



Deep learning based Feature Selection Study

Sanghyuck Lee

2021. 08 24.

Outline

- Instance-Wise Causal Feature Selection
- 중견 연구 과제
- 갑상선 연구 과제
- 산학 협력 과제

Papers for Instance-Wise Feature Selection

Yoon, Jinsung, James Jordon, and Mihaela van der Schaar. "INVASE: Instance-wise variable selection using neural networks." *International Conference on Learning Representations*. 2018.

=> ICLR

Tonekaboni, Sana, et al. "What went wrong and when? Instance-wise feature importance for time-series black-box models." *Advances in Neural Information Processing Systems* 33 (2020).

=> NeurIPS

Panda, Pranoy, Sai Srinivas Kancheti, and Vineeth N. Balasubramanian. "Instance-wise Causal Feature Selection for Model Interpretation." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2021.

=> CVPR

Liyanage, Yasitha Warahena, and Daphney-Stavroula Zois. "Optimum Feature Ordering for Dynamic Instance-Wise Joint Feature Selection and Classification." *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2021.

=> ICASSP

Paper Review

Paper Title: **Instance-wise Causal Feature Selection for Model Interpretation**

Authors: Panda, Pranoy, Sai Srinivas Kancheti, and Vineeth N. Balasubramanian.

Conference: CVPR 2021

Abstract

We formulate a **causal extension** to the **recently** introduced **paradigm of instance-wise feature selection** to **explain black-box** visual classifiers. Our method selects a subset of input features that has the **greatest causal effect** on the model's output. We quantify the causal influence of a subset of features by the Relative Entropy Distance measure. Under certain assumptions this is equivalent to the conditional mutual information between the selected subset and the output variable. The resulting causal selections are sparser and cover salient objects in the scene. We show the efficacy of our approach on multiple vision datasets by **measuring the post-hoc accuracy and Average Causal Effect** of selected features on the model's output.

Introduction of the Paper

Vision Black box model and Interpretability

=> Instance-wise Feature Selection -> Select superpixels or patches

Existing Works: L2X and INVASE

-> have optimization functions which try to maximize mutual information, but mutual information does not always capture causal strength.

"For an explanation to be correct, the selected features should be causally consistent with the model being explained."

"We hypothesize that the **most sparse and class discriminative features** are indeed the most causal features, and they form **good visual explanations**."

Experiments of the Paper

- Use MNIST and CIFAR
- Each Comparison Methods Select patches of images

- Post-hoc accuracy:

$$\frac{1}{|X_{val}|} \sum_{x \in X_{val}} (\operatorname{argmax}(P(y|x)) == \operatorname{argmax}(P(y|x_s)))$$

- Average Causal Effect (ACE):

Average of Individual Causal Effect (ICE)

$$ICE = P(y|x_s) - P(y|x_{random})$$

Methodology

A black box classifier $F : \mathbb{R}^d \rightarrow Y$

A explainer/selector network $E_k : \mathbb{R}^d \rightarrow S_k$, where $S_k = \{e | e \in \{0, 1\}^d, |e| = k\}$

E_k chooses a subset of features (size of subset is fixed as k) for each input that best explains the predictions made by F .

Methodology

Causal Model:

$X \rightarrow Y$

Causal Metric:

Relative Entropy Distance (RED)

is an information theoretic causal strength measure, which is based on performing interventions on the causal graph and exposing the target of the cut/intervened edges with product of marginal distributions.

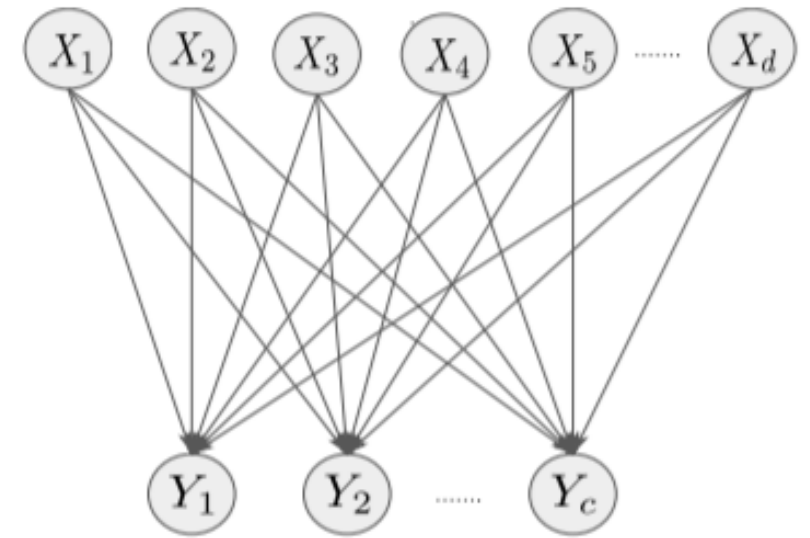


Figure 1. Black-box model as a SCM

intervention 미국·영국 [ɪntəˈvɛnʃən] 영국식 ★ +

1. 명사 사이에 둠, 개재; 조정, 중재
2. 명사 (타국의 내정 등에 대한) 개입, (내정) 간섭; [완곡] 무력 간섭
3. 명사 법 (제3자의) 소송 참가

Methodology

RED's useful properties:

- It can capture the non-linear, complex relationship between input and output.
- The causal strength of an edge $X_i \rightarrow Y_j$ depends on X_i and the other parents of Y_j and on the joint distribution of parents of Y_j . This locality property ensures that we do not have to take into account the causes of X_i when computing causal influence of X_i on Y_j , unless the causes are also immediate causes of Y_j .

⇒ RED is a authors' choice.

Methodology

Lemma 1.1

The causal strength of a subset of links s going from input X to the response variable of the model Y , denoted as CS_s , is given as follows:

$$CS_s = I(X_s; Y|X_{\bar{s}}) \quad (1)$$

Here, X_s denotes the features in set s , $X_{\bar{s}}$ denotes the features not in set s , and $X = X_s \cup X_{\bar{s}}$.

$$\Rightarrow CS_s = I(X_s; Y|X_{\bar{s}}) = -H(Y|X) + H(Y|X_{\bar{s}})$$

$$\Rightarrow \max_s CS_s \equiv \min_s E_{Y, X_{\bar{s}}} [\log(P(Y|X_{\bar{s}}))]$$

where $P(Y|X_{\bar{s}})$ is the explainer's output distribution

Methodology

Two main issues:

1) Approximating $P(Y|X_{\bar{S}})$

to estimate $P(Y|X_{\bar{S}})$, simply use the output of the F ,

$X_{\bar{S}}$: if $s_i=0$, $X_{\bar{S}i} = X_i$, else $X_{\bar{S}i} = 0$.

2) Continuous subset sampling

Use Gumbel Softmax trick:

Define a function g which maps each input feature X_i to a value which indicates its probability of being part of $\bar{S}$.

Learn g via a neural network parameterized by θ .

Estimate $\bar{S}$ by a function $Z(\theta, \zeta)$, where ζ is Gumbel random variables.

Methodology

Final Optimization Function:

$$\begin{aligned}\max_s CS_s &\equiv \min_s E_{Y, X_{\bar{s}}} [\log(P(Y|X_{\bar{s}}))] \\ &= \min_{\theta} E_{X, Y, \zeta} [\log(F(Z(\theta, \zeta) \odot X))] \\ &= \min_{\theta} E_{X, Y, \zeta} \left[\sum_{y=1}^c P(y|X) \log(F(Z(\theta, \zeta) \odot X)) \right]\end{aligned}$$

Experiments of the Paper

- Use MNIST and CIFAR
- Each Comparison Methods Select patches of images

- Post-hoc accuracy:

$$\frac{1}{|X_{val}|} \sum_{x \in X_{val}} (\operatorname{argmax}(P(y|x)) == \operatorname{argmax}(P(y|x_s)))$$

- Average Causal Effect (ACE):

Average of Individual Causal Effect (ICE)

$$ICE = P(y|x_s) - P(y|x_{random})$$

Papers for Instance-Wise Feature Selection

Yoon, Jinsung, James Jordon, and Mihaela van der Schaar. "INVASE: Instance-wise variable selection using neural networks." *International Conference on Learning Representations*. 2018.

=> ICLR

Tonekaboni, Sana, et al. "What went wrong and when? Instance-wise feature importance for time-series black-box models." *Advances in Neural Information Processing Systems* 33 (2020).

=> NeurIPS

Panda, Pranoy, Sai Srinivas Kancheti, and Vineeth N. Balasubramanian. "Instance-wise Causal Feature Selection for Model Interpretation." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2021.

=> CVPR

Liyanage, Yasitha Warahena, and Daphney-Stavroula Zois. "Optimum Feature Ordering for Dynamic Instance-Wise Joint Feature Selection and Classification." *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2021.

=> ICASSP

중견 연구 제안서 작성

- 국내 의료 AI 관련 정책 과제 조사

- 보건복지부가 주관하여 2025년까지 향후 5년간 총 945억원 사용해, 의료데이터셋과 의료 인공지능 기술 개발을 진행함. => 총 3가지 사업으로 요약됨.

1) 중환자 특화 빅데이터 (Korean Medical Information Mart for Intensive Care, K-MIMIC) 구축 및 AI 기반 임상 의사결정 지원 시스템 (Clinical Decision Support System, CDSS) 개발

- 데이터 정제를 통해 데이터 사용자 수요에 따라 제공 가능한 데이터셋 구성
- 확보된 중환자 데이터를 공유 분석하는 플랫폼 구축
- 중환자 CDSS를 위한 설명 가능한 AI 알고리즘 개발: 임상적 인과관계에 근거하여 중증도 평가, 재입원 가능성 등을 설명할 수 있는 AI 알고리즘 개발
- 임상 활용 검증 및 실증 연구

중견 연구 제안서 작성

2) 디지털 병리 기반의 암 전문 AI 분석 솔루션 개발

: 병리 진단이 필수적인 암에 대하여, AI 병리 분석 솔루션 개발

=> 병리 판독 업무 효율화, 환자 진단 및 치료 효과성 제고

- 병리 데이터, 주석 처리를 비롯한 표준화 처리, 의료 기관 간의 이미지 색상 차이, 스캐너 제조사에 따른 차이 등을 위한 표준화 기술 개발

- 설명 가능한 AI 알고리즘 기반 SW 개발, 임상 검증 및 활용 평가

3) 합성 의료 데이터 생성 및 임상 활용 실증 연구

: 민감 정보인 의료 데이터 정보 보호 및 활용성 제고를 위해, 공개 목적의 합성 의료 데이터 생성 및 임상적 유효성 및 활용성 검증 연구 지원

- 시나리오 기반 고품질 데이터셋 구축: 합성 데이터 생성 모델 개발

- 합성 데이터의 임상적 유효성 및 활용성 실증 연구

중견 연구 제안서 작성

요약 및 의견:

현재 국가적 차원에서 관심 있어 하는 내용은,

- 실생활에서 사용 가능, 설명 가능한 의료 AI 개발 => 진단 효율성 증대 및 의료 서비스 격차를 줄임.

가 하나의 최종 목표로 보이고, 이를 위한 기반 사업으로,

- 데이터셋 표준화 기술 및 공유 플랫폼, 합성 의료 데이터 생성 기술 개발 등을 진행하고자 함을 알 수 있었고,

이와 같은, 국가적 차원의 큰 그림에서, 당장 표준화된 데이터셋을 확보하기는 어려울 수 있으나, 전처리 과정과 표준화 방식, 학습 모델 제안 등을 포함한 우리만의 프레임워크가, 실생활에서 어느 정도 사용 가능한 세그멘테이션 성능을 보여줌으로써, 추후 표준화 기술에 대한 방향성 제시나, 국가 표준 세그멘테이션 모형의 초석을 다지는 데에 이바지함이 우리의 역할로 있을 수 있음.

중견 연구 제안서 작성

- 가제:

한글: 의료 분야내 주요 관심 구조물 세그멘테이션을 위한 사전훈련 최신 딥러닝 모델 개발

영문: Development of Pre-trained Deep Learning Model for Segmentation of Structures of Major Interest in the Medical Field

- 연구 목표:

본 과제에서는 몇몇 주요한 해부학적 구조에 대해서 세계 최고 수준의 성능을 갖는 사전학습된 세그멘테이션 모델을 배포함으로써, 다양한 의료분야 응용에 성공적으로 기여하는 것을 최종 목표로 하며, 이를 위해 다음과 같은 연차별 연구 목표를 설정하여 추진한다.

중견 연구 제안서 작성

1. 연구의 목표 (2페이지)

- 최종 연구 목표
- 연차별 목표 및 개요도
- 연차별 연구 개발 내용, 창의성 및 도전성, 추진 전략, 주요 결과물

2. 연구의 필요성 (2페이지)

- 4~5개로 나누어 총 2페이지

3. 연구자의 연구 수행역량 (1페이지)

- 연구실 스펙 및 노하우 등 자랑거리

중견 연구 제안서 작성

4. 연구의 추진 전략 및 방법 (연차별 1.5페이지)

- 연차별 상세 내용 및 주요 기술 설명 (그림 포함하여, 연차당 3가지 정도)

5. 연구결과의 중요성 (0.5페이지)

- 학문적/기술적 파급 효과
- 경제적/사회적 파급 효과

6. 연구기간 및 연구비 적정성 (0.5페이지)

- 추진일정 등
- 연구비 적정성

중견 연구 제안서 작성

연차	데이터셋 이름	촬영 방식	차원	촬영 범위	타겟	샘플수	포함 근거
1	LIDC IDRI	CT	3D	흉부, 폐	폐 결절	1010	국책사업에 언급된 한국인 호발암 3위, 코로나 시국 관련
2	CAU Orbit	CT	3D	머리, 안와 부근	안와내 구조물	1237	갑상선 안병증 치료 계획에 중요 역할
3	ACRIN 6664	CT	3D	결장 부근	결장	825	국책사업에 언급된 한국인 호발암 2위 (대장암)
4	CBIS-DDSM	MG	2D	유방	유방암	1566	국책사업에 언급된 한국인 호발암 5위
5	Prostate MRI US Biopsy	MR	3D	전립선 부근	전립선	1151	국책사업에 언급된 한국인 호발암 7위

변경점: 1) 담낭용종->유방암, 2) 순서 변경: 코로나로 인한 폐 결절 사업성 강조

중견 연구 제안서 작성

폐 결절 세그멘테이션 (Xiao et al., 2020):

악성 폐 결절이 암이 될 확률이 큼 -> 폐 결절의 정확한 조기 검출과 세그멘테이션은 폐암 조기 진단에 매우 중요함.

뿐만 아니라, 코로나와 관련해서, 폐 결절과 코로나 감염 간의 연관성에 대한 내용을 서술할 수 있음.

=> 2D Unet, 3D Unet 등 다양한 변형이 존재

중견 연구 제안서 작성

안와내 구조물 세그멘테이션 (Zhu et al., 2021):

CT를 이용해 측정된 부피적 변화는 갑상선 안병증의 원인, 진행, 재발 연구에서 중요한 역할을 함

수동 세그멘테이션의 작업량을 줄이고, 재현성을 높이는 자동 세그멘테이션 방식 연구

=>U-Net, V-Net 기반 연구 등

중견 연구 제안서 작성

결장 세그멘테이션 (Gonzalez et al., 2021):

세그멘테이션은 결장과 같은 방사선에 민감한 기관에 중요함.

-> 하지만, 시간 소모가 크며, 시간 압박에 의해 오류가 발생하기 쉬움 -> 치료 계획 품질에 영향을 끼침.

- S자 결장은 3차원 공간에 접힌 튜브와 같은 구조이기 때문에, 매우 복잡한 3차원 형태임,

- 환자마다 장기의 크기, 모양 등에서 큰 차이가 있음.

- 소장, 자궁 등 주변 장기 등과 결장을 구별하기 어려움

=> CNN, U-Net 기반

중견 연구 제안서 작성

유방암 세그멘테이션 (Tsochatzidis et al., 2021):

유방암은 여성에게 자주 진단되며, 치명적인 유형의 암임.

유방암 치료 파이프라인에서, 유방 조영술을 통해 병변 위치를 파악하고, 병변의 경계를 세그멘테이션하고, 악성과 양성 병변으로 구별함.

=> CNN 기반 등.

중견 연구 제안서 작성

전립선 세그멘테이션 (Liu et al., 2020):

전립선 질환은 남성에게 흔한 질병이며, MRI에서의 정확한 전립선 세그멘테이션은 질병의 진단 및 치료 계획에 매우 중요함.

=> 3D-CNN, V-Net 기반 등 많은듯.

중견 연구 제안서 작성

초음파 이미지내 담낭 용종 세그멘테이션 (Lian et al., 2017) :

초음파 이미지는, 경제성, 편의성, 우수한 방법성 등의 장점으로, 담석의 임상 진단에서 가장 일반적으로 사용되는 기술임.

-> 담낭 및 담석 크기의 정량화가 필요할 때가 있고, 이를 위해 수동 세그멘테이션을 수행하지만, 이는 노동 집약적이고, 시간소모가 큼

=> 딥러닝 기반 방식 아직 찾지 못함.

- 폐와 같은 메이저 연구와 차이가 있는듯

갑상선 연구

- 새로운 AX 1, 2, 4, 5 세그멘테이션 작업
 - > 금일 완료 예정

산학 협력 프로젝트

- 특허 출원: 금일 '추가 질문이 필요한지 여부'와 '명세서 작성 예상 일정' 수신 예정

Reference

- Gonzalez, Yesenia, et al. "Semi-automatic sigmoid colon segmentation in CT for radiation therapy treatment planning via an iterative 2.5-D deep learning approach." *Medical Image Analysis* 68 (2021): 101896.
- Lian, Jing, et al. "Automatic gallbladder and gallstone regions segmentation in ultrasound image." *International journal of computer assisted radiology and surgery* 12.4 (2017): 553-568.
- Liu, Quande, et al. "MS-Net: multi-site network for improving prostate segmentation with heterogeneous MRI data." *IEEE transactions on medical imaging* 39.9 (2020): 2713-2724.
- Muneeswaran, V., and M. Pallikonda Rajasekaran. "Automatic segmentation of gallbladder using bio-inspired algorithm based on a spider web construction model." *The Journal of Supercomputing* 75.6 (2019): 3158-3183.
- Tsochatzidis, Lazaros, et al. "Integrating segmentation information into CNN for breast cancer diagnosis of mammographic masses." *Computer Methods and Programs in Biomedicine* 200 (2021): 105913.
- Wu, Cong, et al. "Ocular Rectus Muscle Segmentation Based on Improved U-net." *IOP Conference Series: Materials Science and Engineering*. Vol. 533. No. 1. IOP Publishing, 2019.

Reference

Xiao, Zhitao, et al. "Segmentation of lung nodules using improved 3D-UNet neural network." Symmetry 12.11 (2020): 1787.

Zhu, Fubao, et al. "Semantic segmentation using deep learning to extract total extraocular muscles and optic nerve from orbital computed tomography images." Optik 244 (2021): 167551.