

COMPUT 691: Doctoral Comprehensive Examination

An Overview on Review-Based Recommendations

Presented By: Amifa Raj

Supervisor: Michael Ekstrand

Committee Members: Sole Pera,
Casey Kennington, Edoardo
Serra, Leming Qu



BOISE STATE UNIVERSITY

Objectives

- Understand the key concept of review-based recommendations
- Introduce the tasks of recommendations using reviews
- Introduce the common approaches
- Address the limitations
- Identify future works
- Present artifact

Trending Now



Top 10 in the U.S. Today



Netflix Originals



Home

Following

Search

57



auto-rickshaw-wall-painting



People also search for

View



Avengers: Endgame 2019



Avengers: Infinity War 2018

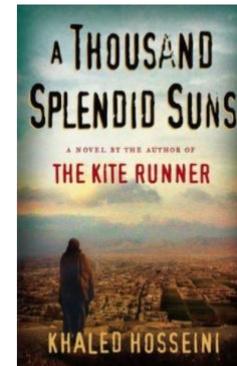
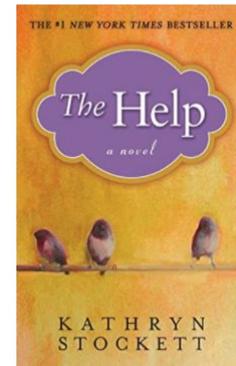


Avengers: Age of Ultron 2015

goodreads

Search books

YOUR INTEREST IN HISTORICAL FICTION



Frequently bought together



This item LEGO Star Wars: The Rise of Skywalker Resistance A Wing Starfighter 75248 Advanced Collectible Starship Model Building Kit (269 Pieces)

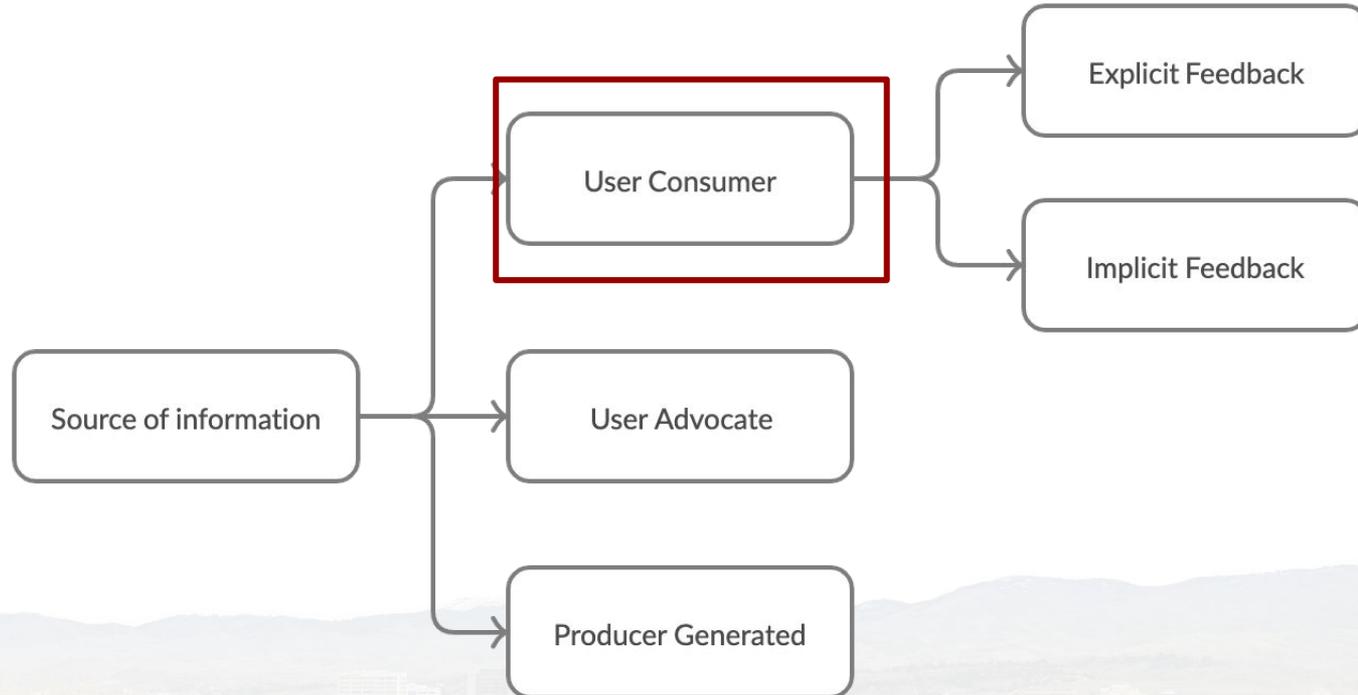
LEGO Star Wars: The Rise of Skywalker Resistance Y-Wing Starfighter 75249 New Advanced Collectible Starship Model Building Kit (578 Pieces)

Compare with similar items

What is Recommender System?

- Recommender Systems are algorithms that suggests **relevant items** to users by **inferring** their information need.
- The purpose of this process is to reduce **information overload** and help users in **decision making process**.

Taxonomy of Source of Item Feature Information



User Consumer

ratings

tags

 **Rory** rated it ★★★★★
Shelves: anti-capitalism, female-author, feminism, nonfiction, politics, social-justice
May 30, 2020
Mostly things I already knew. But many new ideas. Good to have it all in one place

 **Vassiki Chauhan** rated it ★★★★★
Jun 13, 2020
Best book ever. The antidote to the times we live in, transformative reflections on how science can and should be done.

 **Joli Holmes** rated it ★★★★★
Apr 29, 2020
A must-read for anyone interested in or currently pursuing data science. Should be a foundational read for any data ethics class.

Want to Read
Read
Currently Reading
Add shelf

reviews

Explicit Feedback

Implicit Feedback

Screenshots are taken from <https://www.goodreads.com/>

Producer

Item Metadata

Data Feminism

by Catherine D'Ignazio,
Lauren F. Klein (Goodreads Author)

★★★★★ 4.58 · Rating details · 78 ratings · 14 reviews

A new way of thinking about data science and data ethics that is informed by the ideas of intersectional feminism.

Today, data science is a form of power. It has been used to expose injustice, improve health outcomes, and topple governments. But it has also been used to discriminate, police, and surveil. This potential for good, on the one hand, and harm, on the other, make ...more

GET A COPY

Kindle Store \$18.99 Amazon Stores Libraries

Hardcover, 328 pages
Published March 10th 2020 by MIT Press (first published February 21st 2020)

Original Title Data Feminism
ISBN 0262044005 (ISBN13: 9780262044004)
Other Editions (1)



User Advocate

Home My Books Browse Community Search books

Add a New Book

Note: Goodreads has over 12 million books in its database already, so please do a search before adding a book, as it may be a duplicate. Please also carefully read the guidelines to the right, especially the part about what kind of books to add.

title *
sort by title *
author * Add Role
Add new author
isbn isbn 13 Click for ASIN
publisher
published year: month: day:
number of pages
format Other
edition
description

Categorization

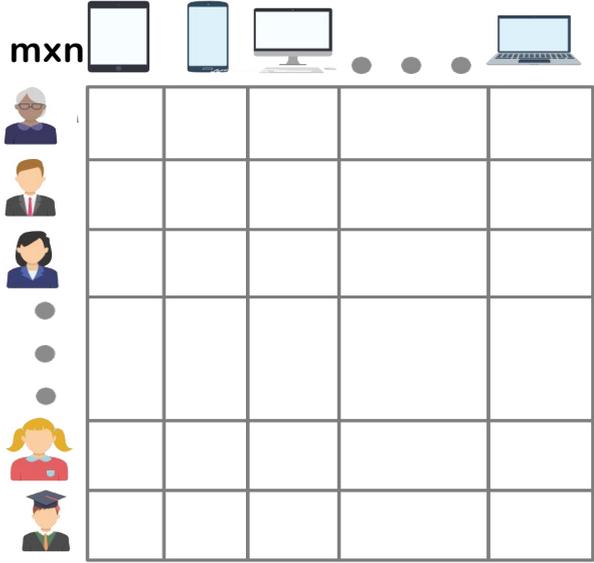
Approaches based on data information

- Collaborative filtering
- Content-based
- Demographic
- Knowledge-based
- Utility-Based
- Hybrid

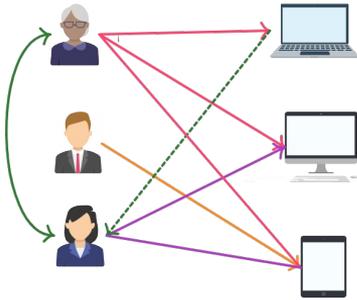
- Personalized
- Non-Personalized

- Top-K
- Prediction

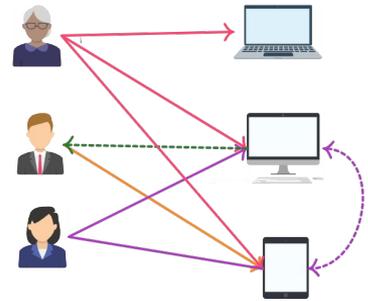
Collaborative Filtering



User-Item Interaction matrix



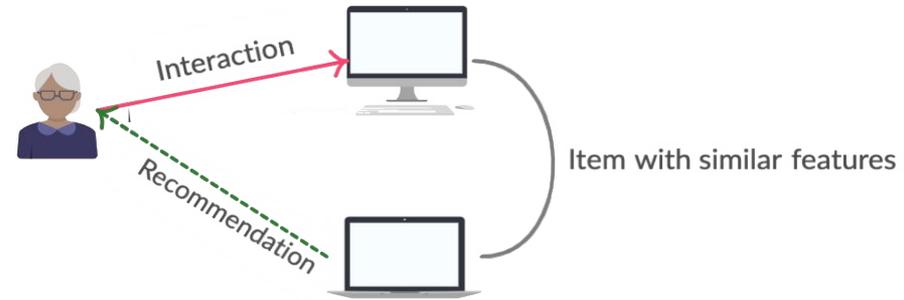
User-User CF



Item-Item CF

Content-based

	Description	Keyboard	...
			-
	...		-
			-
		✓	-
		✓	-



Common Issues



Data sparsity



User Cold Start
(new user)

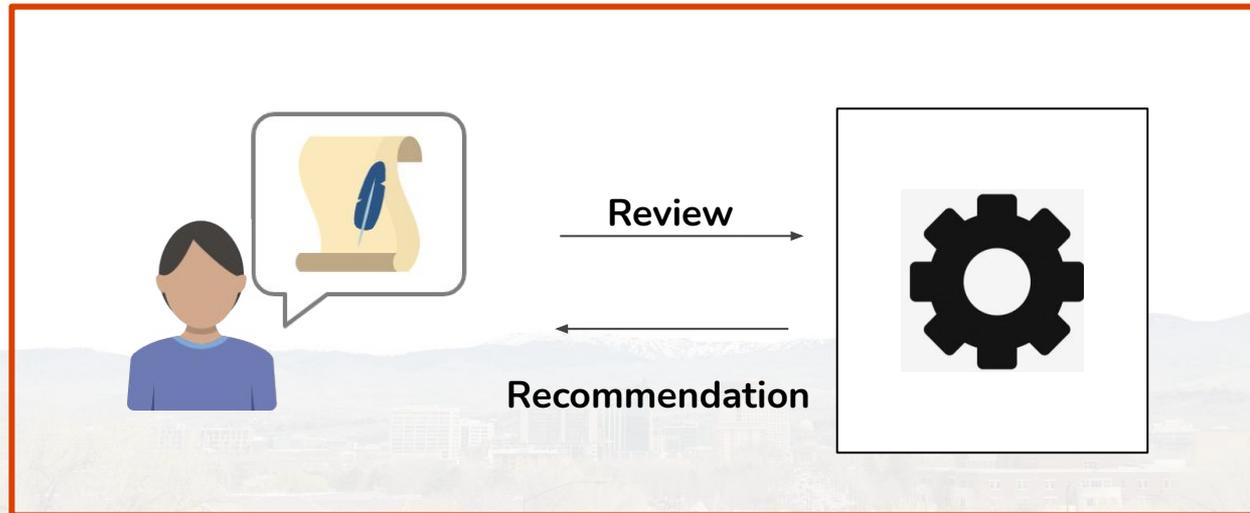


Item Cold Start
(new item)

Cold Start

Review-Based Recommender System

- Unstructured textual representation of user opinion
- Rich source of information
- Captures user preference and emotion for items



Advantages of Using Reviews

- Enriching user profile
- Enriching item profile
- Alleviating rating sparsity problem

I believe I am relatively familiar with **history** in general, and I'm usually not very excited about reading more about it. But this book was something else. **Beautifully written** and **easy to read**, this book just made me want to know more and more about how the author thinks the **world evolved** to what it is today. Revolution by revolution, religion by religion, conception by conception, things were simplified and yet still maintained valid points - and it was **never boring**.



User Profile



<..history, author,...>



<..,history,easy to read, written beautifully, revolution,...>

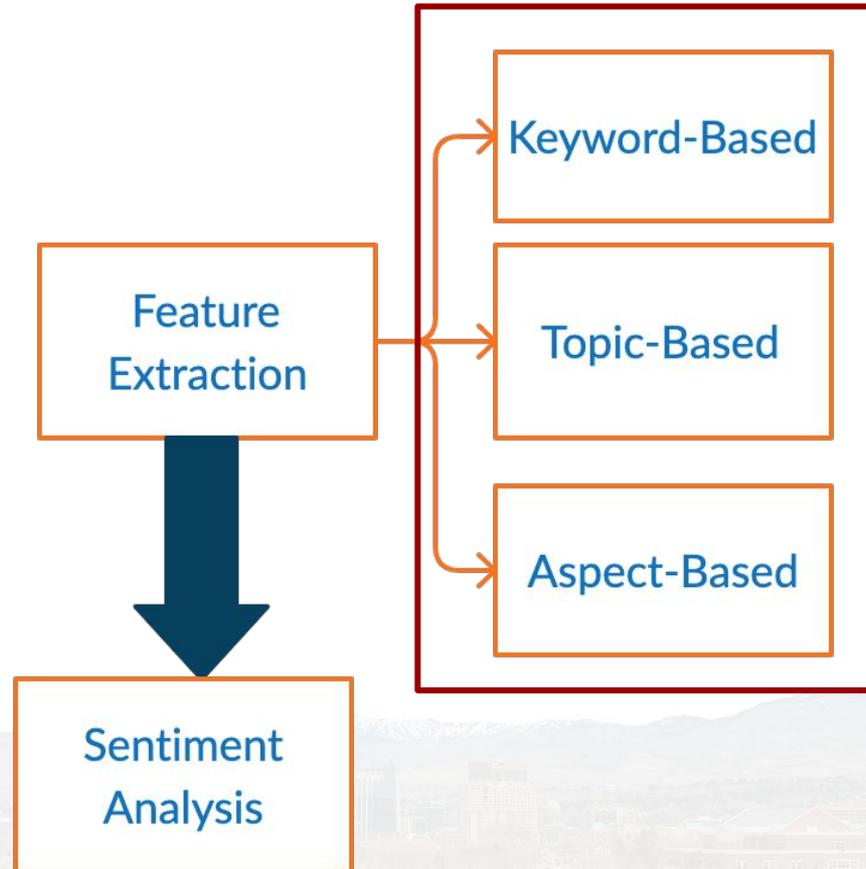
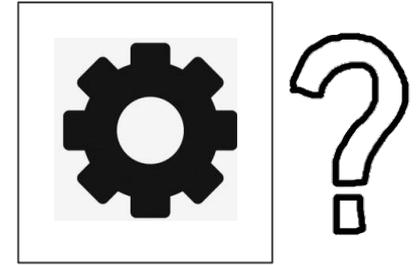
Item Profile



Tasks of Review-Based Recommender System

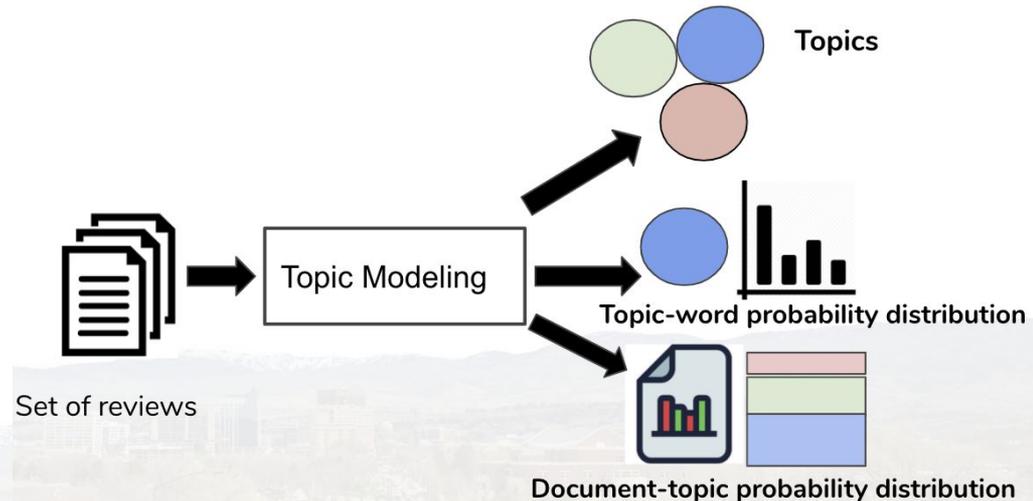
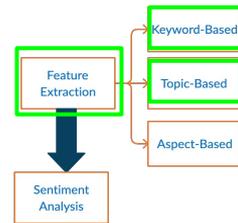
- Rating Prediction Task
 - Infer ratings from reviews
 - Predict user rating for unseen items
- Feature-based profile building
 - User profile building
 - Item profile building
- Determine Review Helpfulness

Opinion Mining Techniques

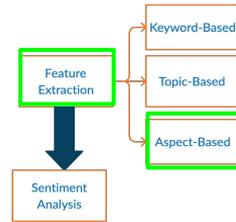


Feature Extraction

- Keyword-Based Feature Extraction
 - Term Frequency-Inverse Document Frequency
 - BM25
- Topic-Based Feature Extraction
 - Latent Dirichlet Allocation
 - PLSI
 - Latent Semantic Analysis



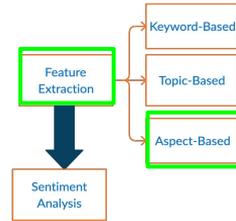
Aspect-Based Feature Extraction



This was great stuff in my opinion. Big blockbuster entertainment with a great sense of humour, a great **cast** and awesome action **sequences**. The impressive thing to me about this was they had so many **characters** here and yet managed to bring them altogether as a team perfectly. The **film** is really well paced, full of good one liners and has heart and the **plot** and **script** are stronger than I thought they would be. This is 4 star entertainment for the whole family

Aspects	Features
Characters	Character, cast
script	set
theme	plot

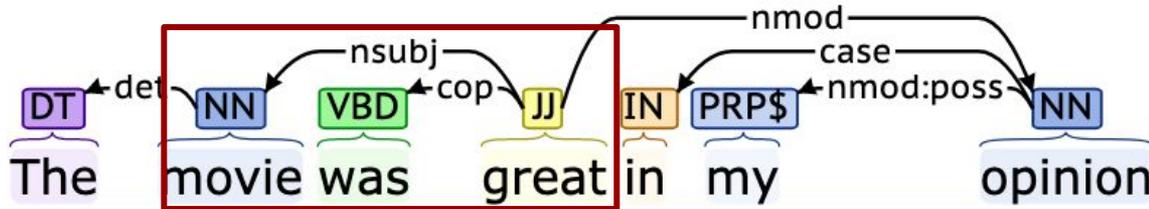
Vocabulary-based (VOC)



POS tagging



Basic Dependency

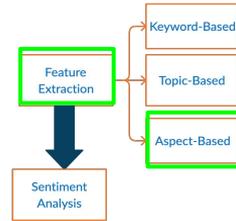


Aspect Opinion tuples (noun, adjective, modifier, isAffirmative)

(movie, great, -, true)

Frequency-Based: SABRE (sab)

- Nouns with highest frequency are considered as aspect
- Two nouns with common lemma are considered to be same
- Domain Dependant



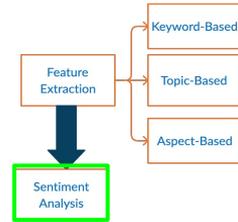
Double Propagation (dp)

- Analyze syntactic relationship (gov,rel,dep) between nouns and adjective iteratively
- Identify aspect and opinion words based on previous iteration
- Starts with a list of known aspect and opinion words

Topic-model-based (lda)

- Latent topic representation of items

Sentiment Analysis



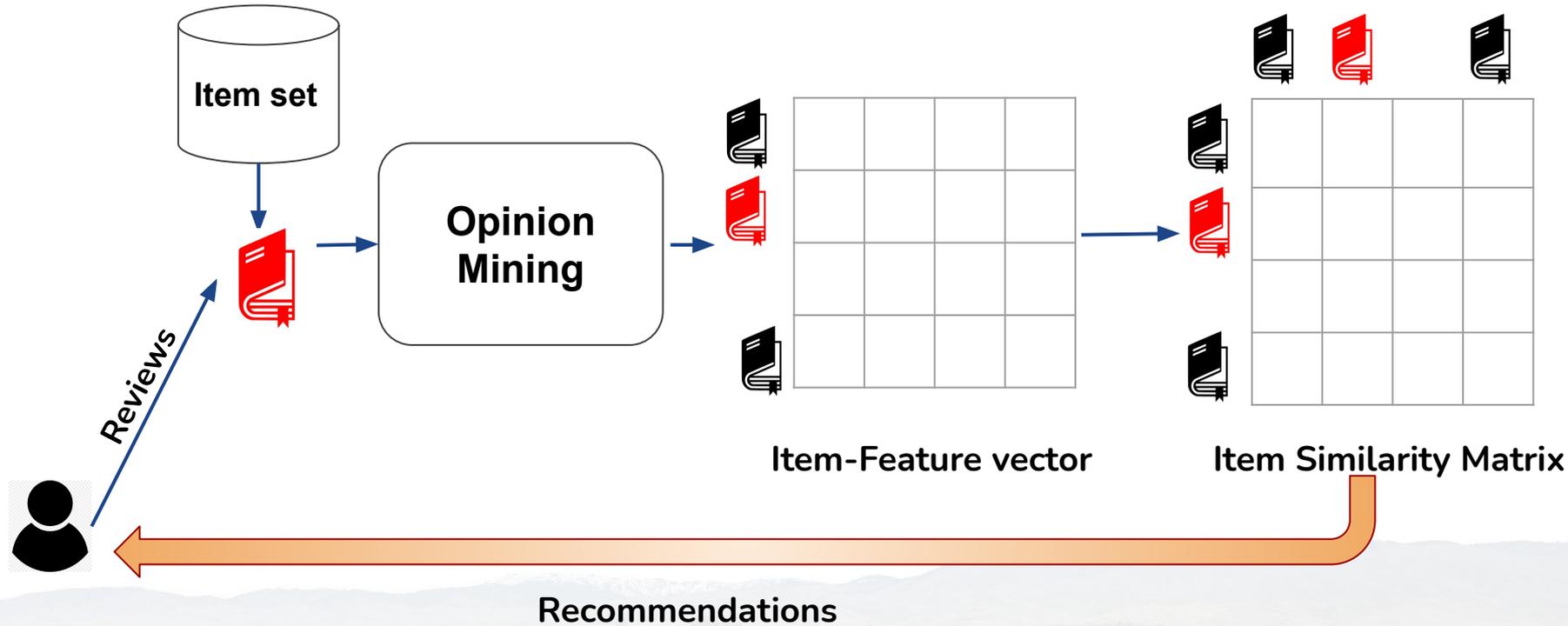
Determining polarity of extracted opinion words:

- Semantic Orientation CaLculator (SO-CAL)
- SentiWordNet : sentiment lexicon analyzer
- SenticNet
- WordNet Affect

The movie was **great** in my opinion.

Adjective/ Opinion word	Polarity/Sentiment Score
great	Positive; +1

Recommendation



Limitations

Customer Reviews

There are no customer reviews yet.

5 star	<input type="text"/>
4 star	<input type="text"/>
3 star	<input type="text"/>
2 star	<input type="text"/>
1 star	<input type="text"/>



Fake Reviews

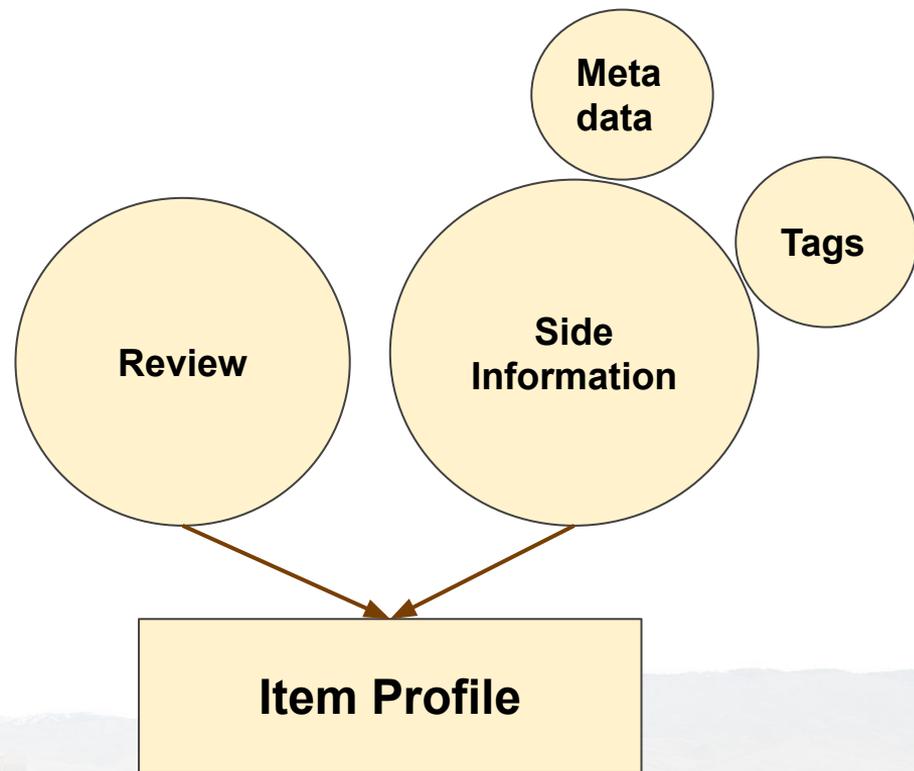


Serendipity

Review Sparsity

Open Problem 1: Combination of Reviews with Other Information

- Alleviate Data Sparsity
- Mitigate Cold Start problems



Open Problem 2: Bias in User Reviews

stereotyping



The New York Times
Book Review
JANUARY 7, 2006



The images are from <https://psmag.com/social-justice/gender-bias-in-book-reviews>,
<http://zalgocometh.blogspot.com/2011/04/like-omg-stereotypes-are-so.html>,
<https://www.nytimes.com/subscription/info/book-review?mode=BookReviewHome>

Open Problem 3: Feature Extraction from Reviews



Reviews in different language



Not following Grammar I am

Summary

- Classification of Recommender System
- Taxonomy of Item Information Sources as Input Data in Recommendations
- Tasks of Review Based Recommender Systems
- Advantages of Using Reviews in Recommendations
- Opinion Mining Techniques
- Limitations of Reviews-Based Recommender Systems
- Future Direction

Artifact

- Implemented
 - TF-IDF
 - LDA
- Recommendation algorithms used to compare
 - User-USER
 - Item-Item
 - BPR (implicit)
 - ALS
 - Popular

Experimental Setup

Dataset

- Steam Review Dataset (user_id, user_url, reviews)

{{'funny': '', 'posted': 'Posted November 5, 2011.', 'last_edited': '', 'item_id': '1250', 'helpful': 'No ratings yet', 'recommend': True, **'review': 'Simple yet with great replayability. In my opinion does "zombie" hordes and team work better than left 4 dead plus has a global leveling system. Alot of down to earth "zombie" splattering fun for the whole family. Amazed this sort of FPS is so rare.'**}

- User-item dataset (user_id, items_count, user_url, items)

#user	#item	#reviews
25458	3682	55313

Experimental Setup

Requirements

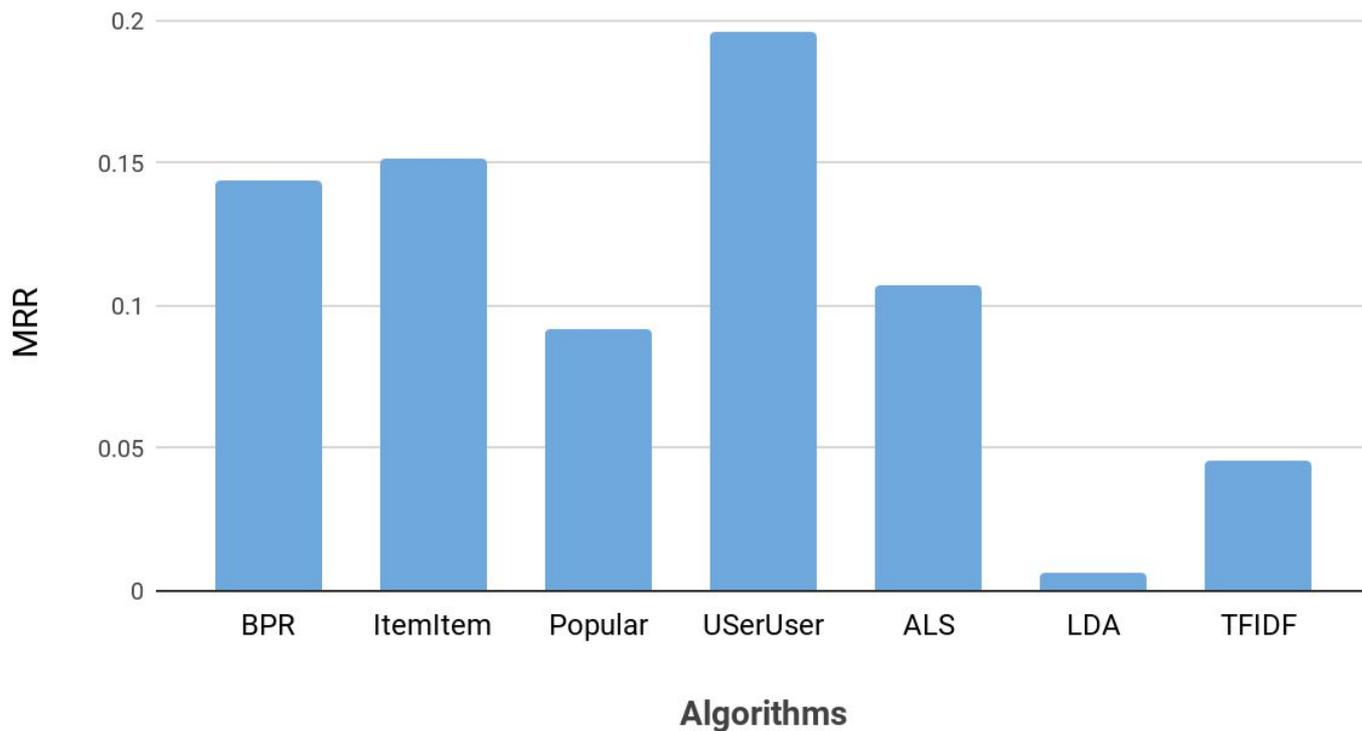
- python
- numpy
- pandas
- nltk
- scikit-learn
- scipy
- matplotlib
- lenskit

Metrics

- Recall
- Mean Reciprocal Rank

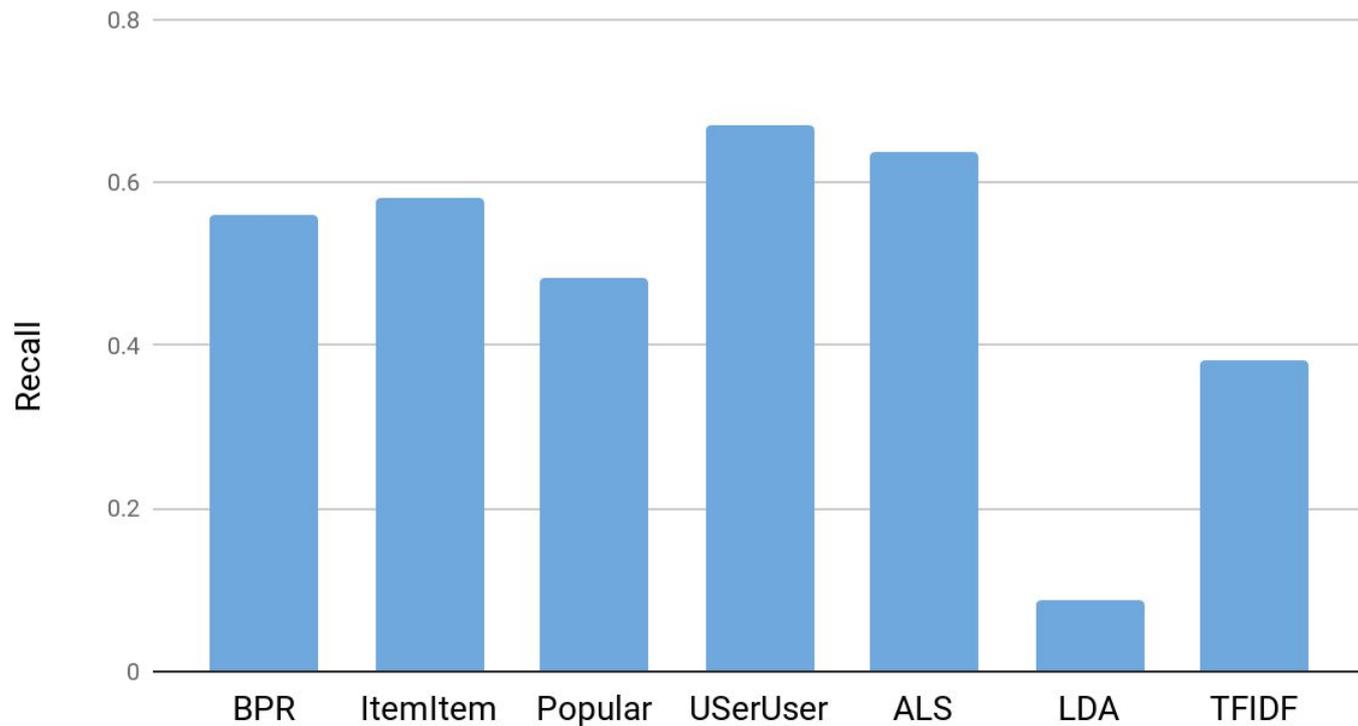
Results

MRR by Algorithms



Results

Recall by algorithms



Discussions

- TF-IDF
 - All extracted keywords were used as features
 - Exact pre-processed word matching
- LDA
 - Topics are unknown
 - words were repeated within the topics

What's Next?

- Implementing more approaches
- Identifying bias in reviews
- Combining reviews with other information

Questions?

