Some minimality results on biresidual and biseparable automata

Hellis Tamm
Institute of Cybernetics, Tallinn

Joint Estonian-Latvian Theory Days
Rakari, Sept 30 - Oct 3, 2010

This talk is based on a paper at LATA 2010

This research is supported by the Estonian Center of Excellence in Computer Science, EXCS, financed by the European Regional Development Fund, and by the Estonian Science Foundation grant 7520.

Outline

- Residual finite state automata (RFSA), canonical RFSA
- Biresidual (biRFSA) and biseparable automata, known properties about their minimality
- Fooling set techniques for the lower bounds of the size of NFAs, their relationship to biRFSA and biseparable automata
- Transition minimality of reversible canonical biRFSAs

Residual finite state automata: motivation

- While there is a unique minimal DFA for every regular language, there may be more than one minimal NFA.
- Residual finite state automata (RFSA) introduced by Denis, Lemay, and Terlutte (2001) are a subclass of NFA with a property similar to the uniqueness of minimal DFA.
- There is a unique RFSA called the canonical RFSA for a given language that is a state-minimal RFSA.
- The size of the canonical RFSA is at least the size of a minimal NFA and at most the size of the minimal DFA.
- Canonical RFSA can be a minimal NFA, even a unique one.
- Learning algorithms for regular languages that derive RFSAs instead of DFAs have been developed.

Definitions

- Let $A = (Q, \Sigma, E, I, F)$ be an NFA where Q is a finite set of states, Σ is an input alphabet, $E \subseteq Q \times \Sigma \times Q$ is a set of transitions, $I \subseteq Q$ is a set of initial states, and $F \subseteq Q$ is a set of final states.

- Given $P \subseteq Q$ and $x \in \Sigma^*$, we denote by $P \cdot x$ the set $P' \subseteq Q$ such that $p' \in P'$ if and only if there is a path from any $p \in P$ to p' labelled by x.

- The left and right languages of a state $q \in Q$ are defined as $L_L(A, q) = \{x \in \Sigma^* \mid q \in I \cdot x\}$ and $L_R(A, q) = \{x \in \Sigma^* \mid \{q\} \cdot x \cap F \neq \emptyset\}$.

Residual finite state automata

- A language $L' \subseteq \Sigma^*$ is a *residual* of a language L if there exists a word $u \in \Sigma^*$ such that $L' = \{v \in \Sigma^* \mid uv \in L\}$.
- An automaton A is a *residual finite state automaton* (RFSA) if for every state q of A, $L_R(A, q)$ is a residual of $L(A)$.
- A residual of a language L is *prime* if it is non-empty and if it cannot be obtained as the union of other residuals of L.
- The *canonical RFSA* of a regular language L is the automaton $A = (Q, \Sigma, E, I, F)$ where Q is the set of prime residuals of L, Σ is an input alphabet, I is the set of prime residuals of L which are included in L, F is the set of prime residuals of L containing the empty word, and for all prime residuals S and S' of L and for all $a \in \Sigma$, $(S, a, S') \in E$ if and only if $aS' \subseteq S$.

Residual finite state automata

- The canonical RFSA is the unique RFSA that has the maximum number of transitions among the set of RFSAs which have the minimum number of states (Denis et al., 2001).
- Any DFA is an RFSA: given a DFA A, for any state q of A, $L_R(A, q)$ is a residual of $L(A)$.
- There are cases where the canonical RFSA is a minimal NFA and much smaller than the minimal DFA.

Biresidual automata (biRFSA)

- BiRFSA is an RFSA such that its reversal automaton is also an RFSA.
- BiRFSAs were introduced by Latteux, Roos, Terlutte (2005) who studied their minimality issues.
- The canonical RFSA of a language accepted by a biRFSA is a biRFSA.
- The canonical biRFSA is a state-minimal NFA (not necessarily unique!)
- Any other state-minimal NFA is a subautomaton of the canonical biRFSA.

Biseparable automata

Latteux et al. (2005) introduced and studied a subfamily of biRFSAs called biseparable automata.

- A trim NFA $A = (Q, \Sigma, E, I, F)$ is called *separable* if for every state $q \in Q$ there is some $u \in \Sigma^*$ such that $I \cdot u = \{q\}$.
- A is *biseparable* if both A and A^R are separable.
- Any biseparable automaton is a canonical biRFSA.
- Any biseparable automaton is a unique state-minimal NFA.
- Since biseparable automata include bideterministic automata as a proper subclass, the last statement improves a similar result for bideterministic automata (HT, Ukkonen, 2003).
- However, there exist unique minimal NFAs which are not biseparable.

Our results

- We consider two lower bound methods for the number of states of NFAs and present two results related to these methods:
  - First, the lower bound provided by the *fooling set technique* is tight for and only for biseparable automata.
  - Second, the lower bound provided by the *extended fooling set technique* is tight for any language accepted by a biRFSA.
- Third result: any reversible canonical biRFSA is transition-minimal.

Lower bound techniques for the size of NFAs

- Fooling set technique (Glaister and Shallit, 1996):

  Let $L \subseteq \Sigma^*$ be a regular language, and suppose there exists a set of pairs $P = \{(x_i, w_i) \mid 1 \leq i \leq n\}$ such that

  (a) $x_i w_i \in L$, for $1 \leq i \leq n$, and

  (b) $x_j w_i \notin L$, for $1 \leq i, j \leq n$, $i \neq j$.

  Then any NFA accepting L has at least n states.

- Extended fooling set technique (Birget, 1992):

  (b') $x_j w_i \notin L$ or $x_i w_j \notin L$, for $1 \leq i, j \leq n$

- Extended fooling set technique may provide a better lower bound.

- Lower bounds obtained by these techniques are not necessarily tight.

Lower bound for biseparable automata

A lower bound provided by the fooling set technique can be tight if and only if L is accepted by a biseparable automaton:

Theorem 1. Let $L \subseteq \Sigma^*$ be a regular language, and let n be the maximum integer such that there exists a set of pairs $P = \{(x_i, w_i) \mid 1 \leq i \leq n\}$ with

(a) $x_i w_i \in L$, for $1 \leq i \leq n$, and

(b) $x_j w_i \notin L$, for $1 \leq i, j \leq n$, $i \neq j$.

Then any NFA accepting L has n states if and only if it is biseparable.

Corollary. Any NFA with n states accepting a language that has a fooling set of size n is a unique minimal NFA for that language.

Lower bound for biresidual languages

The extended fooling set technique provides a tight lower bound for all languages accepted by a biRFSA:

Theorem 2. Let $L \subseteq \Sigma^*$ be a biRFSA language, and let n be the maximum integer such that there exists a set of pairs $P = \{(x_i, w_i) \mid 1 \leq i \leq n\}$ with

(a) $x_i w_i \in L$, for $1 \leq i \leq n$, and

(b) $x_j w_i \notin L$ or $x_i w_j \notin L$, for $1 \leq i, j \leq n$, $i \neq j$.

Then a minimal NFA accepting L has n states.

Note: there are languages other than those accepted by a biRFSA for which the lower bound obtained by this technique is tight.

Transition minimality of reversible biRFSAs

It is known that a canonical RFSA has the maximum number of transitions among the set of RFSAs which have the minimum number of states (Denis et al., 2001).

An NFA is called *reversible* if for any state-label pair (q, a) there is at most one in-transition and one out-transition involving q and a.

Theorem 3. A reversible canonical biRFSA has the minimum number of transitions among all ϵ-NFAs accepting the same language.

To prove this result, the theory of transition-minimal ϵ-NFAs by John (2003) was extended.

Cor. A reversible canonical biRFSA is a transition-minimal NFA.

Cor. A reversible biseparable automaton is transition-minimal.

Conclusions and future work

- RFSAs are an interesting class of automata that deserves more study.
- Since RFSAs are a generalization of DFAs then is it possible to extend some results obtained for DFAs, for RFSAs?
- More connections between lower bound techniques and special automata classes?
- More general theory of transition-minimal ϵ-NFAs?