

**Aoga, John O. R.; Guns, Tias; Schaus, Pierre**

**Mining time-constrained sequential patterns with constraint programming.** (English)

Zbl 1425.68338

Constraints 22, No. 4, 548-570 (2017).

Summary: Constraint Programming (CP) has proven to be an effective platform for constraint based sequence mining. Previous work has focused on standard frequent sequence mining, as well as frequent sequence mining with a maximum ‘gap’ between two matching events in a sequence. The main challenge in the latter is that this constraint can not be imposed independently of the omnipresent frequency constraint. Indeed, the gap constraint changes whether a subsequence is included in a sequence, and hence its frequency. In this work, we go beyond that and investigate the integration of timed events and constraining the minimum/maximum gap as well as minimum/maximum span. The latter constrains the allowed time between the first and last matching event of a pattern. We show how the three are interrelated, and what the required changes to the frequency constraint are. Key in our approach is the concept of an extension window defined by gap/span and we develop techniques to avoid scanning the sequences needlessly, as well as using a backtracking-aware data structure. Experiments demonstrate that the proposed approach outperforms both specialized and CP-based approaches in almost all cases and that the advantage increases as the minimum frequency threshold decreases.

**MSC:**

68T05 Learning and adaptive systems in artificial intelligence

68T10 Pattern recognition, speech recognition

68T20 Problem solving in the context of artificial intelligence (heuristics, search strategies, etc.)

90C59 Approximation methods and heuristics in mathematical programming

Cited in 1 Document

**Keywords:**

data mining; sequential pattern mining; constraint programming; global constraint; gap constraint; span constraint; time constraint

**Software:**

BicSPAM; Oscar; PrefixSpan

**Full Text:** DOI

**References:**

- [1] Aggarwal, C.C., \& Han, J. (2014). Frequent pattern mining. Springer. · Zbl 1297.68010
- [2] Agrawal, R., \& Srikant, R. (1995). Mining sequential patterns, \textit{Proceedings of the eleventh international conference on data engineering, 1995.} (pp. 3-14).
- [3] Antunes, C; Oliveira, AL; Perner, P (ed.); Rosenfeld, A (ed.), Generalization of pattern-growth methods for sequential pattern mining with gap constraints, 239-251, (2003), Berlin · Zbl 1029.68558
- [4] Aoga, JOR; Guns, T; Schaus, P; Frasconi, P (ed.); Landwehr, N (ed.); Manco, G (ed.); Vreeken, J (ed.), An efficient algorithm for mining frequent sequence with constraint programming, 315-330, (2016), Cham
- [5] Aoga, J.O.R., Guns, T., \& Schaus, P. (2017). Mining time-constrained sequential patterns with constraint programming. In Salvagnin, D., \& Lombardi, M. (Eds.), \textit{Integration of AI and OR techniques in constraint programming - 13th international conference, CPAIOR 2017, padova, Italy, June 5 - 8, 2017, Proceedings, Lecture Notes in Computer Science. Springer}. · Zbl 1425.68338
- [6] Ayres, J., Flannick, J., Gehrke, J., \& Yiu, T. (2002). Sequential pattern mining using a bitmap representation, \textit{Proceedings of the 8th ACM SIGKDD international conference on knowledge discovery and data mining, July 23-26, 2002, edmonton, alberta, Canada} (pp. 429-435).
- [7] Batal, I., Fradkin, D., Harrison, J., Moerchen, F., \& Hauskrecht, M. (2012). Mining recent temporal patterns for event detection in multivariate time series data. In \textit{Proceedings of the 18th ACM SIGKDD international conference on knowledge discovery and data mining} (pp. 280-288).

- [8] Beldiceanu, N; Contejean, E, Introducing global constraints in chip, *Mathematical and computer Modelling*, 20, 97-123, (1994) · [Zbl 0816.68048](#)
- [9] Coquery, E., Jabbour, S., Saïs, L., & Salhi, Y. (2012). A sat-based approach for discovering frequent, closed and maximal patterns in a sequence. In Raedt, L.d., Bessière, C., Dubois, D., Doherty, P., Frasconi, P., Heintz, F., & Lucas, P.J.F. (Eds.), *\textit{ECAI 2012 - 20th European Conference on Artificial Intelligence}*. Montpellier, France, August 27-31, 2012, *Frontiers in Artificial Intelligence and Applications*, vol. 242, pp. 258-263. IOS Press}. · [Zbl 1327.68213](#)
- [10] Desai, NAK; Ganatra, A, Efficient constraint-based sequential pattern mining (spm) algorithm to understand customers buying behaviour from time stamp-based sequence dataset, *Cogent Engineering*, 2, 1072,292, (2015)
- [11] Fournier-Viger, P., Wu, C.W., & Tseng, V.S. (2013). Mining maximal sequential patterns without candidate maintenance, *\textit{Advanced data mining and applications}* (pp. 169-180): Springer. · [Zbl 0816.68048](#)
- [12] Guns, T; Nijssen, S; De Raedt, L, K-pattern set mining under constraints, *IEEE Transactions on Knowledge and Data Engineering*, 25, 402-418, (2013)
- [13] Han, J; Pei, J; Yin, Y; Mao, R, Mining frequent patterns without candidate generation: a frequent-pattern tree approach, *Data mining and knowledge discovery*, 8, 53-87, (2004)
- [14] He, J., Flener, P., Pearson, J., & Zhang, W.M. (2013). Solving string constraints: The case for constraint programming, *\textit{International conference on principles and practice of constraint programming}* (pp. 381-397): Springer.
- [15] Henriques, R; Antunes, C; Madeira, SC; Appice, A (ed.); Ceci, M (ed.); Loglisci, C (ed.); Manco, G (ed.); Masciari, E (ed.); Ras, Z W (ed.), *Methods for the efficient discovery of large item-indexable sequential patterns*, 100-116, (2014), Cham
- [16] Henriques, R; Madeira, SC, Bicspam: flexible biclustering using sequential patterns, *BMC Bioinformatics*, 15, 130, (2014)
- [17] Kadioglu, S; Sellmann, M, Grammar constraints, *Constraints*, 15, 117-144, (2010) · [Zbl 1191.68364](#)
- [18] Kemmar, A; Lebbah, Y; Loudni, S; Boizumault, P; Charnois, T, Prefix-projection global constraint and top-k approach for sequential pattern mining, *Constraints*, 22, 265-306, (2017) · [Zbl 06598666](#)
- [19] Kemmar, A; Loudni, S; Lebbah, Y; Boizumault, P; Charnois, T; Pesant, G (ed.), *Prefix-projection global constraint for sequential pattern mining*, 226-243, (2015), Cham
- [20] Kemmar, A., Loudni, S., Lebbah, Y., Boizumault, P., & Charnois, T. (2016). A global constraint for mining sequential patterns with GAP constraint. In Quimper, C. (Ed.), *\textit{Integration of AI and OR techniques in constraint programming - 13th international conference, CPAIOR 2016, banff, AB, Canada, May 29 - June 1, 2016, Proceedings, Lecture Notes in Computer Science}*, (Vol. 9676 pp. 198-215): Springer. · [Zbl 06598666](#)
- [21] Li, C., & Wang, J. (2008). Efficiently mining closed subsequences with gap constraints. In *\textit{Proceedings of the SIAM international conference on data mining, SDM 2008, April 24-26, 2008, atlanta, Georgia, USA}* (pp. 313-322).
- [22] Lu, S., & Li, C. (2004). Aprioriadjust: an efficient algorithm for discovering the maximum sequential patterns. In *\textit{Proc. Intern. Workshop knowl. Grid and grid intell}*.
- [23] Mannila, H; Toivonen, H; Verkamo, AI, Discovery of frequent episodes in event sequences, *Data mining and knowledge discovery*, 1, 259-289, (1997)
- [24] Metivier, J., Boizumault, P., Crémilleux, B., Khiari, M., & Loudni, S. (2011). A constraint-based language for declarative pattern discovery. In *\textit{Data mining workshops (ICDMW), 2011 IEEE 11th international conference on}* (pp. 1112-1119).
- [25] Nègrevergne, B., & Guns, T. (2015). Constraint-based sequence mining using constraint programming. In Michel, L. (Ed.), *\textit{Integration of AI and OR techniques in constraint programming - 12th international conference, CPAIOR 2015, barcelona, Spain, May 18-22, 2015, Proceedings, Lecture Notes in Computer Science}*, (Vol. 9075 pp. 288-305): Springer. · [Zbl 1427.68068](#)
- [26] Oscala Team (2012). Oscala: Scala in OR. Available from <https://bitbucket.org/oscarlib/oscar>.
- [27] Parthasarathy, S., Zaki, M.J., Ogihara, M., & Dworkadas, S. (1999). Incremental and interactive sequence mining. In *\textit{Proceedings of the 8th international conference on information and knowledge management}* (pp. 251-258).
- [28] Pei, J., Han, J., Mortazavi-Asl, B., Pinto, H., Chen, Q., Dayal, U., & Hsu, M.C. (2001). Prefixspan: Mining sequential patterns efficiently by prefix-projected pattern growth. In *\textit{Proceedings of the 17th international conference on data engineering}* (pp. 215-224).
- [29] Pei, J; Han, J; Wang, W, Constraint-based sequential pattern mining: the pattern-growth methods, *Journal of Intelligent Information Systems*, 28, 133-160, (2007)
- [30] Pesant, G. (2004). A regular language membership constraint for finite sequences of variables. In *\textit{International conference on principles and practice of constraint programming}* (pp. 482-495): Springer. · [Zbl 1152.68573](#)
- [31] Pinto, H., Han, J., Pei, J., Wang, K., Chen, Q., & Dayal, U. (2001). Multi-dimensional sequential pattern mining. In *\textit{Proceedings of the tenth international conference on information and knowledge management}* (pp. 81-88).
- [32] Quimper, C.G., & Walsh, T. (2006). Global grammar constraints. In *\textit{International conference on principles and practice of constraint programming}* (pp. 751-755): Springer. · [Zbl 1160.68560](#)
- [33] Régim, J. C. (1996). Generalized arc consistency for global cardinality constraint. In *\textit{Proceedings of the thirteenth national conference on artificial intelligence-volume 1}* (pp. 209-215): AAAI press.
- [34] Rossi, F., Van Beek, P., & Walsh, T. (2006). *Handbook of CP*. elsevier. · [Zbl 1175.90011](#)
- [35] Srikant, R., & Agrawal, R. (1996). *Mining sequential patterns: Generalizations and performance improvements*. Springer.
- [36] Tatti, N., & Cule, B. (2011). Mining closed episodes with simultaneous events. In *\textit{Proceedings of the 17th ACM SIGKDD international conference on knowledge discovery and data mining, KDD '11}* (pp. 1172-1180). New York: ACM. · [Zbl 1260.68346](#)

- [37] Wang, J; Han, J; Li, C, Frequent closed sequence mining without candidate maintenance, *IEEE Transactions on Knowledge and Data Engineering*, 19, 1042-1056, (2007)
- [38] Yan, X., Han, J., & Afshar, R. (2003). Clospan: Mining: Closed sequential patterns in large datasets. In *Proceedings of the 2003 SIAM international conference on data mining* (pp. 166-177): SIAM.
- [39] Zaki, M.J. (1998). Efficient enumeration of frequent sequences. In *Proceedings of the seventh international conference on information and knowledge management* (pp. 68-75): ACM.
- [40] Zaki, M.J. (2000). Sequence mining in categorical domains: incorporating constraints. In *Proceedings of the ninth international conference on information and knowledge management* (pp. 422-429): ACM.
- [41] Zhao, Q., & Bhowmick, S.S. (2003). Sequential pattern mining: a survey. *ITechnical Report CAIS Nanyang Technological University Singapore* pp. 1-26.

This reference list is based on information provided by the publisher or from digital mathematics libraries. Its items are heuristically matched to zbMATH identifiers and may contain data conversion errors. It attempts to reflect the references listed in the original paper as accurately as possible without claiming the completeness or perfect precision of the matching.