**Summary.**

This paper tackles the problem of mining potential high-utility itemsets over uncertain databases. To tackle this problem, the authors propose two algorithms, PHUI-apriori and PHUI-list. The former is extended by the classical Apriori framework, and the latter devises the probability-utility list structure and the depth-first searching framework to avoid the redundant candidate generation. Although the problem is interesting, the paper does not motivate the problem very well. Especially, in the problem definition, the authors directly use the definition of expected support-based frequent itemsets without any justification, which is not very convincing why the definition of probabilistic frequent itemsets is not considered. Furthermore, the proposed solutions are simple combination of existing techniques of both high-utility itemset mining and expected support-based frequent itemset mining. Therefore, the technical novelty of this paper is not high. Finally, experiments only test the synthetic data instead of real uncertain datasets, even if the authors claim to use the three real-life datasets, which actually are synthetic datasets and are generated by the corresponding real deterministic datasets.

**Strong points.**

S1. The problem studied in this paper is interesting.

S2. Experiments are conducted.

**Weak points.**

W1. This paper does not motivate the problem well.

W2. The problem definition needs more justifications.

W3. There are several missing related reference papers.

W4. The technical novelty is not enough.

W5. Experiments only test the synthetic data, and three real-life datasets, which are claimed by the authors, are not real uncertain datasets.

**Detailed Comments**

The details of comments are listed as follows.

D1. As mentioned above, this paper studies an interesting problem of mining potential high-utility itemsets over uncertain databases. However, the definition of this problem is not needs more justifications. Especially, the authors directly use the definition of expected support-based frequent itemsets. Why the definition of probabilistic frequent itemsets is not considered? Furthermore, the definition of probabilistic frequent itemsets can provide the more exact description about uncertainty. Therefore, it would be interesting to consider the definition of probabilistic frequent itemsets rather than the definition of expected-support-based frequent itemsets in this case. Please provide provide more justification what are the differences that the aforementioned two definitions of frequent itemsets over uncertain databases are used in this paper.

D2. In Section I, the motivation example of the transaction including purchase probabilities of customers is not convincing. In particular, it is very confusing how to obtain the purchase probabilities and the corresponding quantities of the items. Furthermore, what applications need to discover the potential high utility itemset over uncertain databases in real life? Please provide more realistic examples for the problem studied in this paper.

D3. This paper ignores several related works of mining high utility itemsets. The missed representative related researches are shown as follows.

[R1] R. Chan, Q. Yang, Y.Shen. Mining High Utility Itemsets. ICDM 2003: 19-26.

[R2] C. Ahmed, S. Tanbeer, B. Jeong, Y. Lee. Efficient Tree Structures for High Utility Pattern Mining in Incremental Databases. IEEE Transactions on Knowledge and Data Engineering, 21(12): 1708-1721, 2009.

[R3] C. Wu, B. Shie, V. Tseng, P. S. Yu. Mining Top-K High Utility Itemsets. KDD 2012: 78-86

[R4] V. Tseng, C. Wu, P. Fournier-Viger, P. S. Yu. Efficient Algorithms for Mining the Concise and Lossless Representation of High Utility Itemsets. IEEE Transactions on Knowledge and Data Engineering, 27(3): 726-739, 2015.

D4. In this paper, I cannot find any correctness and computational analysis of time/space complexity. In particular, I am wondering the space complexity of the PU-list structure. Please provide the analysis of correctness and time/space complexity of proposed algorithms.

D5. Since the time complexity of calculating the expected support of an itemset in uncertain databases is equals to that of computing the support of an itemset in deterministic databases, I am wondering whether the existing studies of mining high utility itemsets can be directly extended to solve this problem studied in this paper. If so, please explain what the new contributions of this paper are. Otherwise, please clarify why the existing studies cannot address this problem.

D6. In Section 5, experiments only test synthetic datasets. Although the authors claim to use the three real-life datasets, they are only real deterministic datasets rather than real uncertain datasets. In fact, the authors only assigned the probability value to each transaction in the real deterministic datasets then transform these deterministic datasets to the uncertain datasets. In other words, all experiments in this paper are not conducted in real uncertain datasets. Therefore, it would be more convincing if real datasets are used.

D7. In the experiments, the authors only compared your own work. However, it would be interesting to compare your work with the variants of other recent related work, e.g., References [12, 14, 15], which can work if the function of checking the utility values is integrated.

D8. In the experimental setting, it is not clear that how the uncertainty is assigned into each transaction. In addition, what probability distributions are followed in the probability assignment? Please clarify the details that the uncertainty is assigned into each deterministic database.

D9. In Table 5, it would be better that the uncertainty information of each dataset should be shown.