

Good Recommender and Topic Modeling Base on good Reads Data

Abstract:

In 2021 and 2022, we have seen a surge of interest in the written word with Publisher's Weekly reporting a 9% increase in the sale of *print* books Year over Year. Interestingly, recommender systems are still behind the times especially that of Good Reads, the largest social network site for book reviewing and sharing. This project endeavors to use that data to create a recommender system that allows for Readers to be shuffled to other books they may love and enjoy.

Design:

Once Data was Cleaned and Explored, used TIDF to create topic modeling from the free form descriptive text. This resulted in a quite a few different rounds of topic modeling (one particularly funny instance where I wound up with topics spanning many different languages before I subset on English only) in order to clean the data appropriately. Then, use cosine similarity to create a recommender system based on book index/ISBN to recommend 3 other books that readers may want to try

Data:

The Data for this project was obtained from the Good Reads API. As it is a particularly messy dataset, it required a bit of aggregating and cleaning in order to derive a clean dataset for NLP. Beginning with ~2MM books dwindles to about 100K in the dataset of consideration given the need for clear Description, book validity, and English Language

Topics:

Topic Examples below.

Topic Number	Top Terms (20)	Genre/Classification
Topic 4	WAR, WORLD, HISTORY, AMERICAN, CENTURY, FIRST, STATE, AMERICA, BATTLE, YEAR, POLITICAL, CITY, GREAT, II, CIVIL, ARMY, MILITARY, ONE, ACCOUNT, BRITISH	American/World History
Topic 8	LIFE, LOVE, WOMAN, MAN, FAMILY, HEART, YOUNG, WOULD, FATHER, COULD, MARRIAGE, MOTHER, CHILD, PASSION, NEVER, DAUGHTER, WIFE, HUSBAND, MEN, BEAUTIFUL	Romance/Family Drama

Topic 2	TEXT, WORK, NEW, STUDENT, STUDY, VOLUME, GUIDE, PROVIDES, CHAPTER, THEORY, INFORMATION, ALSO, RESEARCH, INTRODUCTION, WELL, LANGUAGE, INCLUDES, SOCIAL, DEVELOPMENT, OFFER	Textbook/Reference
Topic 9	BOOK, CHILD, SERIES, PAGE, COLOR, READ, READER, ILLUSTRATION, READING, PICTURE, QUALITY, FULL, PART, ORIGINAL, COVER, FIRST, ILLUSTRATED, WORLD, AUTHOR, CLASSIC	Children's Novels