On Language Equations and Grammar Coalgebras for Context-free Languages

Jurriaan Rot and Joost Winter

Leiden University and Centrum Wiskunde & Informatica

September 2, 2013

Behavioural differential equations

CALCO '11 (Winter/Bonsangue/Rutten): Context-free languages, Coalgebraically...

...behavioural differential equations / Brzozowski derivatives

Example:

$$
\begin{array}{rclcrclcrcl}
o(x) & = & 1 & \quad & x_a & = & xy & \quad & x_b & = & 0 \\
o(y) & = & 0 & \quad & y_a & = & 0 & \quad & y_b & = & 1
\end{array}
$$

Behavioural differential equations

CALCO '11 (Winter/Bonsangue/Rutten): Context-free languages, Coalgebraically. . .

. . . behavioural differential equations / Brzozowski derivatives

Example:

$$\begin{array}{rclcrclcrcl} o(x) & = & 1 & \quad & x_a & = & xy & \quad & x_b & = & 0 \\ o(y) & = & 0 & \quad & y_a & = & 0 & \quad & y_b & = & 1 \end{array}$$

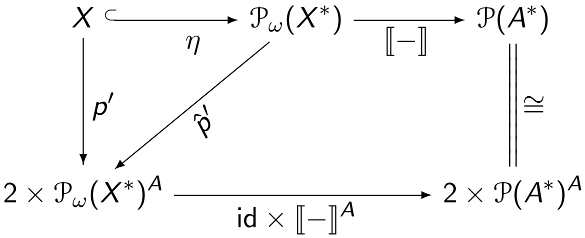
. . . gives . . .

$$\begin{array}{rcl} [\![x]\!] & = & \{a^n b^n \mid n \in \mathbb{N}\} \\ [\![y]\!] & = & \{b\} \end{array}$$

- These systems correspond to coalgebras for the functor $2 \times \mathcal{P}_\omega(-^*)^A \ldots$

- These systems correspond to coalgebras for the functor $2 \times \mathcal{P}_\omega(-^*)^A \ldots$
- ... and can be extended to (infinite) deterministic automata by enforcing

$$
\begin{aligned}
o(x+y) &= o(x) \vee o(y) \qquad & (x+y)_a &= x_a + y_a \\
o(xy) &= o(x) \wedge o(y) \qquad & (xy)_a &= x_a y + o(x) y_a
\end{aligned}
$$



Grammars and the Greibach normal form

Context-free grammars are systems:

$$p : X \to \mathcal{P}_\omega((X + A)^*)$$

A CFG is in *Greibach normal form* iff

$$p(x) \subseteq 1 + A(X + A)^* \qquad \text{for all } x \in X$$

giving an isomorphism:

$$\mathcal{P}_\omega((X + A)^*)_{\mathsf{GNF}} \;\cong\; 2 \times \mathcal{P}_\omega(X^*)^A$$

Hence, grammars in GNF are $2 \times \mathcal{P}_\omega(-^*)^A$-coalgebras.

Correctness via grammar derivations

- CALCO 2011: coalgebraic semantics coincides with classical semantics of context-free languages.
- Shown via leftmost derivations in a grammar.

Correctness via grammar derivations

- CALCO 2011: coalgebraic semantics coincides with classical semantics of context-free languages.
- Shown via leftmost derivations in a grammar.
- Context-free languages can also be seen as (least) solutions to grammars, regarded as systems of equations.
- Question: can we directly relate these systems of equations to coalgebraic semantics?

On solutions

Example:

$$x = 1 + axb$$

Unique solution: $x = \{a^n b^n \mid n \in \mathbb{N}\}$

On solutions

Example:

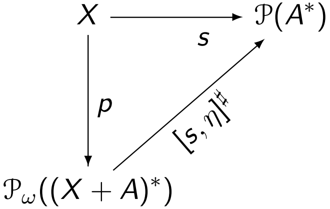
$$x = 1 + axb$$

Unique solution: $x = \{a^n b^n \mid n \in \mathbb{N}\}$

- Such systems of equations (based on the Boolean semiring) always have a least solution.
- If it corresponds to a grammar in GNF, this solution is unique.

Formalizing solutions

Definition: a *solution* is any mapping $s : X \to \mathcal{P}(A^*)$ making the following diagram commute:



$\sharp$: inductive extension based on union and concatenation

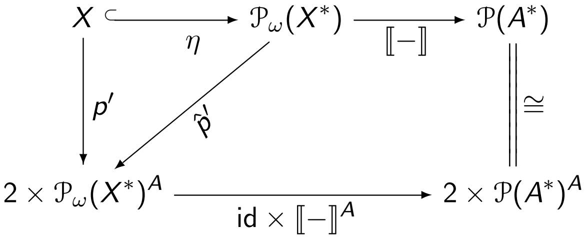
Solutions and GNF

For grammars in GNF, solutions correspond to mappings making the following diagram commute:

$$
\begin{array}{ccc}
X & \xrightarrow[\quad s \quad]{} & \mathcal{P}(A^*) \\
{\scriptstyle p'}\big\downarrow & & \big\| {\scriptstyle \cong} \\
2 \times \mathcal{P}_\omega(X^*)^A & \xrightarrow[\mathrm{id} \times (s^\sharp)^A]{} & 2 \times \mathcal{P}(A^*)^A
\end{array}
$$

A lemma

Coalgebraic semantics diagram:

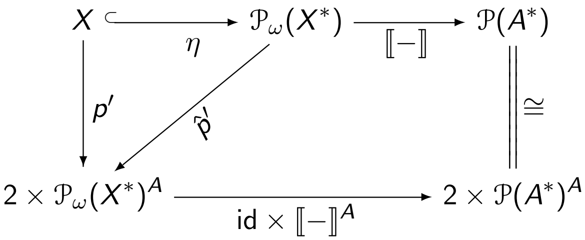


Here $[\![-]\!]$ is an algebra homomorphism, or:

$$[\![x + y]\!] = [\![x]\!] \cup [\![y]\!] \qquad \text{and} \qquad [\![xy]\!] = [\![x]\!][\![y]\!]$$

A lemma

Coalgebraic semantics diagram:



Here $[\![-]\!]$ is an algebra homomorphism, or:

$$[\![x + y]\!] = [\![x]\!] \cup [\![y]\!] \qquad \text{and} \qquad [\![xy]\!] = [\![x]\!][\![y]\!]$$

and hence we obtain:

Lemma

$([\![-]\!] \circ \eta)^\sharp = [\![-]\!]$.

A theorem

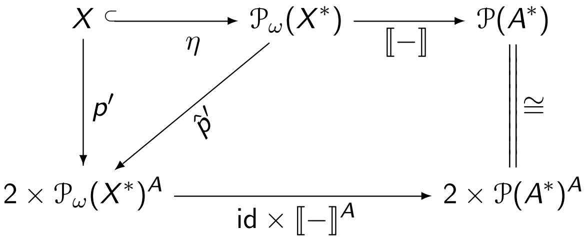
Coalgebraic semantics and classical semantics coincide:

Theorem

Given a (classical) system in GNF, $[\![-]\!] \circ \eta$ is the unique solution.

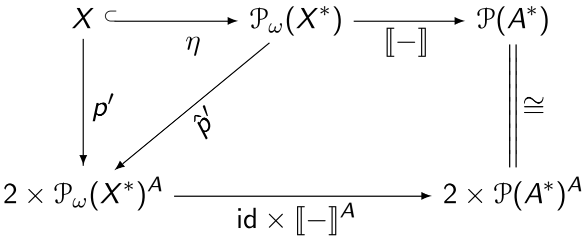
A proof

From the diagram

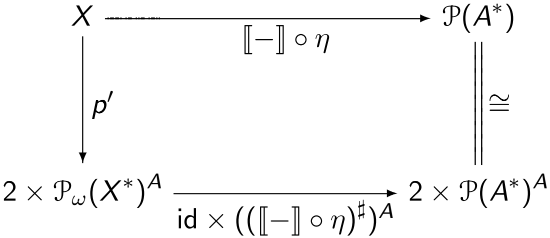


A proof

From the diagram



we obtain (by applying the lemma and deleting the diagonal arrow)



which is precisely the (unique) classical solution diagram for GNF.

A generalization (1)

- Standard generalization: formal languages $\Rightarrow$ formal power series
- Boolean semiring $\mathbb{B} \Rightarrow$ arbitrary semiring K
- generalization of $\mathcal{P}(A^*)$:

$$K\langle\langle A\rangle\rangle := \{f : A^* \to K\}$$

(also a semiring)

A generalization (2)

By applying the following replacements we can generalize our main result to arbitrary commutative semirings:

$$\begin{array}{rcl} 2 & \leftrightarrow & K \\ \mathcal{P}_\omega(-^*) & \leftrightarrow & K\langle - \rangle \\ \mathcal{P}(-^*) & \leftrightarrow & K\langle\langle - \rangle\rangle \end{array}$$

Conclusions and future work

- A more categorical look at the coalgebraic view of context-free languages.
- Essence: diagram manipulation $+$ $[\![-]\!]$ is an algebra morphism.
- Works more generally for power series over a commutative semiring.
- Generalization to other operations, e.g. complement and intersection: straightforward.
- Q: Can these results be further generalized to noncommutative semirings.