Introduction & Course Outline

Ganesh Ramakrishnan[1]

[1]Department of Computer Science and Engineering, IITB, India

CS717, Statistical Relational Learning, 2008

Outline

Synthesis of two areas

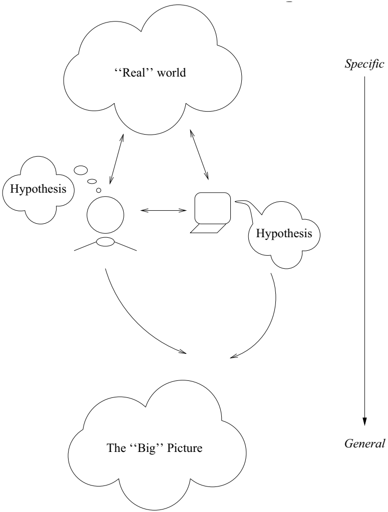
- Statistical Learning and Relational Representation and Reasoning
- Each has a fairly long history of its own

- **The course will attempt to provide a foundational overview of the field while also covering the recent research literature and various applications.**
- **Mathematical maturity, a basic course in statistics, and basic programming skills will be very useful; students without this background should discuss their preparation with me.**

Machines as “intelligent assistants”

- Science
  - Forming and refining new theories
- Engineering
  - Data analysis
  - Training professionals
- Arts
  - Artist/author-specific patterns
  - Suggesting playing method
- “Assistance” should be understandable

2-Way Learning



Traditional Statistical Learning

- Assumes that the data is represented by points in a high-dimensional space. Examples:
  1. Learning to detect a face in an image
  2. Classify an email message as spam or not
- Procedure:
  1. Construct the relevant low-level features (e.g., pixels. filters, words, URLs)
  2. Solve the problem using standard tools for the vector representation
- Flip side: This abstraction hides the rich logical structure of the underlying data, crucial for solving more general and complex problems

Going beyond Traditional SL: Motivation

- Not just answering an isolated yes/no question, but ..
- Producing and manipulating structured representations of the data
  - Detect the face in an image and recognize it as that of man/woman
  - Classify a bunch of emails as a spam or not
  - Web-page classification using hyper-link information

Relational Learning

- Motivation
  - Ability to model dependencies between related instances
  - Desirable to use information about one object to help reach conclusions about other, related objects
- Example: In web data [Chakrabarti et al, 1998]
  - One should be able to propagate information about the topic of a document to documents it has links to and documents that link to it
  - These, in turn. would propagate information to yet other documents
- Underlying intuitions of such *procedural systems* can be given *declarative semantics* using probabilistic graphical models

Relational Learning (contd)

- Early work: Learning deterministic logical concepts
- Earliest System: Winston's arch learning (online-style) system
  1. Trained using a sequence of instances labeled as positive and negative examples of arches
  2. System maintains a 'current' hypothesis, represented as a semantic network
  3. When a new example is presented, the system makes a prediction using the current hypothesis.
     - If the prediction is correct, no changes made to the hypothesis
     - Else, the set of differences between the current hypothesis and the example is identified.
     - If the example was a positive instance, the differences are used to *generalize* the concept
     - If the example is a negative instance, differences are used to *specialize* the concept

Inductive Logic Programming (ILP)

- Area with best track record in relational learning
  - Learning (deterministic) first-order rules from relational data
- Advantages
  1. Rules are human interpretable
  2. Can incorporate background knowledge
- Initial efforts focused on program synthesis from examples and background knowledge
- Recent research: Discovery of useful rules from larger databases
  - Rules often used for prediction
  - May have probabilistic interpretation

ILP: Areas of Application

Science. Protein shape prediction, drug structure activity prediction, . . .

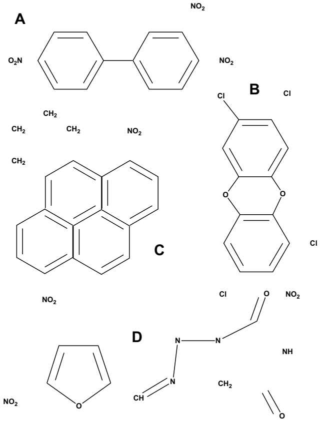
Engineering. Finite element mesh design, satellite fault diagnosis, circuit design, automobile traffic-flow analysis, intelligent software agents for the Internet, reconstruction of simulator models, numerical prediction, . . .

Language. Part-of-speech disambiguation from large real-world corpora, learning grammars, . . .

Music. Analysis of Rachmaninoff's pianoforte performances, . . .

Scientific Application: Mutagenesis

Task. Find rules describing highly mutagenic compounds using molecular structure of chemicals



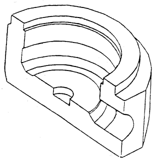
Scientific Application: Mutagenesis (contd)

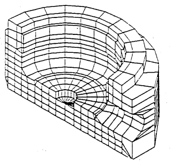
- ILP found new structural alert for poorly-understood subset
- Example rule:

```
A chemical is mutagenic if:
        it contains a double bond conjugated to a
        5-membered aromatic ring via a carbon atom
```

Engineering Application: Mesh-Design

Task. Find correct mesh resolution (partitions) for component parts of a physical object using factors such as the type of edges, boundary conditions and loadings, shape of the structure etc.





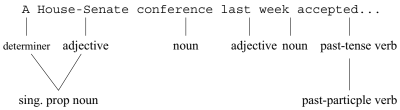
Engineering Application: Mesh-Design (contd)

- ILP derived rules being incorporated into a commercial program for mesh-design
- Example rule:

```
The resolution of an edge A is 2 elements if:
      it is an ``important short'' edge, and
      it is adjacent to an element B, and
      B is completely fixed, and
      B contains no load.
```

Language Application: Tagging

Task. Find rules to eliminate incorrect part-of-speech tags in corpus-based data (1,000,000's of sentences)



Language Application: Tagging (contd)

- ILP rules comparable to best statistical taggers on word-by-word basis
- ILP rules much better on a sentence-by-sentence basis
- Example rule:

```
A word cannot be a past participle if:
    it is sandwiched between a determiner (right) and
    a noun-phrase (left)
```

Music Application: Interpreting Rachmaninoff

Task. Find interpretive rules that may have been used by Rachmaninoff using Ampico recordings of his piano performances (*ca.* 1920, records notes, duration, tempo, dynamics of key pressure, pedalling in a digital form)

Music Application: Interpreting Rachmaninoff (contd)

- ILP rules conform to musical expectations and find some unusual patterns
- Example rules:

```
A note marked ``staccato'' is interpreted by:
      a severe shortening of the note followed by
      lengthening the gap before the next note

A note is played slightly earlier than indicated if:
       it appears at the beginning of a bar
```

ILP: Problems

- Limited treatment of uncertainty
  1. Noise
  2. Incomplete information (e.g., occlusions, misspellings)
- Uncertainty at multiple levels of representation
  - attributes of an object, object's type, the number of objects, the identity of an object (what kind, which, and how many entities are depicted or written about), as well as relationship membership, type and number (which entities are related, how, and how many times)

Complementary Technology

- ILP researchers
  - Developing stochastic and probabilistic representations and algorithms
  - Making ILP techniques more robust to noise and uncertainty
- Traditional ML circles
  - Exploring methods for incorporating relational information
  - Enriching models so that they reflect the relational structure of the domain

Characteristics

- Attempts to represent, reason, and learn in domains with complex relational and rich probabilistic structure
- Other terms: *probabilistic logic learning*, *multi-relational data mining*
- Different dimensions for classifying SRL methods:
  1. Representation Formalism (Logical language)
  2. Probabilistic language
  3. Type of learning
  4. Type of inference
  5. Aggregation

Dimensions for Classification

- Classification based on representation formalism
  1. Full first order logic (e.g., rule-based formalisms)
  2. Prolog/Horn clauses
  3. Frame-based (e.g., objected-oriented) or description logic formalisms
  4. Datalog or Conjunctive database queries
- Classification based on probabilistic semantics
  1. Graphical models
     - Bayesian Networks
     - Markov Networks
     - Dependency networks
     - Logistic Regression
  2. Stochastic logic programs
  3. Stochastic grammars such as PCFGs

Dimensions for Classification (contd)

- Classification based on logical semantics
  1. Least Hebrand Models. E.g. Bayesian logic programs [Kersting & De Raedt, ILP-01]
  2. Ground graphical models. E.g. Markov logic networks [Richardson & Domingos, SRL-04]
- Inference
  1. Pure logical inference
  2. Pure probabilistic inference
     - Marginal/Conditional: MCMC, belief propagation, etc
     - MAP: Graph cuts, weighted satisfiability, etc
  3. ‘Lifted’ probabilistic inference [Pfeffer, 1999]

Dimensions for Classification (contd)

- Learning
  1. Parameter learning
     - Structured representation: Leads to parameter tying/sharing
     - E.g: In a hidden Markov model, because of the Markovian assumption. the parameters determining the next state are the same at each time instance.
     - Enables robust parameter estimation to even be feasible
  2. Generative vs. Discriminative learning
  3. Input to an SRL learning algorithm is most often just a single, richly connected, instance (a set of non-iid instances)
  4. With/without background knowledge

Other Considerations

- Model selection
  1. Many approaches make use of heuristic search through the model space
  2. Language bias to constrain the model search space
     - such as allowing dependencies only among attributes of related instances
     - can lead to exact algorithms
  3. Feature Construction
  4. Statistical predicate invention [Popescul and Ungar, 2003]
- Structural Uncertainty: uncertainty over the relational interpretation
  1. Number uncertainty: distribution over the number of related objects [Koller and Pfeffer, 1998]
  2. Probabilistic logic-based system [Getoor at al, 2002]
  3. Identity uncertainty: modeling uncertainty about the identity of a reference [Pasula and Russell, 2001]
  4. Non-parametric model: Leads to an open-world assumption [Milch et al, 2005]

Course Outline

- Basics of 0-order and First order logic
  - includes types of clauses, syntax and semantics, normal forms, herbrand models,etc.
  - Probabilistic Logic Programming (PLP)
- Inference: Logical and Probabilistic Inference
  - Some details of logical inference including resolution, subsumption theorem, refutation completeness, SLD and SLDNF resolution, answer computation, etc
  - A very brief overview of probabilistic inference (subject matter of Prof. Sunita’s course)
  - First Order Probabilistic Inference

Course Outline

- Inductive Logic Programming: A traditional logic based formalism for Statistical Relational Learning
  - Inductive Learning (in the language of Horn Clauses): Normal and Non-monotonic problem settings
    - The framework for Model inference
    - Inverse Resolution,
    - The Lattice and Cover structure of atoms,
    - The subsumption and implication orders
    - Incorporating Background knowledge
    - Refinement Operators
    - PAC learning
    - Example ILP systems such as GOLEM, FOIL, PROGOL, CLAUDIEN, ICL etc.
    - Inducing classification and regression trees in First Order Logic such as CPROGOL 4.4, TILDE and S-CART
    - Relational learning and boosting
    - Discovery of relational association rules as in WARMR
    - Aleph: An ILP system with a generic framework for implementing many existing ILP systems

Course Outline

- Recent Logic-based Formalisms for Statistical Relational Learning
  - Bayesian Logic Programming
  - Stochastic Logic Programs
  - Markov Logic
  - BLOG: Probabilsitic Models with Unknown Objects
- Induction of Datalog clauses: Efficient induction from hierarchical programs (relational database with multiple tables)
- Other models such as Probabilistic Relational Models, Relational Markov Networks, Relational Dependency Networks, etc

Example Systems

- Probabilistic relational models (Friedman et al, IJCAI-99)
- Stochastic logic programs (Muggleton, SRL-00)
- Bayesian logic programs (Kersting & De Raedt, ILP-01)
- Relational Markov networks (Taskar et al, UAI-02)
- Markov logic networks Richardson & Domingos, SRL-04)

Some Applications of SRL

1. Web search [Brin & Page, WWW-98]
2. Text classification [Chakrabarti et al, SIGMOD-98]
3. Marketing [Domingos & Richardson, KDD-01]
4. Record linkage [Pasula et al, NIPS-02]
5. Gene expression [Segal et al, UAI-03]
6. Information extraction [McCallum & Wellner, NIPS-04]
7. Etc.

Problem Types

1. Collective classification
2. Link discovery
3. Link-based search
4. Link-based clustering
5. Social network analysis
6. Object identification
7. Transfer learning
8. Etc.

References

1. Introduction to Statistical Relational Learning: Edited by Lise Getoor and Ben Taskar, Published by The MIT Press.
   - Many chapters available for public download. Will be placed on the course home page `http://www.cse.iitb.ac.in/~cs717`.
2. Inductive Logic Programming: Techniques and Applications, N. Lavrac and S. Dzeroski. Ellis Horwood, New York, 1994.
   - Available for public download at `http://www-ai.ijs.si/SasoDzeroski/ILPBook/`
3. Course Notes to be made available on the course home page `http://www.cse.iitb.ac.in/~cs717/notes`.
4. Relational Data Mining, Saso Dzeroski and Nada Lavrac, editors, Springer, Berlin, 2001

Grading

- Final Project (in groups of two or three): **30%.** Demo projects are encouraged. You can also take up study projects, but they must be thorough.
  - Project proposal due on February 31st: 5%.
  - Preliminary report due on March 31st: 5%
  - Class presentation of study/demo project by April 10th: 5%
  - Final report of study/demo project by April 10th: 15%
- Homeworks (3–5): **10%** (mix of paper homeworks and programming assignments)
- Midsem: **15%**
- Endsem: **30%**

Grading (contd)

- For selected papers, student are expected to read the papers in advance and submit an independent one-page summary of the paper(s) before the class starts (**15%**). The response papers should be emailed to the instructor by 12:00 noon the day before class. Papers should be at least a half-page and include:
    1. A brief summary of the main contribution of work,
    2. Two or more primary points that critique, praise, or question the findings of the work.

Course Contents and Grading

Course Contents and Grading

Course Contents and Grading

Course Contents and Grading