Quasi-Distances and Weighted Finite Automata

Timothy Ng, David Rappaport, and Kai Salomaa

School of Computing, Queen's University

DCFS 2015, Waterloo, ON

June 27, 2015

Are neighbourhoods of a regular language also regular?
What is the state complexity of the neighbourhood of a regular language?

We use weighted finite automata to help us show the state complexity of the neighbourhood of a regular language (Salomaa, Schofield 2007).

1. Can additive additive WFAs recognize neighbourhoods with respect to additive quasi-distances?

1. Can additive additive WFAs recognize neighbourhoods with respect to additive quasi-distances?
2. Is there a lower bound example for the state complexity of additive WFA languages over an alphabet with a constant number of symbols?

A distance is a function $d : \Sigma^* \times \Sigma^* \to [0, \infty)$ such that

1. $d(x, y) = 0$ if and only if $x = y$
2. $d(x, y) = d(y, x)$
3. $d(x, y) \leq d(x, w) + d(w, y)$

A distance is a function $d : \Sigma^* \times \Sigma^* \to [0, \infty)$ such that

1. $d(x, y) = 0$ if and only if $x = y$
2. $d(x, y) = d(y, x)$
3. $d(x, y) \leq d(x, w) + d(w, y)$

If condition (1) is relaxed to $d(x, y) = 0$ if $x = y$, then d is a quasi-distance.

Islington $\rightarrow$ Eglin_ton

Montréal → Montreal

The neighbourhood of a language $L \subseteq \Sigma^*$ of radius $r \geq 0$ with respect to a distance measure d is the set of all words u with $d(w, u) \leq r$ for some $w \in L$,

$$E(L, d, r) = \{u \in \Sigma^* \mid (\exists w \in L)\, d(w, u) \leq r\}.$$

For which distances are neighbourhoods of regular languages regular for all radii $r \geq 0$?

A distance d on Σ^* is additive if for all factorizations $w = w_1 w_2$, we have for all $r \geq 0$

$$E(\{w\}, d, r) = \bigcup_{r_1 + r_2 = r} E(\{w_1\}, d, r_1) \cdot E(\{w_2\}, d, r_2)$$

A distance d on Σ^* is additive if for all factorizations $w = w_1 w_2$, we have for all $r \geq 0$

$$E(\{w\}, d, r) = \bigcup_{r_1 + r_2 = r} E(\{w_1\}, d, r_1) \cdot E(\{w_2\}, d, r_2)$$

Theorem (Calude, Salomaa, Yu 2002)

Let d be an additive quasi-distance on Σ^* and $L \subseteq \Sigma^*$ be a regular language. Then $E(L, d, r)$ is regular for all $r \geq 0$.

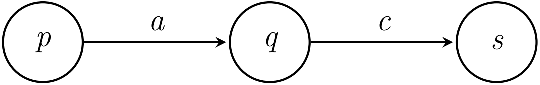
An additive weighted finite automaton is a 6-tuple $A = (Q, \Sigma, \gamma, \omega, q_0, F)$, where

- Q is the set of states
- Σ is the alphabet
- γ is the transition function
- ω is the weight function
- q_0 is the initial state
- F is the set of final states

Theorem (Salomaa, Schofield 2007)

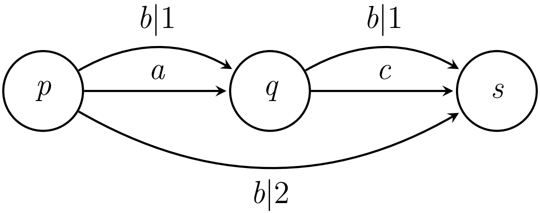
Let A be an NFA, d an additive distance, and $r_0 \geq 0$. We can construct an additive WFA which recognizes the neighbourhood $E(L(A), d, r)$ for any $0 \leq r \leq r_0$.

This involves adding transitions with the appropriate weight.



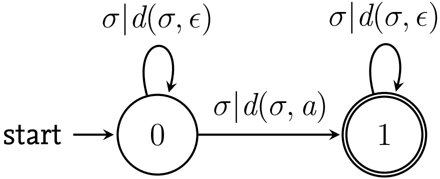
We can do this because neighbourhoods of additive distances are finite.

This involves adding transitions with the appropriate weight.

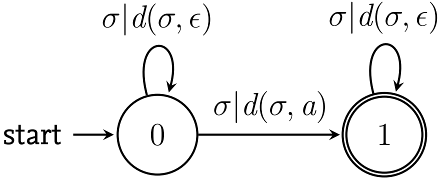


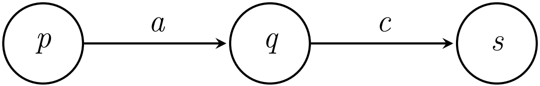
We can do this because neighbourhoods of additive distances are finite.

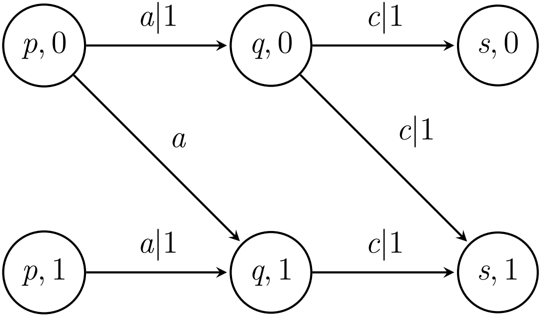
How do we construct an additive WFA for neighbourhoods with respect to quasi-distances?



$\sigma|d(\sigma,\epsilon)$
$\sigma|d(\sigma,\epsilon)$
$\sigma|d(\sigma,a)$
start
0
1


σ|d(σ, ϵ)
σ|d(σ, ϵ)
σ|d(σ, a)
start
0
1


p
a
q
c
s



Compute all-pairs shortest paths and consider the paths with weight at most r.

Theorem

Suppose that L has an NFA with n states and d is a quasi-distance. The neighbourhood of L of radius r can be recognized by an additive WFA having n states within weight bound r.

We can construct an equivalent DFA that recognizes a WFA with weight up to r. This requires at most $(r+2)^n$ states.

We can construct an equivalent DFA that recognizes a WFA with weight up to r. This requires at most $(r+2)^n$ states.

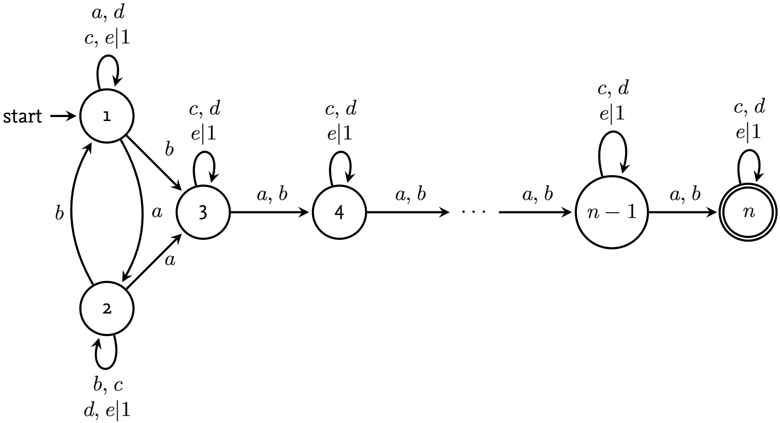
Theorem (Salomaa, Schofield 2007)

Let A be an additive WFA with integer weights and $r \in \mathbb{N}$. The language $L(A, r)$ can be recognized by a DFA having $(r+2)^n$ states.

We only need to keep track of the minimal weight computation that reaches each state.

$$(i_1, i_2, i_3, i_4, i_5, i_6, i_7, i_8)$$

$$(1, 1, 0, r+1, 4, r+1, r+1, r+1)$$



$$ac^{k_n}\, bd^{k_{n-1}}\, ac^{k_{n-2}} \cdots ac^{k_3}\, bd^{k_2}\, c^{k_1}, \text{ if } n \text{ is odd;}$$

$$abd^{k_n}\, ac^{k_{n-1}}\, bd^{k_{n-2}} \cdots ac^{k_3}\, bd^{k_2}\, c^{k_1}, \text{ if } n \text{ is even.}$$

Theorem

For $n, r \in \mathbb{N}$, there exist an n-state WFA with integer weights defined over a five-letter alphabet such that the state complexity of $L(A, r)$ is $(r + 2)^n$.

What is the state complexity of additive neighbourhoods?

- The WFA model implies an upper bound of $(r+2)^n$.
- Is there a matching lower bound?
- What is the state complexity when we consider specific distances?