

Collaborative Crowdsourcing Between Experts and Crowds for Chronological Ordering of Narrative Events

John Joon Young Chung

KAIST
johnr0@kaist.ac.kr

Joseph Jay Williams

National University of Singapore
williams@comp.nus.edu.sg

Juho Kim

KAIST
juhokim@kaist.ac.kr

Abstract

Aligning narrative events in temporal order, a kind of narratological analysis, requires extensive expert labor and is difficult for computers. Those limitations impede narratology research at scale. It is also hard to be accomplished with crowdsourcing because reordering narrative events requires expertise and an understanding of the entire text. We propose a technique that enables crowd workers to do the complex task of narrative event ordering by embedding experts in the workflow who can effectively understand the text and direct crowd workers. In our crowdsourcing workflow, narratology researchers interpret a summary of a story, build instructional scaffolding, and assist a crowd on the reordering task with the scaffolded timeline. Results of a preliminary study show that the scaffolded timeline provided global contexts to crowd workers and enabled them to do the task that requires the overall context. We will improve the transfer of the expert's knowledge with iterations between experts and non-experts and revisions in the scaffolding.

Keyword

Crowdsourcing, Narratology, Expertise

1. Introduction

Narratology theory helps narratology researchers understand, analyze, and evaluate narratives [1]. It helps people understand the structure of narratives and enable them to look at narratives from the perspective of the structure. Ordering narrative events in a chronological timeline is one subtopic of narratological analysis. Narrative researchers find patterns in time, by comparing a time-ordered chronological timeline and an original sequence of events from a narrative text. Even though narratology theory is a major tool for narratology

research, analyzing a large number of long narrative texts in short time is laborious because the procedure of narratological analysis requires a lot of manual annotations and interpretation. Also, it is yet to be aided with automated tools. Most narratives are in a format that is yet to be dealt well with the machine, like natural language, and therefore limited automation has been applied in narratology research [2, 3, 12]. Narratology researchers rather have relied on their own intellectual capabilities when analyzing the chronological order of narrative events.

Crowdsourcing could provide a scalable solution to accelerating narratological analysis since crowd workers' natural language understanding significantly outperforms machines thus far [11]. However, the problem of transferring expertise to the crowd can be an issue in crowdsourcing the narratological analysis. Previous work has embedded experts in crowdsourcing [4, 7], but primarily focused on one-way communication, only conveying experts' point of view to crowd workers. Also, analyzing narrative contents like novels requires crowd workers to get a sense of the global context of the whole narrative material, which is difficult when the narrative is split into micro pieces to make each task approachable to a crowd. Previous research tried to tackle the problem of global context by aggregating local artifacts iteratively to build an artifact with a global viewpoint [13] and making crowds get the sense of the global view by showing diverging examples [6]. The former costs much more money than utilizing one expert because the amount of the necessary work multiplies with iterative summarizations. The latter is not applicable in the domain of narrative analysis because sequential materials like narratives can only be understood with overall context, not part of the materials.

We introduce an interactive crowdsourcing workflow that transfers expert knowledge and interpretation to crowds to enable them to work on the complex task of arranging narrative events in chronological order. The workflow we introduce envisions an iterative communication flow between experts and crowds, making crowd performance gradually better. For the iterative communication, we add a 'not sure' option to know whether the scaffold helps crowds do the task. In the workflow, experts understand global contents by reading an abstracted form of the narrative text, a summary split into chunks which only contain temporally close events. Experts convey their understanding of it by specifying important event chunks. Then they offer a scaffold for ordering narrative events in time order, by aligning important event chunks in time order. Crowd workers do tasks with the scaffold, deciding when event chunks that were not selected by experts happened in the chronological timeline. First, crowds decide which important event chunks come before and after remaining event chunks. Then event chunks with the same surrounding important events will form a group. In the next step, crowds compare pairs of event chunks within the same group to decide which event comes before in chronological timeline. The overall sequence of narrative events can be decided as a result.

In this work, we implemented the collaborative workflow and tested whether the workflow can help crowds do tasks well with an expert scaffold. Also, we tested the effectiveness of including the 'not sure' option in getting crowds' feedback on the difficulty of the task. We tested the workflow with crowd workers from Amazon's Mechanical Turk platform. Results showed that crowd workers could do the task that requires overall context, but the expert scaffold was not enough when the task required more details. We also found that it is challenging to know when crowds are not confident in their work outcome. Future work will improve the communicative and collaborative workflow between experts and crowds by adding components that detect when a crowd-generated outcome is not reliable and by conveying information in ways that do not overload crowd workers.

2. Formative Study

To understand current practices in narratology research and what composes narratological analysis, we conducted a focus group interview with seven narratology researchers at the graduate student level. The interview session took 120 minutes.

Narratology researchers analyze narrative texts by extracting structures before interpreting them. With analysis results, they interpret how structural components affect themes and expressions. Researchers make use of principles in narratological theories [1] to know what composes narratives, how they are structured, and how to analyze them. According to theories, various structural components can be extracted including time, tempo, narration, and event. Researchers usually focus on one aspect of the structure rather than on all components, and decide which component to analyze based on their research interest and prior knowledge of the text. Among them, time is about how the sequence of events expressed in the narrative is different from the sequence of events in chronological order [10].

Researchers noted that the analysis of structural components in a narrative can be different based on researcher's intention and required expertise. For instance, the analysis could be more or less difficult based on which granularity of components they focus on. They could analyze events of the narrative at a high level, which usually requires not much labor but insights, or instead focus on the sentences or words of the narrative, which makes researchers pay more attention to details. Also, they noted that the analysis of narrative structure could yield diverse interpretations because of the subjective nature of the narrative structure.

A major difficulty that narratology researchers face is the laboriousness of analysis. Because narrative structure analysis could hardly be done automatically, researchers analyze the structure manually, which makes the process slow and burdensome. Therefore, doing narratology research at scale is nearly impossible and meaningful research like analyzing general tendency in the history of literature is hard to

be accomplished only with limited resources narratology researchers have.

3. Design Goals

From the expert interview and related work, we identified three design goals for a crowdsourcing workflow for narratological analysis.

SUPPORT COMMUNICATION BETWEEN EXPERTS AND CROWDS

Making a collaborative workflow where experts and crowds can communicate is necessary for crowdsourcing narratological analysis. Based on what a researcher wants to find from the narrative, the practice of narratological analysis can vary a lot, and crowds have to understand her needs. Also, if the analysis requires expertise, crowds need to be informed about what expertise is required.

ENABLE NARRATIVE ANALYSIS AT SCALE IN A SHORT TIME

A tool for narratological analysis needs to alleviate a narratology researcher's burden and facilitate the large-scale analysis, for example by supporting efficient analysis of a long text. Because machine computation is yet to fully analyze the narrative structure, we decided to use crowdsourcing, because the crowd has the capability to do work in distributed ways, enabling faster work compared to a single individual. Also, the crowd has more capability in interpreting textual materials compared to the machine, from simple tasks like annotating emotions in the text[11] to complex task of synthesizing information from texts [6].

SUPPORT CROWDS WITH A GLOBAL UNDERSTANDING OF THE TASK

Crowdsourcing takes advantage of distributed deployment of microtasks, which makes the work fast and efficient [5]. However, tasks that deal with narratives are hard to be split into microtasks because they usually require crowd workers to know the overall picture of the narrative material. Hence, the system should support crowd workers to grasp the global picture of the material.

We decided to build a system that analyzes time order of narrative texts for a couple of reasons: 1) Because time order is one important component of narratological structure, and researchers can come up with various patterns and interpretations from it. 2) According to narratology researchers, time sequence alignment requires less knowledge of narrative theories, which makes it more amenable to crowdsourcing.

4. The System

The crowdsourcing system we introduce aims to analyze narrative text by ordering narrative events in temporal order. In this paper, we built three core steps of the crowdsourcing workflow. In the first step, an expert reads a summary of a narrative, which is split into blocks that contain consecutive flow of time, and identifies major events. The expert also aligns these major event blocks in chronological order. By doing so, a scaffold for future workers can be made, which informs crowd workers of the overall context and the approximate time flow of narrative events. In the second phase, crowd workers determine surrounding major events for all events. Events that fall between the same major events are grouped as events close in temporal distance. In the final phase, events in a group are compared pairwise to determine fine-grained temporal order between them. The final artifact is an event timeline that captures all major and minor events in the narrative material.

4.1 Expert Scaffolding

In this step, an expert builds a scaffolding that enables the crowd to understand the overall picture of the narrative texts. She first reads a summary of narrative text which is chunked into pieces. Each piece contains consecutive events without time leap in between events, making time flow within an event block sequential. After reading the summary, the expert decides which event block is important to understanding the story of the whole text. Then she aligns chosen important events in time order. This activity offers the expert an opportunity to understand the temporal flow of important events in

the narrative. Also, the resulting chronology of important events functions as a hint to future crowd workers, not only giving the abstracted summary of the storyline but also scaffolding workers on the order of events in time.

4.2 Deciding Chronological Positions Relative to Important Events

Figure 1 Interface for the relative chronological position decision task

In the second step, crowds order events not chosen by experts in the time-ordered sequence of the narrative. Crowds estimate the relative temporal position of a non-important event. They will be shown a non-important event, an original sequence of important event chunks and a chronological timeline of important event blocks (Figure 1). In the original sequence, the non-important event shown to each worker is also shown (Figure 1 c). They will get hints on the overall picture of the whole narrative from two artifacts generated by experts. Only important events but not all events are given as a hint, to avoid overloading crowd workers with the excessive amount of information. After deduction, crowds will decide where the non-important event should be placed in the chronological timeline, among places between two important event chunks, or the first or last place (Figure 1 a). It is also possible for crowd workers to select 'not sure' option when they are not sure with their decision (Figure 1 b). After five crowd decisions are collected for each event, they are aggregated with majority voting. Those non-important events with the same aggregated position will be in the same group and will be compared each other in the later step.

4.3 Deciding Chronological Order of Two Events

Figure 2 Interface for the event order decision task

In the last stage of the workflow, crowd workers make pairwise evaluations between non-important events to decide which event comes first in time order. In this step, crowds will be shown two events to compare. They are also shown an expert-generated original sequence and a time-ordered sequence for global understanding and hinting on time order of events (Figure 2). The original sequence includes the pair of events to compare. Two events to compare are from the same event group, which is the result of the previous step (Figure 2 c). After deduction, crowds can decide which event comes first, or they can select the 'not sure' option (Figure 2 a) when they cannot decide.

After all possible pairs from all groups are compared, the overall sequence can be decided. The temporal position of an event in the group can be decided by how many other events in the group happened after the event. If many events happened after the event, it can be said that the event happened earlier in the group [8]. With the time order of events in the group and temporal positions of groups in time order sequence, the overall sequence of the timeline can be decided. In the case when two events have coincided with the same number of following events in the group, the sequence will be decided with the direct pairwise comparison result.

5. The Experiment

In this paper, we hypothesize that crowd workers can yield reliable narrative time structure if afforded with an expert scaffold. To test the effectiveness of the proposed approach, we conducted a preliminary experiment. In the experiment, we used a summary of the movie *Old Boy* as a material. We screened participants to ensure they have not watched the movie. One of the authors chunked the story in 10 pieces (E0 ~ E9) with no time leaps within a piece. In the first step, the expert majored in Korean Literature chose 3 pieces as important and aligned them in time order, making four candidate positions for the second step (Pos 0 ~ Pos 3), which are before, after, and in between important event blocks. Then in the second step, for each event chunk not chosen by the expert, five workers made a decision on the temporal position, resulting in total 35 votes. Each crowd worker could work on only one event block. Workers were paid \$0.8 for the task, and were recruited from MTurk. In the final step, five crowd workers compared each pair of non-important events in the same group and decided which one comes first. Because three groups were generated in the previous step which contains one, four and two event blocks each, there were zero, six and one pair each and seven pairs in total, which resulted in total 35 votes. Each worker could work on one task only. Workers were paid \$0.85 for doing the task and were also recruited from MTurk.

5.1 The Result

Table 1. The gold standard result of time order sequence and the crowdsourced result. Event blocks written in bold-italic are event blocks chosen as important in the first step.

| | | | | | | | | | | |
|---------------|-----------|----|-----------|-----------|-----------|----|----|-----------|-----------|-----------|
| Gold Standard | <i>E4</i> | E0 | <i>E1</i> | E7 | E2 | E3 | E5 | <i>E6</i> | E8 | E9 |
| Crowdsourced | <i>E4</i> | E0 | <i>E1</i> | <i>E2</i> | <i>E7</i> | E3 | E5 | <i>E6</i> | <i>E9</i> | <i>E8</i> |

In the experiment, crowds could yield an accurate result in the first task of making decisions on relative chronological positions of events. However, in the sequence comparison task, which requires more detailed knowledge of narrative events, the result was less reliable. In the former task, crowds result

matched exactly same with the gold standard result. (Table 1) The number of votes tied in the event block 7 (E7), but it was also ambiguous to the expert whether it should be contained in position 1 or 2, coinciding with crowd result. For the latter step, we set the position of event block 7 (E7) as position 2 (Pos 2). In the former step, among seven voted event blocks, in three blocks, the difference in voting number between the most voted option and the second most voted option was less than one. In the latter task, crowds were wrong in getting the order of two event block pairs, resulting in four errors in total. When aggregating, because a pair of events got the same number of preceding events, we decided their order by the pairwise comparison result of those two events. Throughout all cases, crowds only selected option ‘not sure’ only once.

6. Discussion

We could find that expert scaffolding can be helpful to crowds in making a decision on the task that requires overall context, but it was not enough to get a perfect result in the task that requires more specific information. Therefore, the improvement in the scaffold is necessary. Also, the ‘not sure’ option was not adequate to detect crowd uncertainty in the work result. Crowd workers might not have selected it due to the fear that their work might get rejected [9]. Rather, it seemed like they made decisions without certainty, resulting in the small number of vote differences between the most voted options and other options in some decision cases. Therefore, the method for confidence level elicitation needs refinement. A possible candidate would be getting confidence rating (e.g., scale of 1–5) along with their decision. Also, expert scaffolding could be improved by showing more event blocks or highlighting important figures additionally. Future work will design ways to detect the reliability of crowd work to know when they need a more detailed scaffold, and improve the expert scaffold, in a way that does not overload crowds with excessive information.

Acknowledgement

This work was supported by Institute for Information & communications Technology Promotion (IITP) grant funded by the Korean government (MSIT) (No. 2017-0-01217, Korean Language CALL Platform Using Automatic Speech-writing Evaluation and Chatbot).

References

1. Bal, M. *Narratology: Introduction to the theory of narrative*. Toronto : University of Toronto Press (2009).
2. Bögel, T., Strötgen, J., Gertz, M. Computational Narratology: Extracting Tense Clusters from Narrative Texts. In *LREC* (2014), 950–955.
3. Bögel, T., Strötgen, J., Gertz, M. A Hybrid Approach to Extract Temporal Signals from Narratives. In *GSCL* (2015), 106–107.
4. Chan, J., Dang, S., Dow, S. P. Improving crowd innovation with expert facilitation. In *Proc. CSCW 2016*, ACM Press (2016), 1223–1235.
5. Cheng, J., et al. Break it down: A comparison of macro-and microtasks. In *Proc. CHI 2015*, ACM Press (2015), 4061–4064.
6. Hahn, N., et al. The Knowledge Accelerator: Big picture thinking in small pieces. In *Proc. CHI 2016*, ACM Press (2016), 2258–2270.
7. Kohler, R., Purviance, J., Luther, K. Supporting Image Geolocation with Diagramming and Crowdsourcing. In *Proc. HCOMP 2017, AAAI* (2017).
8. Marcus, A., et al. Human-powered sorts and joins. In *Proc. VLDB Endowment*, 5.1 (2011), 13–24.
9. McInnis, B., et al. Taking a HIT: Designing around rejection, mistrust, risk, and workers' experiences in Amazon Mechanical Turk. In *Proc. CHI 2016*, ACM Press (2016), 2271–2282.
10. Prince, G. *Narratology: The form and functioning of narrative*. Walter de Gruyter (1982).
11. Snow, R., et al. Cheap and fast---but is it good?: evaluating non-expert annotations for natural language tasks. In *Proc. EMNLP 2008. ACL* (2008), 254–263.
12. Strötgen, J., Gertz, M., A Baseline Temporal Tagger for all Languages. In *Proc. EMNLP 2015. ACL* (2015), 541–547.
13. Verroios, V., Bernstein, M. S. Context trees: Crowdsourcing global understanding from local views. In *Proc. of HCOMP 2014, AAAI* (2014).