# ESSLLI 2002 Workshop

# *Machine Learning Approaches in Computational Linguistics*

## *Introduction*

Erhard W. Hinrichs, Sandra Kübler

{eh,kuebler}@sfs.uni-tuebingen.de.

Seminar für Sprachwissenschaft

University of Tübingen

Germany

# Motivation for ML in CL

- manually developed NLP systems and language resources for NLP
  - require considerable human effort
  - are often based on limited inspection of the data with an emphasis on prototypical examples
  - often fail to reach sufficient domain coverage
  - often lack sufficient robustness when input data are noisy

# Motivation for ML in CL (2)

- NLP systems and language resources for NLP based on machine learning techniques
  - require less human effort
  - are data-driven and require large-scale data sources
  - achieve coverage directly proportional to the richness of the data source
  - are more adaptive to noisy data

# Machine Learning

- do not give the computer explicit rules
- let it extract knowledge from data
- learning = classification task

# Machine Learning

- do not give the computer explicit rules
- let it extract knowledge from data
- learning = classification task

- from labeled data $\rightarrow$ supervised learning
- from unlabeled data $\rightarrow$ unsupervised learning

# Machine Learning

- do not give the computer explicit rules
- let it extract knowledge from data
- learning = classification task

- from labeled data $\rightarrow$ supervised learning
- from unlabeled data $\rightarrow$ unsupervised learning

- abstract over data $\rightarrow$ eager learning
- do not abstract over data $\rightarrow$ lazy learning

# ML Methods

- supervised learning methods:
  - decision tree learning
  - memory-based learning
  - transformation-based error-driven learning
  - neural networks
  - inductive logic programming
  - maximum entropy learning

# ML Methods

- supervised learning methods:
  - decision tree learning
  - memory-based learning
  - transformation-based error-driven learning
  - neural networks
  - inductive logic programming
  - maximum entropy learning
- unsupervised learning methods:
  - conceptual clustering
  - mimimum description length
  - neural networks

# Example: Part-of-Speech Tagging

- task: find the appropriate POS tag for a word in context
- `They` **`man`** `the boat.` versus `The` **`man`** `in the boat.`
- for English, accuracy > 96 %
- for morphologically rich languages: many POS tags

# Example: Part-of-Speech Tagging

- task: find the appropriate POS tag for a word in context
- `They` **`man`** `the boat.` versus `The` **`man`** `in the boat.`
- for English, accuracy > 96 %
- for morphologically rich languages: many POS tags

Sample instance:

| feature | word - 2 | word - 1 | word | POS tag |
|---|---|---|---|---|
| value | NULL/NULL | They/PRP | man | VB |

# The Learning Problem

- **instance**: a vector of feature values $< f_1, f_2, \ldots, f_n >$ where the values are taken from the discrete or real-valued domain of the $i$th feature

# The Learning Problem

- **instance**: a vector of feature values $< f_1, f_2, \ldots, f_n >$ where the values are taken from the discrete or real-valued domain of the $i$th feature
- let X be the space of possible instances
- let Y be the set of classes

# The Learning Problem

- **instance**: a vector of feature values $< f_1, f_2, \ldots, f_n >$ where the values are taken from the discrete or real-valued domain of the $i$th feature
- let X be the space of possible instances
- let Y be the set of classes
- the goal of the ML system is to learn a **target function** $c : X \to Y$

# The Learning Problem (2)

- **training example**: instance $x \in X$ labeled with the correct class c(x)

# The Learning Problem (2)



- **training example**: instance $x \in X$ labeled with the correct class c(x)
- let D be the set of all training examples
- **hypothesis space**, H: set of functions $h : X \to Y$ of possible definitions

# The Learning Problem (2)

- **training example**: instance $x \in X$ labeled with the correct class c(x)
- let D be the set of all training examples
- **hypothesis space**, H: set of functions $h : X \to Y$ of possible definitions
- the goal is to find an $h \in H$ such that for all $< x, c(x) > \in D$, $h(x) = c(x)$

# CL Problems Approached with ML

- grapheme-phoneme conversion (Stanfill & Waltz 1986, van den Bosch & Daelemans 1993)
- POS tagging (Brants 1998, Cardie 1996, Daelemans et al. 1996)
- PP attachment (Hindle & Rooth 1993, Brill & Resnik 1994, Volk 2001)
- word sense disambiguation (Escudero et al. 2000, Mooney 1996, Veenstra et al. 2000)
- noun phrase chunking (Ramshaw & Marcus 1995, CoNLL 2000)

# Evaluation

- **gold standard**: data against which the ML program is evaluated
- **training set**: data on which the ML program is trained
- **test set**: data on which the performance of the ML program is measured
- **tenfold cross validation**: split data into 10 parts; 10 rounds: use 1 part as test set and remaining parts as training set

# Evaluation Metrics

- **accuracy**: percentage of correctly classified instances from test set
- **recall**: percentage of the items in the gold standard that were found by the ML program
- **precision**: percentage of the items selected by the ML program that are correct

# Workshop Program

| | |
|---|---|
| Mo. | E. Hinrichs, S. Kübler (Tübingen): Introduction |
| | W. Daelemans (Antwerpen): Machine Learning of Language: A Model and a Problem |
| Tu. | M. Rössler (Duisburg): Using Markov models for named entity recognition in German newspapers |
| | P. Osenova, K. I. Simov (Sofia): Learning a token classification from a large corpus |
| We. | O. Streiter (Bolzano): Abduction, induction and memorizing in corpus-based parsing |
| | J. Veenstra, F. H. Müller, T. Ule (Tübingen): Topological field chunking for German |

# Workshop Program (2)

| | |
|---|---|
| Th. | A. Wagner (Tübingen): Learning thematic role relations for wordnets |
| | C. Sporleder (Edinburgh): Learning lexical inheritance hierarchies with maximum entropy models |
| Fr. | P. Lendvai, A. van den Bosch, E. Krahmer (Tilburg), M. Swerts (Eindhoven/Antwerpen): Improving machine-learned detection of miscommunications in human-machine dialogues through informed data splitting |
| | K. Simov (Sofia): Grammar extraction and refinement from an HPSG corpus |
| | Final Discussion |