Optical Flow 기반의 Saccade 탐지를 통한 전정기관 이상 검출과 Dowhy 기반의 연관 관계의 신뢰도 검정

지명진*, 김태현, 김성환
서울시립대학교 과학기술대학원 컴퓨터과학부
wlaudwls2326@uos.ac.kr*, kdtth@daum.net, swkim7@uos.ac.kr

Assurance of HIT (head impulse test, Saccade based Vestibular Anomaly Detection) using Confidence Interval of Optical Flow Comparison on Wasserstein Metric

Myeongjin Ji*, Tae-Hyun Kim, Seong-Whan Kim
*Dept. of Computer Science, University of Seoul

요 약

최근의 기계 학습 (딥러닝)은 기존의 전통적인 통계 분석 방법들에 비해 효율성과 정확도가 높은 장점이 있지만, 처리과정이 블랙박스와 같아 결과 값의 중요한 원인 또는 근거 요인을 찾기 어렵다는 단점을 가지고 있다. 이를 해결하기 위한 최근의 XAI (eXplainable AI) 연구를 기반으로 하여, 본 논문에서는 의료기관에서 전정기관의 이상을 판별하기 위해 수작업으로 이루어지고 있는 HIT (head impulse test) 테스트 결과를 자동화하고, 설득력 있는 신뢰도 검정을 위해, XAI 기반 DoWhy 프레임 워크를 사용하였다. 전정기관 이상으로 의심되는 환자의 동공 움직임을 optical flow 로 추적하고, 정상인과의 Wasserstein metric 의 DoWhy 검증을 통해 전정기관 이상 여부의 신뢰도 구간을 검정한다.

1. 서론

일반적으로 응급실 진료를 받게 되는 어지럼증 환자들의 증상은 매우 다양하지만, 보통 어지럼증을 발생시키는 원인은 전정기관의 이상으로 발생한다. 어지럼증 검진 방법을 위한 대표적인 방법으로는, 머리 움직임 중 발생하는 Saccade 확인을 통해 검진하는 방법이 가능하다 [1]. 환자의 전정기관 여부를 관찰하기 위해, HIT (head impulse test) 를 실행하게 되며, 의사가 환자의 머리를 회전시키고, 머리가 움직이는 경우와 고정시킨 경우 각각에 대한 Saccade 발생을 통해 진단한다 [2]. 본 논문에서는 Saccade 검출을 위한 OF (optical flow)를 검출하고, OF 의 확률분포 (probability distribution function, PDF) 비교를 위한 Wasserstein 거리 기반의 추정치를 통해, DoWhy 신뢰성 검정을 수행함으로써, 의료 분야에 대한 XAI (explainable AI) 사례를 소개한다. 구체적으로, 머리 움직임이 멈춘 후 100ms 이내에 발생하는 동공의 움직임을 관찰하여, 100ms 동안의 OF 를 기록하고, OF 의 PDF 를 정상인의 PDF 와 비교하여, 관측된 동공의 움직임이 전정 기관 이상을 확인할 수 있는 Saccade 인지를 DoWhy 신뢰성 검정을 통해 설명하게 된다. 이를 통해 Saccade 가 전정기관 이상 검출 방법인 HIT 의 신뢰도를 밝혀내는 것을 목표로 한다.

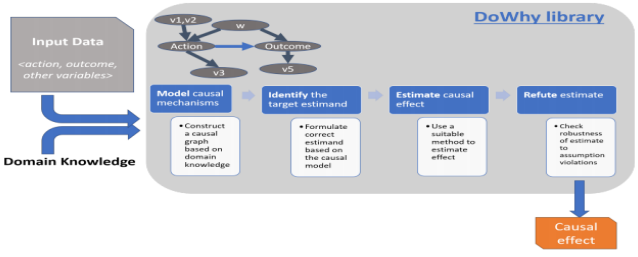
2. 관련연구

본 논문에서는, 영상 내의 움직임 패턴을 측정하기 위해, optical flow (OF) 기법을 사용하였다. OF 는 이전 프레임과 다음 프레임 간의 픽셀이 이동한 방향과 거리를 말하며, OF 를 통해, 영상 내의 물체가 어느 방향으로 얼마만큼 움직였는지 파악할 수 있다. OF 는 연속된 Frame 사이에서 움직이는 물체의 픽셀의 Intensity 는 변함이 없다는 것과 서로 이웃하는 픽셀은 비슷한 움직임을 갖는다는 2 가지 가정을 가지고 계산한다 [3]. OF 측정 방법에는 전체 픽셀을 모두 계산하는 Dense OF 와 일부 픽셀 만을 계산하는 Sparse OF 두 종류가 있다. Sparse OF 보다 밀집된 전체 픽셀에 대하여 계산하는 Dense OF 가 계산 속도는 느리지만 정확도가 증가하기 때문에 Dense OF 를 이용하여 OF 의 측정과 계산을 진행했다.

본 논문에서 OF 의 PDF 간의 거리를 측정하기위해, Wasserstein Distance 를 사용하였다. Wasserstein Distance 는 주어진 Metric 공간에서 확률 분포 X 와 Y

사이의 정의된 거리 함수를 말한다 [4]. 두 확률 분포가 있다면 두 확률분포 X, Y 의 결합확률분포 (joint probability distribution, joint PDF, 확률 변수가 여러 개일 때 이들을 함께 고려하는 확률 분포)가 있는데 이 모든 결합 확률 분포에 대해서 어떤 한 샘플링 한 값에서의 Distance(X, Y) 의 기대 값을 가장 작게 추정한 값을 의미한다. Distance(X, Y)의 기대 값을 가장 작게 추정한다는 의미는 두 확률 분포에 대해서 어떤 한 값인 w 를 Sampling 하여 X(w)와 Y(w)를 추출하고, 이 샘플링 된 2 개 점 간의 거리를 d(X(w), Y(w))로 계산할 수 있고, 샘플링을 계속 할수록 (X, Y)의 결합 확률 분포의 윤곽이 나오게 되는 것을 말한다. 더불어서 두 확률 분포는 결합 확률 분포의 주변 확률 분포 (marginal PDF, 부분 집합에 포함된 변수의 확률 분포)가 된다 [5]. 결합 확률 분포가 두 확률 변수 X, Y 의 연관성 (dependency)을 어떻게 측정하느냐 에 따라 Distance(X, Y)의 분포가 달라지며, Wasserstein Distance 는 이렇게 여러 가지 결합 확률 분포에서 Distance(X, Y)의 기대 값이 가장 작게 나오는 확률 분포를 취하게 된다.

본 논문에서는 HIT 시험을 통한, 정상인과 전정기관 이상이 의심되는 환자의 동공 움직임 관찰로부터 얻어진 데이터로부터 전정기관 이상을 확신할 수 있는 여부를 검정하기 위해, Microsoft 사의 Python 오픈소스 라이브러리인 Dowhy 를 사용한다. Dowhy 는 인과추론 (causality inference)을 위한 강력한 통계를 제공하는 작업에 초점을 맞춘 라이브러리이다. [6]. 그림 1 은 인지 추론 라이브러리인 Dowhy 의 Model, Identify, Estimate, Refute 하는 과정을 나타낸다. Model 단계에서 데이터를 인과 그래프로 인코딩하고 원인 및 질문을 모델링하고, Identify 단계에서는 Model 의 인과 관계를 식별하고 원인을 추정한다. Estimate 단계에서는 식별된 인과 관계에 대하여 추정치를 추정하며, Refute 단계에서는 얻어진 추정치에 대하여 반박을 시도한다.



(그림 1) Dowhy 의 과정. [4]

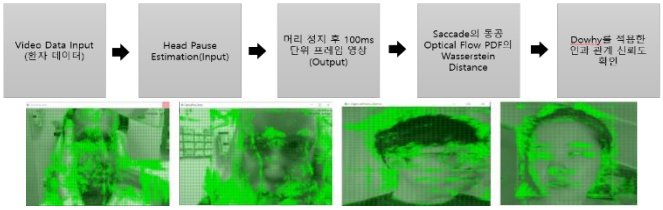
(1) Model 단계에서는 데이터를 인과 그래프로 인코딩하고 원인 및 질문을 모델링한다. 이 단계에서는 각각에 대한 기본 인과 관계 그래픽 모델을 생성하게 된다. 이것은 인과적 가정을 명백하게 하는 역할을 한다 [6]. (2) Identify 단계에서는 그래프 기반 방법을 사용하여 인과 관계를 식별하고 원인을 추정한다. 원인이 되는 그래프를 기반으로 가능한 모든 식별 방법을 Identify 단계에서 찾게 된다. Identify 에서는 Confounder 값이 있는데, Confounder 는 만약 X -> Y 라는 원인과 결과가 있을 경우 Y 의 원인이 X 라고 하고 싶은데 X 와 Y 에 동시에 영향을 주는 제 3 의 변수가 존재한다고 가정하면 Confounding Effect 라는 것이 발생하게 되고 여기서 영향을 주는 변수를 Confounder 라고 한다. Confounding Effect 는 두 개 이상의 변수가 결과에 영향을 끼치는데 각 변수가 얼마만큼 영향을 주는지 구분할 수 없는 효과를 말한다. Confounder 는 식별이 가능한 Observed Confounder 와 식별이 불가능한 Unobserved Confounder 로 나눈다. Identify 는 Estimand 1 (Back-Door), Estimand 2 (Front-Door), Estimand 3 (Instrumental-Variables)로 구분된다. Estimand 는 통계 분석에서 추론의 대상이 되는 추정값을 나타내는 용어이며, Identify 는 일반적으로 Estimand 1 방식을 사용하게 되는데 Back-Door 방식은 어떠한 원인 X 에 대한 결과 Y 가 측정이 가능한 일반적인 원인들이 있을 경우 Conditioning 을 통하여 인과 관계를 식별하는 방식을 의미한다. Estimand 2 인 Front-Door 방법은 Two-stage Linear Regression 을 통한 식별을 한다. 여기서 Two stage 는 계량 경제학의 HeckMan 의 Two-Stage 추정 기법을 의미한다. Estimand 3 인 Instrumental-Variables 방법은 작업 원인과 결과가 관찰되지 않는 경우에도 결과를 예측하도록 한 방법이다. IV 는 Instrument 가 action directly 에만 영향을 미친다는 사실을 활용한다. IV 는 랜덤인 경우와 Exclusion 한 경우 2 가지에 대해서 가정을 하여 진행한다 [6]. (3) Estimate 단계에서는 식별된 추정치를 추정하는데 통계적인 방법을 사용하며 인과 관계를 추정하는데 여기서는 주로 Back Door 기준을 이용하여 지원한다. Estimate 에는 Identified Estimand, Realized Estimand, Estimand Mean Value 값 형식으로 출력되게 되는데 Identified Estimand 는 Estimand Type 을 정의하고 추정치를 식별하고, Realized Estimand 는 실제로 실현된 추정이며, Average Treatment Effect(ATE)의 단위로 분석된 결과를 나타낸다. 여기서 Average Treatment Effect(ATE)는 무작위 실험이나 의료 시험 등에서 Treatment Variable 을 비교하는데 사용되는 측정값을 의미하고 Treatment Variable 은 종속 변수에 미치는 영향을 연구하는 독립변수를 뜻한다. ATE 는 치료에 할당된 단위와 컨트롤에 할당된 단위 사이의 평균 결과의 차이를 측정한다. 무작위 실험 연구에서 ATE 는 Treated unit 과 Untreated Unit 의 비교를 통하여 표본으로부터 추정할 수 있다. 일반적으로 추정 절차에 대

하여 알고자 하는 Causal Parameter 로 인식된다. 마지막으로 Estimand Mean Value 에서 Estimate 에 대한 평균값이 나타난다. [6]

(4) Refute 단계에서는 초기 모델의 가정에 대한 견고성을 테스트하여 얻은 추정치를 반박하는 시도를 하며, 얻어진 추정치를 반박한다. Refute 의 방법은 그림 5 와 같이 총 4 가지가 있으며, Random Common Cause, Unobserved Common Cause, Placebo Treatment, Subset Of Data 로 구성되어 있다. Random Common Cause 는 랜덤 공통 원인 변수를 추가하여 Refute 를 진행하고, Unobserved Common Cause 는 관측되지 않은 공통 원인 변수를 추가하여 Refute 를 진행하며 Placebo Treatment 는 랜덤 변수로 Treatment 를 덮어쓰고 변경하는 과정을 통해 Refute 를 하고, 마지막으로 Subset a Data 는 데이터의 임의의 부분 집합을 제거하여 Refute 를 한다.

3. 동공 움직임에 따른 전정기관 이상 분석 기법

그림 2 는 전체 시스템의 구조이다. 어지럼증을 호소하거나 전정기관에 이상이 있을 만한 의심이 있는 2 명 환자의 관찰영상 (환자 2 명의 fps 23.9, 24, 전체 frame Count 1733, 477)을 입력 받은 후, 환자의 동공을 추적하여 OF 를 계산하고, 머리가 흔들린 후 멈췄을 때 환자의 동공에서 발생되는 Saccade 를 측정한다.

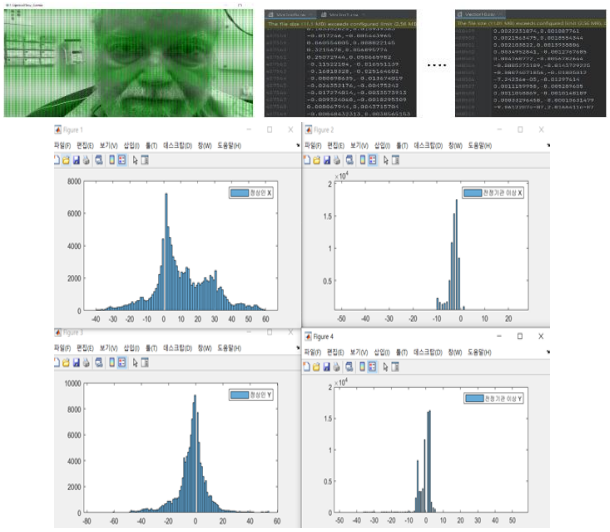


(그림 2) 전체 시스템구조와 실험영상(환자 2 명/정상 2 명)

그림 3 과 같이 왼쪽과 오른쪽으로 머리를 환자의 머리를 회전시키고, 머리의 움직임이 멈춘 직후의 100ms 영상을 저장한 후 저장한 영상에 대하여 Saccade 를 측정하게 된다. 머리가 멈출 때의 조건은 만약 머리가 얼굴의 센터를 벗어나서 왼쪽이나 오른쪽 범위 안에 들 경우 측정한다. 100ms 프레임에 대해 Optical Flow 를 계산한다.

(그림 3) 머리가 멈춘 이후 100ms 구간의 프레임 탐지.

그림 4 는 정상인의 동공움직임을 측정한 OF 와 전전기관 이상으로 의심되는 환자의 OF 값과, OF 값들의 dX 값과, dY 값의 PDF 를 보이고 있다. PDF 는 MATLAB plot Chart 의 histogram 을 사용하였다.



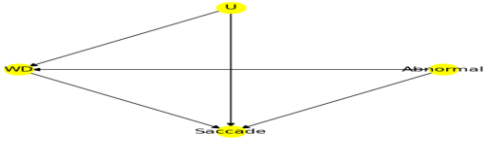
(그림 4) 정상인과 전정기관 이상 환자 OF 값의 PDF.

표 1 은 정상인의 OF 값의 PDF 와 전정기관 이상인 환자의 OF 값의 PDF 사이의 Wasserstein Distance 를 구한 결과이다. Wasserstein Distance 를 구하는 함수는 Python 라이브러리에서 Scipy 라이브러리에 정의되어 있으며, 정상인의 Vector (x, y)와 전정 기관의 이상이 있는 환자의 Vector (x, y)의 값을 Wasserstein Distance(정상인 X, 전정 기관 이상 X)와 Wasserstein Distance(정상인 Y, 전정 기관 이상 Y)로 함수를 구성하여 각각의 X 와 Y 에 대한 Wasserstein Distance 를 구한다.

(표 1) Wasserstein Distance.

Wasserstein Distance 1	15.473658590062719
Wasserstein Distance 2	5.331586162159469
…	
Wasserstein Distance 19	17.179120630323652
Wasserstein Distance 20	7.162499656696441

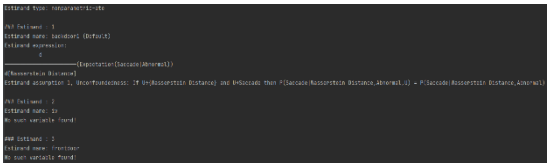
DoWhy 적용을 위한 첫번째 단계로, Model 을 구성하였다. 전정기관의 이상 유무, Saccade 의 유무, Wasserstein Distance 간의 연관성 추론을 위해, DoWhy 의 Casual Model 의 인과 관계 그래프를 다음 그림 5 와 같이 구성하였다.



(그림 5) Dowhy 의 Model 단계

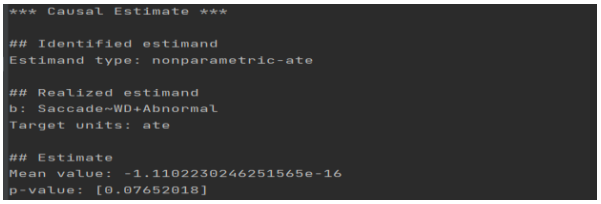
Causal Model 의 인과 관계 구성요소로, Wasserstein Distance 와 Saccade 유무, 전정기관의 이상유무를 입력한다. Model 단계에서는 전정기관의 이상 유무와 Saccade 의 유무 그리고 Wasserstein Distance 의 값들을 입력 받아서 인과 모델 그래프를 그렸다.

DoWhy 적용을 위한 두번째 Identify 단계를 구성한다. 본 논문에서는 원인과 결과에 대한 측정 가능한 일반적인 원인인 Wasserstein Distance 가 있기 때문에, Identify Estimand 1 인 Back-Door 방식을 사용하여 측정한다. Identify 단계에서는 Model 에서 그려진 인과 모델 그래프로 인과 관계를 추정한다. 그리고 Model 의 Estimand 를 결정하는데, Back-door 방식을 이용하였기 때문에 Estimand 1 로 결정하였다.



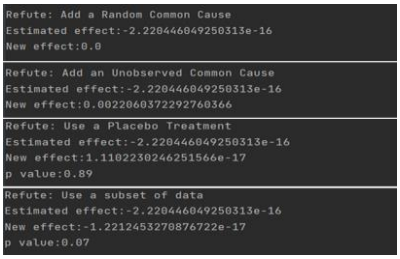
(그림 6) Dowhy 의 Identify 단계

DoWhy 적용을 위한 세번째 Estimate 단계를 구성한다. Estimate 단계에서의 Estimate Mean 결과 값은 -1.1 정도가 나왔고 Estimate 의 p-value 값은 0.07 로 0.1 보다 작게 추정 값이 나왔기 때문에 null hypothesis 가 기각이 될 수 있는 유효한 값을 얻었다.



(그림 7) Dowhy 의 Estimate 단계

DoWhy 적용을 위한 네번째 Refute 단계를 구성한다. Refute 단계는 Estimate 에서의 결과를 4 가지 방법들을 사용하여 추정치를 반박하는데 Placebo Treatment 에서는 p-value 가 0.89 로 null hypothesis 가 기각 될 수 있는 수치보다는 큰 p-value 가 나왔고, 데이터의 임의의 부분 집합을 제거하여 진행하는 Subset Of Data 에서 Refute 를 진행한 결과에서는 p-value 값이 0.07 로 null hypothesis 를 기각 할 수 있는 유효한 값을 얻을 수 있었다.



(그림 8) Dowhy 의 Refute 단계

Dowhy 4 단계를 적용한 결과로 얻어진 Saccade 와 전정기관 이상 간의 연관 관계 신뢰도 값으로 도출된 p-value 결과 값이 0 으로 수렴할수록 Saccade 와 전정기관 이상 간의 연관 관계의 신뢰성은 증가한다고 볼 수 있다. 여기서 p-value 는 null hypothesis 가 올바른 값이라는 전제 하에, 표본에서 실제로 관측된 통계 치와 같거나 더 극단적인 통계치가 관측될 확률이다. [8] p-value 값은 관찰된 데이터가 null hypothesis 와 양립하는 정도를 0 에서 1 사이의 수치로 표현한 것이다. p-value 값이 작을수록 그 정도가 약하다고 보며, 특정 값 보다 작을 경우 null hypothesis 를 기각하고, p-value 값이 클수록 null hypothesis 를 채택한다. 따라서 p-value 값의 결과를 Saccade 와 전정기관 이상 사이의 신뢰성검정의 척도로 볼 수 있다.

4. 결론

머리가 움직였다 멈추는 경우에 순간적으로 (100ms) 발생한 영상에 대하여 OF 방법을 이용하여 Saccade 여부를 탐지하였다. Saccade 발생 여부를 통해, 전정기관 이상 여부를 판단할 수 있다. 본 논문에서는 설명이 가능한 AI 인 XAI 가 적용될 수 있도록 인과 관계의 설명이 가능한 인지 추론 라이브러리인 Dowhy 를 사용하여 Saccade 와 전정기관 이상 간의 연관 관계를 파악하고자 하였으며, 둘 간의 신뢰도를 분석해보고자 하였다.

참고문헌

[1] https://blog.naver.com/dizzy_master/221941152198
[2] https://jamanetwork.com/journals/jamaotolaryngology/article-abstract/2773737?utm_campaign=video&utm_medium=social&utm_source=youtube&utm_content=public.
[3] http://opencv-python-tutroals.readthedocs.io/en/latest/index.html.
[4] Martin Arjovsky, Soumith Chintala, Léon Bottou. (January 2017). "Wasserstein GAN" (PDF).
[5] https://dawoum.ddns.net/wiki/Marginal_distribution
[6] Amit Sharma, Emre Kiciman, et al. DoWhy: A Python package for causal inference. 2019. https://github.com/microsoft/dowhy
[7] Jones, E., Oliphant, T. E., Peterson, P. et al. SciPy: Open source scientific tools for Python (2001−). URL http: //www.scipy.org/.
[8] https://ko.wikipedia.org/wiki/%EA%B7%80%EB%AC%B4_%EA%B0%80%EC%84%A4