

Contingency and Comparison Relation Labeling and Structure Prediction in Chinese Sentences

Hen-Hsen Huang

Department of Computer Science and
Information Engineering,
National Taiwan University,
Taipei, Taiwan
hhhuang@nlg.csie.ntu.edu.tw

Hsin-Hsi Chen

Department of Computer Science and
Information Engineering,
National Taiwan University,
Taipei, Taiwan
hhchen@csie.ntu.edu.tw

Abstract

Unlike in English, the sentence boundaries in Chinese are fuzzy and not well-defined. As a result, Chinese sentences tend to be long and consist of complex discourse relations. In this paper, we focus on two important relations, *Contingency* and *Comparison*, which occur often inside a sentence. We construct a moderate-sized corpus for the investigation of intra-sentential relations and propose models to label the relation structure. A learning based model is evaluated with various features. Experimental results show our model achieves accuracies of 81.63% in the task of relation labeling and 74.8% in the task of relation structure prediction.

1 Introduction

Discourse relation labeling has attracted much attention in recent years due to its potential applications such as opinion mining, question answering, etc. The release of the Penn Discourse Treebank (Joshi and Webber, 2004; Prasad et al., 2008) has advanced the development of English discourse relation recognition (Lin et al., 2009; Pitler et al., 2009; Pitler and Nenkova, 2009; Wang et al., 2010). For Chinese, a discourse corpus is not publicly available yet. Thus, the research on Chinese discourse relation recognition is relatively rare. Most notably, Xue (2005) annotated discourse

connectives in the Chinese Treebank. Our previous work labeled four types of relations, including temporal, contingency, comparison and expansion, between two successive sentences, and reported an accuracy of 88.28% and an F-score of 62.88% (Huang and Chen, 2011). The major issue of our work is the determination of discourse boundaries. Each Chinese sentence is always treated as one of the two arguments in their annotation and many instances of the Contingency and the Comparison remain uncaught.

As suggested by the Penn Discourse Treebank annotation guidelines, an argument is possibly some clauses in a sentence, a sentence, or several successive sentences. In Chinese, the Contingency and the Comparison relations are likely to occur within a sentence. Thus, a lot of the Contingency relations and the Comparison relations are missing from annotation in the corpus used in our previous work, and the classification performance for these two relations, especially the Contingency relation, is especially poor (Huang and Chen, 2011).

In contrast to Chinese inter-sentential discourse relation detection (Huang and Chen, 2011) and the study of English coherence evaluation (Lin et al., 2011), this paper focuses on the Contingency relation and the Comparison relations that occur inside a sentence. In Chinese, the relations usually occur in the sentences which contain many clauses. For example, two relations occur in sample (S1).

(S1) 管理處雖嘗試要讓長期來作為大台北後花園的陽明山區更回歸自然 (“Although the

management office tried to make the Yangmingshan area a more natural environment as the long-term garden of Taipei”), 但隨著週休二日、經濟環境改善 (“But due to the two-day weekend and the improved economic conditions”), 遊客帶來停車、垃圾等間接影響卻更為嚴重 (“The issues of tourists parking, garbage, and other indirect effects become more serious”).

In (S1), the long sentence consists of three clauses, and such a Chinese sentence is expressed as multiple short sentences in English. Figure 1 shows that a *Comparison* relation occurs between the first clause and the last two clauses, and a *Contingency* relation occurs between the second clause and the third clause. An explicit paired discourse marker 雖 (although) ... 但 (but) denotes a Comparison relation in (S1), where the first clause is the first argument of this relation, and the second and the third clauses are the second argument of this relation. In addition, an implicit Contingency relation also occurs between the second and the third clauses. The second clause is the cause argument of this Contingency relation, and the third clause is its effect. It shows a nested relation, which makes relation labeling and relation structure determination challenging.

In Chinese, an explicit discourse marker does not always uniquely identify the existence of a particular discourse relation. In sample (S2), a discourse marker 而 “moreover” appears, but neither Contingency nor Comparison relation exists between the two clauses. The discourse marker 而 has many meanings. Here, It has the meaning of “and” or “moreover”, which indicates an Expansion relation. In other usages, it may have the meaning of “but” or “however”, which indicates a Comparison relation.

(S2) 而大陸經濟開放 10 年以來，其進步更令人刮目相看。 (“Moreover, the progress of mainland is more impressive due to its economic openness for the last 10 years.”)

Note that the relation structure of a sentence cannot be exactly derived from the parse tree of the sentence. Shown in Figure 2 is the structure of sample (S3) based on the syntactic tree generated by the Stanford parser. However, it is clear that the

correct structure of (S3) is the one shown in Figure 3.

(S3) 目前雖然還只能在圖片上讓女性露露臉 (“Although women only appear in the pictures”), 但未來女性的貢獻 (“The contribution of women”), 將是教科書另一個著墨的重點 (“Will be another major focus in textbooks in the future”).

This shows that the Stanford parser does not capture the information that the last two clauses form a unit, which in turn is one of the two arguments of a Comparison relation.

In this work, we investigate intra-sentential relation detection in Chinese. Given a Chinese sentence, our model will predict if Contingency or Comparison relations exist, and determine their relation structure. In Section 2, the development of a corpus annotated with Contingency and Comparison relations is presented. The methods and the features are proposed in Section 3. In Section 4, the experimental results are shown and discussed. Finally, Section 5 concludes this paper.

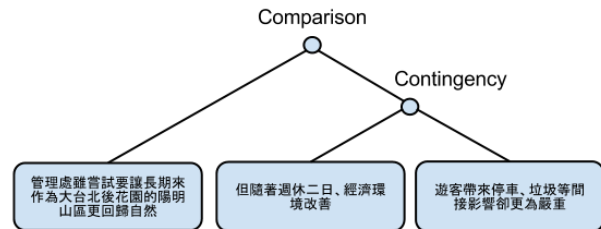


Figure 1: Relation structure of sample (S1).

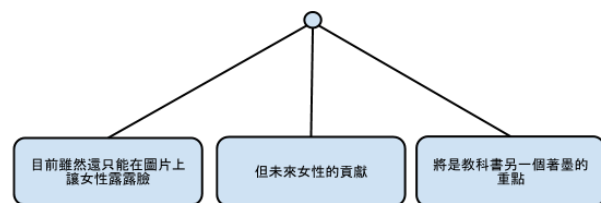


Figure 2: Structure of sample (S3) based on the syntactic tree generated by the Stanford parser.

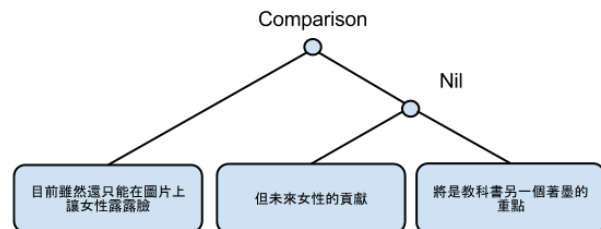


Figure 3: Correct structure of sample (S3)

2 Dataset

The corpus is based on the Sinica Treebank (Huang et al., 2000). A Total of 81 articles are randomly selected from the Sino and Travel sets. All the sentences that consist of two, three, and four clauses are extracted for relation and structure labeling by native Chinese speakers.

A web-based system is developed for annotation. The annotation scheme is designed as follows. An annotator first signs in to the annotation system, and a list of sentences that are assigned to the annotator are given. The annotator labels the sentences one by one in the system. A sentence is split into clauses along commas, and all of its feasible binary tree structures are shown in the interface. The annotator decides if a Contingency/Comparison relation occurs in this sentence. The sentence will be marked as “Nil” if no relation is found. If there is at least one relation in this sentence, the annotator then chooses the best tree structure of the relations, and the second page is shown.

The previously chosen tree structure is presented again, and at this time the annotator has to assign a suitable relation type to each internal node of the tree structure. The relation type includes Contingency “因果”, Comparison “轉折”, and Nil. For example, in sample (S4), its three internal nodes are annotated with three relation types as shown in Figure 4.

(S4) 即使沒有傳承的使命感 (“Even without the sense of mission of the heritage”), 為了尋求更好的治療方式 (“In order to seek better treatments”), 也會驅使這些醫學工作者跨越領域區隔 (“These medical workers will be driven crossing domain areas”), 去尋找資源 (“To find resources”).

The number of feasible relation structures of a sentence may be very large depending on the number of clauses. For a sentence with n clauses, the number of its feasible structures is given as the recursive function $f(n)$ as follows, and the number of its feasible relation structures is $3^{n-1}f(n)$.

$$f(n) = \begin{cases} 1, & n = 1 \\ \sum_{i=1}^{n-1} f(n-i)f(i), & n > 1 \end{cases}$$

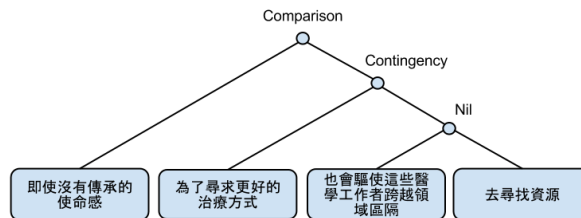


Figure 4: Relation structure of sample (S4).

Explicit/implicit	Relations	2-Clause	3-Clause	4-Clause	Total	%
Explicit	Both	0	5	6	11	0.89%
	Contingency	59	72	45	176	14.31%
	Comparison	41	57	22	120	9.76%
	Nil	269	249	169	687	55.85%
Implicit	Both	0	0	0	0	0.00%
	Contingency	11	8	0	19	1.54%
	Comparison	6	0	0	6	0.49%
	Nil	125	56	4	211	17.15%
All		511	447	272	1,230	100.00%

Table 1: Statistics of the dataset.

For a two-clause sentence, there are only one tree structure and three possible relation tags (Contingency, Comparison, and Nil) for the only one internal node, the root. For a three-clause sentence, there are two candidate tree structures and nine combinations of the relation tags. For a four-clause sentence, there are five candidate tree structures and 27 combinations of the relation tags. There are theoretically 3, 18, and 135 feasible relation structures for the two-, three-, and four-clause sentences, respectively, though only 49 types of relations structures are observed in the dataset.

Each sentence is shown to three annotators, and the majority is taken as the ground-truth. The Fleiss-Kappa of the inter-annotator agreement is 0.44 (moderate agreement). A final decider is involved to break ties. The statistics of our corpus are shown in Table 1. The explicit data are those sentences which have at least one discourse marker. The rest of the data are implicit. A total of 11 explicit sentences which contain both Contingency and Comparison relations form complex sentence compositions. The implicit samples are relatively rare.

3 Methods

To predict the intra-sentential relations and structures, two learning algorithms, the modern implementation of the decision tree algorithm,

C5.0¹, and the support vector machine, SVMlight², are applied. The linguistic features are the crucial part in the learning-based approaches. Various features from different linguistic levels are evaluated in the experiments as shown below.

Word: The bags of words in each clause. The Stanford Chinese word segmenter³ is applied to all the sentences to tokenize the Chinese words. In addition, the first word and the last word in each clause are extracted as distinguished features.

POS: The bags of parts of speech (POS) of the words in each clause are also taken as features. All the sentences in the dataset are sent to the Stanford parser⁴ that parses a sentence from a surface form into a syntactic tree, labels POS for each word, and generates all the dependencies among the words. In addition, the POS tags of the first word and the last word in each clause are extracted as distinguished features.

Length: Several length features are considered, including the number of clauses in the sentence and the number of words for each clause in the sentence.

Connective: In English, some words/phrases called connectives are used as discourse markers. For example, the phrase “due to” is a typical connective that indicates a Contingency relation, and the word “however” is a connective that indicates a Comparison relation.

Similar to the connectives in English, various words and word pair patterns are usually used as discourse markers in Chinese. A dictionary that contains several types of discourse markers is used. The statistics of the connective dictionary and samples are listed in Table 2. An intra-sentential phrase pair indicates a relation which occurs only inside a sentence. In other words, a relation occurs when the two phrases of an intra-sentential pair exist in the same sentence no matter whether they are in the same clause or not. In contrast, an inter-sentential connective indicates a relation that can occur across neighboring sentences. Some connectives belong to both intra-sentential and inter-sentential types. Each connective in each clause is detected and marked with its corresponding type. For example, the phrase 相對

的 “In contrast” will be marked as a connective that belongs to Comparison relation. The number of types and scopes of the connectives in a sentence are used as features.

Dependency: The dependencies among all words in a sentence are used as features. The Stanford parser generates dependency pairs from the sentence. A dependency pair consists of two arguments, i.e., the governor and the dependent, and their types. We are interested in those dependency pairs that are across two clauses. That is, the two arguments of a pair are from different clauses. In our assumption, the clauses have a closer connection if some dependencies occur between them. All such dependency pairs and their types are extracted and counted.

Structure: Recent research work reported improved performance using syntactic information for English discourse relation detection. In the work of Pilter and Nenkova (2009), the categories of a tree node, its parent, its left sibling, and its right sibling are taken as features. In the work of Wang et al. (2010), the entire paragraph is parsed

Relation	Type	#	Samples
Temporal	Single Phrase	41	目前 “now” 之後 “after”
	Intra-Sent Phrase Pair	80	接著...再 “Then...again” 當初...曾 “At first...ever”
	Inter-Sent Phrase Pair	30	當初...後來 “Initially...Later” 最早...緊接著 “At first...Then”
Contingency	Single Phrase	62	如此一來 “As a result” 假設 “If”
	Intra-Sent Phrase Pair	180	如果...則 “If ... then” 無論...都 “Whether ...”
	Inter-Sent Phrase Pair	14	既然...看來 “Since... It seems” 幸而...不然 “Fortunately... otherwise”
Comparison	Single Phrase	34	相對的 “In contrast” 未料 “Unexpectedly”
	Intra-Sent Phrase Pair	38	即使...卻 “Even ... but” 雖然...仍 “Although...still”
	Inter-Sent Phrase Pair	15	雖說...其實 “Although... In fact” 儘管...然而 “Although... However”
Expansion	Single Phrase	182	除此之外 “in addition” 而且 “moreover”
	Intra-Sent Phrase Pair	106	不只...而且 “Not only...but also” 或者...或者 “or...or”
	Inter-Sent Phrase Pair	26	首先...其次 “Firstly...Secondly” 既然...況且 “Since...Furthermore”

Table 2: Statistics of connectives (discourse markers).

¹ <http://www.rulequest.com/see5-unix.html>

² <http://svmlight.joachims.org/>

³ <http://nlp.stanford.edu/software/segmenter.shtml>

⁴ <http://nlp.stanford.edu/software/lex-parser.shtml>

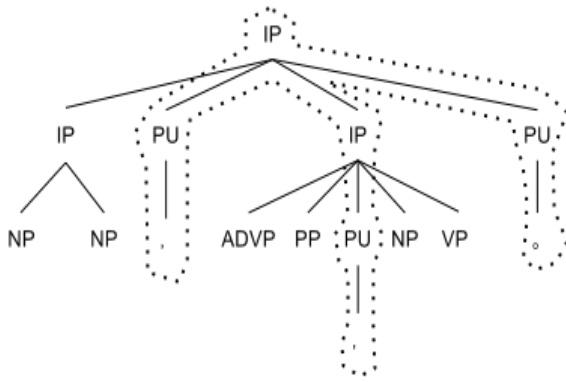


Figure 5: The upper three level sub-tree of (S1) and the punctuation sub-tree of (S1).

as a syntactic tree, and three levels of tree expansions are extracted as structured syntactic features.

To capture syntactic structure, we get the syntactic tree for each sentence using the Stanford parser, and extract the sub-tree of the upper three levels, which represents the fundamental composition of this sentence. In addition, all the paths from the root to each punctuation node in a sentence are extracted. From the paths, the depth of each comma node is counted, and the common parent node of every adjacent clause is also extracted. For example, the upper three level sub-tree of the syntactic tree of (S1) is shown in Figure 5. In addition, the sub-tree in the dotted line forms the structure of the punctuations in the (S1).

Polarity: A Comparison relation implies its two arguments are contrasting, and some contrasts are presented with different polarities in the two arguments. For example, sample (S5) is a case of Comparison.

(S5) 儘管天然環境如此優越，人為的不幸還是叫高棉子民不得好過，遍嚐戰亂的痛楚。
 (“Despite such favorable natural environment, man-made disasters still make the Khmer people unfortunate to suffer from the pain of war.”)

The first clause in (S5) is positive (“favorable natural environment”), while the last two clauses are negative (“unfortunate to suffer from the pain of war”). Besides the connectives 儘管 “despite” and 還是 “still”, the opposing polarity values between the first and the last two clauses is also a strong clue to the existence of a Comparison

relation. In addition, the same polarity of the last two clauses is also a hint that no Comparison relation occurs between them.

To capture polarity information, we estimate the polarity of each clause and detect the negations from the clause. The polarity score is a real number estimated by a sentiment dictionary-based algorithm. For each clause, the polarity score, and the existence of negation are taken as features.

4 Experiments and Discussion

4.1 Experimental Results

All the models in the experiments are evaluated by 5-fold cross-validation. The metrics are accuracies and macro-averaged F-scores. The t-test is used for significance testing.

We firstly examine our model for the task of two-way classification. In this task, binary classifiers are trained to predict the existence of Contingency and Comparison relations in a given sentence. For meaningful comparison, a majority classifier is used as a baseline model, which always predicts the majority class. In the dataset, 72.6% of the sentences involve neither Contingency nor Comparison. Thus, the major class is “Nil”, and the accuracy and the F-score of the baseline model is 72.6% and 42.06%, respectively.

The experimental results for the two-way classification task are shown in Table 3. In the table, the symbol † denotes the lowest accuracy which has a significant improvement over the baseline at $p=0.05$ for the two models. The symbol ‡ denotes the adding of a single feature yields a significant improvement for the model at $p=0.005$.

The performance of the decision tree and the SVM are similar in terms of accuracy and F-score. Overall, the decision tree model achieves better accuracies. In the two-way classification task, the decision tree model with only the Word feature achieves an accuracy of 76.75%, which is significantly better than the baseline at $p=0.05$. For both the decision tree and the SVM, Connective is the most useful feature: performance is significantly improved with the addition of Connective.

Besides the binary classification task, we extend our model to tackle the task of finer classification. In the second task, four-way classifiers are trained

Features	Decision Tree		SVM	
	Accuracy	F-Score	Accuracy	F-Score
Word	†76.75%	58.94%	72.36%	56.54%
+POS	77.15%	61.72%	72.28%	60.53%
+Length	77.15%	61.72%	72.60%	61.09%
+Connective	‡81.63%	71.11%	‡78.05%	69.17%
+Dependency	81.14%	70.79%	77.80%	68.79%
+Structure	81.30%	70.78%	†77.48%	69.08%
+Polarity	81.30%	70.78%	77.64%	69.09%

Table 3: Performance of the two-way classification.

Features	Decision Tree		SVM	
	Accuracy	F-Score	Accuracy	F-Score
Word	†76.50%	34.72%	73.58%	31.54%
+POS	76.99%	36.77%	72.52%	34.44%
+Length	76.99%	36.77%	72.36%	34.54%
+Connective	79.84%	44.08%	‡77.89%	45.26%
+Dependency	79.92%	44.47%	†77.07%	44.42%
+Structure	79.92%	44.47%	77.15%	44.69%
+Polarity	79.92%	44.47%	77.40%	44.80%

Table 4: Performance of the four-way classification.

Features	Decision Tree		SVM	
	Accuracy	F-Score	Accuracy	F-Score
Word	73.66%	3.00%	70.00%	3.62%
+POS	73.66%	3.00%	69.84%	4.29%
+Length	73.66%	3.00%	70.00%	5.08%
+Connective	74.80%	4.90%	74.39%	7.66%
+Dependency	74.72%	4.61%	72.60%	5.60%
+Structure	74.72%	4.61%	73.01%	5.49%
+Polarity	74.72%	4.61%	72.76%	5.23%

Table 5: Performance of the 49-way classification.

Task	Explicit		Implicit	
	Accuracy	F-score	Accuracy	F-score
2-way	77.97%	69.26%	88.98%	50.64%
4-way	76.06%	42.54%	88.98%	31.39%
49-way	71.33%	4.88%	89.41%	1.92%

Table 6: Performances for explicit cases and implicit cases.

to predict a given sentence with four classes: existence of Contingency relations only, existence of Comparison relations only, existence of *Both* relations, and Nil. The experimental results of the four-way classification task are shown in Table 4. Consistent with the results of the two-way classification task, the addition of Connective to the SVM yields a significant improvement at $p=0.005$. The performance between the decision tree and the SVM is still similar, but the SVM achieves a slightly better F-score of 45.26% in comparison with the best F-score of 44.47% achieved by the decision tree.

We further extend our model to predict the full relation structure of a given sentence as shown in Figure 1 and Figure 4. This is a 49-way classification task because there are 49 types of the full relation structures in the dataset. Not only as many as 49-ways, 72.6% of instances belong to the Nil relation, which yields an unbalanced classification problem. The experimental results are shown in Table 5. In the most challenging case, the SVM achieves a better F-score of 7.66% in comparison with the F-score of 4.90% achieved by the decision tree. Connective is still the most helpful feature. Comparing the F-scores of the SVM in the three tasks with the F-scores of the decision tree, it shows that the SVM performs better for predicting finer classes.

4.2 Explicit versus Implicit

We compare the performances between the explicit instances and the implicit instances for the three tasks with the decision tree model trained on all features. The results are shown in Table 6.

The higher accuracies and the lower F-scores of the implicit cases are due to the fact that the classifier tends to predict the sentences as Nil when no connective is found, and most implicit samples are Nil. For example, the relation of Contingency in implicit sample (S6) should be inferred from the meaning of 帶給 “brought”.

(S6) 得天獨厚的地理環境，的確帶給這個百年港埠無窮的財富。（“The unique geographical environment, it really brought the infinite wealth to this hundred-year port.”）

In addition, some informal/spoken phrases are useful clues for predicting the relations, but they are not present in our connective dictionary. For example, the phrase 的話 “if” implies a Contingency relation in (S7). This issue can be addressed by using a larger connective dictionary that contains informal and spoken phrases.

(S7) 想要以自助旅行的方式進行的話，那麼隨團旅遊呢？（“If you want to backpacking, how about an organized tour?”）

We regard an instance as explicit if there is at least one connective in the sentence. However, many explicit instances are still not easy to label

even with the connectives. As a result, predicting explicit samples is much more challenging than the task of recognizing explicit discourse relations in English. One reason is the ambiguous usage of connectives as shown in (S2). The following sentence depicts another issue. The word 但是 “however” in (S8) is a connective used as a marker of an inter-sentential relation. That is, the entire sentence is one of the arguments of an inter-sentential Comparison relation, but it does not contain any intra-sentential relation inside the sentence itself.

(S8) 但是，操一口流利中文的傅吾康則公開批評這種看法。（“However, Fu Wu Kang, who speaks fluent Chinese, openly criticizes this opinion.”）

The fact that connectives possess multiple senses is one of the important reasons for their misclassification. This issue can be addressed by employing contextual information such as the neighboring sentences.

4.3 Number of Clauses

We compare the performance among the 2-clause instances, the 3-clause instances, and the 4-clause instances for the three tasks with the decision tree model trained on all the features. The accuracies (A) and F-scores (F) are reported in Table 7.

Comparing the two-way classification and the four-way classification tasks, the performance of the longer instances decreases a little in relation labeling. Although sentence complexity increases with length, a longer sentence provides more information at the same time. In the 49-way classification, the model should predict the sentence structure and the relation tags from the 49 candidate classes. The performances are greatly decreased because the feasible classes are substantially increased along with the number of clauses.

4.4 Contingency versus Comparison

The confusion matrix of the decision tree model trained on all features for the four-way classification is shown in Table 8. Each row represents the samples in an actual class, while each column of the matrix represents the samples in a predicted class. The precision (P), recall (R),

Task	2-Clause		3-Clause		4-Clause	
	A (%)	F (%)	A (%)	F (%)	A (%)	F (%)
2-way	81.80	66.39	78.52	70.32	79.41	69.32
4-way	79.84	49.98	75.62	42.64	80.88	46.73
49-way	80.23	29.62	70.02	9.56	69.85	2.25

Table 7: Performances of clauses of different lengths.

Actual Class	Predicted Class				Performance		
	Cont.	Comp.	Both	Nil	P (%)	R (%)	F (%)
Cont.	61	3	0	131	81.33	31.28	45.19
Comp.	3	40	0	83	74.07	31.75	44.44
Both	2	4	0	5	0	0	0
Nil	9	7	0	882	80.11	98.22	88.24

Table 8: Confusion matrix of the best model in the 4-way classification.

Feature instance	Category	Usages
The first token in the third clause is the word 但 “but; however”	Word	100%
The first token in the second clause is the word 但 “but; however”	Word	99%
The first token in the third clause is a single connective of Contingency	Connective	98%
The first token in the first clause is the word 由於 “because; due to”	Word	96%
There is at least one word 以免 “in order to avoid” in the entire sentence	Word	95%
The first token in the second clause is the word 而 “moreover; while; but”	Word	94%
The first token in the third clause is a single connective of Comparison	Connective	93%
The second clause contains a single connective of Contingency	Connective	92%
The first token in the second clause is a single connective of Contingency	Connective	91%
The first clause contains a single connective of Contingency	Connective	90%

Table 9: Instances of the top ten useful features for the decision tree model

and F-score (F) for each class are provided on the right side of the table. The class *Both* is too small to train the model, thus our model does not correctly predict the samples in the *Both* class. The confusion matrix shows that the confusions between the classes Contingency and Comparison are very rare. The major issue is to distinguish Contingency and Comparison from the largest class, Nil. The lower recall of the Contingency and Comparison relations also show that our model tends to predict the instances as the largest class.

4.5 Features

The top ten useful feature instances reported by the decision tree model in the 49-way classification are shown in Table 9. Word and Connective provide useful information for the classification. Moreover,

seven of the ten feature instances are about the word or the connective category of the first token in each clause. This result shows that it is crucial to employ the information of the first token in each clause as distinguished features. Certain words, for example, 但 “but; however”, 由於 “because; due to”, and 而 “moreover; while; but” are especially useful for deciding the relations. For this reason, labeling these words carefully is necessary. All the synonyms for each of these words should be clustered and assigned the same category. In addition, a dedicated extractor should be involved in accurately fetching these words from the sentence in order to reduce tokenization errors introduced by the Chinese word segmenter.

The advanced features such as Dependency, Structure, and Polarity are not helpful as expected. One possible reason is that the training data is still not enough to model the complex features. In such a case, the surface features are even more useful.

Sample (S1) shows an interesting case of the use of polarity information. The first clause of (S1) is positive (嘗試要讓長期來作為大台北後花園的陽明山區更回歸自然 “tried to make the Yangmingshan area a more natural state as the long-term garden of Taipei”), the second clause of (S1) is also positive (但隨著週休二日、經濟環境改善 “the two-day weekend and the improved economic conditions.”), while the last clause of (S1) is negative (遊客帶來停車、垃圾等間接影響卻更為嚴重 “the issues of tourists parking, garbage, and other indirect effects”). The polarity of the last clause is opposite to those of the second clause, but they do not form a Comparison relation. Instead, a Contingency relation occurs between the last two clauses. Likewise, the polarities of the first and second clauses are both positive, but a Comparison relation occurs after the first clause. In fact, we realize that this is a complex case after performing an in-depth analysis. Because the last clause plays the role of effect in the Contingency relation, the negative polarity of the last clause makes the last two clauses form a negative polarity. For this reason, a Comparison relation occurs between the first argument with positive polarity and the second argument (i.e., the last two clauses) with negative polarity without a doubt. The polarity diagram of sample (S1) is shown in Figure 6.

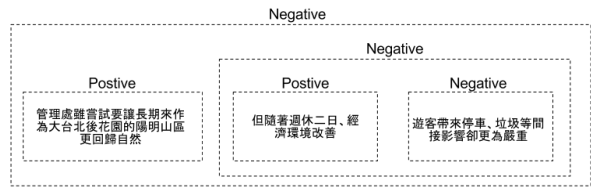


Figure 6: Polarity diagram of (S1).

Overall, the interaction among structure, relation, and polarity is complicated. The surface polarity information we extract by using the sentiment dictionary-based algorithm does not capture such complexity well. A dedicated structure-sensitive polarity tagger will be utilized in future work.

5 Conclusion and Future Work

In this paper, we addressed the problem of intra-sentential Contingency and Comparison relation detection in Chinese. This is a challenging task because Chinese sentences tend to be very long and therefore contain more clauses. To tackle this problem, we constructed a moderate-sized corpus and proposed a learning-based approach that achieves accuracies of 81.63%, 79.92%, and 74.80% and F-scores of 71.11%, 45.26%, and 7.66% in the two-way, the four-way, and the 49-way classification tasks, respectively.

From the experiments, we found that performance could be significantly improved by adding the Connective feature. The next step is to enlarge the connective dictionary automatically by a text mining approach, in particular with those informal connectives, in order to boost performance. The advanced features such as Dependency, Structure, and Polarity are not as helpful as expected due to the small size of the corpus. In future work, we plan to construct a large Chinese discourse Treebank based on the methodology proposed in Section 2 and release the corpus to the public.

Naturally, the intra-sentential relations are important cues for discourse relation detection at the inter-sentential level. How to integrate cues from these two levels will be investigated. Besides, relation labeling and structure prediction are tackled at the same time with the same learning algorithm in this study. We will explore different methods to tackle the two problems separately to reduce the complexity.

References

- Chu-Ren Huang, Feng-Yi Chen, Keh-Jiann Chen, Zhao-ming Gao, and Kuang-Yu Chen. 2000. Sinica Treebank: Design Criteria, Annotation Guidelines, and On-line Interface. In *Proceedings of 2nd Chinese Language Processing Workshop (Held in conjunction with the 38th Annual Meeting of the Association for Computational Linguistics, ACL-2000)*, pages 29-37.
- Hen-Hsen Huang and Hsin-Hsi Chen. 2011. Chinese Discourse Relation Recognition. In *Proceedings of 5th International Joint Conference on Natural Language Processing (IJCNLP 2011)*, pages 1442-1446.
- Aravind Joshi and Bonnie L. Webber. 2004. The Penn Discourse Treebank. In *Proceedings of the Language and Resources and Evaluation Conference*, Lisbon.
- Ziheng Lin, Min-Yen Kan, and Hwee Tou Ng. 2009. Recognizing Implicit Discourse Relations in the Penn Discourse Treebank. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing (EMNLP 2009)*, Singapore.
- Ziheng Lin, Hwee Tou Ng, and Min-Yen Kan. 2011. Automatically Evaluating Text Coherence Using Discourse Relations. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics (ACL 2011)*, pages 997-1006.
- Emily Pitler and Ani Nenkova. 2009. Using Syntax to Disambiguate Explicit Discourse Connectives in Text. In *Proceedings of the ACL-IJCNLP 2009 Conference Short Papers*, pages 13-1, Singapore.
- Emily Pitler, Annie Louis, and Ani Nenkova. 2009. Automatic Sense Prediction for Implicit Discourse Relations in Text. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP (ACL-IJCNLP 2009)*, Singapore.
- Rashmi Prasad, Nikhil Dinesh, Alan Lee, Eleni Miltsakaki, Livio Robaldo, Aravind Joshi, and Bonnie Webber. 2008. The Penn Discourse Treebank 2.0. In *Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC)*.
- WenTing Wang, Jian Su, and Chew Lim Tan. 2010. Kernel Based Discourse Relation Recognition with Temporal Ordering Information. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics (ACL 2010)*, Uppsala, Sweden, July.
- Nianwen Xue. 2005. Annotating Discourse Connectives in the Chinese Treebank. In *Proceedings of the Workshop on Frontiers in Corpus Annotation II: Pie in the Sky*, pages 84-91.