Systematic Derivation of Bounds and Glue Constraints for Time-Series Constraints

Ekaterina Arafailova, Nicolas Beldiceanu, Mats Carlsson, Pierre Flener, M. Andreína Francisco R., Justin Pearson, and Helmut Simonis

7th September 2016

CP 2016











SWEDIS













A subsequence $\langle X_i, \ldots, X_j \rangle$ of $\langle X_0, \ldots, X_m \rangle$ is a valley if the signature of $\langle X_{i-1}, \ldots, X_{j+1} \rangle$ is a maximal word matching '>(>|=)*(<|=)*<'.



A subsequence $\langle X_i, \ldots, X_j \rangle$ of $\langle X_0, \ldots, X_m \rangle$ is a valley if the signature of $\langle X_{i-1}, \ldots, X_{j+1} \rangle$ is a maximal word matching '>(>|=)*(<|=)*<'.

Compositional Time-Series Definition by Multiple Layers of Functions



max_min_valley((5, 5, 6, 4, 6, 6, 4, 2, 4, 4, 1, 1, 1, 1, 3, 0), 4) holds

Space of Time-Series Constraints



Space of Time-Series Constraints ↑Features one surface width min



Main Question

How to go from a compositional constraint definition to compositional combinatorial objects that can be used in different contexts, such as CP, MIP, local search, and data mining?

Main Question

How to go from a compositional constraint definition to compositional combinatorial objects that can be used in different contexts, such as CP, MIP, local search, and data mining?

Time-series constraints are very numerous: we cannot afford to consider each of them separately.

Applications of Time-Series Constraints

- Analysis of output of electric power stations [CP 2013] over multiple days.
- Power management for [Computing, 2016] large-scale-distributed systems.
- Trace analysis for Internet Service Provider. (CeADAR)
- Anomaly detection and error correction (Campus21) in building data.

Use a transducer:

[Constraints, 2016]

- Input sequence: Handle the matching aspect
- Output sequence: Handle the computational aspect

Use a transducer:

[Constraints, 2016]

- Input sequence: Handle the matching aspect
- Output sequence: Handle the computational aspect



Transducer for valley

Use a transducer:

[Constraints, 2016]

- Input sequence: Handle the matching aspect
- Output sequence: Handle the computational aspect



Transducer for valley

Use a **transducer**:

[Constraints, 2016]

- Input sequence: Handle the matching aspect
- Output sequence: Handle the computational aspect



Use a transducer:

[Constraints, 2016]

- Input sequence: Handle the matching aspect
- Output sequence: Handle the computational aspect



Implementing Time-Series Constraints

- CP: Reformulation of automata with accumulators*
- MIP: Linear reformulation of automata with accumulators**

* [Constraints, 2005], ** [CPAIOR 2016]

- Goal: Address the combinatorial aspect of time-series constraints.
- Method: Exploit the compositional nature of time-series constraints at the combinatorial level.

Contents



Compositional Sharp Bounds: Contents



Context and Goal

Consider a $g_f_{\sigma}(\langle X_0, \ldots, X_m \rangle, N)$ time-series constraint, for a regular expression σ , a feature f, and an aggregator g, with every X_i ranging over the same integer interval [a, b].

Context and Goal

Consider a $g_f_\sigma(\langle X_0, \ldots, X_m \rangle, N)$ time-series constraint, for a regular expression σ , a feature f, and an aggregator g, with every X_i ranging over the same integer interval [a, b].

Goal: Derive formulae $\beta_{f,g}^{u}$ and $\beta_{f,g}^{\ell}$ that are **parametrised by** σ and yield respectively sharp upper and lower bounds on the value of *N*.

Compositional Sharp Bounds: Contents



Computing a Sharp Upper Bound for the pair $\langle min, max \rangle$

Consider a max_min_ $\sigma(\langle X_0, \ldots, X_m \rangle, N)$ time-series constraint with every X_i ranging over [a, b].

Goal: Compute a sharp upper bound on the value of *N*.

Computing a Sharp Upper Bound for the pair $\langle min, max \rangle$

Consider a max_min_ $\sigma(\langle X_0, \ldots, X_m \rangle, N)$ time-series constraint with every X_i ranging over [a, b].

Goal: Compute a sharp upper bound on the value of *N*.

Method:

1. Introduce the characteristics of a regular expression σ : the minimal difference between the domain upper bound b and the minima of the subseries, which is called the shift of σ and denoted by Δ_{σ} .

Computing a Sharp Upper Bound for the pair $\langle min, max \rangle$

Consider a max_min_ $\sigma(\langle X_0, \ldots, X_m \rangle, N)$ time-series constraint with every X_i ranging over [a, b].

Goal: Compute a sharp upper bound on the value of *N*.

Method:

- 1. Introduce the characteristics of a regular expression σ : the minimal difference between the domain upper bound b and the minima of the subseries, which is called the shift of σ and denoted by Δ_{σ} .
- **2.** Explore the connection between this characteristics and the sharp upper bound on the value of *N*.

Compositional Sharp Bounds: Contents



Computing a Sharp Upper Bound for max_min_valley

Consider the max_min_valley($\langle X_0, \ldots, X_8 \rangle$, N) time-series constraints with every X_i ranging over [a, b] = [3, 6].

Shift of valley and Its Connection to a Sharp Upper Bound

The shift of valley is the smallest possible difference between b and the minima of valleys, namely 1 and is reached, for example, for valleys whose signature is '><'.</p>

Shift of valley and Its Connection to a Sharp Upper Bound

- The shift of valley is the smallest possible difference between b and the minima of valleys, namely 1 and is reached, for example, for valleys whose signature is '><'.</p>
- ► The minimum of any valley is at least Δ_{valley} , which is 1, away from *b*, which is 6, thus $N \le b \Delta_{valley} = 6 1 = 5$.

Shift of valley and Its Connection to a Sharp Upper Bound

- The shift of valley is the smallest possible difference between b and the minima of valleys, namely 1 and is reached, for example, for valleys whose signature is '><'.</p>
- ► The minimum of any valley is at least Δ_{valley} , which is 1, away from *b*, which is 6, thus $N \le b \Delta_{valley} = 6 1 = 5$.
- This bound is reached for some ground time series.

Shift of valley and Its Connection to a Sharp Upper Bound

- The shift of valley is the smallest possible difference between b and the minima of valleys, namely 1 and is reached, for example, for valleys whose signature is '><'.</p>
- The minimum of any valley is at least Δ_{valley}, which is 1, away from b, which is 6, thus N ≤ b − Δ_{valley} = 6 − 1 = 5.
- This bound is reached for some ground time series.



14 / 24

Compositional Sharp Bounds: Contents



Sharp Upper Bound for the General Case

Theorem

Consider a max_min_ $\sigma(\langle X_0, \ldots, X_m \rangle, N)$ time-series constraint, where σ is one of the 22 regular expressions and every X_i ranges over [a, b], then the sharp upper bound on the value of N is

$$eta_{\texttt{min,max}}^{ extsf{u}} = b - \Delta_{\sigma}$$



Contents



For the ngroup($\langle X_0, \ldots, X_m \rangle, S, N$) constraint, N is the number of maximal subsequences of $\langle X_0, \ldots, X_m \rangle$ whose values are all in S.

For the ngroup($\langle X_0, \ldots, X_m \rangle, S, N$) constraint, N is the number of maximal subsequences of $\langle X_0, \ldots, X_m \rangle$ whose values are all in S.

The ngroup($\langle 1, 0, 1, 1, 0, 0 \rangle$, $\{1\}$, 2) constraint holds.

For the ngroup($\langle X_0, \ldots, X_m \rangle, S, N$) constraint, N is the number of maximal subsequences of $\langle X_0, \ldots, X_m \rangle$ whose values are all in S.

The ngroup($(1, 0, 1, 1, 0, 0), \{1\}, 2$) constraint holds.



For the ngroup($\langle X_0, \ldots, X_m \rangle, S, N$) constraint, N is the number of maximal subsequences of $\langle X_0, \ldots, X_m \rangle$ whose values are all in S.

The ngroup($(1, 0, 1, 1, 0, 0), \{1\}, 2$) constraint holds.



not just a sum ($2 \neq 2 + 1$ **)**...

For the ngroup($\langle X_0, \ldots, X_m \rangle, S, N$) constraint, N is the number of maximal subsequences of $\langle X_0, \ldots, X_m \rangle$ whose values are all in S.

The ngroup($\langle 1, 0, 1, 1, 0, 0 \rangle$, $\{1\}, 2$) constraint holds.



not just a sum $(2 \neq 2+1)$...how to compute the correction term -1?



Automaton for ngroup



Automaton for ngroup







Automaton for ngroup

Glue Matrix for ngroup







Automaton for ngroup

Glue Matrix for ngroup

 $ngroup((1, 0, 1, 1, 0, 0), \{1\}, 2)$ 1, 0, 1, 0, 1, 0 t, t, **u**, **u**, t, t, и, t $ngroup((1, 0, 1), \{1\}, 2)$ $ngroup((0, 0, 1), \{1\}, 1)$

Compositional Glue Matrices for Time-Series Constraints

Goal: For a transducer, derive a glue matrix that is parametrised by a feature and an aggregator.

Generic Glue Matrix for steady sequence = $=^{+}$



 $\begin{array}{l} \mbox{Generic Automaton for} \\ \mbox{g}_f_steady_sequence \end{array}$

Generic Glue Matrix for steady sequence = $=^{+}$

$$\begin{cases} C \leftarrow default_{gf} \\ D \leftarrow id_{f} \\ R \leftarrow default_{gf} \end{cases}$$

$$\begin{cases} \downarrow \neq S \\ \downarrow = S \\ \downarrow \neq S \\ \downarrow \neq S \\ \downarrow \neq S \\ \downarrow = S \\ \downarrow \neq S \\ \downarrow = S \\ \downarrow \neq S \\ \downarrow = S \\ \downarrow$$

$$\begin{array}{c|c} \mathsf{s} & \phi_{g}(\overrightarrow{C},\overleftarrow{C}) & \phi_{g}(\overrightarrow{C},\overleftarrow{C}) \\ \hline \phi_{g}(\overrightarrow{C},\overrightarrow{C}) & \phi_{f}(\overrightarrow{C},\overrightarrow{C},\overrightarrow{D},\overrightarrow{D},\delta_{f}^{i}) \end{array}$$

r

Generic Glue Matrix for $g_f_steady_sequence$

Corresponding Compositional Glue Matrices



Contents



Synergy of Sharp Bounds and Glue Matrices

Consider a $g_f_\sigma(\langle X_0, \ldots, X_m \rangle, N)$ time-series constraint where all X_i range over the same domain.

► Add sharp bounds for every prefix and every reversed suffix of X₀,..., X_m.

Impose glue-matrix implied constraints on every prefix and every reversed suffix of X₀,..., X_m.

The combination of sharp bounds and glue matrices improves propagation.

Conclusion: clear separation between three levels

- Language: compositional description of constraints with aggregator g, feature f, regular expression σ
- Abstract Combinatorial Objects (transducers and formulae)
 - Transducers for each regular expression σ
 - Parametrised bounds for each pair g,f
 - Parametrised glue matrices for each regular expression σ (22)

► Multiple usages (synthesised code: LATEX(154100), Prolog (135267))

- ► **CP**: enhanced propagation (bounds, glue matrices)
- ► Local search: constant time probing (glue matrices)
- Data mining: range of variation of feature

Separate definition from usage, be compositional at all levels

(22) (20)

(bounds)

Forthcoming in late 2016:

Automatically synthesised catalogue of time-series constraints, with all transducers, constraints, bounds, glue matrices, linear invariants, and MiniZinc reformulations.

Forthcoming in late 2016:

Automatically synthesised catalogue of time-series constraints, with all transducers, constraints, bounds, glue matrices, linear invariants, and MiniZinc reformulations.

Thank you for your attention!



 Arafailova, Ekaterina et al. (2016). "Time-series constraints: Improvements and application in CP and MIP contexts". In: *CPAIOR 2016.* Ed. by Claude-Guy Quimper. Vol. 9676. LNCS. Springer, pp. 18–34.

Beldiceanu, Nicolas et al. (2005). "Reformulation of global constraints based on constraints checkers". In: *Constraints* 10.4, pp. 339–362.

- Beldiceanu, Nicolas et al. (2013). "Describing and generating solutions for the EDF unit commitment problem with the ModelSeeker". In: *CP 2013*. Ed. by Christian Schulte. Vol. 8124. LNCS. Springer, pp. 733–748.
- Beldiceanu, Nicolas et al. (2016a). "Towards energy-proportional Clouds partially powered by renewable energy". In: Computing, p. 20. url: https://hal.inria.fr/hal-01340318.
 Beldiceanu, Nicolas et al. (2016b). "Using finite transducers for describing and synthesising structural time-series constraints". In: Constraints 21.1, pp. 22–40.