A Frame-based Sentence Representation for Machine Reading Comprehension

Shaoru Guo\textsuperscript{1,}, Ru Li\textsuperscript{1,2}, Hongye Tan\textsuperscript{1,2}, Xiaoli Li\textsuperscript{3}, Yong Guan\textsuperscript{1} and Hongyan Zhao\textsuperscript{1}

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

\{guoshaoaru0928, guanyong0130\}@163.com, liru@sxu.edu.cn
xlli@ntu.edu.sg, hytan_2006@126.com, hongyanzhao@tyust.edu.cn

Abstract

Sentence representation (SR) is the most crucial and challenging task in Machine Reading Comprehension (MRC). MRC systems typically only utilize the information contained in the sentence itself, while human beings can leverage their semantic knowledge. To bridge the gap, we proposed a novel Frame-based Sentence Representation (FSR) method, which employs Frame semantic knowledge to facilitate sentence modelling. Specifically, different from existing methods that only model lexical units (LUs), a Frame Representation Models, which utilize both LUs in frame and Frame-to-Frame (F-to-F) relations, are designed to model frames and sentences with attention schema. Our proposed FSR method is able to integrate multi-frame semantic information to get much better sentence representations. Our extensive experimental results show that it performs better than state-of-the-art technologies on machine reading comprehension task.

1 Introduction

Machine Reading Comprehension (MRC) requires machines to read and understand a text passage, and answer relevant questions about it. Human beings can easily understand the meaning of a sentence based on their semantic knowledge. For instance, given a sentence Katie bought some chocolate cookies, people know Katie is a buyer, chocolate cookies are goods and belong to Food class etc. Existing machine learning approach, however, face great challenges to address complicated MRC questions, as they do not have above semantic knowledge.

Nevertheless, FrameNet (Baker et al., 1998), as a knowledge base, provides schematic scenario representation that could be potentially leveraged to better understand sentences. It enables the development of wide-coverage frame parsers (Gildea and Jurafsky, 2002; Das et al., 2014), as well as various real-world applications, ranging from event recognition (Liu et al., 2016), information Extraction (Burchardt et al., 2009), question answering (Ofoghi et al., 2009) and narrative schemas (Chambers and Jurafsky, 2010), etc. In particular, Frame (F) is defined as a composition of Lexical Units (LUs) and a set of Frame Elements (FEs). Given a sentence, if its certain word evokes a frame by matching a LU, then it is called Target (T). It is worth mentioning that FrameNet arranges different relevant frames into a network by defining frame-to-frame (F-to-F) relations. Table 1 provides an example of F, FEs, LUs, T and F-to-F.

<table>
<thead>
<tr>
<th>F</th>
<th>Commerce_buy</th>
</tr>
</thead>
<tbody>
<tr>
<td>FEs</td>
<td>Buyer, Goods, ...</td>
</tr>
<tr>
<td>LUs</td>
<td>buy_n, buy_v, buyer_n, purchase_n, ...</td>
</tr>
<tr>
<td>T</td>
<td>[Katie]_Buyer _bought_Commerce_buy</td>
</tr>
<tr>
<td></td>
<td>[some chocolate cookies]_Food</td>
</tr>
<tr>
<td>F-to-F</td>
<td>Commerce_buy—-Shopping—</td>
</tr>
<tr>
<td></td>
<td>Seeking—-Locating</td>
</tr>
</tbody>
</table>

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

How to utilize semantic knowledge from Framenet? We observe the existing works mainly focus on LU vector embedding within a frame (Hermann and Blunsom, 2014; Bojanowski et al., 2017; Glavas et al., 2019), without modeling a frame as a whole. In addition, many sentences could have more than one target words that will evoke multiple frames.
frame semantic knowledge, including both generalizations of LUs and F-to-F relations, to better model sentences. The key contributions of this work are summarized as follows:

1. We propose novel attention based frame representation models, which take full advantage of LUs and F-to-F relations to model sentences with attention schema.

2. We propose a new Frame-based Sentence Representation (FSR) that integrates multi-frame semantic information to obtain richer semantical aggregation for better sentence representation.

3. Our experimental results demonstrate our proposed frame-based sentence representation (FSR) method is very effective on Machine Reading Comprehension (MRC) task.

2 Frame Representation Model

In this section, we present our Frame Representation Model, considering both LUs and F-to-F.

Let $F = \{F_1, F_2, \ldots, F_m, \ldots\}$ represent a set of all frames, where $F_m \in \mathcal{R}^H$ is the representation of $m$-th frame of $F$. Let $U^{F_m} = \{u_1^{F_m}, u_2^{F_m}, \ldots, u_{N}^{F_m}, \ldots\}$ be the LUs set of $F_m$, where $U^{F_m} \in \mathcal{R}^{(H \cdot N)}$, $N$ stands for the total number of LUs in $F_m$, and $u_n^{F_m}$ be the $n$-th LU of $F_m$. $t^{F_m}$ is a target word, matching a LU in $F_m$. We proposed 3 different frame representation models.

2.1 Lexical Units Aggregation Model (LUA)

Lexical Units Aggregation Model (LUA) is a straightforward idea. Given a frame $F_m$, it averages all its underlying LU representation $u_n^{F_m}$ ($u_n^{F_m} \in U^{F_m}$) to represent the frame entirely:

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

2.2 Lexical Units Attention Model (TLUA)

Each frame in above LUA model has the same representation for different sentences, as they do not distinguish the importance of each LU in the frame. To address this issue, we propose TLUA model, utilizing an attention scheme to automatically weight different LUs for the frame, according to target word $T$ in given sentence, shown in Figure 1.

More specifically, we compute the weighted sum of target word $T$’s representation and other LUs’ presentations based on their importance wrt $T$. In other words, we emphasize $T$ as it occurs in the given sentence, which can reduce the potential noise introduced by irrelevant LUs in the same frame.

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

$$\text{att}(u_n^{F_m}) = \frac{\text{esp}(t^{F_m} \cdot u_n^{F_m})}{\sum_{\tilde{U}^{F_m}} \text{esp}(t^{F_m} \cdot u_n^{F_m})}$$

Here, $\tilde{U}^{F_m}$ represents the LUs set of $F_m$ which is not include $t^{F_m}$, and $\tilde{U}^{F_m} \in \mathcal{R}^{H \cdot (N-1)}$.

2.3 Frame Relation Attention Model (FRA)

The key problem in MRC is to analyze semantic relations among multiple sentences. As such, we propose a novel FRA model, which takes advantage of F-to-F relations to get much richer semantic information, shown in Figure 2.

Given frame $F_m$, $F_m^+ = \{F_{m,1}, \ldots, F_{m,w}, \ldots\}$ represents its expanded frames, including all the
frames that can be linked to $F_m$ through F-to-F relation chains in Framenet, with no more than 3 hops to only keep close relations. Note attention schemes have been designed for both intra-frame and inter-frames. Particularly, intra-frame attention focuses on relevant LUs, while inter-frame attention emphasizes relevant frames, avoiding the influence from less relevant but linked frames.

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

$$\text{att}(Z_{m,w}) = \frac{\text{esp}(F_m \cdot F_{m,w})}{\sum_{w=1}^{W} \text{esp}(F_m \cdot F_{m,w})}$$

3. **Frame-based Sentence Representation**

Given a sentence $s = \{x_1, x_2, \ldots, x_k, \ldots\}$ where each $x_k$ is a word, let $T_k$ be the $k$-th frame-evoking target of $s$, and $T_k$ evokes $F_k$ frame. $FE_{ki}$ denotes the $i$-th frame element of $F_k$, and $P_{ki}$ denotes the $i$-th word fulfilling $FE_{ki}$. We define a frame semantic quadruple $c_k = \langle T_k, F_k, FE_{kn}, P_{kn} \rangle$, where $c_k$ represents the $k$-th quadruple of $s$.

3.1 Sentence Semantic Annotations with Multiple Frames

In this paper, we employ SEMAFOR (Das et al., 2014) to automatically process sentences with multiple semantic annotations (Kshirsagar et al., 2015).

Figure 3 provides an example sentence with three $T$, namely **bought**, **some**, **chocolate cookies**. Each $T$ has its evoked semantic frame right below it. For each frame, its $FE$ are shown enclosed in the block where dark grey indicates the corresponding $T$, and the words fulfilling the $FE$s are connected to the corresponding text. For example, $T$ **bought** evokes the **Commerce buy** frame, and has the **Buyer**, **Goods** $FE$s fulfilled by **Katie** and **some chocolate cookies**.

The sentence $s$ in Figure 3 has three quadruples:

1. $c_1 = \langle \text{bought}, \text{Commerce buy}, [\text{Buyer}, \text{goods}], [\text{Katie}, \text{chocolate cookies}] \rangle$
2. $c_2 = \langle \text{some}, \text{Relational quantity}, [\text{Denoted quantity}], [\text{some}] \rangle$
3. $c_3 = \langle \text{chocolate cookies}, \text{Food}, [\text{Food}], [\text{chocolate cookies}] \rangle$

3.2 Frame Integration Representation

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

$$s = \frac{1}{K} \sum_{k=1}^{K} c^s_k$$

$$c^s_k = \phi(c^t_k)$$

Where $K$ represents the total number of quadruples in the sentence. $c^s_k = \phi(c^t_k)$ is max pooling operation, used to form frame set representation $c^s_k$. Finally, we encode sentence information by averaging of all quadruple representations in it.

4. **Experiments**

4.1 Model for MRC

To better analyze the performance of our proposed method on MRC, we apply both BERT (Devlin et al., 2018) and LSTM (Hochreiter and Schmidhuber, 1997) as our neural model. Also, we concatenate given passage and a question as sentence A, while each candidate answer as sentence B. In addition, we apply a linear layer and a softmax layer on the final hidden state, and maximize the log-probability of correct labels during training.

4.2 Datasets for MRC

We employ MCTest (Richardson et al., 2013) to test the system performance of multiple choice machine comprehension task. It consists of two data sets, namely MC160 and MC500.
<table>
<thead>
<tr>
<th>Method</th>
<th>MCTest-160 (%)</th>
<th>MCTest-500 (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Richardson et al. (2013)</td>
<td>69.16</td>
<td>63.33</td>
</tr>
<tr>
<td>Wang et al. (2015)</td>
<td>75.27</td>
<td>69.94</td>
</tr>
<tr>
<td>Li et al. (2018)</td>
<td>74.58</td>
<td>72.67</td>
</tr>
<tr>
<td>Attentive Reader (Hermann et al., 2015)</td>
<td>46.3</td>
<td>41.9</td>
</tr>
<tr>
<td>Neural Reasoner (Peng et al., 2015)</td>
<td>47.6</td>
<td>45.6</td>
</tr>
<tr>
<td>Parallel-Hierarchical (Trischler et al., 2016)</td>
<td>74.58</td>
<td>71.00</td>
</tr>
<tr>
<td>Reading Strategies (Sun et al., 2018)</td>
<td>81.7</td>
<td>82.0</td>
</tr>
<tr>
<td>Bert (Zhang et al., 2019)</td>
<td>73.8</td>
<td>80.4</td>
</tr>
<tr>
<td>Bert+DCMN (Zhang et al., 2019)</td>
<td>85.0</td>
<td>86.5</td>
</tr>
<tr>
<td>FSR</td>
<td>86.1</td>
<td>84.2</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Method</th>
<th>160 (%)</th>
<th>500 (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bert</td>
<td>73.8</td>
<td>80.4</td>
</tr>
<tr>
<td>Bert+LUA</td>
<td>81.7</td>
<td>79.5</td>
</tr>
<tr>
<td>Bert+TLUA</td>
<td>84.6</td>
<td>82.7</td>
</tr>
<tr>
<td>Bert+FRA</td>
<td>86.1</td>
<td>84.2</td>
</tr>
<tr>
<td>bi-LSTM</td>
<td>54.2</td>
<td>49.5</td>
</tr>
<tr>
<td>bi-LSTM+LUA</td>
<td>59.4</td>
<td>57.5</td>
</tr>
<tr>
<td>bi-LSTM+TLUA</td>
<td>61.5</td>
<td>58.2</td>
</tr>
<tr>
<td>bi-LSTM+FRA</td>
<td>62.7</td>
<td>59.6</td>
</tr>
</tbody>
</table>

Table 3: Performance Comparison with Three Different Frame Representation Models

4.3 Experiment Results

Table 2 shows our FSR model achieves 86.1% accuracy on MCTest-160, which is significantly better than all the nine state-of-the-art methods. In addition, it also achieves very competitive results on MCTest-500, i.e., much better than eight existing methods, slightly worse than bert+DCMC model. This is encouraging, as our model is much simpler and more efficient than bert+DCMC, which uses much more sophisticated architecture.

Recall in Section 2, we proposed three different methods, namely, LUA, TLUA, TFA, for frame representation. Table 3 shows their detailed results:

(1) No matter for BERT or bi-LSTM, if we add frame semantic information, the performance improves by several percents, indicating frame information is valuable in semantic understanding.

(2) Comparing TLUA with LUA, TLUA performs better, signifying attention scheme in TLUA can capture semantic information more accurately.

(3) Finally, TFA further improves LUA and TLUA’s performance, as sentences within a passage typically have semantic connections with each other, and it is thus necessary to take advantage of F-to-F relations to enrich semantic information.

4.4 Case Study

For case study, Table 4 shows an example in MCTest, where we are able to answer it correctly. Both Chips, Chocolate cookies belong to the Food frame, while Flowers and Bows evoke two different frames Plants and Accoutrements respectively. The target words Found and Buy in the given passage/question evoke different frames Location and Commerce buy — note in framenet they are connected due to their semantic relations, facilitating us to find answer B) Chocolate cookies.

5 Conclusion

We propose a novel frame-based sentence representation model, which, to our best knowledge, is the first work in this area. Our extensive experimental results demonstrate it works very well for the challenging machine reading comprehension task.
References


