

Mining Sequential Association Rules Efficiently by Using Prefix Projected Databases

Yi-Chun Chen* and Guanling Lee

Department of Computer Science and Information Engineering,
National Dong Hwa University,
Hualien 974, Taiwan
{divien, guanlingl}@gmail.com

Received 4 March 2011; Revised 12 April 2011; Accepted 15 May 2011

Abstract. The mining of sequential patterns has been studied for several years, however, there is little studies pay attention to mining of sequential association rules despite such rules also providing valuable knowledge about many real applications. The sequential association rule represents the concept that a set of items usually occur after a specific order sequence. In this paper, the idea of sequential association rule is introduced and two algorithms, GSAR and PSAR algorithms, are proposed to discover these hidden knowledge. Moreover, experiments are performed on both synthetic and real datasets to show the benefit of our approach.

Keywords: Component, sequential association rule, projected database method, data mining

References

- [1] R. Agrawal, T. Imielinski, A. Swami, "Mining Association Rules between Sets of Items in Large Databases," in *Proceedings of 1993 ACM-SIGMOD International Conference Management of Data(SIGMOD'93)*, Washington, DC, pp. 207-216, 1993.
- [2] R. Agrawal and R. Srikant, "Fast Algorithm for Mining Association Rules," in *Proceedings of 1994 International Conference Very Large Data Bases(VLDB'94)*, Santiago, Chile, pp. 487-499, 1994.
- [3] J. Han, J. Pei, Y. Yin, "Mining Frequent Patterns Without Candidate Generation," in *Proceedings of 2000 ACM SIGMOD International Conference Management of data(SIGMOD'00)*, Dallas, Texas, United States, pp. 1-12, 2000.
- [4] S. K. Tanbeer, C. F. Ahmed, B.S. Jeong, Y.K. Lee, "Efficient Single-pass Frequent Pattern Mining using A Prefix-tree," *Information Sciences*, Vol. 179, No. 5, pp. 559-583, 2009.
- [5] J.S. Park, M.S. Chen, P.S. Yu, "An Efficient Hash-based Algorithm for Mining Association Rules," in *Proceedings of 1995 ACM-SIGMOD International Conference Management of Data(SIGMOD'95)*, San Jose, Ca, pp. 175-186, 1995.
- [6] A. Savasere, E. Omiecinski, S. Navathe, "An Efficient Algorithm for Mining Association Rules in Large Databases," in *Proceedings of 1995 International Conference Very Large Data Bases(VLDB'95)*, Zurich, Switzerland, pp. 432-443, 1995.
- [7] J.Pavon, S. Viana, S.Gomez, "Matrix Apriori: Speeding Up the Search for Frequent Patterns," in *Proceedings of 2006 International Conference Databases and Applications 2006*, Austria, pp. 75-82, 2006.
- [8] J. Pei, J. Han, H. Lu, S. Nishio, S. Tang, D. Yang, "H-mine: Hyper-structure Mining of Frequent Patterns in Large databases," in *Proceedings of 1st IEEE ICDM International Conference Data Mining(ICDM'01)*, San Jose, California, USA, pp. 441-448, 2001.

*Correspondence author

- [9] J. Liu, Y. Pan, K. Wang, J. Han, "Mining Frequent Item Sets by Opportunistic Projection," in *Proceedings of the 8th ACM SIGKDD International Conference Knowledge Discovery and Data Mining(SIGKDD'02)*, Madison, Wisconsin, pp. 229-238, 2002.
- [10] C. K.-S. Leung, Q. I. Khan, T. Hoque, "CanTree: A Tree Structure for Efficient Incremental Mining of Frequent Patterns," in *Proc. 5th IEEE ICDM Int. Conf. Data Mining(ICDM'05)*, Houston, Texas, USA, pp. 274-281, 2005.
- [11] A. Ghoting, G. Buehrer, S. Parthasarathy, D. Kim, A. Nguyen, Y.K. Chen, P. Dubey, "Cache-conscious Frequent Pattern Mining on Modern and Emerging Processors," *International Journal on Very Large Data Bases (VLDBJ)*, Vol. 16, No. 1, pp. 77-96, 2007.
- [12] L. Liu, E. Li, Y. Zhang, Z. Tang, "Optimization of Frequent Itemset Mining on Multiple-core Processor," in *Proceedings of 2007 International Conference Very Large Data Bases(VLDB'07)*, Vienna, Austria, pp. 1275-1285, 2007.
- [13] J. Ayres, J. Gehrke, T. Yiu, J. Flannick, "Sequential Pattern Mining Using A Bitmap Representation," in *Proceedings of 2002 ACM Int. Conf. Knowledge Discovery and Data Mining (SIGKDD'02)*, Edmonton, Alberta, Canada, pp. 429-435, 2002.
- [14] R. Agrawal and R. Srikant, "Mining Sequential Patterns," in *Proceedings of 1995 Int. Conf. Data Engineering (ICDE'95)*, Taipei, Taiwan, pp. 3-14, 1995.
- [15] J. Han, J. Pei, B. Mortazavi-Asl, Q. Chen, U. Dayal, M. C. Hsu, "FreeSpan: Frequent Pattern-Projected Sequential Pattern Mining," in *Proceedings of 2000 Int. Conf. Knowledge Discovery and Data Mining (SIGKDD'00)*, Boston, Massachusetts, USA, pp. 355-359, 2000.
- [16] J. Pei, J. Han, H. Pinto, Q. Chen, U. Dayal, M. C. Hsu, "PrefixSpan: Mining Sequential Patterns Efficiently by Prefix-Projected Pattern Growth," in *Proceedings of 2001 IEEE International Conference Data Engineering (ICDE'01)*, Heidelberg, Germany, pp. 215-224, 2001.
- [17] J. Pei, J. Han, W. Wang, "Constraint-based Sequential Pattern Mining: the Pattern-growth Methods," *Journal of Intelligent Information Systems*, Vol. 28, No. 2, pp. 133-160, 2007.
- [18] M. J. Zaki, "SPADE: An Efficient Algorithm for Mining Frequent Sequences," *Machine Learning Journal*, Vol. 42 No.1-2, pp. 31-60, 2001.
- [19] R. Srikant and R. Agrawal, "Mining Sequential Patterns: Generalizations and Performance Improvements," in *Proceedings of 5th International Conference Extending Database Technology(EDBT'96)*, Avignon, France , pp. 3-17, 1996.
- [20] F. M. Facca and P. L. Lanzi, "Mining Interesting Knowledge from Weblogs: A Survey," *Data & Knowledge Engineering*, Vol. 53, No. 3, pp. 225-241, 2005.
- [21] J. Pei, J. Han, B. Mortazavi-Asl, J. Wang, H. Pinto, Q. Chen, U. Dayal, M.-C. Hsu, "Mining Sequential Patterns by Pattern-growth: The PrefixSpan Approach," *IEEE Transactions on Knowledge and Data Engineering*, Vol. 16, No. 11, pp. 1424-1440, 2004.
- [22] H. Toivonen, "Sampling Large Databases for Association Rules," in *Proceedings of 1996 International Conference Very Large Data Bases(VLDB'96)*, Bombay, India, pp. 134-145, 1996.
- [23] Y.C. Chen and G. Lee, "An Efficient Projected Database Method for Mining Sequential Association Rules," in *Proc. of 5th IEEE International Conference Digital Information Management (ICDIM 2010)*, Thunder Bay, Canada, pp. 274-278, 2010.
- [24] M. J. Zaki website, <http://www.cs.rpi.edu/~zaki/www-new/pmwiki.php/Software/Software#toc32>