AI Backed Crowd-Sourced How-to Manuals

by

Daniel J. Scarafoni

Submitted in Partial Fulfillment

of the

Requirements for the Degree

MASTER'S OF SCIENCE

Supervised by

Professor Philip Guo

Department of Computer Science
Arts, Science & Engineering
Edmund A. Hajim School of Engineering & Applied Sciences

University of Rochester
Rochester, New York

2014

Abstract

So called unskilled jobs such as deli clerks and table waiting in fact require a number of skills to do well; these skills are often learned on the job through experience and mentorship situations. The process could be simplified with instruction written by professionals. Canned Mentorship is a system which allows small groups of professionals to create comprehensive instructions for day-to-day tasks in their jobs for use by newer staff, reducing the need for supervision. This program uses a front-end user interface which allows users to view the instructions as they are being formed and vote on editions and revisions. A back-end server collects the natural language data and parses it to coordinate the groups ideas. The end result is a list of instructions describing a particular task.

Table of Contents

1 Introduction

Crowdsourcing has shown to be an effective means of solving tasks that are computationally difficult, or otherwise impractical, for artificial intelligence programs to solve. (?). There are many instances, of course,where humans are more effective than machines and vice-versa. To pick the most obvious example: humans are more efficient at qualitative, ambiguous tasks, while artificial intelligence (AI) agents tend to be better at more quantitative, direct tasks. Naturally, most tasks do not fall cleanly into one category or another; there are many tasks which cannot be done perfectly by human or AI agents. For this reason, the merging of human and AI is necessary for the completion of a number of tasks.

Consider, for example, the task of building an instruction manual for a common profession. If a restaurant wanted to create a manual for a number of everyday tasks in the restaurant in order to train new employees quickly.

ADD SOMETHING ABOUT THIS BEING ABOUT SKILLED USERS

It would use data gathered from existing employees to solve this task, as the ambiguous and qualitative data of procedural tasks such making food and settling tables is far beyond the current capabilities of artificial intelligence. One experienced employee could be delegated to the task-instruction creation, but there are a number of issues with this method. First, an over-reliance on one individual leaves the system vulnerable to error. An individual may make mistakes when writing instructions. The only solution to this problem would be to have other users check the first users work, which automatically means that multiple users will be needed regardless. Second, many task, especially the qualitative, ambiguous ones previously mentioned, can be done in multiple ways. While the differences may not necessarily be erroneous in nature, they can pose problems. Certain delis, for example, have managers with different policies on the quality of meat used for sandwiches. If a trainee is taught the preference of one manager and not the other, issues can arise if the employee switches managers. Third, there are many instances where individuals have been shown to be slower than crowds (Lasecki and Bigham, 2013; **?**). It it likely that crowd-sourced users will also be more efficient than single users in this domain.

One major issue with gathering instructions from users is dealing with redundancy. Because many people have similar notions of instructions for various tasks, users tend to give similar instructions for a given step. This leads to redundancy. For example, if a group of 3 users were asked to write the first step for making a peanut-butter and jelly sandwich, the users might respond with

1. get two slices of bread
2. get two slices of bread, a knife, peanut butter, and some jelly
3. get the ingredients.

It is obvious in this situation that because suggestion 3 is described completely (and more thoroughly) by suggestion 2, that it is redundant and thus can be removed.

Failure to do so can impede the crowd-sourcing process. First, it introduces a form of noise into the system. With multiple identical sentences, if users are asked to vote on the suggestions, the voting process may be subverted. Consider the same situation as the last example, but the three inputs are

1. get two slices of bread
2. get the ingredients
3. get the ingredients.

If two users believe that option 2 (or equivalently, option 3) is correct, they could vote for either. If the third user votes for option 1, then there is no clear majority from voting, even though the consensus is clear.

This problem ultimately is an issue of natural language processing, and is unique in that it deals with very short documents in a very small corpus. A quick and efficient method of removing and/or minimizing redundancies in the context of this problem is essential to the completion of the task.

In total, the task of building a set of instructions (or a "how-to" manual) efficiently from a group of users requires efficient crowd-sourcing techniques as well as sufficiently sophisticated AI to augment this process. Although there has been much research on using crowds to answer single questions, there has not been much work done in the next level of abstraction: having users answer lists of questions (i.e. instructions) (Lasecki *et al.*, 2013; Bigham *et al.*, 2010).

add objectives, purpose, etc

The remainder of this paper will be organized as follows:

- the Background section will discuss previous work done in both crowdsourcing and redundant sentence removal, as well the unresearched areas where this project will make its contributions

- The Methodology section will describe the Canned Mentorship system in detail, both at an algorithmic and implementation level. It will also descirbe the parameters tested and the experimental setups.

- The results section will describe the experimental results of the study, and discuss the any anomalies or other observations encountered during the testing process.

2 Background

2.1 Crowdsourcing

Crowdsourcing has shown itself to be a very effective computational resource. Formally, it is defined as

> the act of a company or institution taking a function once performed by employees and outsourcing it to an undefined (and generally large) network of people in the form of an open call. (Brabham, 2008).

2.1.1 Current Scope and Usage of Crowdsourcing

Though crowdsourcing covers a very broad range of activities and fields, it is possible to define separate catagories. Brabham defined three separate categories:

1. crowdfunding
2. crowd labor
3. crowd research.

Crowdfunding is the use of crowd-based resources to fund projects. Websites such as gofundme.com, kickstarter, and patreon.com are all crowdfunding website which allow elicit funds from users. This method is particularly useful for freelance artists (Brabham, 2008).

Crowd labor utilizes the crowd for computation and other work-like tasks. Generally, this involves the users doing low effort tasks for low pay. In many situations, this can result in a much cheaper product than non crowdsourced, professional settings. Crowdsourcing is also often used for tasks which are prohibitively difficult or impossible with current computational means.

For example, the stock photo website iStockphoto collects images uploaded by users, and pays them a small compensation for their effort. This website is able to lease the rights to these images for $1, which undercuts professional photographers' prices by over 90%, Crowd labor also extends to a number of other fields, including industrial research (Howe, 2006).

2.1.2 Effective Crowdsourcing

3 Methodology

The efficacy of the program was dependent on its ability to coordinate users and ensure that they were able to work together to generate the final instructions. Thus, it was necessary to not only design an effective workflow, but to also make sure that the program implementation was easy to understand, robust, and close to the theoretical workflow.

3.1 Workflow

Turkomatic ran into difficulties in organizing the workflow. Their system was unable to coordinate workflows, in no small part due to the complexity of the algorithm. Their algorithm used recursive workflow partitioning, and allowed for complex sub-task creation. This led to significant difference of opinion between users, and little consensus was obtained without the use of a single task evaluator who would oversee the task (also known as "derailment" (Kulkarni *et al.*, 2012).

To combat this problem, Canned Mentorship utilizes an iterative, one directional workflow. Instead of a complex, largely unordered workflow of turkomatic, Canned Mentorship makes users create the instruction list in the order it will be executed. This was done to avoid problems with granularity and starvation that were problematic in this system.

3.1.1 In-Order Instruction Addition

CannedMentorship uses a very simple task generating algorithm, with no explicit task subdivision or branching. The system simply asks users to create steps in sequence, from beginning to end, in order. Each step is completed in order, with the latter steps coming directly after former steps chronologically. This avoids the issue of task subdivision. Users do not have to deal with the complex uncertanties of task completion, and do not have to worry about planning the task.

3.1.2 Canned Mentorship Metaworkflow

Previously, three-stage find, fix, verify algorithms have been used successfully in crowdsourcing (Kim, 2013; Bernstein *et al.*, 2010). Canned Mentorship builds off of this system.

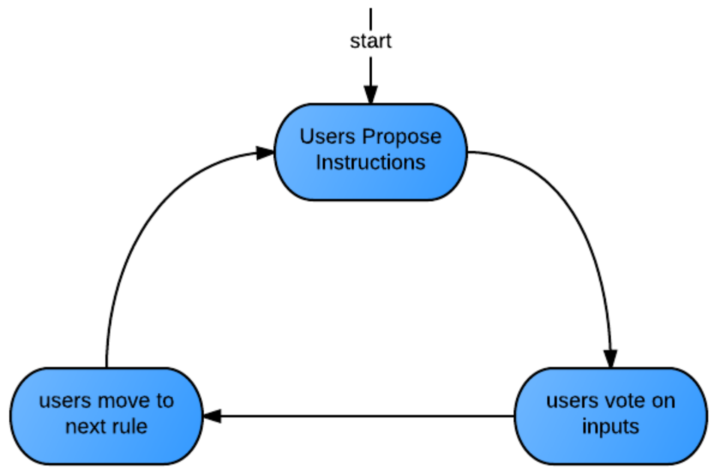
In order to create a rule, users are asked to write down their opinion of what the new rule should be. They send this input, which is typically one or two short sentences. These inputs are then voted on by users. The suggestion with the most votes becomes the next instruction. It is added to the existing list of instructions in the workflow, and the users then begin work on the next instruction.

The actual writing process follows a cycle, illustrated

3.2 Implementation and Webapp details

Our project began by creating a small webapp which coordinated the instruction generating process. This program, also dubbed "Canned Mentorship," consisted of two separate programs:

1. a website interface for coordinating working users
2. a back-end AI script for classifying answers and removing redundant inputs


start
Users Propose Instructions
users vote on inputs
users move to next rule

3.3 System Implementation

3.3.1 The Webapp

The webapp was a program written using Python Flask and was hosted on Heroku webapps. It was designed as a lightweight, simple system which hosted the users for the duration of the task and also coordinated their actions.

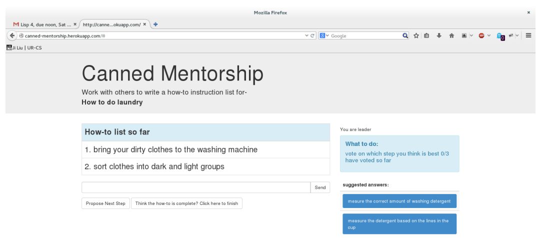


Figure 3.1: Rejection sampling was also considered as a placement mechanism.

As noted earlier, Heroku was chosen as the host location for the webapp ...add more stuff later...

4 Results

5 Conclusion and Discussion

As shown in figure the system is able to handle the simple case for object querying very reasonably. The fastest calculating predicate is `isTouching`, followed closely by `near`. Because these two predicates operate on similar algorithms, it make sense that the two operate in similar time. One potential significance is the fact that `isTouching` operates faster than `near`, and the only tangible difference between the two algorithms is the fact that `isTouching` has a smaller threshold than `near`. It was first speculated that this meant that the internal Blender functions relied on by these predication functions' run-time increased with object distance. However, the more likely explanation is the calculation of the threshold between the two functions. In `near`, the threshold is calculated with a function that operates in linear time. In `isTouching`, it is a constant, predetermined value, and thus requires no overhead to calculate.

The program worked into the Cornell Cup's Haptek team, where it was used to simulate a person navigating a virtual maze. This project set out to equip a room with several Kinect cameras which monitored the movement of a person in the room. A program, receiving input from the cameras, would build a virtual maze in a virtual representation of the room. The person's virtual avatar would be inserted into this simulation.

The person's motion would be tracked with the cameras, and the avatar would move with the user. If the avatar came into contact with one of the virtual walls in the maze, then a buzzing feedback mechanism would be triggered. This allows the user to navigate a virtual maze. Because this is a simulation in 3D space, the project utilized many of the aspects of the IMS project, including several of the predicates, the models for "person" and "wall," and the temporal updating feature of the 3D scene.

In a similar note, the use of Blender has opened up a number of options for future endeavors. Because Blender is a widely-used program with a broad range of applications and widespread compatibility, IMS has the potential to be utilized in any number of studies and situations. For example, many 3D printers currently accept model input in file formats supported by Blender. At this year's Rochack Hackathon, we were able to 3D print one of the models. This model can be seen in figure **??**.

The Blender file for the person was uploaded with minimal changes to the 3D printer at the school, and printed. The entire process of converting the file, uploading it to the printer, and printing the figure took less than an hour. Given the growing importance of 3D printing in the present, and the many applications to 3D environments that IMS offers, the ability to quickly model objects that can be 3D printed could prove to be a valuable feature of the system.

An area of concern is the run-time of the "inside" predication. The run-time of this predication is significantly worse than the others. Because this function runs in $O(n^2)$ time, it is most likely not due to an inefficient algorithm. Rather, this lack of satisfiability is likely related to the legal placement area generated by the placement function in predMethods.py.

Inside is similar to in, differing only in that it requires one object to be inside the mesh (rather than bounding box) of the other. The legal area, however, is generated to encompass the entire bounding box. As such, it may overestimates the legal area by a considerable amount. Shrinking the placement area would potentially solve this problem, though it may prevent placement in legal areas.

Our project was able to successfully query and place entities under predicate constraints. Because the scene was able to pass the original test of constructing the story-based image and deduce the implicit, spatial information in the scene (the person could not see the egg in the nest), the study was a success. The efficacy of rejection sampling in the placement function requires further investigation. Current testing indicates that it improves scene construction by making placements more accurate, though the key issue in this endeavor is the efficiency of the inside predicate.

The placement system creates a complex relationship between predications and entities during placement. An "optimal" order of placement emerges in complex scenes. Deviation from the optimal layout was shown to increase placement time dramatically.

The "optimal" configuration is one in which the entities are placed in order of decreasing number of predication constraints. For example, in the placement of the story scene, the required predications are:

- a person is under a tree
- a nest is in a tree
- an egg is in the nest
- the person can see the nest

In this scene, the nest is the most constrained object, because it is used in a predication with the person (can see), the egg (in), and the tree (in). The least constrained entities are the tree and the person, each involved in only one predication. This forms a "constraint hierarchy."

For an optimal placement run-time, the entities in the scene should be placed from the most to least constrained. Placing the most constrained entities first allows the most legal placement area for the subsequent entities (which, because they are lower in the hierarchy, have naturally less constrained placement areas), which means that the system has to do less sampling and backtracking on average in order to place them. Note that a constraint hierarchy is not necessarily unique for a given scene because multiple entities can be involved in the same number of predications.

The greatest flaw in the system, and the biggest boundary to continued expansion of the project, is the ad-hoc nature of the predicate and object database. Both entities and predicates are created in an ad-hoc fashion; there exist no templates for either, although some share similar features. Because of this, adding new members to either library can be cumbersome. Further, because the predicate functions of each are based on qualitative semantic interpretations of the predicates, it can be quite difficult to write functions that objectively represent the predicates they are meant to.

Our project's success in creating and querying 3D scenes demonstrates both its usability as a situation for reasoning in 3D spaces, and the expressive power of the system in general. Our program was able to quickly and efficiently manage and query over a small library of entities and predicates. The system is black-boxed to outside input, and as such can be used for more than just Epilog. Preliminary work with the 3D printed model and the collaboration with Haptek has shown this. The system, in its current form, holds potential to be of use in a number of projects and experiments that involve 3D space. We hope that the IMS will be put to use as a specialist program, both in Epilog and beyond.

6 Code Tables

obutily.py	
Method	Description
`cast_thru(Object start_obj,Vector endpt, Object end_obj)`	Casts repeated rays from `start_obj` to `endpt`, noting the degradation due to occlusion that occurs along the way stopping when `end_obj` is hit. Returns a value 0-1 indicating the occlusion encountered on the way
`nudge(Object a, Vector pt)`	Returns a points several thousandths closer to a from pt, used in repeated ray casting experiments
`rayCast(Object a, Vector b)`	Shoots a ray from the center of a to b, returns the same as `Scene.ray_cast(start,end)`
`alignMeasure(a, b, top=float(inf), right=float(inf), bottom=float(-inf), left=float(-inf))`	Creates a rectangular prism mesh on top of b that can be used for intersection testing. If no maximum top/bottom/etc... points are specified it will use the most outward points on b's bounding box
`highest(Object obj, char dim)`	Returns the highest vertex in the object (in the dimension (x,y,z) specified)
`lowestPt(Object obj,char dim)`	Returns the lowest vertex in the object (in the dimension (x,y,z) specified)
`getIntersection`	Returns the part of m's mesh that is intersecting with a

`getDiff(Object m, Object a)`	Returns the part of m's mesh that is not intersecting with a
`closest_points(Scene scene, Object a, Object b)`	Returns a tuple containing the closest point on the mesh of a to b, and the closest point to a on b's mesh, in that order
`maxDim(Object [] ary)`	Returns the largest dimension (x,y,z) among all objects in ary
`distance(Vector a, Vector b)`	Returns the distance between a and b (will not work if an object is in the way). Measured without pathfinding.
`glob2Loc(Vector pt, Object obj)`	Returns point pt in `obj`'s local space
`vertsGlob(Object obj)`	Returns an array of `obj`'s vertices in global coordinates
`locs2Glob(Vector[] pts, Object obj)`	Returns the location of the coordinates in `pts` in global coordinates (assuming `pts` are originally in `obj`'s local space)
`loc2Glob(Vecotr pts, Object obj)`	Same as `locs2Glob`, but for only one point
`aInBSpace(Vector pt, Object a, Object b)`	Returns pt (which is in `a`'s local space at the start) in `b`'s local space; works through matrix multiplication
`getVolume(Object obj)`	Returns the volume in BV^3 of the object; underneath this wrapper method are numerous helper methods

Table 6.1: Obutils Methods

predMethods.py			
Predicate	Description	`Query`	`Place()`
`near(A,B)`	determines whether A is close to B	Iterates through the children of a and b and selects the two with the shortest distance between the two meshes. the value returned is the distance relativized to the sizes of the two entities. The result is not proportional to distance, and returns true up to a certain distance and then a result from a steeply graded exponential function thereafter.	Returns a box bounded by the x and y location of the focal object plus and minus the largest dimensions of the two objects.
`under(A,B)`	determines to what extent A is underneath B	draws a temporary object directly under B; the percent volume of A that intersects with this is the return value	If B is the focal object, it returns a bounding area spanning from the bottom of B to 5 BU under B. If A is the focal object, the z boundary is from the top of A to 5 BU above A. In either case, the x any y boundaries are the maximum of the two objects in the respective dimensions.

`in(A,B)`	determines whether A is inside B's bounding box	draws a temporary object around B's bounding box and returns the percent volume of A, or true if more than half of A's volume is in B.	The boundaries in all dimensions are A's location plus or minus B's size in the given dimension.
`inside(A,B)`	determines the extent to which A is inside B	same as in(A,B), but the temporary object is drawn around B's mesh rather than bounding box	The boundaries in all dimensions are A's location plus or minus B's size in the given dimension.
`above(A,B)`	determines the extent to which A is above B	Draws a temporary object in the space above B and returns the percent of A's volume that intersects.	If A is the focal object, it returns a bounding area spanning from the bottom of A to 5 BU under A. If B is the focal object, the z boundary is from the top of B to 5 BU above B. In either case, the x any y boundaries are the maximum of the two objects in the respective dimensions.
`isTouching(A,B)`	determines the extent to which A is touching B	This function works the same as near(A,B), but does not adjust with relative object size and requires objects to be much closer.	Returns a box bounded by the x and y location of the focal object plus and minus the largest dimensions of the two objects divided by ten.

`canSee(A,B)`	determines whether A can see B	(this assumes an "eye" object on A). Ray casting is done from A's eye to points on the meshes of B's children. The percent of successful casts (that reach B) is the value returned. If the cast hits a translucent object, it will continue but will return a reduced value	returns a bounding box surrounding the focal object's position plus or minus 10 BU in all dimensions

Table 6.2: predMethods Methods

Bibliography

Michael S Bernstein, Greg Little, Robert C Miller, Björn Hartmann, Mark S Ackerman, David R Karger, David Crowell, and Katrina Panovich. Soylent: a word processor with a crowd inside. In *Proceedings of the 23nd annual ACM symposium on User interface software and technology*, pages 313–322. ACM, 2010.

Jeffrey P Bigham, Chandrika Jayant, Hanjie Ji, Greg Little, Andrew Miller, Robert C Miller, Robin Miller, Aubrey Tatarowicz, Brandyn White, Samual White, et al. Vizwiz: nearly real-time answers to visual questions. In *Proceedings of the 23nd annual ACM symposium on User interface software and technology*, pages 333–342. ACM, 2010.

Daren C Brabham. Crowdsourcing as a model for problem solving an introduction and cases. *Convergence: the international journal of research into new media technologies*, 14(1):75–90, 2008.

Jeff Howe. The rise of crowdsourcing. *Wired magazine*, 14(6):1–4, 2006.

Juho Kim. Toolscape: enhancing the learning experience of how-to videos. In *CHI'13 Extended Abstracts on Human Factors in Computing Systems*, pages 2707–2712. ACM, 2013.

Anand Kulkarni, Matthew Can, and Björn Hartmann. Collaboratively crowdsourcing workflows with turkomatic. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*, pages 1003–1012. ACM, 2012.

Walter S Lasecki and Jeffrey P Bigham. Interactive crowds: real-time crowdsourcing and crowd agents. In *Handbook of Human Computation*, pages 509–521. Springer, 2013.

Walter S Lasecki, Rachel Wesley, Jeffrey Nichols, Anand Kulkarni, James F Allen, and Jeffrey P Bigham. Chorus: a crowd-powered conversational assistant. In *Proceedings of the 26th annual ACM symposium on User interface software and technology*, pages 151–162. ACM, 2013.