Information Systems 71 (2017) 103–110

Contents lists available at ScienceDirect

Information Systems

journal homepage: www.elsevier.com/locate/is

Deployment strategies for crowdsourcing text creation

Ria Mae Borromeo[a,*], Thomas Laurent[a], Motomichi Toyama[a], Maha Alsayasneh[b], Sihem Amer-Yahia[b], Vincent Leroy[b]

[a] *Keio University, Yokohama, Japan*
[b] *Univ. Grenoble Alpes, CNRS, LIG, F-38000 Grenoble, France*

ARTICLE INFO

Article history:
Received 23 June 2017
Accepted 28 June 2017
Available online 4 July 2017

Keywords:
Crowdsourcing
Deployment strategies
Text creation

ABSTRACT

Automatically generating text of high quality in tasks such as translation, summarization, and narrative writing is difficult as these tasks require creativity, which only humans currently exhibit. However, crowdsourcing such tasks is still a challenge as they are tedious for humans and can require expert knowledge. We thus explore deployment strategies for crowdsourcing text creation tasks to improve the effectiveness of the crowdsourcing process. We consider effectiveness through the quality of the output text, the cost of deploying the task, and the latency in obtaining the output. We formalize a deployment strategy in crowdsourcing along three dimensions: work structure, workforce organization, and work style. Work structure can either be simultaneous or sequential, workforce organization independent or collaborative, and work style either by humans only or by using a combination of machine and human intelligence. We implement these strategies for translation, summarization, and narrative writing tasks by designing a semi-automatic tool that uses the Amazon Mechanical Turk API and experiment with them in different input settings such as text length, number of sources, and topic popularity. We report our findings regarding the effectiveness of each strategy and provide recommendations to guide requesters in selecting the best strategy when deploying text creation tasks.

1. Introduction

Crowdsourcing has been applied to all kinds of tasks ranging from the simplest such as image categorization to more sophisticated ones such as the creation of elaborate text. Although several automatic solutions have been designed for text creation, this task remains difficult for machines as it involves a level of abstraction and creativity that only humans currently possess. Text creation is also challenging for humans because it requires comprehension and edition, two time-consuming operations. That is particularly true for translation, summarization, and narrative writing where inputs of varying length and complexity need to be understood to proceed with a task. In this paper, we explore different deployment strategies for crowdsourcing text creation tasks in a way that optimizes the quality of the produced text, the cost of deployment as well as the latency in obtaining the results. To the best of our knowledge, our work is the first to explore the effectiveness of deployment strategies for crowdsourcing text creation.

We are interested in three text creation tasks: translation, summarization, and narrative writing. It has been shown that for translation, letting workers edit text and correct each others' mistakes in a sequential manner produces higher quality translations than in the case where workers generate independent translations simultaneously [1]. In the case of summarization, it has been shown that automatic methods are not very good at summarizing and merging sentences to generate high-quality summaries [2] and while there are existing tools that can effectively generate narratives, the resulting texts were found to be poor compared to human generated ones [3]. We hence study different crowdsourcing deployment strategies for these text creation tasks.

We define a deployment strategy as a plan on how to carry out a task. It is a combination of three dimensions: *work structure, workforce organization*, and *work style*. Work structure refers to how a task is deployed among workers, which can either be *simultaneous* or *sequential*. Workforce organization refers to how workers are organized to complete a task; it can either be in an *independent* or *collaborative* fashion. Work style distinguishes a *hybrid* approach, where a task is completed by both algorithms and humans, from a *crowd-only* approach, where a task is solely carried out by humans. The combination of those dimensions results in 6 strategies (we do not consider the combination of sequential work structure and collaborative and workforce organization as we interpret collaboration to be simultaneous). For example, in one of our deployment strategies, Sequential-Independent-CrowdOnly

* Corresponding author.
E-mail address: riamae@keio.jp (R.M. Borromeo).

http://dx.doi.org/10.1016/j.is.2017.06.007
0306-4379/

(SEQ-IND-CRO), an initial output is completed by a worker then the output is sent to one worker at a time to improve it. The final result is a single output. In Simultaneous-Collaborative-Hybrid (SIM-COL-HYB), an initial output is generated automatically and then sent to a group of workers who collaborate to improve it.

We developed a tool that uses the Amazon Mechanical Turk[1] (AMT) API to enable a semi-automatic implementation of our proposed strategies. We then performed experiments, testing these strategies by applying them to specific translation, summarization, and narrative writing tasks.

From our empirical study, we found that most workers default to automatic translation tools for translating long text thus, a hybrid work style, where workers are asked to simply improve an automatically translated text, combined with a sequential work structure works best. For shorter text, however, we observed that a simultaneous work structure, which provides more translation options, is more appropriate, and both hybrid and crowd-only work styles perform well.

For summarization tasks, we observed that when given an initial summary to work with, workers tend to improve the syntax of the text rather than its content, making improvements less significant. In this case, we recommend a crowd-only work style combined with a simultaneous work structure.

Although summarization and narrative writing tasks are similar, we found that for popular topics such as soccer, the improvements done to initial texts were not just on syntax but also on content. In this case, we recommend a sequential work structure combined with a hybrid work style since the quality obtained from hybrid and crowd-only work styles are comparable.

The paper is organized as follows. Section 2 reviews the related work. Our proposed strategies and the tasks we apply them to are detailed in Section 3. Our experiments are explained in Section 4. We then discuss our findings in Section 5 and finally conclude in Section 6.

2. Related work

We first review work structure, workforce organization and hybrid human and machine methods, the three task deployment dimensions characterizing our deployment strategies. We then review current approaches for the three types of text creation task that we consider.

2.1. Task deployment dimensions

2.1.1. Work structure

While general-purpose crowdsourcing platforms such as AMT mainly support a simultaneous (i.e. parallel) completion of independent tasks, Little et al. introduced a sequential work structure that involves an iterative workflow paradigm wherein a worker builds on or evaluates the work of another worker [4]. They implemented TurKit, a toolkit that deploys iterative tasks on AMT. TurKit employs a fixed policy that performs improvement tasks until it consumes a given budget.

Studies that compare sequential and simultaneous work structures suggest that the recommended work structure depends on the type of task. For tasks such as writing image descriptions, brainstorming company names, and transcribing blurry text, Little et al. found that using a sequential work structure improves average response quality [5]. Similarly, for a limerick writing task, Andre et al.'s findings reveal the sequential work structure to be more effective as the number of workers collaborating on a task increases [1]. However, for a taxonomy creation task, using sequential work does not yield positive results because the taxonomy grows with every iteration thus making tasks more time-consuming and overwhelming [6]. The sequential work structure also does not fare better for an outline creation task where a tournament workflow, which allows multiple merges of independent parts of an outline in parallel, produces faster, higher quality and more diverse outlines [7].

2.1.2. Workforce organization

This dimension focuses on determining the appropriate set of workers for a specific task. Simple tasks such as labeling an image and judging the sentiment of text, are commonly done by workers independently. However, previous studies show that more complex tasks such as translation [8], workflow design [9], user interface control [10], and article writing [11], are more effectively done by workers collaborating together. Appropriately assigning workers to collaborate on a task, however, is a challenge. Roy et al. models task assignment as an optimization problem and propose adaptive algorithms that consider human factors such as worker expertise, availability, and wage requirement [12]. For a collaborative news document editing task, their framework achieves high-quality and efficient task assignments within a specific budget.

Another way to organize a workforce is based on the known quality of workers. CrowdFlower[2] assigns a contributor level to workers based on their work history. Requesters can specify the required contributor level to be able to complete their tasks. RABJ [13] maintains a tiered worker hierarchy enabling workers to be assigned to tasks based on their performance. MobileWorks [14] hires managers, a particular class of workers who are in charge of recruiting new workers, evaluating potential problems with requester-defined tasks, and resolving task discrepancies. Argonaut [15] uses a predictive model of worker quality to select qualified workers as reviewers of others' work.

2.1.3. Hybrid human and machine methods

Crowdsourcing database systems such as Qurk [16], Deco [17], and CrowdDB [18] combine relational database management systems and crowdsourcing. They follow the basic workflow of query processing, which involves parsing the query, generating one or more query plans, then selecting the best query plan using both humans and machines [19]. The ability of the crowd to provide results to queries that traditional database systems cannot answer, such as those that involve subjective comparisons and unknown or incomplete data, complements the known strengths of database systems.

Another common approach combines automatic methods and crowdsourcing to reinforce each other. For instance, in a structured data extraction task, the Argonaut system uses automated extractors and machine learned classifiers to identify the components of a document then asks reviewers from the crowd to correct the output of the automated extractors [15]. In a sentiment analysis of reviews, Wu et al. [20] uses various machine learning algorithms to classify reviews. If the classifications produced by the algorithms disagree for a particular review, the review is assigned to humans then the results from the algorithms and crowdsourcing are aggregated to derive the outcome. For a web table matching task, CrowdWeb [21] introduces a concept-based approach that maps each column of a web table to the best concept in a well-developed knowledge base and asks the crowd to discern the concepts that are difficult to discern automatically.

In our work, we propose different strategies that combine work structure, workforce organization, and hybrid human and machine plans, for translation, text summarization, and narrative writing

[1] https://www.mturk.com.

[2] http://crowdflower.com.