Is Learning RFSAs Better Than Learning DFAs?

Pedro García[1], José Ruiz[1], Antonio Cano[1], and Gloria Alvarez[2]

[1] Universidad Politécnica de Valencia. Valencia, Spain
[2] Pontificia Universidad Javeriana - Seccional Cali. Cali, Colombia
{pgarcia, jruiz, acano, galvarez}@dsic.upv.es

Abstract. Inference of RFSAs has been recently presented [1] as an alternative to inference of DFAs if the target language has been obtained by a random generation of NFAs. We propose in this paper the algorithm RPNI2, which is a variation of the previous RPNI, that also outputs DFAs as hypothesis. The experiments done using the same data as in [1] show that RPNI2 has an error rate very similar to the rate obtained in the inference of RFSAs, but the size of the hypothesis is substantially smaller.

1 Description of the Algorithms RPNI and RPNI2

The RPNI (Regular Positive and Negative Inference) algorithm can be found in [2]. Definitions and previous works concerning RFSAs and DeLeTe2 algorithm can be found in [1] and in some other previous works of the same authors.

The RPNI (Regular Positive and Negative Inference) algorithm [2] is used for inference of regular languages. It receives a sample of the target language as input and it outputs, in polynomial time, a DFA consistent with the input. This algorithm converges to the minimal automaton of the target language in the limit (i.e. when it has received a characteristic sample as input).

RPNI works merging every state of the Prefix Tree Moore Machine of the sample with the previous ones in lexicographical order and propagates the merges done to keep a deterministic automaton under the condition that it does not accept a negative sample.

Merging states can be seen as a process of enlarging the learning sample, as states that have undefined output in the tree, may now be defined if they can be merged with a state whose output belongs to $\{0, 1\}$.

The main idea of the variation of RPNI that we propose in this paper and we call RPNI2 is the following: If two states p and q can not be merged we try to establish the possible inclusion relation between them (We say that $q \prec q'$ if no word w exists such that $\delta(q, w)$ is a final state whereas $\delta(q', w)$ is not), which will sometimes permit us to define the output associated to some states that were previously undefined (if $q \prec q'$, then if q is final and q' is undefined, q' can be set as final, otherwise if q' is not final and q is undefined, q can be set as final). Except for this variation, RPNI2 works exactly as previous RPNI does and it converges to the minimal DFA of the target language.

J. Farré, I. Litovsky, and S. Schmitz (Eds.): CIAA 2005, LNCS 3845, pp. 343–344, 2006.

2 Results

The aim of the experiments is to compare RPNI2 with DeLeTe2. We have used the samples provided in `http://www.grappa.univlille3.fr/~lemay/`. These samples are generated from 20 state (on average) NFAs, which correspond to 120 state (on average) DFAs.

The table shown below reports the recognition rate and the average size (the number of states of the hypothesis). The error rate of the new algorithm RPNI2 is better than the former RPNI but slightly worse than DeLeTe2. The opposite happens with the description complexity (i.e. number of states) of the output hypothesis. The results obtained by RPNI2 are then better than those of DeLeTe2.

Iden.	RPNI		RPNI2		DeLeTe2	
	Recognition rate	Average size	Recognition rate	Average size	Recognition rate	Average size
er_50	76.36%	9.63	80.03%	16.32	81.68%	32.43
er_100	80.61%	14.16	88.68%	19.24	91.72%	30.73
er_150	84.46%	15.43	90.61%	26.16	92.29%	60.96
er_200	91.06%	13.3	93.38%	27.37	95.71%	47.73
nfa_50	64.8%	14.3	66.43%	30.64	69.80%	71.26
nfa_100	68.25%	21.83	72.79%	53.14	74.82%	149.13
nfa_150	71.21%	28.13	75.69%	71.87	77.14%	218.26
nfa_200	71.74%	33.43	77.25%	88.95	79.42%	271.3

3 Conclusions

Although the experiments are still preliminary, it seems that the slightly better results obtained by DeLeTe2 with respect to RPNI2 do not compensate the fact that the size of the representations obtained by RPNI2 are clearly smaller. A more exhaustive set of experiments should be done in future works.

References

1. Denis, F. Lemay, A., Terlutte, A.: Learning regular languages using rfsas. Theoretical Computer Science **313** (2004) 267–294
2. Oncina, J., García, P.: Inferring regular languages in polynomial updated time. In Pattern Recognition and Image Analysys (1992)