Deep Inside-outside Recursive Autoencoder with All-span Objective

Ruyue Hong Jiong Cai Kewei Tu

School of Information Science and Technology, ShanghaiTech University

Motivation

The original objective function of DIORA is constrained to utilize leaf-level information only. However, higher-level spans also embody meaningful information which could be utilized.

Proposed Solution

We proposed a new objective function that utilizes information among all-level spans and assign weights of spans based on scores.

Empirical Results

Experiment results show that our new training objective performs well on datasets of two languages: English and Japanese. and empirically show that our method achieves improvement in parsing accuracy over the original DIORA.

DIORA

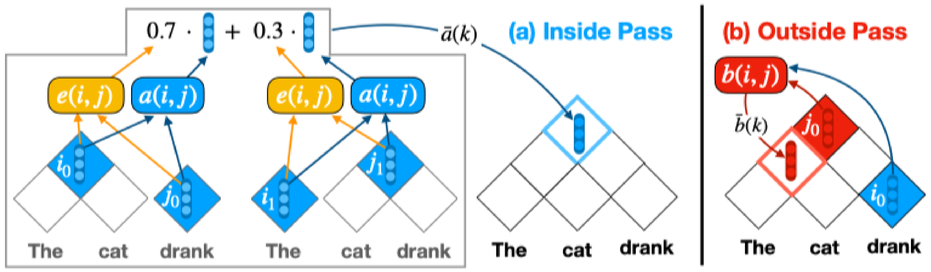

Figure 1. For a span i, denote the inside score by $\bar{e}(i)$, inside vector by $\bar{a}(i)$, outside score by $\bar{f}(i)$ and outside vector by $\bar{b}(i)$.

DIORA(Drozdov et al., 2019) incorporates the inside-outside algorithm into a latent tree chart parser.

- bottom-up inside pass: it computes the inside vector of each span in the input sentence, which captures the phrase information of the inner content in the span.
- top-down outside pass: it computes the outside vector of each span in the input sentence, which models the contextual information of the span.

All-span Objective Function

$$L_x = n \cdot \sum_{i \in \Pi(x)} \frac{w_i}{\sum_{j \in \Pi(x)} w_j} \cdot L_i$$

$$w_i^{\alpha} = \exp(\frac{\bar{e}(i)}{m_i})$$

$$w_i^{\beta} = \exp(\frac{\bar{f}(i)}{(n - m_i + 1) \cdot n})$$

- $\Pi(x)$: the set of all spans in the sentence x.
- L_i: the loss of span i.
- w_i^{α}: weight based on inside scores.
- w_i^{β}: weight based on outside scores.
- m_i: the length of span i.
- n: the length of the sentence.

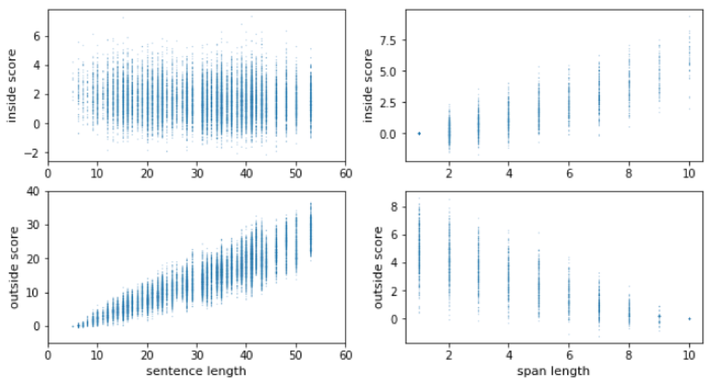

Figure 2. Left: Score distribution of spans of length 5 in sentences of different lengths. Right: Score distribution of spans of different lengths in sentences of length 10.

Inside scores are positively correlated with span length.
Outside scores are negatively correlated with span length.
Outside scores are positively correlated with sentence length.
The influence of a sentence is proportional to its length.

Original Objective Function

max-margin loss:

$$L_x = \sum_{i \in \Phi(x)} \sum_{i^* \in \mathcal{N}(i)} \max(0, 1 - \bar{b}(i) \cdot \bar{a}(i) + \bar{b}(i) \cdot \bar{a}(i^*))$$

softmax loss:

$$L_x = -\sum_{i \in \Phi(x)}^{I} \log(\frac{\exp(\bar{b}(i) \cdot \bar{a}(i))}{Z^*(i) + \exp(\bar{b}(i) \cdot \bar{a}(i))})$$

$$Z^*(i) = \sum_{i^* \in \mathcal{N}(i)} \exp(\bar{b}(i) \cdot \bar{a}(i^*))$$

where $\Phi(x)$ denotes the set of leaf-spans in the sentence x, $\mathcal{N}(i)$ denotes the set of negative samples for span i.

Experiments and Analysis

Dataset: PTB corpus (Marcus et al., 1993) for English.
use section 0-21 for training, 22 for validation, and 23 for testing.

	With Punctuation				No Punctuation			
	F1-10$_{\mu}$	F1-10$_{\max}$	F1-all$_{\mu}$	F1-all$_{\max}$	F1-10$_{\mu}$	F1-10$_{\max}$	F1-all$_{\mu}$	F1-all$_{\max}$
DIORA	55.15 ± 0.86	56.61	44.63 ± 0.46	45.02	60.01 ± 0.40	60.64	**49.31** ± 0.45	**49.73**
DIROA-all	**59.18** ± 4.17	**63.64**	**47.03** ± 4.20	**50.63**	**61.23** ± 2.31	**63.33**	47.95 ± 1.25	49.47
Upper Bound	75.15	-	78.96	-	86.64	-	85.34	-

Table 1. Experimental results on English PTB.

Dataset: KTB corpus (Butler et al., 2012) for Japanese.
shuffle the corpus and take 80% for training, 10% for validation, and 10% for testing.

	With Punctuation				No Punctuation			
	F1-10$_{\mu}$	F1-10$_{\max}$	F1-all$_{\mu}$	F1-all$_{\max}$	F1-10$_{\mu}$	F1-10$_{\max}$	F1-all$_{\mu}$	F1-all$_{\max}$
DIORA	39.33 ± 2.92	42.83	28.93 ± 4.29	32.33	44.02 ± 5.02	**49.18**	35.26 ± 3.10	38.02
DIROA-all	**43.30** ± 5.18	**47.73**	**33.00** ± 3.71	**36.93**	**47.09** ± 1.79	49.17	**36.37** ± 3.58	**41.56**
Upper Bound	61.41	-	62.53	-	67.25	-	67.32	-

Table 2. Experimental results on Japanese KTB.

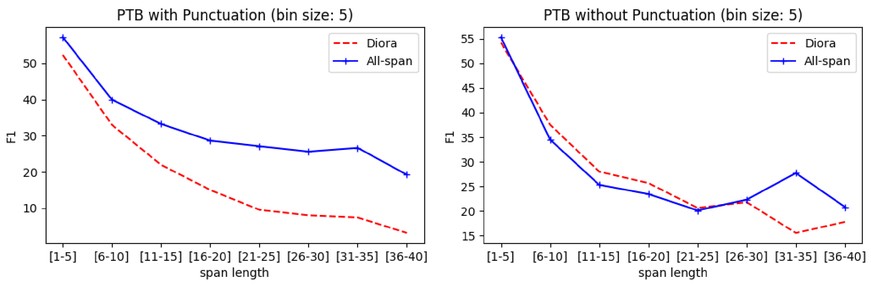

Figure 3. F1 scores on spans of different lengths.

- With punctuation: DIORA-all consistently outperforms DIORA especially on long spans.
- Without punctuation: the two methods have similar accuracy except for an outlier at length range [31-35].