

# A Contribution to Rating and Recommendation Systems: Concepts, Development and Evaluation

Oliver Diestel

March 4, 2013

# Current Situation (Just an assumption)

A screenshot of a Google search results page for the query "Science 2.0". The search bar at the top shows "Science 2.0" and a magnifying glass icon. Below the search bar, there are tabs for "Web", "Bilder", "Maps", "Shopping", "News", and "Mehr". The "Web" tab is selected. The search results list several links. The second result, "Forschungsverbund Science 2.0 - Science 2.0 - Web der ZBW - extern", is highlighted with a red rectangular box. A red arrow points from this box to the ZBW website shown on the right. Other search results include "Science 2.0 : @ The world's best scientists, the Internet's smartest ...", "www.sciencedi.com - Diese Seite übersetzen", "Physics - About Science 2.0 : And ... - Who is Staking You On ... - Psychology", "Science 2.0 - Wikipedia, the free encyclopedia", "Science 2.0 is a somewhat controversial umbrella term, not precisely defined, which describes a range of activities, described by proponents of the term as ...", "Science 2.0 Leibniz Forschungsverbund", "Science 2.0: Social Network für Wissenschaftler | heise online", "Science 2.0 - Leibniz Gemeinschaft", and "Science 2.0 - Is Open Access Science the Future? Scientific ...".

A screenshot of the ZBW (Zentralbibliothek Bonn) website. The header features the ZBW logo and navigation links. The main content area is titled "Leibniz-Forschungsverbund Science 2.0" and describes the network's mission. It includes a section "Über uns" (About us) and a "Kontakt" (Contact) section. The website is designed with a clean, professional layout, featuring a mix of text and images.

User visits the zbw.eu website and searches for the information that he would like to have. Leaves the website.

# Possible Situation



**Forschungsverbund Science 2.0 - Science 2.0 - Wiki der ZBW** - extern  
wiki.zbw.eu | Übersuche Science 2.0  
19.02.2019 - Science 2.0 befasst sich im Grunde mit der Frage, wie das Internet mit seiner zahlreichen (fast 2 Millionen) Forschungs- und ...

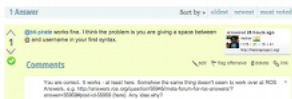
**Science 2.0 - Wikipedia, the free encyclopedia**  
en.wikipedia.org/wiki/Science\_2.0 - Diese Seite übersetzen  
Science 2.0 is a somewhat controversial umbrella term, not precisely defined, which describes a range of activities, described by proponents of the term as ...  
Definitions - History and background - Proliferation of Science 2.0 ... See also

**Science 2.0 Leibniz Forschungsverbund**  
www.leibniz-science20.de  
Willkommen in der Zukunft! 30 Einrichtungen erforschen neue Forschungs- und Publikationsprozesse. Alle Informationen über Forschungsverbünde und ...

**Science 2.0: Social Network für Wissenschaftler | heise online**  
www.heise.de - News - 2009 - KOF 10  
02.03.2009 - Auf ResearchGate, einem Social Network speziell für Wissenschaftler, lassen sich jetzt öffentliche Online-Kartenkarten anzeigen, die alle ...

**Science 2.0 - Leibniz Gemeinschaft**  
www.leibniz-gemeinschaft.de/forschung/leibniz-science20/  
In den fünf Forschungsverbünden schließen sich mehrere Leibniz-Einrichtungen zusammen.

**Science 2.0 - Is Open Access Science the Future? - Scientific ...**  
www.scientificamerican.com/article.cfm? ... - Diese Seite übersetzen  
11.04.2008 - Is research now really free? For all its use, a recent trend in a recent decade?



User visits the website, sees an interesting question from another user, reads the question and the answer, hopefully asks more questions. Goal: Increase the user interaction on the website.

# Q/A System

The screenshot shows the AskWoG website interface. At the top, there are navigation tabs: "questions", "tags", "people", "badges", and "ask a question". Below these is a search bar. The main content area displays a list of questions, each with a title, a brief description, and a set of statistics (votes, answers, views). The questions include topics like "My Gameduino page", "Can You plug in a Wii Nunchuck?", "Gameduino Reference Manual v01", "Due compatibility?", "Can I use the Gameduino on a 60Hz TV?", "Char or sprites with random maze?", "Gameduino with MicroSD card breakout", and "gameduino reads graphics from sd card". To the right of the questions is a "Contributors" section showing a grid of user avatars. At the bottom, there is a "New tool for convert images" link.

The screenshot shows a specific question on the AskWoG website: "Gameduino with MicroSD card breakout". The question text asks for help with a MicroSD breakout board and pin connections. Below the question, there are two answers. The first answer is from a user named "gameduino" and provides a detailed explanation of the pin connections and the importance of the SD card's write speed. The second answer is from a user named "gameduino" and provides a link to a video tutorial. The question has 13 votes and 2 answers. The answers have 1 vote and 0 votes respectively. The question was asked on Oct 18 '12 and the last update was on Dec 22 '12.

General Idea: Use a Question/Answer System to make the information retrieval process public. Use these public questions and answers to increase the user interaction on the website.

# Beginning

What do we have at the beginning?  
Website, Users and Questions

# Rating

- ▶ click
- ▶ active time on page
- ▶ individual click actions

# Rating

- ▶ click
- ▶ active time on page
- ▶ individual click actions

What do we have now?

Questions, users and the information, which user likes which question

# Tagging

Find words that describe the question.



# Tagging

Find words that describe the question.

- ▶ STW Thesaurus for Economics

# Tagging

Find words that describe the question.

- ▶ STW Thesaurus for Economics
- ▶ Stemming: Compute the stem of a word.

# Tagging

Find words that describe the question.

- ▶ STW Thesaurus for Economics
- ▶ Stemming: Compute the stem of a word.

Example rules

"ies"  $\rightarrow$  "y" und "s"  $\rightarrow$  ""

Examples

"libraries"  $\rightarrow$  "library" and "Wikis"  $\rightarrow$  "Wiki"

# Tagging

Find words that describe the question.

- ▶ STW Thesaurus for Economics
- ▶ Stemming: Compute the stem of a word.

Example rules

"ies" → "y" und "s" → ""

Examples

"libraries" → "library" and "Wikis" → "Wiki"

- ▶ Computing Levenshtein distance: Calculate the distance between two words.

Example:

Libraries, Library

Distance: 3

# Tagging

Find words that describe the question.

- ▶ STW Thesaurus for Economics
- ▶ Stemming: Compute the stem of a word.

Example rules

"ies"  $\rightarrow$  "y" und "s"  $\rightarrow$  ""

Examples

"libraries"  $\rightarrow$  "library" and "Wikis"  $\rightarrow$  "Wiki"

- ▶ Computing Levenshtein distance: Calculate the distance between two words.

Example:

Libraries, Library

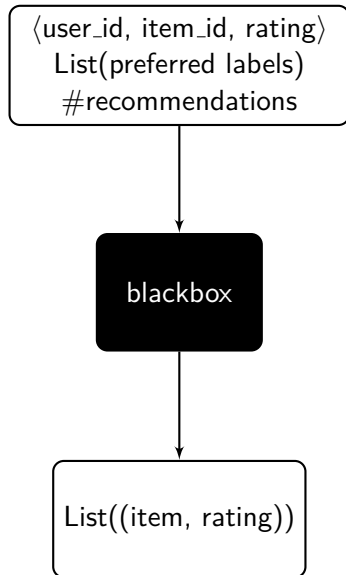
Distance: 3

What do we have now?

Information about which user likes which question and one or more preferred label(s) that describe(s) the subject of the question

# Input/Output

- ▶ Item: id, name, one or more preferred label(s)
- ▶ User: id, name



# Item-Based Algorithm

	Item1	Item2	Item3	Item4	Item5
User1	5	3	4	4	?
User2	3	1	2	3	3
User3	4	3	4	3	5
User4	3	3	1	5	4
User5	1	5	5	2	1

# Item-Based Algorithm

	Item1	Item2	Item3	Item4	Item5
User1	5	3	4	4	?
User2	3	1	2	3	3
User3	4	3	4	3	5
User4	3	3	1	5	4
User5	1	5	5	2	1

## Cosinus Similarity

$$\text{sim}(\vec{a}, \vec{b}) = \frac{\vec{a} \cdot \vec{b}}{|\vec{a}| \cdot |\vec{b}|}$$

## Example

$$\text{sim}(\text{Item5}, \text{Item1}) = \frac{3 \cdot 3 + 5 \cdot 4 + 4 \cdot 3 + 1 \cdot 1}{\sqrt{3^2 + 5^2 + 4^2 + 1^2} \cdot \sqrt{3^2 + 4^2 + 3^2 + 1^2}} = 0.99$$



# Predictions

## Prediction

User  $u$ , Item  $p$ , Rating  $r_{u,p}$

$$pred(u, p) = \frac{\sum_{i \in ratedItems(u)} sim(i, p) \cdot r_{u,i}}{\sum_{i \in ratedItems(u)} sim(i, p)}$$

# Predictions

## Prediction

User  $u$ , Item  $p$ , Rating  $r_{u,p}$

$$\text{pred}(u, p) = \frac{\sum_{i \in \text{ratedItems}(u)} \text{sim}(i, p) \cdot r_{u,i}}{\sum_{i \in \text{ratedItems}(u)} \text{sim}(i, p)}$$

## Example

$$\text{sim}(\text{Item5}, \text{Item1}) = 0.99$$

$$\text{sim}(\text{Item5}, \text{Item2}) = 0.74$$

$$\text{sim}(\text{Item5}, \text{Item3}) = 0.72$$

$$\text{sim}(\text{Item5}, \text{Item4}) = 0.94$$

$$\text{pred}(\text{User1}, I5) = \frac{0.99 \cdot 5 + 0.74 \cdot 3 + 0.72 \cdot 4 + 0.94 \cdot 4}{0.99 + 0.74 + 0.72 + 0.94} = 4.07$$

## What do we have now?

Matrix with less empty fields than before.

	Item1	Item2	Item3	Item4	Item5
User1	5	3	4	4	4
User2	3	1	2	3	3
User3	4	3	4	3	5
User4	3	3	1	5	4
User5	1	5	5	2	1

# Singular Value Decomposition

$$M = \begin{pmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{31} & m_{32} & m_{33} \end{pmatrix}$$

Create a SVD with the matrix  $M = U \cdot \Sigma \cdot V^t$

# Singular Value Decomposition

$$M = \begin{pmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{31} & m_{32} & m_{33} \end{pmatrix}$$

Create a SVD with the matrix  $M = U \cdot \Sigma \cdot V^t$

$$\begin{pmatrix} u_{11} & u_{12} & u_{13} \\ u_{21} & u_{22} & u_{23} \\ u_{31} & u_{32} & u_{33} \end{pmatrix}$$

corresponds to the row  
vectors of matrix  $m$

# Singular Value Decomposition

$$M = \begin{pmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{31} & m_{32} & m_{33} \end{pmatrix}$$

Create a SVD with the matrix  $M = U \cdot \Sigma \cdot V^t$

$$\begin{pmatrix} u_{11} & u_{12} & u_{13} \\ u_{21} & u_{22} & u_{23} \\ u_{31} & u_{32} & u_{33} \end{pmatrix} \quad \begin{pmatrix} \sigma_{11} & 0 & 0 \\ 0 & \sigma_{22} & 0 \\ 0 & 0 & \sigma_{33} \end{pmatrix}$$

corresponds to the row  
vectors of matrix  $u$

diagonal matrix  
with  $\sigma_{ii} > 0$  and  
 $\sigma_{ii} \geq \sigma_{i+1i+1}$

# Singular Value Decomposition

$$M = \begin{pmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{31} & m_{32} & m_{33} \end{pmatrix}$$

Create a SVD with the matrix  $M = U \cdot \Sigma \cdot V^t$

$$\begin{pmatrix} u_{11} & u_{12} & u_{13} \\ u_{21} & u_{22} & u_{23} \\ u_{31} & u_{32} & u_{33} \end{pmatrix}$$

corresponds to the row vectors of matrix  $m$

$$\begin{pmatrix} \sigma_{11} & 0 & 0 \\ 0 & \sigma_{22} & 0 \\ 0 & 0 & \sigma_{33} \end{pmatrix}$$

diagonal matrix  
with  $\sigma_{ii} > 0$  and  
 $\sigma_{ii} \geq \sigma_{i+1i+1}$

$$\begin{pmatrix} v_{11} & v_{12} & v_{13} \\ v_{21} & v_{22} & v_{23} \\ v_{31} & v_{32} & v_{33} \end{pmatrix}$$

corresponds to the column vectors of matrix  $m$

# Low Rank Approximation of M

$$\begin{pmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{31} & m_{32} & m_{33} \end{pmatrix} = \begin{pmatrix} u_{11} & u_{12} & u_{13} \\ u_{21} & u_{22} & u_{23} \\ u_{31} & u_{32} & u_{33} \end{pmatrix} \cdot \begin{pmatrix} \sigma_{11} & 0 & 0 \\ 0 & \sigma_{22} & 0 \\ 0 & 0 & \sigma_{33} \end{pmatrix} \cdot \begin{pmatrix} v_{11} & v_{21} & v_{31} \\ v_{12} & v_{22} & v_{32} \\ v_{13} & v_{23} & v_{33} \end{pmatrix}$$

- ▶ Derive from  $\Sigma$  the matrix  $\Sigma_k$  (with k new rank of M) formed by replacing  $\sigma_{ii}$  with  $i > k$  by zeros.
- ▶ Compute and output  $M_k = U \cdot \Sigma_k \cdot V^T$  as the rank-k approximation to M.



# Low Rank Approximation of M

$$M_2 = \begin{pmatrix} u_{11} & u_{12} & u_{13} \\ u_{21} & u_{22} & u_{23} \\ u_{31} & u_{32} & u_{33} \end{pmatrix} \cdot \begin{pmatrix} \sigma_{11} & 0 & 0 \\ 0 & \sigma_{22} & 0 \\ 0 & 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} v_{11} & v_{21} & v_{31} \\ v_{12} & v_{22} & v_{32} \\ v_{13} & v_{23} & v_{33} \end{pmatrix}$$

- ▶ Derive from  $\Sigma$  the matrix  $\Sigma_k$  (with k new rank of M) formed by replacing  $\sigma_{ii}$  with  $i > k$  by zeros.
- ▶ Compute and output  $M_k = U \cdot \Sigma_k \cdot V^T$  as the rank-k approximation to M.

# Low Rank Approximation of M

$$M_2 = \begin{pmatrix} u_{11} & u_{12} & u_{13} \\ u_{21} & u_{22} & u_{23} \\ u_{31} & u_{32} & u_{33} \end{pmatrix} \cdot \begin{pmatrix} \sigma_{11} & 0 & 0 \\ 0 & \sigma_{22} & 0 \\ 0 & 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} v_{11} & v_{21} & v_{31} \\ v_{12} & v_{22} & v_{32} \\ v_{13} & v_{23} & v_{33} \end{pmatrix}$$

$$M_2 = \begin{pmatrix} u_{11} \cdot \sigma_{11} & u_{12} \cdot \sigma_{22} & u_{13} \cdot 0 \\ u_{21} \cdot \sigma_{11} & u_{22} \cdot \sigma_{22} & u_{23} \cdot 0 \\ u_{31} \cdot \sigma_{11} & u_{32} \cdot \sigma_{22} & u_{33} \cdot 0 \end{pmatrix} \cdot \begin{pmatrix} v_{11} & v_{21} & v_{31} \\ v_{12} & v_{22} & v_{32} \\ v_{13} & v_{23} & v_{33} \end{pmatrix}$$

# Low Rank Approximation of M

$$M_2 = \begin{pmatrix} \textcolor{red}{u}_{11} & \textcolor{green}{u}_{12} & \textcolor{blue}{u}_{13} \\ u_{21} & u_{22} & u_{23} \\ u_{31} & u_{32} & u_{33} \end{pmatrix} \cdot \begin{pmatrix} \textcolor{red}{\sigma}_{11} & 0 & 0 \\ \textcolor{green}{0} & \sigma_{22} & 0 \\ \textcolor{blue}{0} & 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} v_{11} & v_{21} & v_{31} \\ v_{12} & v_{22} & v_{32} \\ v_{13} & v_{23} & v_{33} \end{pmatrix}$$

$$M_2 = \begin{pmatrix} u_{11} \cdot \sigma_{11} & u_{12} \cdot \sigma_{22} & u_{13} \cdot 0 \\ u_{21} \cdot \sigma_{11} & u_{22} \cdot \sigma_{22} & u_{23} \cdot 0 \\ u_{31} \cdot \sigma_{11} & u_{32} \cdot \sigma_{22} & u_{33} \cdot 0 \end{pmatrix} \cdot \begin{pmatrix} v_{11} & v_{21} & v_{31} \\ v_{12} & v_{22} & v_{32} \\ v_{13} & v_{23} & v_{33} \end{pmatrix}$$

# Example SVD

	Item1	Item2	Item3	Item4	Item5	
User1	5	3	4	4	4	=
User2	3	1	2	3	3	
User3	4	3	4	3	5	
User4	3	3	1	5	4	
User5	1	5	5	2	1	

# Example SVD

$$\begin{array}{ccccc}
 & \text{Item1} & \text{Item2} & \text{Item3} & \text{Item4} & \text{Item5} \\
 \text{User1} & 5 & 3 & 4 & 4 & 4 \\
 \text{User2} & 3 & 1 & 2 & 3 & 3 \\
 \text{User3} & 4 & 3 & 4 & 3 & 5 \\
 \text{User4} & 3 & 3 & 1 & 5 & 4 \\
 \text{User5} & 1 & 5 & 5 & 2 & 1
 \end{array} = \begin{pmatrix} 0.544178 & 0.0875457 & 0.303701 & -0.598262 & -0.496039 \\ 0.330582 & 0.270665 & 0.155794 & -0.281254 & 0.845033 \\ 0.517601 & 0.04377 & 0.429477 & 0.737011 & -0.0503875 \\ 0.438285 & 0.373564 & -0.800903 & 0.129864 & -0.100232 \\ 0.366854 & -0.881822 & -0.239996 & -0.0541261 & 0.165165 \end{pmatrix} \cdot \begin{pmatrix} 16.499 & 0 & 0 & 0 & 0 \\ 0 & 4.93905 & 0 & 0 & 0 \\ 0 & 0 & 2.58239 & 0 & 0 \\ 0 & 0 & 0 & 1.20841 & 0 \\ 0 & 0 & 0 & 0 & 0.511218 \end{pmatrix} \cdot \begin{pmatrix} 0.452438 & 0.336841 & 0.410894 & -0.456436 & -0.55197 \\ 0.403968 & -0.531238 & -0.483027 & 0.210155 & -0.526419 \\ 0.43523 & -0.601119 & 0.481498 & -0.122706 & 0.449817 \\ 0.463447 & 0.282982 & -0.586236 & -0.401113 & 0.447854 \\ 0.477391 & 0.403612 & 0.149459 & 0.756011 & 0.12371 \end{pmatrix}^T$$

# Diagram of SVD

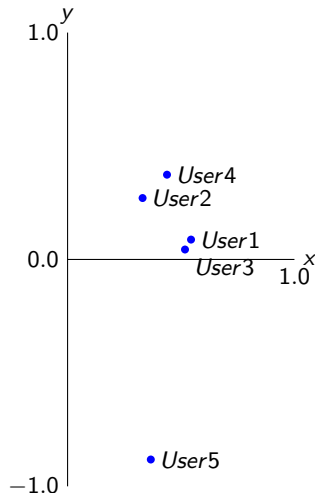
	Item1	Item2	Item3	Item4	Item5
User1	5	3	4	4	4
User2	3	1	2	3	3
User3	4	3	4	3	5
User4	3	3	1	5	4
User5	1	5	5	2	1

$$\begin{pmatrix} 0.544178 & 0.0875457 \\ 0.330582 & 0.270665 \\ 0.517601 & 0.04377 \\ 0.438285 & 0.373564 \\ 0.366854 & -0.881822 \end{pmatrix}$$

# Diagram of SVD

	Item1	Item2	Item3	Item4	Item5
User1	5	3	4	4	4
User2	3	1	2	3	3
User3	4	3	4	3	5
User4	3	3	1	5	4
User5	1	5	5	2	1

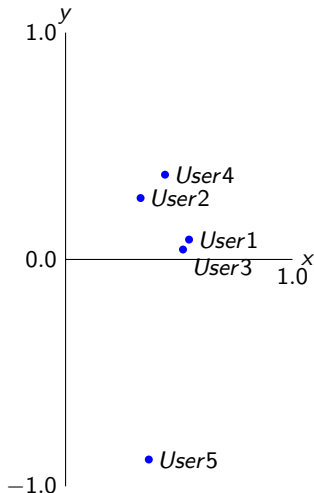
$$\begin{pmatrix} 0.544178 & 0.0875457 \\ 0.330582 & 0.270665 \\ 0.517601 & 0.04377 \\ 0.438285 & 0.373564 \\ 0.366854 & -0.881822 \end{pmatrix}$$



# Diagram of SVD

	Item1	Item2	Item3	Item4	Item5
User1	5	3	4	4	4
User2	3	1	2	3	3
User3	4	3	4	3	5
User4	3	3	1	5	4
User5	1	5	5	2	1

$$\begin{pmatrix} 0.544178 & 0.0875457 \\ 0.330582 & 0.270665 \\ 0.517601 & 0.04377 \\ 0.438285 & 0.373564 \\ 0.366854 & -0.881822 \end{pmatrix}$$



## Recommendations for User

- ▶ Take the k most similar users SU
- ▶ Take the top k unknown items (with the right context) from each user  $\in$  SU
- ▶ Use cosinus similarity for rating weighting



# Is this a good approach?

- ▶ We could have calculated the SVD directly out of the original matrix. Would we get a similar result?

# Is this a good approach?

- ▶ We could have calculated the SVD directly out of the original matrix. Would we get a similar result?
- ▶ Is it better if we use the  $V$  matrix from the SVD to calculate the item similarity and update the SVD afterwards?

# Is this a good approach?

- ▶ We could have calculated the SVD directly out of the original matrix. Would we get a similar result?
- ▶ Is it better if we use the  $V$  matrix from the SVD to calculate the item similarity and update the SVD afterwards?
- ▶ Is it better if we just recommend similar items from the items the user already likes?

# Is this a good approach?

- ▶ We could have calculated the SVD directly out of the original matrix. Would we get a similar result?
- ▶ Is it better if we use the  $V$  matrix from the SVD to calculate the item similarity and update the SVD afterwards?
- ▶ Is it better if we just recommend similar items from the items the user already likes?
- ▶ Should we use cosinus similarity between the users in the first place.

# Evaluation

The test data: I use real data from the zbw econdesk. However, we do not have real user ratings for this data.

- ▶ Every data gets a random quality value from 1 (rather negative rating) to 5 (rather positive rating).
- ▶ Furthermore, I generate 1000 test users. These users will have a rating preferation. So a user might be a person that rates an item more positive or more negative.
- ▶ I will try to evaluate the algorithms with this test data. I might use a movie db as well for this.

# Software Architecture

Scala: Finagle twitter framework

Every part of the software is a service.

The tagger, the rating algorithm and the recommendation algorithms can be used as an individual software.

# Timetable

Start: 18.12.2012

End: 18.06.2013

1st month(18.01): Theory

2nd month(18.02): Theory + Technology

3rd month(18.03): Implementation

4th month(18.04): Implementation + First writings

5th month(18.05): Final thesis

Personal Goal

Technology implemented: 31.03.2013

Diplom Thesis ready: 01.05.2013