

Chapter 6

Conclusions

The AI planning literature contains decades of substantial work on discovering sequences of actions that lead to a given outcome that, similar to this work, is often specified as a goal condition. In this work, we have described an approach to solve problems that at first seems quite similar to those tackled by AI planning; however the main characteristic of the problems of interest are that important assumptions made in AI planning approaches cannot be made in this case. There are, however, significant assumptions made in these works that cannot always be made, such as:

1. The number of actions available to solve the problem is assumed to be relatively small,
2. The causal relationships within the model describing the environment are well understood (e.g., if a block is picked up and put on top of another block, then the rest of the blocks remain unchanged, and the new pile continues to exist until some further action changes it).
3. The effects of actions taken in the environment are well understood (e.g., in the example above, picking up the block only has the effect of the block no longer being on the table).

In this work, we have described an approach to solve problems that at first seems quite similar to those tackled by AI planning; however the main characteristic of the problems of interest are that the above assumptions cannot be made: effects of actions are not clearly understood, nor are the causal relationships between the values of the parameters in the system, and there are a huge number of actions to choose from. Our proposal was therefore to solve these problems in a data-driven manner, where these poorly understood effects and relationships can be gleaned from the data instead of assuming that they are packaged with the input.

In Chap. 1 we began by introducing this problem and the concept of event databases, where attributes are assumed to be either of type *action* or *state*; the former are those that can directly be manipulated (albeit at a certain cost and with certain probability of success), while the latter constitute those that can only be influenced indirectly. Two examples of this kind of data were introduced: school performance Fig. 1.1 and city government Fig. 1.2. In Chap. 2, we formally

introduced the different pieces of this problem: (optimal) state change attempts, effect estimators, cost functions, and probability of effectiveness, illustrating each of them on the datasets from the running examples. We also showed that determining optimal state change attempts is not an easy problem, proving that most interesting versions of the optimization task belong to complexity classes widely believed to be intractable. In Chap. 3, we began by taking a closer look at the different kinds of effect estimators that can be defined, and focused on the *data selection* effect estimator due to its computational properties. Also in this chapter, we introduced the TOSCA algorithm, an approach that combines data selection effect estimators with trie data structures to obtain optimal state change attempts more efficiently. In Chap. 4, we explored how optimal state change attempts can be computed by applying Markov Decision Processes, one of the most widely used models in planning under uncertainty, and showed that, even though it is theoretically possible, the size of the resulting problem prohibits its application in practice. Finally, in Chap. 5 we presented empirical results obtained from running implementations of our algorithms both on synthetic and real-world datasets, showing that the TOSCA algorithm can be applied to event databases with hundreds of thousands, and even millions. of tuples.