Gated End-to-End Memory Networks

Fei Liu, Julien Perez

Improvements

This paper referes to the High-Way networdk idea and introduce the gated mechanism. By this way, it can utilize the information of memory dynamically.

Shotcut Connections

High-Way Network

$y = H(x) \odot T(x) - x \odot C(x)$
Here, T is the transform gate and C is the carry gate.Usually, C = 1 - T. So
$y = H(x) \odot T(x) + x \odot (1 - T(x))$

Residual Network

Residual Netword is a specially case of high-way network. T and C is 1
$y = H(x) + x$
Both of them can relief the gradient vanishing problem.

End-to-End Memory Networks

input context: $x_1, ..., x_n$

context representation: $m_i = A\Phi(x_i)$ $c_i = C\Phi(x_i)$

A and C are two embedding matrics.

Φ is a function that maps the input into a bag of dimension |V|.

question representation: $u = B\Phi(q)$

Attention Counting

$$p_i = softmax(u^T m_i)$$

Output

$$o = \sum_i p_i c_i$$

The next layer question input

$u^{k+1} = o^k + u^k$

The final output

$\hat{a} = softmax(Ww(o^K + u^K))$

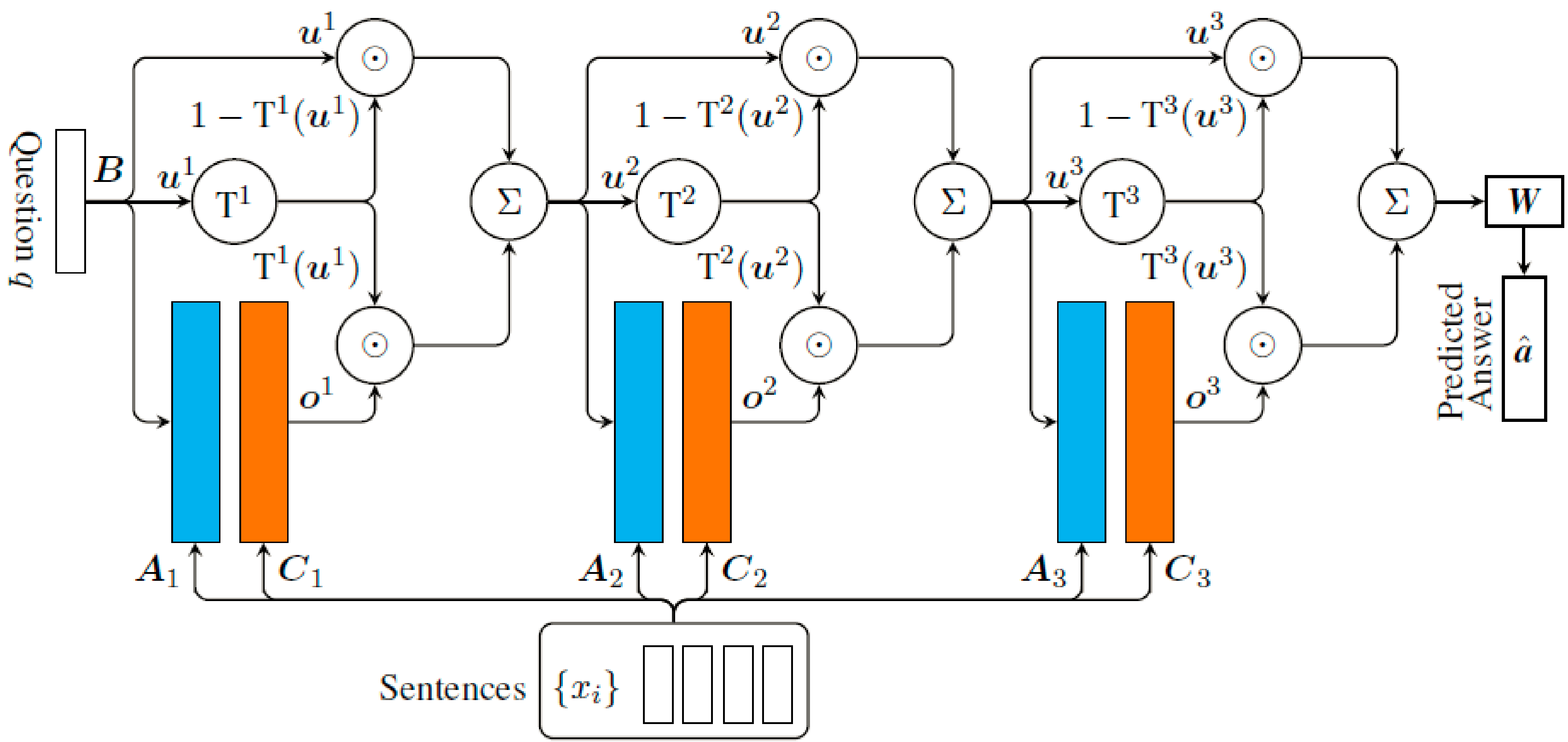
Here, $W \in R^{|V|*d}$

Gated-End-to-End Network

$$u^{k+1} = o^k \odot T^k(u^k) + u^k \odot (1 - T^k(u^k))$$

$$T^k(u^k) = \sigma(W_T^k u^k + b_T^k)$$

Model



Question q
B
u^1
T^1
$1 - T^1(u^1)$
$T^1(u^1)$
o^1
Σ
u^2
T^2
$1 - T^2(u^2)$
$T^2(u^2)$
o^2
u^3
T^3
$1 - T^3(u^3)$
$T^3(u^3)$
o^3
W
Predicted Answer
$\hat{a}$
A_1
C_1
A_2
C_2
A_3
C_3
Sentences
$\{x_i\}$