



**ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH
TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN**

KHOÁ LUẬN TỐT NGHIỆP

**Học có giám sát với dữ liệu có phân bố thay
đổi bằng mô hình dựa trên quan hệ nhân quả**

GVHD: ThS. Trần Trung Kiên và TS. Nguyễn Ngọc Thảo

Nhóm sinh viên thực hiện:

20120032 – Phan Trường An

20120061 – Phạm Dương Trường Đức

Tháng 7/2024

Nội dung

1. Giới thiệu bài toán và một số phương pháp giải quyết đã được đề xuất
2. Mô hình mạng nơ-ron với phương pháp huấn luyện dựa trên quan hệ nhân quả CACM
3. Thực nghiệm
4. Kết luận & hướng phát triển

Nội dung

1. Giới thiệu bài toán và một số phương pháp giải quyết đã được đề xuất
2. Mô hình mạng nơ-ron với phương pháp huấn luyện dựa trên quan hệ nhân quả CACM
3. Thực nghiệm
4. Kết luận & hướng phát triển

Bài toán học có giám sát với dữ liệu có phân bố thay đổi

Cho dữ liệu của K miền huấn luyện, dữ liệu của mỗi miền có dạng $\{(\mathbf{x}_i, y_i)\}_{i=1}^n$ (n : số điểm dữ liệu).

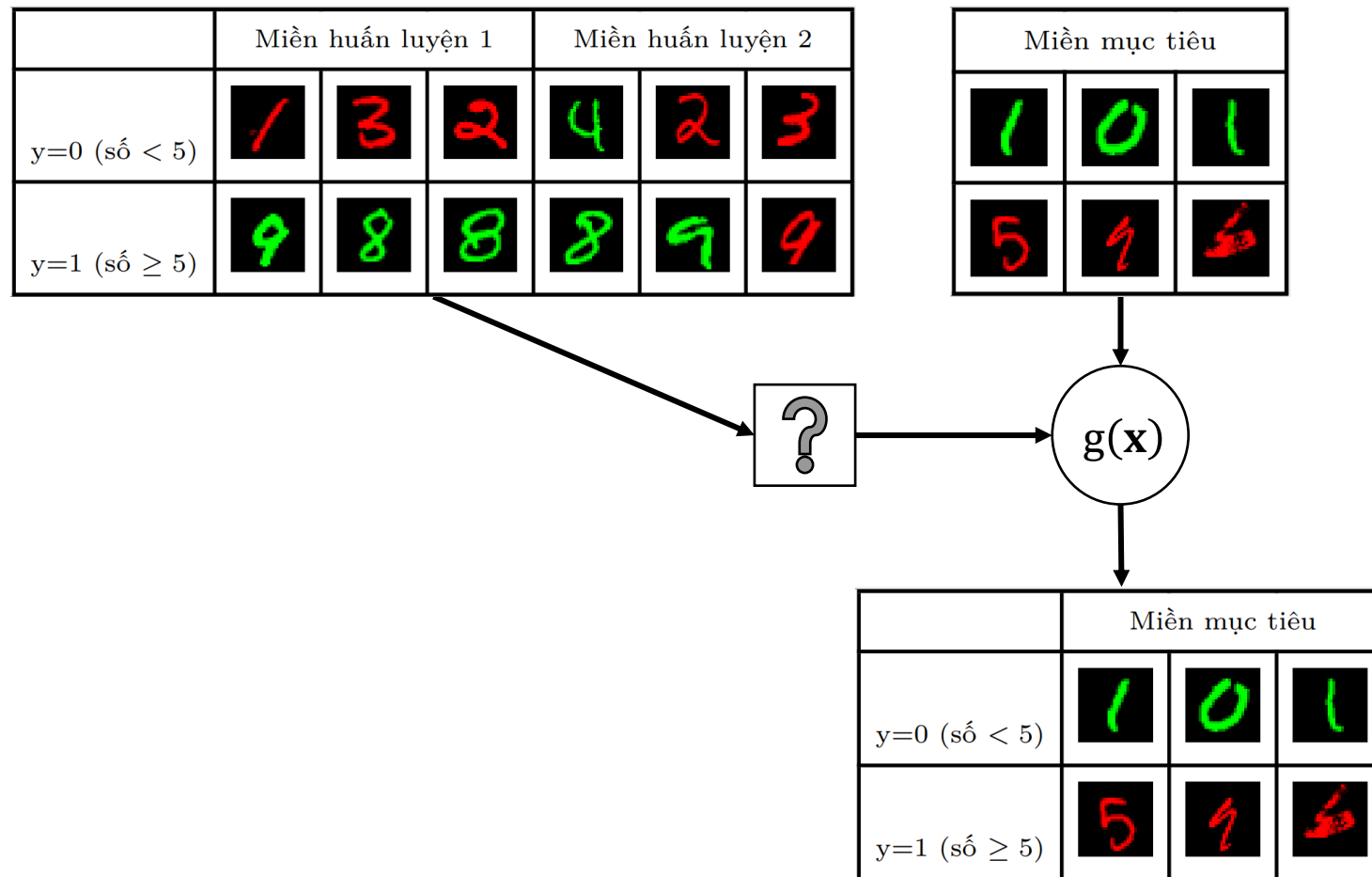
Mỗi miền ứng với một phân bố $P(\mathbf{x}, y) = P(\mathbf{x})P(y|\mathbf{x})$ mà phát sinh ra dữ liệu.

Các miền khác nhau có $P(\mathbf{x})$ khác nhau.

Yêu cầu: Tìm hàm $g(\mathbf{x})$ mà có thể dự đoán tốt y của \mathbf{x} thuộc dữ liệu miền mục tiêu. Miền mục tiêu có phân bố $P(\mathbf{x})$ khác so với các miền huấn luyện.

Minh họa bài toán

Dữ liệu có phân bố $P(x)$ thay đổi



Ý nghĩa

Giúp các mô hình học máy vẫn giữ được hiệu suất tốt khi triển khai ra thực tế và gặp dữ liệu có phân bố thay đổi.

Nội dung

1. Giới thiệu bài toán và một số phương pháp giải quyết đã được đề xuất
2. Mô hình mạng nơ-ron với phương pháp huấn luyện dựa trên quan hệ nhân quả CACM
3. Thực nghiệm
4. Kết luận & hướng phát triển

Empirical Risk Minimization (ERM)

- Tối thiểu độ lỗi trung bình trên tất cả các miền dữ liệu huấn luyện.
- Là phương pháp cơ bản và thường được sử dụng làm baseline.

Correlation Alignment (CORAL) (2016) [1]



















- Căn chỉnh dữ liệu các miền dựa trên hiệp phương sai.
- Sau khi căn chỉnh, dữ liệu của các miền sẽ có phân bố tương tự nhau.

[1] B. Sun, J. Feng, and K. Saenko, “Return of frustratingly easy domain adaptation,” in Proceedings of the AAAI conference on artificial intelligence, 2016.

Causally Adaptive Constraint Minimization (CACM) (2022) [2]

Trong quá trình sinh dữ liệu, dữ liệu có thể chứa các **thuộc tính a**, đặc trưng cho miền tương ứng và ảnh hưởng đến giá trị của đầu vào **x**.

Ví dụ: Thuộc tính **color** trong bộ dữ liệu Colored MNIST.

	Miền huấn luyện 1			Miền huấn luyện 2			Miền mục tiêu		
$y=0$ (số < 5)									
$y=1$ (số ≥ 5)									

Hình 1: Một số mẫu dữ liệu Colored MNIST

[2] J. N. Kaur, E. Kiciman, and A. Sharma, “Modeling the data-generating process is necessary for out-of-distribution generalization,” International Conference on Learning Representations, 2022.

Causally Adaptive Constraint Minimization (CACM) (2022)

- CACM dựa vào **đồ thị nhân quả** tương ứng với **mỗi quan hệ** giữa **nhân** và **thuộc tính** để định hướng quá trình huấn luyện của mô hình.
- Đồ thị nhân quả được xác định dựa trên từng bộ dữ liệu khác nhau.
⇒ CACM có thể hoạt động tốt với nhiều bộ dữ liệu khác nhau.
- CACM là phương pháp chính được tìm hiểu sâu trong khóa luận này.

Nội dung

1. Giới thiệu bài toán và một số phương pháp giải quyết đã được đề xuất
2. Mô hình mạng nơ-ron với phương pháp huấn luyện dựa trên quan hệ nhân quả CACM
3. Thực nghiệm
4. Kết luận & hướng phát triển

Dạng mô hình

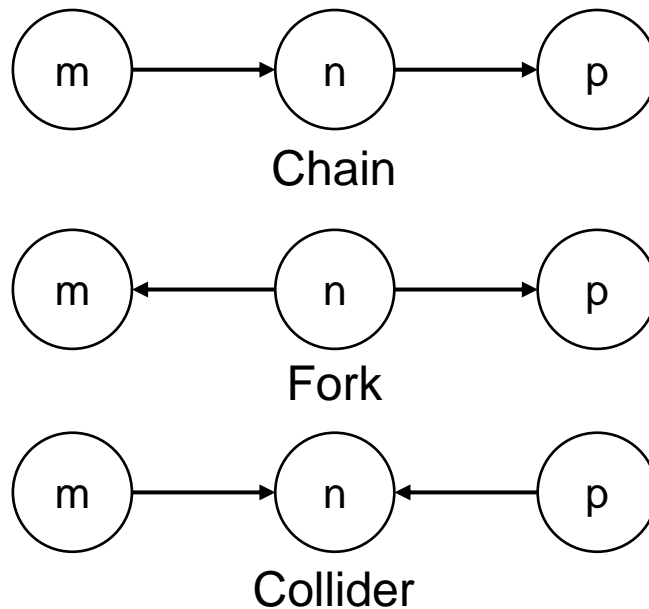
- CACM được áp dụng cho các mô hình mạng nơ-ron.
- Tùy vào từng loại dữ liệu mà sử dụng kiến trúc mô hình phù hợp (CNN cho dữ liệu ảnh, RNN cho dữ liệu văn bản,...).

Huấn luyện mô hình bằng phương pháp CACM

Bước 1: Xác định đồ thị nhân quả ứng với dữ liệu đang xét

Bước 2: Huấn luyện mô hình bằng phương pháp CACM từ dữ liệu huấn luyện và đồ thị nhân quả

Xác định đồ thị nhân quả với dữ liệu đang xét



Hình 2: Ví dụ về đồ thị nhân quả

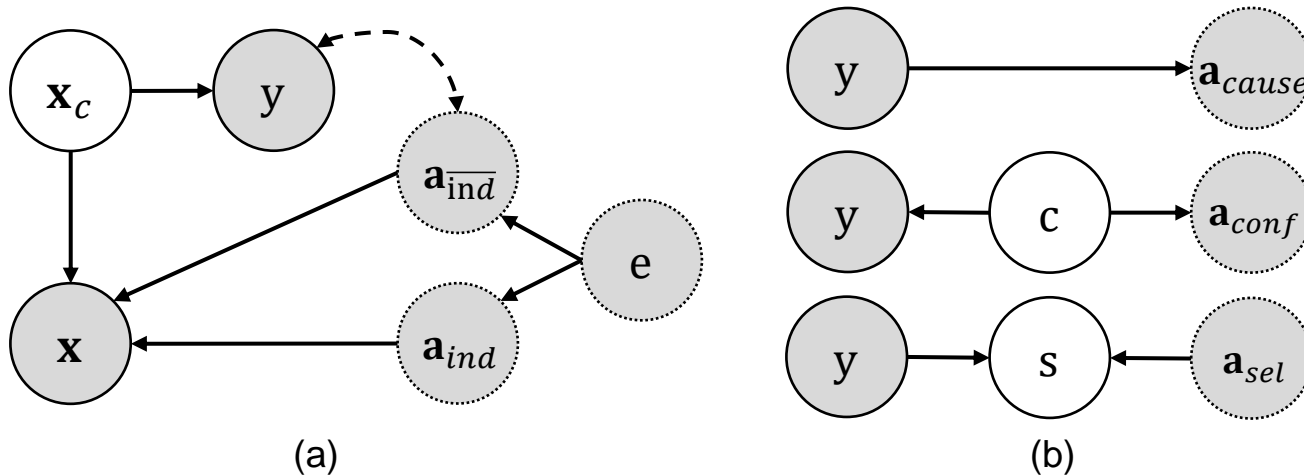
→ : Mối quan hệ nhân quả trực tiếp

Ví dụ:

Trong Chain:

- m: Tập luyện thể dục
- n: Sức khỏe
- p: Tuổi thọ

Xác định đồ thị nhân quả ứng với dữ liệu đang xét

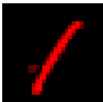










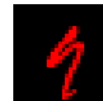


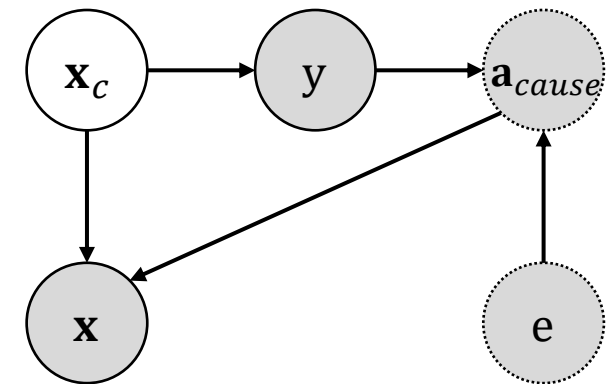
Hình 3: (a) Đồ thị nhân quả thể hiện mối quan hệ giữa các biến trong quá trình sinh dữ liệu. (b) Mối quan hệ cụ thể giữa y và $a_{\overline{ind}}$.

- a : biến đại diện cho các thuộc tính
 $a = a_{ind} \cup a_{\overline{ind}} \cup e$
- e : biến ứng với miền mà dữ liệu được sinh ra ($e = 1, 2, \dots$)
- x_c : tất cả đặc trưng gây ra y ($x_c \rightarrow y$), x_c được rút trích từ x bởi mạng nơ-ron.
- x_c, a cùng gây ra x ($x_c \rightarrow x$ và $a \rightarrow x$)
- $a_{\overline{ind}}$ có thể được chia thành các biến con như $a_{cause}, a_{conf}, a_{sel}$
- \longleftrightarrow : Mối quan hệ tương quan

Xác định đồ thị nhân quả ứng với dữ liệu đang xét

Bộ dữ liệu Colored MNIST













	Train				Test	
	D_1		D_2		D_{Test}	
$y=0$ (số < 5)						
$y=1$ (số ≥ 5)						

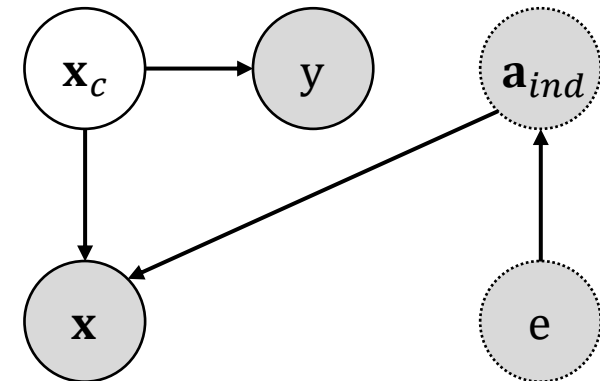


Hình 4: Đồ thị nhận dạng nhân quả ứng với mối quan hệ *Causal*

Xác định đồ thị nhân quả ứng với dữ liệu đang xét

Bộ dữ liệu Rotated MNIST

	Train				Test	
	D_1		D_2		D_{Test}	
$y=0$ (số < 5)						
$y=1$ (số ≥ 5)						



Hình 5: Đồ thị nhận dạng nhân quả ứng với mối quan hệ *Independent*

Huấn luyện mô hình bằng phương pháp CACM từ dữ liệu huấn luyện và đồ thị nhân quả

Tổng quan thuật toán CACM

Đầu vào: Dữ liệu $\{(\mathbf{x}_i, \mathbf{a}_i, y_i)\}_{i=1}^n$ của K miền, đồ thị nhân quả (hoặc mối quan hệ giữa nhãn và thuộc tính)

Đầu ra: Hàm dự đoán $g(\mathbf{x})$.

Pha 1: Rút ra các ràng buộc độc lập từ đồ thị nhân quả.

Pha 2: Áp dụng Regularization penalty sử dụng các ràng buộc đã rút ra.

Huấn luyện mô hình bằng phương pháp CACM từ dữ liệu huấn luyện và đồ thị nhân quả

Pha 1: Rút ra các ràng buộc độc lập từ đồ thị nhân quả.

Với từng thuộc tính, mỗi mối quan hệ giữa a và y sẽ dẫn đến các ràng buộc độc lập khác nhau. Các ràng buộc có thể được rút ra từ đồ thị nhân quả và d-separated [3].

- *Independent:* $\mathbf{x}_c \perp\!\!\!\perp \mathbf{a}_{ind}; \mathbf{x}_c \perp\!\!\!\perp e; \mathbf{x}_c \perp\!\!\!\perp \mathbf{a}_{ind} \mid y; \mathbf{x}_c \perp\!\!\!\perp \mathbf{a}_{ind} \mid e; \mathbf{x}_c \perp\!\!\!\perp \mathbf{a}_{ind} \mid y, e$
- *Causal:* $\mathbf{x}_c \perp\!\!\!\perp \mathbf{a}_{cause} \mid y; \mathbf{x}_c \perp\!\!\!\perp e; \mathbf{x}_c \perp\!\!\!\perp \mathbf{a}_{cause} \mid y, e$
- *Confounded:* $\mathbf{x}_c \perp\!\!\!\perp \mathbf{a}_{conf}; \mathbf{x}_c \perp\!\!\!\perp e; \mathbf{x}_c \perp\!\!\!\perp \mathbf{a}_{conf} \mid e$
- *Selection:* $\mathbf{x}_c \perp\!\!\!\perp \mathbf{a}_{sel} \mid y; \mathbf{x}_c \perp\!\!\!\perp \mathbf{a}_{sel} \mid y, e$

[3] Hayduk, Leslie et al. “Pearl’s D-separation: One more step into causal thinking”. In: Structural Equation Modeling 10.2 (2003), pp. 289–311.

Huấn luyện mô hình bằng phương pháp CACM từ dữ liệu huấn luyện và đồ thị nhân quả

Pha 2: Áp dụng Regularization penalty sử dụng các ràng buộc đã rút ra.

Ứng với các ràng buộc độc lập đã rút ra, thì sẽ chuyển thành một Regularization penalty.

$$\operatorname{argmin}_{\text{NeuralNetWeight}} (\text{TrainingError} + \text{RegPenalty})$$

Nội dung

1. Giới thiệu bài toán và một số phương pháp giải quyết đã được đề xuất
2. Mô hình mạng nơ-ron với phương pháp huấn luyện dựa trên quan hệ nhân quả CACM
- 3. Thực nghiệm**
4. Kết luận & hướng phát triển

So sánh kết quả thí nghiệm với bài báo gốc

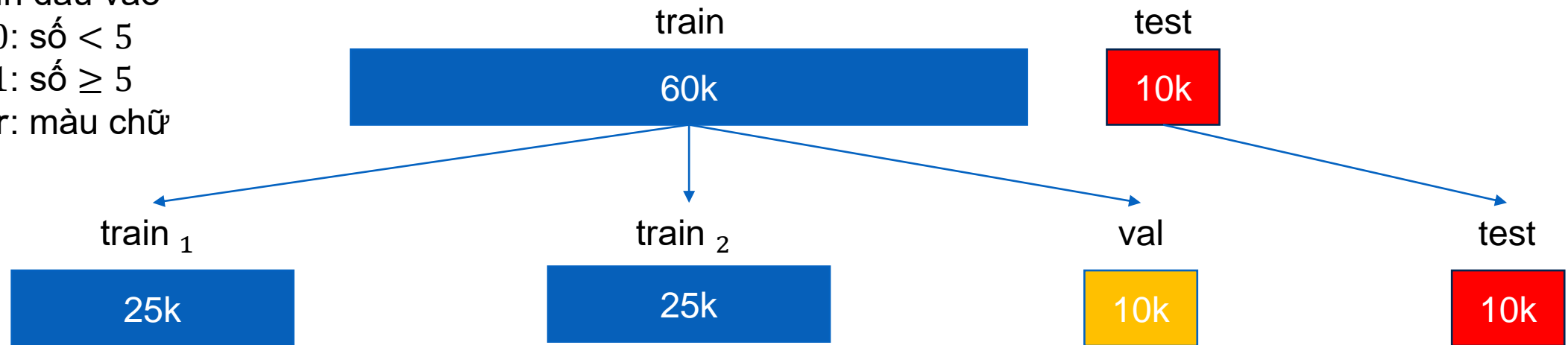
Bộ dữ liệu Colored MNIST

x: ảnh đầu vào

$y = 0$: số < 5

$y = 1$: số ≥ 5

color: màu chữ



train₁

90%:

- $y = 0 \Rightarrow \text{color} = \text{red}$
- $y = 1 \Rightarrow \text{color} = \text{green}$

10%:

- $y = 0 \Rightarrow \text{color} = \text{green}$
- $y = 1 \Rightarrow \text{color} = \text{red}$

train₂

80%:

- $y = 0 \Rightarrow \text{color} = \text{red}$
- $y = 1 \Rightarrow \text{color} = \text{green}$

20%:

- $y = 0 \Rightarrow \text{color} = \text{green}$
- $y = 1 \Rightarrow \text{color} = \text{red}$

val/test

10%:

- $y = 0 \Rightarrow \text{color} = \text{red}$
- $y = 1 \Rightarrow \text{color} = \text{green}$

90%:

- $y = 0 \Rightarrow \text{color} = \text{green}$
- $y = 1 \Rightarrow \text{color} = \text{red}$

So sánh kết quả thí nghiệm với bài báo gốc

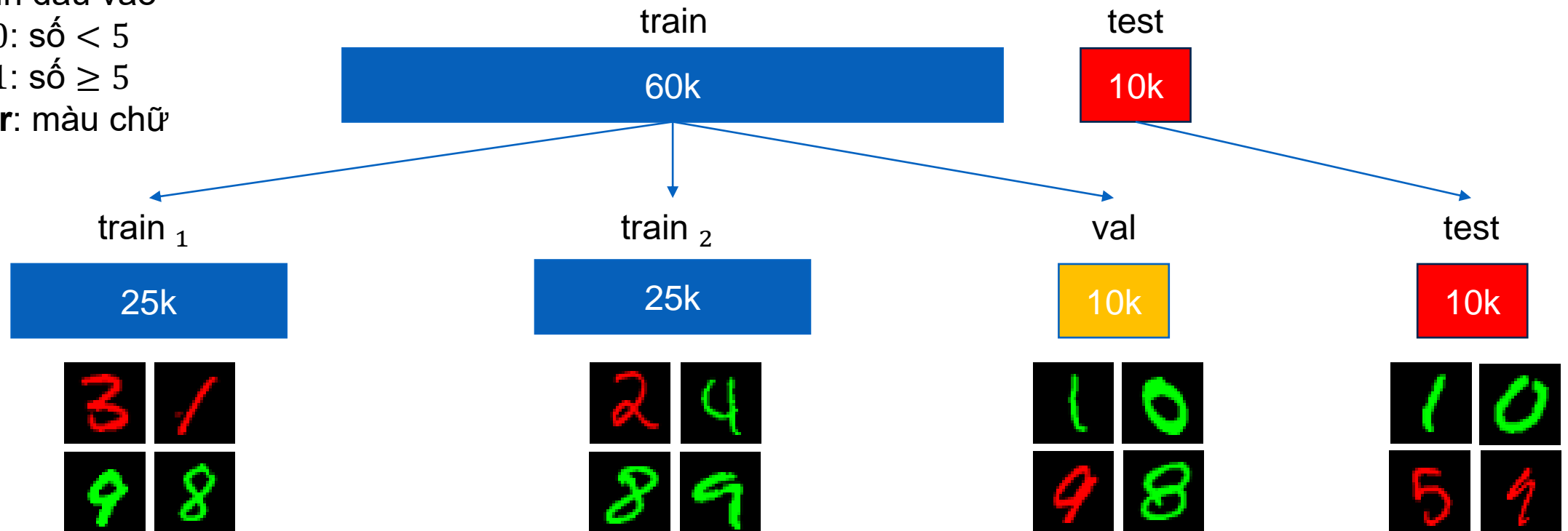
Bộ dữ liệu Colored MNIST

x: ảnh đầu vào

y = 0: số < 5

y = 1: số ≥ 5

color: màu chữ



⇒ $a_{cause} = \text{color}$ (thuộc tính color bị ảnh hưởng bởi nhãn)

Nếu mô hình học tương quan sai của y và **color** thì sẽ có hiệu suất thấp trên tập test.

So sánh kết quả thí nghiệm với bài báo gốc

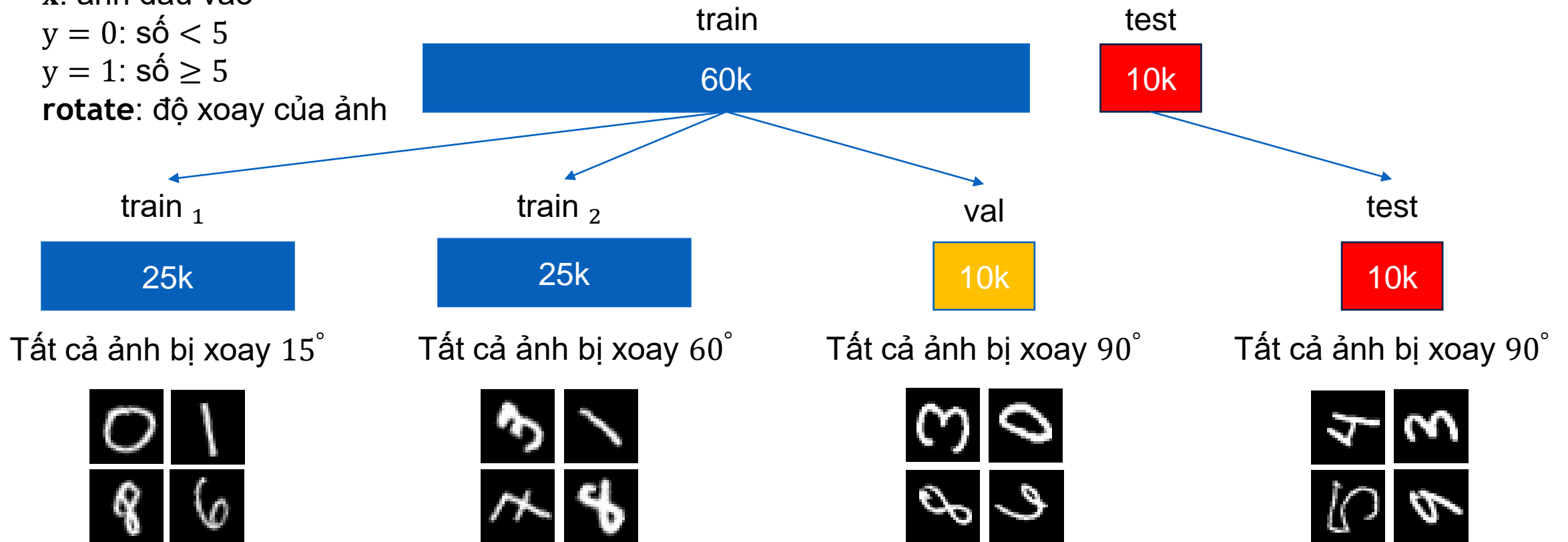
Bộ dữ liệu Rotated MNIST

x: ảnh đầu vào

$y = 0$: số < 5

$y = 1$: số ≥ 5

rotate: độ xoay của ảnh



$\Rightarrow a_{ind} = \text{rotate}$ (thuộc tính rotate độc lập với nhãn)

So sánh kết quả thí nghiệm với bài báo gốc

Bộ dữ liệu Colored+Rotated MNIST

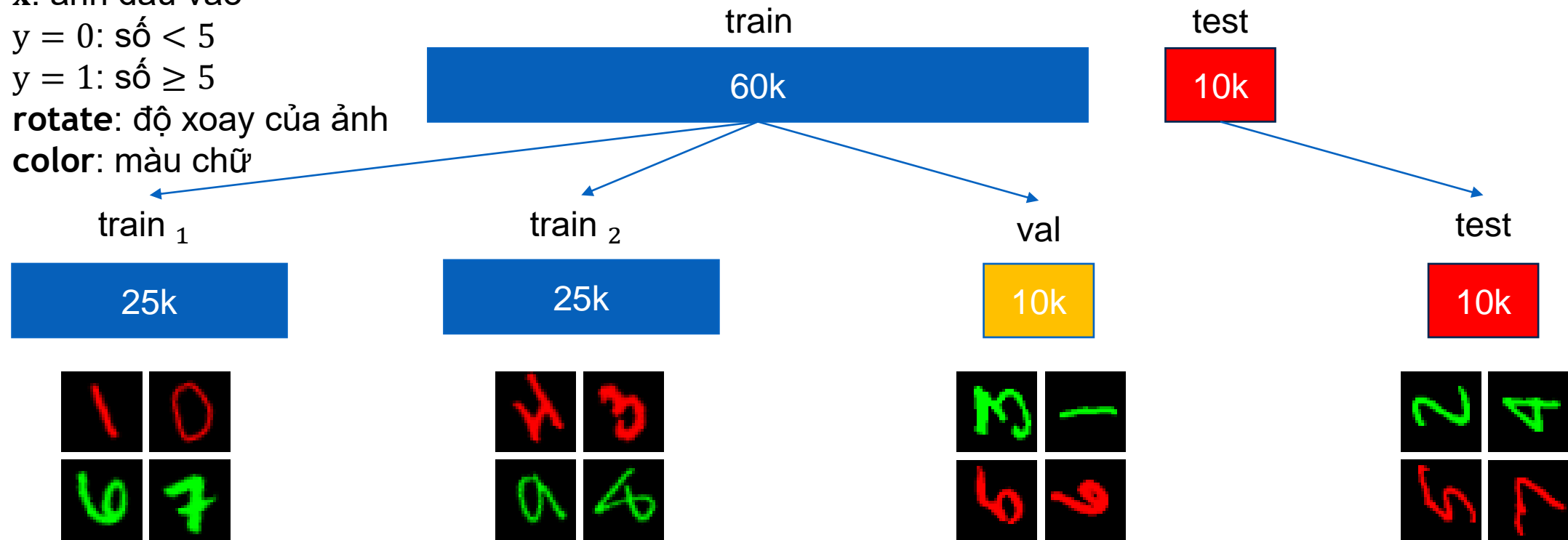
x: ảnh đầu vào

$y = 0$: số < 5

$y = 1$: số ≥ 5

rotate: độ xoay của ảnh

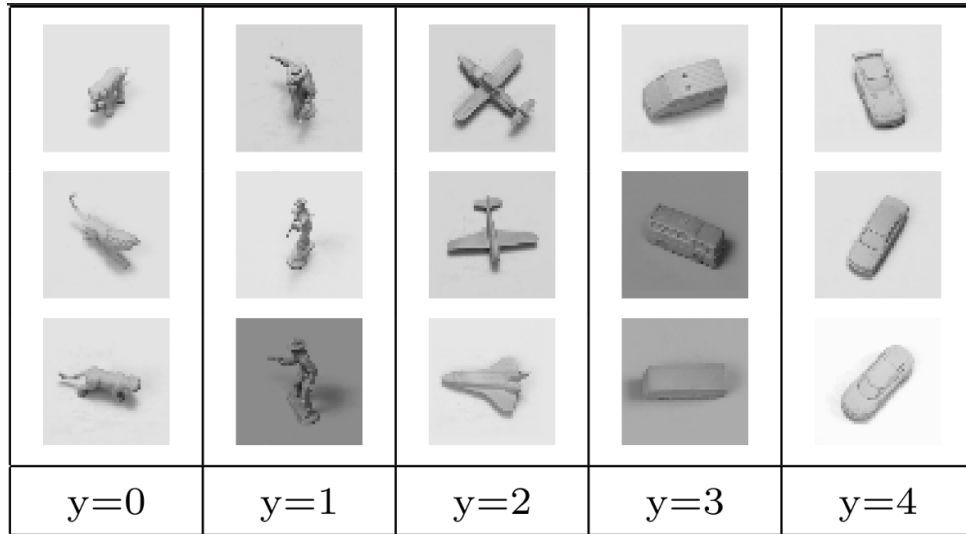
color: màu chữ



⇒ 2 loại phân bố thay đổi trong cùng một bộ dữ liệu.

So sánh kết quả thí nghiệm với bài báo gốc

Bộ dữ liệu small NORB



Hình 6: Một số mẫu dữ liệu **small NORB** gốc

Phân loại 5 lớp: động vật bốn chân, nhân vật, máy bay, xe tải, xe ô tô

3 bộ dữ liệu khác nhau được tạo ra:

- **Lighting small NORB:** $a_{cause} = \text{lighting}$ (lighting bị ảnh hưởng bởi nhãn)
- **Azimuth small NORB:** $a_{ind} = \text{azimuth}$ (azimuth độc lập với nhãn)
- **Lighting+Azimuth small NORB:** thay đổi phân bố đa thuộc tính
 - $a_{cause} = \text{lighting}$
 - $a_{ind} = \text{azimuth}$

So sánh kết quả thí nghiệm với bài báo gốc

Algo	MNIST			small NORB		
	<i>color</i>	Accuracy <i>rotation</i>	<i>col+rot</i>	<i>lighting</i>	Accuracy <i>azimuth</i>	<i>light+azi</i>
Các kết quả trong bài báo gốc						
ERM	30.9 \pm 1.6	61.9 \pm 0.5	25.2 \pm 1.3	65.5 \pm 0.7	78.6 \pm 0.7	64.0 \pm 1.2
CORAL	28.5 \pm 0.8	62.5 \pm 0.7	23.5 \pm 1.1	64.7 \pm 0.5	77.2 \pm 0.7	62.9 \pm 0.3
CACM	70.4 \pm 0.5	62.4 \pm 0.4	54.1 \pm 1.3	85.4 \pm 0.5	80.5 \pm 0.6	69.6 \pm 1.6
Các kết quả từ thí nghiệm của khóa luận						
ERM	30.8 \pm 0.6	61.6 \pm 0.2	25.5 \pm 1.0	70.8 \pm 1.7	77.6 \pm 0.7	64.1 \pm 4.3
CORAL	30.7 \pm 0.3	61.3 \pm 0.9	24.9 \pm 1.6	72.2 \pm 2.5	77.8 \pm 1.3	68.0 \pm 3.2
CACM	72.0 \pm 0.7	61.9 \pm 0.6	49.5 \pm 0.3	86.8 \pm 3.4	77.5 \pm 1.8	76.7 \pm 4.7

Bảng 1: Kết quả thí nghiệm trên các bộ dữ liệu MNIST và small NORB.
Các kết quả ở đây là độ chính xác dự đoán (%) trên tập kiểm tra.

So sánh CACM ràng buộc đúng với CACM ràng buộc sai

(Thí nghiệm mở rộng ngoài bài báo gốc)

Algo	MNIST		
	Accuracy		
	<i>color</i>	<i>rotation</i>	<i>col+rot</i>
CACM ràng buộc đúng	72.0 ±0.7	61.9 ±0.6	49.5 ±0.3
CACM ràng buộc sai	67.2 ±3.4	61.6 ±0.4	27.9 ±1.3

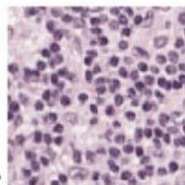
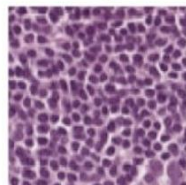
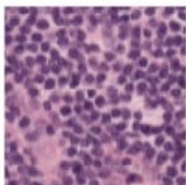
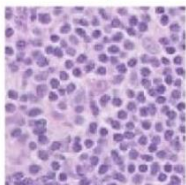
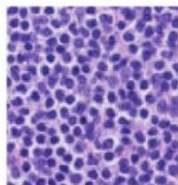
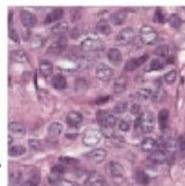
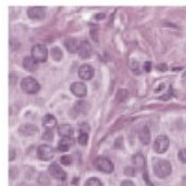
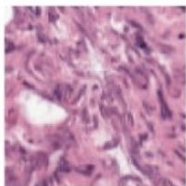
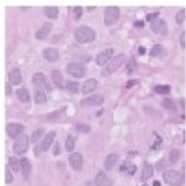
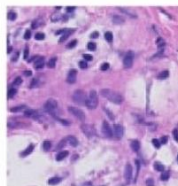
Algo	small NORB		
	Accuracy		
	<i>lighting</i>	<i>azimuth</i>	<i>light+azi</i>
CACM ràng buộc đúng	86.8 ±3.4	77.5 ±1.8	76.7 ±4.7
CACM ràng buộc sai	85.3 ±3.4	78.1 ±1.4	66.1 ±5.0

Bảng 2: Kết quả thí nghiệm CACM trên các bộ dữ liệu MNIST và small NORB với ràng buộc sai. Các kết quả ở đây là độ chính xác dự đoán (%) trên tập kiểm tra.

Áp dụng phương pháp CACM cho dữ liệu ảnh thực tế

(Thí nghiệm mở rộng ngoài bài báo gốc)

Bộ dữ liệu Camelyon17

Train			Val (OOD)	Test (OOD)	
	d = Hospital 1	d = Hospital 2	d = Hospital 3	d = Hospital 4	d = Hospital 5
y = Normal					
y = Tumor					

Phân loại 2 lớp:

- $y = 0$: không có khối u
- $y = 1$: có khối u

Có sự khác biệt trong quá trình **nhuộm tiêu bản** và **thu thập dữ liệu** giữa các bệnh viện.

$\Rightarrow \mathbf{a}_{ind} = \text{hospital}$

Hình 7: Một số mẫu dữ liệu Camelyon17 [4]

[4] Koh, Pang Wei et al. "Wilds: A benchmark of in-the-wild distribution shifts". In: International conference on machine learning. PMLR.2021, pp. 5637–5664

Áp dụng phương pháp CACM cho dữ liệu ảnh thực tế

(Thí nghiệm mở rộng ngoài bài báo gốc)

Algo	Camelyon17	
	Accuracy	
	Validation (OOD)	Test (OOD)
ERM	91.2 \pm 1.5	81.5 \pm 2.6
CORAL	71.0 \pm 18.1	71.8 \pm 5.7
CACM	93.3 \pm 0.5	83.2 \pm2.6

Bảng 3: Kết quả thí nghiệm trên bộ dữ liệu Camelyon17.
Các kết quả ở đây là độ chính xác dự đoán (%).

Nội dung

1. Giới thiệu bài toán và một số phương pháp giải quyết đã được đề xuất
2. Mô hình mạng nơ-ron với phương pháp huấn luyện dựa trên quan hệ nhân quả CACM
3. Thực nghiệm
4. Kết luận & hướng phát triển

4. Kết luận & hướng phát triển

Kết quả đạt được:

- Cài đặt thành công CACM.
- Kết quả của khóa luận tương quan với bài báo gốc.
- Thí nghiệm đánh giá hiệu suất của CACM với ràng buộc sai.
- Thí nghiệm CACM với dữ liệu ảnh thực tế.

4. Kết luận & hướng phát triển

Kết luận:

- CACM vượt trội hơn ERM và CORAL trong đa số trường hợp.
- CACM hoạt động tốt trên dữ liệu có phân bố thay đổi trên nhiều thuộc tính do áp dụng ràng buộc riêng cho từng loại thay đổi.
- Ràng buộc sai làm giảm đáng kể hiệu suất của CACM.
- CACM hoạt động tốt trên dữ liệu ảnh thực tế.

4. Kết luận & hướng phát triển

Hướng phát triển:

- Thí nghiệm CACM với loại dữ liệu khác (âm thanh, văn bản).
- Thí nghiệm để đánh giá điểm mạnh, điểm yếu của CACM.

Tài liệu tham khảo

- [1] B. Sun, J. Feng, and K. Saenko, “Return of frustratingly easy domain adaptation,” in Proceedings of the AAAI conference on artificial intelligence, 2016.
- [2] J. N. Kaur, E. Kiciman, and A. Sharma, “Modeling the data-generating process is necessary for out-of-distribution generalization,” International Conference on Learning Representations, 2022.
- [3] Hayduk, Leslie et al. “Pearl’s D-separation: One more step into causal thinking”. In: Structural Equation Modeling 10.2 (2003), pp. 289–311.
- [4] Koh, Pang Wei et al. “Wilds: A benchmark of in-the-wild distribution shifts”. In: International conference on machine learning. PMLR.2021, pp. 5637–5664

**Cảm ơn Quý Thầy, Cô và
các bạn đã lắng nghe!**

Thuật toán CACM

Đầu vào: Dữ liệu $\{(\mathbf{x}_i, \mathbf{a}_i, y_i)\}_{i=1}^n$, đồ thị nhân quả \mathcal{G} .

Đầu ra: Hàm dự đoán $g(\mathbf{x}) = g_1(\phi(\mathbf{x}))$

$\mathcal{A} \leftarrow$ tập hợp các biến quan sát được trong \mathcal{G} trừ y, e

Pha 1: Rút ra các ràng buộc chính xác

Với từng $\mathbf{a} \in \mathcal{A}$:

Nếu \mathbf{x}_c và \mathbf{a} d-separated thì:

$\mathbf{x}_c \perp\!\!\!\perp \mathbf{a}$ là ràng buộc độc lập hợp lệ

hoặc nếu \mathbf{x}_c và \mathbf{a} d-separated điều kiện \mathbf{a}_s thì:

(với $\mathbf{a}_s \subset \mathcal{A} \setminus \{\mathbf{a}\} \cup \{y\}$)

$\mathbf{x}_c \perp\!\!\!\perp \mathbf{a} | \mathbf{a}_s$ là ràng buộc độc lập hợp lệ

Pha 2: Áp dụng Regularization penalty sử dụng các ràng buộc đã rút ra.

Với từng $\mathbf{a} \in \mathcal{A}$:

Nếu $\mathbf{x}_c \perp\!\!\!\perp \mathbf{a}$ thì

$$RegPenalty_{\mathbf{a}} = \sum_{|e|} \sum_{i=1}^{|\mathbf{a}|} \sum_{j>i} MMD(P(\phi(\mathbf{x})|\mathbf{a}_i), P(\phi(\mathbf{x})|\mathbf{a}_j))$$

hoặc nếu $\mathbf{x}_c \perp\!\!\!\perp \mathbf{a} | \mathbf{a}_s$ thì

$$RegPenalty_{\mathbf{a}} = \sum_{|e|} \sum_{t \in \mathbf{a}_s} \sum_{i=1}^{|\mathbf{a}|} \sum_{j>i} MMD(P(\phi(\mathbf{x})|\mathbf{a}_i, t), P(\phi(\mathbf{x})|\mathbf{a}_j, t))$$

$$RegPenalty = \sum_{\mathbf{a} \in \mathcal{A}} \lambda_{\mathbf{a}} RegPenalty_{\mathbf{a}}$$

$$g_1, \phi = \operatorname{argmin}_{g_1, \phi} L(g_1(\phi(\mathbf{x})), y) + RegPenalty$$

D-separated [3]

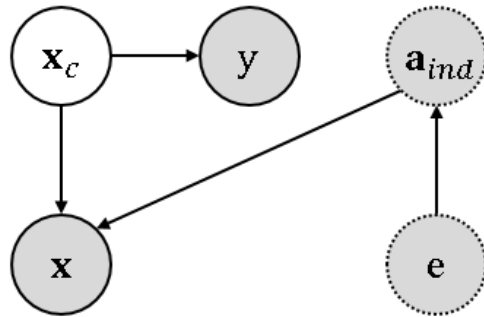
Nếu G là một đồ thị có hướng trong đó A, B và S là các tập hợp đỉnh rời rạc. Một đường đi vô hướng p giữa A và B được xem là bị chặn (d-separated) bởi S khi và chỉ khi có ít nhất một trong hai điều kiện sau thoả mãn:

1. p chứa một chain $m \rightarrow n \rightarrow p$ hoặc một fork $m \leftarrow n \rightarrow p$ sao cho nút nằm giữa n thuộc S .
2. p chứa một collider $m \rightarrow n \leftarrow p$ sao cho nút nằm giữa n và con cháu của n không thuộc S .

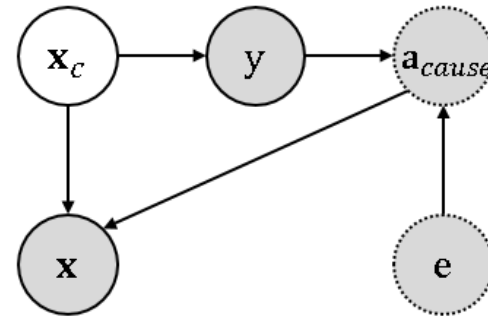
Nếu S chặn mọi đường đi từ một nút trong A đến một nút trong B thì A và B d-separated bởi S nghĩa là A và B độc lập có điều kiện trên S ($A \perp\!\!\!\perp B \mid S$).

[3] Hayduk, Leslie et al. "Pearl's D-separation: One more step into causal thinking". In: Structural Equation Modeling 10.2 (2003), pp. 289–311.

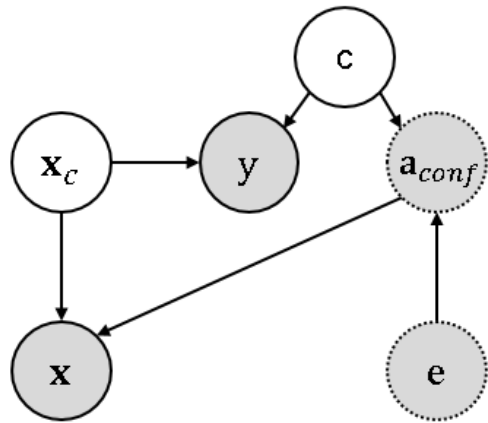
Đồ thị nhận dạng nhân quả ứng với từng mối quan hệ $a - y$



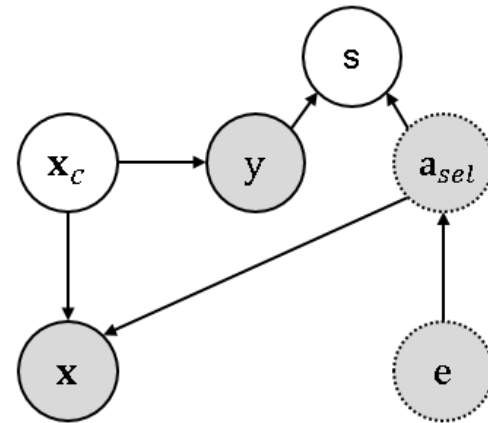
(a) Independent shift



(b) Causal shift



(c) Confounded shift



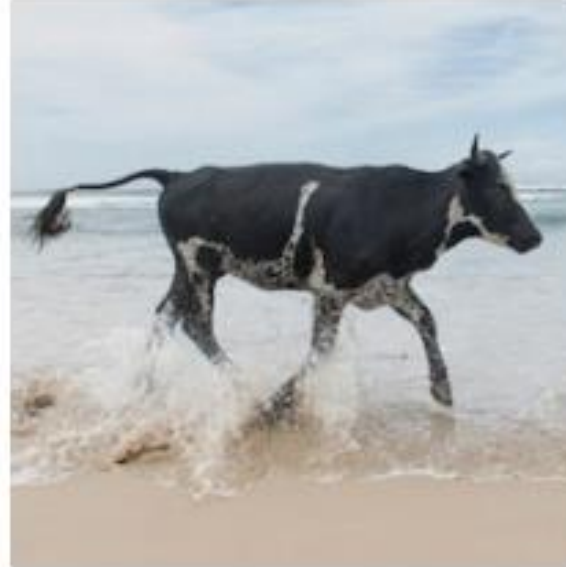
(d) Selection shift

Đồ thị nhận dạng nhân quả

Ví dụ minh họa bài toán



(A) Cow: 0.99, Pasture: 0.99, Grass: 0.99, No Person: 0.98, Mammal: 0.98



(B) No Person: 0.99, Water: 0.98, Beach: 0.97, Outdoors: 0.97, Seashore: 0.97

Miền A là miền huấn luyện

Miền B là miền mục tiêu có hình ảnh trong miền mục tiêu.