Machine Learning for natural language processing

Introduction

Laura Kallmeyer

Heinrich-Heine-Universität Düsseldorf

Summer 2016

Introduction

Goal of machine learning:

- Automatically learn how to assign structure to data.
- "Structure" can be for instance a class label, a POS tag, syntactic structure or semantic structure.
- We learn from data and then apply the resulting model to new data.

Most of the course is based on Jurafsky & Martin (2015); Manning et al. (2008)

Table of contents

Applications

Language Models (LM) are probabilistic models of sentences in a given language, for instance probabilistic models of German sentences. They give us the probability that we encounter a given sentence in German text.

Language Models and Machine Translation

LMs are used in various contexts, for instance in machine translation (MT), where a LM selects a translation among a list of alternatives output by the translation model.

(1) a. Paul a des projets magnifiques pour embellir son appartement.
b. Paul hat Projekte wunderbare um zu verschönern seine Wohnung.
c. Paul hat wunderbare Projekte um zu verschönern seine Wohnung.
d. Paul hat Projekte wunderbare um seine Wohnung zu verschönern.
e. Paul hat wunderbare Projekte um seine Wohnung zu verschönern.

Applications

Various **text classification** tasks, for example according to *topic*, *authorship*, *sentiment*, etc.

Sentiment analysis

Goal: Build a classifier that assigns a class $c \in \{\text{positive}, \text{negative}, \text{neutral}\}$ to documents, for instance to comments on films:

c	document
negative	“total langweilig”
negative	“auf den Film kann man echt verzichten”
negative	“ich habe selten so etwas schlechtes gesehen”
neutral	“ich habe schon bessere Filme gesehen”
neutral	“kann man angucken”
positive	“super Film”
positive	“sicher einer der besten Filme dieses Sommers”
positive	“gute Schauspieler und eine gute Story”

Applications

Tasks assigning categories to words, for instance **Part-of-Speech (POS) Tagging**

POS Tagging

Goal: assign POS tags to words in a text.

For example:

Pierre	Vinken	,	61	years	old	,	will	join	the	board
NNP	NNP	,	CD	NNS	JJ	,	MD	VB	DT	NN

as	a	nonexecutive	director	Nov.	29	.
IN	DT	JJ	NN	NNP	CD	.

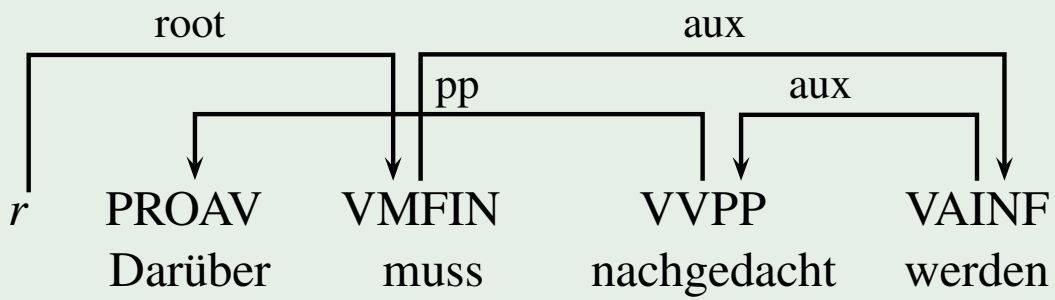
Applications

Tasks assigning structure to sentences, for instance **dependency parsing**

Dependency Parsing

Goal: Create dependencies (directed edges) between words in a sentence and assign dependency labels to these edges.

PROAV	VMFIN	VVPP	VAINF
Darüber	muss	nachgedacht	werden



Supervised vs. unsupervised methods

ML approaches can be distinguished with respect to the type of training data and with respect to the type of model that is learned.

Important distinction concerning training data:
Learning can be based on data containing (parts of) the structure we aim at (**(semi-)supervised learning**) or on data without any information about structure (**unsupervised learning**).

Supervised vs. unsupervised methods

Some examples:

Supervised learning

POS tagging is mostly done with supervised approaches:Traning data:

…	das	muss	man	jetzt	machen	.	…
…	NN	VMFIN	NN	AV	VAINF	PUNCT	…

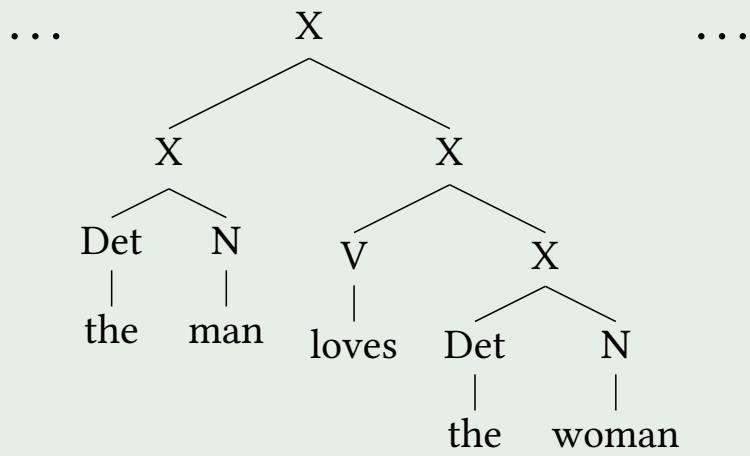
Goal: assign the same types of POS tags to words in new data.

Supervised vs. unsupervised methods

Semisupervised learning

Constituency parsing with latent variables Petrov et al. (2006):

Traning data:



Goal: Learn a probabilistic CFG that generates the same type of structures but with more specific labels instead of X.

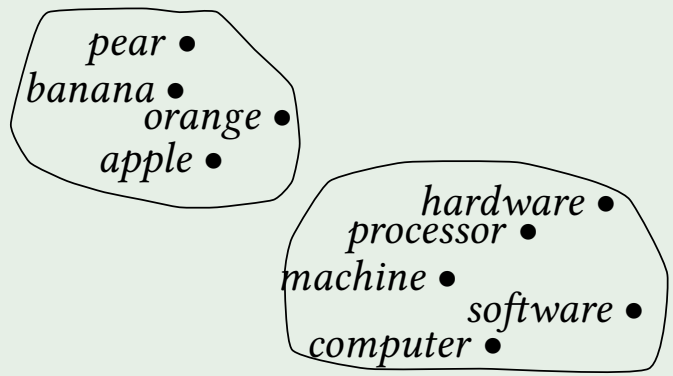
Supervised vs. unsupervised methods

Unsupervised learning

Word sense induction: Create vector representations for words in context and cluster them into word senses.

Input: Raw text.

Output:



In this course, we are mainly concerned with supervised learning.

Topics

Topics to be covered in the course:

- N-Grams and Language Models
- Classification: Naive Bayes
- Classification: k Nearest Neighbors
- Vector Semantics
- Logistic Regression: MaxEnt
- Hidden-Markov Models (HMM)
- POS Tagging
- EM algorithm
- ...

References

Jurafsky, Daniel & James H. Martin. 2015. Speech and language processing. an introduction to natural language processing, computational linguistics, and speech recognition. Draft of the 3rd edition.

Manning, Christopher D., Prabhakar Raghavan & Hinrich Schütze. 2008. *Introduction to information retrieval*. Cambridge University Press.

Petrov, Slav, Leon Barrett, Romain Thibaux & Dan Klein. 2006. Learning Accurate, Compact, and Interpretable Tree Annotation. In *Proceedings of the 21st international conference on computational linguistics and 44th annual meeting of the acl*, 433–440. Sydney.