Supplementary Information

and Complexity of Finite Automata

(Dodatočná informácia a zložitosť konečných automatov)

na získanie vedecko-akademickej hodnosti philosophiae doctor
v odbore doktorandského štúdia: 9.2.1. informatika

Bratislava 2015
Dizertačná práca bola vypracovaná v internej forme doktorandského štúdia na Katedre informatiky Fakulty matematiky, fyziky a informatiky Univerzity Komenského v Bratislave.

**Predkladateľ:** Mgr. Ivan Kováč  
Katedra informatiky  
Fakulta matematiky, fyziky a informatiky  
Univerzity Komenského  
Mlynská dolina  
842 48 Bratislava

**Školiteľ:** Prof. RNDr. Branislav Rovan, PhD.  
Katedra informatiky FMFI UK  
Bratislava

**Oponenti:** .............................................  
.............................................  
.............................................  
.............................................  
.............................................

Obhajoba dizertačnej práce sa koná dňa ............... o ...........  
pred komisiou pre obhajobu dizertačnej práce v odbore doktorandského štúdia  
vymenovanou predsedom odborovej komisie dňa ...............  

v študijnom odbore 9.2.1. informatika  

na Fakulte matematiky, fyziky a informatiky UK, Mlynská dolina, ...............  

**Predseda odborovej komisie:**  
Prof. RNDr. Branislav Rovan, PhD.  
Fakulta matematiky, fyziky a informatiky  
Univerzity Komenského  
Mlynská dolina  
842 48 Bratislava
Introduction

The notion of information is essential in many areas of our lives. We use information almost continuously – we check weather forecast every day, absorb world’s news from various sources, communicate with our family, friends, and co-workers using digital technology. The Internet is literally full of information. But information is not just a string of zeros and ones. Information can be (un)reliable, useful, useless, misleading, redundant, old, etc.

Some aspects of information have been investigated more than others. In particular, Shannon’s concept of information as a string of bits [Sha48] provided a framework useful for studying communication of information through data channels. The notion of size of information was later refined by Solomonoff [Sol64], Kolmogorov [Kol65] and Chaitin [Cha69] using a framework of Kolmogorov complexity.

This early work provided fundamentals for studying basic aspects of information needed in communication systems. However, in recent years, decisions based on information are frequently made by machines. These decisions cannot be made solely on the aspects of information needed in communication, as there are other aspects that are important – for example, having data from a meteorological station from this morning might be useful for forecasting weather in the same location tomorrow, but data collected 10 years ago might not be so interesting for the same purpose.

One of the aspects of information that is worth studying is usefulness of information. Intuitively, information is useful when it reduces complexity of a problem. It is worth noting that usefulness of information is always tied to the problem at hand – information about traffic situation is not very useful when trying to forecast the weather, but it might be useful when computing optimal route from a point $A$ to a point $B$.

In recent research the notion of useful information has been emerging either explicitly – e.g., in [GR08], [Ste10], and [RZ14] – or implicitly, e.g., in study of advice complexity of online algorithms [DKP09, BKK+09], optimization of hard problems with knowledge of an optimal solution for similar instance [BHMW08], etc.

Goals of the Thesis

The thesis aims to continue the research of usefulness of information from two different points of view. The first part is devoted to the study of the computational approach to the usefulness of information for deterministic finite automata. Our approach slightly differs from the one used in [GR08, Ste10], as we consider information, which does not say anything about the currently processed word, but concerns similar word. The second part aims to investigate how information about meaning of states of finite automata can help us to provide better (tighter) bounds on their state complexity.
Results – Automata with Advice

In the first part of the thesis we present a model of automata with advice, where advice represents information about a word similar to the input word. Two notions of admissibility of the contents of the advice tape are defined, in particular \(L\)-constrained admissibility and arbitrary admissibility.

We investigate the families of languages that can be accepted by automata with advice with respect to the two kinds of admissibility, various advice schemes (representing chosen distance measure and advice content, called \(L\)-bit, \(H\)-bit, \(L\)-word, and \(H\)-word) and allowed distance \(k\) between similar words. For a given \(k\), advice scheme \(S\), the family of languages accepted by automata with advice is denoted by \(L_{S,k}\) and \(L_{S,k}\), for arbitrary admissibility and \(L\)-constrained admissibility respectively.

Among the results for the arbitrary admissibility, a characterization of languages belonging to the \(L_{L\text{-bit},k}\) and \(L_{H\text{-bit},k}\) is provided, based on the existence of two regular languages with certain properties. An existence of a regular language for which any advice scheme helps to reduce the number of states (in comparison with the minimal automata) is proved as well. Moreover, we show that \(L_{S,k}\) is the largest subfamily of \(L_{S,k}\) that is closed under complement.

In the case of \(L\)-constrained admissibility, the hierarchies of the families \(L_{S,k}\) are discussed. The respective hierarchies do not collapse, that is

\[
L_{S,0} \supseteq L_{S,1} \supseteq \cdots \supseteq L_{S,k} \supseteq \cdots \supseteq L_{S,\infty} \supseteq R.
\]

Moreover, we provide the relationships between the families on the same level \((k)\) as depicted in Figure 1.

\[
L_{L\text{-bit},k} \subset L_{H\text{-bit},k} \subset \bigcap_{l} L_{l\text{-word},k} \subset L_{L\text{-word},k}
\]

Figure 1: The relationship between \(L_{L\text{-bit},k}\), \(L_{H\text{-bit},k}\), \(L_{L\text{-word},k}\) and \(L_{H\text{-word},k}\), \(k > 0\).

The state complexity of automata with advice for regular languages is addressed in the thesis as well. Examples of language families for which the supplementary information in the form of the advice about similar words is useful are presented.

Results – Semantically Restricted Automata

The model investigated in the second part of the thesis is called reasonable automaton, defined in [HKKŠ13]. A reasonable automaton is a two-way finite automaton for which the states have assigned a meaning in the form of logical formulæ. We are interested in
the state complexity of reasonable automata for a family of problems called one-way liveness (1-LIV for short).

We study how useful it is to have information about the restriction of allowed meanings (logical formulæ). Our results are shown in Table 1 in bold. Other results presented in the table are from [HKKŠ13, Šte14].

<table>
<thead>
<tr>
<th>Predicates</th>
<th>Connectives</th>
<th>Problem</th>
<th>Lower bound</th>
<th>Upper bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>$e(a, b)$</td>
<td>full logic</td>
<td>1-LIV$_2$</td>
<td>$\Omega(n)$</td>
<td>$O(n)$, $O(n^2)$, $O(dn^{2+c+\log_2(dn^c)})$</td>
</tr>
<tr>
<td>$e(a, b)$</td>
<td>conjunction</td>
<td>1-LIV$_2$, 1-LIV$_3$, 1-LIV$<em>4$, 1-LIV$</em>{s(n)}$</td>
<td>$\Omega(2^n)$</td>
<td>$O(2^n)$</td>
</tr>
<tr>
<td>$p(a, b)$</td>
<td>conjunction</td>
<td>1-LIV$_2$, 1-LIV$_k$</td>
<td>$\Omega(2^n)$, $\Omega(2^{(k-1)n})$</td>
<td>$O(2^n)$, $O(k \cdot 2^{(k-1)n})$</td>
</tr>
<tr>
<td>$p(a, b, c)$</td>
<td>conjunction</td>
<td>1-LIV$_2$, 1-LIV$_3$, 1-LIV$<em>4$, 1-LIV$</em>{s(n)}$</td>
<td>$\Omega(n^2)$</td>
<td>$O(n^2)$, $O(n^3)$, $O(dn^{2+c+\log_2(dn^c)})$</td>
</tr>
<tr>
<td>$r(a)$</td>
<td>full logic</td>
<td>1-LIV$_2$</td>
<td>$\Omega(2^n)$</td>
<td>$O(2^n)$</td>
</tr>
<tr>
<td>$r(a)$</td>
<td>conjunction</td>
<td>1-LIV$_2$, 1-LIV$_3$, 1-LIV$_4$</td>
<td>$\Omega(2^n)$</td>
<td>$O(2^n)$</td>
</tr>
<tr>
<td>$r(a)$, $p(a, b, c)$</td>
<td>conjunction</td>
<td>1-LIV$_2$, 1-LIV$_3$, 1-LIV$_4$</td>
<td>$\Omega(n)$</td>
<td>$O(n)$, $O(n^2)$</td>
</tr>
</tbody>
</table>

Table 1: Known bounds for any constant $k$ and polynomial $s$, such that $s(n) \leq dn^c$ for some constants $c, d$. The ones in bold are contained in this thesis. In the case of formulæ constructed through conjunctions and variables $p(a, b)$, an upper bound of $O(n^22^n)$ for 1-LIV$_2$ was previously stated in [HKKŠ13].

**Bibliography**


List of papers of the author related with the thesis


Abstract

This thesis contributes to the research of usefulness of information in formal models of computation. The usefulness of information is investigated in two scenarios. In the first one, deterministic finite automata have access to supplementary information about words similar to the input word, and their computational power and state complexity is studied. The results are grouped according to several types of information provided and types of admissibility of similar words.

In the second scenario, usefulness of information about two-way deterministic finite automata, when proving lower and upper bounds on their state complexity, is studied. Specifically, the model of automata with explicit meanings of states is discussed, and new or improved lower and upper bounds on the state complexity of such automata are provided, when the meaning of states is restricted by supplementary information (e.g., when only existence of edges in some graph can be expressed by the states).

Keywords: supplementary information, usefulness, advice about similar words, finite automata, reasonable automata, formal languages