

# Will You Watch What You Read?

Rachita Hajela  
Cs department, NYU,  
NY, USA  
rs1997@nyu.edu

Anto Loyola  
Cs department, NYU,  
NY, USA  
al4251@nyu.edu

Nikita Keswaney  
Cs department, NYU,  
NY, USA  
nk1664@nyu.edu

**Abstract**--The project attempts at finding out how well a movie will do if it has been adapted from a book of a particular genre i.e. what kind of works should movie makers adapt? Will a movie based on a romantic novel do better than one based on science fiction novel? Or movies based on series of books like Lord of the rings are the best way to top the blockbuster charts? The project also aims at finding the anomalies where the book did exceptionally well but the movie did not and vice versa based on the IMDB ratings for the movies and GoodReads rating for the books.

## I. MOTIVATION

Watching movies and reading books are the favorite pass times of today's world. What happens when both of these are combined? There must be a reason why movies adapted from novels are doing exceptionally well. Five of the top ten films on the American Film Institute's list of the 100 best movies of the last 100 years are adaptations of novels. Indeed we like to watch what we read but this is not true always. There have been many instances where the bubble of hype has been severely burst by the fans of the book and the movie has been criticized a lot.

This analytic aims to find out if a movie will do better if it has been adapted from a book of a particular genre by analyzing the past trends. So that the risk of the movie makers is lowered and they know what kind of works should be adapted. This project also finds the anomaly in the data so as to help movie makers understand what went wrong.

## II. INTRODUCTION

There are over 100 books being turned into movies in 2014 only. On the American Film Institute's list of the 100 best movies of the last 100 years, five of the top ten films are adaptations of novels (The Godfather, Gone With the Wind, Vertigo, Schindler's List, and The Wizard of Oz) while two others come from non-fiction (Raging Bull from the memoir of the same title and Lawrence of Arabia, which is based largely on T.E. Lawrence's own writing). IMDB's fan-voting yields similar results; of the top ten, six are straight adaptations and a seventh (The Godfather II) is derived from a novel. If you look at all-time movie lists, it seems that adaptations do unusually well. So what makes them top the charts? Is there a particular genre which garners more favors from the audiences? It makes a certain amount of sense that

anticipated movies are often adaptations; it's easier to get excited about a story you already know. So this isn't itself an indicator that adaptations makes particularly better movies just that we tend to look forward to adaptations a bit more. So can the risks for the movie makers be mitigated if they adapt books of a particular genre?

## III. RELATED WORK

It is estimated that a third of all films ever made have been adapted from novels, and, if you included other literary forms, such as drama or short stories, that estimate might well be 65 percent or more. Nearly all of the works of classic literature students study in high school have been adapted for film. A narrator mediates the meaning of what we read through his or her point of view. The major difference between film and books is that visual images stimulate our perceptions directly, while written words can do this indirectly. Film is a more direct sensory experience than reading the filmmaker is an independent artist, not a translator for an established author, but a new author in his own right.

Many fiction writers, at one point or another, consider adapting their own work into film. Before a novel can be adapted into screenplay form, all those elements that do not conform to the rules of screenwriting must be eliminated. Ten principles help achieve that. Commerciality is the major concern of film financiers. Movies must conform to a budget. Genre is critical. Movies have a prescribed length. Movies portray a condensed period of time. The hero of a screenplay must pursue a single, visible goal with a clearly implied endpoint. The conflicts a movie hero faces must also be visible. Screenplays may only reveal what the audience will hear and see on the screen. Movies follow a strict structure. When it comes to writing style, a screenwriter's goal must be to create a movie in the reader's mind that is as fast, easy and enjoyable to read as possible. Novels can involve a series of characters, can focus primarily on inner motivation and character arc, or can present heroes who meander through a whole series of events, desires and conflicts. But in a screenplay, the reader must know what specific finish line the hero is hoping to cross.

The craft of adapting a novel and creating a screenplay is one of the ultimate challenges for any writer. The first step is finding a book you truly love and then ask yourself “is this a movie?” The various steps can be summarized as follows:

1. Look before you leap - before you jump into the adaptation, secure the rights.
2. Function equals form - A book can contain three to four times more material than you’ll need for a 120-page screenplay. You need to write a great story, in spite of the fact that it might not be exactly like the book.
3. Rip and read it again – get extra copies of the book, sort the individual pages into plot groupings. For those books that jump in and out of different eras or time zones, physically reconstruct those pages into its own isolated story.
4. Reinvent and Rewrite – Improve by editing, placement and Interaction. You’ll need to have a beginning, middle and end to your screenplay, which may not mirror the book.

**Adapt and Sell** - When you are ready to sell your work, make the reader aware it is an adaptation, but don’t encourage them to read the original work before they read your screenplay. Give the work it’s best possible chance as being reborn as a movie.

Look to capture the essence and spirit of the story. Determine the through-line and major sub-plot of the story and viciously cut everything else. Do not use tons of voiceovers. People pay to watch motion. Characters in novel suffer from over-thinking. When essential plot information is presented only in a character's thought or in the character's internal world, one solution is to give this character a sounding board, another character, to which his thoughts can be voiced aloud. Either adapt an existing character from the novel or create a new one. Move away from direct adaptation toward, "story based upon". Use the brilliant background and characters created by the original author as a platform from which to launch a screen story.

At present, a variety of content in different genres is derived from original sources including not only comic books but novels, games, animations, and movies as well. This paper classifies and analyzes the comic book-based movies that have been produced in Korea until now in terms of success and failure cases. Some of the factors determining the success and failure of the comic book adaptation are: The comic book was cinematized based on an age-long original work, wide difference of viewpoint between the characters in the original work and in the film, Stories in the original work underwent unfaithful adaptation for the screen.

Selecting a popular cartoon to adapt would result into fans of the cartoon becoming interested from the start, but it comes at a risk of meeting the expectation of the audiences being formed even before making of the movie. One of the reasons more and more movies are now based on cartoons is the superior difference in the computer graphic technology giving

completeness and required level of dramatization compared to earlier times. But the work is compared with the original work. Copying the original is not the right path. It is important to make balance the quality of original and freshness of adapted contents.

There is always one imminent question – will the film live up to the book? It has always been difficult to trust Bollywood and often it’s not just the fans who are disappointed. The book is always better mindset persists. But now the Hollywood is changing, it continues to look into the books tied up with preexisting fan base. Screenwriters are expressing the desire to stay true to original work. It’s better if the screenwriters’ attitude resembled that of a fan, and they would want to preserve the heart if the story as much as possible. The difficulty of capturing that essence can be alleviated, though, when authors have a hand in the screenplay. That collaborative effort is becoming increasingly common. Endorsement by the author works well for the fans too. The process, of course, is no easy feat. The patience for the right screenplay stems from the obvious fear of disappointing fans, who don’t shy away from picking apart every detail the studio releases. Screenwriters are lucky if they get to work closely with authors, but some choose to do the heavy lifting themselves.

#### IV. DESIGN

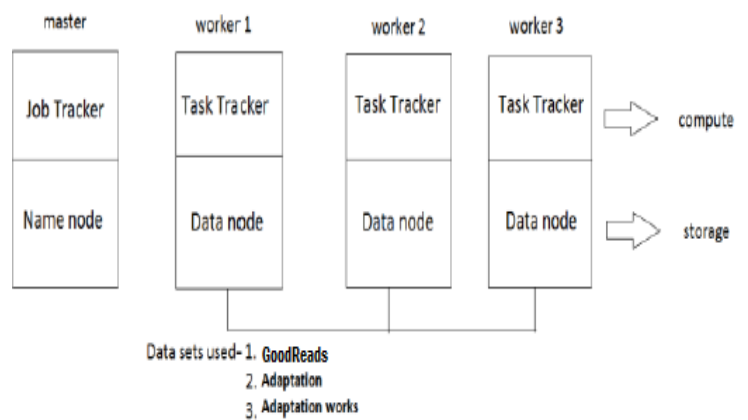


Fig 1: Architecture

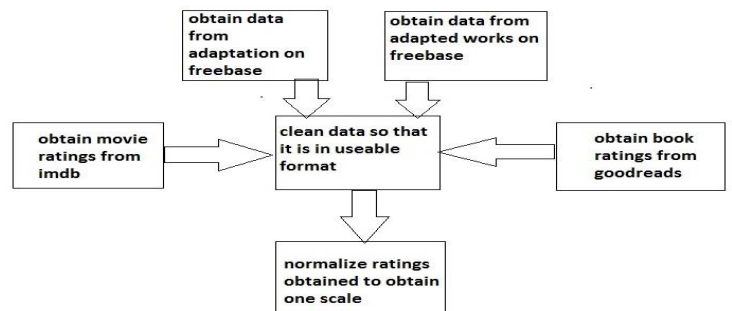


Fig 2 : Data cleaning and normalization

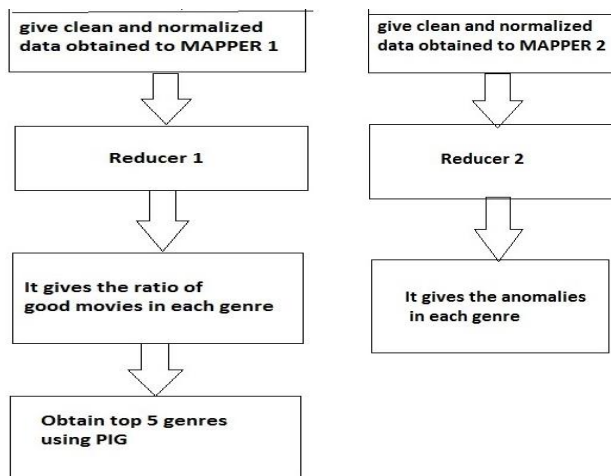


Fig 3: Mapper Reducer 1 and 2

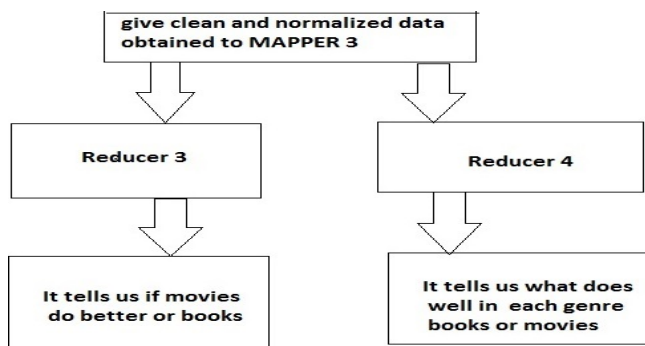


Fig 4: Mapper Reducer 3

The design is as follows:

Data sources:

The data from adaptations and adapted works is used as the major data source. We have mapped names of movies and books to imdb and goodreads to obtain ratings for movies and books.

The core technologies we have used are the freebase api, mapreduce, pig and java.

Freebase api returns json object with a result dataset containing movie information and book information and a cursor which points to consequent dataset which is needed as the size of the result is limited. As multiple http requests are created with updated cursor it can be used to travel through entire dataset. Which is then saved into a text file. Then run through this text file and extract the name and id for mapping as some input have these values interchanged. Then run an imdb search request with name or else id and store movie rating with each line, if rating does not exist throw off the line. Then run goodreads search request and obtain book rating and genre. The data obtained from adapted works and adaptations is combined by name. The ratings of the books and movies are on different scales. So ratings are normalized so that they are comparable on a scale of 10.

This data is input to the map reduce code. There are four jobs running here.

The first mapper manipulates the data and produces the genres of each movie and a count as key value pair if the rating is above the thresh-hold. The reducer then uses the count to find the number of successful moves in each genre. So this produces the genres and their ratio of successful movies. Output from this is given to pig which then provides us with the top five genres that are successful. Key in mapper reducer is genre here.

The second mapper checks if the difference between the movie rating and book rating is above a threshold. If yes, then it writes its corresponding genre and the corresponding better performer (movie or book) as key value pair. The reducer provides us with the genres and the number of anomalies they have. Key in mapper reducer is genre here.

The third mapper checks if for each data entry who has performed better. It provides the better performer (book or movie) and a count value as key-value pair. The reducer then tells us in case of anomalies who performs better, books or movies. Key here is the word book or movie here for the mapper reducer.

The last mapper checks if the difference between the movie rating and book rating is above a threshold. If yes, then it writes its corresponding genre and the corresponding better performer (movie or book) as key value pair. And reducer tells us in each genre who performs better, books or movies. Key in mapper reducer is genre here.

We have also used java to do the following: If you input a genre name it provides you with the top five movies in that genre and also their anomalies.

## V. RESULTS

Fetching data from Freebase was an obstacle. The freebase APIs are rigid. They don't support specific data extraction methods. The documentation is also inadequate. Data in freebase is populated by public so it contains a lot of junk data as well which was difficult to clean. IMDB and GoodReads has restriction on number on http requests one can process per unit time which led to slower extraction of data. The rating scales for IMDB and GoodReads were different. Normalization of the ratings to reach logical results was time consuming and difficult.

We have made use of cloudera to perform our analytic. The core technologies we have used are the freebase api, mapreduce and pig. The data that we required was from freebase (adaptations and adapted works), imdb and goodreads. We required multiple stages to combine the data from various sources, get it in the format needed for input to mapreduce and normalize the data so that the ratings are comparable.

The result obtained are as follows: The top five genres that did the best are: film noir had 100% success rate with all 28 movies with ratings above 7, adult movies had 100% success rate with all 5 adaptations having ratings above 7, short movies have a success rate of 99.5% of 2049 adaptations doing well. Horror movies have a ratio of 97.85% of 2097 movies doing well, and last documentary adaptations have a success ratio of 97.1% of 35 movies doing well.

The genres that had the most anomalies are drama, horror, short, comedy and romance. Drama had 123 anomalies, horror had 118, short had 102, comedy had 64 and romance had 35. With either the book doing well or the movie.

And in all genres in case of anomalies books have performed better than their adaptations. With 460 books doing better than movies and only 228 movie doing better than books.

## VI. FUTURE WORK

The analytic is based on the ratings of books and the ratings of the movies adapted from them. This can be extended by taking into account ratings from different sources. Currently we are only using one source for books and movies each. There are many other areas which make extensive use of ratings. Ratings can be used to find out most watched sports, most popular products etc. Our analytic can be used on anything that depends on customer review. With more time we could have expanded the analytic to work on any of these areas.

## VII. CONCLUSION

We have successfully found the top genre categories for which movie adaptations work well. We have found that even though certain genres like horror, short have done well, they still occur in the top genres with anomalies. So they belong to the risk zone. Top safe genres are categories like film-noir and adult movies which both have 100% success rates and all

movies from these genres have ratings above 7. The analytic provides movie makers enough information to capture the mood of the public.

## ACKNOWLEDGMENT

We would like to thank our Professor Suzanne McIntosh for her immense support in developing the project and helping us out with the various technologies that could be used in the project. We would like to thank freebase.com which provided us with the free data source. We are also grateful to the Cloudera for providing us the VM environment to run our analytic. We are thankful to all the readers and movie watchers who made the effort to rate the books and movies on Goodreads and Imdb.

## REFERENCES

- [1] Chan Ik (Park Chungwoon University), Ho Eun ( Lee Chungwoon University), An Investigation into the Factors Determining the Success of Comic Book-Based Movies, <http://www.aicit.org/AISS/ppl/AISS3692PPL.pdf>
- [2] Ashley Ross, How to turn a book into a movie that isn't terrible, <http://time.com/2815729/book-to-movie-adaptations-screenwriters>
- [3] Dwight C. Douglas How to adapt a novel into a screenplay, <http://www.videomaker.com/article/15701-how-to-adapt-a-novel-into-a-screenplay>
- [4] Michael Hauge, ADAPTATION: Mvies aren't novels, <http://www.storymastery.com/articles-by-industry/articles-for-screenwriters/movies-arent-novels/>
- [5] Lynne Pembroke and Jim Kalergis, From Novel to Screenplay: The Challenges of Adaptation, <http://www.writing-world.com/screen/adaptation.shtml>
- [6] Adaptation from novel to film, [http://d2buyft38glmwk.cloudfront.net/media/cms\\_page\\_media/11/FITC\\_Adaptation\\_1.pdf](http://d2buyft38glmwk.cloudfront.net/media/cms_page_media/11/FITC_Adaptation_1.pdf)