ACL2020

A Frame-based Sentence Representation for Machine Reading Comprehension

Shaoru Guo[1], **Ru Li**[1,2*], **Hongye Tan**[1,2], **Xiaoli Li**[3], **Yong Guan**[1], **Hongyan Zhao**[1] **and Yueping Zhang**[1]

1. School of Computer and Information Technology, Shanxi University, Taiyuan, China
2. Key Laboratory of Computational Intelligence and Chinese Information Processing of Ministry of Education, Shanxi University, Taiyuan, China
3. Institute for Infocomm Research, A*Star, Singapore

{guoshaoru0928,guanyong0130,13233587117}@163.com
{liru,tanhongye}@sxu.edu.cn,xlli@ntu.edu.sg
hongyanzhao@tyust.edu.cn

Main Work

- We propose novel attention-based **frame representation models**, which take full advantage of LUs and F-to-F relations to model frames with attention schema
- We propose a new **Frame-based Sentence Representation** (FSR) method that integrates multi-frame semantic information to obtain richer semantic aggregation for better sentence representation
- Our experimental results demonstrate our proposed frame-based sentence representation (FSR) method is **very effective on Machine Reading Comprehension (MRC) task**.

FrameNet

F	**Commerce_buy**
FEs	Buyer, Goods, ...
LUs	**buy.v**, buy.n, buyer.n, purchase.n,...
T	$[Katie]_{\textit{Buyer}}$ **bought**$_{\textit{Commerce\_buy}}$ $[some\ chocolate\ cookies]_{\textit{Goods}}$
F-to-F	Commerce_buy—-Shopping—Seeking—Locating

Table 1: Example of F, FEs, LUs, T and F-to-F.

Task

Passage	Katie went to the store...She looked around for the *flowers*. She wanted cookies not *chips*. She **found** some ***chocolate cookies***. Katie then looked for a *bow*. ...
Question	What snack did Katie **buy**?
Option	A) Chips B) Chocolate cookies C) Flowers D) Bows
Answer	B
Frame Semantic	{Chips , Chocolate cookies} $\in$ Food {Flowers , Bows}$\notin$Food **Found** and **Buy** have relations, as their frames are connected.

Frame Representation

- Lexical Units Aggregation Model (LUA)

$$F_m = \frac{1}{N} \sum_{U^{F_m}} u_n^{F_m}$$

- Lexical Units Attention Model (TLUA)

$$F_m = t^{F_m} + \sum_{\widetilde{U}^{F_m}} att(u_n^{F_m}) \cdot u_n^{F_m}$$

$$att(u_n^{F_m}) = \frac{exp(t^{F_m} \cdot u_n^{F_m})}{\sum_{u_k^{F_m} \in \widetilde{U}^{F_m}} exp(t^{F_m} \cdot u_k^{F_m})}$$

- Frame Relation Attention Model (FRA)

Frame Representation

- Frame Relation Attention Model (FRA)

$$F_m^* = F_m + \sum_{w=1}^{W} att(F_{m,w}) \cdot F_{m,w}$$

$$att(F_{m,w}) = \frac{exp(F_m \cdot F_{m,w})}{\sum_{k=1}^{W} exp(F_m \cdot F_{m,k})}$$

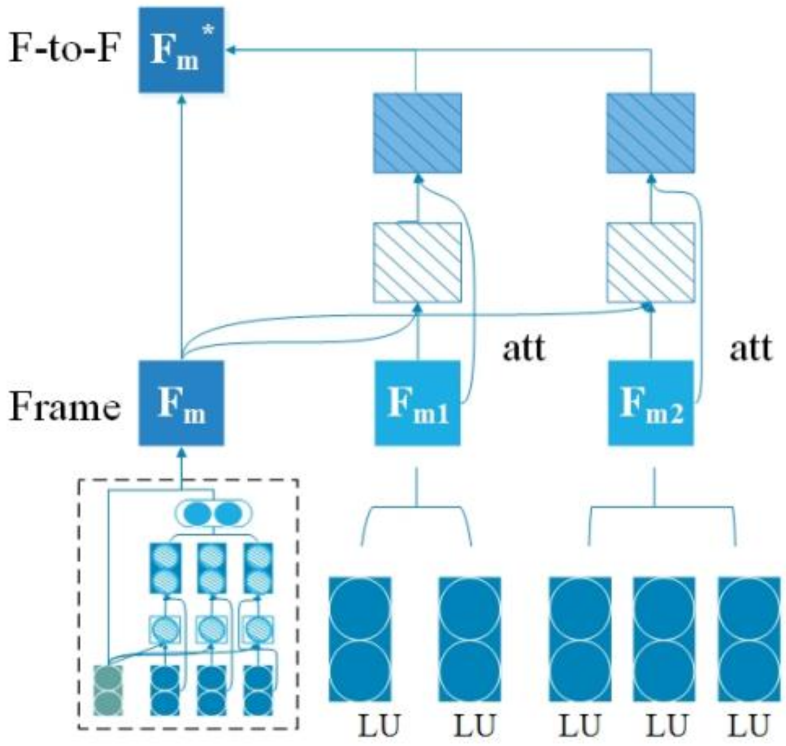

Figure 2: Frame Relation Attention Model.

Frame-based Sentence Representation

- define a frame semantic quadruple

$$c_k = < T_k, F_k, FE_{kn}, P_{kn} >$$

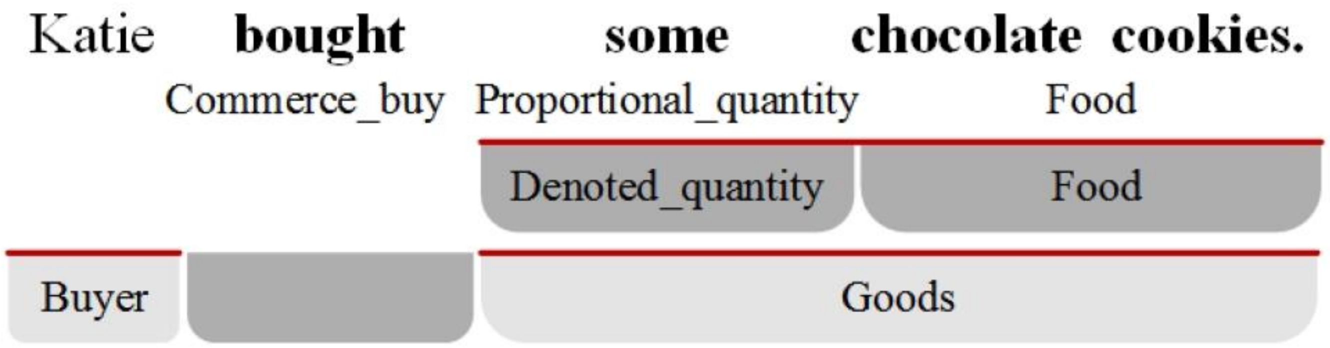

Figure 3: A sentence of FrameNet annotations.

The sentence s in Figure 3 has three quadruples:

1. c_1= <bought, *Commerce_buy*, [Buyer, Goods], [Katie, chocolate cookies]>

2. c_2= < some, *Proportional_quantity*, [Denoted_quantity], [some]>

3. c_3= <chocolate cookies, *Food*, [Food], [chocolate cookies]>

Frame-based Sentence Representation

The sentence s in Figure 3 has three quadruples:

1. c_1= <bought, *Commerce_buy*, [Buyer, Goods], [Katie, chocolate cookies]>

2. c_2= < some, *Proportional_quantity*, [Denoted_quantity], [some]>

3. c_3= <chocolate cookies, *Food*, [Food], [chocolate cookies]>

3.2 Frame Integration Representation

In Figure 4, c_k (k=1, 2, 3) is the input. We first compute its matrix representation c_k^t, with columns denoting different semantic information. Then, we formalize sentence representation as follows:

$$c^s = \mathcal{N}(c^t) \tag{6}$$

$$c^t = \phi(c_k^t, P_k) \quad (k = 1, \ldots, K) \tag{7}$$

Where K represents the total number of quadruples in the sentence. $\phi(c_k^t, P_k)$ is aggregate operation, used to form frame set representation c^t based on the information of P and T in the sequence. Finally, we encode sentence information by neural network models.

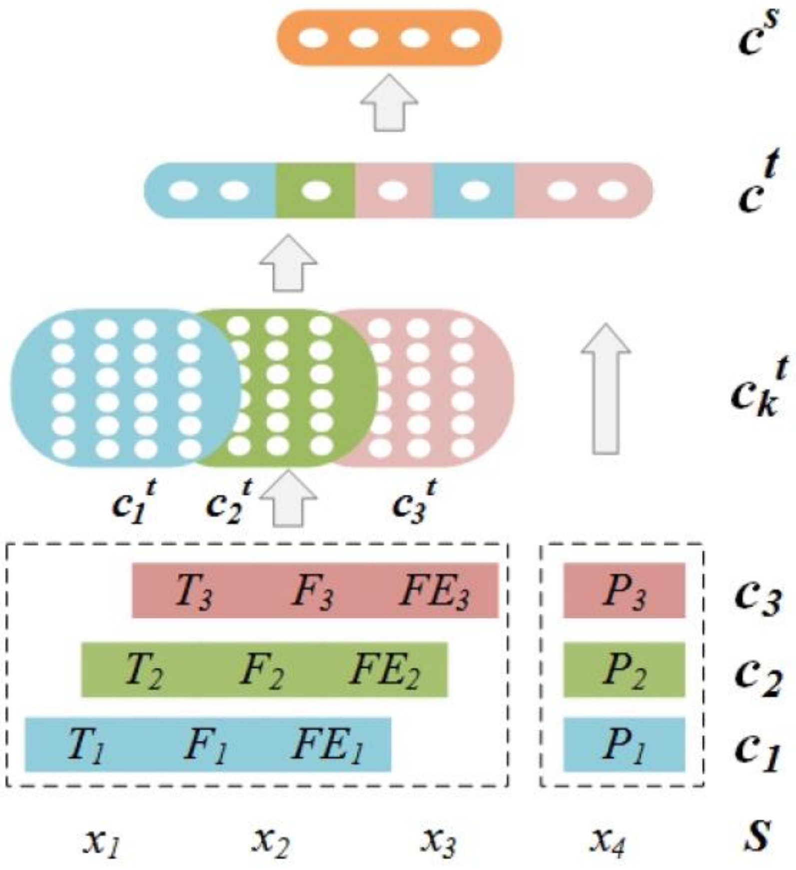

Figure 4: Frame Integration Representation Model.

Models for MRC

- we construct the input :
  - the passage as sequence A, and the concatenation of question and one choice of answer as sequence B
- train stragey:
  - we apply a linear layer and a softmax layer on the final hidden state, and maximize thelog-probability of correct labels during training.

Result

Method	MCTest-160 (%)	MCTest-500 (%)
Richardson et al. (2013)	69.16	63.33
Wang et al. (2015)	75.27	69.94
Li et al. (2018)	74.58	72.67
Attentive Reader (Hermann et al., 2015)	46.3	41.9
Neural Reasoner (Peng et al., 2015)	47.6	45.6
Parallel-Hierarchical (Trischler et al., 2016)	74.58	71.00
Reading Strategies (Sun et al., 2018)	81.7	82.0
Bert (Zhang et al., 2019)	73.8	80.4
BERT+DCMN+ (Zhang et al., 2019)	85.0	86.5
FSR	86.1	84.2

Table 2: The Performance Comparison of 10 Different Models on Two MCTest Datasets.

Result

Method	160 (%)	500 (%)
Bert (Zhang et al., 2019)	73.8	80.4
Bert (Our implementation)	82.5	80.9
Bert+LUA	82.7	79.5
Bert+TLUA	84.6	82.7
Bert+FRA	86.1	84.2
bi-LSTM	54.2	49.5
bi-LSTM+LUA	59.4	57.5
bi-LSTM+TLUA	61.5	58.2
bi-LSTM+FRA	62.7	59.6

Table 3: Performance Comparison with Three Differen-
t Frame Representation Models.