|  |  |
| --- | --- |
| **Masked Reviewer ID:** | Assigned\_Reviewer\_2 |
| **Review:** |  |

|  |  |
| --- | --- |
| **Question** |  |
| Is the submission relevant to the DB track? | Yes |
| Overall Rating | Neutral |
| Top 3 Strengths | 1) The paper reads well 2) Definitions and results clearly stated and correct 3) Experiments support the results claimed. |
| Top 3 Weaknesses | 1)it is not clear which modeling choices are arbitrarily made by the authors and which are common in the literature. [it is clearly mentioned in p2c1, last line.] b) the strength and weaknesses the model are not detailed. [added in the new section] 2) the theoretical results are very easy. Some informal justifications for the heuristics seem rather lousy to me. 3) The results are quite interesting but the originality is not tremendous. |
| Detailed Comments | Summary: This paper investigates the computation of pareto optima for partial orders (maxima in presence of multiple criteria) in a crowdsourcing environment. The emphasis is on minimizing the number of questions before an answer is returned. Here a question is basically a comparison between two items. The comparison can return 3 distinct outputs: the first item is larger, smaller or incomparable.  This paper is the first one to investigate pareto optima for partial orders. This is actually the main novelty of the paper w.r.t. the more classical skyline literature. But the definition of dominance (hence of pareto optima) on partial orders presents some poor characteristics: in particular dominance is not transitive. It is not clear from the paper whether anybody considered such a definition before or whether it is the author's invention, in which case it should be motivated.  [it is mentioned in p2c1 last line that [13] considers the same definition, we also emphasized the citation in p2c2 middle of page]  Another surprising issue is the apparent absence of ground truth: the truth will be considered to be the output of the algorithm; and consequently the paper does not at all consider the issue of obtaining accurate results with their crowdsourcing algorithms. IMHO, this differs from traditional practice in crowd data sourcing in the DB community. [Addressed in the new section]  Furthermore, because of some arbitrary choice on resolving contradiction, this implies that the algorithms, as they are presented, may return different outputs when run on the same data (non deterministic behaviour). This fact is completely ignored by the paper that only investigates the number of comparisons before some result is returned, whatever the result may be. Note that actually exploiting transitive closure increases vulnerability to errors.  [Addressed in the new section]  Details: p2 col2 bottom (also p5 col2 top): 'due to the lack of explicit attribute representation' -> not really; rather due to criteria following a partial order instead of linear order (strict or bucket) [addressed by mentioning the part “, while preference and skyline ….”] p3 col 2 bottom 'preference of multiple agents has always been a fundamental problem in social choice' -> one could also quote much older work like Condorcet [Ref. added] their predication -> prediction [fixed]  p5 def 3: 1) item 3 is improperly defined; y>x is not ruled out IFF \not\exit c'\in C: x>\_c y \in R^+(Q) AND \exist c. x~\_cy\ntoin R^+(Q). And then ii and iii imply i so case i is redundant. [the current definition looks more promising]  All over the paper (except Th1), I would replace 'the proof is omitted due to space limitations and can be found...' with a sentence saying it can be deduced trivially from the definition. Though other reviewers might differ on their definition of trivial so I will not stress the issue if I am the only one of this opinion. It is a bit disappointing that the only non trivial result (theorem 1) is left for the technical report. I had the curiosity to look for such a report on the web and found it. But I think the proof of this theorem could largely be improved-simplified by reformulating explanations. In fact only the part iii-2 is really interesting, and even there explanations lack of rigor. In particular I believe you should stress that you consider questions x?y to be syntactically asymmetric (even though its result is the same as y?x). And if you indeed do so, stick to it. (CURRENTLY YOU SPEAK OF A COMPARISON w\_{m-1}?v that might never be performed if one performed v?w\_{m-1} instead).  You should make clear what you are trying to do (how you ground your reccurence): what does the replacement achieve that guarantees a not-longer question will eventually be obtained that has no non-candidate questions? [reminder: check]  p6 col2 top 'Hence it is a good idea to only select...' This is an overstatement because you do not prove that ALL sequences using only candidates are better. Only that there exists one. You have no guarantee to find it.[I agree, marked the sentence for removal] p6 col2 botttom Rule 2 : This choice is quite arbitrary: one could as well change y>\_c z into y\sim\_c z or remove x\sim\_c z , etc... [addressed in the new section] p8 col1 the independance assumption clearly does not hold. So why make it? [IID assumption is a reasonable assumption] Is the scoring function yours or is it inspired from somewhere? [it is ours] 5.1 What about contradictions? [it also is ours]  Abstract: 'This is the first study on crowdsourcing Pareto optimal object finding' is not actually true -> there have been other such studies, but this one is the first to study crowdsourcing Pareto optima (afaik) FOR PARTIAL ORDERS. Indeed, skylines are the same as pareto optima in the case of total orders.[addressed]  By the way, since you cite the Lofi et al. EDBT13 paper, mentioning that it only deals with missing attribute value, you may also be interested by the PODS'15 paper Skyline Queries with Noisy Comparisons. This paper deals with skyline queries through comparisons in a crowdsourcing framework, the main differences with your paper being that 1) it considers the issue of accuracy 2) it does not consider partial orders.   I am not sure if using the terminology bucket order (especially, without defining it) is a good idea. The term is rarely used. I believe that most people assume total=linear orders to admit ties, and speak of strict orders when this is not the case (cf wikipedia, for instance).[addressed] |
|  |  |

|  |  |
| --- | --- |
| **Masked Reviewer ID:** | Assigned\_Reviewer\_3 |
| **Review:** |  |

|  |  |
| --- | --- |
| **Question** |  |
| Is the submission relevant to the DB track? | Yes |
| Overall Rating | Accept |
| Top 3 Strengths | 1. A novel approach of finding Pareto-optimal objects when explicit attribute values are not available for comparision is presented.  2. The question selection framework and strategies proposed significantly reduce the total number of questions used.  3. The presentation is clear and comprehensive. |
| Top 3 Weaknesses | 1. The rules for resolving unusual contradictions in question outcomes in Section 3.2 are not solid.  2. The general framework describes a sequential execution from question selection to result derivation, which could lead to large latency in practice.   3. The experiments using real crowdsourcing data did not show how the number of attributes will affect the performance. |
| Detailed Comments | 1. The paper is distinguished from related works by addressing the problem of finding Pareto-optimal objects, especially under the scenario where no explicit attribute values are available. It proposes effective question selection strategies, macro-ordering and micro-ordering, that enormously reduce the total number of questions required.   2. Introduction(Section 1) and General Framework(Section 3) are comprehensive with detailed explanation of the weakness of existing work, the motivation and the rationality behind strategies proposed in this paper.The experiments on both simulated data and real data further shows efficiency and scalability of various strategies.  3. In Section 3.2, when dealing with unusual contradictions in question outcomes, the crowdsourcing workers' quality should be considered, instead of directly skipping questions when controversial answers are obtained without any confidence. Although evaluating the cost only by the total number of questions makes sense, the sequential framework can lead to high latency in practice. Asking ask multiple questions concurrently could help, and the current strategies should be adjusted accordingly.  More experiment results based on real data will also be useful to show the effect of the number of attributes, because there can be more than 3 attributes when choosing Pareto-optimal objects in real life. |