Knowledge Discovery for
Clinical Decision Support

Pedro Pereira Rodrigues

CINTESIS & LIAAD – INESC TEC
Faculty of Medicine – University of Porto, Portugal
@ECMLPKDD – Nancy, France
Who am I?

A privileged one, who being educated in machine learning, gets to teach medical students on research methodology and data science ;-)

- MSc (2005) and PhD (2010) on clustering data streams and stream sources.
- Last 6 years involved in medical informatics, clinical research and medical education.

Coordinator of the **BioData - Biostatistics and Intelligent Data Analysis** group of **CINTESIS - Centre for Health Technologies and Services Research** (100+ PhD research unit to start officially in 2015) and collaborator in **LIAAD – INESC TEC** (original research unit since 2003).
Agenda IV

- Uncertainty and evidence-based medicine
- Data science in the EBM loop
- Biostatistics and probabilistic decision support
- Bayesian networks as formalization of uncertainty for decision support
- Toy and real-world examples of Bayesian nets for clinical decision support
- Lessons learned
Uncertainty and Evidence Based Medicine
Uncertainty in clinical decision analysis

- The consequences of a medical decision are uncertain by the time of decision.
- Clinical exam and diagnostic tests are imperfect.
- Therapeutic actions, as well as their risks and benefits, might be vaguely defined or even unknown.
- For a large group of clinical problems, there is no information about clinical trials, or it simply isn't generalizable for the patient.

Evidence-based medicine

Conscient, explicit and criterious use of the best available evidence in clinical decision:

- personal clinical experience;
- best external clinical evidence from quality clinical research;
- values, needs, expectations and individual context of each patient.

Sackett D. et al. (1996)

Evidence based medicine: what it is and what it isn’t

BMJ 312:71-2
Take away message

**M1:** During inference and decision support, uncertainty needs to be reduced.

**S1:** Better focus on the variables that reduce uncertainty the most (e.g. when suggesting a test).
Evidence-based medicine

Conscient, explicit and criterious use of the best available evidence in clinical decision:

- personal clinical experience;
- best external clinical evidence from quality clinical research;
- values, needs, expectations and individual context of each patient.

Sackett D. et al. (1996)

*Evidence based medicine: what it is and what it isn’t*

BMJ 312:71-2
Evidence-based medicine

Conscient, explicit and criterious use of the best available evidence in clinical decision:

- personal clinical experience;
- best external clinical evidence from quality clinical research;
- values, needs, expectations and individual context of each patient.

Sackett D. et al. (1996)

*Evidence based medicine: what it is and what it isn’t*

BMJ 312:71-2
Evidence-based medicine

Conscient, explicit and criterious use of the best available evidence in clinical decision:

- personal clinical experience;
- best external clinical evidence from quality clinical research;
- values, needs, expectations and individual context of each patient.

Sackett D. et al. (1996)

Evidence based medicine: what it is and what it isn’t

BMJ 312:71-2
Clinical information/knowledge/decision

practice

information

generates
Clinical information/knowledge/decision

practice

information

used in

research

generates
Clinical information/knowledge/decision

- Information used in research
- Knowledge generated
- Practice generates information
- Information used in research
- Research generates knowledge
Clinical information/knowledge/decision
Clinical information/knowledge/decision

- Information
- Research
- Knowledge
- Patient
- Decision
- Practice

Information used in research generates knowledge, which is then applied to patient. The patient generates decision, which generates information used in practice.
Clinical information/knowledge/decision

- Information used in research
- Research generates knowledge
- Knowledge applied to patient
- Patient makes decision
- Decision used in practice
- Practice generates information
Where is data science involved?
Clinical information/knowledge/decision

Data Management

- Information
  - Generates
  - Used in
    - Practice
    - Decision
    - Knowledge
    - Research

- Knowledge
  - Generates
  - Applied to
    - Patient

- Decision
  - Generates
Clinical information/knowledge/decision

Knowledge discovery

- Information used in research
- Generates knowledge
- Applies to patient
- Generates decision
- Used in practice

Pedro Pereira Rodrigues - Medical Mining Tutorial
September 2014
Clinical information/knowledge/decision

- Practice
- Information
- Decision
- Research
- Patient
- Representation

Information used in research generates knowledge and is applied to patient. Knowledge generates decision, which is used in practice, and generates information.
Clinical information/knowledge/decision

- Information used in research
- Research generates knowledge
- Knowledge applied to patient
- Patient decision
- Decision used in practice
- Practice generates information
Clinical information/knowledge/decision

- Information used in research
- Research generates knowledge
- Knowledge applied to patient
- Patient generates recommendation
- Recommendation used in decision
- Decision used in practice
- Practice generates information
Clinical information/knowledge/decision

- Information used in research
- Generates knowledge
- Applied to patient
- Generates decision
- Used in practice

Clinical decision support systems
Clinical information/knowledge/decision

REDUCE UNCERTAINTY

- Information used in research generates knowledge.
- Knowledge applied to decision generates patient.
- Decision used in practice generates information.
Clinical information/knowledge/decision

- Information used in research
- Research generates knowledge
- Knowledge applied to patient
- Patient generates decision
- Decision used in practice
- Practice generates information

FORMALIZE UNCERTAINTY

REDUCE UNCERTAINTY
Uncertainty and probability

We use terms such as **frequent**, **possible** or **rare** to express uncertainty.

**Probability** is a numeric expression of the likelihood that an event will occur.

We can then use probability to **express uncertainty without ambiguity**...

... and compute the effect of new information in the probability of disease, using the Bayes theorem.
Knowledge modeling for decision support
Risk and predictive factors

To support clinical decisions, we need to define:

**Outcome** - result variable (diagnosis, prognosis, treatment, etc.)

**Factors** - associated with the outcome (clinical history, demographic, etc.)

- Risk (of developing the disease or worse prognosis)
- Prediction (useful to predict but not necessarily of risk)

**Association** between factors and outcome

**Prevalence/Incidence**

\[ P = \frac{a+c}{n} \]

**Risk ratio**

\[ RR = \frac{a}{a+b} / \frac{c}{c+d} = \frac{a(c+d)}{c(a+b)} \]

**Odds ratio**

\[ OR = \frac{a}{b} / \frac{c}{d} = \frac{ad}{bc} \]

**Sensitivity and specificity of factor as predictor of outcome**

\[ Sens = \frac{a}{a+c}, \ Spec = \frac{d}{b+d} \]

---

**Prevalence/Incidence**

\[ P = \frac{a+c}{n} \]

**Risk ratio**

\[ RR = \frac{a}{a+b} / \frac{c}{c+d} = \frac{a(c+d)}{c(a+b)} \]

These can all be interpreted as (ratios of) conditional probabilities…

**Odds ratio**

\[ OR = \frac{exposition\ odds\ (cases)}{exposition\ odds\ (controls)} = \frac{a/c}{b/d} = \frac{ad}{bc} \]

**Sensitivity and specificity of factor as predictor of outcome**

\[ Sens = \frac{a}{a+c}, \ Spec = \frac{d}{b+d} \]

Clinical decision support

Evidence-based medicine relies on these simple, yet powerful, statistical measures as means for evidence assessment, yielding:

- Easy computation
- Formal representation of uncertainty (probability-based)
- Human-interpretable evidence

(e.g. RR > 1 means increased risk for exposed individuals compared to non-exposed ones)
“The complicated nature of real-world biomedical data has made it necessary to look beyond traditional biostatistics.”

“Bayesian statistical methods allow taking into account prior knowledge when analyzing data, turning the data analysis a process of updating that prior knowledge with biomedical and health-care evidence.”


“Bayesian networks offer a general and versatile approach to capturing and reasoning with uncertainty in medicine and health care.”

Peter Lucas et al. (2004) *Artificial Intelligence In Medicine*

Bayesian networks

**Graph** representation where:

- the attributes are represented by the graph **nodes**, and
- the **arcs** represent dependencies among attributes,
  using **conditional probabilities**.

Easily human-interpretable representation, since it uses a

**probabilistic reasoning similar to the usual uncertainty in human reasoning.**

---


Bayesian networks for clinical decision support

Bayesian networks *intrinsic uncertainty modeling* yields:

- Qualitative interpretation of *associations*
- Formal representation of *uncertainty* (probability-based)
- Human-interpretable *evidence* (a priori risk, a posteriori risk, relative risk, ...)
- Similar to traditional *biostatistics* (remember how measures are based on probabilities?)
- Decision support even with *unobserved variables*. 
Bayesian networks for clinical decision support

Complex research questions can be addressed by the same model:

**Etiology and risk**
Can a visit to China be the cause of patient's SARS?
Can a visit to China (and corresponding acquired SARS) be the cause of patient's dyspnea?

**Diagnosis**
The patient visited China; does he have SARS?
The patient has a high temperature reading; is it SARS?

**Prognosis**
The patient has fever and has visited China; without treatment, is he going to develop dyspnea?
Bayesian networks for clinical decision support

Sample of real examples
Bayesian networks for clinical decision support

2000

24h-prognosis of head-injured ICU patients

2005

Diagnosis of

ventilator-associated pneumonia

Bayesian networks for clinical decision support

Predicting maintenance fluid requirement in ICU

Bayesian networks for clinical decision support

2013

Breast cancer diagnosis

Bayesian networks for clinical decision support

2014

Prognosis of quality of life after ICU stay

Bayesian networks for clinical decision support

Obstructive sleep apnea diagnosis

Temporal modeling of preeclampsia diagnosis

Take away message

**M1:** During inference and decision support, uncertainty needs to be reduced.

**S1:** Better focus on the variables that reduce uncertainty the most (e.g. when suggesting a test).

**M2:** Bayesian models (e.g. networks) are intrinsically modeling uncertainty and can map biostatistics.

**S2:** Consider Bayesian networks (or other probabilistic methods) as models to support clinical decision.
Uncertainty in Modeling

A toy example
Uncertainty in modeling

- You have access to a data set obtained from a cohort of suspected SARS patients, with one of the available variables being “Fever”.
- You learn from your data that “Fever” is associated with SARS.
- Based on expert-knowledge you turn the association into causation.
Uncertainty in modeling

- But the problem lingers:
  - what does “Fever” mean?
  - is it really observed?
- Although unlikely, you may have a reading of less than 37.5º and still have fever (e.g. if controlled with ibuprofen) or a reading of more than 37.5º without actually having fever.
- So, we should not reduce that uncertainty during modeling, rather include it in the model:

```
SARS → Fever → >37.5º
```
Take away message

**M1:** During inference and decision support, uncertainty needs to be reduced.

**S1:** Better focus on the variables that reduce uncertainty the most (e.g. when suggesting a test).

**M2:** Bayesian models (e.g. networks) are intrinsically modeling uncertainty and can map biostatistics.

**S2:** Consider Bayesian networks (or other probabilistic methods) as models to support clinical decision.

**M3:** If what you observe is what you record, it should also be what you model.

**S3:** Better search for the actual meaning (e.g. model temp above 37.5 instead of / along with fever).
Uncertainty in Modeling

A simple but real example
Uncertainty in modeling with expert knowledge

- There are cases where the knowledge discovery process needs to be merged with expert-based modeling and associations gathered from traditional meta-analysis.

- Imagine modeling the association between pneumonia and HIV infeccion, using a Bayesian net.

- The MD presents you a meta-analysis where this association is assessed and confirmed.

- So you can even use the meta-analysis risk assessment to compute the conditional probabilities of your Bayesian net (expert knowledge).
Uncertainty in modeling with expert knowledge

- You now have access to a database and, after the knowledge discovery process, it reveals the same association, so you consider merging the two data sources.

- But the variable HIV in your data is, in fact, given by the application of a standard test (for illustrative purposes, let's consider PCR with 98% sensitivity and 99% specificity).

- So what you end up learning is the association between pneumonia and a positive PCR test result, which is an uncertain expression of HIV (precision may be below 10% for low disease prevalences)...
Uncertainty in modeling with expert knowledge

- But you have information on the association between the standard test and HIV infection...
  (remember that PCR has 98% sensitivity and 99% specificity)

- So the model seems a bit more accurate now...
Uncertainty in modeling with expert knowledge

- But you have information on the association between the standard test and HIV infeccion... (remember that PCR has 98% sensitivity and 99% specificity)

- So the model seems a bit more accurate now...

HIV → PCR

Expert knowledge

PCR → Pneu

HIV → PCR → Pneu
Uncertainty in modeling with expert knowledge

- But you have information on the association between the standard test and HIV infection...
  (remember that PCR has 98% sensitivity and 99% specificity)

- So the model seems a bit more accurate now...

Discovered knowledge
Uncertainty in modeling with expert knowledge

But your expert opinion tells you that is not the PCR test that is associated with pneumonia; it's the HIV infection, so it should look like this, instead:
Uncertainty in modeling with expert knowledge

But your expert opinion tells you that is not the PCR test that is associated with pneumonia; it's the HIV infection, so it should look like this, instead:

If what you observe is what you record, it should also be what you model.
Take away message

M1: During inference and decision support, uncertainty needs to be reduced.

S1: Better focus on the variables that reduce uncertainty the most (e.g. when suggesting a test).

M2: Bayesian models (e.g. networks) are intrinsically modeling uncertainty and can map biostatistics.

S2: Consider Bayesian networks (or other probabilistic methods) as models to support clinical decision.

M3: If what you observe is what you record, it should also be what you model.

S3: Better search for the actual meaning (e.g. model temp above 37.5 instead of / along with fever).

M4: During modeling and knowledge discovery, uncertainty needs to be formalized, not ignored.

S4: Better not dismiss variables' association that include uncertainty (e.g. do not assume PCR=HIV)
Thank you!
Acknowledgements

Cristina Granja
Altamiro Costa-Pereira
Cláudia Camila Dias
Liliana Leite
Cristina Costa-Santos
João Gama
References


References (examples of Bayesian network applications)


