

Literature Survey & Where to find ideas for a baseline

“Enterprise Data Science” (M-DS, WS19/20)

Mahan Ghashghaie, Rupali Sharma, Florian Becker

Outline

Intro

1. Baselines

- a. What is it? Why are they necessary?
- b. Baselines in Data Science

2. Papers

- a. What and where to find them?
- b. Example query
- c. How do we know they are any good?

3. Conferences

- a. How can i find them?
- b. Hirsch Index
- c. NIPS by Example

Baselines

Why do we need baselines?

- Each project is unique: setting, available data, problem to tackle, available features et cetera (e.g. your Master Thesis)
 - Lots of models and algorithms
 - Just apply some models and see what they'll return
 - 1) Time/resource consuming
 - 2) How to evaluate the results?
 - 3) How do we know, that they are actually adding value (e.g. improve our accuracy)?
- > **an initial value for benchmarking could help**

Baseline - a definition

- Techopedia defines a baseline as “ [...] the visible measure or progress and often marks milestones. In other words, a baseline serves as a crucial input for performing analysis to evaluate current performance against anticipated levels for the specific tasks in an established time-phase. [...]”

(<https://www.techopedia.com/definition/6148/baseline>)

How to translate this now to the DS domain?

Baselines in DS

- Main idea still holds: we need a reference for comparison, as simple as possible



- Known task
- Known data set
- Come up with better model

> compare performance model of on well-known data set to others



- New project
- Unknown data set
- Find a baseline algorithm/score or build one yourself

> baseline performance for improvement comparison

Baseline data sets

- Some examples for widely established data sets:
 - Recommender systems using the **Netflix Prize** dataset
 - Speech recognition using the **TIMIT** corpus
 - Natural language understanding using the **GLUE** Benchmark
 - Image Recognition with **MNIST**

(<https://towardsdatascience.com/beating-state-of-the-art-by-tuning-baselines-74ec6ad2cd59>)

- Reproducibility
- https://en.wikipedia.org/wiki/List_of_datasets_for_machine-learning_research (list overview)
- <http://mlr.cs.umass.edu/ml/> (UCI ML Repository)


Baseline models

- Some examples for baseline models:
 - Regression with **Linear Regression/GLM**
 - Classification with **Logistic Regression**
 - NLP using **Naive Bayes**
 - Neural Networks: **CNN**
- Reproducibility
- <https://paperswithcode.com/sota> (more on that later)
- https://en.wikipedia.org/wiki/Outline_of_machine_learning#Machine_learning_algorithms (list of algorithms for different purposes)


Baselines - how to start?

- Keep in mind to start with a stupid model first
- Some example baselines you could fit yourself:
 - Regression: Mean, Median
 - Classification: ZeroR (Mode)
 - “Human-based”
- Then check if you can beat your baseline result with a more complex, sophisticated model
- Iterate by optimizing your methods, data, parameters et cetera and see if you are able to improve
 - If not you may should overthink your problem/approach

Baselines





Alex Gude
@alex_gude





Here is a real use case from work for model improvement and the steps taken to get there:

- Baseline: 53%
- Logistic: 58%
- Deep learning: 61%
- ****Fixing your data: 77%****

Some good ol' fashion "understanding your data" is worth its weight in hyperparameter tuning!

 1,356 8:48 PM - Apr 24, 2019 

 384 people are talking about this 

(via <https://www.jeremyjordan.me/ml-projects-guide/>)

Baselines - what comes next?

- Your baseline is set up, you have a score for comparison and you know the problem
 - You are now ready to use a more advanced model
 - Use a well-studied baseline as approximation

> But where to find them? And how do we know, that there are reliable, accepted and proven?

Papers and Conferences as valuable sources!

Papers

What and where to find them?

- Tangible source: Scientific research papers
- Where to find them?
 - Semantic Scholar: <https://www.semanticscholar.org/>
 - dblp : <https://dblp.uni-trier.de/>
 - Google Scholar: <https://scholar.google.com/>
 - Papers with code: <https://paperswithcode.com/>

Example query:
recommender system netflix
prize

What is the Netflix Prize?

- Open competition (2006-2009)
 - Training (100,000,000 ratings)
 - Qualifying set (2,800,000 ratings)
 - Test (determine winners)
 - Quiz (calculate leaderboard scores)
- Best collaborative filtering algorithm
- Predict user ratings for films based on previous ratings without any other information about the users or films
- Improvement over Netflix's own recommendation system, Cinematch
- Prizes awarded for improvements in the RMSE
 - 1st (2006): 1.06%
 - Last (2009): 10.09%

Semantic Scholar



Semantic Scholar

All Fields

recommender system netflix prize

FAQ

Contact

Sign In

Create Free Account

The Netflix Prize

James Bennett, Stan Lanning • Published 2007

Netflix released a dataset containing 100 million anonymous movie ratings and challenged the data mining, machine learning and computer science communities to develop systems that could beat the accuracy of its recommendation system, Cinematch. We briefly describe the challenge itself, review related work and efforts, and summarize visible progress to date. Other potential uses of the data are outlined, including its application to the KDD Cup 2007.

Share This Paper



1,068 Citations

126 Highly Influenced Papers

573 Cite Background

435 Cite Methods

16 Cite Results



VIEW PDF

SAVE TO LIBRARY

CREATE ALERT

CITE

ABSTRACT

FIGURES

1,069 CITATIONS

2 REFERENCES

RELATED PAPERS

dblp



[home](#) | [browse](#) | [search](#) | [about](#)

recommender system netflix prize

[+] Search dblp ⓘ

[-] powered by CompleteSearch, courtesy of Hannah Bast, University of Freiburg

> Home

☰ ▼ Trier 1

[-] Publication search results ⬇

found 1 match

2009



Frank McSherry, Ilya Mironov:

Differentially Private Recommender Systems: Building Privacy into the Netflix Prize Contenders. KDD

2009: 627-636

[-] Refine list

refine by author

Frank McSherry (1)

Ilya Mironov (1)

refine by venue

KDD (1)

refine by type

Conference and Workshop Papers (1)

refine by year

2009 (1)

Google Scholar

Google Scholar

recommender system netflix prize



Articles

About 9.320 results (0,08 sec)

Any time

Since 2019

Since 2018

Since 2015

Custom range...

Sort by relevance

Sort by date

☒ include patents

☒ include citations

☒ Create alert

Differentially private **recommender systems**: Building privacy into the **netflix prize** contenders

[PDF] psu.edu

F McSherry, I Mironov - Proceedings of the 15th ACM SIGKDD ..., 2009 - dl.acm.org

We consider the problem of producing recommendations from collective user behavior while simultaneously providing guarantees of privacy for these users. Specifically, we consider the **Netflix Prize** data set, and its leading algorithms, adapted to the framework of differential ...

☆ 99 Cited by 605 Related articles All 7 versions

[PDF] The **netflix prize**

[PDF] uic.edu

J Bennett, S Lanning - Proceedings of KDD cup and workshop, 2007 - cs.uic.edu

... The Cinematch **recommendation system** automatically analyzes the accumulated movie ratings weekly using a variant of ... qualifying ratings were withheld and form the basis of the contest scoring **system** ... [2] Resnick, P. Varian, HR, **Recommender Systems**, Communications of ...

☆ 99 Cited by 1761 Related articles All 7 versions

Papers with code

[Browse State-of-the-Art](#)[Follow](#)[Discuss](#)[Trends](#)[About](#)[Log In/Register](#)

Search Results

[Trending](#)[Latest](#)[Greatest](#)[Subscribe](#)

On the Difficulty of Evaluating Baselines: A Study on Recommender Systems

4 May 2019 • 1 code implementation

Numerical evaluations with comparisons to baselines play a central role when judging research in recommender systems.



SOTA for Collaborative Filtering on MovieLens 10M

COLLABORATIVE FILTERING

★ 1,166

Paper

Code

How do we know they are any good?

Just because a modelling technique was proposed more recently doesn't mean it's necessarily going to outperform an older method (even if the results in the paper suggest that it can)

~ Rachael Tatman

- # citations
- Reference list
- Authors and papers with multiple references
- Overlaps between authors, journals, conferences, etc (good thing)
- Journals and conferences the papers were published in

Conferences

Conferences

- Have a focus on subareas in computer science
- How to find a suitable one
 - Online conference aggregators
 - https://scholar.google.com/citations?view_op=top_venues&hl=en&vq=eng
 - <https://www.scimagojr.com>
 - <http://www.guide2research.com>
 - <https://aminer.org/ranks/conf>
 - University curated commendation list
 - <http://www.core.edu.au/conference-portal>
 - Senior colleagues in the same field

Hirsch-Index

- Attempts to measure productivity and citation impact of the publications of an author
- Calculated by:
 - Sorting Publications from highest cited to lowest
 - Choose the number of citations such that the most publications fulfill:
 - $n_{\text{paper_cited}} \leq \text{citation_threshold}$
- Example:
 - <https://scholar.google.de/citations?user=am2ohp0AAAAJ&hl=de>

Example conference: Neural
Information Processing
Systems (NIPS)

NIPS

Conference Information

Submission Deadline	Thursday 23 May 2019	<div>Proceedings indexed by :</div> <div> Neural Information Processing Systems Foundation</div>
Conference Dates	Dec 10, 2019 - Dec 12, 2019	
Conference Address	Vancouver, Canada 🇨🇦	
Conference & Submission Link	https://nips.cc	

Conference Organizers : (*Deadline extended ? Click here to edit* ✎)

Conference Ranking & Metrics (This is a TOP Conference)

Guide2Research Overall Ranking:	2	<div>Google Scholar H5-index:</div> <div>169</div>
<i>Category Rankings</i>		
<i>Machine Learning & Arti. Intelligence</i>	2	
<i>Signal Processing</i>	2	
<i>Computational Theory and Mathematics</i>	1	
Number of Editions:	33 (since 1987)	
CORE 2018 Rating:	A*	
DBPL Proceedings		

Source: <http://www.guide2research.com/conference/nips-2019>

NIPS

← Neural Information Processing Systems (NIPS)



h5-index:169 h5-median:334

#1 Artificial Intelligence

#8 Engineering & Computer Science

Title / Author	Cited by	Year
Faster R-CNN: towards real-time object detection with region proposal networks S Ren, K He, R Girshick, J Sun Proceedings of the 28th International Conference on Neural Information ...	<u>10517</u>	2015
Generative adversarial nets I J Goodfellow, J Pouget-Abadie, M Mirza, B Xu, D Warde-Farley, S Ozair, ... Proceedings of the 27th International Conference on Neural Information ...	<u>10175</u>	2014
Sequence to sequence learning with neural networks I Sutskever, O Vinyals, QV Le Proceedings of the 27th International Conference on Neural Information ...	<u>6991</u>	2014
Two-stream convolutional networks for action recognition in videos K Simonyan, A Zisserman Proceedings of the 27th International Conference on Neural Information ...	<u>2735</u>	2014
How transferable are features in deep neural networks? J Yosinski, J Clune, Y Bengio, H Lipson Proceedings of the 27th International Conference on Neural Information ...	<u>2732</u>	2014
Attention is all you need A Vaswani, N Shazeer, N Parmar, J Uszkoreit, L Jones, A N Gomez, ...	2427	2017

Source: https://scholar.google.com/citations?hl=en&vq=eng&view_op=list_hcore&venue=eqYFflc_uhEJ.2019

NIPS

Publications

Top Cited

Top Downloaded

Authors

Affiliations

Browse by Citation

1

Generative Adversarial Nets.

Cited by 10618

Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron C. Courville, Yoshua Bengio

NIPS (2014)

2

Sequence to Sequence Learning with Neural Networks.

Cited by 7195

Ilya Sutskever, Oriol Vinyals, Quoc V. Le

NIPS (2014)

3

Two-Stream Convolutional Networks for Action Recognition in Videos.

Cited by 2802

Karen Simonyan, Andrew Zisserman

NIPS (2014)

4

How transferable are features in deep neural networks?

Cited by 2802

Jason Yosinski, Jeff Clune, Yoshua Bengio, Hod Lipson

NIPS (2014)

Source: <https://aminerg.org/conference/53a728e520f7420be8bbc4bb>

NIPS



Generative Adversarial Nets

Part of: [Advances in Neural Information Processing Systems 27 \(NIPS 2014\)](#)

[\[PDF\]](#) [\[BibTeX\]](#) [\[Reviews\]](#)

Authors

- [Ian Goodfellow](#)
- [Jean Pouget-Abadie](#)
- [Mehdi Mirza](#)
- [Bing Xu](#)
- [David Warde-Farley](#)
- [Sherjil Ozair](#)
- [Aaron Courville](#)
- [Yoshua Bengio](#)

Conference Event Type: Poster

Abstract

We propose a new framework for estimating generative models via adversarial nets, in which we simultaneously train two models: a generative model G that captures the data distribution, and a discriminative model D that estimates the probability that a sample came from the training data rather than G . The training procedure for G is to maximize the probability of D making a mistake. This framework corresponds to a minimax two-player

How to choose a conference

- use h-index to get an impression
- search for top conferences in your field of interest
- check citations and authors of paper
 - check out comments from reviewers of the paper

Summary

- baselines help in understanding the problem and the data
- compare your developed model (e.g. to established ones)
- a first (self-developed) baseline should be as simple as possible, reproducible
- Start with the well defined platforms (e.g. Semantic scholar) for finding relevant papers
- Look at the 'Related Papers' section in the *aforementioned* platforms

Thank you!

“All models are wrong but some are useful.” (George Box)

literature/sources

- Steffen Rendle, Li Zhang, and Yehuda Koren. *On the difficulty of evaluating baselines: A study on recommender systems*. (<https://arxiv.org/abs/1905.01395v1>)
- Sida Wang, Christopher D. Manning. *Baselines and Bigrams: Simple, Good Sentiment and Topic Classification* (https://nlp.stanford.edu/pubs/sidaw12_simple_sentiment.pdf)
- <https://machinelearningmastery.com/how-to-get-baseline-results-and-why-they-matter/>
- <https://machinelearningmastery.com/why-you-should-be-spot-checking-algorithms-on-your-machine-learning-problems/>
- <https://machinelearningmastery.com/how-to-know-if-your-machine-learning-model-has-good-performance/>
- <https://medium.com/upwork-datascience/managing-data-science-with-v2mom-scrum-and-cross-functional-squads-655b08cf828b>
- <https://towardsdatascience.com/first-create-a-common-sense-baseline-e66dbf8a8a47>
- <https://blog.insightdatascience.com/always-start-with-a-stupid-model-no-exceptions-3a22314b9aaa>
- <https://towardsdatascience.com/beating-state-of-the-art-by-tuning-baselines-74ec6ad2cd59>
- << <https://developers.google.com/machine-learning/guides/rules-of-ml> >>
- << <http://www.jeremyjordan.me/ml-projects-guide/> >>
- <https://medium.com/@i.m.vivek/machine-learning-why-baseline-is-important-cc63c857a56d>
- <https://towardsdatascience.com/machine-learning-general-process-8f1b510bd8af>
- <https://developers.google.com/machine-learning/glossary>
- J.E. Hirsch. *An index to quantify an individual's scientific research output* (https://arxiv.org/PS_cache/physics/pdf/0508/0508025v5.pdf)
- https://en.wikipedia.org/wiki/Netflix_Prize