|  |  |  |
| --- | --- | --- |
| **No** | **Issues** | **Date** |
| 1. | **Title**  I prefer to use word “visualization” instead of “view” as a title. When I first time read the title of Humaira’s paper I thought that word “View” means database’s view not visualization.  Using word “visualization recommendation” is clear and also can increase possibility to be in the first page on the search engine. More people find this work, more people read and more people may cite it. |  |
| 2. | **New dataset**  Looking another dataset to show positive example of diverse views in the set.  Should I looking for medical dataset again? | 4 – 6 June  I tried since last Friday but..  *(I am not sure about this, it easy but hard honestly to find the dataset that interesting but also match for our work)* |
| 3. | **Attributes vs Measure definition**  As my understanding and based on my experiment. While the data type is string it is usually defined as attribute A, while it number which can be measured will be considered as measure M.  However, last meeting you said that ‘Age’ should be in the X axis which means attribute not measure.  It seems this may can leads to us to question, how to define an attribute or measure correctly by data-driven. | 4 – 6 June |
| 4. | **Another distance function for importance score computation**  SeeDB supports a variety of distance functions: Earth Movers Distance, Euclidean Distance, Kullback-Leibler Divergence, and Jenson-Shannon Distance.   SeeDB (paper 2015) used Earth Movers Distance as the default distance but they also discuss results for other distance functions.  I used Euclidean distance function in our experiments because this distance is the easiest one for me. Maybe, I should learn and understand other distances as well.  While the distance function is changed, it may change the maximum importance score bound. | 7 – 8 June |
| 5. | **Maximum importance score bound**  Until now, we do not have any mathematically proven that shows the maximum bound is equal to √2. | 7 – 8 June |
| 6. | **Another distance function for diversity computation**  Why Jaccard distance, is this distance the best option? | 11 – 12 June |
| 7. | **Dealing with dataset which has a hierarchy (OLAP)**  This issue have been discussed since several months ago but not yet solved.  This issue also related to issue no 2. We should looking another dataset that may interesting and has hierarchy. | 11 – 12 June |
| 8. | **Auto-remove similar attributes by applying correlation**  SeeDB has offline and online precomputation mechanism such as removing correlated attributes. While the dataset has so many attributes like superstore. It may has duplicate attributes such as “suburb” and “post code”. We should remove one of those attributes in advanced.  We did not consider yet about this thing. | 11 – 12 June |
| 9. | **Visual Encoding diversity**  Diversity in this work focus on only context of the A, F, M. We did not mention about the types of chart. *I am not sure if this still in the track*. However, different type of chart seems more interesting rather than only present all visualizations in bar chart. In fact, chart types also depends on the data itself. | 13 June  *I am not sure if this still in the track* |
| 10. | **Objective function**  Why use 0.5 as the default of λ parameter? Is 0.5 fair enough to balance between importance and diversity as a default?  While read paper “On Query Result Diversification”, it used 0.7 as the default of λ. | 14 – 15 June |
| 11. | **Max-Min Pruning Greedy**  In each Greedy iteration, unselected view is executed one by one and the minimum bound is updated using the maximum importance score of executed views. It will help DiVE-Greedy to prune more queries. | 14 – 15 June |
| 12. | **Why only Greedy and Swap**  There are several heuristics algorithms for optimization but why we chose Greedy and Swap?  Heuristics generic approaches : greedy, interchange, bound based heuristics (e.g. lagrangean heuristics), tabu search, simulated annealing, population heuristics (e.g. genetic algorithms) | 14 – 15 June |
| 13. | **Query load**  Two types of query load in the experiments:   1. Compare between two subset (e.g., disease vs. no disease) – *targeted/ we know what we want to do* 2. Compare between one subset to whole dataset (get all subsets from dataset then compare to whole dataset)   Do we need to compare among all possible subsets in the dataset?  For instance, Flights dataset has attribute = ‘carrier’ and there are a lot of carrier (e.g., AA, US, XX, UU), each carrier can be one subset.  While we only compare between each subset to whole dataset, it only show the trend of each subset compared to whole dataset.  However, It seems interesting while we compare between each subset to others. But it increases the number of combination significantly.  It might rely mainly on user intention. For instance, user want to compare the performance of carrier AA vs XX in terms of arrival delay. Of course it depends on user want. We did not mention about this scenario in our experiments. This issue is also related to “User intention” issue below. | 18 – 19 June |
| 14. | **User intention**  There are three types of intention by task as follows (Toward Visualization Recommendation System’s paper):   1. Exploratory 2. Comparative 3. Targeted   We did not mention detail about these three types of user intention and how to handle it. | 18 – 20 June |
| 15. | **Prediction Interval**  I really not sure pruning by prediction interval is feasible if our dataset has large number of views, for instance, more than 1000 views. By executing 50 samples of views with number of views more than 1000 does not make sense. While the number of views are very large (e.g., 1000 views), it seems approximation may be promising approach. |  |
| 16. | **Unexplored area (coverage)**  We did experiments mostly on heart dataset which has small size and small number of subset. Flights dataset has so many subsets but it seems not interesting. While we have large dataset such as Flights, even pruning has been used it still take a lot of time.  It seems the sampling approach or approximation is promising, instead of generating all possible views, we use sampling and show to the user the coverage of the results. | 18 – 20 June |
| 17. | **Fixing Reference format**  I need to fix the reference format, especially the author name. | 18 – 20 June |