Online Timed Pattern Matching using Derivatives

Dogan Ulus^{\dagger}, Thomas Ferrère^{\dagger}, Eugene Asarin^{\dagger †}, and Oded Maler^{\dagger}

[†] VERIMAG, Université Grenoble-Alpes & CNRS, France ^{††} IRIF, Université Paris Diderot, France

April 7, 2016

TACAS 2016, Eindhoven

The Problem of Finding All Sprints

- Assume you're a football manager interested in formal methods.
- > You want to **find all sprints** by a player **formally**.
- You have a sprint specification:
 - ► A period of high acceleration followed by a period of high speed for 1 second at least.
- How to solve?

Hint: This is a problem of timed pattern matching.

The Problem of Finding All Sprints

- Assume you're a football manager interested in formal methods.
- > You want to **find all sprints** by a player **formally**.
- You have a sprint specification:
 - ► A period of high acceleration followed by a period of high speed for 1 second at least.
- How to solve?

Hint: This is a problem of timed pattern matching.

The Problem of Finding All Sprints

Given data and pattern, we perform timed pattern matching.



Figure: All sprints by Olcan Adin, Galatasaray v Sivasspor, 16 Jan 2016.

Timed Pattern Matching

Finding all segments of a dense-time signal that satisfy a timed regular expression.

For given an expression φ and a signal w:

$$\mathcal{M}(\varphi,w) = \{ \begin{array}{cc} (t,t') & \mid w[t,t'] \models \varphi \end{array} \}$$

Example:

p

q



Timed Pattern Matching

Finding all segments of a dense-time signal that satisfy a timed regular expression.

For given an expression φ and a signal w:

$$\mathcal{M}(\varphi,w) = \{ (t,t') \mid w[t,t'] \models \varphi \}$$

t'

10

Example:

- A pattern $\varphi = \langle p \cdot q \rangle_{[4,7]}$.
- A signal w over p and q.



- Previously we presented an offline procedure in [UFAM'14].
- We now develop an online procedure to
 - Perform matching on streams.
 - Monitor and alert in real time.
 - Have better memory performance.
- Our online procedure is based on the concept of **derivatives**.

- A recipe is a list of actions to be done sequentially.
- ► To-do list for an eggplant puree:
 - Wash and prick the eggplants with a fork.
 - Bake eggplants for 25 mins.
 - Smash eggplants.
 - Add flour.
 - Add milk and mix it well.
- Each time, you complete an action, you delete the item.
- Do everything correct and you will end up the empty list. (You are missing the point if you think of the puree.)
- Obviously this is an online acceptance procedure.

- A recipe is a list of actions to be done sequentially.
- To-do list for an eggplant puree:
 - Wash and prick the eggplants with a fork.
 - Bake eggplants for 25 mins.
 - Smash eggplants.
 - Add flour.
 - Add milk and mix it well.
- Each time, you complete an action, you delete the item.
- Do everything correct and you will end up the empty list.
 (You are missing the point if you think of the puree.)
- Obviously this is an online acceptance procedure.

- A recipe is a list of actions to be done sequentially.
- ► To-do list for an eggplant puree:
 - Wash and prick the eggplants with a fork.
 - Bake eggplants for 25 mins.
 - Smash eggplants.
 - Add flour.
 - Add milk and mix it well.
- Each time, you complete an action, you delete the item.
- Do everything correct and you will end up the empty list.
 (You are missing the point if you think of the puree.)
- Obviously this is an online acceptance procedure.

- A recipe is a list of actions to be done sequentially.
- To-do list for an eggplant puree:
 - Wash and prick the eggplants with a fork.
 - Bake eggplants for 25 mins.
 - Smash eggplants.
 - Add flour.
 - Add milk and mix it well.
- Each time, you complete an action, you delete the item.
- Do everything correct and you will end up the empty list. (You are missing the point if you think of the puree.)
- Obviously this is an online acceptance procedure.

Recipe Analogy

- A recipe is a list of actions to be done sequentially.
- To-do list for an eggplant puree:
 - Wash and prick the eggplants with a fork.
 - Bake eggplants for 25 mins.
 - Smash eggplants.
 - Add flour.
 - Add milk and mix it well.
- Each time, you complete an action, you delete the item.
- Do everything correct and you will end up the empty list. (You are missing the point if you think of the puree.)
- Obviously this is an online acceptance procedure.

Recipe Analogy

- A recipe is a list of actions to be done sequentially.
- To-do list for an eggplant puree:
 - Wash and prick the eggplants with a fork.
 - Bake eggplants for 25 mins.
 - Smash eggplants.
 - ► Add flour.
 - Add milk and mix it well.
- Each time, you complete an action, you delete the item.
- Do everything correct and you will end up the empty list. (You are missing the point if you think of the puree.)
- Obviously this is an online acceptance procedure.

Recipe Analogy

- A recipe is a list of actions to be done sequentially.
- ► To-do list for an eggplant puree:
 - Wash and prick the eggplants with a fork.
 - Bake eggplants for 25 mins.
 - Smash eggplants.
 - ► Add flour.
 - Add milk and mix it well.
- Each time, you complete an action, you delete the item.
- Do everything correct and you will end up the empty list. (You are missing the point if you think of the puree.)
- Obviously this is an online acceptance procedure.



Definition

The derivative of a language $\mathcal L$ over Σ^* with respect to a word u is defined as

$$D_u(\mathcal{L}) := \{ v \in \Sigma^* \mid u \cdot v \in \mathcal{L} \}.$$

Language Membership

$$w \in \mathcal{L}$$
 iff $\epsilon \in D_w(\mathcal{L})$

- Compute derivatives and check empty word containment.
- Pattern matching is more than that: Membership queries for all subwords.

Derivatives of Regular Expressions

RE syntax:

$$r := \varnothing \mid \epsilon \mid a \mid r_1 \cdot r_2 \mid r_1 \lor r_2 \mid r^*$$

 The derivative of a regular expression r with respect to a letter a can be found recursively by these rewrite rules. (Brzozowski 1964)

Empty Word Extraction

$ u(\varnothing)$	=	Ø	$ u(r_1 \cdot r_2)$	=	$ u(r_1) \cdot \nu(r_2) $
$ u(\epsilon)$	=	ϵ	$ u(r_1 \lor r_2)$	=	$\nu(r_1) \lor \nu(r_2)$
u(a)	=	Ø	$ u(r^*)$	=	ϵ

Derivatives of Regular Expressions

$D_a(\emptyset)$	=	Ø
$D_a(\epsilon)$	=	Ø
$D_a(a)$	=	ϵ
$D_a(b)$	=	Ø

$$\begin{array}{rcl} D_{a}(r_{1} \cdot r_{2}) &=& D_{a}(r_{1}) \cdot r_{2} \lor \nu(r_{1}) \cdot D_{a}(r_{2}) \\ D_{a}(r_{1} \lor r_{2}) &=& D_{a}(r_{1}) \lor D_{a}(r_{2}) \\ D_{a}(r^{*}) &=& D_{a}(r) \cdot r^{*} \end{array}$$

Symbols			a	b	С	b	с
Positions			1	2	3	4	5
1	φ	$\xrightarrow{D_{a}}$	$a^{*}\left(bc ight)^{*}$				

$$\epsilon \in D_a(\varphi) \to a \in \varphi$$

Symbols			a	b	С	b	С
Positions			1	2	3	4	5
1	φ	$\overrightarrow{D_{a}}$	$a^{*}(bc)^{*}$	$\xrightarrow{D_b} c(bc)^*$			

$$\epsilon \notin D_{ab}(\varphi) \to ab \notin \varphi$$

Symbols			a		Ь		с	b	с	
Positions			1		2		3	4	5	
1	φ	$\xrightarrow{D_a}$	$a^{*}(bc)^{*}$	$\overrightarrow{D_{h}}$	$c(bc)^*$	$\xrightarrow{D_c}$	$(bc)^*$			

$$\epsilon \in D_{abc}(\varphi) \to abc \in \varphi$$

Symbols			a		b		с		b	с
Positions			1		2		3		4	5
1	φ	$\xrightarrow{D_a}$	$a^{*}(bc)^{*}$	$\xrightarrow{D_h}$	$c(bc)^*$	$\overrightarrow{D_c}$	$(bc)^*$	$\xrightarrow{D_h}$	$c(bc)^*$	

$$\epsilon \notin D_{abcb}(\varphi) \to abcb \notin \varphi$$

Symbols			a		b		с		b		с
Positions			1		2		3		4		5
1	φ	$\overrightarrow{D_a}$	$a^{*}(bc)^{*}$	$\overrightarrow{D_{b}}$	$c(bc)^*$	$\overrightarrow{D_c}$	$(bc)^*$	$\xrightarrow{D_b}$	$c(bc)^*$	$\overrightarrow{D_c}$	$(bc)^*$

$$\epsilon \in D_{abcbc}(\varphi) \to abcbc \in \varphi$$

Example: Let $\varphi = a^*(bc)^*$ and w = abcbc.

Symbols			a		b		с		b		с
Positions			1		2		3		4		5
1	φ	$\overrightarrow{D_a}$	$a^{*}(bc)^{*}$	$\overrightarrow{D_b}$	$c(bc)^*$	$\overrightarrow{D_c}$	$(bc)^*$	$\overrightarrow{D_b}$	$c(bc)^*$	$\overrightarrow{D_c}$	$(bc)^*$
2			φ	$\overrightarrow{D_{L}}$	$c(bc)^*$	$\overrightarrow{D_c}$	$(bc)^*$	$\overrightarrow{D_{L}}$	$c(bc)^*$	$\overrightarrow{D_c}$	$(bc)^*$
3				0	φ	$\overrightarrow{D_c}$	ø	$\overrightarrow{D_{b}}^{0}$	ø	$\overrightarrow{D_c}^c$	ø
4							φ	$\xrightarrow{D_1}$	$c(bc)^*$	$\xrightarrow{D_{a}}$	$(bc)^*$
5								- 0	φ	$\xrightarrow{D_c}$	ø

Below all segments of w that satify the expression φ :

 $\mathcal{M}(\varphi, w) = \{(1, 1), (1, 3), (1, 5), (2, 3), (2, 5), (4, 5)\}$

Transition to Timed

- Each action takes some time.
- Duration of actions can be constrained to be between m an n time units such that

 $\langle \mathsf{Actions} \rangle_{[m,n]}$

The expression specifying all correct timings for our recipe:

$$\left< \mathsf{Wash} \cdot \left< \mathsf{Bake} \right>_{[23,27]} \cdot \mathsf{Smash} \cdot \mathsf{Flour} \cdot \mathsf{Milk} \right>_{[0,60]}$$

- For online matching such specifications, you have to:
 - Remember how much time passed for actions.
 - Do it for uncountable number of start points.

- Adding absolute timestamps to our recipe.
- Very punctual recipe:
- 6h00 6h05 Wash and prick the eggplants with a fork.
- 6h05 6h30 Bake eggplants.
- 6h30 6h45 Smash eggplants.
- 6h45 6h48 Add flour.
- 6h48 6h50 Add milk and mix it well.
 - ▶ Now, when we complete an action, we put a checkmark.
 - You will end up a fully checkmarked list with stamps.
 - This is more informative: Now we can directly say the puree is cooked between 6h00 and 6h50.
 - ► WHY: It simplifies a lot when we deal with uncountable sets.

- Adding absolute timestamps to our recipe.
- Very punctual recipe:
- 6h00 6h05
- 6h05 6h30 Bake eggplants.
- 6h30 6h45 Smash eggplants.

 \checkmark

- 6h45 6h48 Add flour.
- 6h48 6h50 Add milk and mix it well.
 - Now, when we complete an action, we put a checkmark.
 - You will end up a fully checkmarked list with stamps.
 - This is more informative: Now we can directly say the puree is cooked between 6h00 and 6h50.
 - ► WHY: It simplifies a lot when we deal with uncountable sets.

- Adding absolute timestamps to our recipe.
- Very punctual recipe:
- 6h00 6h05
- 6h05 6h30
- \checkmark 6h30 - 6h45 Smash eggplants.

 \checkmark

- 6h45 6h48 Add flour.
- 6h48 6h50 Add milk and mix it well.
 - Now, when we complete an action, we put a checkmark.

- Adding absolute timestamps to our recipe.
- Very punctual recipe:

./

- 6h00 6h05
- 6h05 6h30 🗸
- 6h30 6h45 🗸
- 6h45 6h48 Add flour.

6h48 - 6h50 Add milk and mix it well.

- Now, when we complete an action, we put a checkmark.
- You will end up a fully checkmarked list with stamps.
- This is more informative: Now we can directly say the puree is cooked between 6h00 and 6h50.
- ► WHY: It simplifies a lot when we deal with uncountable sets.

- Adding absolute timestamps to our recipe.
- Very punctual recipe:

./

 \checkmark

- 6h00 6h05
- 6h05 6h30 🗸
- 6h30 6h45 🗸 🗸
- 6h45 6h48

6h48 - 6h50 Add milk and mix it well.

- ▶ Now, when we complete an action, we put a checkmark.
- You will end up a fully checkmarked list with stamps.
- This is more informative: Now we can directly say the puree is cooked between 6h00 and 6h50.
- ► WHY: It simplifies a lot when we deal with uncountable sets.

- Adding absolute timestamps to our recipe.
- Very punctual recipe:

 \checkmark

- 6h00 6h05 🗸 🗸
- 6h05 6h30
- 6h30 6h45 🗸 🗸
- 6h45 6h48 🗸
- 6h48 6h50 🗸
 - Now, when we complete an action, we put a checkmark.
 - > You will end up a fully checkmarked list with stamps.
 - This is more informative: Now we can directly say the puree is cooked between 6h00 and 6h50.
 - WHY: It simplifies a lot when we deal with uncountable sets.

- Adding absolute timestamps to our recipe.
- Very punctual recipe:

 \checkmark

- 6h00 6h05 🗸 🗸
- 6h05 6h30
- 6h30 6h45 🗸
- 6h45 6h48 🗸
- 6h48 6h50 🗸
 - Now, when we complete an action, we put a checkmark.
 - > You will end up a fully checkmarked list with stamps.
 - This is more informative: Now we can directly say the puree is cooked between 6h00 and 6h50.
 - WHY: It simplifies a lot when we deal with uncountable sets.

- Adding absolute timestamps to our recipe.
- Very punctual recipe:
- 6h00 6h05 🗸
- 6h05 6h30 🗸
- 6h30 6h45 🗸
- 6h45 6h48 🗸
- 6h48 6h50 🗸
 - Now, when we complete an action, we put a checkmark.
 - > You will end up a fully checkmarked list with stamps.
 - This is more informative: Now we can directly say the puree is cooked between 6h00 and 6h50.
 - ▶ WHY: It simplifies a lot when we deal with uncountable sets.

Signals in absolute time

A signal is a piecewise-constant function over Σ:

$$w: [t, t') \to \Sigma$$

starting at t and ending at t' with a duration t' - t.

• The concatenation $w_1 \cdot w_2$ is defined only if w_1 meets w_2 :



Extended Signals

- The special symbol \checkmark extends the alphabet Σ .
- $\blacktriangleright \Sigma_{\checkmark} = \Sigma \cup \{\checkmark\}$
- An extended signal w if $w \in \Sigma_{\checkmark}^{(*)}$.
- Some classes of extended signals:

- Pure Original recipe
- Reduced Fully checkmarked recipe
- Left-reduced Partially checkmarked recipe

Extended Timed Regular Expressions

- Let P a set of propositions.
- Representing extended signal languages over $\Sigma = \mathbb{B}^{|P|}$.
- The syntax:

$$\varphi := \varnothing \mid \epsilon \mid p \mid \checkmark \mid \varphi_1 \cdot \varphi_2 \mid \varphi_1 \lor \varphi_2 \mid \varphi_1 \land \varphi_2 \mid \varphi^* \mid {}^J_K \langle \varphi \rangle_I$$

where $p \in P$ and I, J, K are intervals of $\mathbb{R}_{\geq 0}$.

The semantics:

$$\begin{split} \llbracket p \rrbracket &= \{ w : [t, t') \to \Sigma \mid t < t' \text{ and } \forall t'' \in [t, t'). \ w(t'') \models p \} \\ \llbracket \checkmark \rrbracket &= \{ w : [t, t') \to \checkmark \mid t < t' \} \\ & \dots \\ \rrbracket_J^K \langle \varphi \rangle_I \rrbracket &= \{ w \mid w \in \llbracket \varphi \rrbracket, \ |w| \in I, \ w \neq \epsilon \to (\tau_1(w) \in J \land \tau_2(w) \in K) \} \end{split}$$

Left Reduction



Figure: A left reduction example.

Left Reduction

We introduce left reduction as a position (and duration) preserving derivative operation.

Definition (Left Reduction)

The left reduction of a language \mathcal{L} with respect to u is:

$$\delta_u(\mathcal{L}) := \{ \alpha \gamma w \mid \alpha u w \in \mathcal{L}, \ \alpha \in \checkmark^{(*)} \text{ and } \ w \in \Sigma^{(*)} \}$$

where $\gamma \in \checkmark^{(*)}$ and dom $(u) = \text{dom}(\gamma)$.

Language Membership

$$u \in \mathcal{L}$$
 iff $\gamma \in \delta_u(\mathcal{L})$

Derivatives of Timed Regular Expressions

A symbolic computation of left reduction with respect to all factors of a constant signal v:

Theorem (Derivative Computation)

Given a left-reduced timed regular expression φ and a constant signal $v : [t, t') \mapsto a$, applying the following rules yields an expression ψ such that $\llbracket \psi \rrbracket = \Delta_v(\llbracket \varphi \rrbracket)$.

$$\begin{array}{rcl} \Delta_{v}(\varnothing) &=& \varnothing\\ \Delta_{v}(\varepsilon) &=& \varnothing\\ \Delta_{v}(\epsilon) &=& \varnothing\\ \Delta_{v}(\checkmark) &=& \varnothing\\ \Delta_{v}(\psi) &=& \begin{cases} \Gamma \lor \Gamma \cdot p & \text{if } a \models p \text{ where } \Gamma := \begin{bmatrix} t,t' \\ [t,t'] \end{bmatrix} \langle \checkmark \rangle_{[0,t'-t]} \\ \varnothing & \text{otherwise} \end{cases}$$

$$\begin{array}{rcl} \Delta_{v}(\psi_{1} \lor \psi_{2}) &=& \Delta_{v}(\psi_{1}) \cdot \psi_{2} \lor xt(\psi_{1} \lor \Delta_{v}(\psi_{1})) \cdot \Delta_{v}(\psi_{2}) \\ \Delta_{v}(\psi_{1} \lor \psi_{2}) &=& \Delta_{v}(\psi_{1}) \lor \Delta_{v}(\psi_{2}) \\ \Delta_{v}(\psi_{1} \land \psi_{2}) &=& \Delta_{v}(\psi_{1}) \land \Delta_{v}(\psi_{2}) \\ \Delta_{v}(f_{1}^{K}(\psi)_{I}) &=& \stackrel{K}{J} \langle \Delta_{v}(\psi) \rangle_{I} \\ \Delta_{v}(\psi^{*}) &=& xt(\Delta_{v}(\psi))^{*} \cdot \Delta_{v}(\psi) \cdot \psi^{*} \end{array}$$

Online Timed Pattern Matching

Inputs/Outputs:

- The input φ is a timed regular expression.
- The input $w = w_1 w_2 \dots w_n$ to be read incrementally.
- The procedure yields the set of matches ending in jth segment at each step.

Full Procedure:

- Extract φ to see if the empty word is a match.
- For $1 \le j \le n$ repeat:
 - Update the state by deriving the previous state with respect to w_j and adding a new derivation $\Delta_{w_j}(\varphi)$ to the state for matches starting in j^{th} segment.
 - Extract the state to get matches ending in j^{th} segment.

Example

Symbols Segments		$\begin{array}{l} \{p \land \neg q\} \\ [0,3) \end{array}$	$egin{array}{l} \{p \land q\} \ [3,8) \end{array}$	$ \{ \neg p \land q \} \\ [8, 10) $
[0, 3)	$\langle p \cdot q \rangle_I \qquad \overline{\Delta_w}$	$ \begin{array}{c} \langle \Gamma_1 \cdot q \rangle_I \lor \\ \langle \Gamma_1 \cdot p \cdot q \rangle_I & \overline{\Delta_2} \end{array} $	$\xrightarrow{\langle \Gamma_1 \cdot \Gamma_2 \rangle_I}_{w_2} \bigvee_{\langle \Gamma_1 \cdot \Gamma_2 \cdot q \rangle_I \vee \langle \Gamma_1 \cdot \Gamma_2 \cdot p \cdot q \rangle_I}$	$\xrightarrow{\Delta w_3} \frac{\langle \Gamma_1 \cdot \Gamma_2 \cdot \Gamma_3 \rangle_I}{\langle \Gamma_1 \cdot \Gamma_2 \cdot \Gamma_3 \cdot q \rangle_I} \vee$
[3, 8)		$\langle p \cdot q \rangle_I \qquad \overline{\Delta_i}$	$ \stackrel{\langle \Gamma_2 \rangle_I}{} \bigvee \\ \stackrel{\vee}{} \frac{\langle \Gamma_2 \cdot q \rangle_I}{\langle \Gamma_2 \cdot p \cdot q \rangle_I} \lor $	$\xrightarrow{\Delta_{w_3}} \langle \Gamma_2 \cdot \Gamma_3 \rangle_I \lor$
[8, 10)			$\langle p \cdot q \rangle_I$	$\xrightarrow{\Delta_{w_3}}^{\varnothing}$

$$\Gamma_1 = \begin{bmatrix} 0,3 \\ 0,3 \end{bmatrix} \langle \checkmark \rangle_{[0,3]}$$

$$\Gamma_2 = \begin{bmatrix} 3,8 \\ 3,8 \end{bmatrix} \langle \checkmark \rangle_{[0,5]}$$

$$\Gamma_3 = \begin{bmatrix} 8,10 \\ 8,10 \end{bmatrix} \langle \checkmark \rangle_{[0,2]}$$

Example



▶ At each step, we report segments satisying the expression.

	Of	fline Algor Input Size	ithm e	Online Algorithm Input Size			
Test Patterns	100K	500K	1M	100K	500K	1M	
$\begin{array}{c} p\\ p \cdot q\\ \langle p \cdot q \cdot \langle p \cdot q \cdot p \rangle_{I} \cdot q \cdot p \rangle_{J}\\ (\langle p \cdot q \rangle_{I} \cdot r) \wedge (p \cdot \langle q \cdot r \rangle_{J})\\ p \cdot (q \cdot r)^{*} \end{array}$	0.06/17 0.08/21 0.23/28 0.13/23 0.11/20	0.27/24 0.42/46 1.09/77 0.50/51 0.49/37	0.51/33 0.74/77 2.14/140 1.00/86 0.96/60	6.74/14 8.74/14 28.07/14 15.09/15 11.53/15	29.16/14 42.55/14 130.96/14 75.19/15 52.87/15	57.87/14 81.67/14 270.45/14 148.18/15 110.58/15	

- Execution times/Memory usage (in seconds/megabytes).
- Both are linear for typical inputs.
- Online is 100 times slower but memory usage is constant.
- These numbers are sufficient for many applications.

Discussion

- We presented both theoretical and practical results:
- ▶ We formulated an algebraic approach in the timed theory.
 - The time passed represented by (\checkmark) symbol.
 - Timed derivatives
- ▶ We developed an online timed pattern matching procedure.
- Our procedure consumes a constant segment from the input signal and reports a set of matches ending in that segment.
- Do not worry, we have a tool:
 - sithub.com/doganulus/montre (Soon)
- Applications: Runtime verification, robotics, medical monitoring, ...

Thank you!!