Language Theory and Infinite Graphs

Colin Stirling
School of Informatics
University of Edinburgh

Characterising DPDA

- A DPDA can be converted into a determinized SDG, and configuration $p\alpha$ is converted into sum configuration sum$(p\alpha)$ of the SDG where

  $\mathsf{G}^{\mathsf{c}}(p\alpha) \sim \mathsf{G}^{\mathsf{d}}(\mathrm{sum}(p\alpha))$

- (Conversely, a determinized SDG can be transformed into a DPDA)

- DPDA equivalence problem $=$ to deciding whether two admissible configurations of a determinized SDG are bisimulation equivalent

$$\alpha_1 + \ldots + \alpha_n \sim \beta_1 + \ldots + \beta_m \quad ?$$

Notation

- E, F, G, $\ldots$ range over admissible configurations
- $+$ can be extended: if E and F are admissible, $E \cup F$ is admissible, E and F are disjoint, $E \cap F = \emptyset$, then $E + F$ is the admissible configuration $E \cup F$
- Also use sequential composition, if E and F are admissible, then EF is the admissible configuration $\{\beta\gamma : \beta \in E \text{ and } \gamma \in F\}$
- $E = E_1G_1 + \ldots + E_nG_n$ is a head/tail form, if the head $E_1 + \ldots + E_n$ is admissible and at least one $E_i \neq \emptyset$, and each tail $G_i \neq \emptyset$

  If we write $E = E_1G_1 + \ldots + E_nG_n$ then E is a head/tail form

Congruence

- Decision question $E \sim F$?

- for all j, $G_j \sim H_j$ iff $E_1 G_1 + \ldots + E_n G_n \sim E_1 H_1 + \ldots + E_n H_n$

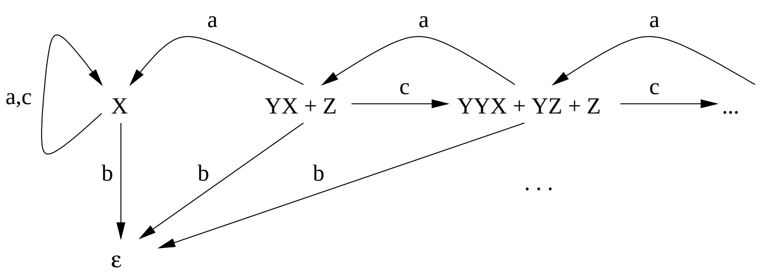
- We can tear apart configurations and replace parts with equivalent parts

Notation

- For $X \in$ S, $w(X)$ is the unique shortest word $v \in$ A$^+$ such that $X \xrightarrow{v} \epsilon$
- $w(E)$ is the unique shortest word for configuration E
- M is the maximum norm of the SDG
- "E following u", written $E \cdot u$, is the unique F such that $E \xrightarrow{u} F$

Example

$X \xrightarrow{a} X \quad Y \xrightarrow{a} \epsilon \quad Z \xrightarrow{a} \emptyset \quad X \xrightarrow{b} \epsilon \quad Y \xrightarrow{b} \emptyset$
$Z \xrightarrow{b} \epsilon \quad X \xrightarrow{c} X \quad Y \xrightarrow{c} YY \quad Z \xrightarrow{c} YZ + Z$



Interested in showing, for instance, that $YX + Z \sim YYX + YZ + Z$

DPDA decision procedure

- Given by a deterministic tableau proof system where goals are reduced to subgoals

- Goals and subgoals have the form $E \doteq F$, is $E \sim F$? E, F admissible

- Rules are (generalisations of) unfold, UNF, and balance rules,

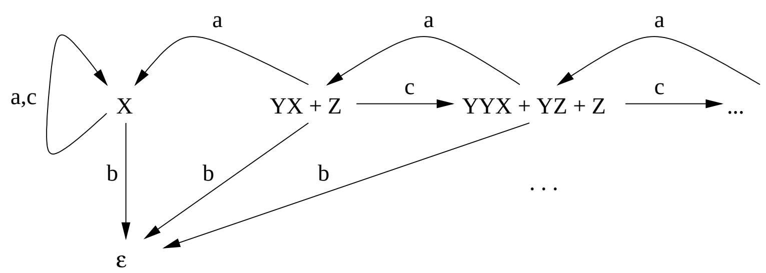
  BAL(L) and BAL(R). (CUT free)

Proof rules: UNF

$$\frac{E \doteq F}{(E \cdot a_1) \doteq (F \cdot a_1) \quad \ldots \quad (E \cdot a_k) \doteq (F \cdot a_k)} \; \mathsf{A} = \{a_1, \ldots, a_k\}$$

Example

$$X \xrightarrow{a} X \quad Y \xrightarrow{a} \epsilon \quad Z \xrightarrow{a} \emptyset \quad X \xrightarrow{b} \epsilon \quad Y \xrightarrow{b} \emptyset$$
$$Z \xrightarrow{b} \epsilon \quad X \xrightarrow{c} X \quad Y \xrightarrow{c} YY \quad Z \xrightarrow{c} YZ + Z$$



$$\frac{YX + Z \doteq YYX + YZ + Z}{X \doteq YX + Z \quad \epsilon \doteq \epsilon \quad YYX + YZ + Z \doteq YYYX + YYZ + YZ + Z}$$

Proof rules: BAL(L)

$$\frac{\begin{array}{rcl} X_1H_1 + \ldots + X_kH_k & \doteq & F \\ & \vdots & \\ E_1H_1 + \ldots + E_kH_k & \doteq & F' \end{array}}{E_1(F \cdot w(X_1)) + \ldots + E_k(F \cdot w(X_k)) \doteq F'} \; \mathrm{C}$$

where C is the condition

1. Each $E_i \neq \varepsilon$ and at least one $H_i \neq \varepsilon$

2. There are $\max\{\,|w(X_i)| \;:\; E_i \neq \emptyset \text{ for } 1 \leq i \leq k\}$ applications of UNF between the top goal and the bottom goal, and no application of any other rule

Proof rules: BAL(R)

$$\frac{\begin{array}{lcl} F & \doteq & X_1H_1 + \ldots + X_kH_k \\ & \vdots & \\ F' & \doteq & E_1H_1 + \ldots + E_kH_k \end{array}}{F' \doteq E_1(F \cdot w(X_1)) + \ldots + E_k(F \cdot w(X_k))} \quad \mathrm{C}$$

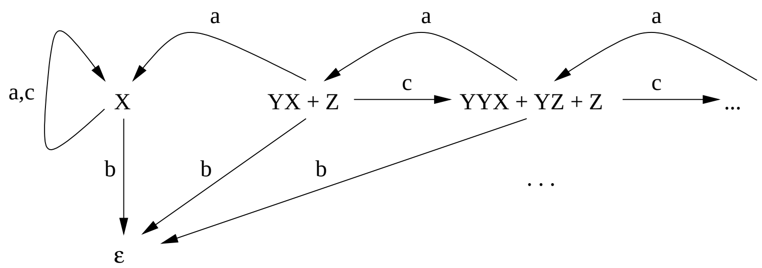
where C is the condition

1. Each $E_i \neq \varepsilon$ and at least one $H_i \neq \varepsilon$

2. There are $\max\{\,|w(X_i)|\ :\ E_i \neq \emptyset \text{ for } 1 \leq i \leq k\}$ applications of UNF between the top goal and the bottom goal, and no application of any other rule

Example

$$X \xrightarrow{a} X \quad Y \xrightarrow{a} \epsilon \quad Z \xrightarrow{a} \emptyset \quad X \xrightarrow{b} \epsilon \quad Y \xrightarrow{b} \emptyset$$
$$Z \xrightarrow{b} \epsilon \quad X \xrightarrow{c} X \quad Y \xrightarrow{c} YY \quad Z \xrightarrow{c} YZ+Z$$



$$\dfrac{YX+Z \doteq YYX+YZ+Z\ (F)}{\cdots \quad \cdots \quad \dfrac{YYX+YZ+Z \doteq YYYX+YYZ+YZ+Z}{YYYX+YYZ+YZ+Z \doteq YYYX+YYZ+YZ+Z}\ \text{BAL(L)}}\ \text{UNF}$$

$w(Y)=a$, $w(Z)=b$ and $(F \cdot w(Y)) = YX+Z$, $(F \cdot w(Z)) = \epsilon$

Procedure

- Start with an initial goal, $E \doteq F$, and deterministically build a proof tree by applying the tableau rules (there is an ordering on the rules)

- Goals are thereby reduced to subgoals

- Need to define final goals. Rules are not applied to final goals

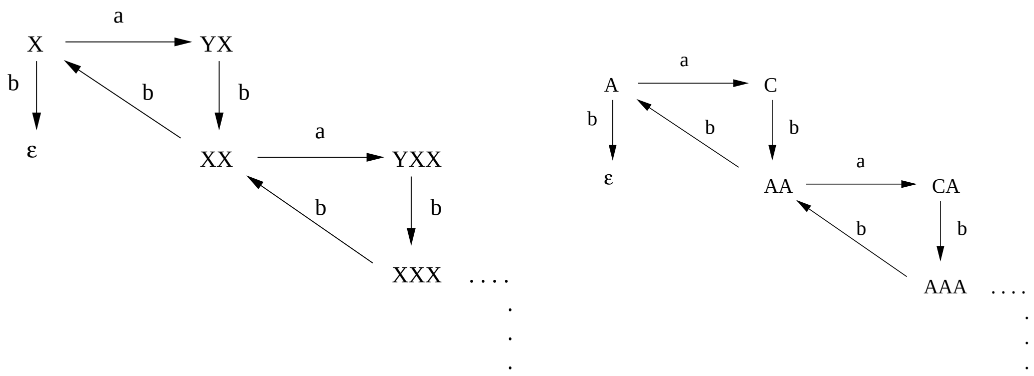
A successful tableau is a finite proof tree all of whose leaves are successful final goals

Otherwise a tableau is unsuccessful

Unsuccessful final goals

$E \doteq F$ (exactly one of E, F is $\emptyset$)

Infinite proof tree ?



$$
\dfrac{
  \dfrac{
    \dfrac{
      \emptyset \doteq \emptyset
      \qquad
      \dfrac{
        \dfrac{YXA^6 \doteq CA^6}{YXXA^5 \doteq CA^6}\ \text{BAL(L)}
        \qquad
        XA^5 \doteq A^6
      }{XXA^5 \doteq AA^6}\ \text{UNF}
    }{YXA^5 \doteq CA^5}\ \text{UNF}
  }{YXX^5 \doteq CA^5}\ \text{BAL(L)}
  \qquad
  X^5 \doteq A^5
}{XX^5 \doteq AA^5}\ \text{UNF}
$$

Extensions

- If $E = E_1G_1 + \ldots + E_nG_n$, $F = F_1H_1 + \ldots + F_mH_m$, then

  F is a tail extension of E provided that for each $i : 1 \leq i \leq m$,

  $$H_i = K_1^iG_1 + \ldots + K_n^iG_n$$

- The extension e is the m-tuple $(K_1^1 + \ldots + K_n^1, \ldots, K_1^m + \ldots + K_n^m)$ without the G_is, the width of e is m, and F is said to extend E with e

- Extensions are matrices, written in a linear notation

- If $E = E_1G_1 + \ldots + E_nG_n$ and $F = F_1G_1 + \ldots + F_nG_n$, then E extends F by $e = (\varepsilon + \emptyset + \ldots + \emptyset, \ldots, \emptyset + \emptyset + \ldots + \varepsilon)$. Identity extension (ε)

Composing extensions

If

$E = E_1G_1 + \ldots + E_lG_l$ and

$E' = E'_1G'_1 + \ldots + E'_mG'_m$ and

$E'' = E''_1G''_1 + \ldots + E''_nG''_n$ and

E' extends E by $e = (J^1_1 + \ldots + J^1_l, \ldots, J^m_1 + \ldots + J^m_l)$ and

E'' extends E' by $f = (K^1_1 + \ldots + K^1_m, \ldots, K^n_1 + \ldots + K^n_m)$,

then E'' extends E by $ef = (H^1_1 + \ldots + H^1_l, \ldots, H^n_1 + \ldots + H^n_l)$

where $H^i_j = K^i_1J^1_j + \ldots + K^i_mJ^m_j$

Example

$$
\begin{array}{lll}
E = YG_1 + ZG_2 & \text{where} & G_1 = X \text{ and } G_2 = \varepsilon \\
E' = YG'_1 + ZG'_2 & \text{where} & G'_1 = YX + Z \text{ and } G'_2 = \varepsilon \\
E'' = YG''_1 + ZG''_2 & \text{where} & G''_1 = YYX + YZ + Z \text{ and } G''_2 = \varepsilon
\end{array}
$$

E' extends E by $e = (Y + Z, \emptyset + \varepsilon)$ and

E'' extends E' by $f = e = (Y + Z, \emptyset + \varepsilon)$

Therefore, E'' extends E by $ef = (YY + (YZ + Z), \emptyset + \varepsilon)$.

Notation

Assume goals with the same heads

$$\text{goal g} \quad (E)\ E_1G_1 + \ldots + E_nG_n \doteq F_1G_1 + \ldots + F_nG_n\ (F)$$

$$\text{goal h} \quad (E')\ E_1H_1 + \ldots + E_nH_n \doteq F_1H_1 + \ldots + F_nH_n\ (F')$$

h extends g by e, if E' extends E by e (and F' extends F by e)

Repeating patterns

Width is 3. Assume $g(i)$, $h(i)$, $i : 1 \leq i \leq 8$

$$E_1 G_1^i + E_2 G_2^i + E_3 G_3^i \doteq F_1 G_1^i + F_2 G_2^i + F_3 G_3^i$$

$$E_1 H_1^i + E_2 H_2^i + E_3 H_3^i \doteq F_1 H_1^i + F_2 H_2^i + F_3 H_3^i$$

Assume extensions e_1, e_2 and e_3 for $g(i)$ and $h(i)$

$$\begin{array}{ccccccccccccccc}
 & & & & & & & e_3 & & & & & & & \\
 & & & e_2 & & & & & & & & e_2 & & & \\
 & e_1 & & & & e_1 & & & & e_1 & & & & e_1 & \\
g_1 & & g_2 & & g_3 & & g_4 & & g_5 & & g_6 & & g_7 & & g_8
\end{array}$$

If each $g(i)$, $1 \leq i \leq 8$, and each $h(i)$, $1 \leq i < 8$, m-true then $h(8)$ is m-true

General case

If there are two families of goals $g(i)$, $h(i)$, $1 \leq i \leq 2^n$

$$E_1 G_1^i + \ldots + E_n G_n^i \doteq F_1 G_1^i + \ldots + F_n G_n^i$$

$$E_1 H_1^i + \ldots + E_n H_n^i \doteq F_1 H_1^i + \ldots + F_n H_n^i$$

and extensions $e_1, \ldots, e_n$ such that for each e_j and $i \geq 0$

$g(2^j i + 2^{j-1} + 1)$ extends $g(2^j i + 2^{j-1})$ by e_j
$h(2^j i + 2^{j-1} + 1)$ extends $h(2^j i + 2^{j-1})$ by e_j.

and each $g(i)$, $i : 1 \leq i \leq 2^n$, and each $h(j)$, $j : 1 \leq j < 2^n$, is m-true, then $h(2^n)$ is m-true

Successful final goals

$$
\begin{array}{ccll}
 & E \doteq F & & \text{the root goal} \\
 & \vdots & & \\
 & \vdots & & \text{there are goals } g(i)\ h(i) \ldots \\
 & \vdots & & h(2^n) \text{ is } E' \doteq F' \\
E \doteq E & E' \doteq F' & &
\end{array}
$$

Decidability

Propositions

- Every tableau is finite
- $E \sim F$ iff the tableau with root $E \doteq F$ is successful

Complexity upper bound is primitive recursive (because of extensions). The number of goals involved may grow with the number of states when applying extension theorem as it is width of a goal that counts

Corollary: $E \sim F$ iff $E \sim_{f(n)} F$

where f is the primitive recursive function, n is the size of the DPDA, E and F

Higher-order pushdown automata

Basic transitions of the form for order n

$$pX \xrightarrow{a} q \,\mathsf{Op} \text{ where } \mathsf{Op} \in \{\mathrm{swap}(\alpha), \mathrm{push}(i), \mathrm{pop}(i)\ 1 < i \leq n\}$$

Stacks of stacks of stacks . . .

Example at order 2

$$\begin{array}{lll}
p'Z \xrightarrow{a} pZ & & \\
pZ \xrightarrow{a} pXZ & pX \xrightarrow{a} pXX & pX \xrightarrow{b} q\,\mathrm{push} \\
qX \xrightarrow{b} q\epsilon & & qZ \xrightarrow{c} r\,\mathrm{pop} \\
rX \xrightarrow{c} r\epsilon & rZ \xrightarrow{\epsilon} r\epsilon &
\end{array}$$

Results/Open Questions

- The languages generable by order 2 pushdown automata are indexed languages (mildly context-sensitive languages) introduced by Aho 1968
- The monadic second-order theory of any graph generated by a higher-order automaton is decidable (Knapik, Niwinski and Urzyczyn 2002)

  Proof nice mixture of Rabin's theorem and geometry of interaction
- Are languages generable by order n, $n > 2$, context-sensitive?
- Is language equivalence decidable for determinisitic higher-order pda ?

  Intimately related to higher-order schema problem open since 1960s

Thanks for listening