IV.CONCLUSION

In this paper, we have studied the problem of top-k high utility item sets mining, where k is the desired number of high utility item sets to be mined. Two efficient algorithms TKU (mining Top-K Utility item sets) and TKO (mining Top-K utility item sets in One phase) are proposed for mining such item sets without setting minimum utility thresholds. TKU is the first two-phase algorithm for mining top-k high utility item sets, which incorporates five strategies PE, NU, MD, MC and SE to effectively raise the border minimum utility thresholds and further prune the search space. On the other hand, TKO is the first one-phase algorithm developed for top-k HUI mining, which integrates the novel strategies RUC, RUZ and EPB to greatly improve its performance. Empirical evaluations on different types of real and synthetic datasets show that the proposed algorithms have good scalability on large datasets and the performance of the proposed algorithms is close to the optimal case of the state-of-the art two-phase and one-phase utility mining algorithms

REFERENCES

[1] K. Chuang, J. Huang, and M. Chen, "Mining top-k frequent patterns in the presence of the memory constraint," VLDB J., vol. 17, pp. 1321–1344, 2008
[2] J. Han, J. Pei, and Y. Yin, "Mining frequent patterns without candidate generation," in Proc. ACM SIGMOD Int. Conf. Manag. Data, 2000, pp. 1–12
[3] J. Han, J. Wang, Y. Lu, and P. Tzvetkov, "Mining top-k frequent closed patterns without minimum support," in Proc. IEEE Int. Conf. Data Mining, 2002, pp. 211–218
[4] Y. Lin, C. Wu, and V. S. Tseng, "Mining high utility itemsets in big data," in Proc. Int. Conf. Pacific-Asia Conf. Knowl. Discovery Data Mining, 2015, pp. 649–661
[5] V. S. Tseng, C. Wu, B. Shie, and P. S. Yu, "UP-Growth: An efficient algorithm for high utility itemset mining," in Proc. ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2010, pp. 253–262
[6] J. Yin, Z. Zheng, L. Cao, Y. Song, and W. Wei, "Mining top-k high utility sequential patterns," in Proc. IEEE Int. Conf. Data Mining, 2013, pp. 1259–1264
[7] S. Krishnamoorthy, "Pruning strategies for mining high utility itemsets," Expert Syst. Appl., vol. 42, no. 5, pp. 2371–2381, 2015.
[8] S. Petrovic, M. Osborne, and V. Lavrenko, "Streaming First Story Detection with application to Twitter," in Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics, 2010, pp. 181–189
[9] S. Kumar, H. Liu, S. Mehta, and L. V. Subramaniam, "From tweets to events: Exploring a scalable solution for Twitter streams," arXiv preprint arXiv:1405.1392, 2014.
[10] C. Aggarwal and K. Subbian, "Event Detection in Social Streams," in Proceedings of SIAM International Conference on Data Mining (SDM), 2012, pp. 624–635.