


Improvements on Goodreads Recommendations

- Using NLP and Unsupervised Learning

Katharina Huang

2017/06/02

ads Home My Books Browse ▾ Community ▾ Search books



Six of Crows (Six of Crows #1)
by Leigh Bardugo (Goodreads Author)

★★★★★ 4.46 · Rating Details · 88,552 Ratings · 15,529 Reviews

Criminal prodigy Kaz Brekker has been offered wealth beyond his wildest dreams. But to claim it, he'll have to pull off a seemingly impossible heist:

Break into the notorious Ice Court
(a military stronghold that has never been breached)

Retrieve a hostage
(who could unleash magical havoc on the world)

Survive long enough to collect his reward
(and spend it)

Kaz needs a crew desp ...more

Hardcover, 462 pages
Published September 29th 2015 by Henry Holt and Company

Original Title Six of Crows
ISBN 1627792120 (ISBN13: 9781627792127)
Edition Language English
LRI <http://www.grishatrilogy.com/six-of-crows/>

Currently Rea... ▾

Rate this book

★★★★★

Preview Listen

N of books	937 (Training 788)
Genre	Alternate history
Stemming	Lancaster Stemmer
Token count	TF-IDF
ngram	1, 2
Exclusion	English stop words
(Min, Max) doc freq	(5, 50%)
Normalize	L2
Platform	MacBook Pro

COMMUNITY REVIEWS

filter | sort: default (?) | Rating Details

(showing 1-30)



Emily May rated it ★★★★★ · review of another edition
Shelves: young-adult, fantasy, arc, 2015

Sep 20, 2014

Trickery is not my native tongue, but I may learn to speak it yet.

I was scared to start Six of Crows. The hype surrounding it has been HUGE, the blurb sounded intriguing, I've been eagerly anticipating it, and I liked but didn't love Shadow and Bone. Honestly, though? I needn't have worried. **This book is fantastic.**

Shadow and Bone has been called fantasy-lite, but there is absolutely nothing about Six of Crows that warrants such a label. It's a sweeping epic tale with six main characters, five [...more](#)

1499 likes · Like · 125 comments · see review



Khanh (the meanie) rated it ★★★★★ · review of another edition

May 22, 2014

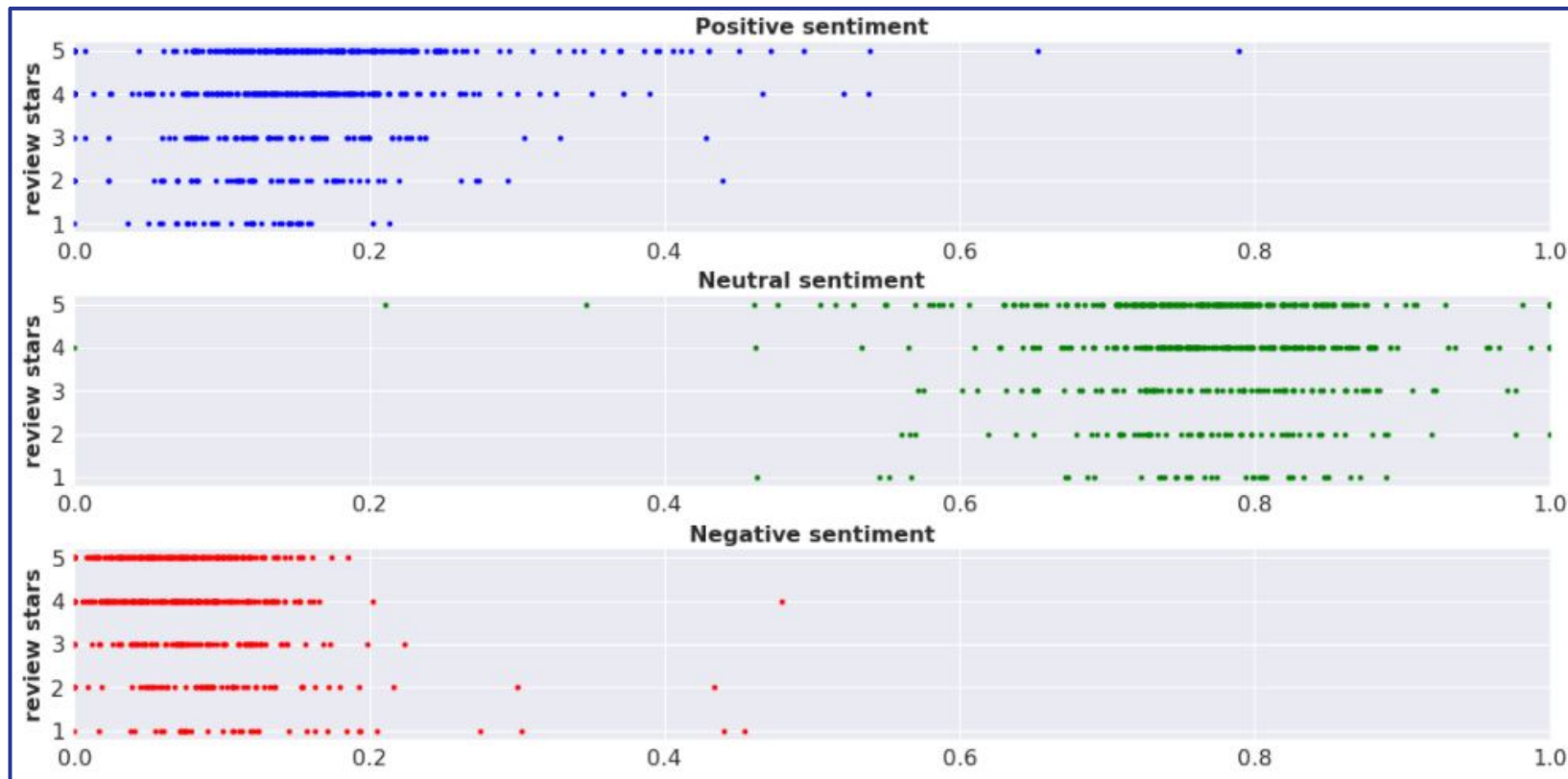
The heat of the incinerator wrapped around Inej like a living thing, a desert dragon in his den, hiding from the ice, waiting for her. She knew her body's limits, and she knew she had no more to give. She'd made a

N of reviews	21278 (Training 19216)
Genre	Alternate history
Stemming	Lancaster Stemmer
Token count	TF-IDF
ngram	1, 2
Exclusion	English stop words, non-English words
(Min, Max) doc freq	(10, 50%)
Normalize	L2
Platform	AWS EC2 c3.8xlarge

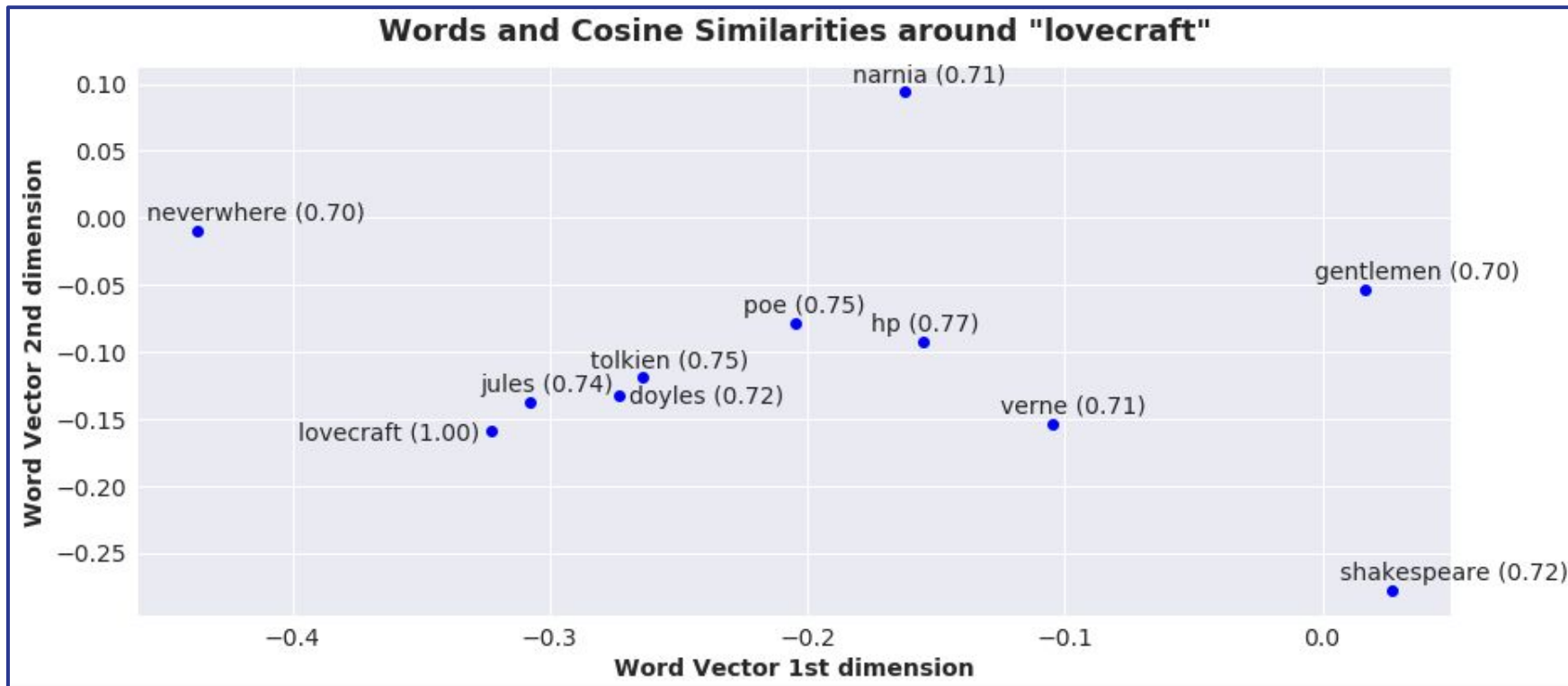
Recommender using reader star ratings



Reader review sentiment analysis

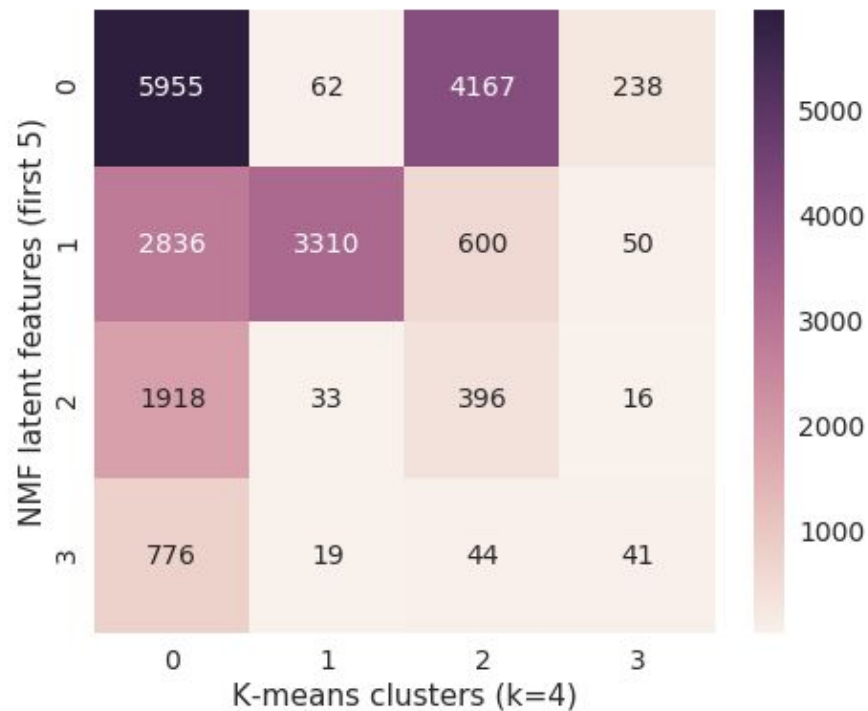
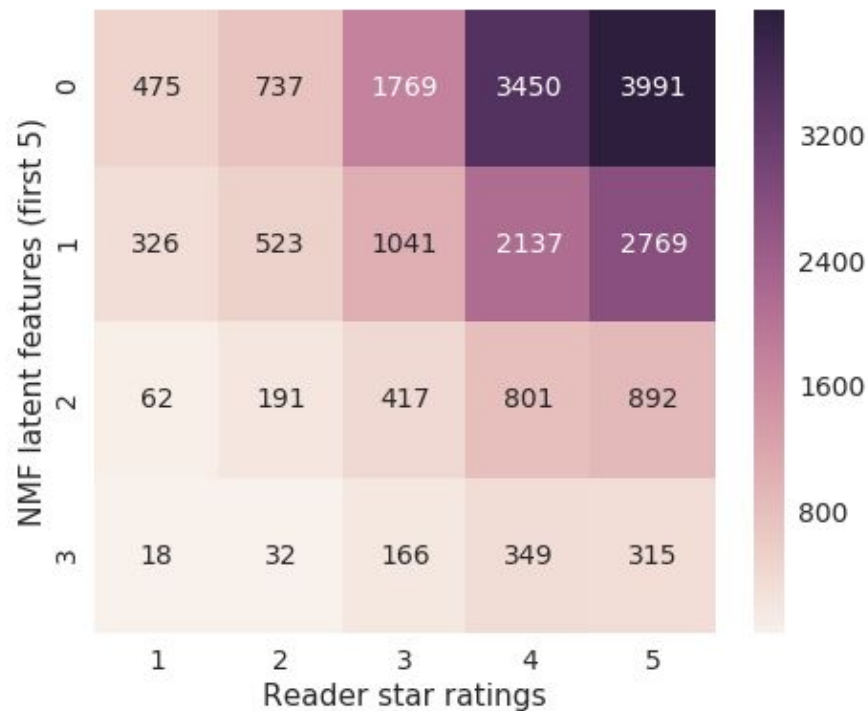


Content-based Filtering with Word2Vec



Improving genre clustering

Comparing different ways to categorize book reviews



Improving genre clustering

K-Means cluster	Top words and books	
0	hist , fun, rachel, harry , di, moor, laur, alex, thursday, jan	
	The Golem and the Jinni, The Giver (638 books)	
1	thi book, did, sery, read thi, lov , auth, felt, thi sery, enjoy , review	
	A Conspiracy of Kings, And I Darken, The Winner's Crime (18 books)	
2	lik, ev, real, just, mor, ther, becaus, story , tim, peopl	
	The Book Thief, The Handmaid's Tale, The Golden Compass (86 books)	
3	la, en, el, es, lo, se, las, par, en el, la la	
	NA (0 books)	

Improving genre clustering

K-Means cluster	Top words and books	Potential topic names
0	hist, fun, rachel, harry , di, moor, laur, alex, thursday, jan	Classic fiction for the general population
	The Golem and the Jinni, The Giver (638 books)	
1	thi book, did, sery, read thi, lov , auth, felt, thi sery, enjoy , review	Sequel fans
	A Conspiracy of Kings, And I Darken, The Winner's Crime (18 books)	
2	lik, ev, real, just, mor, ther, becaus, story , tim, peopl	Good storylines with TV/movie adaptation
	The Book Thief, The Handmaid's Tale, The Golden Compass (86 books)	
3	la, en, el, es, lo, se, las, par, en el, la la	Probably foreign language reviews
	NA (0 books)	

Future Work

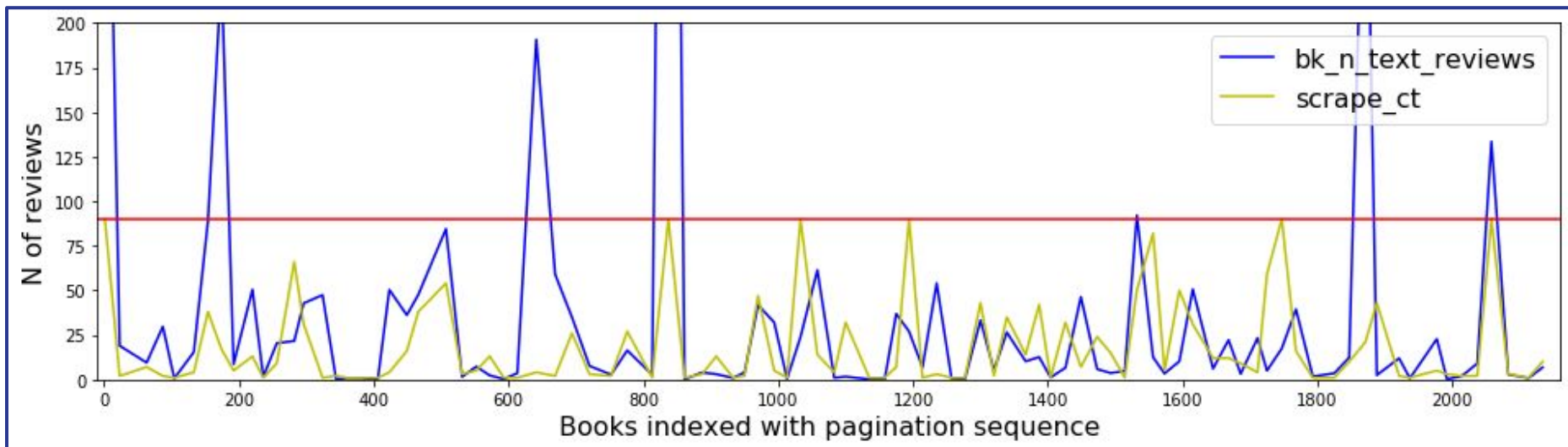
- Tensor factorization (e.g. higher-order SVD) for collaborative-filtering



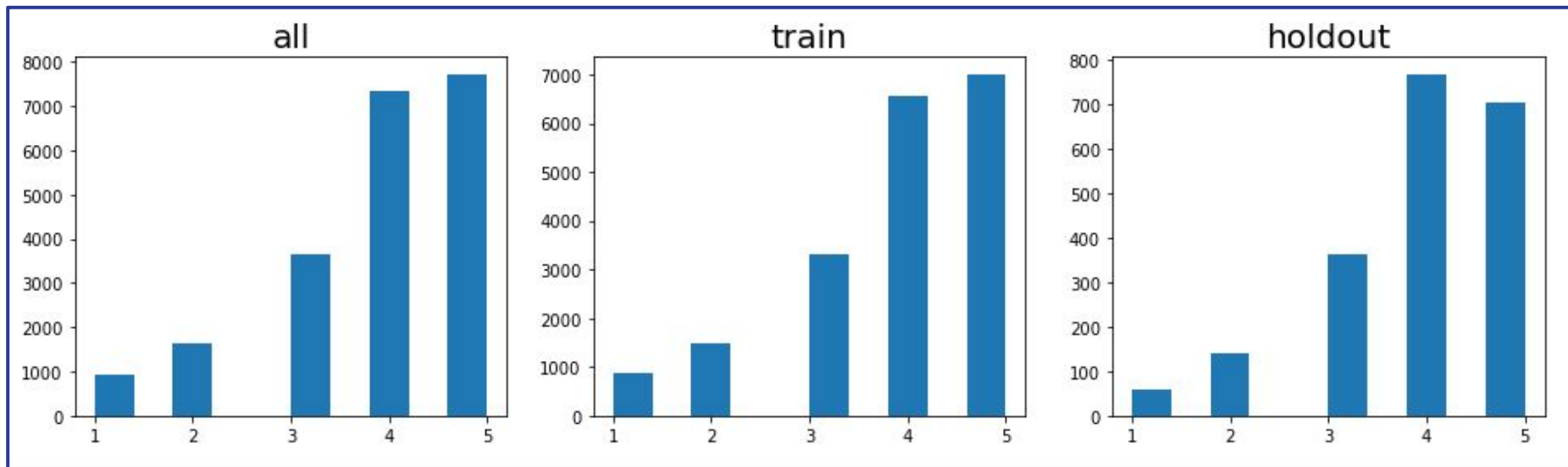
Q & A



Appendix. Scraped data over time



Appendix. Most reviewers scored high



Appendix. Reader review PCA latent features

