A Model for Processing Temporal References in Chinese

Wenjie Li, Kam-Fai Wong
Department of Systems Engineering
and Engineering Management
The Chinese University of Hong Kong
Shatin, N.T., Hong Kong
{wjli, kfwong}@se.cuhk.edu.hk

Chunfa Yuan
Department of Computer Science
and Technology
Tsinghua University,
Beijing, 100084, P.R. China
ycf@s1000e.cs.tsinghua.edu.cn

Abstract

Conventional information systems cannot cater for temporal information effectively. For this reason, it is useful to capture and maintain the temporal knowledge (especially the relative knowledge) associated to each action in an information system. In this paper, we propose a model to mine and organize temporal relations embedded in Chinese sentences. Three kinds of event expressions are accounted for, i.e. single event, multiple events and declared event(s). Experiments are conducted to evaluate the mining algorithm using a set of news reports and the results are significant. Error analysis has also been performed opening up new doors for future research.

1 Introduction

Information Extraction (IE) is an upcoming challenging research area to cope with the increasing volume of unwieldy distributed information resources, such as information over WWW. Among them, temporal information is regarded as an equally, if not more, important piece of information in domains where the task of extracting and tracking information over time occurs frequently, such as planning, scheduling and question-answering. It may be as simple as an explicit or direct expression in a written language, such as "the company closed down in May, 1997"; or it may be left implicit, to be recovered by readers from the surrounding texts. For example, one may know the fact that "the company has closed down before the earthquake", yet without knowing the exact time of the bankruptcy. Relative temporal knowledge such as this where the precise time is unavailable is typically determined by human. An information system which does not account for this properly is thus rather restrictive.

It is hard to separate temporal information (in particular refers to temporal relations in this paper) discovery from natural language processing. In English, tenses and aspects reflected by different verb forms are important elements in a sentence for expressing temporal reference (Steedman, 97) and for transforming situations into temporal logic operators (Bruce, 72). The pioneer work of Reichenbach (Reichenbach, 47) on tenses forms the basis of many subsequent research efforts in temporal natural language processing, e.g. the work of Prior in tense logic (Prior, 67), and of Hwang et al in tense tree (Hwang 92) and temporal adverbial analysis (Hwang 94), etc. Reichenbach argued that the tense system provided predication over three underlying times, namely S (speech time), R (reference time), and E (event time). Later, a multiple temporal references model was introduced by Bruce (Bruce, 72). He defined the set $(S_1, S_2, ..., S_n)$, which is an element of a tense. S_1 corresponds to the time of speech. Each S_i $(i = 2, ..., n-1)$ is a time of reference, and S_n, the time of an event. To facilitate logic manipulation, Bruce proposed seven first order logic relations based on time intervals and a method to map nine English tenses into temporal first order logic expressions[1]. His work laid down the foundation of temporal logic in natural language. These relations were then gradually expanded to nine in (Allen, 81) and further to thirteen in (Allen, 83)[2].

In contrast, Chinese verbs appear in only one

[1] The seven relations are symbolized as $R(A, B)$ for relation R and time intervals A and B, where R includes *before*, *after*, *during*, *contains*, *same-time*, *overlaps* or *overlapped-by*.

[2] *meet*, *met-by*, *starts*, *started-by*, *finishes* and *finished-by* are added into temporal relations.

form. The lack of regular morphological tense markers renders Chinese temporal expressions complicated. For quite a long time, linguists argued whether tenses existed in Chinese; and if they did how are they expressed. We believe that Chinese do have tenses. But they are determined with the assistance of temporal adverbs and aspect auxiliary words. For example, 在 ... 呢 (*being*), 已经 ... 了 (*was/been*) and 要 ... (*will be*) express an ongoing action, a situation started or finished in the past, and a situation which will occur in the future, respectively. Therefore, the conventional theory to determine temporal information based on verb affixations is inapplicable. Over the past years, there has been considerable progress in the areas of information extraction and temporal logic in English (Antony, 87; Bruce, 72; Kaufmann, 97). Nevertheless, only a few researchers have investigated these areas in Chinese.

The objective of our research is to design and develop a temporal information extraction system. For practical and cultural reason, the application target is on-line financial news in Chinese. The final system, referred to as $TICS$ (Temporal Information-extraction from Chinese Sources), will accept a series of Chinese financial texts as input, analyze each sentence one by one to extract the desirable temporal information, represent each piece of information in a concept frame, link all frames together in chronological order based on inter- or intra-event relations, and finally apply this linked knowledge to fulfill users' queries.

In this paper, we introduce a fundamental model of $TICS$, which is designed to mine and organize temporal relations embedded in Chinese sentences. Three kinds of event expressions are accounted for, i.e. single event, multiple events and declared event(s). This work involved four major parts, (1) built temporal model; (2) constructed rules sets; (3) developed the algorithm; and (4) set up the experiments and performed the evaluation.

2 A Model for Temporal Relation Discovery

2.1 Temporal Concept Frame

In IE, it is impossible as well as impractical to extract all the information from an incoming document. For this reason, all IE systems are geared for specific application domains. The domain is determined by a pre-specified concept dictionary. Then a certain concept is triggered by several lexical items and activated in the specific linguistic contexts. Each concept definition contains a set of slots for the extracted information. In addition, it contains a set of enabling conditions which are constraints that must be satisfied in order for the concept to be activated. Due to its versatility, a frame structure is generally used to represent concepts (as shown in Figure 1).

Slots in a temporal concept frame are divided into two types: activity-related and time-related. Activity-related slots provide the descriptions of objects and actions concerning the concept. For example, *company predator*, *company target* and *purchase value* are the attributes of the concept (收购, *TAKEOVER*). Meanwhile, time-related slots provide information related to when a concept begins or finishes, how long does it last and how does it relate to another concept, etc.

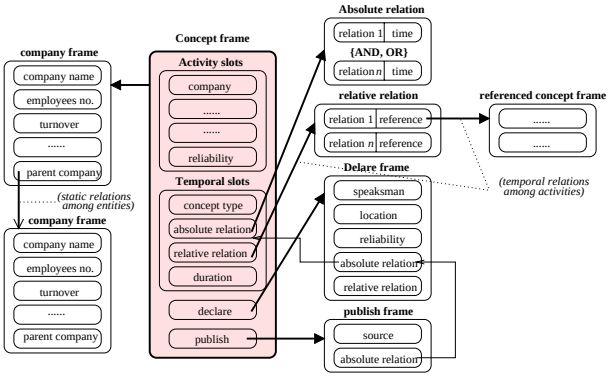


Figure 1: Temporal concept frame construction

2.2 Temporal Relations

The system is designed with two sets of temporal relations, namely *absolute* and *relative* relations. The role of absolute temporal relations is to position situation occurrences on a time axis. These relations depict the beginning and/or ending time bounds of an occurrence or its relevance to reference times, see $TR(T)$ in Section 2.3. Absolute relations are organized by *Time_Line* in the system, see Figure 2.

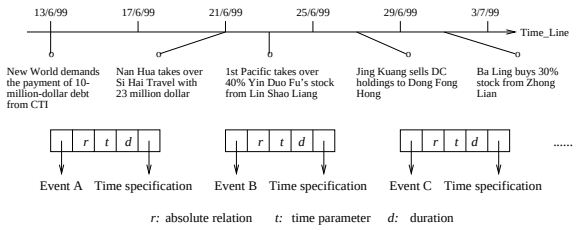


Figure 2: The *Time_Line* organization for absolute relations in $TICS$

In many cases, the time when an event takes place may not be known. But its relevance to

another occurrence time is given. Relative temporal knowledge such as this is manifested by relative relations. Allen has proposed thirteen relations. The same is adopted in our system, see $TR(E_i, E_j)$ in Section 2.3. The relative relations are derived either directly from a sentence describing two situations, or indirectly from the absolute relations of two individual situations. They are organized by *Relational_Chain*s, as shown in Figure 3.

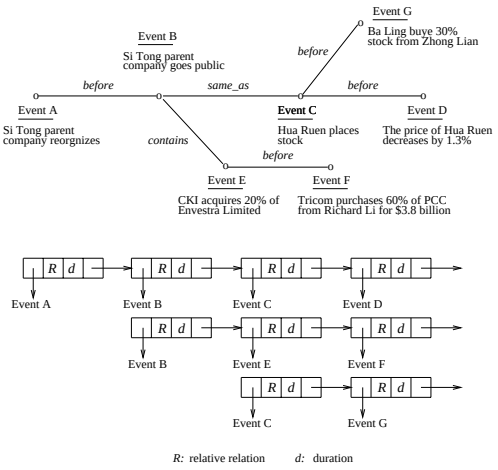


Figure 3: The *Relational_Chain* organization for relative relations in $TICS$

2.3 Temporal Model

This section describes our temporal model for discovering relations from Chinese sentences. Suppose TR indicates a temporal relation, E indicates an event and T indicates time. The absolute and relative relations are symbolized as: $OCCUR(E_i, TR(T))$[3] and $TR(E_i, E_j)$, respectively. The sets of TR are:

$TR(T) = \{ON,\ BEGIN,\ END,\ PAST,\ FUTUER,\ ONGOING,\ CONTINUED\}$

$TR(E_i, E_j) = \{BEFORE,\ AFTER,\ MEETS,\ METBY,\ OVERLAPS,\ OVERLAPPED,\ DURING,\ CONTAINS,\ STAREDBY,\ STARTS,\ FINISHES,\ FINISHEDBY,\ SAME\_AS\}$

For an absolute relation of a single event, T is an indispensable parameter, which includes event time t_e, reference time t_r[4] and speech time t_s:

$T = \{t_e,\ t_r,\ t_s\}$

Some Chinese words can function as the temporal indicators. These include time word (TW), time position word (F), temporal adverb (ADV), auxiliary word (AUX), preposition word (P), auxiliary verb (VA), trend verb (VC) and some special verbs (VV). They are all regarded as the elements of the temporal indicator TI:

$TI = \{TW,\ F,\ ADV,\ AUX,\ VA,\ VC,\ VV,\ P\}$

Each type of the indicators, e.g. TW, contains a set of words, such as $TW = twlist = \{tw_1, tw_2, ...tw_n\}$, with each word having an temporal attribute, indicated by ATT.

The core of the model is thus a rule set R which maps the combinational effects of all the indicators, TI, in a sentence to its corresponding temporal relation, TR,

$$R:\ \ TI \rightarrow \begin{cases} TR(T) \\ TR(E_i, E_j) \end{cases}$$

Regarding to the temporal relations, the language has three basic forms in representation:

- Single event statement: in which only one single event is stated.
- Multiple events statement: in which two or more events are stated.
- declaration statement: in which the event(s) are declared by a person or an organization.

3 Rule Construction

3.1 General Rules for Temporal References (GR)

Some general temporal characteristics, approximations and assumptions are examined to understand and uncover the hidden reference times, or to polish the identified solutions in order to make them more nature. For example, $PAST$(报道日期, *reporting date*) is probably better than ON(日前, *a few days ago*). Or, when no explicit reference time is given, the default value of T, i.e. the "reporting date" (thereafter referred to as RD), would be assumed. It must be noting that the rules given below are not for the immediate use of extracting TR. But they are necessary to design the $TICS$ program.

1. $TR(T)$ (single event) supports the following rules:

(1) Approximation:

$ON(t_e) \oplus (ATT(T) = \text{“}present\text{”}) \Rightarrow ON(\text{RD})$
$ON(t_e) \oplus (ATT(T) = \text{“}past\text{”}) \Rightarrow PAST(\text{RD})$
$ON(t_e) \oplus (ATT(T) = \text{“}future\text{”}) \Rightarrow FUTURE(\text{RD})$

[3] $OCCUR$ is a predicate for the happening of a single event. Under the situations where there are no ambiguity, E_i can be omitted. The $OCCUR(E_i, TR(T)$ is simplified as $TR(T)$.

[4] There maybe exist more than one reference time in a statement.

$PAST(t_r) \oplus (ATT(T) = \text{“}past\text{”}) \Rightarrow PAST(\text{RD})$
$FUTURE(t_r) \oplus (ATT(T) = \text{“}future\text{”}) \Rightarrow FUTURE(\text{RD})$
$TR(t_r) \oplus (ATT(T) = \text{“}present\text{”}) \Rightarrow TR(\text{RD})$
$TR(?) \Rightarrow TR(\text{RD})$

(2) Negation:

$\neg END(t_r) \Rightarrow CONTINUED(t_r)$
$\neg BEGIN(t_r) \Rightarrow FUTURE(t_r)$
$\neg PAST(t_r) \Rightarrow FUTURE(t_r)$
$\neg FUTURE(t_r) \Rightarrow FUTURE(t_r)$

2. $TR(E_i, E_j)$ (multiple events) supports the following rules:

(3) Symmetry:

$BEFORE(E_i, E_j) \equiv AFTER(E_j, E_i)$
$CONTAINS(E_i, E_j) \equiv DURING(E_j, E_i)$
$OVERLAPS(E_i, E_j) \equiv OVERLAPPED(E_j, E_i)$
$STARTS(E_i, E_j) \equiv STAREDBY(E_j, E_i)$
$FINISHES(E_i, E_j) \equiv FINISHEDBY(E_j, E_i)$
$SAME\_AS(E_i, E_j) \equiv SAME\_AS(E_j, E_i)$

3. $TR_s(T)$ and $TR_e(T)$ (declared event) supports the following rules:

(4) $TR_s(t_s) \oplus TR_e(t_r) \Rightarrow TR_e(t_s)$:

$ON(t_s) \oplus TR(?) \Rightarrow TR(t_s)$
$ON(t_s) \oplus ?(?) \Rightarrow ON(t_s)$

(5) $TR_s(t_r) \oplus TR_e(t_r) \Rightarrow TR_e(t_r)$:

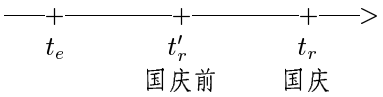
$PAST(?) \oplus PAST(?) \Rightarrow PAST(\text{RD})$
$?(?) \oplus TR(?) \Rightarrow TR(\text{RD})$
$?(?) \oplus ?(?) \Rightarrow PAST(\text{RD})$

3.2 Impact Coefficients of Temporal Indicators (R0)

The combined effect of all the temporal indicators in a sentence determines its temporal relation. However, in different situations, a certain indicator may have different effects. Compared (a) 今天早上他看了报纸 (*This morning, he read the newspaper*) and (b) 昨天我读了两本书 (*I read two books yesterday*), the two sentences are alike as they both embody an indicator 了, which implies $PAST$ in principle. The sole difference is that a definite time present in (b). (a) means the reading is finished at the speech time and the person must have known some news or information from his reading. Thus $TR = PAST(t_s)$ for (a). However, (b) means the event took place yesterday but not before yesterday. Consequently, for (b), $TR = ON$(昨天, *yesterday*) is appropriate. In the database, the impact coefficients are defined for the temporal indicators when T does or does not present in the sentence.

Remark:

It is likely for a sentence to contain two or more indicators. For example, adverbs 已经 and aspectual auxiliary word 了 together express a past tense and they both share the same reference time t_r. The same kind instances include 将...要 (*will*) and 正...着 (*being*) etc. Another example, such as 国庆 前 已经 (*before the National Day, one has already*), however includes two reference times. Here, 国庆 (*National Day*) is t_r, location word 前 (*before*) indicates $PAST$ between t_r and t'_r (i.e. $t'_r < t_r$), and adverb 已经 indicates the same relativity but between t_e and t'_r (i.e. $t_e < t'_r$).



The current proposed algorithm is unable to mine the implicit reference time (i.e. t'_r). But this does not affect the work at all. It does not matter even we cannot discover a relation like $PAST(t'_r)$. As we know $t'_r < t_r$ and $t_e < t'_r$, we can deduce a conclusion such as $t_e < t_r$ by rule $PAST(t'_r) \oplus PAST(t_r) \Rightarrow PAST(t_r)$ (for t_e). Thus, for this example, a relation of $PAST$(国庆, *National Day*) is enough to provide sufficient information to the users. To cater for these cases, we define a general rule: if all the indicators in a sentence indicate the same relation, then it is identified as TR (hereafter this rule together with impact coefficients is referred as R0).

3.3 Rules for Resolving Conflicts (R1)

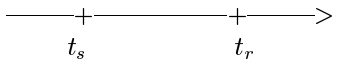
In many situations, the indicators in a sentence may introduce more than one relation. For example, adverbs 已经 (*have already*) and 在 (*being*) indicate $PAST$ and $ONGING$, respectively. But they could be collocated to represent some event which began in the past and continued to the reference time. For example 他 已经 在 看报 (*He has been reading newspaper*). Such a problem is regarded as conflict. In the following, five cases are illustrated with examples. To resolve this conflict, a set of rules are being defined in Table 1 (R1).

Case I: t_e to t_r and t_r to t_s (t_r is unknown)

深信证券市场 (adv仍) (va会) (v是) 香港经济发展的主要动力。(*They believed that the stock market will still be the major motivation for the HK economy development.*)

$TR = CONTINUED$(报导日期), $T = t_s$

t_e
———

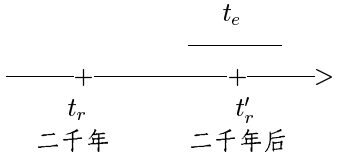


(仍) → $CONTINUED$ (t_e is continued in t_r)
(会) → $FUTURE$ (t_r is future for t_s)
$FUTURE(t_s) \oplus CONTINUED(t_r)$
$\Rightarrow CONTINUED(t_s)$ [5]

A Special Situation of Case I: t_e to t_r and t'_r to t_r (t_r is given)

基本上 (p在) (t二千年) (f后) ，香港在国际金融界 (adv仍) (va会) (v占) 一席位，(*Basically, HK will continue be important in the world global financial sector after 2000.*)

$TR = CONTINUED$(二千年), $T = t_r$

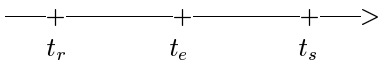


(仍) → $CONTINUED$ (t_e is continued in t'_r)
(后) ... (会) → $FUTURE$ (t'_r is future for t_r)
$FUTURE(t_r) \oplus CONTINUED(t'_r)$
$\Rightarrow CONTINUED(t_r)$ [6]

Case II: t_e to t_r and t_s (t_r is given)

(t八七年) (f以后) 市场 (adv已) (v打下) 良好基础，(*It has laid a good foundation for the market after 1987.*)

$TR = FUTURE$(八七年) and $PAST$(报道日期)

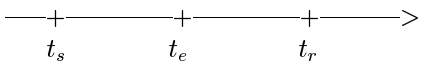


(以后) → $FUTURE$ (t_e is future for t_r)
(已) → $PAST$ (t_e is past for t_s)
$FUTURE(t_r) \oplus PAST(t_s) \Rightarrow FUTURE(t_r)$ [7]

Case III: Composition of Cases I and II
t_e to t_r and t_s, t_r to t_s (t_r is unknown)

但他 (vs估计) ，(p在) (d两个月) (f内) (adv已) (va可) (v落实) ，(*But he estimated that it would be secured within two months.*)

TR = $FUTURE$(报导日期) and PAST(报导日期+两个月), t_s =报导日期

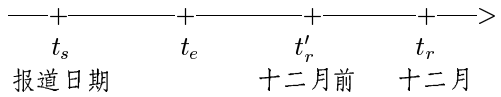


(在) (两个月) (内) → $FUTURE$ (t_r is future for t_s)
(已) → $PAST$ (t_e is past for t_r)
(可) → $FUTURE$ (t_e is future for t_s)
$FUTURE(t_s) \oplus PAST(t_r) \Rightarrow FUTURE(t_s)$

Case IV: t_e to t_s and t'_r, t'_r to t_r (t_r is given)

但他 (vs估计) ，(t十二月) (f前) (adv已) (va可) (v落实) ，(*But he estimated that it would be secured before December.*)

$TR = PAST$(十二月) and $FUTURE$(报道日期), t_s =报导日期

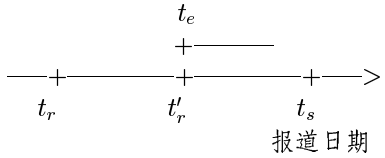


(前) → $FUTURE$ (t'_r is past for t_r)
(已) → $PAST$ (t_e is past for t'_r)
(可) → $FUTURE$ (t_e is future for t_s)
$PAST(t_r) \oplus PAST(t'_r) \Rightarrow PAST(t_r)$ (see R0)

Case V: Multiple implicit reference times

保险业，尤其是一般保险业务，(adv已经) (va要) (vv开始) (v受到) 亚洲金融风暴的影响。(*The insurance business, especially general insurance, has been affected by the Asian financial crisis.*)

$TR = FUTURE$(十二月)



(已经) → $PAST$ (t_r is past for t_s)
(要) → $FUTURE$ (t'_r is future for t_r)
(开始) → $BEGIN$ (t_e is begin for t'_r)
$PAST(t_s) \oplus FUTURE(t_r) \Rightarrow FUTURE(t_s)$
$FUTURE(t_s) \oplus BEGIN(t'_r) \Rightarrow FUTURE(t_s)$

$PAST(t_s) \oplus BEGIN(t_r) \Rightarrow CONTINUED(t_s)$
$PAST(t_s) \oplus END(t_r) \Rightarrow PAST(t_s)$
$PAST(t_s) \oplus FUTURE(t_r) \Rightarrow FUTURE(t_s)$
$PAST(t_s) \oplus ONGOING(t_r) \Rightarrow CONTINUED(t_s)$
$PAST(t_s) \oplus CONTINUED(t_r) \Rightarrow CONTINUED(t_s)$
$FUTURE(t_s) \oplus BEGIN(t_r) \Rightarrow FUTURE(t_s)$
$FUTURE(t_s) \oplus END(t_r) \Rightarrow CONTINUED(t_s)$
$FUTURE(t_s) \oplus PAST(t_r) \Rightarrow FUTURE(t_s)$
$FUTURE(t_s) \oplus ONGOING(t_r) \Rightarrow FUTURE(t_s)$
$FUTURE(t_s) \oplus CONTINUED(t_r) \Rightarrow CONTINUED(t_s)$

Table 1: Rule set R1 for single event statements

3.4 Rules for Discovering the Relevance of Two Events (R2 & R3)

To express two relevant events is straightforward. In general, one of them is treated as the reference event, say E_1, which is expressed by the subordinate clause. Another one, say E_2, i.e. the event concerned, is expression by the main clause. The

[5] See the last rule in Table 1

[6] (See the last rule in Table 3. To fit for this case, t_s is replaced with t_r and t_r is replaced with t'_r in the rule.

[7] See the eighth rule in Table 3. For those rules in Table 3, the parameters t_r and t_s are changeable.

position words (F), such as 前 (*before*) and 后 (*after*), and some special nouns, such as 时候 (*when*) and 期间 (*during*) between the two event expressions play an important role in determining their relevance in time. Also, it is noticed that the impact of $TR(E_2)$ cannot be ignored. Practically, $TR(E_2)$ relates E_2 to t_s or E_1. Especially for the latter, the influence of $TR(E_2)$ is indispensable. The rules for this are being defined in the rule set R2. In addition, some special templates are also necessary for relating two event, which are being defined in the rule set R3, when F is absent.

$(ATT(F) = \text{“}ON\text{”}) \oplus (TR(E_2) = \text{“}PAST\text{”})$
$\Rightarrow (TR(E_1, E_2) = \text{“}BEFORE\text{”})$
$(ATT(F) = \text{“}ON\text{”}) \oplus (TR(E_2) = \text{“}CONTINUED\text{”})$
$\Rightarrow (TR(E_1, E_2) = \text{“}CONTAINS\text{”})$
$(ATT(F) = \text{“}ON\text{”}) \oplus (TR(E_2) = \text{“}FUTURE\text{”})$
$\Rightarrow (TR(E_1, E_2) = \text{“}DURING''\text{)}$
$(ATT(F) = \text{“}ON\text{”}) \oplus (TR(E_2) = \text{“}ONGOING\text{”})$
$\Rightarrow (TR(E_1, E_2) = \text{“}CONTAINS\text{”})$
$(ATT(F) = \text{“}ON\text{”}) \oplus (TR(E_2) = \text{“}BEGIN\text{”})$
$\Rightarrow (TR(E_1, E_2) = \text{“}STARTEDBY\text{”})$
$(ATT(F) = \text{“}FUTURE\text{”}) \oplus (TR(E_2) = \text{“}PAST\text{”})$
$\Rightarrow (TR(E_1, E_2) = \text{“}AFTER\text{”})$
$(ATT(F) = \text{“}FUTURE\text{”}) \oplus (TR(E_2) = \text{“}FUTURE\text{”})$
$\Rightarrow (TR(E_1, E_2) = \text{“}AFTER\text{”})$
$(ATT(F) = \text{“}FUTURE\text{”}) \oplus (TR(E_2) = \text{“}ONGOING\text{”})$
$\Rightarrow (TR(E_1, E_2) = \text{“}AFTER\text{”})$
$(ATT(F) = \text{“}FUTURE\text{”}) \oplus (TR(E_2) = \text{“}BEGIN\text{”})$
$\Rightarrow (TR(E_1, E_2) = \text{“}AFTER\text{”})$
$(ATT(F) = \text{“}FUTURE\text{”}) \oplus (TR(E_2) = \text{“}CONTINUED\text{”})$
$\Rightarrow (TR(E_1, E_2) = \text{“}CONTAINS\text{”})$
$(ATT(F) = \text{“}PAST\text{”}) \oplus (TR(E_2) = \text{“}PAST\text{”})$
$\Rightarrow (TR(E_1, E_2) = \text{“}BEFORE\text{”})$
$(ATT(F) = \text{“}PAST\text{”}) \oplus (TR(E_2) = \text{“}FUTURE\text{”})$
$\Rightarrow (TR(E_1, E_2) = \text{“}BEFORE\text{”})$
$(ATT(F) = \text{“}PAST\text{”}) \oplus (TR(E_2) = \text{“}CONTINUED\text{”})$
$\Rightarrow (TR(E_1, E_2) = \text{“}BEFORE\text{”})$

Table 2: Rule set R2 for two event statements

Templates	Relations
V1 + 了+ V2	*AFTER*
一+ V1 + 就+ V2	*AFTER*
V1 + (的)同时 + V2	*SAME_AS*
V1 + 着+ V2	*SAME_AS*
(一)边+ V1 + (一)边+ V2	*SAME_AS*

Table 3: Templates in rule set R3

4 Algorithm Description

BEGIN:
input temporal statements;
(1) for a single event statement:
IF t_e is found in a temporal statement, let $T = t_e$;
ELSE let $T = t_s$=“报道日期” (reporting date, i.e. the default value);
ENDIF;
DETERMINE(TR):
IF $\bigcap_{ATT(TI_i)\neq\phi} ATT(TI_i) = \psi \neq \phi$, return $TR = \psi$;
ELSEIF $\bigcup ATT(TI_i) = \phi$,
IF $T = t_e$, return TR =“ON”;
ELSE return TR =“$PAST$” (default value);
ENDIF;
ELSE check rule set R1;
IF TR is found, return TR;
ELSE return $TR = \phi$;
ENDIF;
ENDIF;
go to END.
(2) for a declaration statement:
IF t_e is found for v, let $T = t_e$;
ELSE let $T = t_r$;
ENDIF;
IF t_s is found for s, let $t_r = t_s$;
ELSE let $t_r = t_s$ =“报道日期” (reporting date, i.e. the default value);
ENDIF;
do DETERMINE(TR) (for the declared event).
(3) for a multiple event statement or a declared multiple event statement:
IF f is found in a temporal statement, find $ATT(F)$ and $TR^{(E2)}$, then check rule set R2;
IF TR is found, return TR;
ELSE return $TR = \phi$;
ENDIF;
ELSE check R3;
IF TR is found, return TR;
ELSE return TR=“$BEFORE$” (default value);
ENDIF;
ENDIF;
IF one of the events contains time definition (i.e. t), do (1);
ELSE go to END;
ENDIF.
END.

5 Experiment Setup and Error Analysis

943K bytes of test data are collected from one month of financial pages of 《大公报》 (*Ta Kung Bao*). In total, 7924 temporal relations are discovered from the data. The distribution of temporal relations in test data is shown in Table 4. Considering the ultimate objective of this research is to find out the temporal relations embedded in sentences, the focus of the evaluation is therefore to figure out the number of the temporal relations of single event (i.e. $TR(E)$) and of multiple events (i.e. $TR(E_i, E_j)$), which are correctly marked by the program. Table 5 shows the results. Table 6 gives the order of TR classified by the program.

After analyzing the outputs, it was discovered that most errors were due to:

(1) t as a noun modifier;

Since the proposed method does not integrate the mechanism of parsing, the association between a modifier and its corresponding modifyee is not clear. In view of the task engaged, a time expression (indicated by t) could either modify a verb as an adverb, or modify a noun as a restricted modifier. Only the adverb t, determines the temporal reference of the event described by

the verb. Thus, the mistake is unavoidable when a noun modifier t appears in the text.

	Straight	
	Single event	Multiple events
Number	5235	603
Percentage	70.47%	8.12%

	Declared	
	Single event	Multiple events
Number	1507	84
Percentage	20.29%	1.13%

Table 4: Temporal expressions in the test data

TR	No.	Corr. Mark	Accu.
$TR(E)$	6742	6249	92.69%
$TR(E_i, E_j)$	687	643	93.60%
Overall	7429	6892	92.77%

Table 5: Experimental results of temporal relation discovery

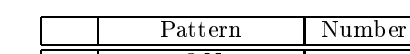
	Pattern	Number	Percentage
1	ON	2087	28.09%
2	$FUTURE$	1728	23.26%
3	$PAST$	1441	19.40%
4	$CONTINUED$	975	13.12%
5	$AFTER$	387	5.21%
6	$ONGOING$	299	4.02%
7	$BEGIN$	139	1.87%
8	$DURING$	128	1.73%
9	$BEFORE$	69	0.93%
10	$BEGIN\&END$	66	0.89%
11	$SAME\_AS$	59	0.79%
12	$CONTAINS$	41	0.55%
13	END	7	0.09%
14	$STARTEDBY$	3	0.04%

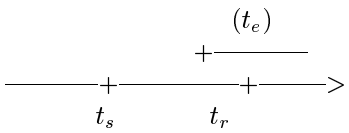
Table 6: TR classified by the program in decending order

(2) Ambiguous rules

All the rules are defined on the basis of indicators' attributes. The majority attributes is taken to be the final inferences. However, some special words may lead to exceptional results. These special words are possible sources of errors. Following is the example of a typical ambiguous rule.

$FUTURE(t_s) \oplus CONTINUED(t_r)$
$\Rightarrow CONTINUED(t_s)$
$\Rightarrow (FUTURE(t_s))$





(a) 集团 (将) (adv继续) 专注 (v发展) 电脑显示器及有关产品，扩阔产品系列。(*The group will continue concentrating on the development of computer monitors, and the related in order to widen the product.*

$TR = CONTINUED$(报导日期)

(b) (vs估计) 欧洲元在长线的走势 (va会) 在经济因素支持下 (adv逐步) (v向好)，(*It is estimated that supported by economic factors, in the long run, Euro will gradually become better,*)

$TR = CONTINUED$(报导日期)

correct: $TR = FUTURE$(报道日期)

reason: The word 逐步 has the essence of CONTINUED, but it is independent to any reference time.

(3) Noisy Annotation

Some errors are resulted from noisy annotations. For example,

- noun or verb?
  切实 (v拉动) (aux了) 国民经济的 (v增长) (?) (*did push the growth of national economy*)
- vs or v?
  (t本周) 半岛豪庭新盘，(va会) (vs公布) (?) (v订价)，(*the price of the new island houses will be announced this week*)
- vv or v?
  (t下周一) 两所会就合并建议 (vv进行) (?) (v表决)，(*the two departments will vote for the suggestion of the mergence on next Monday*)

(4) Coreference

The program searches for a solution within a sentence bounded by a full stop. As the connections between two sentences are ignored, it is incapable to resolve the coreference problem. As such, the following two sentences are all identified as $TR = PAST$, which is acceptable for the first and correct for the second. Nevertheless, since 同时 links the current event to the event described in the last sentence (indicated by ?), a solution of $SAME\_AS$(?, 接受) would be more accurate. Similarly, $BEFORE$(此?,发出) is more proper in the second sentence with 此 refering to the event stated before. The problem of coreference will be studied in our future work.

(a) 另方面，同时 (v接受) 内购登记的粉岭叠茵庭，(*On the other side, the Fenling Dieyinting ac-*

cepts internal registration ...)

(b) 此 (f前) 中国信息产业部 (adv曾) (v发出) 通知，(*Before it, China Information Industry Department sent out a notice.*)

6 Conclusions and Future Work

The issues of mapping linguistic patterns to temporal relations are addressed in the paper. These mapping is preconditioned on the temporal indicators and achieved on a set of pre-defined rules. The mapping mechanism was validated. On 7429 sentences describing temporal relevance, we achieved 92.77% accuracy in average.

These relations will be useful to for information extraction, information retrieval and question-answering application. Once the corresponding frames have been instantiated and their slots filled after temporal natural language processing. The related temporal concepts will be linked together according to their chronological orders, to be applied as the knowledge to fulfill users' queries.

We find two interest questions as our future work.

(1) Reference Time Shift

In the current work, we considered sentences as independent units. The evaluation is also performed on this basis. However, some sentences in a article may be temporally related. They may share the same reference time which is indicated in a preceding sentence or the event time in one sentence servers as a reference point for the next. How to identify whether a reference time is continued from the preceding sentence or is the same as a omitted speech time, and how the reference times shift should be a good topic in the future work.

(2) The focus of Negation

The negation form of a verb may have two focuses. One emphasizes the event, which is expected to become the fact but, still has not yet happened. It implies that the event will take place in the future. Another emphasizes a status where the event didn't happen throughout a specified duration. Is it possible to find out the focus of the negation?

References

Allen J.F., "An Interval-based Represent Action of Temporal Knowledge", In *Proceedings of 7th International Joint Conference On Artificial Intelligent*, 1981, pp221-226.

Allen J.F. et al., "Planning Using a Temporal World Mode", In *Proceedings of 8th International Joint Conference On Artificial Intelligent*, 1983, pp741-747.

Androutsopoulos I., Ritchie G. and Thanisch P., "Time, Tense and Aspect in Natural Language Database Interfaces", cmp-lg/9803002, 22 Mar, 1998.

Antony G. (1987). *Temporal Logics and Their Applications*, Department of Computer Science, University of Exeter, Academic Press.

Bruce B.C., "A Model for Temporal References and its Application in Question-answering Program", *Artificial Intelligence*, 1972, 3, pp1-25.

Glasgow B., Mandell A., Binney D. and Fisher F., "An Information Extraction Approach to Analysis of Free Form Text in Life Insurance application", In *Proceedings of the 9th Conference on Innovative Applications of Artificial Intelligence*, Menlo Park, 1997, pp992-999.

Hwang C.H., Schubert L.K., "Tense Trees as the Fine Structure of Discourse", In *Proceedings of 30th Annual Meeting of the Association for Computational Linguistics*, 1992, pp232-240.

Hwang C.H., Schubert L.K., "Interpreting Tense, Aspect and Time Adverbials: A Compositional, Unified Approach", In *Proceedings of 1st International Conference in Temporal Logic*, Bonn, Germany, July 1994, pp238-264.

Kaufmann M. (1995). Advanced Research Project Agency, In *Proceedings of the Sixth Message Understanding Conference* (MUC-6), USA.

Lehnert W., McCarthy J. etc. "Umass/hughes: Description of the Circus System Used for muc-5", In *Proceedings of the Fifth Message Understanding Conference* (MUC-5), San Francisco, Calif., 1994. pp 277-291.

Prior A.N., *Past, Present and Future*, Oxford, Clarendon Press, 1967.

Reichenbach H., *Elements of Symbolic Logic*, Berkeley CA, University of California Press, 1947.

Soderland S., Aronow D., Fisher D., Aseltine J. and Lehnert W., "Machine Learning of Text Analysis Rules for Clinical Records", Technical Report, TE39, Department of Computer Science, University of Massachusetts, 1995.

Steedman M., *Temporality*, In J. van Benthem and A. ter Meulen, (eds.), *Handbook of Logic and Language*, Elsevier North Holland, 1997, pp895-935.

Wong K.F., Li, W.J. Yuan C.F. and Zhu X.D. "Temporal Representation and Classification in Chinese", submit to *International Journal of Computer Processing of Oriental Languages*, 2000.

李临定, 《现代汉语动词》, 中国社会科学出版社, 1990. (Li Linding, *Mordern Chinese Verbs*, Chinese Social Science Press, 1990, in Chinese)