

The Book Forum: Recommendation Application System using Collaborative Filtering and Autoencoders

مقدمه

این مقاله به بررسی سیستم‌های پیشنهاد دهنده کتاب می‌پردازد که از فیلترینگ مشارکتی و خودرمزگذارها (Autoencoders) استفاده می‌کنند. سیستم‌های پیشنهاد دهنده می‌توانند تجربه خواندن کاربران را با ارائه پیشنهادات شخصی‌سازی شده بهبود بخشند. در این مقاله، نویسندگان به طور خاص بر ترکیب روش‌های فیلترینگ محتوا و فیلترینگ مشارکتی تمرکز کرده‌اند تا دقت و تنوع پیشنهادات را افزایش دهند. سیستم‌های پیشنهاد دهنده کتاب به کاربران کمک می‌کنند تا کتاب‌هایی را بیابند که با علایق و سابقه خواندن آن‌ها همخوانی داشته باشد. این مقاله روش‌های مختلفی از جمله فیلترینگ مشارکتی، فیلترینگ مبتنی بر محتوا، و روش‌های ترکیبی را بررسی می‌کند. استفاده از شبکه‌های عصبی عمیق، به ویژه خودرمزگذارها، نیز در این مقاله بررسی شده است. هدف از این ترکیب روش‌ها افزایش دقت و مقیاس‌پذیری سیستم‌های پیشنهاد دهنده است که در کتابفروشی‌ها و کتابخانه‌های آنلاین مورد استفاده قرار می‌گیرند.

همانطور که گفتیم در این مقاله، سیستم توصیه‌گر کتابی توسعه یافته که از ترکیبی از الگوریتم‌های فیلترسازی مشارکتی (Collaborative Filtering) و خودرمزگذارها (Autoencoders) برای بهبود دقت و کارایی توصیه‌ها استفاده می‌کند. هدف اصلی این سیستم، ارائه توصیه‌های شخصی‌سازی شده به کاربران براساس ترجیحات و سابقه خواندن آن‌ها است.

ساختار و عملکرد سیستم توصیه‌گر

۱. جمع‌آوری داده‌ها (Data Collection)

این سیستم از دو منبع داده اصلی استفاده می‌کند:

داده‌های کاربران: شامل رتبه‌بندی‌ها، نظرات و تعاملات کاربران با کتاب‌ها.

داده‌های کتاب‌ها: شامل اطلاعات متنی مانند عنوان، نویسنده، ژانر، و سایر ویژگی‌های مربوط به کتاب‌ها.

۲. پیش‌پردازش داده‌ها (Data Preprocessing)

تمیز کردن داده‌ها: شناسایی و حذف مقادیر گمشده، داده‌های تکراری و مقادیر پرت.

یکپارچه‌سازی داده‌ها: ترکیب داده‌های کاربران و کتاب‌ها برای ایجاد یک منبع داده جامع.

تبدیل داده‌ها: نرمال‌سازی رتبه‌بندی‌ها و رمزگذاری ویژگی‌های متنی.

مهندسی ویژگی‌ها: انتخاب و استخراج ویژگی‌های مرتبط که می‌توانند به دقت پیش‌بینی‌ها کمک کنند.

تقسیم داده‌ها: تقسیم داده‌ها به مجموعه‌های آموزشی و آزمایشی برای ارزیابی عملکرد مدل.

۳. استفاده از الگوریتم‌ها برای توصیه‌ها (Recommendation Algorithms)

که همان الگوریتم های Collaborative Filtering و Content-based Filtering و Hybrid Approach است که آنها را می‌شناسیم و نیازی به معرفی نیست. ولی الگوریتم خودرمزگذارها (Autoencoders) هست که بصورت :

خودرمزگذارها نوعی شبکه عصبی مصنوعی هستند که برای فشرده سازی و بازسازی داده‌ها طراحی شده‌اند و در سیستم‌های توصیه گر برای یادگیری الگوهای پیچیده در تعاملات بین کاربران و آیتم‌ها، مانند کتاب‌ها، استفاده می‌شوند. ساختار این شبکه‌ها شامل سه بخش اصلی است: لایه ورودی، لایه‌های پنهان، و لایه خروجی. لایه ورودی داده‌هایی مانند رتبه‌بندی‌ها و بازخوردهای کاربران را دریافت می‌کند. لایه‌های پنهان وظیفه فشرده سازی این داده‌ها و یادگیری ویژگی‌های کلیدی را بر عهده دارند. این لایه‌ها تعداد نودهای کمتری نسبت به لایه ورودی دارند تا بتوانند داده‌ها را به یک نمایش فشرده تر تبدیل کنند. در نهایت، لایه خروجی تلاش می‌کند داده‌های ورودی را بازسازی کند تا تفاوت بین ورودی و خروجی به حداقل برسد.

در مرحله فشرده سازی، داده‌های ورودی از طریق لایه‌های پنهان عبور کرده و به یک نمایش فشرده تر تبدیل می‌شوند که اطلاعات کلیدی را حفظ کرده و نویزها و اطلاعات غیرضروری را حذف می‌کند. سپس در مرحله بازسازی، این نمایش فشرده از طریق لایه‌های پنهان به لایه خروجی منتقل می‌شود تا داده‌های ورودی اصلی بازسازی شوند. هدف این است که خروجی تا حد ممکن به ورودی اولیه نزدیک باشد. در طول فرآیند آموزش، خودرمزگذار با استفاده از الگوریتم‌های بهینه سازی مانند پس انتشار خطا آموزش می‌بیند. در این مرحله، وزن‌های شبکه عصبی تنظیم می‌شوند تا خطای بازسازی به حداقل برسد. معیارهایی مانند خطای ریشه میانگین مربعات (RMSE) یا خطای میانگین مربعات (MSE) برای اندازه گیری خطا استفاده می‌شوند.

در سیستم توصیه گر کتاب، خودرمزگذارها با استفاده از داده‌های تعاملی بین کاربران و کتاب‌ها آموزش می‌بینند. به عنوان مثال، ماتریس کاربر-آیتم که نشان‌دهنده تعاملات بین کاربران و کتاب‌ها است، به عنوان ورودی استفاده می‌شود. این ماتریس معمولاً بسیار پراکنده است زیرا هر کاربر تنها تعداد محدودی از کتاب‌ها را رتبه‌بندی کرده است. خودرمزگذارها ویژگی‌های نهفته و مهم در این داده‌ها را یاد می‌گیرند. این ویژگی‌ها می‌توانند شامل ترجیحات پنهان کاربران یا ویژگی‌های مشترک بین کتاب‌های مختلف باشند که به طور مستقیم قابل مشاهده نیستند. پس از آموزش، خودرمزگذار می‌تواند از این نمایش فشرده برای پیش‌بینی رتبه‌بندی‌های گمشده یا تعاملات جدید بین کاربران و کتاب‌ها استفاده کند. به این ترتیب، برای هر کاربر، سیستم می‌تواند کتاب‌هایی را که کاربر هنوز رتبه‌بندی نکرده است، براساس شباهت به کتاب‌های مورد علاقه کاربر و الگوهای یاد گرفته شده توصیه کند.

مزایای استفاده از خودرمزگذارها در سیستم‌های توصیه گر شامل مدیریت بهتر داده‌های پراکنده، کاهش ابعاد داده‌ها و یادگیری ویژگی‌های پیچیده است. خودرمزگذارها می‌توانند با داده‌های پراکنده به خوبی کار کنند و اطلاعات مفید را از داده‌های کم استخراج کنند. این روش می‌تواند ابعاد داده‌ها را کاهش داده و فضای ذخیره سازی و محاسباتی کمتری نیاز داشته باشد، در حالی که همچنان اطلاعات مهم حفظ می‌شود. همچنین، خودرمزگذارها قادرند الگوهای پیچیده و غیرخطی را در داده‌ها یاد بگیرند که ممکن است توسط روش‌های ساده تر قابل تشخیص نباشند.

۴. ارزیابی و بهبود مدل (Model Evaluation and Improvement)

ارزیابی مدل با استفاده از معیارهایی مانند خطای ریشه میانگین مربعات (RMSE)، خطای میانگین مطلق (MAE) و خطای میانگین مربعات (MSE).

بهبود مدل براساس نتایج ارزیابی و انجام تنظیمات لازم برای افزایش دقت و کارایی توصیه‌ها.

ویژگی‌های کلیدی سیستم

شخصی‌سازی توصیه‌ها: این سیستم با تحلیل ترجیحات و سابقه خواندن هر کاربر، توصیه‌های شخصی‌سازی شده‌ای ارائه می‌دهد که به نیازها و سلیقه‌های فردی کاربر پاسخ می‌دهد.

رویکرد هیبریدی: با ترکیب فیلترسازی مشارکتی و فیلترسازی مبتنی بر محتوا، این سیستم می‌تواند دقت توصیه‌ها را افزایش دهد و نقاط ضعف هر یک از روش‌ها را پوشش دهد.

استفاده از خودرمزگذارها: با استفاده از خودرمزگذارها، سیستم می‌تواند الگوهای پیچیده در داده‌های تعاملی را شناسایی کند و توصیه‌های دقیق‌تری ارائه دهد.

پلتفرم انجمن (Forum Platform): این سیستم همچنین شامل یک پلتفرم انجمن است که به کاربران اجازه می‌دهد درباره کتاب‌ها بحث کنند، نظرات خود را به اشتراک بگذارند و از توصیه‌های دیگر کاربران بهره‌مند شوند.

توصیه‌های مبتنی بر زمینه (Context-aware Recommendations): این سیستم می‌تواند با در نظر گرفتن زمینه‌هایی مانند مکان کاربر یا شرایط آب و هوا، توصیه‌های متناسب با وضعیت فعلی کاربر ارائه دهد.

بخش "Literature Survey" در این مقاله به تحلیل و مقایسه تحقیقات و روش‌های موجود در زمینه سیستم‌های توصیه‌گر کتاب می‌پردازد. هدف اصلی این بخش، بررسی روش‌های مختلف برای توصیه کتاب و شناسایی نقاط قوت و ضعف هر یک از این روش‌ها است.

مقایسه روش‌ها و الگوریتم‌ها

انتخاب بهترین الگوریتم برای سیستم توصیه‌گر به چندین عامل بستگی دارد. در این مقاله، روش‌های مختلفی برای ارزیابی و انتخاب الگوریتم‌های مناسب استفاده شده است:

ارزیابی دقت (Accuracy Evaluation)

استفاده از معیارهایی مانند خطای ریشه میانگین مربعات (RMSE)، خطای میانگین مطلق (MAE) و خطای میانگین مربعات (MSE) برای ارزیابی دقت پیش‌بینی‌های مدل. الگوریتمی که کمترین خطا را داشته باشد، به عنوان بهترین الگوریتم انتخاب می‌شود. به عنوان مثال، اگر الگوریتمی با استفاده از معیار RMSE کمترین مقدار خطا را داشته باشد، به عنوان الگوریتم برتر در نظر گرفته می‌شود.

ارزیابی کارایی (Performance Evaluation)

بررسی زمان آموزش و پیش‌بینی مدل‌ها. الگوریتم‌هایی که زمان آموزش و پیش‌بینی کمتری دارند، برای سیستم‌های توصیه‌گر بزرگ مقیاس مناسب‌تر هستند. این معیار به ویژه در سیستم‌های بلادرنگ (real-time) مهم است، زیرا کاربران انتظار دارند توصیه‌ها به سرعت ارائه شوند.

مقایسه با روش‌های موجود: (Comparison with Existing Methods)

مقایسه عملکرد مدل پیشنهادی با مدل‌های موجود در ادبیات. این شامل بررسی نتایج مطالعات قبلی و مقایسه آن‌ها با نتایج مدل پیشنهادی است. به عنوان مثال، اگر مدل پیشنهادی دقت بالاتری نسبت به مدل‌های موجود داشته باشد، می‌توان آن را به عنوان بهبود قابل توجهی در نظر گرفت.

حال با بررسی چندین مطالعه مختلف که به طراحی و پیاده‌سازی سیستم‌های توصیه‌گر کتاب پرداخته‌اند، می‌پردازیم. در جدول I و جدول II، مقایسه‌ای از نتایج و نقاط ضعف الگوریتم‌های مختلف ارائه شده است:

جدول I: تحلیل تطبیقی سیستم‌های توصیه‌گر کتاب

Title	Algorithm	Result	Drawbacks
The design of disciplinary book recommendation system based on android: a view of extra-curricular activities	DAO(Data Access Object) E-R mapping	Searched books will be presented on main interface of application from the database.	Recommends based on basic user information.
Book Recommendation Using Machine Learning Methods Based on Library Loan Records and Bibliographic Information	Machine learning modules based on Support vector machine (SVM), Random Forest, and Adaboost	Data displayed with detail records of students and books that have been previously borrowed from the library	Recommendation is performed only on loan based records and bibliographic information
Research on Personalized Book Recommendation Model for New Readers	Cosine similarity, Euclidean similarity, Jaccard similarity	At specific conditions significant algorithms are used and managed to get desired outcome.	When number of neighbour trend to increase, simultaneously the recommend effect achieves the best effect.
Embedding Model Design for Producing Book Recommendation	Embedding model and, principal component analysis	recommendation is produced by training embedding model to learn the pattern of high-rated books from every user and calculate the preferred books as close as possible	Total accuracy of embedding model is 59%

جدول II: تحلیل تطبیقی فیلترسازی مبتنی بر محتوا و مشارکتی

Title	Algorithm	Result	Drawbacks
Book Recommendation for eLearning Using Collaborative Filtering and Sequential Pattern Mining	Collaborative Filtering & Sequential Data	Provides more accurate recommendations with less bias	Has cold-start and sparsity problem
A Deep Learning Based Collaborative Neural Network Framework for Recommendation System using matrix function	Collaborative filtering and matrix factorization	Recommendation of books after searching for keyword.	Correlation between several users is much larger than the number of user who express similar interest of items. Cold start problem.
Cloud-Based Collaborative Filtering Algorithm For Library Book Recommendation System	collaborative filtering algorithm where books are based on categories and Apriori Algorithm	An automated and dynamic library recommendation system will help the user to choose the best version of the book of his/ her interest within a few seconds depending on the ratings given to that book.	The system needs an active internet connection all the time while accessing.
Book Recommendation system based on Collaborative Filtering and Association Rule Mining for College Students	User-based collaborative filtering and association rule mining.	Book recommendation system which recommends books to users according to their price range and preferred publishers.	The recommendation system is basically used only for students which will recommend textbooks.

نتایج بدست آمده از جداول بالا:

۱. طراحی سیستم توصیه گر کتاب بر پایه Android:

الگوریتم E-R mapping: Data Access Object (DAO)

نتیجه: کتاب‌های جستجو شده در رابط کاربری اصلی اپلیکیشن نمایش داده می‌شوند.
نقاط ضعف: توصیه‌ها بر اساس اطلاعات پایه کاربر انجام می‌شوند.

۲. استفاده از روش‌های یادگیری ماشین برای توصیه کتاب‌ها بر اساس سوابق امانت کتابخانه:

الگوریتم Adaboost، Random Forest، Support Vector Machine (SVM)

نتیجه: نمایش داده‌های کتاب‌ها و سوابق امانت کاربران.
نقاط ضعف: توصیه‌ها تنها بر اساس سوابق امانت و اطلاعات کتاب‌شناسی انجام می‌شوند.

۳. مدل توصیه گر شخصی سازی شده برای خوانندگان جدید:

الگوریتم Jaccard Similarity, Euclidean Similarity, Cosine Similarity

نتیجه: استفاده از الگوریتم های مختلف در شرایط خاص و دستیابی به نتایج مطلوب.
نقاط ضعف: با افزایش تعداد همسایه ها، اثر توصیه ها بهبود می یابد.

۴. طراحی مدل جاسازی برای تولید توصیه های کتاب:

الگوریتم: مدل جاسازی و تحلیل مولفه های اصلی

نتیجه: توصیه ها با استفاده از مدل جاسازی شده و یادگیری الگوهای کتاب های با امتیاز بالا.
نقاط ضعف: دقت کلی مدل جاسازی ۵۹٪ است.

و در ادامه اگر بخواهیم یک مقایسه از الگوریتم ها داشته باشیم بر اساس تحلیل ها و مقایسه های انجام شده در این مقاله، نتایج به دست آمده نشان می دهد که هر یک از الگوریتم ها دارای نقاط قوت و ضعف خاص خود هستند.

فیلترسازی مشارکتی مبتنی بر کاربر و آیتم برای کاربردهایی که داده های کاربران زیاد است، عملکرد خوبی دارند، اما مشکل شروع سرد همچنان وجود دارد.

فیلترسازی مبتنی بر محتوا در مواردی که ویژگی های آیتم ها به خوبی تعریف شده اند، می تواند توصیه های دقیقی ارائه دهد، اما ممکن است کاربران را در حلقه ای از توصیه های مشابه قرار دهد و تنوع را کاهش دهد.

رویکرد هیبریدی با ترکیب دو روش فوق، دقت توصیه ها را افزایش می دهد و مشکلاتی مانند شروع سرد را کاهش می دهد. این روش به عنوان یکی از بهترین رویکردها شناسایی شده است.

خودرمزگذارها با یادگیری الگوهای پیچیده در داده های تعاملی، به ارائه توصیه های دقیق تر کمک می کنند و می توانند به خوبی در سیستم های بزرگ مقیاس مورد استفاده قرار گیرند.

و در نتیجه رویکرد هیبریدی با استفاده از خودرمزگذارها و فیلترسازی مشارکتی به عنوان بهترین روش برای سیستم توصیه گر کتاب شناسایی شده است. این روش با ترکیب قدرت تحلیل داده های تعاملی و ویژگی های آیتم ها، می تواند توصیه های دقیقی ارائه دهد و تجربه کاربری بهتری را فراهم کند. این سیستم می تواند تجربه خواندن کاربران را بهبود بخشد و آن ها را به کشف کتاب ها و نویسندگان جدید ترغیب کند.

آماده سازی داده

در این سیستم، دو فریم داده مجزا وجود دارد: یکی شامل امتیازدهی و بازیابی کاربران و دیگری شامل جزئیات کتاب ها. این داده ها برای ساخت مدل های پیشنهاد دهنده، پردازش و تمیز می شوند تا شامل ویژگی های مرتبط و بدون نقص باشند. پس از آماده سازی، مدل های خودرمزگذار با داده های امتیازدهی آموزش داده می شوند.

نتیجه گیری

سیستم پیشنهادی این مقاله شامل چندین ویژگی است که به کاربران کمک می کند کتاب های مورد علاقه شان را بیابند. این سیستم از فیلترینگ مشارکتی و خودرمزگذارها برای پیش بینی کتاب های مورد علاقه کاربر بر اساس تعاملات گذشته او استفاده می کند. همچنین، یک رویکرد ترکیبی برای پیشنهاد ژانرهای مختلف بر اساس تاریخچه خواندن و امتیازات کاربر به کار گرفته شده است. این سیستم شامل یک فروم برای بحث و تبادل نظر کاربران درباره کتاب ها نیز می باشد. نتایج این مقاله نشان می دهد که استفاده از خودرمزگذارها و فیلترینگ مشارکتی می تواند به طور قابل توجهی دقت پیشنهادات را افزایش دهد. ارزش های MAE ، MSE و $RMSE$ مدل خودرمزگذار به ترتیب 0.0003807988 ، 0.000380 و 0.019514 می باشد که نشان دهنده عملکرد بالای مدل است.

نتایج این مطالعه نشان داد که استفاده از خودرمزگذارها و فیلترسازی مشارکتی می تواند به دقت بالایی در توصیه کتاب ها منجر شود. به طور خاص، مدل خودرمزگذار با استفاده از فیلترسازی مشارکتی موفق به دستیابی به مقادیر پایین خطای ریشه میانگین مربعات ($RMSE$) شده است، که نشان دهنده دقت بالای پیش بینی های مدل است. این مدل توانسته است مشکلاتی مانند مشکل شروع سرد و پراکندگی داده ها را کاهش دهد.

سیستم پیشنهادی با استفاده از خودرمزگذارها و فیلترسازی مشارکتی توانسته به دقت بالایی در توصیه کتاب ها دست یابد. این سیستم با ارائه ویژگی های کاربرپسند مانند پلتفرم انجمن برای بحث و تبادل نظر کاربران درباره کتاب ها، تعامل کاربران را افزایش داده است. آینده این سیستم شامل تحلیل رفتار کاربران و بررسی احساسی ترجیحات آن ها برای ارائه توصیه های شخصی سازی شده تر است. این مطالعه نشان می دهد که استفاده از رویکردهای ترکیبی و پیشرفته می تواند به بهبود چشمگیر سیستم های توصیه کتاب منجر شود. این سیستم می تواند تجربه خواندن کاربران را بهبود بخشد و آن ها را به کشف کتاب ها و نویسندگان جدید ترغیب کند.