Computational Lexical Semantics: Features for WSD

LING 7800-006

Martha Palmer

7/14/2011

1

Machine Learning

- Observations
  - Data! The more the merrier (usually)
- Representations
  - Often raw data is unusable, especially in natural language processing
  - Need a way to represent observations in terms of its properties (features)
- Feature Vector

f_0	f_1							f_n

Supervised Learning

- Given a set of instances, each with a set of features, and their class labels, deduce a function that maps from feature values to labels:

Given:

$$\begin{pmatrix} x_{11}, & x_{12}, x_{13} \ldots x_{1m} \\ x_{21}, & x_{22}, x_{23} \ldots x_{2m} \\ \ldots & \\ x_{n1}, & x_{n2}, x_{n3} \ldots x_{nm} \end{pmatrix} \quad \begin{pmatrix} Y_1 \\ y_2 \\ \ldots \\ y_n \end{pmatrix}$$

Find:

$f(\mathbf{x}) = \hat{y}$

$f(\mathbf{x})$ is called a classifier.
The way and/or parameters of $f(\mathbf{x})$ is chosen is called a classification model.

Supervised Learning

- Stages
  - Train model on data
  - Tune parameters of the model
  - Select best model
  - Evaluate

Naïve Bayes

- Assumes that when class label is known the features are independent:

$$f(\mathbf{x}) = \arg\max_{y} \; p(y) \prod_{i=1}^{m} p(x_i | y)$$

$$P(T, S) = \prod_{i=1}^{n} P(RHS_i \mid LHS_i)$$

Naïve Bayes Dog vs Cat Classifier

- 2 features: weight & how frequently it chases a mouse

mouse chase	weight	label
0.7	55	dog
0.05	15	dog
0.2	100	dog
0.25	42	dog
0.2	32	dog
0.6	25	cat
0.2	15	cat
0.55	8	cat
0.15	12	cat
0.4	15	cat

Given an animal that weighs no more than 20 lbs and chases mouse at least 21% of time, is it a cat or dog?

$$f(dog, w \le 20, m \ge .21) =$$
$$p(dog)\, p(w \le 20 \mid dog)\, p(m \ge 0.21 \mid dog) =$$
$$0.5 \times 0.2 \times 0.4 = 0.04$$

$$f(cat, w \le 20, m \ge .21) =$$
$$p(cat)\, p(w \le 20 \mid cat)\, p(m \ge 0.21 \mid cat) =$$
$$0.5 \times 0.4 \times 0.6 = 0.12$$

So, it's a cat! In fact, naïve Bayes is 75% certain it's a cat over a dog.

Word Sense Disambiguation

- Given an occurrence of a word, decide which sense, or meaning, was intended.
- Example, *run*
  - *run1:* move swiftly (*I ran to the store.*)
  - *run2:* operate (*I run a store.*)
  - *run3:* flow (*A river runs through the farm.)*
  - *run4:* length of torn stitches (*Her stockings had a run.*)

Word Sense Disambiguation

- Categories
  - Use word sense labels (*run1, run2, etc.*)
- Features – describe context of word
  - *near(w)* : is the given word near word w?
  - *pos*: word's part of speech
  - *left(w):* is word immediately preceded by w?
  - etc.

Word Sense Disambiguation

- Categories
  - Use word sense labels (*run1, run2, etc.*)
- Features – describe context of word
  - *near(w)* : is the given word near word [***race, river, stocking***]?
  - *pos*: word's part of speech [**noun or verb**]
  - *left(w):* is word immediately preceded by w?
  - etc.

WSD: Sample Training Data Features

POS	near(race)	near(river)	near(stockings)	Sense#
Verb	No	No	No	run1
Verb	No	No	No	run2
Verb	No	Yes	No	run3
Noun	No	No	Yes	run4

run1: move swiftly (*I ran to the store.*)
run2: operate (*I run a store.*)
run3: flow (*A river runs through the farm.)*
run4: length of torn stitches (*Her stockings had a run.*)

Word Sense Disambiguation

- Given an occurrence of a word, decide which sense, or meaning, was intended.
- Example, *run*
  - *run1:* move swiftly *(I ran to the store. John ran in the race by the river. She's running in heels and stockings!*)
  - *run2:* operate (*I run a store. He runs a river rafting guide service that has an annual race*)
  - *run3:* flow (*A river runs through the farm.)*
  - *run4:* length of torn stitches (*Her stockings had a run. Her sweater had a run.*)

WSD: More instances

POS	near(race)	near(river)	near(stockings)	Sense#
Noun	No	No	No	run4
Verb	No	No	No	run1
Verb	No	Yes	No	run3
Noun	Yes	Yes	Yes	run4
Verb	No	No	Yes	run1
Verb	Yes	Yes	No	run2
Verb	No	No	No	run2
Noun	No	No	Yes	run4

Maybe more kinds of features would help?

SUBJ IOBJ

We serve men

We serve food to men.
We serve our community.
serve —IndirectObject→ men

SUBJ DOBJ

We serve men

We serve organic food.
We serve coffee to connoiseurs.
serve —DirectObject→ men

13

More Features for WSD

Dang & Palmer, SIGLEX-02

- Maximum entropy framework, *p(sense|context)*
- Contextual Linguistic Features
  - Topical feature for W, keywords
    - (determined automatically)
  - Local syntactic features for W:
    - presence of subject, complements, passive?
    - words in subject, complement positions, particles, preps,..
  - Local semantic features for W:
    - Semantic class info from WordNet (synsets, etc.)
    - Named Entity tag (PERSON, LOCATION,..) for proper Ns
    - Words within +/- 2 word window

14

Contribution of features to result

Dang & Palmer, SIGLEX-02

- Maximum entropy framework, *p(sense|context)*
- Contextual Linguistic Features
  - Topical feature for W, keywords : **+2.5%,**
    - (determined automatically)
  - Local syntactic features for W: **+1.5 to +5%,**
    - presence of subject, complements, passive?
    - words in subject, complement positions, particles, preps,..
  - Local semantic features for W: **+6%**
    - Semantic class info from WordNet (synsets, etc.)
    - Named Entity tag (PERSON, LOCATION,..) for proper Ns
    - Words within +/- 2 word window

15