Human-powered Sorts and Joins

Summary

The authors describe how they have integrated crowds into a declarative workflow engine, Qurk that they developed to address the issues that crowdsourcing markets like Amazon’s Mechanical Turk (MTurk) have. The crowdsourcing markets have increasing workflow and thus in order to manage the volume of this workflow effectively, they are looking into redesigning the current database system. The following two specific issues are introduced in the paper: re-implementing database operations such as filtering, sorting, and joining needed due to absence of optimization capability in the existing database system; query optimization. The paper introduces Qurk as a solution to these problems by its being able to run queries over a crowd of workers with crowd-based filter, join, and sort operator that optimize the tasks. The authors particularly focused on how to use humans to compare items and ran experiments to see how join, sort, and combination of join and sort operator implemented in Qurk work. They found that the optimizations they used significantly reduce the overall cost of running sorts and joins on the crowd.

Pros

In the introduction, the authors give a nice background of the research study by clearly explaining issues that the crowdsourcing markets are having and how the system, named Qurk can address the issues. They also provide the very detailed explanation of architecture of the Qurk system with the clear diagram. Lastly, an implementation of each operator with an experiment is structured and presented in a same way that helps the readers understanding and comparing each operator easily.

Cons

In the third paragraph in the section 3.3.4 “Feature Filtering”, the authors hypothesize why combining feature filtering reduces the cost and the error rate, however, the hypothesis is not tested anywhere in the paper. Also, in the section 3.2 “Feature Filtering Optimization”, the authors assume that all filter features are uncorrelated and that the filters do not emit the value UNKNOWN. Making assumption in order to simplify the analysis is not good way to convince the results. Lastly, the authors do not talk about the methodology in the paper in the beginning of the paper.

How to improve the paper further

In the paper, there is no keywords section. Adding keywords section in the paper would be helpful for the reader to what to look for in the paper. Also, in the beginning of the section 5, the title “END TO END Query” doesn’t give the reader a clear meaning in terms of the operator implementations since it is not mentioned in the abstract or introduction. It would be better if the title changes to something like “Complex query of joins and sorts” that could reflect the content under. Lastly, it would be better if the result in the section 5.2 had been presented in a visual representation especially for the feature extraction part.
Max Algorithms in Crowdsourcing Environments

Summary

The authors looked into the issues in retrieving the maximum item from a set in crowdsourcing environments. As a solution, the authors developed parameterized families of max algorithms, proposed strategies to select proper max algorithm parameters under various metrics of conditions, and evaluated the algorithm and strategies in regard to the tradeoff between the following three: quality, cost, and time through simulations and analysis. With results of experiments, the authors provided guidelines for selecting max algorithms and strategies for different scenarios.

Pros

The content of the paper is well organized and logically presented, which helped the reader to understand the concepts and ideas, and to follow the each section easily. Also, in the introduction, what is the problem the researchers want to investigate, how the researchers want to solve the problem, and what the procedures of whole research study will be are clearly described. Lastly, the PRELIMINARIES section provides nice basis knowledge for the readers to understand the following sections.

Cons

In the introduction, the way the researchers introduce their research questions is a little bit wordy. What problems they worked on and what would be the answers to those research questions seem obvious and self-explanatory that the authors didn’t need to use the comparison with using the other questions that they said they would not use (‘instead of just asking the question... we ask...’). Also, the contribution part at the end of the introduction section seems a little bit redundant. Lastly, the discussion and the future work section are missing.

How to improve the paper further

In the section 2.2 Steps, it would be nice if there had been a diagram representation for the process steps. Also in the section 3, it would be better if there is a clear label for the algorithm tables in the case of referring in anywhere in the paper. In the section 4.5 VaryingS, the head for the table should be a kind of strategy not the “algorithm 3” since the VaryingS is one of the strategies. The figure 1, it would be a better representation if the two graphs had been overlaid in a same graph since they have the same X and Y scales. Lastly, in the section 6 Experiments, each result is presented in a box and the description is added right under, which makes a little bit confusing to follow. If they are reorganized in a way that clearly separate them, it would make the result section better and clearer.