Databases and Information Retrieval Integration

TIETS42

Recommendations beyond the Ratings Matrix

Kostas Stefanidis
kostas.stefanidis@uta.fi

Autumn 2016

http://www.uta.fi/sis/tie/dbir/index.html

http://people.uta.fi/~kostas.stefanidis/dbir16/dbir16-main.html
Traditional recommendations

- Only (numerical) ratings

- Recently, numerical ratings & text reviews

![Diagram with ratings and reviews for different users and movies.]
Traditional/ in-the-box recommendations

- The recommender works upon the ratings matrix

- Existing recommenders are closed
  - They profile the user within their application, e.g., Amazon, Facebook.
  - Recommended items are from the available options within their application.

- The user is the only curator of information
  - She explicitly provides ratings/reviews to the system
  - The amount and quality of the input data depends on user’s engagement

- Sparsity problem
  - Users typically rate only a few items

- Cold start problem
  - Dealing with new users

The focus is on the business, not the user!!!
Getting out of the box

- Moving from a passive approach (i.e., waiting for the user to actively participate in the application) to an active approach (detect user/item/ratings traces in the Web)
  - Users follow different paths to fulfill their needs (e.g., Amazon for products, Spotify for music, Foursquare for going out, Google news for news, ...)
  - A more complete user profile can be created by combing information from all these sources.

- In the big data era, we can exploit data *implicitly* given in the context of social media/Web interactions for recommendations → out of the box
  - Allows for better/holistic user profiling
  - Helps solving traditional recommendation issues like sparsity and cold start
  - Generates business value for the involved parties
The envisioned out-of-the-box recommender

Users

Ratings

Concert1  Movie1  Bar1

Anna  0.9  0.8  0.7

Joe  0.9  0.7

Eric  0.9  0.6

Kai  0.7

Items

Web

GeoNames  yago*  G+

DBpedia  Twitter  Freebase

Information Enhancement
Challenges for out-of-the-box recommenders

- Information hunting
  - Data enrichment (What to look for and how)
  - Data integration (How to put everything together)

- Recommendations revisited
  - Cross-domain recommendations
  - Lifelong recommendations and learning
  - Interactive exploration
Goal: Enriching existing data in the recommender and filling missing data through external sources like the Web

All core entities in a recommender can be enriched
- Users
- Items
- Ratings
Information hunting: What to look for and how?

- **Goal:** *Enriching existing* data in the recommender and *filling missing* data through external sources like the Web

- All entities can be enriched
  - Users
    - A user expresses herself differently in different networks
  - Holistic profiling from different social networks

---

**Daniel Lemire**

Autodesign with a Ph.D. who loves two of these three things: computer science, programming and sex on the beach.

Montreal, Canada

Joined November 2007

*Twitter* @lemire

**Daniel Lemire's blog**

Daniel Lemire is a computer science professor at the University of Quebec. His research is focused on software performance and indexing. He is a technocritic.

The genetics of health

*twitter.com/lemire* | *github.com/lemire* | *github.com/DanielLemire* | *linkedIn.com/dlemire*

You can also find me on Twitter as *lemire*, on *GitHub*, on *LinkedIn* and on *Google Scholar*. You can subscribe to the blog by email.

**We are passing the Turing test right on schedule**

In 1950, the brilliant computing pioneer Alan Turing made the famous prediction in his paper *Computing Machinery and Intelligence*:

> I believe that in about fifty years' time it will be possible, to programme computers, with a storage capacity of about $10^9$, to make them play the imitation game so well that an average interrogator will not have more than 70 per cent chance of making the right identification after five minutes of conversation.

---

**Daniel Lemire**

Computer Scientist

Montreal, Canada Area | Research

Current: LICIT Research Center (TELUS-Université du Québec, University of New Brunswick)

Previous: National Research Council of Canada, Acadia University, TechElements Inc.

Education: Université de Montréal - École Polytechnique de Montréal

*LinkedIn.com/dlemire* | *500+ connections*
Goal: Enriching existing data in the recommender and filling missing data through external sources like the Web

All entities can be enriched

- **Items**
  - Enhance item description through information collected from Web pages, ontologies, ...
  - E.g., Wikipedia, Library Thing (for books), IMDB (for movies, actors)

This information can be also used for computing item similarities
Information hunting: What to look for and how?

- Goal: Enriching existing data in the recommender and filling missing data through external sources like the Web
- All entities can be enriched
  - Ratings
    - Trace online references of users to items
      - E.g., Twitter, Facebook, Youtube, ...

---

Dana Hamann @DanHam2 · May 19
Highly recommend reading the alchemist if you're looking for a good book. Inspiring and speaks to your soul! #lifechanging

MagBish Manju
Completed reading The Alchemist for 13 times...both in english and malayalam...so exciting...so motivating....so inspiring....
I realise y i was born.... See More

Kevin James 2 months ago
I don't understand why a lot people hate this song. This song is just great. Pure!

Sleeves were too small as compared to the rest of the dress
By A Happy Mom on May 12, 2016
Disappointed in this dress. 1. For the price, the material/quality isn't what I was expecting. 2. The dress fit fine but the sleeves were extra tight, it seemed like there was a mistake in the way the sleeves were made. 3. There is a note on the dress saying the vibrant colors will fade after 20 hours exposure to the sun.
Data integration: How to put everything together?

- Which pieces of information refer to the same entity?
- Entity resolution aims to identify different descriptions that refer to the same entity
- Numerous knowledge bases on the Web offer comprehensive, machine-readable descriptions of a large variety of real-world entities
  - e.g., DBpedia (extract structured information from Wikipedia)
    - 4.58M things (1.445M persons, 735K places, 411K creative works, 241K organizations, 251K species, 6K diseases)
  - e.g., YAGO (derived from Wikipedia WordNet and GeoNames)
    - >10M entities (like persons, organizations, cities, etc.) and >120M facts about these entities.
  - The size and diversity of the Web offers new challenges for entity resolution
Recommendations revisited

- The connection of a recommender system to the Web offers endless possibilities for data enrichment at all levels (users, items, ratings).

- How can we exploit all this wealth of information?
  - Better recommendations
  - Cross-domain recommendations
  - Lifelong user profiling and learning
  - ...
  - Data/Information as an asset for the company
  - Cross-company recommendations and co-operations

*Data! Data! Data! I can’t make bricks without clay!*  
-Sherlock Holmes, *The Adventure of the Copper Beeches*
Traditionally, single domain recommendations, e.g., only movies.

Cross-domain recommendations: recommendations on items from different domains, e.g., movies & books, job vacancies & educational opportunities.

- Packet recommendations: a central item from the main domain of interest and a set of satellite items possible from different domains.

Cross-company recommendations: Items can belong to different systems, warehouses

- e.g., Polyvore, Zalando
Lifelong recommendations and learning

- Keeping track of user’s history, within and out of the recommender, offers many opportunities for lifelong user profiling and learning
  - Identify changes in habits and tastes
    - Gradual changes over time (concept drifts), e.g. due to ageing.
    - Abrupt changes (concept shifts), e.g., due to moving, expanding the family.
  - Allows for
    - Better recommendations
    - Detecting periodicities in user preferences
    - Contextual recommendations
“There is not enough data to profile a user/item“ is not the case anymore.

There is a huge amount of recommendation-related data in the Web, which calls for reshaping of recommenders.

Data can be enriched from many, diverse sources and at every level allowing for:

- Better recommendations
- Better user profiling
- Putting the emphasis on the user needs

New business opportunities:

- Cross-domain recommendations
- Cross-company recommendations
- Lifelong-recommendations and user learning

Many opportunities, but huge challenges for the research/business domains.
Questions?