Multi-Task Learning for Document Ranking and Query Suggestion

Ahmad et al., ICLR 2018
Speaker : Semin Choi

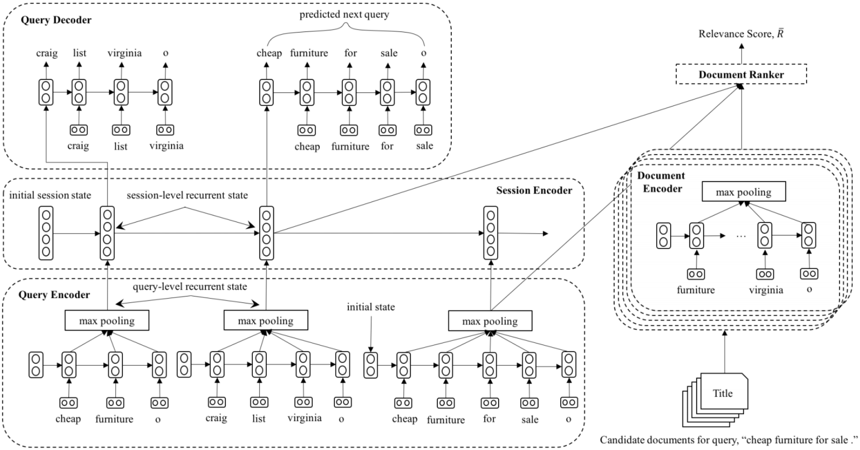
Department of Statistics, Seoul National University, South Korea

April 26, 2018

Multi-Task Neural Session Relevance Framework

- $Q = \{Q_1, ..., Q_n\}$: a sequence of queries.
- $D = \{D_1, ..., D_m\}$: a set of related documents.
- $o = \{o_1, ..., o_m\}$: a set of relavance labels for each document in D.
- A query Q_i and a document D_j consist of a sequence of words:
  - $Q_i = \{w_i^1, ..., w_i^q\}$
  - $D_j = \{w_j^1, ..., w_j^d\}$
- V is the size of vocabulary constructed over queries and relevant documents.

Multi-Task Neural Session Relevance Framework



Multi-Task Neural Session Relevance Framework

Query Encoder

- We follow to adopt a bidirectional LSTM with max-pooling(BiLSTM-max), due to its superior practical performance.
- The encoder composed of forward and backward LSTM reads the sequence reads the sequence in two opposite directions,

$$\overrightarrow{h}_t = LSTM(\overrightarrow{h}_{t-1}, w_i^t), \quad \overleftarrow{h}_t = LSTM(\overleftarrow{h}_{t+1}, w_i^t), \quad h_t = [\overrightarrow{h}_t, \overleftarrow{h}_t]$$

  where $h_t \in \mathbb{R}^{2d_q}$ is the query-level recurrent state, d_q is the dimensionality of the LSTM hidden unit.
- To form a fixed-size vector representation of variable length queries:

$$Q_{i,k} = \max_{1 \leq j \leq q} h_{k,j}, \quad k = 1, ..., 2d_q.$$

  where $Q_{i,k}$ is the k-th element of the latent vector Q_i.

Multi-Task Neural Session Relevance Framework

Document Encoder

- We use the same BiLSTM-max technique that is utilized in encoding queries.
- The only difference is in the dimensionality of the LSTM hidden units.

Session Encoder

- We use a unidirectional LSTM for session encoding.

$$S_i = LSTM(S_{i-1}, Q_i)$$

where $S_i \in \mathbb{R}^{d_s}$ is the session-level recurrent state.

Document Ranker

$$P(y_{ij} = 1 | Q_i, S_{i-1}) = \sigma(D_j^T \tanh(W_r[Q_i, S_{i-1}] + b_r)),$$
$$j = 1, ..., m$$

where $y_{ij} = 1$ if the j-th document is relevant for query Q_i , $W_r \in \mathbb{R}^{(d_q+d_s)\times d_d}$, $b_r \in \mathbb{R}^{d_d}$, and d_q, d_s and d_d are the dimensionality of the query encoder, session encoder and document encoder hidden units.

Multi-Task Neural Session Relevance Framework

Query Recommender

- Basically the query recommender module estimates the probability of the next query $Q_i = \{w_i^1, ..., w_i^q\}$, given all the previous queries up to position $i-1$ in a session.
- Information about all the previous queries represented through a session vector S_i is passed to the query recommender.

$$h_0 = \tanh(W_q S_i + b_q),$$

where $h_0 \in \mathbb{R}^{d_r}$ is the initial recurrent state.

$$h_t = LSTM(h_{t-1}, w_i^t),$$

where h_{t-1} is the previous hidden state, w_i^t is the current query term.

$$P(w_i^t = w | w_i^{1:t-1}, Q_{1:t-1}) = g(W_p h_t + b),$$

where $g(z)_j = \frac{e^{z_j}}{\sum_{k=1}^{K} e^{z_k}}, j = 1, ..., K.$

Learning End-to-end

- The binary cross entropy form the document ranker :

$$L_{1i} := -\frac{1}{m}\sum_{j=1}^{m}\Big\{o_{ij}\log P(y_{ij}=1|Q_i,S_{i-1})$$

$$+(1-o_{ij})\log(1-P(y_{ij}=1|Q_i,S_{i-1}))\Big\}$$

  where o_{ij} represents binary click label for the query i and the document j.

- The regularized negative log-likelihood loss from the query suggestion model :

$$\begin{aligned} L_{2i} \quad := \quad & -\sum_{t}^{q}\log P(w_i^t|w_i^{1:i-1},Q_{1:i-1}) \\ & -\lambda\sum_{w\in V}P(w|w_i^{1:t-1},Q_{1:i-1})\log P(w|w_i^{1:t-1},Q_{1:i-1}) \end{aligned}$$

- The final objective :

$$\sum_i (L_{1i}+L_{2i})$$