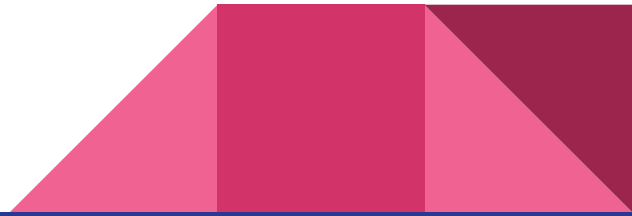
Question Answering using Attention and HMN

By : Manasee Moghey and Sanjog Handique

Topics

1. Problem
2. Dataset
3. Related Work
4. Our Approach
5. Design
6. Experimental Setup
7. Result
8. Sample Output



Problem

- Question answering is crucial task
  - Needs understanding of Natural Language
  - Understanding the world knowledge
- Many solutions exist
  - Select local maxima

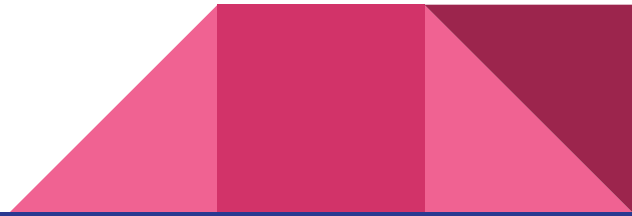
Document : The Rankine cycle is sometimes referred to as a practical Carnot cycle ecause, when an efficient turbine is used, the TS diagram begins to resemble the Carnot cycle.

Question : WHAT IS THE RANKINE CYCLE SOMETIMES CALLED?

Ground truth: practical Carnot cycle
Prediction: practical Carnot cycle

Dataset

- Human annotated datasets had less data but high quality
- Automated datasets are large scaled but low quality
- Using Stanford Question Answering Dataset (SQuAD)
  - Reading comprehension
  - 100,000 questions
  - 500 articles
  - Sources: like Wikipedia
  - 80k trainings set, 10k dev set, 10k test set
  - SQuAD answers can contain long phrases
  - Varied length questions



Related Work

- Statistical QA
  - Rule based algo
  - Relational queries
- Neural QA
  - Encoder decoder
- Bidirectional Attention Flow Mechanism
  - Character level
  - CNN

Our Approach

- Based on Dynamic Coattention Network
  - Question & context encoded into LSTM
  - Coattention matrix is generated using normalised attention
  - Document and question multiplied with co-attention
  - Passed through Bi-LSTM to get output
  - Decoder uses Highway Maxout Networks to recover from local maxima

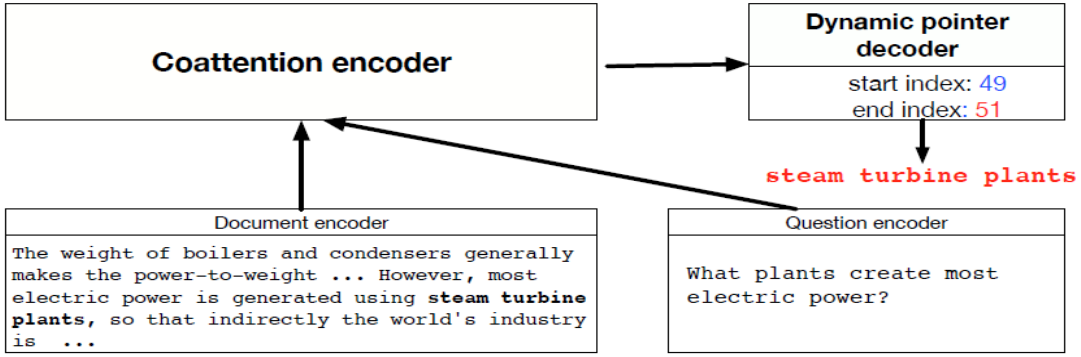
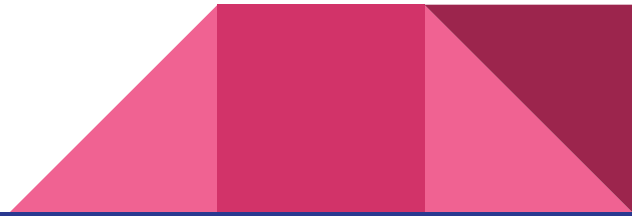


Figure used from :
https://arxiv.org/pdf/1611.01604.pdf

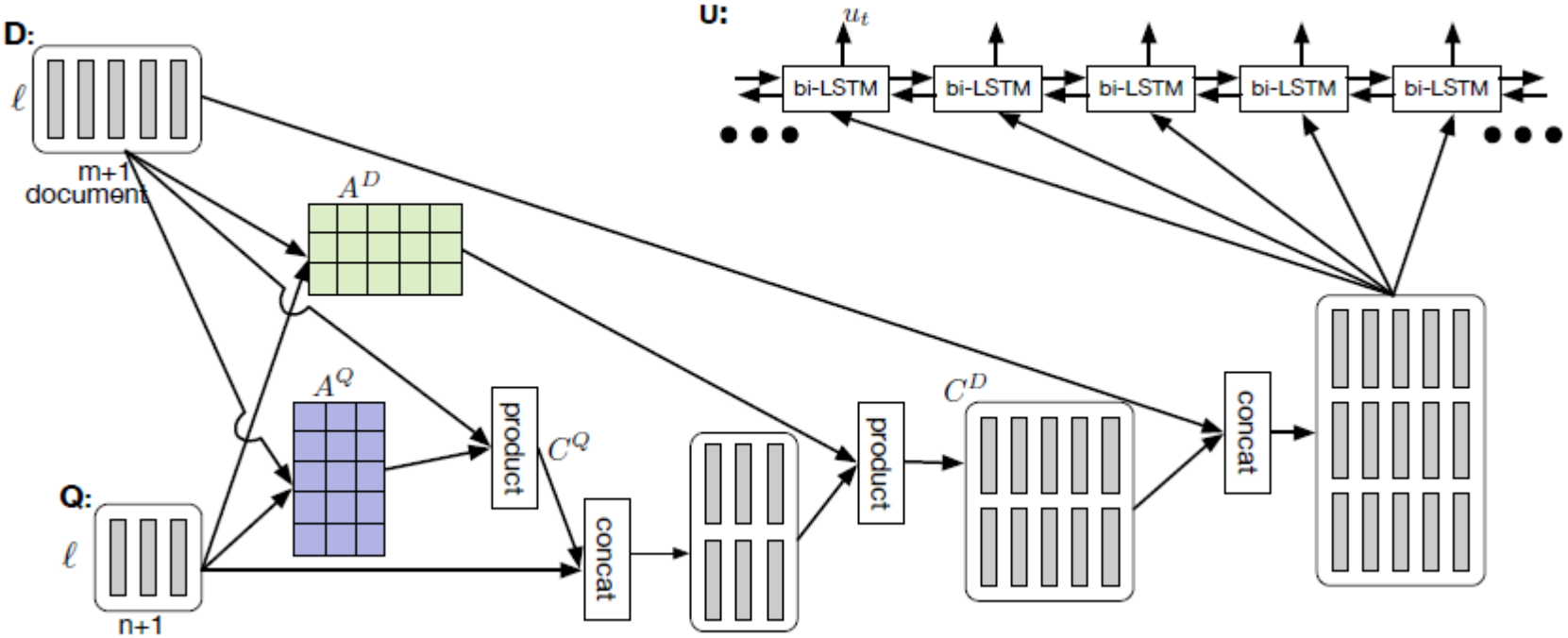
Design

- Preprocessing
- Glove Embeddings : 300 - dim
- Encoder
  - Question and document encoding
  - Co-attention encoding
- Decoder



Encoder

Figure used from : https://arxiv.org/pdf/1611.01604.pdf



Decoder

Figure used from : https://arxiv.org/pdf/1611.01604.pdf

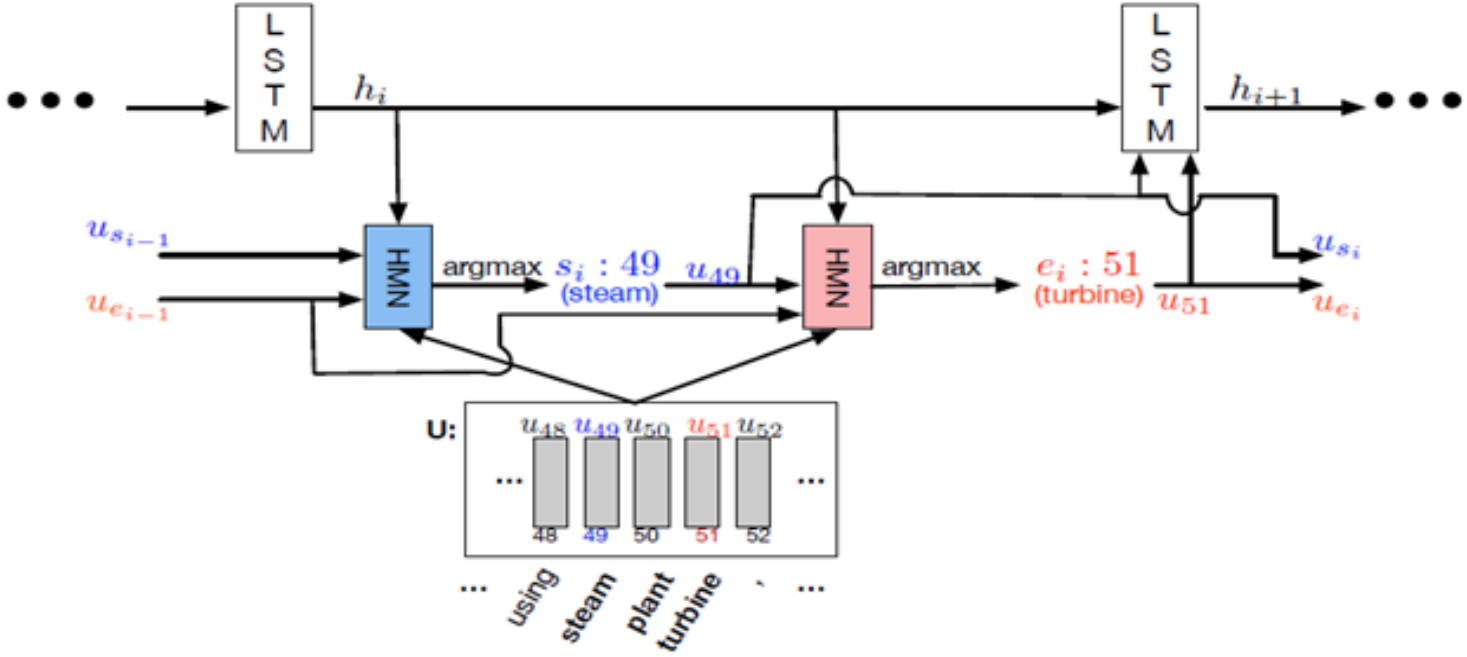


Figure 2. Decoder

Design (Encoder)

Document Matrix

$$D = [d_1, d_2, d_3, d_4, \ldots d_n] \in R^{l \times (m+1)}$$

Encoded Document :

$$d_t = \text{enc-LSTM}\ (d_{t-1}, x_t^d)$$

Encoded Question :

$$q_t = \text{enc-LSTM}\ (q_{t-1}, x_t^q) \in R^{l \times (m+1)}$$

Design (Encoder) contd..

Intermediate Question:

$$Q' = [q_1, q_2, \ldots, q_n] \in R^{l \times (n+1)}$$

Question Representation:

$$Q = \tanh (W^{(Q)}Q' + b^{(Q)}) \in R^{l \times (n+1)}$$

Coattention Encoder

$$L = D^T Q \in R^{(m+1) \times (n+1)}$$

Design (Encoder) Contd..

Affinity Matrix

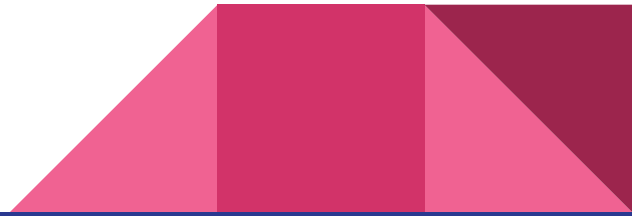
$$A^Q = \text{softmax}(L) \in R^{(m+1) \times (n+1)}$$

$$A^D = \text{softmax}(L) \in R^{(n+1) \times (m+1)}$$

Attention contexts

$$C^Q = D\,A^Q \in R^{l \times (m+1)}$$

$$C^D = Q\,A^D \in R^{l \times (m+1)}$$



Design (Encoder) contd..

Summeries:

$$C^D = [Q ; C^Q] A^D \in R^{2l \times (m+1)}$$

Fusing Temporal Info:

$$U_t = \text{Bi-LSTM}(U_{t-1}, U_{t+1}, [d_t ; c_t^D]) \quad \in R^{2l}$$

Where, $U = [u_1, u_2, u_3, \ldots u_m] \in R^{2l \times m}$

Design (Decoder)

$$h_i = \text{dec-LSTM}\left(h_{i-1} , [u_{si-1} ; u_{ei-1}] \right)$$

New start and end pos:

$$s_i = \text{argmax}\left(\alpha_{i,} \ldots. \alpha_{m}\right)$$

$$e_i = \text{argmax}\left(\beta_{i,} \ldots. \beta_{m}\right)$$

Where, $\alpha_t = \text{HMN}_{start}(u_t, h_i, u_{si-1}, u_{ei-1})$

Highway Maxout Networks

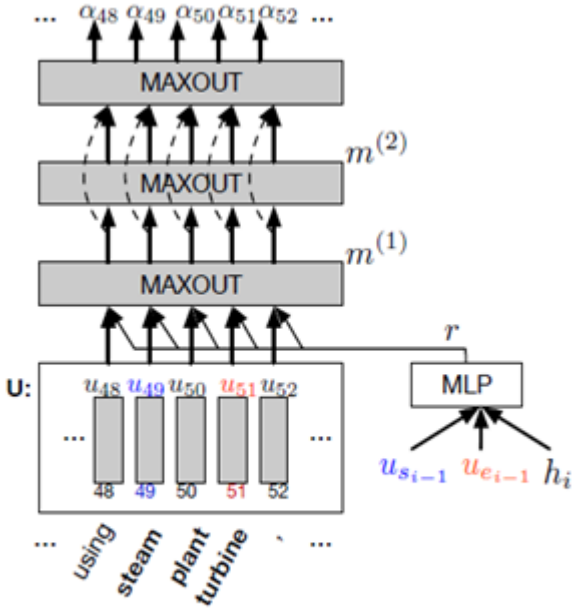


Figure 3. Highway Maxout Network

$$HMN(u_t, h_i, u_{si-1}, u_{ei-1}) = MAX(W^{(3)}[m_t^{(1)}; m_t^{(2)}] + b^{(3)})$$

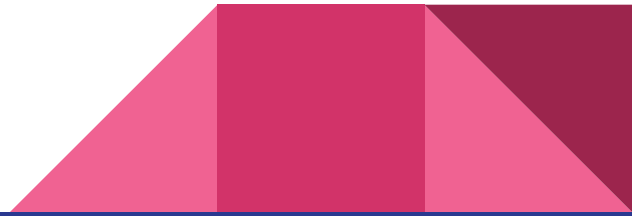
$$r = \tanh(W^{(D)}[h_i^{(1)}; u_{si-1;} u_{ei-1}])$$

$$m_t^{(1)} = MAX(W^{(1)}[u_t; r + b^{(1)})$$

$$m_t^{(2)} = MAX(W^{(2)} m_t^{(1)} + b^{(2)})$$

Experimental Setup

- Split word and created a dictionary
- Word is mapped to glove vectors
- Divided data into 50 epochs (1356 sized)
- Max_len = 600
- Hidden dim = 200
- Batch size = 64
- LR = 0.001
- LSTM assigned zero hidden state
- Softmax cross entropy
- Adam optimizer



Result

Set	F1 (Our Model)	F1 (Baseline)
Dev	8.6	2.3
Test	4.9	1.8

Sample Output 1

Document : British colonial rule introduced Western elements of culture to Burma . Burma 's education system is modelled after that of the United Kingdom . Colonial architectural influences are most evident in major cities such as Yangon . Many ethnic minorities , particularly the Karen in the southeast and the Kachin and Chin who populate the north and northeast , practice Christianity . According to the The World Factbook , the Burman Document 1 : British colonial rule introduced Western elements of culture to Burma . Burma 's education system is modelled after that of the United Kingdom . Colonial architectural influences are most evident in major cities such as Yangon . Many ethnic minorities , particularly the Karen in the southeast and the Kachin and Chin who populate the north and northeast , practice Christianity . According to the The World Factbook , the Burman population is 68 % and the ethnic groups constitute 32 % . However , the exiled leaders and organisations claims that ethnic population is 40 % , which is implicitly contrasted with CIA report (official US report) .

Question:
Do opinions differ on the numbers for that population is considered not a part of the mainstream?
Prediction : According to the The World Factbook , the Burman population is 68 % and the ethnic groups constitute 32 % . However, the exiled leaders and and organisations claims that ethnic population is 40 %.
Groud Truth : exiled leaders and organisations claims that ethnic population is 40 % , which is implicitly contrasted with CIA report (official US report) .

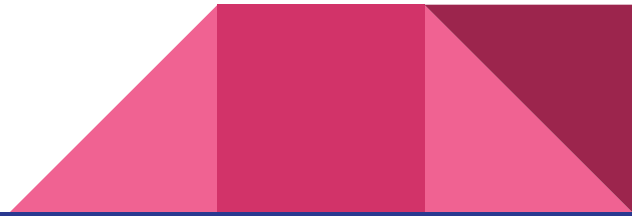
Sample Output 2

Document 2 :
Food is passed from the kitchen to the stube , where the dining room table is placed . Some meals are communal , such as fondue , where a pot is set in the middle of the table for each person to dip into . Other meals are still served in a traditional manner on carved wooden plates . Furniture has been traditionally elaborately carved and in many Alpine countries carpentry skills are passed from generation to generation .

Question :
The food is passed from the kitchen to where ?
Prediction : Food is passed from the kitchen to the stube , where the dining room table is placed . Some meals are communal , such as fondue , where a pot is set in
Groud Truth : the stube

Conclusion

- DCN has proved good results on question answering tasks.
- Need to run more iterations to see the actual results.



References

1. P. Rajpurkar, J. Zhang, K. Lopyrev, and P. Liang, “Squad: 100,000+ questions for machine comprehension of text,” arXiv preprint arXiv:1606.05250, 2016.

2. S. Hochreiter and J. Schmidhuber, “Long short-term memory,” Neural computation, vol. 9, no. 8, pp. 1735–1780, 1997.

3. C. Xiong, V. Zhong, and R. Socher, “Dynamic coattention networks for question answering,” CoRR, vol. abs/1611.01604, 2016.

4. M. Seo, A. Kembhavi, A. Farhadi, and H. Hajishirzi, “Bidirectional attention flow for machine comprehension,” arXiv preprint arXiv:1611.01603, 2016.

5. R. K. Srivastava, K. Greff, and J. Schmidhuber, “Highway networks,” arXiv preprint arXiv:1505.00387, 2015.

Thank you