 주어질 경우, 실제 순위와 추정한 순위가 유사한지를 모형의 평가 지표로 하며 순위 상관계수를 사용하여 나타낸다. 대표적인 순위 상관계수로는 스피어만 순위 상관계수(Spearman’s rank correlation coefficient, \(\rho\))와 켄달 순위 상관계수(Kendall rank correlation coefficient, \(\tau\))가 있다. 본 논문에서는 모형들의 성능을 비교하기 위해 스피어만 순위 상관계수를 사용하였다. 스피어만 순위 상관계수는 순위 상관관계를 측정하는 비모수적인 방법으로 두 순위 사이의 상관관계를 측정한다. 스피어만 순위 상관계수는 다른 상관계수들과 마찬가지로 \(\rho\)는 -1에서 1 사이의 값을 가진다. 1은 두 가지 순위가 완벽히 일치함을 나타내며, -1은 두 가지 순위가 반대의 방향으로 완벽히 일치함을 나타낸다. 0은 두 순위 간 전혀 상관관계가 없음을 나타낸다. 따라서 1에 가까울수록 양의 상관관계가 크고 -1에 가까울수록 음의 상관관계가 큼을 의미한다.

2. 평가대상자의 역량 값 설정
   제안된 모형에는 사람의 능력을 표현하는 2 가지 관점이 있다. 첫 번째는
식 (3.5)의 \(\lambda_i\)와 같이 개인 고유의 역량 값만 능력으로 보는 방법이다. \(\lambda_i\) 는 노력 없이 얻어지는 선천적인 능력이다. 예를 들어 운동신경, IQ 등을 생각할 수 있다. 두 번째는 식 (3.5)의 \(\lambda_i + \sum_{m=1}^{p} x_{im} \beta_m\)로 고유의 역량 값 (\(\lambda_i\))에 개인의 사회적인 속성(\(\sum_{m=1}^{p} x_{im} \beta_m\))을 더한 값까지 개인의 능력으로 보는 방법이다. 사회적인 속성\(x_i\)에는 기본적으로 나이, 학력, 연봉, 경력 등이 들어갈 수 있는데, 이는 사회적으로 생성된 값으로 개인의 노력이나 사회생활과 자연스레 얻어지는 것들로 이 역시 평가대상자들의 능력을 형성하는데 의미 있는 역할을 한다고 판단하였다. 이후 위에서 언급한 두 가지 역량 값을 다음과 같이 간단히 표기하였다.

a) \(\lambda\) : 개인 고유의 역량 값

b) \(\lambda + X\beta\) : 사회적 능력을 포함한 역량 값

따라서 두 가지 관점에 따라 \(\lambda\)값과 \(\lambda + X\beta\)값을 기준으로 실제 순위와 모형 3.1~모형 3.3의 추정 순위를 도출하며, 실제 순위와 추정하여 구한 순위들의 상관계수를 측정하여 모형의 성능 평가를 실시한다.

### 4.2.1. 모형 비교

평가자의 수(\(M\))를 조절하는 \(p_1\)은 0.3이고, 평가자 당 평가 수(\(r\))를 조절하는 \(p_2\)는 0.06일 때, 위에서 언급한 두 가지 관점 별로 모형 비교를 실시했다. 시뮬레이션은 50회 진행하였으며 각 모형 별로 계산된 상관계수의 평균으로 두 순위 간의 일치성을 판단하였다. 결과는 표 4.1과 같다. 표 4.1은 각 관점 당 1. 실제 순위와 브래들리-데리 모형으로 추정한 순위(\(\hat{\lambda}_{3,1}\))의 상관계수, 2. 실제 순위와 평가대상자의 정보를 포함한 모형으로 개인의 능력을 고유 역량 값으로만 추정한 순위(\(\hat{\lambda}_{3,2}\))의 상관계수, 3. 실제 순위와 평가대상자의 정보를 포함한 모형으로 개인의 능력을 고유 역량 값과 사회적 능력의 합으로 추정한
Table 4.1: 역량 값의 관점에 따른 모형의 성능 비교

<table>
<thead>
<tr>
<th></th>
<th>λ</th>
<th>λ + Xβ</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>0.1678</td>
<td>0.2120</td>
</tr>
<tr>
<td>2.</td>
<td>0.2026</td>
<td>0.2042</td>
</tr>
<tr>
<td>3.</td>
<td>0.0929</td>
<td>0.1544</td>
</tr>
<tr>
<td>4.</td>
<td>0.6363</td>
<td>0.6395</td>
</tr>
<tr>
<td>5.</td>
<td>0.5924</td>
<td>0.6648</td>
</tr>
</tbody>
</table>

순위(\(\hat{λ}_{3,2} + X\hat{β}_{3,2}\))의 상관계수, 4. 실험 순위와 평가대상자, 평가자의 정보를 포함한 모형으로 개인의 능력을 고유 역량 값으로만 추정한 순위(\(λ_{3,3}\))의 상관계수, 5. 실제 순위와 평가대상자, 평가자의 정보를 포함한 모형으로 개인의 능력을 고유역량값으로만 추정한 순위(\(\hat{λ}_{3,3} + X\hat{β}_{3,3}\))의 상관계수를 나타냈다. 결과적으로 \(λ + Xβ\)를 기준으로 구한 실제 순위와 \(\hat{λ}_{3,3} + X\hat{β}_{3,3}\)를 기준으로 구한 추정 순위에 대한 상관계수가 가장 컸으며, \(λ\)를 기준으로 구한 실제 순위와 \(\hat{λ}_{3,2} + X\hat{β}_{3,2}\)의 추정 순위에 대한 상관계수가 가장 작았다. 실제 순위와 추정 순위의 관점에 따라 상관계수가 더 높아지기도 하고 낮아지기도 하지만 전반적으로 모형 3.3의 상관계수가 상당히 크다. 모형의 복잡도를 고려하여도 다른 모형과 비교하면 때 크게는 6배, 평균적으로 3배가 넘는 결과 값으로 순위를 매우 효과적으로 추정하였다고 볼 수 있다.

4.2.2. 모형 3.3이 유 효한 평가자 수와 평가자 당 평가 수

위의 모의실험에서 \(p_1\)과 \(p_2\)가 각각 0.3, 0.06일 때, 모형 3.3이 효과적인 모형이라는 결과를 얻었다. 하지만 \(p_1\)과 \(p_2\)는 평가 시에 유동적으로 선택을 할 수 있는 값이므로 \(p_1\)과 \(p_2\)의 범위를 확장시켜 생각해 보겠다. 모든 평가대상자들이 평
Figure 4.1: $p_1, p_2$에 따른 모형 3.3 상관계수

가를 받는 동시에 자신을 제외한 모든 평가대상자들의 쌍을 비교한다면 그로 도출된 순위가 가장 높은 정확도를 보일 것이다. 하지만 이는 경제적인 입장에서 거의 불가능하다. 따라서 평가자의 수와 평가자 당 평가 수가 많지 않아도 상관계수의 값이 높은 지점에서 평가를 실시한다면 정확도는 물론 경제적인 이점을 얻을 수 있을 것이다.

그림 4.1은 사람의 능력을 보는 관점 별로 $p_1, p_2$에 따른 실제 순위와 모형 3.3 순위의 상관계수를 도식화 한 것이다. 빨간색으로 갈수록 상관계수가 높고 파란 부분으로 갈수록 상관계수가 낮다. 전반적으로 오른쪽 그래프가 왼쪽의 그래프보다 빨간색 부분이 많지만 전반적인 상관계수의 분포는 유사하다. 또한 두 그래프 모두 $p_1$은 0.2에서 0.3사이, $p_2$은 0.3 근처에서 주변의 색과는 다르게 상관계수가 상당히 높은 것으로 추정되었다. 이처럼 평가대상자 수가 고정되어 있는 경우 경제성을 고려한 $p_1, p_2$를 설정하여 실험 계획을 할 수 있다.
그림 4.2는 \(p_1\)과 \(p_2\)를 확장시켜 모형의 성능 차이를 도식화한 것이다. 모형별로 계산된 상관계수를 빼주어 각 포인트별로 모형의 정확도를 보여준다. 절댓값이 클수록 성능 차이가 많이 나고 절댓값이 0에 가까울수록 성능 차이가 없다고 판단할 수 있다. 그림 4.2의 (a)와 (b)는 전반적으로 뚜렷한 색을 띈다. 또한 모형 3.3의 상관계수에서 모형 3.1의 상관계수를 빼 (c), (d)와 모형 3.3의 상관계수에서 모형 3.1의 상관계수를 빼 (e), (f)의 그래프의 추세가 거의 일치하는 것을 볼 수 있다. 이는 두 모형 간 성능 차이가 크게 없음을 시사한다. 반면 그림 4.2의 (c)와 (d)는 \(p_1\)과 \(p_2\)가 증가할수록 붉은 색을 띈다. 이는 \(p_1\)과 \(p_2\)가 증가할수록 모형의 복잡성을 비교적 많이 반영함을 의미한다.
Figure 4.2: $p_1, p_2$에 따른 각 모형 간 상관계수 차이

(a) $\hat{\lambda}_{3.2} - \hat{\lambda}_{3.1}$

(b) $(\hat{\lambda}_{3.2} + \hat{X}_3\hat{\beta}_{3.2}) - \hat{\lambda}_{3.1}$

(c) $\hat{\lambda}_{3.3} - \hat{\lambda}_{3.1}$

(d) $(\hat{\lambda}_{3.3} + \hat{X}_3\hat{\beta}_{3.3}) - \hat{\lambda}_{3.1}$

(e) $\hat{\lambda}_{3.3} - \hat{\lambda}_{3.2}$

(f) $(\hat{\lambda}_{3.3} + \hat{X}_3\hat{\beta}_{3.3}) - (\hat{\lambda}_{3.1} + \hat{X}_3\hat{\beta}_{3.3})$
Chapter 5

결론

본 논문에서는 주관적 기준을 가진 쌍 비교 분석을 할 때, 평가자의 주관을 제거시키는 확률모형을 제안하였다. 가장 기본적인 브래들리-테리 모형에 평가자와 평가대상자의 정보를 포함시키고, 평가자와 평가대상자의 상호작용을 평가자 의존 공변량(varying coefficient)으로 모형화하였다. 모의실험 결과, 개인의 역량을 보는 두 가지 관점 모두에서 실제 순위와 모형 3.3의 추정 순위가 다른 모형들의 상관계수보다 두 배 가량 큰 값으로 상관관계가 가장 높았다. 또한 고정된 평가대상과 개인의 속성 하에서 적당히 높은 상관관계를 갖지만 평가자 수와 평가자 당 평가 수가 적은 부분이 존재하여 경제적으로 이점을 갖는 구간을 찾을 수 있었다. 위의 모형은 Lasso로 상당수의 변수가 선택 및 축소되었지만 여전히 복잡한 모형이다. 따라서 고차원 분석에서 로지스틱 회귀분석이 갖는 약점을 여전히 존재한다. 이를 개선하는 방법에 대한 연구가 이루어진다면 본 논문의 제안 모형보다 효과적으로 평가자의 주관을 제거하는 모형을 도출할 수 있을 것으로 사료된다.
References


Abstract

Yoobin Jung
The Department of Statistics
The Graduate School
Seoul National University

The Bradley-Terry model is a probability model that can predict the rank of items through paired comparisons. If there is no objective standards of valuation, appraiser makes a decision subjectively in comparison.

In this thesis, I introduce the Bradley-Terry model using varying coefficient for fair evaluation. Proposed model includes information of appraiser and appraisee unlike simple Bradley-Terry model. And this model models interactions between appraiser and appraisee as varying coefficient. It can remove the appraiser’s subject. Also I can find the optimal number of appraisers and evaluation per appraiser through simulations.

Keyword: Bradley-Terry model, Pairwise comparison analysis, Varying coefficient

Student Number: 2014-20283