

Advertising on the Web

On-line algoritmi

- Off-line algoritam: ulazni podaci su dostupni na početku, algoritam može pristupati podacima u bilo kom redosljedu, na kraju se saopštava rezultat obrade
- On-line algoritam: ulazni podaci nijesu u cjelosti dostupni, primjer stream-mining algoritmi, on-line algoritam saopštava rezultat poslije svakog elementa toka podataka

Adword model

- Prikazivanje reklama kao rezultat postavljenog search-query
- Of-line rješenje: kreira se „prozor“ za mjesec dana, dostupni su podaci o „ulozima“ za svaki se, search term i budžeti oglašivača, algoritam maksimizuje dobit search mašine i broj impression-a za svakog oglašivača, impression je *one display of the ad with the download of the page by some user*
- Realno moramo imati on-line algoritam, neposredno poslije svakog upita moramo selektovati koje reklame prikazujemo
 - Možemo koristiti već dostupne informacije: preostali budžet, click-through rate itd.
 - Ne znamo ništa o upitima koji tek treba da pristignu u sistem

Primjer 1

- Dva proizvođača A i B
- A plaća 10 centi za ključnu riječ „chesterfield“
- B plaća po 20 centi za ključne riječi „chesterfield“ i „sofa“
- Budžeti za A i B su 100 eura
- Kog proizvođača treba prikazati ako je upit „chesterfield“?

Greedy algoritmi

- Algoritam generiše rezultat za svaki ulazni podatak tako što maksimizuje neku funkciju
- Prethodni primjer: algoritam prikazuje reklamu proizvođača sa većim ulogom
- Uporediti off-line i on-line algoritme za prethodni primjer

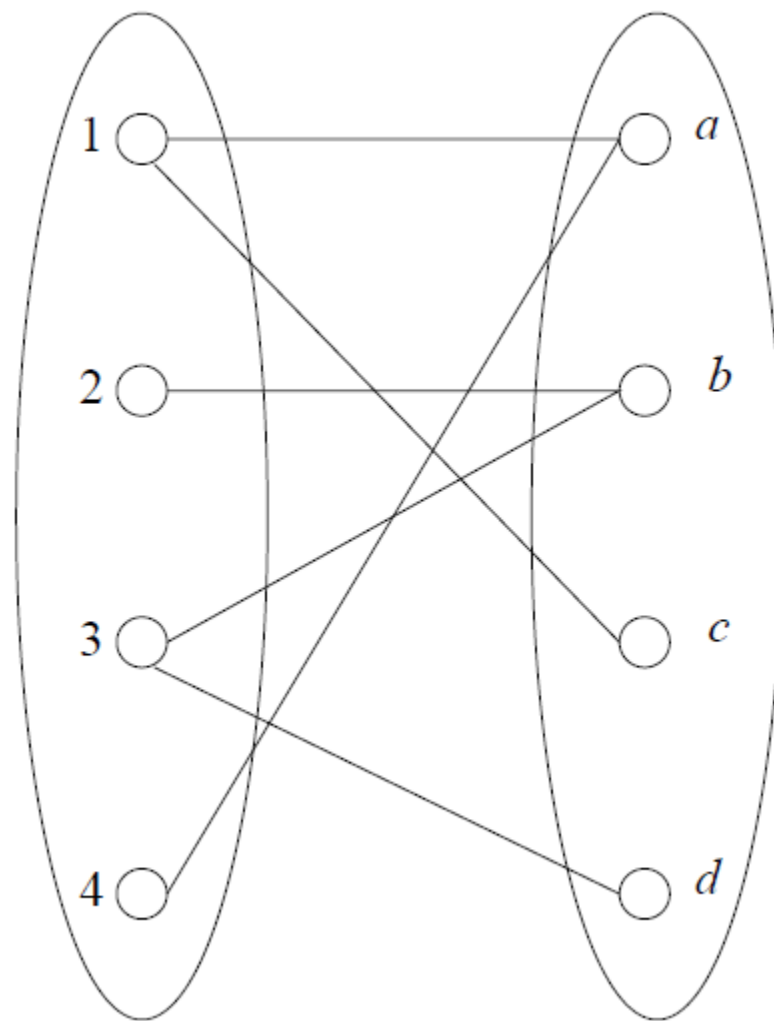
The competitive ration

- Competitive ratio c je konstanta manja od 1 za koju važi da je rezultat on-line algoritma najmanje $c * \text{rezultat optimalnog off-line algoritma}$
- Competitive ratio za prethodni primjer je $\frac{1}{2}$

The matching problem

- matching ads to search queries
- bipartitni graf, matching je podskup grana takav da se nijedan čvor ne pojavljuje više od jednom, perfect matching ako se svaki čvor grafa pojavljuje (particije grafa moraju biti sa jednakim brojem elemenata), maximal matching je matching sa najvećim brojem grana
- Every perfect matching is maximal?

Primjer 2



Greedy algoritam za maximal matching

- Off-line algoritam kvadratne složenosti
- On-line algoritam kada nije poznat čitavi graf unaprijed
- Algoritam: razmatraju se grane grafa u određenom redosljedu, grana (x, y) se dodaje ako x i y nijesu do tok trenutka uključeni u matching, inače se grana preskače
- Primjer 2, leksikografski poredak grana $(1, a), (1, c), \dots, (4, a)$, rezultat $\{(1, a); (2, b); (3, d)\}$, nijesmo našli maximal matching
- Competitive ratio je $1/2$

Početak

- Kompanija Overture uvela je 2000-tih „novi“ model pretraživanja web-a, oglasivači definišu uloge na ključne riječi, kada korisnik pokrene upit sa nekom od ključnih riječi u rezultatu se prikazuju svi oglasivači sa ulozima i to u redosljedu highest-bid-first, ako korisnik klikne na link onda oglasivač plaća ulog
- Google Adwords
 - Ograničeni broj oglasivača se prikazuje u rezultatu
 - Definisani su mjesečni budžeti svakog oglasivača
 - Redosljed prikazivanja na osnovu click-through rate * bid

Adwords problem

- Ulazni podaci:
 - Za svakog oglasivača vrijednosti uloga za ključne riječi
 - Click-through rate da svaki par oglasivač-upit
 - Mjesečni budžeti
 - Ograničenje na broj reklama koje se prikazuju
- Izlaz za svaki search query je skup reklama
 - Veličina rezultata je ograničena ulaznim parametrom
 - Prikazuju se samo oglasivači koji su definisali uloge za taj search-query
 - Prikazuju se samo oglasivači sa dovoljnim preostalim budžetom

Greedy algoritam

- Uprošćavanja
 - Prikazuje se samo jedna reklama u rezultatu
 - Svi oglasivači imaju jednake budžete
 - Svi click-through rate su jednaki
 - Svi ulozi su 0 ili 1
- Greedy algoritam će za postavljeni upit izabrati bilo kojeg oglasivača koji ima ulog 1 za taj upit
- Competitive ratio je $\frac{1}{2}$

Primjer 3

- Dva oglasivača A i B
- Dva upita x i y
- Budžeti po 2 eura
- Oglasivač A ulaže samo na x, oglasivač B na x i y
- Neka je sekvenca upita xxyy
- Kolika je zarada optimalnog off-line algoritma?
- Koji je najgori scenario za greedy algoritam?

The Balance algoritam

- Competitive ratio je $\frac{3}{4}$
- Prikazuje se oglašivač sa definisanim ulogom na search-query i koji ima najveći preostali budžet
- Kolika je dobit Balance algoritma u primjeru 3?
- Competitive ratio je $\frac{3}{4}$ ako postoje samo dva oglašivača
- Competitive ratio je 0.63 kada postoji više oglašivača

Proširenje

- Eliminirati pretpostavku da je ulog 0 ili 1, dati prednost oglasivaču sa većim ulogom
- Iskoristiti budžete za što veći broj oglasivača
- Algoritam: oglasivači A_i , ulozi x_i , f_i procenat neutrošenog budžeta za A_i , maksimizuje se funkcija $x_i^*(1-e^{-f_i})$

Implementacija

- Matching bids and search queries
 - Equality of sets
- Complex matching problem: match many bids (skup ključnih riječi) against many documents (značajno veći skup riječi, kao što je mejl, članak iz novina, objava na društvenim mrežama itd)
- Pretpostavka je da postoji veliki broj bid-ova, kao i da u sistem dopijeva stotine dokumenata u sekundi
- Bid se predstavlja kao lista riječi, dodatno ima
 - Status – broj riječi sa početka liste prisutnih u tekućem dokumentu
 - Početak liste je uređen leksikografski, drugi dio: the most frequent words last

Nastavak

- Bidovi se smještaju u heš tabelu, ključ je prva riječ, statusi su 0
- Još jedna heš tabela sadrži kopije bidova koji su djelimično pronađeni u dokumentu
- Procesiranje dokumenta
 - Sortirati riječi u dokumentu
 - Za svaku riječ w
 - Naći sve bidove koji imaju ključ w u tabeli sa parcijano prepoznatim bidovima
 - Za svaki takav bid, ako je w posljednja riječ, dodati bid u rezultat
 - Ako w nije posljednja riječ, inkrementira se status, heš po riječi sa pozicije $\text{status}+1$
 - Naći sve bidove koji počinju sa w
 - Ako je bid sa samo riječju w , pridruži ga rezultatu
 - Inače, dodaj bid sa statusom 1 u tabelu parcijalno prepoznatih

Skica rješenja

