Structured Prediction for PoS Tagging

Id : 305248791

User : chris91

All models were tested on the last 10% percent of the data. The accuracy is a fraction between 0 and 1, And was calculated according to this formula: $$accuracy = \frac{\#\ test\ set\ words\ with\ correct\ tag}{\#test\ set\ words}$$

model	Trained on 10% of the data	Trained on 25% of the data	Trained on 90% of the data
Baseline	0.8680135514790943	0.8971161791439431	0.9192117005453644
HMM	**0.9014047264914891**	**0.9249049743843992**	**0.9426954222442572**
MEMM with basic feature function (transmission + emission)	0.7617749132374814	0.8121880680879193	0.8680300776731119
MEMM with feature function for “ing” suffix + transmission + emission	0.7553049082796232	0.8172781358453148	0.8693439100975046
MEMM with feature function for “ly” suffix + transmission + emission	0.7456949264584366	0.8166584035696579	0.8696331184928111
MEMM with feature function for words begin with an upper case letter + transmission + emission	0.757676417121137	**0.828524210874235**	**0.8707155842009585**
MEMM with feature function for “ed” suffix + transmission + emission	0.7516774086927781	0.8191042802842505	0.8615931251032887

The hmm model performed best on all training percentages, reaching 94.26% when trained on 90% of the data.

The feature for words that begin with an upper case letter achieved best results from all other feature mappings in the memm model reaching 87.07% when trained on 90% of the data. In general when trained on 25% and 90% of the data the memm models with additional features performed better than the basic memm model with feature function for transmission and emission only.

Sampling 5 sentences by hmm:

[['*START*', 'Operations', '*rare*', ',', 'learning', 'of', 'the', '*rare*', ',', 'Hibor', 'rates', 'and', 'Circuit', 'other', 'rights', 'of', 'competing', 'the', 'after', 'you', 'to', 'take', 'of', 'his', 'consumers', 'only', 'in', 'Africa', '"', 'consultant', 'reruns', '*END*'],

['*START*', 'who', 'also', 'of', 'a', 'months', "'ll", 'come', 'imaginative', 'for', 'Taking', 'shoulder', 'into', 'he', 'heads', 'a', 'stock', '-RCB-', '*END*'],

['*START*', 'Theater', 'enemy', 'whose', '*rare*', 'organization', 'restructuring', 'shortages', '*END*'],

['*START*', 'an', 'rural', '%', 'must', 'work', 'much', 'damaged', 'needs', 'as', 'This', 'bowling', 'does', ',', 'this', '%', 'has', ',', '"', 'of', 'little', 'money', 'and', 'were', 'a', 'Britain', 'and', 'group', 'is', 'Urban', 'jumped', 'of', 'White', '*END*'],

['*START*', 'so', ':', 'would', 'give', 'despite', 'the', 'nation', 'common', 'arbitrator', 'The', 'deficit', 'despite', 'world', 'expected', 'peasant', 'through', 'further', 'five-member', 'output', 'earned', 'to', 'boost', 'the', 'with', 'bill', 'also', 'is', 'million', 'alternatives', 'in', 'both', 'diversity', ',', 'the', '20', '*END*']]

The samples don't have a meaning when you read them, although one senses when reading some of the sentences that they do have a grammatical structure. Much like the sentence "**Colorless green ideas sleep furiously** " by Noam Chomsky, Although the sentence is grammatically correct, no obvious understandable meaning can be derived from it.

שאלה 1.1

בהינתן פרמטרים t ו-e של מודל HMM, נצטרך למצוא מיפוי פיצ'רים שקול ופרמטר w עבור המודל הלוג-ליניארי.

נגדיר את w להיות בגודל $MN+M^2$ כאשר M הוא מספר המצבים האפשריים ו-N מספר התצפיות האפשריות. האינדקסים ה-M^2 הראשונים תואמים ל-M^2 המעברים בין המצבים האפשריים ולכל מעבר מ-i ל-j נגדיר את הערך $w(i,j)$ להיות $\log(t_{ij})$.

האינדקסים ה-$M\cdot N$ תואמים לכל ה-emissions האפשריים של התצפיות ממצב מסוים. הערך $w(i,x_j)$ יהיה $\log(e_{i,x_j})$. נוודא שאכן יתקיים השוויון.

$$Pr(y_t=i \mid y_{t-1}=j, x_t=x_j) = e^{\log(t_{i,j})+\log(e_{i,x_j})}$$

$$= t_{i,j} \cdot e_{i,x_j}$$

שאלה ב,ד:

כן, נוכל לבטא את המודל כמודל לוג-ליניארי.
ההתנייה על המשתנים האחרים משמעותה
שווקטורים Φ יהיו ארוכים יותר. ומספר הפרמטרים
יהיה $M^3+M\cdot N$ כאשר גודל הממד של x הוא M
וגודל הממד של y הוא N.

שאלה 1.4

אילוץ: $\sum_k t_{i,k}=1$.

נגדיר לגרנז'יאן:

$$L(t,e,\lambda)=\sum_{i=1}^{N}\sum_{t=1}^{T^{(i)}}\log\left(t_{y_{t-1}^{(i)},y_t^{(i)}}\right)+\sum_{i=1}^{N}\sum_{t=1}^{T^{(i)}}\log\left(e_{y_t^{(i)},x_t^{(i)}}\right)-\lambda\left(\sum_k t_{ik}-1\right)$$

הערה: בבניית הלגרנז'יאן לא נתתי ״חסם״ לאילוץ על $e_{i,j}$ כי לא צריך אותו בהוכחה, הגזירה היא לפי $t_{i,j}$ והאילוץ יתאפס בגזירה.

$$\frac{\partial L}{\partial t_{i,j}}=\frac{\#(y_i\to y_j)}{t_{i,j}}-\lambda=0$$

$$(*)\quad t_{i,j}=\frac{\#(y_i\to y_j)}{\lambda}$$

נשתמש באילוץ $\sum_k t_{i,k}=1$ כדי למצוא את λ

$$\sum_k t_{i,k}=\sum_k\frac{\#(y_i\to y_k)}{\lambda}=1$$

$$\lambda=\sum_k\#(y_i\to y_k)$$

נציב את λ ב-(*)

$$\boxed{\hat{t}_{i,j}=\frac{\#(y_i\to y_j)}{\sum_k\#(y_i\to y_k)}}$$

1.3 : Slice

$$E(y_t \mid y_{t-1}, x_{1:T}; \theta) = \sum_{y_t} y_t \cdot p(y_t \mid y_{t-1}, x_{1:T}; \theta) \quad \text{.1}$$

$$= \sum_{y_t} y_t \cdot \frac{p(y_t, y_{t-1}, x_{1:T}; \theta)}{p(y_{t-1}, x_{1:T}; \theta)}$$

$$= \sum_{y_t} y_t \cdot \frac{\frac{1}{z_1} e^{-E(y_t, y_{t-1}, x_{1:T}; \theta)}}{\frac{1}{z_2} e^{-E(y_{t-1}, x_{1:T}; \theta)}}$$

$$z_1 = \sum_{y_t, y_{t-1}, x_{1:T}} e^{-E(y_t, y_{t-1}, x_{1:T}; \theta)}$$

$$z_2 = \sum_{y_{t-1}, x_{1:T}} e^{-E(y_{t-1}, x_{1:T}; \theta)}$$

שאלה 1.3

2.

$$l(x;\theta) = \log p(x) = \log \sum_y P(x,y)$$

$$= \log\left(\sum_y \frac{e^{-E(x,y)}}{Z}\right) = \log\left(\sum_y e^{-E(x,y)}\right) - \log Z =$$

$$\left[Z = \sum_{x'}\sum_{y'} e^{-E(x',y';\theta)}\right] \quad \log\left(\sum_y e^{-E(x,y)}\right) - \log\left(\sum_{x'}\sum_{y'} e^{-E[x',y']}\right)$$

$$\frac{\partial l}{\partial \theta} = \frac{1}{\sum_y e^{-E(x,y)}} \cdot \frac{\partial}{\partial \theta} \sum_y e^{-E(x,y)} - \frac{1}{\sum_{x'}\sum_{y'} e^{-E(x',y')}} \frac{\partial}{\partial \theta} \sum_{x'}\sum_{y'} e^{-E(x',y')}$$

$$= -\frac{1}{\sum_y e^{-E(x,y)}} \cdot \sum_y e^{-E(x,y)} \cdot \partial E(x,y) + \frac{1}{\sum_{x'}\sum_{y'} e^{-E(x,y)}} \sum_{x'}\sum_{y'} e^{-E(x',y')} \cdot \frac{\partial E(x',y')}{\partial \theta}$$

נשים לב ש: $e^{-E(x,y)} = Z\, p(x,y)$

$$\frac{\partial l(x;\theta)}{\partial \theta} = -\frac{1}{Z \sum_y P(x,y)} \cdot Z \sum_y P(x,y) \frac{\partial E(x,y)}{\partial \theta} + \frac{1}{Z} \cdot Z \sum_{x'}\sum_{y'} P(x',y') \frac{\partial E(x',y')}{\partial \theta}$$

$$= -\frac{1}{\sum_y P(x,y)} \sum_y P(x,y) \frac{\partial E(x,y)}{\partial \theta} + \sum_{x'}\sum_{y'} P(x',y') \frac{\partial E(x',y')}{\partial \theta}$$

נשים לב ש: $\sum_y P(x,y) = P(x)$

ונשתמש בהגדרה של הסתברות מותנית $p(y|x) = \frac{p(x,y)}{p(x)}$

$$\frac{\partial \log P(x;\theta)}{\partial \theta} = \frac{-1}{P(x)} \sum_y P(x,y) - \sum_y \frac{P(x,y)}{P(x)} \frac{\partial E(x,y)}{\partial \theta} + \sum_{x'} \sum_{y'} P(x',y') \frac{\partial E(x',y')}{\partial \theta}$$

$$= -\sum_y P(y|x) \frac{\partial E(x,y)}{\partial \theta} + \sum_{x'} \sum_{y'} P(x',y') \frac{\partial E(x',y')}{\partial \theta}$$

$$= \underbrace{-E_{P(y|x)}\left[\frac{\partial E(x,y)}{\partial \theta}\right]}_{E_{x\sim D}\left[\frac{\partial E(x;\theta)}{\partial \theta}\right]} + \underbrace{E_{P(x',y')}\left(\frac{\partial E(x',y')}{\partial \theta}\right)}_{E_{x\sim P_\theta}\left[\frac{\partial E(x;\theta)}{\partial \theta}\right]}$$

$$\Rightarrow E_{x\sim D}\left[\frac{\partial \log P_\theta(x)}{\partial \theta}\right] = E_{x\sim P_\theta}\left[\frac{\partial E(x;\theta)}{\partial \theta}\right] - E_{x\sim D}\left[\frac{\partial E(x;\theta)}{\partial \theta}\right]$$

נ"ל.