PGM Assignment - 2

SURAJ TRIPATHI

2015MCS2354

1) Implemented Message Passing over a Junction Tree and Loopy Belief Propagation with the option of switching between the described 4 models

Program Outputs : Character-wise, word-wise Accuracy, Data Log-Likelihood and Running Time.

2) Experiments

- Comparing Inference Algorithms :

Dataset Data-tree	Character-wise accuracy	Word-wise accuracy	Data Log-likelihood	Running Time (sec)
MP	67.56	16.67	-6.60	1.00
LBP	67.56	16.67	-11.39	0.82

Dataset Data-treeWS	Character-wise accuracy	Word-wise accuracy	Data Log-likelihood	Running Time
MP	66.56	16.84	-7.23	0.52
LBP	66.99	16.30	-12.59	0.31

Dataset Data-loops	Character-wise accuracy	Word-wise accuracy	Data Log-likelihood	Running Time
MP	58.99	7.14	-7.67	1.00
LBP	56.83	7.14	-12.03	0.63

Dataset Data-loopsWS	Character-wise accuracy	Word-wise accuracy	Data Log-likelihood	Running Time
MP	68.05	13.07	-7.05	1.50
LBP	68.51	16.15	-13.27	0.82

Message passing and Belief propagation both gives almost equivalent results on every metric (character-wise, word-wise accuracy and Data Log-Likelihood), Belief propagation algorithm in general, takes less time than Message Passing algorithm.

Belief propagation converges within 10-15 iterations with stopping criteria as 0.001(maximum difference between beliefs of clique node).

For Large Networks, Belief Propagation is far more useful than Message passing algorithm because Time complexity of Message passing algorithm is a factor of maximum clique size in induced graph, whereas Time complexity of BP depends on maximum clique size of original markov Network(It's a factor of D^2 if D is the number of values a variable can take in graph using Bethe Cluster Graph because here all nodes are defined over edges).

- Adding Sophistication in Factors:

MP (message passing) over clique tree for the loopy-WS dataset for each of the four set of factor models.

Message Passing over Clique tree:

MP Data-loopsWS	Character-wise accuracy	Word-wise accuracy	Data Log-likelihood	Running Time (Sec)
OCR	60.03	8.46	-10.06	0.01
OCR + TRANS	66.04	11.53	-7.52	0.42
OCR + TRANS + SKIP	57.25	10.00	-6.48	0.40
OCR + TRANS + SKIP + PAIRSKIP	68.05	13.07	-7.05	1.00

Loopy Belief Propagation:

LBP Data-loopsWS	Character-wise accuracy	Word-wise accuracy	Data Log-likelihood	Running Time
OCR	60.03	8.46	-12.64	0.01
OCR + TRANS	66.51	11.54	-12.51	0.40
OCR + TRANS + SKIP	66.35	12.31	-12.72	0.50
OCR + TRANS + SKIP + PAIRSKIP	68.51	16.15	-13.27	0.62

Belief propagation and Message passing algorithm gives similar results in every case. It gives a indication that Belief propagation is able to converge even with less number of factors. It shows that Belief propagation is robust In nature and It gives similar performance as compared to exact inference algorithms like message passing.

3. MAP Inference :

- Comparing Inference Algorithms :

Algorithm Data-tree	Character-wise accuracy	Word-wise accuracy	Data Log-likelihood	Running Time (sec)
MP	77.21	52.38	-6.52	0.20
LBP	77.47	52.38	-11.69	0.10

Algorithm Data-treeWS	Character-wise accuracy	Word-wise accuracy	Data Log-likelihood	Running Time (sec)
MP	74.17	45.65	-7.32	0.20
LBP	75.05	45.10	-13.17	1.02

Algorithm Data-loops	Character-wise accuracy	Word-wise accuracy	Data Log-likelihood	Running Time (sec)
MP	66.18	25.00	-7.85	1.20
LBP	58.27	21.42	-12.55	1.00

Algorithm Data-loopsWS	Character-wise accuracy	Word-wise accuracy	Data Log-likelihood	Running Time (sec)
MP	75.00	50.76	-7.57	1.00
LBP	78.08	54.61	-14.79	1.02

- Adding Sophistication in Factors:

MP(MAP)

MP Data-loopsWS	Character-wise accuracy	Word-wise accuracy	Data Log-likelihood	Running Time
OCR	60.03	8.46	-7.82	0.01
OCR + TRANS	74.38	34.62	-9.77	0.10
OCR + TRANS + SKIP	73.76	33.84	-9.77	0.50
OCR + TRANS + SKIP + PAIRSKIP	75.00	50.76	-7.57	1.00

LBP(MAP)

LBP Data-loopsWS	Character-wise accuracy	Word-wise accuracy	Data Log-likelihood	Running Time
OCR	60.03	8.46	-12.64	0.01
OCR + TRANS	76.85	43.84	-12.65	0.50
OCR + TRANS + SKIP	76.69	49.23	-13.08	1.00
OCR + TRANS + SKIP + PAIRSKIP	78.08	54.61	-14.79	1.02

MAP performance is performs better than Marginal inference approach. The reason of better performance in case of MAP is because it captures the effect of other variables already predicted in the word thus giving better performance in predicting the whole word together(We are utilising knowledge of already predicted characters of word in MAP) whereas in case of Marginal Inference we are looking at the max marginal of individual variable of the word, not taking consideration of already predicted variables of word.