

Tracking Human Process Using Crowd Collaboration to Enrich Data

David Pellow and Maxine Eskenazi

Language Technologies Institute, Carnegie Mellon University
Pittsburgh PA USA
dpellow@cs.cmu.edu, max@cs.cmu.edu

Abstract

A rich source of data that has been largely ignored in crowdsourcing is the processes that humans use to accomplish a task. If we can capture this information, we could model it for automatic processing and use it to better understand the phenomena being modeled. Using crowd collaboration to trace workers' process produces a rich dataset that can be mined for new insights. We tested this approach on the task of sentence simplification and show the sources and types of additional information that we have obtained.

Introduction

Human computation gets tasks done well at relatively low cost in record time. While aiming at one final result, the majority opinion, little attention is paid to the workers' process in arriving at that result. We believe that the use of human process information enhances the results.

Crowdsourcing tasks elicit human judgments for tasks such as annotations of images. Recently, more complex tasks use human creativity to generate content. For this, collaborative crowdsourcing (CC) was developed, where workers work together rather than in parallel. For example, Ambati et al. (2012) show that CC can produce better translations than traditional techniques.

In our CC experiment we embed an Etherpad (etherpad.org) editor in an Amazon Mechanical Turk (AMT) Human Intelligence Task (HIT) in which workers collaborate to simplify sentences. Etherpad is an online real-time collaborative editing platform which allows multiple workers to work together on the same task at the same time. We can observe the partial results and chat generated over time through the collaboration process. Our reasoning is that this information can enrich our dataset and disambiguate some findings.

Crowd Collaboration

We use Kittur's (2010) methodology, embedding an Etherpad editor in a HIT, and have workers collaborate real-time. Kittur noted that some passages in poetry translation take more iterations to complete than others, seeming to be more difficult to translate. He noted that this type of meta-information could provide valuable feedback for translation research. This is the basis of our approach.

worker	original	change	interpretation
W1	<i>under the age of 18</i>	<i>under 18</i>	deletion
W2	<i>behind-the-wheel skill-building</i>	<i>practice driving</i>	lexical
W3	<i>nighttime driving</i>	<i>driving at night</i>	syntax
W1	<i>you are required to</i>	<i>you must</i>	lexical
W4	<i>including</i>	<i>. This must include</i>	split sentence + lexical

Table 1: Multiple workers' iterations in a simplification task

The present study uses sentence simplification HITs to demonstrate that process information can be used to find which of the sentences need to be simplified, and to collect more useful information about the simplifications generated. To do this, we mine both the final simplifications and the intermediary ones as well as the chat between workers.

Besides providing useful meta-information, workers' process could be used to train automated simplification. Consider the following example whose iterations are in Table 1: Original sentence: *If you are under the age of 18, you are required to complete at least 65 hours of behind-the-wheel skill-building including 10 hours of nighttime driving.*

Final simplification: *If you are under 18, you must complete 65 hours of practice driving. This must include at least 10 hours of driving at night.*

In this example, the process information breaks down each change. Steps that may have seemed ambiguous when seeing only a final result are clarified.

Experimental Conditions

Each HIT included one original sentence in an Etherpad editor. Four workers were assigned to a HIT to come up with the simplification. The editor included a chat window allowing workers to discuss the task. Workers were also instructed to make a group decision to accept a given simplification. Workers could use a comments box for suggestions about the task itself. The Etherpad database stores each timestamped edit as well as the chats and the workers' ids.

Workers had to successfully complete a training session before being allowed access to the task. This session gave examples and counterexamples of various simplification techniques along with their explanations: lexical simplification, reordering, sentence splitting, removing unnecessary information, adding explanations, and making no change to

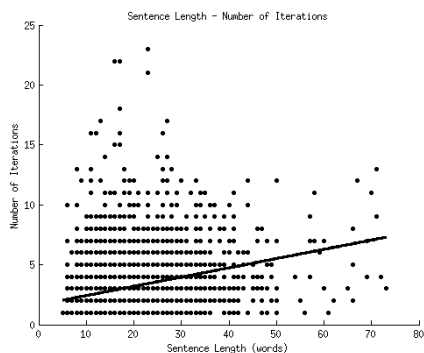


Figure 1: Original sentence lengths plotted against the number of simplification iterations.

	Original	First Iteration	Simplified
Avg. sentence length (# words)	17.04	14.02	13.41
Words not in top 10000	4279 (7.10%)	3933 (7.33%)	3335 (6.76%)

Table 2: Changes in sentence length and word frequency between original sentences, first iteration, and final simplifications.

sentences that are sufficiently simple. The workers’ abilities to apply the simplification techniques was assessed.

Workers were paid 20 cents per HIT, were required to make contributions for work to be accepted, and were instructed to submit only when a consensus had been reached.

3509 sentences from a variety of documents were used of which 200 were flagged as invalid. Multiple simplifications were submitted for 22, giving 3331 valid simplifications.

Simplification Difficulty

To indicate relative difficulty, we tracked the number of iterations for each simplification. Iterations are defined as successive turns that workers take in editing the simplification. We then correlated this measure with sentence length – a surface feature which can be computed directly from the text.

Figure 1 plots the number of iterations for a sentence to be simplified against the length of that sentence. We see a weak correlation between sentence length and number of iterations, with a Pearson’s correlation coefficient of $\rho = 0.29$.

Advantages of Crowd Collaboration

We used the first iteration of each simplification to approximate the result if a worker had been working alone. Table 2 shows a comparison of the sentence lengths and word frequencies (in the Brown corpus) of the original, first iteration and final collaborative result and Table 3 shows the differences in the various changes made. We detect these changes using TERp alignments (Snover et al. 2009), counting contiguous segments of multiple deleted or inserted words as a single simplification operation.

Syntactic Simplifications

We considered two types of syntactic simplification changes: deletions and insertions, detected by TERp. As we

	Original-First Iteration	Original-Final Simplified
Avg. # of deletions per sentence	0.93	1.54
Avg. # of shifts per sentence	0.33	0.63
Avg. # of insertions per sentence	0.23	0.25

Table 3: Changes detected by TERp between parallel original-first iteration and original-final simplification pairs.

# sentences with multiple deletion stages	2660
Avg. # of deletions per sentence in original-final sentence pairs	1.54
Avg. # of deletions per sentence over entire simplification processes	2.33

Table 4: Deletion operations over multiple iterations.

# sentences with multiple insertion stages	1740
Avg. # of insertions per sentence in original-final sentence pairs	0.25
Avg. # of insertions per sentence over entire simplification processes	1.18

Table 5: Insertion operations over multiple iterations.

mentioned, the partial information helps disambiguate some actions. For example, inserting or deleting a large contiguous segment may occur as more than one operation. Similarly, if there are multiple deletion or insertion, they could be interdependent or completely independent of one another. Knowing that the operations occurred in different stages of the simplification allows us to resolve the question.

Tables 4 and 5 report the information about deletion and insertion operations gained from the simplification process.

Conclusions and Future Directions

The results show that features of the workers’ collaborative process can be informative. The positive correlation we found indicates the ability of these measures to predict how difficult a sentence may be to simplify. We have also shown the advantages of using collaboration in complex tasks, and that partial information can produce a richer dataset. Future work should build on these simple features, analyzing chat contents, changed words, contexts and other factors. Eventually we want to model the whole simplification process in order to improve automated simplification.

References

- Ambati, V.; Vogel, S.; and Carbonell, J. 2012. Collaborative workflow for crowdsourcing translation. In *Proc. of the ACM 2012 Conference on Computer Supported Cooperative Work, CSCW ’12*, 1191–1194. ACM.
- Kittur, A. 2010. Crowdsourcing, collaboration and creativity. *XRDS:Crossroads* 17:22–26.
- Snover, M.; Madnani, N.; Dorr, B. J.; and Schwartz, R. 2009. Fluency, adequacy, or hter?: Exploring different human judgments with a tunable mt metric. In *Proc. of the 4th Workshop on Statistical Machine Translation, StatMT ’09*, 259–268. ACL.