Back to the Future: A Logical Framework for Temporal Information Representation and Inferencing from Financial News[1]

Zi Huang [1] Kam-Fai Wong [2] Wenjie Li [3] Dawei Song [4] Peter Bruza [4]

[1] School of Information Technology and Electrical Engineering
University of Queensland, QLD 4072 Australia
huang@itee.uq.edu.au

[2] Department of Systems Engineering
The Chinese University of Hong Kong
Shatin, N.T., Hong Kong
kfwong@se.cuhk.edu.hk

[3] Department of Computing
The Hong Kong Polytechnic University
Hung Hom, Kowloon, Hong Kong
cswjli@comp.polyu.edu.hk

[4] Distributed Systems Technology Centre
University of Queensland, QLD 4072 Australia
{dsong, bruza}@dstc.edu.au

Abstract

Temporal information carries information about changes and time of the changes. Consider a company investing in another company. The former may choose to inject the money gradually with the amount and frequency depending on the performance of the latter. This shows that an event can be completed in multiple steps and at any given time before completion, it is partially completed. Thus, the status of an event at any time could be described by some degree of completion. One can make inference based on such temporal information to predict what event(s) would likely happen next. The prediction could be made not only based on the completed or partially completed events in the past, but also based on the correlation between the events, which have taken place (i.e. *executed events*), and the ones planned (i.e. *planned events*). This process of making inference based on the executed and planned temporal events is described lively as "*Back to the Future*" and can be considered as part of the formally-called *temporal information inference*. Existing temporal information processing frameworks (e.g., temporal database, temporal information extraction, and temporal logic), however, are ineffective for this purpose. This paper defines a novel logical framework for two-dimensional (i.e., executed and planned time lines) temporal information representing and inferencing. An operational model realising the logical framework in financial news data is also addressed.

1. INTRODUCTION

Temporal information is related to changes and times of the changes. In many domains where events occur frequently over time, temporal information is crucial. Inference based on temporal information is a powerful tool for many application domains, such as business. For example, if a company was in disarray and failed to achieve its planned targets, it could consider to declare bankruptcy or to seek for a buyer. For another example, consider a commercial buy-out activity. To reduce its risk, an investor would be keen to find out the track record of the company concerned, e.g. how well did the company perform in the past and how reliable it has been in keeping its promises? Furthermore, the investor may even want to plan in advance the next event(s) based on those, which have already been planned and executed. Both examples involve predication using some known facts. The core of the prediction process is, in fact, based on *temporal information inference*.

Temporal information processing has been investigated for years in three major areas: temporal database and information extraction and temporal logic.

Temporal database (Snodgrass and Ann 1985; Soo 1991) deals with applications which require history of past database states. A temporal database is a database that records time-varying information. For example, once a change is made to an employee's record, the transaction time can be attached in form of a time stamp to both old and new data of the employee. In addition to the two dimensions (row and column) in traditional database, e.g., relational database, temporal database can be considered simply as a cube with the third dimension, being

[1] The work reported in this paper has been funded in part by the Cooperative Research Centres Program through the Department of the Prime Minister and Cabinet of Australia.

represented as various time intervals. Unlike the traditional database, wherein new values overwrite the old ones once rows are updated, the temporal database supports storing and querying time-varying data (Snodgrass 1999). Constraints could be used to establish simple temporal relationships between events, e.g., one transaction must happen prior to another one (Chomicki 1995). Temporal database can be applied to keep track of history of data in a number of domains, e.g. stock market data in financial applications. However, temporal database lacks inference mechanism over temporal information.

Information extraction (IE) involves using natural language processing techniques to extract desirable information from documents, and artificial intelligence techniques to transform the extracted information to some forms of knowledge to facilitate reasoning. In particular, temporal IE aims to extract and maintain time-dependent information, such as temporal events and temporal relationships between them, etc. For example, an event about one company taking over another could be represented as a frame with slots company predator, company target, purchase value, announce date, publishing date, etc. (Li, Wong & Yuan 2002). Temporal information inference is not the focus of IE. However, it could be used to provide the input information required for our research.

Temporal logic is the closest to the objectives of this project. It defines a temporal structure, and based on which a temporal algebra system is defined to reason about events and actions. By far, interval temporal logic (Allen, 1981; Allen and Ferguson 1994) is most widely adopted. In Allen's work, the temporal structure is a simple linear model of time. A time period is the time associated with some event occurring or some property holding. A set of temporal relations, e.g. *before*, *during*, *overlap*, etc., is axiomatized by using a primitive relation: *meets*. Time can then be added into a predicate as an extra argument. Such predicate is referred to as a temporal predicate. Events are represented as event predicates. For example, the event of "Jack lifting the ball onto the table at time t1" would be represented as:

$$\exists e.LIFT(e) \wedge (agent(e) = jack34) \wedge (dest(e) = table5) \wedge (theme(e) = ball26) \wedge (time(e) = t1)$$

or its abbreviation:

$$\exists e.LIFT(jack34, table5, ball26, t1, e)$$

The knowledge about events and reasoning about events can be facilitated by defining a set of axioms using the temporal logic. The traditional logical implication operator (⊃) is used to define the consequences of event occurring. For example, an event of one block being stack on another is described by an event predicate *STACK*(*x*, *y*, *t*, *e*). The axiom

$$\forall x, y, t, e.STACK(x, y, t, e) \supset \exists t'.t : t' \wedge On(x, y, t')$$

represents that the first block will be on the second at the end of the event.

Allen's model assumes a simple linear representation of temporal information rendering it ineffective for inference over events involving interaction between multiple time frames, e.g. executed (i.e. past), planned (i.e. to be executed), and predicted events (i.e. would be executed). Furthermore, an event in classical temporal logic is binary with respect to the time interval. It can either be completed or un-completed at the end of the time period. However, this is too restrictive. In some applications, one may need to keep track of the status of an event at any given time point. For example, an investment of $5,000 may be completed in several time points, e.g. $1,000 at point 1, $3,000 at point 2 and $5,000 at point 3. This phenomenon is quite common in business applications. The degree of completion of an event with reference to a time point is therefore rudimentary.

The primary objective of this paper is to design a novel logic framework for temporal information inference. The new temporal information inference framework would address the following key issues, which have been undermined in the contemporary temporal information processing models:

1. Multi-dimensional representation of temporal events, i.e. executed, planned and predicted events.
2. Association between events and time, i.e., partially completed event with a degree of completion.
3. Various inference mechanisms.

2. TWO-DIMENSIONAL REPRESENTATION AND INFERENCING MODEL

This representation model involves the modelling of events in multiple time dimensions. Unlike many temporal logic models wherein an event is binary with respect to a single time interval, we propose to adopt a finer granularity of time, i.e. time point. We are interested in the status of an event at a given time point, e.g. how far is an event been completed. A time interval consists of a number of time points. It can be modelled using both points of starting and completion. Degree of completion is associated to each event at a specific time point. This is used to cater for events which are partially complete, e.g. at a certain time point, a company invested $1,000, which was only a portion of the total investment of $5,000. In order to formally describe our methodology, we define the following notations:

- *C* is a set of entities (e.g. companies to be investigated), $c_i \in C$
- *T* is a set of time points (e.g. Sept. 11, 2002), $t_i \in T$

- A is a set of *activities*, $a_i \in A$. An activity refers to the action (e.g. buy-out) leading to the occurrence of an event (e.g., company X buys company Y out).
- E is a set of *events*, $e_i \in E$. An event is carried out by an entity; it is composed of one or more activity/activities. To simplify, we assume that an event involves one activity only.
- $e_i = (c_i, t_{i-executed}, t_{i-planned}, degree(a_i))$ represents an event, wherein the entity c_i (e.g. a firm) has completed or due to complete the activity a_i (e.g. buy-out) in some *degree* at time point $t_{i-executed}$ and its planed completion time is $t_{i-executed}$.
- $E(c_i, t_i)$ is a subset of events carried out by an entity c_i at time t_i. Similarly,
- $E(c_i, t_i, t_j)$ is a subset of events carried out by an entity c_i during the time interval $[t_i, t_j]$. When t_i and t_j are missing, $E(c_i)$ refers to a of events carried out by an entity c_i over a time line.
- *progressing*(e, t_i, t_j) represents how much progress have been made to event e at time t_j compared to time t_i.

Note that the real world structure of temporal information is far complicated than the above formal definitions. The focus of this paper is simply helping an investor to watch the activities of a specific company and then make judgement and prediction.

- ***Single time line representation***

Figure 1 depicts a set of events $E(c_i, t_1)$ associated to an entity c_i at time t_1 over a single time line. Each tile in $E(c_i, t)$ is an event.

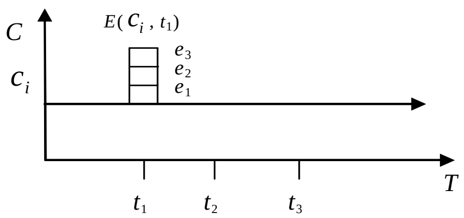

Figure 1. Single time line representation model

- ***Two-dimensional (2D) representation – executed and planned events***

Executed events are those actually completed in any degree at a specific time point, and planned events are those, which are scheduled for completion at some time. Figure 2 presents an example: event e_1 is planned to complete at time point t_1. However, it was actually completed at time t_4. On the other hand, event e_2 has been completed in some degree at time t_2, which was ahead of the planned completion time t_3.

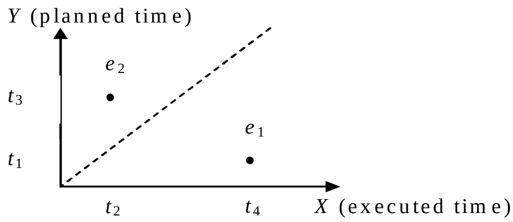

Figure 2. Two-dimensional temporal representation model

- ***Back to the future – Making inference on 2D space***

The following graph can be employed to play with the planned and executed events in order to measure the *reliability* of an entity (e.g. a firm), e.g. based on whether the firm has executed all planned events in time and if not how much has it slipped.

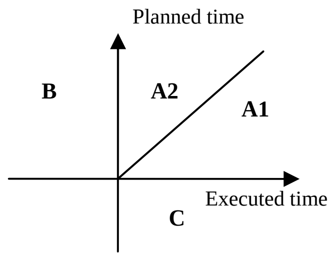

Figure 3. Partition of a 2D space

The events conducted by an entity can be traced and judged by examining which area(s) in Figure 3 one or more of these events fall into:

A (A1 or A2): events planned and executed
B: events planned but not executed
C: events not planned but executed

Notice that the diagonal graph (i.e. $x=y$) in area A represents the special case where an event is executed as planned (with respect to time). By the same token, $x>y$ (i.e. area A1) implies 'over-executed' (or 'under-planned') and $x<y$ (i.e. area A2) 'under-executed' (or 'over-planned').

Example 1 – Tracing a take-over event

Suppose a company A plans to complete an event of taking over another company B, which costs totally \$30M, on March 1. Note that there is no execution time in this frame yet (assume it has been March 20 today). The users who are watching this event may then have to trace backwards on the chain of related events: To successfully complete this take-ever action, company A needs to raise \$20M and borrow \$10M from the bank before the planning date of the take-over activity. Then raising-money event is completed in two steps. Event "Raising money (1)" is the first step: Company A plans to raise 15 million dollars from its sub-companies before Jan. 20 but it achieves this goal earlier on Jan. 1. In the mean time, another event “Raising money (2)” – company A planing to raise some money from company C before Jan 20 (step 2), is actually completed on time. The last step is to deal with the bank. However, the event “Loan” is not going as smoothly as the others – its execution time (March 20) is almost a month after its planning time (Feb 24). The above implicit inference can be shown by the 2D representation model, which gives user a more direct and intuitive view:

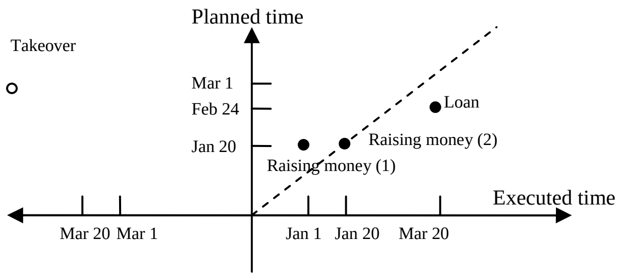

Figure 4. 2D space for example 1

By observing Figure 4, we can easily find that the planning time (Mar. 1) of the “take-over” event has passed – this implies it may have been postponed to a new date after Mar. 20 due to a delay of getting a loan from the bank, and the new date has not been announced yet. All these information will be important for the user to support his decision making process.

3. OPERATIONALIZING THE REPRESENTATION MODEL VIA INFORMATION EXTRACTION

Information Extraction (IE) is an upcoming challenging research area to cope with the increasingly unwieldy volume of distributed information resources, such as World Wide Web (WWW). It is a mapping of short, unstructured natural language texts into predefined, structured representations, or templates, which when filled, represent an extract of key information from original texts. Essentially, IE task involves natural language processing (NLP) technology for extracting the desirable information from the relevant articles.

In this section, we propose a temporal information extraction framework which operationalizes the two-dimensional representation model. We have successfully implemented the proposed framework, which leads to a system namely CHIEF (Chinese Information Extraction from Financial News). However, the NLP techniques involved in system development are not detailed in this paper. Our focus here is the general framework of the operational model, which is event-centred and catering for the occurrence time and other important properties of business events in the financial domains.

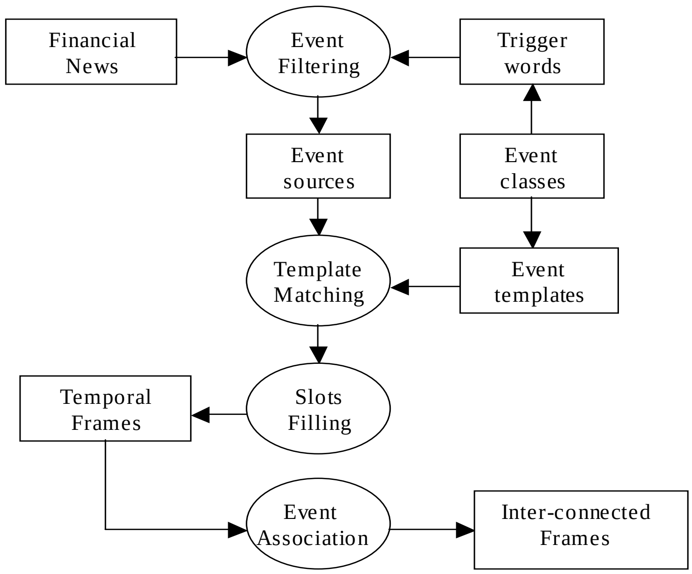

Figure 5. Architecture of the operational model

Figure 5 depicts the architecture of the framework, which consists of four functional modules: event filtering, template matching, slots filling and event association.

3.1 Financial Event Filtering

The system deals with large amount of financial news. It can produce a short-summary of the financial news articles according to specific criteria, rather than presenting the whole article, by eliminating the information that is considered not to be relevant. The availability of summaries of the original articles, rather than the full articles, leads to a marked reduction of the time needed by the operators in the market for analysing the information and, thus, can help reducing the data overload. A set of event classes is defined, such as "takeover", "loss", "bankruptcy", "stock offering", etc., and each of which represents a certain type of financial events. A set of trigger words for each event class is defined to help the system perform filtering. When a sentence contains one of these words, it is selected from the news and added into the summary, which is referred to as "event source" in this paper.

3.2 Templates Matching

Performing template matching on the financial events is the efficient way to parse the construction of the events (Costantino et al. 1996). Each event class has its different formats of expression in natural language. A number of templates are created to describe some commonly used formats. In order to parse the sentence more accurately, one may need to define the templates as completely as possible. Consider the event class "bankruptcy. The most common style is "Company A is bankrupt". On the other hand, the activity "bankruptcy" has many other ways to express. For example,

- "Company A is bankrupt."
- "Bank B announce company A's bankruptcy."
- "Company A applies for bankruptcy".

Although the trigger words for these sentences are the same, difference between the corresponding events can not be ignored. For example the event of a company applying for bankruptcy is much different from the event that a company have been bankrupt.

Through matching the templates and sentences in the event source, the details of different events can be found. These details then fill into frames consisting of a number of information slots (Li et al. 2002, Costantino et al. 1996). The structure of frames is presented next.

3.3 Filling in Slots

Each temporal event is associated to a specific frame which has many slots filled with the details of this event extracted from the event source by the Templates Matching module. All the information filled in the slots can be useful for users to analyse a financial event. The frames identified provide quantitative information that the financial operator takes into account in his decision-making process. The structure of frames is shown in Table 1:

The slots in a frame can be divided into two types: activity related and time-related. The former provides descriptions of the objects and actions for an event. These descriptions are actually the template used in the Template Matching module. For example, "active company", "passive company", and "purchase value" may be the descriptions of a "take-over" event. The time related slots, on the other hand, provides time related properties of the event, e.g., planning time, execution time, etc, which incorporate the two-dimension time line representation model defined in Section2. Note that the other types of time related slots, such as announcement time and publish time, are not used yet in the logical model. They may be taken into account in the future. Different frames can be linked to each other based on either activity or time related slots. This will be addressed in the next section.

Frame: event class	
Slots	Values
Activity-related	
Trigger word	<the trigger word of this event>
Event descriptions (Template used to extract the this event)	<the template used to extract the this event>
Relevant events based on the activity-related information	<hyperlinks to other related frames>
Time-related	
AT	<Announce time>
PlT	<Planning time>
ET	<Execute time>
PuT	<Publish time>
Relevant events based on time-related information	<hyper-links to other related frames>

Table 1. Structure of Frames

3.4 Event Association and Hyper Links

No event happening in this world is isolated from the others. They are always related one another by some factors, such as time, participating company, etc. As a consequence, an insolate frame is not much valuable. The connections among related frames can help us discover the corresponding event associations. These connections are referred to as hyper-links. In the context of financial domain, two types of hyperlinks are paid more attentions: event association by participating company and event temporal association.

An "Event-Company-Event" association links various events conducted by a specific company. On the other hand, the temporal association links events in chronological order. By combining these two types of hyperlinks, one can widely gather the information and make the complex chains among the events. For example, the events which happened to the same company can be linked over a time line. By analysing these events, the user, probably a potential investor, can easily trace the history of a target company. This actually constructs the Figure 1 (single time line model) in Section 2. More complex associations involving both events' planning and execution time can be used to construct the two-dimensional representation model in Figure 2.

Recall the illustrative example (tracing a take-over event) given in Section 2. Its frame structures and participating company based hyper-links are shown in Figure 6, which can then be mapped to the 2D model (figure 4).

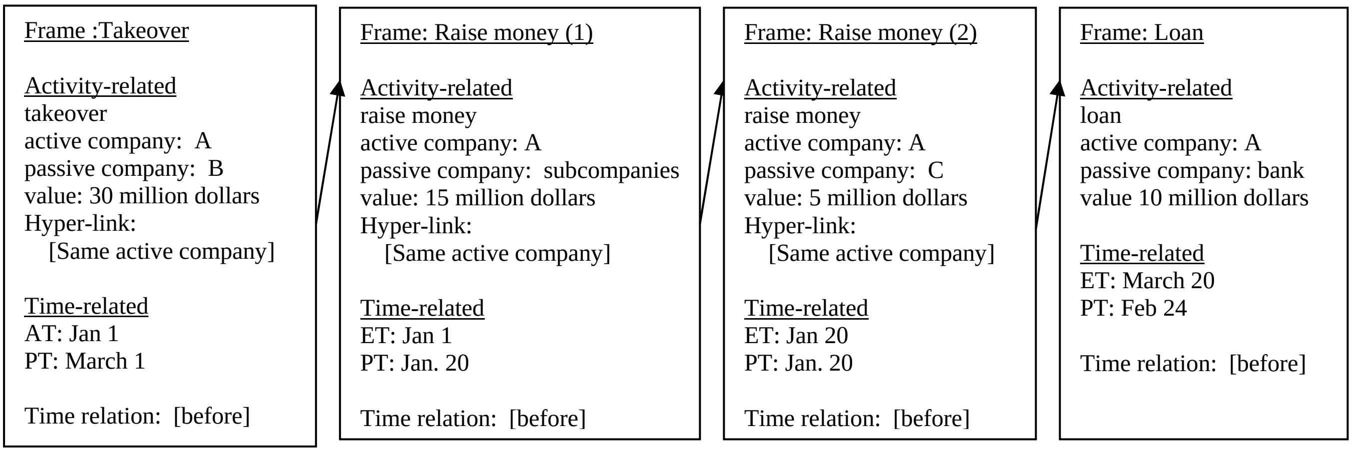

Figure 6. Frames and event associations of Example 1

4. CONCLUSIONS AND FUTURE WORK

We have proposed a novel logical framework for temporal information representing and inferencing. An operational model in financial information processing domain has been addressed. The notion of temporal information inference based on multi-dimensional and partially complete event representation is novel. The research will open up new directions in knowledge engineering (e.g. how to represent temporal events and how to manage them) and temporal logic (e.g. how to make use of the time of executed and planned events to predict future event(s)).

More mechanisms for reasoning about temporal events, i.e. what information could be inferred from an event or composition of events, will be investigated in the future. One possibility is to build up event inferencing on the basis of implicit dependencies between their underlying activities. For example, the activity of *being bankrupt* could be inferred from activity *in-debt*. The inference at event level would be more complicated, as an event involves additional specific parameters, such as entities, time, and degree of completeness. Moreover, an event on different dimensions of time line (e.g. executed and planned events) would play different roles in the inference process. The inference at event level derives what events are implied by an event or a group of events, e.g., $e_1, e_2, e_3 \mid - e_4$. The inferred events can then be used as predictions, i.e., what event will happen next to the company concerned.

References

Allen J.F. (1984) “Towards a General Theory of Action and Time”, *Artificial Intelligence*, 23, 1984, pp123-154.

Allen, J.F. and Ferguson, G. (1994). Actions and Events in Interval Temporal Logic. *Journal of Logic and Computation*, 1994.

Chomicki, J. (1995). Efficient Checking of Temporal Integrity Constraints Using Bounded History Encoding. *ACM Transactions on Database Systems (TODS)*, 20(2):149-186.

Costantino, M., Collingham, R.J., and Morgan, R.G. (1996). Information Extraction in the LOLITA System using Templates from Financial News Articles. In Proceedings of the 18th International Conference on Infoprmation Technology Interfaces (ITI’96), pp. 43-54.

Li, W., Wong. K.F., Yuan, C. (2001). Towards Automatic Temporal Information Extraction, J*ournal of American Society for Information Science (JASIS),* 52(9):748-762.

Snodgrass, R.T. (1999). *Developing Time-Oriented Database Applications in SQL*, Morgan Kaufmann Publishers, Inc., San Francisco, July, 1999.

Snodgrass, R.T. and Ahn, I (1985). Taxonomy of Time in Databases. I*n Proceedings of ACM SIGMOD International Conference on the Management of Data*, pp236-24.

Soo, M. (1991). Bibliography on Temporal Databases. *ACM SIGMOD Record*, 20(1):2-36.