A Star for Part of Speech Tagging

November 11, 2012

1 Heuristic

h(k) = $\left\{min_{t_k..t_N} \sum_{i=k}^{N} P(T = t_i/T = t_{i-1})\right\} + \left\{\sum_{i=k}^{N} min_{t_i} P(T = t_i/S = w_i)\right\}$

2 Admissible Heuristic

Heuristic estimation of cost from node k to goal N

h(k) = $\left\{min_{t_k..t_N} \sum_{i=k}^{N} P(T = t_i/T = t_{i-1})\right\} + \left\{\sum_{i=k}^{N} min_{t_i} P(T = t_i/S = w_i)\right\}$

$forall_{t_k..t_N} \left\{\sum_{i=k}^{N} P(T = t_i/T = t_{i-1})\right\} \geq \left\{min_{t_k..t_N} \sum_{i=k}^{N} P(T = t_i/T = t_{i-1})\right\} - - - - - (1)$

$forall_{t_i} \{P(T = t_i/S = w_i)\} \geq \{min_{t_i} P(T = t_i/S = w_i)\} - - - - - (2)$

$forall_{t_k..t_N} \left\{\sum_{i=k}^{N} P(T = t_i/T = t_{i-1})\right\} \geq \left\{\sum_{i=k}^{N} min_{t_i} P(T = t_i/T = t_{i-1})\right\} - - - - - (3)$

Cost from node k to goal N

$h^*(k) = min_{t_k..t_N} \sum_{i=k}^{N} \{P(T = t_i/T = t_{i-1}) + P(T = t_i/S = w_i)\}$

$h^*(k) = min_{t_k..t_N} \left\{\sum_{i=k}^{N} P(T = t_i/T = t_{i-1}) + \sum_{i=k}^{N} P(T = t_i/S = w_i)\right\}$

Using eqn (3) and (1)

$h^*(k) \geq min_{t_k..t_N} \left\{\sum_{i=k}^{N} P(T = t_i/T = t_{i-1})\right\} + \left\{\sum_{i=k}^{N} min_{t_i} P(T = t_i/T = t_{i-1})\right\}$

$h^*(k) \geq h(k)$

3 Triangle Inequality

h(a) = heuristic estimation of cost from node a to N

h(b) = heuristic estimation of cost from node b to N

Cost(a,b) = Cost from node a to b

h(a) $\leq h(b) + Cost(a,b)$

$$\left\{min_{t_a..t_N} \sum_{i=a}^{N} P(T = t_i/T = t_{i-1})\right\} + \left\{\sum_{i=a}^{N} min_{t_i} P(T = t_i/S = w_i)\right\}$$
$$\leq \left\{min_{t_b..t_N} \sum_{i=b}^{N} P(T = t_i/T = t_{i-1})\right\} + \left\{\sum_{i=b}^{N} min_{t_i} P(T = t_i/S = w_i)\right\}$$
$$+ \{P(T = t_b/T = t_a) + P(T = t_a/S = w_a)\}$$

By removing the common emission probability terms on both side, we get

$$\left\{min_{t_a..t_N} \sum_{i=a}^{N} P(T = t_i/T = t_{i-1})\right\} + \{min_{t_a} P(T = t_a/S = w_a)\}$$
$$\leq \left\{min_{t_b..t_N} \sum_{i=b}^{N} P(T = t_i/T = t_{i-1})\right\}$$
$$+ \{P(T = t_b/T = t_a) + P(T = t_a/S = w_a)\}$$

$\min_{t_a} P(T = t_a/S = w_a) <= P(T = t_a/S = w_a)$ so it can be removed,

$$\left\{min_{t_a..t_N} \sum_{i=a}^{N} P(T = t_i/T = t_{i-1})\right\}$$
$$\leq \left\{min_{t_b..t_N} \sum_{i=b}^{N} P(T = t_i/T = t_{i-1})\right\}$$
$$+ \{P(T = t_b/T = t_a)\}$$

By definition of minimum, the above should be true.

*** In addition running the AStar for five folds did not update the single node in the Closed list.**

4 Results

Time Taken in AStar	$\tilde{6}$3 Seconds
Time Taken in Viterbi	$\tilde{6}$2 Seconds
No of Open Nodes	$\tilde{3}$0 Lacs Nodes
No of Open Nodes(Basic Heu)	$\tilde{3}$1 Lac Nodes
Improvement over basic	18 %

AStar performs same as viterbi because the words seems to be associated with 3 tags on the average. So if the word not associated Tag, negative of log of Emission probablity (edge cost) becomes infinity. If the edge cost is infinity, we have ignored if the edge cost is infinity(no path).

Second experiment is to measure the perfomance ot AStar when words are associated all tags (worst case). We have added small probability to all emission probability. In this case Viterbi performs better.

Time Taken in AStar	$\tilde{1}$.5 hours
Time Taken in Viterbi	$\tilde{2}$0 minutes
No of Open Nodes	$\tilde{1}$50 Lacs Nodes
No of Open Nodes(Basic Heu)	$\tilde{2}$40 Lac Nodes
Improvement over basic	33 %