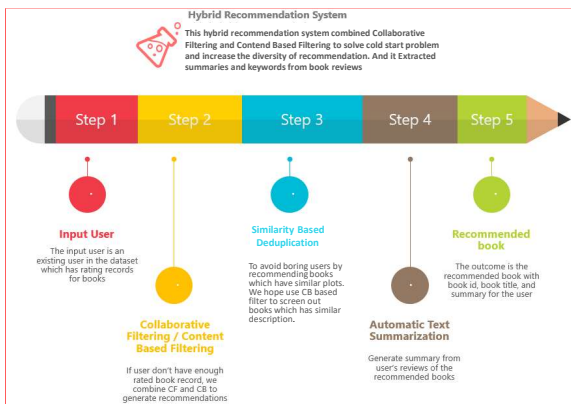


Introduction

- Our team builds a book recommendation system based on **Collaborative Filtering**, **Content Based Filtering** and **Automatic Text Summarization**.
- Goodreads Dataset** from UCSD Book Graph is used.
- Implemented advanced hybrid recommendation system by using Content Based Filtering to solve **Cold Start** problem and increase the **diversity** of recommendation.
- Extracted summaries and keywords from book reviews using **TF-IDF** and **BM25 based Tag method**, **TextRank** Algorithm and **Latent factor models**.
- The application is trained on **Google Colab**

Recommendation System Workflow



Collaborative Filtering and Matrix Factorization

- Provide personalized recommendation based on user's historical rating on books.
- User-based CF: Find other users whose ratings are "similar" to target user's ratings. Using **cosine** and **pearson correlation** to find similarities.
- Matrix Factorization: Using **Stochastic Gradient Descent** method to obtain latent factor in order to find a local minimum of the difference.

Method	Cosine Similarity	Pearson Correlation	Pearson Correlation with weight	Matrix Factorization
MAE	0.6646	0.6928	0.6562	0.6203
RMSE	0.8687	0.9350	0.8625	0.8136

Content Based Filtering

- TF-IDF**, **Word2Vec**, **Sentence2Vec**: three methods are used to produce matrix representation for each books based on their description. Cosine similarities are then calculated.
- Performance of three method are compared, and one of them is chosen to be implemented into the system.

Target Book: 1293847, Season Songs

Description: a collection of twenty-eight poems grouped to represent the four seasons

Word2Vec Recommendation: 268005, Hinge & Sign: Poems

Description: words and forms challenge each other in a collection of new poems and selections from the author's five previous books

TF-IDF Recommendation: 7075534, The Collected Poems of Sara Teasdale

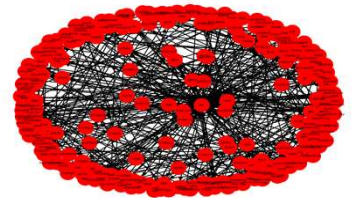
Description: collection poems

Sentence2Vec Recommendation: 15950875, Haiku Reflections: The Four Seasons

Description: a collection of 120 haiku poems, exploring and celebrating the wonder and unique beauty of each of the four seasons.

Automatic Text Summarization

- Motivation:
 - Book review provide a rather unbiased insight into books.
 - When we recommend books to users, users are curious about the reason behind recommendation.
 - Text summarization on book reviews provide such a reason why this specific book is recommended.
 - Manual summarization is laborious. Automatic summarization has fewer biases, is faster, more scalable and more cost-efficient.



TF-IDF/BM25 based Tag Method	Latent factor models	TextRank (with BM25)	TextRank (lemmatization, POS tagging, NER)
'The opposite of a great truth is another truth.' - Niels Bohr I was thinking about Dante the other day and wondering how one could approach him from the angle of a GoodReads review. I do wonder what he would have thought if he had been able to learn that many leading religious figures, even in the early 21st century, reject a large part of science as being somehow unreligious	"Now of course, I am aware that Dante was deeply immersed in the Christian world-view, and Dawkins is famous for being the world's most outspoken atheist. Dante was a Christian to the core of his being, but he was furious with the way the Church was being run; he put several of its leaders, notably Pope Boniface VIII, in Hell.	The opposite of a great truth is another truth." - Niels Bohr I was thinking about Dante the other day and wondering how one could approach him from the angle of a GoodReads review. For Dante, there didn't seem to be any opposition between religious faith and science - they were part of the same thing	The opposite of a great truth is another truth." But it's not quite as crazy as it first may seem. Obviously we would n't have the old geocentric model of the Universe- it would be bang up to date. I think there is now far more material for an ambitious poet to work with than there was in the 14th century. Finally, we reach the Heaven of the Multiverse, and find that we are just one of many different universes.
'dante', 'dawkins', 'heaven', 'universe', 'would', 'religious', 'one', 'wondering', 'could', 'century', 'science', 'able', 'fact', 'astronomy', '"I", 'find', 'paradise', 'bang',	'dante', 'could', 'wondering', 'would', 'heaven', 'one', 'let', 'light', 'might', 'help', 'get', 'find', 'truth', 'paradise', 'thinking', 'like', 'moment'	religious day days world truth wondering wonder wonderful background geocentric fluctuations moment different time mainstream theological write writing today produced human life	ambitious poet, most outspoken atheist, actual fact, more material, heaven, universe, life, same thing, dante, pope, dawkins, obscure theological points, religious man

Recommendation Results

Figure 1: Input User

User1	Book ID	Book	Rating Score
1	764332	Jason and the Golden Fleece	4
2	1519	The Oresteia (Orestes, #1-3)	4
3	1715	Metamorphoses	4
4	12914	The Aeneid	4
5	1371	The Iliad	5
6	1381	The Odyssey	5
7	2696	The Canterbury Tales	4

User2	Book ID	Book	Rating Score
1	34023590	Sad Girls	2
2	25330489	Memories	4
3	18003300	Love & Misadventure	3

Figure 2: Result 1, Naive Recommendation System

Rank	Book ID	Book Title
1	6656	The Divine Comedy
2	15645	Inferno (The Divine Comedy #1)
3	1420	Hamlet
4	23919	The Complete Stories and Poems
5	333706	The Odyssey
6	15997	Paradise Lost
7	88410	Four Quartets
8	42051	The Complete Sonnets and Poems
9	42038	Shakespeare's Sonnets
10	782580	The Complete Poetry and Prose
11	269322	The Raven and Other Poems
12	118389	The Love Song of J. Alfred Prufrock and Other Poems
13	75584	The Collected Poems
14	23913	The Marriage of Heaven and Hell
15	23912	The Complete Poems

Result 2: Hybrid Recommendation System

Figure 4: Recommendation List After Combine CF and CB

Rank	Book ID	Book Title
1	383865	Collected Poems of Emily Dickinson
2	1291581	The Journals of Susanna Moodie
3	968056	The Lady of Shalott
4	644996	The Legend of Nulan: A Folding Book of the Ancient Poem That Inspired the Disney
5	1370171	When We Were Very Young
6	20801853	When We Were Very Young
7	3406085	When We Were Very Young
8	44450	The Portable Dorothy Parker
9	176732	Love Poems from the Japanese
10	160889	Illustrated Treasury of Read-Aloud Poems for Young People
11	35620866	Selected Poems
12	133987	Poetry for Young People: Langston Hughes
13	589454	Climb Into My Lap First Poems to Read Together: First Poems to Read Together
14	7803492	The Canterbury Tales (Barnes & Noble Classics Series)
15	13542892	Madame X

Rank	Book ID	Book Title
1	383865	Collected Poems of Emily Dickinson
2	22151696	Lullabies
3	1291581	The Journals of Susanna Moodie
4	29431881	The Universe of Us
5	968036	The Lady of Shalott
6	23513349	Milk and Honey
7	644996	The Legend of Nulan: A Folding Book of the Ancient Poem That Inspired the Disney
8	25384844	Black Butterfly
9	1370171	When We Were Very Young
10	23434371	Beautiful Chaos
11	20801853	When We Were Very Young
12	13121245	8
13	3406085	When We Were Very Young
14	13105527	I Wrote This For You
15	44450	The Portable Dorothy Parker

Figure 5: Final Result after CB Based Deduplication

Rank	Book ID	Book Title
1	383865	Collected Poems of Emily Dickinson
2	22151696	Lullabies
3	1291581	The Journals of Susanna Moodie
4	29431881	The Universe of Us
5	968036	The Lady of Shalott
6	23513349	Milk and Honey
7	644996	The Legend of Nulan: A Folding Book of the Ancient Poem That Inspired the Disney
8	25384844	Black Butterfly
9	1370171	When We Were Very Young
10	23434371	Beautiful Chaos
11	13121245	8
12	13105527	I Wrote This For You
13	44450	The Portable Dorothy Parker
14	35606568	The Sun and Her Flowers
15	176732	Love Poems from the Japanese