

Uniwersytet Warszawski
Wydział Matematyki, Informatyki i Mechaniki

Grzegorz Szpak

Nr albumu: 319400

Systemy rekomendacyjne oparte o wspólną filtrację

Praca licencjacka
na kierunku MATEMATYKA

Praca wykonana pod kierunkiem
prof. Andrzeja Skowrona / dr. Marcina Szczuki
Instytut Matematyki / Instytut Informatyki

Lipiec 2016

Oświadczenie kierującego pracą

Potwierdzam, że niniejsza praca została przygotowana pod moim kierunkiem i kwalifikuje się do przedstawienia jej w postępowaniu o nadanie tytułu zawodowego.

Data

Podpis kierującego pracą

Oświadczenie autora (autorów) pracy

Świadom odpowiedzialności prawnej oświadczam, że niniejsza praca dyplomowa została napisana przeze mnie samodzielnie i nie zawiera treści uzyskanych w sposób niezgodny z obowiązującymi przepisami.

Oświadczam również, że przedstawiona praca nie była wcześniej przedmiotem procedur związanych z uzyskaniem tytułu zawodowego w wyższej uczelni.

Oświadczam ponadto, że niniejsza wersja pracy jest identyczna z załączoną wersją elektroniczną.

Data

Podpis autora (autorów) pracy

Streszczenie

W pracy przedstawiono analizę systemów rekomendacyjnych opartych na metodzie wspólnej filtracji. Zaprezentowane zostały dwa podstawowe typy wspólnej filtracji - filtracja oparta o użytkowników oraz filtracja oparta o przedmioty. Autor skupia się na implementacji tej metody opartej na algorytmie k - najbliższych sąsiadów. Przedstawione zostają różne warianty poszczególnych składowych algorytmu (wybór funkcji podobieństwa, rozmiar sąsiedztwa, sposób ustalania oceny na podstawie ocen sąsiadów) i ich wpływ na jakość systemu.

Słowa kluczowe

eksploracja danych, systemy rekomendacyjne, wspólna filtracja, algorytm k - najbliższych sąsiadów

Dziedzina pracy (kody wg programu Socrates-Erasmus)

11.1 Matematyka

Klasyfikacja tematyczna

68T05 Learning and adaptive systems

Tytuł pracy w języku angielskim

Recommender systems based on collaborative filtering

Spis treści

1. Wprowadzenie	5
1.1. Systemy rekomendacyjne i ich zastosowanie	5
1.1.1. Typy systemów rekomendacyjnych	5
1.2. The Netflix Prize	5
2. Podstawowe pojęcia	7
2.1. Definicje	7
2.2. Blabalizator różnicowy	7
3. Wcześniejsze implementacje blabalizatora różnicowego	9
3.1. Podejście wprost	9
3.2. Metody wykorzystujące teorię Głombaskiego	9
3.3. Metody wykorzystujące własności fektorów σ	9
4. Teoria fektorów σ-ρ	11
5. Dokumentacja użytkowa i opis implementacji	13
6. Podsumowanie	15
6.1. Perspektywy wykorzystania w przemyśle	15
A. Główna pętla programu zapisana w języku TōFoo	17
B. Przykładowe dane wejściowe algorytmu	19
C. Przykładowe wyniki blabalizy (ze współczynnikami σ-ρ)	21
Bibliografia	23

Rozdział 1

Wprowadzenie

1.1. Systemy rekomendacyjne

Każdego dnia stawiani jesteśmy przed szeregiem wyborów. Którą książkę przeczytać? Jaki film obejrzeć? Którą stronę internetową odwiedzić? Jako że ilość informacji w otaczającym nas świecie rośnie dużo szybciej niż nasze możliwości poznawcze, konieczne stało się stworzenie technologii, która będzie w stanie wybrać informacje najbardziej dla nas użyteczne. Taką rolę we współczesnych aplikacjach internetowych pełnią systemy rekomendacyjne. Bazując na dotychczasowej aktywności użytkownika, systemy rekomendacyjne starają się przewidzieć jego ocenę wobec danego przedmiotu. Terminy "ocena" i "przedmiot" są oczywiście ogólne i w zależności od zastosowania mogą reprezentować różne zjawiska - od wystawienia rzeczywistej oceny filmowi, przez kupno przedmiotu w sklepie internetowym, po fakt dołączenia do grupy na portalu społecznościowym.

Przydatność systemów rekomendacyjnych obrazuje najlepiej tak zwany "fenomen długiego ogona" (ang. *long tail phenomenon*). + ile w google netflix

1.2. Fenomen "długiego ogona"

1.3. Typy systemów rekomendacyjnych

dcdscds

1.4. The Netflix Prize

Blabalizator różnicowy jest podstawowym narzędziem blabalii fetorycznej. Dlatego naukowcy z całego świata prześcigają się w próbach efektywnej implementacji. Opracowana przez prof. Fifaka teoria fetorów σ - ρ otwiera w tej dziedzinie nowe możliwości. Wykorzystujemy je w niniejszej pracy.

Przystępne wprowadzenie do blabalii fetorycznej można znaleźć w pracy Fifaka i Gryzgrzechotalskiego [Fif00]. Dlatego w niniejszym tekście ograniczymy się do przypomnienia pojęć podstawowych.

Praca składa się z pięciu rozdziałów i dodatków. W rozdziale 2 przypomniano podstawowe pojęcia blabalii fetorycznej. Dotychczasowe próby implementacji blablizatora różnicowego zestawiono w rozdziale 3. Rozdział 4 przedstawia teorię Fifaka i wyjaśnia sposób jej wykorzystania w implementacji blabalizatora. W rozdziale 5 przedstawiono algorytm blabalizy

i realizujący go program komputerowy. Ostatni rozdział zawiera przemyślenia dotyczące możliwego wpływu dostępności efektywnej blabalizy numerycznej na rozwój blabalii fetorycznej. W dodatkach umieszczono najciekawszy fragment programu, przykładowe dane i wyniki działania programu.

Rozdział 2

Podstawowe pojęcia

Pojęciem pierwotnym blabalii fetorycznej jest *blaba*. Blabaliści nie podają jego definicji, mówiąc za Ciach-Pfe t-ām Kûn (fooistyczny mędrzec, XIX w. p.n.e.):

Blaba, który jest blaba, nie jest prawdziwym blaba.

tłum. z chińskiego Robert Blarbarucki

2.1. Definicje

Oto dwie definicje wprowadzające podstawowe pojęcia blabalii fetorycznej:

Definicja 2.1.1 *Silny, zwarty i gotowy fetor bazowy nazwiemy skupieniem.*

Definicja 2.1.2 *Fetorem nazwiemy skupienie blaba spełniające następujący aksjomat reperkusatywności:*

$$\forall \mathcal{X} \in Z(t) \exists \pi \subseteq \oint_{\Omega^2} \kappa \leftrightarrow 42$$

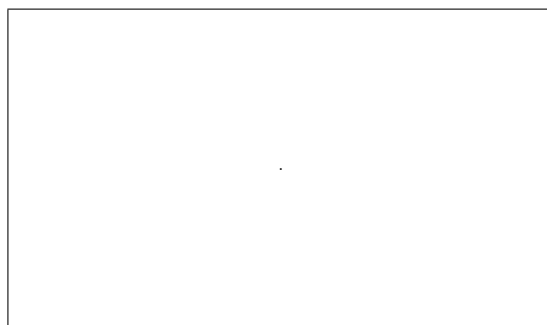
2.2. Blabalizator różnicowy

Teoretycy blabalii (zob. np. pracę [Głomb04]) zadowalają się niekonstruktywnym opisem natury fetorów.

Podstawowym narzędziem blabalii empirycznej jest blabalizator różnicowy. Przyrząd ten pozwala w sposób przybliżony uzyskać współczynniki rozkładu Głombaskiego dla fetorów bazowych i harmoniczych. Praktyczne znaczenie tego procesu jest oczywiste: korzystając z reperkusatywności pozwala on przejść do przestrzeni Λ^∇ , a tym samym znaleźć retroizotonalne współczynniki semi-quasi-celibatu dla klatek Rozkoszy (zob. [Rozk93]).

Klasyczne algorytmy dla blabalizatora różnicowego wykorzystują:

1. dualizm falowo-korpuskularny, a w szczególności
 - (a) korpuskularną naturę fetorów,
 - (b) falową naturę blaba,
 - (c) falowo-korpuskularną naturę gryzmołów;
2. postępującą gryzmozolizację poszczególnych dziedzin nauki, w szczególności badań systemowych i rozcieńczonych;



Rysunek 2.1: Artystyczna wizja blaba w obrazie węgierskiego artysty Josipa A. Rozkoszy pt. „Blaba”

3. dynamizm fazowy stetryczenia parajonizacyjnego;

4. wreszcie tradycyjne opozycje:

- duch — bakteria,
- mieć — chcieć,
- myśl — owsianka,
- parafina — durszlak¹,
- logos — termos

z właściwym im przedziwym dynamizmem.

¹Więcej o tym przypadku — patrz prace Gryzybór-Głombaskiego i innych teoretyków nurtu teoretyczno-praktycznego badań w Instytucie Podstawowych Problemów Blabalii w Fife.

Rozdział 3

Wcześniejsze implementacje blabalizatora różnicowego

3.1. Podejście wprost

Najprostszym sposobem wykonania blabalizy jest siłowe przeszukanie całej przestrzeni rozwiązań. Jednak, jak łatwo wyliczyć, rozmiar przestrzeni rozwiązań rośnie wykładniczo z liczbą fetorów bazowych. Tak więc przegląd wszystkich rozwiązań sprawdza się jedynie dla bardzo prostych przestrzeni lamblialnych. Oznacza to, że taka metoda ma niewielkie znaczenie praktyczne — w typowym przypadku z życia trzeba rozważać przestrzenie lamblialne wymiaru rzędu 1000.

W literaturze można znaleźć kilka prób opracowania heurystyk dla problemu blabalizy (por. [Zen69]). Korzystając z heurystyk daje się z pewnym trudem dokonać blabalizy w przestrzeni o np. 500 fetorach bazowych. Należy jednak pamiętać, że heurystyka nie daje gwarancji znalezienia najlepszego rozwiązania. Fifak w pracy [Fif01] podaje, jak dla dowolnie zadanej funkcji oceniającej skonstruować dane, dla których rozwiązanie wygenerowane przez algorytm heurystyczny jest dowolnie odległe od rzeczywistego.

3.2. Metody wykorzystujące teorię Głombaskiego

Teoria Głombaskiego (zob. [Głomb04]) dostarcza eleganckiego narzędzia opisu przejścia do przestrzeni Λ^∇ . Wystarczy mianowicie przedstawić fetory bazowe wyjściowej przestrzeni lamblialnej w nieskończonej bazie tak zwanych wyższych aromatów. (Formalną definicję tego pojęcia przedstawię w rozdziale poświęconym teorii Fifaka). Podstawową cechą wyższych aromatów jest ulotność. To zaś oznacza, że odpowiednio dobierając współczynniki przejścia do przestrzeni wyższych aromatów można zagwarantować dowolną z góry zadaną dokładność przybliżonego rozwiązania problemu blabalizy.

Oczywiście ze względu na nieskończony wymiar przestrzeni wyższych aromatów koszt poszukiwania współczynników blabalizy jest liniowy ze względu na wymiar wyjściowej przestrzeni lamblialnej.

3.3. Metody wykorzystujące własności fetorów σ

Najchętniej wykorzystywaną przestrzenią wyższych aromatów jest przestrzeń fetorów σ . Fetory σ dają szczególnie prostą bazę podprzestrzeni widłowej. Wiąże się to z faktem, że w tym

przypadku fety harmoniczne wyższych rzędów są pomijalne (rzędu 2^{-t^3} , gdzie t jest wymiarem wyjściowej przestrzeni lamblialnej).

Niestety z fetorami σ wiąże się też przykre ograniczenie: można wykazać (zob. [Fif01, s. 374]), że dla dowolnie dobranej bazy w podprzestrzeni widłowej istnieje ograniczenie dolne w metryce sierpa na odległość rzutu dokładnego rozwiązania problemu blabalizy na podprzestrzeń widłową. Ponieważ rzut ten stanowi najlepsze przybliżone rozwiązanie, jakie można osiągnąć nie naruszając aksjomatu reperkusatywności, więc istnieje pewien nieprzekraczalny próg dokładności dla blabalizy wykonanej przez przejście do przestrzeni fetorów σ . Wartość retroinicjalną tego progu nazywa się *reziduum blabicznym*.

Rozdział 4

Teoria fetorów σ - ρ

Głównym odkryciem Fifaka jest, że fetor suprakowariantny może gryzmolizować dowolny ideał w podprzestrzeni widłowej przestrzeni lamblialnej funkcji Rozkoszy.

Udowodnienie tego faktu wymagało wykorzystania twierdzeń pochodzących z kilku niezależnych teorii matematycznych (zob. na przykład: [Whi25, Spy59, Rozk93, Bea65, Hopp96, Sri64]). Jednym z filarów dowodu jest teoria odwzorowań owalnych Leukocyta (zob. [Leuk00]).

Znaczenie twierdzenia Fifaka dla problemu blabalizy polega na tym, że znając retroizotonalne współczynniki dla klatek Rozkoszy można przeprowadzić fetory bazowe na dwie nieskończone bazy fetorów σ w przestrzeni K_7 i fetorów ρ w odpowiedniej quasi-quasi-przestrzeni równoległej (zob. [Hopp96]). Zasadnicza różnica w stosunku do innych metod blabalizy polega na tym, że przedstawienie to jest dokładne.

Rozdział 5

Dokumentacja użytkowa i opis implementacji

Program przygotowany dla systemu operacyjnego M\$ Windows uruchamia się energicznym dwumlaskiem na jego ikoncie w folderze `\\FIDO\F00\BLABA`. Następnie kolistym ruchem ręki należy naprowadzić kursor na menu `Blabaliza` i uaktywnić pozycję `Otwórz plik`. Po wybraniu pliku i zatwierdzeniu wyboru przyciskiem `OK` rozpocznie się proces blabalizy. Wyniki zostaną zapisane w pliku o nazwie `99-1a.tx.43` w bieżącym folderze.

Rozdział 6

Podsumowanie

W pracy przedstawiono pierwszą efektywną implementację blabalizatora różnicowego. Umiejętność wykonania blabalizy numerycznej dla danych „z życia” stanowi dla blabalii fetorycznej podobny przełom, jak dla innych dziedzin wiedzy stanowiło ogłoszenie teorii Mikołaja Kopernika i Gryzybór Głombaskiego. Z pewnością w rozpoczynającym się XXI wieku będziemy obserwować rozkwit blabalii fetorycznej.

Trudno przewidzieć wszystkie nowe możliwości, ale te co bardziej oczywiste można wskazać już teraz. Są to:

- degryzmolizacja wieńców telecentrycznych,
- realizacja zimnej reakcji lambliarnej,
- loty celulityczne,
- dokładne obliczenie wieku Wszechświata.

6.1. Perspektywy wykorzystania w przemyśle

Ze względu na znaczenie strategiczne wyników pracy ten punkt uległ utajnieniu.

Dodatek A

Główna pętla programu zapisana w języku TōFoo

```
[[foo]{,}[[a3,((([,),{[[]]})),  
  [1; [{,13},[[[11],11],231]]].  
  [13;[!xz]].  
  [42;[{,x},[[2],{'a'},14]]].  
  [br;[XQ*10]].  
, 2q, for, [1,]2, [...].[7]{x}],[(,([,[[1]{123},},},;.112]],  
  else 42;  
  . 'b'.. '9', [[13141],{13414}], 11),  
[1; [[134,sigma],22]].  
[2; [[rho,-],11]].  
)[14].  
, {1234}],. [map [cc], 1, 22]. [rho x 1]. {22; [22]},  
  dd.  
[11; sigma].  
  ss.4.c.q.42.b.ll.ls.chmod.aux.rm.foo;  
[112.34; rho];  
  00111010101010101010101010101011111010010  
[22%f4].  
cq. rep. else 7;  
]. hlt
```


Dodatek B

Przykładowe dane wejściowe algorytmu

α	β	γ
901384	13784	1341
68746546	13498	09165
918324719	1789	1310
9089	91032874	1873
1	9187	19032874193
90143	01938	0193284
309132	-1349	-149089088
0202122	1234132	918324098
11234	-109234	1934

Dodatek C

Przykładowe wyniki blabalizy (ze współczynnikami σ - ρ)

Współczynniki Głombaskiego		ρ	σ	σ - ρ
γ_0	1,331	2,01	13,42	0,01
γ_1	1,331	113,01	13,42	0,01
γ_2	1,332	0,01	13,42	0,01
γ_3	1,331	51,01	13,42	0,01
γ_4	1,332	3165,01	13,42	0,01
γ_5	1,331	1,01	13,42	0,01
γ_6	1,330	0,01	13,42	0,01
γ_7	1,331	16435,01	13,42	0,01
γ_8	1,332	865336,01	13,42	0,01
γ_9	1,331	34,01	13,42	0,01
γ_{10}	1,332	7891432,01	13,42	0,01
γ_{11}	1,331	8913,01	13,42	0,01
γ_{12}	1,331	13,01	13,42	0,01
γ_{13}	1,334	789,01	13,42	0,01
γ_{14}	1,331	4897453,01	13,42	0,01
γ_{15}	1,329	783591,01	13,42	0,01

Bibliografia

- [Bea65] Juliusz Beaman, *Morbidity of the Jolly function*, *Mathematica Absurdica*, 117 (1965) 338–9.
- [Blar16] Elizjusz Blarbarucki, *O pewnych aspektach pewnych aspektów*, *Astrolog Polski*, Zeszyt 16, Warszawa 1916.
- [Fif00] Filigran Fifak, Gizbert Gryzogrzechotalski, *O blabalii fetorycznej*, *Materiały Konferencji Euroblabal* 2000.
- [Fif01] Filigran Fifak, *O fetorach σ - ρ* , *Acta Fetica*, 2001.
- [Głomb04] Gryzybór Głombaski, *Parazytonikacja blabiczna fetorów — nowa teoria wszystkiego*, Warszawa 1904.
- [Hopp96] Claude Hopper, *On some Π -hedral surfaces in quasi-quasi space*, *Omnius University Press*, 1996.
- [Leuk00] Lechosław Leukocyt, *Oval mappings ab ovo*, *Materiały Białostockiej Konferencji Hodowców Drobiu*, 2000.
- [Rozk93] Josip A. Rozkosza, *O pewnych własnościach pewnych funkcji*, *Północnopomorski Dziennik Matematyczny* 63491 (1993).
- [Spy59] Mrowclaw Spyrpt, *A matrix is a matrix is a matrix*, *Mat. Zburp.*, 91 (1959) 28–35.
- [Sri64] Rajagopalachari Sriniswamiramanathan, *Some expansions on the Flausgloten Theorem on locally congested latches*, *J. Math. Soc.*, North Bombay, 13 (1964) 72–6.
- [Whi25] Alfred N. Whitehead, Bertrand Russell, *Principia Mathematica*, Cambridge University Press, 1925.
- [Zen69] Zenon Zenon, *Użyteczne heurystyki w blabalizie*, *Młody Technik*, nr 11, 1969.