Getting Real with Unreal Data: Lessons Learned and the Way Ahead

Thore Graepel
Royal Holloway, University of London
Which word did Lev add?

In this presentation I will give a general introduction to the problem of learning non-vectorial, "unreal", or "unpopular" data.
Outline

• Examples of Unreal Data from the World
• Really Embedding Data
  – Neural Networks
  – Kernel Methods
• Taking unreal Data seriously:
  – Inductive Logic Programming
• Symbolic Measurement Process
Nominal Attribute Vectors

- Simple, logical description
- Hypotheses: decision trees, DNFs, CNFs
- Combinatorial growth in number of attributes
- Hypercube embedding
Ordinal Attribute Vectors

- Example RAE results: Open University 2001
- Popular for questionnaires and psychological experiments

<table>
<thead>
<tr>
<th>Field</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Psychology</td>
<td>4</td>
</tr>
<tr>
<td>Biological Sciences</td>
<td>4</td>
</tr>
<tr>
<td>Chemistry</td>
<td>3</td>
</tr>
<tr>
<td>Physics</td>
<td>3</td>
</tr>
<tr>
<td>Earth Sciences</td>
<td>5</td>
</tr>
</tbody>
</table>
Real Attribute Vectors
Strings: DNA

Thore Graepel, Unreal Data, NIPS 2002
Strings: Text
Strings: Programmes

The Turing Machine

1. Formalism
2. Primitive Recursive
3. Feeling Context Free
4. Grammar of Love
5. I Can't Say Whether I Stop Loving You
6. Decision Problem
Trees: Parse Trees and XML

<!-- ELEMENT sentence (noun_phrase, verb_phrase)>
<!-- ELEMENT noun_phrase (article, noun)>
<!-- ELEMENT verb_phrase (verb, noun_phrase)>
<!-- ELEMENT article (#PCDATA)>
<!-- ELEMENT noun (#PCDATA)>
<!-- ELEMENT verb(#PCDATA)>

<sentence> <noun_phrase> <article> the <noun> girl </noun> </noun_phrase>
<verb_phrase> <verb> likes </verb> <noun_phrase> <article> the </article> <noun> ice cream </noun> </noun_phrase> </verb_phrase> </sentence>
Trees: The Tree of Life

Phylogenetic Tree

One species of Rana: a leaf of the Tree of Life
Subgroups of Rana
The genus Rana
The frog family Ranidae
Pages for subgroups of frogs
Frogs as a whole

Thore Graepel, Unreal Data, NIPS 2002
Graphs: Organic Molecules

- **Methane** $\text{CH}_3$
- **Ethane** $\text{C}_2\text{H}_6$
- **Ethene** $\text{C}_2\text{H}_4$
- **Benzene** $\text{C}_6\text{H}_6$

**ATP**

Energy Transport Molecule

Triphosphate $+$ Sugar $+$ Adenine

Thore Graepel, Unreal Data, NIPS 2002
Graphs: Go Positions
Really Embedding Data

- Most natural approach for NIPS people: Embed your unreal data in real space and apply an SVM (formerly: a neural network)
- Problem I: If possible, could require very high dimensionality for isometric embedding
- Problem II: Generalisation may be bad because compositional structure is neglected
- Problem III: Embedding is a many-to-one mapping: hard to create new class instances
Polyphonic Sequences: Music

Hendrik Purwins

Thore Graepel, Unreal Data, NIPS 2002
Embedding Music: Bach

Bach’s Well-Tempered Clavier II, Fugues, Recording: Glenn Gould
Embedding Music: Chopin

Chopin’s Preludes, Recording: Alfred Cortot

Thore Graepel, Unreal Data, NIPS 2002
Temporal Neural Networks

\[
\begin{align*}
O_1 &= \sum O_1(i)O_1(i) \\
O_2 &= \sum O_2(i)O_2(i) \\
O_N &= \sum O_N(i)O_N(i)
\end{align*}
\]

Input pattern in different time-steps

Thore Graepel, Unreal Data, NIPS 2002
Folding Neural Networks

folding network given by the networks:

compact notation:

encoding layer

context-neurons

recurrent-
g

feed-forward part

leads to the computation

computation induced by g

Barbara Hammer

Thore Graepel, Unreal Data, NIPS 2002
Kernel Methods

• Define kernel function $k(x,x')$ between objects $x$ and $x'$

• Mercer: if $k$ is positive definite, then there exists a feature map $\Phi$ s.t.
  $k(x,x') = \langle \Phi(x), \Phi(x') \rangle$

• Hence, finding a p.d. kernel function provides an isometric embedding in Euclidean space (kernel PCA, MDS)
String Kernels (Watkins 1998, Haussler)

• We can define a kernel between strings $u$ and $v$ by subsequence matching.

• Sum over all possible strings $b$ of length $s$ up to length $r$, weighted by $q^s$, for every co-occurrence of $b$ in the strings $u$ and $v$.

• Calculation can be done efficiently by recursion avoiding the calculations that involve all the potential features.
Example: String Kernel

\[
\begin{array}{cccccc}
U & G & A & T & T & A \\
V & A & B & R & A & C \\
\end{array}
\begin{array}{ccccccc}
& & & & & & C \\
& & & & & & A \\
A & D & A & B & R & A \\
\end{array}
\]

- Consider subsequences of length at most 3
- We have 15 matches for A, one for C, a match for CA and a match for ACA

\[k(u; v) = 15q^1 + q^1 + q^2 + q^3\]
String Kernels: Diagonal Dominance

Thore Graepel, Unreal Data, NIPS 2002
The Fisher Kernel

- Given a probabilistic model $P(x \mid w)$ of data $x$ parameterised by $w$
- Define Fisher score $u_i(x) := \log P(x; w) = @v_i$
- Define Fisher Information Matrix by
  $$I := E_x[u(x)u^T(x)]$$
- Define Fisher Kernel as
  $$k(x_i; x_j) := u(x_i)I^{-1}u^T(x_j)$$
Probabilistic Models

- Fisher kernel provides embedding for objects generated by a probabilistic model
- Example: Markov Model

![Markov Model Diagram]

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>C</th>
<th>G</th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>.2</td>
<td>.4</td>
<td>.3</td>
<td>.1</td>
</tr>
<tr>
<td>C</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Inductive Logic Programming

- Learning Method for data and rules represented in first-order predicate logic
- Learn PROLOG programmes from data and background knowledge
- Fully relational, syntactic approach based on Horn clauses
- Set-covering approaches, general-to-specific search
Consider the rules (horn clauses) for “x is uncle of y”

1. uncle(x,y) :- brother(x,z) parent(z,y)
2. uncle(x,y) :- husband(x,z) sister(z,w) parent(w,y)

• Let “uncle” be the target predicate
• Let “brother”, “sister”, “parent”, “husband” be background predicates
Consider an extended family:

1. uncle(tom, frank), uncle(bob, john)
2. ¬uncle(tom, cindy), ¬uncle(bob, tom)
3. parent(bob, frank), parent(cindy, frank), parent(alice, john), parent(tom, john)
4. brother(tom, cindy)
5. sister(cindy, tom)
6. husband(tom, alice), husband(bob, cindy)
ILP Example III

Tom → uncle → ¬uncle → ¬uncle → parent → husband

Bob → uncle → parent → husband

Frank → ¬uncle → parent

Cindy → parent

John → parent

Alice
ILP: Formal Framework

• **Let:** B, P, N and H be sets of Horn clauses

• **Given:**
  – Background knowledge B
  – Positive examples P
  – Negative examples N

• **Find:** complete and consistent hypothesis H
  – For all p in P: H ⊨ B implies p (completeness)
  – For all n in N: H ⊨ B does not imply n (consistency)
9/11 Data Mining by ILP (Mooney et al. 2002)

• “Contract Killing”: classify killings by motives “threat”, “obstacle”, and “rival”

• Facts as Predicates:
  \[\text{isa}(\text{Murder714}, \text{MurderForHire})\]
  \[\text{perpetrator}(\text{Murder714}, \text{Killer186})\]
  \[\text{victim}(\text{Murder714}, \text{MurderVictim996})\]
  \[\text{deviceTypeUsed}(\text{Murder714}, \text{Pistol}, \text{Czech})\]

• Rules as Hypothesis:
  \[\text{firstDegreeMurder}(A)\]
  \[\text{subEvents}(A, B)\]
  \[\text{performedBy}(B, C)\]
  \[\text{loves}(C, D)\]
Measurement

Definition of Measurement:

Measurement of some attribute of a set of things is the process of assigning numbers or other symbols to the things in such a way that relationships of the numbers or symbols reflect relationships of the attribute being measured. A particular way of assigning numbers or symbols to measure something is called a scale of measurement.

S. S. Stevens
# Scales of Measurement

<table>
<thead>
<tr>
<th>Scale</th>
<th>Perm. Trafo</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nominal</td>
<td>One-to-one</td>
<td>Assignment of numbers to Football players</td>
</tr>
<tr>
<td>Ordinal</td>
<td>Monotone increasing</td>
<td>Moh’s scale of hardness of minerals</td>
</tr>
<tr>
<td>Interval</td>
<td>Affine</td>
<td>Temperature in degree Fahrenheit</td>
</tr>
<tr>
<td>Log-Interval</td>
<td>Power</td>
<td>Fuel efficiency in miles/gallon</td>
</tr>
<tr>
<td>Ratio</td>
<td>Linear scaling</td>
<td>Temperature in degree Kelvin</td>
</tr>
<tr>
<td>Absolute</td>
<td>Identity</td>
<td>Number of children</td>
</tr>
</tbody>
</table>

Thore Graepel, Unreal Data, NIPS 2002
The Peano Axioms: Just Counting

1. There is a natural number $1$
2. Every natural number $a$ has a successor denoted by $a+1$
3. There is no natural number $a$ whose successor is $1$
4. Distinct natural numbers $a$ and $b$ have distinct successors $a+1$ and $b+1$
5. If a property is possessed by $1$ and also by the successor $a+1$ of every natural number $a$ it is possessed by, then it is possessed by all natural numbers.
Number: The Language of Science

Our instruments of detection and measurement, which we have been trained to regard as refined extensions of our senses, are they not like loaded dice, charged as they are with preconceived notions concerning the very things which we are seeking to determine? Is not our scientific knowledge a colossal, even though unconscious, attempt to counterfeit by number the … world disclosed to our senses?

Tobias Dantzig
God made the natural numbers, all the rest is the work of man

Kronecker
Symbolic Representation: ETS

- Replace the natural numbers by an inductive structure that evolves during measurement (ETS)
- Define a class by a class progenitor and a set of associated transformations
- Associate weights with each transformation such that class members can be constructed from the progenitor using low-weight transformations only

Lev Goldfarb
Generative Model of Molecule
Making Molecules

Class Progenitor:

Class Transformation:
Conclusions

• Most data are unreal!
• Often, we can visualise and classify unreal data by embedding them in real space
• Keep in mind the importance of the measurement process, and think about symbolic measurement
• Enjoy the remainder of the workshop!