GHENT UNIVERSITY

FACULTY OF ENGINEERING AND ARCHITECTURE

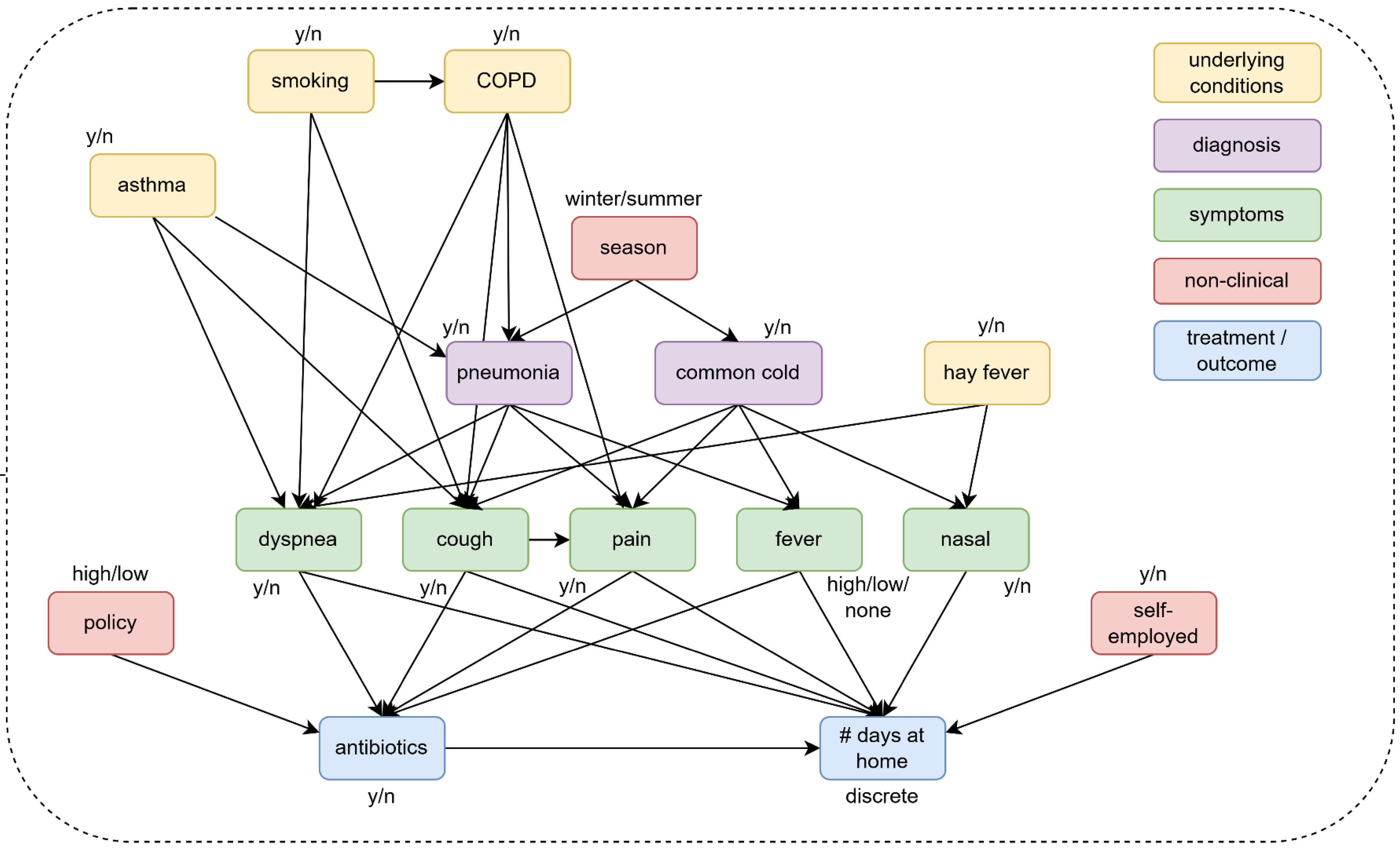
fwo

IDLab INTERNET & DATA LAB

SynSUM – Synthetic Benchmark with Structured and Unstructured Medical Records

Paloma Rabaey, Henri Arno, Stefan Heytens, Thomas Demeester

Bayesian network: expert-defined tabular data generating process



Generate **tabular patient record**

x 10.000

Patient 1	
asthma	yes
smoking	no
COPD	no
hay fever	no
season	winter
pneumonia	no
common cold	no
dyspnea	yes
cough	yes
pain	yes
nasal	no
fever	low
policy	high
self-employed	no
antibiotics	yes
# days at home	5

Descriptor chosen based on cause of symptom (here: asthma)

Probability of mentioning symptom depends on its presence (*)

Always mention background health conditions

(*)

$$\mathcal{P}(ment_{dysp} \mid dysp = yes) = 0.95$$
$$\mathcal{P}(ment_{cough} \mid cough = yes) = 0.95$$
$$\mathcal{P}(ment_{pain} \mid pain = yes) = 0.75$$
$$\mathcal{P}(ment_{fever} \mid fever = low) = 0.7$$
$$\mathcal{P}(ment_{nasal} \mid nasal = no) = 0.1$$

Prompt for clinical text generation

Create a short clinical note related to the following patient encounter.

The following information is known about the patient's symptoms:

- cough: yes,
- respiratory pain: yes,
- dyspnea: yes,

dry
burning pain in windpipe
air hunger

Don't mention anything about the following symptoms:

- fever
- nasal symptoms

The patient currently has the following underlying health conditions, which may or may not be mentioned in the note if relevant:

- asthma

The note has the following structure:
History
<history>
Physical Examination
<physical examination results>

Do not include any suspicions of possible diagnoses in the clinical note (no "assessment" field). You can imagine additional context or details described by the patient, but no additional symptoms. Do not mention patient gender or age. Your notes can be relatively long (around 5 lines or more in history).

Do not add a title. Do not add a final comment after generating the note.

Generate text using **GPT-4o**

Clinical note

History
The patient reports experiencing a persistent dry cough over the past several days, which has been accompanied by a distinct burning pain localized in the windpipe. They describe a pronounced feeling of air hunger and shortness of breath, which is exacerbated during episodes of physical exertion. These symptoms have not been relieved by the use of their usual asthma inhaler. The patient has a history of asthma, which is generally well-controlled with their current medication regimen.

Physical Examination
Upon examination, the patient appears visibly uncomfortable, with noted use of accessory muscles for breathing. Breath sounds are diminished bilaterally, with occasional wheezing heard on auscultation. No stridor or severe respiratory distress is observed, but the patient exhibits an increased respiratory rate. No cyanosis is noted, and heart sounds are normal.

Additional prompt for generating **compact version** "Please write this note in more compact style (using abbreviations and shortcuts), while preserving the content."

Compact clinical note

History
Pt c/o persistent dry cough for several days w/ burning pain in windpipe. Reports air hunger and SOB, worsens w/ exertion. Usual asthma inhaler not effective. Hx of asthma, generally well-controlled w/ meds.

Physical Examination
Pt appears uncomfortable, using accessory muscles for breathing. Diminished breath sounds bilat, occasional wheezing on ausc. No stridor or severe resp distress; ↑ RR noted. No cyanosis, heart sounds normal.

Dataset requirements

1. Fully **synthetic** research dataset.
2. Mixes **structured** tabular data and **unstructured** text.
3. Clinical concepts are connected through a **Bayesian network representing domain knowledge.**
4. Each record is a **static snapshot** of a single patient encounter (no temporal aspect).
5. Text contains **additional context** on some of the encoded tabular variables.

Expert evaluation of clinical notes

We asked 5 general practitioners to perform an in-depth evaluation on a random subset of 30 generated notes. They scored various aspects out of 5.

	Normal				Compact	
	Consis-tency	Realism (hist)	Realism (phys)	Clinical accuracy	Content	Reada-bility
mean	4.69	4.53	4.15	4.92	4.88	4.02
std	0.12	0.21	0.30	0.07	0.10	0.31

Potential uses

1. **Information extraction** from clinical text in the presence of tabular background variables.
2. Automation of **clinical reasoning** over tabular data and text.
3. **Causal effect estimation** in the presence of **textual confounders**.
4. Benchmarking **multi-modal clinical synthetic data** methods.

Symptom predictor baselines

We train symptom predictor baselines that take in either the tabular or textual part of the dataset and predict each of the 5 symptoms. We train each model on 8000 samples, and report F1 score over a test set of 2000 samples.

- **Tabular**: We fit the **Bayesian network** probabilities to the tabular portion of the data and obtain **P(symptom | evidence)** using variable elimination, where evidence is all tabular features except the symptoms.
- **Text**: We turn the note into a **clinical sentence representation** using pre-trained BioLORD embeddings [1], after which we train a **neural text classifier** (two-layer MLP) to predict each symptom label.

	Dyspnea	Cough	Pain	Nasal	Fever
Tabular	0.7153	0.7776	0.1312	0.7146	0.4384
Text (normal)	**0.9617**	**0.9603**	**0.8143**	**0.9628**	**0.9096**
Text (compact)	0.9444	0.9397	0.7940	0.9622	0.9010

[1] F. Remy, K. Demuynck, and T. Demeester, "BioLORD-2023: semantic textual representations fusing large language models and clinical knowledge graph insights," Journal of the American Medical Informatics Association, p. ocae029, 02 2024.

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