Bideterministic automata

Hellis Tamm

Institute of Cybernetics
Tallinn University of Technology, Estonia

Theory Days, Kääriku, January 30 – February 1, 2009

Finite automaton: definitions

An *automaton* $A = (Q, \Sigma, E, I, F)$ is a mathematical model for a finite state machine where Q is a 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 an input of symbols, it goes through a series of states according to its transition function.

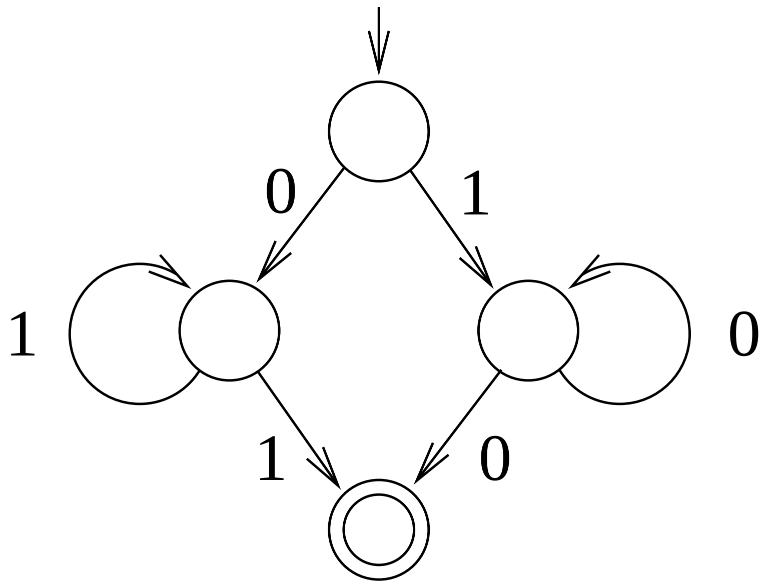
A word $w = a_1 a_2 ... a_n$ is *accepted* by A if there is a sequence of transitions $(q_0, a_1, q_1), (q_1, a_2, q_2), ..., (q_{n-1}, a_n, q_n)$ such that $q_0 \in I$ and $q_n \in F$.

The set of all words accepted by A is the *language* of A, denoted by $L(A)$.

Automata: determinism vs nondeterminism

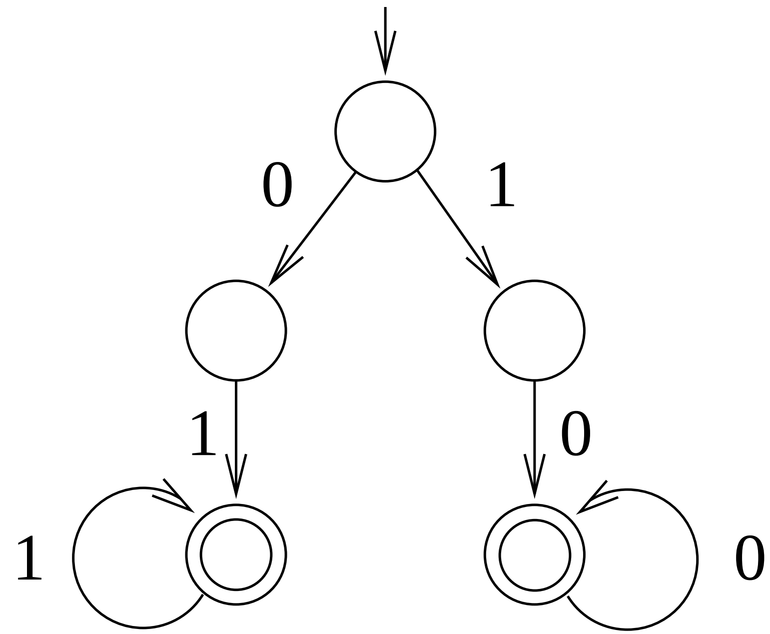
$L(A) = \{01, 011, 0111, ...\} \cup \{10, 100, 1000, ...\}$

nondeterministic automaton



co-deterministic automaton

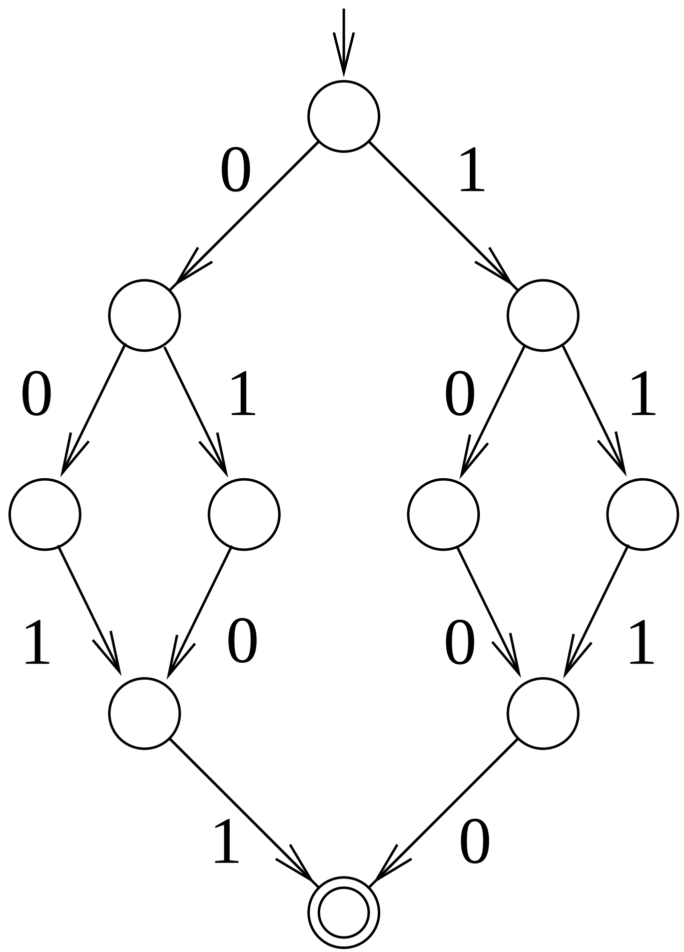
deterministic automaton



co-nondeterministic automaton

Bideterministic automaton

A bideterministic automaton is deterministic and co-deterministic.



State complexity

- The number of states of the minimal deterministic finite automaton (DFA) for a given language can be exponentially larger than the number of states in a minimal nondeterministic automaton (NFA).
- The minimal DFA is unique but there may be several minimal NFAs.
- Many cases where the maximal blow-up of size when converting an NFA to DFA does not occur.
- Some sufficient conditions have been identified which imply that the deterministic and nondeterministic state complexities are the same.

Transition complexity

- While the state-minimal DFA is also minimal with respect to the number of transitions, this is not necessarily the case with NFAs.
- Even allowing one more state in an NFA can produce a considerable reduction in the number of transitions.
- The number of transitions may be even a better measure for the size of an NFA than the number of states.
- Furthermore, allowing ϵ-transitions in an NFA (ϵ-NFAs) it is possible to have automata with even less transitions than NFAs.

Bideterministic automata: minimality results

- A bideterministic automaton is a minimal DFA (easy, by Brzozowski's DFA minimization algorithm: given an automaton A, minimal DFA is obtained by $Det(Rev(Det(Rev(A))))$)
- A bideterministic automaton is a unique state-minimal NFA (HT–Ukkonen 2003)
- A bideterministic automaton is a transition-minimal NFA (HT 2004)
- The necessary and sufficient conditions for a bideterministic automaton to be a unique transition-minimal NFA (HT 2007)
- More generally: a bideterministic automaton is a transition-minimal ϵ-NFA (HT 2007).

Universal automaton

A *factorization* of a regular language L is a maximal couple (with respect to the inclusion) of languages (U, V) such that $UV \subseteq L$.

The *universal automaton* of L is $U_L = (Q, \Sigma, E, I, F)$ where
Q is the set of factorizations of L,
$I = \{(U, V) \in Q \mid \epsilon \in U\}$,
$F = \{(U, V) \in Q \mid U \subseteq L\}$,
$E = \{((U, V), a, (U', V')) \in Q \times a \times Q \mid Ua \subseteq U'\}$.

Fact: universal automaton of the language L is a finite automaton that accepts L.

Automaton morphism and the universal automaton

Let $A = (Q, \Sigma, E, I, F)$ and $A' = (Q', \Sigma, E', I', F')$ be two NFAs. Then a mapping μ from Q into Q' is a *morphism* of automata if and only if $p \in I$ implies $p\mu \in I'$, $p \in F$ implies $p\mu \in F'$, and $(p, a, q) \in E$ implies $(p\mu, a, q\mu) \in E'$ for all $p, q \in Q$ and $a \in \Sigma$.

Known properties:

- Let A be a trim automaton that accepts L. Then there exists an automaton morphism from A into U_L.
- In particular, the minimal DFA and all state-minimal NFAs accepting L are subautomata of U_L.
- At least one transition-minimal NFA is a subautomaton of U_L.

Bideterministic automata: universal and minimal

It can be shown that any bideterministic automaton is the universal automaton for the given language.

Therefore any state-minimal NFA of a bideterministic language is a subautomaton of the bideterministic automaton.

But no strict subautomaton of the minimal DFA can accept the language, therefore bideterministic automaton is the only state-minimal NFA for that language.

Similarly, a bideterministic automaton is a transition-minimal NFA (but not necessarily unique).

Uniqueness of transition minimality

A bideterministic automaton is not necessarily the only transition-minimal NFA for the corresponding language.

The necessary and sufficient conditions for the unique transition-minimality are given by the following theorem:

Theorem. *A trim bideterministic automaton* $A = (Q, \Sigma, E, \{q_0\}, \{q_f\})$ *is a unique transition-minimal NFA if and only if the following three conditions hold:*

(i) $q_0 \neq q_f$,

(ii) $indegree(q_0) > 0$ *or* $outdegree(q_0) = 1$,

(iii) $indegree(q_f) = 1$ *or* $outdegree(q_f) > 0$.

Automaton morphism for a bideterministic language

Let A be a bideterministic automaton and let A' be another automaton accepting the same language.

Since $A = U_{L(A)}$, there exists an automaton morphism μ from A' into A.

Proposition. *μ is surjective.*

Proposition. *There is a transition (p, a, q) of A if and only if there is a transition (p', a, q') of A' such that $p'\mu = p$ and $q'\mu = q$.*

Based on these propositions, it is easy to see that μ defines an automaton transformation from A' to A.

Unambiguous ϵ-NFA

S. John (2003, 2004) has developed a theory to reduce the number of transitions of ϵ-NFAs.

Let A be an ϵ-NFA (Q, Σ, E, I, F) where E is partitioned into two subrelations $E_\Sigma = \{(p, a, q) \mid (p, a, q) \in E, a \in \Sigma\}$ and $E_\epsilon = \{(p, \epsilon, q) \mid (p, \epsilon, q) \in E\}$.

The automaton A is *unambiguous* if and only if for each $w \in L(A)$ there is exactly one path that yields w (without considering ϵ-transitions).

Slices

Let $L \subseteq \Sigma^*$ be a regular language, $U, V \subseteq \Sigma^*$, $a \in \Sigma$.

We call (U, a, V) a *slice* of L if and only if $U \neq \emptyset$, $V \neq \emptyset$ and $UaV \subseteq L$.

Let S be the set of all slices of L.

A partial order on S is defined by:
$(U_1, a, V_1) \leq (U_2, a, V_2)$ if and only if $U_1 \subseteq U_2$ and $V_1 \subseteq V_2$.

The set of *maximal slices* of L is defined by
$S_{max} := \{(U, a, V) \in S \mid$ there is no $(U', a, V') \in S$ with $(U, a, V) < (U', a, V')\}$.

Transition-minimal unambiguous ϵ-NFA

Let $S' \subseteq S$ be a finite slicing of L. In order to read an automaton $A_{S'}$ out of S', each slice from S' is transformed into a transition of $A_{S'}$, and these transitions are connected via states and ϵ-transitions using a follow-relation $\longrightarrow$ which is defined by:
$(U_1, a, V_1) \longrightarrow (U_2, b, V_2)$ if and only if $U_1 a \subseteq U_2$ and $bV_2 \subseteq V_1$

Theorem (S. John). *The three following statements are equivalent for languages $L \subseteq \Sigma^*$ if the slicing S_{max} of L induces an unambiguous ϵ-NFA $A_{S_{max}}$:*
1) *L is accepted by an ϵ-NFA*
2) *$L = L(A_{S'})$ for some finite slicing $S' \subseteq S$*
3) *S_{max} is finite*
Furthermore, $|S_{max}| \leq |S'| \leq |E_\Sigma|$.

Corollary (S. John). *An unambiguous ϵ-NFA $A_{S_{max}}$ has the minimum number of non-ϵ-transitions.*

Transition slice

For each non-ϵ-transition t of an automaton A, we define the *transition slice* of t to be the slice $(U_t, l(t), V_t)$ of $L(A)$ where

– U_t is the set of strings yielded by the paths from an initial state to the source state of t,

– $l(t)$ is the label of t, and

– V_t is the set of strings yielded by the paths from the target state of t to an accepting state.

A bideterministic automaton is a transition-minimal ϵ-NFA

Lemma. *For a bideterministic automaton A, let t_1 and t_2 be two different transitions of A, with the same label $a \in \Sigma$ and with the corresponding transition slices (U_{t_1}, a, V_{t_1}) and (U_{t_2}, a, V_{t_2}). Then $U_{t_1} \cap U_{t_2} = \emptyset$ and $V_{t_1} \cap V_{t_2} = \emptyset$.*

Proposition. *Each transition slice of a bideterministic automaton A is maximal.*

Theorem. *A bideterministic automaton A has the minimum number of transitions among all ϵ-NFAs accepting $L(A)$.*

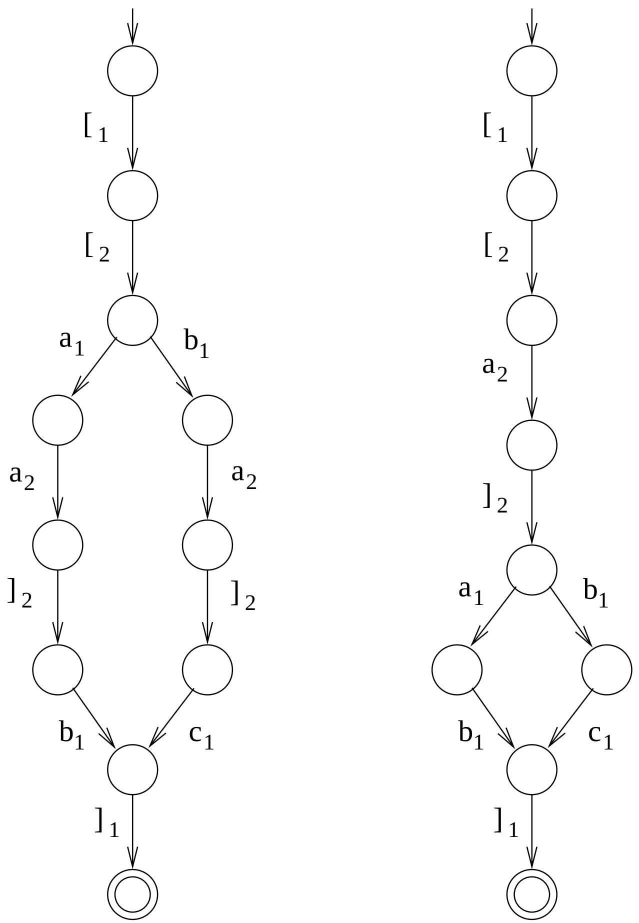
Multitape automaton model

Let us assume that a function $tape : Q \to \{1, ..., n\}$ associates every state of the automaton with a certain tape.

An n-tape automaton is given by a six-tuple $(Q, tape, \Sigma, E, I, F)$ where Q is a finite set of states with a partition into the sets $Q_1, ..., Q_n$ so that $Q_i = \{q \in Q \mid tape(q) = i\}$ for $i = 1, ..., n$, Σ is an input alphabet, $E \subseteq Q \times (\Sigma \cup \{[,]\}) \times Q$ is a set of transitions, $I \subseteq Q$ is a set of initial states and $F \subseteq Q$ is a set of final states.

Bideterministic multitape automata

$L = \{(ab, a), (bc, a)\}$



Strongly bideterministic automata

Let $q \in Q$ and $i \in \{1, ..., n\}$. A transition is called a *future transition* for the state q and tape i if it is the first transition involving this tape on some path that starts from q.

We call a deterministic multitape automaton *strongly deterministic* if for all $q \in Q$, all $i \in \{1, ..., n\}$ and all $a \in \Sigma \cup \{[,]\}$ there is at most one future transition for the state q and tape i with the label a.

An automaton is *strongly bideterministic* if it is strongly deterministic and so is its reversal automaton.

Goal: to show that a strongly bideterministic automaton is both state- and transition-minimal.

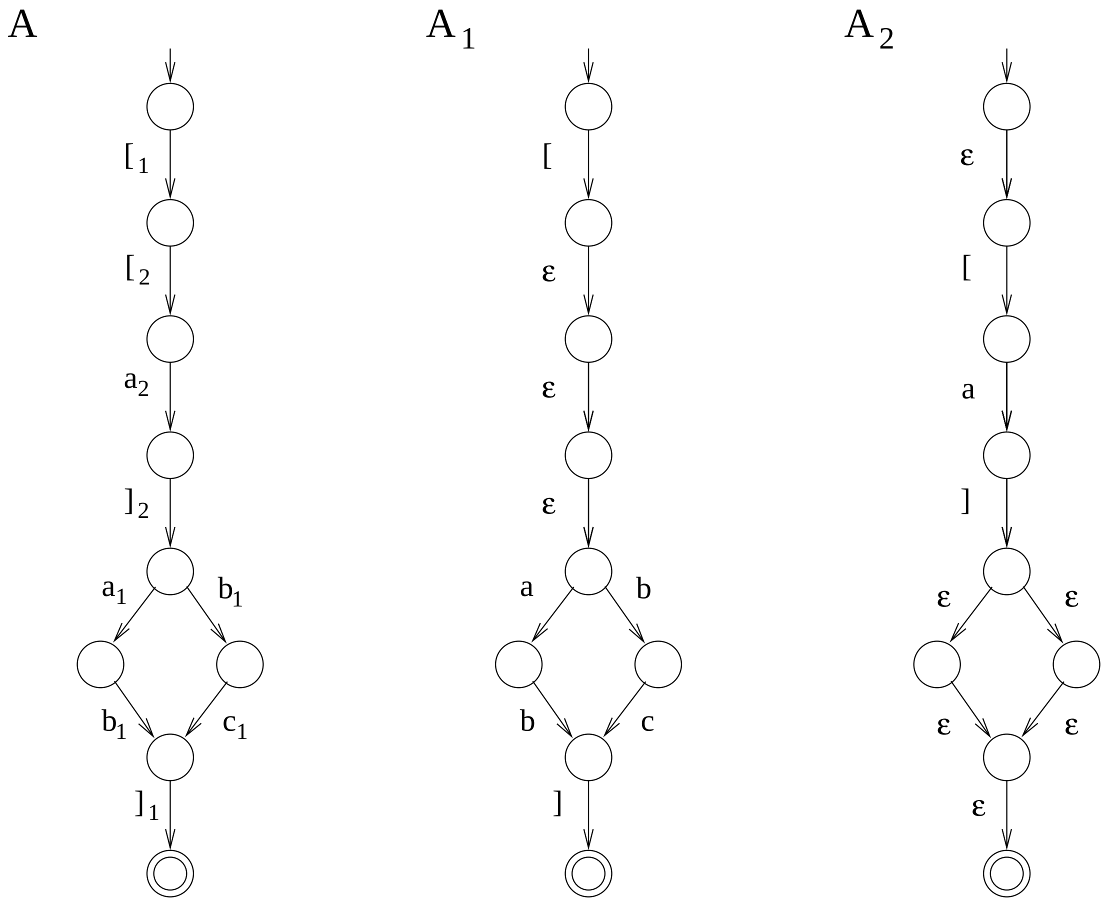
Strongly bideterministic automata

Let $A = (Q, tape, \Sigma, E, I, F)$ be an n-tape automaton.

Let us consider one-tape automata $A_1, A_2, ..., A_n$ such that each $A_i = (Q, \Sigma, E_i, I, F)$, $i = 1, ..., n$, is obtained from A by replacing all transitions that do not involve tape i, by ϵ-transitions, and by discarding the state-tape associations. Then $E_i = E_{i_\Sigma} \cup E_{i_\epsilon}$ where
$E_{i_\Sigma} = \{(p, a, q) \mid (p, a, q) \in E, p \in Q_i\}$ and
$E_{i_\epsilon} = \{(p, \epsilon, q) \mid (p, a, q) \in E, p \in Q \setminus Q_i\}$.
Clearly, each A_i is a one-tape ϵ-NFA accepting the language $L(A_i) = \{[w_i] \mid (w_1, ..., w_i, ..., w_n) \in L(A)\}$.

In the following, our goal is to show that if A is strongly bideterministic then each A_i has a minimum number of non-ϵ-transitions among all ϵ-NFAs for the language $L(A_i)$.

Bideterministic multitape automata



Strongly bideterministic automata

Let A be a strongly bideterministic n-tape automaton with $A_1, A_2, ..., A_n$ being the corresponding one-tape ϵ-NFAs.

The following results hold for all $i \in \{1, ..., n\}$.

Lemma. *Every A_i is unambiguous.*

Lemma. *Consider any A_i. Let t' and t'' be two different transitions of A_i, with the same label $l(t') = l(t'') = a \in \Sigma$ and with the corresponding transition slices $(U_{t'}, a, V_{t'})$ and $(U_{t''}, a, V_{t''})$. Then $U_{t'} \cap U_{t''} = \emptyset$ and $V_{t'} \cap V_{t''} = \emptyset$.*

Lemma. *For any A_i, every transition slice of A_i is maximal.*

Proposition. *Every A_i has the minimum number of non-ϵ-transitions among all ϵ-NFAs accepting $L(A_i)$.*

Strongly bideterministic multitape automata are transition-minimal

Proposition. *A strongly bideterministic multitape automaton A has the minimum number of transitions.*

Proof. Let us consider an n-tape automaton A and the corresponding one-tape ϵ-NFAs $A_1, A_2, ..., A_n$. Suppose that A is not transition-minimal. Let A' be a transition-minimal n-tape automaton accepting the same language with corresponding one-tape ϵ-NFAs $A'_1, A'_2, ..., A'_n$. Clearly, $L(A'_i) = L(A_i)$, $i = 1, ..., n$. Obviously, there must be some $j \in \{1, ..., n\}$ such that A'_j has less non-ϵ-transitions than A_j. Thus, A_j cannot be non-ϵ-transition-minimal which is a contradiction.

Further issues

- There may exist several different strongly bideterministic multitape automata that accept a given language.

  However, it is clear that they all must have the same number of transitions since the corresponding A_i-s have the same number of non-ϵ-transitions.

- State minimality of strongly bideterministic multitape automata.

- Automaton transformations between different strongly bideterministic multitape automata.

- Automaton transformations that produce strongly bideterministic multitape automata.