Graduate Certificate in Big Data Analytics

Introduction to Recommender Systems

Dr. Barry Shepherd
Institute of Systems Science
National University of Singapore
Email: barryshepherd@nus.edu.sg

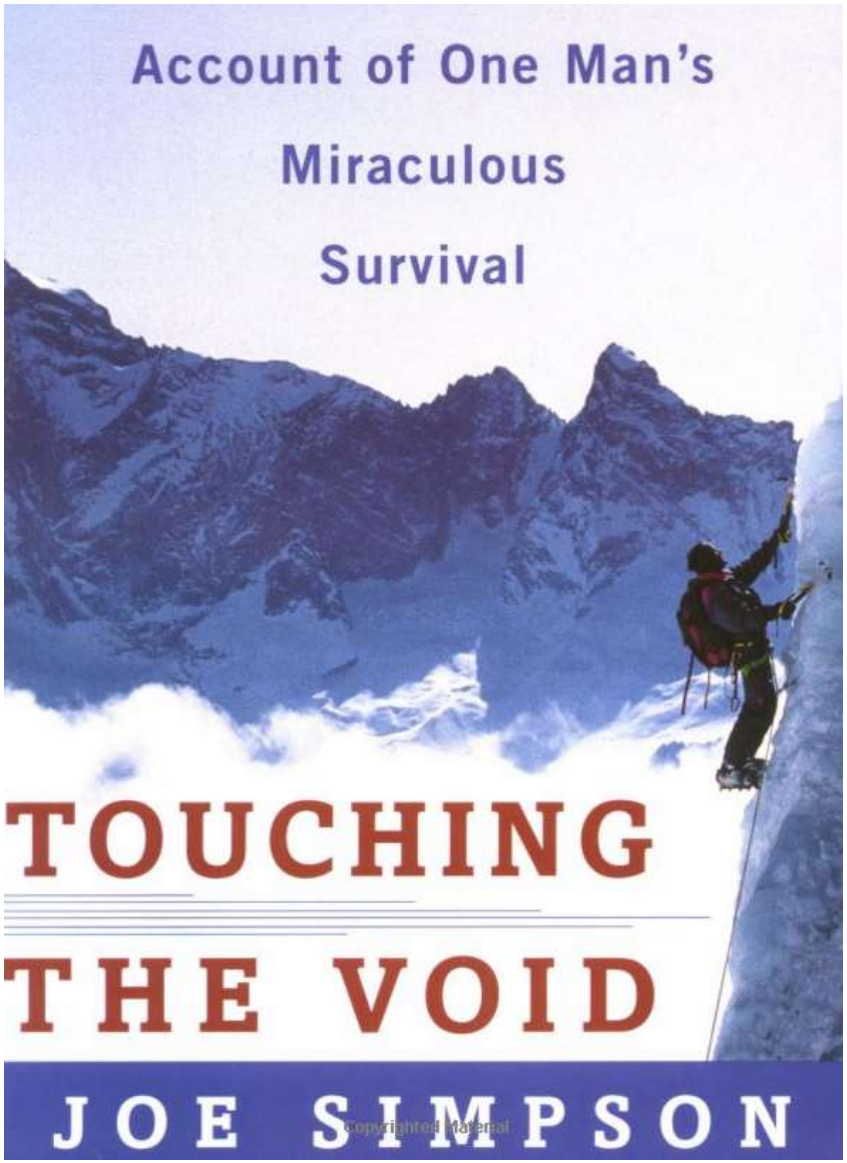
Course Agenda

- Day1
  - Introduction to Recommender Systems
  - Recommender Systems using Association Mining
  - Recommender Systems using Collaborative Filtering
- Day2
  - Recommender Systems and Matrix Factorisation
  - Content-based Recommender Systems
- Day3
  - Handling implicit feedback
  - Advanced & Hybrid Recommender Systems
  - Course conclusion
  - Class quiz

Why Recommender Systems?

In 1988, a British mountain climber named Joe Simpson wrote a book called **Touching the Void**, a harrowing account of near death in the Peruvian Andes. It got good reviews but, only a modest success, it was soon forgotten. Then, a decade later, a strange thing happened. Jon Krakauer wrote **Into Thin Air**, another book about a mountain-climbing tragedy, which became a publishing sensation. Suddenly Touching the Void started to sell again.

WHAT HAPPENED?



Why Recommender Systems?

What happened? In short, Amazon.com recommendations.

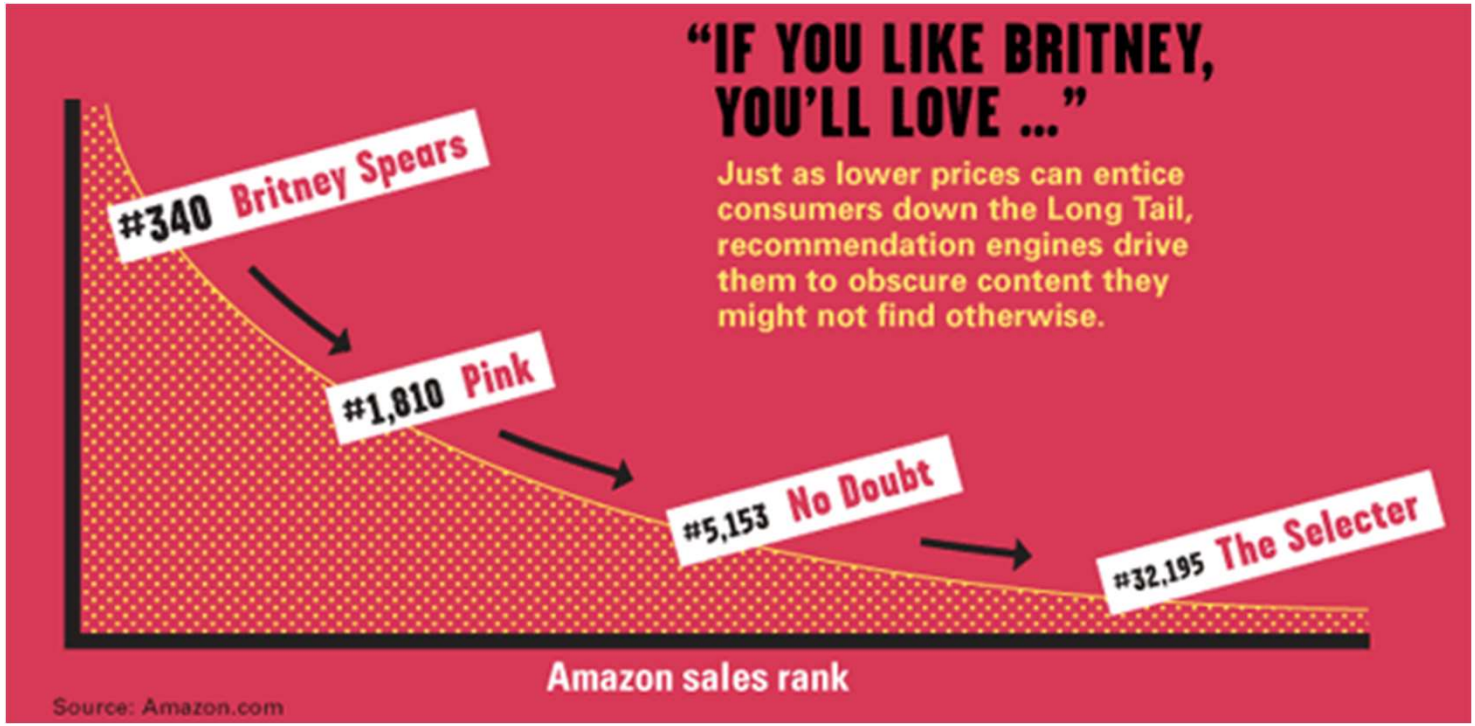
The online bookseller's software noted patterns in buying behavior and suggested that readers who liked "Into Thin Air" ***would also like*** "Touching the Void".

People took the suggestion, agreed wholeheartedly, wrote rhapsodic reviews. More sales, more algorithm-fueled recommendations, and the positive feedback loop kicked in.

When Krakauer's book hit the shelves, Simpson's was nearly out of print. Now Touching the Void outsells Into Thin Air more than two to one.

The Long Tail in E-Retail

- E-retailers can hold huge inventories without needing big (expensive) shops to display them. The result is a large number of products that only sell a small amount each – the long tail!
- How to advertise and market the log tail?



Read the full article at: http://www.wired.com/2004/10/tail

E.g. Amazon.com

12 million products

Amazon sells more than 12 million **products**.

In its quest to be all things to all people, **Amazon** has built an unbelievable catalog of more than 12 million **products**, books, media, wine, and services. If you expand this to **Amazon** Marketplace sellers, as well, the number is closer to more than 350 million **products**.

http://rejoiner.com/resources/amazon-recommendations-secret-selling-online/

Each month more than 197 million people around the world get on their devices and visit Amazon.com.

That's more than the entire population of Russia.

https://www.bigcommerce.com/blog/amazon-statistics/

"Judging by Amazon's success, the recommendation system works. The company reported a 29% sales increase to $12.83 billion during its second fiscal quarter, up from $9.9 billion during the same time last year. A lot of that growth arguably has to do with ***the way Amazon has integrated recommendations*** *into nearly every part of the purchasing process..."*

amazon

E.g. Netflix

How important is Netflix's Recommender System?

80% of stream time is achieved through Netflix's recommender system, which is a highly impressive number. Moreover, Netflix believes in creating a user experience that will seek to improve retention rate, which in turn translates to savings on customer acquisition (estimated $1B per year as of 2016).

https://towardsdatascience.com/deep-dive-into-netflixs-recommender-system-341806ae3b48

Launched in 2006 with a $1 million prize

- *Data* = 100 million movie ratings from 500,000 anonymous users on more than 17,000 movies.
- *Ratings* were on a scale 1 to 5
- *Data format* = (user, movie, rating, time)
- *Goal* = predict user ratings on a test set of 3 million
- *Winner* = first team that can beat their existing recommendation system performance by 10% (as measure by RMSE, root mean square error)

Congratulations!

The Netflix Prize sought to substantially improve the accuracy of predictions about how much someone is going to enjoy a movie based on their movie preferences.

On September 21, 2009 we awarded the $1M Grand Prize to team "BellKor's Pragmatic Chaos". Read about their algorithm, checkout team scores on the Leaderboard, and join the discussions on the Forum.

HOW WAS THE COMPETITION WON?

What is a Recommender System?

Customers who bought this item also bought

Core Java Volume I--Fundamentals (10th Edition) (Core Series)
Cay S. Horstmann
105
Paperback
$33.59 prime

Core Java for the Impatient
› Cay S. Horstmann
24
Paperback
$30.65 prime

Java 8 in Action: Lambdas, Streams, and functional-style programming
› Raoul-Gabriel Urma
80
Paperback
$36.72 prime

Effective Java (2nd Edition)
› Joshua Bloch
233
Paperback
$41.79 prime

Java Concurren
Practice
› Brian Goetz
Paperback
$31.23 prime

What is a Recommender System?

Unpersonalised

A Common Definition

- A Recommender System aims to find and suggest items of likely interest based on the users' preferences
- Most modern Web apps have a recommender system
- Examples:
  - Netflix: TV shows and movies
  - Amazon: products
  - LinkedIn: jobs and colleagues
  - Last.fm: music artists and tracks
  - Facebook: friends

Scoping Recommender Systems

The phrase **Recommender System** is very broad – it *could be* used to describe a very wide range of applications, including common predictive modelling applications. (e.g. replace the word *predicting* with the word *recommending*!)

For example:

- Recommending who to mail ads/promotions to (targeted marketing)
- Recommending the best drug or treatment for a patient
- Recommending the best business partner
- Recommending if a device/equipment needs preventative maintenance (T/F)
- Recommending which insurance claim to investigate further (possible fraud)

One distinguishing feature is the number of possible recommendations that can be made:

Predictive modelling approaches fit well (e.g. logistic regression)

"Recommender System"

Scoping Recommender Systems

- Recommending from a small set of products
  - E.g. A bank wanting to promote its various insurance products (home, car, health, travel) to appropriate customers
  - Typically done by customer segmentation and/or building a customised targeting model for each product
  - Data rich: usually have customer account details + their transaction and activity records

- Recommending from a large inventory of products
  - E.g. On-line retailers like Amazon, Lazada, online newspapers and other content providers (their products are the various webpages, news articles etc)
  - Data poor: often users are anonymous, e.g. have only info about their current browsing session

Predicting User Preferences

*A **recommender system** or a **recommendation system** is a subclass of information filtering system that seeks to predict the "rating" or "preference" a user would give to an item (Wikipedia)*

Data Sources For Recommender Systems

- **Items** ~ the objects to be recommended and their properties
- **Users** ~ their demographics, likes, preferences, goals (direct or inferred)
- **Ratings** ~ interactions between the user and the RS
  - ***Explicit Ratings*** – the user provides an opinion on an item using a rating scale, e.g. 1->5 (numerical) or "agree", "neutral", "disagree" etc. (ordinal)
  - ***Implicit ratings*** – derived from user actions (e.g. views, likes, buys)
  - ***Text comments*** - made by the user about the products
  - ***Tags*** – user assigns a tag to items, e.g. movie is "too long", "bad acting", etc.
- **Context**
  - Mostly used to filter recommendations
  - E.g. Is user a beginner or advanced camera user?
  - E.g. Is user looking for restaurants near their current location?

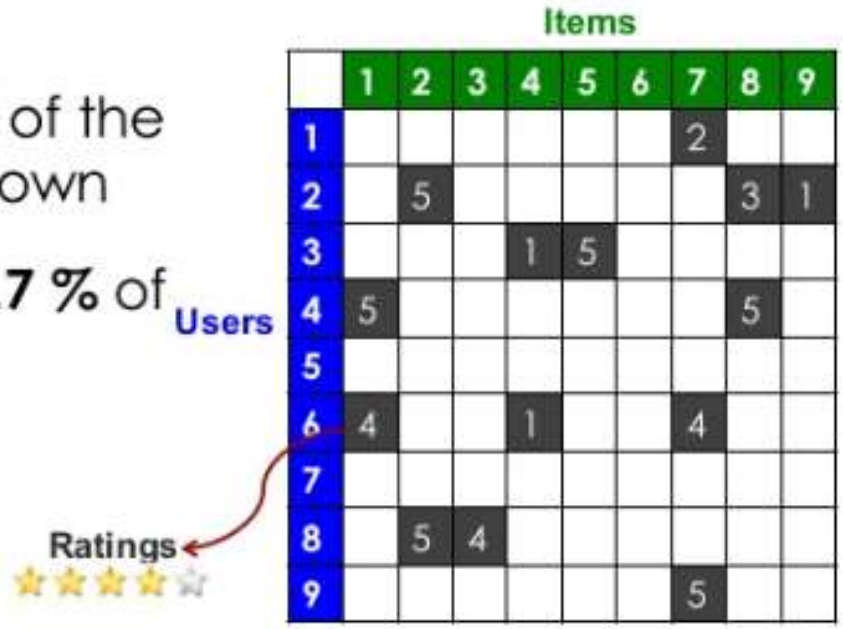
Explicit ratings are high value

Issues: Cold Start & Data Sparsity

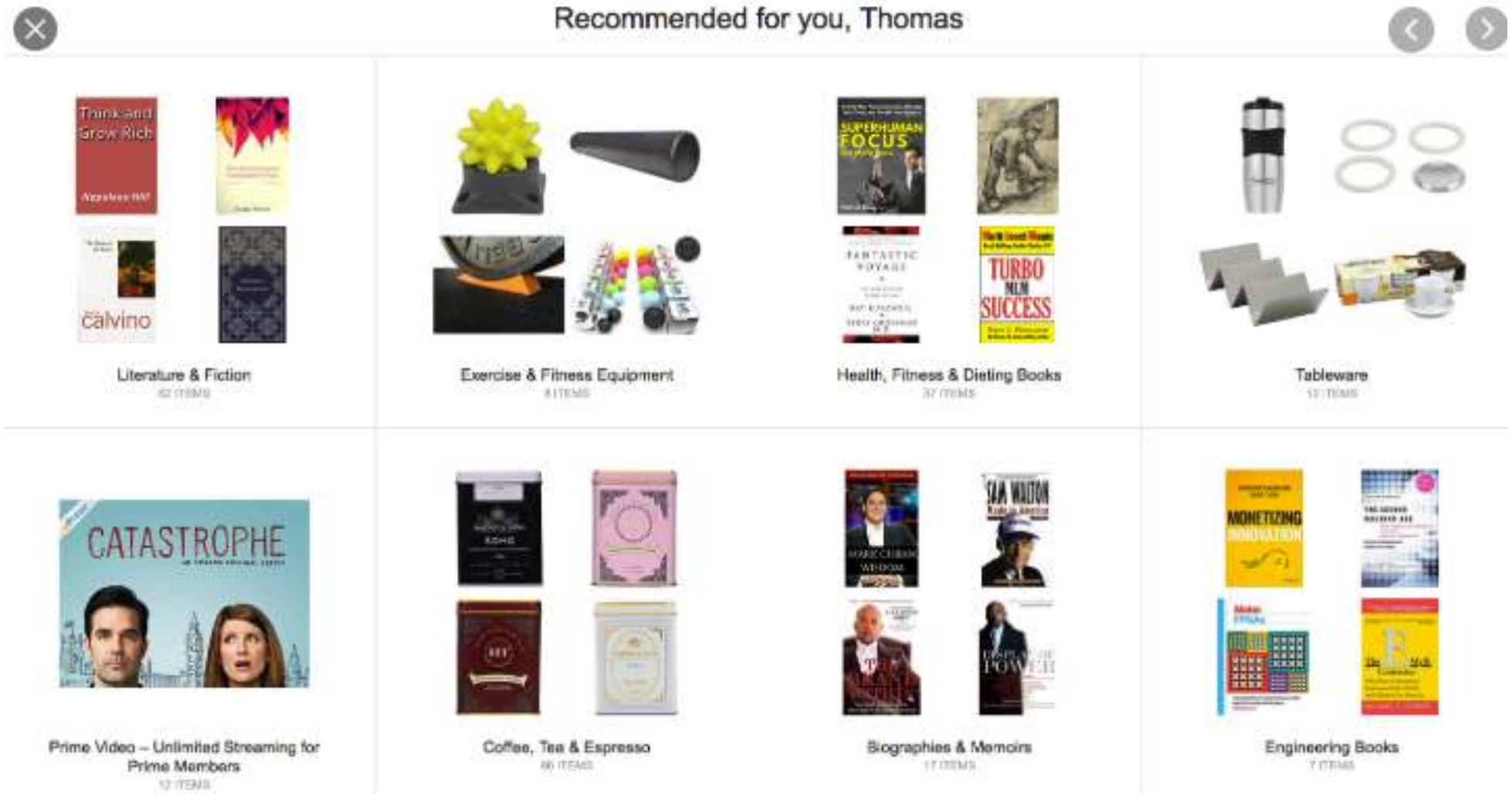
- How do we make personalised recommendations to new users?

- Even for existing users, data is often poor!
  e.g. People often do not bother to rate products and purchases

- In Netflix: **98.8 %** of the ratings are unknown
- In Movielens: **95.7 %** of the ratings are unknown

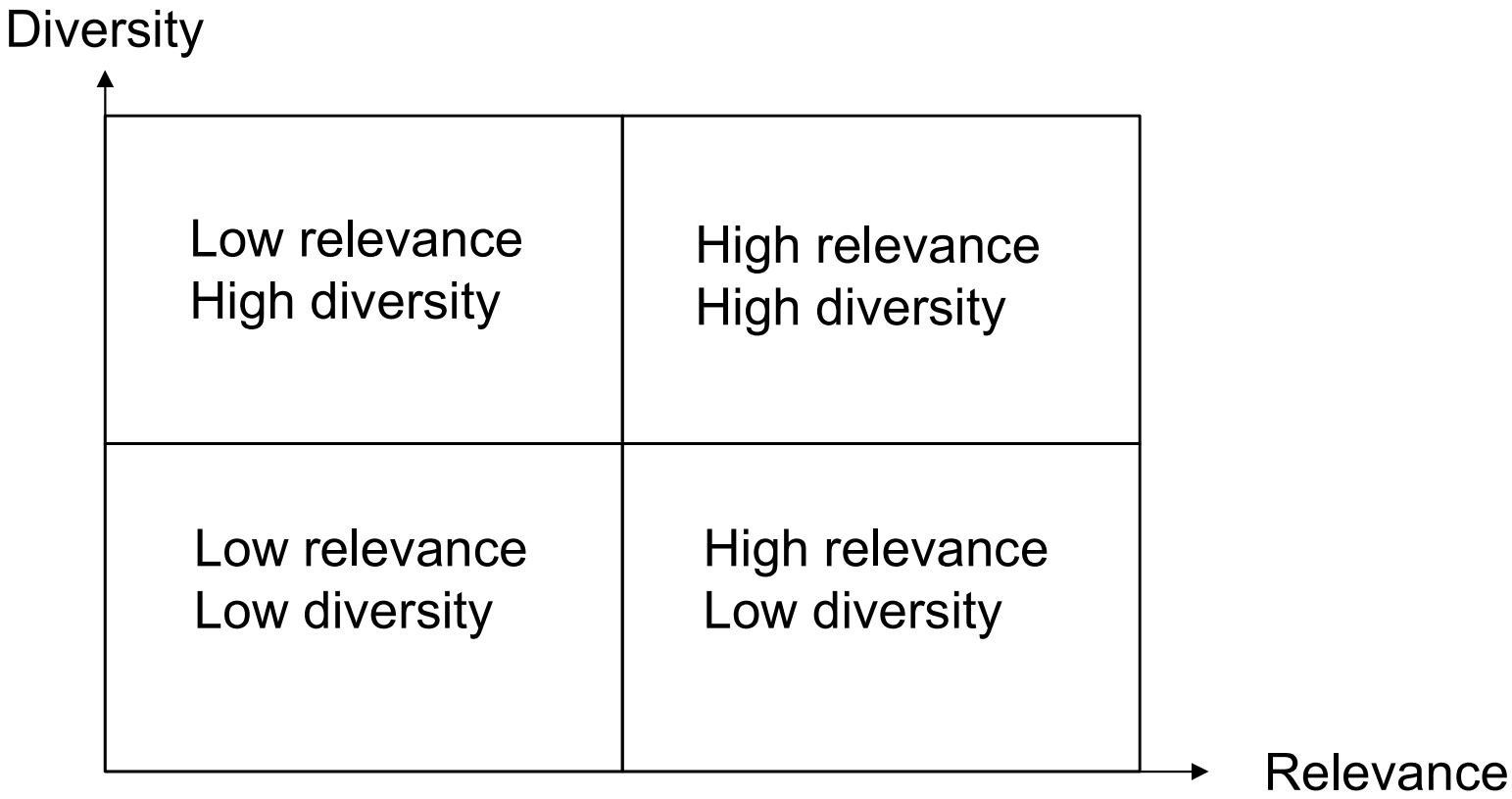


We usually don't recommend just one thing!



Diversity is important too!

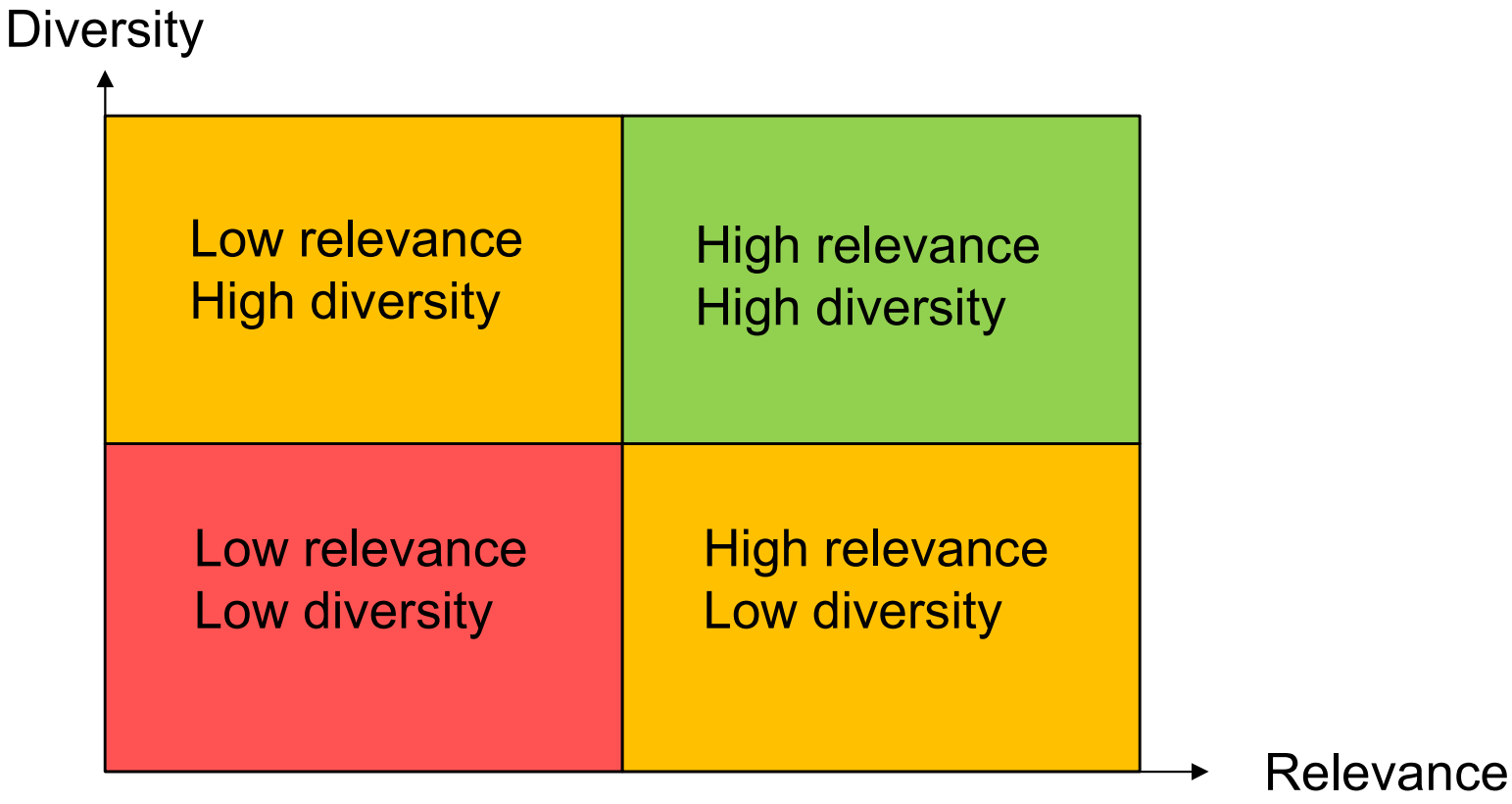
- Unlike predictive modelling, accuracy or precision is not the only criteria for success



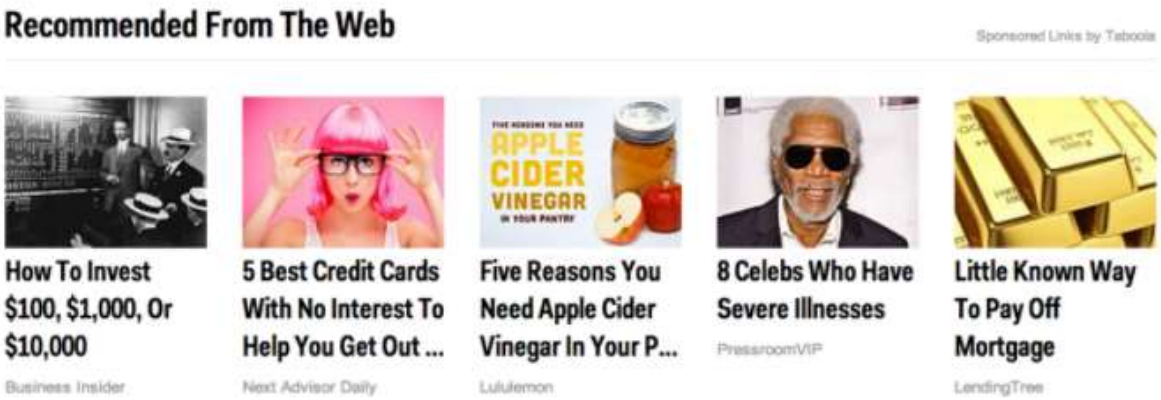
Where do you think your recommendations should be?

Diversity is important too!

- Unlike predictive modelling, accuracy or precision is not the only criteria for success



Recommendations versus Ads



Recommendations should not be perceived as ads!

STUDY: Nearly 80% of Readers Disapprove of 'From the Web' Content Recommendation Widgets

New Arkadium study also finds that 40% blame their negative content recommendation experiences on the hosting site

New York – September 17, 2017 – Arkadium (www.Arkadium.com), the company innovating visual, interactive content for more than 500 of the world's leading publishers, today announced the results of a commissioned study examining consumers' opinions of content recommendation widgets. The study polled 300 people in the U.S. Below are the study's key takeaways.

A majority dislike content recommendation widgets

Seventy-nine percent of respondents either somewhat or strongly disapprove of sites using content recommendation widgets. Consumers are not fooled by these widgets either, as nearly three-quarters (73%) of respondents were aware that articles within these content recommendation widgets were often sponsored and paid for.