마코프랜덤필드를 이용한 설문지에서의 비정상 반응 탐지

김형준[O1]. 김청택[2]. 장병탁[3].
서울대학교 인지과학 협동과정[1]. 서울대학교 심리학과[2]. 서울대학교 컴퓨터공학부[3].
soeque1@snu.ac.kr. ctkim@snu.ac.kr. btzhang@bi.snu.ac.kr.

Detection of Aberrant Response on a Questionnaire using MRFs

HyungJun Kim[O1]. Cheongtag Kim[2]. Byoungtak Zhang[3].
Interdisciplinary Program in Cognitive Science[1]. Department of Psychology[2].
School of Computer Science and Engineering[3],
Seoul National University, Seoul 151-742, Korea

요 약

본 연구는 마코프랜덤필드(MRFs)를 활용하여 설문지에서 무작위반응과 같은 비정상적인 반응 패턴을 탐지하는데 주목적이 있다. 본 연구 모형은 1) 특정 문항에 대한 전체 사람들의 반응율과 2) 이전 문항의 반응과 현재 문항의 반응에 대한 전이를 정문항 및 역문항의 정보와 결합하여 구성하였다. 우선, 본 연구에서는 MRFs가 비정상적인 반응 패턴을 잘 탐지할 수 있음을 확인하기 위해 ROC curve의 AUC(Area Under the Curve)를 제시하였다. 모형을 평가하기 위해 1) 실제 설문 자료를 이용하여 비정상적인 반응 패턴의 비율을 조정해가면서, IRT, SVM, MRFs의 ROC Curve의 AUC를 비교하였고, 2) 더 정밀한 비교를 위해 정상적인 사람들의 반응 패턴 내에서 무선 반응 조작 및 연속 반응 조작을 통하여 IRT, SVM, MRFs의 ROC Curve의 AUC를 비교하였다. 연구 결과, MRFs 모형이 다른 모형에 비해 더 일관되게 비정상적인 반응을 잘 탐지하였다.

1. 서 론

인간 데이터의 경우 많은 잡음(noise)이 존재한다. 본 연구에서는 마코프랜덤필드(Markov Random Fields, MRFs)를 활용하여 '설문지'에서 나타나는 잡음, 즉 비정상적(aberrant) 반응 패턴을 탐지하는 모형을 제안하자고 한다. 설문지에서는 피로 혹은 무관심 등으로 인하여 무선적(random) 반응 또는 한 줄로 응답하기 등의 잡음이 존재할 수 있다. 이러한 잡음이 섞인 개인의 반응을 비정상적(aberrant) 반응 패턴이라고 정의한다.

본 논문에서는 MRFs의 오류 함수를 개인들의 반응들과 설문지 특성을 조화시켜 모형을 구성할 수 있으며, 이를 이용해 비정상적인 반응을 탐지할 수 있음을 제안하려고 한다. 이를 위해 1) 기존에 심리학 및 교육학에서 비정상적인 반응을 탐지하던 문항반응이론, 2) 감독학습이지만 분류정확도가 높은 지지벡터머신, 그리고 3) 본 연구에서 제안하는 MRFs 모형의 분류 변별도를 비교하고자 한다. 이를 위해 세 모형에 대하여 각각 ROC curve의 곡선하면적(Area, Under the Curve, AUC)을 제시하여 마코프랜덤필드 모형이 비정상적인 반응을 안정적으로 잘 탐지할 수 있는지 평가하려 한다.

2. 문항반응이론(IRT)에 근거한 개인 적합도 지수

설문지에서의 비정상 반응 패턴을 탐지하기 위해 가장 흔히 사용되는 방식은 문항반응이론(Item Response Theory, IRT)에 근거한 개인별 로그 우도(log-likelihood = l_0)[1,2]이다. [3]은 이를 개인별로 표준화한 lz가 l_0 보다 더 잘 탐지 할 수 있음을 밝혔다. 보통의 경우 l_z가 -2 보다 작으면 비정상적인 반응패턴으로 분류한다[4].

식 (1)은 이분(binary) 반응 패턴이 주어졌을 때, 2모수 로지스틱 모형에서 k 번째 피험자가 i 번째 문항에서 "Yes"로 응답할 확률이고, 식 (2)는 k 번째 피험자의 개인 적합도(로그 우도, l_0)이다.

$$P(X_{ik} = \text{“}Yes\text{”}) = \frac{1}{1+exp(-\alpha_i(\theta_k - \beta_i))} \quad (1)$$

$$l_o = \sum_i logP_i(\hat{\theta}) \quad (2)$$

문항반응이론은 1) 국지독립성(local independence) 2) 단조 증가성(monotonicity)의 가정이 존재한다. 그런데, 한 번호로 일관된 반응 등의 비정상적인 반응 패턴의 경우 위 두 가지 가정에 위배되므로, 비정상적인 반응 패턴이 많을 경우 모형의 파라미터들을 불안정하게 추정할 수 있다.

3. 마코프랜덤필드(MRFs)를 활용한 개인 적합도 지수

MRFs는 이미지 잡음 제거, 모서리 및 물체 등의 객체 탐지, 이미지 분류 등 컴퓨터 시각 분야에서 다양한 방식으로 활용되어 왔다[5-6].

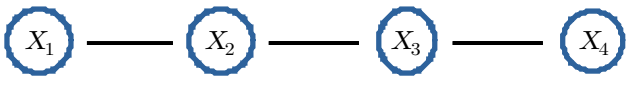

그림 1 이분법적 선형 연쇄 마코프랜덤필드 모형

그림 1에서 X_i는 각 i 번째 문항들에 대한 사람들의 반응들을 나타내고, 반응패턴(X)에 대한 확률은 식 (3) 과 같이 나타낼 수 있다.

$$P(X_1, X_2, \cdots X_I) = \frac{1}{Z}\prod_{i}^{I}\psi_i(X_i)\prod_{i}^{I-1}\psi_{i,j}(X_i, X_j) \quad (3)$$

$$Z = \sum_{X_1}\sum_{X_2}\cdots\sum_{X_I}\prod_{i}^{I}\psi_i(X_i)\prod_{i}^{I-1}\psi_{i,j}(X_i, X_j)$$

보통 설문지를 만들 때, 설문의 집중도 혹은 일관된 응답과 같은 편향을 감소시키기 위해 역문항을 사용한다. '반사회성'을 측정하기 위하여 다음과 같은 4 문항 설문지를 만들었다고 하자.

1번(정문항) 다른 사람들의 욕구나 감정에 둔감하다
2번(역문항) 호감이 느껴지고 괜찮은 인상을 준다
3번(정문항) 공격적으로 행동한다
4번(정문항) 자기중심적이다

반사회성을 가진 사람들은 1번, 3번, 4번 문항에서 "YES"로 응답하도록 기대되며, 2번 문항에서는 "NO"로 응답하도록 기대된다. 따라서, 1번, 3번, 4번 문항은 정문항이라 정의하고, 2번 문항은 역문항이라 정의한다.

3번, 4번 문항은 모두 '반사회성'을 측정하지만, '반사회성'을 측정하는 강도에 따라 $\psi_3(X_3 = Yes) \neq \psi_4(X_4 = Yes)$ 일 수 있다. 따라서, 본 모형에서는 식 (4)의 e^{θ_i}를 문항마다 모두 다르게 추정하였다.

$$\psi_i(X_i) = \left(e^{\theta_i}, e^0\right) \quad (4)$$

식 (5) $\psi_{i,j}(X_i, X_j)$은 2개의 λ_k를 사용하였다. 1) 정문항과 역문항 혹은 역문항과 정문항이 이어질 경우를 '역전이연쇄'라고 정의하고, 정문항과 정문항 혹은 역문항과 역문항이 이어질 경우를 '정전이연쇄'라고 정의한다.

$\psi_{i,j}(X_i, X_j)$ 행렬에서, 1) '역전이연쇄'의 경우에는 $e^{\lambda_{k=1}}$이 e^0보다 큰 값을 가지게 되고, 반대로, 2) '정전이연쇄'의 경우에는 $e^{\lambda_{k=2}}$는 e^0 = 1 보다 작은 값을 가지게 된다.

$$\psi_{i,j}(X_i, X_j) = \begin{pmatrix} e^0 & e^{\lambda_k} \\ e^{\lambda_k} & e^0 \end{pmatrix}, \quad \begin{cases} \lambda_{k=1} & \text{if} \quad \text{역전이연쇄} \\ \lambda_{k=2} & \text{if} \quad \text{정전이연쇄} \end{cases} \quad (5)$$

식 (4)과 식 (5)를 합쳐, 식 (3)으로 나타낼 수 있고, 식 (3)을 최대화 하는 방식으로 (Θ, Λ)의 파라미터들을 추정하게 된다. 추정된 (Θ, Λ)의 파라미터를 이용하여, k 번째 사람의 반응 패턴에 대하여 개인별 우도를 계산할 수 있다. 만약, k 번째 사람이 비정상적이라면 정상적으로 응답한 사람들에 비해 $\psi_i(X_i)$와$\psi_{i,j}(X_i, X_j)$에서 낮은 파라미터값에 해당하는 반응을 선택할 가능성이 높다. 따라서, 비정상적인 사람의 개인별 우도는 낮을 것이다.

4. 실험 결과 및 분석

4.1. 설문 자료

정신과적 진단분류를 위한 측정으로 비정상적인 행동과 증상을 진단하는 목적인 MMPI-2(Minnesota Multiphasic Personality Inventory)의 한국판 자료로서, [7]이 수집한 자료를 재분석하였다. 이 자료는 서울대학교 학부생 225명을 대상으로, 설문지를 주고 정상적으로 설문지에 반응하게 한 뒤, 1주일 뒤에 설문지 없이 답안지만 제시하고, 마음대로 응답을 하도록 하였다. 정상적으로 설문지에 반응한 사람들을 정상 응답자로 정의하고, 마음대로 응답한 사람들을 비정상 응답자로 정의한다.

MMPI-2 설문지는 총 567문항으로 구성되어있으며, 주로 우울, 반사회성 등 여러 임상척도들이 존재하며, 특이점은 비정상적인 응답 패턴을 탐지하기 위해 개발된 무선반응 비일관성 척도(Variable Response Inconsistency, VRIN)와 고정반응 비일관성 척도(True Response Inconsistency, TRIN)가 존재한다. 2 종류의 왜곡 지표는 총 111개의 문항들로서, 설문지 군데군데에 섞여 있다. MMPI-2 설문지는 "YES" 혹은 "NO"의 이분(binary) 응답으로 반응하도록 구성되어 있다.

본 논문에서는 앞서 소개한 IRT 모형과 MRFs 모형, 그리고 SVM 모형의 분류 정확도를 ROC curve의 곡선하면적(Area Under the Curve, AUC)으로 서로 비교하고, 나아가 모의 실험을 통해 각 처치에 따라 세 모형의 분류 정확도가 어떻게 달라지는지를 비교한다. 모든 분석은 training set을 2/3로, test set을 1/3로 나누어 분석하였고, SVM의 varinace와 cost는 1로 고정하였다.

4.2. 분석 결과

4.2.1 VRIN, TRIN, IRT, MRFs, SVM 비교

아래 표 1에서 비정상 비율은 (비정상 응답자 수) / (전체 사람 수)를 말한다. 비정상 비율을 25%, 20%, 15%, 10%, 5%로 변경해가면서, 2개의 왜곡지표(VRIN, TRIN)와 세 모형(IRT, MRFs, SVM)에 대하여 분류의 변별력을 나타내는 ROC curve의 AUC가 제시되어 있다.

표 1. VRIN, TRIN, IRT, MRFs, SVM의 평균 AUC(총 567문항), (괄호 안은 표준편차, 총 10회 시행)

비정상 비율	VRIN	TRIN	왜곡 지표	IRT 2PL	MRFs	SVM
25%	.976 (.02)	.699 (.06)	포함	.861 (.012)	.996 (.002)	.998 (.001)
			제거	.819 (.011)	.996 (.002)	.999 (.002)
20%	.973 (.05)	.663 (.08)	포함	.901 (.005)	.998 (.001)	1 (0)
			제거	.858 (.013)	.996 (.003)	.998 (.002)
15%	.982 (.02)	.671 (.25)	포함	.961 (.003)	.998 (.002)	.999 (.002)
			제거	.944 (.003)	1 (0)	.999 (.002)
10%	.973 (.06)	.692 (.15)	포함	.974 (.011)	.998	1 (0)
			제거	.901 (.012)	1 (0)	.999
5%	.956 (.09)	.786 (.19)	포함	.994 (.010)	1 (0)	1 (0)
			제거	1 (0)	1 (0)	1 (0)

표 1 을 보면, IRT의 경우 비정상 비율이 많아질수록 AUC가 감소하고 있다. 또한, MRFs와 SVM은 설문지에서 측정할 수 있는 왜곡 지표(VRIN, TRIN) 보다 더 우수하

였다. 또한, 왜곡 지표(VRIN, TRIN)를 제거하였을 경우, MRFs와 SVM의 AUC는 감소하지 않으나 IRT의 경우 AUC가 감소하였다. 567문항을 모두 분석했을 경우 비정상 비율이 5% 미만이 되면 IRT, MRFs, SVM 3개의 모형 모두 정확률이 높다. 그러나, 문항이 적더라도 효과적으로 무선반응을 탐지할 수 있어야 하므로, VRIN과 TRIN을 제외한 456문항 중에서 문항의 수를 30개로 줄여서 재분석하였다.

표 2. IRT, MRFs, SVM의 평균 AUC (총 30문항)
(괄호안은 표준편차, 총 10회 시행)

비정상 비율	AUC			Balanced ACC			
	IRT	MRFs	SVM	IRT	MRF(1)	MRF(2)	SVM
5%	.934 (.06)	.988 (.01)	.993 (0.1)	.907 (.08)	.938 (.05)	.942 (.04)	.929 (.06)
4%	.962 (.05)	.982 (.02)	.983 (.03)	.858 (.12)	.896 (.10)	.887 (.09)	.904 (.11)
3%	.972 (.03)	.983 (.02)	.987 (.02)	.804 (.12)	.879 (.11)	.877 (.10)	.762 (.07)
2%	.985 (.01)	.995 (.01)	.987 (.02)	.697 (.08)	.732 (.10)	.733 (.09)	.730 (.03)

표 2는 비정상 비율이 5%미만에 대하여 분석한 결과이다. 연속적인 30개의 문항을 총 10회의 시행마다 서로 다른 set을 추출하였다. 표 2를 보면, 적은 문항에서도 MRFs와 SVM의 성능이 IRT보다 더 좋았다. 또한, SVM의 확률, IRT의 표준화된 개인우도(lz), MRF(1)은 로그 우도를 표준화하여 -2보다 낮은 값으로, MRF(2)는 모든 문항에 대한 무선 반응의 로그 우도($I \times log(.5)$), I = 문항수) 이하를 비정상 반응패턴으로 분류하였다. 연구 결과, Balanced ACC((sensitivity + specificity)/2) 를 보면 AUC와 동일한 결과를 보인다.

일반적으로 한 개인이 문항 전체에 대하여 모두 비정상적인 반응을 보일 가능성이 낮으므로, 정상 반응 중에서 문항 일부를 선택하여 비정상적인 반응으로 만드는 두 조건의 조작 실험을 진행하였다. 첫째 조건은 왜곡지표를 제외한 문항들 중에서 연속적인 30문항을 샘플링하고, 조작할 문항 비율을 달리하면서, 개인마다 문항들을 무선적으로 선택하여 무선적으로 반응을 조작하였고, 둘째 조건은 조작할 문항 비율을 달리하면서, 개인마다 문항들을 연속선택하여, 연속적으로 반응으로 조작하였다. 즉, 첫째 조건은 개인 그리고 문항마다 독립적으로 1/2의 확률로"Yes" 혹은 "No"로 조작하였고, 둘째 조건은 개인마다 1/2의 확률로 연속적으로 "Yes" 혹은 "No"로 조작하였다. 두 조건 모두 총 1000회 반복하였다. 분석 결과, 그림 2 의 경우 SVM 모형이 IRT와 MRFs 비해 정확도가 근소하게 낮고, 그림 3의 경우 IRT 모형이 MRFs와 SVM 모형보다 정확도가 낮다.

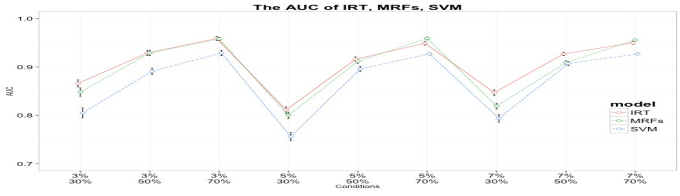

그림 2. 무선 문항 및 무선 반응 조작 시 평균 AUC (95% 신뢰구간)

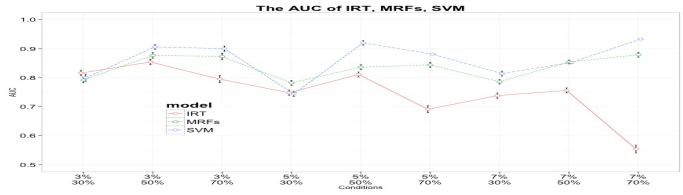

그림 3. 무선 문항 및 무선 반응 조작 시 평균 AUC (95% 신뢰구간)

5. 결론 및 향후 연구

MRFs는 왜곡지표(TRIN,VRIN) 보다 더 높은 정확도를 보이고, 여러 조건들에서 SVM 혹은 IRT보다 안정적인 정확도를 보인다.

본 연구에서 이용한 컷 오프 기준은 (1) 표준화된 로그우도(-2 이하), (2) 전체 문항에 대한 무선 반응의 로그우도이다. (2) 기준의 경우 경우, 부분적으로 무선 응답을 할 경우 기준으로서 적절하지 못하기 때문에 (1)과 (2) 두 가지 기준을 동시에 고려할 필요가 있다. 또한, 본 연구에서는 이산적인(discrete) 변수만을 다루고 있기 때문에, 응답의 거리 등을 활용하여 연속적인 변수로 확장할 필요성이 제기된다.

그럼에도 불구하고, 본 연구에서 제안하는 모형은 다양한 방면에 활용이 가능하다. 첫째, 정상적인 반응 패턴을 학습한 후, 사람들의 반응들을 생성할 수 있다. 둘째, 결측치가 있을 경우 모형을 통해 결측치를 추정하는 데 활용할 수 있다. 셋째, 집단을 가장 잘 대표하는 반응을 추출할 수 있다. 넷째, 비정상적인 정보가 많은 구간을 탐지할 수 있다. 다섯째, 본 연구 모형은 비단 설문지뿐만 아니라, 비정상적인 신호의 흐름 및 기상 상태 등을 탐지하는 곳에 응용할 수 있다.

참고문헌

[1] S. P, Reise & N. G, Waller, "Traitedness andthe assessment of response pattern scalability." *Journal of Personality and Social Psychology. 65. pp.* 143-151. 1993.

[2] M. Levine & D. B. Rubin, "Measuring the appropriateness of multiple choice test scores." *Journal of Educational Statistics. 4. pp.* 269-290. 1979.

[3] F. Drasgow, M. Levine, & E. Williams,"Appropriateness measurement with polytomous item response models and standardized indices." *British Journal of Mathematical and Statistical Psychology. 38. pp.* 67-86. 1985.

[4] P. Ferrando., & E. Chico, "Detecting dissimulation in personality test scores: A comparison between person-fit indices and detection scales." *Educational and Psychological Measurement.* 61(6). *pp.* 997-1012. 2001.

[5] C. Bishop, "*Pattern Recognition and Machine Learning.*" Springer. 2006.

[6] K. Murphy. "*Machine Learning: A Probabilistic Perspective.*" MIT Press. 2012.

[7] 장은경. "반응자 적합도와 마코프 체인을 이용한 한국판 MMPI-2의 반응왜곡 탐지." *서울대학교 석사학위논문.* 2010.