

TOI-CNN: A Solution of Information Extraction on Chinese Insurance Policy

Lin Sun

Zhejiang University City College
50 Huzhou Street, Hangzhou, China
sunl@zucc.edu.cn

Kai Zhang *

Zhejiang University City College
50 Huzhou Street, Hangzhou, China
31601102@stu.zucc.edu.cn

Fule Ji

Zhejiang University City College
50 Huzhou Street, Hangzhou, China
31601087@stu.zucc.edu.cn

Zhenhua Yang

Shanghai Fortune FIS Co., Ltd.
1687 Changyang Road, Shanghai, China
yangzhenhua@fc18.com.cn

Abstract

Contract analysis can significantly ease the work for humans using AI techniques. This paper shows a problem of Element Tagging on Insurance Policy (ETIP). A novel Text-Of-Interest Convolutional Neural Network (TOI-CNN) is proposed for the ETIP solution. We introduce a TOI pooling layer to replace traditional pooling layer for processing the nested phrasal or clausal elements in insurance policies. The advantage of TOI pooling layer is that the nested elements from one sentence could share computation and context in the forward and backward passes. The computation of backpropagation through TOI pooling is also demonstrated in the paper. We have collected a large Chinese insurance contract dataset and labeled the critical elements of seven categories to test the performance of the proposed method. The results show the promising performance of our method in the ETIP problem.

1 Introduction

Automatic contract analysis can gain immediate insight into the content of specific contractual documents in legal or financial areas (Moens et al., 2000). Compared to the traditional method of manually reviewing hundreds of contracts, it is helpful not only manage and access contracts but also significantly free knowledge workers from menial, laborious and often error-prone tasks. The insurance policy is a legal contract that outlines the rights and obligations of the insured and insurer. It consists of a wide variety of different types of insurance coverages to meet specific needs, although most insurance policies are somewhat standardized. Understanding the various types of insurance coverage is time-consuming and error-prone. This paper shows a problem of Element

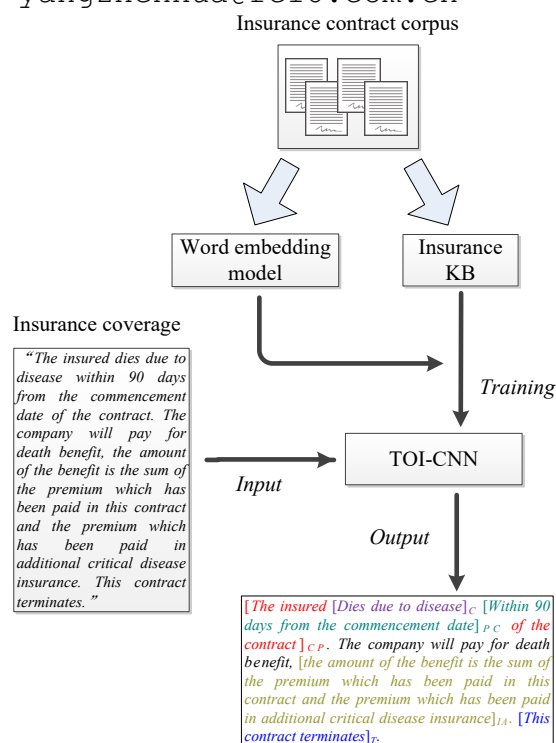


Figure 1: The processing architecture for ETIP.

Tagging on Insurance Policy (ETIP). It can automatically convert a massive amount of insurance policies into structural archives for management and comparison. Due to the vital information highlighted by ETIP, it can also timely provide insurance staff valuable insight into policies, quickly locate requested information and speed up claim processing.

The processing architecture for ETIP is shown in Fig. 1. We have built a large Chinese insurance contract corpus. There are two usages of the corpus. One is for learning word embeddings. In Sec. 5.3, we show the advantage of the insurance-specific corpus over other general language corpora for the training of word embeddings. Another usage of the corpus is to create insurance knowl-

* Kai Zhang is a co-first author.

edge base (KB). Insurance KB, which consists of seven categories of the elements manually labeled by the insurance employees, provides the training data for TOI-CNN model. Specifically, the contributions of this paper can be summarized as follows:

- To our best knowledge, this is the first work on semantic-specific tagging on insurance contracts. Compared to nested NER, not only the type of the elements varies from a short phrase to a long sentence, but also a phrase or clause element could be embedded in other elements.
- We propose a novel TOI-CNN model for the ETIP solution. The advantage of TOI pooling layer is that the elements from the same sentence could share computation and context in the forward and backward passes.
- We have collected 500 Chinese insurance contracts of 46 insurance companies and published the dataset. The experimental results show that the overall performance of TOI-CNN is promising for practical application.

2 Related Work

The work of contract analysis is typically divided into two categories, segmentation and information extraction (IE). Segmentation (Hasan et al., 2008; Loza Mencía, 2009) aims to outline the structure of a conventional text format by annotating title, section, subsection, and so on. Information extraction (Cohen and McCallum, 2004; Piskorski and Yangarber, 2013) focuses on the classification of words, phrases or sentences. Recent works of contract information extraction have addressed recognition of some essential elements in legal documents (Curtotti and McCreath, 2010; Indukuri and Krishna, 2010). Chalkidis et al. (2017) extracted the contract element, types of which are contract title, contracting parties, date, contract period, legislation refs and so on. The extraction method was based on Logistic Regression, SVM (Chalkidis et al., 2017) and BILSTM (Chalkidis and Androutopoulos, 2017) with POS tag embeddings and hand-crafted features. García-Constantino et al. (2017) presented the system called CLIEL for extracting information from commercial law documents. CLIEL identified five element categories similar to the literature mentioned in (Chalkidis

et al., 2017) by rule-based layout detection. Azzopardi et al. (2016) developed a mixture extraction method of regular expressions and named entity to extract information from contract clauses, and provided an intelligent contract editing tool to lawyers. Previous works of contract information extraction always focused on title, date, layout, contracting party, etc.. They are not directly related to the semantics of contracts, and could not provide deep insight into contract understanding. The insurance policies are formal legal documents and usually have general elemental compositions, e.g., coverage, payment, and period. In this paper, we investigate how to interpret insurance clauses. Some examples of ETIP are shown in Sec. 3.

The tasks of information extraction could be Named Entity Recognition (NER) (Nadeau and Sekine, 2007; Ritter et al., 2011), Information Extraction by Text Segmentation (IETS) (Cortez and Da Silva, 2013; Hu et al., 2017), etc.. NER typically recognizes persons, organizations, locations, dates, amounts, etc.. IETS identifies attributes from semi-structured records in the form of continuous text, e.g., product description and ads. The previous IE works on contracts (Azzopardi et al., 2016; Chalkidis et al., 2017) are similar to NER.

Recently researchers pushed the field of NER towards nested representations of named entities. Muis and Lu (2017) incorporated mention separators to capture how mentions overlap with one another. Both of two works relied on hand-crafted features. Ju et al. (2018) designed a sequential stack of flat NER layers that detects nested entities. One bidirectional LSTM layer represented word sequences and CRF layer on top of the LSTM layer decoded label sequences globally. Katiyar and Cardie (2018) presented a standard LSTM-based sequence labeling model to learn the nested entity hypergraph structure for an input sentence. Our ETIP problem is a variant of nested NER, called lengthly nested NER. The type of nested entities varies from phrase to clause. However, in the previous nested NER datasets (Kim et al., 2003; Doddington et al., 2004), the type of nested entities only contains short phrase and the average length is approximately three words.

3 ETIP Problem Statement

In this section, we first give the definition of elements tagging on insurance policy (ETIP) problem. Given an insurance coverage $C =$

(s_1, s_2, \dots, s_n) , where s_i is the i th sentence in C . $s_i = (w_{i,1}, w_{i,2}, \dots, w_{i,m})$, where $w_{i,j}$ is the j th word in sentence s_i . An element e in the coverage C is continuous words in one sentence, denoted as $e = \{(w_{i,s}, w_{i,s+1}, \dots, w_{i,t}), l\}$, where l is the category label of the element e . The goal of ETIP is to find the element e of category l in the coverage C . We define seven categories of insurance clauses listed as follows,

- Cover (C)
- Period of Coverage (PC)
- Condition for Payment (CP)
- Waiting Period (WP)
- Insured Amount (IA)
- Exclusion (E)
- Termination (T).

Here we give a coverage example in ETIP and translate it for English reading paper. One category is represented by one kind of font color.

“被保险人于本合同生效之日90天内因疾病身故，本公司给付身故保险金，其金额为本保险实际缴纳的保险费与本合同所附的重大疾病保险实际缴纳的保险费二者之和，本合同终止。”

The insured dies due to disease within 90 days from the commencement date of the contract. The company will pay for death benefit, the amount of the benefit is the sum of the premium which has been paid in this contract and the premium which has been paid in additional critical disease insurance. This contract terminates.

The extractable elements in the example are listed as follows,

- **C:** *dies due to disease*
- **PC:** *within 90 days from the commencement date*
- **CP:** *The insured dies due to disease within 90 days from the commencement date of the contract*
- **IA:** *the amount of the benefit is the sum of the premium which has been paid in this contract and the premium which has been paid in additional critical disease insurance*
- **T:** *This contract terminates.*

In this example, the sentence of CP in red contains the other two elements which are C in purple and PC in green respectively. It is the challenge of ETIP, a general element tagging problem, which allows that the elements of various length could be overlapped. We demonstrate other examples in ETIP along with English translation as follows, where $[\]_{tag}$ is a category tag labeling the range of an element.

1. [我们向您退还 [本合同终止时] $_T$ 的现金价值] $_{IA}$
[We refund you the cash value when [this contract terminates] $_T$] $_{IA}$
2. [等待期是指本合同生效后 [平安人寿不承担保险责任] $_E$ 的一段时间] $_{WP}$
[The waiting period refers to the period of [no obligation for insurance benefits from Ping An life insurance] $_E$ after the contract takes effect] $_{WP}$
3. [若被保险人在本合同生效之日起180日 ([这180日的时间段称为“等待期”] $_{WP}$) 内 [身故] $_C$] $_{CP}$
[If [the insured died] $_C$ within 180 days ([this 180-day period is called "waiting period"] $_{WP}$) from the commencement date of the contract] $_{CP}$
4. [主合同的保险费 [自给付保险金后的首个保险费约定支付日起] $_{CP}$ 将按被保险人投保年龄的费率及基本保险金额支付] $_{IA}$
[The insurance benefits of the main contract will be paid [from the date of the first premiums paid after the payment of the insurance benefits] $_{CP}$ according to the premiums rate of the insured's age and the basic insurance amount] $_{IA}$

To illustrate phrasal level of an element, we define a metric, called Element Length Ratio (ELR),

$$ELR = \frac{element_length}{sentence_length}. \quad (1)$$

For example, $ELR(C) = 4/16 = 0.25$, $ELR(PC) = 7/16 = 0.44$, $ELR(CP) = 1$ in the previous example. Tab. 1 in experiment section will list the statistics of ELR.

4 TOI-CNN Architecture

Fig. 2 illustrates the TOI-CNN architecture. TOI-CNN takes as input an entire sentence and a set of elements. The network first processes the whole sentence with one convolutional layer (Conv+Relu in Fig. 2) to yield 36 feature maps. Then for each element, the TOI pooling layer extracts a fixed-length feature vector from the feature map. Each feature vector is fed into a sequence of fully connected (fc1) layer that finally connects the output layer, which produces softmax probability estimates over K element classes plus a non-element class.

4.1 The Convolutional Layer

We use the CNN model (Kim, 2014) with pre-trained word embedding (Mikolov et al., 2013) for the convolutional layer. w_i is i -th word in the sentence and is represented as the k -dimensional word embedding vector. The dimension of the input layer is $n \times k$ (padding zeros when the length of the sentence is less than n). Our neural network

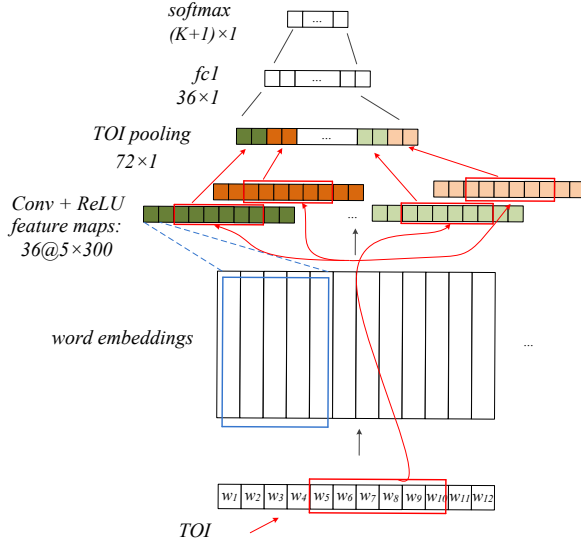


Figure 2: TOI-CNN architecture.

consists of one convolution layer with ReLU activation and one TOI pooling layer, which replaces the max pooling layer of CNN in general. The convolution layer has a set of filters of size $h \times k$ and produces p feature maps of size $(n-h+1) \times 1$.

4.2 The TOI Pooling Layer

The TOI pooling layer uses max pooling to convert the features inside any valid region of an interesting window into a small feature map with a fixed length of L (e.g., $L = 2$ in Fig. 2). Fig. 2 describes the TOI pooling in detail using red lines and rectangles. The TOI pooling layer extracts text region of the elements from the feature maps of the convolutional layer. The TOI region is shown as a red rectangle in the input sentence, shown in Fig. 2. The corresponding TOI window in the feature maps is connected by red curved lines. The length of the TOI window becomes shorter because of the narrow convolution.

TOI max pooling works by dividing the TOI window of length rl into L sub-windows of size $\lfloor rl/L \rfloor$ and then max-pooling the values in each sub-window into the corresponding cell of TOI pooling layer. For example, $rl = 6$ in Fig. 2 and hence the sub-window of size $6/2 = 3$ produce a cell of the pooling layer. Pooling operator is applied independently in each feature map channel. The pooling results of all feature maps are sequentially arranged into a vector, which is followed by the fully connected layer ($fc1$).

4.3 Training Samples of TOI-CNN

In TOI-CNN, training samples include two categories: 1) TOI ground truths, 2) negative sliding windows. A sliding window is defined as negatives, or called non-element class, if IoU (Intersection over Union) with all ground truths of a sentence is less than a threshold th_s . Function $IoU(a, b)$, measuring how much overlap occurs between two text strings a and b , is defined as,

$$IoU(a, b) = \frac{length(a \cap b)}{length(a \cup b)}. \quad (2)$$

4.4 Backpropagation through TOI Pooling Layer

The network is modified to take two data inputs: a set of sentences and a list of training samples in those sentences. Each training sample is given as a one-hot encoding label $p = (0, \dots, p_j = 1, \dots, 0)$ with a class j . The cross entropy loss L is,

$$L = -\log p_j^*, \quad (3)$$

where p^* is the output of the softmax layer. Then, we present the derivative rules in backpropagation through the TOI pooling layer.

Let x_i be the i -th activation input into the TOI pooling layer and let $y_{s,j}$ be the layer's j -th output from the s -th training sample. The TOI pooling layer computes $y_{s,j} = \max(x_i), x_i \in \mathcal{W}_{s,j}$, where $\mathcal{W}_{s,j}$ is the j -th input sub-window over which max-pooling outputs $y_{s,j}$. Due to overlaps between training samples, a single x_i may be assigned to several different outputs $y_{s,j}$. Let $\mathcal{M}(x_i)$ be the set of $y_{s,j}$ that x_i activates in the TOI pooling layer.

Finally, the TOI pooling layer's backwards function computes partial derivative of the loss function with respect to input variable x_i as follows,

$$\frac{\partial L}{\partial x_i} = \sum_s \sum_j \frac{\partial L}{\partial y_{s,j}}, \quad y_{s,j} \in \mathcal{M}(x_i). \quad (4)$$

The partial derivative $\partial L / \partial y_{s,j}$ is accumulated if $y_{s,j}$ is activated by x_i in TOI max-pooling. In backpropagation, the partial derivatives $\partial L / \partial y_{s,j}$ are already computed by the backwards function of the layer on top of the TOI pooling layer.

5 Experiments

5.1 ETIP Dataset

We collected 500 Chinese insurance contracts, which include life, disability, health, property,

home, and auto insurance, where 350 contracts are regarded as the corpus for training word embeddings (Mikolov et al., 2013) and the other 150 contracts are manually labeled for element tagging testing. The maximum nested level is three in ETIP. The dataset is available online (<https://github.com/ETIP-team/ETIP-Project/>) without author information. This project cooperated with an information solution provider of China Pacific Insurance Co., Ltd. (CPIC). Tab. 1 shows the number (N), average length (\bar{L}) and average element length ratio (\overline{ELR}) of seven categories in ETIP dataset. CP and IA are the two largest categories in the dataset. \overline{ELR} of C, PC and E are 0.12, 0.63 and 0.76 respectively, which means that they are usually a phrase or clause embedded in a sentence and C is a 2-3 word phrase. \overline{ELR} of CP, IA, and T are nearly 1.0, which denotes that they are always sentences.

Category ID	N	\bar{L}	\overline{ELR}
Cover (C)	618	2.6	0.12
Waiting Period (WP)	21	17.6	0.91
Period of Coverage (PC)	186	20.0	0.63
Condition for Payment (CP)	1295	25.5	0.98
Insured Amount (IA)	1068	27.3	0.99
Exclusion (E)	25	12.9	0.76
Termination (T)	398	9.2	0.97

Table 1: Statistics of seven categories in ETIP.

5.2 Experimental Settings

Chinese texts are tokenized with Jieba (Jieba, 2017) or NLPIR (NLPIR, 2018). 300-dimensional word vectors are trained on our insurance corpus. The size of the input layer in the CNN model is 60×300 , and zeros are padded if the length of the training sample is less than 60. The kernel size of the convolution layer is 5×300 , and the size of the feature maps is 36. the fixed length of TOI pooling layer output is $72 = 2 \times 36$.

The 150 labeled contracts are split into five equal folds, and we use the evaluation procedure in 5-fold cross-validation. Dealing with imbalanced data, the small categories, e.g., WP, E, and PC, are oversampled. The size of mini-batches is 4 sentences, randomly sampling up to 48 negative sliding windows from each sentence. We implement TOI-CNN using PyTorch and run Adam (Kingma and Ba, 2014) optimizer for 50 training epochs with the learning rate of 0.0001.

In the detection, we apply a greedy non-maximum suppression within and between classes

simultaneously if two sliding windows positionally intersect but they have no inclusion relation. In within-class suppression, a sliding window is rejected if its length is shorter than the other one. In between-class suppression, a sliding window is rejected if its softmax score is lower than the other one. In performance evaluation, a sliding window is recognized as true positive if IoU over a ground truth is larger than th_p and the predicted label is the same as the ground truth.

5.3 Word Embedding Comparison

350 contracts in ETIP Dataset are regarded as the corpus for training word embeddings (Mikolov et al., 2013). The augmented word2vec model trained by our insurance contract corpus can improve the similarities of the insurance synonyms compared to the models trained by other corpora, e.g., Baidu Encyclopedia (Baidu, 2018), Wikipedia_zh (Wikipedia, 2018), People’s Daily News (People’s Daily, 2018). Cosine similarity between word vectors of insurance synonyms is shown in Tab. 2. The Chinese words are translated into English by Google Translate. Tab. 2 shows that the insurance corpus can greatly improve the word embedding similarity between insurance synonyms compared with other corpora.

5.4 Performance of TOI-CNN on ETIP

Tab. 3 shows the confusion matrix computed by TOI-CNN with Jieba word segmentation, where $th_s = 0.5$ and $th_p = 0.8$. The confusion matrix has eight categories, where seven of them are the categories shown in Tab. 1 and the eighth one is negative. Each row of the matrix corresponds to an actual class, and each column of the matrix corresponds to a predicted class. The neg. in the rightmost column denote the ground truths which have been removed from the final candidates, i.e., false negatives. The neg. in the bottom row denote those final candidates who are not the real elements of seven categories, i.e., false positives. PC is more susceptible to negative sliding windows than other categories because PC is always a kind of time description and easily disturbed by other time descriptions in the insurance contracts. Condition for Payment (CP) and Insured Amount (IA) could be confused with each other, because CP sometimes includes coverage amount descriptions like IA.

Tab. 4 shows the results of precision (P), recall (R) and F1 score on seven categories when $th_p =$

	ETIP	Baidu Encyclopedia	Wikipedia_zh	People’s Daily News
缴纳(pay) vs 交付(deliver)	0.786	0.457	0.344	0.457
解除(release) vs 撤销(cancel)	0.701	0.347	0.311	0.408
期间(period) vs 有效期(validity period)	0.475	0.247	0.249	0.215
投保人(insured) vs 您(you)	0.752	0.355	0.384	0.247
成立(established) vs 生效(effective)	0.555	0.354	0.335	0.426

Table 2: Examples of word embedding similarity between insurance synonyms.

	C	WP	PC	CP	IA	E	T	neg.
C	519	0	0	0	0	0	0	99
WP	0	11	5	2	0	0	0	3
PC	0	0	108	1	0	0	0	77
CP	0	0	3	1171	33	0	0	89
IA	0	0	0	31	982	1	0	54
E	0	0	0	0	4	16	0	5
T	0	0	0	3	0	1	380	14
neg.	118	0	46	82	9	2	21	/

Table 3: Confusion matrix result of TOI-CNN.

0.8 and word segmentation={Jieba, NLPPIR}. The overall performance of Jieba and NLPPIR are approximately equal in TOI-CNN model. In TOI-CNN training, we create negative sliding windows using IoU threshold th_s (see Sec. 4.3). Fig. 3 shows F1 score comparison of negative samples generated with varied th_s when $th_p = 0.8$. The performance of $th_s = 0.5$ is better than that of others and close to that of $th_s = 0.6$.

	Jieba			NLPPIR		
	P	R	F1	P	R	F1
C	0.811	0.834	0.823	0.30	0.807	0.819
WP	1.00	0.524	0.688	1.00	0.476	0.645
PC	0.667	0.581	0.621	0.659	0.586	0.620
CP	0.908	0.902	0.905	0.910	0.901	0.905
IA	0.955	0.920	0.937	0.951	0.924	0.937
E	0.800	0.640	0.711	0.762	0.640	0.696
T	0.945	0.954	0.950	0.941	0.930	0.936
Avg.	0.887	0.878	0.883	0.890	0.866	0.878

Table 4: Evaluation results of TOI-CNN on seven categories.

Tab. 5 shows the comparison of F1 scores with nested NER models, where $th_p = 0.8$. We use Jieba for word segmentation and POS tagging. We compare with two public nested NER models, Muis and Lu (2017) and Ju et al. (2018). We have tuned the hyper-parameters of the baselines for the best performance. We chose the best hyper-parameters via Bayesian optimization provided by (Ju et al., 2018). We set the l2-regularization parameter ($\lambda = 0.01$) and the number of Brown clusters ($n = 140$) in (Muis and Lu, 2017). Our TOI-CNN outperforms other models in C, CP, E, and T, and get an excellent result in the overall performance.

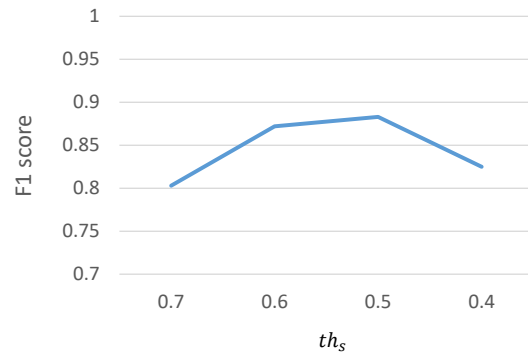


Figure 3: Comparison of negative samples generated with varied th_s .

6 Conclusion

This paper has presented a way of how to tag insurance policies. Seven critical elemental categories in the insurance policy are identified. We collected a large Chinese insurance contract corpus, labeled the samples with seven categories and published the dataset. The proposed TOI-CNN method can effectively solve the overlapping elements extraction problem. The overall performance of TOI-CNN is better than that of the probabilistic graphical models, especially for the overlapped phrases and clauses. Our method can accurately extract the vast majority of Chinese insurance policies according to the pre-defined categories.

7 Acknowledgments

This work was supported by Zhejiang Provincial Natural Science Foundation of China (No.

	C	WP	PC	CP	IA	E	T	Avg.
Muis and Lu (2017)	0.612	0.645	0.759	0.805	0.947	0.711	0.933	0.832
Ju et al. (2018)	0.808	0.757	0.788	0.886	0.783	0.664	0.906	0.854
TOI-CNN	0.823	0.688	0.621	0.905	0.937	0.711	0.950	0.883

Table 5: F1 score comparison with other models.

LY17F020008).

References

- Shaun Azzopardi, Albert Gatt, and Gordon J Pace. 2016. Integrating natural language and formal analysis for legal documents. In *10th Conference on Language Technologies and Digital Humanities*, volume 2016.
- Baidu. 2018. Baidu encyclopedia. <https://baike.baidu.com/>.
- Ilias Chalkidis and Ion Androutsopoulos. 2017. A deep learning approach to contract element extraction. In *JURIX*, pages 155–164.
- Ilias Chalkidis, Ion Androutsopoulos, and Achilleas Michos. 2017. Extracting contract elements. In *Proc. of the 16th Int. Conf. on Artificial Intelligence and Law*, pages 19–28.
- William Cohen and Andrew McCallum. 2004. Information extraction and integration: an overview. In *SIGKDD Conference*.
- Eli Cortez and Altigran S Da Silva. 2013. *Unsupervised information extraction by text segmentation*. Springer.
- Michael Curtotti and Eric McCreath. 2010. Corpus based classification of text in australian contracts.
- George R Doddington, Alexis Mitchell, Mark A Przybocki, Lance A Ramshaw, Stephanie Strassel, and Ralph M Weischedel. 2004. The automatic content extraction (ace) program-tasks, data, and evaluation. In *LREC*, volume 2, page 1.
- Matías García-Constantino, Katie Atkinson, Danushka Bollegala, Karl Chapman, Frans Coenen, Claire Roberts, and Katy Robson. 2017. Cliel: context-based information extraction from commercial law documents. In *Proceedings of the 16th edition of the International Conference on Artificial Intelligence and Law*, pages 79–87. ACM.
- Ismael Hasan, Javier Parapar, and Roi Blanco. 2008. Segmentation of legislative documents using a domain-specific lexicon. In *Database and Expert Systems Application, 2008. DEXA’08. 19th International Workshop on*, pages 665–669. IEEE.
- Meng Hu, Zhixu Li, Yongxin Shen, An Liu, Guan-feng Liu, Kai Zheng, and Lei Zhao. 2017. Cnn-iets: A cnn-based probabilistic approach for information extraction by text segmentation. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, pages 1159–1168. ACM.
- Kishore Varma Indukuri and P Radha Krishna. 2010. Mining e-contract documents to classify clauses. In *Proceedings of the Third Annual ACM Bangalore Conference*, page 7. ACM.
- Jieba. 2017. <https://github.com/fxsjy/jieba>. Accessed: 2018-05-28.
- Meizhi Ju, Makoto Miwa, and Sophia Ananiadou. 2018. A neural layered model for nested named entity recognition. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, volume 1, pages 1446–1459.
- Arzoo Katiyar and Claire Cardie. 2018. Nested named entity recognition revisited. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, volume 1, pages 861–871.
- J-D Kim, Tomoko Ohta, Yuka Tateisi, and Jun’ichi Tsujii. 2003. Genia corpus—a semantically annotated corpus for bio-textmining. *Bioinformatics*, 19(suppl.1):i180–i182.
- Yoon Kim. 2014. Convolutional neural networks for sentence classification. *arXiv preprint arXiv:1408.5882*.
- Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.
- Eneldo Loza Mencía. 2009. Segmentation of legal documents. In *Proceedings of the 12th International Conference on Artificial Intelligence and Law*, pages 88–97. ACM.
- Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.
- Marie-Francine Moens, Caroline Uyttendaele, and Jos Dumortier. 2000. Intelligent information extraction from legal texts. *Information & Communications Technology Law*, 9(1):17–26.
- Aldrian Obaja Muis and Wei Lu. 2017. Labeling gaps between muis2017labelingwords: Recognizing overlapping mentions with mention separators.

In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 2608–2618.

David Nadeau and Satoshi Sekine. 2007. A survey of named entity recognition and classification. *Linguisticae Investigationes*, 30(1):3–26.

NLPIR. 2018. <https://github.com/NLPIR-team/NLPIR>. Accessed: 2018-05-28.

People’s.Daily. 2018. News data from people’s daily. <http://data.people.com.cn/>.

Jakub Piskorski and Roman Yangarber. 2013. Information extraction: Past, present and future. In *Multi-source, multilingual information extraction and summarization*, pages 23–49. Springer.

Alan Ritter, Sam Clark, Oren Etzioni, et al. 2011. Named entity recognition in tweets: an experimental study. In *Proceedings of the conference on empirical methods in natural language processing*, pages 1524–1534. Association for Computational Linguistics.

Wikipedia. 2018. Chinese wikipedia. <https://dumps.wikimedia.org/>.