

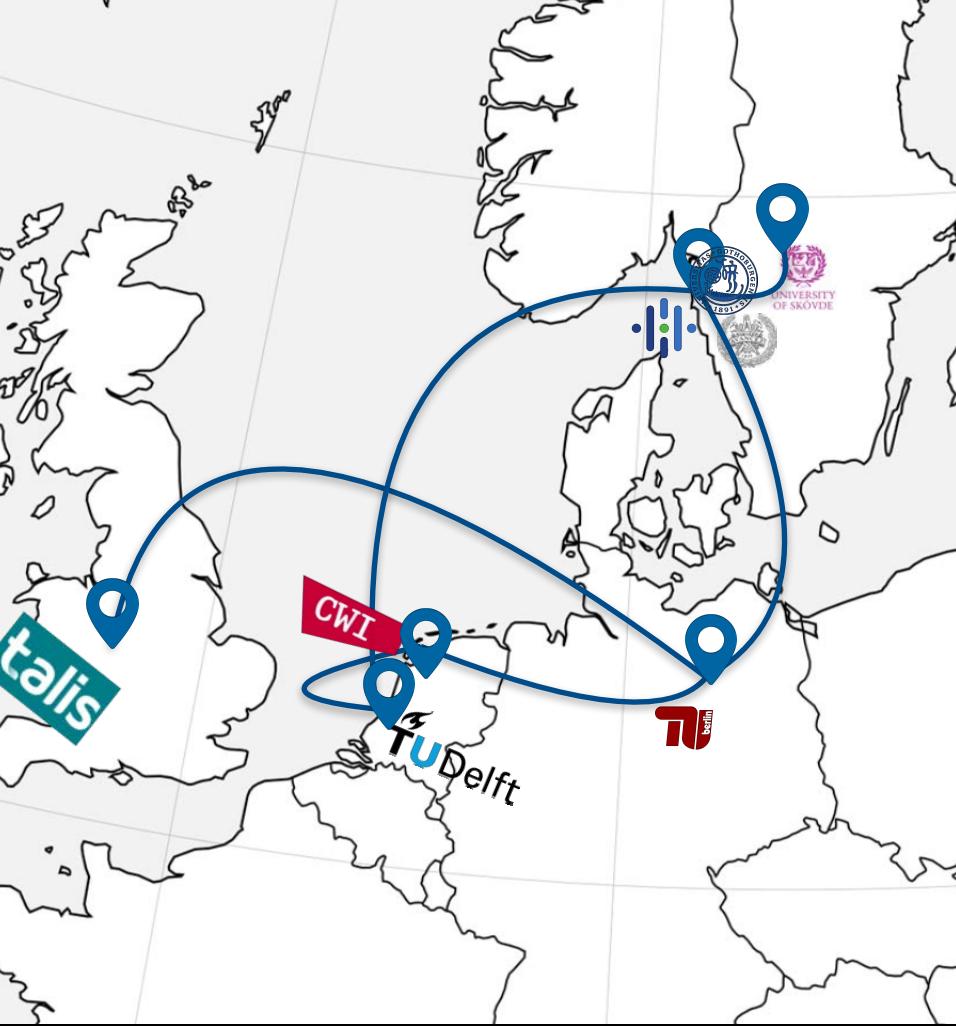


GÖTEBORGS  
UNIVERSITET

# FROM NETNEWS TO ETHICS

## A HISTORICAL OVERVIEW OF RECOMMENDER SYSTEMS

ALAN SAID  
[ALANSAYID@ACM.ORG](mailto:ALANSAYID@ACM.ORG)



# Who am I?

**Present:** Associate Prof, Uni. of Gothenburg, Sweden

**Past:-** Recorded Future, SE

- TU Delft, NL
- Centrum Wiskunde & Informatica, NL
- Talis, EN
- TU Berlin (PhD), DE
- Chalmers University of Tech., SE

**Research:**

- Recommender Systems, Reproducibility, Evaluation, Fairness, Sustainability, Explanations



# What I do in RecSys

**Present:** Chair of ACM RecSys  
Steering Committee 2023-2026

## Past:

- General Chair RecSys 2019
- Creator of RecSys Challenge
- Co-organizer of RecSys  
Summer School 2019 and 2023



- Introduction
- History
- Present
- Future



## ( • Introduction)

- History
- Present
- Future



What is the purpose of  
recommender systems?





A wide-angle photograph of a supermarket aisle. The shelves are packed with a massive variety of products, from small packages to large boxes and jars. The colors are vibrant and diverse, creating a visual representation of the overwhelming amount of information available in a retail environment.

# Information overload



**WIKIPEDIA**  
The Free Encyclopedia

Main page

Contents

Current events

Random article

About Wikipedia

Contact us

Donate

Contribute

Help

Learn to edit

Community portal

Recent changes

Upload file

Tools

What links here

Related changes

Special pages

Article **Talk**

Not logged in Talk Contributions Create account Log in

Read **Edit** View history

Search Wikipedia



# Information overload

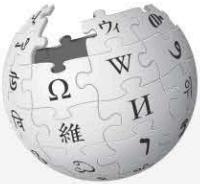
From Wikipedia, the free encyclopedia

*For the album by Alien Sex Fiend, see [Information Overload \(album\)](#).*

**Information overload** (also known as **infobesity**,<sup>[1][2]</sup> **infoxication**,<sup>[3]</sup> **information anxiety**,<sup>[4]</sup> and **information explosion**<sup>[5]</sup>) is the difficulty in understanding an issue and **effectively making decisions** when one has **too much information** (TMI) about that issue,<sup>[6]</sup> and is generally associated with the excessive quantity of daily information. The term "Information overload" was first used in Bertram Gross' 1964 book, *The Managing of Organizations*,<sup>[7]</sup> and was further popularized by Alvin Toffler in his bestselling 1970 book *Future Shock*.<sup>[8]</sup> Speier et al. (1999) said that if input exceeds the processing capacity, information overload occurs, which is likely to reduce the quality of the decisions.<sup>[9]</sup>

In a newer definition, Roetzel (2019) focuses on time and resources aspects. He states that when a decision-maker is given many sets of information, such as complexity, amount, and contradiction, the quality of its decision is decreased because of the individual's limitation of scarce resources to process all the information and optimally make the best decision.<sup>[10]</sup>

The advent of modern **information technology** has been a primary driver of information overload on multiple fronts: in quantity produced, ease of dissemination, and breadth of the audience reached. Longstanding technological factors have been further intensified by the rise of **social media** and the **attention economy**, which facilitates **attention theft**.<sup>[11][12]</sup> In the age of connective digital technologies, **informatics**, the **Internet culture** (or the digital culture), information overload is associated with over-exposure, excessive viewing of information, and input abundance of information and data.



**WIKIPEDIA**  
The Free Encyclopedia

Main page

Contents

Current events

Random article

About Wikipedia

Contact us

Donate

Contribute

Help

Learn to edit

Community portal

Recent changes

Upload file

Tools

What links here

Related changes

Special pages

Article **Talk**

Not logged in Talk Contributions Create account Log in

Read **Edit** View history

Search Wikipedia



# Information overload

From Wikipedia, the free encyclopedia

*For the album by Alien Sex Fiend, see [Information Overload \(album\)](#).*

**Information overload** (also known as **infobesity**,<sup>[1][2]</sup> **infoxication**,<sup>[3]</sup> **information anxiety**,<sup>[4]</sup> and **information explosion**<sup>[5]</sup>) is the difficulty in understanding an issue and effectively making decisions when one has **too much information** (TMI) about that issue,<sup>[6]</sup> and is generally associated with the excessive quantity of daily information. The term "Information overload" was first used in Bertram Gross' 1964 book, *The Managing of Organizations*,<sup>[7]</sup> and was further popularized by Alvin Toffler in his bestselling 1970 book *Future Shock*.<sup>[8]</sup> Speier et al. (1999) said that if input exceeds the processing capacity, information overload occurs, which is likely to reduce the quality of the decisions.<sup>[9]</sup>

In a newer definition, Roetzel (2019) focuses on time and resources aspects. He states that when a decision-maker is given many sets of information, such as complexity, amount, and contradiction, the quality of its decision is decreased because of the individual's limitation of scarce resources to process all the information and optimally make the best decision.<sup>[10]</sup>

The advent of modern **information technology** has been a primary driver of information overload on multiple fronts: in quantity produced, ease of dissemination, and breadth of the audience reached. Longstanding technological factors have been further intensified by the rise of **social media** and the **attention economy**, which facilitates **attention theft**.<sup>[11][12]</sup> In the age of connective digital technologies, **informatics**, the **Internet culture** (or the digital culture), information overload is associated with over-exposure, excessive viewing of information, and input abundance of information and data.



**WIKIPEDIA**  
The Free Encyclopedia

Main page

Contents

Current events

Random article

About Wikipedia

Contact us

Donate

Contribute

Help

Learn to edit

Community portal

Recent changes

Upload file

Tools

What links here

Related changes

Special pages

Article **Talk**

Not logged in Talk Contributions Create account Log in

Read **Edit** View history

Search Wikipedia



# Information overload

From Wikipedia, the free encyclopedia

*For the album by Alien Sex Fiend, see [Information Overload \(album\)](#).*

**Information overload** (also known as **infobesity**,<sup>[1][2]</sup> **infoxication**,<sup>[3]</sup> **information anxiety**,<sup>[4]</sup> and **information explosion**<sup>[5]</sup>) is the difficulty in understanding an issue and effectively making decisions when one has **too much information** (TMI) about that issue,<sup>[6]</sup> and is generally associated with the excessive quantity of daily information. The term "Information overload" was first used in Bertram Gross' 1964 book, *The Managing of Organizations*,<sup>[7]</sup> and was further popularized by Alvin Toffler in his bestselling 1970 book *Future Shock*.<sup>[8]</sup> Speier et al. (1999) said that if input exceeds the processing capacity, information overload occurs, which is likely to reduce the quality of the decisions.<sup>[9]</sup>

In a newer definition, Roetzel (2019) focuses on time and resources aspects. He states that when a decision-maker is given many sets of information, such as complexity, amount, and contradiction, the quality of its decision is decreased because of the individual's limitation of scarce resources to process all the information and optimally make the best decision.<sup>[10]</sup>

The advent of modern **information technology** has been a primary driver of information overload on multiple fronts: in quantity produced, ease of dissemination, and breadth of the audience reached. Longstanding technological factors have been further intensified by the rise of **social media** and the **attention economy**, which facilitates **attention theft**.<sup>[11][12]</sup> In the age of connective digital technologies, **informatics**, the **Internet culture** (or the digital culture), information overload is associated with over-exposure, excessive viewing of information, and input abundance of information and data.

# The volume of data created from 2010 to 2025

200

150

100

50

0

(Zettabytes)



# The volume of data created from 2010 to 2025

200

150

## Psychological process [\[ edit \]](#)

The phenomenon of overchoice occurs when many equivalent choices are available.<sup>[3]</sup>

Making a decision becomes overwhelming due to the many potential outcomes and risks

100

50

0

2010 2011 2012 2013 2014 2015 2016 2017 2018 2019 2020 2021 2022 2023 2024 2025

(Zettabytes)

# The volume of data created from 2010 to 2025

200

150

## Psychological process [\[ edit\]](#)

The phenomenon of overchoice occurs when many equivalent choices are available.<sup>[3]</sup>  
Making a decision becomes overwhelming due to the many potential outcomes and risks

100

50

0

2010 2011 2012 2013 2014 2015 2016 2017 2018 2019 2020 2021 2022 2023 2024 2025



(Zettabytes)



**WIKIPEDIA**  
The Free Encyclopedia

Main page

Contents

Current events

Random article

About Wikipedia

Contact us

Donate

Contribute

Help

Learn to edit

Community portal

Recent changes

Upload file

Tools

What links here

Related changes

Special pages

Article **Talk**

Not logged in Talk Contributions Create account Log in

Read **Edit** View history

Search Wikipedia



# Information overload

From Wikipedia, the free encyclopedia

*For the album by Alien Sex Fiend, see [Information Overload \(album\)](#).*

**Information overload** (also known as **infobesity**,<sup>[1][2]</sup> **infoxication**,<sup>[3]</sup> **information anxiety**,<sup>[4]</sup> and **information explosion**<sup>[5]</sup>) is the difficulty in understanding an issue and **effectively making decisions** when one has **too much information** (TMI) about that issue,<sup>[6]</sup> and is generally associated with the excessive quantity of daily information. The term "Information overload" was first used in Bertram Gross' 1964 book, *The Managing of Organizations*,<sup>[7]</sup> and was further popularized by Alvin Toffler in his bestselling 1970 book *Future Shock*.<sup>[8]</sup> Speier et al. (1999) said that if input exceeds the processing capacity, information overload occurs, which is likely to reduce the quality of the decisions.<sup>[9]</sup>

In a newer definition, Roetzel (2019) focuses on time and resources aspects. He states that when a decision-maker is given many sets of information, such as complexity, amount, and contradiction, the quality of its decision is decreased because of the individual's limitation of scarce resources to process all the information and optimally make the best decision.<sup>[10]</sup>

The advent of modern **information technology** has been a primary driver of information overload on multiple fronts: in quantity produced, ease of dissemination, and breadth of the audience reached. Longstanding technological factors have been further intensified by the rise of **social media** and the **attention economy**, which facilitates **attention theft**.<sup>[11][12]</sup> In the age of connective digital technologies, **informatics**, the **Internet culture** (or the digital culture), information overload is associated with over-exposure, excessive viewing of information, and input abundance of information and data.

A photograph of a stack of four antique books with ornate gold-tooled leather covers. A silver pocket watch hangs from a chain around the neck of the top book. The background is a soft-focus green.

# History of RecSys



Revised 2009 (entry history)

More entries for *recommend*

Nearby entries

Share

Cite

Contribute

Tabbed view



# recommend

VERB<sup>1</sup>

Factsheet

## Etymology

### ▼ Meaning & use

#### CONTENTS

- ▶ 1. † transitive. To praise, extol, or commend (a person)...
- ▶ 2. transitive. To entrust, consign, commit (something) to...
- ▶ 3. † transitive (reflexive). To command (oneself) to (the...)
- ▶ 4. transitive. To mention or present (a thing, course of...)
- ▶ 5. transitive. Chiefly of a quality, circumstance, or thing...
- ▶ 6. transitive. To defend, speak on behalf of. Obsolete. rare.
- ▶ 7. transitive. To offer counsel or advice to someone (to do...)

## Etymology

### Summary

Of multiple origins. Partly a borrowing from French. Partly a borrowing from Latin.

**Etymons:** French *recommander*; Latin *commendare*.

Partly < Anglo-Norman and Middle French *recommander*, variant of *recommender* **recommend v.<sup>1</sup>**, and partly < post-classical Latin *commendare* to commit, entrust (10th cent.; from 11th cent. in British sources), to praise, commend (from 13th cent. in British sources), to commend in prayer (14th cent. in British sources) < classical Latin *re-* **re- prefix** + *commendare* **commend v.** Compare Spanish *recomendar* (late 13th cent.), Portuguese *recomendar* (1387). Compare **recommend v.<sup>1</sup>** (and see discussion at that entry).

Show less

“ Cite

## Meaning & use

**Etymology****▼ Meaning & use****CONTENTS**

- ▶ 1. † transitive. To praise, extol, or commend (a person)...
- ▶ 2. transitive. To entrust, consign, commit (something) to...
- ▶ 3. † transitive (reflexive). To command (oneself) to (the...
- ▶ 4. **transitive. To mention or present (a thing, course of...**
- ▶ 5. transitive. Chiefly of a quality, circumstance, or thing...
- ▶ 6. transitive. To defend, speak on behalf of. Obsolete. rare.
- ▶ 7. transitive. To offer counsel or advice to someone (to do...

**Pronunciation****► Forms****Frequency****► Compounds & derived words****QUOTATIONS**

Hide all quotations

**EARLIER VERSION**

*recommend*, v.<sup>1</sup> in OED Second Edition (1989)

**IN OTHER DICTIONARIES**

c1460 I recomendē to þi meditacioun þe holy passioun of oure lord ihesu.

*Trewe & 12 Frutes* (McClean MS.) (1960) 32 (Middle English Dictionary) [\[...\]](#)

1581 I must recommend vnto them exercise of the bodie.

R. Mulcaster, *Positions* xxxix. 197 [\[...\]](#)

1649 Anent the meason word, which was recommended to the seuerall presbetries for tryall therof.

J. Lamont, *Diary* (1830) 7 [\[...\]](#)

1687 [Biscuits] which were recommended to me, as an excellent thing to drink a mornings draught with.

A. Lovell, translation of J. de Thévenot, *Travels into Levant* i. 96 [\[...\]](#)

1702 He would recommend, and enjoyn the Practice, and Use of both to that [of] his Native Kingdom.

*Clarendon's History of Rebellion* vol. I. ii. 83 [\[...\]](#)

1755 Two Papers in which my Dictionary is recommended to the Public were written by your Lordship.

S. Johnson, *Letter 7 February* (1992) vol. I. 95 [\[...\]](#)

1827 Let me recommend a little of this pike!

B. Disraeli, *Vivian Grey* vol. III. v. xv. 312 [\[...\]](#)

1863 We will conclude by recommending his work to our readers.

*Saturday Review* 16 May 640 [\[...\]](#)

1925 They recommended to him the more even and genial style of John Fiske.

W. Cather, *Professor's House* i. i. 32 [\[...\]](#)

1985 'Can you recommend me a nice hotel?' I was asked.

*Times* 8 April 10/1 [\[...\]](#)

2000 It's a goldmine of information, highly recommended to anyone with an interest in French pop-fic.

*Interzone* October 66/1 [\[...\]](#)

“ Cite

Historical thesaurus ▾

**Etymology****▼ Meaning & use****CONTENTS**

- 1. † transitive. To praise, extol, or commend (a person)...
- 2. transitive. To entrust, consign, commit (something) to...
- 3. † transitive (reflexive). To command (oneself) to (the...
- 4. **transitive. To mention or present (a thing, course of...**
- 5. transitive. Chiefly of a quality, circumstance, or thing...
- 6. transitive. To defend, speak on behalf of. Obsolete. rare.
- 7. transitive. To offer counsel or advice to someone (to do...

**Pronunciation****► Forms****Frequency****► Compounds & derived words****QUOTATIONS**

Hide all quotations

**EARLIER VERSION**

*recommend*, v.<sup>1</sup> in OED Second Edition (1989)

**IN OTHER DICTIONARIES**

c1460 I recomendē to þi meditacioun þe holy passioun of oure lord ihesu.

*Trewe & 12 Frutes* (McClean MS.) (1960) 32 (Middle English Dictionary) ...

1581 I must recommend vnto them exercise of the bodie.

R. Mulcaster, *Positions* xxxix. 197 ...

1649 Anent the meason word, which was recommended to the seuerall presbetries for tryall therof.

J. Lamont, *Diary* (1830) 7 ...

1687 [Biscuits] which were recommended to me, as an excellent thing to drink a mornings draught with.

A. Lovell, translation of J. de Thévenot, *Travels into Levant* i. 96 ...

1702 He would recommend, and enjoyn the Practice, and Use of both to that [of] his Native Kingdom.

*Clarendon's History of Rebellion* vol. I. ii. 83 ...

1755 Two Papers in which my Dictionary is recommended to the Public were written by your Lordship.

S. Johnson, *Letter* 7 February (1992) vol. I. 95 ...

1827 Let me recommend a little of this pike!

B. Disraeli, *Vivian Grey* vol. III. v. xv. 312 ...

1863 We will conclude by recommending his work to our readers.

*Saturday Review* 16 May 640 ...

1925 They recommended to him the more even and genial style of John Fiske.

W. Cather, *Professor's House* i. i. 32 ...

1985 'Can you recommend me a nice hotel?' I was asked.

*Times* 8 April 10/1 ...

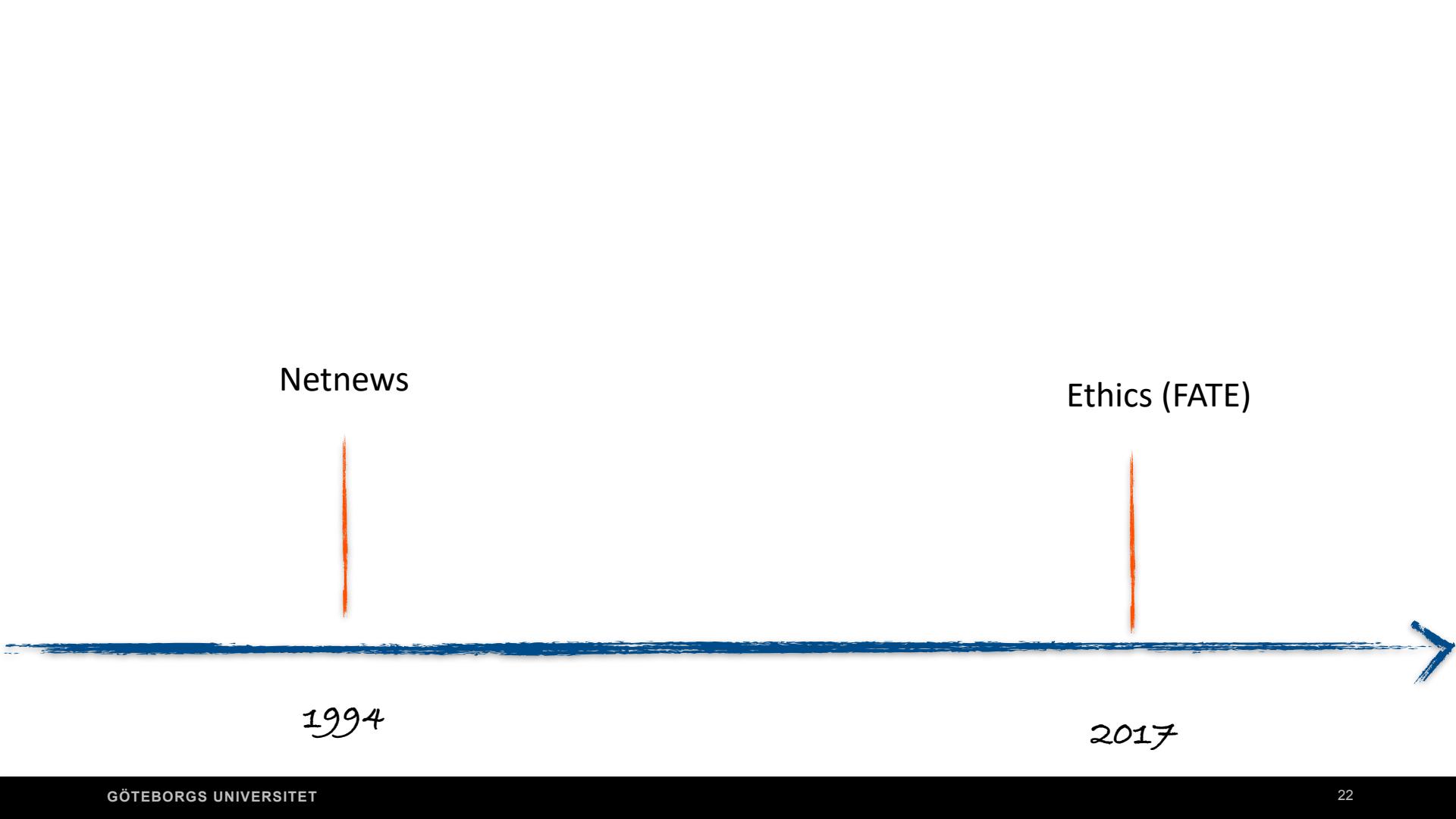
2000 It's a goldmine of information, highly recommended to anyone with an interest in French pop-fic.

*Interzone* October 66/1 ...

Netnews



1994



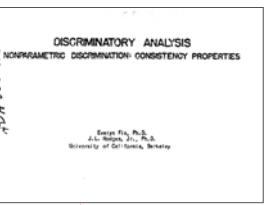
Netnews

Ethics (FATE)

1994

2017





1951

# NONPARAMETRIC DISCRIMINATION: CONSISTENCY PROPERTIES

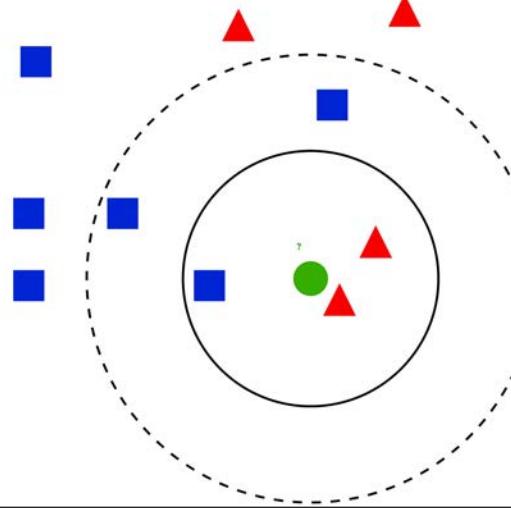


## 1. Introduction

The discrimination problem (two population case) may be defined as follows: a random variable  $Z$ , of observed value  $z$ , is distributed over some space (say,  $p$ -dimensional) either according to distribution  $F$ , or according to distribution  $G$ . The problem is to decide, on the basis of  $z$ , which of the two distributions  $Z$  has.

The problem may be classified in various ways into subproblems. One pertinent method of classification is according to the amount of information assumed to be available about  $F$  and  $G$ . We may distinguish three stages:

- (i)  $F$  and  $G$  are completely known
- (ii)  $F$  and  $G$  are known except for the values of one or more parameters
- (iii)  $F$  and  $G$  are completely unknown, except possibly for assumptions about existence of densities,



An appropriate (positive) constant  $c$  is chosen, and the following rule is observed

If  $\frac{f(z)}{g(z)} > c$ , we decide in favor of  $F$

If  $\frac{f(z)}{g(z)} < c$ , we decide in favor of  $G$

If  $\frac{f(z)}{g(z)} = c$ , the decision may be made in an arbitrary manner

These procedures are known to have optimum properties with regard to control of probability of misclassification (probability of wrong decision). We shall refer to this as the "likelihood ratio procedure," and denote it by  $L(c)$ .

For simplicity, we shall assume throughout the paper that the borderline case  $f(z) = cg(z)$  can be neglected.

# NONPARAMETRIC DISCRIMINATION: CONSISTENCY PROPERTIES



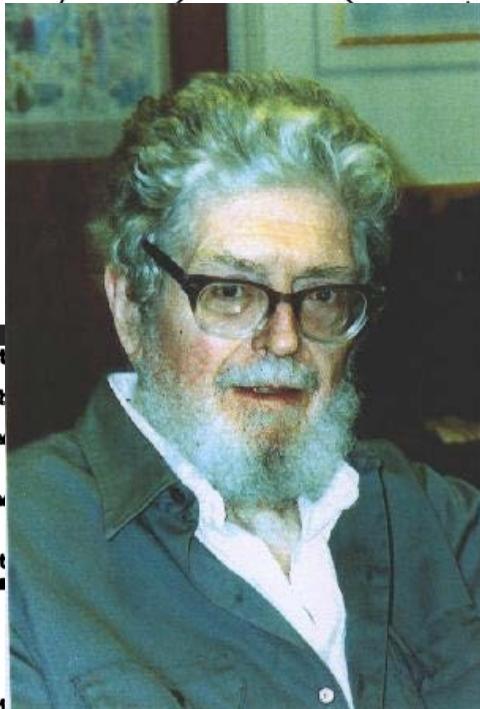
## 1. Introduction

The discrimination problem may be defined as follows. A value  $z$ , is distributed either according to a distribution  $F$  or to a distribution  $G$ . The problem is to decide which of the two distributions is true.

The problem may be divided into two subproblems. One problem is to decide according to the amount of information available about  $F$  and  $G$  at one stage:

- (i)  $F$  and  $G$  are known
- (ii)  $F$  and  $G$  are known up to one or more parameters

(iii)  $F$  and  $G$  are completely unknown, except possibly for assumptions about existence of densities,



An appropriate (positive) decision rule is obtained by the following rule:

If  $\frac{f(z)}{g(z)} > c$ , we decide  $F$ .

If  $\frac{f(z)}{g(z)} < c$ , we decide  $G$ .

If  $\frac{f(z)}{g(z)} = c$ , then we decide  $F$  with probability  $p$  and  $G$  with probability  $1-p$ .

These procedures are called "likelihood ratio tests" in regard to control of the probability of wrong decisions. The rule is called the "likelihood ratio procedure," and denote it by  $L(c)$ .

For simplicity, we shall assume throughout the paper that the borderline case  $f(z) = cg(z)$  can be neglected.

## Nearset Neighbor Pattern Classification

T. M. COVER, MEMBER, IEEE, AND P. E. HART, MEMBER, IEEE

The nearest neighbor decision rule assigns to a test observation  $x_0$  the class of the nearest of the members of a set of training observations. This paper shows that the Bayes probability of error of the nearest neighbor rule is at least as large as the Bayes probability of error of the minimum Bayes procedure. The Bayes probability of error of the minimum Bayes procedure is shown to be at least as large as the Bayes probability of error of the minimum Bayes procedure based on a single prior distribution. Thus the Bayes probability of error of the nearest neighbor rule is at least as large as the Bayes probability of error of the minimum Bayes procedure based on a single prior distribution. Thus the Bayes probability of error of the nearest neighbor rule is at least as large as the Bayes probability of error of the minimum Bayes procedure based on a single prior distribution.

### I. INTRODUCTION

THE CLASSIFICATION problem, that is, the assignment of knowledge which the statistician may have concerning a test observation  $x_0$  to one of a set of categories, has been studied by many authors. A recent survey of the literature on pattern classification was given by Hart [1].

Eugene F. F. P. B.  
J. A. Naujokas, Jr., Ph.D.  
University of California, Berkeley

1967  
1951

# Nearest Neighbor Pattern Classification

T. M. COVER, MEMBER, IEEE, AND P. E. HART, MEMBER, IEEE

**Abstract**—The nearest neighbor decision rule assigns to an unclassified sample point the classification of the nearest of a set of previously classified points. This rule is independent of the underlying joint distribution on the sample points and their classifications, and hence the probability of error  $R$  of such a rule must be at least as great as the Bayes probability of error  $R^*$ —the minimum probability of error over all decision rules taking underlying probability structure into account. However, in a large sample analysis, we will show in the  $M$ -category case that  $R^* \leq R \leq R^*(2 - MR^*/(M-1))$ , where these bounds are the tightest possible, for all suitably smooth underlying distributions. Thus for any number of categories, the probability of error of the nearest neighbor rule is bounded above by twice the Bayes probability of error. In this sense, it may be said that half the classification information in an infinite sample set is contained in the nearest neighbor.

## I. INTRODUCTION

IN THE CLASSIFICATION problem there are two extremes of knowledge which the statistician may possess. Either he may have complete statistical knowledge of the underlying joint distribution of the

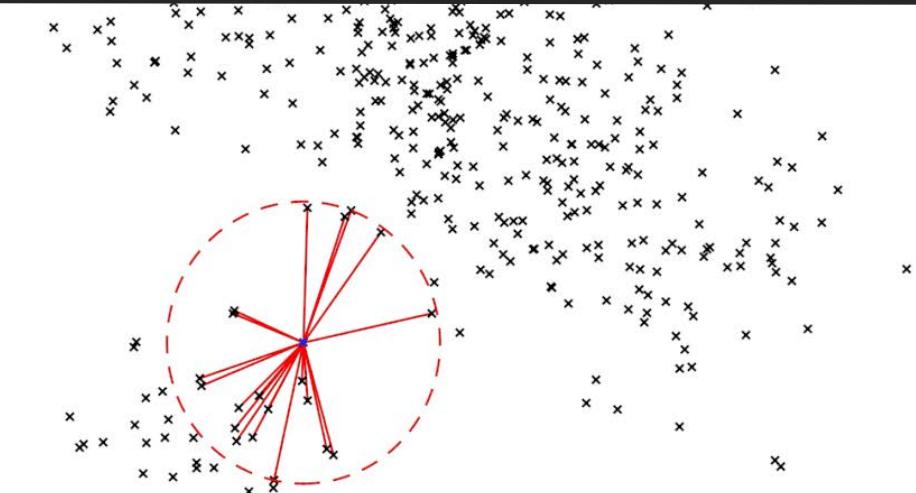
Manuscript received February 23, 1966; revised April 29, 1966. This work has been supported at Stanford University by U. S. Army Electronics Command under Contract DA28-043-AMC-01764(E) and by USAF under Contract AF49(638)1517; and at the Stanford Research Institute, Menlo Park, Calif., by RADC under Contract



th the Department of Electrical Engineering, Stanford, Calif.  
the Stanford Research Institute, Menlo Park,

observation  $x$  and the true category  $\theta$ , or he no knowledge of the underlying distribution of which can be inferred from samples. In the first case, a standard Bayes analysis will yield an optimum procedure and the corresponding minimum (Bayes) probability of error of classification  $R^*$ . In the other case, a decision to classify  $x$  into category  $\theta$  is based only on a collection of  $n$  correctly classified points  $(x_1, \theta_1), (x_2, \theta_2), \dots, (x_n, \theta_n)$ , and the decision is by no means clear. This problem is in general nonparametric statistics and no optimum procedure exists with respect to all under-

If it is assumed that the classified samples are independently identically distributed according to the distribution of  $(x, \theta)$ , certain heuristic arguments may be made about good decision procedures. For example, it is reasonable to assume that observations which are close together (in some appropriate metric) will have the same classification, or at least will have almost the same posterior probability distributions on their respective classifications. Thus to classify the unknown sample  $x$  we may wish to weight the evidence of the nearby  $x_i$ 's most heavily. Perhaps the simplest nonparametric decision procedure of this form is the *nearest neighbor* (NN) rule, which classifies  $x$  in the category of its nearest neighbor. Surprisingly, it will be shown that in the large sample



# Nearest Neighbor Pattern Classification

T. M. COVER, MEMBER, IEEE, AND P. E. HART, MEMBER, IEEE

**Abstract**—The nearest neighbor decision rule assigns to an unclassified sample point the classification of the nearest of a set of previously classified points. This rule is independent of the underlying joint distribution on the sample points and their classifications, and hence the probability of error  $R$  of such a rule must be at least as great as the Bayes probability of error  $R^*$ —the minimum probability of error over all decision rules taking underlying probability structure into account. However, in a large sample analysis, we will show in the  $M$ -category case that  $R^* \leq R \leq R^*(2 - MR^*/(M-1))$ , where these bounds are the tightest possible, for all suitably smooth underlying distributions. Thus for any number of categories, the probability of error of the nearest neighbor rule is bounded above by twice the Bayes probability of error. In this sense, it may be said that half the classification information in an infinite sample set is contained in the nearest neighbor.

## I. INTRODUCTION

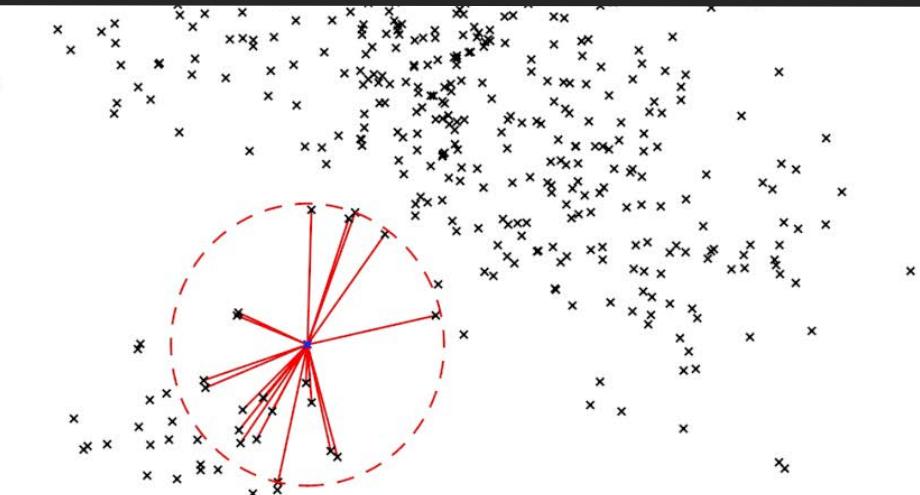
IN THE CLASSIFICATION problem there are two extremes of knowledge which the statistician may possess. Either he may have complete statistical knowledge of the underlying joint distribution of the

Manuscript received February 23, 1966; revised April 29, 1966. This work has been supported at Stanford University by U. S. Army Electronics Command under Contract DA28-043-AMC-01764(E) and by USAF under Contract AF49(638)1517; and at the Stanford Research Institute, Menlo Park, Calif., by RADC under Contract



th the Department of Electrical Engineering, Stanford, Calif.  
the Stanford Research Institute, Menlo Park,

If it is assumed that the classified samples  $(x_i, \theta_i)$  are independently identically distributed according to the distribution of  $(x, \theta)$ , certain heuristic arguments may be made about good decision procedures. For example, it is reasonable to assume that observations which are close together (in some appropriate metric) will have the same classification, or at least will have almost the same posterior probability distributions on their respective classifications. Thus to classify the unknown sample  $x$  we may wish to weight the evidence of the nearby  $x_i$ 's most heavily. Perhaps the simplest nonparametric decision procedure of this form is the *nearest neighbor* (NN) rule, which classifies  $x$  in the category of its nearest neighbor. Surprisingly, it will be shown that in the large sample



## An Algebra for Recommendations

### Using Reader Data as a Basis for Measuring Document Proximity

Jussi Karlgren

Department of Computer and Systems Sciences  
Stockholm University  
Electrum 230  
S-164 40 KISTA, Sweden

jussi@dsv.su.se  
jussi@flash.bellcore.com  
jussi@sc.columbia.edu

In this paper we study the problem of measuring the relevance of a document to a reader. We consider the case where the reader has a set of interests, and the document has a set of topics. We show that it is possible to measure the relevance of the document to the reader by considering the probability that the reader has all the topics in the document. Surprisingly, it will be shown that, in the large sample case, this simple rule has a probability of at least 1/2.

Engineering Faculty  
Jussi Karlgren, Ph.D.  
University of Göteborg, Sweden

1967 1990  
1951



# An Algebra for Recommendations

## Using Reader Data as a Basis for Measuring Document Proximity

Jussi Karlgren

Department of Computer and Systems Sciences  
Stockholm University  
Electrum 230  
S-164 40 KISTA, Sweden

jussi@dsv.su.se  
jussi@flash.bellcore.com  
jussi@sc.columbia.edu



### 3. Defining Proximity

Proximity is defined to be a measure of closeness between documents based on the interest readers of the document have shown it. The main idea is, for every pair or set of documents, to use all their reader's interest grades to define relationships from one document to sets of others.

We will do this by first defining recommendation from a document to another, for one reader. This will correspond to the answer a reader gives to the question "I liked book A. Do you believe I will like B?".

#### 3.1 Proximity of One Document to Others

A reader  $x$  is represented as a vector

$$r(x) = [r_1, \dots, r_i, \dots, r_n],$$

where

$r(x)_i$  = the  $i$ th element in  $r(x)$ ,

#### 3.2.1 The Tendency to Give Good or Bad Grades

Another factor which may have to be taken into account when the constants are given numerical values, is the tendency of readers to give good or bad grades. An "Interesting" grade given by someone who only gives it very seldom might be worth more than an "Interesting" grade given by someone who frequently grades documents "Interesting".

# An Algebra for Recommendations

## Using Reader Data as a Basis for Measuring Document Proximity

Jussi Karlgren

Department of Computer and Systems Sciences  
Stockholm University  
Electrum 230  
S-164 40 KISTA, Sweden

jussi@dsv.su.se  
jussi@flash.bellcore.com  
jussi@sc.columbia.edu



### 3. Defining Proximity

Proximity is defined to be a measure of closeness between documents based on the interest readers of the document have shown it. The main idea is, for every pair or set of documents, to use all their reader's interest grades to define relationships from one document to sets of others.

We will do this by first defining recommendation from a document to another, for one reader. This will correspond to the answer a reader gives to the question "I liked book A. Do you believe I will like B?".

#### 3.1 Proximity of One Document to Others

A reader  $x$  is represented as a vector

$$r(x) = [r_1, \dots, r_i, \dots, r_n],$$

where

$r(x)_i$  = the  $i$ th element in  $r(x)$ ,

#### 3.2.1 The Tendency to Give Good or Bad Grades

Another factor which may have to be taken into account when the constants are given numerical values, is the tendency of readers to give good or bad grades. An "Interesting" grade given by someone who only gives it very seldom might be worth more than an "Interesting" grade given by someone who frequently grades documents "Interesting".

Using Reader Data a

Jussi Karlgren

Department of Computer  
Stockholm University  
Electrum 230  
S-164 40 KISTA, Sweden

jussi@dsv.su.se  
jussi@flash.bellcore.com  
jussi@sc.columbia.edu

# Using **COLLABORATIVE FILTERING** to Weave an Information **TAPESTRY**

David Goldberg, David Nichols, Brian M. Oki, and Douglas Terry

Tapestry is an experimental mail system developed at the Xerox Palo Alto Research Center. The motivation for Tapestry comes from the increasing use of electronic mail, which has led to a large volume of messages and mailing lists containing thousands of documents [2, 7, 12]. One way to handle large volumes of mail is to provide mailing lists, enabling users to subscribe only to those lists of interest to them. However, as illustrated in Figure 1, this leads to a proliferation of lists and a proliferation of users maintaining their own mailing lists.

A better solution is for a user to specify a filter that scans all lists, selecting interesting documents no matter what list they are in. Several mail systems support filtering based on a document's subject line [3, 5, 6, 8]. A more tenable approach for Tapestry is that more intelligent filtering can be done by users having access to the filtering process.

In addition to content-based filtering, the Tapestry system was designed and built to support social-based filtering. Content-based filtering allows users to filter documents to help one another perform filtering by recording their reactions to documents they read. Such reactions may be that a document was particularly interesting (or particularly uninteresting). These reactions can be used by other users who have access to the system through others' filters. One application of annotations is in support of moderated filtering groups.

Eugene Fife, Ph.D.  
J.A. Nichols, Ph.D.  
University of California, Berkeley

1967  
1951 1990  
1992



(a) Electronic mail overload



(c) Conventional filtering



(b) Using distribution lists



(d) Collaborative filtering

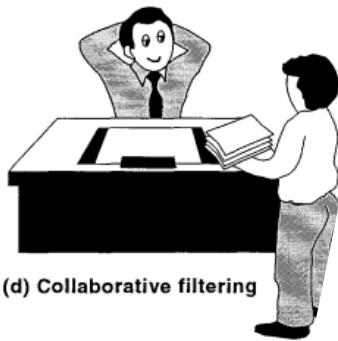
# Using COLLABORATIVE FILTERING to Weave an Information TAPESTRY

David Goldberg, David Nichols, Brian M. Oki, and Douglas Terry

Tapestry is an experimental mail system developed at the Xerox Research Center. The motivation for Tapestry comes from the increasing use of electronic mail, which is resulting in users being inundated by a huge stream of incoming documents [1, 12]. One way to handle large volumes of mail is to provide mailing lists, enabling users to subscribe only to those lists of interest to them. However, as illustrated in Figure 1, a set of documents of interest to a particular user rarely map neatly to existing lists. A better solution is for a user to specify a *filter* that scans all lists, selecting interesting documents no matter what list they are in. Several mail systems support filtering based on a document's contents [3, 5, 6, 8]. A basic tenet of the Tapestry work is that more effective filtering can be done by involving humans in the filtering process.

In addition to content-based filtering, the Tapestry system was designed and built to support *collaborative filtering*. Collaborative filtering simply means that people collaborate to help one another perform filtering by recording their reactions to documents they read. Such reactions may be that a document was particularly interesting (or particularly uninteresting). These reactions, more generally called *annotations*, can be accessed by others' filters. One application of annotations is in support of moderated newsgroups.





(d) Collaborative filtering

# Using COLLABORATIVE FILTERING to Weave an Information TAPESTRY

David Goldberg, David Nichols, Brian M. Oki, and Douglas Terry

Tapestry is an experimental mail system developed at the Xerox Research Center. The motivation for Tapestry comes from the increasing use of electronic mail, which is resulting in users being inundated by a huge stream of incoming documents [1, 12]. One way to handle large volumes of mail is to provide mailing lists, enabling users to subscribe only to those lists of interest to them. However, as illustrated in Figure 1, a set of documents of interest to a particular user rarely map neatly to existing lists. A better solution is for a user to specify a *filter* that scans all lists, selecting interesting documents no matter what list they are in. Several mail systems support filtering based on a document's contents [3, 5, 6, 8]. A basic tenet of the Tapestry work is that more effective filtering can be done by involving humans in the filtering process.

In addition to content-based filtering, the Tapestry system was designed and built to support *collaborative filtering*. Collaborative filtering simply means that people collaborate to help one another perform filtering by recording their reactions to documents they read. Such reactions may be that a document was particularly interesting (or particularly uninteresting). These reactions, more generally called *annotations*, can be accessed by others' filters. One application of annotations is in support of moderated newsgroups.



## GroupLens: An Open Architecture for Collaborative Filtering of Netnews

Paul Reinick\*, Neophytos Iacovou\*\*, Mitch Sachak\*, Peter Bergstrom\*\*, John Riedl\*\*

\* MIT Center for Coordination Science  
Dept. of Electrical Eng. & Comp. Sci.  
32 Memorial Drive  
Cambridge, MA 02139  
(617) 253-0149  
Email: prenick@mit.edu

\*\* University of Minnesota  
Dept. of Computer Science & Engineering  
Minneapolis, Minnesota 55455  
(612) 624-3732  
Email: riedl@cs.umn.edu

### ABSTRACT

Collaborative filters help people make choices based on the opinions of other people. GroupLens is a system for news filtering that uses collaborative filtering to predict which articles will likely be of interest to a user given the user's past history of reading news items. News reader clients display predicted scores and make recommendations to users based on their interests. Rating servers, called Better Bit Bureaus, gather and disseminate ratings from users. A rating server's rating is based on the heuristic that people who agreed in the past with the server's ratings are more likely to agree in the future by entering ratings under pseudonyms, without reducing the effectiveness of the user predictions. The entire architecture is open source and freely available. The Better Bit Bureaus can be developed independently and can interoperate with any news reader client.

**REPORTS** is an electronic bulletin board. The net news system has been around since 1979 [2, 7, 12]. One user is able to subscribe to a list of documents. A better way to share documents is to have them on a document exchange filter. In addition to support for annotations, such reactions, users can use other filters. One application of annotations is in support of moderated newsgroups.

Electronic bulletin boards, social filtering, Usenet, news, user model, selective dissemination of information.

© 1994 International Computer Science Institute. All rights reserved. ISSN 1063-0737. This paper was written while the author was at the University of Minnesota. It is based on work done in part by the Defense Advanced Research Projects Agency (DARPA) and the National Science Foundation (NSF). Part of this work was done while the author was at the Research Institute, Mode Park, Inc., New York.

These reactions, usually generally called annotations, can be accumulated.

This application of annotations is in support of moderated newsgroups.

1967 1951 1990 1992 1994

# GroupLens: An Open Architecture for Collaborative Filtering of Netnews



Paul Resnick\*, Neophytos Iacovou\*\*, Mitesh Suchak\*, Peter Bergstrom\*\*, John Riedl\*\*

\* MIT Center for Coordination Science  
Room E53-325  
50 Memorial Drive  
Cambridge, MA 02139  
617-253-8694  
Email: presnick@mit.edu

\*\* University of Minnesota  
Department of Computer Science  
Minneapolis, Minnesota 55455  
(612) 624-7372  
Email: riedl@cs.umn.edu

## ABSTRACT

Collaborative filters help people make choices based on the opinions of other people. GroupLens is a system for collaborative filtering of netnews, to help people find articles they will like in the huge stream of available articles. News reader clients display predicted scores and make it easy for users to rate articles after they read them. Rating servers, called Better Bit Bureaus, gather and disseminate the ratings. The rating servers predict scores based on the heuristic that people who agreed in the past will probably agree again. Users can protect their privacy by entering ratings under pseudonyms, without reducing the effectiveness of the score prediction. The entire architecture is open: alternative software for news clients and Better Bit Bureaus can be developed independently and can interoperate with the components we have developed.

**KEYWORDS:** Collaborative filtering, information filtering, electronic bulletin boards, social filtering, Usenet, netnews, user model, selective dissemination of information.

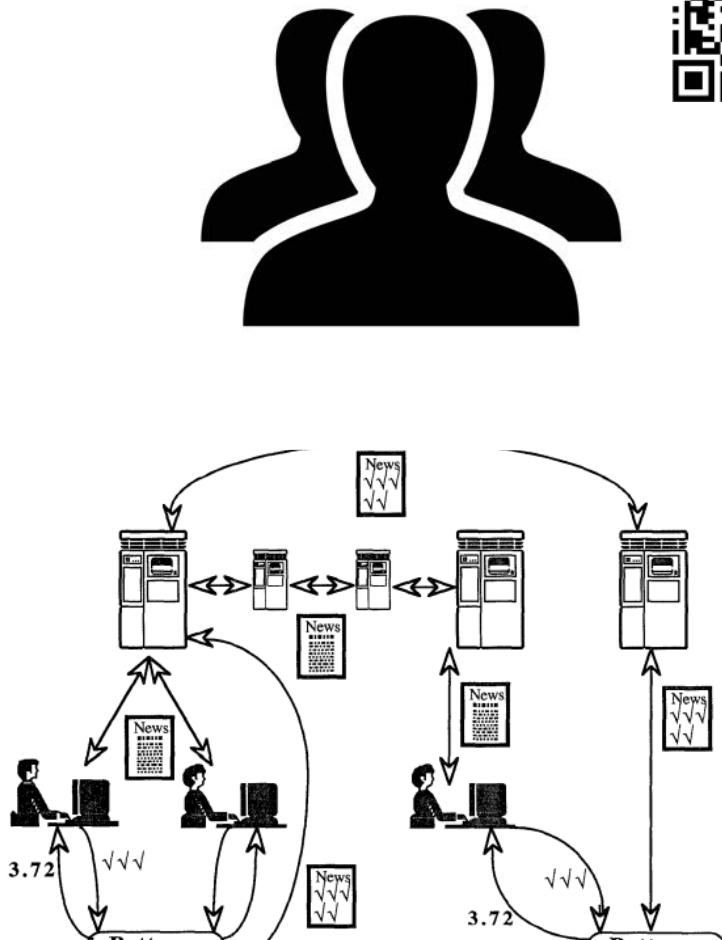
## INTRODUCTION

Computer networks allow the formation of interest groups that cross geographical barriers. Bulletin boards have been an important mechanism for that. Rather than addressing an article directly to a known set of people, the writer posts it in a newsgroup, a public place available to anyone interested in the topic. The Usenet netnews system creates the illusion of a single bulletin board available anywhere in the world. It propagates articles so that, with some delays, an article posted from anywhere in the world is available to everyone else.

newsgroups, with an average traffic of more than 100 per day<sup>1</sup>. The newsgroups carry announcements, quest and discussions. In a discussion, often called a thread article induces replies from several others, each of which may also induce replies. The January 24, 1994 estimated netnews participation indicate that more than 1 million people posted articles in the previous two weeks. To many more "lurkers" who read but do not post. Clearly, a lot of people are getting value from bulletin boards.

In fact, netnews' rapid broadcast nature and wide readership has reshaped the way the computing world works. System administrators depend on netnews to touch with the latest development work, the latest holes, and the latest bug fixes. Researchers use netnews as a way of keeping up-to-date on research directions and important results in between. Many others use netnews just to keep in touch with people around the world, to learn about new recipes, new music, and what life is like in other countries. Over the years netnews has become a principle for sharing among computer users.

Even so, the experience of using netnews is not always satisfying. Almost everyone complains that the noise ratio is too low. Writers cannot easily tell if their comments are valued, except by looking at post responses. Some seem not to care at all, only about their own right to write. Most writers, so that no one article will appeal to everyone in the newsgroup. Each reader ends up siftin-



# GroupLens: An Open Architecture for Collaborative Filtering of Netnews



Paul Resnick\*, Neophytos Iacovou\*\*, Mitesh Suchak\*, Peter Bergstrom\*\*, John Riedl\*\*

\* MIT Center for Coordination Science  
Room E53-325  
50 Memorial Drive  
Cambridge, MA 02139  
617-253-8694  
Email: presnick@mit.edu

\*\* University of Minnesota  
Department of Computer Science  
Minneapolis, Minnesota 55455  
(612) 624-7372  
Email: riedl@cs.umn.edu

## ABSTRACT

Collaborative filters help people make choices based on the opinions of other people. GroupLens is a system for collaborative filtering of netnews, to help people find articles they will like in the huge stream of available articles. News reader clients display predicted scores and make it easy for users to rate articles after they read them. Rating servers, called Better Bit Bureaus, gather and disseminate the ratings. The rating servers predict scores based on the heuristic that people who agreed in the past will probably agree again. Users can protect their privacy by entering ratings under pseudonyms, without reducing the effectiveness of the score prediction. The entire architecture is open: alternative software for news clients and Better Bit Bureaus can be developed independently and can interoperate with the components we have developed.

**KEYWORDS:** Collaborative filtering, information filtering, electronic bulletin boards, social filtering, Usenet, netnews, user model, selective dissemination of information.

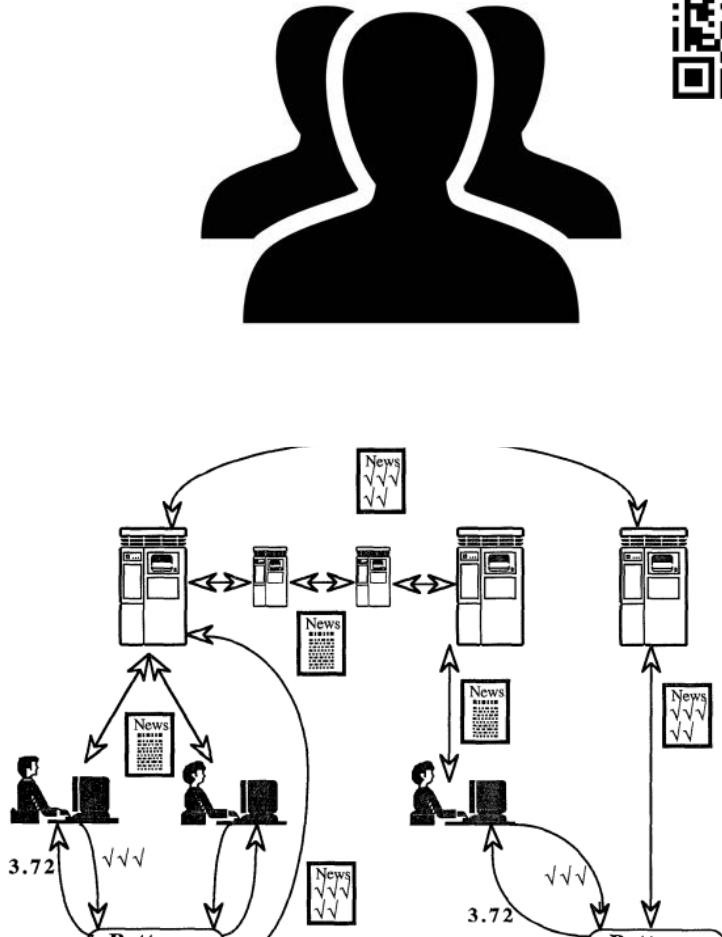
## INTRODUCTION

Computer networks allow the formation of interest groups that cross geographical barriers. Bulletin boards have been an important mechanism for that. Rather than addressing an article directly to a known set of people, the writer posts it in a newsgroup, a public place available to anyone interested in the topic. The Usenet netnews system creates the illusion of a single bulletin board available anywhere in the world. It propagates articles so that, with some delays, an article posted from anywhere in the world is available to everyone else.

newsgroups, with an average traffic of more than 100 per day<sup>1</sup>. The newsgroups carry announcements, quest and discussions. In a discussion, often called a thread article induces replies from several others, each of which may also induce replies. The January 24, 1994 estimated netnews participation indicate that more than 1 million users posted articles in the previous two weeks. To many more "lurkers" who read but do not post. Clearly, a lot of people are getting value from bulletin boards.

In fact, netnews' rapid broadcast nature and wide readership has reshaped the way the computing world works. System administrators depend on netnews to touch with the latest development work, the latest holes, and the latest bug fixes. Researchers use netnews as a way of keeping up-to-date on research directions and important results in between. Many others use netnews just to keep in touch with people around the world, to learn about new recipes, new music, and what life is like in other countries. Over the years netnews has become a principle for sharing among computer users.

Even so, the experience of using netnews is not satisfying. Almost everyone complains that the noise ratio is too low. Writers cannot expect their comments are valued, except by those who post responses. Some seem not to care at all about their own right to write. Most do so that no one article will appeal to everyone in the newsgroup. Each reader ends up siftin-



# Social Computing Research at the University of Minnesota

GroupLens advances the theory and practice of social computing by building and understanding systems used by real people

## ABSTRACT

Collaborative filtering systems have been used to predict user opinions on items such as movies, books, and news articles. No system can make it easy to find what you want to read or watch. Rating systems can disseminate user opinions based on the number of users who have rated an item. These systems will probably never be perfect, but they can be effective if they are designed well.

## KEYWORD

electronic commerce, user modeling

## INTRODUC-

Computer scientists and engineers that cross disciplines are an important part of our article directory. We are interested in new research that addresses the illusion of control that the world has over us. An article published in the journal "Everyone Else" is a good example of this kind of research.

## Featured Research

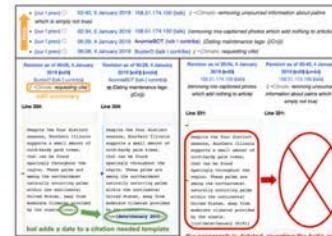
We publish research articles in conferences and journals primarily in the field of computer science, but also in other fields including psychology, sociology, and medicine. See our [blog](#) for research highlights and our [publications](#) page for a comprehensive view of our research contributions. Here are excerpts from recent articles:

### Rethinking Mental Health Interventions: How Crowd-Powered Therapy Can Help Everyone Help Everyone

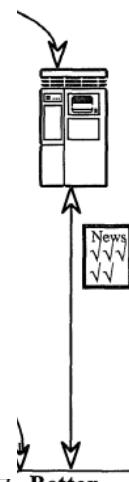


Rates of mental illness continue to rise every year. Yet there are nowhere near enough trained mental health professionals available to meet the need. How can technology create new ways to expand models of delivery for clinically

### Are Bots Ravaging Online Encyclopedias?



Wikipedia is the online encyclopedia that anyone can edit. However, you probably didn't know that "bots" also edit Wikipedia! Read this blog about conflict among bots—from the University of Minnesota's "human-centered computing for social good" REU. [More...](#)



Gr

P

\* M

## ABSTRACT

Collaborative filtering is a technique that uses the opinions of many users to predict what a user will like. It has been used in many applications, such as movie recommendation systems. In this paper, we present a system called MovieLens that provides personalized movie recommendations based on user ratings. We show that MovieLens can make it easier for users to find movies they will like by entering their ratings and interests. We also show that MovieLens can help users discover new movies that they might not have considered before.

## KEYWORD

electronic commerce, user modeling, recommendation systems

## INTRODUC

Computer recommendation systems that cross genres are an important part of article recommendation systems in a news application. Interested in the illusion of the world, I am writing an article for everyone else.

# movielens

Non-commercial, personalized movie recommendations.

[sign up now](#)

or [sign in](#)

## recommendations

MovieLens helps you find movies you will like. Rate movies to build a custom taste profile, then MovieLens recommends other movies for you to watch.

**top picks** see more

based on your ratings, MovieLens recommends these movies

Movie	Year	Length	Rating
Band of Brothers	2001	705 min	8
Casablanca	1942	102 min	8
One Flew Over the Cuckoo's Nest	1975	133 min	8
The Lives of Others	2006	137 min	8
Sunset Boulevard	1950	110 min	4
The Third Man	1949	104 min	8
Pat	1937	87 min	8

**recent releases** see more

movies released in last 90 days that you haven't rated

Movie	Year	Length	Rating
Cantinflas	2014	106 min	8
Felony	2014	88 min	8
What If	2014	102 min	8
Frank	2014	96 min	8
Sin City: A Dame to Kill For	2014	102 min	4
If I Stay	2014	106 min	8
Are We There Yet?	2014	88 min	8

An Algebra for Recommendations

Using Reader Data and Personalized Annotations

Jussi Karlgren

Department of Computer Science  
Stockholm University  
Electrum 230  
S-164 40 KISTA, Sweden

jussi@dsv.su.se  
jussi@flash.bellcorp.com  
jussi@sc.columbia.edu

**GroupLens: A Collaborative Filtering System for Movie Recommendations**

**Paul Resnick\***, Neel Guha  
\* MIT Center for Coordination of Technology and Society  
50 Memorial Drive, Cambridge, MA 02139  
Email: prensc@mit.edu

**ABSTRACT**  
Collaborative filters help people find items that other people like. Given a user's past history of purchases and ratings, these filters predict which articles they will like in the future. New reader clients display predicted item ratings and purchase probabilities. Predictions are based on the user's past history, and also on the user's rating history for other users. By enabling ratings under pseudonyms, users can protect their privacy. The system is currently used by over 200,000 users, and has been deployed on several Web sites.

**Keywords:** Collaborative filtering, electronic bulletin boards, social information model, selective dissemination

\*This work was partially funded by Defense Advanced Research Projects Agency (DARPA) contract AFRL-94-2-0037. This work was done while the author was at the University of Minnesota, and was partially funded by grants from the National Science Foundation and the Minnesota Department of Transportation. Part of this work was done while the author was at the Research Institute for Interactive Media, Media Park, Berlin, Germany.

**REPORTER'S NOTE**  
Collaborative networks allow the fast spread of documents. A benefit of these networks is that users can share a document with others. Such reactions, usually negative, can be used to refine the document further.

**REPORTER'S NOTE**  
Collaborative networks allow the fast spread of documents. A benefit of these networks is that users can share a document with others. Such reactions, usually negative, can be used to refine the document further.

**REPORTER'S NOTE**  
Collaborative networks allow the fast spread of documents. A benefit of these networks is that users can share a document with others. Such reactions, usually negative, can be used to refine the document further.

**REPORTER'S NOTE**  
Collaborative networks allow the fast spread of documents. A benefit of these networks is that users can share a document with others. Such reactions, usually negative, can be used to refine the document further.

## Item-Based Collaborative Filtering Recommendation Algorithms

Badrul Sarwar, George Karypis, Joseph Konstan, and John Reidl

GroupLens Research Group/Army HPC Research Center  
Department of Computer Science and Engineering  
University of Minnesota, Minneapolis, MN 55455

**ABSTRACT**  
Recommendation systems apply knowledge discovery techniques to the problem of making personalized recommendations for information, products or services during a live interaction. Collaborative filtering, one of the most promising such technologies, is achieving widespread success on the Web. In this paper, we discuss item-based collaborative filtering based on user-item rating matrices. We show how item-based filtering handles missing data well, and how it can scale to millions of items. We note that the neighbors like are then recomputed frequently, and we propose an incremental algorithm for collaborative filtering. In traditional collaborative filtering systems the amount of work increases with the number of participants. To address this problem, we propose a hybrid approach that is needed that can quickly produce high quality recommendations. In addition, we propose a hybrid approach that is needed that can quickly produce high quality recommendations. In addition, we propose a hybrid approach that is needed that can quickly produce high quality recommendations.

**In this paper we analyze different item-based recommendation generation algorithms. We look into the impact of item correlation vs. cosine similarity between item vectors and item-based filtering algorithms. We also compare item-based filtering with other models (e.g., weighted sum vs. regression model). Finally, we experimentally evaluate our results and compare them to the state-of-the-art item-based algorithms. Our experiments show that item-based algorithms provide dramatically better recommendations than the best available user-based algorithms.**

**KEYWORDS:** Collaborative filtering, item-based filtering, item-based recommendation generation, item correlation, cosine similarity, user-based filtering, user-based recommendation generation, weighted sum, regression model.

**REPORTER'S NOTE**  
Collaborative networks allow the fast spread of documents. A benefit of these networks is that users can share a document with others. Such reactions, usually negative, can be used to refine the document further.

**REPORTER'S NOTE**  
Collaborative networks allow the fast spread of documents. A benefit of these networks is that users can share a document with others. Such reactions, usually negative, can be used to refine the document further.

**REPORTER'S NOTE**  
Collaborative networks allow the fast spread of documents. A benefit of these networks is that users can share a document with others. Such reactions, usually negative, can be used to refine the document further.

**REPORTER'S NOTE**  
Collaborative networks allow the fast spread of documents. A benefit of these networks is that users can share a document with others. Such reactions, usually negative, can be used to refine the document further.

**REPORTER'S NOTE**  
Collaborative networks allow the fast spread of documents. A benefit of these networks is that users can share a document with others. Such reactions, usually negative, can be used to refine the document further.

**REPORTER'S NOTE**  
Collaborative networks allow the fast spread of documents. A benefit of these networks is that users can share a document with others. Such reactions, usually negative, can be used to refine the document further.

1967  
1951 1990  
1992 1994 2001

# Item-Based Collaborative Filtering Recommendation Algorithms

Badrul Sarwar, George Karypis, Joseph Konstan, and John Riedl

{sarwar, karypis, konstan, riedl}@cs.umn.edu

GroupLens Research Group/Army HPC Research Center  
Department of Computer Science and Engineering  
University of Minnesota, Minneapolis, MN 55455



## ABSTRACT

Recommender systems apply knowledge discovery techniques to the problem of making personalized recommendations for information, products or services during a live interaction. These systems, especially the k-nearest neighbor collaborative filtering based ones, are achieving widespread success on the Web. The tremendous growth in the amount of available information and the number of visitors to Web sites in recent years poses some key challenges for recommender systems. These are: producing high quality recommendations, performing many recommendations per second for millions of users and items and achieving high coverage in the face of data sparsity. In traditional collaborative filtering systems the amount of work increases with the number of participants in the system. New recommender system technologies are needed that can quickly produce high quality recommendations, even for very large-scale problems. To address these issues we have explored item-based collaborative filtering techniques. Item-based techniques first analyze the user-item matrix to identify relationships between different items, and then use these relationships to indirectly compute recommendations for users.

In this paper we analyze different item-based recommendation generation algorithms. We look into different techniques for computing item-item similarities (e.g., item-item correlation vs. cosine similarities between item vectors) and different techniques for obtaining recommendations from them (e.g., weighted sum vs. regression model). Finally, we experimentally evaluate our results and compare them to the basic k-nearest neighbor approach. Our experiments suggest that item-based algorithms provide dramatically better performance than user-based algorithms, while at the same time providing better quality than the best available user-

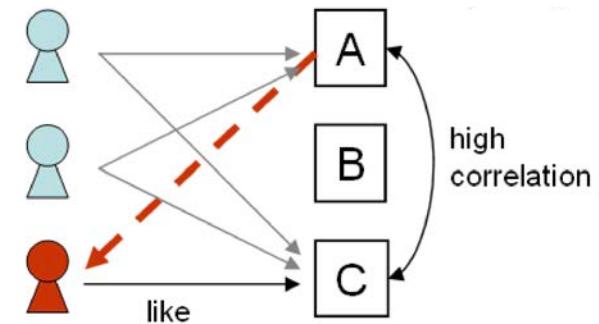
through all the available information to find that which is most valuable to us.

One of the most promising such technologies is *collaborative filtering* [19, 27, 14, 16]. Collaborative filtering works by building a database of preferences for items by users. A user, Neo, is matched against the database to discover *neighbors*, which are other users who have historically had similar taste to Neo. Items that the neighbors like are then recommended to Neo, as he will probably also like them. Collaborative filtering has been very successful in both research and practice, and in both information filtering applications and E-commerce applications. However, there are important research questions in overcoming two fundamental challenges for collaborative filtering recommend

The first challenge is to improve the scalability of collaborative filtering algorithms. These algorithms search tens of thousands of potential neighbors but the demands of modern systems are to search millions of potential neighbors. Further, existing algorithms have performance problems with individual users if the site has large amounts of information. If a site is using browsing patterns as individual user preference, it may have thousands of distinct frequent visitors. These "long tail" users pose a challenge to the number of neighbors that can be searched, further reducing scalability.

The second challenge is to improve the quality of recommendations for the users. Users need to trust the system they use to help them find items they like. They will "vote with their feet" by refusing to use systems that are not consistently accurate.

In some ways these two challenges are related: the less time an algorithm spends searching for neighbors, the more scalable it will be, and the worse



An Algebra for Recommendations

Using Reader Data a

Jussi Karlgren

Department of Computer Science  
Stockholm University  
Electrum 230  
S-164 40 KISTA, Sweden

jussi@dsv.su.se  
jussi@flash.bellcore.com  
jussi@sc.columbia.edu

**GroupLens: A Case Study in Collaborative Filtering**

Paul Resnick\*, Neeraj Ravasi, and J. David Konstan

\* MIT Center for Coordination of Compute and Information Systems  
50 Memorial Drive, Cambridge, MA 02139  
Email: prenick@mit.edu

**ABSTRACT**

Collaborative filters help people find opinions of other people. GroupLens is a system that helps predict what articles will likely interest a user. News reader clients display news items from many sources, and users can rate the news items. Rating servers, called Better Bit Distributors, collect these ratings and base recommendations on them. Ratings are based on the heuristic that people who like similar things tend to like each other. By entering ratings under "readability" effectiveness of the news predictor is improved. The news predictor can now be used to predict what news items a user will like. This predictor can be used to predict what news items a user will like. This predictor can be used to predict what news items a user will like.

**KEYWORDS:** Collaborative filtering, electronic bulletin boards, social information retrieval, selective dissemination of information.

**REPORTS AS A RECOMMENDER SYSTEM**

The most interesting part of the report is the section on "Reports as a recommender system". It discusses how users can use reports to find documents they are interested in. It also discusses how reports can be used to find documents in a document filtering system.

**COMPUTER NETWORKS AS A RECOMMENDER SYSTEM**

The paper also discusses computer networks as a recommender system. It explains how users can use computer networks to find documents they are interested in. It also discusses how computer networks can be used to find documents in a document filtering system.

**REFERENCES**

These references, unless generally called annotations, can be accessed by clicking on the reference number.

1st generation of  
recommender systems

1967  
1951 1990  
1992 1994 2001



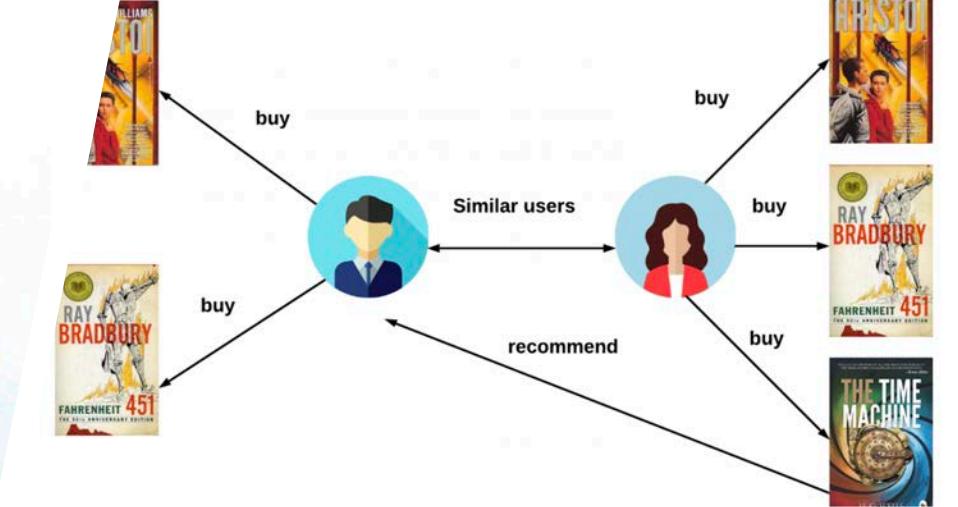
## 1st generation of recommender systems

- 90s-00s
- User/item-based CF, content-based CF, rule-based, knowledge graph-based
- Intuitive algorithms
- Easy to explain



## 1st generation of recommender systems

- 90s-00s
- User/item-based CF, content-based CF, rule-based, knowledge graph-based
- Intuitive algorithms
- Easy to explain



Customers who bought this item also bought

Solatium (Emulations, an urban fantasy series Book 2) Becca Mills ★★★★★ 147 Kindle Edition \$3.99	Wings of Flesh and Bones Cathrina Constantine ★★★★★ 11 Kindle Edition \$3.99	The Dragon Songs Saga: The Complete Quartet: Songs of Insurrection,... JC Kang ★★★★★ 55 Kindle Edition \$0.99	Alone: A sci-fi reverse harem (The Mars Diaries Book 1) Skye MacKinnon ★★★★★ 41 Kindle Edition \$0.99	The Complete Four Worlds Series: An Epic Fantasy Saga: Books 1-4 Angela J. Ford ★★★★★ 15 Kindle Edition \$0.99	Shaxoa's Gift (T Saga Book 2) DelSheree Gladd ★★★★★ 3 Kindle Edition \$3.99

Sponsored products related to this item (What's this?)





# Netflix Prize

COMPLETED

[Home](#) | [Rules](#) | [Leaderboard](#) | [Update](#)

## Leaderboard

Showing Test Score. [Click here to show quiz score](#)

Display top  leaders.

Rank	Team Name	Best Test Score	% Improvement	Best Submit Time
<b><u>Grand Prize - RMSE = 0.8567 - Winning Team: BellKor's Pragmatic Chaos</u></b>				
1	<a href="#">BellKor's Pragmatic Chaos</a>	0.8567	10.06	2009-07-26 18:18:28
2	<a href="#">The Ensemble</a>	0.8567	10.06	2009-07-26 18:38:22
3	<a href="#">Grand Prize Team</a>	0.8582	9.90	2009-07-10 21:24:40
4	<a href="#">Opera Solutions and Vandelay United</a>	0.8588	9.84	2009-07-10 01:12:31
5	<a href="#">Vandelay Industries !</a>	0.8591	9.81	2009-07-10 00:32:20

## An Algebra for Recommendations

## Using Reader Data a

Jussi Karlgren

Department of Computer Science  
Stockholm University  
Electrum 230  
S-164 40 KISTA, Sweden

jussi@dsv.su.se  
jussi@flash.bellcore.com  
jussi@sc.columbia.edu

These annotations, mostly general comments, can be accessed by clicking on the vertical red bars.

## GroupLens: A Case Study

Paul Resnick\*, Neel Guha

\* MIT Center for Coordination of Compute and Communication  
50 Memorial Drive, Room 320  
Cambridge, MA 02139  
Email: prenick@mit.edu

### ABSTRACT

Collaborative filters help people to find items they will like based on the opinions of other people. Given the large number of items available, users will likely be overwhelmed by reading ratings from many people. New reader clients display personalized item recommendations based on collaborative filtering, called Better B Ratings. These recommendations are based on the heuristic that people who have rated the same items tend to have similar tastes. By enabling ratings under a pay-per-view effect of the user prediction system, new readers can quickly produce recommendations for items that they have not yet rated. In this paper we describe how these filters can be developed independently of the underlying recommendation system, and then use these relationships to improve the system.

In this paper we analyze different item rating generation algorithms. We look at how different rating generation methods compare to each other in terms of correlation vs. cosine similarity between users. We also compare different rating generation methods to a baseline of random ratings. Finally, we present a measurement of how closely predicted ratings of movies match subsequent actual ratings. If you develop a system that we say is better than a random rating system, it's probably better than most systems. (If you know what you're doing, you know there would be a catch, right?) Only if you share your method with us and describe to the world how you did it and why it worked, though, would we believe it.

**KEYWORDS:** Collaborative filters, electronic bulletin boards, social item model, selective dissemination

These annotations, mostly general comments, can be accessed by clicking on the vertical red bars.

## Item-Based Collaborative Filtering Recommendation Algorithms

Badrul Sarwar, George Karypis, Joseph Konstan, and John Riedl  
[sarwar, karypis, konstan, riedl]cs.umn.edu  
GroupLens Research Group/Army HPC Research Center  
University of Minnesota

### NETFLIX

**COMPLETED**

**The Netflix Prize Rules**

For a printable copy of these rules, go [here](#).

**OVERVIEW:**

We're quite curious, really. To the tune of one million dollars.

Netflix is all about connecting people to the things they love. To help customers find those movies, we've developed our world-famous collaborative filtering system, Cinematch™. As you'd expect, it's great at predicting whether someone will enjoy a movie based on how much they liked or disliked other movies. We use those predictions to make personal movie recommendations based on each customer's viewing history.

Now there are a lot of interesting alternative approaches to how Cinematch works that we haven't tried. Some are described in the literature, some aren't. We're curious whether any of these can beat Cinematch by making better predictions. Because, well, that's what we do.

So, we thought we'd make it a contest out of finding the answer. It's "easy" really. We provide you with a lot of anonymized rating data from our users. You write a program that takes that data and generates new ratings. Then we'll compare them to ours. That's it. It's a measurement of how closely predicted ratings of movies match subsequent actual ratings. If you develop a system that we say is better than a random rating system, it's probably better than most systems. (If you know what you're doing, you know there would be a catch, right?) Only if you share your method with us and describe to the world how you did it and why it worked, though, would we believe it.

Before money becomes a service loss. We suspect the 10% improvement is pretty tough, but we also think there is a good chance that it can be achieved. It may take months, it might take years. So to keep things interesting, in addition to the Grand Prize, we'll award smaller prizes for milestones along the way. The first one is for the system that achieves the most improvement over the previous year's best accuracy bar on the same qualifying test set. No improvement, no prize. And the third one is for the system that achieves the most improvement over the previous year's best accuracy bar on the same qualifying test set.

There is no cost to enter, no purchase required, and you need not be a Netflix subscriber. So if you know (or want to learn) something about machine learning and recommendation systems, give it a shot. We could make it really worth your while.

**Terms and Conditions in a Nutshell**

- Contest begins October 2, 2006 and continues through at least October 2, 2011.
- Contest is open to anyone, anywhere (except certain countries listed below).

1967  
1951 1990  
1992 1994 2001 2006

## An Algebra for Recommendations

## Using Reader Data a

Jussi Karlgren

Department of Computer Science  
Stockholm University  
Electrum 230  
S-164 40 KISTA, Sweden

jussi@dsv.su.se  
jussi@flash.bellcore.com  
jussi@sc.columbia.edu

These annotations, more generally called annotations, can be accessed by clicking on the document. See also the section "Annotations".

## GroupLens: A

Paul Resnick\*, Neelip

\* MIT Center for Coordination of Compute and Information Systems  
30 Memorial Drive  
Cambridge, MA  
Email: prenick@mit.edu

## ABSTRACT

Collaborative filters help people to find articles they will like in the news reader. News reader clients display news items from their users' reading servers, called Better Browsers. These filters are based on the heuristic that people who have liked one item will like others by entering ratings under previous entries of the user prediction system. News items are selected by the better browser.

Browsers can be developed independently of the news reader.

KEYWORDS: Collaborative filters, electronic bulletin boards, user model, selective dissemination

These annotations are used to support the discussion in the paper "An Algebra for Recommendations".

In addition to helping users find news items, such readers can be used to support users in finding other filters. One application of annotations is in support of moderation of news feeds.

## Item-Based Collaborative Filtering Recommendation Algorithms

Badrul Sarwar, George Karypis, Joseph Konstan, and John Riedl  
[email] [email] [email] [email]  
GroupLens Research Group/Army HPC Research Center  
Depart. of Computer Science, University of Minnesota, Minneapolis, MN 55455



## NETFLIX Prize

### COMPLETED

### The Netflix Prize Rules

For a printable copy of these rules, go here.

#### Overview:

We're quite curious, really. To the tune of one million dollars.

It's all about connecting people to the movies they love. To help customers find those movies, we'd developed our world's

# Recommendation = Rating prediction

1967  
1951 1990  
1992 1994 2001 2006



## An Algebra for Recommendations

## Using Reader Data a

Jussi Karlgren

Department of Computer Science  
Stockholm University  
Electrum 230  
S-164 40 KISTA, Sweden

jussi@dsv.su.se  
jussi@flash.bellcore.com  
jussi@sc.columbia.edu

Report by Jussi Karlgren  
on the Netflix Prize competition  
and its implications for recommendation systems.

Ericsson, Inc., Princeton, NJ,  
University of California, Berkeley

## GroupLens: A

Paul Resnick\*, Neelie

\* MIT Center for Coordination  
of Technology, Policy, and Law  
50 Memorial Drive  
Cambridge, MA  
Email: prenick@mit.edu

### ABSTRACT

Collaborative filters help people make better decisions by combining the opinions of other people. GroupLens researchers have developed several different kinds of collaborative filters that have been used in real-world applications such as movie recommendations, news feeds, and product reviews.

**KEYWORDS:** Collaborative filtering, electronic bulletin boards, social information retrieval, user model, selective dissemination

## Item-Based Collaborative Filtering Recommendation Algorithms

Badrul Sarwar, George Karypis, Joseph Konstan, and John Riedl  
[sarwar, karypis, konstan, riedl]@cs.wisc.edu  
GroupLens Research Group/Army HPC Research Center  
Department of Computer Sciences  
University of Wisconsin-Madison

### NETFLIX



## Recommendation

prediction



[Followup to [hi]]

Ok, so here's what I tell all about how I (know we) got to be tied for third place on the [netflix prize](#). And I don't mean to make a waste of competing in the judge, but rather the actual math and methods. So yes, after reading this post, you too should be able to rank in the top ten or so.

Ur... yesterday's top ten anyway.

1967  
1951 1990  
1992 1994 2001 2006

**Monday, December 11, 2006**

*Netflix Update: Try This at Home*



[Followup to this]

Ok, so here's where I tell all about how I (now we) got to be tied for third place on the [netflix prize](#). And I don't mean a sordid tale of computing in the jungle, but rather the actual math and methods. So yes, after reading this post, you too should be able to rank in the top ten or so.

Ur... yesterday's top ten anyway.

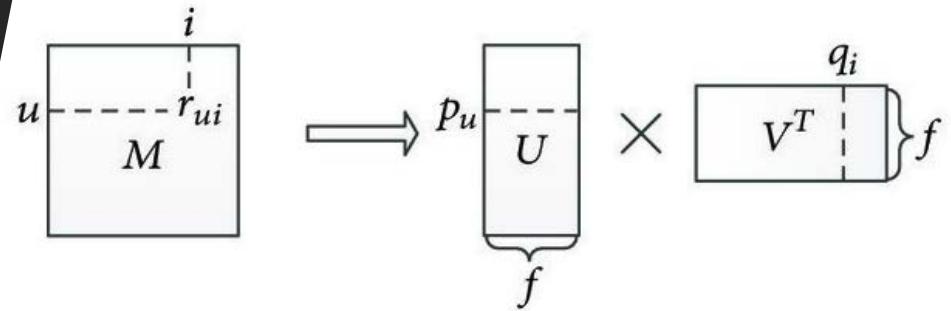
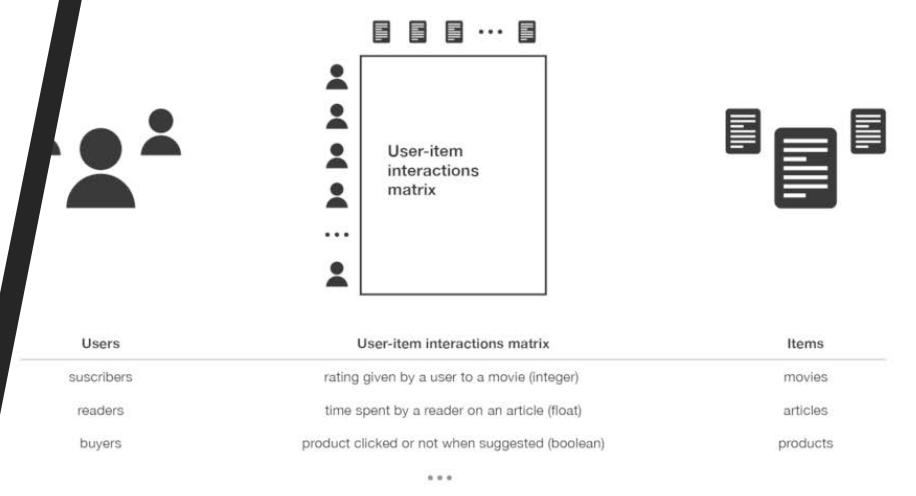
My first disclaimer is that our last submission which tied for third place was only actually good enough for ninth place or so. It landed where it did because, just for giggles and grins, we blended results (50/50) with [Jetrays](#) who had a similar score to us at the time.

Second, my friend Vincent has been manning the runs on his desktop machines, diligently fine tuning and squeezing out every last bit of performance possible with whatever controls I could give him (not to mention learning python so he could write scripts to blend submissions and whatnot). In short, almost all my progress since my last post has been due to other people. In the meantime I've implemented a handful of failed attempts at improving the performance, plus one or two minorly successful ones which I'll get to.

Now for the method to the mathness, beginning with a review of the problem:

Netflix provided 100M ratings (from 1 to 5) of 17K movies by 500K users. These essentially arrive in the form of triplet of numbers: (User,Movie,Rating). E.g., one such entry might be (105932,14002,3). Times 100 million. No go make sense of it. In particular, for (User,Movie,?) not in the database, tell me what the Rating would be—that is, predict how the given User would rate the given Movie.

I'm tempted to get all philosophical on my soap box here and go into ways of thinking about this stuff and modeling vs function mapping approaches, yadda yadda, but I know you all are just here for the math, so I'll that for the next hapless hosteler who asks me what I do for a living.



## An Algebra for Recommendations

Using Reader Data a

Jussi Karlgren

Department of Computer Science  
Stockholm University  
Electrum 230  
S-164 40 KISTA, Sweden

jussi@dsv.su.se  
jussi@flash.bellcore.com  
jussi@sc.columbia.edu



### GroupLens: A Collaborative Filtering System for Movie Recommendation

Paul Resnick\*, Neel Guha  
\* MIT Center for Coordination of Compute and Communication  
50 Memorial Drive, Room 320  
Cambridge, MA 02139  
Email: prenres@mit.edu

#### ABSTRACT

Collaborative filters help people to find items they will like based on the opinions of other people. GroupLens is a system that helps people to find items they will like in the浩大的 space of possible items. New reader clients display movie recommendations from the GroupLens system. Rating servers, called Better Buttons, store user ratings and provide them to the system based on the heuristic that people who like the same things also like each other. The system improves its recommendations by making ratings under "playback" effectiveness of the user prediction. The new prediction is based on the user's rating history and the user's rating of the item. Buttons can be developed independently of the system.

#### KEYWORDS:

Collaborative filters, electronic bulletin boards, user model, selective dissemination

Report is an invited paper.  
The material contained in this report has been submitted to the IEEE International Conference on Data Mining (ICDM) and is being considered for publication in the conference proceedings. Copy right © 2006, IEEE. Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution must be obtained from IEEE. Contact: 317-502-0011.

We are grateful to the anonymous reviewers for their useful comments and suggestions. This work was partially funded by grants from the Swedish Research Council, the Knut and Alice Wallenberg Foundation, the Swedish Research Institute for Information Technology, and the Royal Swedish Academy of Sciences.

George Fox, Ph.D.  
University of Oregon, Eugene

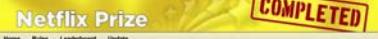
These restrictions, known generally as legal annotations, can affect other filters. One application of annotations is in support of moderate

## Item-Based Collaborative Filtering Recommendation Algorithms

Badrul Sarwar, George Karypis, Joseph Konstan, and John Riedl  
GroupLens Research Group/Army HPC Research Center  
Department of Computer Science  
University of Minnesota



NETFLIX



COMPLETED

#### ABSTRACT

Recommender systems apply knowledge to the problem of making personalized information, products or services due to the lack of time and effort of performing many rounds of trial and error. In the presence of data sparsity, in particular, the amount of work required to find the best recommendations needed that can be done in a reasonable amount of time.

## Recommenders06 - Late Summer School on The Present and Future of Recommender Systems

i On September 12-13, 2006 the Summer School will study the Present and Future of Recommender Systems. The Summer School will be held in the idyllic setting of the Euskalduna Palace in Bilbao, Spain.

## ction

### Introduction

Recommender Systems are applications that provide personalized advice to users about products or services they might be interested in. Recommender Systems are playing a major role in the Digital and Social Networking Revolution and becoming a part of everyday life. They are helping people efficiently manage content overload and dive into the long tail of content discovery. The social prevalence of this can be evidenced by the evolution of, and demand for, personalized radio, television, video and on-line shopping.

The Summer School will bring together the concepts and practices of Recommender Systems and is intended for both researchers (including Ph.D. students) and for professionals and managers who want to benefit from the best information on Recommender Systems and personalization advances. An in-depth introduction to Recommender Systems will be provided.

Participation in the Summer School is by acceptance only. Ten scholarships for promising students will be granted.

### Lecturers will include:

Chris Anderson, Wired Magazine, US  
Todd Beaupre, Yahoo, Inc., US  
Jim Bennett, Netflix, Inc., US  
Dr. Alexander Tsvetkov, Klagenfurt University, Austria  
Dr. Rick Hengartner, MusicStrands, Inc., US  
Prof. Juntae Kim, Dongguk University, Korea  
Kaushal Kurapati, Ask.com, US To the lectures  
Paul Lamere, Sun Microsystems Labs, US  
Dr. Bamshad Mobasher, DePaul University, US  
Mike Mudie, Yahoo, Inc., US



1967  
1951 1990  
1992 1994 2001 2006



# An Algebra for Recommendations

## Using Reader Data as Input

Jussi Karlgren

Department of Computer Science  
Stockholm University  
Electrum 230  
S-164 40 KISTA, Sweden

jussi@dsv.su.se  
jussi@flash.bellcore.com  
jussi@sc.columbia.edu



### GroupLens: A Collaborative Filtering System for Movie Recommendation

Paul Resnick\*, Neel Guha\*

\* MIT Center for Coordination of Compute and Communications  
50 Memorial Drive, Room 320  
Cambridge, MA 02139-1421

Email: prenick@mit.edu  
guha@csail.mit.edu

**ABSTRACT**  
Collaborative filters help people make better decisions by using the opinions of other people. GroupLens.org is a collaborative filtering system that helps people find articles they will like in the digital library of the MIT Media Lab. New reader clients display movie ratings and reviews from other users, and new authors can rate articles based on the heuristic that people who liked the same things are more likely to like each other's work. By enabling ratings under pseudonyms, the system protects the privacy of its users while still providing useful information.

**Keywords:** Collaborative filters, electronic bulletin boards, user model, selective dissemination of information.



These reactions, usually generally called annotations, can be used to refine the user's profile and thus improve the recommendations. One application of annotations is in the support of moderate filters.

## Item-Based Collaborative Filtering Recommendation Algorithms

Badrul Sarwar, George Karypis, Joseph Konstan, and John Riedl  
GroupLens Research Group/Army HPC Research Center  
Department of Computer Science  
University of Minnesota

NETFLIX



### ABSTRACT

Recommender systems apply knowledge to the problem of making personalized information, products or services more interesting. Collaborative filtering based on user ratings has been one of the most effective information and recommendation technologies for several years. Since most users have rated only a small fraction of items, performing many recommendations requires dealing with sparse data and data sparsity. In order to handle data sparsity, in traditional collaborative filtering systems, many more ratings are needed than can be obtained from the user. This paper shows how temporal dynamics can be used to obtain recommendations even from sparse data.

**Keywords:** Collaborative filtering, recommendation systems, content-based filtering, temporal dynamics, item-based filtering, rating prediction, user modeling, selective dissemination of information.

**Abstract**  
Collaborative filters help people make better decisions by using the opinions of other people. GroupLens.org is a collaborative filtering system that helps people find articles they will like in the digital library of the MIT Media Lab. New reader clients display movie ratings and reviews from other users, and new authors can rate articles based on the heuristic that people who liked the same things are more likely to like each other's work. By enabling ratings under pseudonyms, the system protects the privacy of its users while still providing useful information.

**Keywords:** Collaborative filters, electronic bulletin boards, user model, selective dissemination of information.

## Collaborative Filtering with Temporal Dynamics

Yehuda Koren  
Yahoo! Research, Haifa, Israel  
yehuda@yahoo-inc.com



### Introduction

Recommender Systems are applications that are becoming a part of everyday life. They content discovery. The social prevalence television, video and online shopping.

**Abstract**  
Customer preferences for products are shifting over time. Product perception and popularity are constantly changing as new selection criteria are introduced and old ones are removed, and to them to ever redefine their taste. Thus, modeling temporal dynamics should be a key when designing recommender systems or recommendation engines. This paper discusses some of the challenges. Within the eco-system interacting multiple products and consumers, many different characteristics are shifting simultaneously, which makes it hard to predict which products and consumer shifts are durable and associated with a few data instances. This paper discusses how to model temporal dynamics. It is shown that mostly a single concept is tracked. Classical time-window or instance decay approaches cannot work, as they lose too much signal when they decay. Instead, we propose a hybrid approach, which can make better distinctions between transient effects and long-term patterns. The paradigm we offer is creating a model that tracks changes over time, but does not forget the global context of the data. This allows us to exploit the relevant components of all data instances, while discarding only those as being irrelevant.

**Keywords:** The need to model time changes at the level of each individual consumer, while tracking changes at the level of the whole population. This should resort to more accurate techniques than those that suffice for modeling global changes. For example,

temporary effects that have very low impact on future behavior, while capturing longer term trends that reflect the inherent nature of the data. This led to new works on recommendation, which is also related to the field of digital marketing [15, 25]. Modeling temporal changes in customer preferences brings unique challenges. One kind of concept drift in this setting is the emergence of new products and consumers, which are often not well represented in the data. Related to this are seasonal changes, or specific holidays, which lead to temporary changes in consumer behavior. Another type of drift is when the whole population shifts from one taste to another. These shifts are durable and associated with a few data instances. This paper discusses how to model temporal dynamics. It is shown that mostly a single concept is tracked. Classical time-window or instance decay approaches cannot work, as they lose too much signal when they decay. Instead, we propose a hybrid approach, which can make better distinctions between transient effects and long-term patterns. The paradigm we offer is creating a model that tracks changes over time, but does not forget the global context of the data. This allows us to exploit the relevant components of all data instances, while discarding only those as being irrelevant.

Accordingly, we propose new local-temporal collaborative filtering recommendation approaches. Evaluation is made on a large movie

dataset and results show that our approach outperforms state-of-the-art

### Lecturers will include:

Chris Anderson, Wired Magazine, US  
Todd Beaupre, Yahoo, Inc., US  
Jim Bennett, Netflix, Inc., US  
Dr. Alexander Tellerer, Klagenfurt University, Austria  
Dr. Rick Hengartner, MusicStrands, Inc., US  
Prof. Juntae Kim, Dongguk University, Korea  
Kaushal Kurapati, Ask.com, US  
Paul Lamere, Sun Microsystems Labs, US  
Dr. Bamshad Mobasher, DePaul University, US  
Mike Modl, Yahoo!, Inc., US



1951 1967 1990 1992 1994 2001 2006 2008

## An Algebra for Recommendations

### Using Reader Data as Input

Jussi Karlgren

Department of Computer Science  
Stockholm University  
Electrum 230  
S-164 40 KISTA, Sweden

jussi@dsv.su.se  
jussi@flash.bellcore.com  
jussi@sc.columbia.edu

**GroupLens: A Collaborative Filtering System for Movie Recommendations**

**Paul Resnick\***, Neel Guha, and Jussi Karlgren

\* MIT Center for Coordination of Compute and Human Activity, 30 Memorial Drive, Cambridge, MA 02139, USA  
Email: prenres@mit.edu

**ABSTRACT**  
Collaborative filters help people make better decisions by combining the opinions of other people. GroupLens is a system that recommends articles they will like in digital libraries. New reader clients display personalized recommendations from the system. Recommendation servers, called BetterBrowsers, provide recommendations based on the user's history of interactions with the system. Recommendations are generated by mining ratings under the guidance of the user's previous behavior. In this paper we show how recommendations can be generated independently of the user's previous behavior. BetterBrowsers can be developed independently of the user's previous behavior.

**KEYWORDS:** Collaborative filtering, electronic bulletin boards, user model, selective dissemination of information

Copyright © 2006, IUI '06, November 12-15, 2006, Seattle, WA, USA. This material is based upon work supported by the National Science Foundation under grants IIS-0238236 and IIS-0238237. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. This work was partially funded by grants from the National Science Foundation, the Defense Advanced Research Projects Agency, and the National Institute of Standards and Technology. Part of this work was done while Jussi Karlgren was at the Research Institute of the Future, DePaul University.

These reactions, more generally called annotations, can be used to refine filters. One application of annotations is in the support of moderate filters.

## Item-Based Collaborative Filtering Recommendation Algorithms

Badrul Sarwar, George Karypis, Joseph Konstan, and John Riedl  
GroupLens Research Group/HPC Research Center  
Department of Computer Science, University of Minnesota



COMPLETED!

### Collaborative Filtering



On September 12-13, 2006 The Summer School will be held

### Introduction

Recommender Systems are applications that are shifting over perception and popularity are constantly changing an interesting field of research. Including them to ever redefine their tasks. Thus, modeling is needed to be able to predict what items should be key when designing recommender systems. Within the eco-system interacting media and consumers, many different characteristics are shared. These characteristics are often very different and shifts are difficult and associated with a few data instances. This is a challenge for recommendation systems, mostly a single concern is tracked. Classical time-winds decay approaches cannot work, as they lose too much information. Instead, it is better to track the user's interests which can make better distinctions between transient long-term patterns. The paradigm we offer is creating user profiles that change over time, based on each user's data. This allows us to exploit the relevant component information while discarding only what is noise at the same time. Accordingly, we propose a novel collaborative recommendation approach. Evaluation is made on a

### ABSTRACT

Customer preferences for products are shifting over time and popularity are constantly changing an interesting field of research. Including them to ever redefine their tasks. Thus, modeling is needed to be able to predict what items should be key when designing recommender systems. Within the eco-system interacting media and consumers, many different characteristics are shared. These characteristics are often very different and shifts are difficult and associated with a few data instances. This is a challenge for recommendation systems, mostly a single concern is tracked. Classical time-winds decay approaches cannot work, as they lose too much information. Instead, it is better to track the user's interests which can make better distinctions between transient long-term patterns. The paradigm we offer is creating user profiles that change over time, based on each user's data. This allows us to exploit the relevant component information while discarding only what is noise at the same time. Accordingly, we propose a novel collaborative recommendation approach. Evaluation is made on a

set of products and consumer characteristics including their purchase history and demographic information. The results show that our approach outperforms state-of-the-art collaborative filtering models.

Participation in the Summer School is free.

### Lecturers will include:

Chris Anderson, Wired Magazine, US  
Todd Beaupre, Yahoo, Inc., US  
Jim Bennett, Netflix, Inc., US  
Dr. Alexander Felfernig, Klagenfurt University, Austria  
Dr. Rick Houghtaling, MusicStrands, Inc., US  
Prof. Juntae Kim, Dongguk University, Korea  
Kaushal Kurapati, Ask.com, US  
Paul Lamere, Sun Microsystems Labs, US  
Dr. Bamshad Mobasher, DePaul University, US  
Mike Mudie, Yahoo!, Inc., US

**NETFLIX PRIZE**

COMPLETED!

**COLLABORATIVE FILTERING**

**MATRIX FACTORIZATION TECHNIQUES FOR RECOMMENDER SYSTEMS**

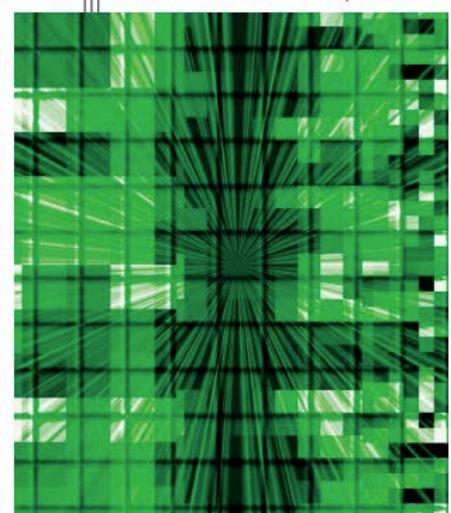
**Fréderic Korke, Yahoo Research**  
Robert Bell and Chris Volinsky, AT&T Labs—Research

**ABSTRACT**  
As the Netflix Prize competition has demonstrated, matrix factorization models are superior to classic nearest-neighbor techniques for producing product recommendations, allowing the incorporation of additional information such as user feedback levels.

**Such systems are particularly useful for entertainment products such as movies, music, and TV shows. Many customers will view the same movie, and each customer is likely to have a different rating for it. This makes it difficult to predict what movie a particular customer prefers willing to indicate their level of satisfaction with a movie. According to our results, matrix factorization levels.**

**RECOMMENDER SYSTEM STRATEGIES**  
Broadly speaking, recommender systems are based on one of two strategies. The *nearest neighbor* approach attempts to find other users who have similar profiles to a given user and then recommend products to a user based on the profiles of these other users. The *matrix factorization* approach attempts to move a variety of special needs and wants that a user has into a profile that can be used to recommend products to the user. The goal of a recommender system is to move a user's profile as close as possible to the user's own profile as possible. This is done by finding other users who have similar profiles to the user and then recommending products to the user based on the profiles of these other users. The *matrix factorization* approach attempts to move a user's profile as close as possible to the user's own profile as possible. This is done by finding other users who have similar profiles to the user and then recommending products to the user based on the profiles of these other users.

1967  
1951 1990  
1992 1994 2001 2006 2008 2009



# MATRIX FACTORIZATION TECHNIQUES FOR RECOMMENDER SYSTEMS

Yehuda Koren, *Yahoo Research*

Robert Bell and Chris Volinsky, *AT&T Labs—Research*

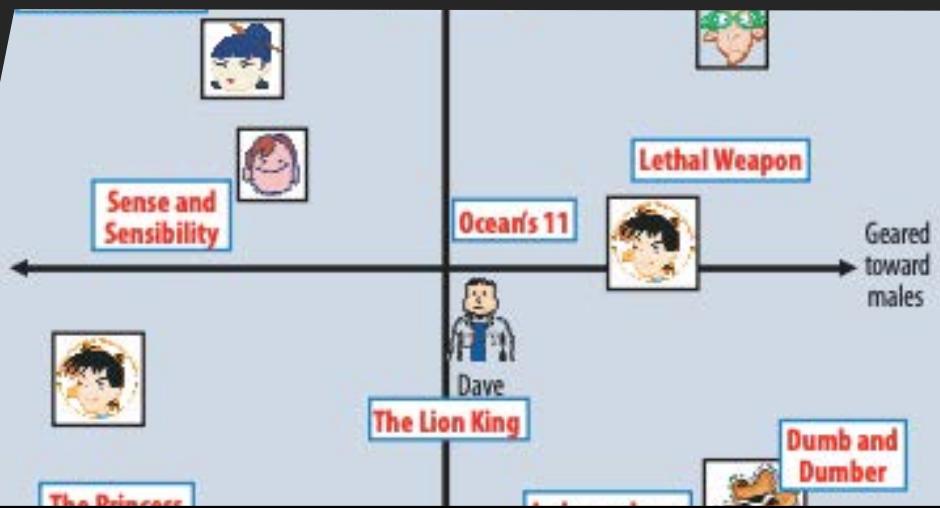
As the Netflix Prize competition has demonstrated, matrix factorization models are superior to classic nearest-neighbor techniques for producing product recommendations, allowing the incorporation of additional information such as implicit feedback, temporal effects, and confidence levels.

**M**odern consumers are inundated with choices. Electronic retailers and content providers offer a huge selection of products, with unprecedented opportunities to meet a variety of special needs and

Such systems are particularly useful for entertainment products such as movies, music, and TV shows. Customers will view the same movie, and each person is likely to view numerous different movies. Customers are proven willing to indicate their level of satisfaction with particular movies, so a huge volume of data is available about which movies appeal to which customers. Companies can analyze this data to recommend particular customers.

## RECOMMENDER SYSTEM STRATEGIES

Broadly speaking, recommender systems fall into one of two strategies. The content-based strategy creates a profile for each user or product based on its nature. For example, a movie profile



## An Algebra for Recommendations

## Using Reader Data a

Jussi Karlgren

Department of Computer Science  
Stockholm University  
Electrum 230  
S-164 40 KISTA, Sweden

jussi@dsv.su.se  
jussi@flash.bellcore.com  
jussi@sc.columbia.edu

Abstract

Introduction

Background

Modeling

Implementation

Conclusion

ACKNOWLEDGMENTS

REFERENCES

BIOGRAPHIES

Abstract

Introduction

Background

Matrix Factorization

Collaborative Filtering

Nearest Neighbors

Conclusion

ACKNOWLEDGMENTS

REFERENCES

BIOGRAPHIES

Abstract

Introduction

Background

Matrix Factorization

Collaborative Filtering

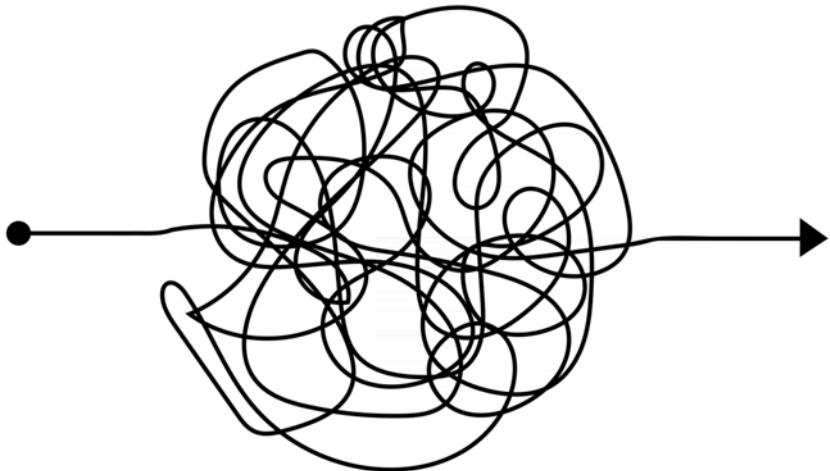
Nearest Neighbors

Conclusion

2nd Generation of recsys

1951 1967 1990 1992 1994 2001 2006 2008 2009

## 2nd generation of recommender systems



- 00s-10s
- Matrix factorization, latent user/item representations
- Less intuitive algorithms
- Hard(er) to explain

## An Algebra for Recommendations

## Using Reader Data a

Jussi Karlgren

Department of Computer Science  
Stockholm University  
Electrum 230  
S-164 40 KISTA, Sweden

jussi@dsv.su.se  
jussi@flash.bellcore.com  
jussi@sc.columbia.edu

GroupLens: A Collaborative Filtering System for Movie Recommendation

**Paul Resnick\***, Neel Guha, and Jussi Karlgren

\* MIT Center for Coordination of Compute and Communications, 32 Memorial Drive, Cambridge, MA 02139, USA  
Email: prenres@mit.edu

**ABSTRACT**  
Collaborative filters help people to find items they will like based on the opinions of other people. GroupLens is a system that helps people find articles they will like in news feeds and news reader clients. News reader clients display news from multiple sources. GroupLens filters based on the user's interests and news sources. News reader clients can also filter news feeds by selecting settings under "privacy" effects of the news provider. News reader clients can also filter news feeds by selecting settings under "privacy" effects of the news provider.

**KEYWORDS:** Collaborative filtering, electronic bulletin boards, user model, selective dissemination of information

These results are preliminary and have not been peer-reviewed. They do not necessarily represent the final version of record. Please use this version carefully, and note that it is subject to change without notice. The definitive version will be available in the ACM Digital Library at the time of publication. © 2006 ACM, Inc. All rights reserved. This work was partially funded by the National Science Foundation (NSF) under grants CCR-0121422 and CCR-0429122. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. This work was partially funded by the National Science Foundation (NSF) under grants CCR-0121422 and CCR-0429122. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

**Item-Based Collaborative Filtering Recommendation Algorithms**

Badrul Sarwar, George Karypis, Joseph Konstan, and John Riedl  
GroupLens Research Group/Army HPC Research Center

**NETFLIX**  
**COMPLETED!**

**COVER FEATURE**

**Collaborative Filtering**

**APR 7**

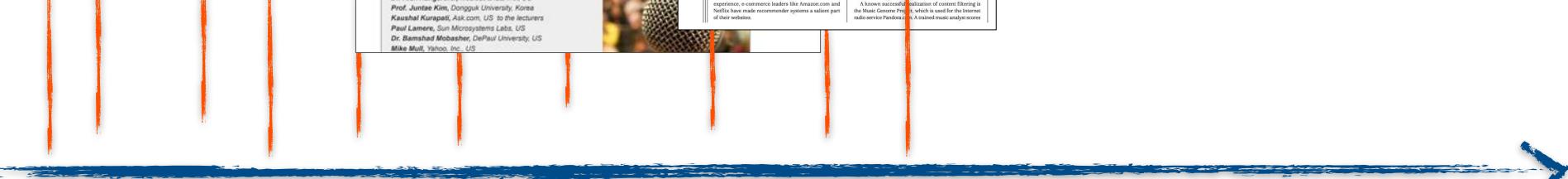
**On September 12-13, 2006 the Summer School will be held**

**ABSTRACT**  
Recommender systems apply knowledge to the problem of making personalized information, products or services that are likely to be of interest to a user. Collaborative filtering based on user ratings has been one of the most effective methods for recommendation over the last ten years since it does not require any domain knowledge about the user or the item. However, collaborative filtering suffers from data sparsity. In order to handle data sparsity, many different techniques have been proposed. In this paper we will discuss what techniques are needed that can handle data sparsity effectively.

**Introduction**  
Recommender Systems are applications that are designed to help users to select items from a large set of choices. A basic recommender system documents no interaction between the user and the system. Such interactions are usually implicit, such as rating a movie or buying a product. Such interactions, more generally called annotations, can be used to refine the recommendations made by the system. One application of annotations is in the support of moderation.

**Lecturers will include:**

Chris Anderson, Wired Magazine, US  
Todd Beaupre, Yahoo, Inc., US  
Jim Bennett, Netflix, Inc., US  
Dr. Alexander Felfernig, Klagenfurt University, Austria  
Dr. Rick Houghtaling, MusicStrands, Inc., US  
Prof. Juntae Kim, Dongguk University, Korea  
Kausik Kurapati, Ask.com, US / the lecturer  
Paul Lamere, Sun Microsystems Labs, US  
Dr. Bamshad Mobasher, DePaul University, US  
Mike Mudie, Yahoo, Inc., US



1967  
1951 1990  
1992 1994 2001  
2006 2008 2009  
2010

# Recommender Systems: We're doing it (all) wrong



A few days back, there was an interesting post by Judy Robertson in the Communications of the ACM blog. The post, entitled "[Stats: We're doing it wrong](#)", builds upon a paper from last year's CHI conference in which they report that more than 90% of the HCI researchers used the wrong statistical tools when analyzing and reporting on likert scale type of data. A Likert scale is a unidimensional scale on which the respondent expresses the level of agreement to a statement - typically in a 1 to 5 scale in which 1 is strongly disagree and 5 is strongly agree.

Here is an excerpt from the post that I think is worth highlighting:

*Likert scales give ordinal data. That it (sic), the data is ranked "strongly agree" is usually better than "agree." However, it's not interval data. You can't say the distances between "strongly agree" and "agree" would be the same as "neutral" and "disagree," for example. People tend to think there is a bigger difference between items at the extremes of the scale than in the middle (there is some evidence cited in Kaptein's paper that this is the case). For ordinal data, one should use non-parametric statistical tests which do not assume a normal distribution of the data. Furthermore, because of this it makes no sense to report means of likert scale data--you should report the mode.*

As Judy, I have to admit that I am not a stats expert myself either. But in the general case I would agree with the previous: likert scale data is ordinal and cannot be treated as interval. However, whether treating it as interval is **always** a mistake or can be accepted under some circumstances is something that I am not sure and relates to the rest of this post.



## Consequences

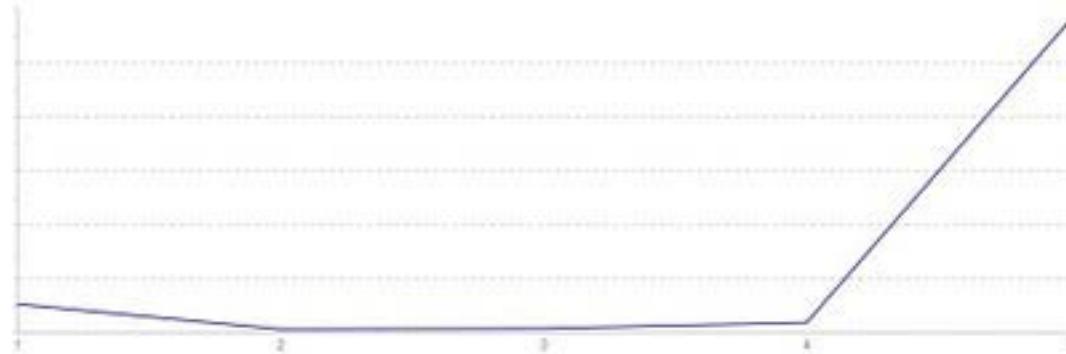
At this point, we can safely say that ratings are ordinal but not interval data. However, they are treated as a continuous interval scale in most of the recommender systems research! Let us stop to think a few of the consequences of ratings not being interval data.

**Distance Measures:** All the neighbor based methods in collaborative filtering are based on the use of some sort of distance measure. The most commonly used are Cosine distance and Pearson Correlation. However, both these distances assume a linear interval scale in their computations! We should conclude that using these distance measures with rating data is wrong. Other measures such as [Spearman's rank correlation](#), do not assume this. But to be honest, I don't remember having read many papers using Spearman.

**Error Measures:** This is my favorite one... The most commonly accepted measure of success for recommender systems is the Root Mean Squared Error (RMSE). But wait, this measure is explicitly assuming that ratings are also interval data! Similar error measures such as MAE also fall in the same trap... banned! So what could we use? Standard Information Retrieval measures such as Precision and Recall do not necessarily assume interval scale on the ratings, although their mapping to recommendation efficiency may also be questioned. Rank-based measures such as [Discounted Cumulative Gain](#) (nDCG) seem like our best bet for now.

### UNDERSTANDING ONLINE STAR RATINGS:

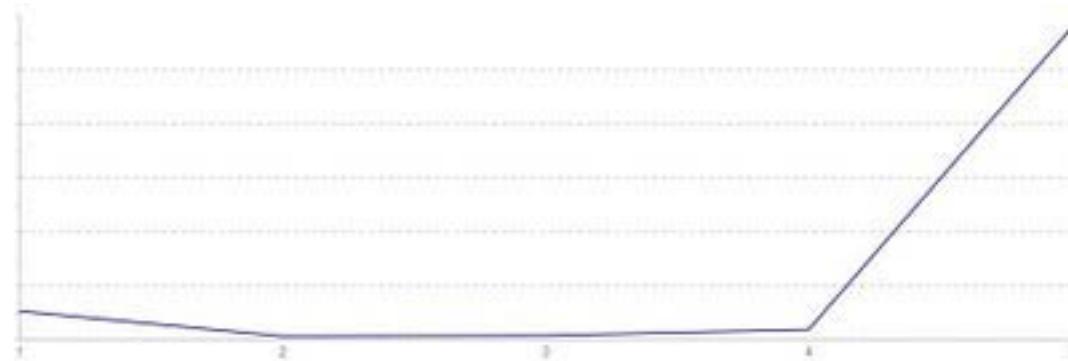




<https://xkcd.com/1098/>

# Five Stars Dominate Ratings

By Shiva Rajaraman  
Group Product Manager  
Sep.22.2009



<https://xkcd.com/1098/>

## An Algebra for Recommendations

## Using Reader Data as Input

Jussi Karlgren

Department of Computer  
Stockholm University  
Electrum 230  
S-164 40 KISTA, Sweden

jussi@dsv.su.se  
jussi@flash.bellcore.com  
jussi@sc.columbia.edu

GroupLens: A Collaborative Filtering System for Movie Recommendation

**Paul Resnick\***, Neel Guha, and Jussi Karlgren

\* MIT Center for Coordination of Compute and Communications, 30 Memorial Drive, Cambridge, MA 02139, USA

Email: [premres@mit.edu](mailto:premres@mit.edu)

**ABSTRACT**  
Collaborative filters help people to find items they will like based on the opinions of other people. GroupLens is a system that uses collaborative filtering to recommend movies to users. New reader clients display movie ratings and reviews from other users, and provide recommendations based on the heuristic that people who liked the same things in the past tend to like new things. Collaborative filters can be improved by extracting ratings under "playback" effects of the user prediction system. We present a model of how such a system can work.

**KEYWORDS:** Collaborative filters, electronic bulletin boards, user model, selective dissemination of information

*This paper is an invited paper for the 13th International Conference on Information and Knowledge Management (CIKM 2004), November 1-5, 2004, Atlanta, Georgia, USA, organized by the Association for Computing Machinery (ACM) Special Interest Group on Information and Knowledge Management. It is also an invited paper for the 2004 Summer School on Recommender Systems, September 12-13, 2004, Stockholm, Sweden. This paper is an extended version of the paper published in the proceedings of the 2004 Summer School on Recommender Systems, Stockholm, Sweden, September 12-13, 2004. The authors would like to thank the anonymous reviewers for their useful comments and suggestions. This research was partially funded by grants from the Swedish Research Council, the Swedish National Institute of Telecommunications, and the Royal Swedish Academy of Sciences.*

**ACKNOWLEDGMENTS**  
Collaborative networks allow the formation of local communities around areas of interest. A basic problem is how to get documents no related to a specific topic available to anyone interested in the topic. The User Interest System makes it possible to search the world's popularity bulletin board available anywhere in the world. It populates bulletin boards with items from an article posted from anywhere in the world to available to anyone interested in the topic.

**Participation in the Summer School is limited to 50 students.**

**These reactions, more generally called annotations, can be used by others. Other applications of annotations is to support moderation of discussions.**

**Europe First Project**  
J.-L. Naupey, A. Léonard, G. R. M. Brézillon  
University of Paris 6, Paris, France

1951 1967 1990 1992 1994 2001 2006 2008 2009 2010

## Item-Based Collaborative Filtering Recommendation Algorithms

Badrul Sarwar, George Karypis, Joseph Konstan, and John Riedl  
<http://www.cs.umn.edu/~sarwar/>  
GroupLens Research Group/Army HPC Research Center  
Depart. of Computer Science, University of Minnesota, Minneapolis, MN 55455

NETFLIX

COMPLETED!

COVER FEATURE

## Collaborative Filtering



i On September 12-13, 2004, the Summer School will be held

APR

Collaborative Filtering

Recommender Systems: We're doing it (all) wrong

A few months ago I wrote a post entitled "What We're Doing Wrong" which builds upon a paper from last year's CIKM conference in which they report that more than 50% of the HCI researchers used the wrong statistical tool for analyzing and reporting on multiple types of data. A Library Science professor has recently written a post entitled "The recommender systems landscape changes from predicting ratings to generating rankings" in which he highlights the fact that many recommender systems researchers have been doing it wrong for years.

Recommender Systems are shifting from predicting ratings to generating rankings.

However, it's important to note that while these two approaches may be the same as "neutral" and "disagree," for example, people tend to think there is a bigger difference between items at the extremes of the scale. In the middle items is something that is cited in Koenen's paper as being what "most people would consider to be a good rating."

Here is an excerpt from the post that I think is worth highlighting:

Likert scales are treated as interval data by the majority of marketing researchers.

However, it's important to note that while these two approaches may be the same as "neutral" and "disagree," for example, people tend to think there is a bigger difference between items at the extremes of the scale. In the middle items is something that is cited in Koenen's paper as being what "most people would consider to be a good rating."

Within the eco-system interacting and consumers, many different characteristics are used when evaluating products. These include: item popularity, price, availability, quality, delivery times, and associated with a few data points, the time when the item was purchased. Most of these are ordinal, and most a single concept is tracked. Classical time-series analysis approaches cannot work, as they lose too much information in the process. For example, which can make better distinctions between traits long term patterns. The paradigm we offer is one of item-based recommendation, which uses item data. This allows us to exploit the relevant context information, while discarding only the irrelevant data. Accordingly, we propose new local collaborative recommendation approaches. Evaluation is made of these methods using the Netflix Prize dataset.

As Judy, I have to admit that I am not a stats expert myself either. But in the general case I would agree with the previous:

Likert scales are ordinal and cannot be treated as interval. However, whether treating it as interval is always a mistake or can be accepted under some circumstances is something that I am not sure and relates to the rest of this post.

So for instance, it is not uncommon to find references where they clearly state that Likert data can be treated as interval. For example, look at what they say in this handbook edited by the FAO.

Likert scales are treated as yielding interval data by the majority of marketing researchers.

Or look at the answer to the question of whether Likert data can be treated as interval in stackexchange.

Is Likert data interval or ordinal?

Answers from stackexchange:

It depends on what you want to do with your product. Of course,

ordinal based metrics require grouping the data.

That might not be available or easy to collect.

A known metric is the "ranked list" which is used for the internet

radio service Pandora. A trained music analysis scores

## An Algebra for Recommendations

### Using Reader Data as Input

Jussi Karlgren

Department of Computer Science  
Stockholm University  
Electrum 230  
S-164 40 KISTA, Sweden

jussi@dsv.su.se  
jussi@flash.bellcore.com  
jussi@sc.columbia.edu

**GroupLens:** A Collaborative Filtering System for Movie Recommendation

**Paul Resnick\***, Neel Guha, Badrul Sarwar, George Karypis, Joseph Konstan, and John Riedl

\* MIT Center for Coordination of Compute and Human Activity, 50 Memorial Drive, Cambridge, MA 02142, USA  
Email: prenres@mit.edu

**ABSTRACT**  
Collaborative filters help people find new items they might like by collecting user ratings on items they have already rated. They can also help people find new items similar to ones they like by mining ratings under the hypothesis that people who like similar items will also like new items based on the similarity of their ratings. In this paper, we introduce the Netflix Prize, a competition to build the best collaborative filter for movie ratings. We also describe how GroupLens has approached the Netflix Prize, and present our system's results.

**KEYWORDS:** Collaborative filtering, recommendation systems, movie ratings, item-based filters, user model, selective dissemination of information

**INTRODUCTION**  
Collaborative networks allow the form of recommendations to change as users interact with them. A basic form of recommendation is to suggest documents most similar to a document of interest. In addition to helping users find what they want, such recommendations can be used to refine other filters. One application of annotations is to moderate other filters.

**Participation in the Summer School is free!**

**GROUPLENS RESEARCH GROUP**  
Badrul Sarwar, George Karypis, Joseph Konstan, and John Riedl  
GroupLens Research Group/Army HPC Research Center  
Department of Computer Science  
University of Minnesota, Minneapolis, MN 55455-0439  
http://www.cs.umn.edu/~groupLens

**GroupLens Research Group**  
Badrul Sarwar, George Karypis, Joseph Konstan, and John Riedl  
GroupLens Research Group/Army HPC Research Center  
Department of Computer Science  
University of Minnesota, Minneapolis, MN 55455-0439  
http://www.cs.umn.edu/~groupLens

1951 1967 1990 1992 1994 2001 2006 2008 2009 2010 2016

## Item-Based Collaborative Filtering Recommendation Algorithms

Badrul Sarwar, George Karypis, Joseph Konstan, and John Riedl  
GroupLens Research Group/Army HPC Research Center  
Department of Computer Science  
University of Minnesota, Minneapolis, MN 55455-0439  
http://www.cs.umn.edu/~groupLens

**ABSTRACT**  
Recommender systems apply knowledge to the problem of making personalized information, products or services more interesting to users. Collaborative filtering based on user ratings is one of the most successful approaches to solving this problem. In recent years, more and more recommender systems have been performing many more complex tasks than just rating prediction. These include rating prediction, item popularity prediction, and item sparsity. In these cases, more sophisticated models are needed that can handle different types of data and make better use of the available data quantity. In this paper, we introduce a new framework for rating prediction that can be applied to all of these tasks. The framework is based on the idea that a recommendation system should be able to learn from previous interactions between users and items, and that this information should be used to predict future interactions.

**ABSTRACT**  
Collaborative filters help people find new items they will likely like based on the opinions of other people. GroupLens Research has developed a system for movie recommendation that uses collaborative filtering to find new items that people will like. The system is based on the hypothesis that people who like similar items will also like new items based on the similarity of their ratings. In this paper, we introduce the Netflix Prize, a competition to build the best collaborative filter for movie ratings. We also describe how GroupLens has approached the Netflix Prize, and present our system's results.

**ABSTRACT**  
Collaborative filters help people find new items they will likely like based on the opinions of other people. GroupLens Research has developed a system for movie recommendation that uses collaborative filtering to find new items that people will like. The system is based on the hypothesis that people who like similar items will also like new items based on the similarity of their ratings. In this paper, we introduce the Netflix Prize, a competition to build the best collaborative filter for movie ratings. We also describe how GroupLens has approached the Netflix Prize, and present our system's results.

**Introduction**

Recommender Systems are application areas that are becoming a part of everyday life. They content discovery. The social prevalence of television, video and online shopping.

The Summer School will bring together researchers (including Ph.D. students) on Recommender Systems and person provided.

Participation in the Summer School is free!

### Lecturers will include:

Chris Anderson, Wired Magazine, US  
Todd Beaupre, Yahoo, Inc., US  
Jim Bennett, Netflix, Inc., US  
Dr. Alexander Felfernig, Klagenfurt University, Austria  
Dr. Rick Houghtaling, MusicStrands, Inc., US  
Prof. Juntae Kim, Dongguk University, Korea  
Kaushal Kurapati, Ask.com, US / the lectures  
Paul Lamere, Sun Microsystems Labs, US  
Dr. Bamshad Mobasher, DePaul University, US  
Mike Mudie, Yahoo, Inc., US

Or look at the answer to this question: What does it mean to say that a recommendation system is based on item-item similarity? If I were to ask you to come up with a recommendation system, what would you do?

It means that the system finds items that are similar to the user's taste. Because recommendations can add some experience, a company needs to have some kind of recommendation system.

So for instance, if a user likes a particular movie, the system will recommend other movies that are similar to it. This is called item-item similarity.

It means that the system finds items that are similar to the user's taste. Because recommendations can add some experience, a company needs to have some kind of recommendation system.

It means that the system finds items that are similar to the user's taste. Because recommendations can add some experience, a company needs to have some kind of recommendation system.



### Collaborative Filtering

On September 12-13, 2006 The Summer School will be held at the University of Minnesota.

### Introduction

Recommender Systems are becoming a part of everyday life. They content discovery. The social prevalence of television, video and online shopping.

The Summer School will bring together researchers (including Ph.D. students) on Recommender Systems and person provided.

Participation in the Summer School is free!

**Recommender Systems: We're doing it (eli.wilson@utdallas.edu)**  
A few weeks ago, I wrote a post entitled "What We're Doing Now" which builds upon a paper from last year. It's not that much different than the paper, but it's a bit more detailed. In this post, I'm going to talk about how recommender systems work, and how they can be improved.

As you may know, there are two main types of recommender systems: content-based and collaborative. Content-based systems are based on the idea that users have certain interests, and that items have certain features. By comparing the user's interests with the item's features, the system can recommend items that are likely to be liked.

Collaborative systems, on the other hand, are based on the idea that users have certain interests, and that items have certain features. By comparing the user's interests with the item's features, the system can recommend items that are likely to be liked.

As you may know, there are two main types of recommender systems: content-based and collaborative. Content-based systems are based on the idea that users have certain interests, and that items have certain features. By comparing the user's interests with the item's features, the system can recommend items that are likely to be liked.

As you may know, there are two main types of recommender systems: content-based and collaborative. Content-based systems are based on the idea that users have certain interests, and that items have certain features. By comparing the user's interests with the item's features, the system can recommend items that are likely to be liked.

As you may know, there are two main types of recommender systems: content-based and collaborative. Content-based systems are based on the idea that users have certain interests, and that items have certain features. By comparing the user's interests with the item's features, the system can recommend items that are likely to be liked.

As you may know, there are two main types of recommender systems: content-based and collaborative. Content-based systems are based on the idea that users have certain interests, and that items have certain features. By comparing the user's interests with the item's features, the system can recommend items that are likely to be liked.

As you may know, there are two main types of recommender systems: content-based and collaborative. Content-based systems are based on the idea that users have certain interests, and that items have certain features. By comparing the user's interests with the item's features, the system can recommend items that are likely to be liked.

As you may know, there are two main types of recommender systems: content-based and collaborative. Content-based systems are based on the idea that users have certain interests, and that items have certain features. By comparing the user's interests with the item's features, the system can recommend items that are likely to be liked.

As you may know, there are two main types of recommender systems: content-based and collaborative. Content-based systems are based on the idea that users have certain interests, and that items have certain features. By comparing the user's interests with the item's features, the system can recommend items that are likely to be liked.

As you may know, there are two main types of recommender systems: content-based and collaborative. Content-based systems are based on the idea that users have certain interests, and that items have certain features. By comparing the user's interests with the item's features, the system can recommend items that are likely to be liked.

As you may know, there are two main types of recommender systems: content-based and collaborative. Content-based systems are based on the idea that users have certain interests, and that items have certain features. By comparing the user's interests with the item's features, the system can recommend items that are likely to be liked.

As you may know, there are two main types of recommender systems: content-based and collaborative. Content-based systems are based on the idea that users have certain interests, and that items have certain features. By comparing the user's interests with the item's features, the system can recommend items that are likely to be liked.

As you may know, there are two main types of recommender systems: content-based and collaborative. Content-based systems are based on the idea that users have certain interests, and that items have certain features. By comparing the user's interests with the item's features, the system can recommend items that are likely to be liked.

As you may know, there are two main types of recommender systems: content-based and collaborative. Content-based systems are based on the idea that users have certain interests, and that items have certain features. By comparing the user's interests with the item's features, the system can recommend items that are likely to be liked.

As you may know, there are two main types of recommender systems: content-based and collaborative. Content-based systems are based on the idea that users have certain interests, and that items have certain features. By comparing the user's interests with the item's features, the system can recommend items that are likely to be liked.

As you may know, there are two main types of recommender systems: content-based and collaborative. Content-based systems are based on the idea that users have certain interests, and that items have certain features. By comparing the user's interests with the item's features, the system can recommend items that are likely to be liked.

## An Algebra for Recommendations

### Using Reader Data as Input

Jussi Karlgren

Department of Computer Science  
Stockholm University  
Electrum 230  
S-164 40 KISTA, Sweden

jussi@dsv.su.se  
jussi@flash.bellcore.com  
jussi@sc.columbia.edu

These annotations, more generally called annotations, can be applied to other filters. One application of annotations is in support of moderation.

Eugene Fix, Ph.D.  
University of California, Berkeley

### GroupLens: A Collaborative Filtering System for Movie Recommendations

Paul Resnick\*, Neel Guha

\* MIT Center for Coordination of Compute and Communication, 30 Memorial Drive, Cambridge, MA 02139, USA  
Email: prenres@mit.edu

#### ABSTRACT

Collaborative filters help people find opinions of other people. GroupLens is a system that recommends articles they will like in the digital library. News reader clients display news from multiple sources based on the headlines that people have written about them. By enabling ratings under pseudonyms, users can express what they think of items without revealing their names. Buttons can be developed independently of the news reader.

#### KEYWORDS:

Collaborative filtering; electronic bulletin boards; user model; selective dissemination of information

These annotations, more generally called annotations, can be applied to other filters. One application of annotations is in support of moderation.

## Item-Based Collaborative Filtering Recommendation Algorithms

Badrul Sarwar, George Karypis, Joseph Konstan, and John Riedl  
GroupLens Research Group/Army HPC Research Center

NETFLIX

COMPLETED!

COVER FEATURE

### Collaborative Filtering



On September 12-13, 2006 The Summer School will be held

#### Introduction

Recommender Systems are applications that provide recommendations to users to help them find products or services of interest. A basic recommender system documents no interaction between the user and the article directly to recommend items. Such systems are available to anyone interested in the topic. The User interest system creates a profile of the user's interests and recommends items to the user. The Summer School will bring together researchers (including Ph.D. students) on Recommender Systems and person provided.

Participation in the Summer School is t

#### Lecturers will include:

Chris Anderson, Wired Magazine, US  
Todd Beaupre, Yahoo, Inc., US  
Jim Bennett, Netflix, Inc., US  
Dr. Alexander Felfernig, Klagenfurt University, Austria  
Dr. Rick Hengartner, MusicStrands, Inc., US  
Prof. Juntae Kim, Dongguk University, Korea  
Kaushtal Kurapati, Ask.com, US to the lectures  
Paul Lamere, Sun Microsystems Labs, US  
Dr. Bamshad Mobasher, DePaul University, US  
Mike Mud, Yahoo, Inc., US

## The recommendation lands

### pre

As Just as

likert scales can be ac

So for instance

example, look at

likert scales

Or look at the answer

intents in products to provide

the user's needs. Recom

mendations are responsible

for helping more than a billion users

discover new products every day

In this paper we will focus on the immense im-

pact deep learning has had on the YouTube video

recommendation system. Figure 1 shows the re-

commendations on the YouTube mobile app home.

Recommending YouTube videos is extremely challenging

because it involves

associating users with matching products. Of course,

products are often not available or easy to collect.

A known solution to this problem is to

use a generative model which is called a

radio service Pandora. A trained music analysis scores

## Deep Neural Networks for YouTube Recommendations

Paul Covington, Jay Adams, Emre Sargin  
Google Mountain View, CA  
pcovington, jaa, msargin@google.com

#### ABSTRACT

YouTube represents one of the largest scale and most sophisticated recommendation systems in existence. In this paper, we describe the system at a high level and focus on the dramatic performance improvements brought by deep neural networks. We also describe the system's two-stage information retrieval dichotomy: first, we detail a deep candidate generation model and then describe a sequence-to-sequence model that performs both item ranking and likelihood scaling. Finally, we discuss the challenges of maintaining a massive recommendation system with enormous user-item interactions.



Figure 1: Recommendations displayed on YouTube mobile app home.

with well-established video can be understood from an exploration/exploitation perspective.

1951 1967 1990 1992 1994 2001 2006 2008 2009 2010 2016

# Deep Neural Networks for YouTube Recommendations

Paul Covington, Jay Adams, Emre Sargin  
Google  
Mountain View, CA  
[{pcovington, jka, msargin}@google.com](mailto:{pcovington, jka, msargin}@google.com)

## ABSTRACT

YouTube represents one of the largest scale and most sophisticated industrial recommendation systems in existence. In this paper, we describe the system at a high level and focus on the dramatic performance improvements brought by deep learning. The paper is split according to the classic two-stage information retrieval dichotomy: first, we detail a deep candidate generation model and then describe a separate deep ranking model. We also provide practical lessons and insights derived from designing, iterating and maintaining a massive recommendation system with enormous user-facing impact.

## Keywords

recommender system; deep learning; scalability

## 1. INTRODUCTION

YouTube is the world's largest platform for creating, sharing and discovering video content. YouTube recommendations are responsible for helping more than a billion users discover personalized content from an ever-growing corpus of videos. In this paper we will focus on the immense impact deep learning has recently had on the YouTube video recommendations system. Figure 1 illustrates the recommendations on the YouTube mobile app home.

Recommending YouTube videos is extremely challenging from three major perspectives:

- **Scale:** Many existing recommendation algorithms proven to work well on small problems fail to operate on our scale. Highly specialized distributed learning algorithms and efficient serving systems are essential for handling YouTube's massive user base and corpus.
- **Noise:** Historical user behavior is often difficult to predict due to the nature of unobservable external factors. Maintain the ground truth of user satisfaction noisy implicit feedback and metadata associated with content without a well defined ontology.
- **Freshness:** YouTube has a very dynamic corpus with

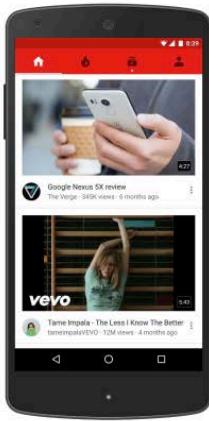
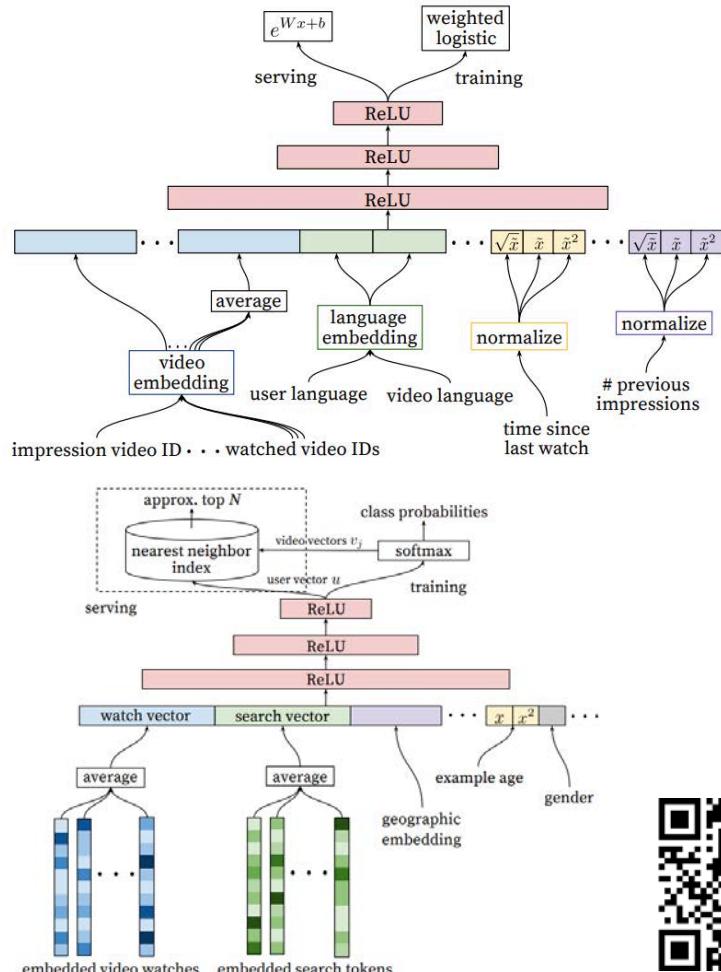


Figure 1: Recommendations displayed on the mobile app home.

with well-established videos can be an exploration/exploitation perspective

- **Noise:** Historical user behavior is often difficult to predict due to the nature of unobservable external factors. Maintain the ground truth of user satisfaction noisy implicit feedback and metadata associated with content without a well defined ontology.
- **Freshness:** YouTube has a very dynamic corpus with



## An Algebra for Recommendations

## Using Reader Data as Input

Jussi Karlgren

Department of Computer Science  
Stockholm University  
Electrum 230  
S-164 40 KISTA, Sweden

jussi@dsv.su.se  
jussi@flash.bellcore.com  
jussi@sc.columbia.edu

**Topology is an important concept in Recommender Systems.** The most well-known recommender system is probably Netflix, which uses a complex matrix factorization algorithm to predict what users will like based on their previous ratings. A better way to think about it is that it's a collaborative filtering system that finds items similar to what you liked in the past. This is done by comparing your rating with other users' ratings and finding items that they also liked. This approach is called "item-based" because it focuses on individual items rather than users. It's also called "collaborative" because it relies on the collective knowledge of many users to make recommendations. In this diagram, we see a network of nodes representing users and items, with edges indicating similarity based on shared preferences. The size of each node represents its popularity or influence in the system. The color of the nodes indicates their genre or category, such as movies or TV shows. The edges between nodes represent the strength of the recommendation, with thicker edges indicating stronger similarity.

## Item-Based Collaborative Filtering Recommendation Algorithms

Badrul Sarwar, George Karypis, Joseph Konstan, and John Riedl  
GroupLens Research Group/Army HPC Research Center



GroupLens Research Group/Army HPC Research Center

NETFLIX

NETFLIX Prize  
Home Rules Leadership Update  
COMPLETED!

COVER FEATURE

## GroupLens: A Case Study

Paul Resnick\*, Neel Guha

\* MIT Center for Coordination of Compute and Commute, 30 Memorial Dr., Cambridge, MA 02139, USA  
Email: [premres@mit.edu](mailto:premres@mit.edu)

## Abstract

Collaborative filters help people to find products or services that others have rated highly. These filters can be used to predict what new users will like, by matching their interests with those of existing users. This paper describes how these filters work and how they can be used to build recommendation systems.

Collaborative filters can be used to predict what new users will like, by matching their interests with those of existing users. This paper describes how these filters work and how they can be used to build recommendation systems.

## Keywords

Collaborative filtering; electronic commerce; user model; selective dissemination of information.

Collaborative networks allow the formation of local clusters of users with similar interests. This makes it easier for users to find products or services that are available to everyone in the system. The User-to-user system offers a more personalized experience than the general search engines available on the web.

In addition to supporting item-based filters, the User-to-user system also supports user-based filters. Such recommendations are based on the user's own history and the history of other users. One application of annotations is to support user-based filters.

These annotations, more generally called annotations, can be used to support user-based filters.



## Collaborative Filtering

**Abstract**  
Recommender systems apply knowledge to the problem of making personalized recommendations, products or services that are likely to be of interest to a user. These systems have been around for many years now and have become increasingly popular in recent years. They are used in various domains such as e-commerce, entertainment, news, social media, etc. In this paper, we propose a framework for building recommender systems based on collaborative filtering, which has been shown to be effective in handling sparse data. We also discuss some challenges in building such systems and propose some solutions to them.

**Introduction**  
Recommender Systems are applications that provide recommendations to users based on their past behavior. They are becoming a part of everyday life. They are used in various domains such as e-commerce, entertainment, news, social media, etc. In this paper, we propose a framework for building recommender systems based on collaborative filtering, which has been shown to be effective in handling sparse data. We also discuss some challenges in building such systems and propose some solutions to them.

**Conclusion**  
In this paper, we proposed a framework for building recommender systems based on collaborative filtering. We also discussed some challenges in building such systems and proposed some solutions to them.

**Keywords**  
Collaborative filtering; electronic commerce; user model; selective dissemination of information.



## Deep Neural Networks for YouTube Recommendations

Paul Covington, Jay Adams, Emre Sargin

Published as a conference paper at ICLR 2016

## SESSION-BASED RECOMMENDATIONS WITH RECURRENT NEURAL NETWORKS

Balázs Hidasi<sup>†</sup>  
Gravity R&D Inc.  
Budapest, Hungary  
[bala.balazs.hidasi@googlemail.com](mailto:bala.balazs.hidasi@googlemail.com)

Alexandros Karatzoglou<sup>†</sup>  
Telefonica Research  
Barcelona, Spain  
[alex.karatzoglou@gmail.com](mailto:alex.karatzoglou@gmail.com)

László Baltrunus<sup>‡</sup>  
Netflix  
Los Gatos, CA, USA  
[lbaltrunas@netflix.com](mailto:lbaltrunas@netflix.com)

Domonkos Tikk<sup>‡</sup>  
Gravity R&D Inc.  
Budapest, Hungary  
[domonkos.tikk@googlemail.com](mailto:domonkos.tikk@googlemail.com)

## ABSTRACT

We apply recurrent neural networks (RNNs) on a new domain, namely session-based systems. Real-life recommendation systems often face the problem of having to base recommendations only on short user histories (e.g. a small sportswear website) instead of long user histories (as in the case of Netflix). In this situation the frequently praised matrix factorization approaches are not accurate. This problem is usually overcome in practice by resorting to item-to-item recommendations, i.e. recommending items that appear to be similar to the whole user history. However, more accurate recommendations can be provided. We therefore propose an RNN-based approach for session-based recommendations. Our approach also considers practical aspects of the task and introduces several modifications to classic RNNs such as a novel loss function that make it more viable for this specific problem. Experimental results on two data sets show marked improvements over widely used approaches.

1951 1967 1990 1992 1994 2001 2006 2008 2009 2010 2016

# SESSION-BASED RECOMMENDATIONS WITH RECURRENT NEURAL NETWORKS

Balázs Hidasi \*

Gravity R&D Inc.  
Budapest, Hungary  
balazs.hidasi@gravityrd.com

Alexandros Karatzoglou

Telefonica Research  
Barcelona, Spain  
alexk@tid.es

Linas Baltrunas †

Netflix  
Los Gatos, CA, USA  
lbaltrunas@netflix.com

Domonkos Tikk

Gravity R&D Inc.  
Budapest, Hungary  
domonkos.tikk@gravityrd.com

## ABSTRACT

We apply recurrent neural networks (RNN) on a new domain, namely recommender systems. Real-life recommender systems often face the problem of having to base recommendations only on short session-based data (e.g. a small sportswear website) instead of long user histories (as in the case of Netflix). In this situation the frequently praised matrix factorization approaches are not accurate. This problem is usually overcome in practice by resorting to item-to-item recommendation i.e. recommending similar items. We argue that by modeling the whole session more accurate recommendations can be provided. We therefore propose an RNN based approach for session-based recommendations. Our approach also considers practical aspects of the task and introduces several modifications to classic RNN such as a ranking loss function that make it more viable for this specific problem. Experimental results on two data-sets show marked improvements over widely used approaches.

## 1 INTRODUCTION

Session-based recommendation is a relatively unappreciated problem in the machine learning and recommender systems community. Many e-commerce recommender systems (

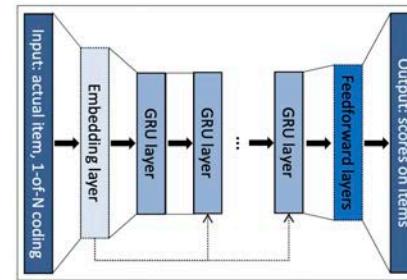


Figure 1: General architecture of the network. Processing of one event of the event stream at once.

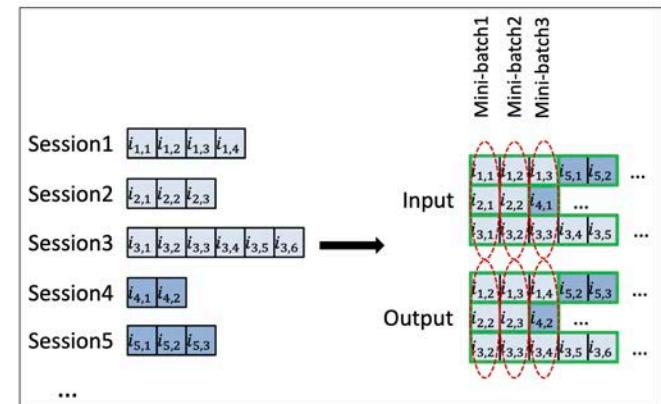


Figure 2: Session-parallel mini-batch creation



# Monolith: Real Time Recommendation System With Collisionless Embedding Table

Zhuoran Liu  
Bytedance Inc.

Leqi Zou  
Bytedance Inc.

Xuan Zou  
Bytedance Inc.

Caihua Wang  
Bytedance Inc.

Biao Zhang  
Bytedance Inc.

Da Tang  
Bytedance Inc.

Bolin Zhu\*  
Fudan University

Yijie Zhu  
Bytedance Inc.

Peng Wu  
Bytedance Inc.

Ke Wang  
Bytedance Inc.

Youlong Cheng<sup>†</sup>  
Bytedance Inc.  
youlong.cheng@bytedance.com



## ABSTRACT

Building a scalable and real-time recommendation system is vital for many businesses driven by time-sensitive customer feedback, such as short-videos ranking or online ads. Despite the ubiquitous adoption of production-scale deep learning frameworks like TensorFlow or PyTorch, these general-purpose frameworks fall short of business demands in recommendation scenarios for various reasons: on one hand, tweaking systems based on static parameters and dense computations for recommendation with dynamic and sparse features is detrimental to model quality; on the other hand, such frameworks are designed with batch-training stage and serving stage completely separated, preventing the model from interacting with customer feedback in real-time. These issues led us to reexamine traditional approaches and explore radically different design choices. In this paper, we present **Monolith**<sup>1</sup>, a system tailored for online training. Our design has been driven by observations of our application workloads and production environment that reflects a marked departure from other recommendations systems. Our contributions are manifold: first, we crafted a collisionless embedding table with optimizations such as expirable embeddings and frequency filtering to reduce its memory footprint; second, we provide an production-ready online training architecture with high fault-tolerance; finally, we proved that system reliability could be traded-off for real-time learning. Monolith has successfully landed in the BytePlus Recommend<sup>2</sup> product.

## ACM Reference Format:

Zhuoran Liu, Leqi Zou, Xuan Zou, Caihua Wang, Biao Zhang, Da Tang, Bolin Zhu, Yijie Zhu, Peng Wu, Ke Wang, and Youlong Cheng. 2022. Monolith: Real Time Recommendation System With Collisionless Embedding Table. In *Proceedings of 5th Workshop on Online Recommender Systems and User Modeling, in conjunction with the 16th ACM Conference on Recommender Systems (ORSUM@ACM RecSys 2022)*. ACM, New York, NY, USA, 10 pages. <https://doi.org/XXXXXX.XXXXXXX>.

## 1 INTRODUCTION

The past decade witnessed a boom of businesses powered by recommendation techniques. In pursuit of a better customer experience, delivering personalized content for each individual user as real-time responses is a common goal of these business applications. To this end, information from a user’s latest interaction is often used as the primary input for training a model, as it would best depict a user’s portrait and make predictions of user’s interest and future behaviors.

Deep learning have been dominating recommendation models [5, 6, 10, 12, 20, 21] as the gigantic amount of user data is a natural fit for massively data-driven neural models. However, efforts to leverage the power of deep learning in industry-level recommendation systems are constantly encountered with problems arising from the unique characteristics of data derived from real-world user behavior. These data are drastically different from those used

# Monolith: Real Time Recommendation System With Collisionless Embedding Table



Zhuoran Liu  
Bytedance Inc.

Leqi Zou  
Bytedance Inc.

Xuan Zou  
Bytedance Inc.

Caihua Wang  
Bytedance Inc.

Biao Zhang  
Bytedance Inc.

Da Tang  
Bytedance Inc.

Bolir  
Fudan T

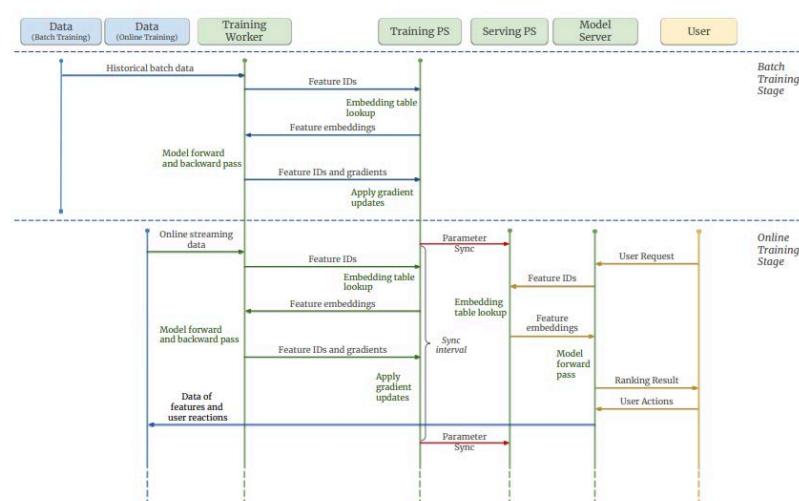


Figure 1: Monolith Online Training Architecture.

## ABSTRACT

Building a scalable and for many businesses dr such as short-videos ra adoption of production sorFlow or PyTorch, th of business demands in sons: on one hand, tweal dense computations for features is detrimental frameworks are design stage completely separa with customer feedback amine traditional appro choices. In this paper, for online training. Ou of our application work reflects a marked depart Our contributions are n embedding table with opt and frequency filtering provide an production-fault-tolerance; finally, traded-off for real-time in the BytePlus Recom

# Monolith: Real Time Recommendation System With Collisionless Embedding Table



Zhuoran Liu  
Bytedance Inc.

Caihua Wang  
Bytedance Inc.

Bolir  
Fudan U

Leqi Zou  
Bytedance Inc.

Biao Zhang  
Bytedance Inc.

## ABSTRACT

Building a scalable and for many businesses such as short-videos raise adoption of production TensorFlow or PyTorch, the business demands in seconds: on one hand, we want dense computations for features is detrimental frameworks are designed stage completely separate with customer feedback. Unlike traditional approaches. In this paper, for online training. One of our application work reflects a marked departure. Our contributions are a collisionless embedding table with optimal frequency filtering provide an production fault-tolerance; finally, traded-off for real-time in the BytePlus Recom-

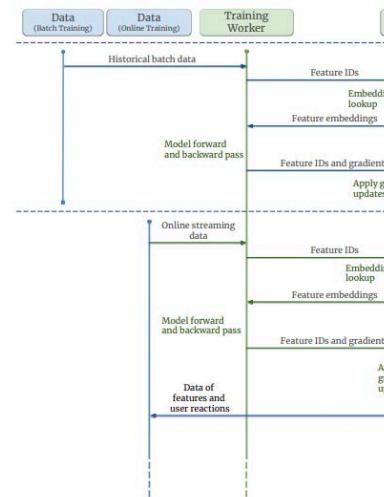


Figure 1: Monolith Online Training Architecture.

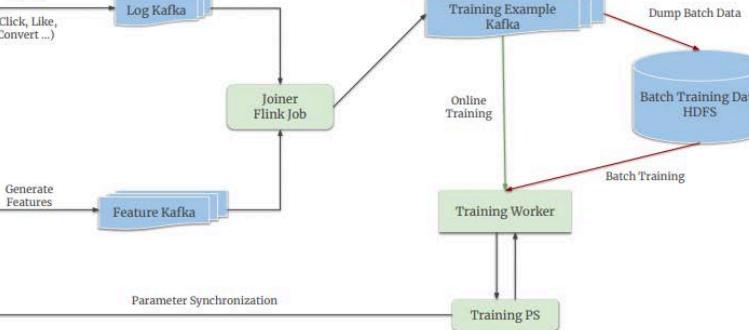


Figure 4: Streaming Engine.

The information feedback loop from [User → Model Server → Training Worker → Model Server → User] would spend a long time when taking the Batch Training path, while the Online Training will close the loop more instantly.

# Monolith: Real Time Recommendation System With Collisionless Embedding Table



Zhuoran Liu  
Bytedance Inc.

Leqi Zou  
Bytedance Inc.

Caihua Wang  
Bytedance Inc.

Biao Zhang  
Bytedance Inc.

Bolir  
Fudan T

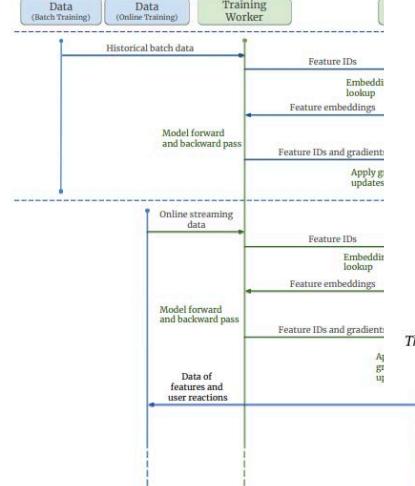


Figure 1: Monolith Online Training Architecture.

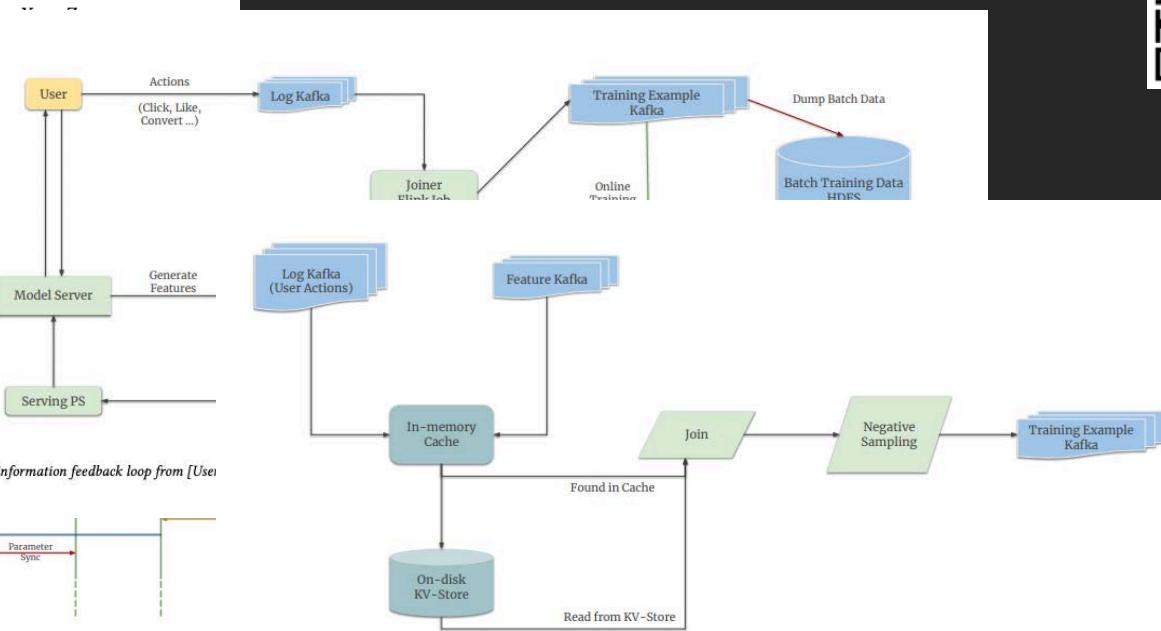


Figure 5: Online Joiner.

# Monolith: Real Time Recommendation System With Collisionless Embedding Table

Zhuoran Liu  
Bytedance Inc.

Leqi Zou  
Bytedance Inc.

Caihua Wang  
Bytedance Inc.

Biao Zhang  
Bytedance Inc.

Bolir  
Fudan U

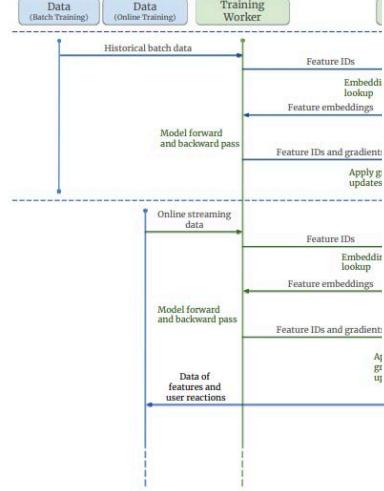


Figure 1: Monolith Online Training Architecture.

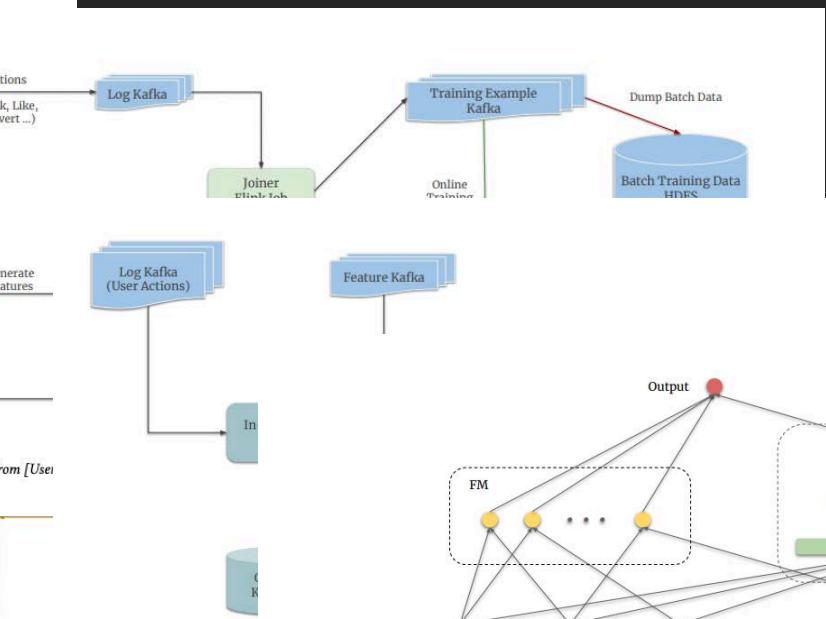
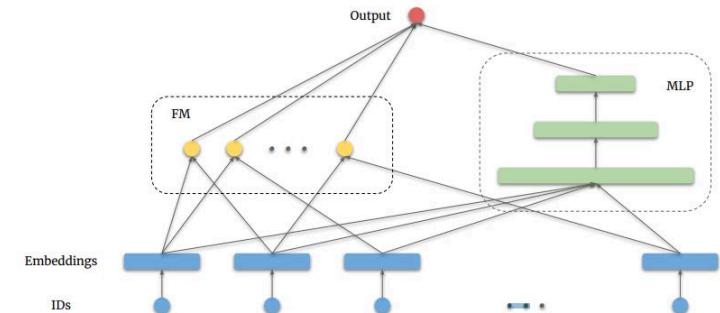


Figure 2: Data Pipeline.



Figure 6: DeepFM model architecture.



# 3rd generation of recsys

An Algebra for Recommendations

Using Reader Data a

Jussi Karlgren

Department of Computer Science  
Stockholm University  
Electrum 230  
S-164 40 KISTA, Sweden

jussi@dsv.su.se  
jussi@flash.bellcore.com  
jussi@sc.columbia.edu

**GroupLens: A Collaborative Filtering System for Movie Recommendation**

**Paul Resnick\***, Neel Guha, Michael J. Huberman, and John Riedl  
\* MIT Center for Coordination of Compute and Communication  
50 Memorial Drive, Cambridge, MA 02139-1493  
Email: prenres@mit.edu

**ABSTRACT**  
Collaborative filters help people to find opinions of other people. GroupLens is a system that recommends articles they will like in movie reviews. New reader clients download past user ratings and use collaborative filtering based on the user's history and other information and recent years' posts and perform many more recommendations. More recommendations were found by increasing data sparsity. In total, the amount of work required to find recommendations that can be used is much less than what was needed for recommendations, even for a small number of users.

**KEYWORDS:** Collaborative filtering, electronic bulletin boards, social user model, selective dissemination of information

\* This research was funded by the Defense Advanced Research Projects Agency (DARPA) under contract AFRL-95-2-0001. This work was also partially funded by the National Science Foundation, by grants CCR-9402650 and CCR-9500316, and by the Defense Advanced Research Projects Agency (DARPA) under contract AFRL-95-2-0001. Part of this work was done while Jussi Karlgren was at the Research Institute of Electrical Communications, Tohoku University, Japan. Part of this work was done while Jussi Karlgren was with the Research Institute, NTT DoCoMo, Japan.

These reactions, more generally called annotations, can be applied to any document. One application of annotations is to support moderation of comments on news sites.

## Item-Based Collaborative Filtering Recommendation Algorithms

Badrul Sarwar, George Karypis, Joseph Konstan, and John Riedl  
GroupLens Research Group/Army HPC Research Center  
Department of Computer Science  
University of Minnesota, Minneapolis, MN 55455-0130 USA

**ABSTRACT**  
Recommender systems apply knowledge to the problem of making personalized information, products or services that are interesting to a user. Collaborative filtering based on user ratings is one of the most popular recommendation approaches. It uses information and recent years' posts and performs many more recommendations. More recommendations were found by increasing data sparsity. In total, the amount of work required to find recommendations that can be used is much less than what was needed for recommendations, even for a small number of users.

**INTRODUCTION**  
Recommender Systems are applications that are becoming a part of everyday life. They content discovery. The social prevalence television, video and online shopping.

The Summer School will bring together researchers (including Ph.D. students) on Recommender Systems and person provided.

Participation in the Summer School is limited.

### Lecturers will include:

Chris Anderson, Wired Magazine, US  
Todd Beaupre, Yahoo, Inc., US  
Jim Bennett, Netflix, Inc., US  
Dr. Alexander Felfernig, Klagenfurt University, Austria  
Dr. Rick Hengartner, MusicStrands, Inc., US  
Prof. Juntae Kim, Dongguk University, Korea  
Kausik Kurupati, Ask.com, US  
The lectures  
Paul Lamere, Sun Microsystems Labs, US  
Dr. Bamshad Mobasher, DePaul University, US  
Mike Mudie, Yahoo!, Inc., US



COMPLETED

## Collaborative Filtering

**On September 12-13, 2006 the 5th Summer School will be held**

**Introduction**  
Recommender Systems are applications that are becoming a part of everyday life. They content discovery. The social prevalence television, video and online shopping.

**Abstract**  
Customer preferences for products are shifting over time and popularity are constantly changing. Recommendation systems include them to ever change their taste. Thus, models should be a key when designing recommendation systems. Within the system interacting and consumers, many different characteristics are used. These characteristics are used to make shifts are delicate and associated with a few data points. This is a challenge for recommendation systems, especially when a single concept is tracked. Classical time-weighted decay approaches cannot work, as they lose too much information. Deep learning approaches appear which can make better distinctions between long term patterns. The paradigm we offer is a deep learning approach which uses a large amount of data. This allows us to exploit the relevant context information while dealing only with a small amount of data. Accordingly, we present a novel, efficient recommendation approach. Evaluation is made on two datasets: MovieLens and Netflix.

**Keynote speakers**  
As I like you can imagine, I can't speak about it.  
So for instance, I can't speak about it.  
I can't speak about it.  
Link to scales  
Or look at the answer  
Information provided by the user  
about his/her interests and frequency to reduce its memory footprint; second, we provide an production-ready online training architecture with high parallelism and expressiveness, so that the system can be trained on the fly. Monolith has successfully landed in the BigData+ Recommended! product.

## Monolith: Real Time Recommendation System With Collisionless Embedding Table

Zhuoran Liu  
Bytedance Inc.  
Caihua Wang  
Bytedance Inc.  
Bolin Zhu\*  
Fudan University  
Leqi Zou  
Bytedance Inc.  
Xuan Zou  
Bytedance Inc.  
Biao Zhang  
Bytedance Inc.  
Yijie Zhu  
Fudan University  
Ke Wang  
Bytedance Inc.  
Youtong Cheng\*  
Bytedance Inc.  
youtong.cheng@bytedance.com

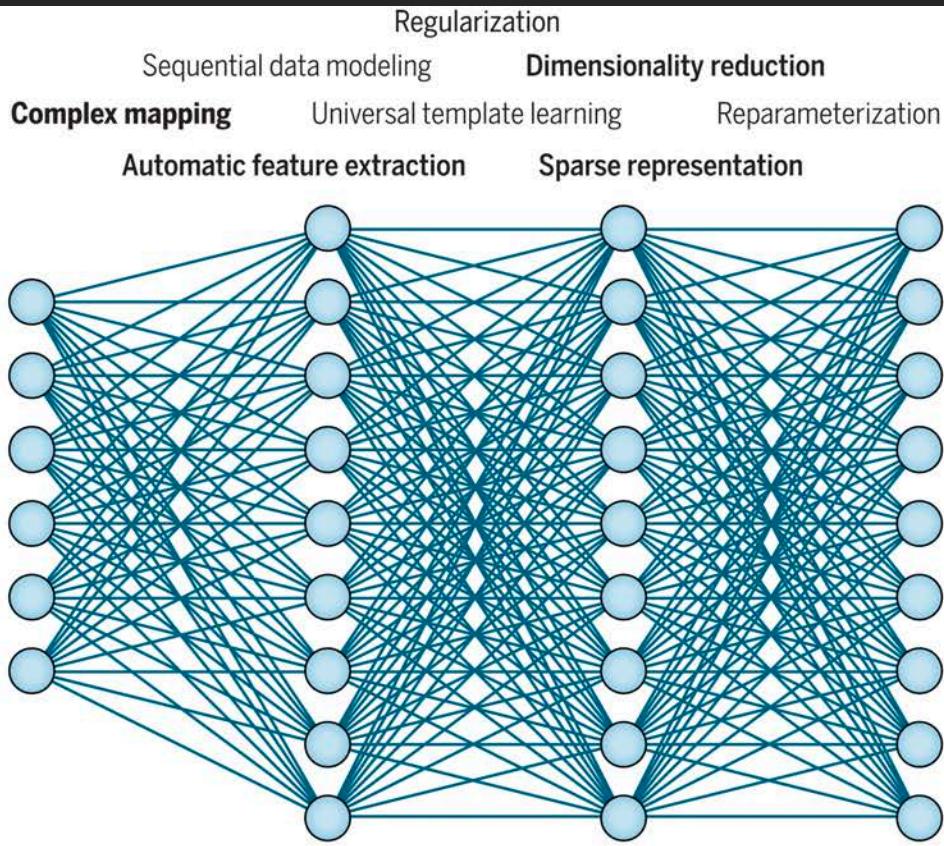
**ACM Reference Format:**  
Zhuoran Liu, Xuan Zou, Caihua Wang, Biao Zhang, Da Tang, Bolin Zhu, Yijie Zhu, Ke Wang, and Youtong Cheng, "Monolith: Real Time Recommendation System With Collisionless Embedding Table", In Proceedings of 3rd Workshop on Online Recommender Systems and User Modeling (OM3U), co-located with ACM Conference on Recommender Systems (RecSys) @ ACM RecSys 2022, New York, NY, USA, 10 pages. <https://doi.org/XXXXXX/XXXXXX>

### 1. INTRODUCTION

The past decade witnessed a boom of businesses powered by recommendation techniques. In pursuit of a better customer experience, delivering personalized content for each individual user as real-time response is a common goal of these business applications. To make informed decisions from user feedback, the primary input is user behavior. One application is recommendation systems system.

Deep learning have been dominating recommendation models [5, 6, 10, 12, 21] as the gigantic scale of user data is a natural fit for the power of deep learning models. However, how to leverage the power of deep learning in industry-level recommendation systems are constantly confronted with problems arising from the nature of recommendation tasks and user behavior. These data are drastically different from those used

1951 1967 1990 1992 1994 2001 2006 2008 2009 2010 2016 2022



## 3rd generation of recommender systems

- 10s-20s
- Deep learning, implicit feedback
- Non-intuitive algorithms
- Even harder to explain

## An Algebra for Recommendations

### Using Reader Data

Jussi Karlgren

Department of Computer  
Stockholm University  
Electrum 230  
S-164 40 KISTA, Sweden

jussi@dsv.su.se  
jussi@flash.bellcore.com  
jussi@sc.columbia.edu

### GroupLens: A Collaborative Filtering System for Movie Recommendation

Paul Resnick\*, Neel Guha

\* MIT Center for Coordination of Compute and Comms, 50 Memorial Drive, Cambridge, MA 02139, USA  
Email: prenres@mit.edu

Badrul Sarwar, George Karypis, Joseph Konstan, and John Riedl  
GroupLens Research Department, University of Minnesota, Minneapolis, MN 55455, USA  
Research Group/Arcy HPC Research Center



NETFLIX

**NETFLIX Prize**  
Home Rules Leaderboard Update  
**COMPLETED!**

#### ABSTRACT

Recommender systems apply knowledge to the problem of matching personalized information, products or services that users are likely to be interested in. Collaborative filtering based on ratings provided by many users is one way to perform many recommendations even from sparse data. In this paper we describe the Netflix Prize competition, which asked researchers to perform many recommendations even from sparse data.

**Collaborative filters help people get opinions of other people. GroupLens has helped people find the articles they will like in the past. News reader clients download news from the web and then base their recommendations on what others have read. News reader clients download news from the web and then base their recommendations on what others have read.**

News reader clients download news from the web and then base their recommendations on what others have read. News reader clients download news from the web and then base their recommendations on what others have read.

**KEYWORDS:** Collaborative filters, electronic bulletin boards, user model, selective dissemination of information.

**寂寞是一种状态，更是一种感觉。电影的寂寞，就是一种孤独感，一种对生活的厌倦和失望，一种对未来的不确定和恐惧。寂寞的人，往往会有强烈的自我封闭感，不愿意与他人接触和交流，但同时又渴望得到关注和认可。寂寞的人，往往会陷入回忆和幻想，无法真正地融入现实生活中。**

**寂寞的人，往往会有强烈的自我封闭感，不愿意与他人接触和交流，但同时又渴望得到关注和认可。寂寞的人，往往会陷入回忆和幻想，无法真正地融入现实生活中。**

## Item-Based Collaborative Filtering Recommendation Algorithms

Badrul Sarwar, George Karypis, Joseph Konstan, and John Riedl  
GroupLens Research Group/Army HPC Research Center

### Collaborative Filtering



COMPLETED!  
Netflix Prize  
Home Rules Leaderboard Update  
COVER FEATURE



### Collaborative Filtering

The 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Seattle, Washington, USA, 12-16 August 2006  
Workshop on Recommender Systems, September 12-13, 2006, Seattle, Washington, USA

**ABSTRACT**  
Customer preferences for products are shifting over time and popularity are constantly changing. Recommendation systems are needed to handle such changes. We present a system that performs item-based collaborative filtering based on a matrix factorization model. The system is able to learn latent factors from the user's history and predict items based on the user's interests. The system is also able to learn user's profile based on their ratings.

**Introduction**  
Recommender Systems are becoming a part of everyday life. They content discovery. The social prevalence, television, video and online shopping.

**Abstract**  
Recommender Systems are shifting over time and popularity are constantly changing. Recommendation systems are needed to handle such changes. We present a system that performs item-based collaborative filtering based on a matrix factorization model. The system is able to learn latent factors from the user's history and predict items based on the user's interests. The system is also able to learn user's profile based on their ratings.

**Introduction**  
Recommender Systems are becoming a part of everyday life. They content discovery. The social prevalence, television, video and online shopping.

**Abstract**  
Recommender Systems are shifting over time and popularity are constantly changing. Recommendation systems are needed to handle such changes. We present a system that performs item-based collaborative filtering based on a matrix factorization model. The system is able to learn latent factors from the user's history and predict items based on the user's interests. The system is also able to learn user's profile based on their ratings.

**Introduction**  
Recommender Systems are becoming a part of everyday life. They content discovery. The social prevalence, television, video and online shopping.

**Abstract**  
Recommender Systems are shifting over time and popularity are constantly changing. Recommendation systems are needed to handle such changes. We present a system that performs item-based collaborative filtering based on a matrix factorization model. The system is able to learn latent factors from the user's history and predict items based on the user's interests. The system is also able to learn user's profile based on their ratings.

## Deep Neural Net

Published as a conference paper at the 2015 ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Seattle, Washington, USA, 12-16 August 2006  
Workshop on Recommender Systems, September 12-13, 2006, Seattle, Washington, USA

### SESSION-DEPENDENT RECOMMENDATION

Balázs Hidasi<sup>1</sup>, Tamás Salakhutdinov<sup>2</sup>, George Bakir<sup>1</sup>  
1 Google Brain Team, Berkeley, CA, USA  
2 University of California, Berkeley, CA, USA  
hidasi.balazs.hidasi@gmail.com, tamas.salakhutdinov@berkeley.edu, bakir@berkeley.edu

László Batrancú<sup>1</sup>, Balázs Hidasi<sup>2</sup>, Tamás Salakhutdinov<sup>2</sup>, George Bakir<sup>1</sup>  
1 Netflix, Los Gatos, CA, USA  
batrancu.laszlo.netflix@gmail.com, hidasi.balazs.netflix@gmail.com, salakhutdinov.tamas.netflix@gmail.com, bakir@berkeley.edu

## Monolith: Real Time Recommendation System With Collisionless Embedding Table

Zhuorun Liu, Leqi Zou, Xuan Bi, Caihua Wang, Biao Zhang, Da Tang, Bolin Chen, Peng Wu, and Yijie Zhu  
YoutuTech, Bytedance Inc.  
Zhuorun Liu, Leqi Zou, Xuan Bi, Caihua Wang, Biao Zhang, Da Tang, Bolin Chen, Peng Wu, and Yijie Zhu  
ACM Reference Format:  
Zhuorun Liu, Leqi Zou, Xuan Bi, Caihua Wang, Biao Zhang, Da Tang, Bolin Chen, Peng Wu, and Yijie Zhu. 2015. Monolith: Real Time Recommendation System With Collisionless Embedding Table. In Proceedings of the 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '15), San Jose, CA, USA, 12–16 August 2015, 161–169. DOI: <https://doi.org/10.1145/2822432.2822510>

**ABSTRACT**  
Building a scalable and real-time recommendation system is vital for many businesses driven by real-time customer feedback, adoption of production-scale deep learning frameworks like TensorFlow or PyTorch, these general-purpose frameworks fall short of business-specific needs, while existing recommendation systems are mostly designed for static datasets and sparse data. This thesis addresses these challenges by proposing a system based on neural networks and dense computations for recommendation with dynamic training stage completely separated, preventing the model from interacting with customer feedback in real-time. These issues led us to recommenders that are explicitly trained and expertly tuned for specific choices. In this paper, we present *Monolith*, a system tailored for online training. Our design has been driven by observations of the field, we found that the traditional recommendation system exhibits a marked departure from other recommendation systems. Our contributions are manifold; first, we crafted a collisionless embedding table to handle collisions in the memory space, second, we provide an production-ready online training architecture with high parallelism and low latency, i.e. real-time training. Monolith's efficiency could be traded-off only for real-time learning. Monolith has successfully landed in the *Bytedance Recommand* product.

Our approach also considers frequent items, which are commonly used with sparse data. Monolith's recommendation system is based on a neural loss function that make it more viable for this specific problem. Experimental results on two data-sets show marked improvements over widely used approaches.

**Deep learning has been dominating recommendation models [5, 6, 10, 12, 21] as the growing size of user data is a natural fit for deep learning models. However, deep learning fails to leverage the power of deep learning in industry-level recommendation systems are constantly faced with problems arising from the massive amount of data derived from the user behavior. These data are drastically different from those used**

1967  
1951 1990 1992 1994 2001 2006 2008 2009 2010 2016 2022







## An Algebra for Recommendations

### Using Reader Data as Input

Jussi Karlgren

Department of Computer Science  
Stockholm University  
Electrum 230  
S-164 40 KISTA, Sweden

jussi@dsv.su.se  
jussi@flash.bellcore.com  
jussi@csc.columbia.edu

**ABSTRACT**  
Collaborative filters help people to find products or articles they will like based on the opinions of other people. Most recommendations are based on user data and few make use of reader data. New reader clients download products from the company's Rating servers, called Better Browsers. These clients rate products based on the heuristic that people who have similar interests are likely to enjoy ratings under purchase effectiveness of the user prediction. Two new rating models are proposed. Both can be derived independently of the user profile.

**KEYWORDS:** Collaborative filtering; electronic bulletin boards; user model; selective dissemination

**TUTORSHIP:** It is in progress.  
**ADVISOR:** Dr. Jussi Karlgren  
**COLLABORATOR:** Prof. Dr. Michael Stoyanovich, University of Illinois at Urbana-Champaign, IL, USA  
**PUBLICATION:** Accepted to appear in the Proceedings of the International Conference on Electronic Commerce (ICEC), July 2006, Seoul, Korea, by the Electronic Commerce Society of Korea. Part of the conference is the Summer School on E-commerce. There is a workshop on Recommendation Systems, held 15-16 July 2006.

These reactions, more generally called annotations, can be used to refine the user profiles. Some applications of annotations in the support of moderation

## GroupLens: A Research Project

**Paul Resnick\***, Neeraj

\* MIT Center for Coordination of Technology, Policy, and Law, 50 Memorial Drive, Cambridge, MA 02139  
Email: pren@mit.edu

Badrul Sarwar, George Karypis, Joseph Konstan, and John Riedl  
GroupLens Research Group/Army HPC Research Center



Home | Rules | Leaderboard | Update

NETFLIX  
PRIZE  
COMPLETED

COVER FEATURE



**Collaborative Fi**

i On September 12-13, 2006 the Summer School will be held

## Introduction

Recommender Systems are application becoming a part of everyday life. They content discovery. The social prevalent television, video and on-line shopping.

The Summer School will bring together researchers (including Ph.D. students) on Recommender Systems and person provided.

Participation in the Summer School is t

## Lecturers will include:

Chris Anderson, Wired Magazine, US  
Todd Beaupre, Yahoo, Inc., US  
Jim Bennett, Netflix, Inc., US  
Dr. Alexander Felfernig, Klagenfurt University, Austria  
Dr. Rick Houghtaling, MusicStrands, Inc., US  
Prof. Juntae Kim, Donguk University, Korea  
Kaushtab Kurapati, Ask.com, US to the lectures  
Paul Lamere, Sun Microsystems Labs, US  
Dr. Bamshad Mobasher, DePaul University, US  
Mike Modl, Yahoo, Inc., US

**ABSTRACT**  
Customer preferences for products are shifting over perception and popularity are constantly changing. This makes it difficult to predict what items are interesting to them to ever redefine their taste. Thus, models should be a key when designing recommender systems. This paper presents several challenges. Within the eco-system interacting with consumers, many different characteristics are involved with the item and the consumer. These shifts are dynamic and associated with a few data points. This paper highlights a few examples of how mostly a single concern is tracked. Classical time-w decay approaches cannot work, as they lose too much information about the consumer's previous purchases which can make better distinctions between transient and long term patterns. The paradigm we offer is a hybrid which decouples recommendation engines from data. This allows us to exploit the relevant context information while decoupling the recommendation engines. Accordingly, we propose two-level collaborative recommendation approaches. Evaluation is made on several real datasets.

**ABSTRACT**  
The recommender system is a field that has been receiving a lot of attention in recent years. However, it is still not clear what the same is happening between item and user. In this paper, we introduce a new type of Likert scales, called "recurrent Likert scales". As just as in classical Likert scales, recurrent Likert scales can be used. So for instance, look at the following example, look at the figure.

As just as in classical Likert scales, recurrent Likert scales can be used. So for instance, look at the following example, look at the figure.



Or look at the answer to this question:  
What do you think the average income is for a person in the U.S.?  
Income is probably a product that provides a good guess about a person's tastes. Because recommendations can add another experience, e-commerce leaders have built recommendation systems as a central part of their websites.

We apply a multi-stage approach to reduce noise and frequency filtering to reduce it provide an production-ready online recommendation system. Our contributions are manifold: first, we propose a model that can be trained off-the-shelf LLMs. The frequent feedback loop is used to evaluate the quality of generated LLMs. Recurrent Likert scales can be used. So for instance, look at the following example, look at the figure.

**1. Introduction**  
Recommender systems play an increasingly important role in our everyday lives. With the emergence of Large Language Models (LLMs) in

1967  
1951  
1990  
1992  
1994  
2001  
2006  
2008  
2009  
2010  
2016  
2022  
2023  
2024



## Articles

About 25 900 results (0,12 sec)

Any time

Since 2024

Since 2023

Since 2020

Custom range...

2023 — 2023

Search

Sort by relevance

Sort by date

Any type

Review articles

 include patents include citations Create alert**Retrieval-augmented recommender system: Enhancing recommender systems with large language models**

[PDF] acm.org

D Di Palma - ... the 17th ACM Conference on Recommender Systems, 2023 - dl.acm.org

... of Large Language Models and prompt engineering. It delves into a comprehensive analysis of the application of LLMs in Recommender Systems. ... as a Recommender System. Finally, ...

☆ Save 99 Cite Cited by 35 Related articles All 4 versions

**Recommender systems in the era of large language models (llms)**

[PDF] arxiv.org

Z Zhao, W Fan, J Li, Y Liu, X Mei, Y Wang... - arXiv preprint arXiv ..., 2023 - arxiv.org

... In this section, we will introduce these two categories that take advantage of language models in recommender systems. These two kinds of recommender systems are illustrated in ...

☆ Save 99 Cite Cited by 154 Related articles All 2 versions

**Leveraging large language models for pre-trained recommender systems**

[PDF] arxiv.org

Z Chu, H Hao, X Ouyang, S Wang, Y Wang... - arXiv preprint arXiv ..., 2023 - arxiv.org

... Our model can also be adapted to other large language models with minor modifications. ... of pretraining recommendation models based on large language models. We introduce a ...

☆ Save 99 Cite Cited by 36 Related articles All 2 versions

**Tutorial on large language models for recommendation**

[PDF] acm.org

W Hua, L Li, S Xu, L Chen, Y Zhang - ... on Recommender Systems, 2023 - dl.acm.org

... introduces Foundation Models such as LLMs for recommendation. We will introduce how recommender system advanced from shallow models to deep models and to large models, ...

☆ Save 99 Cite Cited by 24 Related articles All 5 versions

**Large language models are competitive near cold-start recommenders for language-and item-based preferences**

[PDF] acm.org

S Sanner, K Balog, F Radlinski, B Wedin... - ... on recommender systems, 2023 - dl.acm.org

... large language models (LLMs), we study their use for making recommendations from both item-based and language... of both item-based and language-based preferences elicited from ...

☆ Save 99 Cite Cited by 62 Related articles All 9 versions

**[HTML] Large language models as recommendation Systems in Museums**

[HTML] mdpi.com

G Trichopoulos, M Konstantakis, G Alexandridis... - Electronics, 2023 - mdpi.com

... large language models as recommendation systems for museum visitors. Since the aforementioned models ... that is often present in recommendations for cultural environments (eg, ...

☆ Save 99 Cite Cited by 17 Related articles All 5 versions



## Articles

About 25 900 results (0,12 sec)

Any time

Since 2024

Since 2023

Since 2020

Custom range...

2023 — 2023

Search

Sort by relevance

Sort by date

Any type

Review articles

 include patents include citations Create alert**Large language models in recommender system: Enhancing recommender systems with large language models**

[PDF] acm.org

D Di Palma - ... the 17th ACM Conference on Recommender Systems, 2023 - dl.acm.org

... of Large Language Models and prompt engineering. It delves into a comprehensive analysis of the application of LLMs in Recommender Systems. ... as a Recommender System. Finally, ...

☆ Save 99 Cite Cited by 35 Related articles All 4 versions

**Recommender systems in the era of large language models (llms)**

[PDF] arxiv.org

Z Zhao, W Fan, J Li, Y Liu, X Mei, Y Wang... - arXiv preprint arXiv ..., 2023 - arxiv.org

... In this section, we will introduce these two categories that take advantage of language models in recommender systems. These two kinds of recommender systems are illustrated in ...

☆ Save 99 Cite Cited by 154 Related articles All 2 versions

**Leveraging large language models for pre-trained recommender systems**

[PDF] arxiv.org

Z Chu, H Hao, X Ouyang, S Wang, Y Wang... - arXiv preprint arXiv ..., 2023 - arxiv.org

... Our model can also be adapted to other large language models with minor modifications. ... of pretraining recommendation models based on large language models. We introduce a ...

☆ Save 99 Cite Cited by 36 Related articles All 2 versions

**Tutorial on large language models for recommendation**

[PDF] acm.org

W Hua, L Li, S Xu, L Chen, Y Zhang - ... on Recommender Systems, 2023 - dl.acm.org

... introduces Foundation Models such as LLMs for recommendation. We will introduce how recommender system advanced from shallow models to deep models and to large models, ...

☆ Save 99 Cite Cited by 24 Related articles All 5 versions

**Large language models are competitive near cold-start recommenders for language-and item-based preferences**

[PDF] acm.org

S Sanner, K Balog, F Radlinski, B Wedin... - ... on recommender systems, 2023 - dl.acm.org

... large language models (LLMs), we study their use for making recommendations from both item-based and language... of both item-based and language-based preferences elicited from ...

☆ Save 99 Cite Cited by 62 Related articles All 9 versions

**[HTML] Large language models as recommendation Systems in Museums**

[HTML] mdpi.com

G Trichopoulos, M Konstantakis, G Alexandridis... - Electronics, 2023 - mdpi.com

... large language models as recommendation systems for museum visitors. Since the aforementioned models ... that is often present in recommendations for cultural environments (eg, ...

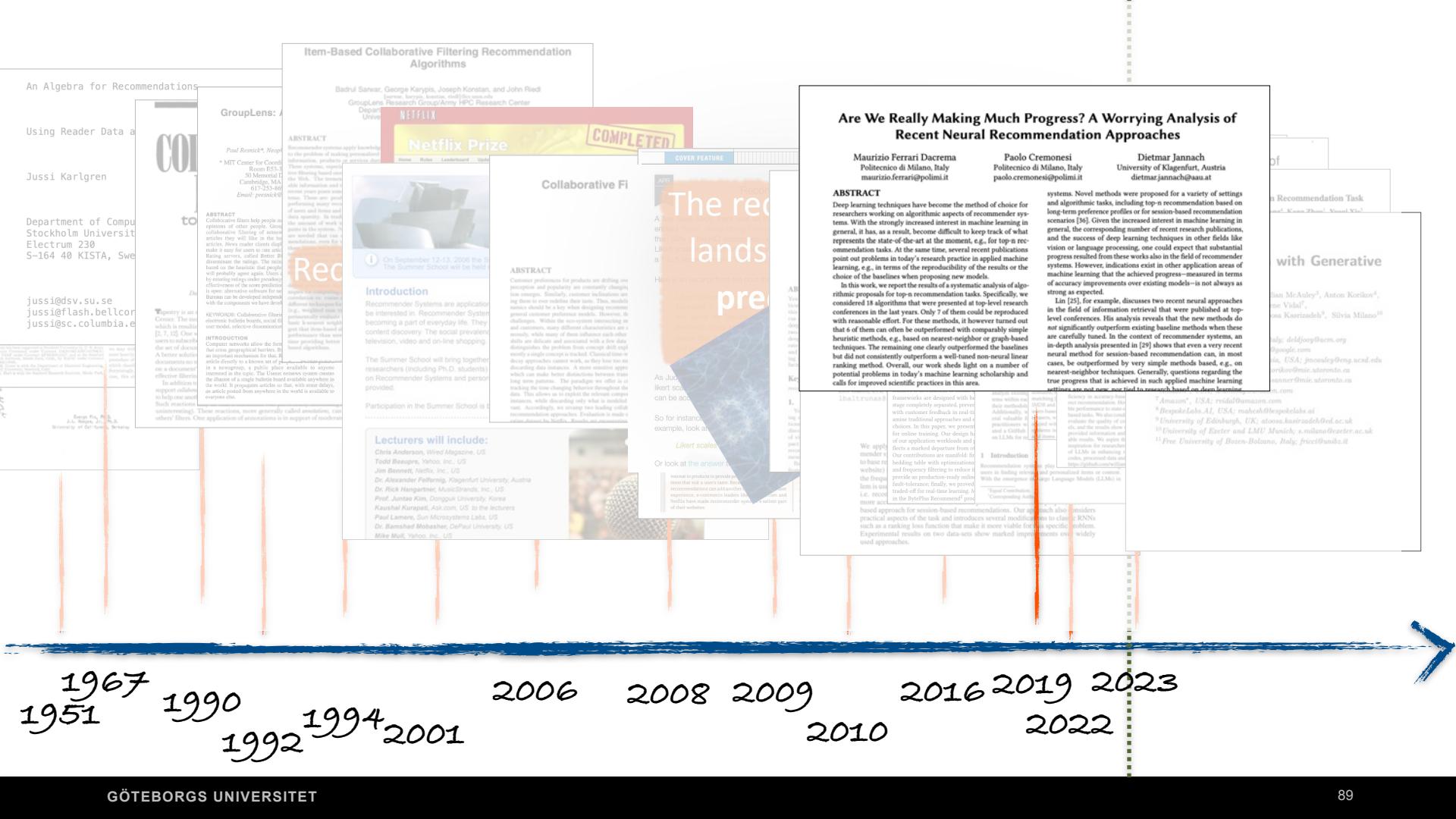
☆ Save 99 Cite Cited by 17 Related articles All 5 versions



## 4th generation of recommender systems

- 20s
- LLMs
- Generative recommendations

Hold on...



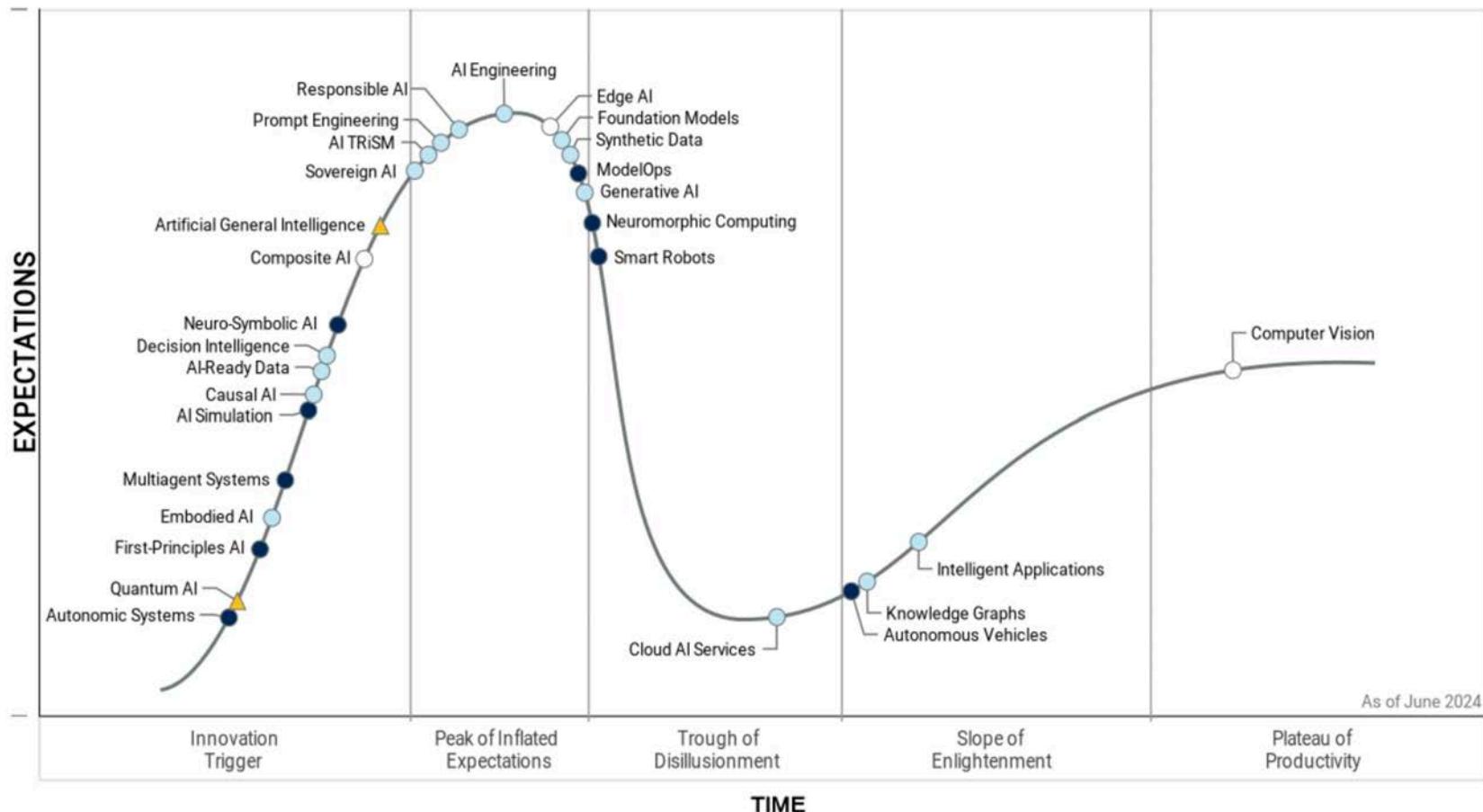




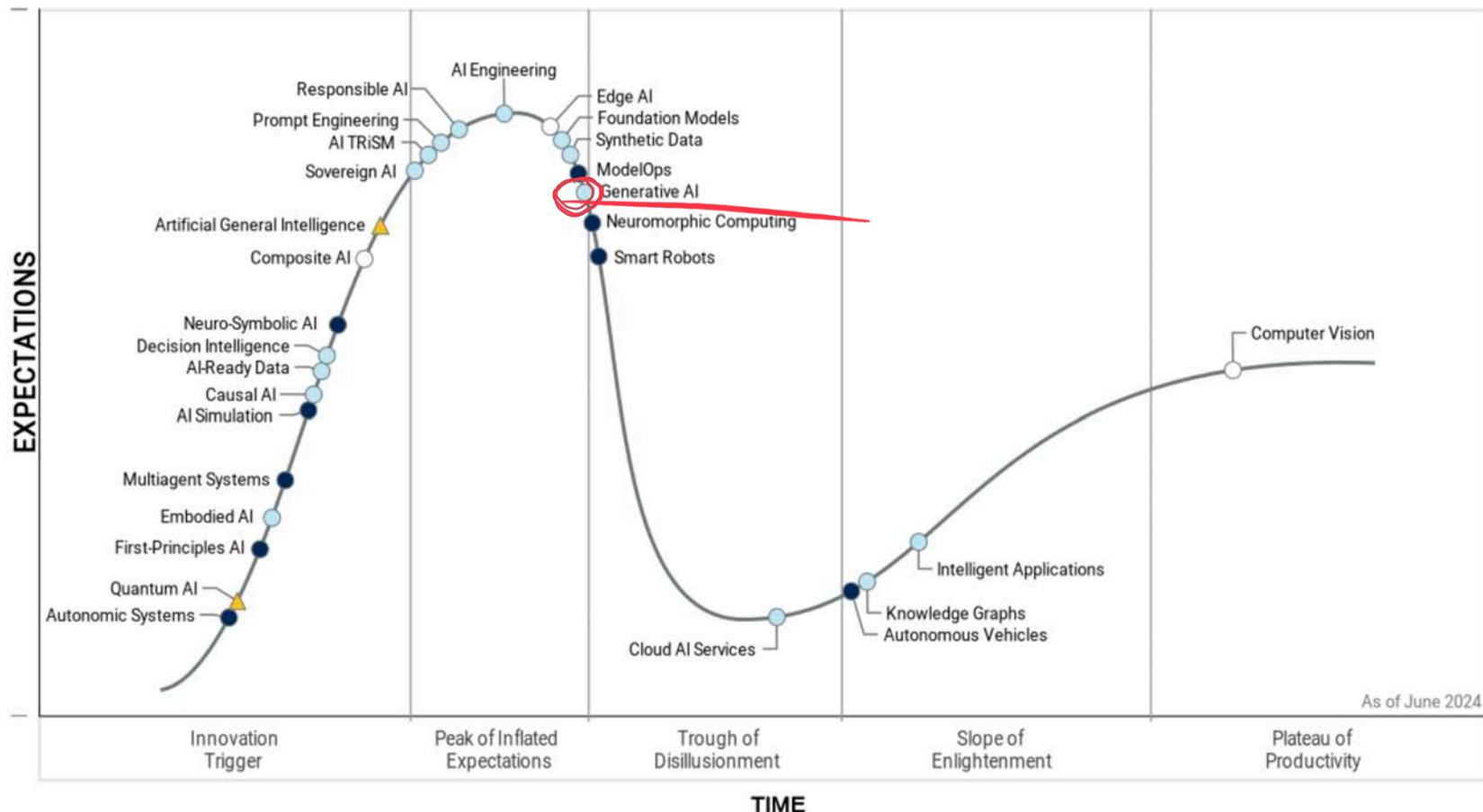


WHAT  
NOW?

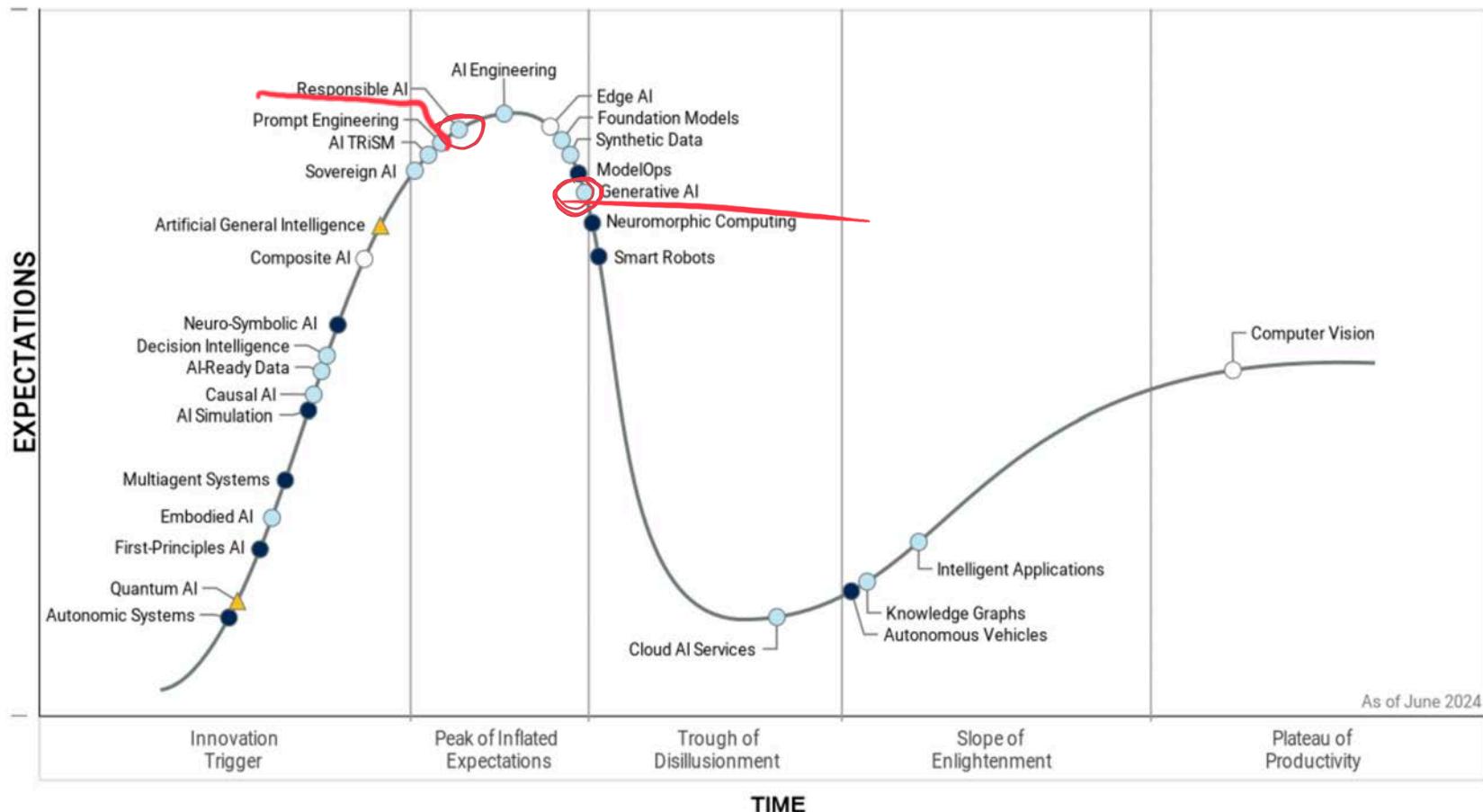
# Hype Cycle for Artificial Intelligence, 2024



# Hype Cycle for Artificial Intelligence, 2024

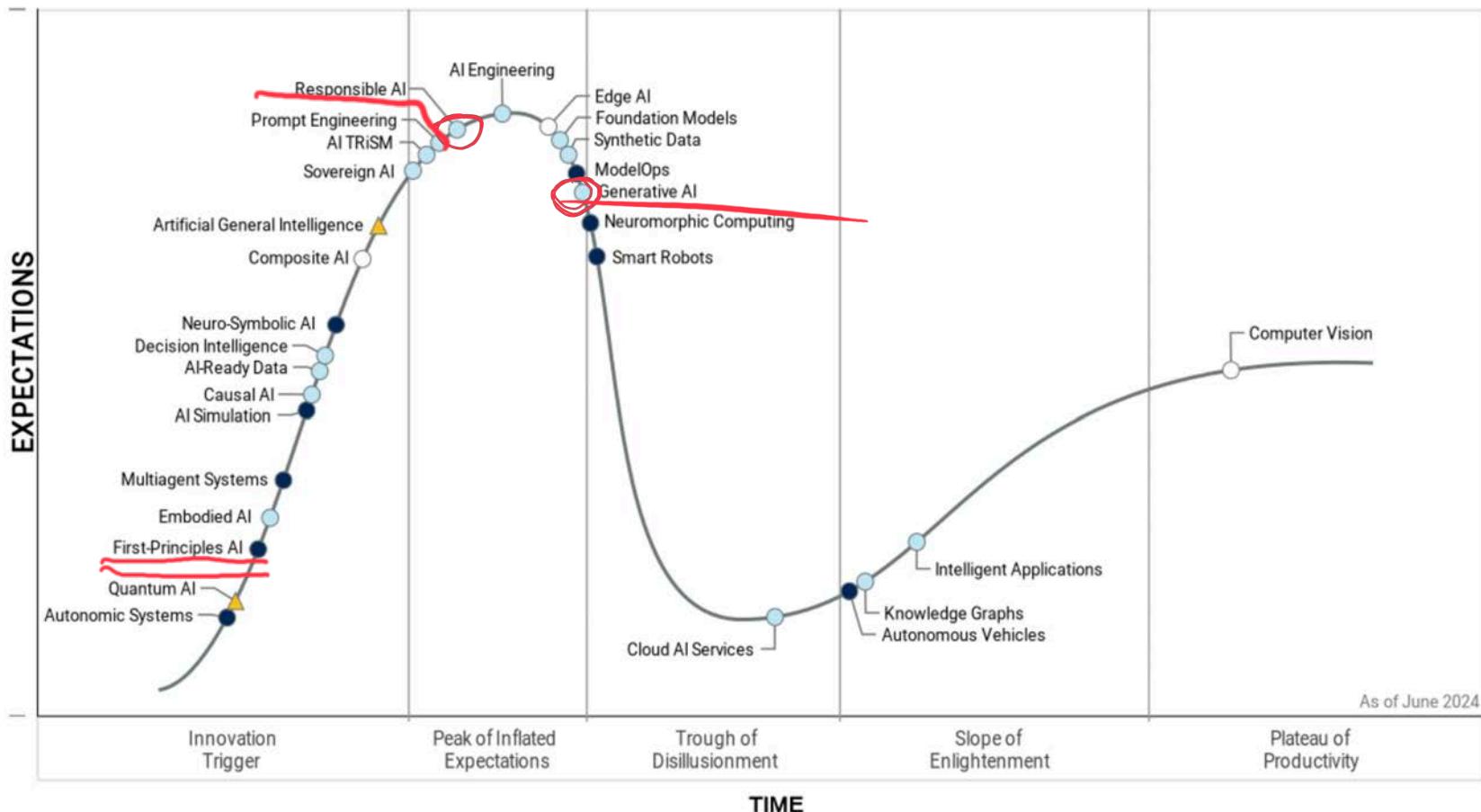


# Hype Cycle for Artificial Intelligence, 2024



Plateau will be reached: ○ <2 yrs. ● 2–5 yrs. ■ 5–10 yrs. ▲ >10 yrs. ✘ Obsolete before plateau

# Hype Cycle for Artificial Intelligence, 2024



Plateau will be reached: ○ <2 yrs. ● 2–5 yrs. ■ 5–10 yrs. ▲ >10 yrs. ✘ Obsolete before plateau

## Shaping Europe's digital future

[Home](#)   [Policies](#)   [News](#)   [Library](#)   [Funding](#)   [Calendar](#)   [Consultations](#)[Home](#) > [Library](#) > Ethics guidelines for trustworthy AI

REPORT / STUDY

## Ethics guidelines for trustworthy AI

On 8 April 2019, the High-Level Expert Group on AI presented Ethics Guidelines for Trustworthy Artificial Intelligence. This followed the publication of the guidelines' first draft in December 2018 on which more than 500 comments were received through an open consultation.

According to the Guidelines, trustworthy AI should be:

- (1) lawful - respecting all applicable laws and regulations
- (2) ethical - respecting ethical principles and values
- (3) robust - both from a technical perspective while taking into account its social environment

### Related topics

[Artificial intelligence](#)[Investing in network and technologies](#)

The Guidelines put forward a set of 7 key requirements that AI systems should meet in order to be deemed trustworthy. A specific assessment list aims to help verify the application of each of the key requirements:

- Human agency and oversight: AI systems should empower human beings, allowing them to make informed decisions and fostering their fundamental rights. At the same time, proper oversight mechanisms need to be ensured, which can be achieved through human-in-the-loop, human-on-the-loop, and human-in-command approaches
- Technical Robustness and safety: AI systems need to be resilient and secure. They need to be safe, ensuring a fall back plan in case something goes wrong, as well as being accurate, reliable and reproducible. That is the only way to ensure that also unintentional harm can be minimized and prevented.
- Privacy and data governance: besides ensuring full respect for privacy and data protection, adequate data governance mechanisms must also be ensured, taking into account the quality and integrity of the data, and ensuring legitimised access to data.
- Transparency: the data, system and AI business models should be transparent. Traceability mechanisms can help achieving this. Moreover, AI systems and their decisions should be explained in a manner adapted to the stakeholder concerned. Humans need to be aware that they are interacting with an AI system, and must be informed of the system's capabilities and limitations.
- Diversity, non-discrimination and fairness: Unfair bias must be avoided, as it could have multiple negative implications, from the marginalization of vulnerable groups, to the exacerbation of prejudice and discrimination. Fostering diversity, AI systems should be accessible to all, regardless of any disability, and involve relevant stakeholders throughout their entire life circle.
- Societal and environmental well-being: AI systems should benefit all human beings, including future generations. It must hence be ensured that they are sustainable and environmentally friendly. Moreover, they should take into account the environment, including other living beings, and their social and societal impact should be carefully considered.
- Accountability: Mechanisms should be put in place to ensure responsibility and accountability for AI systems and their outcomes. Auditability, which enables the assessment of algorithms, data and design processes plays a key role therein, especially in critical applications. Moreover, adequate and accessible redress should be ensured.



# Translation: Internet Information Service Algorithmic Recommendation Management Provisions (Opinion-Seeking Draft)

Major new regulations on the use of recommendation algorithms issued for public comment

August 27, 2021 | Rogier Creemers, Helen Toner, Graham Webster



The Cyberspace Administration of China on Aug. 27 [released](#) a draft "Internet Information Service Algorithmic Recommendation Management Provisions" 《互联网信息服务算法推荐管理规定（征求意见稿）》for public comment, with submissions due Sept. 26.

## TRANSLATION

### Internet Information Service Algorithmic Recommendation Management Provisions (Opinion-seeking Draft)



## 5th generation of recommender systems



## 5th generation of recommender systems

- you decide

# Questions?



Human-centered AI



References used in this talk



RecSys for Social Good SI @ TORS



GÖTEBORGS  
UNIVERSITET

# References

- Fix and Hodges. Discriminatory Analysis - Nonparametric Discrimination: Consistency Properties. (1951) <https://apps.dtic.mil/sti/citations/ADA800276>
- Cover and Hart. Nearest neighbor pattern classification. (1967) <https://ieeexplore.ieee.org/document/1053964>
- Karlgren. An Algebra for Recommendations. (1990) <https://jussikarlgren.wordpress.com/wp-content/uploads/1990/09/algebrawp.pdf>
- Goldberg, Nichols, Oki, and Terry. Using collaborative filtering to weave an information tapestry. (1992) <https://doi.org/10.1145/138859.138867>
- Resnick, Iacovou, Suchak, Bergstrom, and Riedl. GroupLens: an open architecture for collaborative filtering of netnews. (1994) <https://doi.org/10.1145/192844.192905>
- Sarwar, Karypis, Konstan, and Riedl. Item-based collaborative filtering recommendation algorithms. (2001) <https://doi.org/10.1145/371920.372071>
- Funk. Netflix Update: Try This at Home. (2006) <https://sifter.org/simon/journal/20061211.html>
- Koren. Collaborative filtering with temporal dynamics. (2009) <https://doi.org/10.1145/1557019.1557072>
- Koren, Bell, Volinsky. Matrix Factorization Techniques for Recommender Systems. (2009) <https://ieeexplore.ieee.org/document/5197422>
- Amatriain. Recommender Systems: We're doing it (all) wrong. (2011) <https://technocalifornia.blogspot.com/2011/04/recommender-systems-were-doing-it-all.html>
- Rajaraman. Five Stars Dominate Ratings. (2009) <https://blog.youtube/news-and-events/five-stars-dominate-ratings/>
- Covington, Adams, and Sargin. Deep Neural Networks for YouTube Recommendations. (2016) <https://doi.org/10.1145/2959100.2959190>
- Hidasi, Karatzoglou, Baltrunas, and Tikk. Session-based Recommendations with Recurrent Neural Networks. (2016) <https://arxiv.org/abs/1511.06939>
- Liu, Zou, Zou, Wang, Zhang, Tang, Zhu, Zhu, Wu, Wang, Cheng. Monolith: Real Time Recommendation System With Collisionless Embedding Table. (2022) <https://arxiv.org/abs/2209.07663>
- Wu, Zheng, Qiu, Wang, Gu, Shen, Qin, Zhu, Zhu, Liu, Xiong, Chen. A Survey on Large Language Models for Recommendation. (2023) <https://arxiv.org/abs/2305.19860>
- Zhao, Fan, Li, Liu, Mei, Wang, Wen, Wang, Zhao, Tang, Li. Recommender Systems in the Era of Large Language Models (LLMs). (2023) <https://arxiv.org/abs/2307.02046>
- Liu, Liu, Zhou, Ye, Chong, Zhou, Xie, Cao, Wang, You, Yu. LLMRec: Benchmarking Large Language Models on Recommendation Task. (2023) <https://arxiv.org/abs/2308.12241>
- Deldjoo, He, McAuley, Korikov, Sanner, Ramisa, Vidal, Sathiamoorthy, Kasrizadeh, Milano, Ricci. Recommendation with Generative Models. (2024) <https://arxiv.org/abs/2409.15173>
- Ferrari Dacrema, Cremonesi, and Jannach. Are we really making much progress? A worrying analysis of recent neural recommendation approaches. (2019) <https://doi.org/10.1145/3298689.3347058>
- Said and Bellogín. Comparative recommender system evaluation: benchmarking recommendation frameworks. (2014) <https://doi.org/10.1145/2645710.2645746>
- Armstrong, Moffat, Webber, and Zobel. Improvements that don't add up: ad-hoc retrieval results since 1998. (2009) <https://doi.org/10.1145/1645953.1646031>
- Balby Marinho. Explainability. (2023) RecSys Summer School 2023