k-Universality of Regular Languages

Duncan Adamson[1] Pamela Fleischmann[2] Annika Huch[2]
Tore Koß[3] Florin Manea[3] Dirk Nowotka[2]

[1]University of Liverpool, UK

[2]Kiel University, Germany

[3]University of Göttingen, Germany

Theorietag "Automaten und Formale Sprachen" 2023
05.10.2023

Preliminaries

w	b	a	c	b	a	a	b	a	d	a

Preliminaries

d

ba

bc

w $\boxed{b \quad a \quad c \quad b \quad a \quad a \quad b \quad a \quad d \quad a}$

bb

baab

a

c

b

Preliminaries

d

abc

ba

bc

w $\boxed{b\ a\ c\ b\ a\ a\ b\ a\ d\ a}$

bb

baab

a

dd

c

b

Preliminaries

d

~~abc~~

ba

bc

w $\boxed{b \quad a \quad c \quad b \quad a \quad a \quad b \quad a \quad d \quad a}$

bb

baab

a

~~dd~~

c

b

Preliminaries – Subsequences and Universality

Definition

- We call v a subsequence of w, if there exist positions $1 \leq i_1 < i_2 < \ldots < i_k \leq |w|$, such that $v = w[i_1]w[i_2] \cdots w[i_k]$.
- We denote the set of all subsequences (of length exactly k) of w by $\mathrm{Subseq}(w)$ ($\mathrm{Subseq}_k(w)$).
- A word w is k-universal iff $\mathrm{Subseq}_k(w) = \Sigma^k$.
- If v is not a subsequence of w we call it an *absent subsequence*.

Preliminaries – Arch Factorisation

Definition

The universality index $\iota(w)$ is the unique integer such that w is $\iota(w)$-universal but not $(\iota(w)+1)$-universal.

Definition (Arch-Factorisation, Hébrard 1991)

Let $w \in \Sigma^*$. Then $w = \mathsf{ar}_w(1) \cdots \mathsf{ar}_w(\iota(w)) r(w)$ such that $\iota(\mathsf{ar}_w(i)) = 1$, the last letter of $\mathsf{ar}_w(i)$ occurs exactly once in $\mathsf{ar}_w(i)$ and $\iota(r(w)) = 0$. $\mathsf{ar}_w(i)$ are called the *arches of w* and $r(w)$ is called the *rest of w*.

Preliminaries – Finite Automata

Definition

A finite automaton is a 5-tuple $\mathcal{A} = (Q, \Sigma, \delta, q_0, F)$, where Q is a finite set of states, Σ is an alphabet, $\delta : Q \times \Sigma \to 2^Q$ is the transition function, $q_0 \in Q$ is the initial state und $F \subseteq Q$ is a set of final states. If $|\delta(q, a)| \leq 1$ for all $q \in Q$ and $a \in \Sigma$ we call $\mathcal{A}$ *deterministic* (DFA), otherwise we call it *non-deterministic* (NFA).

We generally let $|Q| = n$ and $|\Sigma| = \sigma$.

Subsequence Universality for Languages

Definition

The downward closure of a language L is defined as
$L^{\downarrow} = \{v \in \Sigma^* \mid \exists w \in L : v \in \mathrm{Subseq}(w)\}$.

One could say L is k-universal iff $\Sigma^k \subset L^{\downarrow}$.

A reduction from the Hamiltonian Path Problem yields NP-hardness of the problem to determine the shortest word which is not a subsequence of any word accepted by an FA $\mathcal{A}$. Hence the problem to determine whether $\Sigma^k \subset L^{\downarrow}$ for any regular language L is coNP-hard.

Subsequence Universality for Languages

Definition

- L is k-$\exists$-universal iff there is a word in L which is k-universal.
- L is k-$\forall$-universal iff every word in L is k-universal.

Problem

How efficient can we decide, given a language L and an integer k, whether L is k-$\exists$-universal (k-ESU) or k-$\forall$-universal (k-ASU)?

Universal Universality

For any language L the set $L^{\forall}$ of words occurring as subsequences in all words $w \in L$ is finite ($L^{\forall} = \bigcap_{w \in L} \mathrm{Subseq}(w)$ and $\mathrm{Subseq}(w)$ is finite) but can still be exponential in the length of the shortest word in L.

Universal Universality

Let $\mathcal{A} = (Q, \Sigma, q_0, F, \delta)$ be an NFA. We can decide whether $\mathcal{A}$ satisfies k-ASU in deterministic polynomial time:

(i) For $q, q' \in Q$ we define a relation R_a for every $a \in \Sigma$ such that qR_aq' if and only if there is a state q'' such that there is a path from q to q'' not containing any a and also a transition from q'' to q' labelled by a.

(ii) Let qRq' if and only if there is $a \in \Sigma$ such that qR_aq'.

(iii) Let $Q' = \{q \in Q \mid \text{there is a non-universal path from } q \text{ to } F\}$.

(iv) Let $G = (V, E)$ be a directed graph with $V = Q$ and $(q, q') \in E$ if and only if qRq'.

(v) There is an ℓ-universal word, for an $\ell < k$, accepted by $\mathcal{A}$ if and only if there is a path of length at most $k - 1$ from q_0 to any node corresponding to a state in Q' in G.

Existential Universality

Theorem

We can decide k-ESU in $O^*(n^3 2^\sigma)$ *(where the star only hides poly(*σ*)-factors resulting from arithmetic with large integers).*

Theorem

For $\sigma \in \Omega(\log n)$*, k-ESU is NP-complete (even for* $k = 1$*).*

Existential Universality – FPT

Let $\mathcal{A}$ be a NFA. To check whether $L(\mathcal{A})$ satisfies k-ESU do the following:

(i) Remove non-accessible and non-co-accessible states in $O(n^3)$

(ii) Check whether there is a loop labelled with a 1-universal word, if so accept independently from k.

(iii) Otherwise, for every $q \in Q$, find maximal set V_q of letters occurring in a word β_q which is label of a path from q to q (V_q is unique since the path may contain q more than twice) in $O^*(n^3 2^\sigma)$.

(iv) We can maximise the universality of any word $w \in L(\mathcal{A})$ by pumping β_q^2 for every state q in an accepting path labelled with w.

(v) Determine maximal universality of words in $L(\mathcal{A})$ in $O^*(n^3 2^\sigma)$ with dynamic programming: let $M[\cdot][\cdot]$ be an $n \times 2^\sigma$ matrix such that $M[q_r][V]$ is the maximal universality of a word w labelling a path from q_0 to q_r such that $r(w) = V$.

Existential Universality – NP-membership

Let $\mathcal{A}$ be an NFA with n states over an alphabet of size σ.

Lemma

If $\mathcal{A}$ accepts a k-universal word it also accepts a k-universal word of length at most $kn\sigma - (n-1)(k-1)$

Remark

Let $k > n$, then $\mathcal{A}$ accepts a k-universal word if and only if there is a state q and a path from q to q labelled with a 1-universal word.

Existential Universality – NP-membership

Let $\mathcal{A}$ be an NFA with n states over an alphabet of size $\sigma \in \Omega(\log n)$. To check whether $L(\mathcal{A})$ satisfies k-ESU do the following:

(i) Remove non-accessible and non-co-accessible states.

(ii) Guess non-deterministically whether there is a loop labelled with a 1-universal word of length at most $n\sigma$, if so accept independently from k, if not reject if $k > n$.

(iii) Otherwise, check all words in $\Sigma^{\leq kn\sigma}$ if they are k-universal and accepted by $\mathcal{A}$.

Existential Universality – NP-hardness

Following a proof by Kim, Han, Ko, Salomaa:
Let $G = (V, E)$ a graph with $V = \{v_1, \ldots, v_n\}$. We construct an automaton $\mathcal{A}$ with $n^2 + 2$ states $q_{0,0}, q_f$ and $q_{i,j}$ for $i, j \in [1 : n]$ where $q_{0,0}$ is the initial state, q_f is a failure state, $q_{n,j}$ is final for every $j \in [1 : n]$ and there is a transition from $q_{i,j}$ to $q_{\ell,k}$ labelled with k if and only if $\ell = i + 1$ and either $i = j = 0$ or there is an edge from v_j to v_k in G. Intuitively $q_{i,j}$ represents visiting v_j at the i^{th} step in some path in G.
Then a Hamiltonian Path in G corresponds 1-to-1 to an accepting path labelled with a 1-universal word accepted by $\mathcal{A}$.

Counting and Ranking k-universal Words

Let $L \subset \Sigma^*$ be a formal language.

- The problem of counting words of L is to determine the size of L.
- The problem of ranking a word $w \in L$ is to determine the size of the set $\{v \in L \mid v \prec w\}$ where $\prec$ is an arbitrary ordering of Σ^*, e.g. the length-lexicographic ordering.

Counting and Ranking k-universal Words

In DFAs there is a one-to-one correspondence between paths and words, in NFAs any word can correspond to several paths. For the sake of simplicity we count and rank accepting paths in an FA labelled with k-universal words instead of k-universal words accepted by an FA.

Counting and Ranking k-universal Words

Given an FA $\mathcal{A} = (Q, \Sigma, \delta, q_0, F)$ and two integers k, m we are interested in counting and ranking the accepting paths of length (at most) m labelled with a k-universal word.

We use a dynamic programming approach: Let T be a table of size $n \times (m+1) \times k \times 2^\sigma$ called path table of length m. For any $q \in Q, \ell \in [0, m], c \in [0, k-1]$ and $\mathcal{R} \subset \Sigma$ the entry $T[q][\ell][c][\mathcal{R}]$ denotes the amount of paths π such that π is an ℓ-length path from q_0 to q labelled with a word w such that $\iota(w) = c$ and $\mathcal{R}$ is minimal such that $r(w) \in \mathcal{R}^*$.

Another $n \times (m+1)$ table U, where $U[q][\ell]$ denotes the amount of k-universal paths from q_0 to q of length ℓ, is used to fill T.

T and U allow us to count and rank k-universal paths (of length exactly/at most m) in $\mathcal{A}$.

Type	Length	Complexity
Counting	unrestricted	$O^*(n^4k^32^\sigma)$
Counting	exactly m	$O^*(n^2m^2k2^\sigma)$
Counting	at most m	$O^*(n^2m^2k2^\sigma)$
Ranking	unrestricted	$O^*(n^4k^32^\sigma)$
Ranking	exactly m	$O^*(n^2m^2k2^\sigma)$
Ranking	at most m	$O^*(n^2m^2k2^\sigma)$

Type	Length	Complexity
Counting	unrestricted	$O^*(n^4 k^3 2^\sigma)$
Counting	exactly m	$O^*(n^2 m^2 k 2^\sigma)$
Counting	at most m	$O^*(n^2 m^2 k 2^\sigma)$
Ranking	unrestricted	$O^*(n^4 k^3 2^\sigma)$
Ranking	exactly m	$O^*(n^2 m^2 k 2^\sigma)$
Ranking	at most m	$O^*(n^2 m^2 k 2^\sigma)$

Thank you for listening!