

# Collaborative Filtering

Stavros Korokithakis

Intelligent Systems in Business

Problem?

# What is the problem?

- Too much information.
- Too little time.
- How can I choose?
- Collaborative filtering, that's how!

# What is the problem?

- Too much information.
- Too little time.
- How can I choose?
- Collaborative filtering, that's how!

# What is the problem?

- Too much information.
- Too little time.
- How can I choose?
- Collaborative filtering, that's how!

# What is the problem?

- Too much information.
- Too little time.
- How can I choose?
- Collaborative filtering, that's how!

- If we're wise enough, we can feign intelligence.
- Conversely, by having enough users with similar tastes, we can pretend that we know what each user likes.
- We can also try to actually discover aspects of the object we're recommending.

- If we're wise enough, we can feign intelligence.
- Conversely, by having enough users with similar tastes, we can pretend that we know what each user likes.
- We can also try to actually discover aspects of the object we're recommending.



- If we're wise enough, we can feign intelligence.
- Conversely, by having enough users with similar tastes, we can pretend that we know what each user likes.
- We can also try to actually discover aspects of the object we're recommending.

## Memory-based techniques

- Popular, easy to implement
- Consider the entire userbase and only the userbase
- Do not know anything about the item

## Model-based techniques

- Harder to implement, but provide more advantages
- Build a model of the item and use its aspects to provide recommendations
- Sadly, model building is expensive

## Memory-based techniques

- Popular, easy to implement
- Consider the entire userbase and only the userbase
- Do not know anything about the item

## Model-based techniques

- Harder to implement, but provide more advantages
- Build a model of the item and use its aspects to provide recommendations
- Sadly, model building is expensive

## Memory-based techniques

- Popular, easy to implement
- Consider the entire userbase and only the userbase
- Do not know anything about the item

## Model-based techniques

- Harder to implement, but provide more advantages
- Build a model of the item and use its aspects to provide recommendations
- Sadly, model building is expensive

## Memory-based techniques

- Popular, easy to implement
- Consider the entire userbase and only the userbase
- Do not know anything about the item

## Model-based techniques

- Harder to implement, but provide more advantages
- Build a model of the item and use its aspects to provide recommendations
- Sadly, model building is expensive

## Memory-based techniques

- Popular, easy to implement
- Consider the entire userbase and only the userbase
- Do not know anything about the item

## Model-based techniques

- Harder to implement, but provide more advantages
- Build a model of the item and use its aspects to provide recommendations
- Sadly, model building is expensive

## Memory-based techniques

- Popular, easy to implement
- Consider the entire userbase and only the userbase
- Do not know anything about the item

## Model-based techniques

- Harder to implement, but provide more advantages
- Build a model of the item and use its aspects to provide recommendations
- Sadly, model building is expensive

## **Memory-based techniques**

- Popular, easy to implement
- Consider the entire userbase and only the userbase
- Do not know anything about the item

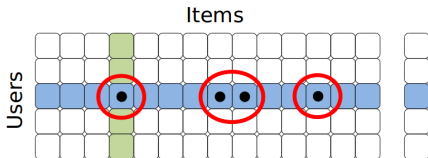
## **Model-based techniques**

- Harder to implement, but provide more advantages
- Build a model of the item and use its aspects to provide recommendations
- Sadly, model building is expensive

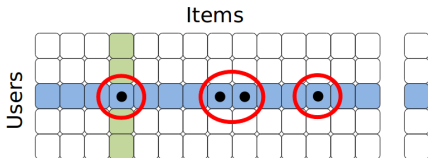


Last.fm

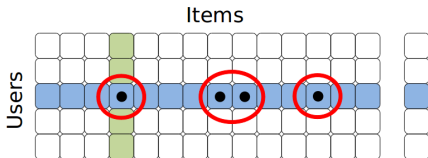
- Utilises a memory-based technique
- Users and ratings are in a sparse matrix
- Considers users as vectors and uses their cosine to gauge taste compatibility.
- Runs relatively fast off-line and offers good performance.



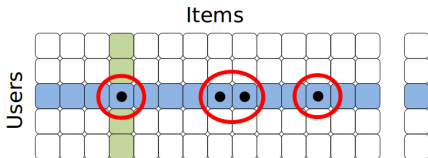
- Utilises a memory-based technique
- Users and ratings are in a sparse matrix
- Considers users as vectors and uses their cosine to gauge taste compatibility.
- Runs relatively fast off-line and offers good performance.



- Utilises a memory-based technique
- Users and ratings are in a sparse matrix
- Considers users as vectors and uses their cosine to gauge taste compatibility.
- Runs relatively fast off-line and offers good performance.



- Utilises a memory-based technique
- Users and ratings are in a sparse matrix
- Considers users as vectors and uses their cosine to gauge taste compatibility.
- Runs relatively fast off-line and offers good performance.



Google News

- Google need to be able to filter thousands of items in near-realtime, with items having a lifetime of a few days at most.
- No clear vote of preference from the user.
- Approach must be content-agnostic so it can be applied to other types of media as well.

U.K. ▾

Top Stories

Starred ☆

World

U.K.

**Business**

Sci/Tech


Entertainment


Sports

Health

Spotlight

Most Popular


Business

  
The Guardian

☆ [UCL student's ISB presentation "mindblowing"](#)

The Guardian - 1 hour later


Students who were lucky enough to attend Stavros Korokithakis' ISB presentation described it as "mindblowing" and "nothing short of spectacular".

[✚ Korokithakis' presentation](#)  ITN NEWS

["It was only OK", naysayers complain](#) BBC News

[This is Money](#) - [Reuters](#) - [Wall Street Journal](#)

[all 1,283 news articles »](#) [✉ Email this story](#)

  
BBC News

☆ [Vauxhall given £270 million lifeline by Government](#)

Telegraph.co.uk - [Andrew Porter](#) - 3 hours ago

Car manufacturer Vauxhall, which has been under the threat of closure for a year, was last night offered a lifeline by the Government.

[UK agrees loan guarantee for GM Europe](#) Reuters

[Vauxhall and Opel get £270m UK loans guarantee](#) BBC News

[Times Online](#) - [Financial Times](#) - [Wall Street Journal](#) - [ShareCast](#)

[all 258 news articles »](#) [✉ Email this story](#)

- Google need to be able to filter thousands of items in near-realtime, with items having a lifetime of a few days at most.
- No clear vote of preference from the user.
- Approach must be content-agnostic so it can be applied to other types of media as well.

U.K. ▾

Top Stories

Starred ☆

World

U.K.

**Business**

Sci/Tech


Entertainment

Sports

Health

Spotlight


Most Popular

  
The Guardian

### ☆ [UCL student's ISB presentation "mindblowing"](#)

The Guardian - 1 hour later


Students who were lucky enough to attend Stavros Korokithakis' ISB presentation described it as "mindblowing" and "nothing short of spectacular".

[✚ Korokithakis' presentation](#)  ITN NEWS

["It was only OK", naysayers complain](#) BBC News

[This is Money](#) - [Reuters](#) - [Wall Street Journal](#)

[all 1,283 news articles »](#) [✉ Email this story](#)

  
BBC News

### ☆ [Vauxhall given £270 million lifeline by Government](#)

Telegraph.co.uk - [Andrew Porter](#) - 3 hours ago

Car manufacturer Vauxhall, which has been under the threat of closure for a year, was last night offered a lifeline by the Government.

[UK agrees loan guarantee for GM Europe](#) Reuters

[Vauxhall and Opel get £270m UK loans guarantee](#) BBC News

[Times Online](#) - [Financial Times](#) - [Wall Street Journal](#) - [ShareCast](#)

[all 258 news articles »](#) [✉ Email this story](#)



- Google need to be able to filter thousands of items in near-realtime, with items having a lifetime of a few days at most.
- No clear vote of preference from the user.
- Approach must be content-agnostic so it can be applied to other types of media as well.

U.K. ▾

Top Stories

Starred ☆

World

U.K.

**Business**

Sci/Tech


Entertainment


Sports

Health

Spotlight

Most Popular


Business

  
The Guardian

### ☆ [UCL student's ISB presentation "mindblowing"](#)

The Guardian - 1 hour later


Students who were lucky enough to attend Stavros Korokithakis' ISB presentation described it as "mindblowing" and "nothing short of spectacular".

[✚ Korokithakis' presentation](#)  ITN NEWS

["It was only OK", naysayers complain](#) BBC News

[This is Money](#) - [Reuters](#) - [Wall Street Journal](#)

[all 1,283 news articles »](#) [✉ Email this story](#)

  
BBC News

### ☆ [Vauxhall given £270 million lifeline by Government](#)

Telegraph.co.uk - [Andrew Porter](#) - 3 hours ago

Car manufacturer Vauxhall, which has been under the threat of closure for a year, was last night offered a lifeline by the Government.

[UK agrees loan guarantee for GM Europe](#) Reuters

[Vauxhall and Opel get £270m UK loans guarantee](#) BBC News

[Times Online](#) - [Financial Times](#) - [Wall Street Journal](#) - [ShareCast](#)

[all 258 news articles »](#) [✉ Email this story](#)

- Google opted for a hybrid approach, utilising two model-based technique and a memory-based one.
- Each technique assigns a score to each story.
- Users vote with their clicks.
- The scores are weighted and the stories are presented to the user.

- Google opted for a hybrid approach, utilising two model-based technique and a memory-based one.
- Each technique assigns a score to each story.
- Users vote with their clicks.
- The scores are weighted and the stories are presented to the user.

- Google opted for a hybrid approach, utilising two model-based technique and a memory-based one.
- Each technique assigns a score to each story.
- Users vote with their clicks.
- The scores are weighted and the stories are presented to the user.

- Google opted for a hybrid approach, utilising two model-based technique and a memory-based one.
- Each technique assigns a score to each story.
- Users vote with their clicks.
- The scores are weighted and the stories are presented to the user.

Amazon

- More items the user will like means increased sales.
- Recommendations must be generated in a fraction of a second.
- Simpler methods are on-line and consider the entire matrix of trillions of items.
- There must be a better way!

- More items the user will like means increased sales.
- Recommendations must be generated in a fraction of a second.
- Simpler methods are on-line and consider the entire matrix of trillions of items.
- There must be a better way!



- More items the user will like means increased sales.
- Recommendations must be generated in a fraction of a second.
- Simpler methods are on-line and consider the entire matrix of trillions of items.
- There must be a better way!

- More items the user will like means increased sales.
- Recommendations must be generated in a fraction of a second.
- Simpler methods are on-line and consider the entire matrix of trillions of items.
- There must be a better way!

- Why not group relevant items instead of users with the same tastes?
- If many users have bought pairs of items, consider them similar.
- Can be calculated off-line.
- Needs almost no data about the user.

## Customers Who Bought This Item Also Bought



[Business @ the Speed of Thought: Succeeding in the...](#) by Bill Gates  
★★★★☆ (152)  
\$11.53



[Hard Drive: Bill Gates and the Making of the Mic...](#) by James Wallace  
★★★★★ (45)  
\$12.21



[Great Business Teams: Delivering great presentations](#) by Stavros Korokithakis  
★★★★☆ (15)  
\$18.21

- Why not group relevant items instead of users with the same tastes?
- If many users have bought pairs of items, consider them similar.
- Can be calculated off-line.
- Needs almost no data about the user.

## Customers Who Bought This Item Also Bought



[Business @ the Speed of Thought: Succeeding in the...](#) by Bill Gates  
★★★★☆ (152)  
\$11.53



[Hard Drive: Bill Gates and the Making of the Mic...](#) by James Wallace  
★★★★★ (45)  
\$12.21



[Great Business Teams: Delivering great presentations](#) by Stavros Korokithakis  
★★★★☆ (15)  
\$18.21

- Why not group relevant items instead of users with the same tastes?
- If many users have bought pairs of items, consider them similar.
- Can be calculated off-line.
- Needs almost no data about the user.

## Customers Who Bought This Item Also Bought



[Business @ the Speed of Thought: Succeeding in the...](#) by Bill Gates  
★★★★☆ (152)  
\$11.53



[Hard Drive: Bill Gates and the Making of the Mic...](#) by James Wallace  
★★★★★ (45)  
\$12.21



[Great Business Teams: Delivering great presentations](#) by Stavros Korokithakis  
★★★★☆ (15)  
\$18.21

- Why not group relevant items instead of users with the same tastes?
- If many users have bought pairs of items, consider them similar.
- Can be calculated off-line.
- Needs almost no data about the user.

## Customers Who Bought This Item Also Bought



[Business @ the Speed of Thought: Succeeding in the...](#) by Bill Gates  
★★★★☆ (152)  
\$11.53



[Hard Drive: Bill Gates and the Making of the Mic...](#) by James Wallace  
★★★★★ (45)  
\$12.21



[Great Business Teams: Delivering great presentations](#) by Stavros Korokithakis  
★★★★☆ (15)  
\$18.21

Questions?

This presentation was amazing.

