An experimental comparison of active learning algorithms on fixed data sets

Du Hyun Cho - 362801 - du.hyun.cho@rwth-aachen.de

March 2025

Abstract

Abstract should be written? or can i ignore this part?

1 Introduction

compare the performance of active DFA learning algorithms when constructing DFAs from examples provided in specific data sets.

2 Preliminaries

2.1 Basic Notation

- I^* is the set of all words, e.g. $aaa, abb, bbca, \ldots \in I^*$,
- A subset $L \subseteq I^*$ of words is called a (formal) language.
- usw.. other notations used for this thesis

2.2 Regular language

no necessary?

2.3 DFA

no necessary?

2.4 Learning

: Overview about active algo for example.. introduce about membership, equiv query

We use for automaton learning L^* algorithm. It is an active automaton learning algorithm Automaton learning is used for, for example construction of a model from an implementation (e.g. implementation of protocol such like TCP). We need for the learning of DFA "membership query" and "equivalence query" For these two queries we need to know what a learner and teacher are. In

this subsection we describe about who they are and what is their role for learning automaton. Assume, L is an unknown language and learner wants to construct automaton for L. So what is the role of learner? The learner is given a finite sample of a language L: examples of words in L and examples of words not in L. For example, the learner can query the language L. First, learner can ask "Is $w \in L$"? for words w that the learner can choose. We call this a "Membership query". And then, for a DFA A, the learner can ask if whether $L = L(A)$. We call this a "Equivalence query". This A is often called a hypothesis.
So what is about a teacher? Above we see membership query and equivalence query. Now we consider a model that is referred to as "minimally adequate teacher" (MAT) model, with the following two types of queries. Assume that w (The Learner can choose) is a word as input and there is a membership query "Is the input $w \in L$?" Then the teacher should answer this question with "yes" or "no". And there is equivalence query "Does an Automaton accept L?" that the learner can choose. Then the teacher can answer with "yes" or "no". The answer "no" means that A is not correct on w for a word w that the teacher can choose. We call this w "counter example". A counterexample is a string in the symmetric difference of the correct language and the guessed language.

2.5 L* Learning

: more precise about active algorithm

The learning algorithm L^* is described as correctly learning the regular set from the minimally adequate Teacher in polynomial time in the minimum number of DFA states for the set and the longest length of any counter example provided by the teacher Now we have a look the algorithm L^*.

To understand this algorithm, we need to know what exactly are observation tables, escaping words and closed observation tables. First, an observation table $B = (R, E, f)$ consists of

- a set $R \subseteq \Sigma^*$ (for the representatives),
- a set $E \subseteq \Sigma^*$ (for the experiments),
- and a function $f\colon (R \cup R \cdot \Sigma) \times E \longrightarrow O$

3 Experiment

with learnlib implement using various dataset simulation

4 Analyse

compare outputs and performance in various perspective.

Algorithm 1 The pseudo-code for L^* is given in Algorithm 1

initialization: an observation table $B = (R, E, f)$ with $R{:=}\{\epsilon\}$ and $E{:=}\{\epsilon\}$
repeat
  while (R, E) is not closed or not consistent **do**
    if (R, E) is not closed **then**
      find $r_1 \in R$, $a \in \Sigma^*$ s.t. $row(r_1 a) \neq row(r)$, for all $r \in R$
      $R{:=}R \cup \{r_1 a\}$
    end if
    if (R, E) is not consistent **then**
      find $r_1, r_2 \in R, a \in A$, and $e \in E$ such that $row(r_1) = row(r_1)$ and $L(r_1 ae) \neq L(r_2 ae)$
      $E{:=}E \cup \{ae\}$
    end if
    Make the conjecture $M(R, E)$
    if The teacher answers with no, with a counter-example c **then**
      $R{:=}R \cup prefixes(c)$
    end if
    Until the teacher answers with yes to the conjecture $M(R, E)$.
return $M(R, E)$

5 Conclusion

Describe limitation..

References