Third Generation
Machine Intelligence

Chris Bishop
Microsoft Research Cambridge

CIP 2010, Elba
The data-driven revolution

Stand-alone → cloud plus client
Applications → services
Isolated data → fusion of diverse data sources
Hand-crafted → solutions learned from data
Exponential growth in stored data

280 Exabytes of data in 2008
Doubling every 18 months
First Generation: 1960s – 1980s

Expert systems

rules obtained from human experts

Within a generation ... the problem of creating ‘artificial intelligence’ will largely be solved

Marvin Minsky (1967)

Combinatorial explosion

General theme: hand-crafted rules
Second Generation: 1990s - present

Neural networks, support vector machines

Difficult to include complex domain knowledge

General theme: black-box statistical models
Why is prior knowledge important?
Third Generation Machine Intelligence

**Question**: how can we integrate domain knowledge with statistical learning?

3 key ideas ...
1. Bayesian learning (in pictures)

Use probabilities to quantify uncertainty

Predictions made by weighted averaging
2. Probabilistic graphical models

Maths (M)

Geometry (G)  Algebra (A)

\[ P(M, G, A) = P(M) \cdot P(G|M) \cdot P(A|M) \]

Graph structure captures domain knowledge
Special cases

Kalman filters & smoothers, hidden Markov models, Markov random fields, conditional random fields

PCA, ICA, factor analysis, linear regression, logistic regression, mixture models
3. Efficient inference

\[
\sum_{x} \sum_{y} xy = x_1 y_1 + x_2 y_1 + x_1 y_2 + x_2 y_2 \\
= (x_1 + x_2)(y_1 + y_2)
\]
Local message-passing

E.g. Kalman filtering, and smoothing, equations are simply special cases of this message-passing framework.
What if distributions are intractable?

True distribution

Monte Carlo

Variational Bayes

Loopy belief propagation

Expectation propagation
Tutorial example: drug trial

Set of participants with a disease
Each given either drug or placebo
Some patients recover, the rest do not

**Question:** is the drug different from the placebo?
Different?

True

Prob. Cure (Drug)

Prob. Cure (Placebo)

False

Prob. Cure (All)

Prob. Cure (Drug)

Prob. Cure (Placebo)

Prob. Cure (All)
Case study 1: Bayesian ranking

Goal: global ranking from noisy partial rankings
Example: rank players using outcomes of games
Conventional approach: Elo (used in chess)
  maintains a single strength value for each player
  cannot handle team games, or more than 2 players
Two-player match outcome model
Multiple team model
TrueSkill™

Xbox 360 Live: launched September 2005
every 360 game uses TrueSkill™ to match players
20 million active users, >2.5 million matches per day
“Planet-scale” application of Bayesian methods
Convergence

![Graph showing convergence of players over the number of games.](image)

- **char (TrueSkill™)**
- **SQLWildman (TrueSkill™)**
- **char (Elo)**
- **SQLWildman (Elo)**

The graph illustrates the level progression of two players, char and SQLWildman, over the number of games played. The TrueSkill™ system, as represented by the blue lines, shows a more consistent and smooth convergence compared to the Elo system, represented by the red lines.
Case study 2: *adPredictor*
One weight per feature value

<table>
<thead>
<tr>
<th>Ad ID</th>
<th>Position</th>
<th>Match Type</th>
<th>ML-1</th>
<th>SB-1</th>
<th>SB-2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1341201</td>
<td>SB-1</td>
<td>Exact Match</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1570165</td>
<td>SB-2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2213187</td>
<td></td>
<td>Broad Match</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9215433</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Uncertainty: Bayesian probabilities

Ad ID
- 1341201
- 1570165
- 2213187
- 9215433

Match Type
- Exact Match
- Broad Match

Position
- ML-1
- SB-1
- SB-2

$p(p_{\text{Click}})$
Predicted click probabilities

Flight 153 from Nov 11 to Nov 17 2008

Predicted click probability

Measured click probability
Principled exploration

average: 25% (3 clicks out of 12 impressions)
average: 30% (30 clicks out of 100 impressions)
Case study 3: MAAS

Child (1527)
- Physiological (759)
  - Cotinine Level (4)
  - Height, Weight, BMI, Fat% (43)
  - Asthma (33)
  - Skin Tests (80)
  - IgE Tests (134)
  - Wheeze (99)
  - Eczema (69)
  - Rhinitis (37)
  - Lung (17)
  - Respiratory (88)
  - Nose (12)
  - Endotoxin (17)
  - Lung function (40)
  - Reactions (6)
  - Sickness e.g. cough, cold (80)
- Treatments (98)
  - Medications (51)
  - Immunizations (28)
  - Sensitizations (19)
- Lifestyle (218)
  - Nursery, Daycare (15)
  - Emotions (58)
  - Visits to Doc (28)
  - Grouping (12)
  - Diet (105)
- Genetic
  - 690k SNPs

Parents (39)
- Socioeconomic
- Asthma and other medical
- Smoking
- Skin Tests + Atopy

Siblings (32)
- Siblings’ medical condition
- Position among sibs,
  - Nursery

Birth (14)
- Height and Weight
- Delivery Type
- Gender
- Breast Feeding
- Grouping
  - Ethnic Group
  - Others

Environment (419)
- Pets (208)
- Home (211)

Others (25)
- Breast Feeding
- Grouping
- Ethnic Group
- Others

~ 1000 children, 2000 variables
Sensitization classes

Asthma Stats:

- 42.2%
- 9.59%
- 11.1%
- 1.66%
- 0.0%

Compare this to 22% using “atopy”

Acknowledgements

Iain Buchan (University of Manchester)
Joaquin Quiñero Candela
Adnan Custovic (University of Manchester)
Thore Graepel
John Guiver
David Heckerman
Ralf Herbrich
Tom Minka
Angela Simpson (University of Manchester)
Markus Svensén
Vincent Tan
John Winn
Summary

We are at the start of a data-driven revolution
New paradigm for machine intelligence built on:

1. a Bayesian formulation
2. probabilistic graphical models
3. fast inference using local message-passing

Infer.NET

Thank you!