A Graph Model of Events Focusing on Granularity and Relations Towards Organization of Collective Intelligence on History

Minoru Naito
Graduate School of Informatics
Kyoto University
Kyoto, Japan
naito@db.soc.i.kyoto-u.ac.jp

Yasuhito Asano
Graduate School of Informatics
Kyoto University
Kyoto, Japan
asano@i.kyoto-u.ac.jp

Masatoshi Yoshikawa
Graduate School of Informatics
Kyoto University
Kyoto, Japan
yoshikawa@i.kyoto-u.ac.jp

***Abstract*—We propose Event Graph Model to organize collective intelligence on history and acquire useful knowledge of events. Our Event Graph Model represents a complicated event of various granularities as a graph composed of nodes corresponding to smaller-grained events and edges corresponding to relations between the events. This model is expected to be useful for finding events whose structures are similar to each other. In other words, our model would be able to distinguish such events from ones similar by vocabulary but different substantially, although previously proposed keyword-based event analyses could not.**

Keywords-Collective intelligence; Model of Events.

I. INTRODUCTION

The investigation of history is important for the future of humankind. As it has been said that history repeats itself, we might be able to infer a solution of a current problem by drawing a lesson from knowledge of past events similar or related to the problem. However, historical knowledge is too massive for common people to find and understand such events.

Automatic analysis and organization of historical knowledge could be helpful for people to find and understand events. Several methods have been proposed to analyze events utilizing news articles [1], [2], [3]. Although these methods are useful for understanding events roughly, they are insufficient for our motivation that people draw a lesson from historical knowledge. For example, the two articles illustrated in Figure 1 represent events similar by vocabulary but different radically; in the left article a famine first occurred and consequently a reform was carried out; in the right article a famine wasted the reform. The methods above might not distinguish them because they regard an article as an event and use few keywords to represent an event. Therefore, a model of events should be able to analyze the granularity of events more flexibly, and organize various kinds of relations between events.

In this paper, we propose an event graph model, abbreviated to EGM, that represents a large event, named *composite event*, in a given article by a graph whose node corresponds to a small event, named *minimal event*, and

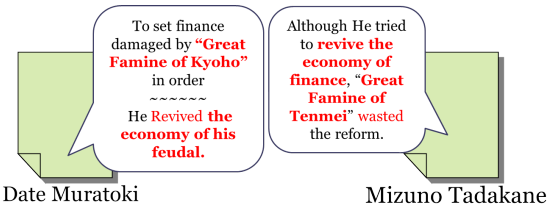

Figure 1. Events that are similar by vocabulary but different.

edge represents the relation between two minimal events. In order to establish EGM, we first define two types of a minimal event which can not be divided on a sentence in the given article: an *event noun phrase* and an *event predicate-argument structure*. An event noun phrase is a noun phrase which can represent an event by itself. If we divide the noun phrase into several elements, then no elements can represent an event. Therefore, it is adequate that we regard an event noun phrase as a minimal event. A predicate-argument structure in a sentence, abbreviated to PAS, is called an event predicate-argument structure, abbreviated to EPAS, if its predicate represents an action related to an event. For example, sentence "Date Muratoki revived the economy of his feudal domain" contains an EPAS because its predicate "revived" represents an action in the economic revival; on the other hand, the predicate of sentence "Tokugawa Ieyasu had been told as a descendant of Nitta Yoshishige" is not regarded as an action of a historical event. If a sentence has two predicates representing actions of events, then we regard the sentence contains two EPAS. We then introduce the following eight kinds of a relation between minimal events: deployment, illustration, parataxis, causality, paradox, progress, means, and situation. In contrast to models dealing with the causality relation between events [1], our eight kinds of relations are expected to find similar events more accurately. We also present an idea how to determine which events have a relation by utilizing the centering theory proposed by Grosz et al. [5], [6]. Figure 2 illustrates an example of composite events in

 ISBN: 978-1-61208-248-6

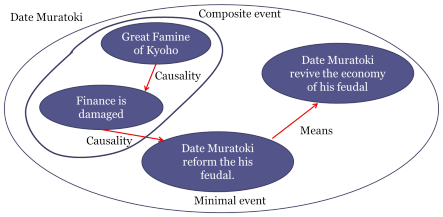

Figure 2. Examples of minimal and composite events.

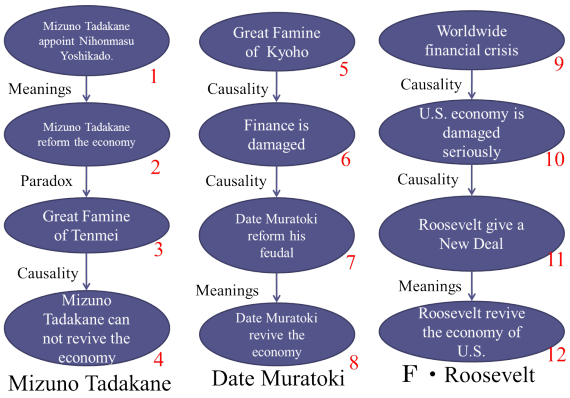

Figure 3. Composite events for Mizuno, Date, and Roosevelt.

our EGM constructed from several sentences in Wikipedia article “Date Muratoki.” Each solid circle represents a minimal event; an arrow represents the relation between two minimal events. Each open circle represents a composite event; note that an arbitrary connected subgraph can be regarded as a composite event despite only two composite events are shown in the figure.

Once we succeed in establishing EGM, then we might be able to find similar events accurately enough for our motivation. For example, Figure 3 portrays three composite events extracted from three Wikipedia articles “Mizuno Tadakane,” “Date Muratoki,” and “Franklin Roosevelt.” The composite events for Mizuno and Date contain similar minimal events: nodes 2 and 7 represent reforms, and nodes 3 and 5 represent famines. However, the orders of the minimal events are different: the reform is before the famine for Mizuno, while the famine is before the reform for Date. Therefore, these events should be regarded as different. On the other hand, the composite event for Date is similar to that for Roosevelt in both of the content of minimal events and their relations. We plan to propose a method for finding events having similar structures by applying techniques based on graph theory including graph isomorphism and topological minor.

As work in progress, we have implemented the extraction of minimal events and the determination which events have a relation. We also have a preliminary experiment on the extraction of minimal events. We use Wikipedia articles about people as a dataset because they contain a plenty knowledge of historical events and the following characteristics useful for acquiring the information of events: the explanation of events are usually described in chronological order; specific expression can be normalized easily using wiki links; attributes of an article are readily identifiable by category; an article has fewer subjective opinions and has many descriptions of objective facts; entity information is included in the article because of the principle of “one entity – one article.” For example, we utilize wiki links to identify event noun phrase in the implementation of the extraction of minimal events.

II. Related Work

Ishii et al. [1] proposed a method for constructing a network of events to illustrate causality relations between events.

Ikeda et al. [2] proposed a method for representation of an event by 5W1H(who, when, where, what, why, how) and predicate information that are extracted from a news article.

Chin-man et al. [3] proposed a method for acquisition of detecting knowledge from collection of past news. They show that analysis of references to the past in news articles allows us to gain a lot of insight into a historical event.

Murakami et al. [4] proposed a Statement Map showing the relation of statements described on the Web by creating a database of knowledge between events using data and a thesaurus of verbs. This system helps users with selection and aggregation of information.

In this way, many earlier studies have been conducted to extract relations between statements to help users. However, few studies have proposed an event-based comparison with consideration of the flow across multiple sentences and statements.

III. Event Graph Model

In this section, we explain the detail of EGM. Because events takes various granularity, we define two types of an event: a minimal event and a composite event.

A. Minimal Event

Minimal events are nodes of a graph in our EGM. A minimal event on a sentence consists of its elements representing an event; if the elements are divided into two or more parts, then any of the parts can not represent an event. We define two types of an minimal event: an event noun phrase and EPAS.

1) Event Noun Phrase: An event noun phrase is a noun phrase which can represent an event by itself. We regard a noun phrase as event noun phrase if it has a wiki link to an article of a category related to events. In this work we use “incident,” “disaster” and “battle” categories. Because a noun phrase itself does not contain a plenty of information about the event, we employ a technique for compensating for such information. It is known that the first sentence of the Wikipedia article of an entity can be regarded as a summary

of the entity. Therefore, we extract an EPAS explained in Section III-A2 from the first sentence of the article linked from the event noun phrase.

2) Event Predicate-Argument Structure: A complex sentence that includes plural predicates is frequently appear on Japanese Wikipedia articles. Therefore, we should not regard a sentence as an event. Instead, we focus on a PAS containing a predicate representing an event in a sentence. One PAS consists of a predicate and terms that have a case relation with the predicate. Let p(c_1:a_2, c_2:a_2, ... c_ℓ:a_ℓ) be a PAS consists of predicate p and ℓ terms, each a_i having case relation c_i with p for $i = 1, 2, ..., \ell$.

We extend the PAS in order to represent a minimal event, because a predicate often corresponds to an action related to an event. However, not all the predicates correspond to such actions. Therefore, we propose a rule-based definition of an EPAS to use only predicates corresponding to such actions. The proposed rules are based on Japanese language. We use a verb or a "sahen noun" as predicates in an EPAS. Here, a sahen noun is a Japanese-specific noun that functions as a verb if it is followed by word "suru", which is the equivalent of "do" in English. We explain our four rules (a)-(d) below; because it is difficult to explain them completely in English, we only describe the summary of them. (a) We do not use a verb that represents hearsay and inference. For example, in sentence "X was said to be serious" of the article for entity X, "said" is not an action of events in which X took part in. (b) We do not use an auxiliary verb. For example, in sentence "He was made to win," "win" would be preferable for representing an action to "made." (c) We use a sahen noun if it is followed by a punctuation. In Japanese, such a representation is considered as an omission of word "suru" explained above. (d) We use a sahen noun if it is a part of an adverbial clause. For example, in adverbial clause "by the attack of enemy," noun "attack" is an action of the enemy.

Our rules, especially (c) and (d), reflect characteristics of Japanese Wikipedia articles. A single sentence in them frequently contains a larger amount of information than that contained in a sentence in usual documents. Therefore, it frequently has omissions of "suru" explained in (c) and adverbial clauses explained in (d). Although our rules use Japanese-specific sahen nouns, we might be able to propose equivalent rules for English Wikipedia articles by utilizing well-known characteristics of English grammars.

Once we determine whether we use predicate p in a given sentence, then we determine the other elements in the EPAS which contains p. We use only two kinds of a case relation, nominative and accusative, because these are necessary and sufficient for determining whether two minimal events have a relation corresponding to an edge in our EGM. We also decide to use keywords which are nouns complement of the predicate but are not nominative and accusative. The keywords help us to compare and determine the connection of minimal events. Therefore, we denote an EPAS containing predicate p by

p(nominative:a_n, accusative:a_a, keywords:k_1, k_2, ...) ,

where word a_n is the nominative of the predicate p, and word a_a is the accusative of p, and words k_1, k_2, ... are the keywords. The nominative and accusative words can be extracted from a given sentence by using morphological analysis tools such as MeCab [7] and CaboCha [8].At the same time, keywords can be detected. Table I depicts an example of EPASs corresponding to three minimal events illustrated in Figure 2.

Table I
EXAMPLES OF A PREDICATE STRUCTURE OF EVENTS.

Predicate	Nominative	Accusative	Keyword
get	finance	damage	-
reform	Date Muratoki	feudal	-
revive	Date Muratoki	economy	feudal

B. Composite Event

A composite event represents a large event in a given article by a graph whose node corresponds to a minimal event and edge represents the relation between two minimal events. As illustrated in Figure 2, a composite event can be represented by a connected subgraph. Every edge has a label representing a kind of the corresponding relation. A causal relation exists between minimal events "Great Famine of Kyoho" and "finances damaged" in the Figure.

1) Relation between Minimal Events: There are studies about discourse structure analysis used to examine the relation between a sentence and another sentence [6]. Various kinds of relations are defined in each study. However, these kinds of relations are not sufficient for represent the relations between minimal events which might have smaller granularity than a sentence, that is, which can be appear more than twice in a single sentence. Therefore, we extend ideas used in the discourse structure analysis in order to capture the relations between minimal events.

We decide to use eight kinds of relations which are classified into three categories: deployment, illustration and parataxis in structural relation category; causality, progress, and paradox in semantic relation category; means and situation in complementary relation category. Structural relations are those necessary to capture the structure of a document. Wikipedia articles about historical figures often describe actions of a person in chronological order. Therefore, these relations are important to compare the chronological structures of events. We consider that the causality, progress, and paradox of events are interpretation of history by human subjective. Therefore, semantic relations would be the most important kind of relations for our motivation. A complementary relation is a particular relation between minimal

 ISBN: 978-1-61208-248-6

events in the same sentence; one event gives complementary information about the other.

2) Determination of Unity between Minimal Events: We here summarize our idea for determining which minimal events have a relation by utilizing centering theory proposed by Grosz [5]. The centering theory can be used for extracting the relation between adjacent sentences. Umezawa et al. [6] proposed a system named DIA that performs discourse structure analysis extend the centering theory so that not only the adjacent sentences but also neighbor sentences are used. We apply this extended centering theory to extraction of the relation between minimal events.

The centering theory uses the nominative words and accusative words mainly in sentences to determine the relation between them. Therefore, we can apply it to determine the relation between two minimal events, each of them is represented by an EPAS which includes nominative and accusative words. The extended centering theory assigns a score representing how much the probability there is a relation between the pair. If the nominative words of the events in a pair are the same, then the highest score is assigned. On the other hand, if there is no common words in them, the score is lowest. In addition to the scoring used in the extended centering theory, we assign a high score to the pair of minimal events extracted in the same sentence because they would have a relation with a high probability. We omit the details for space.

IV. EVALUATION

A. Selection of predicate

In order to make sure the effects of our rules defined in Section 3 that select verbs and sahen nouns as predicates, we apply the rules to 5 Wikipedia articles about historical Japanese people including “Date Muratoki” and “Takeda Motoshige.” Table II shows the ratio of predicates excluded by our rules. The “Precision” column indicates the ratio of predicates which are not EPASs actually to the ones excluded by our rules. The “Recall” column indicates the ratio of ones which are not EPASs and are excluded by the rules correctly to all the verbs and sahen nouns. Most of the predicates excluded by our rules are actually not EPASs. Therefore, our rules effective in precision. On the other hand, the recall is not high sufficiently. We consider that rules should be added for proper exclusion.

Table III shows the number of relations extracted by the extended centering theory (base line) and our method. Our method can find more relation than the baseline.

B. Predicate Search

In order to make sure the benefits of our approach with event graph model, we compare three simple methods of search from Wikipedia articles of Japanese people.

Queries of search are composite events that is organized by two minimal events. We use four queries: (a) He became a priest and renamed; (b) He succeed to a house because of retirement of his father; (c) He is helped to the lord, and returned to secular life; (d) He killed his men and he lost trust. We use three simple methods for search of these queries: AND search of two predicates; search with the information of predicates and the edge of the predicates in our event graph model; search with the information of predicates, the edge and its label in our event graph model. For this experiment, we use 500 Wikipedia articles of Japanese feudal lords. We also use the Japanese Wordnet [9] for comparison of two predicates.

Table IV, V shows the precision and recall of these methods. Because there is not result on query (c) by use edge and label, there is a blank column in tables. In Table V, because query (d) has an only 1 correct article, the line of query (d) is omitted.

Table IV shows there are many noisy results in baseline method because two predicates are not always related in articles. On the other hand, our methods show high scores in precision. It means our method is effective in getting knowledge of composite events.

However, Table V shows the methods show low scores in recall. In most cases, it is caused by different written forms, illegal connection of minimal events and labels. In addition,

Table II
RESULTS OF APPLING OUR RULES OF PREDICATES.

Predicate	Excluded predicates	precision	recall	example
Verbs	3.7 %	87.5 %	36.8 %	said estimated
Sahen nouns	80.7 %	98.6 %	89 %	attack recovery

Table III
THE NUMBER OF RELATIONS BETWEEN MINIMAL EVENTS.

Article Name	Minimal events	Base line	Our method
Date Muratoki	46	40	43
Nasu Takasuke	56	27	32
Takeda Motoshige	68	46	60

Table IV
PRECISION OF PREDICATE SEARCH BY EACH METHODS.

Query Name	Baseline	Use edge	Use edge and label
a	66.7 %	100 %	100 %
b	20 %	95.2 %	100 %
c	37.5 %	100 %	-
d	2.4 %	100 %	100 %

Table V
RECALL OF PREDICATE SEARCH BY EACH METHODS.

Query Name	Use edge	Use edge and label
a	40 %	40 %
b	66.7 %	43.3 %
c	66.7 %	-

 ISBN: 978-1-61208-248-6

there is a case that some minimal events exist between two minimal events of query. To resolve these problems, the method of comparison of graph models is needed.

V. Conclusion and Future Work

We proposed an event graph model in order to organize collective knowledge of history and find useful knowledge from it. Focusing on the granularity and relation between events especially, we proposed a minimal event and a composite event which represents relations among minimal events. We actually found minimal events and determined whether each pair of minimal events has a relation using Wikipedia articles about people.

A candidate of future work is to propose a method for determining the kind of the relation of minimal events. We also plan to construct a method for comparing composite events by varying the granularity. We consider that graph theory, including graph isomorphism and topological minor, would be useful for our plan.

References

[1] Hiroshi Ishii, Qiang Ma, and Masatoshi Yoshikawa. Incremental Construction of Causal Network from News Articles, Journal of Information Processing, Vol. 20, No.1, pp. 207-215, 2012 [accessed December, 2012]

[2] Takahiro Ikeda, Akitoshi Okumura, and Kazunori Muraki. Information Classification and Navigation Based on 5W1H of the Target Information, Computational Linguistics and the Association for Computational Linguistics 1998, Montreal, pp. 571-577, 1998 [accessed December, 2012]

[3] Ching-man Au Yeung and Adam Jatowt. Studying how the past is remembered: towards computational history through large scale text mining, Conference on Information and Knowledge Management Proc. of the 20th, Glasgow, pp.1231-1240, 2011 [accessed December, 2012]

[4] Koji Murakami, Eric Nichols, Junta Mizuno, Yotaro Watanabe, Shouko Masuda, Hayato Goto, Megumi Ohki, Chitose Sao, Suguru Matsuyoshi, Kentaro Inui, and Yuji Matsumoto. Statement map: reducing web information credibility noise through opinion classification, Analytics for Noisy Unstructured Text Data 2010, Toronto, pp. 59-66, 2010 [accessed December, 2012]

[5] Barbara J Grosz, Scott Weinstein, and Aravind K Joshi. Centering: A Framework for Modeling the Local Coherence of Discourse, Association for Computational Linguistics, 21, Cambridge, Massachusetts, pp. 203-225, 1995 [accessed December, 2012]

[6] Toshiyuki Umezawa and Minoru Harada. Discourse Structure Analysis System DIA Based on Centering Theory and Object Knowledge, Journal of Natural Language Processing 18(1), pp. 31-56, 2011 [accessed August, 2011]

[7] MeCab: Yet Another Part-of-Speech and Morphological Analyzer, http://mecab.sourceforge.net/ [accessed December, 2011]

[8] CaboCha: Yet Another Japanese Dependency Structure Analyzer, http://code.google.com/p/cabocha/ [accessed October, 2012]

[9] Japanese Wordnet (v1.1) (c) NICT, 2009-2010, http://nopwww.nict.go.jp/wn-ja/index.en.html [accessed December, 2012]

 ISBN: 978-1-61208-248-6