Critique of Human-powered Sorts & Joins

Summary
Marcus et al. (2011) describe using Amazon’s Mechanical Turk (MTurk) systems utilizing crowds for database sorts and joins. Crowdsourcing activities for these types of queries are typical costly in order to ensure accuracy and reliability. The authors present an interface for optimizing these functions in order to determine cost-effectiveness of various methods (simple join, naïve batching, and smart batching). Simple batching presents the MTurk responder, the one agreeing to respond to the request, with a simple interface with a dichotomous yes or no question regarding similarity of images. Naïve batching is similar in nature in that responders are presented with a dichotomous yes or no question, but there are multiple images presented in pairs to which the responder must answer. Smart batching presents the responder with several images to which they must draw connecting lines between images that are similar. Smart batching was quite successful in terms of both accuracy and cost reduction from $67 to about $3.

Furthermore, Marcus et al. (2011) presented an interface for sorting images. The comparison sort interface displayed two groups of random sized boxes to which the responder must rank in order within the designated groups. The ratings sort interface displayed a subset of 10 random sized boxes from a dataset of 40 other boxes. Responders were asked to numerically rank two of the boxes in size given the subset of other boxes. The authors found “that while significantly cheaper, ratings achieve sort orders close to but not as good as comparisons” (Marcus et al., 2011, p. 20).

Critique

Pros
This paper presented several different interfaces to facilitate sorting and joining using the crowd via MTurk. The findings illustrated that using varying interfaces and batching techniques one can drastically reduce the cost of human intelligence tasks (HITs) while maintaining accuracy and reducing latency. Additionally, the authors presented the interfaces clearly and accurately so that others might extend or practically apply the techniques that were outlined herein. Finally, the paper was well organized with a logical flow through the various sections and subsections making the paper easily understood.

Cons
In contrast, there was a disconnect between the various sorting algorithms investigated and the interfaces shown to responders. The authors discussed comparison and rating sorting interfaces but proceeded to present findings regarding hybrid and other experimental algorithms. Additionally, while the authors noted the dataset for the sorting tasks there was no such dataset given for the join tasks. Finally, it was difficult to locate the summary paragraphs for each of the tasks (sorting and joining) as they were not centrally located at the end as one might expect.
**Improvement**
In order to improve this paper, I would suggest moving all of the summary paragraphs to a central location within the document and adding subsections for sort and join tasks. Additionally, adding an addendum containing the (partial) dataset for the join tasks would be beneficial.

**References**

Critique of Max Algorithms in Crowdsourcing Environments

Summary
Venetis et al. (2012) provide strategies and a framework for retrieving the maximum item from a dataset using crowdsourcing. They take a theoretical approach to describe a parameterized family of algorithms in order to select the item believe to be the maximum or best fit. Additionally, they evaluate how these algorithms function with respect to three parameters: quality, cost, and execution time. These three parameters are evaluated separately in order to gauge their individual effectiveness at achieving the goal – that of selecting the perceived max item. The framework developed by the authors presents a conceptualized model that is abstract such that it is domain agnostic for practical application across varying domains.

Critique

Pros
Venetis et al. (2012) did an exceptional job explaining the algorithms used in their research such that it could be repeated and extended. Additionally, it is appreciated that the assumptions made were explicitly stated such that one can make judgment with regard to the validity or lack thereof. Finally, the logical progression of the dialog presented makes the paper easily understood.

Cons
Unfortunately, it was difficult to understand the experimental method by which the results were observed. There was no discussion with regard to the experimental or simulation environment that the authors discussed. Additionally, this lack of environment description causes portions of the paper to be unrepeatable and therefore unverifiable. Finally, the lack of description also calls into question the validity, reliability and accuracy of the research.

Improvement
The major issues with this paper could be easily solved with an accurate description of the experimental process by which the empirical results were obtained. This would allow further researchers to validate the results and give more credence to the results presented by the authors.

References