Part-Of-Speech (POS) Tagging:
Feature Classification
Evaluation of Results

Recall HMM

- So an HMM POS tagger computes the tag transition probabilities (the A matrix) and word likelihood probabilities for each tag (the B matrix) from a (training) corpus
- Then for each sentence that we want to tag, it uses the Viterbi* algorithm to find the path of the best sequence of tags to fit that sentence.
- This is an example of a sequential classifier, and let's look at how this is related to a more traditional classifier, which we might call a feature-based classifier.
  - Classification task: given a word in a sentence, what is its POS tag?

- The **Viterbi algorithm** is a dynamic programming **algorithm** for finding the most likely sequence of hidden states – called the **Viterbi** path – that results in a sequence of observed events in hidden Markov models. The optional lecture showed how this algorithm could efficiently find the HMM tag sequence.

Comparison of HMM and feature-based classifiers

- Recall that HMM (and n-gram) taggers are sequential classifiers that use the previous sequence of tags as information:

  $word_{n-1}$ … $word_{-2}$ $word_{-1}$ word

  tag_{n-1} … tag_{-2} tag_{-1} XX

  – In order from left to right, use information from previous tags (tag prior probabilities) and word (word likelihood probabilities) to predict the next tag in the sequence

- Instead a feature-based classifier is looking just at the word and properties/features of the surrounding words

  $word_{-2}$ $word_{-1}$ word $word_{+1}$ $word_{+2}$

  XX

  – Assign a tag XX to the word

Evaluation: Is our POS tagger any good?

- Answer: we use a manually tagged corpus, which we will call the "Gold Standard"
  - We run our POS tagger on the gold standard and compare its predicted tags with the gold tags
  - We compute the accuracy (and other evaluation measures)
- Important: 100% is impossible even for human annotators.
  - We estimate humans can do POS tagging at about 98% accuracy.
  - Some tagging decisions are very subtle and hard to do:
    - Mrs/NNP Shaefer/NNP never/RB got/VBD around/RP to/TO joining/VBG
    - All/DT we/PRP gotta/VBN do/VB is/VBZ go/VB around/IN the/DT corner/NN
    - Chateau/NNP Petrus/NNP costs/VBZ around/RB 250/CD
  - The "Gold Standard" will have human mistakes; humans are subject to fatigue, etc.

How can we improve our tagger?

- What are the main sources of information for our HMM POS tagger?
  - Knowledge of tags of neighboring words
  - Knowledge of word tag probabilities
    - *man* is rarely used as a verb….
- Unknown words (words not occurring in the training corpus) can be a problem because we don't have this information
- And we are not including information about the features of the words

Feature-based Classifiers

- A feature-based classifier is an algorithm that will take a word and assign a POS tag based on features of the word in its context in the sentence.
- Many algorithms are used for these traditional classifiers, just to name a few
  - Naïve Bayes
  - Maximum Entropy (MaxEnt)
  - Support Vector Machines (SVM)

  - We'll be covering lots more about classifiers later in the course.

Features of words

- Can do surprisingly well just looking at a word by itself:
  - Word — the: the → DT (determiner)
  - Lowercased word — Importantly: importantly → RB (adverb)
  - Prefixes — unfathomable: un- → JJ (adjective)
  - Suffixes — Importantly: -ly → RB
tangential: -al → JJ
  - Capitalization — Meridian: CAP → NNP (proper noun)
  - Word shapes — 35-year: d-x → JJ
- These properties can include information about the previous or the next word(s)
  - The word *be* appears to the left — pretty → JJ
- But not information about tags of the previous or next words, unlike HMM

Development process for features

- The tagged data should be separated into a training set and a test set.
  - The tagger is trained on the training set and evaluated on the test set
    - May also hold out some data for development
  - Evaluation numbers are not prejudiced by the training set
- If our feature-based tagger has errors, then we improve the features.
  - Suppose we incorrectly tag *as* as IN in the phrase *as soon as*, when it should be RB:

PRP	VBD	IN	RB	IN	PRP	VBD	.
They	left	as	soon	as	he	arrived	.

  - We could fix this with a feature that include the next word.

Overview of POS tagger Accuracies

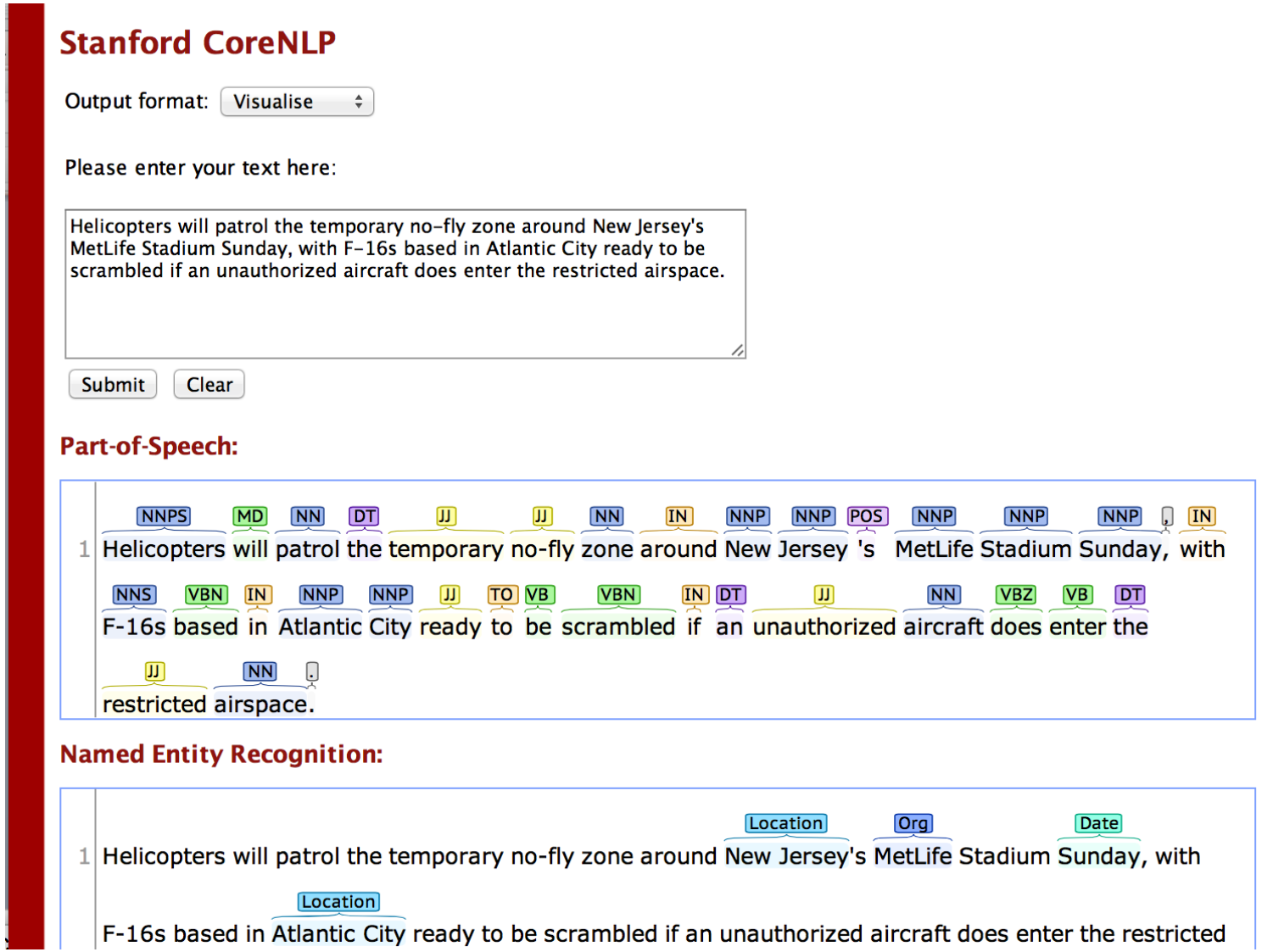
- List produced by Chris Manning
- Rough accuracies: all words / unknown words
  - Most freq tag: ~90% / ~50%
  - Trigram HMM: ~95% / ~55%
    - HMM with trigrams
  - Maxent P(t|w): 93.7% / 82.6%
    - Feature based tagger
  - MEMM tagger: 96.9% / 86.9%
    - Combines feature based and HMM tagger
  - Bidirectional dependencies: 97.2% / 90.0%
  - Upper bound: ~98% (human agreement)

Most errors on unknown words

POS taggers with online demos

- Many pages list downloadable taggers (and other resources) such as this page from the Stanford NLP group and George Dillon at U Washington
  - http://nlp.stanford.edu/software/tagger.shtml
  - http://faculty.washington.edu/dillon/GramResources/
- There are not too many on-line taggers available for demos, but here are some possibilities:
  - The Stanford online parser demo includes POS tags: http://nlp.stanford.edu:8080/parser/ http://nlp.stanford.edu:8080/corenlp/
  - Illinois (UIUC) tagger demo from the Cognitive Computation Group
  - http://cogcomp.cs.illinois.edu/demo/pos/?id=4 (colors!)

Stanford NLP demo



Stanford CoreNLP

Output format: Visualise

Please enter your text here:

Helicopters will patrol the temporary no-fly zone around New Jersey's MetLife Stadium Sunday, with F-16s based in Atlantic City ready to be scrambled if an unauthorized aircraft does enter the restricted airspace.

Submit Clear

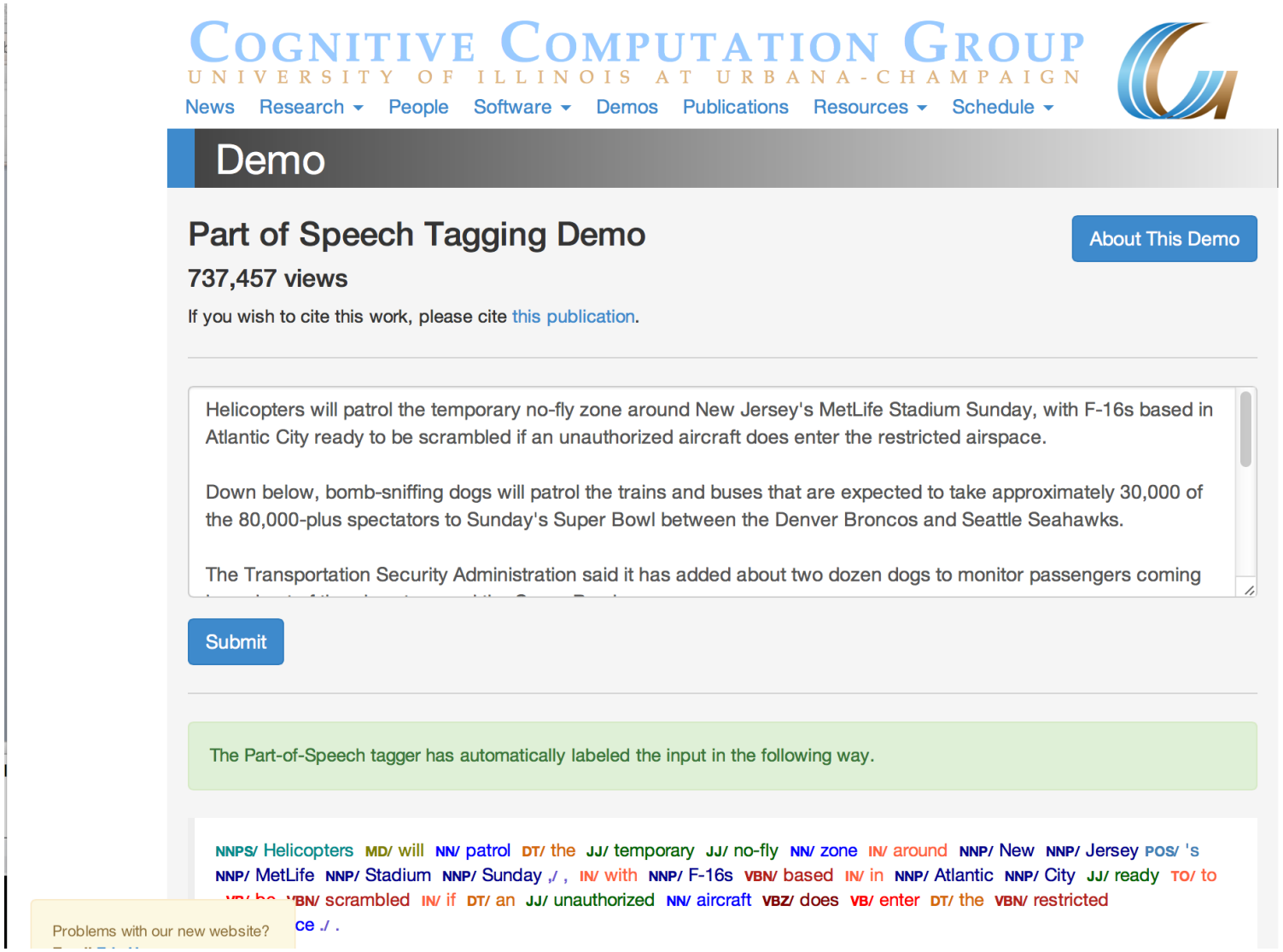
Part-of-Speech:

1 Helicopters/NNPS will/MD patrol/NN the/DT temporary/JJ no-fly/JJ zone/NN around/IN New/NNP Jersey/NNP 's/POS MetLife/NNP Stadium/NNP Sunday/NNP ,/, with/IN F-16s/NNS based/VBN in/IN Atlantic/NNP City/NNP ready/JJ to/TO be/VB scrambled/VBN if/IN an/DT unauthorized/JJ aircraft/NN does/VBZ enter/VB the/DT restricted/JJ airspace/NN ./.

Named Entity Recognition:

1 Helicopters will patrol the temporary no-fly zone around New Jersey/Location 's MetLife/Org Stadium Sunday/Date, with F-16s based in Atlantic City/Location ready to be scrambled if an unauthorized aircraft does enter the restricted

UIUC demo


COGNITIVE COMPUTATION GROUP
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN
News Research People Software Demos Publications Resources Schedule
Demo
Part of Speech Tagging Demo
About This Demo
737,457 views
If you wish to cite this work, please cite this publication.
Helicopters will patrol the temporary no-fly zone around New Jersey's MetLife Stadium Sunday, with F-16s based in Atlantic City ready to be scrambled if an unauthorized aircraft does enter the restricted airspace.
Down below, bomb-sniffing dogs will patrol the trains and buses that are expected to take approximately 30,000 of the 80,000-plus spectators to Sunday's Super Bowl between the Denver Broncos and Seattle Seahawks.
The Transportation Security Administration said it has added about two dozen dogs to monitor passengers coming
Submit
The Part-of-Speech tagger has automatically labeled the input in the following way.
NNPS/ Helicopters MD/ will NN/ patrol DT/ the JJ/ temporary JJ/ no-fly NN/ zone IN/ around NNP/ New NNP/ Jersey POS/ 's
NNP/ MetLife NNP/ Stadium NNP/ Sunday ,/ , IN/ with NNP/ F-16s VBN/ based IN/ in NNP/ Atlantic NNP/ City JJ/ ready TO/ to
BN/ scrambled IN/ if DT/ an JJ/ unauthorized NN/ aircraft VBZ/ does VB/ enter DT/ the VBN/ restricted
ce ./ .
Problems with our new website?

Conclusions

- Part of Speech tagging is a doable task with high performance results
  - In addition to the standard text POS taggers discussed here, there are now POS tag systems and taggers developed for social media text.
- Contributes to many practical, real-world NLP applications and is now used as a pre-processing module in most systems
- Computational techniques learned at this level can be applied to NLP tasks at higher levels of language processing