Chinese Tense Labelling and Causal Analysis

Hen-Hsen Huang, Chang-Rui Yang, and Hsin-Hsi Chen
Department of Computer Science and Information Engineering
National Taiwan University
No. 1, Sec. 4, Roosevelt Road, Taipei, 10617 Taiwan
{hhhuang, cjyang}@nlg.csie.ntu.edu.tw; hhchen@ntu.edu.tw

Abstract

This paper explores the role of tense information in Chinese causal analysis. Both tasks of causal type classification and causal directionality identification are experimented to show the significant improvement gained from tense features. To automatically extract the tense features, a Chinese tense predictor is proposed. Based on large amount of parallel data, our semi-supervised approach improves the dependency-based convolutional neural network (DCNN) models for Chinese tense labelling and thus the causal analysis.

1   Introduction

Causal analysis plays a crucial role in the applications such as event extraction (Hashimoto et al., 2012; 2014), causality inference (Tanaka et al., 2012), question-answering (Oh et al., 2013), and motivation identification (Nguyen et al., 2015). Compared to English, the topic of causal analysis in Chinese is rarely touched. In this work, we explore the role of tense information in Chinese causal analysis. As pointed by Mirza (2014), the causal relation and temporal information is correlated. In a causal relation, the cause intuitively precedes its effect. In other words, the tense information could be useful features in the tasks of causal analysis.

Two tasks of causal analysis are investigated in this study: causal type classification and causal directionality identification. The Chinese discourse relation corpus, Chinese Discourse Treebank (CDTB) (Li et al., 2014), is adopted as our dataset. In CDTB, six types of causality relations, **Purpose**, **Background**, **Hypothetical**, **Inference**, **Condition**, and **Cause-Result**, are defined.

A discourse relation connects two arguments. In the case of causality, one of the two arguments (e.g., *arg*1) presents a situation, and it is causally affected by the other argument (e.g., *arg*2). For example, the first part of the sentence (S1) shows a reason, and the second part, which is underlined, is its effect.

> (S1) 由於產能不足，<u>國內自給率不到四成，大部分要仰賴進口。</u> (Because of in-sufficient capacity, <u>the domestic self-sufficiency rate is less than 40, most rely on imports.</u>)

The direction of *arg*1 and *arg*2 is reversible. Like (S1), the reason is described in the former argument in most cases, and the effect is presented in the latter argument. However, (S2) shows a counter-example that presents the effect in the former part. To exactly extract the cause and the effect in natural language, causal directionality identification is required.

> (S2) <u>西藏銀行部門積極調整信貸結構</u>，以確保農牧業生產等重點產業的投入，加大對工業，能源，交通，通信等建設的正常資金供應量。(<u>Tibet banking sector actively adjust credit structure</u> in order to ensure the input of agricultural production and other key industries, and increase the industrial, energy, transportation, communications, construction of the normal supply of funds.)

There is no tense annotation in the CDTB. For this reason, we select all the samples of causality re-

lation from CDTB, and manually label the tense for each argument as ground-truth for the two tasks. To automatically extract the tense features, a Chinese tense predictor is required. The grammatical tense in English explicitly denotes the temporal information for a given text. In Chinese, however, the temporal information is communicated with aspect particles such as 了 (le) and 着 (zhe) and temporal adverbials such as 现在 ("now") and 明天 ("tomorrow") (Xue et al., 2008; Ge et al., 2015). In other words, it is more challenging to determine the tense in Chinese text. Thus, we propose a semi-supervised algorithm that learns to label tense information in Chinese text. With UM-Corpus, a large English-Chinese parallel corpus aligned at sentence-level (Tian et al., 2014), we generate a pseudo-labelled Chinese tense corpus by deriving the tense information from their English counterpart. Dependency-based convolutional neural network (DCNN) is trained to predict Chinese tense. We incorporate the semi-supervised Chinese tense predictor in the tasks of causal type classification and causal directionality identification. The experimental results are compared with the supervised approach and the ideal situation where human-labelled information is available.

The contribution of this paper is three-fold: (1) we transfer the tense information from English sentence to its Chinese counterpart based on sentence-aligned English-Chinese parallel corpus, (2) we train Chinese tense predictor with DCNN and use it to label tense markers on a Chinese sentence, and (3) we apply the tense information to identify causal type and causal directionality of a sentence. The rest of this paper is organized as follows. Section 2 surveys the related work. Section 3 describes the experimental materials. Section 4 shows our approach to Chinese tense labelling. Section 5 illustrates the use of tense information in causal type and causal directionality identification. Section 6 concludes this paper.

2 Related Work

Causal analysis attracts much attention in AI community for years. A variety of issues have been explored. One of the hottest topic is event analysis, where causal information plays a crucial role (Do et al., 2011; Riaz and Girju, 2013; 2014; Mirza and Tonelli, 2014; Kives et al., 2015). Other applications include generation of event causality hypotheses (Hashimoto et al., 2015), motivation identification (Nguyen et al., 2015), causality detection and extraction (Hashimoto et al., 2012; Mihaila and Ananiadou, 2013), causal inference (Tanaka et al., 2012), question answering (Oh et al., 2013), and future scenario generation (Hashimoto et al., 2014). The correlation between temporality and causality is studied by Mirza (2014) and Mirza and Tonelli (2014).

Unlike English, no grammatical tense is available in Chinese. Various approaches are explored to address the topic of Chinese tense prediction. Liu et al. (2011) propose an unsupervised method for Chinese tense labelling by learning from a Chinese-English parallel corpus. Zhang and Xue (2014) deal with Chinese tense inference by training a supervised model with various linguistic features on a Chinese tense corpus (Xue and Zhang, 2014). Following the unsupervised method by Liu et al. (2011), we develop a semi-supervised model that benefits from a large amount of data labelled by an accurate English tense predictor.

Neural networks such as recurrent neural network (RNN) and convolutional neural network (CNN) are very popular in NLP community. Kim (2014) releases a sentence classifier with convolutional neural network (CNN), where a sentence is represented as a sequence of word vectors (Mikolov et al., 2013). Based on Kim's work, Ma et al. (2015) propose the dependency-based CNN (DCNN) by adding the structure information features to the sentence representation. In this work, we employ DCNN for Chinese tense classification under supervised, unsupervised, and semi-supervised learning.

3 Linguistic Resources

Three types of corpora are used in this work. Section 3.1 describes the corpus for Chinese causal analysis. Section 3.2 and Section 3.3 introduce the corpora for developing our Chinese tense predictor.

3.1 Chinese Causality Corpus

There are few resources for Chinese causal analysis. In this work, we extract instances labelled with causality relation in the Chinese Discourse Treebank (CDTB) (Li et al., 2014) as the basis of our causality dataset. Similar to the English discourse corpus, e.g., Penn Discourse Treebank (PDTB) (Prasad

et al., 2008), CDTB is a Chinese corpus annotated with discourse information. A type of discourse relation is given to a pair of text spans (arguments). For instances of the explicit discourse relation, the connectives (discourse markers) are also annotated.

CDTB does not provide the information of causal directionality. Here we manually label the directionality for each instance extracted from CDTB. Table 1 summarizes the six types of the causality relation. The distributions of explicit/implicit and directionality are shown. Cause-Result, which appears more than 50%, is the majority. Most cases are implicit except Hypothetical and Condition. In terms of directionality, 73.1% of instances are in the direction of Reason-Effect. Furthermore, all the instances of Hypothetical, Inference, and Condition are Reason-Effect. In contrast, 78% of Purpose instances are Effect-Reason. In general, Chinese speakers tend to express the reason before the effect. We release the annotated tense corpus as a resource for NLP community.[1]

Causal Type	Number of Instances	Explicit or Implicit	Number of Instances	%	Directionality	Number of Instances	%
Purpose	332	Explicit	162	48.8%	Reason-Effect	73	22.0%
		Implicit	170	51.2%	Effect-Reason	259	78.0%
Background	127	Explicit	4	3.1%	Reason-Effect	98	77.2%
		Implicit	123	96.9%	Effect-Reason	29	22.8%
Hypothetical	69	Explicit	55	79.7%	Reason-Effect	69	100.0%
		Implicit	14	20.3%	Effect-Reason	0	0.0%
Inference	38	Explicit	3	7.9%	Reason-Effect	38	100.0%
		Implicit	35	92.1%	Effect-Reason	0	0.0%
Condition	71	Explicit	37	52.1%	Reason-Effect	71	100.0%
		Implicit	34	47.9%	Effect-Reason	0	0.0%
Cause-Result	677	Explicit	200	29.5%	Reason-Effect	612	90.4%
		Implicit	477	70.5%	Effect-Reason	65	9.6%
Total	1,314	Explicit	461	35.1%	Reason-Effect	961	73.1%
		Implicit	853	64.9%	Effect-Reason	353	26.9%

Table 1: Statistics of the causality relations in Chinese causality corpus.

3.2 Chinese Tense Corpus

Human-annotated and machine-generated Chinese tense corpora will be used to learn Chinese tense predictor. The human-annotated Chinese tense corpus was developed by Xue and Zhang (2014). Based on a word-aligned Chinese-English parallel treebank, tense, modality, eventually, and event types are manually annotated. Due to the copyright issue, only a subset of data is available for us. For every event, one of the seven tenses is labelled: "Past", "Present", "Future", "Relative Past", "Relative Present", "Relative Future", and "None". We convert all the relative tenses to absolute ones. Finally, total 3,358 instances are extracted. Figure 1 shows the distribution of the human-annotated dataset used in the experiments.
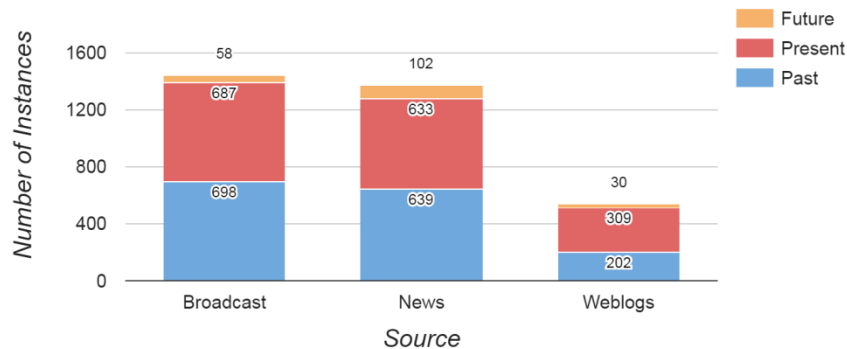

Figure 1: Distribution of instances extracted from the human-labelled Chinese tense corpus.

---

[1] http://nlg.csie.ntu.edu.tw/nlpresource/chinese_causality

3.3 English-Chinese Parallel Corpus

In contrast to the human-annotated Chinese tense corpus, a large English-Chinese parallel corpus, UM-Corpus (Tian et al., 2014), is adopted to label Chinese sentences with tense information. In UM-Corpus, text from eight domains are collected and aligned at sentence-level. A total of 2,215,000 sentences are released. How to develop the machine-generated Chinese tense corpus will be described in Section 4.2.

4 Chinese Tense Labelling

Because grammatical tense is inherent in an English sentence, tense prediction is relatively easier. A large amount of pseudo-labelled data can be generated by tense mapping between English-Chinese parallel sentences. Section 4.1 shows a rule-based tense predictor to determine the tense in the English side. Section 4.2 specifies how to transfer the tense information to its Chinese counterpart by bilingual verb alignment. Section 4.3 proposes a dependency-based convolutional neural network (DCNN) to predict Chinese tense.

4.1 Rule-based English Tense Predictor

Based on the definition of the Stanford typed dependencies[2], we develop a rule-based English tense predictor. For each of the 18 combinations among tenses, voices, and aspects, Table 2 presents the rules in the tense determination. Figure 2 illustrates the dependency tree of the sentence "He was being punished", where the verb "punished" is tagged as VBN (past participle verb), and its dependents contain aux(was/VBD) and auxpass(being/VBG). According to the rules in Table 2, the tense of this sentence is past, the voice is passive, and the aspect is progressive.

Tense	Voice/Aspect	Verb POS	Dep. Auxiliary Verb	Sample
Present	Active/Simple	VB, VBP, VPZ		I write.
	Active/Progressive	VBG	aux(am/VBP)	I am writing.
	Active/Perfect	VBN	aux(have/VBP)	I have written.
	Passive/Simple	VBN	auxpass(is/VBZ)	He is punished.
	Passive/Progressive	VBN	aux(is/VBZ), auxpass(being/VBG)	He is being punished.
	Passive/Perfect	VBN	aux(has/VBZ), auxpass(been/VBN)	He has been punished.
Past	Active/Simple	VBD		I wrote.
	Active/Progressive	VBG	aux(was)-VBD	I was writing.
	Active/Perfect	VBN	aux(had)-VBD	I had written.
	Passive/Simple	VBN	auxpass(was/VBD)	He was punished.
	Passive/Progressive	VBN	aux(was/VBD), auxpass(being/VBG)	He was being punished.
	Passive/Perfect	VBN	aux(had/VBD), auxpass(been/VBN)	He had been punished.
Future	Active/Simple	VB	aux(will/MD)	I will write.
	Active/Progressive	VBG	aux(will/MD), aux(be/VB)	I will be writing.
	Active/Perfect	VBN	aux(will/MD), aux(have/VB)	I will have written.
	Passive/Simple	VBN	aux(will/MD), auxpass(be/VB).	He will be punished.
	Passive/Progressive	VBN	aux(will/MD), aux(be/VB), auxpass(being/VBG)	He will be being punished.
	Passive/Perfect	VBN	aux(will/MD), aux(have/VB) auxpass(been/VBN)	He will have been punished

Table 2: Rules for English tense prediction with the information of POS tagging and dependency parsing.
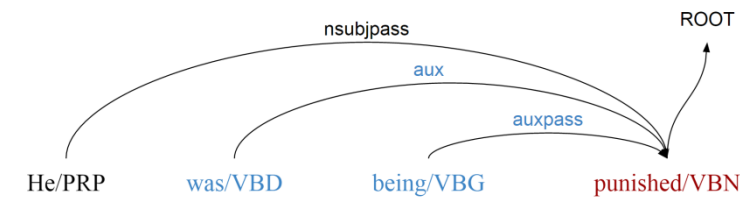

Figure 2: Dependency tree of the sentence "He was being punished".

We evaluate the performance of the English tense predictor on the dataset from the NTHU Academic Writing Database[3]. In this dataset, 1,171 English sentences are carefully annotated with linguistic information such as tense, voice, aspect, and argumentative zone. Our rule-based tense predictor achieves an accuracy of 91.98%. Error analysis shows that most wrongly labelled instances are due to the errors of POS tagging and dependency parsing. (S3) shows an example. The verb (VB) "image" is wrongly labelled as a noun (NN) by the Stanford tagger. Our rule-based English tense predictor is released as a tool[4].

```
(S3) "When combined with multiphoton excitation, both schemes
can image thick samples with three-dimensional optical section-
ing and much improved resolution."
```

4.2   Machine-generated Chinese Tense Corpus

As described in Section 3, total 2,215,000 English-Chinese parallel sentences are released in UM-Corpus. We perform Chinese word segmentation, POS tagging, and dependency parsing for the Chinese sentences with Stanford CoreNLP (Manning et al., 2014). UM-Corpus is aligned at sentence level, but a sentence may contain multiple verbs. For a sentence with multiple verbs, we employ the alignment tool GIZA++[5] to align English verbs with their Chinese counterparts (Och and Ney, 2003). However, not all cases are perfectly aligned. In the example shown in Figure 3, the English verb "go" is wrongly aligned with two Chinese tokens 那回 ("that time") and 去 ("go") because the Chinese word segmenter does not correctly separate 那 ("the") and 回 ("back"). To reduce the noise, we remove all the instances that fail to align. As a result, we obtain 615,521 Chinese instances with tense information as the pseudo-labelled corpus.

Table 3 shows the statistics of this corpus. On the one hand, instances of the *present* tense, which occupy 63.75%, are the majority. On the other hand, only 6.7% of the instances are with the *future* tense. Among all domains, the odd distribution of Law is observable. About 49.09% of the instances in the Law domain are in *future* tense because most legal provisions are made to regulate what will happen in the future. Microblog is the smallest domain, i.e., only 954 instances are found.
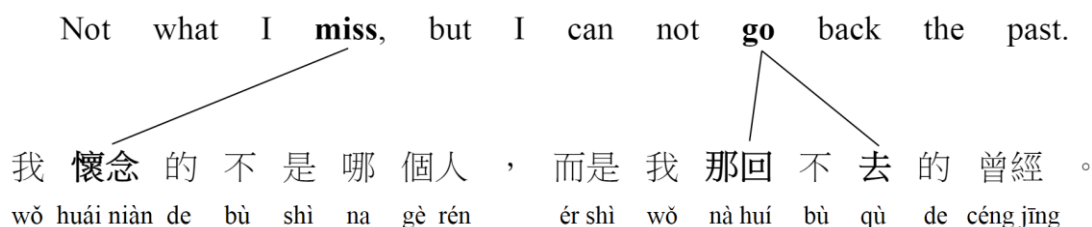

Figure 3: An imperfectly aligned case where the English verb "go" is aligned with two Chinese tokens due to word segmentation error.

4.3   DCNN-Based Chinese Tense Predictor

Tense labelling for a given sentence is a task of sentence classification. In this work, we employ the dependency-based convolutional neural network (DCNN) as the classifier (Ma et al., 2015). Based on the sentence classifier with CNN (Kim, 2014), the DCNN gains improvement by incorporating the information of linguistic structure. In addition to a sequence of word vectors like the skip-gram (Mikolov et al., 2013), the outcome of dependency parsing such as ancestor paths and siblings are added to the sentence representation. In this work, the skip-gram is trained on the Tagged Chinese Gigaword (CGW) corpus 2.0 (Graff et al., 2005; Huang, 2009), and a Chinese word is represented as a vector with a dimension of 400.

---

[3] http://writing.wwlc.nthu.edu.tw/writcent
[4] http://nlg.csie.ntu.edu.tw/nlpresource/english_tense_predictor
[5] http://www.statmt.org/moses/giza/GIZA++.html

Domains	Past		Present		Future		Total
	#	%	#	%	#	%	
Education	51,906	32.67%	98,954	62.28%	8,022	5.05%	158,882
Laws	1,370	4.56%	13,930	46.35%	14,754	49.09%	30,054
Microblog	155	16.25%	743	76.94%	56	5.87%	954
News	50,768	34.79%	88,812	60.86%	6,350	4.35%	145,930
Science	12,222	19.75%	46,065	74.45%	3,586	5.80%	61,873
Spoken	25,924	33.37%	48,313	62.19%	3,447	4.44%	77,684
Subtitles	25,735	29.68%	57,064	65.82%	3,898	4.50%	86,697
Thesis	14,416	26.97%	38,543	72.11%	488	0.98%	53,447
Total	182,496	29.65%	392,424	63.75%	40,601	6.70%	615,521

Table 3: Distribution of the machine-labelled Chinese tense corpus.

4.3.1 Unsupervised Learning for Chinese Tense Labelling

In the setting of unsupervised learning, we train the DCNN classifier on the machine-generated Chinese tense corpus, and test on the human-annotated Chinese tense corpus. The support vector machine (SVM) with RBF kernel and the random forest (RF) classifiers are also trained as baseline models. The hyperparamters of both classifiers are adjusted with grid search. The McNemar test is applied for significance testing at p=0.05. Table 4 shows the results in accuracies in the order of domain size. In general, the more the data, the better the performance. All the three models trained on the tiny Microblog dataset are superior to those trained on Law, the relatively larger dataset, because of the odd distribution of the Law domain. The DCNN significantly outperforms the other two models in most domains except for Microblog and Subtitles. DCNN with the data from all domains achieves the highest accuracy of 68.62% in the unsupervised approach. The performances of SVM and RF with all data are slightly decreased. That confirms the selection of pseudo data is crucial for traditional classifiers (Liu et al., 2011). In contrast, the DCNN model is not affected by this issue. That shows the high discriminative ability of the neural network model.

Domains	Number of Instances	DCNN	SVM	RF
Microblog	954	48.62%	50.14%	49.45%
Law	30,054	43.28%	41.06%	40.75%
Thesis	53,447	54.95%	53.81%	49.97%
Science	61,873	57.70%	56.69%	52.56%
Spoken	77,384	65.07%	62.90%	60.38%
Subtitles	86,697	55.43%	56.27%	56.63%
News	145,930	66.80%	64.38%	62.36%
Education	158,882	67.91%	64.70%	62.22%
All Domains	615,521	**68.62%**	62.20%	61.57%

Table 4: Experimental results of learning from pseudo-labelled data by domains.

4.3.2 (Semi-)Supervised Learning for Chinese Tense Labelling

This section evaluates our model under supervised and semi-supervised learning. Five-fold cross validation is performed on the 3,358 genuine instances. For each fold, one fifth of 3,358 genuine instances (human-annotated) are used for testing, and four-fifth of 3,358 genuine instances and various amounts of pseudo-labelled (machine-generated) data are used for training. Table 5 compares the accuracies of

Settings	# Genuine Data	# Pseudo Data	DCNN	SVM	RF
Supervised	3,358	0	66.77%	64.79%	65.92%
Unsupervised	0	615,521	68.62%	62.20%	61.57%
Semi-Supervised	3,358	10,000	68.00%	66.10%	62.85%
	3,358	20,000	68.59%	66.33%	61.99%
	3,358	100,000	67.97%	66.60%	63.80%
	3,358	300,000	68.56%	66.42%	64.13%
	3,358	615,521	**69.64%**	65.86%	63.83%

Table 5: Comparison of supervised, unsupervised, and semi-supervised learning for Chinese tense labelling.

supervised, unsupervised, and semi-supervised training. Our model gains improvement by adding the genuine instances to large pseudo-labelled corpus. Compared to supervised training, adding the pseudo-labelled data increases the performance of DCNN up to 69.64%. SVM is also improved under semi-supervised training. RF is a counter-example that performs best under supervised training, but still does not compete with DCNN.

5 Causal Analysis

The DCNN-based Chinese tense predictor is used to label the tense features to the instances in the causal corpus. Sections 5.1 and 5.2 confirm if the two tasks of causal analysis gain improvement from tense information. This work focuses on the correlation between tense information and causal analysis in Chinese text. The bag-of-word SVM classifiers with or without tense features are experimented to verify if the tense information improves the two tasks of causal analysis. The tense features consist of six binary values: *arg1-is-past*, *arg1-is-present*, *arg1-is-future*, *arg2-is-past*, *arg2-is-present*, and *arg2-is-future*. Three sources of tense features are compared: labelled by the supervised model (M_{super}), labelled by the semi-supervised model (M_{semi}), and labelled by human (M_h). Refer to Section 4.3.2, the supervised model is the DCNN-based Chinese tense predictor trained on 3,358 genuine data. The semi-supervised model is the DCNN-based Chinese tense predictor trained on the combination of 3,358 genuine and 615,521 pseudo data. The model with human-labelled tense, M_h, is an ideal model since human-labelled information is unavailable in real applications. Five-fold cross validation is performed. The hyperparameters are adjusted for the SVM (RBF) classifier with grid search. The McNemar test is applied for significance testing at p=0.05.

5.1 Causal Type Classification

In the task of causal type classification, the model predicts one of the six causal types for a given argument pair. The performances measured in accuracy and macro F-score are given in Table 6. Compared to the model with only Word feature (M_w), tense information indeed improves the performance of this task. M_h is significantly superior to M_w at *p*=0.05. Furthermore, it is surprising that M_{semi} competes with M_h.

Model	M_w		M_{super}		M_{semi}		M_h	
	Accuracy	F-Score	Accuracy	F-Score	Accuracy	F-Score	Accuracy	F-Score
Explicit	75.48%	44.74%	76.35%	44.93%	76.57%	46.48%	**77.00%**	**46.63%**
Implicit	59.78%	30.84%	60.60%	29.09%	**62.36%**	**32.48%**	62.25%	29.71%
Overall	65.28%	35.71%	66.11%	34.64%	67.33%	**37.38%**	**67.41%**	35.64%

Table 6: Experimental results of causal type classification.

The confusion matrices of M_h and M_{semi} are shown in Tables 7 and 8, respectively. M_h tends to predict an instance to Cause-Result, the largest type of the six. In contrast, M_{semi} is fairer that more instances are classified to minor types.

(S4) is an example which is correctly classified to Cause-Result by M_{semi}, but wrongly classified to Background by M_h. The part of effect is underlined, while the rest is the part of reason. This instance shows the grey zone between Cause-Result and Background. By definition, Cause-Result holds on a stronger factually cause-effect relation.

Types	Purpose	Background	Hypothetical	Inference	Condition	Cause-Result
Purpose	66.57%	0.00%	0.30%	0.00%	1.51%	31.63%
Background	6.30%	19.69%	0.00%	0.79%	0.00%	73.23%
Hypothetical	13.04%	0.00%	49.28%	0.00%	1.45%	36.23%
Inference	10.53%	7.89%	0.00%	5.26%	5.26%	71.05%
Condition	21.13%	1.41%	1.41%	0.00%	21.13%	54.93%
Cause-Result	6.20%	4.73%	0.74%	0.44%	0.89%	87.00%

Table 7: Confusion matrix of the model with human-labelled tense features.

Types	Purpose	Background	Hypothetical	Inference	Condition	Cause-Result
Purpose	72.29%	0.60%	0.30%	0.00%	1.20%	25.60%
Background	6.30%	31.50%	0.00%	0.79%	1.57%	59.84%
Hypothetical	17.39%	0.00%	47.83%	0.00%	1.45%	33.33%
Inference	10.53%	15.79%	0.00%	5.26%	5.26%	63.16%
Condition	28.17%	1.41%	1.41%	0.00%	22.54%	46.48%
Cause-Result	9.45%	6.20%	1.33%	0.44%	0.74%	81.83%

Table 8: Confusion matrix of the model with the tense features labelled by our semi-supervised tense predictor.

(S4) 僅中國陸上三大天然氣最富集的四川盆地，近四十多年來，已累計生產一千六百三十三億立方米天然氣。<u>基本上解決了成都、重慶等一批大中城市的民用燃料，並形成以天然氣為原料的中國最大的維尼龍生產線四川維尼龍廠。</u> (Sichuan Basin, the only place with the three major natural gas resources in China, nearly forty years, has produced a total of 163.3 billion cubic meters of natural gas. <u>This basically provided domestic fuel for Chengdu, Chongqing and other cities, and found the Sichuan Vinylon plant, China's largest production line of Vinalon using natural gas as raw materials</u>)

5.2 Causal Directionality Identification

In the task of causal directionality identification, the binary classifier predicts one of the two direction (i.e., Reason-Effect and Effect-Reason) for a given argument pair. Refer to Table 1, 73.1% of instances in the direction of Reason-Effect, and no instances in the direction of Effect-Reason are found in the Hypothetical, Inference, and Condition types. Thus, only the performances of Purpose, Background, and Cause-Result are reported in Table 9. The results are consistent with the task of causal type classification. The ideal model M_h achieves the best performance and significantly outperforms M_w (p=0.05), and M_{semi} is second.

Model	M_w		M_{super}		M_{semi}		M_h	
	Accuracy	F-Score	Accuracy	F-Score	Accuracy	F-Score	Accuracy	F-Score
Purpose	87.04%	78.31%	88.25%	81.23%	88.85%	82.19%	91.26%	86.04%
Background	77.16%	43.55%	77.16%	43.55%	77.16%	43.55%	78.74%	50.39%
Cause-Result	90.84%	54.52%	90.84%	53.29%	90.84%	55.68%	91.13%	59.18%
Overall	88.18%	60.52%	88.53%	60.35%	88.71%	62.06%	89.76%	66.03%

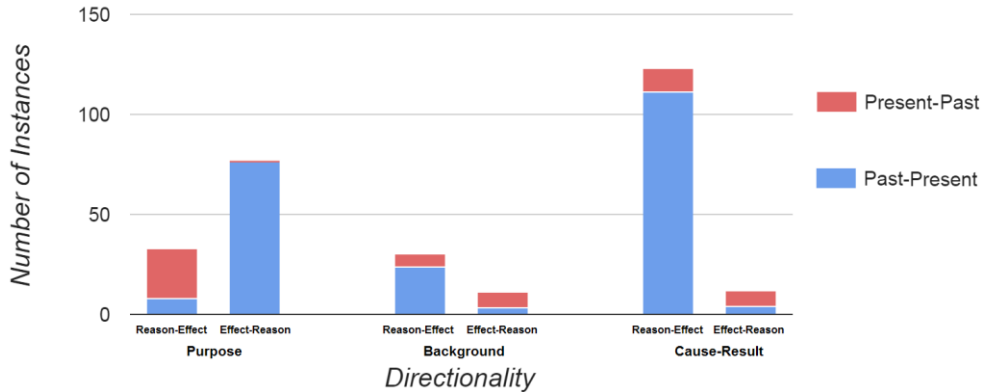Table 9: Experimental results of causal directionality identification.

Figure 4: Relationship between causal directionality and chronology.

Figure 4 presents the relationship between causal directionality and chronology. Due to the sparseness of the tense of future, only the two transitions, Past to Present (forward) and Present to Past (reverse), are shown. The direction of Reason-Effect is the majority in the types of Background and Cause-Result, where the reason and the effect of most instances happen in the order of chronology. The type of Purpose is different. As the instance of Purpose shown in (S5), where the part of effect is underlined, while the rest is the part of reason. In the case of Purpose, the reason usually happens after the effect because the reason is the goal, and the effect is the manner to achieve to goal. The statistics reflects the special natural of the Purpose type.

> (S5)<u>香港特別行政區行政長官董建華今日（星期二）與四萬名信眾出席佛教界慶祝香港回歸祈福大會</u>，為香港的繁榮安定及世界和平祝禱。(<u>Today (Tuesday), Hong Kong Chief Executive Tung Chee-hwa and forty thousand faithful attended the Buddhist blessing event to celebrate the return of Hong Kong,</u> for the prosperity and stability of Hong Kong and the world peace.)

6 Conclusion

This work investigates the role of tense information in Chinese causal analysis. We annotate the tense information on CDTB, and propose an approach that learns from parallel data for Chinese tense labelling. Our semi-supervised approach improves the performance of the DCNN and SVM models. The best model achieves an accuracy of 69.64% in Chinese tense labelling, while its outcome is useful information for the tasks of causal analysis.

Experimental results confirm the causal analysis tasks gain improvement from the tense features. Furthermore, we observe the high discriminative ability of the neural network model when the pseudo-labelled data are added to training set. Linguistics phenomena about causality and chronology are discussed with the evidence of data. We release the annotated tense corpus and a high performance rule-based English tense predictor for NLP community.

7 Acknowledgements

References

Quang Xuan Do, Yee Seng Chan, Dan Roth. 2011. Minimally Supervised Event Causality Identification. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, pages 294–303, Edinburgh, Scotland, UK.

Tao Ge, Heng Ji, Baobao Chang, and Zhifang Sui. 2015. One Tense per Scene: Predicting Tense in Chinese Conversations. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing* (*Short Papers*), pages 668–673, Beijing, China.

David Graff, Ke Chen, Junbo Kong, and Kazuaki Maeda. 2005. Chinese Gigaword Second Edition LDC2005T14. Web Download. Philadelphia: Linguistic Data Consortium.

Chikara Hashimoto, Kentaro Torisawa, Julien Kloetzer, and Jong-Hoon Oh. 2015. Generating Event Causality Hypotheses through Semantic Relations. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*, pages 2396- 2403.

Chikara Hashimoto, Kentaro Torisawa, Julien Kloetzer, Motoki Sano, Istvan Varga, Jong-Hoon Oh, and Yutaka Kidawara. 2014. Toward Future Scenario Generation: Extracting Event Causality Exploiting Semantic Rela-

tion, Context, and Association Features. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, pages 987–997, Baltimore, Maryland, USA.

Chikara Hashimoto, Kentaro Torisawa, Stijn De Saeger, Jong-Hoon Oh, and Jun'ichi Kazama. 2012. Excitatory or Inhibitory: A New Semantic Orientation Extracts Contradiction and Causality from the Web. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pages 619–630, Jeju Island, Korea.

Chu-Ren, Huang. 2009. Tagged Chinese Gigaword Version 2.0 LDC2009T14. Web Download. Philadelphia: Linguistic Data Consortium.

Yoon Kim. 2014. Convolutional Neural Networks for Sentence Classification. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing* (EMNLP), pages 1746–1751, Doha, Qatar.

Christopher Kives, Stephen G. Ware, and Lewis J. Baker. 2015. Evaluating the Pairwise Event Salience Hypothesis in *Indexter*. In *Proceedings of the Eleventh AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment* (*AIIDE-15*), pages 30-36.

Yancui Li, Wenhe Feng, Jing Sun, Fang Kong, and Guodong Zhou. 2014. Building Chinese Discourse Corpus with Connective-driven Dependency Tree Structure. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing* (*EMNLP*), pages 2105–2114, Doha, Qatar.

Feifan Liu, Fei Liu, and Yang Liu. 2011. Learning from Chinese-English Parallel Data for Chinese Tense Prediction. In *Proceedings of the 5th International Joint Conference on Natural Language Processing*, pages 1116–1124, Chiang Mai, Thailand.

Mingbo Ma, Liang Huang, Bing Xiang, and Bowen Zhou. 2015. Dependency-based Convolutional Neural Networks for Sentence Embedding. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing* (*Short Papers*), pages 174–179, Beijing, China.

Christopher D. Manning, Mihai Surdeanu, John Bauer, Jenny Finkel, Steven J. Bethard, and David McClosky. 2014. The Stanford CoreNLP Natural Language Processing Toolkit. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 55-60.

Claudiu Mihaila and Sophia Ananiadou. 2013. What causes a causal relation? Detecting Causal Triggers in Biomedical Scientific Discourse. In *Proceedings of the ACL Student Research Workshop*, pages 38–45, Sofia, Bulgaria.

Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Distributed Representations of Words and Phrases and their Compositionality. In *Advances in Neural Information Processing Systems*, 26, pages 3111-3119.

Paramita Mirza. 2014. Extracting Temporal and Causal Relations between Events. In *Proceedings of the ACL 2014 Student Research Workshop*, pages 10–17, Baltimore, Maryland USA.

Paramita Mirza and Sara Tonelli. 2014. An Analysis of Causality between Events and its Relation to Temporal Information. In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, pages 2097–2106, Dublin, Ireland.

Dong Nguyen, Tijs van den Broek, Claudia Hauff, Djoerd Hiemstra, and Michel Ehrenhard. 2015. #SupportTheCause: Identifying Motivations to Participate in Online Health Campaigns. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 2570–2576, Lisbon, Portugal.

Franz Josef Och and Hermann Ney. 2003. A Systematic Comparison of Various Statistical Alignment Models. *Computational Linguistics*, 29(1):19-51.

Jong-Hoon Oh, Kentaro Torisawa, Chikara Hashimoto, Motoki Sano, Stijn De Saeger, and Kiyonori Ohtake. 2013. Why-Question Answering using Intra- and Inter-Sentential Causal Relations. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, pages 1733–1743, Sofia, Bulgaria.

Rashmi Prasad, Alan Lee, Nikhil Dinesh, Eleni Miltsakaki, Geraud Campion, Aravind Joshi, and Bonnie Webber. 2008. Penn Discourse Treebank Version 2.0 LDC2008T05. Web Download. Philadelphia: Linguistic Data Consortium, 2008. https://catalog.ldc.upenn.edu/LDC2008T05

Mehwish Riaz and Roxana Girju. 2013. Toward a Better Understanding of Causality between Verbal Events: Extraction and Analysis of the Causal Power of Verb-Verb Associations. In *Proceedings of the SIGDIAL 2013 Conference*, pages 21–30, Metz, France.

Mehwish Riaz and Roxana Girju. 2014. In-depth Exploitation of Noun and Verb Semantics to Identify Causation in Verb-Noun Pairs. In *Proceedings of the SIGDIAL 2014 Conference*, pages 161–170, Philadelphia, U.S.A.

Shohei Tanaka, Naoaki Okazaki, and Mitsuru Ishizuka. 2012. Acquiring and Generalizing Causal Inference Rules from Deverbal Noun Constructions. In *Proceedings of COLING 2012: Posters*, pages 1209–1218, COLING 2012, Mumbai, India.

Liang Tian, Derek F. Wong, Lidia S. Chao, Paulo Quaresma, Francisco Oliveira, Shuo Li, Yiming Wang, and Yi Lu. 2014. UM-Corpus: A Large English-Chinese Parallel Corpus for Statistical Machine Translation. In *Proceedings of the 9th International Conference on Language Resources and Evaluation* (*LREC*), Reykjavik, Iceland.

Nianwen Xue, Zhong Hua, and Kai-Yun Chen. 2008. Annotating "tense" in a Tense-less Language. In *Proceedings of the Fifth International Conference on Language Resources and Evaluation*, pages 3461-3466, Marrakech, Morocco.

Nianwen Xue and Yuchen Zhang. 2014. Buy one get one free: Distant Annotation of Chinese Tense, Event Type, and Modality. In *Proceedings of the 9th International Conference on Language Resources and Evaluation* (*LREC*), Reykjavik, Iceland.

Yuchen Zhang and Nianwen Xue. 2014. Automatic Inference of the Tense of Chinese Events Using Implicit Linguistic Information. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1902–1911, Doha, Qatar.