Mining Historical Data to build constraint viewpoint

Gilles Raymond, Christian Bessière and Joel Quinqueton

Microsoft Research, CNRS, LIRMM, Université Montpellier II

Introduction

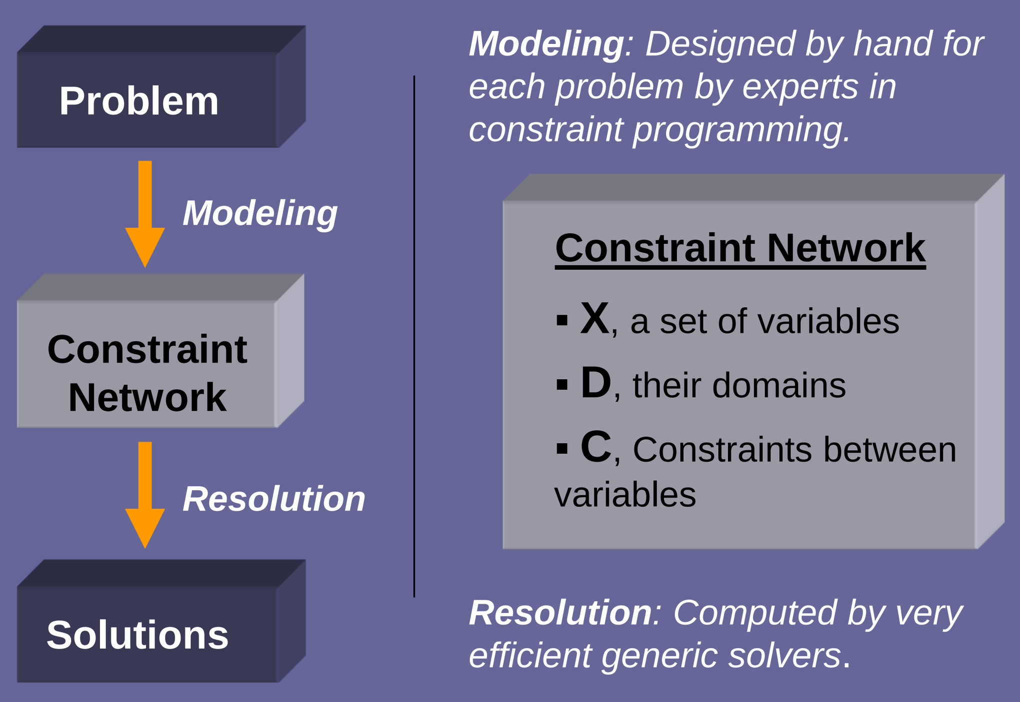
The **Constraint Satisfaction Problem** approach has encountered convincing success in solving real problems in the industry world. However, despite its wide use on some industrial tasks, the constraint satisfaction technology is still reserved for experts in this domain.

If the design of the Constraint network could be partially automated this technology would be applied to a greater number of tasks including consumer software.

Some promising results have been reached in such automated modeling. However all these works suppose that a part of the network is already known. We present here a fully automated approach to build the very first elements of the Constraint network.

CSP Background

The constraint network is a formal description of a specific problem that allow a generic resolution.



A **Constraint Viewpoint** is a constraint network without constraint. It is able to describe any possible instance of the problem.

Constraint Viewpoint

- **X**, a set of variables
- **D**, their domains

First step in automated modeling: Build a Constraint Viewpoint from historical data

Viewpoint modeling main features

- One variable to describe one line in a table.
- Specialization of the root viewpoint variables relying on constant value schemes.
- At certain condition we can assure that the viewpoint built can represent any solution of the target problem.

Modeling Viewpoints

Our first step in constraint network modeling is to build a Viewpoint which will be able to describe the whole solutions of our target problem.

To do so we process known solutions of problem close to that we want to model. This historical data is a set of tables each representing one solution.

Time	Subject	Class	Room
Monday AM	Science	ClassB	Room1
Mon			
Mon			

Time	Subject	Class	Room
Monday AM	English	ClassA	Room1
Mon			
Mon			

Time	Subject	Class	Room
Monday AM	Science	ClassB	Room1
Monday AM	Math	ClassC	Room2
Monday PM	Science	ClassB	Room1
...	...	...	...

Sample history for a school timetable problem. Each table represents a valid solution to the problem encountered in previous years.

A trivial modeling: The root viewpoint

One variable for each lessons

Year 2000

Time	Subject	Class	Room
Monday 8h	English	Class5	Room1
Monday 8h	Math	Class3	Room3
Monday 8h	Science	Class6	RoomTP
...	...	...	...

→ X1 = (Monday_8h, English, Class5, Room1)
→ X2 = (Monday_8h, Math, Class3, Room3)
→ X3 = (Monday_8h, Science, Class6, RoomTP)
→ Xn = (... , ... , ... , ...)

Year 2001

Time	Subject	Class	Room
Monday 8h	Math	Class3	Room2
Monday 10h	Math	Class4	Room1
Monday 10h	English	Class3	Room2
...	...	...	...

→ X1 = (Monday_8h, Math, Class3, Room2)
→ X2 = (Monday_10h, Math, Class4, Room1)
→ X3 = (Monday_10h, Science, Class3, Room2)
→ Xn = (... , ... , ... , ...)

D(Xi) = D(Time) x D(Subject) x D(Class) x D(Room)

Specializing the root viewpoint variables:

Representations in the root viewpoint

In year 2000

Monday PM	Math	ClassC	Room1	**Xi** = **(MondayAM, Math, ClassC, Room1)**

In year 2001

Monday PM	Science	ClassB	Room1	**Xj** = **(MondayAM, Science, ClassB, Room1)**

In year 2002

Monday PM	Science	ClassA	Room1	**Xk** = **(MondayAM, Science, ClassA, Room1)**

Representations with a specialised variable

In year 2000

Monday PM	Math	ClassC	Room1	**X**(MondayAM, Room1) = **(Math, ClassC)**

In year 2001

Monday PM	Science	ClassB	Room1	**X**(MondayAM, Room1) = **(Science, ClassB)**

In year 2002

Monday PM	Science	ClassA	Room1	**X**(MondayAM, Room1) = **(Science, ClassA)**

D(X(MondayAM, Room1)) = D(Subject) x D(Class)

Homogeneous viewpoints

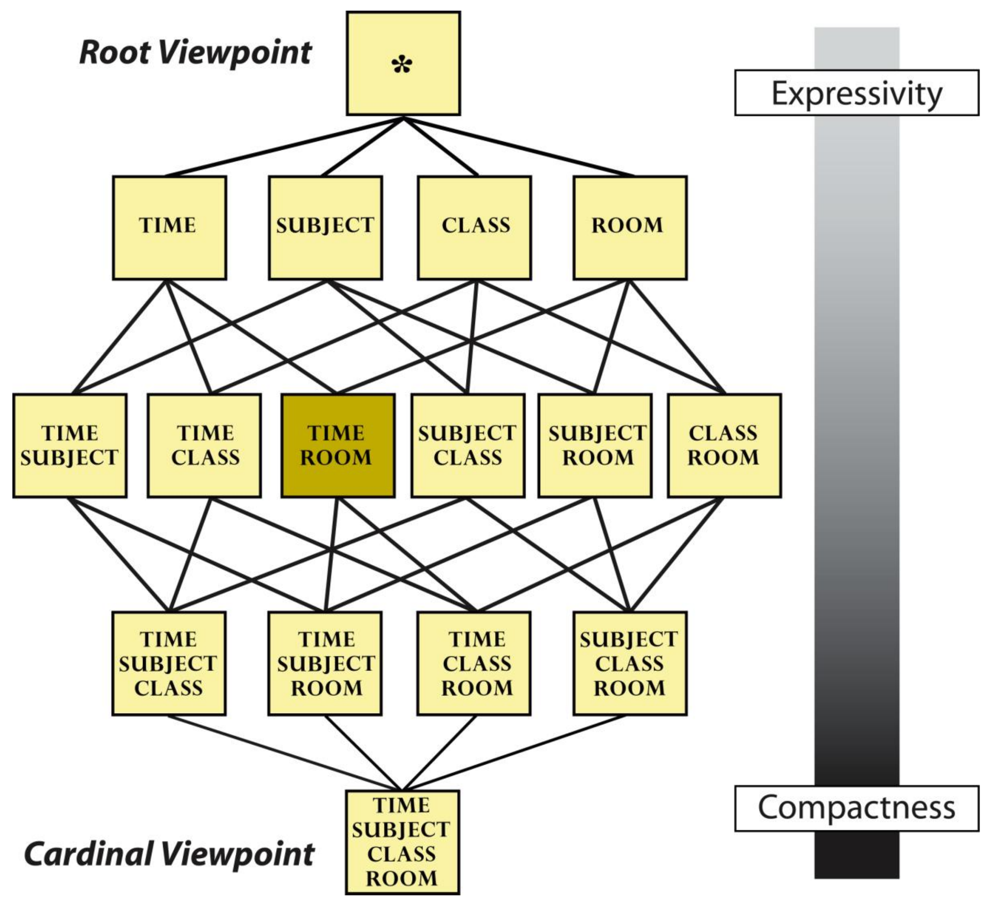
School Timetable Year 1

Time	Subject	Class	Room
Monday AM	English	ClassA	Room1
Monday AM	Math	ClassB	Room2
Monday PM	Math	ClassC	Room1
Monday PM	Science	ClassA	Room2
Tuesday AM	Science	ClassC	Room1
Tuesday AM	English	ClassB	Room2
...	...	...	...

Solution for Year 1

X1 (Monday AM, Room1) = **(English, ClassA)**
X2 (Monday AM, Room2) = **(Math, ClassB)**
X3 (Monday PM, Room1) = **(Math, ClassC)**
X4 (Monday PM, Room2) = **(Science, ClassA)**
X5 (Tuesday AM, Room1) = **(Science, ClassC)**
X6 (Tuesday AM, Room2) = **(English, ClassB)**
Xn (..., ...) = **(... , ...)**

D(Xi(Time, Room)) = D(Subject) x D(Class)



Homogeneous viewpoints can be arranged in a lattice based on the scheme of attributes they are built from.

Constant Heterogeneous viewpoints

School Timetable Year 1

Time	Subject	Class	Room
Monday AM	English	ClassA	Room1
Monday AM	Math	ClassB	Room2
Monday PM	Math	ClassC	Room1
Monday PM	Science	ClassA	Room2
Tuesday AM	Science	ClassC	Room1
Tuesday AM	English	ClassB	Room2
Tuesday PM	Science	ClassB	Room1
TuesdayPM	English	ClassA	Room2
ThursdayAM	Science	ClassC	Room1
ThursdayAM	Math	ClassB	Room2
Thursday PM	Science	ClassB	Room1
Thursday PM	Math	ClassA	Room2
Friday AM	English	ClassC	Room1
Friday AM	Math	ClassA	Room2
Friday PM	Science	ClassA	Room1
Friday PM	Math	ClassC	Room2

Solution for Year 1

X1 (ClassA) = **(Monday AM, English, Room1)**
X2 (Monday AM, Math, Room2) = **(ClassB)**
X3 (Math, Room1) = **(Monday PM, ClassC)**
X4 (ClassA) = **(Monday PM, Science, Room2)**
X5 (Science) = **(Tuesday AM, ClassC, Room1)**
X6 (Tuesday AM, English, Room2) = **(ClassB)**
X7 (ClassB) = **(Tuesday PM, Science, Room1)**
X8 (ClassA, Room2) = **(Tuesday PM, English)**
X9 (Science, ClasseC, Room1) = **(Thursday AM)**
X10(ClassB) = **(Thursday AM, Math, Room2)**
X11(Thursday PM, ClassB, Room1) = **(Science)**
X12(ClassA) = **(Thursday PM, Math, Room2)**
X13(English, Room1) = **(Friday AM, ClassC)**
X14(Math, ClassA, Room2) = **(Friday AM)**
X15(Room1) = **(Friday PM, Science, ClassA)**
X16(ClassC) = **(Friday PM, Math, Room2)**

School Timetable Year 2

Time	Subject	Class	Room
Monday AM	Science	ClassB	Room1
Monday AM	Math	ClassC	Room2
Monday PM	Science	ClassB	Room1
Monday PM	Math	ClassC	Room2
Tuesday AM	Science	ClassC	Room1
Tuesday AM	English	ClassA	Room2
Tuesday PM	Math	ClassB	Room1
TuesdayPM	Science	ClassA	Room2
ThursdayAM	English	ClassC	Room1
ThursdayAM	Math	ClassA	Room2
Thursday PM	Math	ClassB	Room1
Thursday PM	English	ClassA	Room2
Friday AM	Science	ClassA	Room1
Friday AM	English	ClassB	Room2
Friday PM	Science	ClassC	Room1
Friday PM	Math	ClassA	Room2

Solution for Year 2

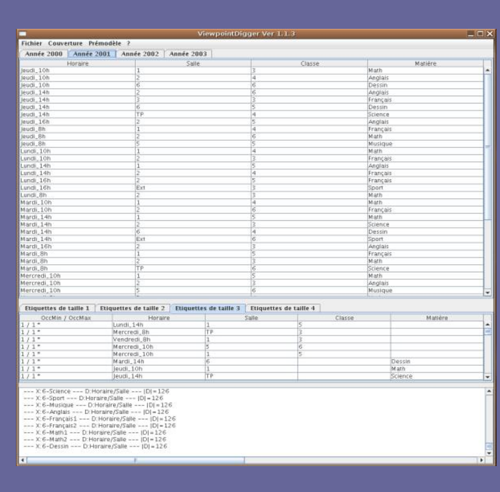
X10(ClassB) = **(Monday AM, Science, Room1)**
X2 (Monday AM, Math, Room2) = **(ClassC)**
X15(Room1) = **(Monday PM, Science, ClassB)**
X16(ClassC) = **(Monday PM, Math, Room2)**
X9 (Science, ClasseC, Room1) = (**Tuesday AM)**
X6 (Tuesday AM, English, Room2) = **(ClassA)**
X3 (Math, Room1) = **Tuesday PM, ClassB)**
X4 (ClassA) = **(Tuesday PM, Science, Room2)**
X13(English, Room1) = (**Thursday AM, ClassC)**
X14(Math, ClassA, Room2) = **(Thursday AM)**
X11(Thursday PM, ClassB, Room1) = **(Math)**
X1 (ClassA) = **(Thursday PM, English, Room2)**
X12(ClassA) = **(Friday AM, Science, Room1)**
X7 (ClassB) = **(Friday AM, English, Room2)**
X5 (Science) = **(Friday PM, ClassC, Room1)**
X8 (ClassA, Room2) = **(Friday PM, Math)**

D(Xi(ClassA)**)** = **D(Time) x D(Subject) x D(Room)**

D(Xi(Tuesday AM, English, Room2)**)** = **D(Class)**

Results

- Our method extracts every constant value schemes present in the historical data. The theoretical complexity is **O(n!)**, where n is the number of attributes of the table.
- Our algorithm use the properties of the inclusion lattice of the constant elements to reduce the practical complexity.
- The ***ViewpointDigger*** tool, realised in Java, does implements this optimized algorithm to process historical data and build viewpoints.
- We obtain satisfying homogeneous viewpoints on typical problems such as **n-queens**, **timetable design**, **sudoku** …



Screenshots from the ViewpontDigger tool, research of homogeneous constant viewpoints.

Conclusion

We have designed an automatic method to extract different kind of viewpoints from historical data (Heterogenous viewpoints, homogeneous, …).

This first step in automated Constraint Network modeling is very promising and offers several possible evolutions for the continuation:

- Build and exploitation of heterogenous viewpoints
- Addition of user interaction to correct learning mistakes.
- Use of these viewpoints as Constraint Network modeling help for human designers.
- Merging several viewpoints in a redundant viewpoint to increase relevance of the model.
- Use of existing constraint learning techniques to complete the Constraint Network.