

# How to read a Paper

ISMLL

Dr. Josif Grabocka

## Outline



How to read a paper

Common paper structure

Finding additional material



### How to read a paper

- ► Like novel or newspaper stories, scientific articles needs to be read differently.
- Since they are not books designed for students sometimes they are not self contained and requires some research to be fully understood.
- Understand a paper for a researcher means to be able to implement the described algorithm.



### How to read a paper

- ► Skim
- ▶ Re-read
- ► Analyze
- Summarize

Hildesheim, April 2019 3 / 30



#### Skim

First get the "Big picture" by reading the title, abstract, and introduction carefully: this will tell you the major findings and why they matter.

- Quickly scan the article without taking notes: focus on headings and subheadings
- Note the publishing date and conference/journal
- Note terms and parts you don't understand. Only with the bigger picture you will understand how much it is necessary to investigate something.



#### Re-read

Read the article again, asking yourself questions such as:

- ▶ What problems is the study trying to solve?
- Are findings well supported by evidence?
- ► Is the study repeatable? (i.e. is the article self contained?)
- If you do not understand take some time to find a brief explanation of what you are not understanding (one-two sentences).
- ► Is the paper innovative?



### Interpret

- ► Examine graphs and tables carefully
- ► Try to interpret data first before looking at captions
- When reading the discussion and results look after key issues and new findings
- ► Make sure you have distinguished the main points. If not go over the text again.



#### Summarize

- ► Take notes and underline key points: it improves reading
- ▶ Decide what part of the paper needs to be expanded and how much.

Hildesheim, April 2019



#### Common paper structure

- ► Abstract
- ► Introduction
- ► State of the art
- ► Algorithms explanation
- Experiments
- ► Conclusions and future work
- References

Hildesheim, April 2019 8 / 30



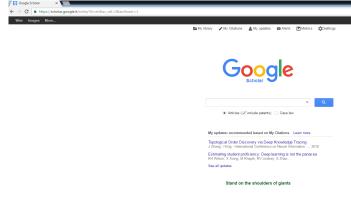
Let's take this paper as an example:

"Huang, S., Wang, S., Liu, T. Y., Ma, J., Chen, Z., and Veijalainen, J. (2015, August). Listwise Collaborative Filtering. In Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval (pp. 343-352). ACM."

Go to: www.scholar.google.com/

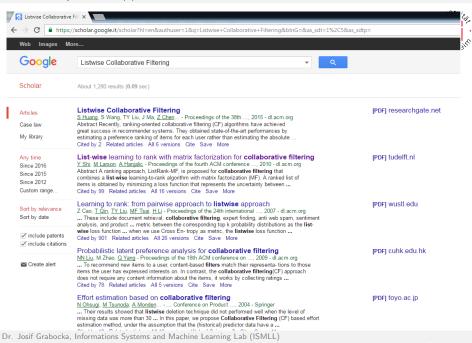
Hildesheim, April 2019 9 / 30





About Google Scholar Privacy Terms Go to G

Hildesheim, April 2019



How to read a Paper Common paper structure

Hildesheim, April 2019 11 / 30



#### **Abstract**

- Brief introduction to the topic
- ▶ Brief introduction to paper achievements
- ► Brief summary of the experiments

Hildesheim, April 2019 12 / 30



#### Introduction

- ► Introduction to the topic
- ► Introduction of the main concepts
- ► Introduction of the main state of the art methods
- ► State of the art limitations
- Hypotheses
- **▶** Contributions



## The Importance of Hypotheses

- ▶ It is not enough to describe some new technique or system, some claim about it must be stated and evaluated
- In experimental research, hypotheses typically take one of these two forms:
  - ► Technique/system X automates task Y for the first time
  - ► Technique/system X automates task Y better, along some dimension, than each of its rivals
- ▶ In theoretical papers, the hypotheses are the statements of theorems and the supporting evidence is their proofs



## The Importance of Hypotheses

Technique/system X automates task Y better, along some dimension, than each of its rivals, where the dimensions are typically:

- ▶ **Behavior:** X has a higher success rate or produces better quality outputs than Y
- ► Coverage: X is applicable to a wider range of examples then Y
- **Efficiency:** X is faster or uses less space then Y
- **Dependability:** X is more reliable, safe or secure than its rivals
- **Maintainability:** X is easier to adapt and extend than its rivals
- **Usability:** Users find X easier to use than its rivals

Hildesheim, April 2019 15 / 30



## State of the art / Related work

- ▶ Is a broad and shallow account of the field, which helps to place the contribution of the paper in context
- ► What are the rival approaches?
- ▶ What are the drawbacks of each?
  - ▶ One sentence per method. Is it clear enough?
- ► How has the battle between different approaches progressed?
- ► What are the major outstanding problems?

Hildesheim, April 2019 16 / 30



## **Algorithm Explanation**

- First the authors introduce the algorithm from which they derived the new algorithm
- ► Then, the new algorithm is explained
- ► Contains:
  - ► Formulas
  - Pseudo code



#### Algorithm 1: The ListCF Algorithm

```
Input: An item set I, a user set U, and a rating matrix
                 R \in \mathbb{R}^{M \times N}. A set of rated items I_u \subseteq I by
                each user u \in U. The maximal number of
                 iterations maxIteration and error threshold \epsilon.
     Output: A ranking \hat{\tau}_u of items for each user u \in U.
  1 for u \in U do
      for v \in U and u \neq v do
               P_u, P_v \leftarrow \text{TopKProDist}(I_u, I_v, R) /* Eq. 1 */
  \mathbf{3}
              sim(u, v) \leftarrow Similarity(P_u, P_v) /* Eq. 2 */
  5
          end
          N_u \leftarrow \text{SelectNeighbors}(\{sim(u, v)\}_{v \in U/u})
 7 end
 8 for u \in U do
          t = 1
10
          repeat
11
               \varepsilon = 0
12
               Initialize (\varphi_n^0)
              for g \in G_{t}^{T_u} do
13
                   \phi_{u,g}^t \leftarrow \text{Update}(N_u, sim, R) /* Eq.8 */
\varepsilon + = \sqrt{\sum (\varphi_{u,g}^t - \varphi_{u,g}^{t-1})^2}
14
15
16
               end
17
               t \leftarrow t + 1
          until t > maxIteration \text{ or } \varepsilon < \epsilon:
18
          for t \in T_u do
19
               P(t) \leftarrow \text{Aggregation}(\{\varphi_{u,g}\}_{g \in \mathcal{G}_{b}^{T_{u}}})
20
21
          end
          \hat{\tau}_u \leftarrow \text{Ordering}(\{P(t)\}_{t \in T_u})
23 end
```

Hildesheim, April 2019 18 / 30



#### **Experiments**

- ▶ Dataset Explanation
  - ► What are the available information?
  - ▶ What are the available statistics? E.g. number of users, items, sparsity etc.
- ► Evaluation protocol
  - ► How is the error of the algorithm computed?
  - Are there any other quantitative success measures?
- Experiments
  - Are the results statistically significant?

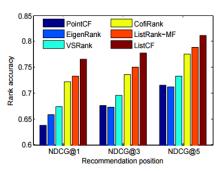
Hildesheim, April 2019 19 / 30



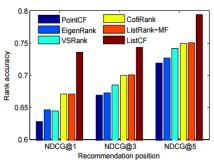
Table 2: Statistics on the three datasets

Table 2: Statistics on the three datasets.				
	Movielens-1M	EachMovie	Netflix	
#users	6,040	36,656	429,584	
#items	3,952	1,623	17,770	
#ratings	1,000,209	2,580,222	99,884,940	
#ratings/user	165.6	70.4	232.5	
#ratings/item	253.1	1589.8	5621.0	
sparsity	93.7%	95.7%	98.7%	





(a) Movielens-1M



(b) EachMovie

Dr. Josif Grabocka, Informations Systems and Machine Learning Lab (ISMLL)

Hildesheim, April 2019 21 / 30



## **Conclusions**

- Repeats the contributions pointing out specifically how the paper addressed it
- ► Include future works



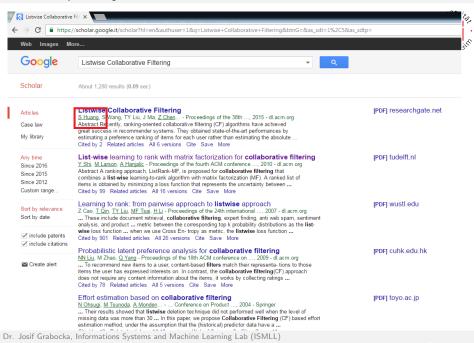
## Finding additional material

- ► If you don't understand something..
- ► This is not a book, it happens...
  - ► Try to pose yourself a specific questions
  - ► Look online



### Finding additional material

- ► A book explaining the algorithms
- ► A PhD thesis
- Tutorials
- Highly related state of the art papers



How to read a Paper Finding additional material

Hildesheim, April 2019 25 / 30





#### Sheng-Lung Huang National Taiwan University Biomedical imaging, fiber, laser, crystal Verified email at ntu.edu.tw

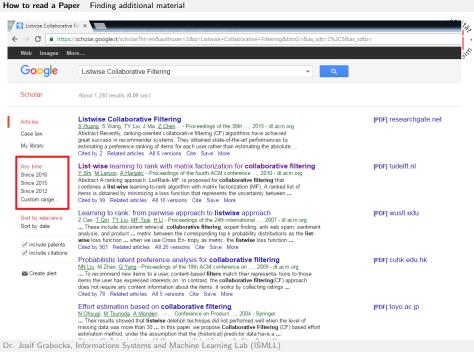
Title 1–20	Cited by	Year
Strain engineering and one-dimensional organization of metal-insulator domains in single-crystal vanadium dioxide beams J Cao, E Enkein, V Srimivsan, W Fan, S Huang, H Zheng, JWL Yim, Nature nanotechnology 4 (11), 732-737	266	2009
Contrasting patterns of retinoblastoma protein expression in mouse embryonic stem cells and embryonic fibroblasts. P Savatier, S Huang, L Szekely, KG Wiman, J Samarut Oncogene 9 (3), 809-818	248	1994
Flooding-induced membrane damage, lipid oxidation and activated oxygen generation in corn leaves B Yan, Q Dai, X Liu, S Huang, Z Wang Plant and soil 179 (2) 361-368	227	1996

# Get my own profile Citation indices All Since 2011 Citations 15313 10955 h-index 49 39 110-index 408 298

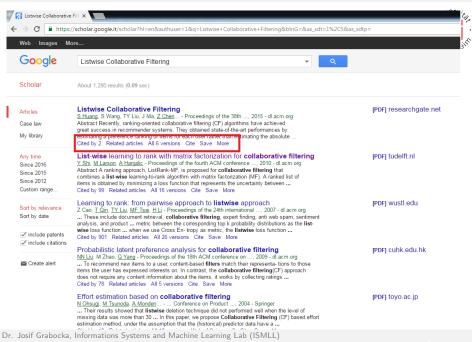
Follow •

 ${\sf Dr.\ Josif\ Grabocka,\ Informations\ Systems\ and\ Machine\ Learning\ Lab}\ ({\sf ISMLL})$ 

Hildesheim, April 2019 26 / 30

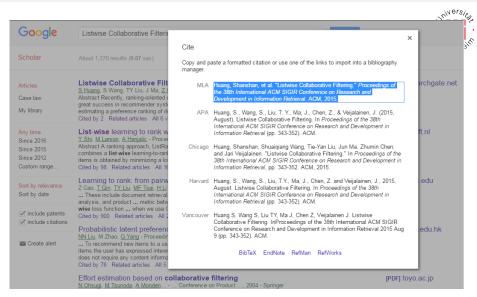


Hildesheim, April 2019 27 / 30



How to read a Paper Finding additional material

Hildesheim, April 2019 28 / 30



Dr. Josif Grabocka, Informations Systems and Machine Learning Lab (ISMLL)

Hildesheim, April 2019 29 / 30