Forgetting Automata

Lakshmanan K

School of Computer Science & Engineering,
VIT, VELLORE - 632 014, Tamilnadu
klakshma@vit.ac.in

Outline of the Talk

1. Introducing
   - Forgetting automata
   - Different families of Forgetting automata
2. Relationship with
   - Among the families of forgetting automata
   - Chomsky families of languages
3. Some Open Problems

Forgetting Automata

- Introduced in 1992 by Jančar, Mráz and Plátek.
- Main objective was to to model the so called ‘analysis by reduction’, a technique used in linguistics.
- Also motivated by combining *list* and *erase* automata.
- This is a Linear Bounded Automata (LBA) with restriction in overwriting symbols.
  - Erase: The cell content is erased and replaced by * (partial information is retained).
  - Delete: The whole cell is deleted (information is lost)
  - No other rewriting is possible.
  - R/W head can move in either direction unless restricted.

Forgetting automata: Formal definition

A Forgetting automaton is $M = (Q, \Sigma, \delta, \triangleleft, \triangleright, *, q_0, O, F)$ where

- Q is a finite set of states
- Σ is the input alphabet
- δ is the transition function
- $\triangleleft, \triangleright$ are left and right borders respectively and $\triangleleft, \triangleright, * \notin \Sigma$
- O is the set of operations
- $q_o \in Q$ is the initial state
- F is the set of final states.
- $\triangleleft, \triangleright$ cannot be erased/deleted. Only * and $a \in \Sigma$ can be.

◁	a	a	a	a	b	b	b	▷

Working of Forgetting Automata

The transition function $\delta : Q \times (\Sigma \cup \{\triangleleft, \triangleright, *\} \rightarrow 2^{\Sigma \times O}$ with

- MV_L, MV_R - moving the head one item to the left/right.
- ER_L, ER_R - erasing the content of the scanned item with * and moving the head to one item left/right, respectively
- DL_L, DL_R - deleting the item from the list moving the head one item to the left/right respectively.

example 1(Erase Right)

◁	**a**	a	a	b	b	b	▷

$[q_0, a] \rightarrow [q_1, ER_R]$

◁	*	**a**	a	b	b	b	▷

example 2 (Delete Left)

◁	a	a	a	**b**	b	b	▷

$[q_i, b] \rightarrow [q_j, DL_L]$

◁	a	a	**a**	b	b	▷

Families of Forgetting Automata

figuration $w_1 q w_2$ with $q \in Q, w_1 w_2 \in \triangleleft(\Sigma \cup \{*\})^* \triangleright$.

cceptance An input word is accepted by F if there is a computation that starts from the initial configuration and achieves a configuration with the control unit being in a final state.

families $[O]$ where $O \in \{MV_L, MV_R, ER_L, ER_R, DL_L, DL_R\}$, we denote the class of languages recognized by F-automata using operations from O. For convenience, $MV = MV_R, MV_L$.

choices Total choices $2^6 - 1 = 63$ families

3 cases Non determinism, determinism, Unary alphabet

total 3x63 $=$ 189 families of languages $\subsetneq$ CSL

Results among themselves

1. $[MV_R, ER] = [MV_R, ER_L]$ (leave out ER_R)
   Proof: $ER_R \equiv ER_L \vdash MV_R \vdash MV_R$.
2. $[MV_L, ER] = [MV_L, ER_R]$ (leave out ER_L)
3. $[MV_R, DL] = [MV_R, DL_L]$ (leave out DL_R)
   Proof: $DL_R \equiv DL_L \vdash MV_R$.
4. $[MV_L, DL] = [MV_L, DL_R]$ (leave out DL_L)
5. $[MV, ER] = [MV, ER_L] = [MV, ER_R]$ (head is free to move)
6. $[MV, DL] = [MV, DL_L] = [MV, DL_R]$
7. $[ER_R, DL]$ and $MV_R \notin O$ then
   $[ER_R, DL] = [ER_R, DL_L]$ (leave out DL_R)
   Proof: $DL_R \equiv DL_L \vdash ER_R$

Relationship with Chomsky Families

$$REG = O, \; O \in \{[MV], [ER], [MV_L, ER], [MV_R], [ER_R], [DL_R]\}$$

Relationship with Chomsky Families

$REG = O,\ O \in \{[MV], [ER], [MV_L, ER], [MV_R], [ER_R], [DL_R]\}$

Classes between REG and CFL: $[ER_R, DL] \to [MV_R, DL]$

candidate language: $L_{ab} = \{a^n b^n \mid n \geq 1\} \notin REG$

Proof: $\{a^n b^n\} \in [ER_R, DL]$

Replace all a by * (ER_R)and read the first b, delete it and go back (i.e., Del-left) and delete the last * and move right. Repeat it.

Relationship with Chomsky Families

$REG = O, \; O \in \{[MV], [ER], [MV_L, ER], [MV_R], [ER_R], [DL_R]\}$

Classes between REG and CFL: $[ER_R, DL] \rightarrow [MV_R, DL]$
candidate language: $L_{ab} = \{a^n b^n \mid n \geq 1\} \notin REG$

Proof: $\{a^n b^n\} \in [ER_R, DL]$
Replace all a by * (ER_R)and read the first b, delete it and go back (i.e., Del-left) and delete the last * and move right. Repeat it.

$L_{pl} = \{wcw^R : w \in \{a, b\}^*\} \in CFL \cup LIN - [ER_R, DL]$.
Proof: Unable to differentiate between a and b after they are replaced by $*$.

Relationship with Chomsky Families

$REG = O, \; O \in \{[MV], [ER], [MV_L, ER], [MV_R], [ER_R], [DL_R]\}$

Classes between REG and CFL: $[ER_R, DL] \rightarrow [MV_R, DL]$
candidate language: $L_{ab} = \{a^n b^n \mid n \geq 1\} \notin REG$

Proof: $\{a^n b^n\} \in [ER_R, DL]$
Replace all a by * (ER_R)and read the first b, delete it and go back (i.e., Del-left) and delete the last * and move right. Repeat it.

$L_{pl} = \{wcw^R : w \in \{a, b\}^*\} \in CFL \cup LIN - [ER_R, DL]$.
Proof: Unable to differentiate between a and b after they are replaced by $*$.
Note: $L_{ba} = \{w \in \{a, b\}^* \mid |w|_a = |w|_b\} \in [ER_R, DL]$

Relationship with CFL and more

1. $CFL = [MV_R, ER], [MV_R, ER_R, DL]$
2. Classes between REG and CSL, but still not CFL are $[ER, DL_R] \rightarrow [ER, DL] \rightarrow [MV_L, ER, DL]$

Relationship with CFL and more

1. $CFL = [MV_R, ER], [MV_R, ER_R, DL]$
2. Classes between REG and CSL, but still not CFL are $[ER, DL_R] \to [ER, DL] \to [MV_L, ER, DL]$
   Proof: $REG \to [ER, DL_R]$ and $\{a^n b^n\} \in [ER, DL_R]$

   $L_{pl} = \{wcw^R : w \in \{a, b\}^*\} \in CFL - ([ER_R, DL] \wedge [MV_L, ER, DL])$.
   Then, $\{a^{2^n} \mid n \geq 0\} \in [ER, DL_R] - CFL$.
   The proof idea is to shorten the length between sentinels just by deleting alternate symbols (repeatedly) (its unary).

Relationship with CFL and more

1. $CFL = [MV_R, ER], [MV_R, ER_R, DL]$
2. Classes between REG and CSL, but still not CFL are
   $[ER, DL_R] \rightarrow [ER, DL] \rightarrow [MV_L, ER, DL]$
   Proof: $REG \rightarrow [ER, DL_R]$ and $\{a^n b^n\} \in [ER, DL_R]$

   $L_{pl} = \{wcw^R : w \in \{a, b\}^*\} \in CFL - ([ER_R, DL] \wedge [MV_L, ER, DL])$.
   Then, $\{a^{2^n} \mid n \geq 0\} \in [ER, DL_R] - CFL$.
   The proof idea is to shorten the length between sentinels just by deleting alternate symbols (repeatedly) (its unary).
3. Classes strictly between CFL and CSL:
   $[MV_R, ER, DL_R] \rightarrow [MV_R, ER, DL] \rightarrow [MV, ER, DL]\&[MV, ER]$

Relationship with CFL and more

1. $CFL = [MV_R, ER], [MV_R, ER_R, DL]$
2. Classes between REG and CSL, but still not CFL are $[ER, DL_R] \to [ER, DL] \to [MV_L, ER, DL]$
   Proof: $REG \to [ER, DL_R]$ and $\{a^n b^n\} \in [ER, DL_R]$

   $L_{pl} = \{wcw^R : w \in \{a, b\}^*\} \in CFL - ([ER_R, DL] \wedge [MV_L, ER, DL])$.
   Then, $\{a^{2^n} \mid n \geq 0\} \in [ER, DL_R] - CFL$.
   The proof idea is to shorten the length between sentinels just by deleting alternate symbols (repeatedly) (its unary).
3. Classes strictly between CFL and CSL:
   $[MV_R, ER, DL_R] \to [MV_R, ER, DL] \to [MV, ER, DL] \& [MV, ER]$
   Proof: $CFL \subsetneq [MV, ER]$ and $[MV_R, ER, DL_R]$.
   One can construct the automata to accept $L_{abc} = \{a^n b^n c^n \;\; n \geq 1\}$.
   idea: Read a a, replace by $*$, move forward and find a b, replace by $*$ and move forward until find a c and replace it by $*$. Repeat this.

Some Open Problems

1. Whether there is a CFL not in $[MV, DL]$?
2. $[MV, DL] \subsetneq [MV, EL]$? (later resolved for det. case)
3. Will determinism has any effect on these automata? (see above)
4. When alphabet is Unary, what is the relationship among the familes of languages?
5. Relationship with 1-counter and 2-counter automata.

Jančar P., Mráz F., Plátek M. (1992), Forgetting automata and the Chomsky hierarchy, in Proc. SOFSEM '92, Ždiar, Slovakia, November 1992, pp. 41–44.

Jančar P., Mráz F., Plátek M.(1993), A Taxonomy of forgetting automata,

Jančar P., Mráz F., Plátek M. (1996), Forgetting automata and context-free languages, Acta Informatica, **33**, 409–420.

Glockler J. (2007), Forgetting automata and Unary languages, Int. J. of Foundations of Computer Science, **18(4)**, 813–827.