

Homework 2: Part-of-Speech Tagging with HMMs and CRFs

Ankit Goyal
ankit@cs.utexas.edu
Natural Language Processing

March 8, 2015

1 Introduction

In this experiment we analyze the performance of Hidden Markov Models (HMMs) and Conditional Random Fields (CRFs) for POS tagging problem. Part of Speech Tagging involves reading a piece of text in some language and assigning part of speech to each word (or other token), such as noun, verb, adjective, etc.

Hidden Markov Models: Hidden Markov models are generative models. Input to HMM are tokens of observed symbols and it infers the corresponding pos tags. It models the joint distribution of labels and observations. The parameters estimated in HMM are the transition from one state to another (where states are POS tags) and observation probabilities (probability of a token in that POS tag).

CRFs: CRFs are a type of discriminative undirected probabilistic graphical model. To build a Conditional Random Field, we define a bunch of feature functions (which can depend on the entire sentence, a current position, and nearby labels), assign them weights, and add them all together, transforming at the end to a probability.[1]

Each feature function is a function that takes input as: a sentence, the position i of a word in the sentence, the label l_i of the current word and the label l_{i-1} of the previous word. and It outputs a real valued number.

HMM vs CRM : CRFs can model a much richer set of label distributions since HMMs are necessarily local in nature (because they are constrained to binary transmission and emission feature function, which forces that each word can depend only on a current label and each label to only depend on a previous label). CRFs can define functions with more context and this can predict better.

2 Experiments:

- Since mallet requires the training data in a fixed format, the first task was to convert the given datasets to the format required by the mallet.
- For the WSJ two sets of tests were run, in SHORT test, the model was trained on 00 and it was tested on 01 and in LARGE set, the model was trained on 00, 01 and tested on 02 and 03.
- Tests were run with different features i.e., only considering morph features and then considering morph features with suffix features.

ATIS				
	Training Accuracy	Testing Accuracy	OOV Accuracy	Time
HMM	0.8865	0.8659	0.2916	00:00:07
CRF	0.9982	0.9163	0.3333	00:01:14
CRF + morph	0.9983	0.9299	0.3750	00:01:59
CRF + morph + suffix	0.9981	0.9357	0.4583	00:02:25
WSJ				
HMM SHORT	0.8617	0.7848	0.3794	00:01:14
HMM LARGE	0.8870	0.8333	0.3917	00:06:28
CRF SHORT	0.9837	0.7916	0.4721	02:08:44
CRF LARGE	0.9938	0.84262	0.50518	05:32:10
CRF + morph SHORT	0.9935	0.8323	0.5900	02:25:19
CRF + morph LARGE	0.9878	0.8605	0.6060	05:01:17
CRF + morph + suffix SHORT	0.9902	0.8772	0.7630	01:57:55
CRF + morph + suffix LARGE	0.9915	0.9030	0.7665	06:28:00

3 Observations:

1. **Test Accuracy:** Test accuracy for CRF is always better than test accuracy for HMM. The reason for that is that the generative model has higher asymptotic error and HMMs are less prone to overfitting than CRFs. As we increase the amount of data the test accuracy increases for CRF significantly.
2. **Training Accuracy:** Training accuracy for CRF is higher than HMM because CRF can overfit the model and thus the training accuracy is higher for CRF.
3. **Test Accuracy for OOV** Test accuracy of OOV words is significantly less than as compared to seen tokens. Since their observation probability will be very less than compared to seen tokens.
4. **Adding orthographic features:** I added two types of orthographic features. Morphological Features: caps, hyphen and number. And suffix features: able, ed, er, ly, ing, full, less, ity, or, ion. With increase in orthographic features, the time required to train CRF didn't increase proportionally to the number of features added. However the accuracy increased with the introduction of orthographic features across all phases.
5. **Runtime:** Due to the nature of algorithms as expected CRF is slower than HMM. The CRF is about 15-100 times slower than HMM. This is because CRF runs perform more number of computations, however CRF is more accurate than HMM which could be more important than the runtime for lots of the applications.
6. **Number of iterations:** Number of iteration for HMM to converge are less than the number of iterations required in CRF. It is due to the complexity of the models. CRF needs to learn about more parameters and thus need more number of iterations. Both the models observe an increase in accuracy with more number of iterations.

4 References:

1. <http://blog.echen.me/2012/01/03/introduction-to-conditional-random-fields/>