Katedra Podstaw Informatyki

Jacek Cichoń
jacek.cichon@pwr.edu.pl

Politechnika Wrocławska
Wydział Informatyki i Telekomunikacji
Katedra Podstaw Informatyki

21 marca 2023

O Katedrze

Kadra

1. kadra naukowo - badawcza: 23 osoby
2. tytuł profesora: 3 osoby
3. stopień doktora habilitowanego : 1 osoba
4. pozostale osoby: stopień doktora

Nasz paradygmat

Algorytmika = Esencja informatyki

O co chodzi ?

Hello world dla Big Data

```
topWords :: String → [(String,Int)]
topWords txt n = take ((sortByCount . group . sort . split . clear) txt) n
```

Analiza

- clear : $O(|txt|)$
- split : $O(|txt|)$
- sort : $O(|txt| \log(|txt|))$
- ...

Czy się zajmujemy?

Główne kierunki badań

1. kryptografia i bezpieczeństwo komputerowe
2. optymalizacja
3. algorytmy probabilistyczne
4. kombinatoryka
5. big data
6. sieci komunikacyjne (np. Ad Hoc)
7. ...

Ze względu na specyfikę spotkania nie omawiam działalności naukowej w matematyce i fizyce.

Kryptografia I

Opis problematyki

1. bezpieczne urządzenie: czarna skrzynka, która powinna uniemożliwić wyciek prywatnych kluczy kryptograficznych,
2. konsekwencja: brak dostępu/inspekcji środka urządzenia
3. ale: **dostawca urządzenia może zaimplementować urządzenie w zdradliwy sposób, tak aby klucze prywatne niezauważenie wyciekały**

Rozwiązanie w modelu Dawida i Goliata:

1. uniemożliwienie wycieku nawet gdy adwersarz zna stan generatora pseudolosowego w urządzeniu
2. efektywne odróżnienie podpisów cyfrowych utworzonych przez adwersarza, który złamał klucz publiczny
3. ...

Kryptografia - II

Opis problematyki

elastyczny blockchain

1. odtworzenie blockchaina gdy uległ zniszczeniu
2. zbalansowanie nienaruszalności/uaktualnień/weryfikowalności

Przykład

1. rozwinięcie izraelskiego schemat umożliwiającego odtworzenie całej struktury z tranzakcji, własność failstop: osoba oszukująca automatycznie ujawniłaby swoje klucze prywatne
2. VRBC – blockchain umożliwiający modyfikacje w określonych polach bez konieczności przebudowy blockchaina, wraz z efektywnymi metodami sprawdzenia poprawności

Kryptografia - III

Liderem grupy z KPI zajmującej się tymi zagadnieniami jest prof. Mirosław Kutyłowski. Tą tematyką zajmują się Ł. Krzywiecki, P. Błaśkiewicz, P. Kubiak, A.Lauks-Dutka, P. Kozieł, Marcin Słowik, Marta Słowik, J. Cichoń, M.Zawada.

Podsumowanie

1. szereg pragmatycznych rozwiązań przeciw wrogiej kryptografii
2. schematy realizujące ochronę danych osobowych i identyfikację elektroniczną
3. badania/publikacje ściśle związane z regulacjami europejskimi (GDPR, eIDAS, eID,...)

Konkretna kryptografia

Budowane schematy w ścisłym związku z wyzwaniami e-gov i transformacji cyfrowej. Ścisła interakcja z rzeczywistymi wyzwaniami!!!

Odporna optymalizacja - I

Opis problematyki

1. $\mathcal{X}$ - przestrzeń dopuszczalnych rozwiązań
2. $\mathcal{U}$ - przestrzeń scenariuszy (realizacji parametrów zadania), np. postaci $U_i = \{0, 2, 4, 10\}$ lub $U_i = [a, b]$
3. jeden z głównych celów

$$\min_{x \in \mathcal{X}} \max_{s \in \mathcal{U}} F(x, s)$$

Przykład

1. $\mathcal{X}$ - przestrzeń wszystkich ścieżek w grafie od s to t
2. $\mathcal{U} = (U_e, e \in E)$, $U_e = \{s_e^1, s_e^2\}$ (niepewne koszty łuków)
3. zadanie minimaksowe $\min_{x \in \mathcal{X}} \max_{s \in \mathcal{U}} \sum_{e \in x} s_e$ jest NP - trudne

Odporna optymalizacja - II

Przykład

1. INPUT: $(n, (C_i)_{i=1,\ldots,n}, k)$, gdzie $|C_i| = 2$, tj. $|\mathcal{U}| = 2$ (dwa scenarusze)
2. $\mathcal{X} = \{T \subseteq \{1, \ldots, n\} : \mathrm{card}(T) = k\}$
3. CEL: zadanie minimaksowe

Twierdzenie

1. *dla* $|\mathcal{U}| = 2$ *zadanie jest już NP-trudne*
2. *dla żadnego* $\epsilon > 0$ *nie istnieją wielomianowe* $O(\log^{1-\epsilon}(|\mathcal{U}|))$ *-aproksymacyjne algorytmy*
3. *istnieje* $O(\log(|\mathcal{U}|))$ *- aproksymacyjny algorytm wielomianowy*

Odporna optymalizacja - III

Liderem grupy z KPI zajmującej się tymi zagadnieniami jest prof. dr hab. Paweł Zieliński.

Podsumowanie

1. szereg klasycznych wyników w odpornej optymalizacji
2. szereg praktycznych algorytmów aproksymacyjnych
3. około 10 dobrych publikacji w ciągu ostatnich 5 lat.

Precyzja

Zbudowane algorytmy mają udowodnioną skuteczność !!!
To nie są heurystyki bądź metaheurystyki !!!

Algorytmy probabilistyczne - I

Złożoność komunikacyjna algorytmu "maxima propagation"

1. $\mathcal{G} = (V, E)$ - sieć rozproszona modelowana za pomocą **spójnego** grafu połączeń
2. $X : V \to \mathbb{R}$ obserwowalny parametr
3. rozważamy algorytm propagowania największej z zaobserwowanej obserwacji

Twierdzenie

1. *Niech $\mathcal{G}$ będzie dowolnym spójnym grafem o n wierzchołkom. Niech X będą losowo, niezależnie przydzielone z odcinka $[0, 1]$. Wtedy*

$$2 - \frac{1}{n} \leqslant E[MC_x] \leqslant \sum_{k=1}^{n} \frac{1}{k} \approx \ln n$$

2. *otrzymane wyniki pozwalają na konstruowanie dużej klasy algorytmów dla sieci rozproszonych o małej złożoności komunikacyjnej*

Algorytmy probabilistyczne - II

Sieci rozproszone

1. $\mathcal{G} = (V, E)$ - sieć rozproszona modelowana za pomocą grafu połączeń
2. $X : V \to \mathbb{R}$ obserwowalny parametr
3. CEL (przykład): wyznacz podstawowe parametry statystyczne, takie jak $\min_{v \in V} X(v)$,$\max_{v \in V} X(v)$, $\frac{1}{|V|} \sum_{v \in V} X(v)$

Rozwiązanie

1. za pomocą metody **extreme propagation** wyznacz losową próbkę
2. zbuduj przybliżony histogram
3. za pomocą tego histogramu, przy ponownym użyciu extreme propagation i liczników probabilistycznych wyznacz średnią wartość

Algorytmy probabilistyczne - III

Badaniem algorytmów probabilistycznych w KPI zajmują się J. Cichoń, Z. Gołębiewski, M. Kutyłowski, M. Sulkowska, M. Zawada,

Podsumowanie

1. szereg nowych algorytmów probabilistycznych
2. szereg praktycznych algorytmów aproksymacyjnych
3. około 15 dobrych publikacji w ciągu ostatnich 5 lat.

Precyzja

Zbudowane algorytmy mają udowodnioną skuteczność !!!
W większości wypadków mają uniwersalny charakter (np. są prawdziwe dla dowolnych rozkładów probabilistycznych)!!!

Kombinatoryka - I

Definicja

1. $\mathcal{S}_n$ = liczba nieetykietowanych planarnych binarnych drzew o n liściach
2. $\mathrm{sym}(T)$ = liczba takich wewnętrznych wierzchołków t, że lewa i prawe poddrzewa t są izomorficzne
3. $c_n = \frac{1}{(n-1)^2} \sum_{k=1}^{n-1} c_k c_{n-k}$

Twierdzenie

1. $E[\mathrm{sym}(S_n)] = n \cdot (0.3725 \pm \frac{1}{10^5})$
2. $E[H[S_n]] \approx 1.109 \cdot n$
3. *istnieje algorytm kompresji klasy $\mathcal{S}$ asymptotycznie optymalny.*

Kombinatoryka - II

Zespół

D. Bojko, J. Cichoń, Z. Gołębiewski, R. Kapelko, M. Sulkowska,

Big Data - I

Strumienie danych

1. obserwujemy intensywny strumień danych $(x_t)_{t=0,\ldots,n}$, mamy dane $T > 0$
2. chcemy mieć losową próbkę danych z ostatnich T obserwacji $x_{n-T+1}, x_{n-T+2}, \ldots, x_n$

Rozwiązanie

1. użyj dwie niezależne kopie diabelskich schodów długości T
2. użyj większy ze wskaźników
3. otrzymasz wskaźnik jednorodnie rozłożony na zbiorze $\{n - T + 1, \ldots, n\}$

Big Data - II

Szkice danych

1. obserwujemy strumień danych $\mathcal{A} = ((x_t, w_t))_{t=0,\ldots,n}$, w_i = waga elementu x_i
2. chcemy zbudować szkic $S_{\mathcal{A}}$ strumienia małych rozmiarów o dobrych własnościach algebraicznych; $|S_{\mathcal{A}}|_w$ = suma wag unikalnych elementów ,

Wyniki

Opracowano metodę konstrukcji szkiców, która

1. umożliwia dokładne szacowanie

$$|S_{\mathcal{A}} \cap S_{\mathcal{B}}|_w, \quad |S_{\mathcal{A}} \cup S_{\mathcal{B}}|_w, \quad |S_{\mathcal{A}} \setminus S_{\mathcal{B}}|_w$$

2. dla dla strumienia rozmiaru 10^9: rozmiar $S_{\mathcal{A}} \approx$ 10kB; dokładność rzędu 10 %

Big Data - III

Zagadnieniami Big Data w KPI zajmują się D. Bojko, J. Cichoń, K. Gotfryd, J. Lemiesz, M. Kutyłowski, M. Zawada.

Podsumowanie

1. On the algebra of data sketches, Jakub Lemiesz, 2021, 47th International Conference on Very Large Data Bases, publikacja z roku 2023
2. Badane zagadnienia: szkice, próbkowanie, estymowanie rozmiaru sieci

Metodologia

Zbudowane algorytmy mają udowodnioną poprawość
Przeprowadzone symulacje służą tylko do sprawdzenia, czy gdzieś w obliczenia nie wkradł się błąd!!!

Dydaktyka - I

Kierunki

1. Informatyka algorytmiczna, studia I stopnia
2. Informatyka algorytmiczna, studia II stopnia
   1. specjalność Algorytmika
   2. specjalność Computer Security (w języku angielskim)

Specyfika

1. solidna dawka matematyki niezbędnej do rozumienia i analizy algorytmów
2. główny element studiów I stopnia: Algorytmy i Struktury Danych
3. Języki Formalne i Teoria Translacji, Systemy Wbudowane

Dydaktyka - II

Znajomość języków programowania

1. Wszyscy absolwenci: podstawowe, powszechnie używane (w tym Ada)
2. Do wyboru: programowanie funkcyjne (Haskell, ML, F#), programowanie w logice: Prolog

Co powinni umieć absolwenci

1. umiejętność znajdowania krytycznych punktów aplikacji (czyli wyszukiwanie miejsc do sensownej optymalizacji)
2. optymalizacja parametrów stosowanych algorytmów
3. znajomość (wybranych) najnowszych algorytmów z aktualnych konferencji algorytmicznych

Potrzeby

Co potrzebujemy do badań

1. Główne narzędzia: kartka papieru, długopis, tablica, kreda lub pisak
2. Programy komputerowe: np. Mathematica - do przyśpieszania i sprawdzania poprawności obliczeń matematycznych
3. Komputery: głównie do sprawdzania poprawności modeli (symulacje), do spisywania prac naukowych
4. Trochę środków na współpracę międzynarodową (Austria, Chiny, Francja, Izrael, USA)

Potrzeby

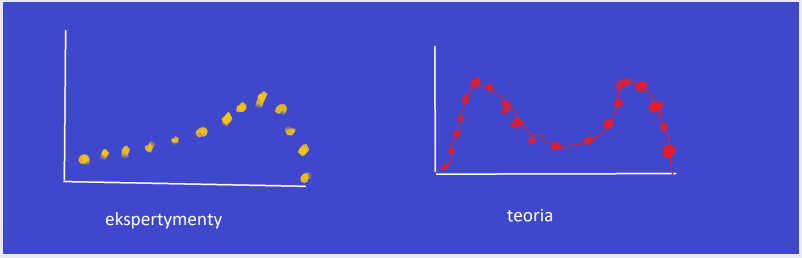
Co potrzebujemy do badań

1. Główne narzędzia: kartka papieru, długopis, tablica, kreda lub pisak
2. Programy komputerowe: np. Mathematica - do przyśpieszania i sprawdzania poprawności obliczeń matematycznych
3. Komputery: głównie do sprawdzania poprawności modeli (symulacje), do spisywania prac naukowych
4. Trochę środków na współpracę międzynarodową (Austria, Chiny, Francja, Izrael, USA)

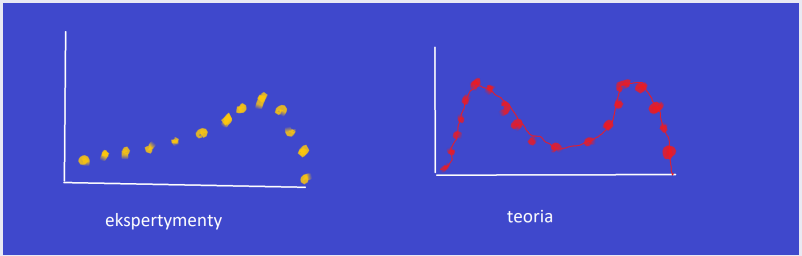
Wniosek

Stosunkowo mało potrzebne są nam granty.
Głównie na wyjazdy na konferencje zagraniczne i współpracę międzynarodową.

Hipotetyczna sytuacja



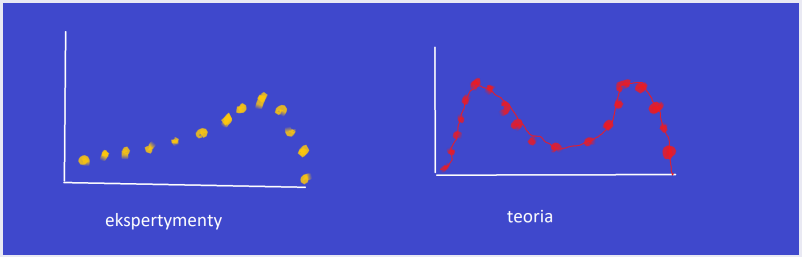
Hipotetyczna sytuacja



Najbardziej prawdopodobna przyczyna

Pomyliliśmy się w obliczeniach !!!
Trzeba na nowo wziąć się do pracy.

Hipotetyczna sytuacja



Najbardziej prawdopodobna przyczyna

Pomyliliśmy się w obliczeniach !!!
Trzeba na nowo wziąć się do pracy.

To już koniec