



Fakulta matematiky, fyziky a informatiky
Univerzity Komenského v Bratislave



Mgr. Ivan Kováč

Autoreferát dizertačnej práce

SUPPLEMENTARY INFORMATION AND COMPLEXITY OF FINITE AUTOMATA

(DODATOČNÁ INFORMÁCIA A ZLOŽITOSŤ KONEČNÝCH AUTOMATOV)

na získanie vedecko-akademickej hodnosti philosophiæ 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 Rován, 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 Rován, 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 \mathcal{S} , the family of languages accepted by automata with advice is denoted by $\mathcal{L}_{\mathcal{S},k}^{\vee}$ and $\mathcal{L}_{\mathcal{S},k}$, for arbitrary admissibility and *L-constrained admissibility* respectively.

Among the results for the arbitrary admissibility, a characterization of languages belonging to the $\mathcal{L}_{L-bit,k}^{\vee}$ and $\mathcal{L}_{H-bit,k}^{\vee}$ 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 $\mathcal{L}_{\mathcal{S},k}^{\vee}$ is the largest subfamily of $\mathcal{L}_{\mathcal{S},k}$ that is closed under complement.

In the case of *L-constrained admissibility*, the hierarchies of the families $\mathcal{L}_{\mathcal{S},k}$ are discussed. The respective hierarchies do not collapse, that is

$$\mathcal{L}_{\mathcal{S},0} \supsetneq \mathcal{L}_{\mathcal{S},1} \supsetneq \cdots \supsetneq \mathcal{L}_{\mathcal{S},k} \supsetneq \cdots \supsetneq \mathcal{L}_{\mathcal{S},\infty} \supsetneq \mathcal{R}.$$

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

$$\mathcal{L}_{L-bit,k} \begin{array}{c} \supsetneq \\ \supsetneq \end{array} \begin{array}{c} \mathcal{L}_{H-bit,k} \\ \mathcal{L}_{L-word,k} \end{array} \begin{array}{c} \supsetneq \\ \supsetneq \end{array} \mathcal{L}_{L-word,k}$$

Figure 1: The relationship between $\mathcal{L}_{L-bit,k}$, $\mathcal{L}_{H-bit,k}$, $\mathcal{L}_{L-word,k}$ and $\mathcal{L}_{H-word,k}$, $k > 0$.

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].

Predicates	Connectives	Problem	Lower bound	Upper bound
$e(a, b)$	full logic	1-LIV ₂ 1-LIV ₃ , 1-LIV ₄ 1-LIV _{s(n)}	$\Omega(\mathbf{n})$	$O(n)$ $O(n^2)$ $O(dn^{2+c+\log_2(dn^c)})$
$e(a, b)$	conjunction	1-LIV ₂ 1-LIV _k	$\Omega(2^n)$ $\Omega(2^{(k-1)n})$	$O(2^n)$ $O(k \cdot 2^{(k-1)n})$
$p(a, b)$	conjunction	1-LIV ₂ 1-LIV _k	$\Omega(2^n)$	$O(2^n)$ $O(k \cdot n \cdot 2^n)$
$p(a, b, c)$	conjunction	1-LIV ₂ , 1-LIV ₃ 1-LIV ₄ 1-LIV _{s(n)}	$\Omega(n^2)$	$O(n^2)$ $O(n^3)$ $O(dn^{2+c+\log_2(dn^c)})$
$r(a)$	full logic	1-LIV ₂	$\Omega(2^{\frac{n}{2}})$	$O(2^{\frac{n}{2}})$
$r(a)$	conjunction	1-LIV ₂ , 1-LIV ₃ , 1-LIV ₄	$\Omega(2^n)$	$O(2^n)$
$r(a),$ $p(a, b, c)$	conjunction	1-LIV ₂ 1-LIV ₃ , 1-LIV ₄	$\Omega(n)$	$O(n)$ $O(n^2)$

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 $O(n^2 2^n)$ for 1-LIV₂ was previously stated in [HKKŠ13].

Bibliography

- [BHK15] Maria Paola Bianchi, Juraj Hromkovič, and Ivan Kováč. On the size of two-way reasonable automata for the liveness problem. In *Proceedings of the 19th International Conference Developments in Language Theory, DLT 2015*, to appear in 2015.
- [BHMW08] Hans-Joachim Böckenhauer, Juraj Hromkovič, Tobias Mömke, and Peter Widmayer. On the Hardness of Reoptimization. In Geffert et al. [GKB⁺08], pages 50–65.
- [BKK⁺09] Hans-Joachim Böckenhauer, Dennis Komm, Rastislav Královič, Richard Královič, and Tobias Mömke. On the advice complexity of online problems. In Yingfei Dong, Ding-Zhu Du, and Oscar H. Ibarra, editors, *Algorithms and Computation, 20th International Symposium, ISAAC 2009, Honolulu, Hawaii, USA, December 16-18, 2009. Proceedings*, volume 5878 of *Lecture Notes in Computer Science*, pages 331–340. Springer, 2009.

- [BL77] P. Berman and A. Lingas. On complexity of regular languages in terms of finite automata. Technical report, Institute of Computer Science, Polish Academy of Sciences, Warsaw, 1977.
- [Cha69] Gregory J. Chaitin. On the length of programs for computing finite binary sequences: statistical considerations. *J. ACM*, 16(1):145–159, 1969.
- [DKP09] Stefan Dobrev, Rastislav Kráľovič, and Dana Pardubská. Measuring the problem-relevant information in input. *ITA*, 43(3):585–613, 2009.
- [ESY84] Shimon Even, Alan L. Selman, and Yacov Yacobi. The complexity of promise problems with applications to public-key cryptography. *Information and Control*, 61(2):159 – 173, 1984.
- [GKB⁺08] Viliam Geffert, Juhani Karhumäki, Alberto Bertoni, Bart Preneel, Pavol Návrat, and Mária Bieliková, editors. *SOFSEM 2008: Theory and Practice of Computer Science, 34th Conference on Current Trends in Theory and Practice of Computer Science, Nový Smokovec, Slovakia, January 19-25, 2008, Proceedings*, volume 4910 of *Lecture Notes in Computer Science*. Springer, 2008.
- [Gol06] Oded Goldreich. On promise problems: A survey. In Oded Goldreich, Arnold L. Rosenberg, and Alan L. Selman, editors, *Essays in Memory of Shimon Even*, volume 3895 of *Lecture Notes in Computer Science*, pages 254–290. Springer, 2006.
- [GR08] Peter Gaži and Branislav Rován. Assisted Problem Solving and Decompositions of Finite Automata. In Geffert et al. [GKB⁺08], pages 292–303.
- [HKKŠ13] Juraj Hromkovič, Rastislav Kráľovič, Richard Kráľovič, and Richard Štefanec. Determinism vs. nondeterminism for two-way automata: Representing the meaning of states by logical formulæ. *Int. J. Found. Comput. Sci.*, 24(7):955–978, 2013.
- [HMU01] John E. Hopcroft, Rajeev Motwani, and Jeffrey D. Ullman. *Introduction to automata theory, languages, and computation*. Addison-Wesley Longman Publishing, second edition, 2001.
- [HS66] J. Hartmanis and R.E. Stearns. *Algebraic structure theory of sequential machines*. Prentice-Hall international series in applied mathematics. Prentice-Hall, 1966.
- [HU69] John E. Hopcroft and Jeffrey D. Ullman. *Formal languages and their relation to Automata*. Addison-Wesley Longman Publishing, 1969.

- [Kol65] A. N. Kolmogorov. Three approaches to the quantitative definition of information. *Problems of Information Transmission*, 1(1):1–7, 1965.
- [Kol74] A.N. Kolodin. Two-way nondeterministic automata. *Cybernetics*, 10(5):778–785, 1974.
- [Krá10] Richard Královič. *Complexity Classes of Finite Automata*. PhD thesis, ETH Zürich, Switzerland, 2010.
- [LR11] Pavel Labath and Branislav Rován. Simplifying DPDA Using Supplementary Information. In Adrian Horia Dediu, Shunsuke Inenaga, and Carlos Martín-Vide, editors, *LATA*, volume 6638 of *Lecture Notes in Computer Science*, pages 342–353. Springer, 2011.
- [Moo71] F.R. Moore. On the bounds for state-set size in the proofs of equivalence between deterministic, nondeterministic, and two-way finite automata. *Computers, IEEE Transactions on*, C-20(10):1211–1214, Oct 1971.
- [RS59] M. O. Rabin and D. Scott. Finite automata and their decision problems. *IBM J. Res. Dev.*, 3(2):114–125, 1959.
- [RZ14] Branislav Rován and Marek Zeman. Modeling time criticality of information. *Inf. Process. Lett.*, 114(3):147–151, 2014.
- [Sha48] C. E. Shannon. A mathematical theory of communication. *The Bell Systems Technical Journal*, 27:379–423, 623–656, 1948.
- [Sol64] R.J. Solomonoff. A formal theory of inductive inference. Part I. *Information and Control*, 7(1):1 – 22, 1964.
- [SS78] William J. Sakoda and Michael Sipser. Nondeterminism and the size of two way finite automata. In *Proceedings of the Tenth Annual ACM Symposium on Theory of Computing*, STOC '78, pages 275–286, New York, NY, USA, 1978. ACM.
- [Ste10] Ľuboš Steskal. *On Usefulness of Information: A Computational Approach*. PhD thesis, Comenius University, Bratislava, Slovakia, 2010.
- [Šte14] Richard Štefanec. *Semantically restricted two-way automata*. PhD thesis, Comenius University, Bratislava, Slovakia, 2014.

List of papers of the author related with the thesis

- [BHK15] Maria Paola Bianchi, Juraj Hromkovič, and Ivan Kováč. On the size of two-way reasonable automata for the liveness problem. In *Proceedings of the 19th International Conference Developments in Language Theory, DLT 2015*, to appear, 2015.

Abstrakt

Táto práca prispieva k výskumu užitočnosti informácie pre formálne výpočtové modely. Užitočnosť informácie skúmame v dvoch variantoch. V prvom skúmame výpočtovú silu a stavovú zložitosť deterministických konečných automatov, ktoré majú prístup k dodatočnej informácii o podobnom slove. Výsledky sú rozdelené podľa niekoľkých typov informácie a prípustnosti podobných slov.

V druhom variante skúmame užitočnosť informácie o dvojsmerných deterministických konečných automatoch, pri dokazovaní dolných a horných hraníc ich stavovej zložitosti. Presnejšie, hovoríme o modeli automatov s explicitne daným významom stavov. Dosiahnuté výsledky sú nové alebo vylepšujú známe výsledky o stavovej zložitosti takýchto automatov, keď je význam stavov obmedzený dodatočnou informáciou (napríklad, ak stavy môžu vyjadrovať len existenciu hrán v nejakom grafe).

Keywords: informácia, užitočnosť, rada o podobných slovách, konečné automaty, rozumné automaty, formálne jazyky

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