Deconfounding age effects when detecting dementia

Presented by Zining Zhu

CSC2541 presentation

Jan 30, 2020

Agenda

Background: Detecting dementia

What is dementia?

- Impaired memory and cognitive capacity.
- $\approx$1 in 3 elderly suffer from dementia in US [1]
- Caused by e.g., Alzheimer's Disease.

No definitive cure yet. We can at best detect them.

Background: Detecting dementia

How to detect dementia? E.g. Winterlight pipeline:

- Let users describe pictures.
- Compute *linguistic features* about descriptions.
- Train classifiers.

However this approach has an imperfectness.

Problem: The confounder age effect

Deconfounding age effects when detecting dementia

Linguistic features include those describing:

- Memory capacity (e.g., complexity of sentence)
- Cognitive sharpness (e.g., pause time and rate)
- Linguistic knowledge (e.g., vocab richness)

Dementia could impact them, but age can, as well!

Problem: The confounder age effect

Deconfounding age effects when detecting dementia

Causal diagrams between age A, features X, and dementia D:

$$A \rightarrow X \leftarrow D$$

Classifier models learn $P(D, X)$, but they are actually:

$$P(D, X, A) = \sum_X \sum_A P(D, X) P(X, A)$$

The $P(X, A)$ term is causing confounding.

Problem: The confounder age effect

Deconfounding age effects when detecting dementia

What does confounding on age mean?

- The classifier might first infer age based on features, and then predict dementia $\hat{D}$.
- E.g., If age $A > 80$, then predict $\hat{D} = 1$.
- This improves accuracy, but this is "spurious accuracy".
- We'll show classifiers are able to do so.

Problem: The confounder age effect

Deconfounding age effects when detecting dementia

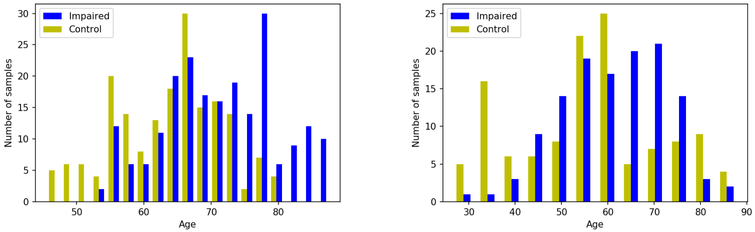

Figure: Aging histogram for DementiaBank (left) and Famous People (right). For DemBank, almost all data samples of $A > 80$ are positive. For FP: $65 < A < 73$ contain a lot of positive.

Problem: The confounder age effect

Deconfounding age effects when detecting dementia

	Mean abs error (years)
DementiaBank	15.5 ± 1.3
Famous People	14.3 ± 2.5

Table: DNNs could indeed infer ages to certain accuracy.

- On DB: $93.6 \pm 0.0\%$ elder-than-80 seniors are classified as dementia.
- On FP: $80.9 \pm 0.0\%$ 65-to-73 years old seniors are classified as dementia. [1]

[1] All on 5-fold cross validations, 10 runs with different random seeds.

Problem: The confounder age effect

How to remove the confounding effect?

1. Residualization (on X or D)
2. Inverse probability weighting
3. Propensity score matching

The traditional statistical approaches are either inferior to our approach (1,2), or not applicable in this scenario (3).

A fair representation learning approach

Deconfounding age effects when detecting dementia

In this paper, we:

- Deconfound with fair representation learning (i.e., learn $P(D, do(X))$ not $P(D, X)$)
- Evaluate with a modified equalized opportunity score.
- Experiments are on two datasets (DemBank and Famous People).

A fair representation learning approach

Deconfounding age effects when detecting dementia

How to let classifiers learn $P(D, X)$ given only $P(D, X, A)$?

- Still learn $P(D|X, A)$, but let the representations *indistinguishable* across A.
- Then our representations are clean of A.
- This "indistinguishable" idea looks like GAN [2].

A fair representation learning approach

Deconfounding age effects when detecting dementia

In Generative Adversarial Network:

- A generator $G(.)$ tries to generate images that are *indistinguishable from* true images.
- A discriminator $D(.)$ tries to tell them apart.
- The G and D networks are optimized iteratively.

A fair representation learning approach

Deconfounding age effects when detecting dementia

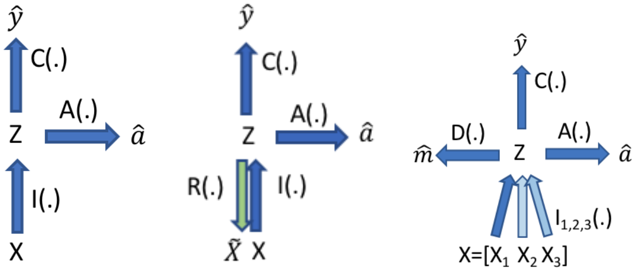

Figure: Model structures. From left to right: age-indep-simple, age-indep-AE, age-indep-CN. *-AE is used by [3], *-CN is inspired by [4].

A fair representation learning approach

Deconfounding age effects when detecting dementia

Evaluate models on accuracy and $\Delta_{eo}^{(N_a)}$.

$$\Delta_{eo}^{(N_a)} = \sum_{a=1}^{N_a} \left| p_a - \hat{p} \right| + \sum_{a=1}^{N_a} \left| n_a - \hat{n} \right|,$$

- There are N_a age groups.
- $p_a - \hat{p}$ is false positives in group a.
- $n_a - \hat{n}$ is false negatives in group a.

A fair representation learning approach

Deconfounding age effects when detecting dementia

Comments on the metric:

- When everyone has the same age: $\Delta_{eo}^{N_a} = 0$.
- Extension of equalized odds in [3] and [5].
- Our main contributions: apply to age.
- $\frac{1}{2}\Delta_{eo}^{2}$ is different from $\frac{1}{4}\Delta_{eo}^{4}$, so don't normalize.

A fair representation learning approach

Deconfounding age effects when detecting dementia

Datasets: DementiaBank (DB) [6] and Famous People (FP) [7].

	N. Samples (pos/neg)	Age
DB	213 / 182	68.26 ± 9.00
FP	124 / 121	59.25 ± 13.60

Table: Demographic information about the DementiaBank (DB) and Famous People (FP) datasets.

Results

Deconfounding age effects when detecting dementia

Overview of our results:

1. Our approaches does better than statistical adjustments.
2. Our approaches is not faraway from theoretical upper bound.

Results

Deconfounding age effects when detecting dementia

Deconfounding	DementiaBank			Famous People		
	Accuracy	$\Delta_{eo}^{(2)}$	$\Delta_{eo}^{(5)}$	Accuracy	$\Delta_{eo}^{(2)}$	$\Delta_{eo}^{(5)}$
Raw features	.77±.05	0.17±0.14	0.94±0.22	.65±.06	0.37±0.18	1.66±0.75
Res-linear	.74±.03	0.21±0.16	1.08±0.38	**.69±.04**	0.27±0.19	1.72±0.74
Res-quadratic	.74±.03	0.16±0.08	0.84±0.34	.66±.07	0.32±0.17	1.49±0.57
IPW-adjust	.70±.03	0.11±0.07	**0.67±0.18**	.63±.08	0.32±0.15	1.87±0.49
*-simple	.75±.06	**0.08±0.07**	0.80±0.28	.64±.06	0.22±0.14	1.38±0.50
*-autoencoder	.75±.05	0.11±0.08	0.88±0.24	.64±.07	0.21±0.16	**1.27±0.47**
Optimize-EO	.68±.03	**0.08±0.04**	0.91±0.27	.64±.05	**0.16±0.15**	1.35±0.47

Figure: 2. Our approaches vs others. Highlighted stats outperform traditional approaches (Residualization and IPW-adjust)

Limitations

Deconfounding age effects when detecting dementia

- Size of datasets reduce validity.
- Hard to generate large, synthetic datasets.
- Having to look at group size division introduces parameter.

Conclusion

Deconfounding age effects when detecting dementia

Summary of our contributions:

1. Identify the $A \rightarrow X \leftarrow D$ confounding problem.
2. Propose fair representation learning to address it.
3. Propose an evaluation metric $\Delta_{eo}^{(N_a)}$
4. Show superior performances of our models.

Conclusion

Deconfounding age effects when detecting dementia

For details please check out our paper on arxiv.

Deconfounding age effects with fair representation learning when assessing dementia

Zining Zhu[1,2], **Jekaterina Novikova**[1], **Frank Rudzicz**[3,2,4,5,1]
[1]Winterlight Labs [2]University of Toronto Computer Science
[3]Li Ka Shing Knowledge Institute, St Michael's Hospital
[4]Vector Institute for Artificial Intelligence [5]Surgical Safety Technologies Inc
{zining, jekaterina}@winterlightlabs.com, frank@cs.toronto.edu

Figure: arxiv 1807.07217

References

Deconfounding age effects when detecting dementia

Alzheimer's Association.
Alzheimer's disease facts and figures.
Alzheimer's & dementia, 2018.

Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio.
Generative adversarial nets.
In *NeurIPS*, 2014.

David Madras, Elliot Creager, Toniann Pitassi, and Richard Zemel.
Learning adversarially fair and transferable representations.
In *ICML*, 2018.

Zining Zhu, Jekaterina Novikova, and Frank Rudzicz.
Detecting cognitive impairments by agreeing on interpretations on linguistic features.
NAACL, 2019.

Rich Zemel, Yu Wu, Kevin Swersky, Toni Pitassi, and Cynthia Dwork.
Learning Fair Representations.
In *ICML*, pages 325–333, 2013.

Harold Goodglass and Edith Kaplan.
The assessment of aphasia and related disorders, 1983.

Aparna Balagopalan, Jekaterina Novikova, and Frank Rudzicz.
Early prediction of Alzheimer's disease from spontaneous speech.
NeurIPS ML4H, 2018.