Information sheet of the paper "MASS-CSP: Mining with Answer Set Solving for Contrast Sequential Pattern Mining"

Gioacchino Sterlicchio[1*] and Francesca Alessandra Lisi[2*]

[1]DMMM, Polythecnic University of Bari, Bari, Italy.
[2]DIB and CILA, University of Bari Aldo Moro, Bari, Italy.

*Corresponding author(s). E-mail(s): g.sterlicchio@phd.poliba.it;
FrancescaAlessandra.Lisi@uniba.it;

1 What is the claim of the paper?

The paper claims to offer a new, effective, and flexible way to mine contrast sequential patterns by leveraging the power of declarative programming and advanced solvers. More precisely, it describes *MASS-CSP (Mining with Answer Set Solving - Contrast Sequential Patterns)*, a declarative approach based on Answer Set Programming (ASP) to support the Contrast Sequential Pattern Mining (CSPM) task.

MASS-CSP benefits from advantages in handling complex pattern preferences and constraints due to the declarative nature of ASP. In particular, the ASP framework allows for a simple and elegant encoding of the CSPM task and of its variants obtained by restricting the form of patterns.

MASS-CSP can be applied to a wide number of CSPM problem instances in real-world or realistic contexts, by exploiting efficient ASP solvers such as Clingo and by imposing effective constraints such as gap and span for sequences.

2 Why is this an important contribution to the machine learning literature?

The CSPM task is a pattern mining problem with significant real-world applications in discovery science. It goes beyond traditional temporal analysis on sequential patterns and focuses also to identify patterns that are significantly different between groups (e.g., healthy vs. diseased patients, normal and attack traces). This capability is crucial

to gain insight about anomaly, outlier or concept drift into the underlying processes and making informed decisions. CSPM is inherently complex due to the combination of sequential and contrast aspects. The MASS-CSP approach tackles this complexity head-on, offering a viable solution where previous methods might struggle.

MASS-CSP leverages the declarative framework of ASP. This framework allows users to specify what they want to achieve (the problem and constraints) rather than how to achieve it (the algorithm). This shift is significant because it makes problem formulation more intuitive and accessible, especially for complex tasks like CSPM. Thus, the work helps bridge the gap between the logical, rule-based world of declarative programming and the data-driven world of machine learning. It demonstrates how logical reasoning can be effectively applied to extract meaningful patterns from data.

One of the key strengths of MASS-CSP is its ability to handle a wide range of constraints on patterns. Users can easily express complex preferences and restrictions using ASP syntax, making the approach highly adaptable to different domains of science and specific analytical needs. The approach moves towards a richer query language for pattern mining. Instead of being limited to pre-defined algorithms, users can essentially "query" the data for patterns that meet their specific criteria, opening up new possibilities for exploratory data analysis.

MASS-CSP provides a solid foundation for future research in Declarative Pattern Mining (DPM). It opens up avenues for exploring more advanced constraints, integrating other machine learning techniques, and developing more sophisticated solvers tailored for pattern mining tasks.

In summary, MASS-CSP can be considered an important contribution to the machine learning literature, and in particular to the topics covered by this Special Issue, because it introduces a novel, effective, and flexible approach to a complex and relevant problem in discovery science. It advances the use of declarative programming in machine learning, offers a powerful framework for handling constraints, and has the potential to impact various application domains. It represents a significant step forward in the field of pattern mining and demonstrates the growing power of combining logical reasoning with data-driven analysis.

3 What is the evidence you provide to support your claim?

To support our claim, we have conducted an empirical evaluation on a set of real-world attacks to the 4G-LTE network. The MASS-CSP evaluation focuses on understanding the efficiency, in terms of time requirements, of pattern extraction and its variant with condensed representations (closed and maximal patterns) and constraints on patterns (gap and span). The other side of evaluation is the approach itself. In fact we want to understand how the task formalized with ASP is empirically efficient in time and space. Experiments provide evidence that, to reduce patterns explosion, condensed representations and pattern constraints can help. In particular, condensed representations are more time-memory consuming compared to the basic formulation of MASS-CSP. The gap constraint reduces pattern set, and time-memory requirements, while the span constraint does not lead to significant improvement of memory consumption.

As regards effectiveness, we would like to add that the ASP encoding is rather easy to understand due to the readability of the first-order logic languages. This emphasizes a significant advantage of declarative pattern mining: it demands less development time for clearly specified tasks than procedural approaches. The flexibility of the ASP language and its solvers implies that incorporating new constraints into our framework is straightforward, requiring only a few rules.

4 What papers by other authors make the most closely related contributions, and how is your paper related to them?

Our paper proposes an ASP-based approach to the task of mining contrast sequential patterns. Consequently, the works closest to ours use ASP as a framework for pattern mining. Guyet et al. [1] explore a preliminary approach to solve the sequential pattern mining problem using ASP and compare it with a specialized algorithm. Gebser et al. [2] adopt ASP to extract condensed representations of sequential patterns, focusing on closed, maximal, and skyline patterns. For our work, we have adapted the ASP encoding of closed and maximal patterns from the previous work to the case of contrast sequential patterns. Additionally, Samet et al. [3] propose a method for mining meaningful rare sequential patterns with ASP. Guyet et al. [4] propose two representations of embeddings (fill-gaps versus skip gaps) for ASP sequential pattern mining. The results show that the fill-gaps strategy is better on real problems due to lower memory consumption. Based on this result, we have chosen the fill-gaps encoding as the starting point for our approach to address the task of CSPM and span and gap constraints. The task of mining negative sequential patterns in ASP is addressed by Besnard and Guyet [5].

5 Have you published parts of your paper before, for instance in a conference?

Our paper provides an overall overview of the work we have done over the past two years on the CSPM task with ASP. Some parts of the paper have been published in conference proceedings before. In particular, the ASP encoding of MASS-CSP (reported in Appendix B for self-containment of the paper) has been presented and discussed in the following publications:

- In [6], we address the CSPM task for the first time and provide the ASP encoding of the basic MASS-CSP. The approach is evaluated through a scalability test and a comparison with a hybrid ASP-based approach.
- In [7], we consider condensed representations of contrast sequential patterns and evaluate the extended ASP encoding of MASS-CSP on the same data as in [6] for a comparative evaluation.
- In [8], we apply MASS-CSP to find patterns that characterize different attacks on the 4G LTE network. The experiments show that a declarative approach is feasible in

this context, and that the implementation of span and gap constraints in MASS-CSP makes the search for patterns more efficient and effective.

For this special issue, we have conducted a more extensive evaluation of MASS-CSP, running several experiments on the same data used in [9]. In particular, we have discussed the grounding phase of ASP to understand the behavior of condensed representations and gap and span constraints.

References

[1] Guyet, T., Moinard, Y., Quiniou, R.: Using answer set programming for pattern mining. arXiv preprint arXiv:1409.7777 (2014)

[2] Gebser, M., Guyet, T., Quiniou, R., Romero, J., Schaub, T.: Knowledge-based sequence mining with ASP. In: IJCAI 2016-25th International Joint Conference on Artificial Intelligence, p. 8 (2016). AAAI

[3] Samet, A., Guyet, T., Negrevergne, B.: Mining rare sequential patterns with ASP. In: ILP 2017-27th International Conference on Inductive Logic Programming (2017)

[4] Guyet, T., Moinard, Y., Quiniou, R., Schaub, T.: Efficiency analysis of ASP encodings for sequential pattern mining tasks, 41–81 (2018)

[5] Besnard, P., Guyet, T.: Declarative mining of negative sequential patterns. In: DPSW 2020-1st Declarative Problem Solving Workshop, pp. 1–8 (2020)

[6] Lisi, F.A., Sterlicchio, G.: Mining contrast sequential patterns with ASP. In: Basili, R., Lembo, D., Limongelli, C., Orlandini, A. (eds.) AIxIA 2023 - Advances in Artificial Intelligence - XXIInd International Conference of the Italian Association for Artificial Intelligence, AIxIA 2023, Rome, Italy, November 6-9, 2023, Proceedings. Lecture Notes in Computer Science, vol. 14318, pp. 44–57. Springer, Cham (2023). https://doi.org/10.1007/978-3-031-47546-7_4 . https://doi.org/10.1007/978-3-031-47546-7_4

[7] Sterlicchio, G., Lisi, F.A.: Condensed representations for contrast sequential pattern mining in ASP. In: Angelis, E.D., Proietti, M. (eds.) Proceedings of the 39th Italian Conference on Computational Logic, Rome, Italy, June 26-28, 2024. CEUR Workshop Proceedings, vol. 3733. CEUR-WS.org, Germany (2024). https://ceur-ws.org/Vol-3733/short1.pdf

[8] Sterlicchio, G., Lisi, F.A.: Detecting patterns of attacks to network security in urban air mobility with answer set programming. In: Endriss, U., Melo, F.S., Bach, K., Diz, A.J.B., Alonso-Moral, J.M., Barro, S., Heintz, F. (eds.) ECAI 2024 - 27th European Conference on Artificial Intelligence, 19-24 October 2024, Santiago de Compostela, Spain - Including 13th Conference on Prestigious Applications of Intelligent Systems (PAIS 2024). Frontiers in Artificial Intelligence and

Applications, vol. 392, pp. 1285–1292. IOS Press, Amsterdam, Netherlands (2024). https://doi.org/10.3233/FAIA240626 . https://doi.org/10.3233/FAIA240626

[9] Sterlicchio, G., Lisi, F.A.: Detecting Patterns of Attacks to Network Security in Urban Air Mobility with Answer Set Programming. Zenodo (2024). https://doi.org/10.5281/zenodo.13135192 . https://doi.org/10.5281/zenodo.13135192