Connectionist and Statistical Language Processing

Vorbesprechung

Matthew Crocker and Frank Keller

crocker@coli.uni-sb.de keller@coli.uni-sb.de

Office: Gebäude 17.1, Raum 1.15

but please e-mail first

Overview

- Contents of the course
  - Subject matter
- Recommended background
- Structure of the course
  - Lectures, labs, and exercises
- Assessment
  - Exercises
  - Klausur
- Resources
  - Software
  - Main texts
  - Readings

What the course is about

- In computational linguistics, we spend much time formalising linguistic knowledge for subsequent computational analysis:
  - Linguistic theories, descriptions & representations
    - What is syntactic, semantics ...
    - What are linguistic categories ...
    - What are the rules governing linguistic behaviour ...
  - Algorithms
    - How can we use rules and knowledge to analyse and interpret language
- But people aren't given explicit rules and knowledge:
  - Some knowledge might be 'innate' (rationalist)
  - Much is clearly learned from the linguistic environment (empiricist)
  - **Goal:** model the *human* capacity to learn and process language?
- Machine learning of natural language:
  - Cognitively plausible approaches: connectionist
  - Computational approaches: statistical language learning

The Rationalist View

- The Innateness Hypothesis:
  - Poverty of Stimulus: children do not receive sufficient information in their early linguistic experience to achieve their linguistic competence.
    - They are not taught grammar explicitly
    - They do not receive sufficient "negative" stimuli (Gold's Proof)
  - Therefore, the child must be endowed with innate knowledge of grammatical principles and representations
  - Since a child can acquire any language, the innate rules of language must hold across language.
    - <u>Prediction: many linguistic universals</u>
  - Principles and Parameters (Chomsky, 1981;1986; 1992)
    - A set of language universal principles, with finite parametric variation
    - Learning = acquisition of lexicon + parameter setting
- Rational = Theory Driven
- Competence-Performance distinction

Two “Empiricist” approaches

- Connectionist Models:
  - Inspired directly by computational mechanisms in the **brain**
    - Computation is undertaken by networks of simple neuron-like units
  - Nodes compute an output **activation** based on weighted sum of inputs
  - **Learning** involved gradual adjustment of weights
  - Representations are **distributed** across the net, not explicit
  - No distinction between representation and processing
- Hypothesis: such networks have new/different learning abilities
  - Sensitive to order of acquisition, context, and frequency
  - Perhaps detailed innate linguistic knowledge is not required
- Statistical Models:
  - More abstract mathematical models: also maintain symbolic rules
  - Properties, predictions are better understood
- Equivalence of these approaches ?

The Aims of the Course

- Understanding of basic connectionist models:
  - Architecture: the shape, number of nodes, connections, representations
  - Behaviour: categorisation, prediction, association, memory, attractors
  - Learning: different learning rules, limitations on learning, starting small
  - Properties: distributed representation, sensitive to frequency, fault & noise tolerance, generalisation ...
- The abilities of connectionist models for language:
  - Learning phonology, morphology, lexical category, syntax
  - Models of human behaviour (language acquisition, reading times, aphasia)
- The limitations of connectionism:
  - Accessing explicit representations
  - Scaling up
- Statistical approaches to learning linguistic knowledge
- The relationship between connectionist and statistical approaches

Course contents

- Introduction: Stochastic Language Learning
  - Connectionism and the brain
  - The appeal of connectionism
  - Overview of connectionism in language processing
  - Basic connectionist models: nodes and activations
- Part I: Connectionist Models
  - Simple connectionist models and their properties: The perceptron
  - Multi-layer perceptrons: feed-forward networks and internal representations
  - The encoding problem: Localist and distributed representations
  - Generalisation, association, and translational invariance
- Part II: Connectionist Models of Language
  - Modelling acquisition of the English Past-Tense and reading aloud
  - Processing sequences: Simple recurrent networks
  - Modelling acquisition of hierarchical syntactic knowledge

Contents continued

- Part III: Statistical language learning
  - Sparse data and smoothing
    - N-grams, maximum likelihood estimation, held-out estimation,crossvalidation, Good-Turing estimation [Manning & Schütze, ch. 6]
  - Decision trees
    - Classification problems, attribute selection, inductive bias ,overfitting [Mitchell, ch. 3]
  - Bayesian learning
    - Bayes theorem, maximum likelihood vs. least-squared error, MDL, naive Bayes classifier [Mitchell, ch. 6]
  - Clustering
    - Hierarchical vs. non-hierarchical clustering, k-means clustering,EM-based clustering [Manning and Schütze, ch. 14]
  - Applications in lexical acquisition
    - Subcategorization, attachment preferences, selectional restrictions, semantic similarity [Manning and Schütze, ch. 8]

What the course is *not* about

- Computational neuroscience:
  - the development of faithful computational models of different neuron types in the brain.
- Overview of connectionist architectures:
  - There are many, many interesting architectures which do interesting things.
  - We only consider a few of them!
- Practical applications of connectionist techniques

Recommended background

- Students should have completed the Grundstudium:
  - Ideally: Mathematische Grundlagen III: Statistische Methoden
- Mathematical background:
  - General facility for math and statistics is helpful, but specifics will be covered in the lectures:
    - Basic statistics
    - Linear algerbra (vectors and matrixes)
    - Differentiation (Analysis)
- No programming required:
  - Connectionist simulations will be done using the Tlearn simulator
  - Statistical models will using existing tools and Unix level commands

Structure of the Course

- 2 meetings each week:
  - Vorlesung (Di 14-16, room 2.11):
  - Übung (Mi 14-16, CIP Raum):
    - Lab session: to work through simulation exercises (CIP Raum)
    - Statistical learning exercises
  - Need to schedule 1 or 2 additional meetings in January.

- Course review: 5 February
- Discussion Session: 6 February

- <u>Course mark will be based entirely on the final Klausur</u>
  - Klausur: 12 Feb. 2002 @ 14:00 - 16:00
  - In order to take the Klausur, students must satisfy <u>all</u> of the following requirements:
    - Satisfactory completion of all tutorial exercises (e.g. simulations)
      - These can mostly be completed during the lab/tutorial sessions
      - Exercises will be graded as pass/fail
    - Miss no more than 1 exercise session

- Responsible for all material covered in lectures, exercises, assigned readings

Resources

- Main texts:
  - MacLeod, Rolls & Plunkett (199?) *Introduction to Connectionist Modelling of Cognitive Processes*. Oxford University Press.
  - Plunkett & Elman (1997) *Exercises in rethinking innateness*. MIT Press.
  - Tom M. Mitchell. 1997. *Machine Learning*. New York: McGraw-Hill.
  - Manning and Schütze. 1999. *Foundations of Statistical Natural Language Processing*. Cambridge, MA: MIT Press.
- Supplementary reading:
  - Elman, Bates, Johnson, Karmiloff-Smith, Parisi & Plunkett (1996) *Rethinking innateness*. MIT Press.
  - Seidenberg (1997) Language Acquisition and Use: Learning and Applying Probabilistic Constraints. *Science*, vol. 275.
- Selected Articles:
  - Chater & Christianson (1999). Connectionism and Natural Language Processing. In Garrod & Pickering (eds): *Language Processing*. Psychology Press, UK.
  - Elman (1990) Finding structure in time. *Cognitive Science*, 14, 179-211.
  - Steedman (1999) Connectionist Sentence Processing. *Cognitive Science*,23(4), 615:634.