AI - 705 Recommendation Systems
Project Paper - Book Recommendation System using Transformers

Tejas Sharma
IMT2020548

Jainav Sanghvi
IMT2020098

Harsh Shah
IMT2020006

Chaitanya Manas Cheedella
IMT2020053

Abstract—**Whether one wishes to start reading books as a fresh reader, or wants to build upon their existing passion for reading books, we have found that the common bottleneck is the lack of personalized quality recommendations which capture the user's individual tastes. Extending upon this problem we found that similar analogous problems such as research paper recommender systems, movie recommender systems, etc. also generally rely on more collaborative filtering based approaches in existing solutions. Through this paper, we wish to not only make a content-focused approach to this problem, but also create an easily reusable and expandable pipeline for all problems of a similar category.**

Keywords - Content based recommendation, Transformers, Word2Vec, Web-Scraping

1. Introduction and Philosophy

The paper shall mainly focus on our findings with respect to building a content based recommendation systems for books, but also shall serve as a guide to building systems for analogous problems as well using the same pipeline, as we shall demonstrate by also building a small scale research paper recommendation system model as well.

To use our pipeline, it is essential to pick a problem where the item to be recommended has a deeper underlying theme which can be personalized to a user. For example, books have genres like action, comedy and adventure, but each individual book revolves around deeper concepts such as an impossible survival, rags to riches, evolution of a particular character arc, etc. It is easy to see that these themes purely depend on the actual content of the book. While it is practically impossible to analyze the whole book to capture this, the summary or description of the book can help us extract this information to an extent. This along with some additional prior knowledge such as the book's author, how liked a book is by the people, it's genres, etc. adds more *context* to our information about each book. This deeper knowledge then enables us to make a better judgement about a person's taste in books, and hence we can recommend new books with themes along the same lines of books they already like.

Essential parts of our pipeline include -

- Building a customized dataset : Using techniques such as web-scraping, we need to extract information relevant to our needs, i.e. the columns we need (eg. for books, we use description, author, genres, liked ratio, etc.)
- Picking the right model : We use *transformers*, a modern neural network based model to obtain vector embeddings for text based inputs. Even amongst them, picking the right transformer can help improve recommendations (eg. SBERT for movie/book based systems, and SciBERT for scientific information like research papers).
- Creating modified inputs to the transformers (**our Novelty**): We must maximize the information that we can provide to our model to obtain embeddings, by smartly modifying our input data to give more context about the information.

2. Dataset Creation

The first major task at hand while creating a pipeline which can be extended to other domains is to obtain a dataset which is similar to other tasks. As stated above, our approach requires us to find a dataset which has a description for each of the items.

To actually demonstrate the extendable nature of our pipeline, we built two datasets, one each for books and research papers. Here are some of the analogies we draw in the columns of our datasets for books and research papers (in that order) is as follows-

- Description/Abstract
- Book title/Paper title
- Book author/Paper author
- Book genres/Paper keywords
- Ratings/Number of citations

We decided to create the dataset using web-scraping. Given a url, we can use selenium and beautiful soup to scrape the data from the webpage corresponding to the url. For the book dataset we used the Goodreads platform, and for the research papers, we used the IEEE website.

Our books dataset consisted of 50,000 books across various genres and authors. Since it would be extremely time consuming to scrape so many books, we found an existing dataset which satisfies our requirement - The Comprehensive Literary Greats dataset.

Columns of the dataset : 'id', 'title', 'series', 'author', 'ratings', 'description', 'genres', 'numPages', 'firstPublished', 'awards', 'numRatings', 'ratings By Stars'
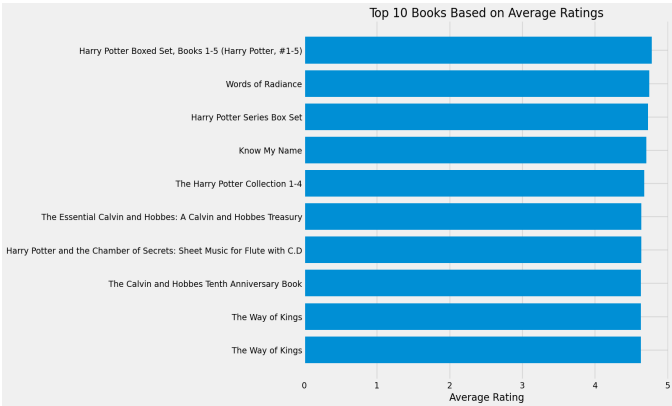
2.1. Dataset Analysis

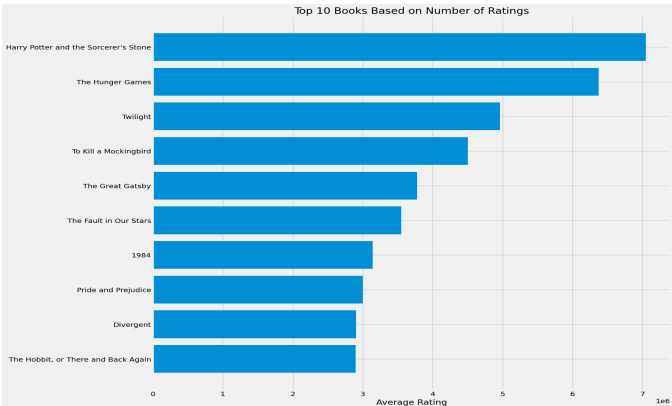Figure 1. Top 10 books based on average ratings

Figure 2. Top 10 books based on average ratings

For our research paper data set, we noticed a common trend in existing datasets, where they contained only information such as author of the papers, the awards it has received, the conference published in, etc. This information is clearly not indicative of the content of the respective papers, and hence we needed to scrape more appropriate data columns to build our dataset. We even went a step ahead to extract the conclusion of all the research papers in addition to the abstracts to obtain more information about the content each paper covers.
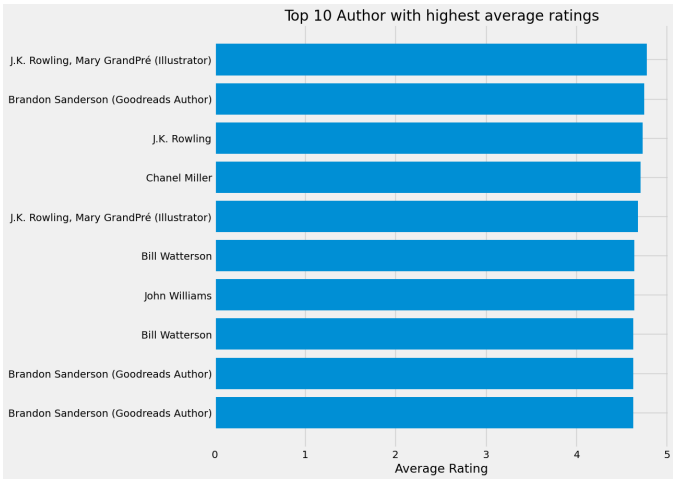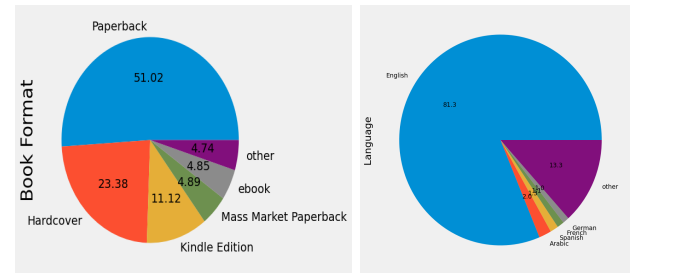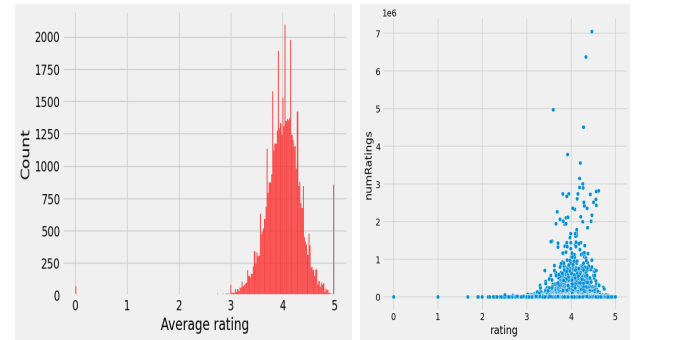

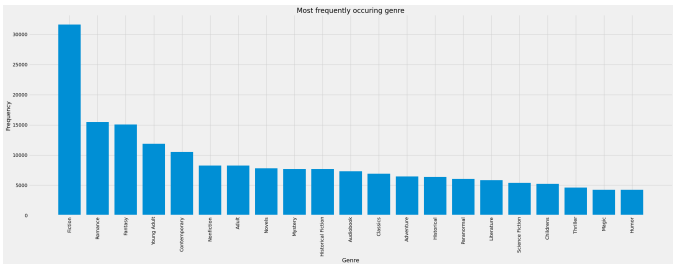Figure 3. Top 10 authors with highest average ratings

Figure 4. Most frequently occuring genre

Columns of the dataset : 'Title', 'Abstract', 'Conclusion', 'Authors', 'Paper Citations', 'Keywords'.

3. Algorithm

Our algorithm aims to provide personalized book recommendations to users based on their reading preferences and interests. To achieve this, we first collect basic information from the user, such as their favorite genres, and book length, to gain prior knowledge about their reading habits. This however can be easily extended collect other data such as preferred authors etc. to extract as much prior knowledge as we can from the user. Using this information, we select the highest-rated books in each genre, sort them by ratings, and suggest the top five books to the user.

Our algorithm however, mainly aims at recommending books to the user where the themes match the user's pre-existing taste in books. We aim to go beyond author and genre based recommendations by using Deep Learning and NLP techniques to analyze the underlying themes in books, mainly using Transformers.

Algorithm :

1) Create the input to be given to the transformer for each book.
2) Pass each book's data through the transformer to obtain the embeddings of all the books of the dataset.
3) Collect user information regarding the books they have read and liked in the past. It is likely that these books do not exist currently in our dataset, and hence we scrape the data corresponding to them from the Goodreads platform.
4) Repeat 1 and 2 to then obtain the embeddings for theese books. Append this to your dataset for future use.
5) Using cosine similarity, determine which books closely capture the themes of the user's books and recommend the top results.

We used multiple techniques to eventually build up to the best method to obtain embeddings from our input data. Below is our sequential experimentation. To start off, the input sentence we gave to the embedding model was the concatenated string of description, author and genre. As a testing metric, we use the book 'The Martian' as our user input.

3.1. Word2Vec - CBOW and Skip Gram

We initially began with the Word to Vec model. Using the Gensim library, we used the pre-built models for both the Skip Gram as well as the Continuous bag of words approach.

We first iterate through the dataset and carry out the following steps -

- Take the input sentence and break it down into words

- Remove stop words and buffer words
- Pass them through the Word2Vec models to obtain the embedding for each word
- Take the average of all the embeddings of each word in the sentence to obtain the sentence embedding

We then use cosine similarity to compare the similarities of the user's book with all the books in the dataset, and return our top 5 recommendations based on it.

title	description	author
Reboot	five years ago wren connolly was shot three ti...	Amy Tintera (Goodreads Author)
The Watchers	the watchers is the first book in the watchers...	Lynnie Purcell (Goodreads Author), Holly Purce...
Midnighter and Apollo	midnighter and apollo are back with a vengeanc...	Steve Orlando (Writer), Fernando Blanco (Artist)
Times Square Red, Times Square Blue	if one street in america can claim to be the m...	Samuel R. Delany
Strength of the Pack	a marine werewolf and his commander bring lege...	Kendall McKenna (Goodreads Author)
The Devil's Son	chief of police brady simone is determined to ...	Jennifer Loren (Goodreads Author)

Figure 5. Recommendations for "The Martian" after CBOW embeddings

genres
['Young Adult', 'Dystopia', 'Science Fiction',...
['Young Adult', 'Fantasy', 'Paranormal', 'Ange...
['Comics', 'Graphic Novels', 'LGBT', 'Dc Comic...
['Nonfiction', 'Queer', 'History', 'New York',...
['M M Romance', 'Paranormal', 'Military Fictio...
['Paranormal', 'Romance', 'Dark', 'Paranormal ...

Figure 6. Genres of the recommended books above

title	description	author
Reboot	five years ago wren connolly was shot three ti...	Amy Tintera (Goodreads Author)
The Watchers	the watchers is the first book in the watchers...	Lynnie Purcell (Goodreads Author), Holly Purce...
Midnighter and Apollo	midnighter and apollo are back with a vengeanc...	Steve Orlando (Writer), Fernando Blanco (Artist)
Times Square Red, Times Square Blue	if one street in america can claim to be the m...	Samuel R. Delany
Strength of the Pack	a marine werewolf and his commander bring lege...	Kendall McKenna (Goodreads Author)

Figure 7. Recommendations for "The Martian" after Skip Gram embeddings

genres
['Young Adult', 'Dystopia', 'Science Fiction',...
['Young Adult', 'Fantasy', 'Paranormal', 'Ange...
['Comics', 'Graphic Novels', 'LGBT', 'Dc Comic...
['Nonfiction', 'Queer', 'History', 'New York',...
['M M Romance', 'Paranormal', 'Military Fictio...
['Paranormal', 'Romance', 'Dark', 'Paranormal ...

Figure 8. Genres of the recommended books above

We observe that the recommendation for both the Skip-Gram and CBOW models are identical. Additionally, we clearly see that the recommendations aren't up to the mark. 'The Martian' covers the genres of 'Science Fiction', 'Humour', 'Thriller', etc. which are far from the recommended genres seen above. This is however an expected result considering that Word 2 Vec models (Both Skip Gram as

well as CBOW) fail to capture the context of sentences and hence can not provide same embeddings to different sentences which have the same meaning. Hence, we now switch to a transformer based embedding.

3.2. Transformers 1 - Concatenated Strings

We used the SBERT (Sentence BERT) transformer, which is a pre-trained model that specializes on analyzing sentences and providing appropriate embeddings based on the context. To obtain embeddings, we simply passed each sentence/paragraph corresponding to each individual book into our model, and we used cosine similarity again as a metric to evaluate and identify the books similar to the user's input by comparing embeddings iteratively.

title	description	author
The Prime of Miss Jean Brodie / The Girls of S...	the brevity of muriel sparks novels is equaled...	Muriel Spark
The Monkey Wars	the controversy over the use of primates in re...	Deborah Blum
The Malloreon, Vol. 1: Guardians of the West /...	discover the magic of the malloreon–david eddi...	David Eddings
The Martian Chronicles	the strange and wonderful tale of man's experi...	Ray Bradbury
I'll Never Let You Go	kevin is debonair well educated loyal and will...	Leonard Anderson Jr.
Олень на заводской территории	олень на заводской территории это один из расс...	Kurt Vonnegut Jr., Курт Воннегут, Н. Пальцев (...

Figure 9. Recommendations for "The Martian" using SBERT

genres
['Fiction', 'Classics', 'Literature', 'Novels'...
['Science', 'Nonfiction', 'History', 'Biology'...
['Fantasy', 'Fiction', 'High Fantasy', 'Epic F...
['Science Fiction', 'Fiction', 'Classics', 'Sh...

Figure 10. Genres of the recommended books above

We see a drastic improvement in results with books much more similar to the martian being recommended, such as 'The Martian Chronicles'. However, we still see multiple unrelated genres of books being pushed into our recommendations such as 'Literature', 'History' and 'Classics'.
Upon some analysis we concluded that this was mainly due to the way our input was being presented to the transformer. When we merely just concatenate the Genres and Authors of a particular book at the end of the book's description, they do not mix well with the context of the rest of the paragraph as they appear as random unrelated words. When compared with embeddings of all other books, some of these random words find a match in unrelated books - Eg. The book 'The Prime of Miss Jean Broadie' which is our top recommendation has the genres of 'Fiction', 'Classics', 'Novels' in it which are also present as sub genres of our main book 'The Martian' which might have pushed it up the similarity ladder.
We realized that our main aim must be to focus on the book descriptions and capture the story underlying in each of these books while making comparisons.

3.3. Transformers 2 - Description Only

Using the same SBERT model, we decided to now obtain fresh embeddings by passing only the description of each book into the transformer as input.

title	description	author
The Martian Chronicles	the strange and wonderful tale of man's experi...	Ray Bradbury
Red Planet	jim marlow and his strangelooking martian frie...	Robert A. Heinlein
Union	there are no holes in space none to hide in no...	Jolea M. Harrison (Goodreads Author)
Planet of Adventure	stranded on the distant planet tschai young ad...	Jack Vance
Deepsix	in the year 2204 tragedy and terror forced a s...	Jack McDevitt
The Creatures That Time Forgot	from planet stories mad impossible world sunbl...	Ray Bradbury

Figure 11. Recommendations for "The Martian" using SBERT : Description only as input

genres
['Fiction', 'Classics', 'Literature', 'Novels'...
['Science', 'Nonfiction', 'History', 'Biology'...
['Fantasy', 'Fiction', 'High Fantasy', 'Epic F...
['Science Fiction', 'Fiction', 'Classics', 'Sh...

Figure 12. Genres of the recommended books above

As suspected in our earlier observations, removing the concatenated words which add randomness to our input gives us our best embeddings and thereby best recommendations thus far. This can be verified seeing the descriptions of the books and their Genres, with multiple of them being Mars related and having similar genres to our target book.

Though this model provides a great improvement in the recommendations, we wished to further improve our model by incorporating the other data of our data set. However, we needed to do it in a better way than concatenation of random words to our book description, in a way that can provide more context to the transformer about the data we are adding.

3.4. Transformers 2 - Improvised Description

Our novelty in this project, was coming up with the solution to this problem. We added data like the book author, book genres, book ratings, etc. in the form of meaningful english sentences. For example, for a book with description "Tom met harry today", with author 'Harvey Specter' covering the genres of 'action' and 'adventure', the final input corresponding to this book which we would pass to our transformer is " Tom met harry today. The author of this book is Harvey Specter. The genres of this book are action and adventure". We did this for each book of our dataset as well as the user input to obtain our final and best set of embeddings.

We yet again see many books related to 'The Martian' and many of the same genres as well. To prove our point

The Creatures That Time Forgot	from planet stories mad impossible world sunbl...	Ray Bradbury
The Martian Chronicles	the strange and wonderful tale of man's experi...	Ray Bradbury
Red Planet	jim marlow and his strangelooking martian frie...	Robert A. Heinlein
Лунна дъга	роман по черному следу это первая часть фанта...	Sergey Pavlov, Сергей Павлов, Анастасия Цонева...
Summary of The Martian: by Andy Weir | Include...	summary of the martian by andy weir includes ...	NOT A BOOK
Mars Trilogy	all three volumes of the worldwide bestselling...	Kim Stanley Robinson
2001: A Space Odyssey	on the moon an enigma is uncovered\r\nso great...	Arthur C. Clarke
Voyage	the space mission of a lifetime an epic saga o...	Stephen Baxter
Green Mars	in the nebula award winning red mars kim stanl...	Kim Stanley Robinson

Figure 13. Recommendations for "The Martian" using SBERT : Description improvized

genres
['Science Fiction', 'Fiction', 'Classics', 'Sh...
['Science Fiction', 'Fiction', 'Young Adult', ...
🗋
['Science Fiction', 'Fiction', 'Fantasy', 'Adv...
['Science Fiction', 'Fiction', 'Space Opera', ...

Figure 14. Genres of the recommended books above

that this indeed is a better set of recommendations than our other models so far.

Result Analysis. We try to make our argument about the quality of our recommendations based on a multiple factors -

1) Author Context: Our fifth recommendation, "The summary of the Martian by Andy Wier" clearly should have been one of the top recommendations of any system based on similarity of books. Due to us adding the context regarding the author of our books in the input, we believe that this helped in finally recommending this book.

2) Description Context : Let us take the description of the top recommendation of our improvised model to compare it to 'The Martian' itself.

   - The Creatures That Time Forgot : From planet stories, mad impossible world sunblasted by day coldwracked by night and life condensed by radiation into eight days. Simon eyed the ship if he only dared reach it and escape - but it was more than half an hour distant, perhaps the limit of life itself. From the author of fahrenheit 45,1 the martian chronicles and the illustrated man originally published in the fall 1946 issue of planet stories it was later reprinted under the title frost and fire.
   - The Martian : Six days ago astronaut mark watney became one of the first people to walk on mars now he's sure he'll be the first person to die thereafter a dust storm nearly kills him and forces his crew to evacuate while thinking him dead mark finds himself stranded and completely alone with no way to even signal earth that he's alive—and even if he could get word out his supplies would be gone long before a rescue could arrive chances are though he won't have time to starve to death the damaged machinery unforgiving environment or plainold "human error" are much more likely to kill him first but mark isn't ready to give up yet drawing on his ingenuity his engineering skills — and a relentless dogged refusal to quit — he steadfastly confronts one seemingly insurmountable obstacle after the next will his resourcefulness be enough to overcome the impossible odds against him'

The martian is the story of an astronaut who gets stuck on Mars, left behind by his team after a thunderstorm, and the story revolves around his survival and escape from the planet. Similarly, in 'The creatures that time forgot', we can see from the lines "Simon eyed the ship if he only dared reach it and escape" and from "more than half an hour distant, perhaps the limit of life itself" that the theme of this story as well revolves around an astronaut striving for survival and escape from a foreign planet, which very closely mirrors the main theme in the Martian. This being the top recommendation in our improvised model as compared to being fourth one he list in our description-only model, highlights our improvement in recommendations.

3) Genre Context : Let us once again compare the genres of 'The Martian' along with the top recommendations of both the description-only model as well as our improvised model.

   - The Martian : Science Fiction, Fiction, Adventure, Audiobook, Space, Adult, Science Fiction Fantasy, Thriller, Humor, Survival.
   - The Martian Chronicles (Top rec of Description-only) : Science Fiction, Fiction, Classics, Short Stories, Space, American, Literature, Fantasy, Science Fiction Fantasy, Audiobook.
   - Creatures That Time Forgot (Top rec of Improvised) : Science Fiction, Short Stories, Science Fiction Fantasy, Classics, Fiction

Here, we observe that the genres of 'The Martian Chronicles' do contain some of the genres of our target book, but however they also contain several others too which are unrelated to 'The Martian'. This is unlike 'The Creatures that Time Forgot' whose set of genres is a proper subset of 'The Martian'. This detail identified possibly due to our increased context input in the form of genres helped our improvised model pick 'The creatures that Time Forgot' over 'The Martian Chronicles'.

Another example to prove that our transformers model is far superior to Word2Vec or any other method to obtain embeddings can be seen with reference to the input book 'Harry potter'.

title	description	author
The Adventures of Jewel Cardwell	what would you do if you knew you'd inherited …	Fumi Hancock (Goodreads Author)
Thunderlight	vibrant scale swathed wings… abilities that ca…	Adrienne Woods (Goodreads Author)
The Immortal Lover	down through the ages of human history his nam…	A. White (Goodreads Author)
Consequences	every action has consequenceswaking in an unfa…	Aleatha Romig (Goodreads Author)
Emily Post's Etiquette	features twenty new chapters that cover such a…	Peggy Post, Emily Post
The Liberation of Gabriel King	gabriel king was a born chicken hes afraid of …	K.L. Going (Goodreads Author)

Figure 15. Recommendations for "Harry Potter volume 1" using CBOW

Harry Potter and the Chamber of Secrets	ever since harry potter had come home for the …	J.K. Rowling, Mary GrandPré (Illustrator)
Harry Potter and the Prisoner of Azkaban	harry potters third year at hogwarts is full o…	J.K. Rowling, Mary GrandPré (Illustrator)
Harry Potter Schoolbooks Box Set: Two Classic	there is an alternate cover edition for this i…	J.K. Rowling
The Harry Potter Collection 1-4	the exciting tales of harry potter the young w…	J.K. Rowling, Mary GrandPré (Illustrator)
James Potter and the Hall of Elders' Crossing	what's it like to be the son of the most famou…	G. Norman Lippert (Goodreads Author)
James Potter and the Curse of the Gatekeeper	a summer of change brings james sirius potter …	G. Norman Lippert (Goodreads Author)

Figure 16. Recommendations for "Harry Potter volume 1" using Transformers

Yet again, the Word2Vec models fail to recommend the easiest of related books, which are the sequels of the first book in the series. The transformers however do the task without any problem.

4. Recommending Books Based on Combined User Read Book Embeddings

In addition to our approach of recommending top books for each user read book, we also experimented with a different method of recommending books based on the combined embeddings of all the books the user has read. This approach involves computing the combined book embedding in two different ways: (1) finding the average of all the user read books and (2) concatenating the embeddings of all the user read books.

We generated embeddings for each book using the transformers based method with the improvised description approach, which involves creating meaningful English sentences for each book using various attributes such as title, author, description, genres, and ratings.

With this approach, we aim to capture the user's preferences and reading habits as a whole, instead of just considering their top read books. By combining the embeddings of all the user's read books, we obtain a representation of the user's overall reading interests, which can be used to recommend books that match their preferences more accurately.

When testing this approach with books such as "The Diary of a Wimpy Kid," "How Data Happens," and "The Martian," our system successfully recommended books with genres such as humor, young adult, science fiction, space, and science. However, since our dataset has a higher proportion of books in the science fiction category, the recommendation

title	description	author	genres
The Martian Chronicles	the strange and wonderful tale of man's experi…	Ray Bradbury	['Science Fiction', 'Fiction', 'Classics', 'Sh…
Red Planet	jim marlow and his strangelooking martian frie…	Robert A. Heinlein	['Science Fiction', 'Fiction', 'Young Adult', …
Union	there are no holes in space none to hide in no…	Jolea M. Harrison (Goodreads Author)	[]
Planet of Adventure	stranded on the distant planet tschai young ad…	Jack Vance	['Science Fiction', 'Fiction', 'Fantasy', 'Adv…
Deepsix	in the year 2204 tragedy and terror forced a s…	Jack McDevitt	['Science Fiction', 'Fiction', 'Space Opera', …
The Creatures That Time Forgot	from planet stories mad impossible world sunbl…	Ray Bradbury	['Science Fiction', 'Short Stories', 'Science …
The Creatures That Time Forgot	from planet stories mad impossible world sunbl…	Ray Bradbury	['Science Fiction', 'Short Stories', 'Science …
Starship Titanic	sur la planète blérontin on sapprête à inaugur…	Terry Jones, Marie-Catherine Caillava (Transla…	['Science Fiction', 'Fiction', 'Humor', 'Fanta…
Lost Moon: The Perilous Voyage of Apollo 13	april 1970 the glory days of the apollo space …	Jim Lovell, Jeffrey Kluger, Fred Sanders (Narr…	['Nonfiction', 'History', 'Space', 'Science', …
Doctor Who: Apollo 23	for a few moments this afternoon it rained on …	Justin Richards	['Doctor Who', 'Science Fiction', 'Fiction', '…

Figure 17. Recommendation for the the user with books read=Dairy of a wimpy kid,

results may reflect this skewness, leading to more recommendations in this genre.

Our experiments showed that this approach is effective in providing book recommendations that align with the user's overall reading preferences, considering both the topics and genres of the books they have previously read. We also found that the concatenated embedding method performed slightly better than the average embedding method.

Overall, this approach provides an alternative way of recommending books that takes into account the user's entire reading history and can be useful in cases where the user has a diverse range of reading interests.

5. Extendable recommendation pipeline

In order to have a larger social impact, we have designed our system to be extendable beyond just book recommendations. In fact, we have already demonstrated the potential for extension by building a small scale research paper recommendation system using the same pipeline as our book recommendation system.

For this recommendation system, we were able to create a dataset for research papers with columns such as 'Title', 'Abstract', 'Conclusion', 'Authors', 'Paper Citations', and 'Keywords'. for this we collected 5 research papers from each of the domains following: Blockchain, VLSI, IOT and Deep Learning. This dataset was then used to train a recommendation model similar to the book recommendation system.

Similar to our approach with books, we added data such as paper citations and keywords in the form of meaningful English sentences to provide more context about the information.Hence,we used the Transformers based approach with Improvised Description to create the embedding of each research paper. For example, for a research paper with abstract "This paper proposes a new algorithm for optimizing network routing", with author 'John Smith', with conclusion 'this paper presents a new algorithm that significantly improves network routing efficiency' and keywords 'networks' and 'optimization', the final input corresponding to this paper which we would pass to our transformer is "This paper proposes a new algorithm for optimizing network routing. The author of this paper is John Smith. The keywords of this paper are networks and optimization.In conclusion, this paper presents a new algorithm that significantly improves network routing efficiency".

Title	Abstract	Conclusion of the paper	Authors	Paper Citations	Keywords
Leakage current mechanisms and leakage reducti...	High leakage current in deep-submicrometer reg...	With the continuous scaling of CMOS devices, l...	K. Roy; S. Mukhopadhyay; H. Mahmoodi-Meimand	1606	, CMOS integrated circuits , leakage currents...
The flipped voltage follower: a useful cell fo...	In this paper, a basic cell for low-power and/...	In this paper, a cell coined as FVF has been r...	R.G. Carvajal; J. Ramirez-Angulo; A.J. Lopez-M...	470	Voltage , Circuit synthesis , Consumer electro...
A Systematic Review of Compressive Sensing: Co...	Compressive Sensing (CS) is a new sensing moda...	Introduction of CS has revolutionized many are...	Meenu Rani; S. B. Dhok; R. B. Deshmukh	270	Sensors , Transforms , Mathematical model , Sp...
A low-power low-noise CMOS amplifier for neura...	There is a need among scientists and clinician...	An 80-µ W fully integrated CMOS biosignal ampl...	R.R. Harrison; C. Charles	1283	Low-noise amplifiers , Low-frequency noise , P...
SegNet: A Deep Convolutional Encoder-Decoder A...	We present a novel and practical deep fully co...	We presented SegNet, a deep convolutional netw...	Vijay Badrinarayanan; Alex Kendall; Roberto Ci...	8360	Decoding , Neural networks , Training , Comput...

Figure 18. Recommendation for the paper "A 10-bit 50-MS/s SAR ADC With a Monotonic Capacitor Switching Procedure" using transformers

Here, we observe that our recommendation system successfully suggests more VLSI domain papers with similar topics, such as CMOS circuits, based on the input paper. This demonstrates the versatility of our model and its ability to effectively handle other problems beyond book recommendations.

By using the same pipeline as our book recommendation system, we were able to create an extendable recommendation pipeline for research papers. Our system leverages the deep learning capabilities of transformers to create vector embeddings for text-based inputs. By modifying the input data to give more context about the information, we maximize the information that we can provide to our model to obtain embeddings.

In the future, we believe that our system can be extended even further beyond just recommending books and research papers. By using similar techniques, our pipeline can be adapted to recommend other types of content such as movies, music, and even products. Our approach can be applied to any problem where the item to be recommended has a deeper underlying theme that can be personalized to a user. For example, it could be extended to recommend movies by adding relevant data columns such as movie synopsis, director information, and genre, and using them to create embeddings. We hope that our work serves as a guide to building systems for analogous problems using the same pipeline to help find content similar to what they enjoy, regardless of the medium.

6. Future work

As of now, the model works only on the basis of data that is currently available, i.e. it is a Content based filtering. Every time a user needs a recommendation, he/she needs to provide the books that he liked reading in the past. We are not building a separate user profile for each new user. There is no way to remember the past inputs by a user. This does not allow the recommendations to be personalised to a specific user without getting the inputs from him. We can add the previous books read by the user to a database to reduce the requirement of taking input from the user every time.

Along with implementing Collaborative filtering, we can also add a feature which recommends a book to the user based on the sequence in he/she has read the books till now by using RNN/LSTM architecture. This particularly helps in the case in which the user wants to get recommendation based on the order in which they have read the books.(By giving more recommendation weightage to recently read book than past read books)

We could also perform clustering on the embeddings that we generate for each book. We can use this to find books which are similar to a specific book and recommend books from that cluster.

There are several other small details which can be worked upon to improve the recommendation model:

- Use of a better transformer which is pre-trained specifically on for scientific text. This can significantly boost the level of recommendations made by the model in the case of research paper.
- Currently we are making use of the already available transformer which is trained on a large corpus of text data. We can fine tune it by taking the pre-trained model and training it again on our database to get better results.
- Use of a wider range of data, by including more relevant columns to build our recommendation system. Also, the way we pass data to the transformer by adding additional string can be tested in different orders.

7. Conclusion

In conclusion, our experimentation showed that Word2Vec methods, including Bag of Words and Skipgram, produced similar results in terms of embeddings. However, they were not closely related to the input book, as Word2Vec is designed to predict outputs based on similar words, which can be problematic when the same words have different meanings in different contexts.

To overcome this shortcoming, we experimented with Transformers, which take into account the context of the background while producing the output. We observed that the Transformer gave better book recommendations, especially when we provided additional information about the specific columns of the text.

Furthermore, in our experiments with the Recommending Books Based on Combined User Read Book Embeddings approach, we found that the concatenated embedding method outperformed the average embedding method. These findings suggest that the use of Transformers and the concatenated embedding method can significantly improve the accuracy of book recommendations, and may be particularly useful for users with diverse reading interests.