prof. Damian Niwiński
Instytut Informatyki UW

Recenzja rozprawy doktorskiej pana Tomasza Steifera pt. *Computable Prediction of Infinite Binary Sequences with Zero-One Loss*

Jednym z najciekawszych problemów podjętych przez matematykę XX wieku jest próba wyjaśnienia zjawiska losowości. Andriej N. Kołmogorow, którego książka z 1933 r. dała aksjomatyczne podstawy nowoczesnemu rachunkowi prawdopodobieństwa, w latach 1960-tych zaproponował nowe podejście do pojęcia losowości, którego rdzeniem jest proces algorytmiczny. Jego idee (a także innych wybitnych matematyków, m.in. Claude'a Shannona, Alonzo Churcha, Pera Martina-Löfa i in.) dały asumpt do rozwoju fascynującej dziedziny badań, w której matematyczną interpretację zyskują takie intuicyjne opozycje jak losowość *vs.* struktura, złożoność *vs.* prostota, chaotyczność *vs.* przewidywalność. Dziedzina ta obejmuje wiele nurtów rozwijanych współcześnie przez różne społeczności naukowe (m.in. logików matematycznych, probabilistów, teoretyków informatyki), przy czym szczególnie cenne wydaje się przerzucanie mostów i łączenie potencjału różnych teorii.

Rozprawa Tomasza Steifera ambitnie wchodzi w ten nurt badań, który do tej pory (według mojego rozeznania) był w Polsce mało obecny. Tematem pracy jest zagadnienie algorytmicznej (nie)przewidywalności kolejnych bitów ciągu losowego i w szczególności związki tej własności ze złożonością badanego ciągu w sensie hierarchii arytmetycznej. Jak stwierdzają w swojej monografii Downey i Hirschfeldt (ref. [12] w pracy), spośród trzech paradygmatów określających algorytmiczną losowość nieprzewidywalność jest być może tym najbardziej intuicyjnym (obok paradygmatu złożonościowego i teorio-miarowego). Rozprawa pana Steifera wnosi do tego zagadnienia wiele ciekawych wyników – czasem nieoczekiwanych, a także otwiera ścieżki do dalszych badań.

Wyniki rozprawy

Praca poprzedzona jest wstępem i obejmuje trzy części. Część pierwsza (*Preliminaries*) przedstawia podstawowe pojęcia używane w pracy zaczerpnięte z teorii obliczeń, arytmetyki i teorii prawdopodobieństwa. Wprowadzone jest też kluczowe dla rozprawy pojęcie predyktora.

Część druga (*Prediction of individual sequences*) licząca 30 stron stanowi głowny trzon pracy i przynosi najważniejsze wyniki. Ogólnie interesujemy się przewidywaniem kolejnych bitów nieskończonego ciągu x przy pomocy funkcji obliczalnych operujących na słowach skończonych, a więc w szczególności na prefiksach ciągu x.

W podrozdziale 4.1 autor wprowadza i porównuje ze sobą trzy pojęcia nieprzywidywalności określanej tu jako stochastyczność (*stochasticity*). Począwszy od najogólniejszego, jest to stochastyczność w sensie Ker-I Ko, w sensie Churcha i w sensie Martina-Löfa. Autor przytacza znane wyniki, m.in. konstrukcję Ville'a ciągu stochastycznego w sensie Churcha, przy pomocy której rozróżnia dwa pierwsze pojęcia. W podrozdziale 4.2 różne warianty stochastyczności są scharakteryzowane w języku martyngałów. Podrozdział 4.3 przynosi pierwszy istotny oryginalny wynik autora, a zarazem jest to przerzucenie mostu między zagadnieniem stochastyczności a teorią obliczeń. Autor stawia pytanie, jak „proste", czy inaczej jak bardzo konstruktywne – w sensie stopni Turinga i hierarchii arytmetycznej – mogą być ciągi stochastyczne w sensie powyższych definicji. (Złożoność ciągu bitów rozumiemy tu jako złożoność zbioru, którego jest on funkcją charakterystyczną.) Ze znanych faktów wynika, że złożoność ciągu stochastycznego w najsilniejszym sensie Martina-Löfa może sięgnąć (w dół) klasy Δ^0_2, ale nie Σ^0_1. Nie wyklucza to jednak istnienia ciągu o złożoności Σ^0_1 stochastycznego w (słabszym) sensie Ko. Autor obala to przypuszczenie wykazując, że żaden ciąg stochastyczny w sensie Ko nie może leżeć w klasie Σ^0_1 (*Theorem 92*), ani nawet na żadnym skończonym poziomie hierarchii Jerszowa (*Theorem 100*). Dowód polegający na konstrukcji odpowiedniego predyktora jest bardzo pomysłowy.

W rozdziale 5 (*Unstable prediction*) autor podejmuje pytanie o złożoność arytmetyczną ciągów A, które nie są stochastyczne, ale odznaczają się nieregularnością innego rodzaju: dla żadnego predyktora

f średnia liczba błędów nie jest określona (ponieważ $\liminf_{n\to\infty} \zeta(f, A_1^n) < \limsup_{n\to\infty} \zeta(f, A_1^n)$). Ciągi takie określane są jako *not error stable*; powiedzmy: o niestabilnym błędzie. Na wstępie autor wskazuje własność teorio-rekurencyjną (*weak genericity*), która implikuje niestabilność błędu. Główny wynik rozdziału (*Theorem 106*) pokazuje, że ciągi o niestabilnym błędzie można znaleźć już w klasie Σ_1^0. Jest to być może najtrudniejsze i najbardziej oryginalne twierdzenie pracy. Wynik otrzymany jest w dwóch krokach: najpierw autor wykazuje istnienie odpowiedniego ciągu o ograniczonym stopniu Turinga (*Theorem 105*), co już jest nietrywialne, a następnie „poprawia" konstrukcję do efektywnej (*Theorem 106*) używając w bardzo subtelny sposób techniki priorytetów znanej z dowodu twierdzenia Friedberga-Mucznika.

W kolejnym rozdziale 6 (*Optimaility*) autor zajmuje się porównywaniem predyktorów ze względu na wartość błędu $\zeta(f, A)$. Oczywiście z uwagi na symetrię możemy założyć, że ta wartość leży w przedziale $[0, \frac{1}{2}]$. Autor wykazuje, że dowolna liczba rzeczywista w tym przedziale może być wartością $\zeta(f, A)$ dla *optymalnego* predyktora pewnego ciągu A. Z drugiej strony optymalny predyktor może dla pewnego ciągu A nie istnieć, nawet jeśli istnieje „optymalna wartość" r, która może być dowolnie blisko przybliżana przez $\zeta(f, A) > r$.

Część drugą rozprawy zamyka rozdział 7, w którym autor omawia pewne alternatywne podejścia do zagadnienia przewidywalności i porównuje je z rezultatami swojej pracy. Pojęcie przewidywalności zaproponowane przez Kohtaro Tadakiego generuje klasę ciągów, która leży ściśle pomiędzy klasą ciągów stochastycznych w sensie Churcha a analogiczną klasą w sensie Ko; ścisłość tych zawierań autor wykazuje konstruując dwa nietrywialne kontrprzykłady. Autor konstruuje też przykład wykazujący rozbieżność omawianych pojęć w innym aspekcie: ciąg o niestabilnym błędzie, który w sensie Tadakiego jest przewidywalny z błędem 0. Z kolei autor rozważa pojęcie „zgrubnej obliczalności" (*coarse computability*), jakie zaproponowali Carl G. Jockusch Jr. i Paul E. Schupp. W sensie tej definicji ciąg obliczany jest dla – może nie wszystkich, ale – dla stosunkowo gęsto rozłożonych bitów, co w naturalny sposób prowadzi do konstrukcji predyktora. Autor wykazuje jednak, znów przez pomysłowy kontrprzykład, że taki predyktor może nie być optymalny nawet jeśli metoda zgrubnego obliczania była optymalna.

Omawiając wyniki tej części doktoratu warto podkreślić erudycję i zmysł syntezy autora. Pojęcie stochastyczności w sensie Ko nie było dotąd (wg mojej wiedzy) badane z punktu widzenia złożoności arytmetycznej; we wspomnianej wyżej monografii Downeya i Hirschfeldta praca Ko (skądinąd wysoko cytowana) nie jest nawet wspomniana, podobnie zresztą jak podjeście Tadakiego i zgrubna obliczalność. Tymczasem wszystkie te pojęcia zestawione razem stanowią przekonujący zrąb matematycznej teorii przewidywalności.

Trzecia część rozprawy (*Prediction in probabilistic framework*) przerzuca kolejny most: między algorytmiczną teorią losowości a teorią prawdopodobieństwa. W rozdziale 8 (*Randomness and nonuniform measure*) autor poszukuje efektywnych wariantów znanych twierdzeń probabilistycznych, co jest rozumiane jako zastąpienie kwantyfikatora „dla prawie wszystkich ciągów" przez „dla wszystkich ciągów losowych" (w określonym sensie). Z uwagi na postać wyjściową rozważanych twierdzeń, wcześniejsze definicje losowości zostają rozszerzone na ciągi bitów nieskończone w obie strony rozumiane tu jako funkcje $\mathbb{Z} \to \{0, 1\}$[1]. Autor przytacza znane wyniki; na przykład znana równoważność testów Martina-Löfa i testów Sołowieja (*Solovay tests*) może być interpretowana jako efektywna wersja Lematu Borela-Cantellego. Z kolei autor przedstawia efektywną wersję twierdzenia Dooba o zbieżności martyngałów i twierdzenia ergodycznego Breimana. Rezultaty te są niewątpliwie ciekawe, choć dla czytelnika przychodzącego z obszaru logiki i teorii obliczeń (jak obecny recenzent) przydałoby się jakieś wyjaśnienie motywacji dla takich efektywnych wariantów twierdzeń i umieściło je w szerszym kontekście. W szczególności nasuwa się pytanie (nieco w duchu Martina-Löfa): czy „typowe" twierdzenie z probabilistyki da się zawsze uczynić efektywnym?

Rozdział 9 (*Lower bounds for zero-one loss*) przedstawia dolne ograniczenie na błąd predykcji wynikające z teorii miary prowadzące do optymalnej strategii, która jednak na ogół nie jest obliczalna. W kolejnym rodziale 10 (*Universal prediction*) autor pokazuje, że obliczalna strategia optymalna istnieje dla stacjonarnych procesów ergodycznych; dokładniej, strategia ta jest optymalna dla każdej miary i ciągów losowych w sensie tej miary. Dowód tego twierdzenia wykorzystuje zaawansowane fakty z teorii procesów ergodycznych, ale też oryginalne pomysły autora.

Ostatni rozdział 11 (*Optimality*) przedstawia probabilistyczną wersję wyniku z rozdziału 6 (pod tym samym tytułem) z części drugiej, a mianowicie istnienie ciągu, dla którego optymalną wartość (równą 0)

[1] Czyli nieco inaczej niż w teorii automatów na bi-nieskończonych słowach, gdzie nie wyróżnia się punktu 0.

można – w tej wersji twierdzenia prawie na pewno – dowolnie przybliżać, ale nie można jej osiągnąć.

Uwagi o prezentacji

Praca napisana jest zwięźle, co ogólnie uważam za zaletę. Dowody podane są precyzyjnie. Odniosłem jednak wrażenie, że autor założył znaczną wiedzę czytelnika — nieco więcej wyjaśnień intuicyjnych, a zwłaszcza przykładów dla wprowadzanych nowych pojęć uczyniłoby lekturę łatwiejszą. Pewnym mankamentem jest kilka drobnych nieścisłości, które dostrzegłem i wszystkie wyjaśniłem w korespondencji z autorem, jednak z obowiązku recenzenta odnotowuję poniżej.

Definition 64. Powinnno być $S_f(x,n) = \sum_{i<n} x_{s_f(x,n)+1}$ (reguła selekcji wskazuje **następny** bit).

Stwierdzenie w tekście poprzedzającym *Proposition 90*: *…the class of 1-random sequences is, in fact,* Π^0_1 jest chyba na wyrost (?); Downey i Hirschfeldt stwierdzają jedynie, że klasa ta jest Σ^0_2 (por. [12], *Proposition 8.1.1*), co jednak wystarcza dla *Proposition 90* (które w cytowanym w tym miejscu [12] jest obecne jako *Proposition 8.1.2*).

W *Lemma 99* zapewne powinno być *for all* $n \in \mathbb{N}$ (nie x), a odnośnik w dowodzie *Theorem 100* prowadzi do *Lemma 99* (nie *Lemma 1*).

Definition 102. Miało być, że zbiór A spotyka (*meets*) każdy **gęsty** zbiór słów klasy Σ^0_1. (Zbiór słów jest gęsty, jeśli każde słowo posiada rozszerzenie w tym zbiorze.) Wszelako zbiór konstruowany w dowodzie *Proposition 103* ma tę własność, więc dowód jest poprawny.

W dowodzie *Proposition 123* w punkcie *Verification* notacja powinna być $\lim_{n\to\infty} \frac{\sharp_1(A_1^n)}{n} = 0$.

Definition 128. Miało być $\liminf_{n\to\infty} \frac{\sharp\{i<n : f(i)=A_i\}}{n} = r$.

Powyższe usterki mogą być łatwo poprawione i nie obniżają wartości pracy.

Jeszcze uwaga co do języka angielskiego. Nie uważam się wprawdzie za eksperta, ale znana mi praktyka jest taka, że w tekstach matematycznych dominuje czas teraźniejszy (*present tense*); czas przyszły używany jest jedynie wtedy, gdy specjalnie chcemy zaznaczyć, że coś zostanie zdefiniowane lub udowodnione dopiero później. Tymczasem autor używa czasu przyszłego niemal przez *default*, co zwłaszcza w definicjach (*we will say*, *let… will be*) jest raczej niespotykane. Zachęcam autora do zbadania tej kwestii[2].

Ocena pracy i konkluzja

Jak już argumentowałem powyżej, rozprawa doktorska Tomasza Steifera zawiera szereg trudnych i ważnych wyników matematycznych. Autor wykazał się biegłością w technikach teorii obliczeń, a także twórczą znajomością innych dziedzin, m.in. rachunku prawdopodobieństwa i teorii ergodycznej. Szczególnym walorem pracy jest – jak wspomniałem – przerzucanie mostów między teoriami i umiejętność łączenia rozproszonych pojęć i idei w spójny obraz. Należy też podkreślić ambitny wybów tematu, który w naszym kraju nie ma wielu prekursorów. Wyniki pracy zostały częściowo opublikowane w dwóch pracach z dr. Dariszuem Kalocińskim (promotorem pomocniczym) w czasopiśmie (z listy A) *Mathematical Logic Quarterly* i na konferencji 2019 IEEE International Symposium on Information Theory; konferencja ta ma w rankingu CORE rangę B (*good*), ale tu również należy docenić wejście polskich autorów w nową społeczność naukową.

Podsumowując, stwierdzam, że praca pana Tomasza Steifera spełnia w zupełności ustawowe i zwyczajowe wymagania stawiane rozprawom doktorskim i wnoszę o dopuszczenie doktoranta do dalszych etapów przewodu doktorskiego.

Cieszyn, 17 czerwca 2020 — Damian Niwiński

[2] Zob. np. https://ell.stackexchange.com/questions/97918/we-prove-or-we-will-prove-in-mathematical-papers.