

Collaborative Recommendations (UPCV) Using Public Library Loan Data

Third Telenor-IDI Seminar, Trondheim, Dec. 12, 2013
Ville Ollikainen
VTT Technical Research Centre of Finland

“Public Library” = HelMet

- Helsinki Metropolitan Area Libraries
 - Helsinki, Espoo, **Vantaa**, Kauniainen
- 63 libraries + 14 institutional libraries
- 900 employees
- 3.400.000 items
- over 1.000.000 customers
- 17.000.000 loans / year
- 17.000.000 visits / year (5.000.000 on the net)

=> a fairly large library

This is what I have for sale.

Content Based Recommendations?

Make an apple pie; also get some donuts.



- Red
- Orange
- Green
- Round
- Rectangular
- ...
- Red
- Orange
- Green
- Round
- Rectangular
- ...

I like:

- Red
- Yellow
- Orange
- Pink
- Fuchsia
- Green
- Round
- Rectangular
- Pentagonal
- Torus
- Wedge
- ...



This is what I have for sale.

Content Based Recommendations?

Make an apple pie; also get some donuts.



- Red
- Orange
- Green
- Round
- Rectangular
- ...
- Red
- Orange
- Green
- Round
- Rectangular
- ...

Recommended

I like:

- Red
- Yellow
- Orange
- Pink
- Fuchsia
- Green
- Round
- Rectangular
- Pentagonal
- Torus
- Wedge
- ...



~~Content Based Recommendations?~~

Make an apple pie; also get some donuts.



- Red
- Orange
- Green
- Round
- Rectangular
- ...
- Red
- Orange
- Green
- Round
- Rectangular
- ...
- Red
- Orange
- Green
- Round
- Rectangular
- ...

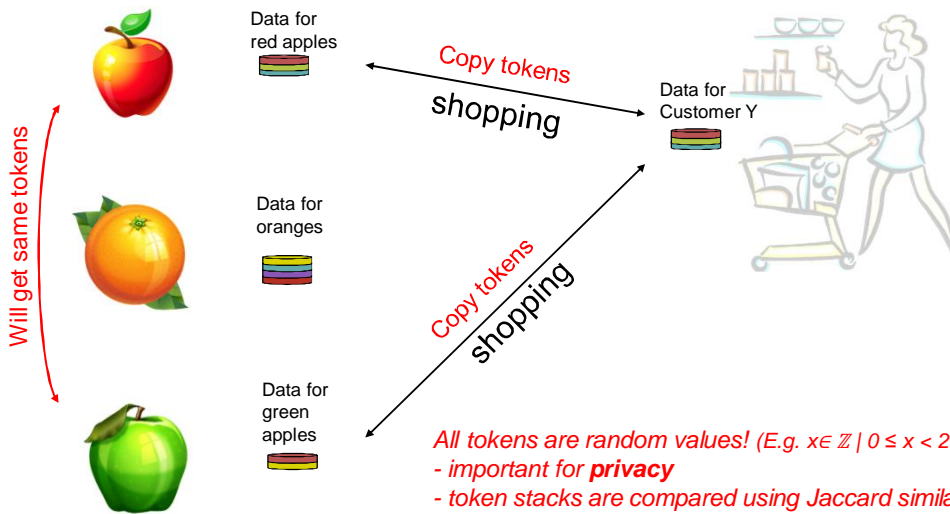
I like:

- Red
- Yellow
- Orange
- Pink
- Fuchsia
- Green
- Round
- Rectangular
- Pentagonal
- Torus
- Wedge
- ...



? **Not recommended!**

UPCV: a Novel Approach for Collaborative Filtering



HeIMet Data

- Four files
 - All books in 2 files
 - Other items in one file (mainly A/V-material, was not included in the study)
 - Transaction data from automated return system

- Loan data September 12. – October 2. 2013
 - 107,985 loan returns

- Secure processing of loan data
 - Truecrypt volume, opened only while processing

Data Preprocessing (1/2)

- Lots of inconsistencies in item data format
 - Created over time for human reading, not machine reading (e.g. "xxx11"="xxx11")

- Merging titles/book series => 420,497 **hash** values => **Item ID's**
 - Author + beginning of title (delimited by any special character) => a string
 - Cleaning the string by replacing digits by a pound sign (#) and removing spaces
 - Forming a 64-bit **hash** from the cleaned string => **Item ID**
 - E.g. numerous "Waltari,MikaSinuheegyptiläinen" titles => always the same ID

- "Waltari, Mika";"Sinuhe egyptiläinen : viisitoista kirjaa lääkäri Sinuhen elämästä n. 1390-1335 e.Kr. / Mika Waltari [kuv. Björn Landström]

- "Waltari, Mika";"Sinuhe egyptiläinen : viisitoista kirjaa lääkäri Sinuhen elämästä n. 1390-1335 e.Kr"

- "Waltari, Mika";"Sinuhe egyptiläinen : viisitoista kirjaa lääkäri Sinuhen elämästä n. 1390-1335 e.Kr. / Mika Waltari"

- "Waltari, Mika";"Sinuhe egyptiläinen. [J]: viisitoista kirjaa lääkäri Sinuhen elämästä n. 1390-1335 e.Kr. / Mika Waltari"
...och så vidare...

Data Preprocessing (2/2)

- 2,602,575 individual copies (barcodes) in book data
 - References from the barcodes to **Item ID's** (N=420,497; average ~6 copies/ID)
 - Added by **location information**: children's shelves – other shelves
 - E.g. "h56aI" -> adult, "e10II" -> child
 - (some titles had a copy in both children's shelves and other shelves)
- The barcode in loan data was replaced by the Item ID
 - **Each loan was associated with a title/book series, not to an individual copy**
- If the user loaned a book from children's shelf, the library card number was concatenated by '0', otherwise '4' => User ID
 - Adults loan books for their children, too => **different people behind one card =>** converted to **different user ID's**

Training UPCV Recommendations

- User ID – Item ID pairs were shuffled into random order
 - Temporal order was not relevant, because the loan data was from a short period
 - 15,784 User ID's
 - 24,350 Item ID's
- First half of these pairs was used for training
 - We used the same procedure for *Bonnier book club data* (January 2013)
 - Other half exists for quantitative evaluation (TBD)
- UPCV was trained with one (1) batch run without any tweaking, resulting 9,477 recommendations
 - (Also in random order)

Case Bonnier Book Club (January 2013)

| Person # | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q |
|----------|-----|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | | | | | | | | | | | | | | | | | |
| 2 | 1 | | | | | | | | | | | | | | | | |
| 3 | 2 | | | | | | | | | | | | | | | | |
| 4 | 3 | | | | | | | | | | | | | | | | |
| 5 | 4 | | | | | | | | | | | | | | | | |
| 6 | 5 | | | | | | | | | | | | | | | | |
| 7 | 6 | | | | | | | | | | | | | | | | |
| 8 | 7 | | | | | | | | | | | | | | | | |
| 9 | 8 | | | | | | | | | | | | | | | | |
| 10 | 9 | | | | | | | | | | | | | | | | |
| 11 | 10 | | | | | | | | | | | | | | | | |
| 12 | 11 | | T | | | | | | | | | X | | | | | |
| 13 | 12 | | | | | | | | | | | | | | | | |
| 14 | 13 | | | | T | | | | | | | | | | | | |
| 15 | 14 | | | | | | | | | | | | | | | | |
| 16 | 15 | | | | | | | | | | | | | | | | |
| 17 | 16 | | | | | | | | | | | | | | | | |
| 18 | 17 | | | | | | | | | | | | | | | | |
| 19 | 18 | | | | | | | | | | | | | | | | |
| 20 | 19 | | | | X | | | | | | | | | | | | |
| 21 | 20 | | | | | | | | | | | | | | | | |
| 22 | 21 | | | | | | | | | | | | | | | | |
| 23 | 22 | | | | | | | | | | | | | | | | |
| 24 | 23 | | | | | | | | | | | | | | | | |
| 25 | 24 | | | | | | | | | | | | | | | | |
| 26 | 25 | | | | T | | | | | | | | | | | | T |
| 27 | 26 | | | | | | | | | | | | | | | | |
| 28 | 27 | | X | | X | | | | | | | | | | | | |
| 29 | 28 | | | | | | | | | | | | | | | | |
| 30 | 29 | | | | | | | | | | | | | | | | |
| 31 | 30 | | | | | | | | | | | | | | | | |
| 32 | 31 | | | | | | | | | | | | | | | | |
| 33 | 32 | | | | | | | | | | | | | | | | |
| 34 | 33 | | | | | | | | | | | | | | | T | T |
| 35 | 34 | | | | | | | | | | | | | | | | |
| ... | ... | | | | | | | | | | | | | | | | |

1575 users ↓

→ 1041 books

Results

- We picked randomly seven users from the recommendation data
 - **Qualitative evaluation**... next 7 slides
- **On the top**: book loans in the training data (blue)
- **On the bottom** five recommendations (dark red)
 - If the User ID belonged to a child, only those books were recommended which had majority of copies on children's shelves, and vice versa
- Translated book titles in *italics*, my comments in brackets []
 - comments based on book descriptions, reviews, etc

User #1, ID 128354 (adult)

16 book loans in training data:

- Riikka Tanner: Linda kulkee keskellä tietä [*memoir*]
- Maija Asunta-Johnston: Verkkorakkautta varttuneille [*romantic fiction*]
- Jarkko Sipilä: Seinää vasten [*crime fiction*]
- Risto Lindstedt: Moottoripyörä ja miehen mieli : erilainen matkakirja [*travelling*]
- Ingald Andersson: Kotipihan puuterassit [*home and gardening*]
- Peter Sheridan: 47 ruusua : muistelma [*memoir*]
- Eino Niskala: Puutalon korjaus [*home and gardening*]
- Helene S. Lundberg: Näppärät säilytystilat : kauniita ja käytännöllisiä ratkaisuja kodin... [*home and gardening*]
- Anders Hallgren: Vanha koira [*pets and animals*]
- Markku Ropponen: Kuhala ja kevään ensi ruumis [*crime fiction*]
- Håkan Nesser: Ewa Morenon juttu : rikosromaani [*crime fiction*]
- Kjell Eriksson: Julmien tähtien alla [*crime fiction*]
- Pirkko Vekkieli: Kristiina Halkola : jos rakastat [*memoir*]
- Karin Alvtgen: Häpeä [*thriller*]
- Seppo Jokinen: Viha on paha vieras [*crime fiction*]
- Karen Lyager Horve: Voimaharjoittelua naisille : unelmien keho kotona harjoittelemalla [*fitness*]

| | |
|-------------------------|---|
| Reijo Mäki: | Kuoleman kapellimestari [<i>crime fiction</i>] |
| Martin Walker: | Poliisimestari Bruno [<i>crime fiction</i>] |
| Katarina Eskola: | Tyttö Pitkän sillan molemmin puolin : Elsa Eklundin tytönpäiväkirjat... [<i>memoir</i>] |
| Markku Ropponen: | Kuhala ja puhelu kiusaajalta [<i>crime fiction</i>] |
| Seppo Jokinen: | Kuka sellaista tekisi [<i>crime fiction</i>] |

User #2, ID 166080 (child)

8 book loans in training data:

- [Aku Ankan taskukirja 370]: Tukka hyvin [*Donald Duck*]
- Agatha Christie: Appointment with Death [*crime fiction*]
- Kati Jaakonen: Hellapoliisi [*cooking*]
- [Aku Anka, Taskarin tähtihetket 1]: Viihteen mestari [*Donald Duck*]
- [Aku Anka, Taskarin tähtihetket 3]: Kilpamenttien kuningas [*Donald Duck*]
- Jussi Nikkinen, Timo Rothovius, Petri Sahlström: Arvopaperisijoittaminen [*stock markets*]
- [Ankan elämää]: Pätkätöissä [*Donald Duck*]
- [Aku Ankan taskukirja esittää]: Teillä tietymättömillä [*Donald Duck*]

| | |
|--------------------------------------|---|
| [Aku Anka]: | Ihan Ankkana [<i>Donald Duck</i>] |
| [Aku Anka]: | Ankkalinnan matkaopas [<i>Donald Duck</i>] |
| [Aku Ankan teemataaskari 19]: | Kaksoiselämää [<i>Donald Duck</i>] |
| Odile Hellman-Hurpoil: | Hirmu-Riku [<i>fiction, ~11 yrs</i>] |
| Timothy Knapman: | My adventure island [<i>fiction, 5-7 yrs</i>] |

User #3, ID 264160 (child)

2 book loans in training data:

- Gösta Knutsson: Pekka Töpöhännän kolmoset [*Pelle Svanslös*]
- Gary W. Priester ja Gene Levine: Uskotko silmiäsi?: *Hidden Treasures: 3-D Stereograms* [-10y+]

| | |
|---|---|
| <i>Sinikka Nopola:</i> | Heinähattu ja Vilttitossu... [-5-8y] |
| <i>R. L. Stine:</i> | Kolme kauheaa toivetta [-10y+] |
| <i>[Aku Ankan taskukirja 401]:</i> | Tuplanolla ja uhka taivaalta [<i>Donald Duck</i>] |
| <i>Ulf Stark & Linda Bondestam:</i> | Diktatorn [-4-7y] |
| <i>Virpi Penna:</i> | Perhosten polku [-4-7y] |

User #4, ID 214440 (child)

7 loans in training data:

- Erin Hunter: *Into the Wild* [-9-15y]
- Herge: Seitsemän kristallipalloa [*The Adventures of Tintin*]
- Marja Luukkonen: Sumumetsän kasvatit [*juvenile, fantasy*]
- toim. Viia Järvinen: Aikakirjat. 3 [*witches, juvenile*]
- Merja Jalo: Vaunukello [*juvenile, a ghost story*]
- Natsumi Ando & Miyuki Kobayashi: Kitchen princess. 1 [*juvenile manga*]
- Merja Jalo: Noiduttu huone [*juvenile, a ghost story*]

| | |
|------------------------------------|---|
| <i>Kanata Konami:</i> | Chi's sweet home. 8 [<i>juvenile manga</i>] |
| <i>[Aku Ankan taskukirja 345]:</i> | Tähtihetkiä [<i>Donald Duck</i>] |
| <i>Jeff Kinney:</i> | <i>Wimpy Kid 4: Dog Days</i> [<i>cartoon fiction</i>] |
| <i>Martin Widmark:</i> | Timanttien arvoitus [<i>juvenile detective</i>] |
| <i>Natsumi Ando:</i> | Kitchen princess. 5 [<i>juvenile manga</i>] |

User #5, ID 110744 (adult)

3 book loans in training data:

- William Fischer et. al: *Prague...* [*City Map Guide*]
- Nancy Thayer: *The Moon Shell Beach* [*romantic fiction*]
- Mark Baker: *Frommer's Prague day by day* [*City Guide*]

| | |
|--|--|
| Hannu Mäkelä: | Pushkinin enkeli [<i>historical fiction</i>] |
| Ilkka Vauhkonen, Peter Holmström: | Sisätaudit [<i>medical book</i>] |
| [Puolustusvoimat]: | Suomen puolustusvoimien... [<i>military</i>] |
| Lucy Edge: | Yoga School Dropout [<i>spiritual adventure in India</i>] |
| [Marianne Mehling]: | Ateena ympäristöineen [<i>cultural travel book</i>] |

User #6, ID 187534 (adult)

3 book loans in training data:

- Tuula T. Matintupa: *Kuolema kulkee kesässä* [*thriller*]
- Agatha Christie: *Sparkling Cyanide* [*crime fiction*]
- Agatha Christie: *Ten Little Niggers* [*crime fiction*]

| | |
|-------------------------|---|
| Staffan Skott: | Romanovs [<i>history</i>] |
| Agatha Christie: | The Third Girl [<i>crime fiction</i>] |
| Tess Gerritsen: | Ice Cold [<i>crime fiction</i>] |
| Hannu Mäkelä: | Pushkinin enkeli [<i>historical fiction</i>] |
| Niina With: | Taisit narrata, Stella [<i>chick lit</i>] |

User #7, ID 133530 (child)

1 book loan in training data:

- Gunilla Bergström: *God natt Alfons Åberg* [~4-8y]

| | |
|--------------------|---|
| Debbie Martin: | <i>Animal Homes</i> [flap book w info ~2-10y] |
| Tove Appelgren: | <i>Vesta-Linnéa och gosnosen</i> [~4-8y] |
| [Walt Disney]: | Bambi |
| Gunilla Bergström: | <i>Listigt, Alfons Åberg</i> [~4-8y] |
| Sven Nordqvist: | <i>När Findus var liten och försvann</i> [~3-15y] |

Conclusions

- **It works!**
 - Same genres exist in both training data and top five recommendations, sometimes even same authors
 - **Similar results were found in Bonnier book club study**
 - Ollikainen, Mensonen, Tavakolifard... in JPMTR 2013 (in publishing process)
 - Looks like there is **no particular cold start problem**
 - Processing loan data (exchanging tokens) in milliseconds
 - Recommendations in few seconds (**coverage ~24,000 books**), 1 processor core
- Algorithm taken from *Bonnier book club* experiment as such
 - Bonnier experiment in turn inherited the algorithm from previous experiments
 - There has been lots of development since
 - ⇒ Results/ speed can be optimized further
- Discreet recommendations? "On the shelf now:"

More Information

Ville Ollikainen

Senior Scientist

ville.ollikainen@vtt.fi

Caj Södergård

Research Professor

caj.sodergard@vtt.fi

Raimo Launonen

Principal Scientist, Team Leader

raimo.launonen@vtt.fi

VTT Technical Research Centre of Finland

+358 20 722 111

P.S. Recommended Reading

WILMINGTON, Del., Dec. 10, 2013 (GLOBE NEWSWIRE) -- **InterDigital**, Inc. (Nasdaq: **IDCC** [[FREE Stock Trend Analysis](#)]), a wireless research and development company, today announced that its Innovation Partners unit has signed a collaboration agreement with VTT Technical Research Centre of Finland, a multi-technological applied research organization in Northern Europe.

In today's hyper-connected world, collection of personal information by merchants, devices, advertisers and technology ecosystem players has become so pervasive and subtle that many users do not even realize that it is happening. Data has become the fuel of the digital gold rush and it sits at



VTT luo teknologiasta liiketoimintaa