Learning Deterministic Weighted Automata with Queries and Counterexamples

Gail Weiss, Yoav Goldberg, and Eran Yahav

Expanding the Observation Table

\( O_{P,S} \) is expanded until it is closed and consistent, i.e.:

- **Closed:**
  - For every \( p \in P, \sigma \in \Sigma \) there exists \( p' \in P \) such that \( O_{P,S}(p,\sigma) \approx_{t} O_{P,S}(p') \)

- **Consistent:**
  - For every \( p_1, p_2 \in P, \sigma \in \Sigma \) if \( O_{P,S}(p_1,\sigma) \approx_{t} O_{P,S}(p_2,\sigma) \) then \( p_1 \approx_{t} p_2 \)

Building a Hypothesis

- \( P \) is partitioned into a partitioning (clustering) \( C \) which satisfies determinism and \( t \)-equality, i.e.:

  - **Determinism:**
    - For every \( c \in C, p_1, p_2 \in c, \sigma \in \Sigma \) if there exist \( c_1, c_2 \in C \) such that
      - \( p_1, p_2 \in c_1 \)
      - \( p_1, p_2 \in c_2 \)
    - Then \( c_1 = c_2 \)

  - **\( t \)-equality:**
    - For every \( c \in C, p_1, p_2 \in c \)
      - \( O_{P,S}(p_1) \approx_{t} O_{P,S}(p_2) \)

This is watched in case the clustering algorithm puts together prefixes that are not \( t \)-equal to each other (recall that \( t \)-equality is non-transitive)

This is remedied by splitting violating cliques

Building:

- \( C \) defines the states of the hypothesis, and its transitions and weights are mapped according to the observations in \( O_{P,S} \)

Try it yourself:
https://github.com/tech-srl/weighted_lstar