



Univerzita Komenského v Bratislave
Fakulta matematiky, fyziky a informatiky
Katedra informatiky

Modeling Time Criticality of Information

mgr. Marek Zeman

na získanie akademického titulu *philosophiae doctor*
v odbore doktorandského štúdia: 9.2.1 Informatika

Bratislava, 2014

Úvod

Témou predkladanej práce je pojem informácia. Dôležitým medzníkom v skúmaní tohto pojmu je štúdia Clauda Shannona, kde skúmal množstvo informácie generovanej náhodným zdrojom s motiváciou kvantifikovať objem prostriedkov potrebných na jej prenos. Medzi neskoršie teórie patrí napríklad Kolmogorovská zložitosť, v ktorej sa kvantifikuje množstvo informácie v konečných objektoch.

Šírenie informácií sa vďaka internetu v nedávnej minulosti výrazne zintenzívnilo. Keďže množstvo dostupných údajov je obrovské, často používame automatické systémy, ktorú informácie spracovávajú, robia na jej základe rozhodnutia a prípadne produkujú nové informácie. Preto sa aj náš pohľad na informáciu rozširuje a vzniká potreba sústrediť sa na jej užitočnosť, použiteľnosť, validitu, aktuálnosť a iné aspekty. Tieto vlastnosti informácie musíme skúmať s prihliadaním na výpočtový kontext ich použitia.

Ako reakcia na tento vývoj sa v informatike vykonáva výskum, ktorý aspekty informácie skúma z abstraktnejšieho hľadiska. Formálne aspekty tohto pojmu sú témou prác [GR08] [LR11], v ktorých sa autori sústreďujú na užitočnosť prídavnej informácie vo formálnych výpočtových modeloch. Pokusom o abstraktný a ucelený pohľad na užitočnosť informácie je práca [Ste10]. V spomenutom výskume sa aplikuje výpočtový prístup: informácia je užitočná, ak ju vieme využiť na zníženie zložitosti problému, ktorý potrebujeme riešiť.

Predkladaná dizertačná práca je pokračovaním týchto výskumných aktivít. Jej témou je aspekt aktuálnosti informácie. Využitelnosť konkrétnej informácie môže byť rôzna v závislosti od momentu, kedy ju získame. Aktuálnosť teda interpretujeme ako užitočnosť v priebehu nejakého procesu. Našou úlohou je skúmať vývoj užitočnosti počas tohto procesu. Budeme sa pridŕžať výpočtového prístupu a merať množstvo výpočtových prostriedkov, ktoré vieme po získaní informácie ušetriť jej využitím.

Ciele

Cieľom práce je vytvoriť formálny model aktuálnosti prídavnej informácie a navrhnúť metódu na kvantifikovanie jej užitočnosti v priebehu výpočtu. Jedným z možných postupov je začať analýzou tohto pojmu v širokom kontexte a následne vytvoriť abstraktný formálny model. Ten by sa potom bolo možné aplikovať do rôznych konkrétnejších scenárov, kde by sme skúmali ďalšie vlastnosti.

Keďže pojem aktuálnosti doposiaľ nebol skúmaný, lepším sa nám javil prístup, v ktorom začneme modelovaním tohto pojmu v jednom alebo niekoľkých zjednodušených výpočtových scenároch. Po obdržaní základného prehľadu o vlastnostiach aktuálnosti informácie bude možné spoločné črty zovšeobecniť. Ako vhodný jednoduchý výpočtový model sme zvolili deterministické konečné automaty.

Cieľom práce je navrhnúť vhodný model pojmu aktuálnosti prídavnej informácie obdržanej konečným automatom. Po obdržaní prídavnej informácie je možné v spracovávaní vstupu pokračovať v inom konečnom automate. Našou úlohou je definovať spôsob merania aktuálnosti a skúmať vlastnosti tohto pojmu. Jedným z cieľov je analyzovať možné priebehy aktuálnosti informácie v priebehu konkrétneho

výpočtu.

Na základe neformálnej analýzy navrhujeme viacero variant využitia prídavnej informácie. V práci budeme tieto varianty porovnávať a analyzovať ich vlastnosti. Sústreďíme sa na skúmanie formálneho základu pojmu informácia. Našou úlohou nie je skúmať konkrétne výpočtové problémy.

Výsledky

Na úvod našej práce sme spravili neformálnu analýzu pojmu aktuálnosť, pri ktorej sme využili existujúce výsledky o pojme užitočnosti informácie. Ako výpočtový model sme zvolili jednosmerné deterministické konečné automaty. Zložitosť výpočtového problému je počet stavov najmenšieho automatu, ktorý ho rieši.

Automat A má za úlohu zistiť, či $w \in L(A)$. Pomocná informácia prichádza vo forme faktu, že spracovávané slovo w patrí do regulárneho jazyka L_{adv} . V momente získania tejto informácie môžeme vo výpočte na slove w pokračovať v novom automate A' , ktorý môže využiť pomocnú informáciu. Vyžadujeme od neho, aby korektné dokončil riešenie problému $w \in L(A)$, avšak len pre vstupy, ktoré patria do L_{adv} . Toto zľavenie z požiadaviek môže spôsobiť, že automat A' je jednoduchší ako automat A . Pri vyhodnocovaní aktuálnosti pomocnej informácie uvažujeme najjednoduchší možný A' pre daný moment.

Na celú situáciu sa môžeme pozeráť tak, že niekto nám paralelne pripravuje informáciu $w \in L_{adv}$ a v nejakom momente nám ju poskytne. Naším zámerom bolo skúmať, ako sa mení užitočnosť tejto informácie v rôznych momentoch výpočtu. V našom modeli uvažujeme aj možnosť, že sa počas prípravy zistí fakt $w \notin L$. Skúmame využitie oboch prípadov a pri vyhodnocovaní užitočnosti prídavnej informácie v danom momente používame horší z týchto prípadov. Aktuálnosť je v našom škálovaní hodnota medzi 0 (vrátane) a 1.

Existuje viacero možností definície prídavnej informácie. V prípade vyššie v každom momente môžeme získať rovnakú pomocnú informáciu – fakt, či $w \in L_{adv}$. Tento variant preto nazývame *statická* prídavná informácia. Ďalším typom pomocnej informácie je, či doposiaľ nespracovaná časť vstupného slova patrí do pomocného regulárneho jazyka L_{adv} . Keďže táto informácia sa môže v priebehu výpočtu meniť, voláme ju *dynamická*. V práci uvádzame porovnanie týchto dvoch variantov. V niektorých aspektoch sa zhodujú, v iných sa líšia.

Vývoj aktuálnosti v priebehu výpočtu môže byť komplikovaný. Jedným z hlavných výsledkov práce je, že aj pre konkrétnu jednoduchú pomocnú informáciu môže graf aktuálnosti pri riešení určitého problému obsahovať ľubovoľne veľa lokálnych maxím, ktoré môžu dosahovať (skoro) ľubovoľné hodnoty. Tento výsledok je platný pre statickú aj pre dynamickú pomocnú informáciu.

V ďalšej časti práce sme definovali meranie aktuálnosti pomocnej informácie, v ktorom sme zahrnuli aj jej zložitosť. Aj keď môže niekedy prídavná informácia pomôcť, jej príprava môže byť pridrahá a nemusí sa oplatiť. Tento variant nazývame *efektívna aktuálnosť*. V navrhnutom spôsobe merania môžu byť hodnoty aktuálnosti racionálne čísla menšie ako 1. Podobne ako v predchádzajúcich možnostiach,

aj v tomto prípade môže byť vývoj aktuálnosti komplikovaný. Ak sa ale stane, že aktuálnosť klesne pod 0 a pomoc je pridrahá, potom zostane pridrahá až do konca výpočtu.

Literatúra

- [Ant02] Luís Filipe Coelho Antunes. *Useful Information*. PhD thesis, Faculdade de Ciências da Universidade do Porto, 2002.
- [ATW+08] P. Adriaans, P. Thagard, J. Woods, , and D.M. Gabbay, editors. *Philosophy of information*. North-Holland, 2008.
- [Cal10] C.S. Calude. *Information and Randomness: An Algorithmic Perspective*. Texts in Theoretical Computer Science. an EATCS Series. Springer, 2010.
- [ESY84] Even, Selman, and Yacobi. The complexity of promise problems with applications to public-key cryptography. *Information and Control*, 61 no. 2:159—173, 1984.
- [GR08] Peter Gaži and Branislav Rován. Assisted problem solving and decompositions of finite automata. In *Proceedings of the 34th conference on Current trends in theory and practice of computer science, SOFSEM'08*, pages 292–303, Berlin, Heidelberg, 2008. Springer-Verlag.
- [HMU01] John E. Hopcroft, Rajeev Motwani, and Jeffrey D. Ullman. *Introduction to automata theory, languages, and computation - (2. ed.)*. Addison-Wesley series in computer science. Addison-Wesley-Longman, 2001.
- [Huf52] David Albert Huffman. A method for the construction of minimum-redundancy codes. In *Proceedings of the IRE 40, no. 9*, pages 1098–1101, 1952.
- [Huf54] David Albert Huffman. The synthesis of sequential switching circuits. Technical Report 274, Massachusetts institute of technology, Research laboratory of electronics, 1954.
- [Kap06] Christos A. Kapoutsis. *Algorithms and lower bounds in finite automata size complexity*. PhD thesis, Cambridge, MA, USA, 2006. Advisor – Michael Sipser.
- [Kol65] Andrey N. Kolmogorov. Three approaches to the quantitative definition of information. *Problems in Information Transmission*, 1965.
- [Kra10] Richard Kralovič. *Complexity Classes of Finite Automata*. PhD thesis, ETH Zurich, 2010.
- [LR11] Pavel Labath and Branislav Rován. Simplifying deterministic pushdown automata using supplementary information. In *Proceedings of the 5th international conference on Language and automata theory and applications, LATA'11*, pages 342–353, Berlin, Heidelberg, 2011. Springer-Verlag.
- [LV08] Ming Li and Paul M.B. Vitányi. *An introduction to Kolmogorov complexity and its applications*. Springer Publishing Company, Incorporated, 2008.

- [Mea55] G. M. Mealy. A method for synthesizing sequential circuits. *Bell System Technical Journal*, 34(5):1045–1079, 1955.
- [MF71] A. R. Meyer and M. J. Fischer. Economy of description by automata, grammars, and formal systems. In *Proceedings of the 12th Annual Symposium on Switching and Automata Theory (SWAT)*, pages 188–191, Washington, DC, USA, 1971. IEEE Computer Society.
- [Moo56] Edward F. Moore. Gedanken-experiments on sequential machines. *Automata Studies*, 34:129–153, 1956.
- [RS59] Michael O. Rabin and Dana Scott. Finite automata and their decision problems. *IBM Journal of Research and Development*, 3, 1959.
- [RZ14] Branislav Rován and Marek Zeman. Modeling time criticality of information. *Information Processing Letters*, 114(3):147 – 151, 2014.
- [Sha48] Claude Shannon. A mathematical theory of communication. *Bell Systems Tech. Journal*, 27:379–423, 623–656, 1948.
- [Sol64] Ray J. Solomonoff. A formal theory of inductive inference, part 1 and part 2. *Information and Control*, 7, 1964.
- [SS78] William J. Sakoda and Michael Sipser. Nondeterminism and the size of two way finite automata. In *Proceedings of the 10th Annual ACM Symposium on Theory of Computing (STOC)*, pages 275–286, New York, 1978. ACM Press.
- [Ste10] Luboš Steškal. *On Usefulness of Information: A Computational Approach*. PhD thesis, Comenius University, Bratislava, 2010.
- [Top09] Flemming Topsøe. Game theoretical optimization inspired by information theory. *J. of Global Optimization*, 43:553–564, April 2009.

Abstract

In this thesis, we continue the ongoing research on formal treatment of attributes of information realized at the department by studying the aspect of time criticality of information. Same piece of information can be more or less useful depending on the time it is received. We base our studies on the research concerning usefulness of advisory information measured by the decrease in complexity of a computational problem we need to solve.

We use finite automata setting, observing the usefulness of advisory information when received at various points during the computation on a given input word. We define the basic notions and study possible behaviour of time criticality.

We analyze two types of advice. In the “static” case, the supplementary information concerns entire input instance while in the “dynamic” case, the supplementary information concerns only the unprocessed part of the input word. We also analyzed the case where the complexity of the advice is considered and model a situation where the complexity of the advice is too expensive.

Despite the simplicity of the computational model used, we have shown that the time criticality may exhibit interesting properties.

Vlastné publikácie autora

- Branislav Rován and Marek Zeman. Modeling time criticality of information. *Information Processing Letters*, 114(3):147 – 151, 2014.