Obligation and Prohibition Extraction Using Hierarchical RNNs

ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS

Ilias Chalkidis, Ion Androutsopoulos, and Achilleas Michos
Department of Informatics, AUEB, Greece (http://nlp.cs.aueb.gr)
Cognitiv+ Ltd., London, UK (http://www.cognitivplus.com)

cognitiv+

The Task - Examples

No.	Gold Class	Sentences/Clauses
1	Obligation None	The Supplier *is obliged to* meet and comply with the Approved Requirements. Details shall be determined in the individual contracts.
2	Prohibition Obligation	*No* Provider staff *will* provide services to any Customer Competitor. Provider *will* take such measures to prevent these actions.
3	Prohibition	Provider *is not entitled* to suspend this Agreement prior to the lapse of the fifth year.
4	Oblig./Prohib. List Intro Obligation List Item Prohibition List Item	The Supplier *shall*: (*a*) only process the Personal Data in accordance with Client's written instructions; (*b*) *not* transfer any Personal Data to any other third parties;
5	Oblig./Prohib. List Intro Obligation List Item Prohibition List Item Prohibition List Item	The Receiving Party *will*: (*i*) keep the Confidential Information secret and confidential; (*ii*) *not* disclose the Confidential Information to any person other than in accordance with Clauses 13.3; and (*iii*) *not* use the Confidential Information other than for the purposes of this Agreement.
6	Oblig./Prohib. List Intro Prohibition List Item Prohibition List Item None	A Party *shall not* directly solicit the employment of: (*i*) in the case of Client, Supplier's employees engaged in the provision of the Services, (*ii*) in the case of Supplier, Client's employees engaged. Nothing in this section will restrict either Party's right to recruit.

Datasets

Sections selected from 100 random English service agreements

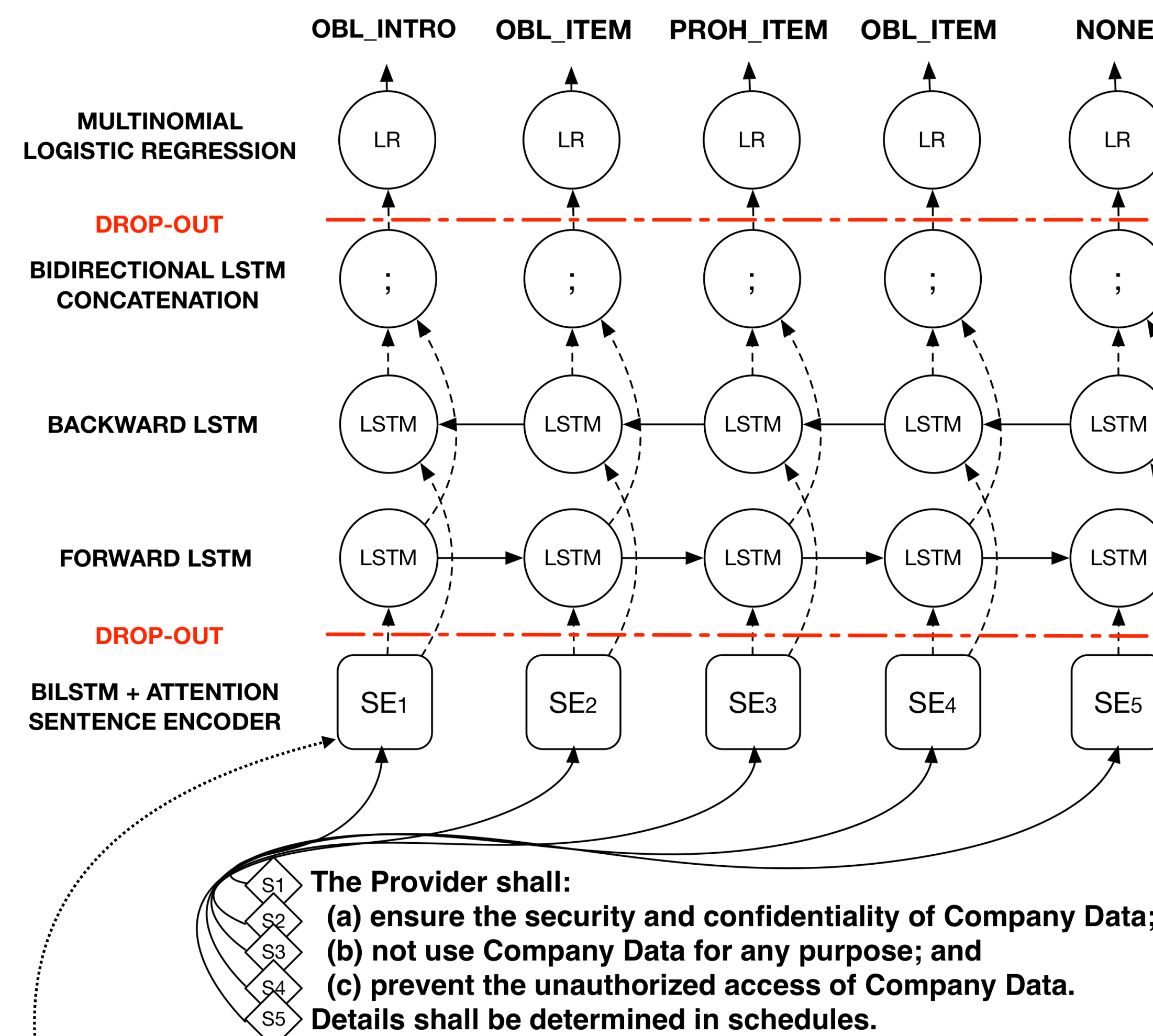|	Train Sections	Development Sections	Test Sections
Sections	6,385	1,595	1,420
Gold Class	**Train Sentences**	**Development Sentences**	**Test Sentences**
None	15,401	3,905	4,141
Obligation	11,005	2,860	970
Prohibition	1,172	304	108
Obl. List Begin	828	203	70
Obl. List Item	2888	726	255
Proh. List Item	251	28	19
Total	**31,545**	**8,036**	**5,563**

Token Representation

- 200-dimensional word embeddings obtained by applying word2vec to approx. 750K English contracts.
- 25-dimensional POS tag embeddings obtained by applying word2vec to 50K English contracts.
- 5-dimensional token shape embeddings (e.g., all capitals, first letter capital, all digits) obtained by applying word2vec to 50K English contracts.
- Each token is represented by the concatenation of its word, POS and shape embeddings. Unknown tokens are mapped to pre-trained POS-specific 'unk' embeddings (e.g., 'unk-n', 'unk-vb').
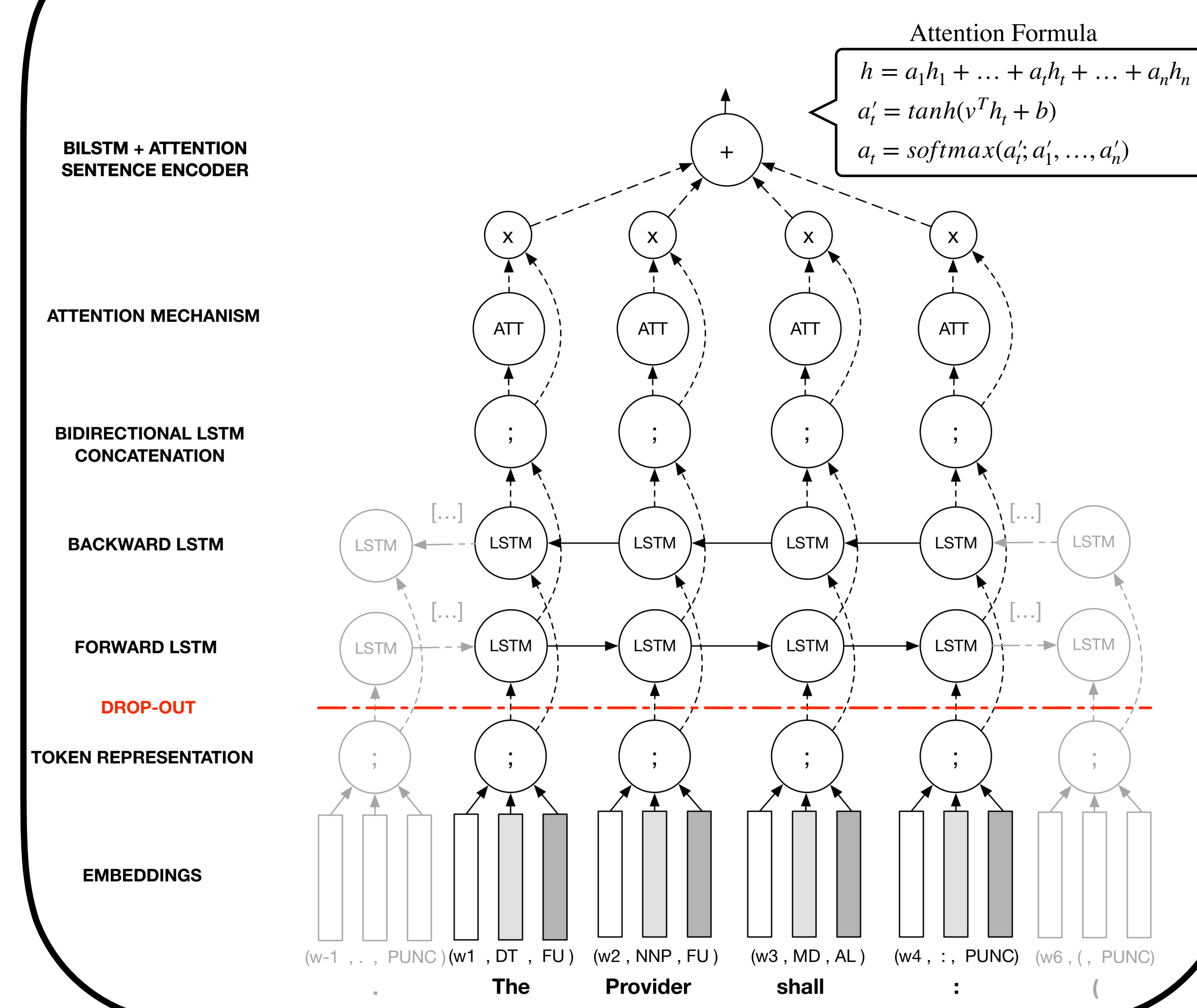
Methods

- **BILSTM:** The BILSTM chain is fed with the token embeddings of a single sentence (or clause) at a time. The last hidden states of the forward and backward LSTM are concatenated and fed to a multinomial LR layer.
- **BILSTM-ATT:** The sentence (or clause) is represented by the weighted sum of the (concatenated forward/backward) states of the BILSTM chain. The weights are produced by a single-layer self-attention mechanism (Attention Formula, "Sentence Encoder" diagram). The weighted sum is fed to a multinomial LR layer.
- **X-BILSTM-ATT:** Same as BILSTM-ATT, but the BILSTM chain is fed with the token embeddings not only of the sentence being classified, but also of the previous and following tokens (faded parts of "Sentence Encoder" diagram). The self-attention mechanism still considers the BILSTM states of the current sentence only.
- **H-BILSTM-ATT:** The hierarchical BILSTM classifier considers all the sentences (or clauses) of an entire section. Each sentence is converted to a sentence embedding using BILSTM-ATT as a sentence encoder (without its LR layer). A second BILSTM reads the sentence embeddings and produces context-aware sentence embeddings, which are fed to a multinomial LR layer.

Hierarchical Sentence Classifier

Sentence Encoder

Attention Formula

$$h = a_1 h_1 + \ldots + a_t h_t + \ldots + a_n h_n$$
$$a'_t = tanh(v^T h_t + b)$$
$$a_t = softmax(a'_t; a'_1, \ldots, a'_n)$$

Attention Scores (heatmap) of BILSTM-ATT

The Supplier is obliged to meet and comply with the Approved Requirements.
Details shall be determined in the individual contracts.

No Provider staff will provide services to any Customer Competitor.
Provider will take such measures to prevent these actions.

The Supplier shall:
(a) only process the Personal Data in accordance with Client's written instructions;
(b) not transfer any Personal Data to any other third parties;

The Receiving Party will:
(i) keep the Confidential Information secret and confidential;
(ii) not disclose the Confidential Information to any person other than in accordance with Clauses 13.3; and
(iii) not use the Confidential Information other than for the purposes of this Agreement.

A Party shall not directly solicit the employment of:
(i) in the case of Client, Supplier's employees engaged in the provision of the Services, and
(ii) in the case of Supplier, Client's employees engaged.
Nothing in this section will restrict either Party's right to recruit.

Evaluation - Training Statistics

	BILSTM				BILSTM-ATT			
Gold Class	P	R	F1	AUC	P	R	F1	AUC
None	0.95	0.91	0.93	0.98	0.97	0.90	0.93	**0.99**
Obligation	0.75	0.85	0.79	0.86	0.75	0.88	0.81	0.86
Prohibition	0.67	0.62	0.64	0.75	0.74	0.75	0.74	0.80
Obl. List Begin	0.70	0.86	0.77	0.81	0.71	0.85	0.77	0.83
Obl. List Item	0.53	0.66	0.59	0.64	0.48	0.70	0.57	0.60
Proh. List Item	0.59	0.35	0.43	0.50	0.61	0.55	0.59	0.62
Macro-average	0.70	0.70	0.70	0.74	0.73	0.78	0.74	0.78
Micro-average	0.90	0.88	0.88	0.94	0.90	0.88	0.89	0.96

	X-BILSTM-ATT				H-BILSTM-ATT			
Gold Class	P	R	F1	AUC	P	R	F1	AUC
None	0.96	0.90	0.93	0.98	**0.97**	**0.96**	**0.97**	**0.99**
Obligation	0.75	0.87	0.81	0.88	**0.87**	**0.92**	**0.90**	**0.96**
Prohibition	0.65	0.75	0.70	0.74	**0.84**	**0.83**	**0.84**	**0.90**
Obl. List Begin	0.72	0.75	0.74	0.80	**0.90**	**0.89**	**0.89**	**0.93**
Obl. List Item	0.49	0.78	0.60	0.66	**0.85**	**0.94**	**0.89**	**0.94**
Proh. List Item	**0.83**	0.50	0.62	0.67	0.80	**0.84**	**0.82**	**0.92**
Macro-average	0.73	0.76	0.73	0.79	**0.87**	**0.90**	**0.89**	**0.94**
Micro-average	0.90	0.88	0.89	0.94	**0.95**	**0.95**	**0.95**	**0.98**

Network	Training Time	Parameters
BILSTM	5h 30m	1,278M
BILSTM-ATT	8h 30m	1,279M
X-BILSTM-ATT	25h 40m	1,279M
H-BILSTM-ATT	2h 30m	1,837M

Other Recent Publications

(nlp.cs.aueb.gr/publications.html)

- I. Chalkidis and I. Androutsopoulos, "A Deep Learning Approach to Contract Element Extraction". Proceedings of the *30th International Conference on Legal Knowledge and Information Systems (JURIX 2017)*, Luxembourg, pp. 155-164, 2017.
- I. Chalkidis, I. Androutsopoulos and A. Michos, "Extracting Contract Elements". Proceedings of the *16th International Conference on Artificial Intelligence and Law (ICAIL 2017)*, London, UK, pp. 19-28, 2017.