Quantitative Modelling of Adverse Outcome Pathway for Risk Assessment

WANG GAO (INERIS UTC)

SOUS DIRECTION DE

GHISLAINE GAYRAUD (LMAC UTC)
FRÉDÉRIC Y. BOIS (INERIS)
Overview

I - Introduction
  ◦ I.1 Toxicology in general
  ◦ I.2 Project Eu-ToxRisk21
  ◦ I.3 AOP and qAOP
  ◦ I.4 Objectives of my thesis

II - Probabilistic Model
  ◦ II.1 BN : Bayesian Networks
  ◦ II.2 DBN : Dynamic Bayesian Networks
  ◦ II.3 Model family for qAOP

III - Learning problem

IV - Conclusion
I. Introduction
I.1. Toxicology in general

- Toxicology
  - Application domains: Cosmetics, Drug, etc.

- Objective: Risk assessment

- Approaches (Different types of routines)
  - *In vivo* test: Animal experiments
  - Alternative methods
    - *In vitro* test (Classic) - fast
    - *In vitro* test (advanced): High throughput screening - very fast
    - *In silico* test: computational modelling and simulation - we are working on this
High-throughput screening

Robotic tools

384, 1536, 3456 wells

- Fast
- Low-cost
I.2. Project Eu-ToxRisk21


**Organisation:**
- Long range 2016-2020 research program
- 41 international research teams (40 EU + 1 USA) from 13 countries
- 14 Work packages with sub-objectives
  - WP 10: Computational Modelling for Risk assessment

**Motivation:**
- Partially replace animal experimentations
- Improve the predictive methods based on high-throughput toxicity tests
15 Case studies

Liver
Kidney
Lung
Nervous system
Etc.

15 Cases are studied in the program Eu-ToxRisk21
I.3. AOP and qAOP

- **EU-ToxRisk21**
  - An Integrated European ‘Flagship’ Programme Driving *Mechanism-based* Toxicity Testing and *Risk Assessment* for the 21st century

- **AOP**: Adverse outcome pathway – qualitative tool for *mechanism description*

- **qAOP**: quantitative AOP for *risk assessment*

**Remark**: Each case study corresponds to a set of AOPs.
AOP and qAOP

AOP: Adverse outcome pathway

- Knowledge Exchange
- A qualitative tool for the structural representation of causal relationships (Dose-response)
  - Starting from a molecular perturbation (MIE)
  - Through key events (KE)
  - Arriving at an adverse outcome (AO)
- In this case: AO = Steatosis (fat liver)
- qAOP = quantitative AOP:
  - Probabilistic model
  - Predictive capacity: Decision Support Tool
A richer version of AOP: AOP Network

- Each node may have more than one parent.
- One node can be involved in different AOPs

9 AOPs are hidden here
1.4. Objectives of my thesis

Objectives:

❖ Probabilistic modelling of AOP = qAOP
❖ => Prediction of AO

Action plan:

❖ Fixed structure for mini AOP
  ❖ Modelling of the strength cause-effect
  ❖ Prediction
❖ Structural learning: more ambitious
II. Probabilistic Model
II.1. BN : Bayesian Networks

Def : BN : \((\mathcal{G}, P_{\mathcal{G}})\) with

- \(\mathcal{G} : \) BN Structure : DAG directed acyclic graph \(\mathcal{G}(V, E)\)
  - \(V : \) Set of vertices, nodes : variables \(V = \{X_i| i \in 1:N\}\)
  - \(E : \) Set of directed edges : causality relationships
- \(P_{\mathcal{G}} : \) multivariate distribution over \(V\)
  - specified as set of local conditional probability distribution (CPDs) associated with \(\mathcal{G}\)’s nodes.

\[
P_{\mathcal{G}}(X) = \prod_{i=1:N} P_{\mathcal{G}}(X_i | \text{Par}(X_i))
\]

with \(X = (X_1, ...X_i, ...X_N)\) system stat vector
BN : Bayesian Networks

\[ P_G(X) = \prod_{i=1:N} P_G(X_i | \text{Par}(X_i)) \]

Local conditional distributions (CPD) and the DAG completely determine the joint distribution

Example: steatosis case

\[ P(X, L, C, F, T, S) = P(X)P(L|X)P(C|L)P(F|L)P(T|C, F)P(S|T) \]

Problem in the context of qAOP: Child node = function (Parent node(s), time)
II. 2. DBN : Dynamic Bayesian Networks

Extension of BNs to handle temporal models

Assumptions:
- the timeline discretised into a set of time slices:
  \[ X(t_0), X(t_1), \ldots, X(t_m), \text{ with } m \text{ the number of observations} \]
- Markov assumption for a dynamic system over the template variables \( X : \forall i \in \mathbb{N}^+ \):
  \[ X(t_{i+1}) \perp X(t_0: (t_{i-1})) | X(t_i) \]

Two types of dependency:
- Inter time-slice dependency (between time-slices)
- Intra time-slice dependency (in the same time-slice)
DBN: Dynamic Bayesian Networks

Markov assumption:

\[ X(t_{i+1}) \perp X(t_0:(t_i-1)) \mid X(t_i) \]

Compact definition of the joint probability distribution in DBN

\[ P(X(t_0:t_m)) = P(X(t_0)) \prod_{i=1}^{m} P(X(t_i) \mid X(t_{i-1})) \]

Example: Linear Dynamic System

Classic linear DBN

\[ \mathbb{E}[C_{t_i}] = \alpha + \beta_{prev} C_{t_{i-1}} + \beta_{curr} P_{t_i} \]

AIGM - TOULOUSE - 14 DEC. 2017
Insights about qAOP

Data visualization DEMO: Real data for kidney disease case study

Dynamic 3D plot GSH-DCF-Time (online) (local)

❖ $S_C$: Stationary state (Saturation level) of child node, denote:
❖ $S_C = f(P_t)$: Dependence of $S_C$ on $P_t$
❖ $C_t$: Child node activity at time $t$
❖ $C_t$ converges to $S_C$ over time
II.3. Model family for qAOP: Embryonic form

\[ S_C[P_t] - \mathbb{E}[C_t] = (S_C[P_t] - C_{t-h})e^{-\nu h} \]

- \( C_t \): Child node activity at time \( t \), (observed)
- \( P_t \): Parent node(s) activity at time \( t \), I could be a vector (observed)
- \( S_C[P_t] \): stationary state of child node given its parent(s) (unobserved)

Questions:

\[ S_C[P_t] = ? \]

\( h \in \mathbb{R}^+ \): non regular observation?

\( \nu > 0 \): to ensure the convergence of \( C_t \) towards \( S_C(P_t) \)
Model family for qAOP: Linear model

\[
\begin{cases}
S_c[P_t] - \mathbb{E}[C_t] = (S_c[P_t] - C_{t-h})e^{-\nu h} \\
S_c[P_t] = \beta P_t + \beta_0 \\
h \in \mathbb{R}^+
\end{cases} \quad (M_L)
\]

**Assumption**: Stationary stat of Child node is a **linear function** of parent node(s)

**Remark**: The classic linear DBN model is a special case of \((M_L)\)

- \(S_c[P_t] = \frac{\beta_{cur} P_t}{1-\beta_{prev}} + \frac{\alpha}{1-\beta_{prev}}\)
- \(h = 1\)
- \(e^{-\nu h} = e^{-\nu} = \beta_{prev}\)

\[
\mathbb{E}[C_{ti}] = \alpha + \beta_{prev} C_{ti-1} + \beta_{cur} P_{ti}
\]
Model family for qAOP : Sigmoid model

\[
\begin{align*}
S_C[P_t] - \mathbb{E}[C_t] &= (S_C[P_t] - C_{t-h})e^{-\nu h} \\
S_C[P_t] &= \alpha_{\max} \text{Sigmoid}[\beta P_t + \beta_0]  \\
&\quad \quad h \in \mathbb{R}^+ \\
\end{align*}
\]

With
- Sigmoid\([x]\) = \frac{1}{1+e^{-x}} - \frac{1}{2}
- Odd
- Bounded \([-1,1]\)

Assumption : The stationary state of Child node is a **sigmoid function** of parent node(s)
III. Inference
III.1. Input – Steatosis (real)

Real public domain databases (available before my thesis)

- ToxCast
- TG-Gate

DEMO : Steatosis real data (use [online DEMO link](#))

Problems :

The experiments are not designed for qAOP modelling.

- Very few number of data : 3 endpoints
- Discretization assumption may fail because the measurements of the system state taken at intervals that are regularly spaced with a predetermined time granularity $\Delta t$
- observations on time $t = 2h, 8h$ and $24h$
Input – Steatosis (simulated)

Data simulated from pharmacokinetics models based on ODE

Three virtual experimentation conditions:

- One injection of chemical $X$, without decreasing feature $\forall i \in I$, $C_X^{(t_i)} = C_X^{(0)}$

- One injection of chemical $X$, with decreasing feature. $\forall i \in I$, $C_X^{(t_i)} < C_X^{(t_{i-1})}$

- Four injection of chemical $X$, with decreasing feature : $\frac{C_X^{(0)}}{4}$ each time

online DEMO : Steatosis Generated data
Input – Kidney disease (real)

Data of kidney disease from Eu-ToxRisk21
- 5 doses
- 8 replicates
- 103 observations endpoints
  - measured on time $t_i = 1 + 15 \times (i - 1)$ minutes $\forall i \in 1:10$
III.2. Parameters learning: linear model

\[
\begin{align*}
S_C[P_t] - \mathbb{E}[C_t] &= (S_C[P_t] - C_{t-h})e^{-vh} \\
S_C[P_t] &= \beta P_t + \beta_0 \\
&\quad \quad (M_L)
\end{align*}
\]

Parameter estimation based on observations

Frequentist approach

\[\hat{\theta}_{ML} = \arg\max_{\theta \in \Theta} L(\theta | \mathcal{D}) = \arg\max_{\theta \in \Theta} P(\mathcal{D} | \theta)\]

Bayesian approach

\[\hat{\theta}_{MAP} = \arg\max_{\theta \in \Theta} P(\mathcal{D} | \theta) \pi(\theta)\]

- Algorithm MCMC under the probabilistic programming language “stan”
III.3. Implementation: Simulated data

Parameter learning for a local conditional model:

\[ P(FAS \mid LXR, t) \]

Result:

❖ Sigmoid model works significantly better with small set of data
❖ The linear model works as well as Sigmoid with large set of data
Implementation: Kidney disease (real)
IV. Conclusion
Conclusion

❖ Proposition of qAOP model family: Linear model, Sigmoid model

❖ Application on Steatosis (real), steatosis (simulated), kidney disease (real).
   ❖ Steatosis (real): Not fit well, not enough data.
   ❖ Steatosis (simulated):
     ❖ The sigmoid model fits better when only few data are available
     ❖ The Linear model fits as well as the sigmoid model when more data are available
   ❖ Kidney disease (real):
     ❖ Linear models can well fit the database.

Future steps:

❖ Test sigmoid model on Kidney disease data
❖ Test model performance (behaviour) on simulated data with more nodes in AOP
❖ qAOP network problems: hierarchical DBN.
Reference


Thank you for your attention