Ant Colony Optimization for Optimal Process Mining in Healthcare with Event Attributes

Omar Rifki[1], Zhihao Peng[2], Lionel Perrier[3], Vincent Augusto[4] , and Xiaolan Xie[4]

[1] Univ. Littoral Côte d'Opale, UR 4491, LISIC, F-62100 Calais, France,
[2] Univ. Picardie Jules Verne, LTI, EA 3899, 80025 Amiens, France,
[3] Léon Bérard Cancer Centre, GATE UMR 5824, University of Lyon, 69008 Lyon, France,
[4] Centre CIS, CNRS, UMR 6158 LIMOS, Mines Saint-Étienne, Université Clermont Auvergne, 42023 Saint-Étienne, France, xie@emse.fr.

Mots-clés : *Optimisation, healthcare, process mining.*

1 Introduction

Personal activities nowadays generate vast amounts of data. These trails are called event logs. Although data mining, big data, and machine learning techniques provide rich tools for knowledge discovery, they perform poorly when it comes to process-oriented data. The field of process mining is filling this gap [3]. Discovering clinical processes and patient care pathways, and studying their conformance is even more salient in the medical domain, which allows gaining significant insights into the treatment trajectory of various types of patients, and studying the effectiveness of clinical protocols, among others [1]. Peng et al. [2] proposed a method to generate an optimal process model, *i.e.* a representation of the sequence of steps and the different paths a process can take, which is more adapted to the care pathways of cancer patients by : (i) being multi-layered thus allowing repetitions of health events, (ii) having transitions augmented with health states between the events, (iii) using a new relevance measure to quantify its worth. Dynamic programming and a constructive local search-like algorithm are used for the optimization part. We study an extension of this work by allowing a more flexible representation which : (a) associates a set of attributes to each health event, (b) chooses the attribute values in a tractable manner with much or less precision, and (c) incorporates the attribute allocation within the process model optimization framework.

2 Problem modeling

The problem statement follows [2]. Each trace of the event log is mapped in a process model into what is termed an *event game*, obtained by dynamic programming. To address point (c), we incorporate the attribute choices in the event game and the relevance measure. The relevance of each health event is a weighted sum between the relevance of the event label, and that of its attributes, which is in turn an average of the relevance of its elements. Concerning point (b), we use a tree structure to represent the different values an attribute can take, with the following property : a value in level n generalizes and is less precise than the values of its children nodes in level $n + 1$. An example is given below for the attribute Sarcoma type. A precision function is associated with the tree, such that it is equal to one for the leaves of the higher level, and becomes increasingly smaller when going up in the tree, with the lowest precision given to the root node. The relevance of an attribute value v for a node in a process model is thus equal to the number of traces passing by this node, which have an attribute value similar to or children of v in the corresponding tree, multiplied by the precision of v. The application to cancer patients shows that this design is flexible enough to output less precise values for the attributes when there is not a dominant value. Finally, for point (a), we rely on the practitioners' expertise to affect a number of attributes to the health events considered in [2]. More technical details will be given during the talk.

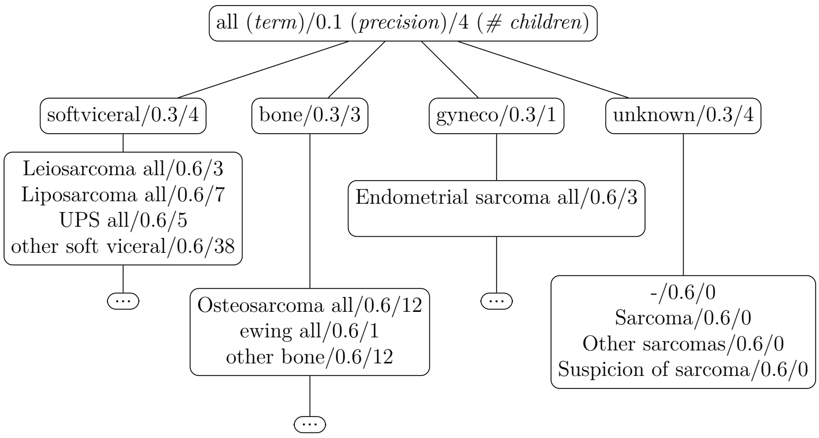

Fig. An example of the health attribute tree.

3 Solution algorithm

We propose an ant colony optimization algorithm to solve the problem. At each iteration, each ant moves in a graph composed of KB nodes plus an initial starting node, where K is the number of layers (an input of the problem), and B is the number of health labels considered, leaving pheromone trails behind. A movement means adding an event to the solution, thus ants construct process models from scratch. The transitions are optimized at the end of each addition by repairing the solution if needed. For the heuristic information, we rely on the constructive local search-like algorithm of [2] upgraded to handle the attributes' relevance. This latter is based on the notion of marginal relevance to select nodes to be added to the process model. For the algorithm update rule, Max-Min Ant System (MMAS) is shown to perform better. The ant colony optimization has a slight superiority compared to the constructive algorithm in terms of solution quality. This is mainly due to the stochasticity induced by the algorithm.

4 Conclusion and perspectives

This work is a follow-up study of Peng et al. [2], where heath attributes are introduced to the problem modeling with a flexible and adaptive precision. We propose an ant colony optimization algorithm, which adds some stochasticity to the constructive local search-like algorithm of [2], upgraded to handle health attributes. The perspectives of the work are multiple and include : (1) adding nodes with the same health label in the same layer but with different attributes, (2) relaxing the constraint of the maximum number of layers K, (3) instead of having one attribute value, selecting a set of values, as practitioners might be interested to know the second and the third rank choice of the attribute value.

Références

[1] Jorge Munoz-Gama, Niels Martin, Carlos Fernandez-Llatas, Owen A Johnson, Marcos Sepúlveda, Emmanuel Helm, Victor Galvez-Yanjari, Eric Rojas, Antonio Martinez-Millana, Davide Aloini, et al. Process mining for healthcare : Characteristics and challenges. *Journal of Biomedical Informatics*, 127 :103994, 2022.

[2] Zhihao Peng, Vincent Augusto, Lionel Perrier, and Xiaolan Xie. Optimal process mining of traces with events and transition attributes with application to care pathways of cancer patients. *IEEE Transactions on Automation Science and Engineering*, 2023.

[3] Wil Van Der Aalst. Process mining. *Communications of the ACM*, 55(8) :76–83, 2012.