Context-Aware End-To-End Relation Extracting From Clinical Texts With Attention-Based Bi-Tree-GRU

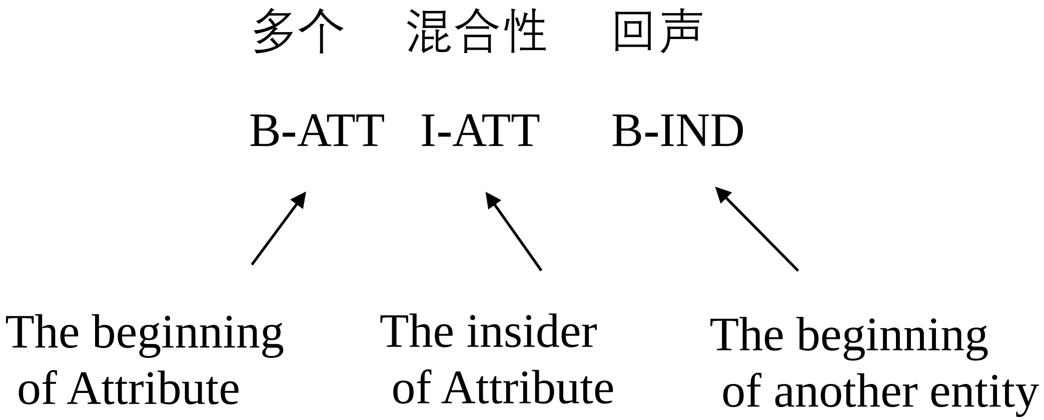
Dehua Chen, Yunying Wu, Jiajin Le, Qiao Pan
Donghua University, Shanghai, China

Tagging Scheme:

(B) Beginning

(I) Insider

(O) Outsider



Entity:

(LOC)Anatomical Location: right lobe /left lobe/narrow isthmus…

(IND)Index： echos/ nodule/size…

(ATT)Attribute： mixed/ multiple…

Relation:

(Loc-Ind)Location-Index

(Ind-Att) Index-Attribute

(Ind-Sub Ind) Index-Sub Index

(U)Unknown

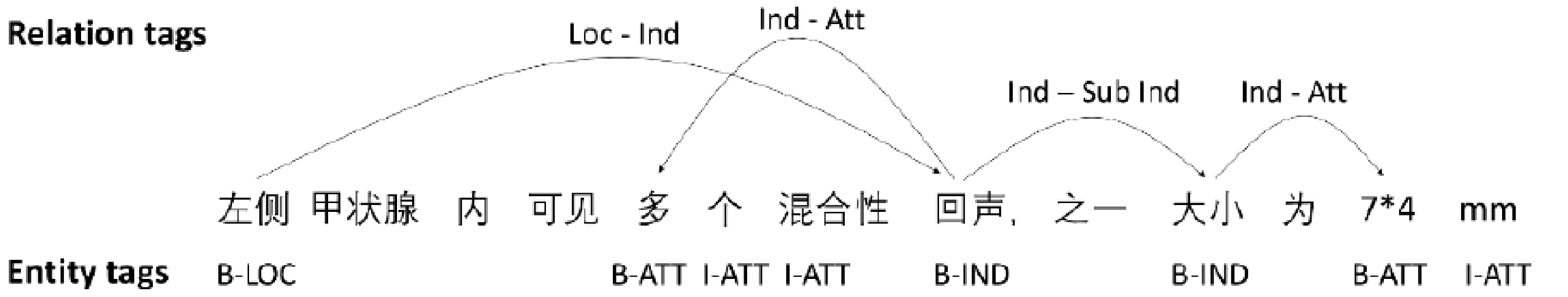


Fig. 1. The annotation of an example sentence from the thyroid ultrasound reports. Based on the word segmentation with correction by completing the vocabulary set, words are annotated as Anatomical Location (LOC), Index (IND) or Attribute (ATT) in BIO (Beginning, Insider, Outsider) tagging scheme. The relation tags are annotated as Location-Index (Loc-Ind), Index-Attribute (Ind-Att), Index-Sub Index (Ind-Sub Ind) or Unknown.

Correct type of entity

Correct boundary of entity → **Complicated** in Chinese Clinical Texts

Correct relation

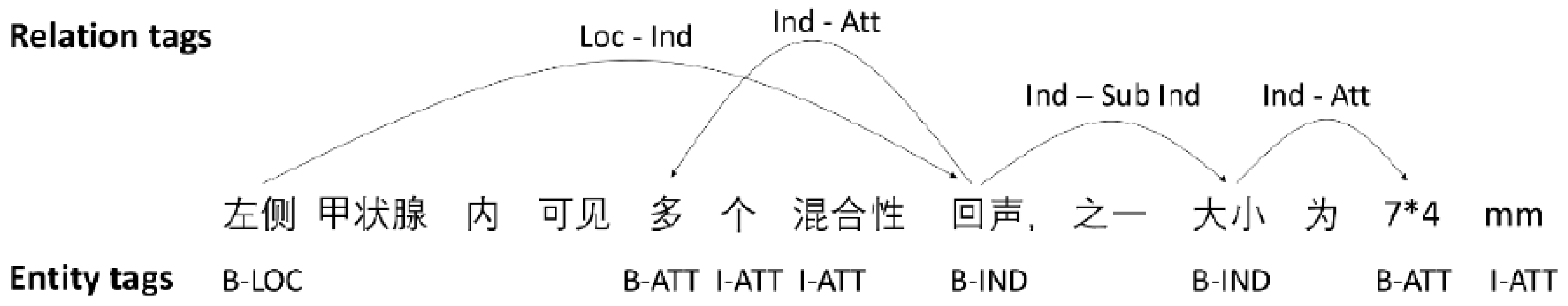


Fig. 1. The annotation of an example sentence from the thyroid ultrasound reports. Based on the word segmentation with correction by completing the vocabulary set, words are annotated as Anatomical Location (LOC), Index (IND) or Attribute (ATT) in BIO (Beginning, Insider, Outsider) tagging scheme. The relation tags are annotated as Location-Index (Loc-Ind), Index-Attribute (Ind-Att), Index-Sub Index (Ind-Sub Ind) or Unknown.

Input Representation

Preparation 1:

A sentence

↓

Dependency Parsing Tree

Preparation 2:

Three elements of words

(word/Part Of Speech/Dependency Relation)

Dependency Relation resulted from Dependency Parsing Tree

↓

Embeddings

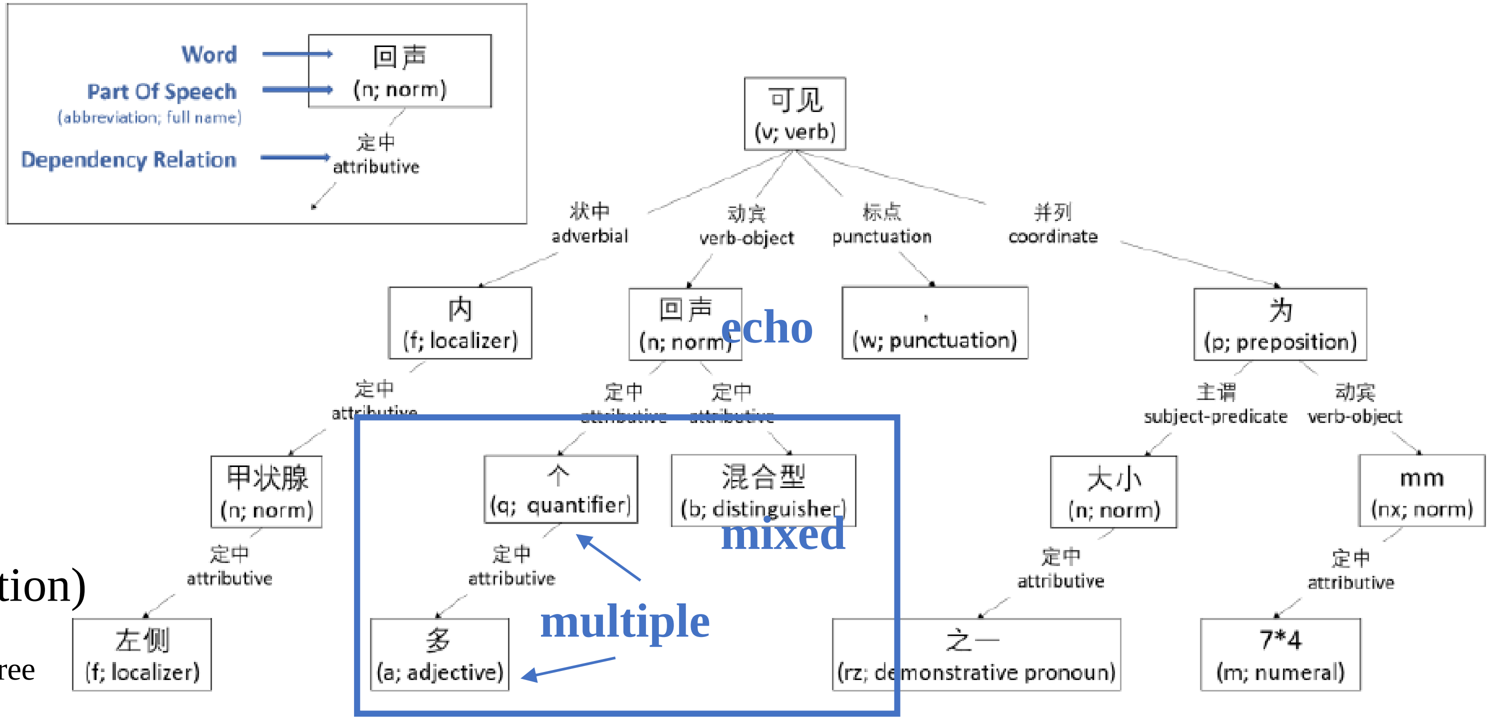


Fig. 2. An example of dependency parsing tree. A rectangular indicates a word node and part of speech tag (POS) is under the word, the strings covering the arrows denote the syntactic dependency relation.

Clinical Entity Extraction

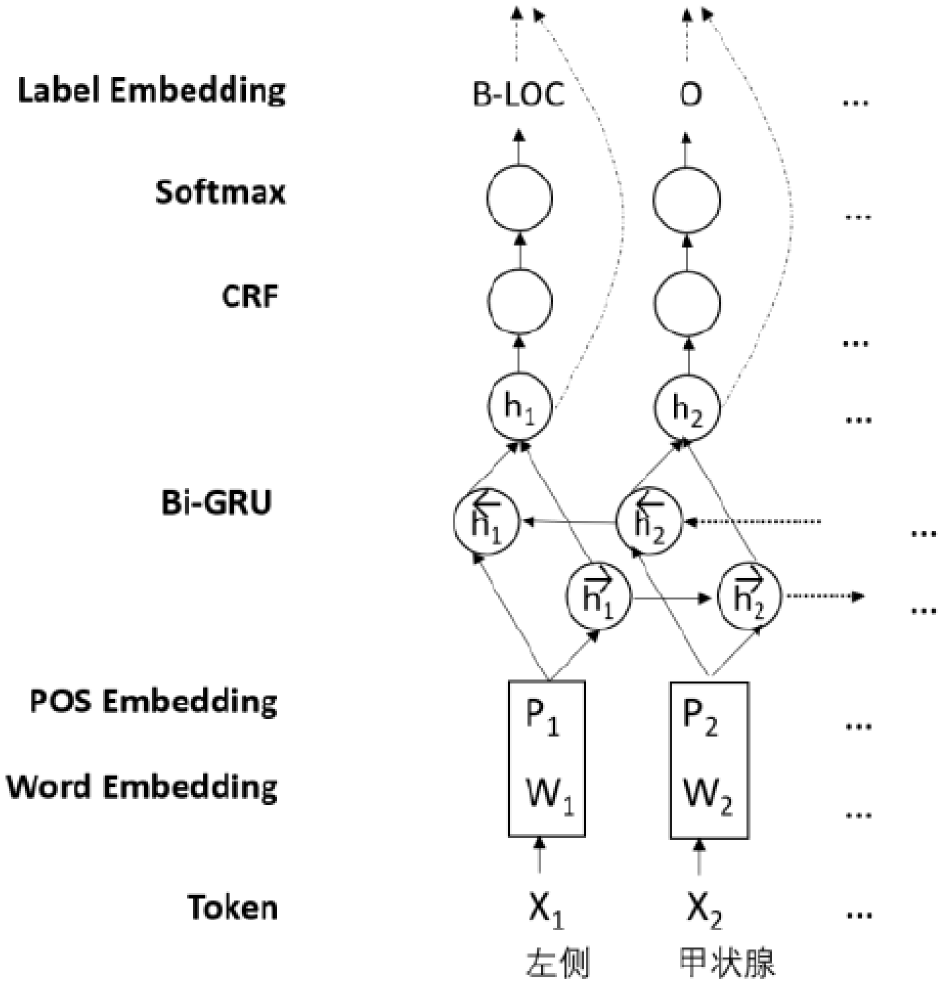


Fig. 3. The architecture of clinical entity recognition. The concatenations of words and POS tags are fed into Bi-GRU with CRF on top and softmax layer outputs the predictions of entity types.

Clinical Relation Extraction

Sub-Sentence Level Attention

Entity Level Attention

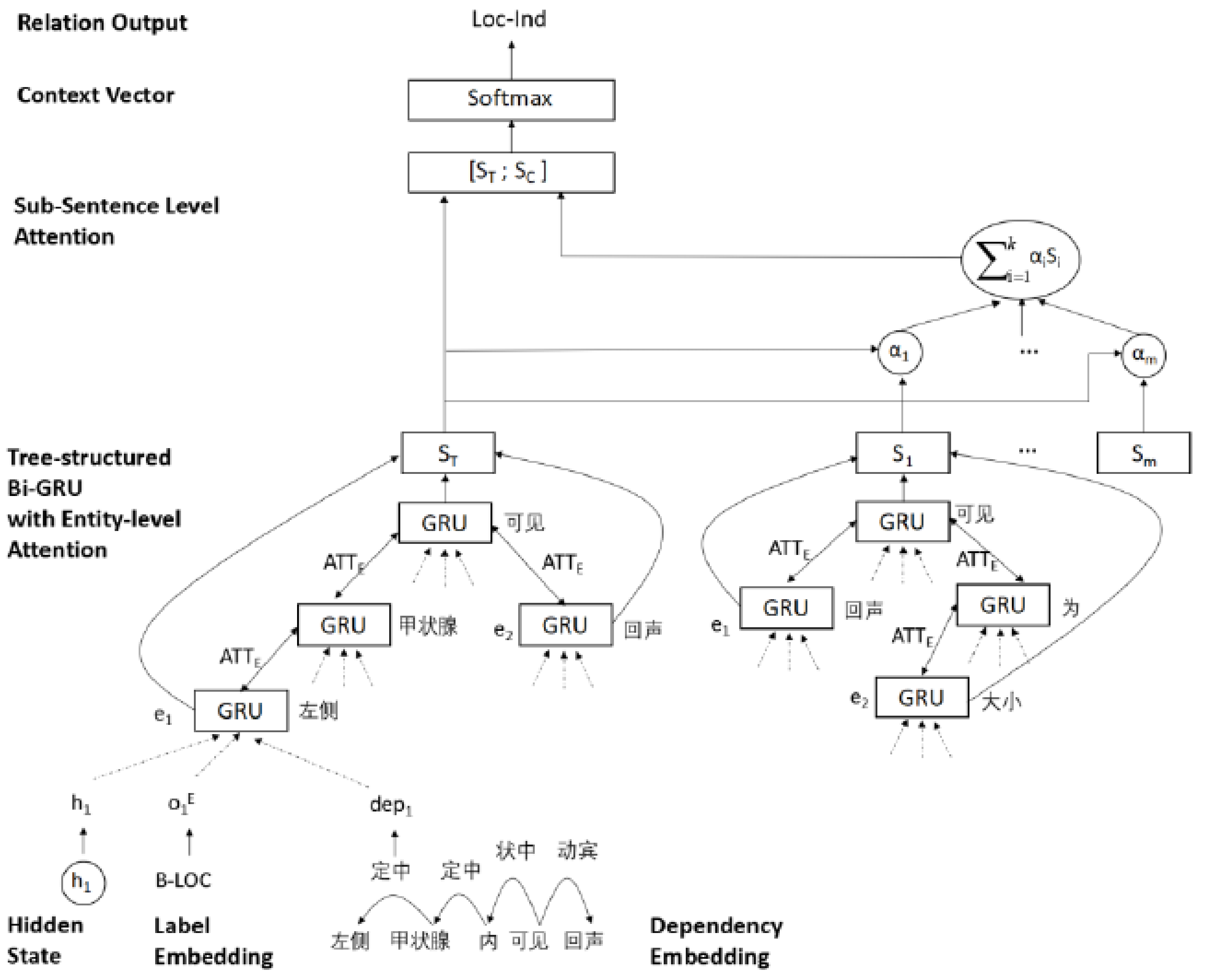


Fig. 4. The architecture of relation classification. The concatenations of hidden states from Bi-GRU, the label embeddings from clinical entity recognition and the syntactic dependency embeddings from dependency parsing tree are fed into Bi-Tree-GRU with entity-level and sub sentence-level attention and finally softmax predicts the relations between entity pairs.

Entity Level Attention

1.The representation vectors of the children nodes

$$R_t = [r_{t1}, r_{t2}, \dots r_{tN}]$$

2.A weighted sum of its children embeddings

$$M_t = \tanh(\mathrm{W_h} H_t)$$

Attention Layer $\longrightarrow$ $$\alpha = softmax(\mathrm{W}_\alpha^T M_t)$$

$$ch_t = H_t \alpha^T$$

3.GRU (Gated Recurrent Unit)

$$\mathrm{z}_t^{(R)} = \sigma(\mathrm{W}_z^{(R)} emb_t^R + \mathrm{U}_z^{(R)} ch_t + \mathrm{b}_z^{(R)})$$

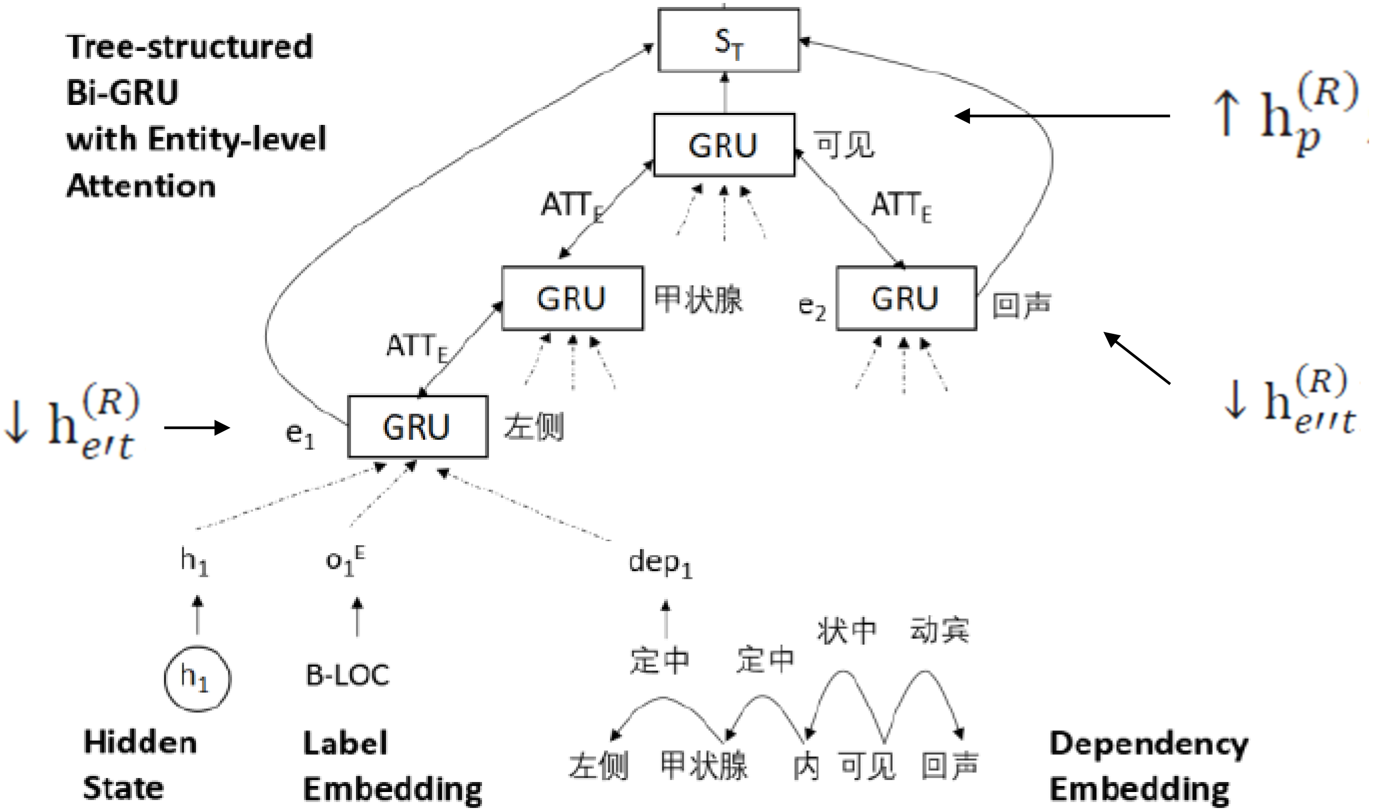
$$\mathrm{r}_t^{(R)} = \sigma(\mathrm{W}_r^{(R)} emb_t^R + \mathrm{U}_r^{(R)} ch_t + \mathrm{b}_r^{(R)})$$

$$\widetilde{\mathrm{h}_t^{(R)}} = \tanh(\mathrm{W}_h^{(R)} emb_t^R + \mathrm{U}_h^{(R)} (\mathrm{r}_t^{(R)} \odot \mathrm{h}_{t-1}^R) + \mathrm{b}_h^{(R)})$$

$$\mathrm{h}_t^{(R)} = (1 - \mathrm{z}_t^{(R)}) \odot \mathrm{h}_{t-1}^{(R)} + \mathrm{z}_t^{(R)} \odot \widetilde{\mathrm{h}_t^R}$$

4.bidirectional output vector

$$\mathrm{S}_p = [\uparrow \mathrm{h}_p^{(R)}; \downarrow \mathrm{h}_{e't}^{(R)}; \downarrow \mathrm{h}_{e''t}^{(R)}]$$



Sub-Sentence Level Attention

k context sub-sentences

$$S = \{S_1, S_2, \dots, S_k\}$$

To capture the relationship between the target sub-sentence and it context sub-sentences

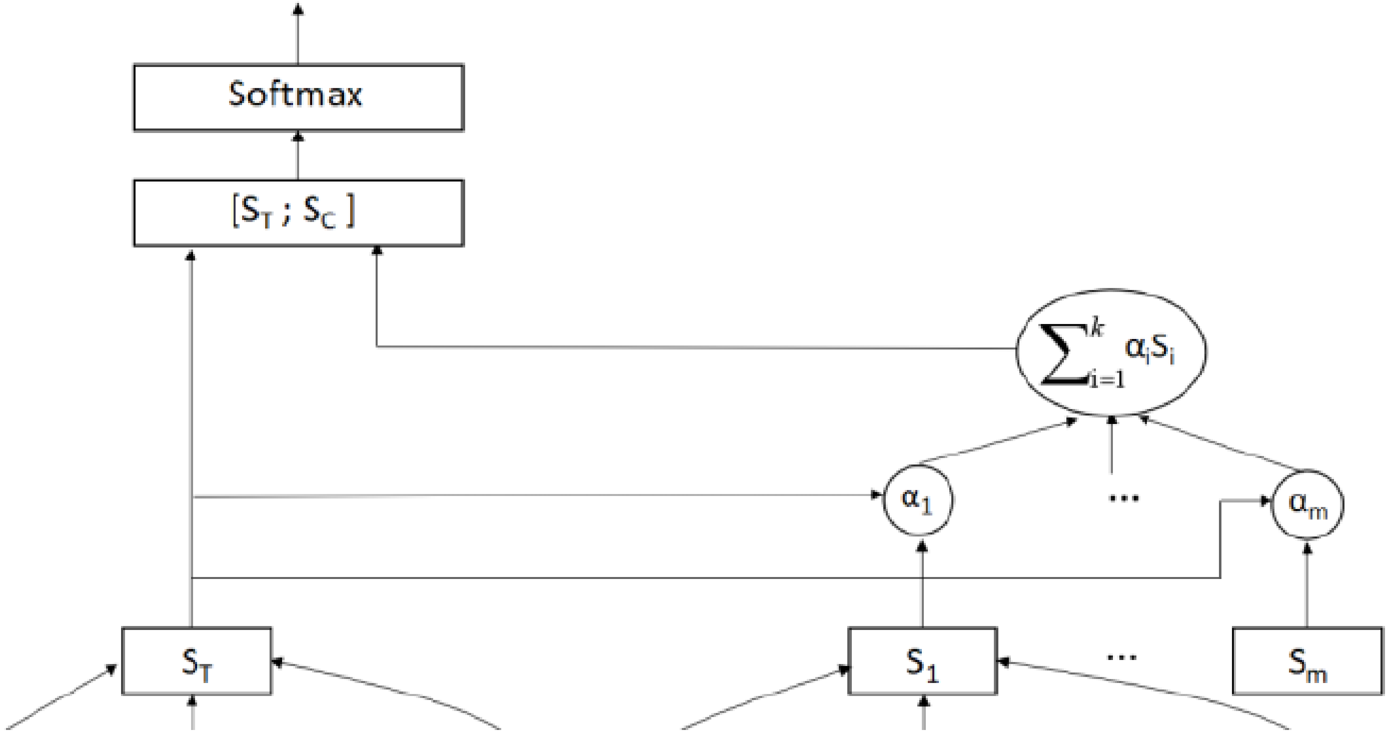
$$G(S_i, S_T) = S_i{}^T W_\alpha^{(S)} S_T$$

$$\alpha_i^{(S)} = softmax(G(S_i, S_T))$$

$$S_c = \sum_{i=1}^{k} \alpha_i^{(S)} S_i$$

$$S^* = [S_T; S_c]$$

Target Sentence Context Sentences



Dataset

ultrasonic reports, X-ray/CT reports, Puncture reports, pathology reports of thyroid and mammary gland fromRuijin Hospital.

True Positive (TP)

the number of entity types are identified as correct and boundaries are matched in NER or the numbers of correct relation types in RE.

False Positive (FP)

the number of incorrectly identified entities or relations that do not meet the above conditions.

False Negative (FN)

the number of unidentified entities or relations.

Comparison of different **entity-level attention** on the whole thyroid dataset

Type	Method	P	R	F1
No attention	SP-Tree	76.5	77.0	76.7
	SubTree	79.4	76.9	78.2
	FullTree	78.1	78.0	78.0
Attention	SP-Tree	82.2	84.2	83.2
	SubTree	82.1	85.6	83.8
	FullTree	82.5	84.0	83.3

Comparison of different **sub sentence-level attention** on the whole thyroid dataset

Method	P	R	F1
No Attention	76.4	76.9	76.7
Simple Attention [7]	79.5	82.1	80.8
Context Attention	82.5	84.0	83.3

Shortest Path Tree(SP-Tree)[1] only consists of the nodes on the shortest path in dependency parsing tree between the target entity pairs
SubTree selects the nodes in the subtree under the lowest common ancestor of the target entity pair
FullTree take all the nodes into the entity-level attention.

"Simple Attention"[2] simply uses a weighted sum of all the sub-sentences including a waited classification of relation in one sentences and not distinguish the target pairs from other context sub-sentences

1.Miwa M., Bansal M.: End-to-End Relation Extraction using LSTMs on Sequences and Tree Structures[C]. Meeting of the Association for Computational Linguistics. 1105-1116 (2016).
2.Zhou P., Shi W., Tian J., Qi Z., Li B., Hao H., et al: Attention-Based Bidirectional Long Short-Term Memory Networks for Relation Classification[C]. Meeting of the Association for Computational Linguistics. 207-212 (2016).