Operational Calculus of Modified Subset Construction

Mirzakhmet Syzdykov[1]
[1]Satbayev University, Almaty, Kazakhstan
[1]mspmail598@gmail.com
[1]ORCID ID: https://orcid.org/0000-0002-8086-775X

Abstract. We present the continuation of studying Extended Regular Expression (ERE) on the view of modified subset construction within the overridden operators like intersection, subtraction, and re-written complement. As before we have stated that in this case the complexity has a decreasing nature and tendency. We will give the strict definition of the operational part of this modified subset construction which is due to Rabin and Scott. The complexity of algorithm remains a magnitude less than NP-hard problems for which we have given the strict proof of equivalence in the prior work, so this work continues the studying of the comparable proof for a variety of problems to be computationally complex, however, explainable in terms of unified approach like operational calculus. In this calculus the general points of research are given to the representation of modified subset construction with at least two operands which are to be computed by subset construction and in terms of complexity of the effective algorithm they are computed using modified subset construction.

Keywords: subset construction, extended regular expressions, modification, operations, calculus.

Introduction. The subset or Rabin-Scott construction which was full described in [1] represents conservative system of choosing between determinism and non-determinism in both aspects, however, lacks the efficiency of complexity in case of deterministic machine operating on the finite set of states, thus, it's obvious that it will lead the number of states as well as number of operations to grow exponentially in time of $O(2^n)$.

The latter case isn't limited to the usage of the classical Thompson algorithm [2], which is less complex and requires asymptotic explosion of complexity in $O(m \cdot n)$, where m is the number of symbols in sought or input string and n is the number of elements in matched regular expression. To the present time Thompson's constructions weren't used for extended regular expression matching.

Samuel C. Hsieh showed a more effective algorithm for intersection operator [3], however, it's still NP-hard as its complexity can be denoted by $O(n^t)$, where n is the average size of length of operands in ERE and t is the number of &-operators.

We have shown that ERE for intersection problem can be computed on both deterministic and non-deterministic finite automata, NFA and DFA respectively [4]. We have also introduced the sliced model of computation for our algorithm which tends to be in magnitude faster by applying operational calculus [5].

Our algorithm for NFA or DFA supersedes previous results [6, 7] which operate on the cross-product construction of the DFA by applying the operational calculus in the form of the operational logic for the set of operands to be performed in-time and in-memory. The non-emptiness intersection problem was shown to be NP-hard for sparse set of automata [7], however, we give another argument towards our conjecture of equivalence of complexity classes.

Aho-Corasick trees [8] and Lempel-Ziv-Welch [9] streams for regular expression matching are also discussed in this article as well as the "P versus NP" conjecture [10] for common case of the problem of deciding whether the intersection of the given languages is empty or not.

References

1. Rabin M. O., Scott D. Finite automata and their decision problems //*IBM journal of research and development*. – 1959. – Т. 3. – №. 2. – С. 114-125.
2. Thompson K. Programming techniques: Regular expression search algorithm //*Communications of the ACM*. – 1968. – Т. 11. – №. 6. – С. 419-422.
3. Hsieh S. C. Product construction of finite-state machines //*Proc. of the World Congress on Engineering and Computer Science*. – 2010. – С. 141-143.
4. Syzdykov M. Deterministic automata for extended regular expressions //*Open Computer Science*. – 2017. – Т. 7. – №. 1. – С. 24-28.

5. Syzdykov M. et al. INTRODUCING PQI-OPERATOR IN THEORY OF COMPUTATIONAL COMPLEXITY //*ADVANCED TECHNOLOGIES AND COMPUTER SCIENCE*. – 2022. – №. 3. – С. 4-9.
6. de Oliveira Oliveira M., Wehar M. On the fine grained complexity of finite automata non-emptiness of intersection //*Developments in Language Theory: 24th International Conference*, DLT 2020, Tampa, FL, USA, May 11–15, 2020, Proceedings. – Cham : Springer International Publishing, 2020. – С. 69-82.
7. Fernau H., Hoffmann S., Wehar M. Finite automata intersection non-emptiness: Parameterized complexity revisited //*arXiv preprint* arXiv:2108.05244. – 2021.
8. Zha X., Sahni S. Highly compressed Aho-Corasick automata for efficient intrusion detection //*2008 IEEE Symposium on Computers and Communications*. – IEEE, 2008. – С. 298-303.
9. Bille P., Fagerberg R., Gørtz I. L. Improved approximate string matching and regular expression matching on Ziv-Lempel compressed texts //*ACM Transactions on Algorithms (TALG)*. – 2009. – Т. 6. – №. 1. – С. 1-14.
10. Cook S. The importance of the P versus NP question //*Journal of the ACM (JACM)*. – 2003. – Т. 50. – №. 1. – С. 27-29.