Bigger Thinking Through Micro-Tasks

Aniket Kittur
Joseph Chee Chang
Nathan Hahn
Carnegie Mellon University
Pittsburgh, PA, USA
nkittur@cs.cmu.edu
josephcc@cmu.edu
nhahn@cmu.edu

Ji Eun Kim
Bosch Corporate Research
Pittsburgh, PA, USA
JiEun.Kim@us.bosch.com

Abstract

Crowdsourcing offers a powerful new paradigm for accomplishing work through microtasks. However, real world tasks are often interdependent, preventing them from being easily decomposed and distributed. While existing crowdsourcing approaches offer solutions to task decomposition, no systems offer a distributed solution to the “big picture” context required to manage the task. In this position paper, we discuss the challenges in crowdsourcing a particularly complex, interdependent tasks: information synthesis. We introduce two interleaved systems that accomplish this feat, known as the Knowledge Accelerator (KA) and Alloy, and introduce a set of design patterns for complex, interdependent crowdsourcing. Overall, we believe these approaches represent a step towards a future of big thinking in small packages, in which complex and interdependent cognitive processes can be scaled beyond individual cognitive limitations by distributing them across many individuals.

Author Keywords

Crowdsourcing, coordination, context, work, micro-tasks, markets, crowd

Introduction

Crowdsourcing is a powerful mechanism for accomplishing work online. By decomposing and distributing the cognitive work of an individual,

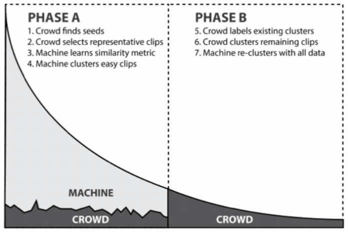

Figure 1: A conceptual overview of the Alloy system. In the first phase, crowd workers identify seed clips to train a machine learning model, which is used to classify the "head" of the distribution. In the second phase, crowd workers classify the more difficult items in the "tail". A machine learning backbone provides a consistent way to connect worker judgments in different phases.

crowdsourcing can provide a larger pool of resources more quickly and with lower transaction costs than through traditional work. However, much work in the real world is not amenable to crowdsourcing because of the difficulty in decomposing tasks into small, independent units. As noted by many researchers [1,2,3], decomposing tasks ranging from writing an article to creating an animated film often results in pieces that have complex dependencies on each other. Take for example the goal of writing even a simple article about growing tomatoes. At the lowest level, each sentence must be coherent and align with the other sentences in the paragraph. At a higher level, each paragraph within the article must fit together as well, and sections need to have proper transition and flow. Moving to even higher levels, the article must have an appropriate set of topics (e.g., appropriate soil, sunlight, watering, pruning) that are coherent and comprehensive. Information from different sources should be appropriately synthesized and cited while reducing redundancies and bias. Supporting this type of work requires having a big picture view of different pieces at different scales and ensuring they all fit together.

Accomplishing this big picture thinking through small tasks is challenging because it means that each person can only have a limited view of the bigger picture. As a result, many of the applications of crowdsourcing have been limited to simple tasks such as image labeling where each piece can be decomposed and processed independently. Those approaches that do crowdsource tasks requiring big picture thinking – such as volunteer communities such as Wikipedia, open source software, or paid crowd work approaches such as flash teams [2] or Turkomatic [3] – have relied on a heavily invested contributor such as a moderator or an experienced contributor to maintain the big picture. For example, in Wikipedia a large proportion of the work is done by a small group of heavily invested editors, and the quality of an article is critically dependent on there being a small number of core editors who create and maintain a big picture structure for more peripheral members to contribute effectively.

A reliance on a single or a small number of individuals to maintain the big picture creates a bottleneck on the size and complexity of task amenable to crowdsourcing, and also results in brittleness: if the person maintaining the big picture leaves, it can cause serious problems for the group task. This is a real problem that online production communities are facing; for example, Wikipedia has identified as a key challenge that it is losing core editors faster than it can attract and grow new ones. Enabling the production of complex artifacts through many contributors making small contributions might thus have implications in reducing individual bottlenecks in microtask markets; in volunteer crowdsourcing efforts such as Wikipedia or friendsourcing in which many small contributions are readily available; or self-sourcing in which the crowd within could accomplish complex tasks in small increments (e.g., waiting for the bus) without needing to load the entire task context into working memory.

Over the past three years we have built a series of prototype systems aimed at engaging with the challenge of supporting big picture context through small contributions on the task of information synthesis. During this process we have encountered innumerable dead ends and promising but ultimately low value approaches; for example, at one point we

Question	N	Score
Q1: *How do I unclog my bathtub drain?*	116	0.292 *
Q2: *How do I get my tomato plants to produce more tomatoes?*	177	0.420 *
Q3: *What are the best attractions in LA if I have two little kids?*	158	-0.044
Q4: *What are the best day trips possible from Barcelona, Spain?*	98	-0.109
Q5: *My Worcester CDi Boiler pressure is low. How can I fix it?*	139	0.878 *
Q6: *2003 Dodge Durango has an OBD-II error code of P440. How do I fix it?*	138	0.662 *
Q7: *2005 Chevy Silverado has an OBD-II error code of C0327. How do I fix it?*	135	0.412 *
Q8: *How do I deal with the arthritis in my knee as a 28 year old?*	139	0.391 *
Q9: *My Playstation 3 has a solid yellow light, how do I fix it?*	119	0.380 *
Q10: *What are the key arguments for and against Global Warming?*	138	0.386 *
Q11: *How do I use the VIM text editor?*	138	0.180
* = significant at $p < 0.01$ after Bonferroni correction		

Table 1. Average difference between the KA output and top websites for the eleven evaluated questions (positive indicates higher ratings for KA, negative indicates higher ratings for the competing website). Each rating was an aggregate of 6 questions on a 7-point Likert scale.

built an entire subsystem to support “conductors” whose job was to focus on the global context, which failed for interesting reasons. It is possible that this “negative” knowledge may prove at least as useful in the workshop as the design patterns and systems we will also bring as fodder for discussion.

As for the latter, we have been working on two primary systems (KA and Alloy). KA takes as input an arbitrary question (e.g., “how do I grow better tomatoes”) and uses human judgments and computation to output a synthesized digest which finds and vets relevant online sources, extracts relevant information segments, clusters them into appropriate topics, synthesizes knowledge within each topic, and maintains context within and between topics. To avoid high-load bottlenecks we set a limit on each microtask of $1. To accomplish this we introduce a new set of design patterns for supporting big picture context, including ways to provide context before action, and to leverage workers’ natural inclination to pick “tasks of least resistance” to increase agency, quality control, and context simultaneously. In an evaluation we find the system’s output is rated higher overall than the top five Google results for a variety of questions, suggesting that we can achieve competitive performance for a complex and interdependent task entirely through small contributions (see Table 1).

Alloy expands on the clustering phase of KA, addressing the problem of clustering complex, rich text snippets when no worker can process the whole dataset. Alloy improves on previous crowd clustering methods in providing greater global context along with increased efficiency by introducing two new concepts. First, it introduces a “sample and search” pattern that increases context by changing the crowd’s task from classifying a fixed subset of items to actively sampling and querying the entire dataset. Second, it improves efficiency by introducing a novel modular approach we call “cast and gather” which employs a machine learning backbone to stitch together different types of crowd judgment tasks. For example, it uses initial crowd judgments to help a machine learning algorithm cluster high-confidence unlabeled items in the head of the distribution, and then uses later crowd judgments to improve the quality of machine clustering by covering edge cases in the tail of the distribution.

We hope the design choices and patterns embodied in the KA and Alloy prototype systems may be useful for other system designers aiming to accomplish complex cognitive tasks without the bottleneck of requiring an individual having the full global context of the system. Overall, we believe these approaches represent a step towards a future of big thinking in small packages, in which complex and interdependent cognitive processes can be scaled beyond individual cognitive limitations by distributing them across many individuals.

References

1. Aniket Kittur, Jeffrey V Nickerson, Michael Bernstein, Elizabeth Gerber, Aaron Shaw, John Zimmerman, Matt Lease, and John Horton. 2013. The future of crowd work. In *CSCW 2013*.
2. Daniela Retelny, Sebastien Robaszkiewicz, Alexandra To, Walter S Lasecki, Jay Patel, Negar Rahmati, Tulsee Doshi, Melissa Valentine, and Michael S Bernstein. 2014. Expert crowdsourcing with flash teams. In *UIST 2014*.
3. Anand Kulkarni, Matthew Can, and Bjorn Hartmann. 2012. Collaboratively crowdsourcing workflows with turkomatic. In *CSCW 2012*.