FACULTY OF ENGINEERING AND ARCHITECTURE

IDLab INTERNET & DATA LAB

Neural Bayesian Network Understudy

Paloma Rabaey, Cedric De Boom, Thomas Demeester

Motivation

Bayesian networks (BN) have many desirable properties for **decision-making in healthcare**, but their practical adoption is limited due to their inability to deal with data inadequacies. Neural networks (NN) have their own potential, but lack interpretability.

Bayesian

- \+ **Generative**: no input/output distinction
- \+ Explicitly encode **domain knowledge** in the form of (causal) structure
- − Difficulties with continuous nodes
- − Cannot deal with unstructured data (text, images)

Neural

- − Discriminative: input/output distinction
- − No knowledge of (causal) structure
- − Not interpretable
- \+ **Flexible**: learn any input/output relation
- \+ Learn useful representation of **unstructured data** (text, images)

→ **Combine strengths** of both approaches

Desired properties

Proposed model: **neural understudy (NU)** of **BN** to approximate reasoning capabilities

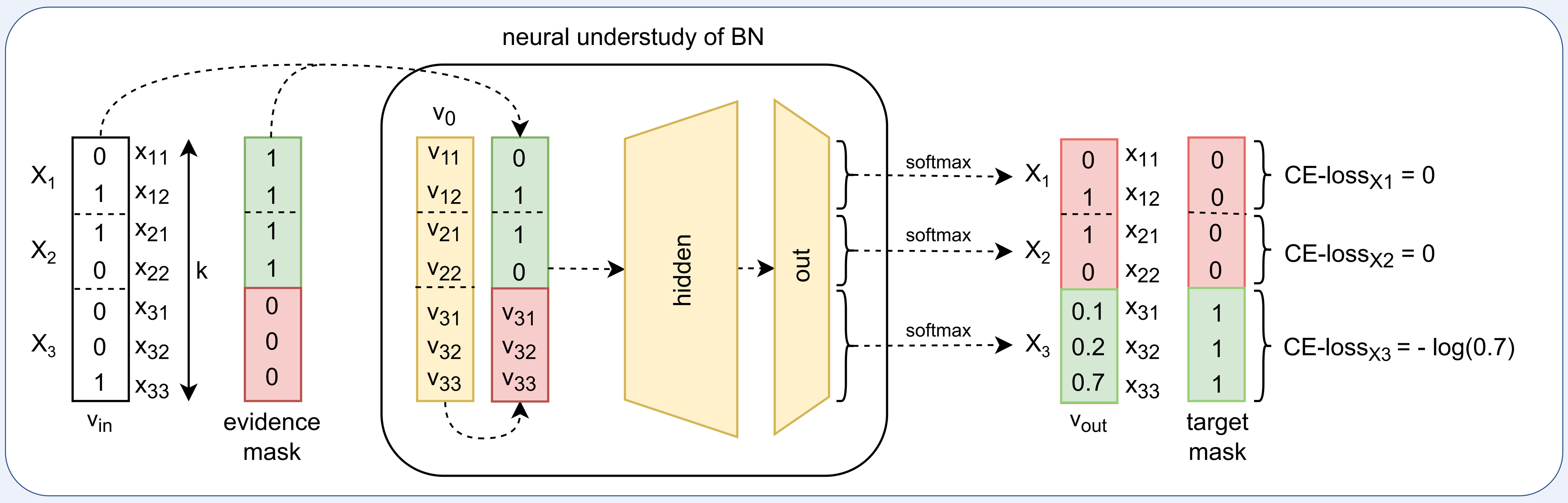
1. Trained on discrete data samples to infer the **probability** of **target** variables **conditioned** on any set of observed **evidence**

$$P(X|\mathcal{E}=e) \qquad \mathcal{E} \subset \mathcal{V} = \{X_1, X_2, \ldots, X_N\} \qquad X \in \mathcal{T} = \mathcal{V} \setminus \mathcal{E}$$

2. Incorporates **causal structure** knowledge from Directed Acyclic Graph (DAG) to improve predictions and interpretability
3. Future work: extend NU w/ continuous and **unstructured data** nodes

Neural architecture and training

Property 1: Train NU to infer conditional probabilities



Example: model receives random sample $\{X_1 = x_{12}, X_2 = x_{21}, X_3 = x_{33}\}$ from training set

- Random mask divides variables into evidence $\mathcal{E} = \{X_1, X_2\}$ and targets $\mathcal{T} = \{X_3\}$
- Model is tasked to predict $P(X_3|X_1 = x_{12}, X_2 = x_{21})$

Minimize **training loss** $\mathcal{L}^{\mathcal{T}} = \frac{1}{|\mathcal{T}|}\sum_{X_i \in \mathcal{T}} -\log\,\hat{p}_{ij}$, with $\hat{p}_{ij}$ predicted prob for target class *j* of X_i

Property 2: Incorporate causal structure

DAG describes independence relations (IRs) of the form $X \perp Y \mid \mathcal{C}$

1. **REG**: inject IRs through regularization

$$\text{Minimize } \mathcal{L} = \mathcal{L}^{\mathcal{T}} + \alpha\,\mathcal{L}^{\mathcal{R}}$$

$$\mathcal{L}^{\mathcal{R}} = \text{MSE}(p(X \mid Y{=}y, C{=}c),\ p(X \mid Y{=}y', C{=}c))$$

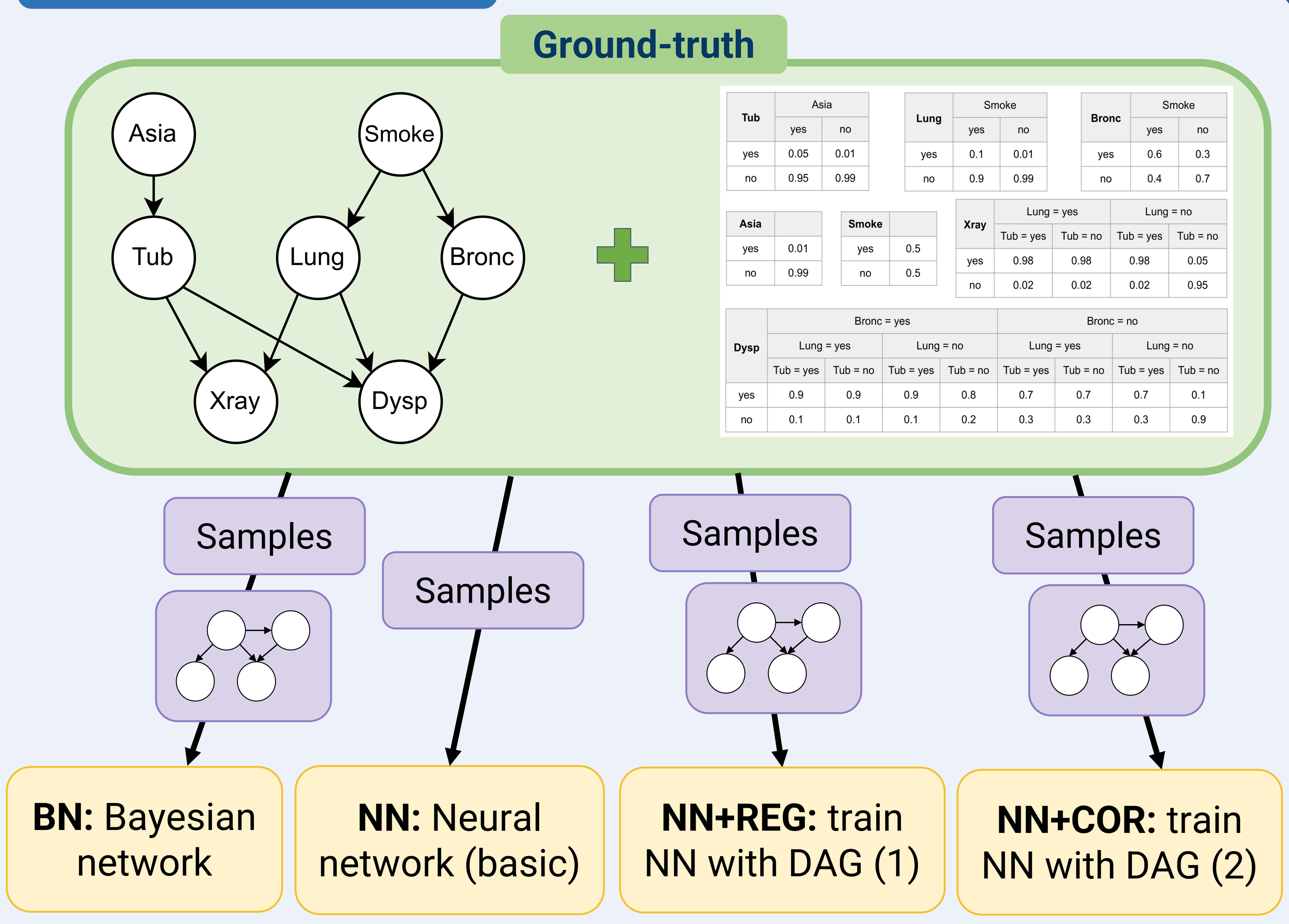
2. **COR**: inject IRs through evidence corruption

For given $\mathcal{E}$ and $\mathcal{T}$ find IR such that...

$(\{Y\} \cup \mathcal{C}) = \mathcal{E}$ and $X \subset \mathcal{T}$ OR $(\{X\} \cup \mathcal{C}) = \mathcal{E}$ and $Y \subset \mathcal{T}$

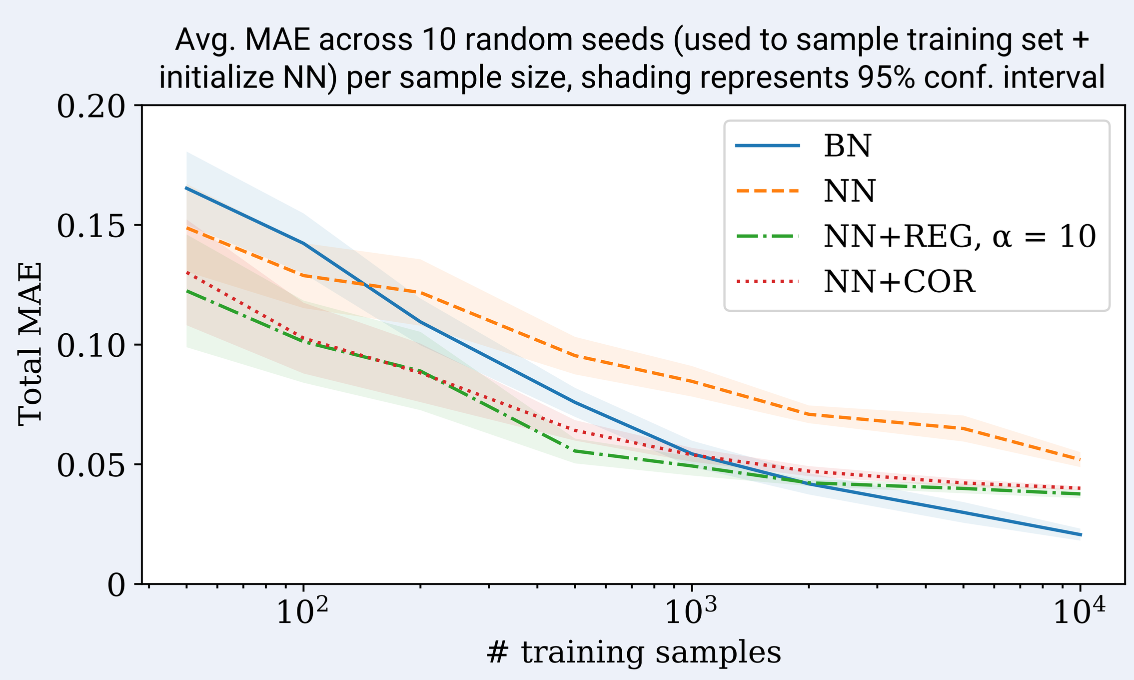
Teach model to ignore Y (X) given X (Y) by randomly corrupting Y (X)

Evaluation strategy



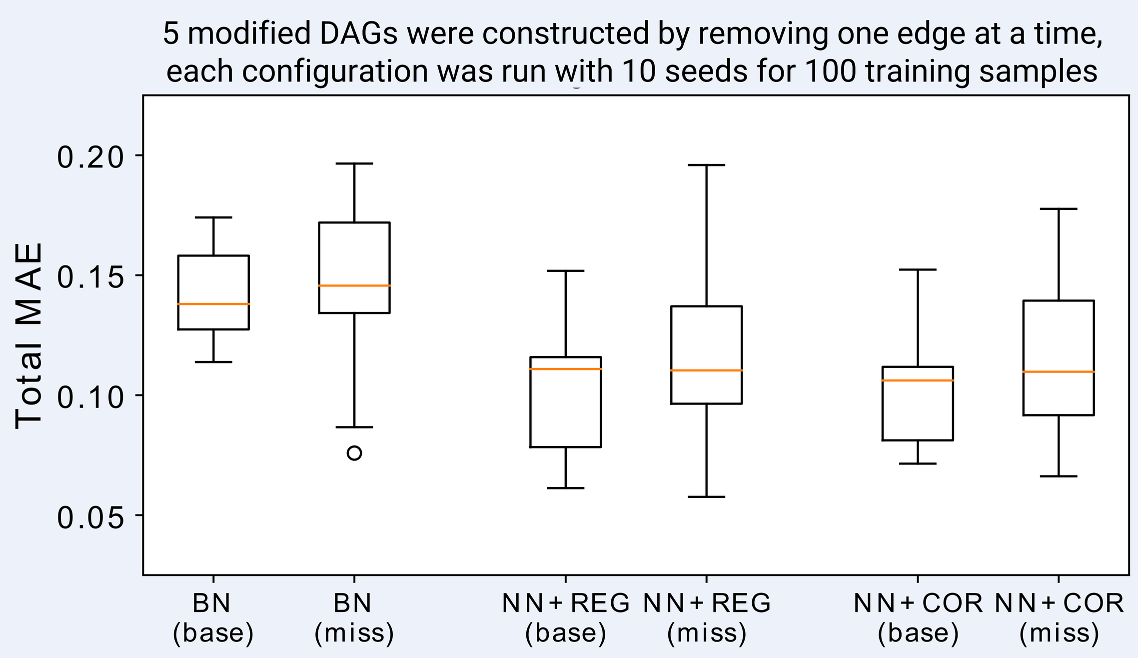
Total Mean Absolute Error (MAE): Build test queries with ground-truth probabilities of target variables, for any particular assignment of evidence variables. Compare with any model's predicted target distribution by calculating MAE between distributions.

Proof-of-concept results



RQ1: Performance of neural understudy NN is able to make approximate predictions of conditional probabilities for arbitrary set of evidence vars.

RQ2: Training NN with causal structure Training neural understudy with independence relations extracted from DAG results in similar performance compared to BN counterpart.



RQ3: Robustness to DAG miss-specification: When an incomplete DAG (one edge randomly removed) is passed to the models, performance of all models becomes less stable across sample sets.

paloma.rabaey@ugent.be
github.com/prabaey/NBN-understudy
/in/paloma-rabaey-3a29091a2

GHENT UNIVERSITY fwo imec