

The Inductive Constraint Programming Loop

Christian Bessiere¹, Luc De Raedt², Tias Guns², Lars Kotthoff³,
Mirco Nanni⁴, Siegfried Nijssen², Barry O’Sullivan³, Anastasia
Paparrizou¹, Dino Pedreschi⁴, and Helmut Simonis³

¹CNRS, University of Montpellier, France

²DTAI, KU Leuven, Belgium

³Insight, University College Cork, Ireland

⁴University of Pisa, Italy

Abstract

Constraint programming is used for a variety of real-world optimization problems, such as planning, scheduling and resource allocation problems. At the same time, one continuously gathers vast amounts of data about these problems. Current constraint programming software does not exploit such data to update schedules, resources and plans. We propose a new framework, which we call the *inductive constraint programming loop*. In this approach data is gathered and analyzed systematically in order to dynamically revise and adapt constraints and optimization criteria. Inductive Constraint Programming aims at bridging the gap between the areas of data mining and machine learning on the one hand, and constraint programming on the other.

1 Introduction

Machine Learning/Data Mining (ML/DM) and Constraint Programming (CP) are central to many application problems. ML is concerned with learning functions/patterns characterizing some training data whereas CP is concerned with finding solutions to problems subject to constraints and possibly an optimization function.

The problem with current technology is that the problems of data analysis and constraint satisfaction/optimization have almost always been studied independently and in isolation. Indeed, there exist a wide variety of successful approaches to analysing data in the field of ML, DM and statistics, and at the same time, advanced techniques for addressing constraint satisfaction and optimization problems have been developed in the CP community. Over the past decade a limited number of isolated studies on specific cases has indicated that significant benefits can be obtained by connecting these two fields

[EF01, XHHL08, DGN08, BHO09, KBC10, CJSS12], but so far a truly general, integrated and cross-disciplinary approach has been missing.

CP technology is used to solve many types of constraint satisfaction and optimization problems, such as in power companies generating and distributing electricity, in hospitals planning their surgeries, and in public transportation companies scheduling buses. Despite the availability of effective and scalable solvers, current approaches are still unsatisfactory. The reason is that when using CP technology to solve these applications, the constraints and criteria, that is, the *model*, must be statically specified. However, in reality this model often needs to be revised over time. The revision can be needed to reflect changes in the environment due to external events that impact the problem. The revision can also be needed because the execution of the solution generated by the model has modified the characteristics of the problem. Finally the revision can be needed simply because the original model did not capture correctly the problem. Observing the impact of the solution allows us to correct or improve the model. Therefore, there is an urgent need for improving and revising a model over time *based on data* that is continuously gathered about the performance of the solutions and the environment they are used in.

The CP community has extended the basic constraint satisfaction and optimization problems to better tackle changing environments. The *dynamic* constraint satisfaction approach [DD88] allows the addition/retraction of constraints from the initial model. But this approach does not predict the changes from data, but rather the addition/retraction of constraints is performed by the user. The *online/stochastic* constraint programming approach [Wal02, BH04] offers a framework to deal with unknown future events, such as customer requests. It builds a finite set of *future scenarios*, e.g. using sampling from a known distribution, and the optimization problem is then defined over each of the scenarios. The framework does not capture ways of using data, other than for the prediction of possible scenarios of events. In constrained-based planning, the *conditional temporal* problem approach [TVP03] extends standard temporal constraint satisfaction by adding observation nodes and attaching labels to all nodes to indicate the situation(s) in which each will be executed. This extension permits the construction of conditional plans that are guaranteed to satisfy complex temporal constraints. This makes it possible to dynamically adapt the plan in response to the observations made during execution. However, it does not allow learning from experience from data, such as unsuccessful plans. Even further from CP technology, a *truth maintenance system* [Doy79] is a knowledge representation approach to recording and maintaining the reasons for program beliefs. The name truth maintenance is due to the ability of these systems to maintain consistency between old and current beliefs through a revision mechanism. In order to choose their actions, reasoning programs make assumptions and subsequently revise their beliefs when discoveries contradict these assumptions. Truth maintenance systems do not contain any constraint optimization or learning capabilities.

In general, exploiting gathered data to modify and adjust any aspect of a model is difficult and labor intensive with state-of-the-art solvers. As a conse-

quence, the data that is being gathered today, in order to monitor the quality of the produced solutions and to help evaluate the effect of possible adjustments to the constraints or optimization criteria, is not fully exploited when changes in a schedule or plan are needed. Hence, schedules and plans that are produced are often suboptimal. This, in turn, leads to a waste of resources. Instead of using data passively, data should be actively analysed in order to discover and update the underlying regularities, constraints and criteria that govern the data.

In this paper, we propose and formalize the new framework of inductive constraint programming. This framework is based on what we call the *inductive constraint programming loop*, which is an interaction between a machine learning component (ML) and a constraint programming component (CP). The ML component observes the world and extracts patterns. The CP component solves a constraint satisfaction or optimization problem using these patterns whose solution is applied to the world. We assume the world changes over time, possibly due to the impact of applying our solution. This process is repeated in a loop. Inductive constraint programming will serve the long-term vision of easier-to-use and more effective tools for resource optimization and task scheduling.

2 Background

In this section we introduce the basic concepts used later in the paper. We briefly define and explain what is a constraint problem and a learning problem.

2.1 Constraint problem

The central notion in constraint programming is the *constraint*. A constraint is a Boolean function whose scope is a set of (integer) variables. Depending on whether the function returns true or false for a given input assignment of its variables, the constraint accepts or rejects the assignment. For instance, the constraint $X_1 + X_2 = X_3$ specifies that any combination of values for variables X_1, X_2 and X_3 has to be such that the sum of X_1 and X_2 equals X_3 . Based on the notion of constraint, we define the concept of a constraint network and a solver.

A *constraint network* $N = (X, D, C, f)$ is composed of: a set X of variables taking values from domain D . These variables are subject to constraints in the set C . The optional evaluation function f takes as input an assignment on X and returns a cost for it. A solution (optionally *best* solution) of N is an assignment in D^X satisfying all the constraints in C (optionally minimizing f). A *solver* takes as input a constraint network and returns a solution/best solution or failure in case no solution satisfying all the constraints exists. If we take as an example the well-known Sudoku problem, a constraint network expressing it could be the following. The variables are the cells, namely $X = \{X_1 \dots X_{81}\}$. Each X_i represents the digit in this cell, and as such, it takes a value from $D_i = \{1..9\}$. For every prefilled cell i , X_i is assigned the corresponding prefilled digit. For all the 810 pairs of cells i and j that belong to the same row, column

```

1 array [1..9,1..9] of 0..9: start; %% initial board 0 = empty
2 array [1..9,1..9] of var 1..9: puzzle;

3 % fill initial board
4 constraint forall(i,j in 1..9 where start[i,j] > 0)(
5   puzzle[i,j] = start[i,j] );

6 % All different in rows
7 constraint forall (i in 1..9) (
8   alldifferent( [ puzzle[i,j] | j in 1..9 ] ) );

9 % All different in columns.
10 constraint forall (j in 1..9) (
11   alldifferent( [ puzzle[i,j] | i in 1..9 ] ) );

12 % All different in sub-squares:
13 constraint forall (i, j in 1..3)(
14   let { int: a = (i-1)*3; int: b = (j-1)*3 } in
15   alldifferent( [ puzzle[a+i1, b+j1] | i1,j1 in 1..3 ] ) );

16 solve satisfy;

```

Figure 1: Sudoku in pseudo-Minizinc.

or block, a constraint $X_i \neq X_j$ is put in C . Alternatively, C can be composed of 27 global constraints `alldifferent(S)`, for the 27 sets S of 9 variables in the same row, column, or block.

There exist several languages/formats for specifying a constraint problem to be given to a solver for solving. Figure 1 expresses Sudoku as a constraint program using a pseudo-MiniZinc language [MS14]. Line 1 defines an input matrix `start` containing the prefilled cells of the Sudoku. Line 2 defines the matrix `puzzle` of variables that will contain the solution of the Sudoku. Lines 4–5 put equality constraints between the prefilled cells in the input matrix `start` and the matrix of variables `puzzle`. Lines 7–8 post an `alldifferent` constraint on every row of `puzzle`. `alldifferent($x_i \mid i \text{ in } 1..n$)` is a global constraint that specifies that variables $x_1..x_n$ must all take different values. Lines 10–11 do the same for the columns. The constraint specified in Lines 13–15 is a bit more tricky as it has to play with the indices of the subsquares to post the `alldifferent` constraints on the variables of every subsquare in `puzzle`. Finally, line 16 calls the solver on the instance.

2.2 Learning problem

In machine learning, the goal is to learn a hypothesis that explains the observed data, and thus is able to predict future data. The data typically consists of a set of training examples E , which are assumed to be independent and identically

distributed. Different learning methods differ largely in the type of examples to learn from, and the type of hypothesis they want to learn. The most popular learning setting is supervised learning, where each example in E is accompanied by a label that should be predicted.

Considering the sudoku example again, one might wish to learn how long it takes to solve a sudoku for a typical user, based on features of the sudoku such as the number of empty cells, how many cells contain the same number, the average number of choices left in the cells, etc. The training examples would then consist of these features for a certain sudoku, with as label the time it took a particular user to solve it. One can then search, for example, for a linear function over the features that best predicts the labels, or for a *decision tree* that does so.

More formally, we define the learning problem as follows. A *learning problem* $L = (E, H, t, \text{loss})$ is composed of a set E of examples, a hypothesis space H , the target function t that one wants to learn, and a loss function $\text{loss}(E, h, t)$ that measures the quality of a hypothesis $h \in H$ with respect to dataset E and the target hypothesis t . The goal is to find a hypothesis that minimizes the loss. This is a very general definition that encompasses both supervised and unsupervised settings, including clustering classification and regression.

For example, for linear regression, the data would be real-valued data $E \subset \mathbb{R}^d$ with real-valued labels identified by target function t , where $\forall \mathbf{e} \in E : t(\mathbf{e}) \in \mathbb{R}$. The goal is then to learn a linear function $h_{\mathbf{c}} : E \rightarrow \mathbb{R}$ with coefficients \mathbf{c} that minimizes the sum of squared errors between the predicted value and the observed value: $\text{loss}(E, h_{\mathbf{c}}, t) = \sum_{\mathbf{e} \in E} |h_{\mathbf{c}}(\mathbf{e}) - t(\mathbf{e})|^2 = \sum_{\mathbf{e} \in E} |\mathbf{e} \cdot \mathbf{c} - t(\mathbf{e})|^2$. Many other loss functions and hypothesis spaces have been defined in the literature.

More concretely for the sudoku solving time example, the target function would return the running time to solve the sudoku. The data would be three dimensional when using the number of empty cells, the average frequency of the numbers and the average number of choices left in each cell. The learned function could give, for example, high weight to the number of empty cells (more empty is longer to solve) and negative weight to the average frequency the numbers appear (higher frequency is easier).

Moreover, a range of machine learning methods such as (linear) regression and support vector machines can be expressed as standard optimization problems (often unconstrained), where the goal is to find an assignment to function parameters such that the loss is minimized.

For the linear regression with a sum of squared error (least squares regression), the optimization problem is defined as follows:

$$\underset{\mathbf{c}}{\text{minimize}} \quad \sum_{\mathbf{e} \in E} |\mathbf{e} \cdot \mathbf{c} - t(\mathbf{e})|^2$$

In case of *ridge* regression, the loss function includes a regularization component, to avoid fitting the given examples too exactly, by restricting the capacity of

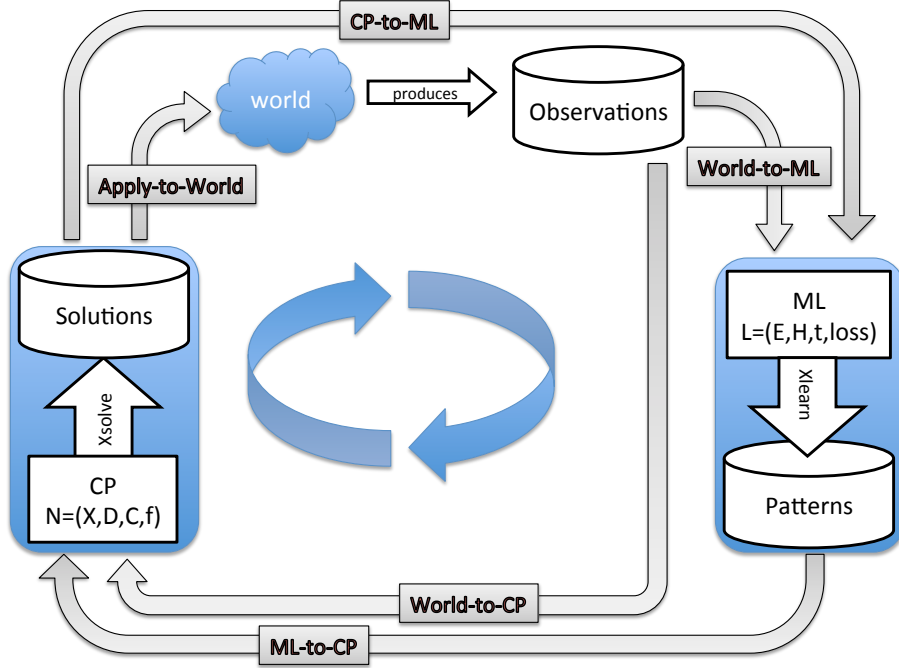


Figure 2: The inductive constraint programming loop

the weights:

$$\begin{aligned}
 & \underset{\mathbf{c}}{\text{minimize}} && \sum_{\mathbf{e} \in E} |\mathbf{e} \cdot \mathbf{c} - t(\mathbf{e})|^2 \\
 & \text{subject to} && \sum_i |\mathbf{c}_i|^2 \leq 1
 \end{aligned}$$

Similarly but more complex, support vector machine learning can also be seen as solving a convex optimization problem [STS11].

In practice, one does not typically use generic optimization techniques for such problems, but more specialized and scalable solving methods that exploit specific properties of the optimization problem.

3 Inductive Constraint Programming Loop

The inductive constraint programming loop will cope with changes in the world by iteratively solving a learning problem and a constraint problem. The loop is composed of several components that interact with each other through writing and reading operations. A visualization of the loop is given in Figure 2. We introduce each of the elements in the loop in turn.

The CP component is composed of a constraint network $N = (X, D, C, f)$ (f is optional), a constraint solver **Xsolve**, and a **Solutions** repository. **Xsolve** generates solutions of N , or good/best solutions of N according to f , that it writes in the **Solutions** repository. In case **Xsolve** is not able to produce any solution to be applied to the world, the CP component notifies the ML component by sending information about the failure.

The ML component is composed of a learning problem $L = (E, H, t, loss)$, a learner **XLearn**, and a **Patterns** repository. **XLearn** learns hypotheses t (typically one) and writes them in the **Patterns** repository.

The World component is composed of a world W , an evaluation function *eval_world*, and a **Observations** repository. The world W can have its own independent behavior, dynamically changing under the effect of time and the effect of applying solutions from the **Solutions** repository. The solutions are evaluated by the *eval_world* function and this feedback is stored in the **Observations** repository.

Now that we have defined the basis of the inductive constraint programming loop, we need to define the way the CP component, the ML component, and the world interact with each other. They interact through a set of reading/writing functions.

An *inductive constraint programming loop* is composed of a world $(W, eval_world)$, a CP component $(N, Xsolve)$, and an ML component $(L, XLearn)$. The loop uses the following channels of communication:

- function **World-to-ML** reads data and evaluations from the **Observations** repository and updates the learning problem L , that will be used by **XLearn** to learn a hypothesis h ;
- function **CP-to-ML** is used to send feedback from the previous iteration of the CP component to the ML component, e.g. when **Xsolve** cannot find any satisfactory solution to be applied to the world;
- function **World-to-CP** reads data from the **Observations** repository that can be used to directly update the constraint network N used by **Xsolve**;
- function **ML-to-CP** reads patterns from the **Patterns** repository and updates the constraint network N used by **Xsolve** to produce solutions;
- function **Apply-to-World** takes solutions in the **Solutions** repository and applies them to the world, if possible.

The following pseudo code demonstrates how these communication channels are used in the inductive constraint programming loop:

Algorithm 1 Pseudo code of a loop cycle using the components.

```

function CYCLE(Observations, optional Solutions)
  repeat
     $L_o \leftarrow \text{World-to-ML}(\text{Observations})$ 
     $L_p \leftarrow \text{CP-to-ML}(\text{Solutions})$ 
     $L \leftarrow \text{constructL}(L_o, L_p)$ 
    Patterns  $\leftarrow \text{applyXlearn}(L)$ 

     $N_o \leftarrow \text{World-to-CP}(\text{Observations})$ 
     $N_p \leftarrow \text{ML-to-CP}(\text{Patterns})$ 
     $N \leftarrow \text{constructN}(N_o, N_p)$ 
    Solutions  $\leftarrow \text{applyXsolve}(N)$ 
  until Apply-to-World(Solutions)
end function

```

Initially, **World-to-ML** is used to gather training data for the ML component. These data can be feedback from previous executions of solutions of the CP component on the world. The solution of the previous cycle can also directly be used as well, through **CP-to-ML**. This is especially useful if the previous solution could not be applied to the world, for example because the learned patterns lead to an inconsistency. Using the output of **World-to-ML** and **CP-to-ML**, the learning problem L can then be constructed, specific to the learner at hand. Next, the learner is applied to L and patterns are obtained. These patterns can be weights of an objective function, constraints, or any other type of structural information that is part of the CP problem.

A similar process then happens for the CP component: the network is constructed using the output of **World-to-CP** and **ML-to-CP**, after which the solving method is used and solutions are obtained.

These solutions are then applied to the world using **Apply-to-World**. As mentioned before, it may be that the particular solution (or non-solution) is not applicable to the world. In that case, a new iteration of the loop is started immediately which bypasses the world. Otherwise the solutions are applied to the world, after which a new cycle with new observations can be started.

We can observe that there is no direct link between the ML component and the world. Our framework is indeed devoted to solving combinatorial problems such as scheduling and routing, revising them based on feedback from the world; it does not aim to only classify or predict events in the world.

4 Illustrative Example

To illustrate the inductive constraint programming loop we will use a scheduling setting that occurs in hospitals. This setting includes an ML component, a CP component and a world component.

We will first describe the CP component. In this component we focus on a task scheduling problem. The treatment of a patient typically involves the

execution of various tasks on this patient, such as executing scans, taking blood tests, operating on the patient, physiotherapy sessions, and so on. These tasks need to be executed in a well-defined order, and require the use of the resources of the hospital for a certain amount of time. The overall scheduling problem is how to schedule these tasks in the shortest amount of time possible, using the limited resources of the hospital. Therefore, important parameters of this scheduling problem include the resources available in the hospital and the tasks that need to be executed. For each task, it is important which resources need to be used, how many such resources are needed, and for how long they need to be used.

Whereas for many patients it is clear which procedures need to be followed before the patient can be discharged from the hospital, this is not the case for the duration of these tasks: depending on parameters such as age or health conditions, a certain task may take much longer for one patient than for another.

The goal of the ML component is to address this challenge: its role is to predict how long a task is estimated to take for a patient. This involves solving a regression problem as identified earlier: for each given task for a patient, the properties of the task and the patient, together with similar historic data and the resulting durations, are used to predict the task duration, which is a real number.

The world component executes the schedules; it produces data about patients and observations concerning the true durations of tasks.

Clearly, as the tasks are executed in the hospital, the predicted durations may differ from the actual durations. Furthermore, new patients, and hence new tasks, arrive. This means that the hospital needs to schedule tasks on a regular basis. The patient data that is collected during each such iteration can here be used to improve the quality of the predicted task durations. This makes it a good example of the inductive constraint programming loop. Within this loop, we can distinguish the following components and functions:

- function **World-to-ML** reads historical patient data and historical task durations for these patients; furthermore, it reads the patients that are currently in the hospital and the tasks that need to be executed for these patients;
- the ML component predicts the durations for the tasks that need to be executed, using the historical data;
- function **ML-to-CP** reads the learned durations and updates the CP network accordingly;
- function **World-to-CP** reads the tasks that need to be executed from the world, as well as the resources available in the hospital;
- the CP component solves the updated scheduling problem;
- function **Apply-to-World** applies the resulting schedule in the world.

In this example, the function **CP-to-ML** is not used; it could be used, for instance, if there is a preference to schedule nurses and doctors in similar teams or with similar load or time-breaks from day-to-day.

Both components can be formalized using a CP language, such as the MiniZinc language mentioned earlier. Figure 3 shows MiniZinc code for the task scheduling problem. In this model, the parameters of the problem setting are reflected as follows:

- the **dur** array represents the durations of all the tasks, as predicted by the ML component (line 3);
- the **prev** array indicates for each task which task needs to be executed before this task; note that we assume that there is a dummy first task that precedes all tasks (line 4);
- the **cap** array represents the capacity of the resources available (line 7);
- the **use** array represents how many resources of each type need to be used to execute a certain task (line 8).

The variables that need to be found are the **start** variables (line 11), which indicate at which times the tasks need to be executed. The constant **max_time** represents the latest time at which a task may still start, this could be specified for each task separately as well.

The constraints are twofold:

- the constraint on line 14 is a **cumulative** constraint; for a given resource, it ensures for each time point that the use of the resource is within the capacity bound of that resource. Note that the **cumulative** constraint is a built-in constraint available in the MiniZinc language. Constraints that can involve any number of variables are called *global* constraints. They can capture complex structural constraints of the problem. Global constraints are an essential part of the efficiency of CP models.
- the constraint on line 17 ensures that a task only executes after the task that should precede it has finished.

The optimization criterion is to minimize the makespan, that is, to assign the **start** variables so that the total amount of time used by the schedule is minimum (line 19).

To predict the durations of the tasks in the hospital, a regression problem needs to be solved. Many different models can be made for this regression problem, each corresponding to learning a different type of regression model. Arguably the most simple regression model is the linear model, in which the task duration prediction is based on a linear combination of the characteristics of the patient on which the task is executed.

The problem of learning such a regression model is formalized in Figure 4. Variables **X** and **Y** represent the training data, where **X** contains the descriptive

```

1  % Tasks: duration and precedence
2  int: nbTasks; set of int: Tasks = 1..nbTasks;
3  array[Tasks] of int: dur;
4  array[Tasks] of int: prev;

5  % Resources: capacity and use
6  int: nbRes; set of int: Res = 1..nbRes;
7  array[Res] of int: cap;
8  array[Res,Task] of int: use;

9  % Variables: start times
10 int: max_time;
11 array[Tasks] of var 0..max_time: start;

12 % Resource capacities
13 constraint forall(r in Res) (
14     cumulative(start, dur, use[r], cap[r]) );

15 % Precedence between tasks
16 constraint forall(t in Tasks) (
17     start[t] > (start[prev[t]] + dur[prev[t]]) );

18 % Minimize the amount of time
19 solve minimize max(t in Tasks) (start[t]+dur[t]);

```

Figure 3: The hospital scheduling problem in pseudo-Minizinc.

attributes of various tasks and Y the historical durations of these tasks; variable W represents the weights of the features that we are learning.

Based on these weights, we can calculate an error for the predictions; line 10 calculates a weighted linear combination for each training observation, using the weights W ; this prediction is used in line 12 to calculate an error for each observation. Line 15 minimizes the error over all observations, where line 17 defines that the errors for the individual observations are combined by summing the squared errors. Whereas this problem is formulated as a generic constraint optimization problem, many machine learning toolkits exist that have highly optimized algorithms for this problem.

The scheduling model and the machine learning model together define both components of the inductive constraint programming loop. We have demonstrated how a declarative, unified language could be used to model both the learning problem and the solving problem. While a single language for both the learning and solving components is an appealing prospect, it is not a requirement for the applicability of the inductive constraint programming loop.

```

1  % Dimension of the input data.
2  int: N; % Number of observations
3  int: M; % Dimension of observations

4  % Input data: observed data (X) and target labels (Y)
5  array[1..N, 1..M] of float: X;
6  array[1..N] of float: Y;

7  % Weights to fit (W[M+1] is constant term)
8  array[1..M+1] of var float: W;

9  % Calculate predictions and errors
10 array[1..N] of var float: Est =
11   [ sum(j in 1..M) (W[j]*X[i,j]) + W[M+1] | i in 1..N ];
12 array[1..N] of var float: Err =
13   [ Est[i] - Y[i] | i in 1..N ];

14 % Minimize the squared error
15 solve minimize norm2(Err);

16 % Auxiliary functions for computing the 2-norm
17 function var float: norm2(array[int] of var float: W) =
18   sum(j in index_set(W)) ( W[j]*W[j] );

```

Figure 4: The hospital learning problem in pseudo-Minizinc.

5 Other Examples of Applications

We now briefly describe a number of other problems that can be captured in the inductive constraint programming loop. For each of them, we define the problem, and then define the interactions between the world, the ML component, and the CP component. The three problems we describe (optimizing bus schedules, car pooling, and energy-aware data centers) are real world problems that can be expressed in a neat and efficient way through the inductive constraint programming loop.

5.1 Optimizing bus schedules

In order to improve human mobility, the region of Pisa plans to take into account information about the trajectories of people taking their car in order to improve the public transportation system. The problem is composed of two parts. The first one consists in tracking the GPS localization of cell phones to understand the way people commute in the Pisa region. The second part consists in optimizing bus schedules to meet as much as possible the requirements of these people. The problem is dynamic as the implementation of the generated bus schedule will affect the way people commute, which should be

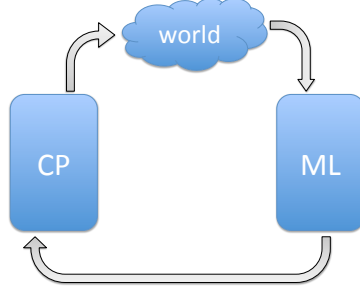


Figure 5: The loop of the bus schedule problem

observed again, and so on. Such a problem can be represented in the inductive constraint programming loop framework.

The world is observed through the GPS localization of cells phones in the region of Pisa and is represented by a set of trajectories of individuals. An evaluation of the traffic quality is provided by the *eval_world* function. *eval_world* is based on a measure of the amount of traffic jams generated by the traffic. All these observations are stored in the **Observations** repository. The function **World-to-ML** reads observations that are given as input to the ML component. The ML component uses these observations to learn patterns from the trajectories and from the quality evaluation of the traffic. These patterns can be periodical traffic densities, commuting patterns of users that travel along specific routes on a regular basis, etc. These learned patterns on the trajectories/time slots are written in the **Patterns** repository. The CP component contains a constraint network that models the problem of generating good bus schedules for the region of Pisa, that is, bus schedules that cover as much as possible trajectories of people at the time they need them. It is parameterized by the weight of each trajectory/time slot. The values of these parameters are computed by the function **ML-to-CP** based on the input from **Patterns**. The output of the CP component is a new bus schedule that is written in the **Solutions** repository to be applied to the world by function **Apply-to-World**. The process can loop for ever. Figure 5 shows the loop solving this problem.

5.2 Carpooling

The carpooling application is aimed at proposing carpooling matches to a set U of users participating in the service, based on their actual mobility (the trips they performed with their private cars) and any information about which kind of match proposals are likely to be accepted by a user. This problem can be modeled in the inductive constraint programming loop framework.

The world is observed through a set of raw trajectories for each user, de-

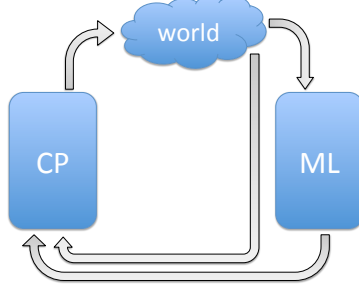


Figure 6: The loop of the carpooling, energy-aware data center, and portfolio selection problems

scribing the recent mobility of that user. The *eval_world* function returns the response of each user to the previous carpooling solution offered to her (accepted, rejected, will reject next time). All these data are written in the **Observations** repository. The function **World-to-ML** simply reads this information and sends it to the ML component, which will perform a data mining analysis of the individuals' mobility to infer their routine trips, which are then compared to produce a temporally labeled weighted oriented graph $G = (U, E, t, p)$, where there exists an edge $(u, v) \in E$ if and only if user v can give a lift to user u . Each edge e is labeled with the time $t(e)$ the lift can take place, and a probability $p(e)$ of having the match proposal being accepted by both u and v . The graph G is written in the **Patterns** repository. **ML-to-CP** reads this graph and encodes it as a constraint network. The observation of the world also provides other information useful to the carpooling system, e.g. the number of passengers that can be hosted in the users vehicle. The function **World-to-CP** reads this information and directly sends it to the CP component, which will add the relevant constraints in the network. For each user u , a constraint $c(u)$ defining the maximum capacity of the vehicle of user u will be added. The CP component solves the network to find the best solution. This solution is a carpooling assignment that maximizes the reduction of cars and the likelihood of acceptance by users. It is written in the **Solutions** repository, then proposed to the users. Figure 6 shows the loop solving this problem.

5.3 Energy-aware Data Centers

The aim is to improve the energy-efficiency of data-centers. Consider a cloud computing service, where customers contract to run computing services (tasks) throughout the day. Tasks are assigned to machines within the data center and require a certain amount of resources for the duration which they run. The aim is to schedule these tasks such that the overall cost of energy used is minimized.

However this is complicated by the fact that large electricity consumers, like a data center, will typically pay a variable price for their electricity, which is not known in advance. In Ireland for example, the actual price on its SEMO market is not known until four days after. The price may also fluctuate significantly throughout the day, which provides the opportunity to reduce the energy use during peak periods and instead perform the work during cheaper periods. This requires a forecast of the price ahead of time and to produce a schedule of the tasks based on the forecasted price data. The price of electricity and its dependence on, e.g. the amount of sun and wind, also changes over time depending on the energy sources used and their operating costs. For this reason, we believe a loop in which new prediction models are regularly built is essential.

This application was the focus of the first inductive CP Challenge¹ and is modeled in the inductive constraint programming loop as follows. The world consists of a number of elements: the wide range of factors which affect the energy market, like weather conditions, producers/consumers of electricity, etc; and customers of the data center who contract the various workloads. **World-to-CP** gives to the CP component the tasks to be scheduled at the next turn; **World-to-ML** takes input from the world to produce a hypothesis h modeling the electricity price. **ML-to-CP** incorporates this forecast to produce a solution to the scheduling problem minimizing the forecast energy cost. **Apply-to-World** takes this schedule and applies it to the world. As time progresses and the world changes **World-to-ML** will need to evolve the forecast model, and subsequently the schedule, to take account of factors affecting the energy price. Figure 6 shows the loop solving this problem.

Here, the machine learning component is first applied and then the outcome is used by the CP program. An alternative, proposed by Tulabandhula and Rudin [TR13], is to do the machine learning while taking the operational cost (the outcome of the CP problem with the learned weights) into account. This can be achieved by making the operational cost a part of the loss function of the ML problem. One can then repeatedly iterate between solving the ML and CP component, before applying the schedule in the world.

6 Academic Examples

Here we present two additional examples of problems: constraint acquisition and algorithm selection. These are existing academic problems that can be seen with a new eye through the inductive constraint programming loop.

6.1 Constraint acquisition

Modeling a problem as a constraint network requires expertise in constraint programming. If we want novices to use constraint programming, we need automatic constraint acquisition systems that assist the user in modeling the problem. CONACQ is such a system [BCOP07]. CONACQ interacts with a user

¹<http://iconchallenge.insight-centre.org/>

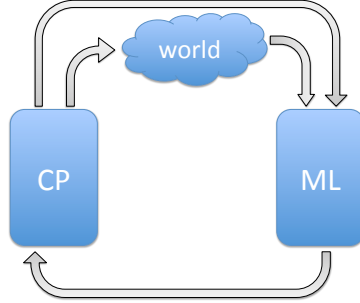


Figure 7: The loop of the constraint acquisition problem

to learn a target constraint network that represents the constraint satisfaction problem of the user; CONACQ does not learn optimization functions. We describe how CONACQ can be implemented as an instance of the inductive constraint programming loop.

The world involves a set of examples defined on a set of objects/variables X . An example is an assignment of a value to each of the variables in X . Examples are produced by the CP component or by the world itself. The evaluation function *eval_world* is the user herself, who evaluates the quality of examples. In CONACQ the quality is either true or false, depending on if the example is a solution of the target constraint network or not. The examples and their evaluation are written in the **Observations** repository. **World-to-ML** simply reads the classified examples from **Observations** and gives them to the ML component. The hypotheses space H used by the ML component is defined by the language of constraints used to express the target network. When reading examples from **Observations**, the ML component updates – if needed – the hypothesis h in the form of a set of constraints that correctly classify the examples. These learned constraints are written in the **Patterns** repository. Based on the learned constraints, the ML component generates other constraints that implement the followed query strategy (e.g., near-miss). These constraints together with the learned constraints are sent to the CP component via the **ML-to-CP** function. The CP component solves the network built with all the constraints provided by the ML component and generates new solutions that are stored in the **Solutions** repository. **Apply-to-World** sends the solutions from the repository to the world to be classified by the user/*eval_world* function. In case the constraint network has no solution that can be provided to the world, the **CP-to-ML** function notifies the ML component that it was not able to generate a satisfactory query, possibly with some reasons of failure such as an inconsistent set of constraints. Figure 7 shows the loop solving this problem.

6.2 Portfolio and algorithm selection

An algorithm portfolio contains a number of algorithms or solvers that are all suitable for solving the same kind of problem, but have different performance characteristics. The aim is to, given a problem instance to solve, determine the best solver for that particular problem, where “best” is defined according to an application-specific metric. Algorithm portfolios have been shown to achieve significant performance improvements over individual algorithms [XHHL08]. One prominent application area of these techniques is constraint programming [HKMO14].

Within the inductive constraint programming loop, this problem can be modeled as a data mining problem that takes its data from runs of CP solvers. The world consists of the performance data of CP solvers on CP instances. *World-to-ML* reads this performance information. The ML component learns a model that describes the predicted performance of CP solvers on CP instances and helps determine the best CP solver for a particular CP instance. When a *new* CP instance to be solved appears in the world, *World-to-ML* and *World-to-CP* read it. *World-to-ML* sends it to the ML component, which classifies it and writes the classification in the *Patterns* repository. *World-to-CP* sends the CP instance to be solved to the CP component. *ML-to-CP* reads in the *Patterns* repository the classification provided by the ML component to decide which solver to use. The selected solver then solves the instance and generates a new data point in the world through the *Apply-to-World* function. This new world can lead the ML component to update the predictive model, and so forth (e.g., [BBP15]). Figure 6 shows the loop solving this problem.

7 Conclusion

After a brief introduction to constraint programming and machine learning, we have introduced the framework of inductive constraint programming. The key idea in the inductive constraint programming loop is that the CP and ML components interact with each other and with the world in order to adapt solutions to changes in the world. This is an essential requirement in problems that change under the effect of time, or problems that are influenced by the application of a previous solution. It is also very effective for problems that are only partially specified and where the ML component learns from observation by applying a partial solution, e.g. in the case of constraint acquisition. We have presented multiple examples of the use of inductive constraint programming loop in real-world problem settings. Many other settings exist, and as the frequency with which learning methods are used to produce schedules and other operational plans, the need for a framework that can adapt to changes in the world will increase.

Acknowledgments

This work was funded by the European Commission under the project Inductive Constraint Programming, contract number FP7-284715. We thank the members of the consortium for the intense collaborations and the project reviewers for their insightful comments.

References

- [BBP15] A. Balafrej, C. Bessiere, and A. Paparrizou. Multi-armed bandits for adaptive constraint propagation. In *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence (IJCAI 2015)*, pages 290–296, Buenos Aires, Argentina, 2015. AAAI Press.
- [BCOP07] C. Bessiere, R. Coletta, B O’Sullivan, and M. Paulin. Query-driven constraint acquisition. In *Proceedings of the Twentieth International Joint Conference on Artificial Intelligence (IJCAI’07)*, pages 44–49, Hyderabad, India, 2007.
- [BH04] R. Bent and P. Van Hentenryck. Online stochastic and robust optimization. In *Proceedings of the 9th Asian Computing Science Conference (ASIAN 04)*, pages 286–300, Chiang Mai, Thailand, 2004. Springer.
- [BHO09] C. Bessiere, E. Hebrard, and B. O’Sullivan. Minimising decision tree size as combinatorial optimisation. In *Proceedings of the 15th International Conference on Principles and Practice of Constraint Programming (CP 2009)*, pages 173–187, Lisbon, Portugal, 2009. Springer.
- [CJSS12] E. Coquery, S. Jabbour, L. Saïs, and Y. Salhi. A sat-based approach for discovering frequent, closed and maximal patterns in a sequence. In *Proceedings of the 20th European Conference on Artificial Intelligence (ECAI 2012)*, pages 258–263, Montpellier, France, 2012. IOS Press.
- [DD88] R. Dechter and A. Dechter. Belief maintenance in dynamic constraint networks. In *Proceedings of the 7th National Conference on Artificial Intelligence (AAAI-88)*, pages 37–42, St. Paul, MN, 1988. AAAI Press / The MIT Press.
- [DGN08] 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 (KDD)*, pages 204–212, Las Vegas, Nevada, 2008. ACM.
- [Doy79] J. Doyle. A truth maintenance system. *Artif. Intell.*, 12(3):231–272, 1979.

- [EF01] S.L. Epstein and E.C. Freuder. Collaborative learning for constraint solving. In *Proceedings of the 7th International Conference on Principles and Practice of Constraint Programming (CP 2001)*, pages 46–60, Paphos, Cyprus, 2001. Springer.
- [HKMO14] B. Hurley, L. Kotthoff, Y. Malitsky, and B. O’Sullivan. Proteus: A hierarchical portfolio of solvers and transformations. In *Proceedings of the 11th International Conference on Integration of AI and OR Techniques in Constraint Programming (CPAIOR 2014)*, pages 301–317, Cork, Ireland, 2014. Springer.
- [KBC10] M. Khiari, P. Boizumault, and B. Crémilleux. Constraint programming for mining n-ary patterns. In *Proceedings of the 16th International Conference on Principles and Practice of Constraint Programming (CP 2010)*, pages 552–567, St. Andrews, Scotland, 2010. Springer.
- [MS14] K. Marriott and P.J. Stuckey. A minizinc tutorial. <http://www.minizinc.org/downloads/doc-latest/minizinc-tute.pdf>, 2014.
- [STS11] J. Shawe-Taylor and S. Sun. A review of optimization methodologies in support vector machines. *Neurocomputing*, 74(17):3609 – 3618, 2011.
- [TR13] T. Tulabandhula and C. Rudin. Machine learning with operational costs. *Journal of Machine Learning Research*, 14:1989–2028, 2013.
- [TVP03] I. Tsamardinos, T. Vidal, and M. E. Pollack. CTP: A new constraint-based formalism for conditional, temporal planning. *Constraints*, 8(4):365–388, 2003.
- [Wal02] T. Walsh. Stochastic constraint programming. In *Proceedings of the 15th European Conference on Artificial Intelligence (ECAI’2002)*, pages 111–115, Lyon, France, 2002. IOS Press.
- [XHHL08] L. Xu, F. Hutter, H. H. Hoos, and K. Leyton-Brown. Satzilla: Portfolio-based algorithm selection for SAT. *J. Artif. Intell. Res. (JAIR)*, 32:565–606, 2008.