Overview

Learnable representations for languages

Alexander Clark

Department of Computer Science
Royal Holloway, University of London

August 2010
ESSLLI, 2010

Outline

Course Overview

First things first

- Call me Alex.
- Please ask questions: discussion is good.
- Some material is "mature"; some is quite new. There may be errors or omissions on the slides.
- I will post updated slides on the website.

Outline of the course

Methodological part

Learnability is a central concern of linguistics.

- Learnability and linguistic theory.
- Efficient learnability should be central to language theory
- Design representations to be intrinsically learnable
- Primitives based on objective features of the language

Technical part: recent rapid progress

Some specific models:

- Regular based on the residual languages
- Context-free based on the syntactic monoid
- Non context-free representations based on the Syntactic Concept Lattice

Lecture 1
Fundamentals

- Learnability and linguistics
- Why is it so important?
- What does learnability mean? How can we formalise it?
- Why is learnability hard?
  - Information theoretic problems
  - Complexity problems
- The Argument from the poverty of the stimulus

Joint work with Shalom Lappin

Lecture 2
Regular languages

- Classical theory: Work out all of the problems of inference in this small domain
- Myhill-Nerode theorem and the right congruence
- Minimal DFA
- Three learning algorithms:
  - Positive data only – reversible languages
  - Query learning – Angluin's LSTAR algorithm
  - Stochastic learning of PDFAs.

Lecture 3

Congruence based approaches

- Context free inference based on the syntactic congruence
- Close to classic ideas of structuralist learning
- Three learning algorithms:
  - Positive data only – substitutable languages
  - Query learning – congruential/ NTS languages
  - Stochastic learning of unambiguous NTS languages.
- Not quite powerful enough for linguistic description

Lecture 4
Lattice based approaches

- Much richer structures
- Much richer class of languages
- Slightly (Mildly) context-sensitive
- Only partial results – FG 2009, CoNLL 2010
  - Positive data and MQs

Lecture 5

Option 1

- MCFGs: Derivations which operate on more than one string at a time
- Input and Output
- Two parts of a string with a gap in:
    - Discontinuous constituents

Joint work with Ryo Yoshinaka

Lecture 5

Option 1

- MCFGs: Derivations which operate on more than one string at a time
- Input and Output
- Two parts of a string with a gap in:
    - Discontinuous constituents

Joint work with Ryo Yoshinaka

Option 2

General algorithms for grammatical inference