UCL Université catholique de Louvain

École polytechnique de Louvain (EPL)

epl

CP based Sequence Mining on the cloud using spark

Dissertation presented by
Cyril DE VOGELAERE

for obtaining the Master's degree in
Computer Science

Supervisor(s)
Pierre SCHAUS

Reader(s)
John AOGA, Guillaume DERVAL

Academic year 2016-2017

Contents

List of Figures

List of Tables

List of Algorithms

Abstract

TODO - Half a page to a page of content should be enough

1 Introduction

TODO Prefix Projection Incremental Counting propagator mathematics Sequence of Symbols Sequence of Sets of Symbols

2 Sequential Pattern Mining

2.1 Sequential Pattern Mining Background

2.1.1 Definitions and Concepts

2.1.2 Sequences of Symbols VS Sequences of Set of Symbols

2.2 Existing specialised approaches

2.2.1 Prefix-Span

2.2.2 apriori

2.2.3 GSP

2.2.4 cspade

2.2.5 FP-Growth

Should I talk about FP-Growth ? you didn't mention it in your sheet of paper

TODO - ADD REFERENCES, CA MANQUE ENCORE

References

[1] M. Zaharia, M. Chowdhury, T. Das, A. Dave, J. Ma, M. McCauley, M. J. Franklin, S. Shenker, and I. Stoica, "Resilient distributed datasets: A fault-tolerant abstraction for in-memory cluster computing," in *Proceedings of the 9th USENIX conference on Networked Systems Design and Implementation*, pp. 2–2, USENIX Association, 2012.

[2] OscaR Team, "OscaR: Scala in OR," 2012. Available from `https://bitbucket.org/oscarlib/oscar`.

[3] N. R. Mabroukeh and C. I. Ezeife, "A taxonomy of sequential pattern mining algorithms," *ACM Computing Surveys (CSUR)*, vol. 43, no. 1, p. 3, 2010.

[4] A. Kemmar, S. Loudni, Y. Lebbah, P. Boizumault, and T. Charnois, "Prefix-projection global constraint for sequential pattern mining," in *International Conference on Principles and Practice of Constraint Programming*, pp. 226–243, Springer, 2015.

[5] A. Kemmar, S. Loudni, Y. Lebbah, P. Boizumault, and T. Charnois, "A global constraint for mining sequential patterns with gap constraint," in *International Conference on AI and OR Techniques in Constriant Programming for Combinatorial Optimization Problems*, pp. 198–215, Springer, 2016.

[6] T. Guns, S. Nijssen, and L. De Raedt, "Itemset mining: A constraint programming perspective," *Artificial Intelligence*, vol. 175, no. 12-13, pp. 1951–1983, 2011.

[7] J. O. Aoga, T. Guns, and P. Schaus, "An efficient algorithm for mining frequent sequence with constraint programming," in *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pp. 315–330, Springer, 2016.

[8] J. Pei, J. Han, B. Mortazavi-Asl, J. Wang, H. Pinto, Q. Chen, U. Dayal, and 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.

[9] Z. Yang, Y. Wang, and M. Kitsuregawa, "Lapin: effective sequential pattern mining algorithms by last position induction for dense databases," *Advances in Databases: Concepts, Systems and Applications*, pp. 1020–1023, 2007.

[10] Z. Yang and M. Kitsuregawa, "Lapin-spam: An improved algorithm for mining sequential pattern," in *Data Engineering Workshops, 2005. 21st International Conference on*, pp. 1222–1222, IEEE, 2005.

[11] F. Bonchi and C. Lucchese, "Extending the state-of-the-art of constraint-based pattern discovery," *Data & Knowledge Engineering*, vol. 60, no. 2, pp. 377–399, 2007.

[12] L. De Raedt, T. Guns, and S. Nijssen, "Constraint programming for itemset mining," in *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 204–212, ACM, 2008.

[13] L. De Raedt, T. Guns, and S. Nijssen, "Constraint programming for data mining and machine learning," in *Proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence (AAAI-10)*, pp. 1671–1675, 2010.

[14] L. D. Raedt and A. Zimmermann, "Constraint-based pattern set mining," in *proceedings of the 2007 SIAM International conference on Data Mining*, pp. 237–248, SIAM, 2007.

[15] M. J. Zaki, "Sequence mining in categorical domains: incorporating constraints," in *Proceedings of the ninth international conference on Information and knowledge management*, pp. 422–429, ACM, 2000.

Acronyms

PPIC Prefix Projection Incremental Counting propagator. 4

SoS Sequence of Symbols. 4

SoSS Sequence of Sets of Symbols. 4

Rue Archimède, 1 bte L6.11.01, 1348 Louvain-la-Neuve **www.uclouvain.be/epl**