View Reviews

Paper ID
31

Paper Title
Continuous Monitoring of Pareto Frontiers on Partially Ordered Attributes for Many Users

Track Name
Regular Papers

REVIEWER #1

REVIEW QUESTIONS

1. Relevance: Is this paper relevant to EDBT?
Yes

2. Correctness: Is this paper technically correct?
Yes

3. Originality: Are the ideas and solutions novel?
Medium - some aspects of the work are novel.

4. Presentation: Is the paper readable and well presented?
Fair - the presentation has minor issues and should be improved. However, overall the presentation is acceptable.

5. Impact: What is the impact of this paper to the database community?
Medium - the paper is incremental and/or will have limited impact with respect to citations or follow-up work.

6. Strong Points: Please list the three strongest points of the paper:
S1. The paper is well written (for most parts) and the authors explore an interesting problem.

S2. The proposed algorithms are well designed and nicely leverage properties of the partial order and pareto frontier.

S3. The experimental section validates the claims made earlier in the paper.

7. Weak points: Please list the three weakest points of the paper:
W1. The authors do not fully explore properties of the problem studied. In particular, the situation of the work in the context of incremental algorithms is missing.

W2. It is left vague as to whether the proposed algorithms have certain optimality conditions, e.g., whether they only recompute the absolute minimum of information.

W3. The explanation of the approximate common preference tuples is inadequate. This should be revised substantially.

8. Review summary: Please summarize the key points in terms of contributions, novelty, technical depth, presentation, and impact that lead to your overall rating:
See comments above and detailed comments below.

9. Detailed Comments: Please provide detailed comments (contributions, pros and cons, concerns, errors, etc) to the authors:
Overall, I liked the problem considered in this paper. However, from a conceptual point of view, one would have expected to see the problem of dynamically maintaining the target users in terms of the general theory of incremental algorithms. For example, is this problem bounded in the incremental sense? Is there a notion of optimality? That is, does the proposed method only recomputes the absolute necessary parts of the frontier?

When going to approximation, it is a bit unclear what the holy grail would be? That is, can the approximation problem be phrased more generally in terms of some approximation desiderata and again, show that the proposed algorithm satisfies those. In summary, a more in depth exploration could have been carried out.

Said this, the underlying ideas are sound and natural. While most parts of the paper are reasonably well written and clearly illustrated by examples, Section 6.1., which deals with the notion of approximate common preference tuples is very hard to digest: It took me several passes to understand what definition 6.1. means. I suggest that the authors start by describing the algorithm, give a detailed examples, and only then, if needed, formalize the algorithmic notion of approximate common preference tuples.

The parts related to algorithms FilterThenVerify and its approximate counterpart are built on interesting observations. By contrast, the parts related to the similarity functions and sliding windows are a bit tedious to read and of less technical depth. This could perhaps be all trimmed down a bit? Instead, a more general discussion as to what properties the similarities function must satisfy for the algorithm to work would be more interesting.

A general remark is that it seems that an underlying assumption is that the user specify enough information to be of user when new objects arrive. What happens if insufficient information is stored in the partial orders? Is it realistic to assume that user specify these orders up front, without knowing which objects will arrive later on? I feel that this needs more discussion and/or experimental validation that this is indeed a realistic assumption.

A suggestion is to also include the complexity analysis of FilterThenVerity (similarly to what was done for the Baseline approach) showing the theoretical advantage compared to the Baseline approach.

10. Confidence: How confident are you about your review?
Medium - I am familiar with the area.

11. Overall rating
Weak Accept - I would not mind this paper being accepted, but will not fight for acceptance.

REVIEWER #2

REVIEW QUESTIONS

1. Relevance: Is this paper relevant to EDBT?
Yes

2. Correctness: Is this paper technically correct?
Yes

3. Originality: Are the ideas and solutions novel?
Medium - some aspects of the work are novel.

4. Presentation: Is the paper readable and well presented?
Fair - the presentation has minor issues and should be improved. However, overall the presentation is acceptable.

5. Impact: What is the impact of this paper to the database community?
Medium - the paper is incremental and/or will have limited impact with respect to citations or follow-up work.

6. Strong Points: Please list the three strongest points of the paper:
1. The paper has adequate examples and explanations.

7. Weak points: Please list the three weakest points of the paper:
1. The problem requires further motivation and justification.
2. The technical contribution of the work is not that interesting. For example, the authors discuss approximation but it is not clear if they have any provable guarantees.
3. Experimental results are weak.

8. Review summary: Please summarize the key points in terms of contributions, novelty, technical depth, presentation, and impact that lead to your overall rating:
This paper formulates the problem of continuous monitoring
of Pareto frontiers: given a number of users and continuously
arriving new objects, for each newly arrived object,
discover all users for whom the object is Pareto-optimal. The studied problem is intellectually interesting, although I am not sure how practical it is (no real datasets in the experiment corroborates my concern).

Additionally, it is not clear why and how this problem is different from Skyline over data streams [15]. [15] is cited in the paper, but no discussion is presented.

9. Detailed Comments: Please provide detailed comments (contributions, pros and cons, concerns, errors, etc) to the authors:
The authors study the problem of continuous object dissemination, which
is formalized as finding the users who approve a new object in Pareto-optimality. The authors design efficient algorithms to solve the problem and conduct experimental analysis.

One fundamental concern that remains unaddressed is, it is not clear why and how this problem is different from Skyline over data streams [15]. [15] is cited in the paper, but no discussion is presented. Why these works could not be used to solve the problem, or at least implement as baselines.

Even though the paper is well written in general, I feel the authors should make some effort to simplify it. For example, Section 5 has a long discussion on similarity measures. It is not clear if all these discussion is needed.

The technical contribution of the work requires some clarifications. It is not clear how the presented theorems and proofs are helping in the overall solution design. Similar comments go for the approximate solution which does not seem to have any guarantee.

Experimental analysis requires an overhaul.
1. The authors should at least have some simulation results that truthfully mimic the problem settings.
2. Appropriate baseline algorithms (such as, [15]) should be implemented.
3. The authors vary the number of dimension just between [2-4]. This is clearly not adequate for a large scale study.

10. Confidence: How confident are you about your review?
Medium - I am familiar with the area.

11. Overall rating
Weak Reject - I would prefer this paper being rejected, but will not fight for rejection.

REVIEWER #3

REVIEW QUESTIONS

1. Relevance: Is this paper relevant to EDBT?
Yes

2. Correctness: Is this paper technically correct?
Yes

3. Originality: Are the ideas and solutions novel?
Medium - some aspects of the work are novel.

4. Presentation: Is the paper readable and well presented?
Good - the presentation is good, all aspects (motivation, methodology, solution, evaluation, and impact) are described well.

5. Impact: What is the impact of this paper to the database community?
Medium - the paper is incremental and/or will have limited impact with respect to citations or follow-up work.

6. Strong Points: Please list the three strongest points of the paper:
1/ The paper falls in a well studied research area, but the problem definition and setting is quite novel, and the paper ultimately proposes interesting ideas.

2/ The paper is well written, with detailed examples making it easy to follow.

3/ Despite the lack of real data, the authors make their best by constructing realistic semi-synthetic datasets.

7. Weak points: Please list the three weakest points of the paper:
1/ The general research area is not that fresh as it has been extensively studied in the past.

2/ There exist many experimental parameters that could have been investigated, but are not.

3/ The presentation of certain notions is a bit convoluted, and could be simplified.

8. Review summary: Please summarize the key points in terms of contributions, novelty, technical depth, presentation, and impact that lead to your overall rating:

The paper deals with the problem of maintaining skylines over streaming objects, where the preferences in each attribute are partial orders (PO). It makes three contributions:
- it considers the problem of clustering users based on their PO preferences, and defines appropriate similarity metrics
- it investigates the implications of maintaining a skyline for groups of users with PO preferences
- the previous point becomes less trivial when the PO preferences for a group are approximated, i.e., by not including all pairwise preference relationships
- the cases of streaming data and sliding windows are covered.

Among them, the first three are the strongest contributions, with the case of streaming data being largely based on previous work. Overall, the technical depth and novely is quite substantial. The impact of this work is however less dramatic, as it follows a long line of papers in this research direction. Presentation is excellent.

9. Detailed Comments: Please provide detailed comments (contributions, pros and cons, concerns, errors, etc) to the authors:

- In the second half of Line 6 in UpdateParetoFrontierU in Algorithm 2, the update of C_o' should be removed.

- The discussion of how to construct the approximate preferences for a group is a bit complex, although the underlying procedure is not. Consider revising.

- It seems that the similarity metrics in Sec. 6.3 are more elaborate than those in Sec. 5, since they consider the frequency of pairwise preferences. Can't the Sec. 6.3 metrics be used instead of those in Sec.5? What is then the implication — see also next comment.

- There are interesting experimental settings that are not considered.
- The first is the case of the different similarity metrics. What is the implication of using each?
- The parameters theta1 and theta2 of the approximate skylines are not investigated at all, in terms of running time and approximation quality.
- The parameter h of the clustering algorithm is investigated only in terms of approximation quality. What is the running time for the different h values used?

- A general comment about concepts. It is argued throughout the paper that the skyline operator is not the same as the Pareto Frontier, which is not accurate. I understand that the authors mean that the majority of skyline paper deal with totally ordered attributes. But it is should be noted that the handful of papers that discuss PO attributes also call the skylines. Moreover, the ideas and algorithms for the streaming/sliding window case in this paper do not really differ from those for the totally ordered case.

10. Confidence: How confident are you about your review?
High - I am an expert in the area of the paper.

11. Overall rating
Accept - I will support acceptance of this paper.