

POSUDEK VEDOUCÍHO/OPONENTA BAKALÁŘSKÉ PRÁCE

Název: Jak funguje vyhledávač Google

Autor: Jaromír Vaněček

SHRNUTÍ OBSAHU PRÁCE

Práce se zabývá algoritmem PageRank na ohodnocení důležitosti internetových stránek.

CELKOVÉ HODNOCENÍ PRÁCE

Téma práce. Zadání práce bylo, stručně, seznámit se s principem PageRanku (a relevantní matematickou teorií) a aplikovat výsledky v jiných oblastech. První část se podařilo docela dobře zpracovat, pojetí aplikační části je neuspokojivé.

Vlastní příspěvek. Vlastním příspěvkem je přehledné shrnutí relevantní matematiky z více zdrojů a doplnění některých detailů důkazů. Práce je napsaná čtivě a je docela dobrým úvodem do problematiky. Aplikace na fotbalovou ligu je rovněž vlastním příspěvkem autora. Tuto část ale, jak jsem zmínil, nepovažuji za kvalitní.

Matematická úroveň. Formulace definic, tvrzení a důkazů je (většinou) korektní. Důkazy jsou ale někde dost těžkopádné (příklady uvádím níže).

Práce se zdroji. Nemám vážné výhrady (ale viz bod 16).

Formální úprava. Dobrá.

PŘIPOMÍNKY A OTÁZKY

Připomínky jsou řazeny podle čísla strany. Za podstatné považuji zejména body 15–18.

1. Předposlední odstavec na straně 8 mi přijde zavádějící. Žádný vyhledávač neřadí výsledky náhodně, nejen Google. Úspěch Googlu může být mimo jiné v tom, že řazení je lepší než má konkurence.
2. V druhém odstavci na straně 10 se říká „Následující formulace přesně říká, jak toho docílíme“. Formulace přitom neodpovídá přesně vzorci na následující straně.
3. Na straně 13 je psáno „...můžeme využít náhodnou procházku. Tento termín má svou matematickou definici, kterou uvádět nebudeme, ale budeme ho používat intuitivně“. Co část 3.2?
4. Na straně 14 by bylo vhodné uvést jak vektory v_0, v_1, \dots vypadají pro výslednou matici G a výsledek okomentovat.
5. Definice kladné (nezáporné) matice na straně 15 je velmi neobratně napsaná.
6. Věta „Pro matici A s vlastními čísly \dots vypadá Jordanův kanonický tvar \dots “ měla být uvedena až za následujícím odstavcem, kde se (pouze implicitně) vysvětluje, co to znamená. Navíc je tvrzení nepravdivé (nebo přinejmenším velmi nepřesně formulované). Jak je to správně?
7. V části 3.2. se explicitně neříká, o jaký proces jde v případě Googlu. Lze si to sice domyslet z poznámky před definicí 3.3., ale podrobnější zdůvodnění by bylo lepší.

8. Poznámka nad tvrzením 3.9: jaké je zdůvodnění toho, že důležitost počítáme součtem? Nebylo to ani předtím nikde zmíněno, diskuze spíše odpovídala limitě.
9. V lemmatu 3.11 chybí předpoklad, že S je nezáporná
10. V druhém odstavci na straně 20 má asi být matice sousednosti (ne vzdálenosti). Co je váha hrany?
11. V posledním odstavci na straně 20 se tvrdí, že X_{kl} jsou čtvercové. To není nutně pravda.
12. Důkaz implikace zprava doleva v lemmatu 3.16 indukci je extrémním příkladem neefektivity. Dále nevidím, jak důsledek 3.17. snadno plyne z lemmatu. Proč nedokázat přímo tento důsledek?
13. K důkazu Perronovy věty. Vzhledem k tomu, jak podrobně jsou rozepsány některé detaily (jako lemma 3.16), bylo by také vhodné rozepsat, proč limita k -tých mocnin Jordanových buněk pro $|\lambda| < 1$ je nulová matice. Dále na straně 24 uprostřed se dělí v_1 a není zdůvodněno, proč $v_1 \neq 0$.
14. Lemma 3.23. je nejasně formulováno. Dokazujete, že každý vlastní vektor je roven *ec*. (Navíc formulace „vypadá jako“ není dobrá, vyskytuje se i jinde v textu.)
15. Diskuze mocniné metody od strany 27 dole je zavádějící a neefektivní. Předně, trik metody není umocňování matice, tomu se právě metoda vyhýbá! V příkladech se znovu odvozuje něco, co už bylo odvozeno jinde. Jak diskuzi zefektivnit?
16. Jaký software jste využil pro výpočty?
17. Část 4 zaměřenou na Synot ligu považuji za špatnou. Výsledkům ligy se nějakým způsobem přiřadí řada matic a spočítá se Perronův vektor. Chybí jakékoliv zdůvodnění, proč by měl Perronův vektor pro nějakou z matic jakkoliv korespondovat s kvalitou týmu (je zajímavé, že vysvětlení v případě Google z předchozích kapitol je docela pěkné). V některých případech je dokonce očividné, že výsledek nebude kvalitě odpovídat vůbec (alespoň při nevyrovnaném počtu utkání). Při obhajobě uveďte příklad.
18. Kapitola “Dodatečná pozorování” působí opravdu dodatečně. Ekvivalence není správně definovaná. Pozorování 5.2. je nepravdivé (proč?). Poslední věta posledního odstavce důkazu je „Pro každou koncovou třídu máme množinu lineárně závislých vektorů, koncových tříd je l , tedy $\dim(\text{Ker}(H - I)) \geq 1$ “. Tomu neumím dát žádný rozumný smysl. Následující příklad mi přijde zmatený a nadbytečný.

ZÁVĚR

Práci považuji za dostatečnou a doporučuji ji k obhajobě.

Libor Barto
Katedra algebry
5.6.2015