# New interpretation and generalization of the Kameda-Weiner method

Hellis Tamm
Institute of Cybernetics
Tallinn University of Technology

Estonian-Latvian Theory Days, Lilaste, Oct 16, 2016

# Main results

- We present a reinterpretation of the Kameda-Weiner method of finding a minimal nondeterministic finite automaton (NFA) of a language, in terms of atoms of the language.
- We introduce a method to generate NFAs from a set of languages, and show that the Kameda-Weiner method is a special case of it.
- Our method provides a unified view of the construction of several known NFAs (e.g. canonical RFSA, the átomaton, and others).

# Quotients and atoms

Let $L$ be a regular language over an alphabet $\Sigma$.

The left quotient of a language $L$ by a word $w$ is the language $w^{-1}L = \{x \in \Sigma^* \mid wx \in L\}$.

Let $K_0, \ldots, K_{n-1}$ be the quotients of $L$.

An atom of $L$ is any non-empty language of the form

$$\widetilde{K_0} \cap \widetilde{K_1} \cap \cdots \cap \widetilde{K_{n-1}},$$

where $\widetilde{K_i}$ is either $K_i$ or $\overline{K_i}$.

# Quotients and atoms

Let $L$ be a regular language over an alphabet $\Sigma$.

The left quotient of a language $L$ by a word $w$ is the language $w^{-1}L = \{x \in \Sigma^* \mid wx \in L\}$.

Let $K_0, \ldots, K_{n-1}$ be the quotients of $L$.

An atom of $L$ is any non-empty language of the form

$$\widetilde{K_0} \cap \widetilde{K_1} \cap \cdots \cap \widetilde{K_{n-1}},$$

where $\widetilde{K_i}$ is either $K_i$ or $\overline{K_i}$.

- Any quotient $K_i$ of $L$ (including $L$ itself) is a union of atoms.
- Atoms define a partition of $\Sigma^*$.
- Atoms are the classes of the left congruence of $L$: for $x, y \in \Sigma^*$, $x$ is equivalent to $y$ if for every $u \in \Sigma^*$, $ux \in L$ if and only if $uy \in L$.

# The átomaton

Let $K_0 = L$ be the initial quotient of $L$.
Let $A = \{A_0, \dots, A_{m-1}\}$ be the set of atoms of $L$.

- An atom is initial if it has $K_0$ (rather than $\overline{K_0}$) as a term.
- Let $I_A \subseteq A$ be the set of initial atoms.
- An atom is final if it contains $\varepsilon$.
- There is exactly one final atom $A_{m-1}$.
- The átomaton of $L$ is the NFA $\mathcal{A} = (A, \Sigma, \alpha, I_A, \{A_{m-1}\})$, where $A_j \in \alpha(A_i, a)$ if $A_j \subseteq a^{-1}A_i$.

## Some properties of the átomaton

- The language accepted by $\mathcal{A}$ is $L$.
- The (right) language of state $A_i$ of $\mathcal{A}$ is the atom $A_i$.
- The reverse automaton $\mathcal{A}^R$ of $\mathcal{A}$ is a minimal DFA for $L^R$ (the reverse language of $L$).
- The determinized automaton $\mathcal{A}^D$ of $\mathcal{A}$ is a minimal DFA of $L$.
- If $\mathcal{D}$ is a minimal DFA of $L$, then $\mathcal{A}$ is isomorphic to $\mathcal{D}^{RDR}$.

# Kameda-Weiner matrix

Kameda and Weiner (1970) used minimal DFAs for a language $L$ and its reverse $L^R$, to form a matrix, and based on the grids in this matrix, a minimal NFA was found.

- Trimmed minimal DFA $\mathcal{D}^T$ of $L$ with a state set $Q$.
- By Brzozowski's theorem, $\mathcal{D}^{RDT}$ is trim minimal DFA of $L^R$ with a state set $S \subseteq 2^Q \setminus \emptyset$.
- Form a matrix with rows corresponding to states $q_i$ of $\mathcal{D}$, and columns, to states $S_j \in S$ of $\mathcal{D}^{RDT}$.
- The $(i, j)$ entry is 1 if $q_i \in S_j$, and 0 otherwise.

# Quotient-atom matrix

- We use $\mathcal{D}^{RDRT}$, the trim átomaton of $L$, instead of $\mathcal{D}^{RDT}$, since the state sets of these automata are the same.
- The states of the minimal DFA correspond to quotients, and the states of the átomaton correspond to atoms of $L$.
- Interpret rows of the matrix as quotients, and columns as atoms of $L$ (exc. the empty quotient and the atom $\overline{K_0} \cap \cdots \cap \overline{K_{n-1}}$, if they exist).
- We call this matrix the quotient-atom matrix of $L$.
- Then the $(i, j)$ entry is 1 if and only if $A_j \subseteq K_i$.

# Grids and cover of the quotient-atom matrix

- A grid $g$ of the matrix is the direct product $g = P \times R$ of a set $P$ of quotients with a set $R$ of atoms, such that every atom in $R$ is a subset of every quotient in $P$.
- If $g = P \times R$ and $g' = P' \times R'$ are two grids, then $g \subseteq g'$ if and only if $P \subseteq P'$ and $R \subseteq R'$.
- A grid is maximal if it is not contained in any other grid.
- A cover is a set $G = \{g_0, \ldots, g_{k-1}\}$ of grids, such that every pair $(K_i, A_j)$ with $A_j \subseteq K_i$ belongs to some grid $g_i$ in $G$.

# NFA minimization by the Kameda-Weiner method

Let $f_G$ be the function that assigns to every non-empty quotient $K_i$, the set of grids $g = P \times R$ from a cover $G$, such that $K_i \in P$.

The constructed NFA is $\mathcal{N}_G = (G, \Sigma, \eta_G, I_G, F_G)$,
where $G$ is a cover consisting of (maximal) grids,
$I_G = f_G(K_0)$ is the set of grids involving the initial quotient $K_0$,
$g \in F_G$ if and only if $g \in f_G(K_i)$ implies that $K_i$ is a final quotient, and
$\eta_G(g, a) = \bigcap_{K_i \in P} f_G(a^{-1} K_i)$ for a grid $g = P \times R$ and $a \in \Sigma$.

# NFA minimization by the Kameda-Weiner method

Let $f_G$ be the function that assigns to every non-empty quotient $K_i$, the set of grids $g = P \times R$ from a cover $G$, such that $K_i \in P$.

The constructed NFA is $\mathcal{N}_G = (G, \Sigma, \eta_G, I_G, F_G)$,
where $G$ is a cover consisting of (maximal) grids,
$I_G = f_G(K_0)$ is the set of grids involving the initial quotient $K_0$,
$g \in F_G$ if and only if $g \in f_G(K_i)$ implies that $K_i$ is a final quotient, and
$\eta_G(g, a) = \bigcap_{K_i \in P} f_G(a^{-1}K_i)$ for a grid $g = P \times R$ and $a \in \Sigma$.

It may be the case that $\mathcal{N}_G$ does not accept the language $L$.

A cover $G$ is called legal if $L(\mathcal{N}_G) = L$.

To find a minimal NFA of a language $L$, the method tests the covers of the matrix in the order of increasing size to see if they are legal.

The first legal NFA is a minimal one.

# Reinterpretation of the Kameda-Weiner method

Let $R$ be a set of atoms and let $U(R) = \bigcup_{A_j \in R} A_j$.

**Theorem**

*Let $G = \{g_0, \ldots, g_{k-1}\}$ be a cover consisting of maximal grids $g_i = P_i \times R_i$, and let $\mathcal{N}_G = (G, \Sigma, \eta_G, I_G, F_G)$ be the corresponding NFA, obtained by the Kameda-Weiner method. It holds that*

- $g_i \in I_G$ *if and only if* $U(R_i) \subseteq L$,
- $g_i \in F_G$ *if and only if* $\varepsilon \in U(R_i)$,
- $g_j \in \eta_G(g_i, a)$ *if and only if* $U(R_j) \subseteq a^{-1}U(R_i)$ *holds, for any* $g_i, g_j \in G$ *and* $a \in \Sigma$.

# Reinterpretation of the Kameda-Weiner method

Let $R$ be a set of atoms and let $U(R) = \bigcup_{A_j \in R} A_j$.

**Theorem**

*Let $G = \{g_0, \ldots, g_{k-1}\}$ be a cover consisting of maximal grids $g_i = P_i \times R_i$, and let $\mathcal{N}_G = (G, \Sigma, \eta_G, I_G, F_G)$ be the corresponding NFA, obtained by the Kameda-Weiner method. It holds that*

- *$g_i \in I_G$ if and only if $U(R_i) \subseteq L$,*
- *$g_i \in F_G$ if and only if $\varepsilon \in U(R_i)$,*
- *$g_j \in \eta_G(g_i, a)$ if and only if $U(R_j) \subseteq a^{-1}U(R_i)$ holds, for any $g_i, g_j \in G$ and $a \in \Sigma$.*

We note that essentially the same approach to the Kameda-Weiner method which uses projections of grids, consisting of subsets of the state set of the DFA $\mathcal{D}^{RDT}$ (corresponding to sets of atoms), was presented by Champarnaud and Coulon (IJFCS, 2005).

# Generating automata by a set of languages

Let $L$ be a regular language, and let $K = \{K_0, \ldots, K_{n-1}\}$ be the set of quotients of $L$.

**Definition**

A set $\{L_0, \ldots, L_{k-1}\}$ of languages is a cover of the quotients of $L$, or simply, a cover for $L$, if every quotient $K_j$ of $L$ is a union of some $L_i$'s.

Note that $L$ itself is a union of some $L_i$'s, because $L = \varepsilon^{-1} L$.

# Generating automata by a set of languages

Let $L$ be a regular language, and let $K = \{K_0, \ldots, K_{n-1}\}$ be the set of quotients of $L$.

### Definition

A set $\{L_0, \ldots, L_{k-1}\}$ of languages is a cover of the quotients of $L$, or simply, a cover for $L$, if every quotient $K_j$ of $L$ is a union of some $L_i$'s.

Note that $L$ itself is a union of some $L_i$'s, because $L = \varepsilon^{-1}L$.

We define the NFA based on a cover $\{L_0, \ldots, L_{k-1}\}$ as follows:

### Definition

The NFA generated by a cover $\{L_0, \ldots, L_{k-1}\}$ for $L$ is defined by $\mathcal{G} = (Q, \Sigma, \delta, I, F)$, where $Q = \{q_0, \ldots, q_{k-1}\}$, $I = \{q_i \mid L_i \subseteq L\}$, $F = \{q_i \mid \varepsilon \in L_i\}$, and $q_j \in \delta(q_i, a)$ if and only if $L_j \subseteq a^{-1}L_i$ holds for all $q_i, q_j \in Q$ and $a \in \Sigma$.

# Some properties of a generated automaton

Let $\mathcal{G} = (Q, \Sigma, \delta, I, F)$ be generated by a cover $\{L_0, \ldots, L_{k-1}\}$ for $L$.
We denote the (right) language of state $q_i$ of $\mathcal{G}$ by $L_{q_i}(\mathcal{G})$.

## Proposition

*The following properties hold:*

- $L_{q_i}(\mathcal{G}) \subseteq L_i$ *for every* $q_i \in Q$.
- $L(\mathcal{G}) \subseteq L$.

# Some properties of a generated automaton

Let $\mathcal{G} = (Q, \Sigma, \delta, I, F)$ be generated by a cover $\{L_0, \ldots, L_{k-1}\}$ for $L$.
We denote the (right) language of state $q_i$ of $\mathcal{G}$ by $L_{q_i}(\mathcal{G})$.

**Proposition**

*The following properties hold:*

- $L_{q_i}(\mathcal{G}) \subseteq L_i$ *for every* $q_i \in Q$.
- $L(\mathcal{G}) \subseteq L$.

**Proposition**

*If* $a^{-1}L_i$ *is a union of* $L_j$*'s for every* $L_i$ *and* $a \in \Sigma$*, then* $\mathcal{G}$ *accepts* $L$.

However, this condition is not necessary for generated NFA to accept $L$.

# Examples

## Example

Consider the set $K = \{K_0, \ldots, K_{n-1}\}$ of quotients of $L$ as a cover for $L$.
The NFA $\mathcal{G}_K$, generated by the set $K$, is the saturated version of the minimal DFA of $L$.
Since for every quotient $K_i$ and $a \in \Sigma$ there exists some quotient $K_j$ such that $a^{-1}K_i = K_j$, $\mathcal{G}_K$ accepts $L$.

# Examples

## Example

Consider the set $K = \{K_0, \ldots, K_{n-1}\}$ of quotients of $L$ as a cover for $L$.
The NFA $\mathcal{G}_K$, generated by the set $K$, is the saturated version of the minimal DFA of $L$.
Since for every quotient $K_i$ and $a \in \Sigma$ there exists some quotient $K_j$ such that $a^{-1}K_i = K_j$, $\mathcal{G}_K$ accepts $L$.

## Example

Consider the set $K' \subseteq K$ of prime quotients of $L$, that is, those non-empty quotients of $L$ which are not unions of other quotients, as a cover for $L$.
The NFA $\mathcal{G}_{K'}$ generated by $K'$ is known as the canonical residual finite state automaton (canonical RFSA) of $L$.
Since every quotient is a union of prime quotients, so is $a^{-1}K'_i$, for a prime quotient $K'_i$ and $a \in \Sigma$. Thus, $\mathcal{G}_{K'}$ accepts $L$.

# Example: generating the átomaton

## Example

Consider the set $A = \{A_0, \ldots, A_{m-1}\}$ of atoms of $L$. The set $A$ is a cover for $L$, because every quotient of $L$ is a union of atoms.
The NFA $\mathcal{G}_A$, generated by the set $A$, is the átomaton of $L$.
It is known that for every atom $A_i$ and $a \in \Sigma$, $a^{-1}A_i$ is a union of atoms.
Thus, $\mathcal{G}_A$ accepts $L$.

# Kameda-Weiner method as a special case of generating an NFA

The following theorem shows that the NFA minimization method presented by Kameda and Weiner is a special case of generating an NFA:

**Theorem**

*Let $G = \{g_0, \dots, g_{k-1}\}$ be a set of maximal grids, with $g_i = P_i \times R_i$, forming a cover of the quotient-atom matrix of $L$. The NFA $\mathcal{N}_G$, obtained by the Kameda-Weiner method using $G$, is isomorphic to the NFA $\mathcal{G}_U$, generated by the set $U = \{U(R_0), \dots, U(R_{k-1})\}$.*

# Generality of our method

**Theorem**

*If there is a trim NFA accepting $L$, with the set $\{L_0, \ldots, L_{k-1}\}$ of languages of its states, then the NFA generated by the cover $\{L_0, \ldots, L_{k-1}\}$ for $L$ is such an NFA.*

If one is interested in finding an NFA for a given language, such that the states of that NFA correspond to certain languages, our method can be used to generate such an NFA if it exists.

If the generated NFA is not such an NFA, then it does not exist.

# Conclusions

- Our theory shows that atoms of regular languages have an important role in the Kameda-Weiner method.
- We hope that our contributions provide a useful insight into the problem of NFA minimization.
- We also think that the introduced method of generating NFAs is of interest on its own as shown by the examples.
- This method provides a unified view of the construction of several known NFAs.