Implicit Knowledge Completion Method Using Relevance Calculation of Distributed Word Representations

Sho Takishita[1,2], **Rafal Rzepka**[1,2], **Kenji Araki**[1]
[1]Faculty of Information Science and Technology, Hokkaido University, Japan
[2]RIKEN AIP, Japan
{sho.takishita, rzepka, araki}@ist.hokudai.ac.jp

Abstract

In this paper, we propose a method to automatically discover implicit knowledge for events. In recent approaches researchers calculate relevance between two events appearing in the same document from news articles using distributed representations for machine learning. Since such methods handle only explicitly written information, it is very difficult for a machine to, for example, automatically answer questions about context. We acquire temporal events and calculate their implicit relationship by relevance using distributed representation of words. Likewise, relevant places are also added by using similar algorithm to enrich text with implicit knowledge. Our proposed method utilizes raw text corpus, does not require prior knowledge, and is task-independent. Experiments were performed to show temporal and semantic naturalness of the proposed knowledge completion process.

1 Introduction

Experiences common to human beings allow them to exclude obvious information during the process of generating and understanding natural language. We unconsciously augment missing contextual pieces using background (commonsense) knowledge, and its scarcity remains one of the biggest obstacles for artificial intelligence. For example, when we hear "he took a shower" statement, usually occurring sub-events or states as "he took off clothes" and "he wants to wash his body", "he is breathing", "he is standing", etc. are obvious *without saying* to us, but not to machines lacking similarly rich experiences. Although many methods for acquiring commonsense knowledge have been proposed, they are limited to simple knowledge units (concepts, "events" in this paper) as *Having a breakfast*–HasSubevent–*Reading a newspaper* [Speer and Havasi, 2012] which only partially contribute to natural language understanding tasks and are not sufficient for machine learning techniques dealing with longer event chains impeding planning and predicting longer behavioral patterns. Although new powerful, big text data-based stochastic methods are being proposed [Wu *et al.*, 2018], it is rather difficult to sufficiently learn what is not explicitly written. This phenomenon of omitting obvious details leads also to difficulty in generating event sequences in natural order limiting state of the art techniques to short-range predictions as answering questions or choosing correct agents, reasons or consequences in tasks such as Winograd Schema Challenge [Levesque *et al.*, 2012]. Current computing power and increasing storage capacity allows enriching already existing sentences with implicit knowledge, even if a simple five word sentence becomes several times longer. This research addresses two aspects of implicit knowledge completion: a) automatic retrieval of information related to events and calculating their order; b) automatic retrieval of information about locatives – common places where usually an event occurs. We propose a set of simple methods for both generating longer event chains and adding implicit knowledge which in future should not only lead to enhancing machine learning in deeper understanding of natural languages but also allow easier explainability of text-based decisions and predictions which remains difficult for current statistical approaches. In this paper we present our methods for both combining event pairs into longer chains and the evaluation performed on event chains consisting of four events. The results show that our proposed method is able to combine knowledge retrieved from other sentences and complete implicit knowledge about places and events without a big loss (0.29) of generality.

2 Related Work

Our aim is to automatically generate textual representation of natural sequences of events similar to manually crafted Schankian scripts [Schank and Abelson, 1977], a goal which is yet to be reached due to the absence of implicit (tacit) knowledge in sentences describing acts. We think that eventually agreeable behavioral patterns will be collected simultaneously also from other sensory input as vision [Vedantam *et al.*, 2015], however the text remains the easiest mean to convey and manipulate world knowledge. Obtaining longer patterns enhances reasoning, predicting following events and discovering abnormalities, but existing methods utilize costly annotations [Yao and Huang, 2018] or massive data-based language models. Such models can generate highly natural follow-ups to an input sentence via heavy use of transformers and attention [Wu *et al.*, 2018; McCann *et al.*, 2018], but the naturalness of output carries the same problem of omitting facts obvious for humans. When applied to co-reference anal-

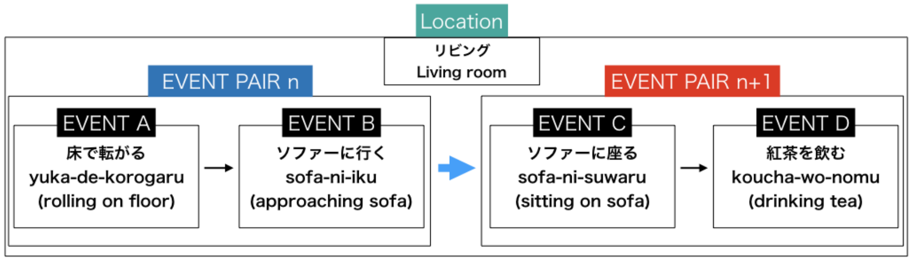

Figure 1: Example of events formed into pairs and chains with common location.

ysis and discourse analysis, longer and enriched event knowledge should allow to infer relations between actors, patients, themes and other semantic roles more precisely, but to the authors' best knowledge no method extends script generation beyond a pair of events. The problem of omitted semantic details is clearly visible in highly contextual languages as Japanese which is our choice for experiments. Similarly to other languages, research on unsupervised generation of event chains in Japanese language is limited to two elements, e.g. *plant seeds* $\rightarrow$ *water soil* (research examples are given in a separate subsection). These methods acquire knowledge from events appearing in the same sentence, but actions such as "crossing a street" are rarely accompanied by sub-events as "after the light went green" and "cars stopped and waited" in one sentence. Since there are not many cases that three or more events are included in one sentence, we decided to combine pairs of events from various sentences into longer event chains and extend zero-anaphora solutions for pronouns to add location information. Most closely related research examples on event extraction and completion are given below.

2.1 Event Extraction and Completion (English)

Several methods for event knowledge acquisition from texts have been proposed, however they usually target languages with fewer contextual omissions such as English. Probably the closest approach to ours is proposed by Chambers and Jurafsky [Chambers and Jurafsky, 2008] who use co-references to acquire events related to specific subjects in an event chain and calculate frequent event relations into a more general event chain. However, they use texts with temporal relationship annotations limiting their research to the set prepared beforehand. Granroth and Stephen [Granroth-Wilding and Clark, 2016] utilize data from [Chambers and Jurafsky, 2008] to apply machine learning for calculating time series relationship of two events. The evaluation is performed by predicting an event in a chain after masking it. Recently, deep learning methods have been used to detect events [Feng *et al.*, 2018; Tozzo *et al.*, 2018] or recognize temporal relationship between events [Lin *et al.*, 2018] but they do not address script-like event chains generation nor created implicit knowledge completion.

2.2 Event Extraction and Completion (Japanese)

The above-mentioned problem is explained by Huang and Kurohashi [Huang and Kurohashi, 2017] who also limit their approach to one pair of events in an argument identification task. Research on Japanese unsupervised event knowledge acquisition is based on co-occurrence frequency and uses temporal particles and conjunctions [Fujita *et al.*, 2011], or adds dependency relations and weights [Higashiyama *et al.*, 2017]. Abe et al. [Abe *et al.*, 2008] acquired single events from the Japanese Web corpus using bootstrapping. These studies rely heavily on the boundaries of one sentence as the retrieval condition, and simplicity of acquired events causes a problem of being too general to be fairly evaluated by evaluators whose imagination easily fills up contextual blanks.

3 Implicit Knowledge Completion Using Distributed Word Representations

In this session, we describe our algorithm for event chain construction and implicit knowledge completion. Its key modules include temporal event extraction, events relevance calculation, implicit event knowledge concatenation and implicit knowledge complementation.

3.1 Task Description

First, I will explain the task of this paper in knowledge construction. The goal of this time is to build a highly natural event chain to understand human behavior. Extracting time-series relationships for the web corpus and combining them, we construct an event chain. In addition to that, it is challenging to embed the place where the action is performed to event knowledge, and mechanically embed information that is not specified.

3.2 Temporal Event Pair Extraction

In the first step, temporal event pairs are extracted from a blog corpus [Ptaszynski *et al.*, 2012]. We assumed that the order of experiences described in blog texts naturally represents time series, and we retrieve basic event pairs ordered accordingly with verb phrase appearance. For example, in the case of sentence "They were reading articles and writing blogs.", the module acquires a pair "read article(s) $\rightarrow$ write blog(s)" (Japanese does not distinguish singular from plural).

3.3 Events Relevance Calculation

Events retrieved from a noisy raw corpus naturally include semantically irrelevant data. To tackle this problem we calculate relevance between the events distributed word represent and acquire semantically related words between events. The relevance calculation is shown in Equation (1).

$$similarity(W_{Am}) = \frac{\sum_{n=0}^{N-1} cosine(W_{Am}, W_{Bn})}{N} \quad (1)$$

For the set A of words included in event A and the word W_{Am}, the set B of words included in event B and the word W_{Bn} (W_{Am} is the m^{th} word of word set A and W_{Bn} is the n^{th} word of word set B, N is word count of word set B). The relevance is calculated by Equation (1) and a word is acquired if the relevance is higher than the experimentally set threshold R. The relevance value of a word not included in the distributed representation is set to 0.

3.4 Combining Event Pairs

In this module, event pairs are combined and event chains are constructed. By retrieving events from not only within the same documents but also from all remaining documents, we were able to acquire implicit event candidates which do not appear in a single sentence. When event B of event pair AtoB and event C of event pair CtoD are highly relevant, event B and event C are considered likely to occur in the same situation, and to appear in a common chains of acts. Therefore, we calculate the semantic relevance between each event pair and all remaining ones using Equation (2), and event chains are generated by linking event pairs when the relevance is higher than threshold R_{bc}.

$$\begin{aligned} & similarity(Event_A, Event_B) \\ = & \frac{\sum_{m=0}^{M-1} \sum_{n=0}^{N-1} cosine(W_{Am}, W_{Bn})}{MN} \end{aligned} \quad (2)$$

This threshold is calculated as follows. First, the relevance between events B and C is calculated to determine whether there is a semantic connection between these two events. Next, the degree of relevance between events AC, AD, BC, BD, is calculated. This condition is used to prevent generating unrelated event chains when only relevance between events B and C is calculated. Therefore, if relevance for all pairs is higher than thresholds R_{bc} , R_{ac} , R_{ad} , R_{bd} , an event chain is generated.

3.5 Implicit Knowledge Completion

A generated event is limited to verb phrase containing an object, but other information as actor, place or tool are omitted. As described in Introduction, we also aim at simulating implicit knowledge completion unconsciously processed by humans. For the first step of this simulation, we chose adding locations and by calculating relevance between events and places included in each event and by setting thresholds we attempted to generate most possible places for every generated event chain. A formula for a location completion is shown in Equation (3).

$$similarity(W_{place}, Event) = \frac{\sum_{n=0}^{N-1} cosine(W_{place}, W_n)}{N} \quad (3)$$

4 Performance Evaluation

We performed a manual evaluation of the results obtained in each process using online survey. Three graduate students in their twenties (native speakers of Japanese) evaluated generated event pairs and chains for generality and naturalness of semantic relations.

▷ level of generality in event chains
Generality level is evaluated from "commonly occurring" to "impossible" (1 to 5 on the Likert scale). Both event chains and event completion are evaluated with the same method, as well as the relevance of each event pair in the acquired event chain.

▷ naturalness of semantic relationship between events
Identical 1-5 using the Likert scale was used. In addition, we also evaluate all events combinations.

▷ generality level of location completion
Automatically added locations for event pairs and event chains were evaluated for generality in a given place on 1-5 Likert scale.

In this research we use Word2vec [Mikolov *et al.*, 2013] trained on Wikipedia [Suzuki *et al.*, 2016] and ConceptNet Numberbatch [Speer *et al.*, 2017] for word embeddings. Since Word2vec learns from sentences, it depends on the co-occurrences within the text. On the other hand, ConceptNet Numberbatch embeddings are based on node distances from commonsense knowledge ontology ConceptNet represented as a knowledge graph containing created manually facts. We compare effectiveness of both distributed word representations.

4.1 Dataset

In order to acquire the event knowledge, we use Japanese web blog corpus YACIS[Ptaszynski *et al.*, 2012] containing 309,765 articles. If two events occur in one article and their semantic relevancy is confirmed, they are treated as an event pair. From the corpus, we acquire two event pairs from all articles preserving their chronological order, and an event chain is generated from those pairs. To remove noisy pairs we used experimentally set the initial AB pair retrieval threshold to 0.3, deleted stopwords and events had to contain verb phrases. The table 1 shows the evaluation results for both embedding models used for calculations.

Table 1: Number of event pairs and chains depending on embedding model.

Models	Event Pairs	Event Chains
Word2vec	19,536	31,323
NumberBatch	8,727	13,423

Table 2: Evaluation of semantic relation naturalness evaluation between event pairs.

Models\part	AB	AC	AD	BC	BD	CD	Chain
Word2vec	3.6	3.4	3.4	2.7	2.86	4.2	3.0
NumberBatch	2.73	3.57	3.13	3.33	2.7	2.83	2.57

Table 3: Completion Evaluation depending on embedding model.

Models	AB with location	AB only	CD with location	CD only	Chain	Chain with location
Word2vec	2.53	3.6	4.0	4.2	2.43	3.0
NumberBatch	2.23	2.73	2.77	2.83	2.23	2.57

4.2 Effectiveness of Combining Event Pairs into Chains

Event chains generated from retrieved temporal events have been evaluated. Twenty event chains which had places added in the completion process were chosen randomly but presented without locations. The thresholds were different for each embedding model, and they were set experimentally. In addition to the overall generality of the event chain, naturalness of semantic relationship between the events was judged. The results are shown in Table 2.

4.3 Generality of Implicit Knowledge Completion

Location automatically added to acquired event pairs and event chain has been evaluated. Candidate place words for completion were restricted to ones a) included in the place category in the dictionary of JUMAN++[Morita *et al.*, 2015], morphological analyzer for Japanese, and b) included in ConceptNet Numberbatch. We added locations to a) event pairs AB and CD that we previously acquired from text and b) to event chains that were generated. Generality of both cases was evaluated and the results are shown in Table 3.

5 Results and Conclusions

From the experiment results, the generality of the event pairs which do not appear in the same sentence during chain construction appeared as high as the pairs retrieved from a single sentence. As expected, the average evaluation score of event chains was lower (0.29 point) than event pairs. This was caused by cases when one of events in an event pair was unnatural and overall evaluation score for the whole event chain automatically decreased. Experimental results also show that knowledge completion (with "place" as an example) achieved a lower degree of semantic relevance than a chain without common location added. However, in our opinion, tacit knowledge completion could lead to enriching machine understanding of situations through more exact context information. This is difficult to obtain when human annotators implicitly understand an event and their evaluation is less precise. The more information is given to evaluator, the evaluation becomes stricter, which can improve the difficult task of common sense evaluation.

In this paper, we proposed an event generation and completion method which, due to distributed word representation, does not depend on word frequency or temporal connectors for calculating semantic relation among events used in previous works. Furthermore, by automatic addition of place information to event chains we managed to enrich contextual knowledge about an event, trying to mimic implicit knowledge augmentation known from human thinking process.

6 Future work

In near future we plan to augment more implicit knowledge considering not only place information, but also other semantic roles, sentiments or social relations. We are also experimenting with automatic generation of script labels for event chains obtained in this research. We will also utilize machine learning for extending number of chains beyond four event pairs and for enriching implicit knowledge by generative approaches. If a rich completion is successful, we want to compare efficiency of language models trained on natural texts and the same text enriched by our completion methods.

7 Acknowledgements

This work was supported by JSPS KAKENHI Grant Number 17K00295.

References

[Abe *et al.*, 2008] Shuya Abe, Kentaro Inui, and Yuji Matsumoto. Acquiring event relation knowledge by learning cooccurrence patterns and fertilizing cooccurrence samples with verbal nouns. In *Proceedings of the Third International Joint Conference on Natural Language Processing: Volume-I*, 2008.

[Chambers and Jurafsky, 2008] Nathanael Chambers and Dan Jurafsky. Unsupervised learning of narrative event chains. *Proceedings of ACL-08: HLT*, pages 789–797, 2008.

[Feng *et al.*, 2018] Xiaocheng Feng, Bing Qin, and Ting Liu. A language-independent neural network for event detection. *Science China Information Sciences*, 61(9):092106, Aug 2018.

[Fujita *et al.*, 2011] Motoyasu Fujita, Rafal Rzepka, and Kenji Araki. Evaluation of utterances based on causal knowledge retrieved from blogs. In *Proceedings of the*

International Conference Artificial Intelligence and Soft Computing (ASC 2011), pages 294–299, 2011.

[Granroth-Wilding and Clark, 2016] Mark Granroth-Wilding and Stephen Clark. What happens next? event prediction using a compositional neural network model. In *Thirtieth AAAI Conference on Artificial Intelligence*, 2016.

[Higashiyama *et al.*, 2017] Shohei Higashiyama, Kunihiko Sadamasa, Takashi Onishi, and Yotaro Watanabe. Event relation acquisition using dependency patterns and confidence-weighted co-occurrence statistics. In *2017 Federated Conference on Computer Science and Information Systems (FedCSIS)*, pages 339–345. IEEE, 2017.

[Huang and Kurohashi, 2017] Yin Jou Huang and Sadao Kurohashi. Improving shared argument identification in japanese event knowledge acquisition. In *Proceedings of the Events and Stories in the News Workshop*, pages 21–30, 2017.

[Levesque *et al.*, 2012] Hector Levesque, Ernest Davis, and Leora Morgenstern. The winograd schema challenge. In *Thirteenth International Conference on the Principles of Knowledge Representation and Reasoning*, 2012.

[Lin *et al.*, 2018] Chen Lin, Timothy Miller, Dmitriy Dligach, Hadi Amiri, Steven Bethard, and Guergana Savova. Self-training improves recurrent neural networks performance for temporal relation extraction. In *Proceedings of the Ninth International Workshop on Health Text Mining and Information Analysis*, pages 165–176, 2018.

[McCann *et al.*, 2018] Bryan McCann, Nitish Shirish Keskar, Caiming Xiong, and Richard Socher. The natural language decathlon: Multitask learning as question answering. *arXiv preprint arXiv:1806.08730*, 2018.

[Mikolov *et al.*, 2013] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space. *CoRR*, abs/1301.3781, 2013.

[Morita *et al.*, 2015] Hajime Morita, Daisuke Kawahara, and Sadao Kurohashi. Morphological analysis for unsegmented languages using recurrent neural network language model. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 2292–2297, 2015.

[Ptaszynski *et al.*, 2012] Michal Ptaszynski, Pawel Dybala, Rafal Rzepka, Kenji Araki, and Yoshio Momouchi. Yacis: A five-billion-word corpus of Japanese blogs fully annotated with syntactic and affective information. In *Proceedings of The AISB/IACAP World Congress*, pages 40–49, 2012.

[Schank and Abelson, 1977] R. Schank and R. Abelson. *Scripts, plans, goals and understanding: An inquiry into human knowledge structures*. Lawrence Erlbaum Associates, Hillsdale, NJ., 1977.

[Speer and Havasi, 2012] Robert Speer and Catherine Havasi. Representing general relational knowledge in conceptnet 5. In Nicoletta Calzolari (Conference Chair), Khalid Choukri, Thierry Declerck, Mehmet Uğur Doğan, Bente Maegaard, Joseph Mariani, Jan Odijk, and Stelios Piperidis, editors, *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*, Istanbul, Turkey, may 2012. European Language Resources Association (ELRA).

[Speer *et al.*, 2017] Robyn Speer, Joshua Chin, and Catherine Havasi. Conceptnet 5.5: An open multilingual graph of general knowledge, 2017.

[Suzuki *et al.*, 2016] Masatoshi Suzuki, Koji Matsuda, Satoshi Sekine, Naoaki Okazaki, and Kentaro Inui. Multi-label classification of wikipedia articles into fine-grained named entity types. In *Proceedings of the Twenty-second Annual Meeting of the Association for Natural Language Processing*, 2016.

[Tozzo *et al.*, 2018] Alex Tozzo, Dejan Jovanovic, and Mohamed Amer. Neural event extraction from movies description. In *Proceedings of the First Workshop on Storytelling*, pages 60–66, 2018.

[Vedantam *et al.*, 2015] Ramakrishna Vedantam, Xiao Lin, Tanmay Batra, C. Lawrence Zitnick, and Devi Parikh. Learning common sense through visual abstraction. In *Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV)*, ICCV '15, pages 2542–2550, Washington, DC, USA, 2015. IEEE Computer Society.

[Wu *et al.*, 2018] Xing Wu, Shangwen Lv, Liangjun Zang, Jizhong Han, and Songlin Hu. Conditional BERT contextual augmentation. *CoRR*, abs/1812.06705, 2018.

[Yao and Huang, 2018] Wenlin Yao and Ruihong Huang. Temporal event knowledge acquisition via identifying narratives. *arXiv preprint arXiv:1805.10956*, 2018.