MACHINE LEARNING TECHNIQUES FOR STRUCTURED AND UNSTRUCTURED DATA FUSION

LJK - PhD students half day

Thibaud Rahier, first year PhD student

December 17, 2015

PHD supervisors:
Florence Forbes and Stephane Girard (MISTIS team at INRIA)
Sylvain Marié (Analytics for solutions team at Schneider Electric)

OUTLINE

Research project presentation

Overview of Probabilistic Graphical Models

Next steps

RESEARCH PROJECT PRESENTATION

Structured data

- Time series from housing data (power consumption, temperature...)
- Metadata surrounding these time series (location and type of measurement device, category of physical quantity measured...)

Unstructured data

- Diverse web information sources (tweets, weather, satellite pictures...)
- Expert (human) knowledge on the field

Industrial applications:

Regression

- Virtual sensing
- Forecasting
- Key performance indicator (KPI) prediction

Classification

- Fault detection (one class classification), diagnosis (multiclass classification)
- Predictive maintenance, predictive diagnosis

Clustering

- e.g. Typical days profiling

CURRENT EXPLORATION: PROBABILISTIC GRAPHICAL MODELS

Technical advantages of PGMs

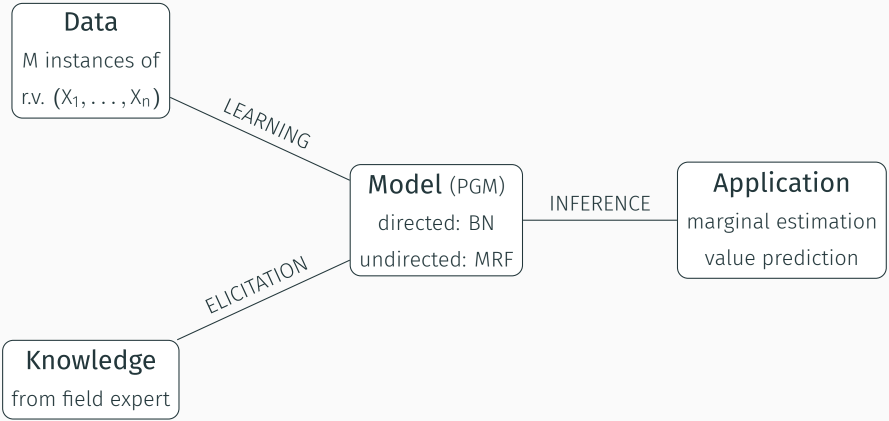
- Broad category of models (Hidden Markov Models, Markov Random Fields, Bayesian Networks, Deep Belief Networks...)
- Easy handling of missing data / heterogeneous data (metadata)
- Several ways to incorporate human knowledge
- Generative models: same model can be used for diverse applications

A very active research field

- Recent reference books:
  - Probabilistic Graphical Models (Koller and Friedman, 2012)
  - A Probabilistic Approach on Machine Learning (Murphy, 2012)
- Benchmarks keep getting better in several recent papers
- Programming libraries are currently being developed

OVERVIEW OF PROBABILISTIC GRAPHICAL MODELS

THEORETICAL FRAMEWORK AND VOCABULARY



Inference

- Marginal inference: For $i \in \{1, \ldots, n\}$, evaluate $P(X_i)$.
- MAP inference: For $i \in \{1, \ldots, n\}$, evaluate $\text{argmax}_{x_i \in Val(X_i)} P(x_i)$

Learning

- Parameter learning (for a given structure): maximize likelihood
- Structure learning:
  - structure scoring: likelihood-based
  - structure exploration (optimisation in super-exponential space)

Knowledge incorporation

- Priors in Bayesian approach for both learning tasks
- Starting points proposals for heuristic algorithms (particularly interesting in super-exponential structure spaces)

NEXT STEPS

Programming:

Define a simple application to be run on a real dataset

Bibliographic exploration:

Sum-product networks (sometime inference is possible in linear time), cutting-edge variationnal methods

Short term goal:

Build a PGM that incorporates expert knowledge and metadata, and that competes with Schneider Electric's current benchmark algorithms (SVM...) on a given problem (KPIs prediction?)

Thank you

Two main categories of graphical models:

Directed (Bayesian networks) Joint distribution decomposes in a product of local conditional distributions:

$$P(X_1, \dots, X_n) = \prod_{i=1}^{n} P(X_i | Pa_{X_i})$$

Ex: Dynamic Bayesian networks, Kalman filters, Hidden Markov models

Undirected (Markov random fields) Joint distribution decomposes in a product of local potentials:

$$P(X_1, \dots, X_n) = \frac{1}{Z} \prod_{c} \phi_c(X_c)$$

Ex: Ising model, Deep Belief Networks

We often use the Gibbs representation: for each clique c, $\phi_c(x_c) = \exp(-E_c(x_c))$ where E_c is the energy function of clique c

What we want to infer

- Marginal inference. For $i \in \{1, \ldots, n\}$, evaluate

$$P(X_i)$$
$$P(X_i \mid E = e)$$

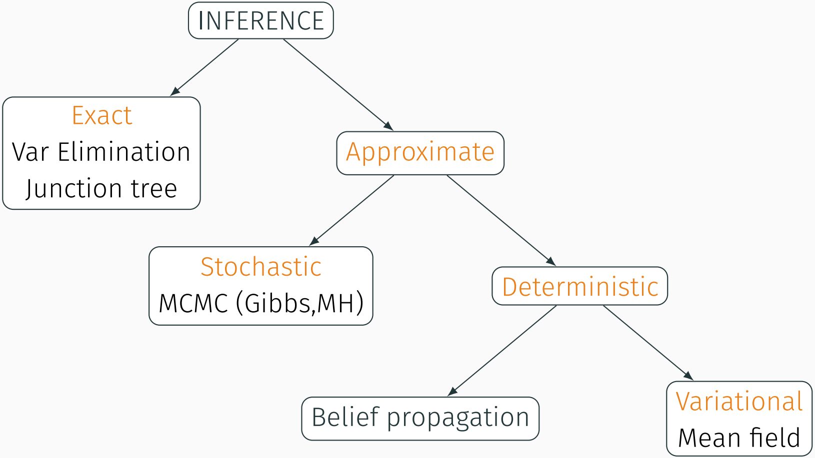
- MAP inference For $i \in \{1, \ldots, n\}$, evaluate

$$\operatorname{argmax}_{x_i \in Val(X_i)} P(x_i)$$
$$\operatorname{argmax}_{x_i \in Val(X_i)} P(x_i \mid E = e)$$

Complexity

For both directed and undirected graphs: NP-Hard problem in general: number of computations is exponential in number of variables

Different inference algorithms:



Log-likelihood function

- Bayesian networks

$$l(\theta) = \sum_{m=1}^{M} \sum_{i=1}^{n} \log \left(P(X_i = x_i[m] \mid Pa_{X_i} = \mathbf{u}_i[m]) \right)$$

- Markov random fields

$$l(\theta) = \sum_{m=1}^{M} \sum_{c} \log \left(\phi_c(x_c[m]) \right) - M \log(Z(\theta))$$

LEARNING (2/2)

Table: Maximum likelihood estimation in 4 main situations

	Complete data unimodal likelihood	**Incomplete data** multimodal likelihood
BNs	sufficient statistics tractable optimization in closed form	specific algorithms (Expectation-Maximization) inference
MRFs	no closed-form optimization (partition function) inference	both challenges part. funct. & missing data inference

Priors in Bayesian approach

for both learning tasks (θ is the vector of parameters, $l(\theta)$ the corresponding log-likelihood, D the data and G the graph structure)

- Parameter learning: maximum a posteriori:

$$\hat{\theta}^{MAP} = \mathrm{argmax}_\theta \; P(\theta|D) = \mathrm{argmax}_\theta \; l(\theta) + \log(P(\theta))$$

- Structure learning (Bayesian score):

$$\mathrm{score}_B(G : D) = \log(\underbrace{P(D|G)}_{\int P(D|\theta, G)P(\theta|G)d\theta}) + \log(P(G))$$

Starting points in optimization algorithms

Starting point proposals for heuristic algorithms (particularly interesting in super-exponential structure spaces) can be very useful