Lecture 05 Probabilistic Graphical Models: From Knowledge Representation to Graph Learning

Andreas Holzinger
VO 709.049 Medical Informatics
23.11.2016 11:15-12:45

1. Adjacency matrix: simplest form of computational graph representation, in which or for 1 denotes whether or not there is a directed edge from one node to another (in graph theory adjacent nodes in a graph are linked by an edge).

2. Artifacts: not only a noise disturbance, which is contaminating and influencing the signal (e.g., artifacts) but also data which is wrong, however interpreted as to be reliable, consequently may lead to a wrong decision.

3. Computational graph representation: e.g., by adjacency matrices.

4. Data fusion: data integration techniques that analyze data from multiple sources in order to develop insights into patterns that are more efficient and potentially more accurate than if they were developed by analyzing a single source of data. Signal processing techniques can be used to implement some types of data fusion (e.g., combined sensor data in Ambient Assisted Living).

5. Global Distance Test (GDT): a measure of similarity between two protein structures with identical amino acid sequences but different tertiary structures. It is most commonly used to compare the results of protein structure prediction to the experimentally determined structure as measured by X-ray crystallography or protein NMR.

6. Graph theory: study of mathematical structures to model relations between objects from a certain collection.

7. Graphs: a hypothetical structure consisting of a series of nodes connected by weighted edges (graphs can be directed/undirected and stoichiometric/non-stoichiometric regarding interaction classes).

8. Glossary:
   - ANS = American National Standards Institute
   - CD = Cardiac development
   - CDA = Clinical Data Architecture
   - CHD = Congenital heart disease
   - CMM = Cumulative module mining
   - DIR = Dossier Patient integre = integrated patient record
   - E = Edge
   - EPR = Electronic Patient Record
   - G(U) = Graph
   - GI = gastrointestinal
   - HER = Health Record
   - HIC = Health Level 3
   - KEGG = Kyoto Encyclopedia of Genes and Genomes
   - NP = Nondendritic monophasic transient
   - OWL = Web Ontology Language
   - PPI = Protein-Protein interaction
   - SOM = Standardized Markup Language
   - TFI = Transcription factor
   - TG = Target Gene
   - V = Vertex
   - XML = Extensible Markup Language

9. Learning Goals: at the end of this lecture you...:
   - have an idea of the complexity of data in biomedical informatics
   - are aware of the enormous importance of graphs (= network structures) and graph theory
   - have seen some application examples of network structures from both macro- and micro-levels and are fascinated about it;
   - have a rough overview about some basics of how to get point clouds out of data sets
   - have an understanding of some challenges of network science

10. Agenda for today:
   - 00 Reflection – follow-up from last lecture
   - 01 Reasoning under Uncertainty
   - 02 Where do graphs come from?
   - 03 Why are graphs so awesome?
   - 04 Knowledge Representation in Networks
   - 05 Graphs: Concepts, Metrics, Measures
   - 06 Example: Graphs from Natural Images
   - 07 Graphical Model Learning
Reflection Quiz

1. Identify the key concepts and their relationships in the graph.

2. Explain the role of prior and evidence in the calculation of posterior.

3. Describe the difference between causal and probabilistic inference.

4. Discuss the limitations of data sets in healthcare applications.

5. Analyze the implications of the "human-in-the-loop" approach.

Expected Utility Theory $E(U|d)$ Neumann-Morgenstern

For a single decision variable $d$, an agent can select $D = d$ for any $d \in \text{dom}(D)$.

The expected utility of decision $D = d$ is

$$E(U|d) = \sum_{x_1, \ldots, x_n} P(x_1, \ldots, x_n | d) U(x_1, \ldots, x_n, d)$$

An optimal single decision is the decision $D = d_{\text{max}}$ whose expected utility is maximal:

$$d_{\text{max}} = \arg \max_{d \in \text{dom}(D)} E(U|d)$$


Graphical models are graphs where the nodes represent random variables and the links represent statistical dependencies between variables. This provides us with a tool for reasoning under uncertainty.

What are Probabilistic Graphical Models?

- PGM can be seen as a combination between
- Graph Theory + Probability Theory + Machine Learning
- One of the most exciting AI advances in the last decades
- Compact representation for exponentially-large probability distributions
- Example Question: “Is there a path connecting two proteins?”
- Path $(X, Y) = \text{edge}(X, Y)$
- Path $(X, Y) = \text{edge}(X, Y), \text{path}(X, Y)$
- This can NOT be expressed in first-order logic
- Need a Turing-complete fully-fledged language


Key Challenges

- Medicine is an extremely complex application domain – dealing most of the time with uncertainties – > probable information!
- Key: Structure learning and prediction in large-scale biomedical networks with probabilistic graphical models
- Causal and Probabilistic inference.
- Uncertainties are present at all levels in health related systems
- Data sets from which ML learns are noisy, mislabeled, atypical, etc. etc.
- Even with data of high quality, gauging and combining a multitude of data sources and constraints in usually imperfect models of the world requires us to represent and process uncertain knowledge in order to make viable decisions.
- In the increasingly complicated settings of modern science, model structure or causal relationships may not be known a-priori (1).
- Approximating probabilistic inference in Bayesian belief networks is NP-hard (2) -> here we need the “human-in-the-loop” (3).


01 Reasoning under Uncertainty
Dealing with uncertainty in the real world

- The information available to humans is often imperfect - imbalanced - imprecise - uncertain.
- This is especially in the medical domain the case.
- An human agent can cope with deficiencies.
- Classical logic permits only exact reasoning:
  - IF A is true THEN A is non-false and
  - IF B is false THEN B is non-true
- Most real-world problems do not provide this exact information, mostly it is inexact, incomplete, uncertain and/or un-measurable!

Reasoning under uncertainty

- Take patient information, e.g., observations, symptoms, test results, -omics data, etc. etc.
- Reach conclusions, and predict into the future, e.g. how likely will the patient be re-admissioned
- Prior = belief before making a particular observation
- Posterior = belief after making the observation and is the prior for the next observation - intrinsically incremental

\[ p(x_t|y_t) = \frac{p(y_t|x_t)p(x_t)}{\sum p(x_t, y_t)p(x_t)} \]

Two types of decisions (Diagnosis vs. Therapy)

- Type 1 Decisions: related to the diagnosis, i.e. computers are used to assist in diagnosing a disease on the basis of the individual patient data. Questions include:
  - What is the probability that this patient has a myocardial infarction on the basis of given data (patient history, ECGs, …)?
  - What is the probability that this patient has acute appendicitis, given the signs and symptoms concerning abdominal pain?

- Type 2 Decisions: related to therapy, i.e. computers are used to select the best therapy on the basis of clinical evidence, e.g.:
  - What is the best therapy for patients of age x and risks y, if an obstruction of more than z% is seen in the left coronary artery?
  - What amount of insulin should be prescribed for a patient during the next 5 days, given the blood sugar levels and the amount of insulin taken during the recent weeks?

Leonhard Euler (1707-1783) in 1736 ...

252 years later: Belief propagation algorithm

\[ G_1 \rightarrow (V_1) \rightarrow (V_2) \rightarrow (V_3) \rightarrow (V_4) \rightarrow G_2 \]

\[ \lambda(V_1) \rightarrow \mu(V_2) \rightarrow \lambda(V_3) \rightarrow \mu(V_4) \]


227 years later ... the "Nobel-prize in Computer Science"

2) Where do Graphs come from?
Our World in Data (1/2) – Macroscopic Structures

03 Why are graphs so awesome?

Our World in Data – Microscopic Structures


Getting Insight: Knowledge Discovery from Data

TU

First yeast protein-protein interaction network

Light blue = known proteins
Orange = disease proteins
Yellow ones = not known yet

Complexity Problem: Time versus Space

P versus NP and the Computational Complexity Zoo, please have a look at https://www.youtube.com/watch?v=V5k0Mh4x2

Two thematic mainstreams in dealing with data...

Time

Space

e.g. Entropy

e.g. Topology

Fagiola & Bourke (2012) Klein Bottle


But our question was: Where do graphs come from?

- Parametric models
  - given as direct input (point cloud data sets)
  - Given as properties of a structure, e.g. biological networks
  - Given as a representation of information (e.g. Facebook data, viral marketing, etc.,...)

- Nonparametric models
  - We extract the graph from other data [1]
  - We learn the structure from samples and infer
  - Flat vector data, e.g. similarity graphs
  - Encoding structural properties (e.g. smoothness, independence, ...)


http://www.nobelprize.org/nobel_prizes/chemistry/laureates/2013/

http://news.harvard.edu/gazette/story/2013/10/nobel_price_awarded_2013/

NRC S1 15 Omega Centauri by Edmund Miller (1877, ESO, Macao, Chile)

First human protein-protein interaction network

Matrix contains many sparse elements - in this case it is computationally more efficient to represent the graph as an adjacency list.

Electronic patient records remain an unexplored, but potentially rich data source for example to discover correlations between diseases.

Homology modeling is a knowledge-based prediction of protein structures.

- In homology modeling a protein sequence with an unknown structure (the target) is aligned with one or more protein sequences with known structures (the templates).
- The method is based on the principle that homologue proteins have similar structures.
- Homology modeling will be extremely important to personalized and molecular medicine in the future.

05 Graphs: Basic concepts, metrics and measures

- In order to understand complex biological systems, the three following key concepts need to be considered:
  - (i) emergence, the discovery of links between elements of a system because the study of individual elements such as genes, proteins and metabolites is insufficient to explain the behavior of whole systems;
  - (ii) robustness, biological systems maintain their main functions even under perturbations imposed by the environment; and
  - (iii) modularity, vertices sharing similar functions are highly connected.
- Network theory can largely be applied for biomedical informatics, because many tools are already available.
06 Example: How do you get point cloud data from natural images?
Example Watershed Algorithm

Algorithm 4.3 Watershed transformations (with topological distance based on image integration via the Distance-Metric-Function graph algorithm)

1. Procedure ComputeWatershed()
   2. INPUT: Image, a gray-scale image with gradient magnitude.
   3. OUTPUT: Labeled image I on V.
   4. WHILE I ≠ 0 DO
      5. For all pixels p ∈ V DO
         6. If p is a local minimum wrt. a
         7. Add p to the catchment basin C.
         8. End if
   9. End while


Slide 5-20 Graphs from Images: Voronoi -> Delaunay


Watershed 4 Steps

- 1) Transformation into a topographic map
- 2) Finding local minima
- 3) Finding catchment basins
- 4) Erecting watersheds

Example: Watershed Segmentation of the human cortex

07 Graphical Model Learning
Learning Graphical Models from data

- Remember: GM are a marriage between probability theory and graph theory and provide a tool for dealing with our two grand challenges in the biomedical domain:

  **Uncertainty and complexity**

- The learning task is two-fold:
  1. Learning unknown probabilities
  2. Learning unknown structures


Example for Graphical Model Learning

<table>
<thead>
<tr>
<th>Patient</th>
<th>Hb</th>
<th>Russia</th>
<th>Smoker</th>
</tr>
</thead>
<tbody>
<tr>
<td>Florian</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Tamás</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Matthias</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Benjamin</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Dimitris</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Rows are independent during learning and inference

Asthma can be hereditary
- Friends may have similar smoking habits
- Augmenting graphical model with relations between the entities – Markov Logic

2.1 Asthma ⇒ Cough
3.5 Smokes ⇒ Cough
2.1 Asthma(x) ⇒ Cough(y)
3.5 Smokes(x) ⇒ Cough(y)
1.9 Smokes(x) ∧ Friends(x,y) ⇒ Smokes(y)
1.5 Asthma(x) ∧ Family(x,y) ⇒ Asthma(y)

Future Outlook

The future is in integrative MI, i.e. combining relational databases, ontologies and logic with probabilistic reasoning models and statistical learning – and algorithms that have good scalability

Conclusion and Future Challenges
Questions

Sample Questions

- What is the primary idea of a graphical model learning algorithm?
- Where do graphs come from in the medical domain?
- Where do decision trees originally come from?
- What are probabilistic graphical models?
- Why is the topic “reasoning under uncertainty” so important for the health domain?
- Why was MYCIN not a success in the clinical domain?
- What was the core essence in MYCIN?
- What is the principle of GAMUTS?
- Which two types of decisions do clinicians execute?
- What is the goal of network medicine?
- What is a true PPI topology?
- Why are structural homologies so important?
- What is the vision of personalized medicine?
- What does robustness in the context of complex biological systems mean?
- How do you get point cloud data from a natural image?
- Why is graphical model learning so interesting for medical problems?

Appendix

Key Problems

- Automated Machine Learning algorithms need much training data – focus is on adjusting model parameters without fully understanding the data that the learning algorithm is modeling [1]
- Weakly structured data [4]


Recommended Books

- Bayesian reasoning and machine learning, Cambridge University Press.
- Probabilistic graphical models: principles and techniques, MIT press.

Slide 5-19: Watershed Principle

- Catchment basins:
  - treating an image as a height field or landscape, regions where the rain would flow into the same lake
  - Start flooding from local minima, and label ridges wherever differently evolving components meet

Slide 5-15 Graphs from Images: Voronoi ↔ Delauney

- (a) Initial set of points.
- (i) Delaunay Triangulation Graph.
Example: Cell based therapy (1) (Heart transplantation)

Ventricular muscle
Dilated ventricle architecture
Conduction system of the heart
Pacemaker cell

Das et al. 2011

Chen et al. 2003


Example: Cell based therapy (2) (Heart transplantation)

Ventricular muscle
Dilated ventricle architecture
Conduction system of the heart
Pacemaker cell

Das et al. 2011

Chen et al. 2003


Example: Network Generated by Gene Duplication

High Modularity
(Modularity = 0.472, Seeded Modularity = 0.29)

Different colors represent different modules identified by Girvan and Newman algorithm.


Wing & Zhang (2007)

Genome-Phenome association in complex diseases

Pleiotropic effects


Additional Reading

Causality

Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference

Graphical Models for Genetics, Genomics, and Postgenomics